A BASIS FOR THE SELECTION OF A RESPONSE SURFACE DESIGN*

G. E. P. Box
Princeton University
AND
NORMAN R. DRAPER
University of North Carolina and Imperial Chemical Industries

The general problem is considered of choosing a design such that

(a) the polynomial $f(\xi) = f(\xi_1, \xi_2, \cdots, \xi_k)$ in the k continuous variables $\xi' = (\xi_1, \xi_2, \cdots, \xi_k)$ fitted by the method of least squares most closely represents the true function $g(\xi_1, \xi_2, \cdots, \xi_k)$ over some "region of interest" R in the ξ space, no restrictions being introduced that the experimental points should necessarily lie inside R; and

(b) subject to satisfaction of (a), there is a high chance that inadequacy of $f(\xi)$ to represent $g(\xi)$ will be detected.

When the observations are subject to error, discrepancies between the fitted polynomial and the true function occur:

(i) due to sampling error (called here "variance error"), and

(ii) due to the inadequacy of the polynomial $f(\xi)$ exactly to represent $g(\xi)$ (called here "bias error").

To meet requirement (a) the design is selected so as to minimize J, the expected mean square error averaged over the region R. J contains two components, one associated entirely with variance error and the other associated entirely with bias error.

There is a class of designs which satisfy requirement (a). To meet requirement (b) we select from this class a subclass for which the "non-centrality term" in the expectation of the residual sum of squares in the analysis of variance is large. This leads to a sensitive test of goodness of fit.

In this paper the theory is applied to the particular case where $f(\xi)$ is a polynomial of first degree and $g(\xi)$ a polynomial of second degree; that is, the experimenter is hopefully fitting a first degree equation over the region R in the circumstances where the true function is really quadratic. The somewhat unexpected conclusion is reached that, at least in the cases considered, the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize bias alone*.

Particular examples of the class of optimal designs derived are fractional by replicated two-level factorial designs (in which no two-factor interaction is confounded with the main effect) with added center points.

It is proved (Appendix 1) that if a polynomial of *any* degree d_1 is fitted by the method of least squares over *any* region of interest R in the k variables, when the true function is a polynomial of any degree $d_2 > d_1$, then the bias averaged over R is minimized for all values of the coefficients of neglected terms, by making the moments of order $d_1 + d_2$ and less of the design points equal to the corresponding moments of a uniform distribution over R.

* Paper read at the 31st Session of the International Statistical Institute in Brussels 6th Sept. 1958. Material originally presented at the I.M.S. Summer Statistical Institute, July 1957, and prepared as *Technical Report* No. 23, August 1958 in connection with research sponsored by the Office of Ordnance Research, U. S. Army; Statistical Techniques Research Group, Princeton University, Contract No. DA 36-034-ORD 2297.

1. INTRODUCTION

SUPPOSE a functional relationship

$$\eta = g(\xi_1, \xi_2, \cdots, \xi_k; \theta_1, \theta_2, \cdots, \theta_p) = g(\xi, \theta) \tag{1}$$

exists between a response η and k continuous variables $\xi_1, \xi_2, \cdots, \xi_k$. Suppose further that in order to elucidate certain aspects of this relationship, measurements of η are to be made for each of N combinations of levels of the variables

$$\xi_u' = (\xi_{1u}, \xi_{2u}, \cdots, \xi_{ku}), \quad (u = 1, 2, \cdots, N) \tag{2}$$

1.1 *The Problem.* The problem of experimental design here considered is the choice of the *design matrix* $\boldsymbol{D}$ of N rows and k columns. This matrix specifies the levels of all the factors for all the measurements to be made and its u-th row is given by the elements of ξ_u' in (2).

The considerations that influence the choice of $\mathbf{D}$ are different in different circumstances. Two particular cases that arise are:

(i) when, the form of the true functional relationship $\eta = g(\xi, \boldsymbol{\theta})$ (which is of course not necessarily linear in the parameters $\boldsymbol{\theta}$ or in the variables ξ) being assumed known, the object is to estimate the *parameters* $\theta_1, \cdots, \theta_p$; and

(ii) when, the form of the true functional relationship being unknown, the object is to approximate within a given region R of the k-dimensional space of the variables, the function $g(\xi, \boldsymbol{\theta})$ by some graduating function $f(\xi_1, \xi_2, \cdots, \xi_k; \beta_1, \beta_2, \cdots, \beta_l)$. The function $f(\xi, \beta)$ would often be a polynomial in which case β would be the $l \times 1$ vector of polynomial coefficients.

The two different objectives lead to different types of design. Designs appropriate for (i) may be called *designs for estimating parameters;* designs appropriate for (ii) may be called *designs for exploring a response surface* (the surface involved being that defined by the function (1) in the $(k+1)$-dimensional space of the response η, and the variables $\xi_1, \xi_2, \cdots, \xi_k$. The first problem has been discussed [11, 9, and 7]. This paper, however, is specifically concerned with the second problem and not with the first and it is supposed that the functional form $g(\xi, \theta)$ is unknown and is to be graduated locally by a polynomial $f(\xi, \beta)$ of degree d_1 in ξ. In accordance with an already established notation [8, 6] a design suitable for fitting a polynomial of degree d_1 is called a design of *order* d_1.

Now, in practice, the nature of the variables whose levels are represented by ξ_{iu} will change from one application of a design to another. In one case, for example, ξ_i may refer to a temperature reading and in another to the dosage of a drug. Therefore, it is useful to define the general design in terms of "standardized" variables x_i which in any particular application are related linearly to the ξ_i by a transformation

$$x_{iu} = \frac{\xi_{iu} - \xi_i(0)}{S_i} \tag{3}$$

In what follows we shall choose $\xi_i(0)$ to be the mean of the ξ_{iu} so that

$$\sum_{u=1}^{N} x_{iu} = 0. \tag{4}$$

The scale factor S_i is a convenient constant relating the standardized levels x_i in the general design to the levels of the actual variables ξ_i to be used in a particular application.

1.2. *The Requirements.* In a recent paper [6], it was suggested that, when the function was to be graduated by a polynomial, suitable requirements for a response surface design were as follows:

(a) The design should allow the graduating polynomial of chosen degree d_1 to represent the true function as well as possible within the region of interest;
(b) It should allow a check to be made on the representational adequacy of the polynomial;
(c) It should not contain an excessively large number of experimental points;
(d) It should lend itself to "blocking";
(e) It should form a nucleus from which a satisfactory design of higher order could be built in case the polynomial of degree d_1 proved representationally inadequate.

1.3 *The Necessity for Considering Bias as Well as Variance.* Most approaches to the theory of experimental design have been concerned only with the errors arising from "sampling" variation. The graduating function is assumed to be capable of providing a perfect representation and the expectation of a fitted value $\hat{y}(\mathbf{x})$ is supposed equal to $\eta(\mathbf{x})$ the true value. In practice, of course, this specification is inadequate because a graduating function such as a polynomial will always fail, to some extent, to represent the true function. In reality therefore there are not one, but two, possible sources of discrepancy between the true function $\eta(\mathbf{x})$ and the fitted graduating function $\hat{y}(\mathbf{x})$. The first occurs because of sampling error and the second because of the inadequacy of the graduating function. We refer to the first as the "variance error" and the second as the "bias error."

Derivation of statistical designs as if "variance error" were the only cause of discrepancy has often led to conclusions which are at odds with the experimenter's natural intuition. Consideration of why this happens helps us to see how we ought to proceed. Usually there will be a large "operability region," O, of unknown, or vaguely known, extent within which it is *possible* to carry out experiments and within this, at a given stage of experimentation, a smaller "region of immediate interest," R. In this paper it is assumed that R is entirely within O and the boundary of O is never reached by any experimental point. Within the region of immediate interest R, the experimenter may feel it is reasonable to represent the response function by, for example, a polynomial of first or second degree, although he may know that such a representation would be quite inadequate over the whole operability region O.

Assuming a particular model, the variance of $\hat{y}(\mathbf{x})$ at some point $\mathbf{x}$ will normally decrease as the size of the design is increased, so that if variance error

is treated as the only kind of discrepancy we are led to the conclusion that in order to obtain a good representation over R we ought to take as large a design as possible covering the *whole operability region* O. This result is at odds with common sense and is only reached because we ignore the decreased ability of the simple graduating function to represent the real relationship as wider and wider regions of the space of the variables are considered. It may be argued that to avoid this difficulty one could simply choose the design points to "cover" the limited region R. Against this however is the consideration that it might turn out that, even where bias was taken into account, to obtain a close fit within R it was best to locate some points outside R. Alternatively it might be best to confine all the design points to a region much smaller than R.

2. METHOD

Clearly what we require is some way in which the apparent added precision obtainable by making the design larger may be balanced against the loss of representational accuracy. So far we have mentioned only the choice of size of the design. In practice, the values of other quantities (such as λ_4 for a second order rotatable design [6]), which determine the distribution of the design points in the space of the variables, have to be decided upon and again similar questions arise.

The object of the present paper is to consider a general theory which will meet the first two requirements (a) and (b) listed above when both "variance error" and "bias error" are taken into account. To develop this theory we use a principle which might with profit be adopted more often in statistical investigations. This is that, rather than suppose, as is usually done, that the assumed model is always correct, we suppose that it is always to some extent incorrect. The general theory which is developed is applied to the particular problem of choosing a first order design, that is a design which is appropriate when the polynomial fitted is of first degree in $x_1, \cdots, x_k$.

2.1 *Interpretation of Requirement* (*a*). We shall interpret requirement (a), that the design should allow a graduating polynomial of specific degree d_1 to represent the true function as well as possible within the region of interest R, in the following way.

We suppose that, although it is assumed that within R a polynomial of degree d_1 will be adequate, in fact the true function is a polynomial of higher degree $d_2 > d_1$ and that this polynomial of higher degree provides an *exact* representation within the whole operability region O. Let $\hat{y}(\mathbf{x})$ be the value of the estimated response at the point $\mathbf{x}$ where this value is obtained by the least squares fitting of the polynomial of degree d_1 using a design containing N points. Let $\eta(\mathbf{x})$ be the true response at this point given exactly by the polynomial of higher degree d_2 and let σ^2 be the experimental error variance. Then we should like to choose the design so as to minimize

$$J = \frac{N}{\sigma^2} \int_R E[\hat{y}(\mathbf{x})] - \eta(\mathbf{x})]^2 d\mathbf{x} \Big/ \int_R d\mathbf{x} \tag{5}$$

where $dx = dx_1, dx_2 \cdots dx_k$. Thus J is the mean squared deviation from the true response, averaged over the region R and normalized with respect to the number of observations and the variance. To illustrate the division into "variance error" and "bias error," (5) can be written

$$J = V + B \tag{6}$$

$$= \frac{N\Omega}{\sigma^2}\int_R V[\hat{y}(\mathbf{x})]d\mathbf{x} + \frac{N\Omega}{\sigma^2}\int_R [E\hat{y}(\mathbf{x}) - \eta(\mathbf{x})]^2 d\mathbf{x} \tag{7}$$

where

$$\Omega^{-1} = \int_R d\mathbf{x}.$$

It will be noted that the expression V is the *variance function* of the reference [6] averaged over the region R. A similar criterion has been employed in recent work by David and Arens [10], and by Folks [12].

2.2 *Interpretation of Requirement* (*b*). We shall interpret requirement (b), that the design should allow a check to be made on the representational accuracy of the assumed polynomial in the following way.

We suppose that a test for lack of fit is to be made by the use of an analysis of variance in which the residual sum of squares

$$S_R = \sum_{u=1}^{N} (\hat{y}_u - y_u)^2 \tag{8}$$

is compared with the experimental error variance. This test may involve the comparison of S_R either with a prior value σ^2 of the experimental error variance, supposed to be known exactly, or with some independent estimate s^2. In either case, a parameter which determines the power of the test for goodness of fit will be the quantity

$$\sum_{u=1}^{N} [E(\hat{y}_u) - \eta_u]^2 = E(S_R) - \nu\sigma^2 \tag{9}$$

where ν is the number of degrees of freedom on which the residual sum of squares is based. While our ultimate object should be to make the power of the test as large as possible, in any particular instance in which ν is assumed fixed, an intermediate objective will be to make the expectation of S_R large.

We shall interpret requirement (b) therefore as implying that the design should be chosen so as to make $E(S_R)$ large. It seems reasonable to regard requirement (a) as being of major importance so that in practice we shall proceed by first attempting to find the class of designs which minimizes J in (5) and then attempting to satisfy criterion (b) by selecting from this class, a sub-class which makes large the expected value of S_R in (8).

3. APPLICATION: CHOOSING A DESIGN FOR FITTING A STRAIGHT LINE

We first illustrate the theory in the simple case $d_1 = 1$, $d_2 = 2$, $k = 1$. Here the situation is that a relationship

$$\eta = g(\xi) \tag{10}$$

is supposed to exist between the response η and a variable ξ such as temperature. The exact nature of the function $g(\xi)$ is unknown but the experimenter feels that he can usefully graduate this relationship by a linear equation in ξ over a limited region R within O extending from ξ' to ξ''. We assume that over some wider operability region O it can in fact be represented *exactly* by a quadratic equation in ξ. For simplicity and without loss of generality we make the transformation

$$x = \frac{2\xi - \xi' - \xi''}{\xi'' - \xi'} \tag{11}$$

so that in terms of x the region R extends from -1 to $+1$. Thus the fitted linear expression intended to approximate the true function over the region R is

$$\hat{y}(x) = b_0 + b_1 x \tag{12}$$

while the true relationship over the whole region O is assumed to be

$$\eta(x) = \beta_0 + \beta_1 x + \beta_{11} x^2 \tag{13}$$

We shall assume in what follows that

$$\sum_{u=1}^{N} x_u = 0,$$

which implies that the experimental design is centrally located in the region of interest R. We then apply our theory to decide what is the best distribution of the experimental points subject only to this restriction.

3.1 *Requirement (a); true model quadratic.* Denote the second and third moments of the design points respectively by

$$c = [11] = N^{-1} \sum_{u=1}^{N} x_u^2 \tag{14}$$

$$[111] = N^{-1} \sum_{u=1}^{N} x_u^3 \tag{15}$$

Then

$$V = \frac{N}{\sigma^2} \int_{-1}^{1} V\{\hat{y}(x)\} dx \Big/ \int_{-1}^{1} dx \tag{16}$$

$$= \frac{1}{2} \int_{-1}^{1} \left(1 + \frac{x^2}{c}\right) dx = 1 + \frac{1}{3c} \tag{17}$$

Now with $\eta(x)$ given by the quadratic model (13) the expected values of b_0 and b_1 in the fitted linear equation (12) are

$$E(b_0) = \beta_0 + c\beta_{11} \tag{18}$$

$$E(b_1) = \beta_1 + \frac{[111]}{c}\beta_{11}. \tag{19}$$

Therefore

$$E\hat{y}(x) = E(b_0) + E(b_1)x$$

$$= \beta_0 + \beta_1 x + \beta_{11}\left[c + \frac{[111]}{c}x\right], \tag{20}$$

whence, with $\alpha_{11}^2 = N\sigma^{-2}\beta_{11}^2$,

$$B = \frac{1}{2}\alpha_{11}^2 \int_{-1}^{1} \{E\hat{y}(x) - \eta(x)\}^2 dx \tag{21}$$

$$= \frac{1}{2}\alpha_{11}^2 \int_{-1}^{1}\left[c - x^2 + \frac{[111]}{c}x\right]^2 dx \tag{22}$$

$$= \alpha_{11}^2\left[c^2 - \frac{2c}{3} + \frac{1}{5} + \frac{[111]^2}{3c^2}\right]. \tag{23}$$

Combining (17) and (23) we obtain

$$J = V + B$$

$$= \left[1 + \frac{1}{3c}\right] + \alpha_{11}^2\left[c^2 - \frac{2c}{3} + \frac{1}{5} + \frac{[111]^2}{3c^2}\right] \tag{24}$$

The quantity V in the first bracket is the contribution from sampling error and the remaining quantity B is the contribution from bias. From (24) it is seen immediately that whatever the values of c and α_{11} the expression is minimized for changes in $[111]$ when this third moment is zero. In the subsequent discussion therefore we suppose that $[111]=0$.

With $[111]$ set equal to zero

$$J = V + B = \left[1 + \frac{1}{3c}\right] + \alpha_{11}^2\left[\left(c - \frac{1}{3}\right)^2 + \frac{4}{45}\right] \tag{25}$$

and the only design characteristic contained in this expression is c. It will be noted that while the part V does not contain any of the parameters β, the part B, which expresses the integrated bias, contains $\alpha_{11}^2 = N\beta_{11}^2\sigma^{-2}$ a "standardized" measure of the quadratic curvature. The value of c which minimizes the whole expression J depends, therefore, on the size of α_{11}^2.

Now the sampling error in $\hat{y}$ is proportional to $\sigma/\sqrt{N}$ so that the quantity

$$\alpha_{11} = \beta_{11}/(\sigma/\sqrt{N}) \tag{26}$$

is a measure of the ratio of the quadratic curvature to the sampling error. When α_{11} is very small, the curvature is very small compared with the sampling error, and the optimum c tends to infinity. On the other hand, when α_{11} tends to infinity, the curvature is very large compared with the sampling error and c tends to the value $\frac{1}{3}$.

The quantity $\sqrt{c}$ is the root mean square deviation of the experimental points from the origin. It is a convenient measure of the spread of the design points which, if the x's were random variables, would correspond to the standard deviation σ_x of x.

In the case where there is no experimental error and all the discrepancy arises from bias, that is from the inadequacy of the linear model, we are instructed therefore to choose the x_u so that the root mean square deviation

$$\sqrt{c} = \frac{1}{\sqrt{3}} = 0.58 \tag{27}$$

To summarize: if we believed implicitly in the assumption of linearity over an indefinitely wide region then we should minimize V alone, which would lead us to allow the bounds of the design to extend as far as possible. At the other extreme, if with R extending from $-\theta$ to $+\theta$ we knew that errors which occurred in the observations were negligible but we were doubtful about the assumption of linearity, then we should minimize B alone, which would lead us to limit the range of the observations so that the root mean square deviation $\sqrt{c}=0.58\theta$.

It will be seen that the two extreme cases lead to widely different conclusions about the optimal choice of the spread of the design. To make further progress we must obtain some idea of what might be a "typical value" for α_{11} in the usual situation where variance and bias *both* occur. Now it might be expected that if the experimenter were graduating a function by a polynomial over a particular limited region R, he would try to choose the degree of polynomial so that, over the region considered, the average size of error arising from the bias part was at least no larger than the average size of error arising from the variance part and, in a typical case, things might perhaps be arranged so that these contributions were about equal in magnitude. To see what this would imply for the present example we write (25) in the form

$$J = V(c) + \alpha_{11}^2 B(c). \tag{28}$$

For every value of α_{11} there will exist a corresponding minimizing value of c and corresponding values for $V=V(c)$ and $B=\alpha_{11}^2 B(c)$. We shall proceed by choosing c from these minimizing values so that the value of α_{11} makes $V=B$.

Now in general we can choose c from the minimizing values so that the value of α_{11} makes $V=gB$ where g is any desired positive constant by minimizing the product $\{V(c)\}^g B(c)$.

For, let c^* be the value of c that satisfies

$$\frac{\partial}{\partial c}\left[\{V(c)\}^g B(c)\right] = 0, \tag{29}$$

that is, at $c=c^*$,

$$g\{V(c)\}^{g-1}V'(c)B(c) + \{V(c)\}^g B'(c) = 0 \tag{30}$$

Provided the product $g\{V(c)\}^{g-1}B(c)B'(c)\neq 0$, we have

$$\frac{1}{g}\frac{V(c^*)}{B(c^*)} = -\left[\frac{V'(c)}{B'(c)}\right]_{c=c^*} = \alpha_{11}^{*2}, \quad \text{say.} \tag{31}$$

Thus equation (29) provides the values c^* of c, and α^* of α such that

$$\text{(a)} \quad V'(c) + \alpha_{11}{}^{*2}B'(c) = \frac{\partial}{\partial c}(V + B) = 0 \tag{32}$$

$$\text{(b)} \quad V(c) - g\alpha_{11}{}^{*2}B(c) = V - gB = 0 \tag{33}$$

Putting $g = 1$ we find that the optimal values of $\sqrt{c}$ when $V = B$ is 0.62, when the value of α_{11} is 4.49. This compares with the value $\sqrt{c} = 0.58$ when variance is completely ignored.

We are led therefore to a somewhat remarkable conclusion. This is that the optimal design in a "typical" situation in which the influence of bias and variance are equal is very nearly the same as that obtained *when variance is ignored completely* and the experiment is designed to minimize bias alone.

It is admitted that the concept, that the experimenter would typically arrange matters so that the average errors in the predicted response due to sampling variation and to bias were about the same, is not very precise and in practice considerable deviations from it would be expected to occur. It is seen from Table 630 however, that even in the case where the variance contri-

TABLE 630. OPTIMAL VALUES FOR $\sqrt{c}$ WITH ASSOCIATED VALUES OF α_{11}

$\sqrt{c}$	α_{11}	
∞	0	(variance contribution completely dominant)
0.72	1.82	(variance contribution four times that of bias)
0.62	4.49	(variance and bias contributions equal)
0.58	∞	(bias contribution completely dominant).

bution is *four times* the bias contribution the resulting root mean square value $\sqrt{c}$ is very close to the value obtained when variance is ignored completely and entirely different from the value obtained when bias is ignored. It will be seen in the following sections that similar results are obtainable for the case of k variables. Almost all previous investigators (notable exceptions being Hotelling [14], and more recently David and Arens [10], and Folks [12]), have been principally concerned with the minimization of variance. The recent work raises the possibility that this preoccupation with variance is mistaken and, if in some design problems a simplification of the situation had to be made, it might be better to ignore sampling variation rather than to ignore bias.

3.2 *The Effect of Bias When the True Model Is a General Polynomial.* We have seen that the situation where all the discrepancy arises from bias may be of considerable importance since, in the case studied above, it approximates to the situation in which *both* variance and bias occur. It is of some interest therefore to consider the case where a linear function is fitted and all the discrepancy arises from bias but the true function is *not* necessarily quadratic but is a polynomial of any degree d_2 whatever. It appears as a special case of a theorem proved in Appendix 1 that the integral B which measures the average error arising from bias alone is minimized when all the moments of the design

up to that of order (d_2+1) are equal to the moments of a uniform distribution over the region R. For example with the true model quadratic, all moments up to order 3 must be the same as those of a uniform distribution over the range $[-1, 1]$. This means that $\sqrt{c}$ the root mean square deviation of the design points from the center of the region must be equal to 0.58 and the third moment $[111]$ to zero: a result obtained already. In general, when the true model is a polynomial of degree d the i-th moment

$$N^{-1}\sum_{u=1}^{N} x_u{}^i = \begin{cases} 0 & i \text{ odd} \\ (i+1)^{-1} & i \text{ even} \end{cases} \quad i \leqq d_2 + 1. \tag{34}$$

This is intuitively a very reasonable result; in particular it implies that if bias only had to be considered and *nothing whatever* were known of the nature of the function $\eta(x)$ except that it could be represented by some polynomial having an indefinitely large number of terms, we would do best by spreading the design points *evenly* over the region R. It is thought that a completely general polynomial in which high order terms are given the same weight as lower order terms is not a particularly realistic model. Usually it could be safely assumed that the function would have smoothness properties which would place greater emphasis on lower order terms so that, except in the case of complete ignorance as to the nature of the function, even spacing of the design points is not necessarily indicated as the most desirable solution, even for the case where bias is completely dominant.

3.3 *Requirement (b); True Model Quadratic.* Having selected the best value of c for the design by use of criterion (a), we introduce criterion (b).

To satisfy this criterion we must make large

$$\sigma^{-2}E(S_R) - \nu = \sigma^{-2}\sum_{u=1}^{N}[E(\hat{y}_u) - \eta_u]^2, \tag{35}$$

For this example of the fitting of a straight line with the true function quadratic we obtain, by use of (12), (13) and (35),

$$\sigma^{-2}E(S_R) - \nu = \alpha_{11}{}^2c^2\left[\frac{[1111]}{c^2} - \frac{[111]^2}{c^3} - 1\right].$$

Now criterion (a) requires that $[111]=0$. It is interesting to note that criterion (b) requires this independently; for in order that (36) be maximized $[111]$ must be equal to zero. We see then from (36) that, with $\sqrt{c}$ set equal to some suitable value decided by criterion (a) and $[111]$ set equal to zero, criterion (b) aditionally requires that the fourth moment coefficient should be made large.

It can be shown that the design of this kind that maximizes the fourth moment is obtained by placing $(N-2)$ of the N points at the center of the interval and the remaining two points symmetrically about this center and equidistant from it. Intuitively such an arrangement is not particularly acceptable. The reason can be seen if we remember that in our example it has been assumed that the true function is *exactly* represented by a quadratic polynomial. If this were not the case, and in fact a cubic term had to be included, then from Section 3.2, minimization of bias alone would require that the fourth moment

[1111] should be $\frac{1}{5}$. This implies that the measure of "kurtosis" $[1111]/c^2$ for the design points should be 1.8 —a small value. The practical implication is that to get good detectability of quadratic departure from the assumed model, the design should be chosen so that the fourth moment was not small but the value should not be taken so large as to cause serious cubic bias in B. The precise value chosen will depend on one's relative anxiety to get good graduation on the one hand and good detectability of quadratic lack of fit on the other.

The conclusions to be drawn from the application of criteria (a) and (b) to this simple case therefore are as follows. The design should be selected so that:

(i) the third moment of the distribution of the design points is zero;
(ii) the size of the design is such that the root mean square distance of the design points from the origin is approximately $.6\theta$, with 2θ the width of the region over which a linear approximation is required. (This is very close to the value which would be appropriate if there were *no experimental error* and is completely contrary to the conclusion obtained when bias is ignored.)
(iii) In order that a quadratic tendency in the true model can be readily detected, the fourth moment of the design should not be small.

4. APPLICATION: CHOOSING A DESIGN FOR FITTING A PLANE

Suppose now that there are k variables $\xi_1, \xi_2, \cdots, \xi_k$, that the nature of the true relationship $\eta = g(\xi, \theta)$ is unknown but the experimenter believes it can usefully be graduated by a linear relationship over some region R of immediate interest.

The fitted function is then

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k \tag{36}$$

or, in matrix notation,

$$\hat{y} = \mathbf{x}_1'\mathbf{b}_1 \tag{37}$$

where

$$\mathbf{b}_1' = (b_0, b_1, \cdots, b_k)$$
$$\mathbf{x}_1' = (1, x_1, \cdots, x_k)$$

The true relationship which applies over the whole operability region O is assumed to be some polynomial function which, for the moment, we will suppose is of some *unspecified* degree d_2 in the k variables.

$$\eta = \beta_0 + \beta_1x_1 + \cdots + \beta_kx_k + \beta_{11}x_1^2 + \cdots + \beta_{kk}x_k^2 + \beta_{12}x_1x_2 + \cdots + \beta_{(k-1)k}x_{k-1}x_k + \beta_{111}x_1^3 + \cdots \tag{38}$$

or, in matrix notation,

$$\eta = \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_2 \tag{39}$$

where

$$\boldsymbol{\beta}_1' = (\beta_0, \beta_1, \cdots, \beta_k)$$
$$\mathbf{x}_1' = (1, x_1, \cdots, x_k)$$

$$\beta_2' = (\beta_{11}, \beta_{22}, \cdots, \beta_{kk}; \beta_{12}, \beta_{13}, \cdots, \beta_{(k-1)k}; \beta_{111}, \cdots)$$
$$\mathbf{x}_2' = (x_1^2, x_2^2, \cdots, x_k^2; x_1x_2, x_1x_3, \cdots, x_{k-1}x_k; x_1^3, \cdots)$$

and each of the vectors β_2' and $\mathbf{x}_2'$ contains p terms, where

$$p = \binom{d_2 + k}{d_2} - (k + 1)$$

We have from (7)

$$J = \Omega \frac{N}{\sigma^2} \int_R E[\hat{y}(\mathbf{x}) - \eta(\mathbf{x})]^2 d\mathbf{x} = V + B \tag{40}$$

with

$$V = N\Omega \int_R \mathbf{x}_1'(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\mathbf{x}_1 d\mathbf{x} \tag{41}$$

and

$$B = \frac{N}{\sigma^2} \Omega \int_R \beta_2'[\boldsymbol{A}'\mathbf{x}_1 - \mathbf{x}_2][\mathbf{x}_1'\boldsymbol{A} - \mathbf{x}_2']\beta_2 d\mathbf{x} \tag{42}$$

where

$$\boldsymbol{X}_1' = [\mathbf{x}_{11}, \cdots, \mathbf{x}_{1u}, \cdots, \mathbf{x}_{1N}] \tag{43}$$

is a $(k+1)\times N$ matrix with $\mathbf{x}_{1u}' = (1, x_{1u}, x_{2u}, \cdots, x_{ku})$,

$$\boldsymbol{X}_2' = [\mathbf{x}_{21}, \cdots, \mathbf{x}_{2u}, \cdots, \mathbf{x}_{2N}] \tag{44}$$

is a $p\times N$ matrix with

$$\mathbf{x}_{2u}' = (x_{1u}^2, x_{2u}^2, \cdots, x_{ku}^2; x_{1u}x_{2u}, \cdots, x_{(k-1)u}x_{ku}; x_{1u}^3, \cdots);$$

and

$$\boldsymbol{A} = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2 \tag{45}$$

is the $(k+1)\times p$ "alias matrix." This last matrix (for example, [8]) has, for its elements, quantities which measure the extent to which thee stimates $\boldsymbol{b}_1$ are biased by higher order coefficients in accordance with the equation

$$E(\boldsymbol{b}_1) = \beta_1 + \boldsymbol{A}\beta_2. \tag{46}$$

Making the necessary substitutions we obtain

$$\Omega^{-1}V = \sum_{i=0}^{k}\sum_{j=0}^{k} c^{ij} \int_R x_i x_j d\mathbf{x} \tag{47}$$

$$\Omega^{-1}B = \boldsymbol{\alpha}_2'\boldsymbol{A}'\left(\int_R \mathbf{x}_1\mathbf{x}_1' d\mathbf{x}\right)\boldsymbol{A}\boldsymbol{\alpha}_2 - 2\boldsymbol{\alpha}_2'\left(\int_R \mathbf{x}_2\mathbf{x}_1' d\mathbf{x}\right)\boldsymbol{A}\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_2'\left(\int_R \mathbf{x}_2\mathbf{x}_2' d\mathbf{x}\right)\boldsymbol{\alpha}_2 \tag{48}$$

where

$$\alpha_2 = \frac{1}{\sigma/\sqrt{N}}\,\beta_2$$

and c^{ij} are the elements of the matrix $\{c^{ij}\} = \{c_{ij}\}^{-1} = \mathbf{C}^{-1} = N(\mathbf{X}_1'\mathbf{X}_1)^{-1}$.

We note in passing that, if we were concerned only with minimizing the average squared bias B, then as a special case of the result of Appendix 1 *whatever* the nature of the region R and *whatever* the degree d_2 of the polynomial describing the true response, for all values of β_2 the bias term B is minimized when all the moments of the design up to order d_2+1 are made equal to the moments of a uniform distribution over R. This result has been still further generalized to include weighting functions, in unpublished work by C. L. Mallows. In particular this implies, as for a single variable, that, if only bias had to be considered and if the rather unrealistic assumption were made that nothing whatever was known of the nature of the function over the region R, then we should do best by spreading the points evenly over R.

4.1 *Choice of the Region R.* The specific results we obtain, whether from minimizing V, B or the whole integral J, will of course depend on how we define the region R. In the present paper we shall suppose that considerations of strategy (for example [1]) dictate the sequential exploration of subregions R entirely contained within O. Such a strategy is often appropriate, for example, in experiments designed to find and to explore a region in which some response or responses have optimal values (for example [8]). To give satisfactory expression to the experimenter's desire to use designs which symmetrically generate information, it seems reasonable to choose R to be a symmetric region in the coordinate system *currently* believed by the experimenter to be most appropriate. Of course the experimenter's ideas as to what *is* the best coordinate system in which to work will almost certainly change as the investigation proceeds. For example suppose he began his investigation with a simple 2^2 factorial in variables x_1 and x_2. This would seem to imply that he currently expected that the response surface could be conveniently represented in an orthogonal coordinate system (x_1, x_2) with the scales of measurement for the two variables proportional to the step size in the factorial design. In fact (and unknown at this time to him) when so represented the response surface might be highly unsymmetrical; for example it might have a ridge-like appearance. In such a case, as he proceeded iteratively from one group of experiments to another, the information built up about the response surface would probably lead him by a process such as is already described [4] to employ, in later stages, transformations and changes of scales and metrics in the variables in terms of which the response surface could be more symmetrically and simply described. At any given stage he would work with that coordinate system which his experience had so far led him to believe would provide the simplest and most symmetric representation of the response surface. Later experience would usually show that his ideas were capable of improvement and would lead to modifications.

We shall try to select designs therefore which generate information symmetrically in that coordinate system currently thought to be best and we will interpret "symmetric" to mean that the region R measured in this coordinate

system is a sphere. Adopting this convention the particular case which we study further in this paper is that where a *linear* function is fitted over a *spherical* region R and the true model involving $p=\frac{1}{2}k(k+1)$ extra constants is in fact *quadratic* over the whole of the operability region O.

4.2 *Requirement (a); True Model Quadratic and Region R Spherical.* We assume that the "center of gravity" of the design points is at the center of the region R, defined by

$$\sum_{i=1}^{k} x_i^2 \leqq 1$$

where

$$x_i = \frac{\xi_i - \bar{\xi}_i}{S_i}, \tag{49}$$

Now for such a region

$$\int_R x_1^{\delta_1} x_2^{\delta_2} \cdots x_k^{\delta_k} dx_1 \cdots dx_k$$

$$= \frac{\Gamma\left(\frac{\delta_1+1}{2}\right)\Gamma\left(\frac{\delta_2+1}{2}\right)\cdots\Gamma\left(\frac{\delta_k+1}{2}\right)}{\Gamma\left\{\frac{\sum_{i=1}^{k}(\delta_i+1)}{2}+1\right\}} \tag{50}$$

unless any δ_i is odd, when the value of the integral is zero.

Hence

$$\Omega \int_R \mathbf{x}_1\mathbf{x}_1' d\mathbf{x} = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & (k+2)^{-1}I_k \end{array}\right] \tag{51}$$

and with $\mathbf{x}_2'$ now the vector of quadratic variables

$$(x_1^2, \cdots, x_k^2; x_1x_2, \cdots, x_{k-1}x_k),$$

$$\Omega \int_R \mathbf{x}_2\mathbf{x}_1' d\mathbf{x} = (k+2)^{-1}\left[\begin{array}{c|c} j_k & 0 \\ \hline 0 & 0 \end{array}\right] \tag{52}$$

$$\Omega \int_R \mathbf{x}_2\mathbf{x}' d\mathbf{x} = (k+2)^{-1}(k+4)^{-1}\left[\begin{array}{c|c} 2I_k + j_kj_k' & 0 \\ \hline 0 & I_{p-k} \end{array}\right] \tag{53}$$

where $j_k' = [1\ 1\ \cdots\ 1]$ and I_k is the unit matrix of order k.

4.2a *Vanishing of the Third Order Moments.* Now with $\boldsymbol{X}_2$ a matrix of quadratic variables, the alias matrix A appearing in (48) can be partitioned after the first row as follows:

$$A = \left[\begin{array}{c} A_1 \\ \hline A_2 \end{array}\right] \tag{54}$$

where $A_1 = (c_{11}, \cdots, c_{kk}; c_{12}, \cdots, c_{(k-1)k})$ is a $1 \times p$ row vector of second moments, with

$$c_{ij} = N^{-1} \sum_{u=1}^{N} x_{iu} x_{ju}$$

and A_2 is a $k \times p$ matrix of linear combinations of third moments only. Using (50) to obtain the values of the integrals in the equation (47) for V and substituting (51), (52), (53) and (54) in the equation (48) for B, the elements of the integral $J = V + B$ for this case of a plane fitted to a truly quadratic surface are found to be

$$V = 1 + \sum_{i=1}^{k} c^{ii}/(k+2) \tag{55}$$

$$B = \frac{1}{k+2} \sum_{g=1}^{k} \left\{ \sum_{i=1}^{k} \sum_{j=i}^{k} \alpha_{ij} \sum_{h=1}^{k} c^{gh}[hij] \right\}^2 + \left\{ \sum_{i=1}^{k} \sum_{j=i}^{k} \alpha_{ij}(c_{ij} - \delta_{ij}/(k+2)) \right\}^2$$
$$+ \frac{2(k+2) \sum_{i=1}^{k} \alpha_{ii}^2 + (k+2) \sum_{i=1}^{k} \sum_{j=i+1}^{k} \alpha_{ij}^2 - 2\left(\sum_{i=1}^{k} \alpha_{ii} \right)^2}{(k+2)^2(k+4)} \tag{56}$$

where

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \qquad \text{and the} \qquad \alpha_{ij} = \frac{\beta_{ij}}{(\sigma/\sqrt{N})}$$

measure the size of the second order constants β_{ij} relative to the sampling error. Now third moments $[hij]$ enter J only in the first term of B and from the form of this term we see that, if $\{c_{gh}\}$ is non-singular, J is minimized with respect to the $[hij]$ for general α_{ij} and whatever the c^{gh}, if all the $[hij]$ vanish. We proceed therefore in what follows by assuming that all third moments have these optimal zero values.

With third moments all zero, the expression J contains only the elements of $\boldsymbol{\alpha}_2$ and the second moments c_{ij} of the design. If the x's were random variables these c_{ij} would be the variances and covariances of the k-variate distribution of the x's and they will be so called. Now, whereas V does not contain $\boldsymbol{\alpha}_2$, the integrated bias B is a quadratic form in the elements of $\boldsymbol{\alpha}_2$. In general then, if contributions from bias and variance are considered simultaneously, values of c_{ij} which minimize J will depend on the elements of $\boldsymbol{\alpha}_2$, that is they will depend on the size of the quadratic parameters β_{ij} relative to the sampling error. Before studying this general case further we first consider the extreme cases of completely dominant variance and completely dominant bias.

4.2b *Case of Completely Dominant Variance.* In the case of completely dominant variance, the model is assumed to fit perfectly and there is no contribution from the component B. The integral J is minimized by minimizing V. In general, designs chosen to minimize V will extend to the limits of the operability region O.

For example suppose the size of the design is limited by limiting the variances of the k-variate distribution of points so that

$$c_{ii} \leqq \gamma_i \quad i = 1, \cdots, k \tag{57}$$

with γ_1 suitably chosen constants. It is shown [13, 16, 19, 2] that V is minimized when

$$\left.\begin{aligned} c_{ii} &= \gamma_i \quad i = 1, \cdots, k \\ c_{ij} &= 0 \quad \text{all } i, j \end{aligned}\right\} \tag{58}$$

Thus in these circumstances we should use a first order *orthogonal* design, that is a design which is such that the column vectors of the design matrix have zero inner products one with another and with a vector of ones. This orthogonal design would be chosen to have largest possible values for the "variances" of the x's.

4.2c *Case of Completely Dominant Bias.* In the other extreme case of completely dominant bias, where the model is not assumed to fit perfectly and the variance contribution approaches zero, J is minimized by minimizing B. From inspection of equation (56) with the $[hij]=0$, or from the general result of Appendix 1, the minimizing values for the c_{ij} are seen to be

$$\left.\begin{aligned} c_{ii} &= c = (k+2)^{-1} \quad i = 1, \cdots, k \\ c_{ij} &= 0 \quad \text{all } i, j \end{aligned}\right\} \tag{59}$$

Thus to minimize bias alone over the spherical region R, the design must be orthogonal with third moments all zero and the "variances" of the x's all equal to $(k+2)^{-1}$. What is implied concerning the *size* of the design can be seen as follows. If

$$r_u = \left(\sum_{i=1}^{k} x_{iu}^2 \right)^{1/2}$$

is the distance of the u-th design point from the center of the region R, then for the optimal design

$$kc = \frac{k}{k+2} = \sum_{i=1}^{k} c_{ii} = N^{-1} \sum_{u=1}^{N} \sum_{i=1}^{k} x_{iu}^2 = N^{-1} \sum_{u=1}^{N} r_u^2 = \overline{r^2} = \dot{r}^2 \tag{60}$$

where $\dot{r}$ is the root mean square distance of the experimental points from the origin and so equals $\sqrt{c}$ if $k=1$. Thus for minimal bias $\dot{r}$ the root mean square distance of the experimental points from the center of R must be $\sqrt{k/(k+2)}$ times the radius of R.

It is interesting to see that the orthogonality condition arises here as a result of minimizing *bias alone* over a spherical region. In previous investigations the conclusion that an orthogonal design was optimal has usually been arrived at

by minimizing only variance. The result concerning minimization of bias alone is of some practical interest in certain problems of numerical analysis [4, 3] where there are no sampling errors but planarity assumptions may not be completely justified.

4.2d *Intermediate Case of Contributions from Both Variance and Bias.* In practice usually both variance and bias occur simultaneously. Now for *known* α_{ij} the optimum values of the c_{ij} could readily be found; usually however, the α_{ij} are unknown. In these circumstances we can make some progress by proceeding in the following way. In practice we usually do not know the nature of the response surface with which we are dealing; in particular we do not know the orientation of the response surface with respect to the design. We proceed therefore by taking the average value of J over all orthogonal rotations of the response surface denoting the average thus obtained by $\tilde{J}$. Since V does not contain any of the constants of the response function, such a process leaves the value of (55) unchanged. The quantity B however is a quadratic form in the elements of the vector $\boldsymbol{\alpha}_2$ or equivalently of the vector $\boldsymbol{\beta}_2$. The average value of B over all rotations of the surface is obtained by substituting the averaged values over all rotations of the products $\alpha_{gh}\alpha_{ij}$ which occur in (56). The details of the averaging are given in Appendix 2. Using these averaged values, and with the optimal value of zero substituted for all the third moments we obtain

$$\begin{aligned}\tilde{J} &= V + \tilde{B} \\ &= \left[1 + (k+2)^{-1}\sum_{i=1}^{k} c^{ii}\right] + a\sum_{i=1}^{k}\left(c_{ii} - \frac{1}{k+2}\right)^2 \\ &\quad + b\sum_{i=1}^{k}\sum_{j=1}^{k}\left(c_{ii} - \frac{1}{k+2}\right)\left(c_{jj} - \frac{1}{k+2}\right) + 2(a-b)\sum_{i=1}^{k}\sum_{j=1}^{k} c_{ij}^2 \\ &\quad + k(k+2)^{-2}(k+4)^{-1}[k(k+3)a - (k-1)(k+4)b], \qquad (61)\end{aligned}$$

where

$$a = \widetilde{\alpha_{ii}^2} = N\sigma^{-2}\widetilde{\beta_{ii}^2}, \qquad b = \widetilde{\alpha_{ii}\alpha_{jj}} = N\sigma^{-2}\widetilde{\beta_{ii}\beta_{jj}}$$

Now V is minimized with respect to the c_{ij} $(i \neq j)$ when $\boldsymbol{C}=N^{-1}(\boldsymbol{X}'\boldsymbol{X})$ is diagonal, i.e. when

$$c_{ij} = 0.$$

But these are also the values of c_{ij} $(i \neq j)$ which minimize $\tilde{B}$; hence minimization of J for variations in c_{ij} $(i \neq j)$, irrespective of the values of the c_{ii}, is achieved when

$$c_{ij} = 0. \qquad (62)$$

This implies that

$$c^{ii} = c_{ii}^{-1}. \qquad (63)$$

After substituting (62) and (63) in (61), differentiating with respect to the c_{ii} $(i=1, 2, \cdots, k)$ and equating the result to zero, we obtain k equations the i-th one of which is

$$-(k+2)^{-1}c_{ii}^{-2}+2a\left(c_{ii}-\frac{1}{k+2}\right)+2b\sum_{j=1}^{k}\left(c_{jj}-\frac{1}{k+2}\right)=0. \tag{64}$$

Subtracting the g-th such equation from the i-th then gives

$$(k+2)^{-1}c_{ii}^{-2}c_{gg}^{-2}(c_{ii}^{2}-c_{gg}^{2})+2a(c_{ii}-c_{gg})=0 \tag{65}$$

which, since the c_{ii} are positive, implies that

$$c_{ii}=c_{gg}=c \quad (i,g=1,2,\cdots,k) \tag{66}$$

We are thus led to the conclusion that the criterion J averaged for all orientations of the response surface is minimized when the design is a first order orthogonal design with all third moments zero and with the variances of the x_i all equal.

4.2c *Summary of Results.* We have shown then that if the surface is truly quadratic over the operability region O and a linear model is fitted over a smaller spherical region within O defined by $\sum x_i^2 \leq 1$:

(i) $V+B$ is minimized, for all β_{ij} and c_{ij}, when the third moments $[ijk]$ of the design are chosen to be zero.

(ii) V alone is minimized, for O defined by $c_{ii} \leq \gamma_i$, when the design is chosen to be first order orthogonal with $c_{ii}=\gamma_i$, that is with the design of maximum possible size.

(iii) B alone is minimized when the design is chosen to be first order orthogonal with all third order moments zero and $c_{ii}=1/(k+2)$.

(iv) $V+B$ is minimized if values of β_{ij} averaged over all rotations are substituted, when the design has all third order moments zero and is first order orthogonal with the c_{ii} all equal.

These conclusions suggest that we should accept that the best design to use in the circumstances implied by our assumptions is first order orthogonal with the c_{ii} equal and with all third order moments zero, that is such that

$$\begin{aligned} c_{ii} &= c_{jj} = c \quad (\text{all } i \text{ and } j); \\ c_{ij} &= 0 \qquad\quad (i \neq j); \\ [ijk] &= 0. \end{aligned} \tag{67}$$

Adopting the terminology used before [8] we shall call designs of this class first order orthogonal designs *of type B.*

4.3 *Choice of Optimal Size for First Order Orthogonal Design of Type B.* The question remains of the optimal size of the first order orthogonal type B design. This involves the choice of c or equivalently of $\dot{r}=\sqrt{kc}$, the root mean square distance of the experimental points from the centre of the design. We have seen that for minimization of V alone the size of the design should be as large as possible whereas for minimization of B alone $\dot{r}$ the root mean square distance of the experimental points from the center of R should be $\sqrt{k/(k+2)}$, that is smaller than the radius of R. We need to reach some compromise in the practical situation where neither the contribution from variance nor that from bias can be ignored.

The variance and bias functions for any first order orthogonal design of type B in the general case where no assumptions are made concerning the nature of the β_{ij} may be found by setting $c_{ii}=c_{jj}=c$ (all i and j); $c_{ij}=0$ $(i\neq j)$ and $[ijk]=0$ (all i and j) in (55) and (56). We find

$$V = 1 + \frac{k}{k+2}\,\frac{1}{c} \tag{68}$$

$$B = \Big(\sum_i \alpha_{ii}\Big)^2\Big(c - \frac{1}{k+2}\Big)^2 + \frac{2(k+2)\sum_i \alpha_{ii}^2 + (k+2)\sum_{i<j}\sum \alpha_{ij}^2 - 2\Big(\sum_i \alpha_{ii}\Big)^2}{(k+2)^2(k+4)}. \tag{69}$$

Now consider a symmetric $k\times k$ matrix $\boldsymbol{\alpha}$ for which the i-th diagonal element is α_{ii} and for which the element occupying the intersection of the i-th row and j-th column is $\frac{1}{2}\alpha_{ij}$ $(i\neq j)$ and $\alpha_{ij}=\alpha_{ji}$. We have

$$\operatorname{tr}\boldsymbol{\alpha} = \sum_i \alpha_{ii}, \qquad \operatorname{tr}\{\boldsymbol{\alpha}^2\} = \sum_i \alpha_{ii}^2 + \frac{1}{4}\sum_{i\neq j}\sum \alpha_{ij}^2.$$

Writing

$$\theta = \operatorname{tr}\{\boldsymbol{\alpha}^2\} \qquad \varphi = (\operatorname{tr}\boldsymbol{\alpha})^2/\operatorname{tr}\{\boldsymbol{\alpha}^2\} \tag{70}$$

we have finally for any first order orthogonal design of type B

$$J = V + B = \Big\{1 + \frac{k}{(k+2)c}\Big\} + \theta\Big\{\varphi\Big(c - \frac{1}{k+2}\Big)^2 + \frac{2(k+2-\varphi)}{(k+2)^2(k+4)}\Big\} \tag{71}$$

Remembering that $\alpha_{ij}=\beta_{ij}\sqrt{N}/\sigma$, the matrix $\boldsymbol{\alpha}$ is seen to be a simple multiple of a symmetric matrix whose diagonal elements are the quadratic coefficients β_{ii} and whose off-diagonal elements are one-half the interaction coefficients β_{ij} of the true surface. Suppose the k latent roots of this latter matrix are $\lambda_1, \lambda_2, \cdots, \lambda_k$. Then

$$\theta = \operatorname{tr}\{\boldsymbol{\alpha}^2\} = \frac{N}{\sigma^2}\Big\{\sum_{i=1}^k \beta_{ii}^2 + \frac{1}{2}\sum_{i=1}^k\sum_{j=i+1}^k \beta_{ij}^2\Big\} = \frac{N}{\sigma^2}\sum_{i=1}^k \lambda_i^2 \tag{72}$$

$$\varphi = (\operatorname{tr}\boldsymbol{\alpha})^2/\operatorname{tr}\{\boldsymbol{\alpha}^2\} = (\textstyle\sum \beta_{ii})^2 \Big/ \Big\{\sum_{i=1}^k \beta_{ii}^2 + \frac{1}{2}\sum_{i=1}^k\sum_{j=i+1}^k \beta_{ij}^2\Big\} = (\textstyle\sum \lambda_i)^2/\sum \lambda_i^2 \tag{73}$$

Now the λ_i's in the expression above are simply the coefficients in

$$\sum_{i=1}^k \lambda_i X_i^2$$

the canonical form containing no product terms, into which the quadratic part of the true model

$$\sum_{i=1}^{k} \sum_{j=i}^{k} \beta_{ij} x_i x_j$$

can be transformed by orthogonal rotation. These λ_i thus represent the quadratic effects of the canonical variable X_i in the new coordinate system and their signs and magnitudes determine the characteristics of the true response surface. For example, if all the λ_i are negative, the surface has a maximum with ellipsoidal contours, the length of the i-th principal axis being proportional to λ_i^{-1}. If one or more of the λ_i approaches zero, a line, plane or hyperplane of maxima results, while if certain of the λ_i are positive various minimax situations occur.

The quantity θ is therefore seen to be an overall measure of the magnitude of the quadratic tendency of the surface relative to the sampling error. The quantity φ on the other hand is a homogeneous function of degree zero in the λ's, being independent both of the sampling error and of the absolute magnitude of the λ's. It measures the "state of conditioning" (for example [5]) of the quadratic surface as evidenced by the variation among the λ's. In fact

$$\varphi = k/(1 + C_\lambda^2) \tag{74}$$

where

$$C_\lambda = \sqrt{\frac{\sum_i (\lambda_i - \bar{\lambda})^2}{k}} \Big/ \bar{\lambda} \tag{75}$$

is the coefficient of variation of the λ's.

If all the λ's were equal and of the same sign corresponding to the best state of conditioning of the surface (which would then have spherical contours) C_λ would be zero and φ would take its maximum value of k. At the other extreme, if the λ's were of mixed signs and $\sum \lambda_i = 0$, C_λ would be infinite and φ would take its minimum value of zero. If it so happened that p of the λ's were equal and of the same sign and the remainder were zero the value of φ would be equal to p. Thus in this case p would be a measure of the number of non-redundant canonical variables. The latter situation approximates to that found in problems where the eventual objective is the location and exploration of maxima. Here, the most common situation is that the true response surface is approximated by a system in which a point, line, plane, or space of maxima occur. For such examples p of the λ's would be negative and the remainder zero.

The optimal value of c and hence of $\dot{r} = (kc)^{1/2}$, the root mean square distance of the experimental points from the center of the design, can of course be calculated for any given values of θ and φ by finding that value which minimizes J in (71). As we have seen however, the optimal value of $\dot{r}$ can take any value between ∞ and $\sqrt{k/(k+2)}$ depending on the values chosen for θ and φ. To make further progress we need to determine what might be typical values for these constants and to do this we may proceed in a manner similar to that

adopted in the one-dimensional case. We have seen that the constant φ measures the state of conditioning of the response surface and values of particular interest lie between the limits $1 \leq \varphi \leq k$. Now for any fixed value of φ the integral J is of the form:

$$J = V(c) + \theta B(c)$$

As before, by minimizing the product $\{V(c)\}^{\varphi} B(c)$ we can choose c and hence $\dot{r}$ from these minimizing values so that $V = gB$, that is so that the value of θ makes the contribution $V = V(c)$ from variance any desired multiple g of the contribution $B = \theta B(c)$ from bias. The case $g = 1$ is of particular interest since it is plausible that an experimenter attempting to graduate a function over a particular limited region might make his choices of the size of region and degree of polynomial so that the average size of error arising from bias was about the same as that which he thought might arise from variance.

Table 642 shows the value of $\dot{r}$ for $\varphi = k, k-1, \cdots, 2, 1$ and 0.2; for $k = 2, 3$ and 5, for the four cases.

V	Variance completely dominant
$V = 4B$	Variance contribution four times that of bias
$V = B$	Variance and bias contributions equal
B	Bias completely dominant.

Considering first those entries for which the variance and bias contributions are equal, it is seen once more that over the important range $1 \leq \varphi \leq k$ the values for $\dot{r}$ are very little larger than those obtained if the variance contribution is ignored entirely, and that furthermore, even if the variance contribution is made as large as *four times* the bias contribution, the value of $\dot{r}$ still remains comparatively small.

As ϕ approaches the value 0 the optimal value of $\dot{r}$ becomes larger and eventually goes to infinity. The reason for this apparent anomaly is easily seen. If θ is not infinite, a zero value for φ implies that $\sum \lambda_i = 0$. Now $\sum \lambda_i = \sum \beta_{ii}$ is proportional to the bias in the estimate b_0 of the constant term β_0. With the particular class of designs considered b_0 is the only biased estimate, all the other estimates $b_1, b_2, \cdots, b_k$ being completely unbiased by second order terms. It follows that for the particular case in which $\sum \lambda_i = 0$ and hence $\varphi = 0$, *no second*

TABLE 642. OPTIMAL VALUE FOR $\dot{r}$ WITH ASSOCIATED VALUE OF $\sqrt{\theta}$

k	ϕ	0.2		1		2		3		4		5	
		$\dot{r}$	$\sqrt{\theta}$	$\dot{r}$	$\sqrt{\theta}$	$\dot{r}$	$\sqrt{\theta}$	$\dot{r}$	$\sqrt{\theta}$	$\dot{r}$	$\sqrt{\theta}$	$\dot{r}$	$\sqrt{\theta}$
2	V	∞	0	∞	0	∞	0						
	$V = 4B$	1.34	1.54	1.00	2.00	0.84	3.07						
	$V = B$	0.99	4.73	0.79	6.23	0.74	8.10						
	B	0.71	∞	0.71	∞	0.71	∞						
3	V	∞	0	∞	0	∞	0	∞	0				
	$V = 4B$	1.36	3.11	0.99	4.80	0.87	6.56	0.82	8.75				
	$V = B$	1.01	9.58	0.84	12.55	0.80	15.18	0.79	18.97				
	B	0.77	∞	0.77	∞	0.77	∞	0.77	∞				
5	V	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
	$V = 4B$	2.05	1.89	1.61	1.87	1.33	2.59	1.09	4.70	0.97	7.32	0.91	10.65
	$V = B$	1.30	9.02	1.01	12.15	0.93	14.57	0.89	17.04	0.87	20.20	0.86	25.13
	B	0.85	∞	0.85	∞	0.85	∞	0.85	∞	0.85	∞	0.85	∞

order bias exists in the estimate $\hat{y}$ and consequently the optimal design (which is now that which minimizes variance alone) is of *infinite* size. This case $\sum\lambda_i=0$ is of course very atypical and we are left with the conclusion that in cases likely to be met in practice the optimal orthogonal design of type B should usually be chosen so that $\dot{r}$ is somewhat greater than $\sqrt{k/(k+2)}$ but rather less than unity.

4.4 *Requirement (b); True Model Quadratic.* In accordance with the plan outlined at the beginning of the paper, we will choose from among that class of designs which satisfy requirement (a) that sub-class which best meets requirement (b). It will be recalled that this latter requirement is designed to ensure that the experimental arrangement should allow a check to be made on the representational accuracy of the assumed class of polynomials. The requirement is met by choosing the expected value of the residual sum of squares S_R to be large. Before going ahead with this plan we make a small digression to show that if, following the indications of requirement (a), we choose the first order design to be orthogonal, then requirement (b) *independently* implies that all third order moments should be zero.

4.4a *Independent Vanishing of Third Order Moments.* Writing

$$\mathbf{x}_{1u}' = (1, x_{1u}, x_{2u}, \cdots, x_{ku}) \tag{77}$$

$$\mathbf{x}_{2u}' = (x_{1u}^2, x_{2u}^2, \cdots, x_{ku}^2, x_1x_2, x_1x_3, \cdots, x_{k-1}x_k) \tag{78}$$

we have from (9)

$$E(S_R) - \nu\sigma^2 = \sum_u \beta_2'(A'\mathbf{x}_{1u} - \mathbf{x}_{2u})(\mathbf{x}_{1u}'A - \mathbf{x}_{2u}')\beta_2$$

$$= \beta_2'A'\sum_u \mathbf{x}_{1u}\mathbf{x}_{1u}'A\beta_2 - 2\beta_2'\sum_u \mathbf{x}_{2u}\mathbf{x}_{1u}'A\beta_2 + \beta_2\sum_u \mathbf{x}_{2u}\mathbf{x}_{2u}'\beta_2 \tag{79}$$

Now if the design is orthogonal $N^{-1}\sum_u \mathbf{x}_{1u}\mathbf{x}_1' = N^{-1}(X_i'X_1)$ is a $(k+1)\times(k+1)$ diagonal matrix with first diagonal element unity and remaining diagonal elements equal to c. The matrix

$$N^{-1}\sum_u \mathbf{x}_{2u}\mathbf{x}_{1u}' = N^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1 = [\boldsymbol{M}_2 \mid \boldsymbol{M}_3]$$

is $p\times(k+1)$ and may be partitioned after the first column into two submatrices M_2 and M_3, with M_2 a column vector containing p elements the first k of which are equal to c and the remainder to zero, and M_3 having each element a third order moment. Finally the elements of the matrix

$$N^{-1}\sum_u \mathbf{x}_{2u}\mathbf{x}_{2u}' = N^{-1}\boldsymbol{X}_2'\boldsymbol{X}_2 = \boldsymbol{M}_4$$

are all fourth order moments. Substitution in equation (79) and division by N defines a quantity

$$F = N^{-1}\{E(S_R) - \nu\sigma^2\} = -\beta_2'\boldsymbol{M}_2\boldsymbol{M}_2'\beta_2 - c^{-1}\beta_2'\boldsymbol{M}_3\boldsymbol{M}_3'\beta_2 + \beta_2'\boldsymbol{M}^4\beta_2 \tag{80}$$

$$= -c^2\left\{\sum_{g=1}^{k}\beta_{gg}\right\}^2 - c^{-1}\sum_{g=1}^{k}\left\{\sum_{h=1}^{k}\sum_{i=h}^{k}\beta_{hi}[ghi]\right\}^2$$

$$+ \sum_{g=1}^{k}\sum_{h=g}^{k}\sum_{i=1}^{k}\sum_{j=i}^{k}\beta_{gh}\beta_{ij}[ghij]. \tag{81}$$

Third order moments are involved only in the second term whence it is seen that for any first order orthogonal design, F is maximized with respect to third order moments for all values of β_{hi} when all these third order moments are zero.

4.4b *Size of Fourth Moments.* We now return to the main theme. From the infinite class of first order orthogonal designs which have all third order moments zero and for which $\dot{r}$ is fixed at some specific value chosen on criterion (a), we have to choose a sub-class which makes large

$$F = -c^2\left\{\sum_{i=1}^{k}\beta_{ii}\right\}^2 + \sum_{g=1}^{k}\sum_{h=g}^{k}\sum_{i=1}^{k}\sum_{j=i}^{k}\beta_{gh}\beta_{ij}[ghij] \tag{82}$$

We notice that the only quantities which determine the design and which are at our choice are the fourth moments $[ghij]$. In general the optimal values of these moments depend upon the values of the elements of β_2 and usually these elements are unknown. We can however obtain a design which is "good on the average" by arguing as before that, in a case where all orientations of the response surface with respect to the chosen coordinate system are equally likely, it is reasonable to average F over all rotations. Using equations (2.5), (2.6), (2.7), (2.18) and (2.19) from Appendix 2, we obtain

$$\frac{N}{\sigma^2}\bar{F} = \frac{1}{k(k+2)}\theta(\varphi+2)\left\{\sum_{i=1}^{k}[iiii] + \sum_{i\neq j}^{k}\sum^{k}[iijj]\right\} - \theta\varphi c^2 \tag{83}$$

$$= \frac{\theta}{k}\left\{\frac{\varphi+2}{k+2}\frac{\sum_u r_u^4}{N} - \frac{\varphi\dot{r}^4}{k}\right\} \tag{84}$$

where

$$r_u = \left\{\sum_{i=1}^{k} x_{iu}^2\right\}^{1/2}$$

is the distance of the u-th experimental point from the center of the region R and

$$\left\{N^{-1}\sum_{u=1}^{N} r_u^2\right\}^{1/2} = \sqrt{kc} = \dot{r}$$

is the root mean square distance.

The above expression can be written

$$\frac{N}{\sigma^2}\bar{F} = \dot{r}^4\frac{\theta}{k}\left\{\frac{\varphi+2}{k+2}G(r) - \frac{\varphi}{k}\right\} \tag{85}$$

where $G(r)$ is a fourth moment coefficient

$$\frac{\sum r_u^4}{N} \Big/ \dot{r}^4$$

of the distances r_u.

Also we can write

$$G = 1 + \{C_{r^2}\}^2 \tag{86}$$

where C_{r^2} is the coefficient of variation of the r_u^2. C_{r^2} takes its minimal value of zero and $G(r)$ its minimal value of unity when all the experimental points are equidistant from the center of R. We can see therefore from (85) that the detectability of quadratic discrepancy from the assumed planar model is increased as the distances of the experimental points from the center of the region R have greater and greater "percentage variation."

With this in mind we can rewrite the expression in the form

$$\frac{N}{\sigma^2}\tilde{F} = \dot{r}^4 \frac{\theta}{k}\left\{\frac{2(k-\varphi)}{k(k+2)} + \frac{\varphi+2}{k+2}\{C_{r^2}\}^2\right\} \tag{87}$$

Now, since $0 \leq \varphi \leq k$, both terms within the bracket are necessarily non-negative. For designs in which the points are all equally spaced from the center the second term vanishes. For given k and with $\dot{r}$ fixed by criterion (a) the size of $\sigma^{-2}\tilde{F}$ then depends only on the value of $\theta(k-\varphi)$ and in particular takes the value zero when $\varphi = k$. To see the reason for this we note that when $\varphi = k$ the contours of the response surface are spheres and the response function is represented by a second degree expression containing quadratic terms only and no interaction terms. Now if all points of the design are equally spaced from the center it is easily shown that F in equation (82) and hence the expected value of the residual sum of squares contains only interaction terms. Second order effects of a purely quadratic nature are therefore undetectable.

As φ becomes smaller we encounter surfaces which cannot (except in particular orientations) be represented by quadratic terms alone and consequently the power to detect interaction terms becomes on the average more and more valuable to us. When the points are not all equally spread from the center, $\tilde{F}$ is non-zero even if $\varphi = k$, because unequal values for the r_u render pure quadratic effects detectable.

5. SOME DESIGNS WHICH MEET THE REQUIREMENTS

We now consider one particular way of generating first order designs which satisfy requirements (a) and (b) and which, in accordance with requirement (c), do not contain an excessively large number of points.

5.1 *Requirement (a).* Designs satisfying requirement (a) must be first order orthogonal with third order moments zero. One simply class of designs of this sort can be generated as follows. Suppose k factors are to be investigated in N trials where N is even and $k \leq \frac{1}{2}N$. We first write down any $(\frac{1}{2}N \times k)$ matrix $\boldsymbol{Z}_1 = \{x_{ij}\}$ having orthogonal columns so that

$$\sum_{u=1}^{N/2} x_{iu}x_{ju} = 0, \quad (i \neq j = 1, 2, \cdots, k) \tag{88}$$

$$\sum_{u=1}^{N/2} x_{iu}^2 = \tfrac{1}{2}Nc,$$

the value of c being chosen on criterion (a). We now take for our design the $N \times k$ matrix

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \hline -\boldsymbol{Z}_1 \end{bmatrix} \tag{89}$$

This resulting arrangement satisfies the requirements for a first order orthogonal design and (compare [8, Appendix 2]) the third moments of the design are all zero because the $N \times \frac{1}{2}k(k+1)$ matrix $\boldsymbol{Z}_2$ of independent variables $x_1^2, x_2^2, \cdots, x_k^2, x_1x_2, \cdots, x_{k-1}x_k$ is of the form

$$\begin{bmatrix} \boldsymbol{Z}_2 \\ \hline \boldsymbol{Z}_2 \end{bmatrix}$$

and consequently the matrix of third order moments is null since

$$\left[\begin{array}{c|c} \boldsymbol{Z}_1' & -\boldsymbol{Z}_1' \end{array}\right] \cdot \begin{bmatrix} \boldsymbol{Z}_2 \\ \hline \boldsymbol{Z}_2 \end{bmatrix} = \boldsymbol{Z}_1'\boldsymbol{Z}_2 - \boldsymbol{Z}_1'\boldsymbol{Z}_2. \tag{90}$$

Now [2], if the levels of the factors are completely unrestricted and at our choice and $k < \frac{1}{2}N$, the original k vectors in the matrix $\boldsymbol{Z}_1$ may be chosen to be orthogonal to a set of up to $\frac{1}{2}N - k$ further vectors which correspond to extraneous systematic effects which it is desired to eliminate. These further vectors may for example represent block contrasts or they may be orthogonal polynomials corresponding to possible time trends. Finally, "spherical randomisation" [2] may be employed to ensure that normal theory may be validly applied in subsequent statistical analysis. The design $\boldsymbol{D}$ obtained by replicating $\boldsymbol{Z}_1$ with reversed signs will preserve all these properties and will possess the additional one (arising from the property that all third order moments are zero) that no estimate of a first order effect is biased by a term of second order, that is by a quadratic or interaction term.

With k *equal* to $\frac{1}{2}N$ we obtain the designs of this kind which allow the investigation of the maximum number of factors. These particular designs are such that the distances of the experimental points from the origin are all equal, for since $\boldsymbol{Z}_1$ is a square orthogonal matrix

$$\sum_{i=1}^{k} x_{iu}^2 = r_u^2 = kc \quad (u = 1, 2, \cdots, N) \tag{91}$$

Of particular interest are the designs of this kind which employ only two levels. It was shown by Plackett and Burman [16] that a square $m \times m$ orthogonal matrix whose first column consisted of $+1$'s, and whose remaining columns contained an equal number of $+1$'s and -1's, could be formed for m any multiple of 4. Using such a design for the matrix $\boldsymbol{Z}_1$, first order designs of type B suitable for testing up to four factors in eight trials, eight factors in 16 trials, twelve factors in 24 trials, etc., can be generated by replicating the basic design with reversed signs.

When m is a power of two, the original Plackett and Burman design and the design of type B derived by replicating it with reversed signs are each fractionally replicated 2^k factorials. The requirement that all third order moments are zero is equivalent to choosing the fractional replicates so that no three-factor interaction is included in the alias sub-group. That is to say, so that no two factor interaction is confounded with a main effect. Two level fractional factorials of this kind, thus provide a particular class of designs which satisfy requirement (a). These designs were derived some time ago [17, 18, and 8]

and have been extensively used (for example [15]). They can readily be generaated in the manner illustrated in the example which follows.

Suppose we require a design to test eight factors in 16 experiments. We first generate an 8×8 orthogonal matrix whose first column, which we associate with the variable x_1, consists of $+1$'s and whose remaining seven columns, associated with the variables $x_2, x_3, \cdots, x_8$, consist of equal numbers of -1's and $+1$'s. These last seven columns can be obtained by writing out as column vectors the contrasts associated with three factors x_2, x_3, x_4 run in a 2^3 factorial design together with their interactions. These interaction contrasts are then associated with the remaining factors, for example $x_5=x_2x_3$, $x_6=x_2x_4$, $x_7=x_3x_4$, $x_8=x_2x_3x_4$. We thus obtain for the matrix Z_1

$$Z_1 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The first order design of type B is then obtained by replicating this matrix Z_1 with all signs reversed.

5.2 *Requirement* (*b*). We now need to try to satisfy the requirement that departures from the assumed model, which occur because the true function is quadratic rather than linear, should be readily detectable.

The experimental designs we have discussed, in which k factors are tested in $2k$ trials by replicating a square $k\times k$ orthogonal matrix with reversed signs, necessarily have points all equispaced from the origin. It should be noted however that even if the points of the design are all equispaced from the origin the expected value of the residual sum of squares although *minimal* is not necessarily *small* except for special types of response surfaces. The arrangements considered do in fact provide quite sensitive tests for all but the rather exceptional kinds of departure from linearity in which the true response functions are described in the original coordinate system by quadratic constants β_{ii} alone. These particular kinds of departures from linearity only become detectable when C_{r^2}, the coefficient of variation of the squared radial distances of experimental points, is non-zero. If the basic design is to be modified by the addition of extra points, the greatest increase in C_{r^2} can be achieved by adding n_0 extra points at the origin. As an example, consider the $n=16$ point factorial design of type B already discussed in Section 5.1. With n_0 additional points at the center we obtain a design containing $N=n+n_0$ points in all.

Using this design the least squares estimates of the coefficients in the linear model have expected values which, on the assumption that the model is truly quadratic, are of the form

$$E(b_0) = \beta_0 + c\sum_{i=1}^{8}\beta_{ii}$$

$$E(b_i) = \beta_i \qquad (i = 1, 2, \cdots, 8)$$

The expected value of the residual sum of squares which has $7+n_0$ degrees of freedom is

$$E(S_R) = c^2 n_0 \frac{N}{n}(\beta_{11} + \beta_{22} + \cdots + \beta_{88})^2 + c^2\frac{N^2}{n}\{(\beta_{12} + \beta_{35} + \beta_{46} + \beta_{78})^2$$
$$+ (\beta_{13} + \beta_{25} + \beta_{47} + \beta_{68})^2 + (\beta_{14} + \beta_{26} + \beta_{37} + \beta_{58})^2$$
$$+ (\beta_{15} + \beta_{23} + \beta_{67} + \beta_{48})^2 + (\beta_{16} + \beta_{24} + \beta_{38} + \beta_{57})^2$$
$$+ (\beta_{17} + \beta_{28} + \beta_{34} + \beta_{56})^2 + (\beta_{18} + \beta_{27} + \beta_{36} + \beta_{45})^2\} + (7 + n_0)\sigma^2.$$

The residual sum of squares S_R based on $(7+n_0)$ degrees of freedom can be divided into two parts, the first part containing $7+1=8$ degrees of freedom associated exclusively with terms measuring components due to possible lack of fit and the second part S_e containing n_0-1 degrees of freedom associated exclusively with a measure of pure error obtained from the variation of the observations recorded at the center point. The sum of squares for lack of fit having 8 degrees of freedom can be further sub-divided into 8 separate components $S_1, S_2, S_3, \cdots, S_8$ each associated with a single degree of freedom. If we denote the n_0 observations at the center conditions by $y_{01}, \cdots, y_{0u}, \cdots, y_{0n_0}$, their average by $\bar{y}_0$, and the average of the $n=16$ observations in the fractional factorial design by $\bar{y}_d$, we can write a detailed analysis of this residual sum of squares as follows.

The seven sums of squares $S_2, S_3, \cdots, S_8$ correspond to the seven contrasts within each of which four interaction terms are confounded (for example [8, p. 14]). Even if no center points were added, so that S_1 did not appear, a fairly powerful test of the assumption of linearity would usually be possible, provided

(i) an independent estimate of σ^2 was available;
(ii) the contours of the true response surface in the eight-dimensional space of the variables were not too nearly spherical;
(iii) the "canonical axes" of the true system were not oriented too closely parallel to the coordinate axes of the variables;
(iv) the interaction terms were not such that the estimated linear functions of them were all zero.

Conditions (ii) and (iii) are essentially concerned with the possibility that the true surface can be described entirely by the quadratic terms β_{ii}, none of which appear in the expected value of the residual sum of squares when $n_0=0$. With $n_0>0$ an additional term containing

$$\sum_{i=1}^{8}\beta_{ii}$$

appears. As seen from the detailed analysis in Table 649, this is associated with the isolation of the contrast $\bar{y}_d-\bar{y}_0$, the difference between the average of the

TABLE 649. ANALYSIS OF RESIDUAL SUMS OF SQUARES

Sums of Squares	D/F	Expected Values of Mean Squares
$S_1 = n_0\left(\frac{n}{N}\right)(\bar{y}_d - \bar{y}_0)^2$	1	$c^2 n_0 \frac{N}{n}\left\{\sum_{i=1}^{8} \beta_{ii}\right\}^2 + \sigma^2$
$S_2 = \frac{n}{c^2N^2}\left\{\sum_{u=1}^{N} y_u x_{1u} x_{2u}\right\}^2$	1	$c^2 \frac{N^2}{n}(\beta_{12} + \beta_{35} + \beta_{46} + \beta_{78})^2 + \sigma^2$
$S_3 = \frac{n}{c^2N^2}\left\{\sum_{u=1}^{N} y_u x_{1u} x_{3u}\right\}^2$	1	$c^2 \frac{N^2}{n}(\beta_{13} + \beta_{25} + \beta_{47} + \beta_{68})^2 + \sigma^2$
$\vdots$		
$S_8 = \frac{n}{c^2N^2}\left\{\sum_{u=1}^{N} y_u x_{1u} x_{8u}\right\}^2$	1	$c^2 \frac{N^2}{n}(\beta_{18} + \beta_{27} + \beta_{36} + \beta_{45})^2 + \sigma^2$
$S_e = \sum_{u=1}^{n_0} (y_{0u} - \bar{y}_0)^2$	$n_0 - 1$	σ^2

center points and the average of the points in the fractional factorial design. The expected value of this difference is

$$c\frac{N}{n}\sum_{i=1}^{8}\beta_{ii}.$$

As n_0 is increased above the value unity, the experimental arrangement becomes more and more sensitive to discrepancies associated with the overall criterion of curvature

$$\sum_{i=1}^{8}\beta_{ii}.$$

In addition it becomes possible to isolate a sum of squares based on $n_0 - 1$ degrees of freedom which, on the basic assumption of homogeneity of the error variance, measures pure error. This makes possible tests of departures from the linear model based solely on the internal evidence supplied by the design.

6. DISCUSSION

The class of designs to which we have been led, of which the fractional factorial discussed above is a member, seems to be excellently suited to the task at hand, indeed designs of exactly this type have for some time actually been applied in response surface studies. For the case of first order designs studied here the property of orthogonality which has repeatedly arisen is equivalent to that of "rotatability," a concept introduced previously [6]. This agreement between the present theory and what has seemed desirable on an intuitive basis suggests that the present formulation of the problem is reasonably well conceived. By happy circumstance designs of the type discussed not only achieve the first three requirements of Section 1.2 (graduate the function as accurately as possible, allow check of representational accuracy of assumed form of polynomial, do not contain excessively large numbers of points) but also these arrangements ([6] for details) can form a nucleus upon which a satisfactory

design of order 2 can be built in case the assumed degree of polynomial proves inadequate, and they are ideally suited to become part of blocking arrangements in these larger designs.

Much remains to be done and work is in progress on such topics as the effect of cubic bias in first order designs, the extension of the present ideas to second order designs, the effect of changing the criterion to the minimization of maximum mean square error instead of minimization of average mean square error. The modifications necessary when a measure of the absolute value of the response is not important, but only the change in response from one point to another in the space of the variables, are also under consideration.

APPENDIX 1

A General Result Concerning Averaged Squared Bias. Suppose a polynomial of degree d_1 is fitted by the method of least squares over any region of interest R in the space of the variables when the true function is a polynomial of degree d_2. Then B the squared bias averaged over the region R is minimized for all values of the coefficient of neglected terms, by making the moments of the design up to order d_1+d_2 equal to the corresponding moments of a uniform distribution over the region R.

We suppose that the fitted model is

$$\mathcal{Y}(\mathbf{x}) = \mathbf{x}_1'\boldsymbol{b}_1 \tag{1.1}$$

with $\boldsymbol{b}_1$ the vector of least squares estimates, while the true model is

$$\eta(\mathbf{x}) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_2 \tag{1.2}$$

where

$$\mathbf{x}_1' = (1; x_1, \cdots, x_k; x_1^2, \cdots, x_k^2; x_1x_2, \cdots; x_1^3, \cdots) \tag{1.3}$$

$$\boldsymbol{\beta}_1' = (\beta_0; \beta_1, \cdots, \beta_k; \beta_{11}, \cdots, \beta_{kk}; \beta_1\beta_2, \cdots; \beta_{111}, \cdots) \tag{1.4}$$

contain all terms up to order d_1, and $\mathbf{x}_2'$ and $\boldsymbol{\beta}_2'$ are similar vectors containing all terms from order d_1+1 up to order $d_2(d_2>d_1)$. With observations at N points we can define the matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ by

$$\boldsymbol{X}_1 = \begin{bmatrix} \mathbf{x}_{11}' \\ \vdots \\ \mathbf{x}_{1u}' \\ \vdots \\ \mathbf{x}_{1N}' \end{bmatrix} \quad \boldsymbol{X}_2 = \begin{bmatrix} \mathbf{x}_{21}' \\ \vdots \\ \mathbf{x}_{2u}' \\ \vdots \\ \mathbf{x}_{2N}' \end{bmatrix} \tag{1.5}$$

where the column vectors which make up $\boldsymbol{X}_1$ are assumed to be linearly independent.

Now the component of the integral J arising from bias alone is B, where

$$\frac{\sigma^2}{N}B = \Omega\int_R [E\mathcal{Y}(\mathbf{x}) - \eta(\mathbf{x})]^2 dx = \Omega\int_R \boldsymbol{\beta}_2'(\mathbf{x}_1'A - \mathbf{x}_2')'(\mathbf{x}_1'A - \mathbf{x}_2')\boldsymbol{\beta}_2 dx \tag{1.6}$$

$$= \boldsymbol{\beta}_2'\Delta\boldsymbol{\beta}_2 \tag{1.7}$$

where

$$A = (X_1'X_1)^{-1}X_1'X_2 \tag{1.8}$$

and

$$\Delta = A'\mathfrak{y}_1 A - \mathfrak{y}_2'A - A'\mathfrak{y}_2 + \mathfrak{y}_3 \tag{1.9}$$

with

$$\mathfrak{y}_1 = \Omega \int_R x_1 x_1' dx \tag{1.10}$$

$$\mathfrak{y}_2 = \Omega \int_R x_1 x_2' dx \tag{1.11}$$

$$\mathfrak{y}_3 = \Omega \int_R x_2 x_2' dx. \tag{1.12}$$

Now write $M_1 = X_1'X_1$ and $M_2 = X_1'X_2$ so that

$$A = M_1^{-1}M_2. \tag{1.13}$$

Then

$$\Delta = (\mathfrak{y}_3 - \mathfrak{y}_2'\mathfrak{y}_1^{-1}\mathfrak{y}_2) + M_2'M_1^{-1}\mathfrak{y}_1 M_1^{-1}M_2 - \mathfrak{y}_2'M_1^{-1}M_2 - M_2'M_1^{-1}\mathfrak{y}_2 + \mathfrak{y}_2'\mathfrak{y}_1^{-1}\mathfrak{y}_2 \tag{1.14}$$

$$= (\mathfrak{y}_3 - \mathfrak{y}_2'\mathfrak{y}_1^{-1}\mathfrak{y}_2) + (M_1^{-1}M_2 - \mathfrak{y}_1^{-1}\mathfrak{y}_2)'\mathfrak{y}_1(M_1^{-1}M_2 - \mathfrak{y}_1^{-1}\mathfrak{y}_2) \tag{1.15}$$

$$= \Delta_1 + \Delta_2 \quad \text{(say)} \tag{1.16}$$

Now it is shown below that Δ_1 is positive semi-definite as is $\mathfrak{y}_1$. It follows that whatever the value of β_2, $\beta_2'\Delta\beta_2$ is minimized when $M_1^{-1}M_2 = \mathfrak{y}_1^{-1}\mathfrak{y}_2$ and in particular when $M_1 = \mathfrak{y}_1$ and $M_2 = \mathfrak{y}_2$. Now the elements of M_1 and M_2 include all the moments of the design up to order d_1+d_2 while the elements of $\mathfrak{y}_1$ and $\mathfrak{y}_2$ include, in corresponding positions, all the moments of the region R up to order d_1+d_2. The stated result follows.

To show that $\mathfrak{y}_1$ and Δ_1 are positive semi-definite

$$0 \leqq \int_R [(x_1' \mid x_2')\beta]^2 dx \tag{1.17}$$

$$= \beta' \left\{ \int_R \begin{pmatrix} x_1 \\ \hline x_2 \end{pmatrix} (x_1' \mid x_2') dx \right\} \beta \tag{1.18}$$

$$= \Omega^{-1}\beta' \begin{pmatrix} \mathfrak{y}_1 & \mathfrak{y}_2 \\ \mathfrak{y}_2' & \mathfrak{y}_3 \end{pmatrix} \beta; \tag{1.19}$$

whence $\mathfrak{y}_1$ and

$$\begin{pmatrix} \mathfrak{y}_1 & \mathfrak{y}_2 \\ \mathfrak{y}_2' & \mathfrak{y}_3 \end{pmatrix}$$

are both positive semi-definite; but

$$\begin{pmatrix} \mathfrak{y}_1 & 0 \\ 0 & \Delta_1 \end{pmatrix} = \begin{pmatrix} \mathfrak{y}_1 & 0 \\ 0 & \mathfrak{y}_3 - \mathfrak{y}_2'\mathfrak{y}_1^{-1}\mathfrak{y}_2 \end{pmatrix} = T' \begin{pmatrix} \mathfrak{y}_1 & \mathfrak{y}_2 \\ \mathfrak{y}_2' & \mathfrak{y}_3 \end{pmatrix} T \tag{1.20}$$

where

$$T = \begin{pmatrix} I_1 & -\mathfrak{y}_1^{-1}\mathfrak{y}_2 \\ 0 & I_3 \end{pmatrix} \tag{1.21}$$

whence Δ_1 is positive semi-definite also.

APPENDIX 2

Rotational Average of the Constants in the Second Order Response Surface. To appreciate the character of the rotational average which is being taken, consider first the special case where $k=2$. We have for the quadratic part of the model

$$Q = \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1x_2. \tag{2.1}$$

Now if we make an orthogonal rotation to new coordinates X_1 and X_2

$$\begin{aligned} x_1 &= X_1 \sin\theta + X_2 \cos\theta \\ x_2 &= X_1 \cos\theta - X_2 \sin\theta \end{aligned} \tag{2.2}$$

then

$$\begin{aligned} Q = {} & \{\beta_{11}\sin^2\theta + \beta_{22}\cos^2\theta + \beta_{12}\sin\theta\cos\theta\}X_1^2 \\ & + \{\beta_{11}\cos^2\theta + \beta_{22}\sin^2\theta - \beta_{12}\sin\theta\cos\theta\}X_2^2 \\ & + \{2(\beta_{11}-\beta_{22})\sin\theta\cos\theta + \beta_{12}(\cos^2\theta - \sin^2\theta)\}X_1X_2 \end{aligned} \tag{2.3}$$

The quantities we are interested in are the values of the coefficients of X_1^2, X_2^2 and X_1X_2 averaged over values of θ from 0 to 2π.

When k is greater than 2 we can deduce the form of the averages as follows. With $\beta_2' = (\beta_{11}, \beta_{22}, \cdots, \beta_{kk}, \beta_{12}, \cdots, \beta_{k,k-1})$ the elements we wish to average are those of the matrix $\beta_2\beta_2'$. Now if the rotational averages of the elements are substituted in this matrix to give $\widetilde{\beta_2\beta_2'}$ then $x_2'\widetilde{\beta_2\beta_2'}x_2$ must be a function of

$$\sum_{i=1}^{k} x_i^2$$

only.

This implies that

$$\widetilde{\alpha_2\alpha_2'} = N\sigma^{-2}\widetilde{\beta_2\beta_2'} = \left[\begin{array}{cccc|c} a & b & b & \cdots & b \\ b & a & & & \\ \vdots & & & & \\ b & & & a & \\ \hline & & 0 & & \begin{matrix} 2(a-b) & & \\ & \ddots & \\ & & 2(a-b) \end{matrix} \end{array}\right] \tag{2.4}$$

That is

$$N\sigma^{-2}\widetilde{\beta_{ii}^2} = \widetilde{\alpha_{ii}^2} = a \quad (i = 1, 2, \cdots, k) \tag{2.5}$$

$$N\sigma^{-2}\widetilde{\beta_{ii}\beta_{jj}} = \widetilde{\alpha_{ii}\alpha_{jj}} = b \quad (i \neq j = 1, 2, \cdots, k) \tag{2.6}$$

$$N\sigma^{-2}\widetilde{\beta_{ij}^2} = \widetilde{\alpha_{ij}^2} = 2(a-b) \quad (i \neq j = 1, 2, \cdots, k) \tag{2.7}$$

with the remaining rotational averages zero. The quantities a and b are those substituted in equation (61).

To determine the nature of these quantities consider again the symmetric matrix referred to in section 4.3 with diagonal elements β_{ii}, off-diagonal elements $\frac{1}{2}\beta_{ij}$ and latent roots $\lambda_1, \lambda_2, \cdots, \lambda_k$. We can express the a's and b's of equation (2.4) in terms of the λ's as follows.

In all orthogonal rotations the trace of any power of the matrix of section 4.3 remains constant. It follows in particular that

$$\sum_i \lambda_i = \sum_i \beta_{ii}$$

that is

$$\Big(\sum_i \lambda_i\Big)^2 = \sum_i \beta_{ii}^2 + \sum_{i \neq j} \sum \beta_{ii}\beta_{jj}. \tag{2.8}$$

Also

$$\sum_i \lambda_i^2 = \sum_i \beta_{ii}^2 + \frac{1}{4} \sum_{i \neq j} \sum \beta_{ij}^2 \tag{2.9}$$

Now since (2.8) and (2.9) are true in every rotation, these equations are also true for the rotational averages. That is

$$(\sum \lambda_i)^2 = k\widetilde{\beta_{ii}^2} + k(k-1)\widetilde{\beta_{ii}\beta_{jj}} \tag{2.10}$$

$$\sum \lambda_i^2 = k\widetilde{\beta_{ii}^2} + \tfrac{1}{4}k(k-1)\widetilde{\beta_{ij}^2} \tag{2.11}$$

and, using (2.5), (2.6) and (2.7)

$$N\sigma^{-2}\Big(\sum_i \lambda_i\Big)^2 = ka + k(k-1)b \tag{2.12}$$

$$N\sigma^{-2} \sum_i \lambda_i^2 = ka + \tfrac{1}{2}k(k-1)(a-b) \tag{2.13}$$

$$= \tfrac{1}{2}k\{(k+1)a - (k-1)b\}. \tag{2.14}$$

Whence

$$\frac{\sigma^2}{N} a = \{(\sum \lambda_i)^2 + 2\sum \lambda_i^2\}/k(k+2) \tag{2.15}$$

$$\frac{\sigma^2}{N} b = \{(k+1)(\sum \lambda_i)^2 - 2\sum \lambda_i^2\}/(k-1)k(k+2). \tag{2.16}$$

Whence, with

$$\theta = \frac{N}{\sigma^2} \sum_i \lambda_i^2 \quad \text{and} \quad \varphi = \Big(\sum_i \lambda_i\Big)^2 / \sum_i \lambda_i^2, \tag{2.17}$$

$$a = \frac{\theta(\varphi + 2)}{k(k+2)} \tag{2.18}$$

$$b = \frac{\theta[(k+1)\varphi - 2]}{(k-1)k(k+2)}. \tag{2.19}$$

REFERENCES

[1] Box, G. E. P., "Integration of Techniques in Process Development," *Transactions of the 11th Annual Convention of the American Society for Quality Control*, 1957.
[2] Box, G. E. P., "Multifactor Designs of First Order," *Biometrika*, 39 (1952), 49–57.
[3] Box, G. E. P., "Use of Statistical Methods in the Elucidation of Basic Mechanisms," *Proceedings of the 30th Session of the International Statistical Institute*, Stockholm, 1957.
[4] Box, G. E. P. and Coutie, G. A., "Application of Digital Computers in the Exploration of Functional Relationships," Proceedings of the Institution of Electrical Engineers, 103, Part B, Supplement No. 1 (1956), 100–7.
[5] Box, G. E. P. and Hunter, J. S., "A Confidence Region for the Solution of a Set of Simultaneous Equations with an Application to Experimental Design," *Biometrika*, 41 (1954), 190–9.
[6] Box, G. E. P. and Hunter, J. S., "Multi-factor Experimental Designs for Exploring Response Surfaces," *Annals of Mathematical Statistics*, 28 (1957), 195–241.
[7] Box, G. E. P. and Lucas, H. L., "Design of Experiments in Non-Linear Situations," Statistical Techniques Research Group (Princeton University), Technical Report No. 15, submitted for publication in *Biometrika*.
[8] Box, G. E. P. and Wilson, K. B., "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society* (Series B), 13 (1951), 1–45.
[9] Chernoff, H., "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24 (1953), 586–602.
[10] David, H. A. and Arens, Beverly E., "Optimal Spacing in Regression Analysis," Technical Report No. 38, Department of Statistics and Statistical Laboratory, Virginia Polytechnic Institute, 1958.
[11] Elfving, G., "Optimum Allocation in Linear Regression Theory," *Annals of Mathematical Statistics*, 23 (1952), 255–62.
[12] Folks, J. L., "Comparison of Designs for Exploration of Response Relationships," paper read at Chicago at the 18th Annual Meeting, American Statistical Association, December, 1958.
[13] Hotelling, H., "Some Improvements in Weighting and Other Experimental Techniques," *Annals of Mathematical Statistics*, 15 (1943), 297–306.
[14] Hotelling, H., "The Experimental Determination of the Maximum of a Function," *Annals of Mathematical Statistics*, 12 (1941), 20–45.
[15] Hromi, John D., "Application of Fractional Replications in the Food Industry," presented at the Biometric Society Joint Meeting with the American Statistical Association, Atlantic City, 1957.
[16] Plackett, R. L. and Burman, J. P., "The Design of Multifactorial Experiments," *Biometrika*, 33 (1946), 305–25.
[17] Rao, C. R., "Factorial Experiments Derivable from Factorial Arrangements of Arrays," *Journal of the Royal Statistical Society*, Supplement, 9 (1947), 128–30.
[18] Rao, C. R., "On Hypercubes of strength *d:* a system of confounding any factorial experiments," *Bulletin of the Calcutta Mathematical Society*, 38 (1946), 67–78.
[19] Tocher, K. D., "A Note on the Design Problem," *Biometrika*, 39 (1952), 109–17.