FILE

Department of Statistics

UNIVERSITY OF WISCONSIN

Madison, Wisconsin

Technical Report No.10

July, 1962

THE CHOICE OF A SECOND ORDER ROTATABLE DESIGN

George E. P. Box and Norman R. Draper

This research was supported primarily by the United States Navy through the Office of Naval Research, under Contract Nonr-1202(17), Project NR 042 222. Partial support to the second named author was provided by the United States Army under Contract DA-11-022-ORD-2059 at the Mathematics Research Center, U. S. Army, Madison, Wisconsin.

1. INTRODUCTION

1.1 General Remarks

It frequently happens that an experimenter is interested in exploring a functional relationship

$$\eta = \eta(\xi_1, \xi_2, \ldots, \xi_k) .$$

Sometimes, the actual functional form is known. Quite frequently it is not; in this case, useful information about the nature of the actual relationship in some particular region R of the $\underline{\xi}$ space can often be obtained by approximating the relationship by a graduating function $g(\underline{\xi}, \underline{\beta})$ where $\underline{\beta}$ is a vector of adjustable constants.

The graduating functions which have usually been employed have been polynomials in the variables $\underline{\xi}$, though other types of functions might be of value on occasion. The problem of experimental design which arises in the fitting of a graduating function has been discussed in reference [1] and will be briefly restated here.

We desire to choose a design matrix $\underline{D}$ of N rows and k columns which will specify the levels of the k variables to be run in the N experiments. Denote the u-th row of this matrix by $\underline{\xi}_u'$. This vector has as elements the levels $(\xi_{1u}, \xi_{2u}, \ldots, \xi_{ku})$ of the k factors to be employed in the u-th experiment, u=1, 2, . . . , N.

Our primary objective will be to choose these levels so that when the graduating function $g(\underline{\xi})$ is fitted by least squares, it will closely represent the true function $\eta(\underline{\xi})$, within the region of interest R.

Subject to the satisfaction of our primary objective, we shall also require that the factor levels be such that there is a high chance that the inadequacy of the graduating function $g(\underline{\xi}, \underline{\beta})$ to represent $\eta(\underline{\xi})$ will be detected.

As will be seen, a subclass of all possible designs can be selected which will satisfy our primary requirement. We can then make use of our secondary requirement to make a selection of a particular design from this subclass.

We now define how we shall interpret "region of interest," "closely represent" and "detection of inadequacy of model."

1.2 Interpretation of "region of interest"

Let us call the region in the $\underline{\xi}$ space in which experiments can actually be performed, the operability region O. In practice, this region is usually large and bounded although its limits are often known only vaguely. For example in chemical experiments, there will often be conditions of temperature or pressure which will be too severe to use safely on the apparatus. In biological work, certain combinations of drugs for therapeutic use will produce death. For some applications, the experimenter may wish to explore the whole region O, but this is comparatively rare. Usually a particular group of experiments is used to explore a rather limited region of interest R entirely contained within the operability region O. Frequently, experiments are conducted sequentially and a group of experiments designed for the exploration of one current region of interest may lead to a further set exploring a different region. Often, an alternative statement of the problem would be that it is desired to explore the nature of a functional relationship "in the neighborhood" of a point P. The latter statement is perhaps closer to the real desires of some experimenters with its implication that the situation is one of a falling off of interest at points more and more distant from P rather than of equal interest at all points within R and no interest outside R and within O.

As we shall indicate briefly below, by dealing with various types of weight functions in specified regions (for example if R is a k-dimensional sphere centered at the origin we can use weights which are functions of distance of a point in space from the origin), these various possible desires can be combined into a unified treatment. For the immediate purposes of this paper, however, we shall soon revert to consider the "interest within R, no interest outside R" formulation; while this may not suit all tastes, a great many experimental investigations are undertaken with this thought in mind and the formulation is thus not unrealistic.

1.3 Interpretation of "closeness"

Let $\hat{y}(\underline{\xi})$ denote the response estimated by the graduating function at the point $\underline{\xi}$. Then we desire to choose $\underline{D}$ so that the difference $\hat{y}(\underline{\xi})-\eta(\underline{\xi})$ will be small over the region of interest R. The measure of closeness which we shall use at a particular point $\underline{\xi}$ is:

$$E[\hat{y}(\underline{\xi}) - \eta(\underline{\xi})]^2$$

Over the whole region we may use the average

$$\Omega \int_R E[y(\underline{\xi}) - \eta(\underline{\xi})]^2 \, d\underline{\xi} \qquad (1.3.1)$$

where

$$\Omega^{-1} = \int_R d\underline{\xi}$$

In certain circumstances we might wish more weight given to errors at one value of $\underline{\xi}$ than at another. We may therefore generalize the concept above by introducing a weight function $W(\underline{\xi})$ such that

$$\int_O W(\underline{\xi}) \, d\underline{\xi} = 1$$

Our measure will then take the form

$$\int_O W(\underline{\xi}) \, E[\hat{y}(\underline{\xi}) - \eta(\underline{\xi})]^2 \, d\underline{\xi}$$

The previous formulation is a special case of this as is easily seen by setting

$$W(\underline{\xi}) = \begin{cases} \Omega & \text{in } R \\ 0 & \text{elsewhere} \end{cases}$$

It is desirable that we should be able to compare designs which do not contain the same number of points and that our criterion of closeness should be independent of the variance σ^2 of the observations, which we assume to be constant. Thus we shall choose as our measure of closeness

$$J = \int_O w(\underline{\xi}) \, E[\hat{y}(\underline{\xi}) - \eta(\underline{\xi})]^2 \, d\underline{\xi} \qquad (1.3.2)$$

where

$$w(\underline{\xi}) = NW(\underline{\xi})/\sigma^2$$

Writing

$$\hat{y}(\underline{\xi}) - \eta(\underline{\xi}) = \{\hat{y}(\underline{\xi}) - E\hat{y}(\underline{\xi})\} + \{E\hat{y}(\underline{\xi}) - \eta(\underline{\xi})\}$$

we can split J into two parts

$$J = V + B$$

where V is the average weighted variance,

$$V = \int_O w(\underline{\xi})[\hat{y}(\underline{\xi}) - E\hat{y}(\underline{\xi})]^2 \, d\underline{\xi} \qquad (1.3.3)$$

and B is the average squared bias,

$$B = \int_O w(\underline{\xi}) \, [E\hat{y}(\underline{\xi}) - \eta(\underline{\xi})]^2 \, d\underline{\xi} \qquad (1.3.4)$$

In what follows we shall suppose that the graduating function is a polynomial of degree d_1 in $\underline{\xi}$,

$$g(\underline{\xi}) = \underline{\xi}_1' \, \underline{\beta}_1$$

where the vector $\underline{\xi}_1$ contains p_1 elements, all of which are powers and products of ξ_i, $i=1, 2, \ldots, k$, of order d_1 or less.

The true functional form over the whole operability region O is assumed to be a polynomial of degree d_2 in $\underline{\xi}$.

$$\eta(\underline{\xi}) = \underline{\xi}_1{}' \underline{\beta}_1 + \underline{\xi}_2{}' \underline{\beta}_2$$

where $\underline{\xi}_2'$ contains p_2 elements, all of which are powers and products of $\xi_1, \xi_2, \ldots, \xi_k$ of order d_2 or less but greater than d_1.

Corresponding to any design matrix $\underline{D}$, there will exist an N x p_1 matrix X_1 which has $\underline{\xi}_{1u}'$ as its u-th row, whose elements are powers and products of order d_1 or less of the elements of the vector $\underline{\xi}_u$.

There will also be a matrix $\underline{X}_2$ with u-th row $\underline{\xi}_{2u}'$ whose elements are the powers and products of orders $(d_1+1), \ldots, d_2$ of the elements of the vector $\underline{\xi}_u$.

Let

$$\underline{M}_{11} = N^{-1}\underline{X}_1{}' \underline{X}_1$$

$$\underline{M}_{12} = N^{-1}\underline{X}_1{}' \underline{X}_2$$

$$\underline{M}_{22} = N^{-1}\underline{X}_2' \underline{X}_2 \qquad (1.3.5)$$

It will be seen that the elements of these matrices are of the form

$$N^{-1}\sum_{u=1}^{N} \xi_{1u}^{\alpha_1} \xi_{2u}^{\alpha_2} \ldots \xi_{ku}^{\alpha_k}$$

They will thus be referred to as the moments of the design points and will be said to be of order α if

$$\alpha = \alpha_1+\alpha_2+\ldots+\alpha_k .$$

Then $\underline{M}_{11}, \underline{M}_{12}, \underline{M}_{22}$ are matrices of moments of the experimental design.

Also write

$$\underline{\mu}_{11} = \int w(\underline{\xi})\underline{\xi}_1 \underline{\xi}_1{}' d\underline{\xi}$$

$$\underline{\mu}_{12} = \int w(\underline{\xi}) \underline{\xi}_1 \underline{\xi}_2{}' d\underline{\xi}$$

$$\underline{\mu}_{22} = \int w(\underline{\xi}) \underline{\xi}_2 \underline{\xi}_2{}' d\underline{\xi} \qquad (1.3.6)$$

where all integrals are taken over the region O.

The elements of these matrices are of the form

$$\int_O w(\underline{\xi})\, \xi_1^{\alpha_1} \xi_2^{\alpha_2} \ldots \xi_k^{\alpha_k}\, d\underline{\xi}$$

and this is a moment of the weight function of order α where $\alpha = \alpha_1 + \alpha_2 + \ldots + \alpha_k$. Now, proceeding exactly as in a previous paper (1959) by the same authors, we obtain

$$J = \text{Trace}\ [\underline{\mu}_{11} \underline{M}_{11}^{-1}]$$
$$+ \underline{\beta}_2' [(\underline{\mu}_{22} - \underline{\mu}_{12} \underline{\mu}_{11}^{-1} \underline{\mu}_{12}) + (\underline{M}_{11}^{-1} \underline{M}_{12} - \underline{\mu}_{11}^{-1} \underline{\mu}_{12})' \underline{\mu}_{11} (\underline{M}_{11}^{-1} \underline{M}_{12} - \underline{\mu}_{11}^{-1} \underline{\mu}_{12})]\ \underline{\beta}_2$$
$$= V + B \quad , \quad \text{say,} \qquad (1.3.7)$$

and, as our first objective, we shall choose the design matrix $\underline{D}$ in such a way that this quantity is a minimum. Our formulation will thus ensure that the graduating function will closely represent the true function $\eta(\underline{\xi})$, in the way we have described, after a suitable weight function has been chosen.

We see from the above expression that if we write $J = V + B$, V does not contain $\underline{\beta}_2$ at all and depends only on $\underline{D}$, while B depends on both $\underline{D}$ and $\underline{\beta}_2$. Mathematically

$$J(\underline{D}, \underline{\beta}_2) = V(\underline{D}) + B(\underline{D}, \underline{\beta}_2) \ .$$

Thus, minimization of J depends on what value we assign to $\underline{\beta}_2$. We shall return to this point later.

1.4 Interpretation of "detection of inadequacy of model"

We shall 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

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

where y_u are the actual observations, 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

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

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, this will be equivalent to making the expectation of S_R large.

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

1.5 Choice of R as a spherical region

Up to this point we have said nothing about the shape of the region of interest R. Previous remarks would apply no matter what the shape of the region. We shall particularize at this point and choose R to be a spherical region, a choice which appears (to the authors at least) intuitively reasonable in many situations for the following reasons.

R can be of any shape one can imagine and, given any *particular* region, the theory can be applied in a way similar to the way shown here. However, it is impossible to forsee all conceivable choices of R and in order to develop results we must make a reasonable assumption. Two reasonable assumptions are

(i) R is spherical or ellipsoidal, that is, some deformation of a sphere attained because of change of scale, so that mathematically only a sphere need be considered (this case we shall treat).

(ii) R is cuboidal, or is some deformation of a k-dimensional cube attained because of change of scale, so that mathematically only a cube need be considered. (We shall not treat this case but will remark how it would affect succeeding paragraphs. Instead of being later led to rotatable designs where all odd moments are zero and even moments bear certain relationships to one another as given by Box and Hunter (1958), we should be led to "rectangular" designs, in which all odd moments are zero and even moments bear certain (other) relationships to one another. As a possible example in certain circumstances: Instead of obtaining, as for case (i), a second order rotatable design with ratio (pure fourth moment)/(mixed fourth moment) = 3, we should obtain a symmetrical design with all odd moments zero but with the ratio (pure fourth moment)/(mixed fourth moment) = 1.8.

From these two reasonable formulations we select the first for further development. It is, in the authors' opinion, probably the one more frequently in an experimenter's mind.

1.6 Reasons for the consideration of rotatable designs only.

We now intend to consider only rotatable designs and this choice is closely related to our choice of a spherical region R as we shall now explain.

In their previous paper (1959) the authors showed that no matter what the shape of R, a sufficient condition for the bias B alone to be minimized is that the moments of the design should be equal to the moments of the region R up to and including order (d_1+d_2). It is clear from an inspection of equation (1.3.7) above that a <u>necessary</u> <u>and</u> <u>sufficient</u> condition for the minimization of B alone is simply

$$\underline{M}_{11}^{-1}\underline{M}_{12} = \underline{\mu}_{11}^{-1}\underline{\mu}_{12}$$

since the first term of B is always positive as was previously shown (1959). (This result is extremely interesting, incidentally, when interpreted in a numerical analysis situation. The details will be found in Appendix 1.) This implies of course that a sufficient condition is $\{\underline{M}_{11} = \underline{\mu}_{11}\ ,\ \underline{M}_{12} = \underline{\mu}_{12}\}$ which is just a statement that moments of the design equal moments of the region R up to and including order (d_1+d_2).

If the region R is spherical it follows that designs which will minimize bias B <u>only</u> are <u>rotatable</u> designs of certain orders which depend on d_1 and d_2. If $d_1+d_2 = 2m$, say, then the appropriate design is an m-th order rotatable design. If $d_1+d_2 = 2m+1$, then the appropriate design is m-th order rotatable with moments of order $(2m+1)$ all zero.

A spherical weight function has no effect whatsoever on this conclusion; since rotatability is entirely dependent on ratios between moments of the same order. These ratios, which are attained for spheres, must therefore be attained for shells, hence attained for any collection of shells and it follows that a spherical weight function, which merely attaches weights to various spherical shells, cannot affect these ratios.

The the adoption of a spherical region R leads to the conclusion that bias B alone is minimized by a rotatable design of appropriate order, no matter what (spherical) weight function is considered.

This persuades us to consider only rotatable designs when both V and B enter into consideration. Of course we have still to determine the values of the design parameters to be used for any given situation as well as the particular rotatable design.

1.7 <u>Choice of weight function</u>

We shall choose, in what follows the weight function originally introduced, namely

$$W(\underline{\xi}) = \begin{cases} \Omega \text{ in } R \\ 0 \text{ elsewhere .} \end{cases}$$

It <u>might</u> be more appropriate in some applications to choose a weight function which decreased as we moved away from the center of the region. In that case, questions which would naturally arise would be "How quickly should $W(\underline{\xi})$ fall off?" "At what point should $W(\underline{\xi})$ be made zero?" "Should the rate of fall-off vary for successive 'zones' as we move from the center of R?"

It is clear that the weight function could be chosen in numerous ways and for any particular weight function the problem could be treated as it will be below. Clever choice of the weight function might also contribute to the ease with which J can be minimized but we shall not discuss this point further here.

Choice of the weight function as given above corresponds to the choice made in the previous paper on this topic. Thus this paper will give results which are a logical extension of those previously found.

1.8 Recapitulation of previous paper

In previous work (1959) it was assumed that $\eta(\underline{\xi})$ was a quadratic polynomial in k factor variables $\xi_1, \xi_2, \ldots, \xi_k$ and that the graduating function $g(\underline{\xi})$ was a linear function of these same variables. It emerged, somewhat surprisingly that, in typical experimental situations, choice of the design depended far more on the effect of bias error than on variance error. Moreover, designs suitable when both variance and bias contributions contributed about equally to the total error were close to designs suitable for the "all bias, no variance"situation and completely different from those suitable for the "all variance" situation on which most previous conclusions have been based.

2. THE PROBLEM AND ITS SOLUTION

2.1 Assumptions in the present paper

We shall assume that $\eta(\underline{\xi})$ is a cubic polynomial and $g(\underline{\xi})$ is a quadratic polynomial in $\xi_1, \xi_2, \ldots, \xi_k$. In other words, $d_1 = 2$, $d_2 = 3$. Hence, $d_1 + d_2 = 5$ and we shall consider designs which are second order rotatable with fifth moments zero, for the reasons which were described earlier. We shall also assume that the variables $\underline{\xi}$ have been suitably transformed to variables $\underline{x}$ in such a way that the center of the design is at the origin $(0, 0, \ldots, 0)$ and the scale is such that the region R is the k-dimensional unit sphere. This is achieved by a transformation of the type $x_i = \{\xi_i - \xi_i(0)\}/s_i$ where $\xi_i(0)$ and s_i are suitably chosen.

The graduating function $g(\underline{x})$ is

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + b_{11}x_1^2 + \ldots + b_{kk}x_k^2 + b_{12}x_1x_2 + \ldots + b_{k-1,k}x_{k-1}x_k .$$

or, in matrix notation

$$\underline{\hat{y}} = \underline{x}_1'\underline{b}_1$$

where

$$\underline{b}_1' = (b_0\,;\, b_1 \ldots b_k;\, b_{11}, \ldots, b_{kk};\, b_{12}, \ldots, b_{k-1,k})$$

$$\underline{x}_1' = (1;\, x_1 \ldots x_k;\quad x_1^2, \ldots, x_k^2;\, x_1x_2, \ldots, x_{k-1}x_k) .$$

The true relationship which applies over the whole operability region O is assumed to be the cubic polynomial function

$$\eta = \beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_kx_k + \beta_{11}x_1^2 + \ldots + \beta_{kk}x_k^2 + \beta_{12}x_1x_2 + \ldots + \beta_{k-1,k}x_{k-1}x_k$$

or, in matrix notation

$$\eta = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$$

where $\underline{x}_1$ is as above, $\underline{\beta}_1$ is defined like $\underline{b}_1$ and where

$\underline{\beta}_2' = (\beta_{111}, \beta_{122}, \ldots, \beta_{1kk}; \beta_{222}, \beta_{211}, \ldots, \beta_{2kk}; \ldots \beta_{kk,k-1}; \beta_{123}, \beta_{124}, \ldots, \beta_{k-2,k-1,k})$

$\underline{x}_2' = (x_1^3, x_1x_2^2, \ldots, x_1x_k^2; x_2^3, x_2x_1^2, \ldots, x_2x_k^2; \ldots x_{k-1}x_k^2; x_1x_2x_3, x_1x_2x_4, \ldots, x_{k-2}x_{k-1}x_k)$.

Exactly as in the previous paper (1959) we have $J = V + B$ where

$$V = N\Omega \int_R \underline{x}_1' (\underline{X}_1' \underline{X}_1)^{-1} \underline{x}_1 \, d\underline{x}$$

and

$$B = N\sigma^{-2}\Omega \int_R \underline{\beta}_2' [\underline{A}' \underline{x}_1 - \underline{x}_2][\underline{x}_1' \underline{A} - \underline{x}_2'] \underline{\beta}_2 \, d\underline{x}$$

where now

$$\underline{X}_1' = [\underline{x}_{11}, \ldots, \underline{x}_{1u}, \ldots, \underline{x}_{1N}]$$

is a $\frac{1}{2}(k+1)(k+2)$ by N matrix with $\underline{x}_{1u}' = (1, x_{1u}, x_{2u}, \ldots, x_{ku}, x_{1u}^2, \ldots, x_{ku}^2, x_{1u}x_{2u}, \ldots, x_{k-1,u}x_{ku})$

$$\underline{X}_2' = [\underline{x}_{21}, \ldots, \underline{x}_{2u}, \ldots, \underline{x}_{2N}]$$

is a $k(k+1)(k+2)/6$ by N matrix with $\underline{x}_{2u}' = (x_{1u}^3, x_{1u}x_{2u}^2, \ldots, x_{1u}x_{ku}^2; x_{2u}^3, \ldots; x_{1u}x_{2u}x_{3u}, \ldots, x_{k-2,u}x_{k-1,u}x_{ku})$ and $\underline{A} = (\underline{X}_1'\underline{X}_1)^{-1}\underline{X}_1'\underline{X}_2$ is the $\frac{1}{2}(k+1)(k+2)$ by $k(k+1)(k+2)/6$ "alias matrix." This last matrix has, for its elements, quantities which measure the extent to which the estimates $\underline{b}_1$ are biased by higher order coefficients in accordance with the equation

$$E(\underline{b}_1) = \underline{\beta}_1 + \underline{A}\underline{\beta}_2 .$$

By making the necessary substitutions we can evaluate V and B as follows.

2.2 Evaluation of B

Writing $\underline{\alpha}_2 = \underline{\beta}_2 N^{\frac{1}{2}}/\sigma$ we see that

$$\Omega^{-1}B = \underline{\alpha}_2' A'(\int_R \underline{x}_1\underline{x}_1' \, d\underline{x}) \underline{A}\underline{\alpha}_2 - 2\underline{\alpha}_2'(\int_R \underline{x}_2\underline{x}_1' \, d\underline{x}) \underline{A}\underline{\alpha}_2 + \underline{\alpha}_2'(\int_R \underline{x}_2\underline{x}_2' \, d\underline{x}) \underline{\alpha}_2 \qquad (2.2.1)$$

A straightforward series of calculations will provide the following results. Since the design is second order rotatable with fifth order moments zero

$$\underline{A} = \frac{\lambda_4}{\lambda_2}\left[\begin{array}{c|c|c|c|c} \underline{0} & \underline{0} & & \underline{0} & \underline{0} \\ \hline 3\,1\ldots 1 & & & & \\ & 3\,1\ldots 1 & & & \\ & & \ldots & & \underline{0} \\ & & & 3\,1\ldots 1 & \\ \hline \underline{0} & \underline{0} & & \underline{0} & \underline{0} \end{array}\right]$$

where $3\lambda_4 N = \sum_{u=1}^{N} x_{iu}^4 = 3\sum_{\substack{u=1\\ i\neq j}}^{N} x_{iu}^2 x_{ju}^2$ and $\lambda_2 N = \sum_{u=1}^{N} x_{iu}^2$ are the parameters of the second order rotatable design. The columns of $\underline{A}$ correspond to the elements of $\underline{x}_2'$ and the rows of $\underline{A}$ correspond to the elements of $\underline{x}_1$. Let us denote this fact by saying that A is $(\underline{x}_1)(\underline{x}_2')$. Only k^2 elements of A are non-zero, and these are shown. They occupy the second, third, ..., (k+1)th rows. In the second row they are in the first k columns, ..., in the (k+1)th row they are in the columns numbered (k^2-k+1) to k^2. Evaluation of any simple special case will quickly show the reader how these numbers arise, if it is not at once obvious. The divisions in $\underline{A}$ correspond to the semicolons in the $\underline{x}$-vectors mentioned.

Further straightforward calculations will show that

$$\Omega\int_R \underline{x}_1\underline{x}_1'\, d\underline{x} = \left[\begin{array}{c|c|c|c} 1 & \underline{0} & u\underline{j}_k' & \underline{0} \\ \hline \underline{0} & u\underline{I}_k & \underline{0} & \underline{0} \\ \hline u\underline{j}_k & \underline{0} & v(2\underline{I}_k + \underline{j}_k\underline{j}_k') & \underline{0} \\ \hline \underline{0} & \underline{0} & \underline{0} & v\underline{I}_p \end{array}\right]$$

where $u(k+2) = v(k+2)(k+4) = 1$, and as usual, $\underline{I}_k$ denotes the k by k unit matrix and $\underline{1}_k$ is a column vector of ones. This matrix is shape $(\underline{x}_1)(\underline{x}_1')$ in our notation, the divisions again corresponding to the semicolons, and, since $\frac{1}{2}(k+1)(k+2)$ is the number of elements in $\underline{x}_1$, $p = \frac{1}{2}(k+1)(k+2) - 2k - 1 = \frac{1}{2}k(k-1)$.

Similarly, we can show that

$$\Omega\int_R \underline{x}_2\underline{x}_1'\, d\underline{x} = v\left[\begin{array}{c|cccc|c} & \begin{matrix}3\\1\\\vdots\\1\end{matrix} & & & & \\ & & \begin{matrix}3\\1\\\vdots\\1\end{matrix} & & & \\ \underline{0} & & & \ddots & & \underline{0} \\ & & & & \begin{matrix}3\\1\\\vdots\\1\end{matrix} & \\ \hline \underline{0} & & & \underline{0} & & \underline{0} \end{array}\right]$$

where $v(k+2)(k+4) = 1$. This matrix is of exactly the same dimensions and is similar element-wise to the transpose $\underline{A}'$ of $\underline{A}$.

Again similarly we find that

$$\Omega\int_R \underline{x}_2\underline{x}_2'\, d\underline{x} = w\left[\begin{array}{c|c|c|c|c} \underline{G}_1 & & & & \\ \hline & \underline{G}_2 & & & \\ \hline & & \ddots & & \\ \hline & & & \underline{G}_k & \\ \hline & & & & \underline{I}_q \end{array}\right]$$

where
$$G_i = \begin{bmatrix} 15 & 3 & 3 & \cdots & 3 \\ 3 & 3 & 1 & \cdots & 1 \\ 3 & 1 & 3 & \ddots & 1 \\ \vdots & & & & \vdots \\ 3 & 1 & 1 & \cdots & 3 \end{bmatrix}, \text{ all } i,$$

and $w(k+2)(k+4)(k+6) = 1$. The matrix divisions correspond to the semicolon divisions in the vector $\underline{x}_2$. Hence, since each G_i is a k by k matrix, $q = k(k^2+3k+2)/6-k^2 = k(k-1)(k-2)/6$.

The values of the four matrices we have just quoted must now be substituted in equation (2.2.1) for B.

Making these substitutions, carrying out the appropriate matrix multiplications and collecting the terms element-wise in the matrices we find that the bias contribution B is given by

$$B = \underline{\alpha}_2' \, \underline{Q} \, \underline{\alpha}_2 ,$$

where
$$\underline{Q} = \left[\begin{array}{c|c|c|c|c} \underline{Q}_1 & & & & \\ \hline & \underline{Q}_1 & & & \\ \hline & & & & \\ \hline & & & \underline{Q}_1 & \\ \hline & & & & \underline{Q}_2 \end{array}\right],$$

$$\begin{array}{cc} & \begin{array}{ccccc} \alpha_{i11} & \alpha_{i22} & \alpha_{i33} & \cdots & \alpha_{ikk} \end{array} \\ \underline{Q}_1 = & \left[\begin{array}{ccccc} A & E & E & \cdots & E \\ E & C & D & \cdots & D \\ E & D & C & \cdots & \vdots \\ & \vdots & \vdots & C & D \\ E & D & \cdots & D & C \end{array}\right] \begin{array}{c} \alpha_{i11} \\ \alpha_{i22} \\ \alpha_{i33} \\ \vdots \\ \alpha_{ikk} \end{array} \end{array}$$

and

$$\underline{Q}_2 = \underline{I}_q/(k+2)(k+4)(k+6) = w\underline{I}_q$$

The α_{ijj} indicate the positions of the elements of $\underline{Q}_1$ and show how the quadratic form will arise. The elements of $\underline{Q}_2$ will be multiplied by terms like $\alpha_{ij\ell}$ where i, j and ℓ are all different.

If we define

$$\theta = 3\lambda_4/\lambda_2 ,$$
$$U = [\theta - 3/(k+4)]^2/9(k+2),$$
$$W = 1/(k+2)(k+4)^2(k+6);$$

then

$$A = 9U + 6(k+1)W$$
$$E = 3U - 6W$$
$$C = U + 2(k+3)W$$
$$D = U - 2W$$

We can now carry out the tedious, but not difficult, evaluation of the quadratic form and we find, eventually, that

$$B = PU + [(k+4)Q - 2P]\ W$$

where U and W are as defined above and where

$$NP/\sigma^2 \equiv (3\beta_{111}+\beta_{122}+\ldots+\beta_{1kk})^2+\ldots+(3\beta_{kkk}+\beta_{k11}+\ldots+\beta_{k,k-1,k-1})^2$$

$$NQ/\sigma^2 \equiv 2(3\beta_{111}^2+\beta_{122}^2+\ldots+\beta_{1kk}^2)+\ldots+2(3\beta_{kkk}^2+\beta_{k11}^2+\ldots+\beta_{k,k-1,k-1}^2)$$
$$+(\beta_{123}^2+\ldots+\beta_{k-2,k-1,k}^2) .$$

It is shown in appendix 2 that P and Q are both invariant under rotation. This means that given a true relationship of any particular kind, the bias will be independent of the orientation of the contours.

2.3 Evaluation of V

The matrix $(\underline{X}_1'\underline{X}_1)^{-1}$ is found from the formulae given in Box and Hunter (1958) for the inverse of certain matrices which frequently arise in response surface work. If we pre-multiply this inverse by $\underline{x}_1'$, post-multiply by $\underline{x}_1$ and carry out the appropriate integration we find

$$V = \frac{1}{c} + \frac{3(k-1)}{2(k+4)\theta c} + \frac{(k+2)(k+4)\,\theta c + 3 - 2(k+4)\,\theta}{(k+4)\,c[(k+2)\,\theta - 3kc]}$$

where $\theta = 3\lambda_4/\lambda_2$, as before and $c = \lambda_2$.

2.4 Minimization of J

Altogether, then, we have

$$\begin{aligned} J &= V + B \\ &= \frac{1}{c} + \frac{3(k-1)}{2(k+4)\theta c} + \frac{(k+2)(k+4)\theta c + 3 - 2(k+4)\,\theta}{(k+4)\,c[(k+2)\,\theta - 3kc]} \\ &\quad + P[\theta - 3/(k+4)]^2/9(k+2) \\ &\quad + [(k+4)Q - 2P]/(k+2)(k+4)^2(k+6) \end{aligned} \tag{2.4.1}$$

and we should like to choose c and θ, which is equivalent to choosing λ_2 and λ_4 in order to minimize J.

Suppose now that we fix θ. Then B is fixed and V depends only on c. Thus, it is possible to choose c as a function of θ, so that, <u>for each fixed</u> θ, V (and thus J over all) takes on the lowest possible value. This gives us J(θ) in terms of θ alone, after we substitute the appropriate value for c = c(θ) and we can then minimize J in terms of θ if we are given the values of P and Q which are functions of the β_{ijk}. Thus, for each pair of values (P, Q) we can choose θ so that the linked pair [c(θ), θ] give rise to a minimum value of J.

In fact, the minimizing design parameters $c(\theta)$ and θ depend only on P since Q enters only in the constant portion of B. However, as Q varies, for fixed P, the ratio $g = V/B$ changes since the amount of B changes as Q does, once P is fixed.

As a matter of practical calculation we shall not specify P and Q and then find the design which minimizes J. Computationally, it is far simpler to specify a value of θ, then determine c as a function of θ, $c(\theta)$, so that V is minimized and, finally, see for what value of P this design would be best, i.e., which P would give these specified $c(\theta)$ and θ as the ones which minimize J. We can then vary Q and see the effect of changes in Q, for fixed P, or the ratio $V/B = g$. Since the design depends only on the value of P, it will not change as we vary Q.

We shall carry out the numerical calculations outlined above for the cases $k = 1, 2, 3, 4$ and 5; first we obtain the necessary formulae in terms of general k.

2.5 Obtaining designs which will minimize J

We refer back to equation (2.4.1) for

$$J = V(c, \theta) + B(\theta, P, Q) \tag{2.5.1}$$

Fix θ and remember that P and Q, though their values are unknown, are constants. Then we must choose $c = c(\theta)$ so that V (and hence J) are minimized for this θ; thus we get

$$\frac{\partial V}{\partial c}(c, \theta) = 0$$

A lengthy, straightforward evaluation will give

$$c(\theta) = \theta \left\{ \frac{\{6[2(k+4)\theta + 3(k+1)][(k+2)^2(k+4)\theta^2 - 6k(k+4)\theta + 9k]\}^{\frac{1}{2}} - 3k[2(k+4)\theta + 3(k+1)]}{3[2\theta^2(k+2)(k+4) - 6k(k+4)\theta - 9k(k-1)]} \right\} \tag{2.5.2}$$

and we can, in principle, substitute this value in equation (2.5.1) so that now

$$J = V(\theta) + B(\theta, P, Q) \ .$$

Fortunately, as mentioned above, it is not necessary actually to make the substitution. We now differentiate J with respect to θ. Differentiating equation (2.5.1) with respect to θ, remembering that c is a function of θ, and equating the result to zero we obtain

$$\frac{\partial V}{\partial c} \cdot \frac{dc}{d\theta} + \frac{\partial V}{\partial \theta} + \frac{dB}{d\theta} = 0$$

or

$$\frac{\partial V}{\partial \theta} + \frac{dB}{d\theta} = 0 \ , \quad \text{since } \frac{\partial V}{\partial c} \equiv 0 \ .$$

Since

$$\frac{\partial V}{\partial \theta} = - \frac{3}{(k+4)c} \quad \frac{k-1}{2\theta^2} + \frac{k(k+2)(k+4)c^2 - 2k(k+4)c + (k+2)}{[(k+2)\theta - 3kc]^2}$$

and

$$\frac{dB}{d\theta} = \frac{2}{9(k+2)} \left(\theta - \frac{3}{k+4}\right) P$$

it follows that the equation $\partial V/\partial\theta + dB/d\theta = 0$ implies

$$P = \frac{27(k+2)}{2c[(k+4)\theta - 3]} \quad \frac{k-1}{2\theta^2} + \frac{k(k+2)(k+4)c^2 - 2k(k+4)c + (k+2)}{[(k+2)\theta - 3kc]^2} \qquad (2.5.3)$$

where $c = c(\theta)$ as given in equation (2.5.2).

Thus, if we select a value for k and then a value for θ, we can use equations (2.5.2) and (2.5.3) to tabulate sets of values of (θ, c, P). Although obtained in that order, they can be interpreted as follows. A design which minimizes J must necessarily have moments related by $c = c(\theta)$, as in equation (2.5.2), when there is a contribution from V. We shall choose the appropriate "all-bias" design so that it, too, has moments related by the equation $c = c(\theta)$. Then, for a given P, we can find the appropriate value for θ from equation (2.5.3). Additionally, for a given Q, we can evaluate B for the situation being considered (V will already be available) and the ratio V/B can be determined. We now have to use these calculations to arrive at some general conclusions about the correct design to use in "typical" situations which might arise.

2.6 The calculations and their interpretation

The calculations described above have been performed for $k = 1, 2, 3, 4$ and 5 and for sufficiently many representative values of P and Q so that the behavior over all possible (P, Q) can be predicted.

In the case $k = 1$, P and Q are both multiples of β_{111} which is the only cubic coefficient and so Q is fixed if P is given and the possible situations are $0 \leq P \leq \infty$, i.e., one-dimensional. Table 1 shows figures for the best design moment values in various situations, i.e., for various possible values of P. For the all-bias situation we should choose $c^{\frac{1}{2}} = 0.606, \lambda = 1.632$ where $\lambda = \theta/c = 3\lambda_4/\lambda_2^2$, and is independent of scale. As situations arise where the bias contribution to J becomes smaller and smaller we see that the best design moment values increase. For example, when $V = 8B$ approximately, the best design is such that $c^{\frac{1}{2}} = 0.7$, $\lambda = 2.0$ approximately, which is quite close to the appropriate design for the "all-bias" situation and far from the appropriate "all-variance" design which is, as always, the largest possible (denoted in the table by infinite moments, but in practice as large as possible until restricted by the operability region O). When $V = B$ approximately, a situation we can regard as "typical," (as described in our earlier paper) the best design is such that approximately, $c^{\frac{1}{2}} = 0.621$ $\lambda = 1.669$, very close to the "all-bias" figures.

When $k \geq 2$, the possible situations are $0 \leq P \leq \infty$, $0 \leq Q \leq \infty$, i.e., two-dimensional. The best design in a given situation depends only on P, however, though the value of Q (when P is considered fixed) affects the relative values of V and B and hence affects $g = V/B$. The graph, Figure 1 which applies to the case $k = 2$, is typical of the results for each $k \geq 2$. The appropriate design for the all-bias situation has $c^{\frac{1}{2}} = 0.515$, $\lambda = 1.887$. Over a very large range of possible values of P, the optimum design changes only slightly, the moments becoming, progressively, slightly larger compared with those for the best design for the all-bias situation. Only when quite small values of P are postulated do the moments of the best design increase appreciably and, of course, as in all cases, when there is no contribution from bias at all ($P = 0$) the best design is the largest possible. Note that a situation can arise where P is very small or zero and, at the same time, Q is large; thus, the total bias could be quite large (because of the size of Q) but the appropriate best design would be the largest possible because only the V part of J can be affected by altering the moments. In such a case, the lack of fit would be large but the coefficients would be either unbiased ($P = 0$) or not very biased (P small).

This last-mentioned set of circumstances ($P = 0$, Q large) is somewhat unlikely and we should like, now, to consider what "typical" situations might arise in practice. As in previous work (1959) we shall regard as a "typical" situation one in which V and B are, approximately, of the same size, i.e., $g = V/B = 1$, approximately. Figure 1 shows several curves on which g is constant and we shall consider what designs are appropriate as P and Q vary in such a way that the point (P, Q) stays on a curve $g =$ constant. The same sort of conclusions we arrived at when considering P alone are true here too.

Assuming now that the point (P, Q) is on the line g = 1, we can say:

When P and Q are "very large" or infinite the appropriate best design is the all-bias design $c^{\frac{1}{2}} = 0.515$, $\lambda = 1.887$. As P and Q become smaller, the best design has moments which become larger than, but not very different from, those of the all-bias design. Until quite small values of P and Q are examined the appropriate best design is only slightly more spread out than the all-bias design and is completely different from the all-variance design which is the largest possible as always.

As is evident from Figure 1, even when the variance contribution V is several times greater than the bias contribution B, the same conclusions still hold. Curves for g = V/B = 2, 4 and 8 all lie between the g=1 curve and the line P =(k+2)Q which is the boundary of possible points (P, Q) since, as can easily be seen after a little calculation, $(k+2)Q \geq P$.

Thus, overall, we can conclude that in most situations, and even in circumstances where V is expected to be several times as large as B, the appropriate experimental designs to use to minimize J have moments slightly larger than the moments of the appropriate all-bias design. As a practical matter in situations where no information about the possible sizes of V and B exists, about 10% greater is suggested by the authors. Tables 1 through 5 consist of selected values from larger computations and contain specific recommendations for the cases k = 1, 2, 3, 4 and 5. (The suggested values are obtained as follows: Take 110% times the value of $c^{\frac{1}{2}}$ when $P = \infty$. Select the table entry (with $P \neq \infty$) which has a value of $c^{\frac{1}{2}}$

nearest to this calculated number. It is of course possible to find the exact pair $(c^{\frac{1}{2}}, \lambda)$ which has $c^{\frac{1}{2}}$ equal to 110% of $c^{\frac{1}{2}}$ when $P = \infty$, by use of equation (2.5.2). However, since the "10% greater" is a guide and not an exact figure, this is hardly worthwhile.)

2.7 Use of the secondary consideration to select a design

Even now we do not have a specific design, but only a certain subset of designs, since the requirements "second order rotatable, with zero fifth moments, and λ_2 and λ_4 of a given size" can be satisfied by a number of designs for every value of k. How, then, do we select an individual design from all those with the correct sized moments? To do this, we appeal to our second criterion which specifies that our selected design should make large the quantity

$$\sum_{u=1}^{N} \{E(\hat{y}_u) - \eta_u\}^2 = E(S_R) - \nu\sigma^2 = NF, \text{ say.}$$

As explained in our earlier paper (1959), this quantity can be written

$$\begin{aligned} NF &= \underline{\beta}_2' \underline{A}' \underline{X}_1' \underline{X}_1 \underline{A} \underline{\beta}_2 - 2\underline{\beta}_2' \underline{X}_2' \underline{X}_1 \underline{A} \underline{\beta}_2 + \underline{\beta}_2' \underline{X}_2' \underline{X}_2 \underline{\beta}_2 \\ &= -\underline{\beta}_2' \underline{X}_2' \underline{X}_1 \underline{A} \underline{\beta}_2 + \underline{\beta}_2' \underline{X}_2' \underline{X}_2 \underline{\beta}_2 \end{aligned} \qquad (2.7.1)$$

since one term cancels directly.

The matrix $N^{-1}(\underline{X}_2' \underline{X}_1 \underline{A})$ is square and of dimension k(k-1)(k-2)/6. It consists of a number of submatrices down the main diagonal. The first k of these are of dimension k by k and have the form

(continued on page 30)

Table 1: Best design moment values for various P, when k=1.

P	V	B	g	$c^{\frac{1}{2}}$	λ
∞	2.961	∞	0	0.606	1.632
5785	2.917	14.714	0.198	0.610	1.640
1728	2.835	4.446	0.638	0.617	1.657
933	2.763	2.457	1.125	0.623	1.674
602	2.697	1.640	1.644	0.629	1.691
426	2.638	1.208	2.183	0.635	1.709
245	2.534	0.777	3.260	0.647	1.745
131	2.409	0.508	4.737	0.662	1.801
60	2.254	0.340	6.621	0.685	1.896
30	2.124	0.262	8.100	0.709	2.009
20	2.058	0.234	8.795	0.723	2.084
0	1.000	0	∞	∞	∞

Suggested values for "unknown" situation: $c^{\frac{1}{2}} = 0.667, \lambda = 1.820$

Table 2: Best design moment values for various P, when k=2.

P	V	$c^{\frac{1}{2}}$	λ
∞	5.936	0.515	1.887
3896	5.297	0.535	1.923
1709	4.912	0.549	1.957
957	4.605	0.562	1.993
604	4.353	0.574	2.032
411	4.144	0.585	2.072
295	3.966	0.596	2.112
169	3.681	0.615	2.194
88	3.368	0.640	2.316
33	2.966	0.683	2.549
15	2.680	0.725	2.798
0	1.000	∞	∞

Suggested values for "unknown" situation: $c^{\frac{1}{2}} = 0.562, \lambda = 1.993$

Note: In this table B is not shown because it depends on both P and Q. The above moment values are appropriate for P no matter what Q may be, since it cannot affect the choice of the design; see text.

Table 3: Best design moment values for various P, when k = 3

P	V	$c^{\frac{1}{2}}$	λ
∞	9.920	0.462	2.062
2346	7.093	0.508	2.170
1463	6.599	0.521	2.212
986	6.193	0.533	2.254
702	5.852	0.544	2.298
397	5.310	0.564	2.385
249	4.897	0.583	2.471
118	4.307	0.616	2.638
87	4.087	0.630	2.717
45	3.671	0.664	2.906
21	3.238	0.709	3.181
0	1.000	∞	∞

Suggested values for "unknown" situation: $c^{\frac{1}{2}} = 0.508, \lambda = 2.170$

Note: In this table B is not shown because it depends on both P and Q. The above moment values are appropriate for P no matter what Q may be, since it cannot affect the choice of the designs; see text.

Table 4: Best design moment values for various P, when k = 4

P	V	$c^{\frac{1}{2}}$	λ
∞	14.907	0.414	2.189
2062	8.549	0.493	2.384
1442	7.976	0.505	2.430
615	6.725	0.537	2.566
392	6.138	0.556	2.654
189	5.301	0.589	2.822
121	4.856	0.612	2.941
58	4.231	0.651	3.163
30	3.754	0.690	3.399
15	3.348	0.734	3.674
9	3.071	0.772	3.923
0	1.000	∞	∞

Suggested values for "unknown" situation $c^{\frac{1}{2}} = 0.493$, $\lambda = 2.384$

Note: In this table B is not shown because it depends on both P and Q. The above moment values are appropriate for P no matter what Q may be, since it cannot affect the choice of the designs; see text.

Table 5: Best design moment values for various P, when k = 5

P	V	$c^{\frac{1}{2}}$	λ
∞	20.898	0.328	2.286
2433	10.106	0.477	2.540
1762	9.407	0.489	2.588
1319	8.814	0.501	2.634
1014	8.306	0.511	2.680
797	7.865	0.521	2.725
518	7.139	0.540	2.813
300	6.323	0.566	2.939
144	5.398	0.603	3.136
59	4.502	0.653	3.421
12	3.372	0.758	4.072
0	1.000	∞	∞

Suggested values for "unknown" situation: $c^{\frac{1}{2}} = 0.477$, $\lambda = 2.540$

Note: In this table B is not shown because it depends on both P and Q. The above moment values are appropriate for P no matter what Q may be, since it cannot affect the choice of the designs; see text.

$$\frac{\lambda_4^2}{\lambda_2}\begin{bmatrix} 9 & 3 & \cdots & 3 \\ 3 & 1 & \cdots & 1 \\ \vdots & \vdots & & \\ 3 & 1 & \cdots & 1 \end{bmatrix}$$

All other elements, apart from the k^3 mentioned above, are zero, in the matrix $N^{-1}(\underline{X}_2'\,\underline{X}_1\,\underline{A})$.

Thus,

$$\underline{\beta}_2'\,\underline{X}_2'\,\underline{X}_1\,\underline{A}\,\underline{\beta}_2 = \lambda_4^2\lambda_2^{-1}\,\{9(\beta_{111}^{\,2}+\ldots)+6(\beta_{111}\beta_{122}+\ldots)+(\beta_{122}^{\,2}+\ldots)+2(\beta_{122}\beta_{133}+\ldots)\}$$

$$= \lambda_4^2\lambda_2^{-1}\,\{(3\beta_{111}+\beta_{122}+\ldots+\beta_{1kk})^2+\ldots+(3\beta_{kkk}+\beta_{k11}+\ldots+\beta_{k,\,k-1,\,k-1})^2\}$$

$$= \lambda_4^2\lambda_2^{-1}P.$$

The matrix $N^{-1}(\underline{X}_2'\,\underline{X}_2)$ is also square and of dimension $k(k-1)(k-2)/6$. It consists of (k+1) submatrices down the main diagonal. The first of these is of the form

$$N^{-1}\begin{bmatrix} \sum x_{1u}^6 & \sum x_{1u}^4x_{2u}^2 & \cdots & \sum x_{1u}^4\,x_{ku}^2 \\ \sum x_{1u}^4x_{2u}^2 & \sum x_{1u}^2x_{2u}^4 & \cdots & \sum x_{1u}^2x_{2u}^2x_{ku}^2 \\ \vdots & & & \\ \sum x_{1u}^4\,x_{ku}^2 & \sum x_{1u}^2\,x_{2u}^2x_{ku}^2 & \cdots & \sum x_{1u}^2x_{ku}^4 \end{bmatrix}$$

and the second, third, . . . , down to the k-th are similar but with the obvious variation in suffices. The (k+1)-th matrix is of diagonal form with terms such as

$N^{-1}\sum x_{1u}^2 x_{2u}^2 x_{3u}^2$. All summations are over u. The pattern of suffices can be seen by imagining the vector $\underline{\beta}_2$ to be written out along the top and side of the matrix $\underline{X}_2'\,\underline{X}_2$. Apart from the elements already mentioned, all elements of $\underline{X}_2'\,\underline{X}_2$ are zero.

To make further progress on our examination of F we shall make use of a concept introduced in our previous paper (1959), namely that, since the orientation of our true surface with respect to our design is unknown, we shall average the value of F over all orthogonal rotations, denoting the average value by $\tilde{F}$. Only the second term on the right-hand side of equation (2.7.1) for F is affected by this rotation; as we have seen above,the first term, which is a multiple of P is independent of the rotation. The details of the averaging process are given in Appendix 3. Our result is that

$$\tilde{F} = (P+Q)\sum_{u=1}^{N} r_u^6 / Nk(k+2)(k+4) - \lambda_4^2\lambda_2^{-1}P .$$

Since λ_2 and λ_4 are already determined by the work given above, for given (or suggested) values of P and Q, only the quantity $\sum_u r_u^6$ is capable of allocation as far as the design is concerned. It follows that our requirement that the design should be such that $\tilde{F}$ is made large implies that $\sum_u r_u^6$ should be large.

It should be noted that this conclusion is based on the assumption that the true model is of exactly third order; if this were indeed so we should choose a design which maximizes $\sum_u r_u^6$. However, if bias coefficients of fourth and higher order were present in the true model,we should find our treatment led to a rather different conclusion - that $\sum_u r_u^6$ should be kept small, at some suggested value, and that a quantity of even higher order be made large. Thus as a hedge against the possibility of presence of higher order terms we shall, as a practical matter, require $\sum_u r_u^6$ to be large but not necessarily maximal. This will achieve an intuitively reasonable compromise for the choice of design, as we shall see in the example which follows.

3. APPLICATION OF THE SOLUTION TO PARTICULAR PROBLEMS

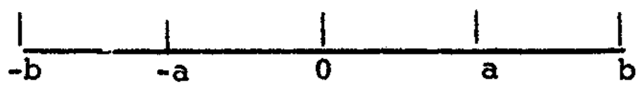
3.1 An application of the results to the case k = 1

When k = 1 we have

$$\hat{y} = b_0 + b_1 x + b_{11}x^2 \ ,$$

$$\eta = \beta_0 + \beta_1 x + \beta_{11}x^2 + \beta_{11} x^3 \ .$$

Suppose, for the purposes of this example, that we wished to have a design with N = 10 points and with not more than five distinct levels of x, one of the five being the origin and the others symmetrically placed as in the sketch



How would we allocate the ten points to the five levels to get a design which would be judged best by the criteria used in this paper?

The appropriate design for the all-bias situation is one for which $c^{\frac{1}{2}} = 0.606$, $\lambda = 1.632$. However, if we felt that, probably, V and B would be of the same size, we could choose $c^{\frac{1}{2}} = 0.62$, $\lambda = 1.67$; or if we felt that variance error would probably be about eight times as big as bias error we could choose $c^{\frac{1}{2}} = 0.70$, $\lambda = 2.00$.

For any particular choice of moment values, several ten point designs at five or less levels are possible and the situation is shown in Tables 6 and 7. Note that both levels a and b must be used in order to satisfy the specified values of $c^{\frac{1}{2}}$ and λ.

Table 6: Possible Designs when $c^{\frac{1}{2}} = 0.62, \lambda = 1.67$

Number of points at					Value of		Number Proportional to
-b	-a	0	a	b	a	b	$\sum r_u^6$
1	4	0	4	1	0.478	1.009	142
1	3	2	3	1	0.567	0.984	125
1	2	4	2	1	0.790	0.828	75
2	1	4	1	2	0.777	0.815	68
2	3	0	3	2	0.358	0.880	62
2	2	2	2	2	0.452	0.873	60

All the above designs have the same (c, λ).

Table 7: Possible Designs when $c^{\frac{1}{2}} = 0.70$, $\lambda = 2.00$

Number of points at					Value of		Number Proportional to $\sum r_u^6$
-b	-a	0	a	b	a	b	
1	4	0	4	1	0.500	1.225	45
1	3	2	3	1	0.587	1.210	42
1	2	4	2	1	0.755	1.166	39
2	3	0	3	2	0.303	1.054	37
2	2	2	2	2	0.376	1.053	37
2	1	4	1	2	0.553	1.047	36

All the above designs have the same (c, λ).

We now make use of our secondary consideration to choose a particular design from whichever group we decided to use. We recall that our theory told us to choose a design which makes large the quantity $\sum r_u^6$.

If we thought that only cubic bias need be guarded against we would select the design which gave the largest value of $\sum r_u^6$. However, if, as usual, we would like to hedge a little against the possibility of higher order bias we would like $\sum r_u^6$ to be large but not necessarily maximal, for as we indicated in our earlier paper, minimization of higher order biases would require $\sum r_u^6$ to be kept small and for an even higher order quantity to be made large. Thus, as a compromise in our present situation we would probably select the second design in each table as the appropriate one to use in the circumstances we have assumed in this paper.

Note especially that the "evenly spaced" design (2, 2, 2, 2, 2) is not a particularly good one in our assumed circumstances, depite the fact that it might seem the natural arrangement to choose.

3.2 An application to certain rotatable designs for $2 \leq k \leq 5$.

Consider, in k dimensions, a rotatable design consisting of a cube plus octahedron plus center points, the basic central composite rotatable design. This consists of $N = (2^k + 2k + n_0)$ points where n_0 is the number of center points. Let the points of the cube be $(\pm a, \pm a, \ldots, \pm a)$; of the octahedran $(\pm b, 0, \ldots, 0), \ldots, (0, 0, \ldots, \pm b)$. Necessarily, $b = 2^{k/4}a$ for rotatability. Thus to achieve given values of $c^{\frac{1}{2}}$, λ for this design we must allocate values to a and to n_0 since $c = (2^k + 2^{k/2})\, a^2/N$ and

$$\lambda = 3\lambda_4/\lambda_2{}^2 = 2^k\, a^4/c^2 N = N/(1+2^{k/2})^2$$

It is thus clear that, if we choose a value for n_0, λ is fixed, hence c is fixed since the (c, λ) pairs are linked, and thus a is fixed. We can examine possible variations of this type of design, therefore, by considering various numbers of center points to be added. The upper portion of Table 8 shows the parameters of the designs which result when $0 \leq n_0 \leq 12$, for the case k = 2. Recalling the results of our theoretical work we see that for an "all-bias, no variance" situation we are told to use a design for which $(\lambda, c^{\frac{1}{2}}) = (1.887,\ 0.515)$. This would be approximately achieved by using a cube of half-side a = 0.565, an octahedron with b = 0.799 and two center points, as we see from Table 8. However, if we felt that the situation to be investigated was not all bias we should want to use a bigger design. We can see from Table 2 that pairs $(\lambda, c^{\frac{1}{2}})$ between (1.887, 0.515) and (2.798, 0.725) are suitable design "sizes" for possible values of P between infinity and 15, where

$$P = N\ \{(\beta_{111}+\beta_{122})^2 + (\beta_{222}+\beta_{122})^2\}/\sigma^2 \ ,$$

and where $N = 8 + n_0$. The extreme suitable designs for these extremes of P are, respectively ($n_0 = 2$, a = 0.565, b = 0.799) and ($n_0 = 7$, a = 0.996, b = 1.408). The region of interest R, it should be remembered, is the unit circle. We can thus observe that, as we expect less and less effect from the biases of the coefficients we add more and more center points to the composite design and place the points further and further from the origin, even outside R the region of interest. (Recall that for two factors, $E(b_1) = \beta_1 + (\beta_{111} + \beta_{122})$ and $E(b_2) = \beta_2 + (\beta_{222} + \beta_{112})$ are the expectations of the estimates of the linear coefficients in a fitted quadratic model when cubic terms exist and this particular type of design is used. Hence P is proportional to the sum of squares of biases in the estimates of the linear coefficients). This is extremely reasonable. As we become surer of our model (i.e. bias is thought to be small) we spread out the design; but since variance error becomes a greater and greater part of the total discrepancy between fitted and actual model we add more center points to provide a better estimate of the error variation. On the other hand, if we doubt our model (i.e. bias is thought to be large) and we believe variance error to be a small part of the total discrepancy between fitted and actual model we contract our design into the region of interest R and use only enough center points (two or three, say) to provide some estimate of σ^2.

Succeeding tables, Tables 8 through 11 contain calculations, similar to those described, for the cases $2 \leq k \leq 5$ and for two types of rotatable designs, (a) cube plus octahedron plus center points and (b) cube plus doubled octahedron plus center points. Similar comments apply to all cases. Note that for designs (b) where the octahedron is doubled, rotatability implies $b = 2^{(k-1)/4} a$.

Table 8: Parameter values for certain rotatable designs when k = 2

(a) Cube plus octahedron (8 points) plus n_0 center points.

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.628	1.500	0.628	0.880
1	0.578	1.688	0.613	0.867
2	0.505	1.875	0.565	0.799
3	0.583	2.063	0.684	0.967
4	0.627	2.250	0.768	1.086
5	0.663	2.438	0.846	1.196
6	0.696	2.625	0.921	1.303
7	0.727	2.813	0.996	1.408
8	0.757	3.000	1.070	1.514
9	0.785	3.188	1.145	1.619
10	0.813	3.375	1.220	1.725
11	0.840	3.563	1.295	1.832
12	0.867	3.750	1.371	1.939

(b) Cube plus doubled octahedron (12 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.618	1.544	0.689	0.820
1	0.583	1.673	0.677	0.805
2	0.528	1.802	0.636	0.756
3	0.538	1.930	0.671	0.798
4	0.582	2.059	0.749	0.891
5	0.614	2.188	0.814	0.968
6	0.640	2.316	0.874	1.040
7	0.665	2.445	0.933	1.109
8	0.688	2.574	0.990	1.177
9	0.709	2.702	1.046	1.244
10	0.730	2.831	1.102	1.311
11	0.751	2.960	1.158	1.377
12	0.770	3.088	1.214	1.444

(Note: The recommended "all bias" design for k =2 has λ= 1.887, $c^{\frac{1}{2}}$= 0.515.)

Table 9: Parameter values for certain rotatable designs when k = 3

(a) Cube plus octahedron (14 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.536	1.802	0.542	0.912
1	0.507	1.930	0.531	0.894
2	0.453	2.059	0.490	0.825
3	0.514	2.188	0.573	0.964
4	0.548	2.316	0.630	1.059
5	0.577	2.445	0.681	1.145
6	0.603	2.574	0.730	1.228
7	0.628	2.702	0.778	1.309
8	0.651	2.831	0.826	1.389
9	0.673	2.960	0.873	1.468
10	0.694	3.088	0.920	1.548
11	0.715	3.217	0.967	1.627
12	0.735	3.346	1.014	1.706

(b) Cube plus doubled octahedron (20 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.522	1.875	0.583	0.825
1	0.493	1.969	0.565	0.799
2	0.457	2.063	0.535	0.757
3	0.503	2.156	0.603	0.853
4	0.532	2.250	0.651	0.921
5	0.555	2.344	0.694	0.981
6	0.576	2.438	0.734	1.038
7	0.595	2.531	0.773	1.093
8	0.613	2.625	0.811	1.147
9	0.631	2.719	0.849	1.201
10	0.648	2.813	0.887	1.254
11	0.664	2.906	0.924	1.307
12	0.680	3.000	0.961	1.359

(Note: the recommended "all bias" design for k = 3 has λ= 2.062, $c^{\frac{1}{2}}$= 0.456)

Table 10: Parameter values for certain rotatable designs when k = 4

(a) Cube plus octahedron (24 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.475	2.000	0.475	0.950
1	0.459	2.083	0.469	0.937
2	0.421	2.167	0.438	0.876
3	0.450	2.250	0.478	0.955
4	0.479	2.333	0.517	1.034
5	0.502	2.417	0.551	1.103
6	0.522	2.500	0.584	1.167
7	0.541	2.583	0.615	1.229
8	0.558	2.667	0.645	1.290
9	0.575	2.750	0.675	1.349
10	0.591	2.833	0.704	1.408
11	0.607	2.917	0.733	1.466

Please Note: Because of the later addition of tables, there is no page numbered 41.

Table 11: Parameter values for certain rotatable designs when k = 5

(a) Cube plus octahedron (42 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.430	2.149	0.423	1.007
1	0.421	2.200	0.419	0.997
2	0.402	2.252	0.406	0.965
3	0.397	2.303	0.405	0.963
4	0.422	2.354	0.435	1.034
5	0.440	2.405	0.458	1.089
6	0.455	2.456	0.479	1.139
7	0.469	2.507	0.499	1.187
8	0.482	2.559	0.518	1.232
9	0.495	2.610	0.537	1.277
10	0.507	2.661	0.555	1.321
11	0.518	2.712	0.573	1.363
12	0.529	2.763	0.591	1.406

(b) Cube plus doubled octahedron (52 points) plus n_0 center points

n_0	$c^{\frac{1}{2}}$	λ	a	b
0	0.428	2.167	0.445	0.890
1	0.419	2.208	0.440	0.880
2	0.403	2.250	0.428	0.855
3	0.389	2.292	0.416	0.832
4	0.413	2.333	0.447	0.893
5	0.430	2.375	0.468	0.936
6	0.443	2.417	0.487	0.975
7	0.456	2.458	0.505	1.010
8	0.467	2.500	0.522	1.045
9	0.478	2.542	0.539	1.078
10	0.488	2.583	0.555	1.110
11	0.498	2.625	0.571	1.142
12	0.508	2.667	0.587	1.173

(Note: The recommended "all bias" design for k = 5 has $\lambda = 2.286$, $c^{\frac{1}{2}} = 0.328$.)

APPENDIX I

A problem in numerical analysis (see section 1.6)

Suppose we have a function $\eta(\underline{x})$, known exactly without error. Suppose we wish to approximate to this function over a region R by a polynomial of form $\hat{\eta}(\underline{x}) = \underline{x}_1' \underline{\gamma}_1$, say, and of order d_1.

We shall choose $\underline{\gamma}_1$ so that the integral defined by

$$\sum = \Omega \int_R (\eta(\underline{x}) - \hat{\eta}(\underline{x}))^2 \, d\underline{x} \qquad \text{is minimized .}$$

Now

$$\sum = \Omega \int_R (\eta(\underline{x}) - \underline{x}_1' \underline{\gamma}_1)^2 \, d\underline{x} \; .$$

Thus

$$\frac{\partial \sum}{\partial \underline{\gamma}} = \Omega \int_R \underline{x}_1 \{\eta(\underline{x}) - \underline{x}_1' \underline{\gamma}_1\} \, d\underline{x} = 0$$

implies that

$$\Omega \int_R \underline{x}_1 \eta(\underline{x}) \, d\underline{x} = \{\Omega \int_R \underline{x}_1 \underline{x}_1' \, d\underline{x}\} \underline{\gamma}_1 \; .$$

Using earlier definitions on the right-hand side and calling the left-hand side $\underline{\mu}_{1\eta}$ by analogy, we can write this as

$$\underline{\mu}_{1\eta} = \underline{\mu}_1 \underline{\gamma}_1 \; .$$

Therefore,

$$\underline{\gamma}_1 = \underline{\mu}_1^{-1} \underline{\mu}_{1\eta}$$

no matter what η may be, assuming that $\underline{\mu}_1$ is non-singular.

Let us now apply this to the case where

$$\eta(\underline{x}) = \underline{x}_1' \underline{\beta}_1 + \underline{x}_2' \underline{\beta}_2 \qquad \text{, a polynomial of order } d_2 \text{, with}$$

$\underline{x}_1$ as before,

then
$$\underline{\gamma}_1 = \mu_1^{-1} \{\Omega \int_R \underline{x}_1 (\underline{x}_1' \underline{\beta}_1 + \underline{x}_2' \underline{\beta}_2)\, d\underline{x}\}$$
$$= \mu_1^{-1}\{\underline{\mu}_1 \underline{\beta}_1 + \underline{\mu}_2 \underline{\beta}_2\}$$
$$= \underline{\beta}_1 + \underline{\mu}_1^{-1}\underline{\mu}_2 \underline{\beta}_2$$
$$= \underline{\beta}_1 + \underline{\mathcal{A}}\underline{\beta}_2 , \qquad (\text{where } \underline{\mathcal{A}} = \underline{\mu}_1^{-1}\underline{\mu}_2)$$

What is the value of the average integrated discrepancy between the fitted and true model in this case? It is, of course, minimal and has value

$$\sum_{(min)} = \Omega \int_R \{\eta(\underline{x}) - \hat{\eta}(\underline{x})\}^2 d\underline{x} , \quad (\text{where } \hat{\eta}(\underline{x}) = \underline{x}_1' \underline{\gamma}_1)$$
$$= \Omega \int_R \{\underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2 - \underline{x}_1'\underline{\beta}_1 - \underline{x}_1'\underline{\mathcal{A}}\underline{\beta}_2\}^2 d\underline{x}$$
$$= \Omega \int_R (\underline{x}_2' - \underline{x}_1'\underline{\mathcal{A}})'(\underline{x}_2' - \underline{x}_1'\underline{\mathcal{A}})\, d\underline{x}$$
$$= \Omega \int_R \{\underline{x}_2\underline{x}_2' - \underline{x}_2\underline{x}_1'\underline{\mathcal{A}} - \underline{\mathcal{A}}'\underline{x}_1\underline{x}_2' + \underline{\mathcal{A}}\,\underline{x}_1\underline{x}_1'\underline{\mathcal{A}}\}\, d\underline{x}$$
$$= \underline{\mu}_3 - \underline{\mu}_2'\underline{\mu}_1^{-1}\underline{\mu}_2 - \underline{\mu}_2'\underline{\mu}_1^{-1}\underline{\mu}_2 + \underline{\mu}_2'\underline{\mu}_1^{-1}\underline{\mu}_1\underline{\mu}_1^{-1}\underline{\mu}_2$$
$$= \mu_3 - \underline{\mu}_2'\underline{\mu}_1^{-1}\underline{\mu}_2$$

Thus, if we knew the polynomial $\eta(\underline{x}) = \underline{x}_1'\underline{\beta}_1 + \underline{x}_2'\underline{\beta}_2$ <u>exactly</u> and fitted $\hat{\eta}(\underline{x}) = \underline{x}_1'\underline{\gamma}_1$ to it to minimize $\sum$, then $\sum_{(min)}$ is the value of the average integrated discrepancy or bias which we cannot avoid; it is the smallest bias we can achieve.

But, if we fit, not to a known function, but to a function whose value is known only at a few points, then we know that the bias which arises is

$$B_{(min)} = \sum_{(min)} + (\underline{A} - \underline{\mathcal{A}})'\, \underline{\mu}_1(\underline{A} - \underline{\mathcal{A}})$$

as shown in Box and Draper (1959).

Thus by choosing $\underline{A}$, which is at our disposal and involves design moments, in such a way that $\underline{A} = \underline{\mathcal{A}}$ we can make

$$B_{(min)} = \sum_{(min)}$$

Thus even if we do not know our polynomial function $\eta(\underline{x})$ exactly, we can choose the points at which to evaluate $\eta(\underline{x})$ in such a way that the bias incurred when we graduate by $\hat{\eta}(\underline{x})$ is exactly what it would be (and no more!) in the situation where we know the function exactly at every point.

APPENDIX 2

To show that P and Q are invariant under rotation (see section 2.2)

We know that

$$P = (3\alpha_{111} + \alpha_{122} + \ldots + \alpha_{1kk})^2 + \ldots + (3\alpha_{kkk} + \alpha_{k11} + \ldots + \alpha_{k,k-1,k-1})^2$$

$$Q = 2(3\alpha^2_{111} + \alpha^2_{122} + \ldots + \alpha^2_{1kk}) + \ldots + 2(3\alpha^2_{kkk} + a^2_{k11} + \ldots + \alpha^2_{k,k-1,k-1})$$
$$+ (\alpha^2_{123} + \ldots + \alpha^2_{k-2,k-1,k})$$

where $\alpha^2_{ijk} = N\beta^2_{ijk}/\sigma^2$.

For the proof that these quantities P and Q are invariant under rotations of the surface about the origin we shall make use of the matrix direct product which has the following properties (Marcus, 1960).

$$\underline{A}*\underline{B} = \begin{bmatrix} b_{11}\underline{A} & b_{12}\underline{A} & \cdots & b_{1n}\underline{A} \\ b_{21}\underline{A} & b_{22}\underline{A} & \cdots & b_{2n}\underline{A} \\ \vdots & & & \\ b_{m1}\underline{A} & b_{m2}\underline{A} & \cdots & b_{mn}\underline{A} \end{bmatrix}$$

$$(A * B)' = A' * B'$$
$$(A * B) * C = A * C + B * C$$
$$(A * B)(C * D) = AC * BD$$

Let $\underline{H}$ be the matrix of an orthogonal transformation taking $\underline{x}$ into $\underline{X}$ by $\underline{x} = \underline{H}\,\underline{X}'$

where

$$\underline{x}' = (x_1, x_2, \ldots, x_k)$$
$$\underline{X}' = (X_1, X_2, \ldots, X_k)$$

Now

$$\underline{x}' * \underline{x}' = \{x_1(x_1, x_2, \ldots, x_k); \ldots; x_k(x_1, x_2, \ldots, x_k)\}$$

and

$$\underline{x}' * \underline{x}' * \underline{x}' = [x_1\{x_1(x_1, x_2, \ldots, x_k); \ldots; x_k(x_1, x_2, \ldots, x_k)\}, \ldots, x_k\{x_1(x_1, x_2, \ldots, x_k) ; \ldots; x_k(x_1, x_2, \ldots, x_k)\}]$$

Write

$$\underline{\beta}' = \{\beta_{111}, \tfrac{1}{3}\beta_{112}, \tfrac{1}{3}\beta_{113}, \ldots, \tfrac{1}{3}\beta_{11k}; \tfrac{1}{3}\beta_{112}, \tfrac{1}{3}\beta_{122}, \tfrac{1}{6}\beta_{123}, \ldots\}$$

Then

$$\underline{\beta}'(\underline{x}' * \underline{x}' * \underline{x}') = \beta_{111}x_1^3 + \beta_{122}x_1^2x_2 + \ldots$$
$$= \underline{x}_2' \underline{\beta}_2 \ .$$

where the right hand side consists of all the cubic terms, with appropriate coefficients, of the cubic response surface model. Employing the transformation $\underline{x}' = H\underline{X}'$ we can write

$$\underline{\beta}'(\underline{x}' * \underline{x}' * \underline{x}') = \underline{\beta}'(\underline{H}\,\underline{X}' * \underline{H}\,\underline{X}' * \underline{H}\,\underline{X}')$$
$$= \underline{\beta}'(\underline{H} * \underline{H} * \underline{H})(\underline{X}' * \underline{X}' * \underline{X}')$$
$$= \underline{B}'(\underline{X}' * \underline{X}' * \underline{X}') \ ,$$

say, where $\underline{B}' = \underline{\beta}'(\underline{H} * \underline{H} * \underline{H})$ is the new "β vector" for the transformed coordinates. Thus, for these transformed coordinates, Q would be given by

$$Q = \underline{B}'\ \underline{B}$$
$$= \underline{\beta}'(\underline{H} * \underline{H} * \underline{H})(\underline{H}' * \underline{H}' * \underline{H}')\ \underline{\beta}$$
$$= \underline{\beta}'(\underline{H}\,\underline{H}' * \underline{H}\,\underline{H}' * \underline{H}\,\underline{H}')\ \underline{\beta}$$
$$= \underline{\beta}'(\underline{I} * \underline{I} * \underline{I})\ \underline{\beta}$$
$$= \underline{\beta}'\ \underline{\beta} \ ,$$

using the fact that, since H is orthogonal, $\underline{H}\,\underline{H}' = \underline{I}$, the unit matrix of appropriate dimension. Hence Q is invariant under rotation of axes (or equivalently under rotation of the response surface relative to the axes).

Now let $\underline{u}_i' = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ be a 1 by k vector with unity in the i-th place and zeros elsewhere. Further, let $\underline{u}' = (\underline{u}_1', \underline{u}_2', \ldots, \underline{u}_k')$ so that $\underline{u}'$ is a 1 by k^2 vector with k unities in the appropriate positions and zeros elsewhere. Finally, let $\underline{U} = \underline{u} * \underline{I}$ where $\underline{I}$ is a k by k unit matrix. Then $\underline{U}$ is a k^3 by k matrix consisting of submatrices $\underline{u}$ in the diagonal and zeros elsewhere as follows:

$$\underline{U} = \begin{bmatrix} \underline{u} & & & \underline{0} \\ & \underline{u} & & \\ & & \ddots & \\ \underline{0} & & & \underline{u} \end{bmatrix} \quad \text{where } \underline{u} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \\ \vdots \\ \underline{u}_k \end{bmatrix}, \qquad \underline{u}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

In these terms, and recalling that $\underline{B}' = \underline{\beta}'(\underline{H} * \underline{H} * \underline{H})$, a little thought will show that for the transformed coordinates

$$P = \underline{B}'\,\underline{U}\,\underline{U}'\,\underline{B}$$

$$= \underline{\beta}'(\underline{H} * \underline{H} * \underline{H})(\underline{u} * \underline{I})(\underline{u}' * \underline{I})(\underline{H}' * \underline{H}' * \underline{H}')\,\underline{\beta}$$

$$= \underline{\beta}'\,\{(\underline{H} * \underline{H})\,\underline{u} * \underline{H}\}\{\underline{u}'(\underline{H}' * \underline{H}') * \underline{H}'\}\,\underline{\beta}$$

(We here use the fact that, since $\underline{H}$ is orthogonal $\{\underline{H} * \underline{H}\}\,\underline{u} = \underline{u}$.)

$$= \underline{\beta}'(\underline{u} * \underline{H})(\underline{u}' * \underline{H}')\,\underline{\beta}$$

$$= \underline{\beta}'(\underline{u}\,\underline{u}' * \underline{H}\,\underline{H}')\,\underline{\beta}$$

$$= \underline{\beta}'(\underline{u}\,\underline{u}' * \underline{I})\,\underline{\beta}$$

$$= \underline{\beta}'(\underline{u} * \underline{I})(\underline{u}' * \underline{I})\,\underline{\beta}$$

$$= \underline{\beta}'\,\underline{U}\,\underline{U}'\,\underline{\beta}$$

Thus, P also is invariant under the transformation $\underline{H}$.

APPENDIX 3

The average of F over all orthogonal rotations of the response surface (see section 2.7)

We need consider only $\underline{\beta}_2' \underline{X}_2' \underline{X}_2 \underline{\beta}_2$ since the other portion of NF is unaffected by the rotation.

Write

A = average value of β^2_{111} and similar terms ,

B = average value of $\beta_{111}\beta_{122}$ and similar terms ,

C = average value of β^2_{122} and similar terms ,

D = average value of $\beta_{122}\beta_{133}$ and similar terms ,

E = average value of β^2_{123} and similar terms .

Only three of these quantities are independent and two relations exist between them as will be shown.

Then

$$F' = \text{Rotational average of } \underline{\beta}_2' \underline{X}_2' \underline{X}_2 \underline{\beta}_2$$

$$= A \sum_i \sum_u x^6_{iu} + 2B \sum_i \sum_u x^4_{iu} (r^2_u - x^2_{iu})$$

$$+ C \sum_i \sum_u x^2_{iu} (\sum_j x^4_{ju} - x^4_{iu})$$

$$+ D \sum_i \sum_u x^2_{iu} (\sum_{\substack{j \neq i \\ \ell \neq i \\ j \neq \ell}} x^2_{ju} x^2_{\ell u})$$

$$+ (E/6) \sum_{i \neq j \neq \ell} \sum_u x^2_{iu} x^2_{ju} x^2_{\ell u}$$

$$= A \sum_u r_u^6$$

$$+ (2B + C - 3A)\{ \sum_i x_{iu}^4 (r_u^2 - x_{iu}^2) \}$$

$$+ \{D + (E/6) - A\}\{ \sum_u \sum_{i \neq j \neq \ell} x_{iu}^2 x_{ju}^2 x_{\ell u}^2 \} ,$$

where $r_u^2 = x_{1u}^2 + \ldots + x_{ku}^2$.

Since this is a rotational average it is necessarily a function of the r_u^2 only. A little thought will convince the reader that this necessarily implies the relationships

$$2B + C - 3A = 0$$

$$D + (E/6) - A = 0$$

It follows that

$$F' = A \sum_u r_u^6 \ .$$

We now recall our earlier definition

$$P = (3\beta_{111} + \beta_{122} + \ldots + \beta_{1kk})^2 + (k-1) \text{ similar terms}$$

$$Q = 2(3\beta_{111}^2 + \beta_{122}^2 + \ldots + \beta_{1kk}^2) + (k-1) \text{ similar terms} + (\beta_{123}^2 + \ldots).$$

Hence, averaging over all rotations, we see that

$$P = k \{9A + 6(k-1)B + (k-1)C + (k-1)(k-2)D \}$$

$$Q = k \{6A + 2(k-1)C \} + k(k-1)(k-2)E/6$$

and

$$P + Q = k\{15A + 6(k-1)B + 3(k-1)C + (k-1)(k-2) D + (k-1)(k-2) E/6 \}$$

$$= k\{ (k+2)(k+4)A + 3(k-1)(2B+C-3A) + (k-1)(k-2)(D+[E/6]-A) \}$$

$$= k(k+2)(k+4) A ,$$

because of the relationships mentioned above.

Substituting for A_3 we find

$$F' = (P + Q) \left(\sum_u r_u^6\right) / k(k+2)(k+4)$$

and so

$$\widetilde{F} = (P + Q) \left(\sum_u r_u^6\right) / Nk(k+2)(k+4) - \lambda_4^2 \lambda_2^{-1} P .$$

REFERENCES

Box, G. E. P. and Hunter, J. S. (1957), "Multifactor experimental designs for exploring response surfaces," Annals of Mathematical Statistics, Volume 28, 1957, pp. 195-241.

Box, G. E. P. and Draper, Norman R., (1959) "A basis for the selection of a response surface design," Journal of the American Statistical Association, Volume 54, 1959, pp. 622-654.

Marcus, Marvin, (1960) "Basic theorems in matrix theory," National Bureau of Standards, Applied Mathematics Series, No. 57.

FIGURE I: THE BEST SECOND ORDER ROTATABLE DESIGN PARAMETERS AND THE RATIO g = V/B FOR VARIOUS VALUES OF P AND Q, WHEN k = 2