A PROPERTY OF THE METHOD OF STEEPEST ASCENT

By Charles Henry Johnson and John Leroy Folks

Oklahoma State University

1. Introduction. Box and Wilson (1951) proposed the method of steepest ascent as part of a general technique for attaining optimal operating conditions. The experimental points are coded and the path of steepest ascent is calculated in the space of the coded points. This path is decoded and further experiments are performed along it until it is felt that no further appreciable gain can be realized. It is well known that the path of steepest ascent is not scale invariant. In general, the decoded path and the path of steepest ascent calculated from the uncoded points, are distinct.

This procedure of moving along the gradient of a fitted plane has appeal since we can expect a plane to provide a reasonable fit of the surface locally. It would seem that we might gain efficiency by revising the path after each observation. Therefore, suppose that as each additional experiment is performed, it is combined with all previous experiments to calculate a revised path by fitting a plane. The results of Plackett (1950) for inclusion of additional observations, have been put in a form useful for this purpose by Box and Wilson (1951). For certain designs, if an additional point is taken along the path of steepest ascent in the space of the coded variables, the revised path is the same as the initial path. In the space of the uncoded variables, if an additional point is taken along the decoded path, the revised path of steepest ascent is the same as the initial path. Indeed, these statements remain true however many additional experiments are performed.

2. Notation. Let X_1 be the $N_1 \times (p + 1)$ initial design matrix; X_2 be the $N_2 \times (p + 1)$ design matrix of the N_2 additional points, $B' = [b_0, b_1, \cdots b_p]$ be the vector of coefficients of the plane fitted to the initial N_1 responses; $A' = [a_0, a_1, \cdots, a_p]$ be the vector of coefficients of the plane fitted to the $N_1 + N_2$ responses, and Y_2 be the $N_2 \times 1$ vector of responses at the N_2 additional points. With this notation the expression for A, (see Box and Wilson) is written

$$A = B + J'G\Delta \tag{1}$$

where

$$J = X_2(X_1'X_1)^{-1}, \quad G = (I_{N_2} + JX_2'), \quad \Delta = Y_2 - X_2B.$$

3. Results. Consider the three conditions:

$$X_2 = (X_{20}, X_{21}) = (X_{20}, HB_1'T^2) \tag{a}$$

Where $X_{20} = \text{col}\ (1, \cdots, 1)$, $H = \text{col}\ (h_1, \cdots, h_{N_2})$, $B_1 = \text{col}\ (b_1, b_2, \cdots, b_p)$, and $T = \text{diag}\ (t_1, \cdots, t_p)$.

Received 1 August 1963.

(b) $$X_1' X_1 = \begin{bmatrix} w & 0 \\ 0 & T^2 v I_p \end{bmatrix}$$

(c) $$a_i/a_j = b_i/b_j, \qquad i, j = 1, 2, \cdots, p.$$

Condition (a) states that the N_2 additional points lie on the decoded path of steepest ascent. Condition (b) is a restriction on the initial experimental points and is satisfied by several well known designs including the 2^p factorial designs with center at the origin as well as the rotatable designs. Condition (c) states the proportionality which results in the initial and revised paths being identical.

THEOREM. *Any two of the conditions,* (a), (b), *and* (c), *imply the third.*

PROOF.

1. *Proof that* (a) *and* (b) *imply* (c).

$$J = X_2(X_1' X_1)^{-1} = (X_{20}, HB'T^2) \begin{bmatrix} w^{-1} & 0 \\ 0 & T^{-2} v^{-1} I_p \end{bmatrix}$$

so that

(2) $$J' = \begin{bmatrix} X_{20}' w^{-1} \\ B_1 H' v^{-1} \end{bmatrix}.$$

Partitioning A and B and substituting J into (1) gives

(3) $$\begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} + \begin{bmatrix} X_{20}' w^{-1} \\ B_1 H' v^{-1} \end{bmatrix} G\Delta$$

where A_0 and B_0 are 1×1 and A_1 and B_1 are $p \times 1$.

Hence

(4) $$A_1 = B_1 + B_1 v^{-1} H' G\Delta = B_1(1 + v^{-1} H' G\Delta) = kB_1$$

where k is the scalar $1 + v^{-1} H' G\Delta$.

Condition (c) follows immediately from (4).

2. *Proof that* (a) *and* (c) *imply* (b). Let $(X_1' X_1)^{-1}$ be partitioned as

$$(X_1' X_1)^{-1} = \left[\begin{array}{c|ccc} c_{00} & c_{01} & \cdots & c_{0p} \\ \hline c_{10} & c_{11} & \cdots & c_{1p} \\ \vdots & & & \\ c_{p0} & c_{p1} & \cdots & c_{pp} \end{array}\right] = \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix}.$$

Then

$$J = (X_{20}, HB_1' T^2) \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix}$$

so that

$$J' = \begin{bmatrix} C_{00} \; X'_{20} + C_{01} \; T^2B_1 \; H' \\ C_{10} \; X'_{20} + C_{11} \; T^2B_1 \; H' \end{bmatrix}. \tag{5}$$

Since (c) implies that $A_1 = kB_1$, (1) may be written

$$\begin{bmatrix} A_0 \\ kB_1 \end{bmatrix} = \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} + \begin{bmatrix} C_{00} \; X'_{20} + C_{01} \; T^2B_1 \; H' \\ C_{10} \; X'_{20} + C_{11} \; T^2B_1 \; H' \end{bmatrix} G\Delta. \tag{6}$$

Hence

$$kB_1 = B_1 + C_{10}X'_{20}G\Delta + C_{11}T^2B_1H'G\Delta. \tag{7}$$

Since (7) is true for all B_1, $C_{10} = 0$. Then (7) becomes

$$B_1(kI - H'G\Delta C_{11}T^2) = B_1 \tag{8}$$

so that $kI - H'G\Delta C_{11}T^2 = I$. It follows that

$$[(k - 1)I/H'G\Delta]T^{-2} = C_{11}$$

so that C_{11}^{-1} is a scalar multiple of T^2 as in (b).

3. *Proof that* (b) *and* (c) *imply* (a). (1) may be written

$$\begin{bmatrix} A_0 \\ kB_1 \end{bmatrix} = \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} + \begin{bmatrix} w^{-1} & 0 \\ 0 & T^{-2}v^{-1}I_p \end{bmatrix} \begin{bmatrix} X'_{20} \\ X'_{21} \end{bmatrix} G\Delta. \tag{9}$$

Then $kB_1 = B_1 + T^{-2}v^{-1}I_pX'_{21}G\Delta$ so that $v(k - 1)T^2B_1 = X'_{21}G\Delta$ and

$$(G\Delta)'X_{21} = rB'_1T^2 \tag{10}$$

where $r = v(k - 1)$. That is

$$(G\Delta)'X_{21} = r(b_1t_1^2b_2t_2^2b_3t_3^2 \cdots b_pt_p^2).$$

Then every element of the jth column of X_{21} contains $b_jt_j^2$. That is, the elements of X_{21} have the form $x_{ij} = a_{ij}b_jt_j^2$. Since (10) is true for all $G\Delta$, $a_{ij} = a_{ik}$ and it follows that X_2 has the form stated in Condition (a).

Discussion. Note that when $T = I_p$, the theorem is a statement about the situation in the coded variables. It is interesting that when $T_p \neq I_p$, the decoded path is not the path of steepest ascent in the uncoded variables. However, if points are run along this path and a plane is fitted to the total observations, the revised path is the same as the initial path.

REFERENCES

Box, G. E. P., and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc. Ser. B* **13** 1–45.

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika* **37** 149–157.