CHAPTER 6

The Use of Steepest Ascent to Achieve Process Improvement

If we attempted to fit an empirical function such as a polynomial over the whole *operability region* (that is, over the whole region within which the studied system could be operated) a very complex function would usually be needed. The fitting of such a function could involve an excessive number of experiments. The exploration of the whole operability region is, however, almost never a feasible or sensible objective. First, the extent of this region is almost never known and second, the experimenter can often safely dismiss as unprofitable whole areas of the region where experiments could, theoretically, be conducted. What he usually wants to do, with good reason, is to explore a smaller subregion *of interest* which often changes as the investigation progresses. He can expect that, within this smaller subregion, a fairly simple graduating equation will be representationally adequate.

Now the mathematical procedure of fitting selects, from all possible surfaces of the degree fitted, that which approximates the responses *at the experimental points* most closely (in the least squares sense). The features of the fitted surface at points remote from the region of the experimental design need not, and usually would not, bear any resemblance to the features of the actual surface. We can therefore expect our approximation to be useful only in the immediate neighborhood of the current experimental region. If, for example, we were already close to a maximum, we might represent its main features approximately by fitting, to a suitable locally placed set of experimental points, an equation of only second degree, possibly in transformed variables. However, particularly if the system were being investigated for the first time, starting conditions would often not be very close to such a maximum. Thus, the experimenter often first needs some *preliminary procedure*, to bring him to a suitable point where the second degree equation can most usefully be employed. *One* such preliminary procedure is the one factor at a time method. An alternative which,

in the authors' opinion, is usually more effective and economical in experiments (at least in the fields of application in which it has been used) is the "steepest ascent" method. This preliminary procedure, which may later be followed by the fitting of an equation of second degree (possibly in transformed variables), is the basis of a sequential experimentation method which we discuss in this chapter and which has proved extremely effective in a wide variety of applications.

6.1. WHAT STEEPEST ASCENT DOES

For illustration, consider the investigation of a chemical system whose percentage yield depends on the three predictor variables time, temperature, and concentration; a possible situation is shown in Figure 6.1. The overall objective is to find settings of the variables which will result in improved yields. Suppose we are at an early stage of investigation when considerable improvement is possible. Then the planar contours of a first degree equation can be expected to provide a fair approximation in the immediate region of a point such as P, which is far from the optimum. A direction at right angles to these contour planes is the direction of *steepest ascent*, if it points toward higher yield values. (The opposite direction is that of *steepest descent*, needed for *reducing*, and so improving, a response such as "impurity" for example.) In practice, the first degree approximating equation must be estimated via experiments. This may be done by running a design such as that indicated by the dots around the point P.

Exploratory runs performed along the path of steepest ascent and indicated by stars in Figure 6.1 will normally show increasing yield along the path. The best point found, or an interpolated estimated maximum point on the path could then be made the base for a new first-order design from which a further advance might be possible.

The direction of ascent we shall calculate is the steepest when the response surface is scaled in the units of the design as it is, for example, in Figure 6.1. The effects of scale changes are discussed in Appendix 6A.

It will usually be found that, after one or perhaps two applications of steepest ascent, first-order effects will no longer dominate and the first-order approximation will be inadequate. At this stage, further steepest ascent progress will not be possible and more sophisticated second-order methods, discussed in Chapters 7 and 9, should be applied. The steepest ascent technique is particularly effective for the initial investigation of a new process. Such an investigation typically takes place on the laboratory scale.

For an already operating process arrived at after much development work, designs performed in the vicinity of the currently known best process

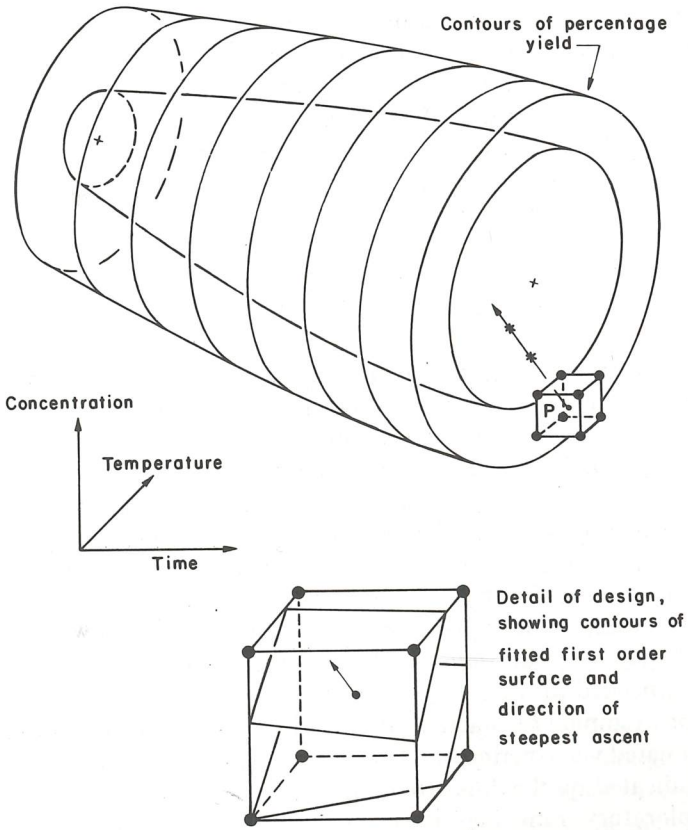

FIGURE 6.1. A stage of the steepest ascent procedure.

may show hardly any first-order effects at all, because these have already been exploited to obtain the current conditions. In this situation then, no initial large gains may be possible using steepest ascent but, again, the more sophisticated second-order methods of Chapter 9 may still lead to substantial improvement.

6.2. AN EXAMPLE. IMPROVEMENT OF A PROCESS USED IN DRUG MANUFACTURE

In laboratory experiments on a reaction taking place between four reagents A, B, C, and D, the amount of A was kept constant but five factors, time of reaction (ξ_1), temperature of reaction (ξ_2), amount of B (ξ_3), amount of C

TABLE 6.1. Factors ξ_i and coded levels x_i

ξ_i	Selected levels of x_i: Lower(−1)	Upper(+1)	
ξ_1	6 h	10 h	so that $x_1 = (\xi_1 - 8)/2$
ξ_2	85°C	90°C	so that $x_2 = (\xi_2 - 87.5)/2.5$
ξ_3	30 cm^3	60 cm^3	so that $x_3 = (\xi_3 - 45)/15$
ξ_4	90 cm^3	115 cm^3	so that $x_4 = (\xi_4 - 102.5)/12.5$
ξ_5	40 g	50 g	so that $x_5 = (\xi_5 - 45)/5$

(ξ_4), and the amount of D (ξ_5) were varied as in Table 6.1. A 2_{V}^{5-1} design with **I = 12345** was used and the design and resulting percentage yields, y, were as shown on the left of Table 6.2.

On the assumption that the second degree polynomial

$$y = \beta_0 + \sum_{i=1}^{5} \beta_i x_i + \sum_{i=1}^{5} \beta_{ii} x_i^2 + \sum_{i<j}^{5}\sum^{5} \beta_{ij} x_i x_j + \varepsilon$$

is adequate to represent the response locally, least-squares estimates of the

TABLE 6.2. 2^{5-1} design, responses, and the Yates' analysis

Run order and number	$x_1x_2x_3x_4x_5$	y	Yates' analysis (1)	(2)	(3)	(4)	Half-effect (5)[a]	Effect name
16	− − − − +	51.8	108.1	213.2	420.3	914.8	$\bar{y}$ = 57.1750	0
2	+ − − − −	56.3	105.1	207.1	494.5	−53.6	−3.3500	1
10	− + − − −	56.8	112.1	242.0	−19.5	−34.6	−2.1625	2
1	+ + − − +	48.3	95.0	252.5	−34.1	2.2	0.1375	12
14	− − + − −	62.3	122.1	−4.0	−20.1	4.4	0.2750	3
8	+ − + − +	49.8	119.9	−15.5	−14.5	−12.0	−0.7500	13
9	− + + − +	49.0	132.9	−16.8	3.5	−24.2	−1.5125	23
7	+ + + − −	46.0	120.1	−17.3	5.7	−30.6	−1.9125	45
4	− − − + −	72.6	4.5	−3.0	−6.1	74.2	4.6375	4
15	+ − − + +	49.5	−8.5	−17.1	−1.5	−14.6	−0.9125	14
13	− + − + +	56.8	−12.5	−2.2	−11.5	5.6	0.3500	24
3	+ + − + −	63.1	−3.0	−12.3	−0.5	9.2	0.5750	35
12	− − + + +	64.6	−23.1	−13.0	−14.1	16.6	1.0375	34
6	+ − + + −	67.8	6.3	29.4	22.5	4.0	0.6875	25
5	− + + + −	70.3	3.2	29.4	22.5	4.0	0.2500	15
11	+ + + + +	49.8	−20.5	−23.7	−53.1	−75.6	−4.7250	5

[a]Except for $\bar{y}$.

TABLE 6.3. First- and second-order contributions to corrected total sum of squares

Source	df	SS	MS
First order	5	956.905	191.38
Second order	10	150.785	15.08
Total, corrected	15	1107.690	

five first-order coefficients β_i and the 10 two-factor interaction coefficient β_{ij} are

$$\begin{aligned} b_1 &= -3.35, & b_{12} &= 0.14, & b_{24} &= 0.35, \\ b_2 &= -2.16, & b_{13} &= -0.75, & b_{25} &= 0.69, \\ b_3 &= 0.28, & b_{14} &= -0.92, & b_{34} &= 1.04, \\ b_4 &= 4.64, & b_{15} &= 0.25, & b_{35} &= 0.58, \\ b_5 &= -4.73, & b_{23} &= -1.51, & b_{45} &= -1.91. \end{aligned}$$

It will be recalled that, because each of the x variables has been allocated the coded levels -1 and $+1$, the least-squares values of the b's will be one-half of the values obtained via Yates' method in Table 6.2. In this experiment, some of the estimated second-order effects are not particularly small compared with those of first order. On the whole, however, the first-order effects are considerably larger in magnitude than those of second order, as evidenced, for example, by the analysis of variance in Table 6.3. As an approximation, it was decided to assume tentatively that second degree effects could, for the present, be ignored and to combine the second-order effects to provide an "estimate of error" based on 10 df. From this, the standard error of each estimate is $(15.08/16)^{1/2} = 0.971$. Note that the approximation is conservative in the sense that, if second-order effects are not negligible, this calculation will tend to *inflate* the estimate of error variance. Under these assumptions we have

$$\begin{aligned} b_1 &= -3.35 \pm 0.97, \\ b_2 &= -2.16 \pm 0.97, \\ b_3 &= 0.28 \pm 0.97, \\ b_4 &= 4.64 \pm 0.97, \\ b_5 &= -4.73 \pm 0.97. \end{aligned}$$

The estimated direction of steepest ascent, when the variables are scaled in units of the design, then follows the vector of coefficient values $[-3.35, -2.16, 0.28, 4.64, -4.73]$. The length of this vector is $[(-3.35)^2 + (-2.16)^2 + \cdots + (-4.73)^2]^{1/2} = 7.74$. A vector of unit length in the direction of steepest ascent therefore has coordinates $-3.35/7.74 = -0.43$, $-2.16/7.74 = -0.28$, and so on, that is, $[-0.43, -0.28, 0.04, 0.60, -0.61]$. If we follow the steepest ascent path then, for a change of -0.43 units of the first variable we should change the second variable by -0.28 units, change the third variable by $+0.04$ units, and so on. Exploratory runs were chosen at the conditions shown in Table 6.4, obtained by multiplying the unit length vector by 2, 4, 6, and 8, respectively. These particular multiples were chosen because the first of these is close to the periphery of the experimental region. It corresponds quite closely with run number 4 which, in fact, gave the highest recorded yield of 72.6, while the spacing separating the runs E_2, E_4, E_6, and E_8 is about the same as that separating the upper and lower levels of the factorial design. Note that the decoded conditions may be obtained by inverting the formulas in Table 6.1 so that

$$\xi_1 = 2x_1 + 8,$$

$$\xi_2 = 2.5x_2 + 87.5,$$

$$\xi_3 = 15x_3 + 45,$$

$$\xi_4 = 12.5x_4 + 102.5,$$

$$\xi_5 = 5x_5 + 45.$$

Because of its closeness to the conditions of run number 4, the experiment E_2 was not actually run. The observed yields in Table 6.4 are plotted in Figure 6.2, from which it was concluded that a conditional maximum along the path of steepest ascent was located close to E_6.

TABLE 6.4. Points on the path of steepest ascent in coded and decoded units

	Coded conditions					Decoded conditions					Observed yields
	x_1	x_2	x_3	x_4	x_5	Time	Temperature	Amount B	Amount C	Amount D	
E_2:	−0.86	−0.56	0.08	1.20	−1.22	6.3	86.1	46.2	117.5	38.9	72.6 (run 4)
E_4:	−1.72	−1.12	0.16	2.40	−2.44	4.6	84.7	47.4	132.5	32.8	85.1
E_6:	−2.58	−1.68	0.24	3.60	−3.66	2.8	83.3	48.6	147.5	26.7	82.4
E_8:	−3.44	−2.24	0.32	4.80	−4.88	1.1	81.9	49.8	162.5	20.6	80.8

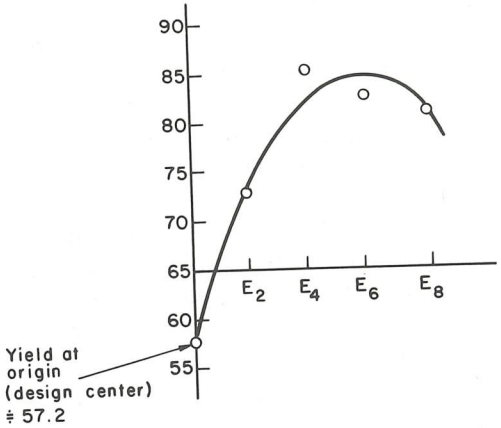

FIGURE 6.2. Yields found along the path of steepest ascent. The yield at the origin is approximated by the average 57.2 of all the 16 runs. The steepest ascent procedure has thus been very effective, moving us to a new set of conditions yielding over 80%.

6.3. GENERAL COMMENTS ON STEEPEST ASCENT

Design

It will be remarked that in the example, the fractional factorial employed was a 2_{V}^{5-1}, a design of resolution V, requiring 16 runs. It might have been possible to make progress with an eight-run design but in this case the experimenters decided on a somewhat conservative course, partly because of their feeling that the first-order (i.e., main) effects might not be particularly large compared with the experimental error.

Steepest Ascent or Descent?

In this example, the *objective function* was yield and the objective was to increase it. Other objective functions that might be important in chemical investigations would be *unit cost* and *level of impurity*, both of which we would wish to *decrease* rather than increase. This would require a path of steepest *descent*, given by changing all the signs from those of the path of steepest ascent at any point. Alternatively, the same effect can be achieved by changing the sign of the objective function and following the path of steepest ascent. Thus, essentially, all problems can be set up as *ascent* problems.

When Will Ascent Methods Work?

If we are looking for, say, a maximum yield, it would be wasteful to investigate *in any detail* regions of *low* yield. Typically, the main features in such regions are the first-order (i.e., main) effects which can point a direction of ascent up a surface. Once first-order effects become small compared to those of second order, or compared to the error, or to both, it may be necessary to switch to a second-order approximation. The appropriate designs can then be chosen to cover somewhat larger regions.

Checks on the adequacy of the first-order approximation or, equivalently, on the need for second-order terms, are supplied by:

1. Examining individual interaction contrasts, which are not used for estimation of first-order effects.
2. An overall curvature check.

A useful overall curvature check is supplied by adding a number of center points to a two-level fractional factorial or factorial design. The order in which the complete design, including center points, is carried out should be random. Thus, a set of randomly replicated experimental runs from which a pure error estimate of the basic variation in the data can be calculated is obtained. The overall curvature check is supplied by the contrast

$$c = \{\text{average response in two-level factorial runs}\} - \{\text{average response for runs at the center}\}.$$

It may be shown that, if the true response function is of second degree, c supplies an unbiased estimate of the sum of the pure quadratic effects, namely

$$c \to \sum_{i=1}^{k} \beta_{ii}.$$

For a surface that contained a minimax, the β_{ii} would be of different signs and theoretically $E(c)$ could be zero even when large curvatures β_{ii} occurred. However, minimax surfaces are rare in practice and typically the β_{ii} will be of the same sign.

There will nevertheless be intermediate situations where it is unclear whether or not the first-order approximation is good enough to allow progress via steepest ascent. In such cases, it is usually worthwhile to calculate the path of steepest ascent anyway, and to make a few trials along it to find out.

A discussion of the effect of experimental error on the estimation of the direction of steepest ascent is given in Section 6.4.

The First-Order Design as a Building Block

Suppose that, possibly after previous steepest ascent application(s), a new first-order design has just been completed. The appropriate checks, possibly augmented by tentative local exploration, indicate that no further advance by steepest ascent is likely and that we need to proceed with a second-order exploration. It will often be possible to incorporate the first-order design just completed as a building block of the second-order design now required. Consider, for instance, the chemical process example discussed above. Suppose the 2^{5-1} design had shown large second-order effects (i.e., two-factor interactions) and that these appeared to be reasonably well estimated. Then the addition of further points to the first-order design in a manner we describe in more detail in Sections 9.2 and 13.8 could convert it into a second-order design suitable for fitting a full second degree equation. The results from this second-order fit might then permit further response improvement.

Steepest Ascent as a Precursor to Further Investigation

Steepest ascent is rarely an end in itself. It is mainly of value as a preliminary procedure, to move the region of interest to a neighborhood of improved response, worthy of more thorough investigation.

Steepest Ascent Classroom Simulation

The steepest ascent procedure can be realistically simulated for teaching purposes; see, for example, Mead and Freeman (1973).

6.4. A CONFIDENCE REGION FOR THE DIRECTION OF STEEPEST ASCENT

In the example of Section 6.2 the first-order coefficients were estimated as

$$\begin{aligned} b_1 &= -3.35 \pm 0.97, \\ b_2 &= -2.16 \pm 0.97, \\ b_3 &= 0.28 \pm 0.97, \\ b_4 &= 4.64 \pm 0.97, \\ b_5 &= -4.73 \pm 0.97, \end{aligned} \tag{6.4.1}$$

where the standard error of each estimate (0.97) was based on an estimated variance having 10 df. Retaining now the convention that we consider the variables as scaled in the units of the design, the vector

$$(-3.35, -2.16, 0.28, 4.64, -4.73) \tag{6.4.2}$$

provides the estimated best direction of advance. Assuming that a first degree response equation provides an adequate model, we may now ask how much in error due to sampling variation this direction might be. This question can be answered in a conventional way by obtaining a confidence region for the direction of steepest ascent (Box, 1955, p. 211). This confidence region will turn out to be a cone (or hypercone depending on the number of dimensions involved) whose axis is the estimated direction vector.

Suppose there are k variables and we seek a confidence region for the *true* direction of steepest ascent as defined by its direction cosines $\delta_1, \delta_2, \ldots, \delta_i, \ldots, \delta_k$. If $b_1, b_2, \ldots, b_k$ are the estimated first degree effects, the expected values of these quantities are proportional to $\delta_1, \delta_2, \ldots, \delta_k$ so that

$$E(b_i) = \gamma\delta_i, \qquad i = 1, 2, \ldots, k \tag{6.4.3}$$

where γ is some constant. Now if we think of this relationship as a regression model in which the b_i are responses and the δ_i are the levels of a single-predictor variable, then γ is the "regression coefficient" of $b_1, \ldots, b_k$ on $\delta_1, \delta_2, \ldots, \delta_k$. The required region is supplied by those elements $\delta_1, \delta_2, \ldots, \delta_k$ which just fail to make the residual mean square significant compared with $V(b_i) = \sigma_b^2$ at some desired level α. That is, for those δ's which satisfy

$$\frac{\left\{\sum b_i^2 - \left(\sum_{i=1}^{k} b_i\delta_i\right)^2 \Big/ \sum \delta_i^2\right\} \Big/ (k-1)}{s_b^2} \le F_\alpha(k-1, \nu_b), \tag{6.4.4}$$

where s_b^2 is an estimate of σ_b^2 and ν_b is the number of degrees of freedom on which this estimate s_b^2 is based. For the particular case where $k = 2$ we have an application of Fieller's theorem (Fieller, 1955) of which the present development is one extension. Because all the quantities in the foregoing inequality are known except for the values of the δ's, this expression defines a set of acceptable δ's, thus a set of acceptable vectors, and hence a confidence region for the direction of steepest ascent.

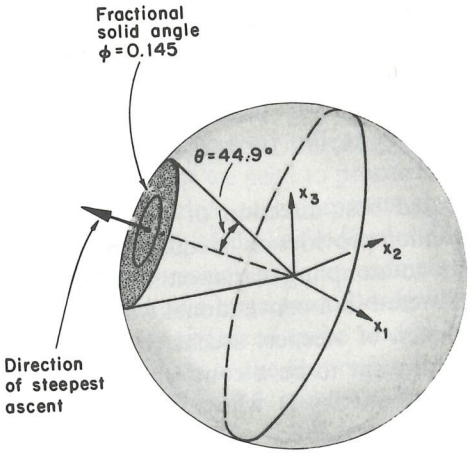

FIGURE 6.3. Direction of steepest ascent with confidence cone around it.

As an illustration we can again use the process example of Section 6.2. For simplicity and for the time being, we ignore the last two variables and treat the problem as if there had been only three variables ($k = 3$). The confidence region then takes the form of a cone in three dimensions about the estimated direction of steepest ascent as in Figure 6.3. For this reduced example, we have

$$\sum b_i^2 = (-3.35)^2 + (-2.16)^2 + (0.28)^2 = 15.9665,$$

$$\sum b_i\delta_i = -3.35\delta_1 - 2.16\delta_2 + 0.28\delta_3,$$

$$\sum \delta_i^2 = 1,$$

$$(k-1) = 2, \qquad s_b^2 = 0.97, \qquad F_{0.05}(2,10) = 4.10, \tag{6.4.5}$$

and the confidence region is thus defined by the cone with apex at the origin and such that all points $(\delta_1, \delta_2, \delta_3)$ a unit distance away from the origin satisfy

$$15.9665 - (-3.35\delta_1 - 2.16\delta_2 + 0.28\delta_3)^2 \le 2 \times 0.97 \times 4.10, \tag{6.4.6}$$

that is,

$$(-3.35\delta_1 - 2.16\delta_2 + 0.28\delta_3)^2 \ge 8.0125. \tag{6.4.7}$$

When there are more than three variables, the preparation of a diagram like that of Figure 6.3 is not possible. However, in practical applications, all we usually need to know is whether the direction of steepest ascent has been determined accurately enough for us to proceed. A good indication is supplied by the magnitude of the solid angle of the confidence cone about the estimated vector. Consider again Figure 6.3. Suppose the cap on the unit sphere centered at the origin has a surface which is a fraction ϕ of the total area of the sphere; then we shall say that the confidence cone subtends a fractional solid angle ϕ. Associated with the confidence cone we might also consider another angle, the semiplane angle θ at the vertex between a line on the surface of the cone at the origin and the axis of the cone. For a given number of dimensions, there will be a one-to-one correspondence between the fractional solid angle ϕ and the semiplane angle θ at the vertex.

This correspondence is given by tables of the t distribution. These tables are equivalent to a listing of the fractional solid angle 2ϕ (which corresponds to the double-tailed probability in the tables) in terms of the function $t = (k-1)^{1/2}\cos\theta$. The fractional solid angle ϕ may therefore be obtained by inverse interpolation in the tables of t. Specifically, in terms of the quantities we have defined,

$$\sin\theta = \left\{\frac{(k-1)s_b^2 F_\alpha(k-1, \nu_2)}{\sum_{i=1}^{k} b_i^2}\right\}^{1/2} \tag{6.4.8}$$

and

$$t_{2\phi}(k-1) = (k-1)^{1/2}\cot\theta$$

$$= \left\{\frac{\sum_{i=1}^{k} b_i^2}{s_b^2 F_\alpha(k-1, \nu_2)} - (k-1)\right\}^{1/2} \tag{6.4.9}$$

In the simplified example above in which we have pretended that there were only three variables we have

$$\sin^2\theta = \frac{\{2F_{0.05}(2,10)\}0.97}{15.9665} = \frac{2(4.10)(0.97)}{(15.9665)} = 0.4982, \tag{6.4.10}$$

whence $\theta = 44.9°$. Thus, the confidence region for this example consists of a

cone about the direction of steepest ascent with a semiplane angle at the vertex of $\theta = 44.9°$. Applying Eq. (6.4.9) with $\theta = 44.9°$, $k = 3$, we have, since $\cot\theta = 1.003$,

$$t_{2\phi}(2) = 1.418. \tag{6.4.11}$$

Interpolation in the t tables gives $\phi = 0.145$ approximately. Thus the 95% confidence cone excludes about 85.5% of the possible directions of advance.

This kind of statement is available to us however many variables are involved. Thus, without necessarily being able to visualize the situation geometrically, we can appreciate the size of the confidence region simply by calculating ϕ.

For the original five-variable example, we would set $k = 5$, $\nu_2 = 10$ to obtain

$$t_{2\phi}(4) = \sqrt{\frac{59.8690}{0.97 \times 3.48} - 4} = 3.71. \tag{6.4.12}$$

Interpolation in the t tables yields $\phi = 0.01$ approximately. Thus, in this instance the 95% confidence cone for the direction of steepest ascent excludes about 99% of possible directions of advance. The appropriate direction is, therefore, known with some considerable accuracy.

6.5. STEEPEST ASCENT SUBJECT TO A CONSTRAINT

In some problems, we cannot proceed very far in the direction of steepest ascent before some constraint is encountered. It is then of interest to explore the modified direction of steepest ascent subject to the condition that the constraint is not violated. For example, in Figure 6.4 we show a 2^3 factorial experiment in the space of x_1, x_2, and x_3. As usual, x_1, x_2, and x_3 are standardized and relate to real variables such as temperature, ξ_1, time, ξ_2, and so on, by such equations as $x_i = (\xi_i - \xi_{i0})/S_i$. It may happen that the path of steepest ascent leads to a region of inoperability. Suppose that, at least as a local approximation, the inoperability region is bounded by the plane

$$a_0 + a_1x_1 + a_2x_2 + \cdots + a_kx_k = 0. \tag{6.5.1}$$

A vector perpendicular to this plane is then $\mathbf{a} = (a_1, a_2, \ldots, a_k)'$. Let the unmodified vector of steepest ascent be given by $\mathbf{b} = (b_1, b_2, \ldots, b_k)'$. Then the ith element of the modified vector direction of steepest ascent is

$$e_i = b_i - ca_i \qquad i = 1, 2, \ldots, k, \tag{6.5.2}$$

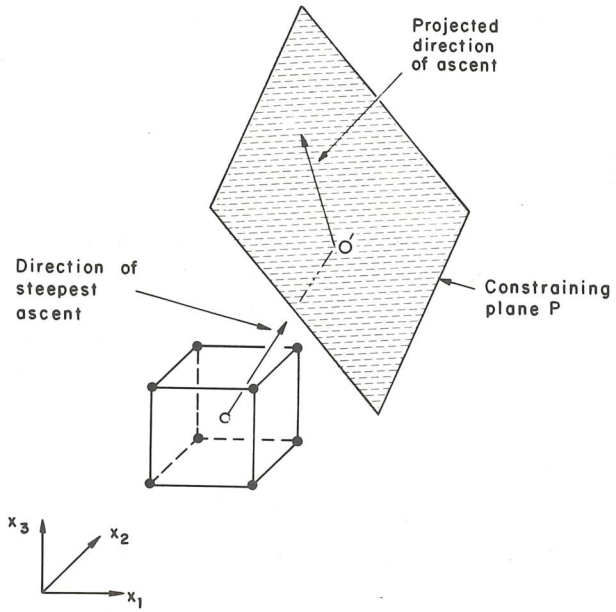

FIGURE 6.4. Steepest ascent subject to a constraint.

where

$$c = \frac{\sum b_i a_i}{\sum a_i^2}. \tag{6.5.3}$$

It will be recognized that the e_i are the residuals obtained by regressing the elements of the steepest ascent vector on to the coefficients which represent the plane, the quantity c being the appropriate estimated regression coefficient. The above calculation gives the *direction* of the modified vector only. We also need to calculate the point O' where the initial vector of steepest ascent hits the constraining plane P and the new direction takes over. Clearly any point on the initial direction of steepest ascent will have coordinates

$$x_1 = \lambda b_1, \qquad x_2 = \lambda b_2, \qquad \ldots, x_k = \lambda b_k, \tag{6.5.4}$$

for some λ. Consequently, the required point is that which, while on the path of steepest ascent, also lies on the plane $a_0 + a_1x_1 + a_2x_2 + \cdots + a_kx_k = 0$. Thus, it corresponds to the value λ_0 which satisfies

$a_0 + (a_1b_1 + a_2b_2 + \cdots + a_kb_k)\lambda_0 = 0$, that is,

$$\lambda_0 = \frac{-a_0}{\sum a_i b_i}. \tag{6.5.5}$$

An Example

The activity of a certain chemical mixture depends upon the proportion of three major ingredients A, B, and C. The mixture usually also contains an inactive diluent which is added to bring the total volume to a fixed value. A 2^3 factorial design was run to determine the effects of the ingredients on the activity y of the mixture. The lower and upper levels of the ingredients A, B, and C were as follows:

Ingredients	Amount (cm^3) Lower	Upper
A	15	25
B	10	30
C	20	40

(6.5.6)

In each case sufficient diluent was added to bring up the total volume to 100 cm^3. Thus our standardized variables are

$$x_1 = \frac{(A-20)}{5}; \qquad x_2 = \frac{(B-20)}{10}, \qquad x_3 = \frac{(C-30)}{10}. \tag{6.5.7}$$

In this example, the interactions were found to be fairly small compared with the main effects and their standard errors, and the activity y of the resulting mixture was adequately represented by the linear relationship

$$\hat{y} = 12.2 + 3.8x_1 + 2.6x_2 + 1.9x_3. \tag{6.5.8}$$

Thus, the estimated direction of steepest ascent runs from the origin $x_1 = x_2 = x_3 = 0$ along the vector (3.8, 2.6, 1.9). In the original variables, the origin corresponds to the mixture of 20 cm^3 of A, 20 cm^3 of B, and 30 cm^3 of C and, if we follow the direction of steepest ascent from this point, we shall clearly need to keep increasing the amounts of all the active

ingredients. However, we are working with a system having the constraint

$$A + B + C \le 100. \tag{6.5.9}$$

Because $A = 20 + 5x_1$, $B = 20 + 10x_2$, and $C = 30 + 10x_3$, this implies that

$$20 + 5x_1 + 20 + 10x_2 + 30 + 10x_3 \le 100. \tag{6.5.10}$$

Thus, the constraining plane has the equation

$$-30 + 5x_1 + 10x_2 + 10x_3 = 0. \tag{6.5.11}$$

It follows that

$$\lambda_0 = -(-30)/\{5(3.8) + 10(2.6) + 10(1.9)\} = 0.469. \tag{6.5.12}$$

Thus the path of steepest ascent just meets the constraining plane at the point where

$$x_1 = 0.469(3.8) = 1.78, \qquad x_2 = 0.469(2.6) = 1.22,$$
$$\text{and} \quad x_3 = 0.469(1.9) = 0.89, \tag{6.5.13}$$

that is, at the point O' with coded coordinates (1.78, 1.22, 0.89). To check, we note that

$$5(1.78) + 10(1.22) + 10(0.89) = 30.00 \tag{6.5.14}$$

so that this point is therefore just on the constraining plane. We now wish to calculate the modified direction of steepest ascent which is the projection of this direction in the plane $-30 + 5x_1 + 10x_2 + 10x_3 = 0$. Regressing the initial steepest ascent vector on to the coefficients of the constraining plane, we find that

$$c = \frac{3.8(5) + 2.6(10) + 1.9(10)}{5^2 + 10^2 + 10^2} = 0.284. \tag{6.5.15}$$

Thus the modified vector direction of steepest ascent from O' has elements

$$e_1 = 3.8 - 0.284(5) = 2.38,$$
$$e_2 = 2.6 - 0.284(10) = -0.24,$$
$$e_3 = 1.9 - 0.284(10) = -0.94. \tag{6.5.16}$$

Further progress in this direction of constrained steepest ascent will therefore lie along the line having coded coordinates

$$x_1 = 1.78 + \mu(2.38), \qquad x_2 = 1.22 - \mu(0.24),$$

$$\text{and} \quad x_3 = 0.89 - \mu(0.94), \tag{6.5.17}$$

where μ is some suitably chosen multiplier.

Some specimen calculations along the unmodified direction of steepest ascent and then along the modified direction of steepest ascent are shown in Table 6.5.

The first set of conditions listed in the table is for the center point. Conditions two and three lie on the steepest ascent path $x_1 = 3.8\lambda$, $x_2 = 2.6\lambda$, and $x_3 = 1.9\lambda$ where λ is 0.2 and 0.4. At the fourth set of conditions, the steepest ascent vector reaches the constraining plane ($\lambda = \lambda_0 = 0.469$). The remaining conditions listed are on the modified direction of steepest ascent defined in the foregoing display with $\mu = 0.2$ and 0.4. It will be noticed that the predicted activity continues to rise, but at a considerably reduced rate, on the modified path of ascent.

The elements comprising the vector of the path of steepest ascent are, of course, often subject to fairly large errors and, in any case, the linear approximation is probably inaccurate at points not close to the region of the design. The type of calculation given above should not therefore be regarded as an exact prediction but rather as supplying an experimental path worthy of further exploration. In some cases the constraint is not known exactly *a priori*, as in this example, but must also be estimated. It often happens that, in a given set of experimental runs, a number of responses may be measured, one of which, the principal response function, represents an objective function such as yield which it is desired to maximize and the

TABLE 6.5. Points on the original and modified directions of steepest ascent

Point number	Total (cm^3)	x_1	x_2	x_3	Predicted activity	Comments
1	70	0	0	0	12.2	Center O
2	82.8	0.76	0.52	0.38	17.2	Unconstrained
3	95.6	1.52	1.04	0.76	22.1	path
4	100.0	1.78	1.22	0.89	23.8	Point O'
5	100.0	2.26	1.17	0.70	25.2	Constrained
6	99.95	2.73	1.12	0.51	26.5	path

others are measures of physical characteristics of the product—purity, color, viscosity, odor, and so on. In those cases where the objective of the experiment is to maximize the principal response, the constraining relations may themselves be estimated from the auxiliary responses. We have seen an example of this situation earlier where we were required to maximize the strength of a dyestuff, subject to the attainment of suitable levels of hue and brightness.

To understand constraining relations between three variables, it is very worthwhile to examine appropriate sets of three-dimensional models. Computer graphics have now greatly simplified such exploration. One can say, of course, that the geometrical models do no more than express the related algebra and so consequently algebraic manipulation without the accompanying models is enough. In practice, however, it seems that there is much less likelihood of missing important aspects of the problem and of making outright mistakes, if graphical and geometrical illustrations are used whenever possible.

For an excellent pictorial representation of steepest ascent using five-factor simplex designs, see Gardiner and Cowser (1961).

APPENDIX 6A. EFFECTS OF CHANGES IN SCALES ON STEEPEST ASCENT

In this book, we have tried to emphasize that the outcome of an investigation depends critically on a number of crucial questions which are not directly dependent on the data and which the experimenter must decide from knowledge and experience. Moreover, these decisions are typically modified as the investigation unfolds. For example, the experimenter decides *which* factors to include, the *region* of the factor space in which the experiments are to be carried out, and the *relative scales* of measurement of the factors. Such questions are quite outside the competence of the statistician, although he ought always to ask appropriately probing questions to ensure that the experimenter has properly considered all the relevant issues, and the statistician may need to point out, after the data are available, that certain suppositions of the experimenter appear to be of doubtful validity. Granted all this, it is the role of the statistician to design the best experiment within the framework currently believed to be most appropriate by the experimenter and, in particular, in the scaling of the variables the experimenter regards as most applicable at any given time. It has to be remembered that a subject as concrete and mathematically satisfying as experimental design is actually embedded in a morass of uncertainty, uncertainty due to the possibilities that the experimenter might choose wrong variables, might explore the wrong region, or might use scaling that was inappropriate. However, the gloomy view that successful experimentation is a matter of purest luck is lightened by two circumstances; first, that experimenters often *do*

know a very great deal about the system they are studying and second, that, because most investigations are conducted sequentially, the experimenter does not need to guess exactly right but need only guess sufficiently right to place himself on one of the many possible paths that will lead adaptively to a satisfactory answer. In particular, the correction of any grossly unsuitable scaling of the predictor variables will normally occur as the experimental iteration proceeds.

The problem of scaling the variables in steepest ascent is one area where the general indeterminacy of experimentation does not lie conveniently hidden but makes itself manifest. In order to appreciate the problem, consider the following: Suppose the scale factors that one experimenter adopts are $s_1, s_2, \ldots, s_k$ and that the scale factors adopted by a second experimenter are $s_1' = a_1 s_1$, $s_2' = a_2 s_2, \ldots, s_k' = a_k s_k$. (For example, if in a two-level factorial or fractional factorial design, one experimenter changed temperature by $s_1 = 20°$ and another by $s_1' = 10°$, then $a_1 = s_1'/s_1 = 0.5$.) Suppose further that the direction of steepest ascent as measured in the first experimenter's scales has direction cosines proportional to the elements $1, 1, \ldots, 1$. (We can make this assumption without loss of generality because we could always rotate the axes of our x space so that the direction, whatever it may be, lies in this particular direction in new coordinates.) In this case, the direction of steepest ascent appropriate to the second experimenter would have direction cosines proportional to $a_1^2, a_2^2, \ldots, a_k^2$. Figure 6A.1 shows the situation for $k = 3$.

The fact that the direction of steepest ascent was not invariant to scale change was pointed out in the original response surface paper by Box and Wilson (1951). In the discussion of the paper, this point was taken up by N. L. Johnson who emphasized the dependence of the elements of the direction on the square of the

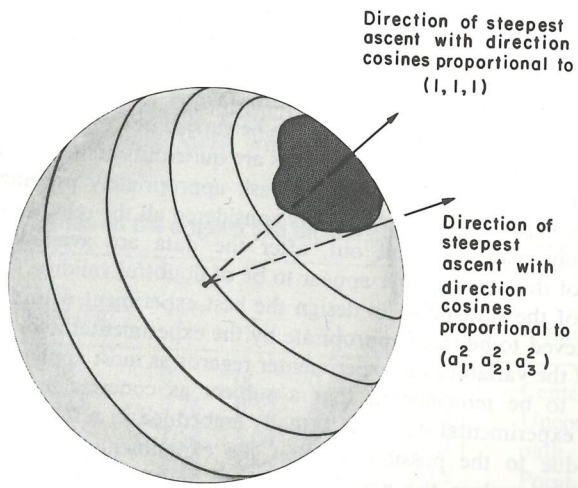

FIGURE 6A.1. Two directions of steepest ascent, one of which traces a closed curve on the unit sphere about the other (see text).

amount by which the scale factor was altered. In this same discussion it was also argued that it might be better simply to rely on the signs of the effects, which are of course scale invariant.

We can, to some extent, answer the question of how much *additional* useful information is contained in the direction of steepest ascent compared with using only the signs of the effects by examining a brief, unpublished investigation undertaken by G. E. P. Box and G. A. Coutie some years ago.

Let us choose some measure $G(|a|)$ of how widely different the second experimenter's scales are from those adopted by the first experimenter. Because multiplying all the a's by some constant factor will not change the actual situation, we need to choose, for G, some suitable homogeneous function of degree zero in the a's. One such suitable function is $G = |a|_{\max}/|a|_{\min}$. If we adopt this criterion, we shall regard all values of the a's that produce the same value, G_0 say, of G as representing a degree of scaling equally discordant with the original one. Such values of the a's will trace out a closed curve on the unit sphere as shown in Figure 6A.1. Thus, a measure of how much uncertainty is introduced in the direction of steepest ascent for a scaling as discordant as, or less discordant than, that represented by $G = G_0$ will be represented by the content of the closed curve bounded by $G = G_0$ on the unit sphere, taken as a proportion of the whole surface of the unit sphere. In the three-dimensional case of Figure 6A.1, this is simply proportional to the surface of the cap enclosed by the curve $G = G_0$ on the unit sphere.

Now let us write $a_i^2 = y_i$, then $G^4 = a_{\max}^4/a_{\min}^4 = y_{\max}^2/y_{\min}^2$ and the inequality $G < G_0$ is equivalent to the inequality $G^4 < G_0^4$. It can be shown that the required value is

$$\phi = \frac{1}{(2\pi)^{k/2}} \int_R \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{k} y_i^2\right\} dy, \qquad (6A.1)$$

where the integral is taken over the region R such that $y_{\max}^2/y_{\min}^2 < G_0^4$. This is the same integral as would arise in a comparison of variances each having 1 df using Hartley's (1950) $F_{\max}$ criterion. H. A. David (1952) gives the upper and lower 5 and 1% points of $F_{\max}$ for a set of k variance estimates each based on ν df for $\nu \geq 2$. The test required here is for $\nu = 1$ which is not given in the tables presumably because it is usually of very little importance. The appropriate values for this integral have, however, been calculated for $\nu = 1$ to give Table 6A.1.

As an example of the use of Table 6A.1, consider the following, in which we shall leave aside the question of sampling error. Suppose we were concerned with $k = 5$ variables. Let us further suppose that experimenters may be expected to differ in their ideas about scaling by a factor G representing the ratio of the largest to the smallest modifying factor. Finally let us suppose a reference experimenter has so scaled the variables that the elements of the steepest ascent direction are proportional to $1, 1, 1, 1, 1$ in the scaling adopted. A knowledge of all the signs will reduce the uncertainty concerning the direction of advance by a factor of $1/2^5 = 1/32 = 0.03125$. If $G = 2$, the uncertainty will be reduced still further by a factor of 0.167 (obtained from Table 6A.1 with $k = 5$, $G = 2$) giving an overall uncertainty of

TABLE 6A.1. Table of the reduction of the "percentage of uncertainty" due to changes of scale in a particular *n*-tant

Value of G	Values of k				
	2	3	4	5	6
$\sqrt{2}$	41.0	15.5	5.6	2.0	0.7
2	68.8	45.4	28.0	16.7	11.6
3	85.7	69.8	55.8	43.9	33.8
4	93.8	83.8	75.1	67.7	61.4

0.167(0.03125) = 0.0052 for $G = 2$; by a factor of 0.439 if $G = 3$ giving an overall uncertainty of 0.439(0.03125) = 0.014 and by a factor of 0.677 for $G = 4$ giving an overall uncertainty of 0.677(0.03125) = 0.021. Our certainty of knowledge of the direction is thus 1 minus these figures or 0.9948, 0.986, and 0.979 for $G = 2$, 3, and 4, respectively.

In summary, the present authors feel that the direction of steepest ascent correctly distills the first degree information concerning possible directions of improvement, by combining the experimenter's prior knowledge concerning the nature of the response relationship appropriately with the available data. This direction does and should change for different experimenters because different prior opinions must enter the problem.

EXERCISES

6.1. An experimenter begins a steepest ascent procedure on two variables (X_1, X_2) at the current central point (90, 20) and performs five runs with the response results provided:

X_1	80	100	80	100	90
X_2	10	10	30	30	20
y	11	0	29	6	12

Code (X_1, X_2) sensibly to variables (x_1, x_2) and fit a first-order (planar) model $\hat{y} = b_0 + b_1x_1 + b_2x_2$ to the data. Determine the direction of steepest ascent. The experimenter now performs six more runs:

X_1	64.5	47.5	39	30.5	43.25	34.75
X_2	38	50	56	62	53	59
y	43	58	72	62	65	68

Which of these runs lie on the path of steepest ascent you determined earlier? The experimenter decides now to combine the two runs $(X_1, X_2, y) = (43.25, 53, 65)$ and $(34.75, 59, 68)$ with six more, namely these:

X_1	34.75	43.25	39	39	39	39
X_2	53	59	56	56	56	56
y	71	68	71	72	72	73

Fit a first-order model to those eight runs and use the repeat observations to test for lack of fit. Plot, as points on a diagram, all the runs performed so far, with their y values attached. What should the experimenter do next? Go off in a new direction of steepest ascent? Fit a second-order surface? Or what?

6.2. Two coded variables $x_1 = (X_1 - 99)/10$, $x_2 = (X_2 - 17)/20$ are examined, and the data of Table E6.2 are obtained. Is the point $(X_1, X_2) = (59, 67)$ on the path of steepest ascent?

TABLE E6.2

x_1	x_2	y
−1	−1	11
1	−1	2
−1	1	29
1	1	6
0	0	12

6.3. In a steepest ascent investigation to examine the effects of three variables, the initial design of Table E6.3 was employed.

The model $\hat{y} = 4.7 + 0.25x_1 + x_2 + 0.5x_3$ is fitted. The following restriction exists:

$$4X_1 + 5X_2 + 6X_3 \leq 371.$$

Is the point $(X_1, X_2, X_3) = (36.40, 27.10, 15.00)$ on the restricted path of steepest ascent? (Allow for the fact that round-off errors may occur unless adequate figures are retained.)

6.4. A mixture of three ingredients X_1, X_2, X_3 (plus inerts to make up to 100%) is being examined. Steepest ascent techniques are being applied to improve the response variable, percentage yield. The following fitted equation is obtained from the first set of runs made:

$$\hat{y} = 9.925 - 4.10x_1 - 9.25x_2 + 4.90x_3,$$

TABLE E6.3

Uncoded			Coded		
X_1	X_2	X_3	x_1	x_2	x_3
30	10	6	−1	−1	−1
40	10	6	1	−1	−1
30	20	6	−1	1	−1
40	20	6	1	1	−1
30	10	16	−1	−1	1
40	10	16	1	−1	1
30	20	16	−1	1	1
40	20	16	1	1	1
35	15	11	0	0	0
35	15	11	0	0	0
35	15	11	0	0	0
35	15	11	0	0	0

where

$$x_1 = X_1 - 8, \qquad x_2 = X_2 - 14, \qquad x_3 = 2(X_3 - 7).$$

If the restriction $X_1 + X_2 + X_3 \geq 26.4$ applies, is the point $X_1 = 7.11$, $X_2 = 11.29$, $X_3 = 8.00$ on the path of steepest ascent? (There will undoubtedly be a little rounding error in your calculations, remember.)

6.5. A four-dimensional steepest ascent path is obtained from the fitted coefficients $b_1 = 1$, $b_2 = 2$, $b_3 = 3$, $b_4 = 2$. If we have to keep inside or on the restricting boundary

$$-4 + x_1 + x_2 + x_3 + x_4 = 0,$$

answer this question: Is the point $(X_1, X_2, X_3, X_4) = (-5/2, 10, 45/2, 10)$ on the adjusted path of steepest ascent? Assume $x_i = (X_i - 5)/5$, $i = 1, 2, 3, 4$.

6.6. A manufacturer of circular saw blades has two empirical first-order equations for blade life in hours (η_1, observed as y_1) and for unit blade cost in cents (η_2; y_2) in terms of the manufacturing time (coded to x_1) and the hardness of the steel used (coded to x_2). These are

$$\hat{y}_1 = 20 + x_1 + x_2,$$

$$\hat{y}_2 = 50 + 4x_1 + 2x_2,$$

and they are valid only for $-2 \leq x_i \leq 2$. If the unit blade cost must be kept below 54 cents and the blade life must exceed 21.5 h, can the manufacturer produce blades? If yes, explain how he must operate. If no, explain why he cannot.

CHAPTER 7

Fitting Second-Order Models

In Chapter 6 we saw how, in appropriate situations, a fitted first degree approximation to the response function could be exploited to produce process improvement. In this chapter, we begin to consider the use of quadratic (second-order) approximating functions.

7.1. FITTING AND CHECKING SECOND DEGREE POLYNOMIAL GRADUATING FUNCTIONS

To fix ideas, we discuss in more detail an example first introduced in Chapter 2 concerning the behavior of worsted yarn under cycles of repeated loading. For simplicity in the earlier analysis, only 8 of the full set of 27 runs were utilized. These 8 runs are distinguished by superscript b's in Table 7.1, where the complete data are set out.

Reconsideration of the Textile Example

It will be recalled that, in this example, the three inputs were

$$\xi_1 = \text{the length of test specimen (mm)},$$

$$\xi_2 = \text{the amplitude of load cycle (mm)},$$

$$\xi_3 = \text{the load (g)},$$

which were conveniently coded in terms of

$$x_1 = \frac{(\xi_1 - 300)}{50}, \qquad x_2 = (\xi_2 - 9), \qquad x_3 = \frac{(\xi_3 - 45)}{5}. \quad (7.1.1)$$