DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin

Technical Report No. 25

March 1964

Some Notes on Non-Linear Estimation

by

George E.P. Box

These notes on non-linear estimation were written in 1956 and were informally circulated at Princeton. We have had many requests for copies and it seems worthwhile to reissue this as a technical report.

SOME NOTES ON NON-LINEAR ESTIMATION

G. E. P. Box

In a recent paper (ref. 1) the iterative nature of experimentation has been emphasised. In this process of iteration there are two essential components:-

(a) The devising of experiments suggested by the investigator's appreciation of the situation to date and designed to elucidate it further;

(b) the examination of results of experiments performed to date in the light of all background knowledge available, with the object of postulating theories susceptible of test in future experimentation.

The main contribution of statistics to experimentation is in the provision of tools to assist the experimentor in these tasks.

Now the iteration which occurs in experimental investigations goes on simultaneously in a number of different spheres. Thus there will usually be a movement in the space of the factors in the location of the centre of interest as the experiments proceed to regions of better experimental conditions. Simultaneously there will tend to be a steady improvement in the choice of relative scales, metrics and transformations on the basis of which the variables are considered, and in the type of design which is used to study the relationship. Finally as experimentation proceeds a better understanding of the theoretical basis of the equation under study may result. An example of such a development is given in ref. 1. It is mainly with the last type of iteration that we are concerned in the present paper.

Frequently, either after a period of empirical experimentation (as is the case in ref. 1) or initially, some theoretical mechanism for the process under study may be postulated. (It should be clearly understood that we are not here necessarily concerned with an exact theoretical picture; it is doubtful whether any physical system has ever been described exactly. What is required is a description of the mechanism

sufficiently close to the truth to assist understanding and to allow useful, but not necessarily exact, extrapolation.) Suppose that for some system under study a theoretical mechanism has been put forward whose consequences can be expressed in terms of the functional relationships

$$\eta_r = f_r(\xi_1, \xi_2, \ldots, \xi_k; \theta_1, \theta_2, \ldots, \theta_p) \qquad (r = 1, 2, \ldots, m) \qquad \ldots\ldots 1$$

connecting responses $\eta_1, \eta_2, \ldots, \eta_m$ with a number of variables $\xi_1, \xi_2, \ldots, \xi_k$ in which $\theta_1, \theta_2, \ldots, \theta_p$ are unknown parameters. For instance in chemical problems $\eta_1, \eta_2, \ldots, \eta_m$ may be the yields of various substances resulting from a chemical reaction; $\xi_1, \xi_2, \ldots, \xi_k$ operating variables such as time, temperature and concentration; and $\theta_1, \theta_2, \ldots, \theta_p$ chemical rate constants, activation energies, etc.

Occasionally the functions of equation (1) can be obtained explicitly, but more frequently no such expression is available and the functions are described by differential equations or in some other implicit form. A very important tool, therefore, to enable progress concerning the theoretical nature of the system to be made would be one which, given a set of experimental data, would allow the adequacy of the assumed mechanism in describing the data to be checked and which, on the assumption that the representational adequacy of the system was satisfactory, would enable estimates of the unknown parameters and their standard errors to be determined.

Now where the functional form can be written explicitly and is linear in the unknown constants $\theta_1, \theta_2, \ldots, \theta_p$ (as in the case, for example, where this function is a polynomial in the ξ's with coefficients

$\theta_1, \theta_2, \ldots, \theta_p$) well known methods are available for doing all the things mentioned above. In particular estimates of the parameters and their standard errors may be obtained by Gauss's method of least squares, or equivalently, if we assume normality by the method of maximum likelihood. Again assuming normality, a test of the adequacy of the assumed model may also be made by comparing the size of the residual sum of squares with some measure of pure error such as that obtainable from randomised replication. However, in common situations, even when the functional form is known explicitly, it is seldom linear in the parameters and more frequently it is not explicitly known.

A method has recently been described (ref. 2) whereby using numerical methods on a digital computer a test of the adequacy of the assumed model and estimates of the parameters and their standard errors can be obtained for virtually any system whether the equations are explicit or not. The way in which this method works is perhaps best illustrated by an example. A fuller account will be found in ref. 2.

Example 1.

This concerns a consecutive chemical reaction in which a reactant M_1 (concentration η_1) changed at a rate proportional to its concentration to M_2 which then changed at a different rate to M_3.

$$M_1 \xrightarrow{k_1} M_2, \quad M_2 \xrightarrow{k_2} M_3$$

This can be represented by the simultaneous differential equations

$$\frac{d\eta_1}{dt} = -k_1\eta_1$$

$$\frac{d\eta_2}{dt} = k_1\eta_1 - k_2\eta_2 \qquad \ldots\ldots 2$$

$$\frac{d\eta_3}{dt} = k_2\eta_2$$

with boundary conditions $\eta_1 = 100\%$, $\eta_2 = 0$, $\eta_3 = 0$ at time $t = 0$. Six duplicate experiments were performed in random order in which measurements y_2 of η_2 were made at the times specified in Table 1.

Table 1.

(time in mins.)	10	20	40	80	180	320
Yield of M_2 (y_2)	19.2	14.4	42.3	42.1	40.7	27.1
	14.0	24.0	30.8	40.5	46.4	22.3

It was required

(a) to ascertain whether the form of the differential equations was consistent with the data;

(b) to estimate the parameters k_1 and k_2; and

(c) to estimate the accuracy with which these values were determined.

The method used was as follows: Guessed values of the unknown constants k_1 and k_2 were substituted in the differential equations, which were then integrated using the step-by-step method due to Runge and Kutta, as described by Gill (ref. 3). By this means 6 values $\eta(k_1, k_2)$ of the yields at the 6 times were calculated. The extent by which these values differed from the observed values was measured by the sum of squares of discrepancies,

$S^2(k_1, k_2) = \Sigma [y-\eta (k_1, k_2)]^2$. The problem was then that of proceeding via a series of trial values to these quantities k_1, k_2 which made the sum of square S^2 a minumum. This was done by a method precisely analogous to that which has been employed in the exploration of response surfaces except that calculations on the computer were substituted for experiments. Initially a first-order design of trial parameter values was carried out, and if it appeared that first-order effects were dominant the direction of steepest descent was followed till no further decrease in the sum of squares of discrepancies was obtained. The process was then repeated. When a point was reached at which it appeared that

first-order effects were no longer large compared with those of second-order, a second-order design was carried out. From this either immediately or after one or two further iterations, which might involve transformation, the minimum was reached. The course of such a calculation is shown in figure 1. On this figure the approximate contours of S^2 were also shown. (These would not normally of course be known. They have been obtained in this instance by carrying out a grid of computations and drawing smooth curves through the points.)

The grid on which these contours are based is as follows:-

Values of S^2 for various θ_1, θ_2

$\theta_1 = 3 + \log k_1$

$\theta_2 = 3 + \log k_2$	0.75	1.00	1.25	1.5
1.25	2,232.09	1,372.88	1,450.04	2,337.43
1.00	1,290.10	450.84	698.48	1,750.89
0.75	790.15	138.82	721.91	2,156.28
0.50	1,034.53	864.87	1,921.95	3,748.56

Conditioning.

In this particular example the method is seen to converge rapidly, as is to be expected from the nature of the surface of sums of squares of discrepancies, which can be seen to be very well-conditioned. This state of affairs cannot be expected to be general, however, and in many problems if the sum of squares surface were plotted in terms of the original parameters, extensive ridges would be found.

An example of a poorly conditioned surface arises from consideration of the simple case where it is desired to fit the expression $y = b_0 + b_1 x$. Suppose $x = 9, 10, 11$; then the matrix of independent variables will be

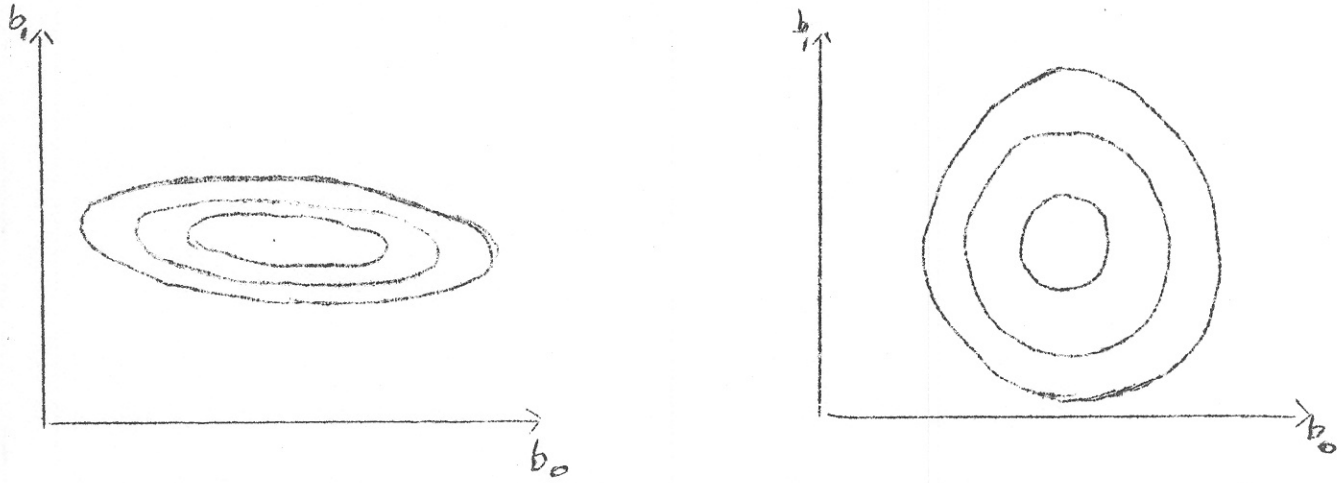
$$\underline{X} = \begin{bmatrix} 1 & 9 \\ 1 & 10 \\ 1 & 11 \end{bmatrix}, \text{ so that } \underline{X}'\underline{X} = \begin{bmatrix} 3 & 30 \\ 30 & 302 \end{bmatrix}.$$ Thus $\underline{X}'\underline{X}$ is very nearly singular, and the sum of the squares surface

$$S^2 = S^2_{min} + n(b_0 - \hat{b}_0)^2 + 2\,\Sigma x(b_0 - \hat{b}_0)(b_1 - \hat{b}_1) + \Sigma x^2(b_1 - \hat{b}_1)^2$$

consists of a very long attenuated ridge. If an attempt was made to minimize by first-order steepest descent, or by "one-factor-at-a-time," one would have an extremely long and painful task. However if the model is rewritten as

$$\begin{aligned} y &= (b_0 + b_1\bar{x}) + b_1(x - \bar{x}) \\ &= A + b\,x \end{aligned}$$

then for this model $\underline{X}'\underline{X} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ and almost any system of numerical minimization would rapidly converge. Diagrams showing the appearance of the sums of squares surfaces in the two cases are shown below.



The above example is a linear one, but suppose one were fitting $y = ae^{bx}$. This could be written in the alternative form

$$\ln y = \ln a + bx$$

If the coefficient of variation of y was not large, then the S-surface for

ln y would be similar to the S-surface for y. Unless the mean of x was close to zero, therefore, a ridge-like minimum would result as before. When fitting $y = ae^{bx}$ and other similar expressions, it would usually be better to consider $y = Ae^{b(x-\bar{x})}$ where $A = ae^{b\bar{x}}$. In general, whenever this sort of expression occurs it will be better to work in terms of transformed parameters. However, cases will frequently arise where the nature of the transformation that should be employed to enable the surface to be thrown into a state of satisfactory conditioning will not be apparent.

One method for finding a linear transformation in which the surface will, locally at least, be better conditioned, is to fit a second-degree equation and then to define new parameters $\gamma_1, \gamma_2, \ldots, \gamma_p$, which are orthogonal linear functions of $\theta_1, \theta_2, \ldots, \theta_p$, chosen so that the second degree equation can be written

$$S^2 = S^2_{min} + \sum_{i=1}^{p} \gamma_i^2 \quad .$$

This can, of course, be done in a number of different ways (Gram-Schmidt, canonical reduction, etc.).

In the example above of fitting $y = a + bx$ (where, of course, this procedure will be quite unnecessary) the transformed parameters obtained by this method would be <u>equivalent</u> to $b_0 + b\bar{x}$ and b. (Equivalent to and not equal to because the transfor nation is obviously not unique, and any orthogonal transformation of the canonical variables, of which the transformation to $b_0 + bx$ and b would be one, would do as well.) We could condition the surface by this method, and then proceed by means of steepest descent till we came close to the minimum or ran into trouble.

If a second-degree fitting method is to be used, however, alternatives would be either to follow the <u>second-order</u> formula for steepest descent, or to use the second-degree equation to estimate the minimum direct. An iterative procedure could be based on either method, but the second is

probably better since it is less dependent upon choice of scales.

Concerning this latter method, in which the second-degree equation is used to estimate the minimum it is probable that, where the minimum is far from the region of the "design," its direction is indicated more precisely than its position. For this reason a procedure may be used in which exploratory points along the direction are computed and the estimated minimum used as a new starting point. It is convenient to employ three such points which can consist of a point at the predicted minimum, a point half-way and a point the same distance on the other side. Interpolation between these values provides a starting point for a new second-order design. It is, of course, not necessary to employ designs with minimum variance properties in this connection. It is simpler to use those illustrated below. The designs are of 1st and 2nd order respectively and are for two parameters; the extension to other cases will be obvious.

1st order	2nd order
0 0	0 0
1 0	1 0
0 1	0 1
	2 0
	0 2
	1 1

The latter design was used in what was probably the first example of this type of technique in ref. 4.

Example 2.

This concerns the fitting of the equation $y = ce^{kt}$. Again in order to draw the type of surface encountered, a grid of calculations has been performed. It was assumed that the observations of y at times

	t =	.125	.25	.5	1.0	2.0	4.0
were	y =	2.093	1.650	1.717	1.245	.683	.285

The sums of squares of descrepancies for various values of c and k are then:-

x_2	c	x_1 k	-2 .25	-1 .3536	0 .50	1 .7071	2 1.0
2	2.8280		4583.6	2902.0	1624.1	868.1	652.7
1	2.3782		1825.3	893.3	316.8	163.9	441.9
0	2.0		658.4	224.4	89.7	300.7	862.4
-1	1.6817		502.7	382.0	514.8	934.4	1636.6
-2	1.4140		952.0	1023.8	1312.7	1833.3	2589.7

The contours of the surface obtained from this grid are shown in figure 4, together with the second-order designs employed. For convenience the calculations are conducted in terms of the standardized parameters x_1 and x_2 where

$$x_1 = 6.64 \log k + 2$$
$$x_2 = 13.28 \log c - 4$$

The progress of the calculations is shown below.

Starting at the point $k = .7071$, i.e. $x_1 = 1$, where $S^2 = 163.9$, $c = 2.3782$, $x_2 = 1$

a second-order design of the type described in the previous section was calculated. The values of S^2 were:-

x_2	x_1 1.0	1.1	1.2
1.2	224.4		
1.1	189.8	193.8	
1.0	163.9	172.7	185.8

From which the estimated minimum was calculated to be at the point (-0.3293, -0.0242). Values of S^2 were calculated along the line passing through this

point as follows:-

x_1	1	0.3353	-0.3293	-0.9940
x_2	1	0.4879	-0.0241	-0.5363
S^2	163.9	71.4	97.4	230.3

By interpolation with these 4 values, the lowest value of S^2 on the line was found to be $S^2 = 67.2$, at the point $x_1 = .158$, i.e. at $k = .528$
$x_2 = .351$, $c = 2.125$.

A new second-order design was now carried out with this point as base-point, and gave an estimated minimum at the point $x_1 = 0.2183$. Values $x_2 = 0.2911$
S^2 calculated along the line to this point were:-

x_1	.1581	.1882	.2183	.2484
x_2	.3513	.3212	.2911	.2610
S^2	67.48	66.13	66.84	69.15

By interpolation from these 4 values, the lowest value of S^2 on this line was found to be $S^2 = 66.11$, at the point $x_1 = .1918$, i.e. $k = .5344$
$x_2 = .3176$, $c = 2.1131$.

An Alternative Procedure

Consider fitting a function relationship

$$\eta_u = \eta(\xi_u , \theta)$$

where θ is to be estimated. Then if there are N observations, $u = 1, 2, \ldots, N$, the above equation will define the locus of all possible solutions in the N-dimensional space of the η_u . For example, if $N = 3$, the locus might appear like that shown in figure 5.

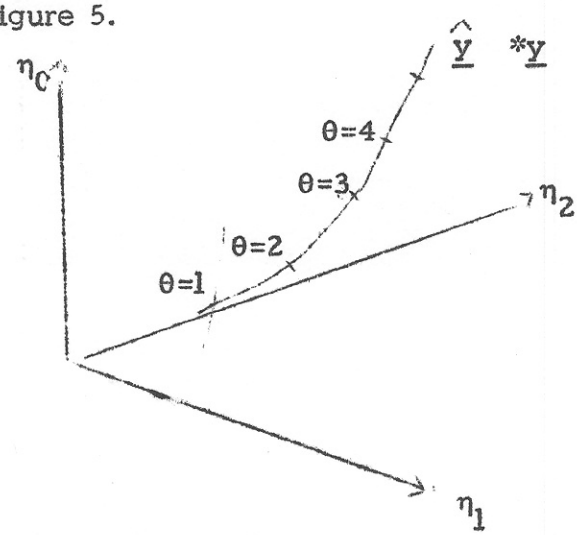
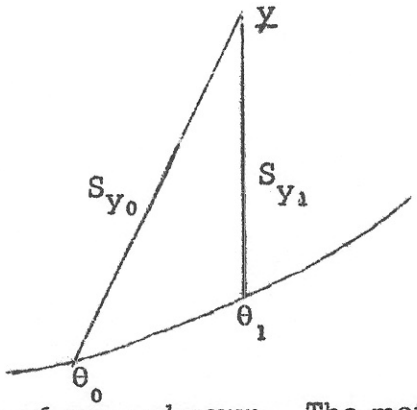


Fig. 5

We have a set of N observations defining a point $\underline{y}$, indicated by the star in figure 5, and according to the principle of least squares we adopt that estimate $\hat{\theta}$ of θ corresponding to the point $\hat{\underline{y}}$ on the line which is such that the distance between the points $\underline{y}$ and $\hat{\underline{y}}$ is as small as possible, where $\hat{\underline{y}} = \{y_u\}$ and $\hat{y}_u = \eta(x_u, \hat{\theta})$.

In the case of linear estimation the solution locus is a straight line (or for more than one parameter a plane or hyperplane) and the contours of constant θ in this line (plane or hyperplane) are equidistant. In the non-linear case this is no longer true, but we proceed by supposing it to be approximately true locally.



Consider the simple case of one unknown. The method of solution outlined in ref. 2 used two procedures which we can call the first-order (steepest descent) procedure and the second-order procedure.

(a) The first-order steepest descent procedure would involve the determination of at least two values of S (or for k parameters, k + 1 values of S). From this we could determine which to move, but not how far to move.

(b) The second-order procedure would involve determination of at least 3 values of S (or for k parameters, $1/2(k + 1)(k + 2)$ values). From this we could determine the position of the minimum (exactly if the problem were really linear, approximately otherwise).

Both procedures use only the distances between the point $\underline{y}$ and the

points corresponding to the θ's; not only do we know these distances but we also know the distance between the parameter points. Thus knowledge of the location of the points y, θ_0 and θ_1 enable us to determine the three distances, S_{y0}, S_{y1} and S_{01}. For

$$S_{y0} = \{\Sigma [y - \eta(\theta_0)]^2\}^{1/2}$$

$$S_{y1} = \{\Sigma [y - \eta(\theta_1)]^2\}^{1/2}$$

$$S_{01} = \{\Sigma [\eta(\theta_0) - \eta(\theta_1)]^2\}^{1/2}$$

and making the assumption of linearity knowledge of these quantities is clearly suffiecient to fix the position of $\underline{y}$ relative to the solution locus, and consequently to provide an estimate $\hat{\theta}$. In the general case of k parameters, from (k + 1) points calculated on the solution locus there will be $1/2(k + 1)(k + 2)$ quantities available, viz. the (k + 1) distances from $\underline{y}$ to the calculated points on the solution locus, and the $1/2k(k + 1)$ distances between them. These are just sufficient on the linear assumption to fix the position of the point $\underline{y}$ relative to the solution locus and consequently to enable the k parameters $\theta_1, \theta_2, \ldots, \theta_k$ to be estimated. It can be seen that, although this method would involve the calculation of only (k + 1) points, it should nevertheless be of comparable accuracy to that obtainable with a second-order procedure using the previous method, which would involve the calculation of $1/2(k + 1)(k + 2)$ points.

In practice the simplest designs will again be those in which a base point is first calculated and then each parameter is varied in turn. Suppose the calculated values at these points are $\eta_1, \eta_2, \ldots, \eta_k$. The situation for 3 observations and two parameters is illustrated in figure 6.

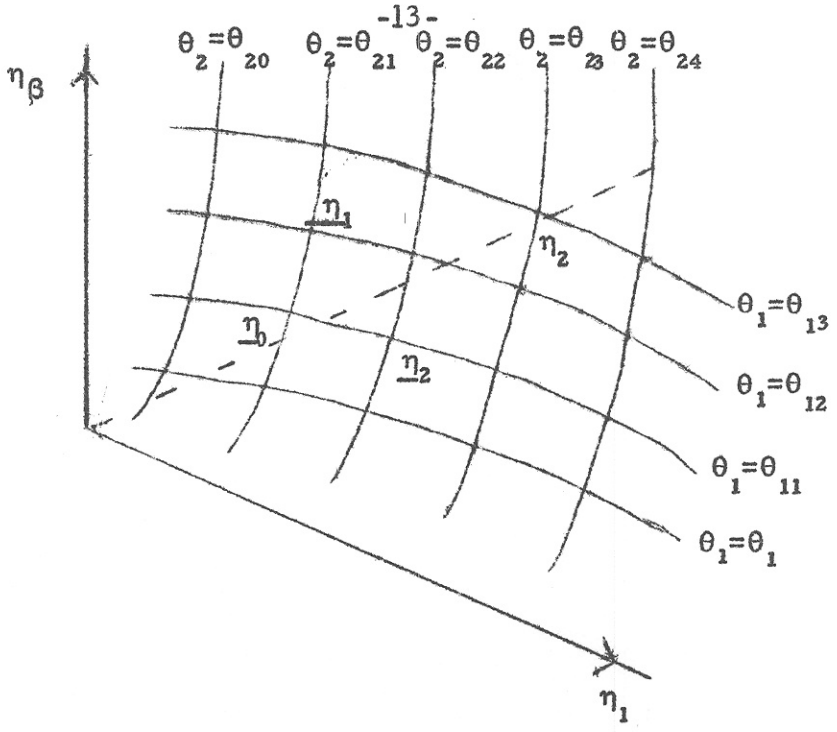


Fig. 6

The approximation that we are making is now seen to be capable of expression in a familiar form. Geometrically the method of least squares (multiple regression) consists of selecting a linear combination

$$\hat{\underline{w}} = \hat{\phi}_1 \underline{z}_1 + \hat{\phi}_2 \underline{z}_2 + \cdots + \hat{\phi}_k \underline{z}_k$$

such that the vector $\hat{\underline{w}}$ is the projection of the vector of observations $\underline{w}$ in the plane of the vectors $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$. What we require for the numerical method of minimization outlined above is to find a vector $\hat{\underline{w}}$ which is the projection of the vector $\underline{y} - \underline{\eta}_0$ on the plane of the vectors $\underline{\eta}_1 - \underline{\eta}_0, \underline{\eta}_2 - \underline{\eta}_0, \ldots, \underline{\eta}_k - \underline{\eta}_0$. Consequently we obtain what we need by performing a multiple regression of $\underline{w}$ on $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$, where $\underline{w} = \hat{\underline{y}} - \eta_0$, $z_i = \eta_i - \eta_0$ $(i = 1, 2, \ldots, k)$, the coordinates of the solution then being given by

$$\hat{\theta}_i = (1 - \hat{\phi}_i)\,\theta_i + \hat{\phi}_i \theta_i' \qquad (i = 1, 2, \ldots, k)$$

where θ_i and θ_i' are the two levels of θ_i used in the design, and $\hat{\phi}_i$ is

the ith regression coefficient. Normally this procedure will need to be repeated a number of times before sufficient accuracy is obtained. When this point is reached the approximate standard errors for the quantities $\hat{\phi}_i$, and hence for $\hat{\theta}_i$ may be calculated from the usual least squares formulas treating the $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$ as independent variables. Similarly the residual sum of squares for $\underline{w}$ will be a close estimate of the residual sum of squares for y and will be distributed approximately as $\chi^2 \sigma^2$ with $n-k$ degrees of freedom if the model is representationally adequate. (The quantity σ^2 is the experimental error variance.) When some other estimate of experimental error variance is available, comparison of the estimates provides a test of the adequacy of the model in the usual way.

As a first instance of the application of the method, consider again the example given in ref. 2. Using three of the points in the first design, and denoting the calculated value at the points (1.18, 1.18), (1.20, 1.18) and (1.18, 1.20) by $\underline{\eta}_0$, $\underline{\eta}_1$, $\underline{\eta}_2$, and the observed values by $\underline{y}$, we obtain

t (mins.)	$\underline{y}$	$\underline{\eta}_0$	$\underline{\eta}_1$	$\underline{\eta}_2$
10	16.600	13.010	13.574	12.963
20	19.200	22.365	23.253	22.206
40	36.550	33.047	34.115	32.579
80	41.300	36.077	36.719	35.066
160	43.550	21.498	21.274	20.316
320	24.700	3.817	3.573	3.412

Whence we calculate the quantities

t (mins)	$w = y - \eta_0$	$z_1 = \eta_1 - \eta_0$	$z_2 = \eta_2 - \eta_0$
10	3.590	.564	-.047
20	-3.165	.888	-.159
40	3.503	1.068	-.468
80	5.223	.642	-1.011
160	22.057	-.224	-1.182
320	20.883	-.244	-.405

and the matrix of sums of squares and products for w, z_1 and z_2 is as follows:

	w	z_1	z_2
w	985.067,001	-3.727,610	-41.114,341
z_1	-3.727,610	2.769,140	-0.952,998
z_2	-41.114,341	-0.952,998	2.829,784

whence the normal equations are:

$$2.769,140\ \phi_1 - 0.952,998\ \phi_2 = -3.727,610$$
$$-0.952,998\ \phi_1 + 2.829,784\ \phi_2 = -41.114,341$$

and $\phi_1 = -7.1783$, $\phi_2 = -16.9466$

providing a first estimate for the minimum $(\hat{\theta}_1, \hat{\theta}_2)$ of (1.0364, 0.8411).

A second design was now carried out in the region of this point. Each parameter was varied by the same amount as previously, the 3 points being (1.04, 0.84), (1.06, 0.84) and (1.04, 0.86). Denoting now the calculated values at these points by η_0, η_1 and η_2, we obtain

t (mins.)	y	η_0	η_1	η_2
10	16.600	10.020	10.473	10.003
20	19.200	18.330	19.111	18.269
40	36.550	30.684	31.831	30.477
80	41.300	43.058	44.245	42.468
160	43.550	42.670	43.097	41.458
320	24.700	21.492	21.112	20.184

Whence we calculate the quantities:

t (mins.)	$w = y-\eta_0$	$z_1 = \eta_1-\eta_0$	$z_2 = \eta_2-\eta_0$
10	6.5797	0.4530	-0.0172
20	0.8696	0.7802	-0.0617
40	5.8665	1.1470	-0.2066
80	-1.7576	1.1871	-0.5892
160	0.8796	0.4267	-1.2126
320	3.2082	-0.3802	-1.3074

The matrix of sums of squares and products is :

	w	z_1	z_2
w		7.457,062	-5.604,270
z_1	7.457,062	3.865,361	-1.012,682
z_2	-5.604,270	-1.012,682	3.573,636

from which the normal equations are:

$$3.865,361\ \phi_1 \quad -1.012,682\ \phi_2 = 7.457,062$$
$$-1.012,682\ \phi_1 \quad +3.573,636\ \phi_2 = -5.604,270$$

and $\phi_1 = 1.6401$

$\phi_2 = 1.1035$

A better estimate of the parameters (θ_1, θ_2) is therefore

$$\theta_1 = 1.04 + (1.6401)(0.02) = 1.0728$$

$$\theta_2 = 0.84 - (1.1035)(0.02) = 0.8179$$

These may be compared with the values $(\theta_1, \theta_2) = (1.072, 0.819)$ obtained in the same example in ref. 2.

An example containing more parameters, in which these techniques have been studied, is that discussed in ref. 1. The primary purpose of this paper was to demonstrate the way in which an empirical relationship might lead to the deduction of a more fundamental theoretical description of a physical system. The method used for fitting the theoretical function was crude, but sufficient for this purpose. The data are valuable, however, for exemplifying the techniques we have discussed.

The example was one in which M_1 and M_2 were assumed to react with M_3, which then further reacted with more M_1 to form M_4, in accordance with the scheme:

$$M_1 + M_2 \rightarrow M_3, \qquad M_1 + M_3 \rightarrow M_4$$

If η_1, η_2, η_3 and η_4 are the concentrations of M_1, M_2, M_3 and M_4 relative to the initial concentrations C_{20} of M_2, α, β and ρ are unknown constants and T is the absolute temperature, on assumptions set out in ref. 1, the progress of the reaction is represented by the following set of differential equations:

$$\frac{d\eta_1}{dt} = -2C_{20}\alpha\eta_1(\rho\eta_2 + \eta_3)\exp(-\beta/T)$$

$$\frac{d\eta_2}{dt} = -\rho C_{20}\alpha\eta_1\eta_2\exp(-\beta/T)$$

$$\frac{d\eta_3}{dt} = C_{20}\alpha\eta_1(\rho\eta_2 - \eta_3)\exp(-\beta/T)$$

$$\frac{d\eta_4}{dt} = C_{20}\alpha\eta_1(\rho\eta_2 + \eta_3)\exp(-\beta/T)$$

with the boundary conditions $\eta_1 = \eta_{10}$, $\eta_2 = 1$, $\eta_3 = 0$ at time $t = -t_0$, where in the treatment discussed in ref. 1, t_0 was assumed to be zero.

Experiments were performed in which the temperature T, the ratio of the initial concentration of the starting materials, η_{10}, and the time "on temp." elapsing before the reaction was stopped, t, were varied. Although the initial concentration of M_2, C_{20}, could have been varied, it was in fact held constant in these experiments. The data is set out in Table 2.

Table 2

Experiment	Levels of Variables			Observed Yield η_3
	T (°C.)	c (%)	t (hr.)	
1	162	25	5	0.459
2	162	25	8	0.533
3	162	30	5	0.575
4	162	30	8	0.588
5	172	25	5	0.606
6	172	25	8	0.580
7	172	30	5	0.586
8	172	30	8	0.524
9	167	27.5	6.5	0.569
10	177	27.5	6.5	0.554
11	157	27.5	6.5	0.469
12	167	32.5	6.5	0.575
13	167	22.5	6.5	0.550
14	167	27.5	9.5	0.589
15	167	27.5	3.5	0.503
16	177	20	6.5	0.611
17	177	20	6.5	0.629
18	160	34	7.5	0.600
19	160	34	7.5	0.606

The time recorded is the time on temperature. In practice a preliminary heat-up period occurs during which a small amount of reaction probably takes place. It seemed appropriate to allow for this approximately by assuming that the effective time on temperature was $t-t_o$ and t_o is some negative quantity to be estimated, representing the effective time for which the reaction had already been proceeding at the commencement of the period studied. There are therefore four unknown parameters to be estimated, α, β, ρ and t_o. Applying the conclusions of an earlier section, in which methods for improving the conditioning of the sums of squares sur-

face are discussed, we should prefer to consider the parameters $\gamma = \text{Ln}\,\alpha - \bar{x}\beta$, and β, rather than α and β, where $\bar{x}$ is the average reciprocal absolute temperature.

The example was first considered from the point of view of the earlier method of estimation. To obtain some insight into the situation, an extensive grid of calculations was performed, the results of which are shown in figure 7. (see following pages) These 16 contour diagrams show sections of the 4-dimensional contour surfaces for S^2 regarded as a function of γ, β, Ln ρ and t_o. It will be seen that this estimation surface is reasonably well-conditioned and should allow rapid convergence by the methods of descent. The very different situation which would have occurred if the transformation $\gamma = \text{Ln}\alpha - \bar{x}\beta$ had not been made can be seen if an attempt is made to replot these graphs in terms of α rather than γ. The square sections in γ and β shown in figure 7, become parallelograms, whose width are 35 times their height. As can be imagined, the contours whose axes are reasonably parallel to the axes of the parameters in the transformed variables, become extremely attenuated diagonal ridges, so that the application of any simple procedure of minimization in terms of the untransformed parameters would have been fruitless.

Method 2 has been applied to this case with the following results:

In ref. 1, an approximate method, based not on the value of η_3 alone, but on the observed values of both η_2 and η_3, gave a point which

corresponds to the values

$$\gamma = -6.06$$

$$\beta = 10091$$

$$\text{Ln } \rho = 1.224$$

where t_o was assumed to be zero. Working with the values of η_3 only this point gave a sum of squares of discrepancies of 230.2 . The design in this neighbourhood was carried out as follows:

First Design

Levels of parameters

	Base	Variant		
γ	-6.45	-6.25	whence	$\gamma = -6.45[1 - \phi_1) - 6.25\ \phi_1$
β	11000	14000		$\beta = 11000(1 - \phi_2) + 14000\ \phi_2$
$L_n \rho$	1.21	1.30		$\text{Ln } \rho = 1.21\ (1 - \phi_3) + 1.30\ \phi_3$
t_o	1	2		$t_o = (1 - \phi_4) + 2\ \phi_4$

The sums of squares and products corresponding to the quantities w, z_1, z_2, z_3, z_4, for this design were as follows:

	w	z_1	z_2	z_3	z_4
w		3814	967	350	2322
z_1	3814	13790	-2284	5838	9369
z_2	967	-2284	3982	-3531	-1886
z_3	350	5838	-3531	7976	4930
z_4	2322	9369	-1886	4930	6798

This leads to the estimates

whence the first improved approximation is

$\phi_1 = 0.5846$ $\gamma = -6.3331$

$\phi_2 = 0.4367$ $\beta = 12,310.1$

$\phi_3 = 0.0386$ $\text{Ln } \rho = 3.3651$

$\phi_4 = 0.3709$ $t_0 = 0.6291$

for which the value of S^2 is estimated as 90.00.

Second Design

Using this estimated point as a new base, a second design was performed. As the accuracy of the calculations was extremely high, an interval of only 1/100th of the previous interval was used.

Levels of parameters

	Base	Variant		
γ	-6.331	-6.331	whence	$\gamma = -6.3331\ (1-\phi_1) - 6.3311\ \phi_1$
β	12,310.1	12,340.1		$\beta = 12,310.1(1-\phi_2) + 12340.1\ \phi_2$
$\text{Ln } \rho$	1.2134577	1.214,3577		$\text{Ln } \rho = 1.2134577(1-\phi_3) + 1.2143577\ \phi_3$
t_0	0.6291	0.6391		$t_0 = 0.6291(1-\phi_4) + 0.6391\ \phi_4$

The sums of squares and products corresponding to the quantities w, z_1, z_2, z_3, z_4 for this design are

	z_1	z_2	z_3	z_4
w	-15.031,344	4.282,350	-12.531,670	-11.851,217 ($\times 10^{-6}$)
z_1	159.562,904	-15.477,099	65.052,009	120.334,432 ($\times 10^{-8}$)
z_2		40.975,777	-32.737,568	-13.554,382 ($\times 10^{-8}$)
z_3			72.830,157	57.995,115 ($\times 10^{-8}$)
z_4				96.386,239 ($\times 10^{-8}$)

This leads to the estimates

$$\phi_1 = -5.982,795$$

$$\phi_2 = -3.114,475$$

$$\phi_3 = -15.586,066$$

$$\phi_4 = 3.467,003$$

whence a second improved estimate of the parameters is

$$\gamma = -6.3451$$

$$\beta = 12216.7$$

$$\text{Ln } \rho = 1.19943$$

$$t_0 = 0.6638$$

for which the residual sum of squares is 89.6788.

Higher order approximations

So far we have proceeded by employing an iteration based on the idea that the η's were locally linear in the ϕ's. The solution locus of figure 6 is then assumed to be locally approximated by a plane on which the contours of ϕ_1 and ϕ_2 are equally spaced straight lines. It is clearly possible to use a higher order approximation, in which the solution locus is represented by some curved surface and a logical next approximation supposes that the η's are locally quadratic in the ϕ's rather than linear.

It will not of course happen necessarily that a higher order approximation will provide a better iteration rule than a lower order one.

The relative values of different iteration rules will depend on the relative amount of computation that has to be done and the convergence and rate of convergence of the procedures. Representation of the solution locus by a surface which adequately represents it locally is important however for another reason. Such a representation makes it possible to carry out an investigation in any given case of the associated distribution theory

concerning tests of significance, tests of goodness of fit, and confidence intervals.

Second and higher order approximations are most readily arrived at by noting that the solution we have used is equivalent to representing the function by a first order Taylor's Series expansion in the θ where derivatives are replaced by differences.

The first order procedure uses an iteration rule which consists of fitting successively by least squares

$$\hat{y}_u - \eta_{uo} = \sum_{i=1}^{k} \phi_i z_i$$

$$\text{i.e.} \qquad \hat{y}_u = \eta_{ou} + \sum_{i=1}^{k} \frac{\hat{\theta}_i - \theta_i}{\hat{\theta}_i' - \theta_i} (\eta_{iu} - \eta_{ou}) \qquad \text{(a)}$$

where $\eta_{iu} = \eta(\theta_1, \ldots, \hat{\theta}_i', \ldots, \theta_k; \xi_u)$, and

where after each iteration the solution $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_i, \ldots, \hat{\theta}_k$ replaces the starting solution $\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_k$.

If we let $\hat{\theta}_i' - \theta_i$ tend to zero this becomes

$$\hat{y}_u = \eta_{ou} + \sum_{i=1}^{k} (\theta_i - \theta_i) \frac{\partial \eta_u}{\partial \theta_i} \qquad \text{(b)}$$

where the derivatives are taken at the 'starting' values $\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_k$.

In practice we should usually work with the form (a) rather than (b). In both cases we are involved in the evaluation of $N(k+1)$ quantities but with (a) the function to be evaluated $\eta(\xi, \theta)$ is the same all the way through the calculation whereas with (b) $k+1$ different functions _ $\eta(\xi, \theta)$ and its k partial derivatives - have to be evaluated for each of the N trials. This involves extra labour in differentiating the function but more importantly where a digital computer is employed, it would be necessary to programme the ma-

chine for the k+1 different functions. Furthermore, where the function $\eta(\xi, \theta)$ is not given explicitly but appears, for example, as the solution of a set of differential equations which must be evaluated by some numerical subroutine such as the Runge-Kutta, the derivatives could not in general be explicitly obtained.

The second order approximation corresponding to (b) is

$$\hat{y}_u = \eta_{ou} + \sum_{i=1}^{k} \delta_i \frac{\partial \eta_u}{\partial \theta_i} + \frac{1}{2} \sum_{i=1}^{k} \sum_{i=j}^{k} \delta_i \delta_j \frac{\partial^2 \eta_u}{\partial \theta_i \partial \theta_j}$$

where $\delta_t = \hat{\theta}_t - \theta_t$ and the derivatives are taken at some fixed value $\theta_1, \theta_2, \ldots, \theta_k$.

Fitting by least squares we obtain the k normal equations

$$[w,k] = \sum_g \delta_g \{[g,i] - [w,g\ i]\} + \frac{1}{2} \sum_g \sum_h \delta_g \delta_h \{[i,\ g\ h] + 2[g,\ i\ h]\} + \frac{1}{2} \sum_g \sum_h \sum_j \delta_g \delta_h \delta_j \{[gh,\ i\ j]\}$$

$$i = 1, 2, \ldots k.$$

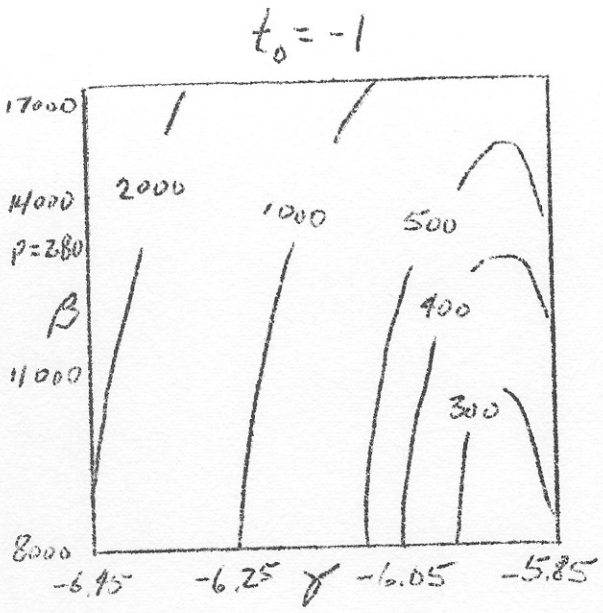
where $w_u = y_u - \eta_{ou}$ $[w,i] - \sum_u w_u \left\{ \frac{\partial \eta_u}{\partial \theta_i} \right\}$

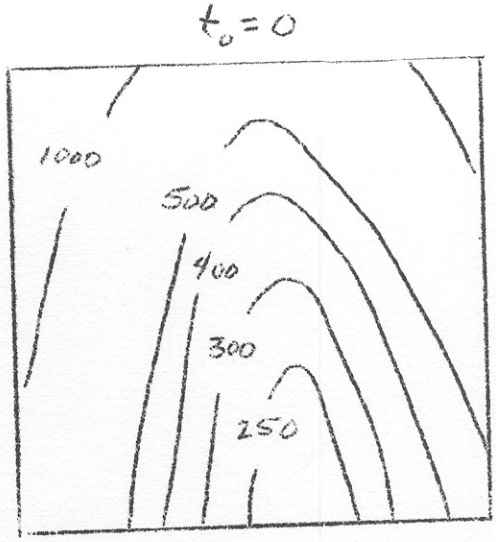
$$[i,g\ h] = \sum_u \left(\frac{\partial \eta_u}{\partial \theta_i} \right) \left(\frac{\partial^2 \eta_u}{\partial \theta_g \partial \theta_h} \right) \text{ and so on.}$$

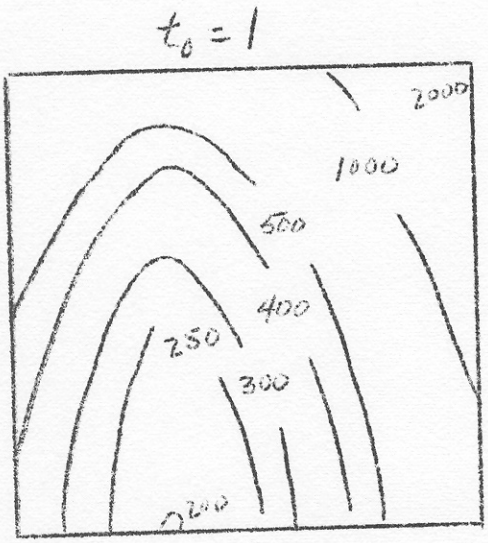
As before in practice the derivatives are evaluated numerically using the type of second order design suggested for the second order minimization procedure.

G. E. P. Box

Computer calculations carried out by G.A. Coutie


$t_0 = -1$
17000
14000
$p = 280$
β
11000
8000
2000
1000
500
400
300
-6.45
-6.25
γ
-6.05
-5.85


$t_0 = 0$
1000
500
400
300
250


$t_0 = 1$
2000
1000
500
400
250
300
200

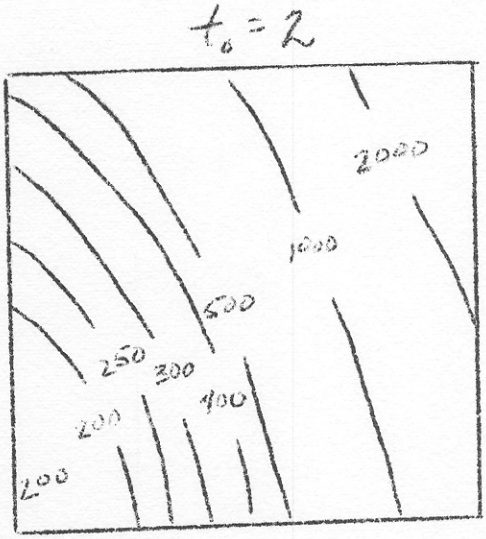
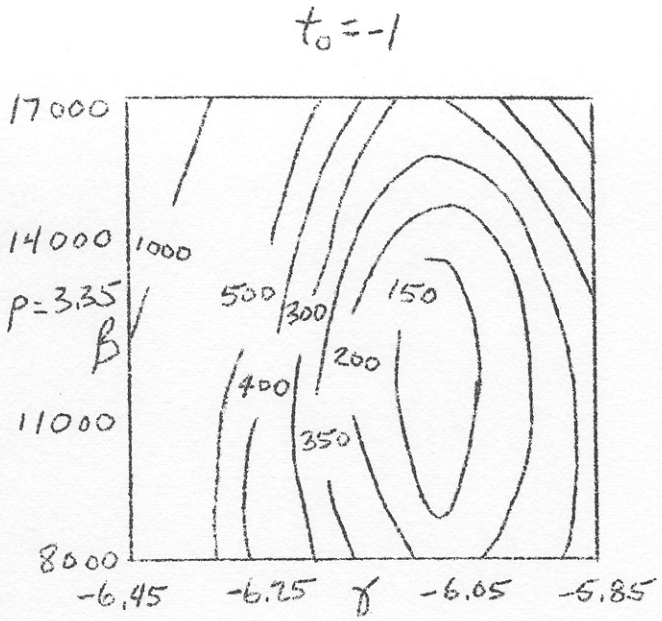
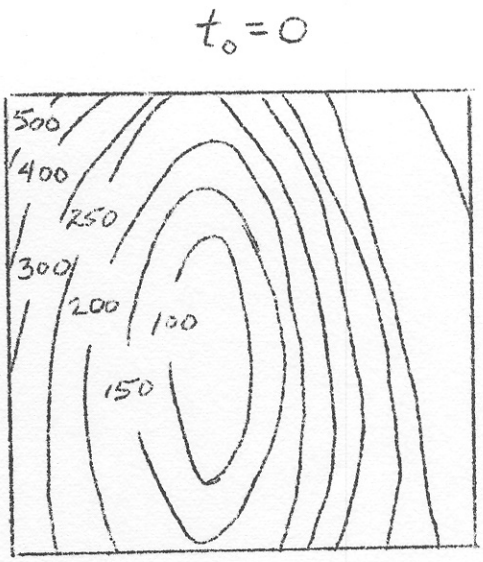
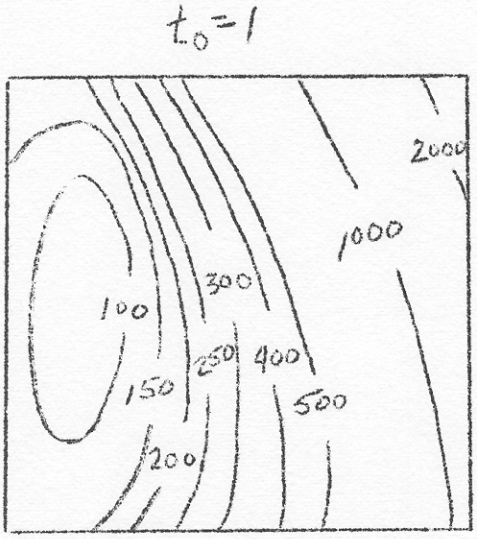

$t_0 = 2$
2000
1000
500
250
300
400
200
200

FIGURE 7


$t_0 = -1$
17000
14000
p=3.35
B
11000
8000
1000
500
300
150
400
200
350
-6.45
-6.25
γ
-6.05
-5.85


$t_0 = 0$
500
400
250
300
200
100
150


$t_0 = 1$
2000
1000
300
100
250
400
150
500
200

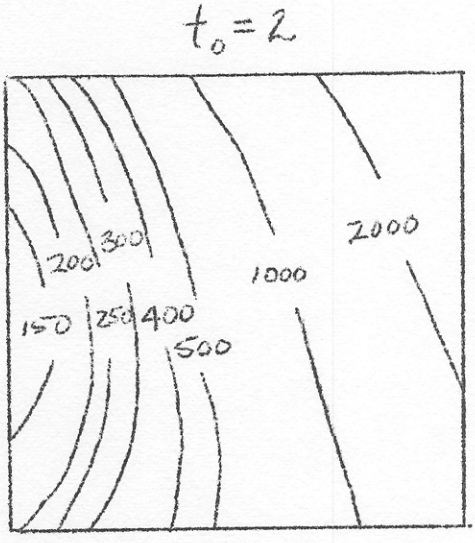
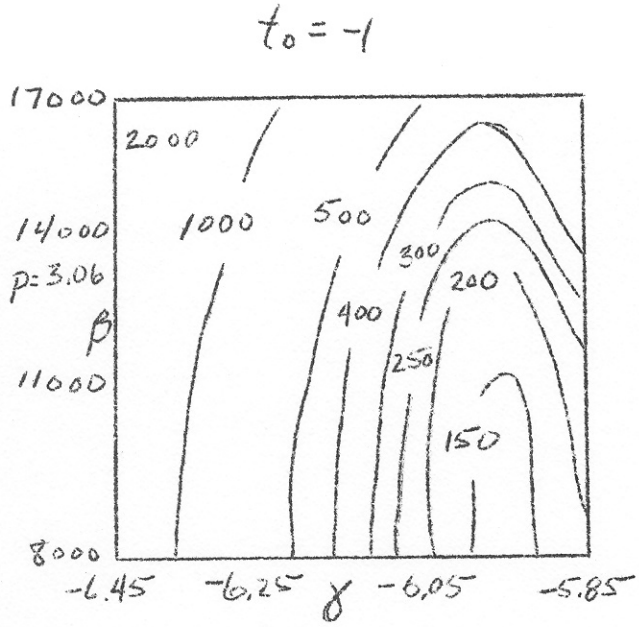
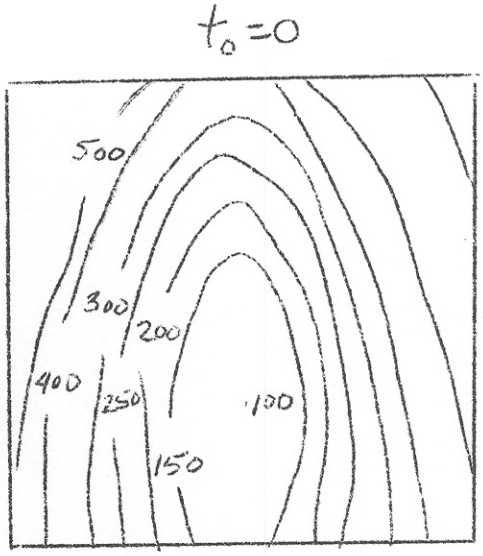
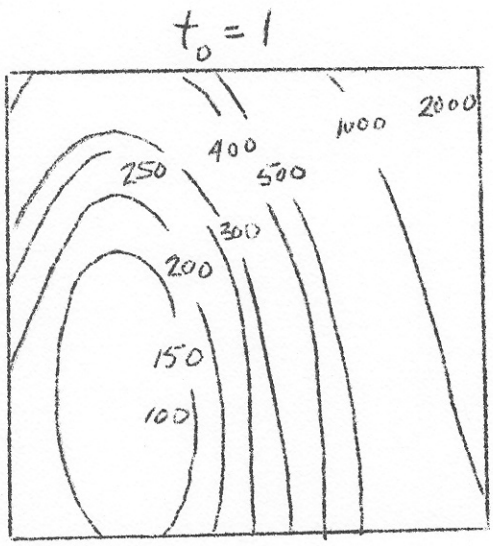

$t_0 = 2$
2000
200
300
1000
150
250
400
500

FIGURE 7


$t_0 = -1$
17000
14000
p=3.06
β
11000
8000
2000
1000
500
400
300
250
200
150
-6.45
-6.25
γ
-6.05
-5.85


$t_0 = 0$
500
300
200
400
250
100
150


$t_0 = 1$
2000
1000
400
500
250
300
200
150
100

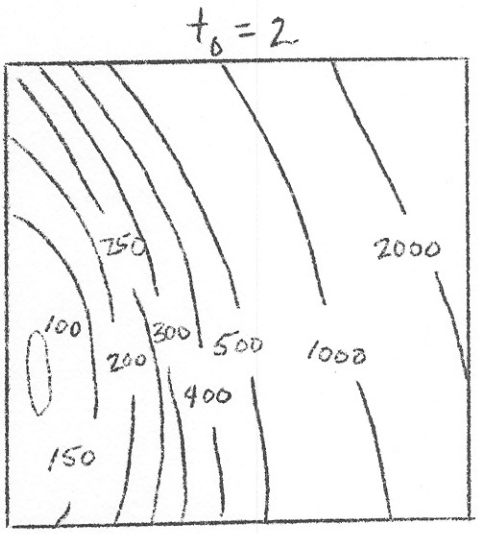

$t_0 = 2$
250
2000
100
300
500
1000
200
400
150

FIGURE 7

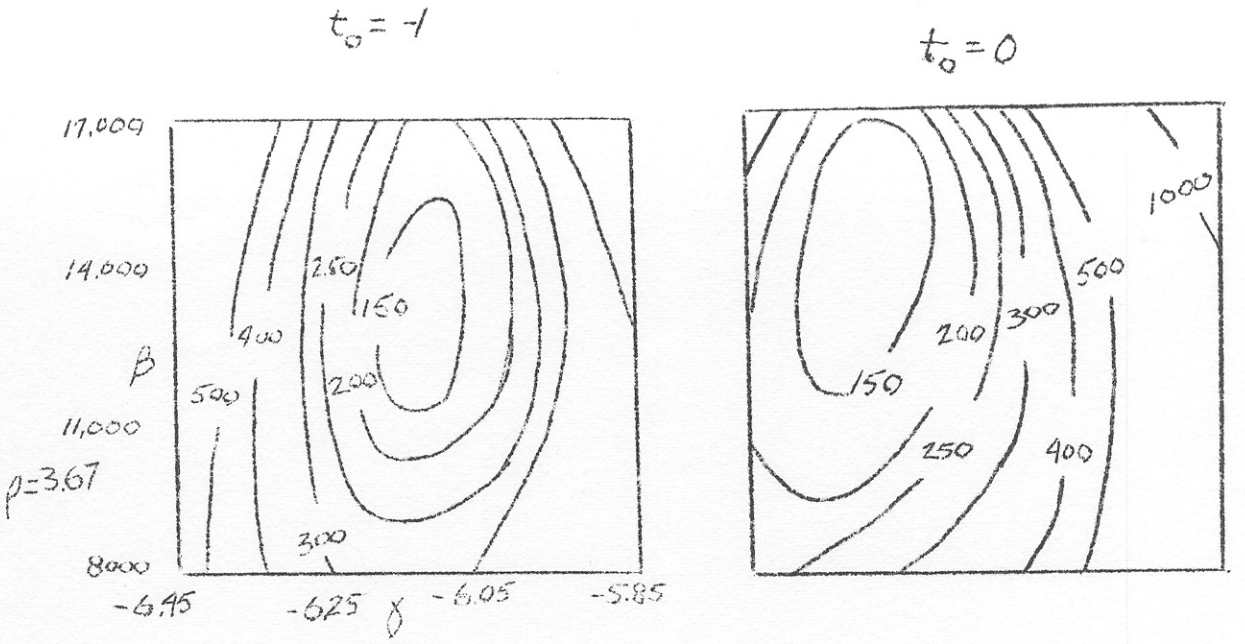

$t_0 = -1$
17,000
14,000
11,000
8000
β
ρ=3.67
-6.45
-6.25
-6.05
-5.85
δ
250
150
400
500
200
300
$t_0 = 0$
1000
500
200
300
150
250
400
$t_0 = 1$
2000
1000
150
250
200
500
300
400
$t_0 = 2$
300
2000
250
1000
400
500

FIGURE 7

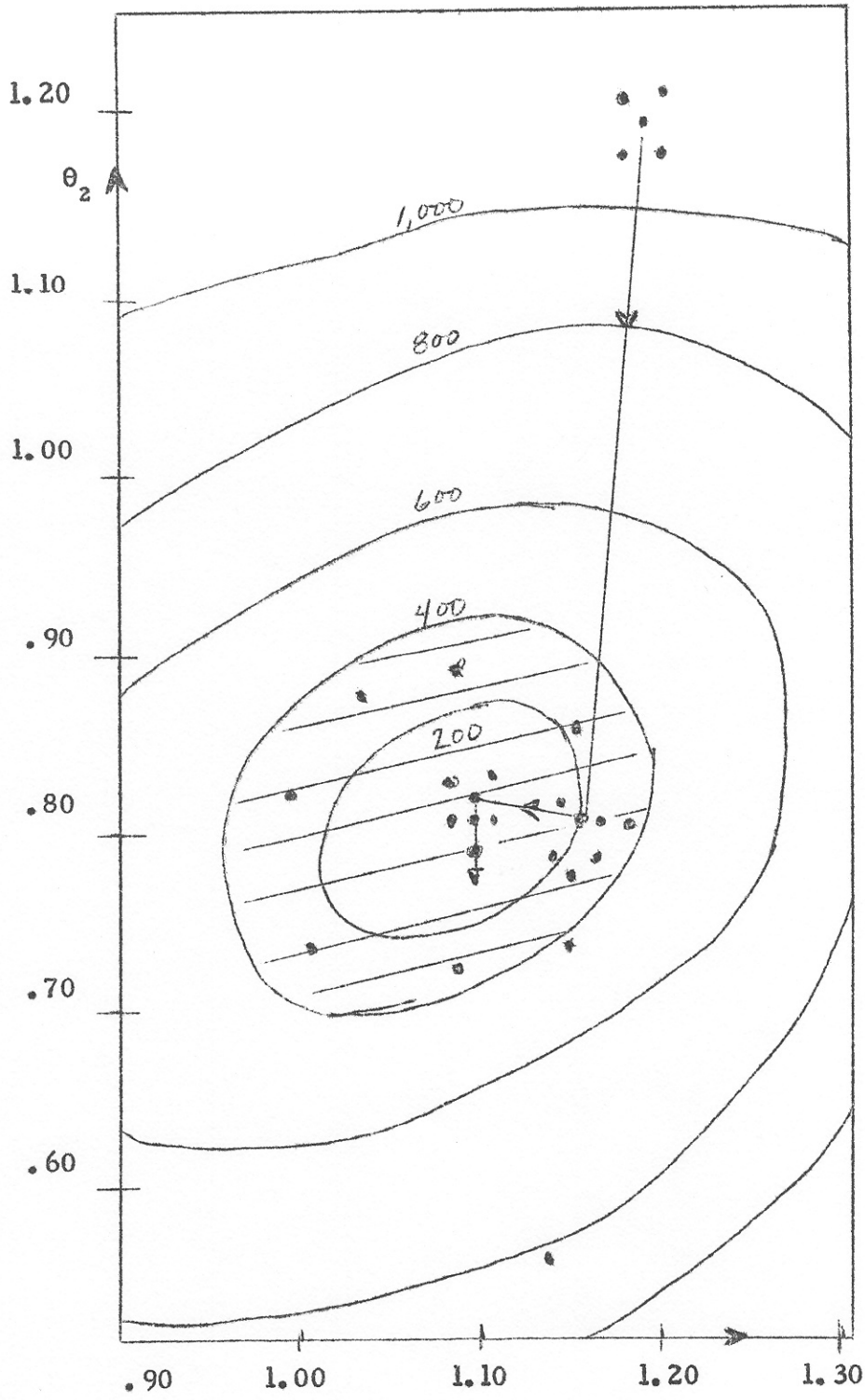


FIGURE 1: COURSE OF LEAST SQUARES CALCULATION, CONTOURS AND CONFIDENCE INTERVAL.

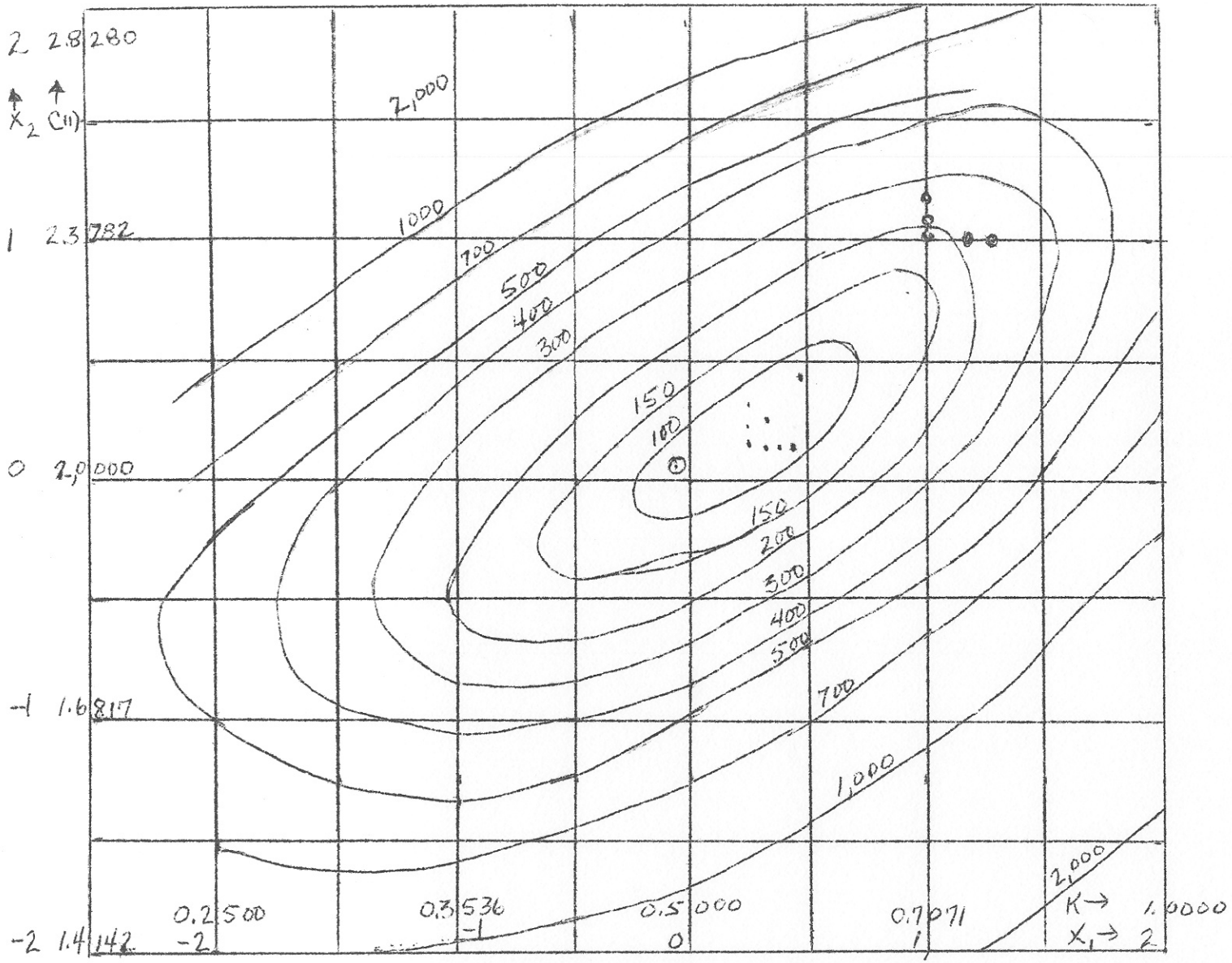

2 2.8280
X_2 (11)
1 2.3782
0 2.0000
-1 1.6817
-2 1.4142
2,000
1000
700
500
400
300
150
100
150
200
300
400
500
700
1,000
2,000
0.2500
0.3536
0.5000
0.7071
K→ 1.0000
X_1→ 2
-2
-1
0
1

FIGURE 4

REFERENCES:

1. BOX, G.E.P. and YOULE, P.V.: "The Exploration and Exploitation of Response Surfaces: An Example of the Link between the Fitted Surface and the Basic Mechanism of the System", Biometrics, 1955, 11, p. 287.

2. BOX, G.E.P. and COUTIE, G.A.: "Application of Digital Computers in the Exploration of Functional Relationships," Proceedings of the Institute of Electrical Engineers, 103, Pt. B, Supplement No. 1, (1956), pp. 100-107.

3. GILL, S.: "A Process for the Step-by-Step Integration of Differential Equations in an Automatic Digital Computing Machine", Proceedings of the Cambridge Philosophical Society, 47, p. 96.

4. KOSHAL, R.S.: "Application of the Method of Maximum Liklihood to the Improvement of Curves Fitted by the Method of Moments", Journal of the Royal Statistical Society, 1933, 96, p. 303.