THE CONDITION OF THE DETERMINANT SURFACE IN THE SPACE OF THE EXPERIMENTAL DESIGN LEVELS AS A FUNCTION OF THE INITIAL PARAMETER "GUESTIMATES" IN NONLINEAR PARAMETER ESTIMATION

Dwight E. Spence, University of Pennsylvania

Introduction: Professors G.E.P. Box and H.L. Lucas (1959) considered the matter of selecting those factor levels (ξ) which would result in the estimation of a given nonlinear (w.r.t. parameters) model whose parameter estimates (θ) would have the property of smallest variance. For least squares estimates of these parameters the criterium would entail minimizing the absolute value of the determinant $\mathrm{Cov}(\theta)$, which is approximately the generalized variance. The Variance-Covariance matrix $\mathrm{cov}(\theta)$ is proportional to $(\underline{F}'\,\underline{F})^{-1}$ where $\underline{F}$ is the pseudo-design matrix (Gallant, 1975) and is a function of both the parameters (θ) and the factor levels (ξ). Used to explain this criterium are the relationships between the parameter space, the sample space and the solution locus, which constitutes a subset of sample space, as follows. For P dimensional points θ in parameter space, $P(\theta)$ defines a transformation to N dimensional sample space constituting a subset of sample space, i.e., the solution locus. Using an orthogonal transformation of sample space coordinates, solution locus points in the near neighborhood of a point $P(\theta)$ may be reduced to P dimensionality. Whe $P(\theta)=P(\theta_0)$, θ_0 being the true and unknown values, the determinant $|(\underline{F}'\underline{F})^{-1}|$ equals the square of the Jacobian of the transformation of these reduced-dimension points (near $P(\theta)$) to image points in parameter space. Thus, minimization of the absolute value of $|(\underline{F}'\underline{F})^{-1}|$ implies minimization of the volume covered by these points mapped to parameter space.

Determinant Surface:
The primary illustration of the above involves a two stage chemical reaction: $A \rightarrow B \rightarrow C$. Box and Lucas concern themselves with the model $\eta = \theta_1/(\theta_1-\theta_2)\{e^{-\theta_2\xi} - e^{-\theta_1\xi}\}$ where η is the percentage yield of substance B. Given the restriction that the number of data points to be selected equal the number of parameters to be estimated, note that the problem of minimizing the absolute value of $|(\underline{F}'\underline{F})^{-1}|$ reduces to maximizing the absolute value of $|\underline{F}|$ which is of course still a function of both parameters (θ) and the factor levels (ξ). To proceed toward the selection of the desired factor levels initial estimates of the parameters must be chosen. This brings us to the focal question on which the present paper is based: What are the consequences of choosing one set of parameter estimates (θ^*) over one of the many other possible sets? In this particular example the parameters are interpreted as indicating the rates of chemical reactions. Often there may not be even this much information on which to base the "questimates" (θ^*). The initial estimates used by Box and Lucas are $\theta_1^* = .7$ and $\theta_2^* = .2$. Utilizing a computer and an extensive grid search technique, we can plot the contours of the response surface of $|\underline{F}^*|$ in the space of the factor levels, i.e., ξ_1 and ξ_2. Figure A4 shows, as did Box and Lucas, that the result of grid search is a multi-extremum surface with stationary points at $|\underline{F}^*| \doteq .039$ and $|\underline{F}^*| \doteq -.81$. If instead we consider $\theta_1^* = .2$ and $\theta_2^* = .7$, we get through grid search the contour surface indicated in Figure B4. For this case and in the region considered here, the determinant response surface is simpler having only one stationary point, a valley at $|\underline{F}^*| \doteq -.1176$. In these two cases the consequent factor levels, the indicated numerical values of ξ_1 and ξ_2, differed fairly negligibly as we will speak more about later. If we go on to consider initial parameter estimates of $\theta_1^* = 1.0$ and $\theta_2^* = 0.0$, figure C4 shows that the result is a rising ridge type surface indicating that, within the region under consideration, the absolute value of $|\underline{F}^*|$ is maximized at $|\underline{F}^*| = -12.0$. Finally, figure D4 displays the results of using $\theta_1^* = .5$ and $\theta_2^* = .4$ as the starting estimates; and again we find a multi-extremum surface with a peak at $|\underline{F}^*| = .000093$ and a valley at $|\underline{F}^*| = -.3770$. Evidently, reasonably small changes in the initial estimates of the parameters result in quite a variety of determinant surfaces.

Experimental Design:
Should circumstances require that one attempt to locate these experimental design points, i.e., ξ_1 and ξ_2 which optimize $|\underline{F}^*|$, by the use of numerical methods (Box & Wilson, 1951) such as steepest descent (ascent), different results can be expected depending on the nature of the particular determinant surface. Figures A3, B3, C3, and D3, reveal the approximated determinant contour surfaces for each of the four sets of initial estimates and based on quadratic equations whose data points are presented in Table 1. (Note how rapidly the approximation deteriorates as we move away from the center of the selected data points.) Table 1 also indicates the four sets of factor level coordinates, ξ_1 and ξ_2, obtained as a result of calculating the stationary points of the corresponding quadratic equations. With respect to $\theta_1^* = 1.0$ and $\theta_2^* = 0.0$ the values of ξ_1 and ξ_2 are dictated in

part by the limits of the experimental region since the quadratic equation in this case indicates a stationary point far outside the experimental region considered here. The graphs shown in figures B_1, B_2, B_3, and B_4 (as well as those in figures A_1, A_2, A_3, and A_4) when available provide important information concerning the selection of the data points on which the quadratic equations are generated. In each case of graphs of type B, graphs in the space of the partial derivatives ($D_2\eta$), the objective is to determine the triangular simplex of greatest volume such that one vertex is at the origion; the remaining two vertices indicate values of ξ_1 and ξ_2 which define a point very near (or, ideally, exactly equal to) the desired stationary point. This information would thus eliminate the need for most all of the steepest descent (ascent) methodolgy.

Finally, Table 1 (Bottom) gives the results of translating and orthogonally rotating the x_1 and x_2 axes by way of canonical reduction of selected quadratic equations (Bradley, 1958). The x_1 and x_2 coordinate values are simple functions of ξ_1 and ξ_2 as also indicated in Table 1. The graph of figure A_3 illustrates these results with canonical axes X_1 and X_2 centered at the stationary point.

REFERENCES

Box, G.E.P. and Lucas, H.L. (1959): "Design of Experiments in Nonlinear Situations," Biometrika, 46 , 77-90.

Box, G.E.P. and Wilson, K.B. (1951): "On the Experimental Attainment of Optimum Conditions," Journal of the Royal Statistical Society, Series B, Vol. XIII, No. 1.

Bradley, Ralph A. (1958): "Determination of Optimum Operating Conditions by Experimental Design," Industrial Quality Control, 15, 16-20.

Gallant, A.R. (1975): "Nonlinear Regression," American Statistican, Vol. 29, 73-81.

TABLE 1

The MODEL, under the assumption of normality:

$$\eta = \theta_1/(\theta_1-\theta_2)\{\exp(-\theta_2\xi)-\exp(-\theta_1\xi)\}$$

$$\eta_i^* = \eta_i+\epsilon_i,\ E(\eta_i^*)=\eta_i \text{ and } E(\epsilon_i\epsilon_j) = \begin{cases}\sigma^2 & i=j\\ 0 & i\neq j\end{cases}$$

$$\phi_1=\partial\eta/\partial\theta_1=[1/\theta_1-1/(\theta_1-\theta_2)][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_2\xi)-[1/\theta_1-1/(\theta_1-\theta_2)-\xi][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_1\xi)$$

$$\phi_2=\partial\eta/\partial\theta_2=[1/(\theta_1-\theta_2)-\xi][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_2\xi)-[1/(\theta_1-\theta_2)][\theta_1/(\theta_1-\theta_2)]\exp(-\theta_1\xi)$$

	A	B	C	D
INITIAL PARAM. ESTIMATES:	$\theta_1^*=.7, \theta_2^*=.2$	$\theta_1^*=.2, \theta_2^*=.7$	$\theta_1^*=1.0, \theta_2^*=0.0$	$\theta_1^*=.5, \theta_2^*=.4$
OPTIMAL LEVELS INDICATED:	$\xi_1=1.23, \xi_2=6.86$	$\xi_1=1.61, \xi_2=5.93$	$\xi_1\doteq 1.0, \xi_2=36.0$	$\xi_1=1.37, \xi_2=5.32$

A.

The following data points

ξ_1	ξ_2	Δ
1.10	6.5	-.803 336
1.20	6.5	-.807 992
1.30	6.5	-.806 709
1.10	7.0	-.804 758
1.10	7.5	-.798 920
1.20	7.0	-.809 632
1.23*	6.87*	-.810 413*

$x_1=5(\xi_1-1.1)$
$x_2=(\xi_2-6.5)$

resulted in a Quadratic equation of the form

$$\Delta=-.8033-.01531x_1-.01013x_2+.01195x_1^2+.01454x_2^2-.00067x_1x_2$$

B.

The following data points

ξ_1	ξ_2	Δ
1.6	4.5	-.108 211
1.7	4.5	-.106 370
1.8	4.5	-.104 087
1.6	5.5	-.117 023
1.7	5.5	-.116 342
1.6	6.5	-.116 487
1.65*	6.0*	-.117 618*

$x_1=5(\xi_1-1.6)$
$x_2=(\xi_2-5.5)$

resulted in a Quadratic equation of the form

$$\Delta=-.11715-.00033x_1-.00415x_2+.00519x_1^2+.00491x_2^2-.0009x_1x_2$$

C.

The following data points

ξ_1	ξ_2	Δ
.8	25.0	-8.62712
.9	25.0	-8.78190
1.0	25.0	-8.82911
.8	30.0	-10.42443
.9	30.0	-10.61147
.8	35.0	-12.22175
1.0*	35.0*	-12.50790*

$x_1=5(\xi_1-.8)$
$x_2=(\xi_2-25.0)/10$

resulted in a Quadratic equation of the form

$$\Delta=-8.62487-.43559x_1-3.61406x_2+.2331x_1^2+.01793x_2^2-.08416x_1x_2$$

D

The following data points

ξ_1	ξ_2	Δ
1.3	4.5	-.363 893
1.4	4.5	-.362 252
1.5	4.5	-.358 417
1.3	5.0	-.375 072
1.4	5.0	-.374 863
1.3	5.5	-.375 811
1.39*	5.34*	-.377 209*

$x_1=5(\xi_1-1.3)$
$x_2=(\xi_2-4.5)$

resulted in a Quadratic equation of the form

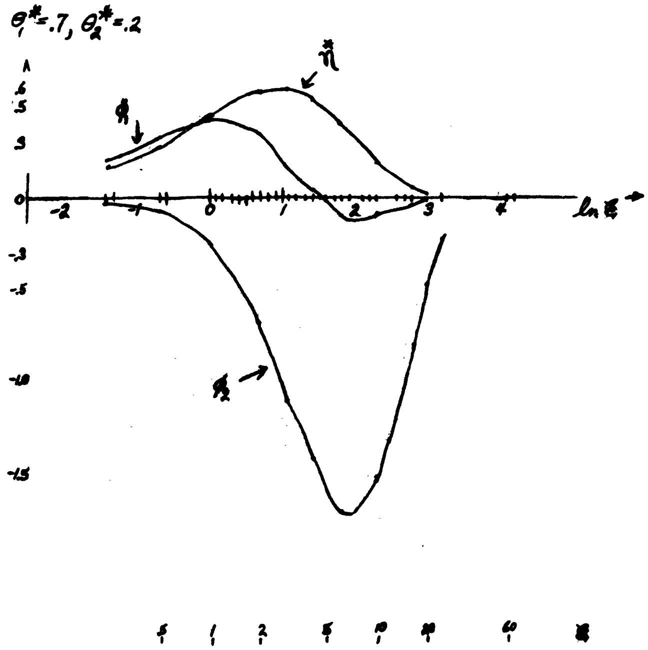
$$\Delta=-.36387+.00089x_1-.03315x_2+.00455x_1^2+.02124x_2^2-.0049x_1x_2 .$$

Reduction of the above to canonical form gives the following:

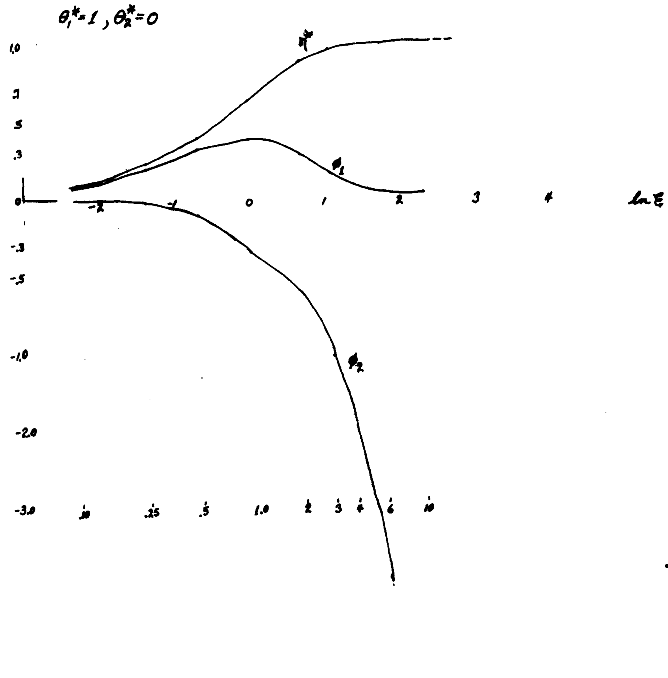
	A	B	C	D
Eigenvalues:	$B_1=.0146$	$B_1=.0055$?	$B_1=2.410$	$B_1=.0216$
	$B_2=.0119$	$B_2=.0046$	$B_2=.0099$	$B_2=.0042$
Eigenvectors:	$\begin{bmatrix} .1262 \\ -.9920 \end{bmatrix}_1$		$\begin{bmatrix} .9827 \\ -.1853 \end{bmatrix}$	
	$\begin{bmatrix} .9920 \\ .1262 \end{bmatrix}_2$		$\begin{bmatrix} .1853 \\ .9827 \end{bmatrix}$	

*Asterik values are the coordinates of the stationary point based on the quadratic equation generated from the initial six data points.

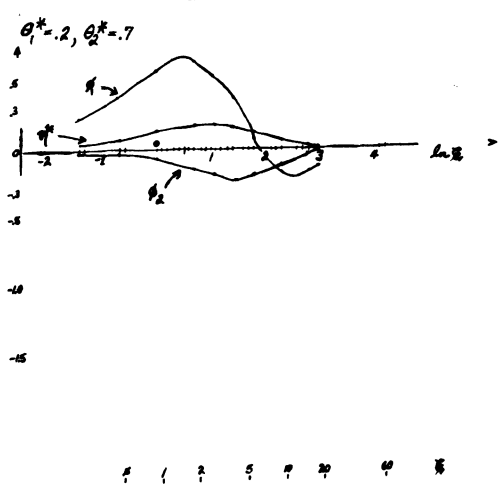
A_1



C_1

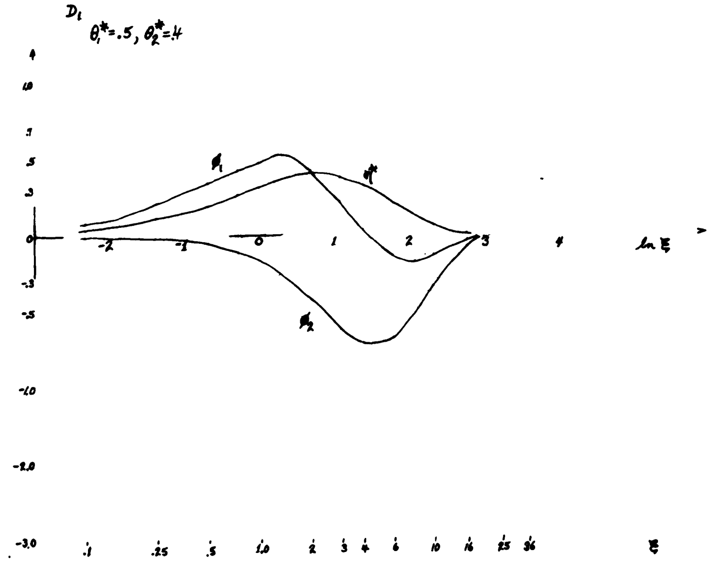


B_1



$$\hat{\eta} = \frac{\theta^*}{\theta_1^* - \theta_2^*}\left\{e^{-\theta_2^* \xi} - e^{-\theta_1^* \xi}\right\}$$

D_1



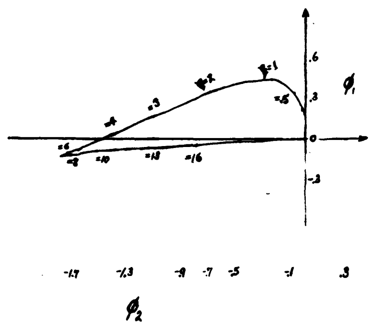
A_2

7, $\theta_2^* = .2$

Design Locus:

$$\phi_1(\xi) = [.8 + 1.4\xi]e^{-.7\xi} - .8e^{-.2\xi}$$

$$\phi_2(\xi) = [2.8 - 1.4\xi]e^{-.2\xi} - 2.8e^{-.7\xi}$$

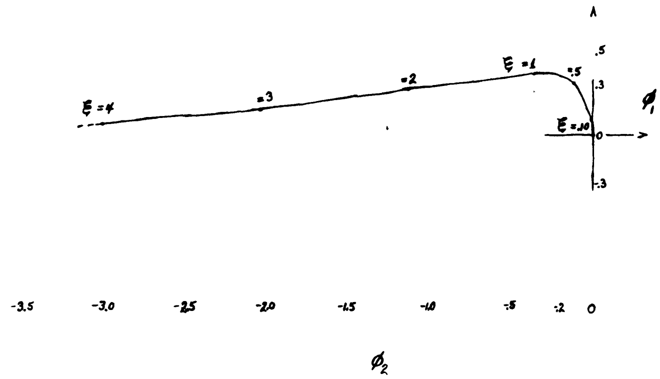


C_2

$\theta_1^* = 1,\ \theta_2^* = 0$

$$\phi_1(\xi) = \xi e^{-\xi}$$

$$\phi_2(\xi) = 1 - \xi - e^{-\xi}$$



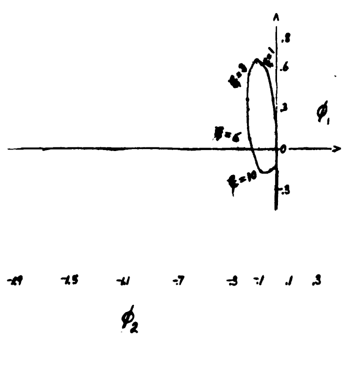
B_2

$\theta_1^* = .2,\ \theta_2^* = .7$

Design Locus:

$$\phi_1(\xi) = [2.8 - .4\xi]e^{-.2\xi} - 2.8e^{-.7\xi}$$

$$\phi_2(\xi) = [.8 + .4\xi]e^{-.7\xi} - .8e^{-.2\xi}$$

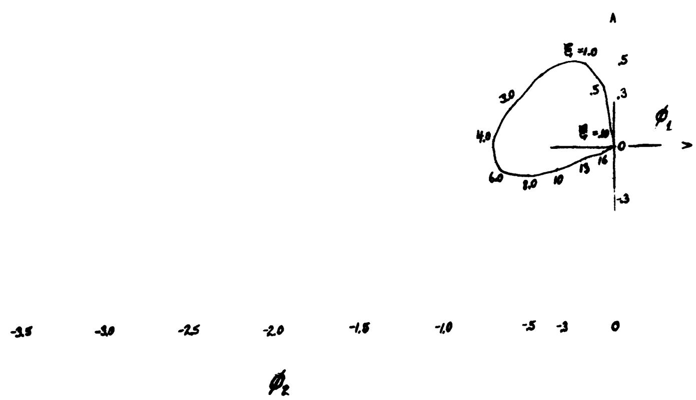


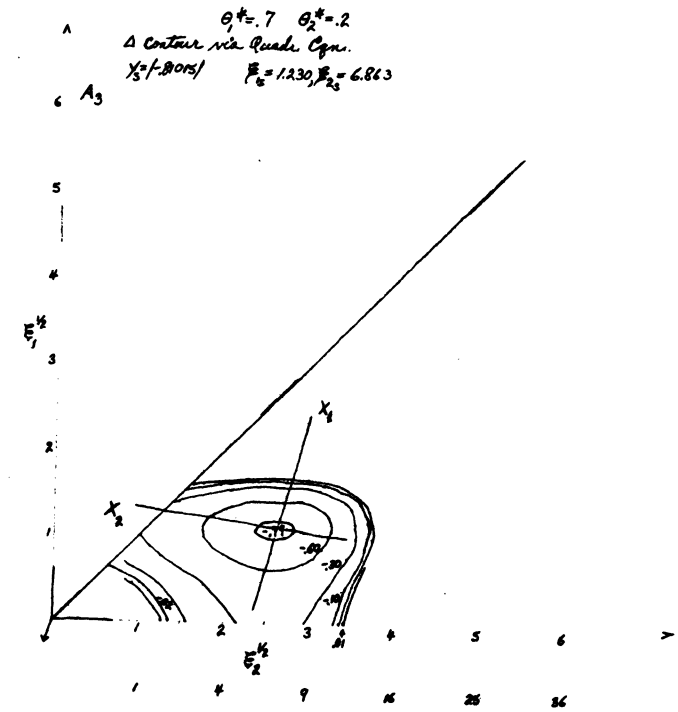
D_2

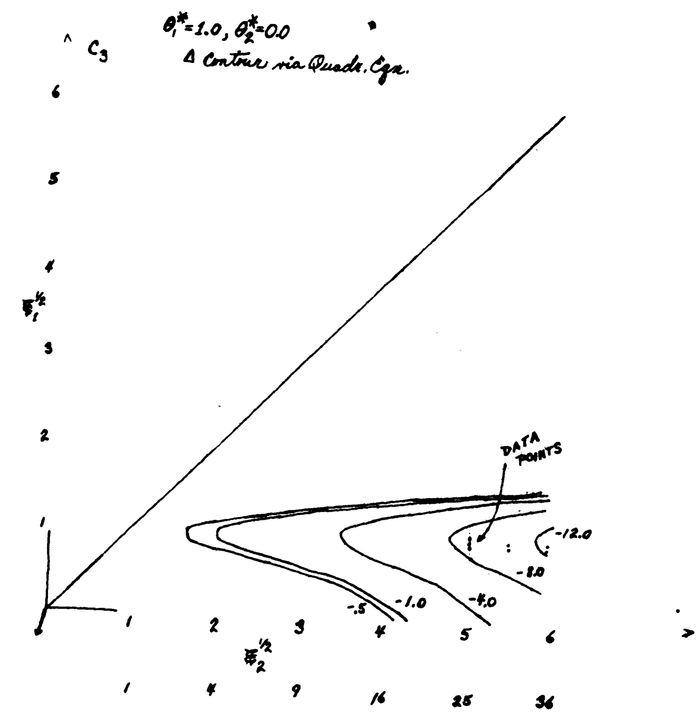
$\theta_1^* = .5,\ \theta_2^* = .4$

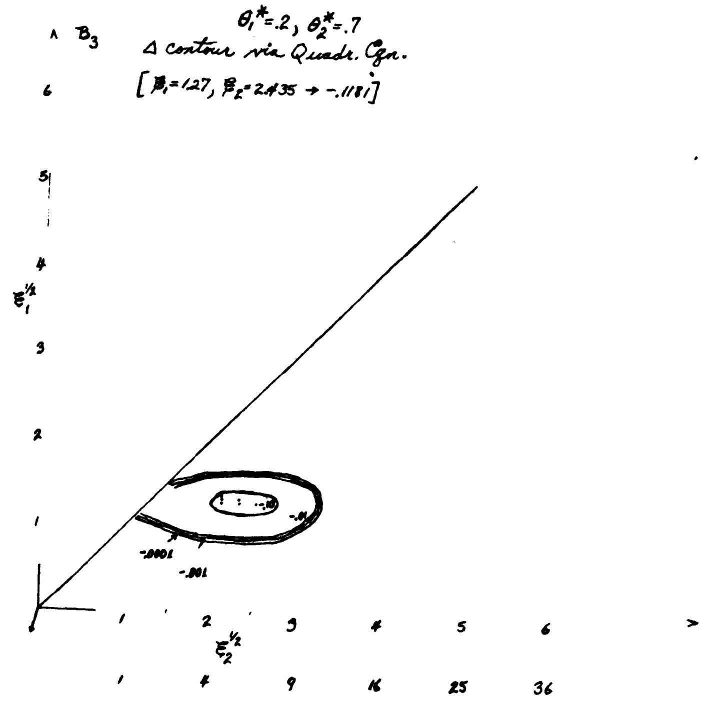
$$\phi_1 = (40 + 5\xi)e^{-.5\xi} - 40e^{-.4\xi}$$

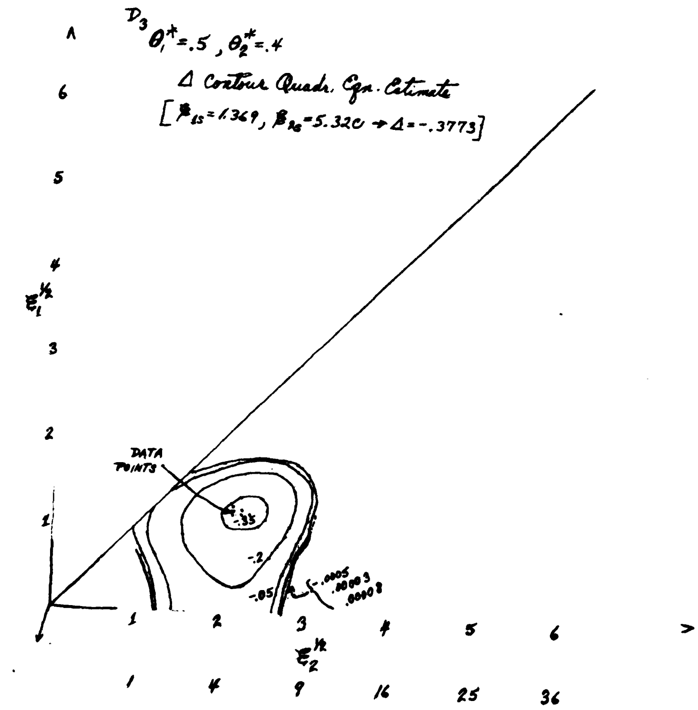
$$\phi_2 = (50 - 5\xi)e^{-.4\xi} - 50e^{-.5\xi}$$

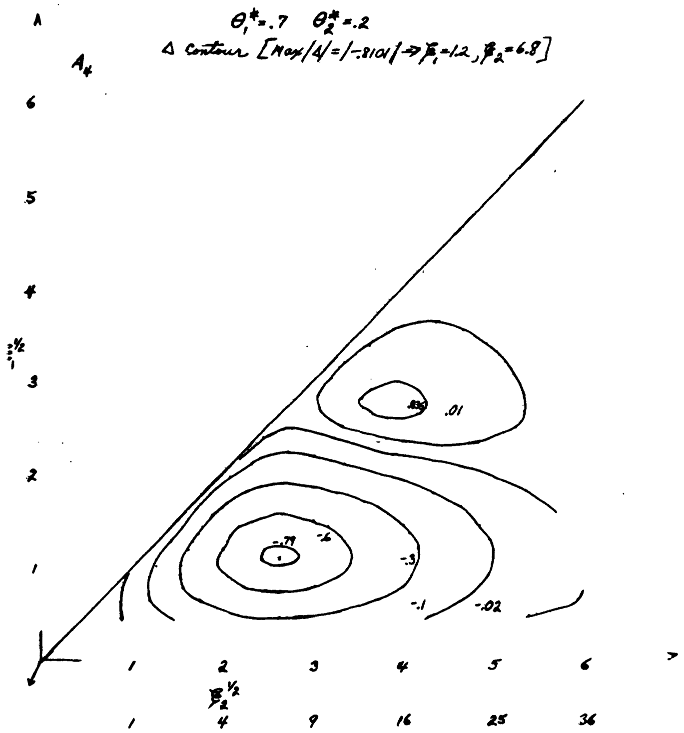


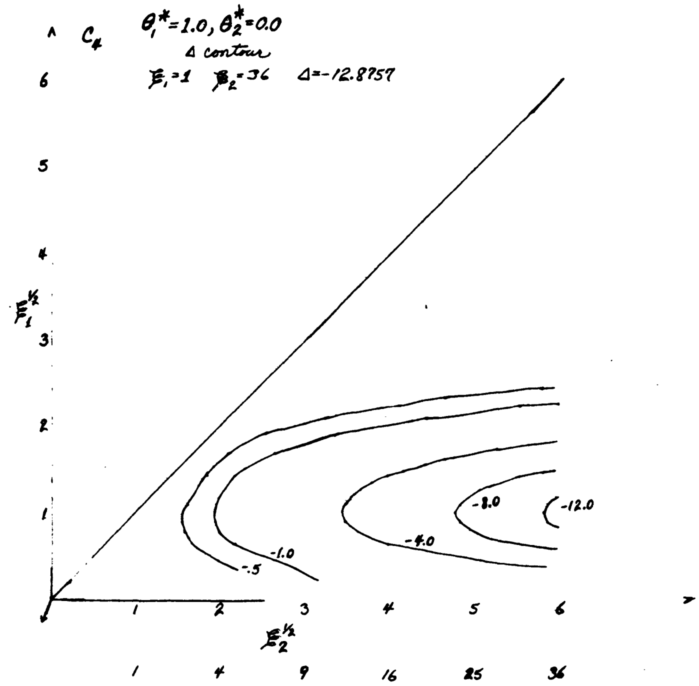

θ₁* = .7 θ₂* = .2
Δ contour via Quadr. Cqn.
A3
χ₁
χ₂
1 4 9 16 25 36

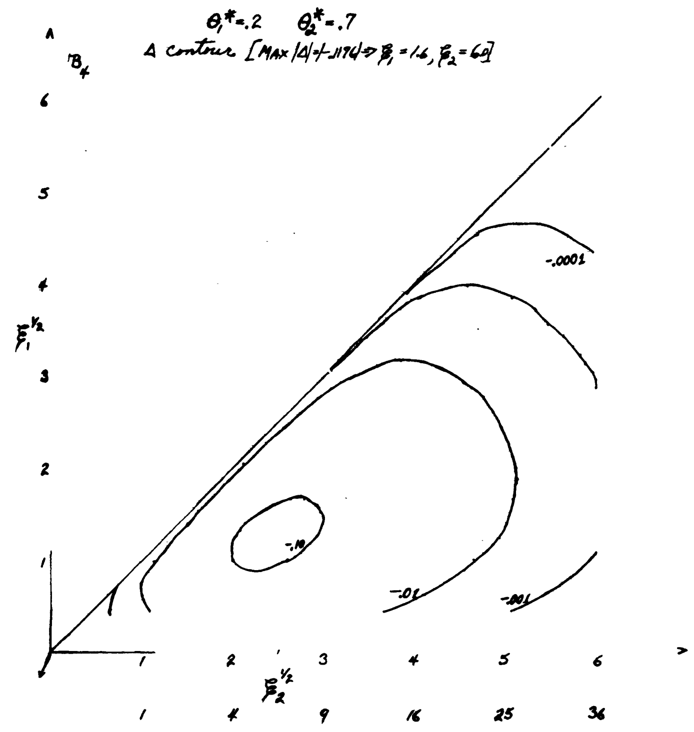

C3
θ₁* = 1.0, θ₂* = 0.0
Δ Contour via Quadr. Cqn.
DATA POINTS
-12.0
-8.0
-4.0
-1.0
-.5
1 4 9 16 25 36

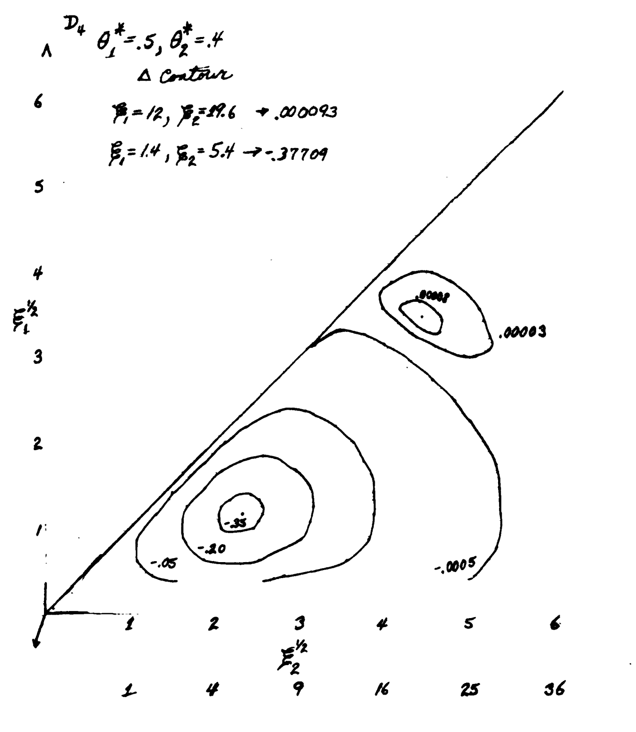

B3
θ₁* = .2, θ₂* = .7
Δ contour via Quadr. Cqn.
-.0001
-.001
1 4 9 16 25 36


D3
θ₁* = .5, θ₂* = .4
Δ Contour Quadr. Cqn. Estimate
DATA POINTS
-.33
-.2
-.05
1 4 9 16 25 36


A4
θ1* = .7 θ2* = .2
Δ contour [Max |Δ| = |-.8101| ⇒ ξ1 = 1.2, ξ2 = 6.8]
.01
-.79
-.6
-.3
-.1
-.02
ξ1^½
ξ2^½
1 2 3 4 5 6
1 4 9 16 25 36


C4
θ1* = 1.0, θ2* = 0.0
Δ contour
ξ1 = 1 ξ2 = 36 Δ = -12.8757
-8.0
-12.0
-4.0
-1.0
-.5
ξ1^½
ξ2^½
1 2 3 4 5 6
1 4 9 16 25 36


B4
θ1* = .2 θ2* = .7
Δ contour [MAX |Δ| = |-.1194| ⇒ ξ1 = 1.6, ξ2 = 6.0]
-.0001
-.10
-.01
-.001
ξ1^½
ξ2^½
1 2 3 4 5 6
1 4 9 16 25 36


D4
θ1* = .5, θ2* = .4
Δ contour
ξ1 = 12, ξ2 = 19.6 → .000093
ξ1 = 1.4, ξ2 = 5.4 → -.37709
.00008
.00003
-.35
-.20
-.05
-.0005
ξ1^½
ξ2^½
1 2 3 4 5 6
1 4 9 16 25 36