DEPARTMENT OF STATISTICS

University of Wisconsin

Madison, Wisconsin

Technical Report No. 21

October, 1963

Sequential Design of Experiments for
Non-Linear Models

by

G. E. P. Box and William G. Hunter

SEQUENTIAL DESIGN OF EXPERIMENTS FOR NON-LINEAR MODELS

by

G. E. P. Box and William G. Hunter

1. Objective of the Experimenter

Suppose that an experimenter is interested in studying a particular system for which there exists a mathematical model $\eta = f(\underline{\theta}, \underline{\xi})$, non-linear in the parameters $\underline{\theta}$, which relates a measurable response η to the controllable variables $\underline{\xi}$. The objective of the experimenter may be (1) to obtain an estimate of a response η over some particular region of interest in the space of the variables or (2) to determine the underlying physical mechanism of the phenomenon under investigation.

Mathematically, we could say for problem (2) the whole object is to discover the nature of the function $f(\underline{\theta}, \underline{\xi})$. In practical situations we can never know this completely. However, we shall say that we have an adequate theoretical model when we have derived from a consideration of the mechanism a function which closely predicts the results of actual experiments.

In problem (1), which has come to be called the response surface problem, it is useful but not essential to employ such a theoretical model. [2, 3, 4, 10] In many circumstances even though no theoretical model is available, perfectly good empirical approximations can be obtained by fitting a polynomial or some other flexible graduating function over the region of interest. [1, 5, 6, 15]

Empirical models are, however, of limited value when the aim is to develop a suitable mechanistic theory. The search for underlying physical mechanisms constitutes a major portion of effort in a number of scientific fields. To engineers, for example, basic mechanism studies are of interest principally because a deeper understanding makes it possible to cope with engineering design problems in a more intelligent and useful manner than would be possible if the mechanism were entirely unknown.

In what follows we will be concerned with a particular aspect of this second objective of trying to elucidate the mechanism. Such mechanism studies consist essentially of two steps: (i) establishing an adequate form for the theoretical model and then (ii) determining precisely the values of its parameters.

Step (i) is the model-building problem which has been discussed in references [7], [12], [13], and [17]. Further facets of this important problem are currently being investigated. In this paper we suppose that step (i) has been accomplished and the form of the theoretical model is therefore known. The problem which confronts the experimenter now is the evaluation of the physical parameters (e.g., rate constants in chemical kinetics examples). The problem of statistical <u>analysis</u> of data in these situations has been discussed by Box [3]. The purpose of this paper is to consider the problem of <u>generation</u> of data, i.e., the statistical design of experiments.

Bayes' Theorem

The essential machinery we shall use in drawing inferences from data is the well-known formula due to the Rev. Thomas Bayes

$$p_N(\underline{\theta}|\underline{y}, \underline{D}) = p_N(\underline{\theta}|\underline{y}) = \frac{p_o(\underline{\theta}) \, L(\underline{\theta}|\underline{y})}{\int p_o(\underline{\theta}) \, L(\underline{\theta}|\underline{y})} \tag{1}$$

where $p_N(\underline{\theta}|\underline{y}, \underline{D})$ is the posterior distribution of the parameters $\underline{\theta}$ after N observations $\underline{y}$ have been obtained; $P_o(\underline{\theta})$ is the prior distribution that exists at stage N=O, i.e., before any observations are available from the experimental program; and $L(\underline{\theta}|\underline{y})$ is the likelihood function.

The dependence of the posterior distribution on the design matrix $\underline{D}$ as well as the experimental results $\underline{y}$ is made explicit when it is written as $p_N(\underline{\theta}|\underline{y}, \underline{D})$. For convenience we shall write the posterior distribution as $p_N(\underline{\theta}|\underline{y})$ suppressing $\underline{D}$; however, the dependence of $p(\underline{\theta}|\underline{y})$ on $\underline{D}$ should always be borne in mind.

2. Designs for Parameters

If experiments are not carefully planned the experimental points may be so situated in the space of the variables that the estimates which can be obtained for the parameters $\underline{\theta}$ are not only imprecise but also highly correlated. Once the data are collected a statistical analysis, no matter how elaborate, can do nothing to remedy this unfortunate situation. However, by selecting a suitable experimental design in advance these shortcomings can often be overcome.

The problem of designing experiments in non-linear situations has received comparatively little attention. Some possible approaches have been suggested by Box and Lucas [8], Chernoff [11], Elfving [16], Kiefer [18, 19], Stone [23], and Wald [25]. In the next section of this paper we present a Bayesian approach to the problem.

Box and Lucas [8] proceed by attempting to choose $\underline{D}$ in such a manner that the volume of the approximate confidence region for $\underline{\theta}$ is minimized, or, equivalently, under suitable assumptions, trying to choose $\underline{D}$ to minimize the volume in the parameter space which contains a given percentage of the posterior distribution. If the experimental errors are approximately Normally distributed and the response relationship is approximately linear in the vicinity of the least squares estimates $\hat{\underline{\theta}}$ then the volume of this region is proportional to the reciprocal of the determinant $\Delta = |X'X|$ where $X = \{x_{ru}\}$ and

$$x_{ru} = \left\{ \frac{\partial f(\theta, \xi_u)}{\partial \theta_r} \right\}_{\underline{\theta}=\hat{\underline{\theta}}} \tag{2}$$

Unfortunately, since we do not know the values of $\hat{\underline{\theta}}$ in advance, we do not know the derivatives x_{ru} on which the design is to be based. In most cases, however, some knowledge of the size of the θ's will be available and it was suggested [8] that preliminary guesses $\underline{\theta}^o$ should be made, and that the derivatives should be determined at these values $\underline{\theta}^o$ instead of $\hat{\underline{\theta}}$. The resulting determinant $\Delta^o = |X^{o'} X^o|$ is an explicit function of the settings of the experimental variables $\underline{\xi}$. It is therefore possible to find (perhaps analytically but, in any event, numerically) those values for $\underline{\xi}$ which maximize the determinant Δ^o.

At first sight it may seem strange that in order to use this scheme one must initially have estimates of the parameters since, after all, it is the purpose of the experiment to obtain such estimates. Actually, however, this is merely an example of the fact that any experimental design uses the

experimenter's beliefs about the situation being studied. It is thus efficient depending on whether the experimenter turns out to be nearly right.

In general, the more one knows initially the better he can design experiments. As has been pointed out by Box [3], if nothing is known about the experimental situation then strictly speaking no experiment can be planned. Or, as Daniel [14] has stated, "All experimental plans reflect what you know, what you think you know but don't, what you don't know, and what you think you don't know but do."

3. A Criterion for Design

It is usually the case in the study of physical systems that experiments can be performed sequentially; that is, information from previous experimental results can be used in planning further experiments. If this procedure is adopted all the available information about the parameters $\underline{\theta}$ after N experiments have been performed is contained in the posterior distribution function $p_N(\underline{\theta}|, \underline{y})$, and a careful analysis of the estimation situation involves a thorough study of this function.

To decide at what values the variables should be set in further experiments we select those which will yield the most desirable posterior distribution, or equivalently, will produce the most desirable modification of the present posterior distribution. Ideally, barring purely technical difficulties, one would display the various possible posterior distributions which could result from different choices of the experimental conditions and let the experimenter select that which he thought best.

This general approach to planning experiments involves no restrictions with regard to the distribution of errors, the form of the response relationship, the nature of the prior distribution, or the definition of "best posterior distribution" that can be considered. In many common situations, however, actually plotting the posterior distribution for every combination of experimental conditions for a multiparameter system would be virtually impossible. Fortunately, to do this would often be unnecessary for by making a set of plausible assumptions the posterior distribution could be completely described by a few summary statistics.

4. Assumptions

We shall consider, specifically, the situation where information from the previous N experiments is available in planning the $N+1^{th}$ and experiments are planned one at a time. Such a procedure will usually be most economical when it can be adopted.

Suppose that a posterior distribution $p_N(\underline{\theta})$ has been calculated after N observations have been obtained, and that at this stage $\hat{\underline{\theta}}_N$ are the maximum likelihood estimates of $\underline{\theta}$. Suppose that we are about to take a further observation y_{N+1} whose

true value is given by a known function of the settings for the

k variables $\underline{\xi}_{N+1}$

$$\eta_{N+1} = f(\underline{\theta}, \underline{\xi}_{N+1}) \ . \tag{3}$$

We now make two assumptions:

(1) that the y_u's are distributed Normally and independently

as

$$p(y_u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_u - \eta_u)^2} \qquad u = 1, 2, \ldots, N+1 \tag{4}$$

with mean η_u and common variance σ^2, and

(2) that for a region in the $\underline{\theta}$ space sufficiently close to the

maximum likelihood estimates $\hat{\underline{\theta}}_N$,

$$f(\underline{\theta}, \underline{\xi}_{N+1}) = f(\hat{\underline{\theta}}_N, \underline{\xi}_{N+1}) + \sum_{i=1}^{p} (\theta_i - \hat{\theta}_{iN}) \, x_{N+1}^{(i)} \tag{5}$$

where

$$x_{N+1}^{(i)} = \left[\frac{\partial f(\underline{\theta}, \underline{\xi}_{N+1})}{\partial \theta_i} \right]_{\underline{\theta} = \hat{\underline{\theta}}_N}$$

Then

$$y_{N+1} - \eta_{N+1} = y_{N+1} - f(\underline{\theta}, \underline{\xi}_{N+1}) = r_{N+1} - \underline{x}_{N+1}'(\underline{\theta} - \hat{\underline{\theta}}_N) \tag{6}$$

where we are now letting

$$r_{N+1} = y_{N+1} - f(\hat{\underline{\theta}}_N, \underline{\xi}_{N+1}). \tag{7}$$

If the observation y_{N+1} were actually available then Bayes'

formula would give for the posterior distribution of $\underline{\theta}$

$$p_{N+1}(\underline{\theta}|y_{N+1}) = \frac{p_N(\underline{\theta})\ L(\underline{\theta}|y_{N+1})}{\int p_N(\underline{\theta})\ L(\underline{\theta}|y_{N+1})d\underline{\theta}} \tag{8}$$

$$= \frac{p_N(\underline{\theta})\ e^{-\frac{1}{2\sigma^2}\{y_{N+1}-f(\theta,\underline{\xi}_{N+1})\}^2}}{\int p_N(\underline{\theta})\ e^{-\frac{1}{2\sigma^2}\{y_{N+1}-f(\theta,\underline{\xi}_{N+1})\}^2}d\underline{\theta}} \tag{9}$$

$$= \frac{p_N(\underline{\theta})\ e^{-\frac{1}{2\sigma^2}\{r_{N+1}-\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)\}^2}}{\int p_N(\underline{\theta})e^{-\frac{1}{2\sigma^2}\{r_{N+1}-\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)\}^2}d\underline{\theta}} \tag{10}$$

$$= \frac{p_N(\underline{\theta})\ e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'\underline{x}_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)]}}{\int p_N(\underline{\theta})e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'\underline{x}_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_{N+1})]}d\underline{\theta}} \tag{11}$$

This expression is true for any prior distribution $p_N(\underline{\theta})$ but to make further progress we need to be more specific about it.

The Principle of Precise Measurement

One difficulty associated with Bayes' theorem has been the question of what to take for the prior distribution. In most experimental situations this is not as troublesome as it might seem. Consider two cases with regard to the parameters θ of the response relationship itself. (1) The prior distribution is nearly constant over a region where the likelihood function has an appreciable value and the prior distribution outside of this region does not become sufficiently great so that its contribution is appreciable when combined with the likelihood function; i.e., the likelihood function dominates the prior distribution; and vice versa (2) where the prior distribution dominates the likelihood function.

In the second case almost all of the information about θ will come from the prior distribution and very little, if any, will be coming from the data. In most instances of this second kind there would be little motivation for carrying out experiments since knowledge already available is so much more precise than any that could be expected from the data. Consequently, in experimental situations case (1) is the one that usually occurs.

Since the prior distribution is virtually constant over the range where the likelihood is appreciable it is spoken of as being locally uniform. In this case it is not necessary to know the exact mathematical form of the prior distribution since it cancels out from both the numerator and denominator in Bayes' formula. The posterior distribution is then very nearly proportional to the likelihood. The principle of precise measurement [22] refers to this

situation in which most of the information comes from the data and not the prior distribution.

As Box and Tiao [9] have pointed out, our assumption that the prior distribution is locally uniform is appropriate in those situations where, if we were to compute a sensible confidence region for the parameters, then we could state honestly that a priori any point in this region would have been about as acceptable as any other. This serves to indicate that the assumption of a locally uniform prior distribution is of rather general application since a statement similar to the one above could be made in most experimental situations.

Returning to Equation 11 and appealing to the principle of precise measurement just described, we could regard the initial prior distribution as being locally uniform. This is a particularly innocuous assumption in this instance since certainly after a few observations have been taken the effect of any moderate non-uniformity a priori would have become negligible.

To obtain the posterior distribution $p_{N+1}(\underline{\theta}|\underline{y})$ by using Bayes' formula, the posterior distribution at the Nth stage $p_N(\underline{\theta}|\underline{y})$ can be used as the prior distribution $p_N(\underline{\theta})$ for the $N+1^{th}$ stage.

5. Obtaining $p_N(\underline{\theta})$

The probability density for the first N observations is

$$p(\underline{y}) = p(y_1, \ldots, y_N) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \, e^{-\frac{1}{2\sigma^2} \sum_{u=1}^{N} (y_u - \eta_u)^2} \qquad (12)$$

Using the linearity assumption (2) and Equation 7,

we have

$$y_u - \eta_u = y_u - f(\underline{\theta}, \underline{\xi}_u) = r_u - \underline{x}'_u (\underline{\theta} - \hat{\underline{\theta}}_N) . \qquad (13)$$

The likelihood function $L(\underline{\theta}|\underline{y})$ for the parameters $\underline{\theta}$ can then be written as

$$L(\underline{\theta}|\underline{y}) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \, e^{-\frac{1}{2\sigma^2} \{R_N - X_N(\underline{\theta} - \hat{\underline{\theta}}_N)\}' \{R_N - X_N(\underline{\theta} - \hat{\underline{\theta}}_N)\}} \qquad (14)$$

where

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} = \begin{bmatrix} y_1 - f(\hat{\underline{\theta}}_N, \underline{\xi}_1) \\ y_2 - f(\hat{\underline{\theta}}_N, \underline{\xi}_2) \\ \vdots \\ y_N - f(\hat{\underline{\theta}}_N, \underline{\xi}_N) \end{bmatrix}$$

$$\underline{\theta} - \hat{\underline{\theta}} = \begin{bmatrix} \theta_1 - \hat{\theta}_1 \\ \theta_2 - \hat{\theta}_2 \\ \vdots \\ \theta_p - \hat{\theta}_p \end{bmatrix}$$

$$X_N = \begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \vdots \\ \underline{x}_N' \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(p)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(p)} \\ & \vdots & & \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(p)} \end{bmatrix}$$

and

$$x_j^{(i)} = \left[\frac{\partial f(\underline{\theta}, \underline{\xi}_j)}{\partial \theta_i} \right]_{\underline{\theta} = \hat{\underline{\theta}}_N} .$$

Now the likelihood (14) is maximized if and only if

$$R'_N X_N = \underline{0} . \tag{15}$$

Consequently,

$$L(\underline{\theta} \mid \underline{y}) = \frac{1}{(\sqrt{2\pi}\sigma)^N} e^{-\frac{1}{2\sigma^2} \{R'_N R_N + (\underline{\theta} - \hat{\underline{\theta}}_N)' C_N (\underline{\theta} - \hat{\underline{\theta}}_N)\}} \tag{16}$$

where $C_N = X'_N X_N$.

Using Bayes' formula

$$p_N(\underline{\theta} \mid \underline{y}) = \frac{p_0(\underline{\theta}) L(\underline{\theta} \mid \underline{y})}{\int p_0(\underline{\theta}) L(\underline{\theta} \mid \underline{y}) d\underline{\theta}} \tag{17}$$

But if $p_0(\underline{\theta})$ is locally uniform as we have assumed

$$p_N(\underline{\theta} \mid \underline{y}) = \frac{L(\underline{\theta} \mid \underline{y})}{\int L(\underline{\theta} \mid \underline{y}) d\underline{\theta}} \tag{18}$$

$$\int L(\underline{\theta}|\underline{y}) \, d\underline{\theta} = \frac{e^{-\frac{1}{2\sigma^2} R_N' R_N}}{(\sqrt{2\pi} \, \sigma)^N} \int e^{-\frac{1}{2\sigma^2} (\underline{\theta}-\hat{\underline{\theta}}_N)' C_N (\underline{\theta}-\hat{\underline{\theta}})} \, d\underline{\theta} \qquad (19)$$

Using the well-known integral

$$\int e^{-\frac{1}{2\sigma^2} \underline{x}' A \underline{x}} \, d\underline{x} = \frac{(\sqrt{2\pi}\sigma)^p}{|A|^{\frac{1}{2}}} \qquad (20)$$

$$\int L(\underline{\theta}|\underline{y}) \, d\underline{\theta} = \frac{e^{-\frac{1}{2\sigma^2} R_N' R_N}}{(\sqrt{2\pi}\sigma)^N} \frac{(\sqrt{2\pi} \, \sigma)^p}{|C_N|^{\frac{1}{2}}} \qquad (21)$$

Hence, by substituting Equations 16 and 21 into 18, we obtain the posterior distribution of θ after N observations.

$$p_N(\underline{\theta}|\underline{y}) = \frac{|C_N|^{\frac{1}{2}}}{(\sqrt{2\pi}\sigma)^p} \; e^{-\frac{1}{2\sigma^2}(\underline{\theta}-\hat{\underline{\theta}}_N)' C_N (\underline{\theta}-\hat{\underline{\theta}}_N)} \qquad (22)$$

This expression can be used as the prior distribution $p_N(\underline{\theta})$ for the N+1th stage.

6.   The Posterior Distribution $p_{N+1}(\underline{\theta}|\underline{y})$

The posterior distribution after N+1 observations can, of course, be obtained by writing N+1 for N in Equation 22   However, for our purposes we wish to express this distribution in terms of the information available to the experimenter at stage N, the contemplated levels $\underline{\xi}_{N+1}$ of the experimental variables, and the observation y_{N+1}. This can conveniently be done by deriving $p_{N+1}(\underline{\theta}|\underline{y})$ from $p_N(\underline{\theta})$ by a further application of Bayes' theorem by substituting Equation 22 into 11. Then

$p_{N+1}(\underline{\theta}|\underline{y}_{N+1})$

$$= \frac{e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'C_N(\underline{\theta}-\hat{\underline{\theta}}_N)+(\underline{\theta}-\hat{\underline{\theta}}_N)'\underline{x}_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)]}}{\int e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'C_N(\underline{\theta}-\hat{\underline{\theta}}_N)+(\underline{\theta}-\hat{\underline{\theta}}_N)'\underline{x}_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)]} d\underline{\theta}}$$

(23)

$$= \frac{e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'C_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)]}}{\int e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'C_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)]} d\underline{\theta}}$$

(24)

$$= \frac{|C_{N+1}|^{\frac{1}{2}}}{(\sqrt{2\pi}\sigma)^p} e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N)'C_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)-2r_{N+1}\underline{x}'_{N+1}(\underline{\theta}-\hat{\underline{\theta}}_N)+r_{N+1}\underline{x}'_{N+1}C_{N+1}^{-1}\underline{x}_{N+1}r_{N+1}]}$$

(25)

$$= \frac{|C_{N+1}|^{\frac{1}{2}}}{(\sqrt{2\pi}\sigma)^p} e^{-\frac{1}{2\sigma^2}[(\underline{\theta}-\hat{\underline{\theta}}_N-C_{N+1}^{-1}\underline{x}_{N+1}r_{N+1})'C_{N+1}(\underline{\theta}-\hat{\underline{\theta}}-C_{N+1}^{-1}\underline{x}_{N+1}r_{N+1})]}$$

(26)

But

$$C_{N+1}^{-1} = (X'_{N+1}X_{N+1})^{-1} = (X'_N X_N + \underline{x}_{N+1}\underline{x}'_{N+1})^{-1} = (C_N + \underline{x}_{N+1}\underline{x}'_{N+1})^{-1} \quad (27)$$

$$= [C_N \{I + C_N^{-1} \underline{x}_{N+1}\underline{x}'_{N+1}\}]^{-1} \quad (28)$$

$$= (I + C_N^{-1} \underline{x}_{N+1}\underline{x}'_{N+1})^{-1} C_N^{-1} \quad (29)$$

We now employ a useful matrix result mentioned by Tocher [24],

$$(I_p + AB)^{-1} = I_p - A(I_q + BA)^{-1} B \tag{30}$$

where A is a $p \times q$ matrix and B is a $q \times p$ matrix. The advantage of this formula is that where q is less than p the size of the matrix that needs to be inverted is smaller on the right hand side than on the left. In our particular case $q = 1$ so that $I_q + BA$ is a scalar.

Letting

$$A = C_N^{-1} \underline{x}_{N+1}$$

$$B = \underline{x}'_{N+1} ,$$

we obtain

$$C_{N+1}^{-1} = I - C_N^{-1} \underline{x}_{N+1} [1 + \underline{x}'_{N+1} C_N^{-1} \underline{x}_{N+1}]^{-1} \underline{x}'_{N+1} \ C_N^{-1} \tag{31}$$

$$= \left\{ I - \frac{C_N^{-1} \underline{x}_{N+1} \underline{x}'_{N+1}}{1 + \underline{x}'_{N+1} C_N^{-1} \underline{x}_{N+1}} \right\} C_N^{-1} \tag{32}$$

$$= C_N^{-1} - \frac{C_N^{-1} \underline{x}_{N+1} \underline{x}'_{N+1} C_N^{-1}}{1 + \underline{x}'_{N+1} C_N^{-1} \underline{x}_{N+1}} \tag{33}$$

$$= C_N^{-1} - \frac{t t'}{1 + t' \underline{x}_{N+1}} \tag{34}$$

where t is a $p \times 1$ vector $C_N^{-1} \underline{x}_{N+1}$.

It follows that

$$\underline{x}'_{N+1} C_{N+1}^{-1} r_{N+1} = \underline{x}'_{N+1} C_N^{-1} r_{N+1} - \frac{\underline{x}'_{N+1} C_N^{-1} \underline{x}_{N+1} \underline{x}'_{N+1} C_N^{-1}}{1 + \underline{x}'_{N+1} C_N^{-1} \underline{x}_{N+1}} r_{N+1} \tag{35}$$

i.e.,

$$\underline{x}'_{N+1} C^{-1}_{N+1} r_{N+1} \doteq \frac{\underline{x}'_{N+1} C^{-1}_N r_{N+1}}{1 + \underline{x}'_{N+1} C^{-1}_N \underline{x}_{N+1}} \tag{36}$$

Thus we may write the posterior distribution after $N+1$ observations entirely in terms of quantities $\hat{\underline{\theta}}_N$ and C_N that are known before the last observation is taken and quantities r_{N+1} and $\underline{x}_{N+1}$ associated with the choice of the final experiment

$$p_{N+1}(\underline{\theta}|\underline{y}) = \frac{\left| C_N + \underline{x}_{N+1} \underline{x}'_{N+1} \right|^{\frac{1}{2}}}{(\sqrt{2\pi}\sigma)^p}$$

$$\times \; e^{-\frac{1}{2\sigma^2} \left\{ \underline{\theta} - \hat{\underline{\theta}}_N - \frac{C^{-1}_N \underline{x}_{N+1} r_{N+1}}{1 + \underline{x}'_{N+1} C^{-1}_N \underline{x}_{N+1}} \right\}' (C_N + \underline{x}_{N+1} \underline{x}'_{N+1}) \left\{ \underline{\theta} - \hat{\underline{\theta}}_N - \frac{C^{-1}_N \underline{x}_{N+1} r_{N+1}}{1 + \underline{x}'_{N+1} C^{-1}_N \underline{x}_{N+1}} \right\}}$$

$$\tag{37}$$

This posterior distribution $p_{N+1}(\underline{\theta}|\underline{y})$ is multinormal with mean

$$\hat{\underline{\theta}}_{N+1} = \hat{\underline{\theta}}_N + \frac{C^{-1}_N \underline{x}_{N+1} r_{N+1}}{1 + \underline{x}'_{N+1} C^{-1}_N \underline{x}_{N+1}} \tag{38}$$

where $\hat{\underline{\theta}}_{N+1}$ is the maximum likelihood estimate at stage $N+1$ and dispersion matrix

$$C^{-1}_{N+1} = \left[C_N + \underline{x}_{N+1} \underline{x}'_{N+1} \right]^{-1} \tag{39}$$

The expression for the dispersion matrix C^{-1}_{N+1} does not depend on y_{N+1} and can be calculated exactly at stage N for any given set of contemplated

experimental conditions $\underline{\xi}_{N+1}$. Since the posterior distribution is multinormal

this dispersion matrix contains all the information concerning the precision of

the estimates that will result from running a particular set of experimental

conditions $\underline{\xi}_{N+1}$. On the basis of C_{N+1}^{-1} the experimenter can therefore choose

the best experimental conditions for the next run.

7. A Portmanteau Criterion

In practice even the calculation of the elements of the dispersion

matrix for a number of different possible experimental conditions may prove to

be too prodigious a task. If there were only three parameters, for example,

there would be three variances and three covariances that would have to be

calculated for each set of contemplated experimental conditions. The number

n of quantities to calculate increases rapidly as p the number of parameters

increases; in fact $n = \frac{1}{2}(p^2+p)$.

It is desirable in many situations to have some kind of overall criterion

that involves the calculation of only a single quantity; however, it is clear

that as soon as one tries to economize by using such a portmanteau criterion,

every possible need cannot be satisfied.

There will be investigations where there is particular interest in one

parameter but less interest in the remaining parameters, and in other circum-

stances there may be a special reason for wanting to minimize the correlation

between a certain pair of estimates. In such

situations one could proceed by calculating the variance and covariance
term which are of special interest in addition to the overall criterion. At any
rate , if an overall criterion is adopted for sequentially planning
experiments, a facility should be provided for being able to look at
these other quantities if the experimenter so desires.

If one overall criterion must be chosen, in the absence of special
needs, it is reasonable to take for the next experiment those conditions
which give the maximum posterior density to the most probable values,
i.e., maximize the posterior density with respect to
both $\underline{\theta}$ and $\underline{\xi}_{N+1}$.

The posterior distribution after N+1 observations will be

$$p_{N+1}(\underline{\theta}|\underline{y}) = \frac{\left|C_{N+1}\right|^{\frac{1}{2}}}{(\sqrt{2\pi}\,\sigma)^p} e^{-\frac{1}{2}(\theta-\hat{\theta}_{N+1})\,C_{N+1}(\theta-\hat{\theta}_{N+1})} \tag{40}$$

The maximum probability density will be at the point $\underline{\theta}=\hat{\underline{\theta}}_{N+1}$ whatever
the settings of $\underline{\xi}_{N+1}$

$$p_{\underline{\theta}} = \max_{\underline{\theta}} \; p_{N+1}(\underline{\theta}|\underline{y}) = \frac{\left|C_{N+1}\right|^{\frac{1}{2}}}{(\sqrt{2\pi}\,\sigma)^p} \tag{41}$$

where σ is a positive constant. (The quantity $p_{\underline{\theta}}$ is necessarily
positive if C_{N+1} is positive definite.) Now $C_{N+1}= C_N +\underline{x}_{N+1}\underline{x}'_{N+1}$.
If we are at stage N, C_N is fixed but $\underline{x}_{N+1}$ is a function of $\underline{\xi}_{N+1}$ and
$\hat{\underline{\theta}}_N$, so settings can be chosen for $\underline{\xi}_{N+1}$ to maximize the determinant

$$\Delta = |C_{N+1}| = |C_N + \underline{x}_{N+1} \underline{x}'_{N+1}| \tag{42}$$

This criterion of maximizing the determinant Δ has previously been suggested in other situations and on other grounds (e.g., Box and Lucas [8], Kiefer [18, 19], and Wald [25]). A number of alternatives for an overall criterion are discussed by Kiefer [18]. Of these, the most important competitor is perhaps the trace of the dispersion matrix which is to be minimized (Elfving [16]). This criterion suffers from the disadvantage that it is not independent of the manner in which the parameters $\underline{\theta}$ are scaled and we shall not consider it further here.

Examination of the Portmanteau Criterion

The portmanteau criterion has some interesting implications which we shall now discuss. To maximize the determinant Δ we can maximize

$$\ell n \Delta = \ell n |C_N + \underline{x}_{N+1} \underline{x}'_{N+1}| \tag{43}$$

$$= \ell n |C_N| \; |I + C_N^{-1} \underline{x}_{N+1} \underline{x}'_{N+1}| \tag{44}$$

$$= \ell n |C_N| + \ell n |I + C_N^{-1} \underline{x}_{N+1} \underline{x}'_{N+1}| \tag{45}$$

But at stage N we are faced with the choice of $\underline{\xi}_{N+1}$ with C_N fixed, so we want to maximize $\ln|I+C_N^{-1}\,\underline{x}_{N+1}\,\underline{x}'_{N+1}|$.

Now if λ is a latent root of a p x p matrix A so that

$$u'\,A = u'\,\lambda \tag{46}$$

then $1 + \lambda$ is a latent root of $I + A$ since

$$u'(I+A) = u' + u'\lambda = u'(1+\lambda) \tag{47}$$

Thus

$$|I+A| = \prod_{i=1}^{p}(1+\lambda_i) \tag{48}$$

$$\ln|I+A| = \sum_{i=1}^{p}\ln(1+\lambda_i)\,. \tag{49}$$

If the λ's are < 1,

$$\ln|I+A| = \sum_{i=1}^{p}\lambda_1 - \sum_{i=1}^{p}\frac{\lambda_i^2}{2} + \sum_{i=1}^{p}\frac{\lambda_i^3}{3} - \ldots \tag{50}$$

$$= \text{tr}\,A - \frac{\text{tr}\,A^2}{2} + \frac{\text{tr}\,A^3}{3} - \ldots \tag{51}$$

where tr A is the trace of A. If the λ's are sufficiently small

$$\ln|I+A| \doteq \text{tr}\,A \tag{52}$$

Since the latent roots of $C_N^{-1}\,\underline{x}_{N+1}\,\underline{x}'_{N+1}$ will be of order $\frac{1}{N}$, maximizing $\ln|I+ C_N^{-1}\,\underline{x}_{N+1}\underline{x}'_{N+1}|$ is equivalent to maximizing

$$\text{tr}\,C_N^{-1}\,\underline{x}_{N+1}\underline{x}'_{N+1} = \text{tr}\,\underline{x}'_{N+1}\,C_N^{-1}\,\underline{x}_{N+1} \tag{53}$$

and $\underline{x}'_{N+1}\,C_N^{-1}\,\underline{x}_{N+1}$ is a 1 x 1 matrix or scalar. Thus

$$\text{tr } \underline{x}'_{N+1} \, C_N^{-1} \, \underline{x}_{N+1} = \underline{x}'_{N+1} \, C_N^{-1} \, \underline{x}_{N+1}$$

$$= c^{11} x_1^2 + c^{22} x_2^2 + \dots + c^{pp} x_p^2 + c_1^{12} x_1 x_2 + \dots c^{p-1, \, p} x_{p-1} x_p$$

$$(54)$$

where c^{ij} and $2\{c_{ij}\}^{-1}$ is the i, j^{th} element of C_N^{-1} and $\{c_{ij}\} = C_N$, i.e.,

$$c_{ij} = \sum_{u=1}^{N} \left[\frac{\partial f(\theta, \underline{\xi}_u)}{\partial \theta_i} \cdot \frac{\partial f(\theta, \underline{\xi}_u)}{\partial \theta_j} \right]_{\underline{\theta} = \hat{\underline{\theta}}_N}$$

and x_i is the i-th element of $\underline{x}_{N+1}$, i.e., $\underline{x}'_{N+1} = (x_1, x_2, \dots, x_i, \dots, x_p)$.

Therefore, to the degree of approximation employed, maximizing the determinant Δ is equivalent to maximizing the quantity

$$c^{11} x_1^2 + c^{22} x_2^2 + \dots + c^{pp} x_p^2 + c^{12} x_1 x_2 + \dots + c^{p-1, \, p} x_{p-1} x_p \, .$$

The terms c^{ii} are proportional to the variances of the estimated parameters $\hat{\theta}_i$ at stage N, and c^{ij} ($i \neq j$) to the covariances.

If there is no correlation between the estimates it is clearly desirable to make the x_i^2's as large as possible. Using the criterion of maximum Δ we are essentially weighting the x_i^2 terms with the corresponding variances $V(\hat{\theta}_i)$, and we are thus giving most weight to those terms x_i^2 which are associated with the estimates $\hat{\theta}_i$ which are known _least precisely_.

To shed some light on the role of the weighting of the cross-product terms consider the case in which there are just two parameters.

Let

$$C_N = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} \quad \text{and} \quad C_N^{-1} = \begin{bmatrix} c^{11} & c^{12} \\ c^{12} & c^{22} \end{bmatrix}$$

If c^{12} is positive (negative), i.e., the correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is positive (negative), then c_{12} is negative (positive). To maximize the cross-product term $c^{12} x_1 x_2$ if c^{12} is negative the quantities x_1 and x_2 will be chosen to be of different sign if possible. If c^{12} is positive, x_1 and x_2 will be chosen to be of the same sign if possible. This means that if correlation exists between $\hat{\theta}_1$ and $\hat{\theta}_2$ at stage N, the overall criterion we have adopted will tend to pick out a set of conditions $\underline{\xi}_{N+1}$ which will cancel out this correlation.

8. Conclusions

Under the assumptions of

(a) Normality and independence of homoscedastic errors

(Equation 4)

(b) Approximate linearity near $\hat{\underline{\theta}}_N$ (Equation 5)

(c) Locally uniform prior distribution $p_o(\underline{\theta})$

(1) The posterior distribution of $\underline{\theta}$ after N+1 observations is multinormal and is therefore completely described by the vector of least squares estimates $\hat{\underline{\theta}}_{N+1}$ and the dispersion matrix C_{N+1}^{-1} .

(2) Information on how the precision of our estimates is improved is supplied by the change in the dispersion matrix

$$C_N^{-1} - C_{N+1}^{-1} = \frac{C_N^{-1} \underline{x}_{N+1} \underline{x}_{N+1}' C_N^{-1}}{1 + \underline{x}_{N+1}' C_N^{-1} \underline{x}_{N+1}} \tag{55}$$

(3) If we are going to use the general design criterion of trying to ensure the best posterior distribution, we can calculate all the elements in the dispersion matrix (or alternatively the changes in these quantities from Equation 55) for any give values $\underline{\xi}_{N+1}$.

(4) If we are going to use the special overall design criterion of trying to choose those conditions which yield the maximum posterior density for the most probable values, we find those settings for $\underline{\xi}_{N+1}$ which maximize the determinant $|C_{N+1}|$.

(5) Maximizing the determinant $|C_{N+1}|$ is approximately equivalent to maximizing the quadratic form

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \text{Cov} (\hat{\theta}_i \hat{\theta}_j) \ x_i x_j$$

where

$$x_i = \left[\frac{\partial f(\underline{\theta}, \underline{\xi}_{N+1})}{\partial \theta_i} \right]_{\underline{\theta} = \hat{\underline{\theta}}_N} .$$

This quadratic form can be regarded as a weighted sum of the squares and cross-products of the $x_i x_j$ terms. A large value of x_i^2 is desirable in increasing the precision with which a particular parameter is estimated; and under certain conditions if this overall criterion is used, most weight is given to those quantities x_i associated with the parameters about which least is known.

9. Example

To illustrate this method for sequential design discussed above we apply it to a constructed example. A chemical reaction of the type

$$R \;\; \rightarrow \;\; P + P_1 \tag{56}$$

is being studied and the true model is

$$\eta \;=\; f(\underline{\theta}, \underline{\xi}) \;=\; \frac{\theta_3 \, \theta_1 \, \xi_1}{1 + \theta_1 \xi_1 + \theta_2 \xi_2} \tag{57}$$

where

η = the true rate of the chemical reaction

ξ_1 = partial pressure of reactant R

ξ_2 = partial pressure of product P

θ_1 = adsorption equilibrium constant for reactant R

θ_2 = adsorption equilibrium constant for product P

θ_3 = effective reaction rate constant.

This model has been reported by Laible [20] to be applicable to a number of catalytic reactions of the type $R \rightarrow P + P_1$ where the reactant R is one of certain tertiary or long chain primary alcohols, the product P is an olefin, and the product P_1 is water.

For realistic values for the parameters $\underline{\theta}$, whose values of course are assumed unknown to the experimenter, the results found by Laible [20] for the catalytic dehydration of n-hexyl alcohol at 550° F. are used, namely,

$$\theta_1 = 2.9 \qquad \theta_2 = 12.2 \qquad \theta_3 = 0.69 \tag{58}$$

We further suppose that the region of operability is defined by those values of ξ_1 and ξ_2 for which

$$0 \leq \xi_1 \leq 3$$

$$0 \leq \xi_2 \leq 3$$

If the true situation, i.e., the situation with no experimental error, were depicted as a contour diagram (See Figure 1) for η with ξ_1 as the abscissa and ξ_2 as the ordinate, then it can be seen directly from Equation 57 that all constant η contour lines are straight lines with the same intercept

$$a = -\frac{1}{\theta_2} \tag{59}$$

but different slopes b depending on the value of η

$$b = \frac{\theta_1}{\theta_2}\left(\frac{\theta_3}{\eta} - 1\right). \tag{60}$$

A contour line for a fixed value η_0 could, therefore, be represented by the equation

$$\xi_2 = -\frac{1}{\theta_2} + \frac{\theta_1}{\theta_2}\left(\frac{\theta_3}{\eta_0} - 1\right)\xi_1 \tag{61}$$

A contour line for a high value η_0 will have a small slope b.

To construct an observation y for a fixed pair of values ξ_1 and ξ_2 a random Normal deviate with standard deviation $= 0.01$ was added to the true value η obtained from Equation 57.

It was supposed that initially a 2^2 factorial design was performed with the following results:

ξ_1	ξ_2	y
1	1	0.126
2	1	0.219
1	2	0.076
2	2	0.126

Whence applying the method of non-linear least squares [21] to these data, we obtain

$$\theta_1 = 10.39 \qquad \theta_2 = 48.83 \qquad \theta_3 = 0.74 \qquad (62)$$

We are now in a position to select the settings of the levels of ξ_1 and ξ_2 for the fifth experiment. In accordance with the criterion we are using we choose those levels which maximize the determinant

$$\Delta_5 = \begin{vmatrix} \Sigma x_{1u}^2 & \Sigma x_{1u}x_{2u} & \Sigma x_{1u}x_{3u} \\ & \Sigma x_{2u}^2 & \Sigma x_{2u}x_{3u} \\ \text{sym.} & & \Sigma x_{3u}^2 \end{vmatrix} = \begin{vmatrix} c_{11}+x_{15}^2 & c_{12}+x_{15}x_{25} & c_{13}+x_{15}x_{25} \\ & c_{22}+x_{25}^2 & c_{23}+x_{25}x_{35} \\ \text{sym.} & & c_{33}+x_{35}^2 \end{vmatrix}$$

$$(63)$$

where each summation goes from u=1 to u =5 and where

$$x_{15} = \left[\frac{\partial f(\underline{\theta}, \underline{\xi}_5)}{\partial \theta_i} \right]$$ are evaluated at $\underline{\theta} = \underline{\hat{\theta}}$, the current least squares

values given in Equation 62   The quantities $c_{ij} = \sum_{u=1}^{4} x_{iu}x_{ju}$ have

fixed and known values.

We suppose that the experimenter wants values for the best

settings of ξ_1 and ξ_2 to the nearest tenth.   In this case it is feasible

to calculate the value of Δ_5 at each point of the 31 x 31 grid for ξ_1 and

ξ_2, this computation being easily programmed for a digital computer.   A

print-out of this kind which can also be set up to produce the whole or

selected parts of the dispersion matrix would normally be presented to

the experimenter for study after each run.   In this case we have assumed

that the conditions which give the maximum value for the determinant are

selected.

The maximum value for the determinant Δ_5 occurs at $\xi_1 = 0.1$ and

$\xi_2 = 0.0$. The fifth experiment was "run" with these settings and the

result was $y_5 = 0.186$. Fitting the first five observations yields

$$\hat{\theta}_1 = 3.11 \qquad \hat{\theta}_2 = 15.19 \qquad \hat{\theta}_3 = 0.79$$

(Upon comparing these estimates to those obtained after four observations, we notice that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are much closer to their true values and $\hat{\theta}_3$ is now slightly farther away from its true value.)

With these current values for $\hat{\underline{\theta}}$ and N = 5 we can maximize Δ_6 with respect to the settings of ξ_1 and ξ_2 for the sixth experiment. The computer print-out after the fifth experiment is summarized in Fig. 2 The maximum determinant Δ_6 occurs now at $\xi_1 = 3.0$ and $\xi_2 = 0.0$. The sixth experiment was "run" there, y_6 was obtained, new estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ were obtained, and so on. The results from thirteen cycles of this kind are summarized in Table 1.

Discussion

(1) The example has demonstrated the feasibility of such a scheme for sequentially designing experiments. In summary, at stage N the experimenter would in general supply the computer with (a) the model, (b) the data, and (c) the current least squares estimates for the parameters and the computer would produce (d) the new least squares estimates, (b) the best conditions for the next experiment, (c) information on the nature of the relationship between the determinant Δ and the conditions $\underline{\xi}_{N+1}$ in the neighborhood of the maximum, and perhaps (d) additional information concerning elements in the dispersion matrix corresponding to particular variances and covariances that are of special interest.

Table 1   Results from Sequential Design Example

Experiment No.	ξ_1	ξ_2	y	θ_1	θ_2	θ_3
1	1.0	1.0	0.126			
2	2.0	1.0	0.219			
3	1.0	2.0	0.076			
4	2.0	2.0	0.126	10.39	48.83	0.74
5	0.1	0.0	0.186	3.11	15.19	0.79
6	3.0	0.0	0.606	3.96	15.32	0.66
7	0.2	0.0	0.268	3.61	14.00	0.66
8	3.0	0.0	0.614	3.56	13.96	0.67
9	0.3	0.0	0.318	3.32	13.04	0.67
10	3.0	0.8	0.298	3.33	13.48	0.67
11	3.0	0.0	0.509	3.74	13.71	0.63
12	0.2	0.0	0.247	3.58	13.15	0.63
13	3.0	0.8	0.319	3.57	12.77	0.63

(2) We notice that the experimental points from the fifth onwards are all contained in three distinct regions of the factor space.   Let us designate these regions as A, B, and C where

A is the region in the neighborhood of $\xi_1 = 0.2$ and $\xi_2 = 0.0$

B "   "   "   "   "   "   " $\xi_1 = 3.0$ " $\xi_2 = 0.0$

C "   "   "   "   "   "   " $\xi_1 = 3.0$ " $\xi_2 = 0.8$

These sites, A, B, and C are shown in Figure   1.   The experimental points in turn fall into the regions

ABABACBAC

Since there are three parameters we would expect that there would

be at least three sites for observation points and, furthermore, an

optimal design might require a different number of points at each site,

as in the case above.

(3) Once again, as is true in most situations in which a maximum

is being sought (e.g., in determining maximum yield conditions

by using response surface methods or,

in finding the maximum likelihood estimates), it is useful to determine

not only the point in the operability region for which the value of the

determinant is maximized but also the nature of the dependence of the

determinant on the variables $\underline{\xi}$ in the neighborhood of this maximum.

A print-out of the type we are using (Figure 2)   indicates far

more than the point at which the determinant is maximized and enables

the experimenter to use informed judgment in selecting his actual runs;

for instance, he might find that the determinant fell off very rapidly in

the direction of the variable ξ_2 indicating that the control of this variable

is of critical importance relative to the others.   Information of this kind

can be of considerable value to the experimenter.

(4) The example further illustrates the fundamental difference between

the problem in which the object is in estimating the response η and

the problem in which the object is to estimate parameters $\underline{\theta}$.   After four

experiments, for example, although the parameter estimates $\hat{\theta}_1$ and $\hat{\theta}_2$

are widely discrepant from the true values, nevertheless the estimates for

η that are produced in the region of the experimental design are in close

agreement with the true values.

The point that is illustrated here is compensating nature of the errors

in $\hat{\theta}_1$ and $\hat{\theta}_2$. Although both estimates are much greater than their corresponding

true values the errors are such that when these estimated values are substituted

into Equation 57 in place of θ_1 and θ_2 the results in terms of estimated

values of η over the region in which the data have been taken are in close

agreement with the true values. Such correlation among the estimates is

common in models such as Equation 57; in particular, this correlation is

characteristic of estimates of parameters in catalytic reaction kinetics since

the models are often of this form.

References

1. Box, G. E. P. (1954), "The exploration and exploitation of response surfaces: some general considerations and examples," Biometrics, 10, 16-60.

2. Box, G. E. P. (1958), "Use of statistical methods in the elucidation of basic mechanisms," Bull. Inst. Intern. De Statistique, 36, 215-225.

3. Box, G. E. P. (1960), "Fitting empirical data," Annals of the New York Academy of Sciences," 86, 792-816.

4. Box, G. E. P., and Coutie, G. A. (1956), "Application of digital computers in the exploration of functional relationships," Proc. I. E. E., 103, Part B, Suppl. No. 1, 100-107.

5. Box, G. E. P., and Draper, N. R. (1959), "A basis for the selection ; of a response surface design," J. Am. Stat. Assoc., 54, 622-54.

6. Box, G. E. P., and Hunter, J. S. (1957), "Multifactor experimental designs for exploring response surfaces," Ann. Math. Stat., 28, 195-241.

7. Box, G. E. P., and Hunter, W. G. (1962), "A useful method for model-building," Technometrics, 4, 301-318.

8. Box, G. E. P., and Lucas, H. L. (1959), "Design of experiments in non-linear situations," Biometrika, 46, 77-90.

9. Box, G. E. P., and Tiao, G. C. (1963), personal communication.

10. Box, G. E. P., and Youle, P. V. (1955), "The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system," Biometrics, 11, 287-323.

11. Chernoff, H. (1953), "Locally optimal designs for estimating parameters," Ann. Math. Stat., 24, 586-602.

12. Cox, D. R. (1961), "Tests of separate families of hypotheses," Proc. Fourth Berkeley Symp., 1, 105-123.

13. Cox, D. R. (1962), "Further results on tests of separate families of hypotheses," J. R. Stat. Soc. (B), 24, 406-424.

14. Daniel, C. (1963), "Factor screening in process development," Industrial and Eng. Chem., 55, 45-48.

15. Davies, O. L. (Ed.) (1956), <u>Design and Analysis of Industrial experiments</u>, Hafner, New York, 495-578.

16. Elfving, G. (1952), "Optimal allocation in linear regression theory," Ann. Math. Stat., $\underline{23}$, 255-262.

17. Hunter, W. G., and Mezaki, R. (1963), "A model-building technique for chemical engineering kinetics," A. I. Ch. E. Journal, to appear.

18. Kiefer, J. C. (1959), "Optimum experimental designs," J. R. Stat. Soc. (B), $\underline{21}$, 272-319.

19. Kiefer, J. C. (1961), "Optimum experimental designs V, with applications to systematic and rotatable designs," Proc. Fourth Berkeley Symp., $\underline{1}$, 381-405.

20. Laible, J. R. (1959), Ph. D. Thesis, University of Wisconsin, Madison, Wisconsin.

21. Meeter, D. A. (1961), "Non-linear estimation program," Numerical Analysis Laboratory, University of Wisconsin, Madison, Wisconsin.

22. Savage, L. J., and others (1962), <u>The Foundation of Statistical Inference</u>, John Wiley and Sons, New York.

23. Stone, M. (1959), "Application of a measure of information to the design and comparison of regression experiments," Ann. Math. Stat., $\underline{30}$, 55-70.

24. Tocher, K. D. (1951), "Discussion on Mr. Box and Dr. Wilson's paper," J. R. Stat. Soc. (B), $\underline{13}$, 39-42.

25. Wald, A. (1943), "On the efficient design of statistical investigations," Ann. Math. Stat., $\underline{14}$, 134-140.
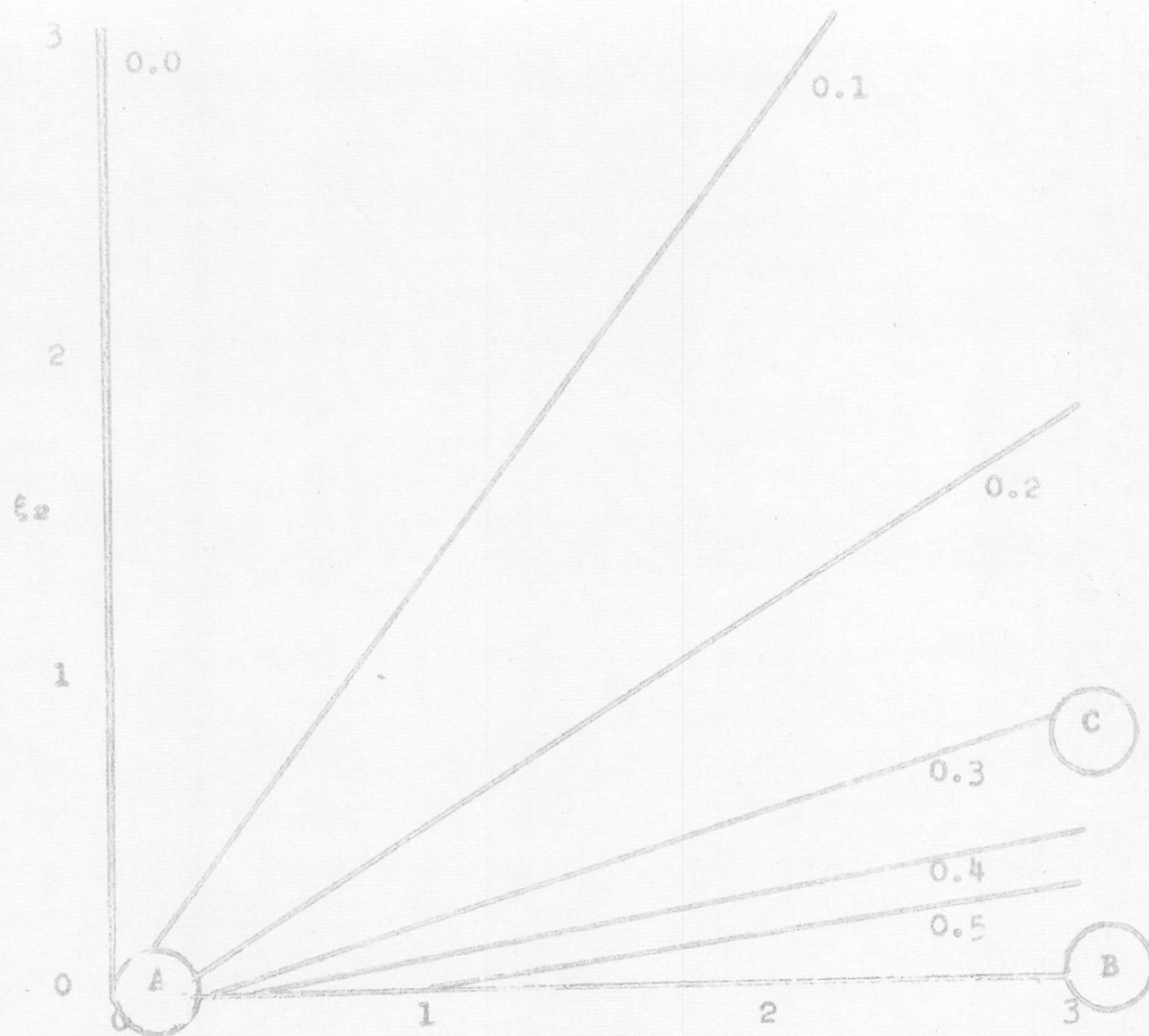
Figure 1.   Contour diagram of η vs. ξ_1 and ξ_2 for the model $\eta = \theta_1\theta_3\xi_1/(1+\theta_1\xi_1+\theta_2\xi_2)$

Figure 2 Condensation of Computer Print-Out of 31×31 Grid of Determinant $\Delta_6 \times 10^{10}$ as a Function of ξ_1 and ξ_2

ξ_2							
3.0	4	4	5	5	5	5	5
2.5	4	4	5	5	6	5	5
2.0	4	5	5	5	5	5	6
1.5	4	5	5	5	5	6	8
1.0	4	5	5	5	8	15	27
0.5	4	5	7	22	57	112	182
0	4	389	1155	1737	2154	2460	2693
	0	0.5	1.0	1.5	2.0	2.5	3.0

ξ_1