Optimizing Methods in Simulation

J. S. Rustagi

Department of Statistics
The Ohio State University
1958 Neil Avenue, Columbus, OH 43210

Office of Naval Research
Department of Navy
Arlington, Virginia 22207

Distribution of this document is unlimited.

Simulation, optimizing techniques, response surface, stochastic approximation, optimal design.

Simulation techniques utilize methods of optimization in several aspects. Validation of simulation models, estimation of models and design of simulation experiments require optimization. Methods of optimization are discussed with special applications to simulation. Criteria of optimization, classical methods of optimization, numerical methods, and optimal search procedures are discussed.
1. **Introduction**

Simulation of various physical, biological or social complex systems allows us to develop elaborate models for them and helps in the process of making valid inferences from them. There are many situations in which systems cannot be easily described in a compact form for analysis and prediction. The modern computer simulations allow us to represent such systems by series of simpler models and thus help us in providing reasonable solutions to complex problems.

A schematic representation of the simulation strategy for developing models of complex systems has been given by Ziegler et al. (1979).

![Diagram](image)

**Figure 1**

The validation of models requires some sort of optimization. One has to provide criteria of optimization and possible techniques to achieve that optimization to
complete the process of validation.

Optimizing techniques are also required in other aspects of simulation experiment such as in their design and ultimate analysis. The method of optimization form a vast body of knowledge spreading over several fields. We have classified optimizing methods broadly in the following categories and have arranged the list of references in that order.

A. Classical optimizing techniques;
B. Numerical procedures;
C. Mathematical programming methods;
D. Stochastic approximation methods;
E. Optimum seeking methods and response surface methodology;
F. Optimal Design Theory;
G. Miscellaneous methods.

In this brief account, we emphasize those optimization techniques which are of potential use in simulation methodology. We shall concentrate here on Optimizing Criteria, Classical Methods, Numerical Methods, Optimal Search Procedures, Response Surface Methods and Optimal Designs of Regression Experiments.

However, technical references are provided on various other optimization techniques for the interested reader.
2. **Optimizing Criteria**

Optimization is basically dependent on the criteria used in a given situation. The same problem may lead to different solutions depending upon the criteria of optimality utilized. The criteria depend on the nature of the problem and are many times dictated by practical considerations. Consider the case of least square estimation of parameters in hypothesized models. The criterion of minimizing the sum of squares of residuals, was dictated more as a mathematical convenience than from heuristic point of view. It allows simple mathematical solution in most cases. However, if the criterion of optimality is chosen to be that of minimizing the sum of absolute deviations of residuals, the mathematical simplification is minimal and recourse has to be made to numerical solutions. It may be highly important to select the "right" criterion of optimality in a given situation.

There does not seem to be a simple and logical approach of choosing among a class of competing criteria of optimality for a given problem. Experience and intuition in a given setting may be the ultimate judge for proper selection. In many situations, however, more is known about the comparative properties of the optimality...
criteria and the experimenter is guided by such considerations to select the appropriate criterion. We shall discuss some of the most commonly used criteria in this section.

Least Squares Criterion

One of the most common criterion used in validation of models is that of least squares. Given the realization of the process from simulations or actual observations, the observed and the expected value under the assumed models are compared. If the sums of squared deviation is minimized, this method provides the unknown parameters of the model. Various other criteria such as sum of absolute deviations or weighted least squares criterion are also in use. The criterion to be chosen heavily depends on the experimental situation.

Example (Milstein (1979))

In a biochemical process, the equations of the process are described by the following

\[
\frac{dx_i}{dt} = f(x, k),
\]

\[x(0) = \xi_i,
\]

\[i = 1, 2, \ldots, l,
\]
and the vector $\chi$ is n-dimensional with nonnegative components, $\kappa$ is a vector of parameters having p unknown components, $f$ is a vector function. The vector $\zeta$ represents the given initial condition. Let the data be given by $y^S(t_r)$ at $r$th time point $t_r$ and let the corresponding value of $x$ be given by $x(\kappa, t_r)$. Let $W_r$ be the matrix of known weights, then a common measure of the discrepancy between the data points $y$ and the trajectories can be the following

$$F(\kappa) = \sum_{s=1}^{M} \sum_{r=1}^{M} [y^S(t_r) - \chi^S(\kappa, t_r)] W_r [y^S(x_r) - \chi^S(\kappa, t_r)]$$

$M$ is the number of points chosen. The object will be to determine the unknown parameters $\kappa$ which can be obtained by using the criterion of minimizing $F(\kappa)$.

A computer algorithm is given in terms of an iterated numerical procedure starting with a first guessed value of $\kappa$ by Milstein (1979).

In the context of design of experiments, which are highly pertinent to the simulation experiments, we discuss a few criteria which are in commonly use.

Consider the model,

$$\chi = X\beta + \epsilon$$
where $y$ is the observation vector in $n$-dimensions, $X$ is an $n \times p$ design matrix, $\beta$, is $p \times 1$ vector of unknown parameters and $\xi$, is $n \times 1$ vector of residuals. If we use least squares method to estimate $\beta$, it is well known that we optimize $\xi' = (y-X\beta)'(y-X\beta)$ leading to the optimal estimates of $\beta$ as given by 
\[ \hat{\beta} = (X'X)^{-1}y \]

In the problem of finding optimum $X$ such that the parameter $\beta$ is estimated optimally, one considers the covariance matrix of $\hat{\beta}$ given by 
\[ V(\hat{\beta}) = (X'X)^{-1}\sigma^2 \]

where $\xi$ is assumed to have means zero and covariance $\sigma^2I$.

By an experimental design, we mean the choice of levels of $X$. Consider the case in one dimension for present and assume that there are $n$ observations available. We are interested in knowing the method of allocation of these observations to the various levels of $x$'s. That is, the problem is find levels $x_1, x_2, \ldots, x_k$ to be repeated $n_1, n_2, \ldots, n_k$ times such that $n_1 + n_2 + \ldots + n_k = n$. The set of $x_i$'s with $n_i$'s is called the design of an experiment. In place of integers $n_i$, we can use fractions $P_1, P_2, \ldots, P_k$. 


with \( \frac{n_i}{n} = p_i \) and \( \sum p_i = 1 \). The collection of \( x_i \)'s with \( p_i \)'s describes generally a discrete probability measure. The theory of optimal design of experiments is concerned with obtaining such a measure so as to optimize some objective function of the parameters in the assumed model for the experiment.

There are several optimality criteria in the case of regression design of experiments and they are given in terms of the matrix \( X'X \). Suppose \( X = (x_1', x_2', \ldots, x_n') \), with \( x_i, i = 1, 2, \ldots, n \) being p-vectors and let \( x_i \in X \).

**Criterion of G-Optimality**

It is also known as the criterion of minimax optimality.

Find \( x_i \) such that

\[
\min_{\xi_i} \max_{\xi \in X} \{X'(X'X)^{-1}X\}
\]

\( i=1, 2, \ldots, n \)

**Criterion of D-Optimality**

In this criterion, we find \( x_i \) such that determinant of the matrix \( X'X \) is maximized. That is, find \( x_i \), such that we have
\[
\max_{\mathbf{x}_i} \det (\mathbf{X}'\mathbf{X})
\]

\textbf{Criterion of A-Optimality}

Find \( \mathbf{x}_i \) such that

\[
\min_{\mathbf{x}_i} \text{trace} (\mathbf{X}'\mathbf{X})^{-1}
\]

\textbf{Criterion of E-Optimality}

This criterion is concerned with finding \( \mathbf{x}_i \) such that

\[
\max_{\mathbf{x}_i} (\min \text{eigenvalue of } \mathbf{X}'\mathbf{X})
\]

Many other kinds of optimality criterion in the context of design of regression experiments have been discussed in the literature, for reference, see Federov (1972).

\textbf{Integrated Mean Square Error Criterion}

Recently Brown (1979) has proposed the integrated mean square error as an optimization criterion in the context of linear inverse.

This criterion has been used in other contexts as well, see Tapia and Thompson (1978). A common measure of discrepancy between the observed and expected value is obtained in terms of mean squared errors (MSE).
Consider the model,
\[
E(Y|x) = \alpha + \beta x
\]
and
\[
V(Y|x) = \sigma^2
\]
Let \( L < x < U \), be the interval of possible \( x \) values. The MSE(\( x \)) is the mean squared error of \( x \) as obtained from \( y \).
Let \( w(x) \) be a weight function. Then Integrated Mean Square Error is defined as
\[
\text{IMSE} = \int_{L}^{U} \text{MSE}(x) \cdot w(x) \, dx
\]
In calibration problems, Brown has shown that optimization of IMSE gives much better results as compared to simply minimizing MSE. In case, no special form of the weight function \( W(x) \) is suggestible from the problem, \( W(x) \) may be taken to be uniform over the range \( (L, U) \).

3. Classical Methods of Optimization

The basic problem of optimization is concerned with finding a value \( x_0 \) in a finite dimensional set \( A \), for which a function \( f(x) \) defined on the set \( A \), attains a maximum or a minimum. If \( A \) is a finite set, the minimizing and maximizing values always exist. They need not exist when \( A \) is not finite.
Suppose \( f(x) = \begin{cases} 1, & x = 0, \\ x, & x > 0. \end{cases} \)

Then the function \( f(x) \) defined on \( x \geq 0 \), the non-negative part of the real line does not have a minimum which can be attained. The ideas of infimum and supremum are introduced to take care of such a situation.

Define supremum of \( f(x) \) or \( \text{sup } f(x) \) by the least value of \( \lambda \) such that

\[
f(x) \leq \lambda \quad \text{for all } x \in A.
\]

Similarly infimum of \( f(x) \) or \( \text{inf } f(x) \) is defined by the largest value \( \lambda \) such that \( f(x) \geq \lambda \).

An important result in this regard is given by the following theorem.

**Theorem 3.1.** If \( f(x) \) is continuous and the set \( A \) is finite and closed interval then \( f(x) \) attain its minimum or maximum (extrema) values in \( A \).

For proof, see any book on calculus, for example, Whittle (1971).

The necessary and sufficient conditions for extrema are given by the following theorems, usually available in standard calculus books.
Theorem 3.2. (Necessary Conditions for an extremum)

If the derivative \( \frac{\partial f}{\partial x} \) exist at an interior point, \( x_0 \), of the set \( A \), and if \( x_0 \) is an extremum point, then \( \frac{\partial f}{\partial x} = 0 \) at \( x = x_0 \).

Define the Hessian of a function \( f(x) \) by the matrix of second order partial derivations as follows.

\[
H = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{bmatrix}
\]

Theorem 3.3. (Sufficient Condition for an extremum)

The sufficient condition that \( f(x) \) has a maximum (minimum) at an interior point \( x_0 \in A \) is that \( H \) exist and be negative definite (positive definite).

The proofs require expanding the function \( f(x) \) with the help of Taylor's theorem using \( H \). For details see, Whittle (1971).
Constrained Optimization

In finding extrema of a function $f(x)$ over the set $A$, these may be additional constraints added such as by the condition, $g(x) = b$. Essentially the constraints introduce a subset of the set $A$ over which $f(x)$ should be optimized. The case when the constraints are introduced by inequalities is dealt with by mathematical programming methods.

The method of Lagrange multipliers has been used extensively for solving constrained optimization problems. The method requires optimizing

$$f(x) + \lambda g(x)$$

where $\lambda$ is some unknown constant. If the number of constraint equation is more than one, Lagrange's method requires optimizing

$$f(x) + \lambda \xi' g(x)$$

where $\xi$ is the vector of function given and the vector $\lambda$ is unknown. For an extensive discussion, see Whittle (1971).
4. Numerical Methods of Optimization

By the very nature of the simulation process, numerical methods are necessary for optimizing techniques for simulation models. In the case of functions of one variable, it may sometimes be easy to graph the function and then obtain the optimizing value. In the case of several variables, the process involves large numbers of calculation and may exceed the limit of computers.

The optimization of functions in many cases reduces to finding the solutions of equations since the extremizing values are given by the derivatives or partial derivatives if they exist. We first consider methods of solving an equation of the type,

\[ f(x) = 0 \]  

(4.1)

General methods for solutions are available in textbooks of numerical analysis, for example see Ralston (1965). We first define Lagrange polynomials which are used in interpolation. Lagrange polynomials of \((n-1)\)-th degree are defined by

\[ \ell_j(x) = \frac{p_n(x)}{(x-a_j)p_n'(a_j)}, \quad j = 1, 2, \ldots, n \]  

(4.2)

where

\[ p_n(x) = (x-a_1)(x-a_2) \ldots (x-a_n) \]  

(4.3)
is a polynomial of nth degree with given constants $a_1, a_2, \ldots, a_n$.

$p_n'(a_j)$ gives the derivative of the polynomial $p_n(x)$ at $a_j$. For example, Lagrange polynomials of order 3 are given by

$$
\ell_1(x) = \frac{(x-a_2)(x-a_3)}{(a_1-a_2)(a_1-a_3)} \quad (4.4)
$$

$$
\ell_2(x) = \frac{(x-a_1)(x-a_3)}{(a_2-a_1)(a_2-a_3)} \quad (4.5)
$$

$$
\ell_3(x) = \frac{(x-a_1)(x-a_2)}{(a_3-a_1)(a_3-a_2)} \quad (4.6)
$$

Iterative procedure for roots of the equation, $f(x) = 0$.

Suppose inverse of the function $f$ exists. Let $y = f(x)$ so that $x = f^{-1}(y) = g(y)$. We are looking for $g(0)$ which will be the root $a$. That is, $g(0) = a$.

The Lagrange interpolation formula gives an approximation for $g(y)$ by $h(y)$, denoted by, $g(y) \approx h(y)$.

$$
h(y) = \sum_{j=1}^{n} \ell_j(y) g(y_j)
$$

$$
= \sum_{j=1}^{n} \ell_j(y) x_{i-j+1} \quad (4.7)
$$

where $g(y_j) = x_{i-j+1}$, given the points, $y_1, y_2, \ldots, y$.

An approximation of $a$ by $x_{i+1}$ is given by $h(0)$. That is,

$$
x_{i+1} = \sum_{j=1}^{n} \ell_j(0) x_{i+1-j} \quad (4.8)
$$
Notice that
\[
\ell_j(0) = \frac{(-1)^ny_1y_2\cdots y_{j-1}y_{j+1}\cdots y_n}{(y_j-y_1)(y_j-y_2)\cdots(y_j-y_{j-1})(y_j-y_{j+1})\cdots(y_j-y_n)}
\]  
(4.9)

The equation (4.9) gives an n-point iteration process. That is given \(x_i, x_{i-1}, \ldots, x_{i-(n-1)}\), we can find \(x_{i+1}\). Or the n-point iteration function is given by
\[
x_{i+1} = F_i(x_i, x_{i-1}, \ldots, x_{i-(n-1)})
\]  
(4.10)

Most iteration procedures use only one point iteration and the same function for iteration. That is,
\[
x_{i+1} = F(x_i)
\]  
(4.11)

There are many methods of iteration. We shall discuss here the most commonly used methods such as those of Newton-Raphson.

**Newton-Raphson Procedure**

In this procedure, we use
\[
x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}
\]  
(4.12)

Using the approximation of \(f'(x_i)\) by
\[
f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}
\]  
(4.13)

we have
\[
x_{i+1} = x_i - \frac{f(x_i)(x_i - x_{i-1})}{f(x_i) - f(x_{i-1})}
\]
The above two-point iteration is known as the secant method.

In the case of several equations, a generalized one-point Newton-Raphson iteration procedure can be similarly described. Let \( x = (x_1, x_2) \) be a two-dimensional vector. Let \( f_1(x) = 0 \) and \( f_2(x) = 0 \) be the two simultaneous equations to be solved. Then the Newton-Raphson iteration requires the following:

\[
\begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}
\end{bmatrix}
^{-1}
\begin{bmatrix}
f_1(x_1) \\
f_2(x_2)
\end{bmatrix}
= x_{i+1} - x_i
\]

\( (4.15) \)

\[ x = x_i \]

**Gradient Method**

Gradient method was introduced by Cauchy in 1847. This method utilizes the gradient of the function \( f(x) \) given by \( \ell(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n})' \). The gradient represents the direction cosines of the normal to the tangent hyperplane at point \( x \) of the surface \( f(x) \). The method utilizes steepest ascent for a maximum and steepest descent for a minimum.
Descent for a minimum, to increase the speed of approach to the optimum. Consider the matrix

\[ d^2 = (\chi - \gamma')' B(\chi - \gamma) \]

where \( B \) is a given matrix and \( \chi \) and \( \gamma \) are any two vectors. Then the direction of steepest ascent is the direction from the point \( x_0 \) to the ellipsoid

\[ (x - x_0)' B(x - x_0) = k^2. \]

The following theorem gives an explicit form for optimization.

**Theorem 4.1.** For a function \( f(x) \), the maximum occurs in the direction \( \delta(x_0) \) given by \( \delta(x_0) = B^{-1}(\ell(x_0)) \) where \( \ell(x_0) \) is the gradient of \( f(x) \) at \( x_0 \). For proof and other relevant material the reader is referred to Crockett and Chernoff (1955).

5. **Optimal Search Procedures**

In optimum seeking methods, the aim is to design the most economic or shortest time consuming procedure. Suppose a function is to be explored over the points \( x_1, x_2, \ldots, x_n \). Let \( 0 \leq x_i \leq 1 \). Consider the following two situations with \( n = 3 \) in Figures 1 and 2, where the values of the function are given by verticle lines.
In Figure 1, the maximum may be in the interval \((0, x_2)\) and in Figure 2, it be in the interval \((x_1, x_3)\). Such an interval is called the interval of uncertainty. In general, the interval of uncertainty is \( (x_{k-1}, x_{k+1}) \). The length of the interval of uncertainty is given by \( l(x_k, k) \).

Several search plans based on the interval of uncertainty are given below.

**Minimax Search.** A plan which minimizes the maximum interval of uncertainty. That is,

\[
\min_{x_1, x_2, \ldots, x_n} \quad \max_{1 \leq k \leq n} \quad l_n(x_k, k).
\]

**Uniform Pairs Search.** It requires that the intervals chosen should be of uniform length. One such plan is to take

\[
x_k = \frac{(1+\epsilon)[k+\frac{1}{2}]}{n+1} - \left(\left\lfloor \frac{k+\frac{1}{2}}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor\right)\epsilon
\]

where \([a]\) denotes the integral part of \(a\).
Other plans including the **Fibonacci Search** or **Sequential Search** plan, which is based on the Fibonacci sequence, and the **Golden Section Search** plan, are also used in practice. For literature on optimum seeking methods, see Wilde (1964). An important class of optimum seeking procedures is concerned with optimizing the regression function in statistics. Such procedures have become known as **Response Surface Methods**. We shall discuss some elements of this methodology in the next section.

6. **Response Surface Methods**

The response surface methodology was developed to solve some problems in chemical investigation. However, its use became universal and in simulation methodology response surface techniques are very commonly used. The problem can be stated as follows. Let a region $R$ of $k$ dimensions be called the **factor space** of with points $x = (x_1, x_2, \ldots, x_k)'$. Let the mean, $\mu$, of a response $y_u$ depend on the factors $x_u$ through the function $\phi$.

$$
\mu_u = \phi(x_u).
$$

(6.1)

Let $y_u$ have variance $\sigma^2$. The problem then is to find a point $x^*$ in the smallest number of experiments so as to
optimize \( u_0 \) over the region \( R^k \).

This classical problem was stated by Hotelling (1941) and Friedman and Savage (1947). Box and Wilson (1951) provided the basic framework to develop optimal response surface designs and their techniques have found considerable use in many applications. Myers (1971) has collected the available material in a book on response surfaces. We discuss elements of response surface methodology based on the paper of Box and Wilson. One of their major contributions was to develop new types of designs in place of complete factorial designs.

Let the distance, \( r \), from the origin to the point \( x \) be Euclidean, with
\[
r^2 = \sum x_i^2.
\] (6.2)
The object here is to choose \( x \) in such a way that
\[
\phi(x) - \phi(0)
\] (6.3)
is maximized with the constraints in (6.2).

Using Lagrange's method, we maximize
\[
\mathcal{L}(x) = \phi(x) - \phi(0) - \frac{1}{2} \lambda \sum x_i^2.
\] (6.4)
The stationary solution is given by equating to zero the partial derivatives with respect to \( x_i \). We have
\[
\lambda x_i = \frac{\partial \phi}{\partial x_i} (x).
\] (6.5)
Squaring and summing over all i and simplifying, we get
\[
\lambda = \frac{1}{n} \left( \sum_{i=1}^{n} \left( \frac{\phi(x)}{\delta x_i} \right)^2 \right)^{1/2}
\]  
(6.6)

That is, the maximizing point should have coordinates proportional to the derivatives of \( \phi \).

Suppose the conditions of Taylor’s expansion for \( \phi(x) \) in the neighborhood of the origin hold, then \( \phi(x) \) can be expanded to linear, quadratic and higher order terms. If we assume that second and higher order terms in the expansion of \( \phi \) are zero, then, \( \phi(x) \) is approximated by a linear function of the following type:
\[
\phi(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k
\]  
(6.7)

Then,
\[
\frac{\partial \phi(x)}{\partial x_i} = \beta_i, \quad i = 1, 2, \ldots, k
\]  
(6.8)

and the optimal \( x_i \) are proportional to \( \beta_i \). Similarly expressions involving coefficients of linear and quadratic terms can be obtained if the Taylor’s expansion of \( \phi(x) \) does not contain third and higher order terms. The move along the derivatives of the response function gives the steepest ascent approach to a maximum.
For the sake of clarity of presentation, we assume $k = 2$. Suppose $\phi(x)$ has third and higher order derivatives zero. Hence we represent $\phi(x)$ as follows:

$$\phi(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2$$

(6.9)

Using the usual least squares theory, the regression equations (6.9) can be estimated by at least six or more points, since there are six unknown constants. As a rule, one would consider a complete 3x3 factorial experiment.
with nine points so as to provide estimates for the quadratic regression (6.9). However, Box and Wilson provided a design, not of the factorial type which has five points on the vertices of a pentagon and the sixth at the origin. Such a design would give the estimates of the coefficients in the regression model and hence about the derivatives. These estimates then can be used to define the path of steepest ascent.

Several designs such as fractional factorials have also been used in response surface techniques and are available in text books on design of experiments, for example, see Kempthorne (1978) and Myers (1971) providing a large number of new designs which are commonly applied in response surface methodology.

7. Optimal Design of Regression Experiments

The theory of optimal design of regression experiments is concerned with choosing the levels of the independent variable $x$ for the model

$$y = f(x)$$

so as to optimize a certain function of parameters to be estimated in the model. We have given several optimality criteria as commonly used in optimal design theory in
Section 2. In simulation studies such criteria assume further importance since the design of a simulation may require several replications in a given problem. There is an extensive literature on optimality of designs. For a recent survey, see Federov (1972). Reviews of various other aspects of optimal designs have been presented more recently in the statistical literature. A review of D-optimality for regression designs has been given by St. John and Draper (1975) with an extensive bibliography.

A typical problem of optimal design theory is of the following type.

**Example:**

Consider the simple linear regression model

\[ y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n \quad (7.1) \]

We assume that the errors \( \epsilon_i \) are uncorrelated and have common variance \( \sigma^2 \). Let

\[ (y_1, y_2, \ldots, y_n)', \quad \theta = (\theta_0, \theta_1)' \quad (7.2) \]

and

\[ \chi = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad (7.3) \]
Using the general linear model and results in Section 2, we find

\[ S = X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \] (7.4)

and

\[ S^{-1} = a \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \] (7.5)

\[ X'Y = \begin{bmatrix} n \sum y_i \\ \sum x_i y_i \end{bmatrix} \] (7.6)

where

\[ a = \frac{1}{n \sum (x_i - \bar{x})^2} \] (7.7)

The estimates are given by

\[ \hat{\theta} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix} = a \begin{bmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ -\sum x_i \sum y_i + n \sum x_i y_i \end{bmatrix} \] (7.8)

and

\[ \text{Cov}(\hat{\theta}_0, \hat{\theta}_1) = -ao^2 \sum x_i \] (7.9)

\[ V(\hat{\theta}_0) = ao^2 \sum x_i^2 \] (7.10)

\[ V(\hat{\theta}_1) = ao^2 n^{-1} \] (7.11)
Suppose \( V(\hat{\theta}_1) \) is to be minimized to obtain optimal \( x_i \)'s. That is, the optimization problem is to maximize
\[
\sum (x_i - \bar{x})^2. \tag{7.12}
\]
Assuming that \( x \)'s are between -1 and 1, the solution to the above problem is that \( x \)'s should be placed at -1 and 1, half at each place to make (7.12) a maximum. For D-optimality, we maximize the determinant of \( S \). That is, again we maximize
\[
n\sum (x_i - \bar{x})^2. \tag{7.13}
\]
Hence the same answer obtains as in minimizing the variance of \( \hat{\theta}_1 \).

Comparisons of Optimality Criteria

G-optimality (minimax optimality) was introduced by Smith (1918) and was exploited by Kiefer and Wolfowitz (1959). Wald (1943) used the criterion of D-optimality in some other context and was so named by Kiefer and Wolfowitz (1959). One of the most important results in optimal design theory is the equivalence and characterizations of G-optimality and D-optimality under various conditions. This was established by Kiefer and Wolfowitz. Recently such results have also been extended to nonlinear models by White (1973). Various computer algorithms
to generate D-optimum designs are available in the literature. Essentially the algorithm of Federov (1972), requires the following steps:

1) Select any non-degenerate starting design,
2) Compute the dispersion matrix,
3) Find the point of maximum variance,
4) Add the point of maximum variance to the design, with measure proportional to its variance
5) Update the design measure.

For further details, the reader is referred to the exposition by St. John and Draper (1975).

8. References

A. Classical Methods


Hancock, Harris (1917). Theory of Maxima and Minima. (Dover, New York, 1960).


Lagrange, Conte Joseph Louise (1759). Researches sur la
méthode de maximis et minimis. Miscellanea
Taurinensis 1: 3-20.

eigenvalues of a matrix and optimality of principal

Sons, New York.

B. Numerical Procedures

Beltrami, E. J. (1970). An Algorithmic Approach to Non-
linear Analysis and Optimization. Academic Press,
New York.

Box, M. J. (1966). A comparison of several current
optimization methods and the use of transformations
in constrained problems. The Computer Journal 2,
66-77.

Businger, P. and Golub, G. H. (1965). Linear least
squares solution by Householder transformation.
Numerische Mathematik. 7: 269-276.

experiments with various optimal estimators.
1: 113-130.


C. Mathematical Programming


Francis, Richard L. (1971). On relationships between the
Neyman-Pearson problem and linear programming.
Optimizing Methods in Statistics. (Editor, J. S.

Hadley, G. (1962). Linear Programming. Addison-Wesley,
Reading, Massachusetts.

Addison-Wesley, Reading, Massachusetts.

Hartley, H. O. (1961). Nonlinear programming by the

Kiountouzis, E. A. (1973). Linear programming techniques

Hill Co., New York.


Rustagi, J. S., Editor (1979). Optimizing Methods in

Rustagi, J. S., Editor (1971). Optimizing Methods in

Rustagi, J. S. (1968). Dynamic programming model of


D. Stochastic Approximation Methods


E. Optimum Seeking Methods and Response Surface Methodology


F. Optimal Design Theory


Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* 12, 1-85.


