5-1967

Design Optimization Using Model Estimation Programming

Richard Kay Brimhall
Utah State University

DESIGN OPTIMIZATION USING MODEL ESTIMATION PROGRAMMING

by

Richard Kay Brimhall

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1967

ACKNOWLEDGMENTS

The material presented in this thesis was developed for application as an optimization procedure at the Wasatch Division of the Thiokol Chemical Corporation.

The development of new ideas and methods of application seldom involves a single individual. Particular thanks are due to Norman L. Anderson for his suggestion on constraining the least squares fit, and to Richard A. Olsen who programmed the problem. Appreciation must also be given to Boyd C. Brinton, Dee Jay Hammon, Dr. Bruce L. Baird, and others of Thiokol Chemical Corporation for their encouragement and advice. Thanks is also extended to Thiokol Chemical Corporation for typing and printing the final manuscript.

The relationship between a student and his major professor is indeed a unique one; appreciation and thanks are given to Dr. Bartell C. Jensen for his patience and understanding during the preparation of the manuscript.

And finally, the love, understanding, and patience of my wife Alice helped me to get past many a minor crisis.

Richard Kay Brimhall

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

ABSTRACT

DESIGN OPTIMIZATION USING

MODEL ESTIMATION PROGRAMMING

by

Richard Kay Brimhall, Master of Science

Utah State University, 1967

Major Professor: Dr. Bartell Jensen
Department: Applied Statistics

Model estimation programming provides a method for obtaining extreme solutions subject to constraints. Functions which are continuous with continuous first and second derivatives in the neighborhood of the solution are approximated using quadratic polynomials (termed estimating functions) derived from computed or experimental data points. Using the estimating functions, an approximation problem is solved by a numerical adaptation of the method of Lagrange. The method is not limited by the concavity of the objective function.

Beginning with an initial array of data observations, an initial approximate solution is obtained. Using this approximate solution as a new datum point, the coefficients for the estimating function are recalculated with a constrained least squares fit which forces intersection of the functions and their estimating functions at the last three observations. The constraining of the least squares estimate provides a sequence of approximate solutions which converge to the desired extremal.

A digital computer program employing the technique is used extensively by Thiokol Chemical Corporation's Wasatch Division, especially for vehicle design optimization where flight performance and hardware constraints must be satisfied simultaneously. (46 pages)

INTRODUCTION

Model estimation programming was developed to obtain extreme solutions to functions subject to constraints, where the function being optimized and the constraint functions are defined by discrete data points. The data points may be developed by computer simulation or experimentation.

The programming method was developed originally as a multivariable flight path optimization technique for trajectory simulation problems. Its successful use in this application and its general nature suggested use in other applications. Subsequently, the method has been successfully incorporated into a solid propellant automated design and performance program developed and used at Thiokol Chemical Corporation's Wasatch Division.

A typical motor design problem on which the program has been used is that of a small, single stage, air-to-air missile where values for motor chamber pressure, nozzle expansion ratio, nozzle half-angle, and average thrust are to be selected so that missile ideal velocity will be maximized subject to the constraints that: nozzle exit diameter $\leq$ D; nozzle length $\geq$ L; and maximum motor thrust-to-weight ratio $\leq a_x$. The numerical solution to this problem is presented later in this paper.

This example involves only four independent variables and three conditions of constraint; however, analytically, there is no reason to restrict the problem size. In fact, problems in the design of multistage vehicles have been successfully solved using 20 independent variables and 15 constraints.

Even with the small example problem, the optimization process becomes complicated, and it is apparent that some systematic procedure must be used. Before this procedure is discussed, however, let us consider the scope of the general programming problem.

REVIEW OF PROGRAMMING METHODS

Usage of the term "programming" implies an optimal allocation of limited resources. Although programming nomenclature was derived from problems in economics, the methods themselves are precise algorithms embracing all the rigor of applied mathematics. Each programming method developed to date solves a class of problems which are a subset of the general programming problem.

The general programming problem is to maximize or minimize an objective function with n independent variables subject to m conditions of constraint. The basic mathematics required for solving the general programming problem are classic; however, the classical method is cumbersome, awkward, and sometimes impossible to use. For this reason optimization remains an art rather than a science.

We will first consider the classical method.

The Classical Method

First consider the simple optimization problem of maximizing

$$y = f(x_1, x_2, \ldots, x_n)$$

The necessary condition for y to have a maximum value at the solution vector X, is that the first partial derivatives of y with respect to each x_j vanish. If such solution points exist, they will describe local maximums, local minimums, or saddle points. The classical method would require the determination of all such

points, testing to determine which sets were local maximums, and then comparing all of the local maximums for the greatest solution which we expect to be the global maximum. Even when the global maximum is apparently found, the function can still be unbounded at some point not detectable by a stationary point. For a simple maximization, there may be infinitely many candidate solutions from which to choose.

For the constrained optimization where m inequality constraints exist, $m \leq n$, the above process must be expanded to 2^m cases for each of the possible combinations of active constraints. If m is greater than n, the number of possible solutions is increased to

$$\sum_{i=0}^{i=n} \binom{m}{i}$$

Although the classical method is extremely useful for many problems, it is impractical to apply the general case because of the large number of possible solutions which must be examined. These difficulties have resulted in the development of programming algorithms for obtaining optimal solutions for restricted problem types. A partial list of programming methods is discussed below to provide background and basis for comparing current methods with Model Estimation programming.

Methods Which Yield Exact Solutions

Of the current methods which yield exact algebraic solutions for the global extremal, we will examine linear programming, quadratic programming, and dynamic programming.

Linear Programming. Linear programming may be used to solve that subset of the general programming problem in which the objective function and each of the constraint functions are represented by linear equations. The solution of the linear programming problem will lie upon the simplex described by the constraint functions. The linear programming problem is solved by use of the simplex algorithms developed in 1947 by Danzig where the solution moves from point to point along the convex hull. Linear programming has the desirable property that any local maximum is a global maximum and a local maximum may be obtained in a finite number of iterations.

Many problems may be described exactly, or closely approximated, by a set of linear functions. Linear programming solutions will optimally select as many non-zero activities as there are constraints.

Quadratic Programming. Quadratic programming is a convenient method for maximizing a quadratic objective function subject to linear constraints. If the objective function is strictly concave, the solution is unique in that the local maximum always exists and it is also the global maximum. In cases where the objective functions are concave (but not strictly concave), a global maximum may be determined only if a bounded solution exists. In those cases where the function is neither concave nor convex, the solution will be a local maximum, but not necessarily a global maximum. The quadratic programming problem may be solved by starting with a basic solution and then moving from point to point on the simplex.

Dynamic Programming. Dynamic programming is an optimization method which may be applied to any integer problem. When dynamic programming is applied with increment intervals to a continuous function, it becomes a problem in approximation and the solution depends upon the size of the unit interval chosen.

For the integer problem, dynamic programming guarantees that an extremal solution will be the global extremal; this desirable property makes dynamic programming an extremely powerful tool for integer problems.

Methods Which Use Optimal Seeking Approximations

Those problems for which the exact methods do not apply must be solved by some approximate method. In general, optimal seeking methods will determine an approximate solution for a local extremal. In most applied problems, the investigator is satisfied with a local extremal solution; if this is not satisfactory, a broad area of the feasible domain must be investigated for other solutions.

Polygonal Approximations. Many nonlinear programming problems may be solved with the use of polygonal approximations. If the problem can be transformed into one in which all of the functions are separable, a solution for the approximation problem may be found by using linear programming techniques. The separable functions are evaluated at a finite number of equally spaced points for each of the independent variables. Each point sampled becomes an independent variable in the linear programming problem.

It is beyond the scope of this review to provide complete details of the method. However, because of the large number of times that the functions must be evaluated, the many independent variables and the numerous linear constraint equations which must be solved, the method appears to be practical only when applied to small problems.

Steepest Path of Ascent. The solution of a general programming problem using the steepest path of ascent is effective for finding a local extremal for many problem classes. Generally the problem is started either in the interior of the feasible domain, or with some variation of the problem where a starting routine is used to find a feasible constraint boundary.

The solution is then moved in finite increments along the boundary of the constraint condition in a direction providing maximum gain to the objective function. The solution will move along this constraint boundary until a new constraint is violated. When the second constraint is encountered, the steepest path will either follow along the intersection of the two constraint conditions, or the path will follow the second constraint and depart from the first. This general procedure is followed until a solution near the local extremal has been obtained.

For those problems where the partial derivatives must be estimated by varying the functions, the method has two disadvantages: first, the partial derivatives are often inaccurate; and second, the required number of variations may result in excessive computation time.

There are many unique forms of the steepest path of ascent solution and, while no attempt is made to catalogue them, it appears that each method has problem class limitations when applied to the general programming problem.

Need for General Non-Linear Programming Methods

Although the algorithms developed for linear programming, quadratic programming, and dynamic programming are well established, there is a need for a general computational method which may be applied to non-linear programming problems. As reported by Saaty [8], many methods have been proposed, and within the framework of limiting assumption, each will solve some subset of the general non-linear programming problem.

This thesis presents a method which partially fills the requirements as a general non-linear programming algorithm. The only limiting assumptions imposed are that the functions and their first and second partial derivatives are continuous.

MODEL ESTIMATION PROGRAMMING

Scope

The general programming problem is to maximize (or minimize) an objective function with n independent variables,

$$y_o = f_o(x_1, x_2, \ldots, x_n), \tag{1}$$

subject to m conditions of constraint, where the constraints are expressed in the form

$$\begin{aligned}
&y_1 = f_1(x_1, x_2, \ldots, x_n);\ y_{L_1} \leq y_1 \leq y_{U_1} \\
&y_2 = f_2(x_1, x_2, \ldots, x_n);\ y_{L_2} \leq y_2 \leq y_{U_2} \\
&\quad\vdots \\
&y_m = f_m(x_1, x_2, \ldots, x_n);\ y_{L_m} \leq y_m \leq y_{U_m}.
\end{aligned} \tag{2}$$

In addition to the constraints on the dependent variables, upper and lower bounds constraints may be imposed on the independent variables, i.e.,

$$
\begin{aligned}
x_{L_1} &\leq x_1 \leq x_{U_1} \\
x_{L_2} &\leq x_2 \leq x_{U_2} \\
&\vdots \\
x_{L_n} &\leq x_n \leq x_{U_n}
\end{aligned}
\tag{3}
$$

The expression of an applied problem (such as that presented in the introduction) in the framework of Eqs (1), (2), and (3) is not difficult, if data can be supplied in the form of discrete data points obtained from computation (or experimentation). Supplying data in discrete point form, however, necessitates some form of an estimating function to convert those data points into the continuous functions required in Eqs (1) and (2).

Estimating Functions. An estimating function is defined as an approximation to an unknown function. Most optimization methods [1] use some form of estimating function, the more common of which are the linear equations obtained by evaluating the first order terms of a Taylor expansion. (Non-linear effects can also be included if the estimating function includes the second order terms.)

For this program, two quadratic forms of estimating functions have been studied. On the surface, an equation of the form

$$\begin{aligned} y_o = {} & a_o + a_{10} x_1 + a_{11} x_1^2 \\ & + a_{20} x_2 + a_{21} x_2 x_1 + a_{22} x_2^2 \\ & + a_{30} x_3 + a_{31} x_3 x_1 + a_{32} x_3 x_2 + a_{33} x_1^2 \\ & + \text{etc.}, \end{aligned} \tag{4}$$

appears to have the desired estimating properties and, in fact, proves adequate for most problems. However, experience has demonstrated that a less complicated form, without the interaction terms,

$$\begin{aligned} y = {} & b_o + b_{10} x_1 + b_{11} x_1^2 \\ & + b_{20} x_2 + b_{22} x_2^2 \\ & \vdots \\ & + b_{no} x_n + b_{nn} x_n^2 \end{aligned} \tag{5}$$

has better numerical convergence properties than Eq (4), because the independent variables are separable, and less data are required for evaluation of the coefficents. Consistent with usage by Hadley [2], a function is defined as separable if it may be expressed as a sum of functions of one variable. Precisely,

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) = {} & f(x_1) + f(x_2) \\ & + \ldots, + f(x_n) \end{aligned} \tag{6}$$

This property of separability will be used in the algorithms which follow.

Subsequent discussion on estimating functions will be limited to equations in the form of Eq (5).

To obtain the data for evaluating the estimating function coefficients, an initial estimate of the value of each of independent variables must be made; these estimates are defined as the estimation solution vector, X. Evaluation of the coefficients for Eq (5) is based upon an array of $4n + 1$ observations. The first $4n$ observations are obtained by varying the independent variables, x_j, one at a time by $\pm \Delta x_j$ and $\pm 2\Delta x_j$. The last observation corresponds to the estimated solution vector, X. The value selected for Δx_j should be small enough to adequately approximate the function; on the other hand, it should be large enough to provide a good estimate of the functions curvature. If Δx_j is too small, the estimating function will appear linear and the convergence of the solution may be retarded. The coefficients are then evaluated with a constrained least squares fit, details of which are discussed in a later section. The usable information of an estimating function is restricted to some finite neighborhood of the estimated solution vector, X; good results are obtained if the use of information is restricted to the interval $X \pm \Delta X$.

With the coefficients evaluated, the estimating functions may be used in an approximation problem.

The Approximation Problem. The approximation problem, like the applied problem, is a special case of the general programming problem defined as Eqs (1), (2), and (3). Specifically, the approximation problem is defined by Eqs (1), (2), and (3) where Eqs (1) and (2) are estimating functions. Even though the applied problem is described by data points, the functions of the approximation problem have the desirable properties of being separable and continuous, and having continuous first and second derivatives.

When the approximation problem is evaluated at the estimated solution vector X, we say that X is a feasible solution to the general programming problem if the inequalities of Eqs (2) and (3) are true. An optimal feasible solution of the approximation problem is an approximate solution for the applied problem. The solution of the applied problem is obtained by repeatedly solving the approximation problem with new estimating functions.

Concavity. A function may be convex, concave, or neither. If for any two arbitrary points, a or b, in n space,

$$\lambda f(a) + (1 - \lambda)f(b) \leq f(\lambda a + (1 - \lambda)b); \quad 0 < \lambda < 1 \tag{7}$$

the function is concave; if Eq (7) is true without equality, the function is strictly concave. When the inequality is reversed, the function is convex; if Eq (7) is true without equality and the inequality is reversed, the function is strictly convex. If Eq (7) is not true with either sense of the inequality for all arbitrary points, a and b, the function is neither concave nor convex.

Because the form of Eq (5) is separable, testing the estimation function concavity is simple, as only the sign of the squared term coefficients need be examined. If all $b_{jj} < 0$, the function is strictly concave; if all $b_{jj} > 0$, the function is strictly convex.

As the estimation functions are separable, we may also determine the concavity of their separated components. If $f(x_j)$ is strictly concave, we define the independent variable x_j to be a concave variable. If $f(x_j)$ is strictly convex, we define the independent variable x_j to be a convex variable.

The approximation problem is solved by the method of Lagrange.

The Lagrange Function

For a large class of problems, the method of Lagrange is a simple and straightforward approach to finding an extreme solution to a function that is subject to constraints. This paper first considers the Lagrange function, G, and later develops a numerical method of its solution.

Let

$$G = f_o(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{i=m} \lambda_i [f_i(x_1, x_2, \ldots, x_n) - L_i] \tag{8}$$

where: G is the function actually maximized or minimized;

f_o is the objective function for which a maximum or minimum solution is desired;

f_i is the ith constraint function;

L_i is the limit at an active constraint boundary (either Y_{Li} or Y_{Ui}) defined in Eq (2) for the ith constraint; and

λ_i is the Lagrange multiplier for the ith constraint.

When Eq (8) is differentiated with respect to each x_j and each λ_i and the partial derivatives are set equal to zero, a system of n + m equations in n + m unknowns is obtained. When solved simultaneously, these equations will result in values for the independent variables which will yield either a maximum, a minimum, or a saddle point of G.

Next we will examine some properties of Eq (8). We will adopt the convention that X*, λ* specifies the coordinates and Lagrange multipliers for a local extremal of the Lagrange function G.

Equation (8) may be expressed functionally as $G = F(X, \lambda)$. If a maximal solution $G = F(X^*, \lambda^*)$ does exist, we may define $H(X) = F(X, \lambda^*)$ as a function of X which is concave and has a global maximum at X*.

The coefficients of H may be evaluated by combining the objective function with the constraint functions. Defining $\lambda_o^* = 1.0$, the coefficients are determined by:

$$H(X) = \sum_{i=0}^{i=m} \lambda_i^* b_{o_i} + x_1 \sum_{i=0}^{i=m} \lambda_i^* b_{1o_i} + x_1^2 \sum_{i=0}^{i=m} \lambda_i^* b_{11_i} + \ldots + x_n^2 \sum_{i=0}^{i=m} \lambda_i^* b_{nn_i} - \sum_{i=1}^{i=m} \lambda_i^* L_i \qquad (9)$$

Because the solution to Eq (8) may have more than one root, and because either the maximum or the minimum solution is the only one of interest, the numerical solution of Eq (8) must have the capability of selecting the proper local extremal. The logic needed to guarantee this capability is derived below for the maximization case.

The multiroot solution of Eq (8) may be avoided if the constraint functions are linearized by redefining the Lagrange function as

$$\hat{G} = f_o(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{i=m} \hat{\lambda}_i \left[g_i(x_i, x_2, \ldots, x_n) - L_i\right] \tag{10}$$

where g_i is a hyperplane tangent to f_i at the estimated solution vector X. Differentiating Eq (10) and setting the partial derivatives equal to zero results in a system of n + m linear equations in n + m unknowns which are easily solved. To guarantee that the solution obtained for Eq (10) is a maximum, we must consider the concavity of f_o. The algorithm for solving Eq (8) will differ according to three conditions of concavity for object function, f_o.

Objective Function, f_o, Is Strictly Concave. In the case where f_o is strictly concave, a local maximum solution of Eq (8) may be obtained by iteration of Eq (10). This is accomplished moving X incrementally in the direction of the solution and then solving for new hyperplanes. The rules for modifying the estimated solution

vector X for each iteration of Eq (10) will be described in the next section. As f_o is concave, then the solution to Eq (10) will always describe a global maximum, and the converged solution of Eq (10) will be a local maximum of Eq (8). This iterative solution converges at least linearly, and in the final steps, it converges quadratically.

Objective Function, f_o, Is Not Strictly Concave. If f_o is not strictly concave, another approach must be taken. As H is concave and yields a global maximum at X*, H will also yield a local maximum to the solution of Eq (8). To use the above principle, we will develop a method to numerically evaluate the coefficients of H.

As f_o is a separable function, we may separate the solution of Eq (10) into two parts; that portion which is strictly concave and that portion which is convex. The strictly concave portion of Eq (10) may be solved if we hold the convex variables constant.

This partial solution provides estimates for the Lagrange multipliers, $\hat{\lambda}$, and provides a method to estimate the combined function H. Letting $\hat{\lambda}_o = 1$, the function H(X) is estimated as:

$$\hat{H}(X) = \sum_{i=0}^{i=m} \hat{\lambda}_i b_{o_i} + x_1 \sum_{i=0}^{i=m} \hat{\lambda}_i b_{1o_i}$$

$$+ x_1^2 \sum_{i=0}^{i=m} \hat{\lambda}_i b_{11_i} \qquad (11)$$

$$+ \ldots + x_n^2 \sum_{i=0}^{i=m} \hat{\lambda}_i b_{nn_i} - \sum_{i=1}^{i=m} \hat{\lambda}_i L_i$$

As $\hat{H}$ is only an estimate of H, the solution vector which maximizes $\hat{H}$ will not necessarily satisfy the constraints. We avoid this inconvenience by maximizing $\hat{H}$ subject to the hyperplane constraints g_i. To do this, we redefine the Lagrange function as:

$$\hat{G} = \hat{H}(x_1, x_2, \ldots, x_n) + \sum_{i=0}^{i=m} \Lambda_i \, [g_i(x_1, x_2, \ldots, x_n) - L_i]. \tag{12}$$

As $\hat{H}$ is an estimate of H, we should expect $\hat{H}$ to be concave; however, if it is not, we will hold all convex variables in $\hat{H}$ constant when solving Eq (12).

We note that $\hat{\lambda}$ was functionally included into $\hat{H}$; therefore, we may conclude that Λ is a measure of the error between λ and $\hat{\lambda}$. We could iterate Eqs (11) and (12) for a precise value of $\hat{\lambda}$ with the hyperplanes g_i (X), and then move X uphill in $\hat{H}$; however, experience has shown that convergence will be obtained if we combine the solution of Eqs (10) and (12) to move X uphill.

Let the solution vector from solving Eq (10) be X_p, $\hat{\lambda}$; and let the solution vector from solving Eq (12) be X_q, Λ; and then classify the independent variables three ways:

Group 1: x_j strictly concave in $\hat{H}$.

Group 2: x_j strictly concave in f_o, but convex in $\hat{H}$.

Group 3: x_j convex in both f_o and $\hat{H}$.

Using these classifications, we must make a new estimate for X. First, we will form a vector X_R, which we will make up according to the above classifications by using: the solution of Eq (12) for those elements in Group 1; the solution of Eq (10) for Group 2; and force x_j uphill in $\hat{H}$ for Group 3.

During the early iterations, the direction of the solution is more meaningful than the values contained in X_R. Convergence will result if we move in the general gradient direction of H(X) and limit the total change of δX to some predetermined step size δ. Using the new value of X, we may repeat the above process until convergence is obtained.

Objective Function, f_o, Is Convex. Before examining this problem, we will briefly examine the two preceding concepts. In the first case, where f_o is strictly concave the solution of Eq (10) always yields unique values of X_p, $\hat{\lambda}$. When we allow X to approach X_p at a controlled rate, the problem will converge to a local extremal.

In the case where f_o is neither concave nor convex, we took advantage of the separability property of Eq (5), and separated the problem into two subproblems. We first solved the portion of Eq (10) which is strictly concave in f_o while holding the remaining independent variables fixed. The Lagrange multipliers, $\hat{\lambda}$, obtained from solving the concave portion of the problem were used to form $\hat{H}$.

Maximizing $\hat{H}$ subject to the linearized constraints, Eq (12), yielded estimates for the concave independent variables. If the approximation problem was unbounded, the independent variables were convex in both f_o and $\hat{H}$. In this case, the solution can only be improved for forcing variables convex in both functions uphill in $\hat{H}$.

These latter principles must be used in solving Eq (8) when f_o is convex.

Let us now examine the problem where f_o is convex in all variables. The numerical solution of this problem is more difficult, and its development is still in process. The following method appears valid; however, it has yet to be proven as a computational procedure.

If X is a feasible solution of $\hat{H}$, we can simply force f_o uphill on its steepest path until one or more constraints become active. Approaching the solution as in the preceding subsection, we assume that if an optimal solution X*, λ* does exist, an equivalent estimating function H(X) may be derived. The major difficulty with this assumption is that there is no direct method for obtaining an estimate for $\hat{\lambda}$ with which we can form $\hat{H}$. A less direct method must be used to estimate $\hat{\lambda}$. In the case of inequality constraints, we know that $\lambda_{i=} \leq 0$ for upper bound constraints and $\lambda_{i=} \geq 0$ for lower bound constraints. [3] The sign of λ_i for equality constraints may be determined by determining an equivalent problem where the constraint is an inequality with either an upper, or a lower bound. Borrowing from the graphical method of O'Brien, [4] an estimate may be made for the Lagrange multipliers λ as:

$$\hat{\lambda}_i = K_s \frac{1}{n} \sum_{j=1}^{j=n} \frac{\frac{\partial f_o}{\partial x_j}}{\frac{\partial f_i}{\partial x_j}} \tag{13}$$

where

$$K_s = \begin{cases} -1 \text{ for an upper bound constraint;} \\ +1 \text{ for a lower bound constraint;} \\ (+1)\ \text{sgn}\left[L_i - g_i(X)\right] \text{ for an equality constraint.} \end{cases}$$

Work to date indicates that if we eliminate the extreme terms from the summation of Eq (13), the prediction for $\hat{\lambda}$ will be improved. These extremes may be eliminated by selecting the ratio, $\frac{\partial f_o}{\partial x_j} \Big/ \frac{\partial f_i}{\partial x_j}$, whose absolute value is

greatest, and then testing to determine if it is within a reasonable limit (say four standard deviations) of the mean and standard deviations of the remaining ratios. If the largest ratio is eliminated, the next largest is tested, and so on. Although this estimate for λ_i appears to be crude, it provides a starting point.

Because of the repeated solutions, good estimates of $\hat{\lambda}$ may be available from previous solutions and the use of Eq (13) for estimation of $\hat{\lambda}$ may be omitted.

With $\hat{\lambda}_i$ known, we now form $\hat{H}$ from Eq (11). If $\hat{H}$ is strictly convex, we force the apparent solution vector X uphill in $\hat{H}$ until a new constraint is violated. If, however, $\hat{H}$ is not convex, we will solve Eq (12) while holding constant those independent variables which are convex in $\hat{H}$. This solution will provide an estimate for the independent variables that are concave in $\hat{H}$, and a vector of Lagrange multipliers Λ.

As $\hat{\lambda}$ was used to estimate $\hat{H}$, then Λ is a measure of the error in the estimation of $\hat{\lambda}$. Using this relationship, we will solve for $\hat{\lambda}$ by iterating Eqs (10) and (11), where

$$\hat{\lambda}_i^{(r+1)} = q_i \Lambda_i^{(r)} + \hat{\lambda}_i^{(r)}, \tag{14}$$

where r is an iteration counter and $0 < q_i \leq 1.0$. If this fails to converge with the chosen value of q_i, then convergence can be obtained if q_i is reduced by one-half each time that

$$\Lambda_i^{(r-1)} \Lambda_i^{(r)} < 0$$

Common to all three solutions is the assumption that a solution X^*, λ^*, exists for Eq (8). This assumption is valid providing that the Kuhn-Tucker [5] conditions for qualifying constraints are met.

If there is a nonqualifying constraint, the function $\hat{H}$ will always be convex for some x_j. During the initial iterations, this x_j will move in one direction. During later iterations it reaches a ridge for which $\frac{\partial G}{\partial L_i}$ does not exist. This will cause an oscillation between $\hat{\lambda}_i$ and x_j in the solution of Eqs (11) and (12). A numerical solution may still be obtained if δ_j is reduced by one-half each time the ratio

$$\frac{\delta x_j^{(r)}}{\delta x_j^{(r-1)}} \approx -1 \tag{15}$$

Having established an algorithm for solving the approximation problem, we will next consider the convergence of the applied problem.

The Applied Problem

As previously stated, the solution to the approximation problem is used as a predicted solution of the applied problem. A block diagram showing the overall process is presented in Figure 1. The applied problem is evaluated initially at $4n + 1$ points to provide data for evaluating the coefficients of the first set of estimating functions. These estimating functions define the approximation problem which is solved to provide a predicted solution of the applied problem. The applied problem is then evaluated at this solution vector.

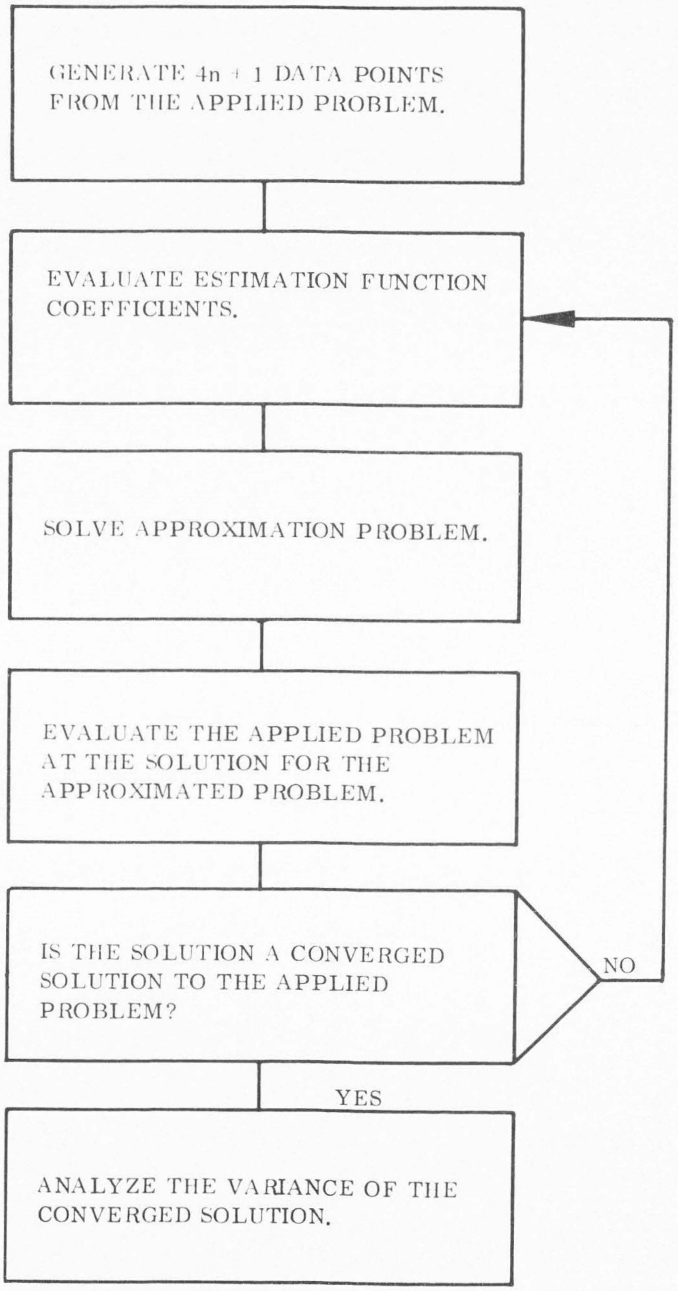

Figure 1. Solution of the Applied Problem

This evaluation provides one additional observation which, when added to the $4n + 1$ data points previously calculated, provides data to re-evaluate the estimating function coefficients.

Predictions and evaluations are repeated until a feasible solution is obtained in which the last approximation problem solution agrees with the applied problem within the prescribed tolerance. Finally, an error analysis is evaluated to establish a probable bound on the remaining gain that has not been detected by the Lagrange function G.

We will next consider the constrained least squares fit used to evaluate the coefficients for the estimating functions.

Constrained Least Squares Estimate. The derivation for the least squares estimate is well documented in literature; however, it is included to introduce the constrained least squares solution.

In the case of least squares estimates, more observations have been made than there are coefficients. The objective is to minimize the square of the differences between values of the estimating function and the observations obtained from sampling the applied problem.

The sum of the squares for these differences may be expressed as:

$$\epsilon^2 = \sum_{i=1}^{i=h} (b_{oo} + b_{10}x_{1i} + b_{11}x_{1i}^2 + \ldots + b_{nn}x_{ni}^2 - y_i)^2 \qquad (16)$$

where there are h observations. Differentiating Eq (16) with respect to each coefficient and setting the derivatives equal to zero yields a set of simultaneous equations from which the least squares coefficients can be determined.

$$\frac{\partial(\epsilon^2)}{\partial b_{oo}} = \Sigma b_{oo} + \Sigma b_{1o}x_1 + \Sigma b_{11}x_1^2 + \ldots + \Sigma b_{nn}x_n^2 - \Sigma y = 0$$

$$\frac{\partial(\epsilon^2)}{\partial b_{1o}} = \Sigma b_{oo}x + \Sigma b_{1o}x_1^2 + \Sigma b_{11}x_1^3 + \ldots + \Sigma b_{nn}x_1x_n^2 - \Sigma x_1y = 0 \qquad (16a)$$

$$\vdots$$

$$\frac{\partial(\epsilon^2)}{\partial b_{nn}} = \Sigma b_{oo}x_n + \Sigma b_{1o}x_1x_n + \Sigma b_{11}x_1^2x_n + \ldots + \Sigma b_{nn}x_n^4 - \Sigma x_n^2y = 0$$

Bringing the coefficients of Eq (16a) outside the summation sign and factoring the coefficients into a vector, this system of equations may be written in the matrix form:

$$X^TX\beta = X^TY, \qquad (17)$$

where X is the coordinate matrix, Y is the observation matrix, and β is the coefficient matrix.

The above derivation was for one dependent variable; however, as all dependent variables have the same form for the same observations, the set of solutions may be collected by adding a column to the coefficient matrix and a column to the Y matrix for each additional dependent variable.

The constrained solution to the least squares estimate is made to force the error between the estimating functions and the applied problem to zero at a selected set of observations. Identifying the coordinate of these observations as vectors P_1, P_2, ..., P_q, the least squares problem may be written subject to q constraints as:

$$E^2 = e^2(b_{oo}, b_{1o} \ldots, b_{nn}) + \sum_{l=1}^{l=q} k_l \left[P(b_{oo}, b_{1o}, \ldots, b_{nn}) - y_l \right] \tag{18}$$

Differentiating E^2, Eq (18), with respect to each coefficient and each Lagrange multiplier, and setting the resulting system equal to zero, a linear system of equations is obtained which is similar to the normal equations for the least squares fit.

Let p be a matrix with one column for each coefficient and one row for each constrained point. The coefficients for the constrained least squares fit are obtained from solving the following matrix equation for β.

$$\left[\begin{array}{c|c} X^TX & P^T \\ \hline P & 0 \end{array}\right] \left[\begin{array}{c} \beta \\ \hline k \end{array}\right] = \left[\begin{array}{c} X^TY \\ \hline P(X_l) \end{array}\right] \tag{19}$$

In applying the constrained least squares fit, we will produce a set of estimating functions which duplicates the applied problem as closely as possible in a particular region of interest. For the first set of estimating functions, the expected solution is at the center of the array; therefore, when the coefficients of the estimating functions are evaluated, we constrain the solution to pass through the initial estimated solution vector, X_0.

For the second set of estimating functions we will constrain the solution to pass through both X_0 and the first predicted solution X_1. When these two points are constrained, the estimating functions, even though separable, will be warped to include the effects of the general interactions which are measured between X_0 and X_1.

For the third and succeeding sets of estimating solutions, the solution for the coefficients is constrained to force the estimating function to pass through the last three observations. This constrained solution warps the model in n-space so that the effects of the n-space interactions projected into a plane defined by these three constrained points are satisfied.

An example of estimating the fourth order polynomial

$$y = -1.35430 + 8.59419x - 2.77672x^2 + 0.33966x^3 - 0.01398x^4 \tag{20}$$

with the quadratic model equation

$$\hat{y} = b_o + b_1x + b_2x^2 \tag{21}$$

will demonstrate the estimating properties of a constrained least squares fit. Sampling the fourth order polynomial at x = 0, 1, 2, ..., 12, least squares fit made for the estimating function coefficients yields

$$\hat{y} = 4.12046 + 1.18704x - 0.114361x^2 \tag{21a}$$

Assuming that the area of interest is at x = 6, a better estimate in the region of x = 6 would have been obtained had we required $\hat{y}$ to equal y at x = 6. Thus, imposing the requirement that $\hat{y} = 5.3340$ when x = 6, a second estimating function is evaluated.

$$\hat{y} = 5.47364 + 0.24984x - 0.03998x^2 \tag{21b}$$

For comparative purposes, we shall consider the two additional estimating functions wherein the values of y at x = 5 and at x = 6 are both satisfied, Eq (21c). In Eq (21d), the function is satisfied at x = 5, x = 6, and x = 7. Thus,

$$\hat{y} = 7.41350 - 0.19797x - 0.01923x^2 \tag{21c}$$

$$\hat{y} = 17.13290 - 3.76163x + 0.30473x^2 \tag{21d}$$

Using the quadratic model equation, four different estimation functions have been derived. The relative merits of these four functions may be observed by comparing the function and its derivatives at x = 6 as tabulated in the following page. Note the close agreement between the fourth order function and its derivatives, and the estimating function (D) and its derivatives. The effects of constraining the function are shown graphically in Figure 2.

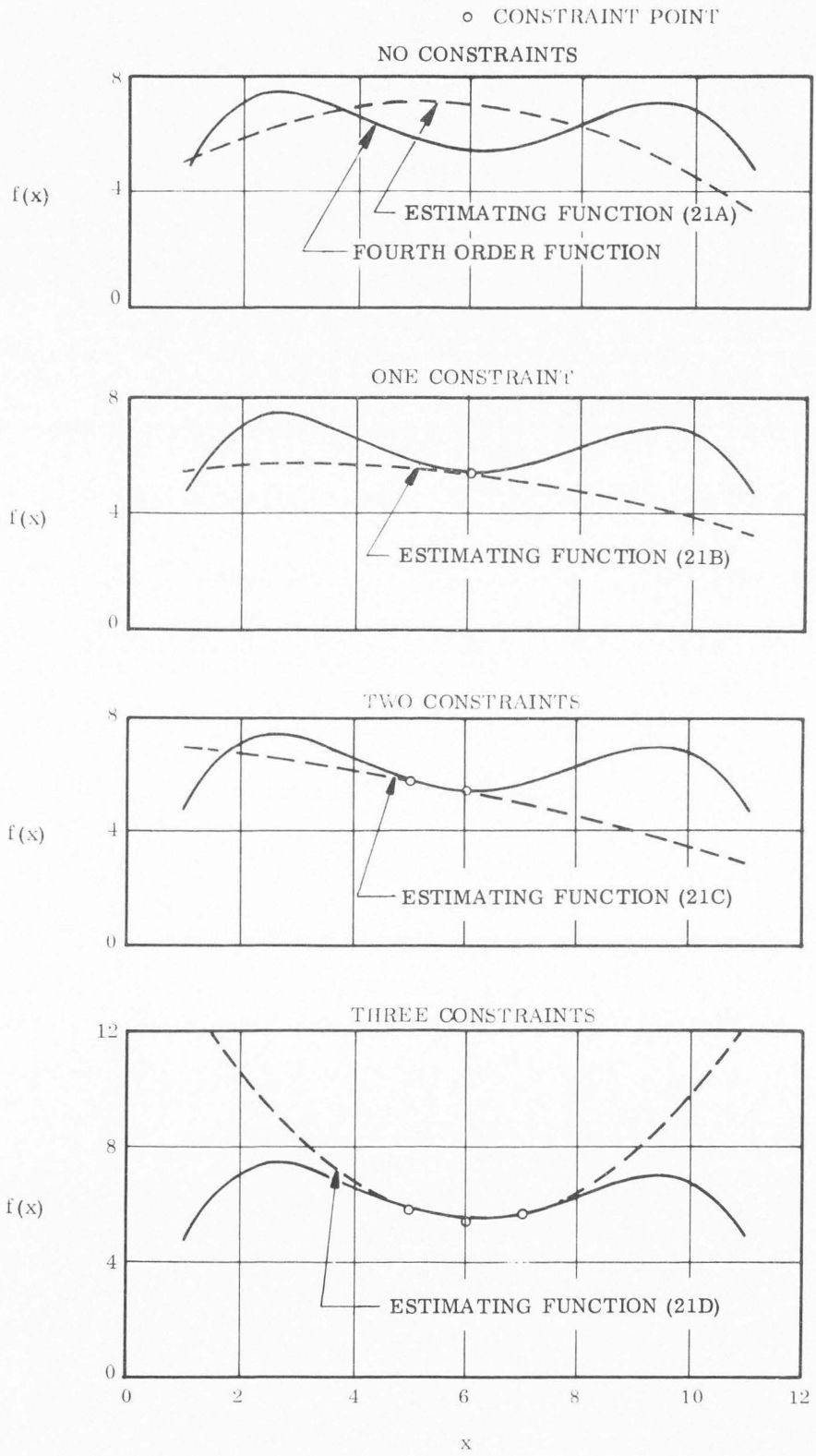

Figure 2. Constrained Least Squares Fit

The example below demonstrates that for the functions considered, the constrained least squares fit improves the estimating quality in the neighborhood of the observation constraints. Although graphical representation of similar effects for functions of several variables would be difficult, numerical sampling verifies that similar results are obtained.

EFFICIENCY OF ESTIMATING FUNCTIONS EVALUATED AT x = 6

	Fourth Order Function	Estimating Functions, Eq (21)			
		(A)	(B)	(C)	(D)
Function	5.53340	7.12574	5.53340	5.53340	5.53340
1st Derivative	-0.10989	-0.18528	-0.22922	-0.42873	-0.10487
2nd Derivative	0.63696	-0.22872	-0.07996	-0.03846	0.60946

Applied Example. Next, let us examine the convergence process of the typical design problem introduced at the beginning of the paper; results are tabulated in Table I. Note that the problem converges randomly. Also note that the poor prediction of iteration No. 7 is followed by a well-behaved prediction for 8. This type of convergence pattern is typical of the method.

Analysis of Variance of the Constrained Extremal

The applied problem is defined to have converged when the solution is feasible and f_o is within the tolerance ΔZ of the predicted value.

We may evaluate the variance of the last approximation problem to determine if the model adequately described a valid solution for the applied problem.

TABLE I

TYPICAL DESIGN PROBLEMS

	Independent Variables				Dependent Variables			
						Constraint		
Iteration No.	Chamber Pressure (psia)	Expansion Ratio	Nozzle Half-Angle (deg)	Sea Level Thrust (lbf)	Ideal Velocity (ft/sec)	Nozzle Length (in.)	Exit Diameter (in.)	Acceleration (g's)
						≥ 10	≤ 7.45	≤ 5
Base	1,500	15.000	15.00	7,000	5,182.93	12.0537	7.5327	4.4229
1	1,355	13.000	16.29	7,096	5,187.36	10.8663	7.46374	4.5095
2	1,243	11.000	16.78	7,148	5,182.34	10.0070	7.2202	4.5632
3	1,066	9.000	18.39	7,203	5,165.88	8.8199	7.1393	4.6312
4	939	7.674	16.64	7,703	5,190.30	9.4505	7.2853	4.9770
5	910	7.768	16.07	7,730	5,207.74	9.9693	7.4653	5.0001
6	930	7.933	16.06	7,737	5,208.61	10.0114	7.4565	5.0002
7	933	7.573	16.36	8,237	5,200.00	9.8025	7.5060	5.3223
8	917	7.798	16.01	7,737	5,207.97	9.9960	7.4507	5.0029
9	956	8.147	16.18	7,744	5,207.90	9.9971	7.4483	4.9999
10	932	7.936	16.07	7,737	5,208.10	10.0007	7.4500	5.0000

First we must develop an analysis of variance for a constrained least squares fit, then apply the variance of the coefficients to the Lagrange function G.

First we will simplify Eq (19) to the form

$$AB = C \tag{22}$$

where $A = \begin{bmatrix} X^T X & P^T \\ P & 0 \end{bmatrix}$

$$B = \begin{bmatrix} \beta \\ k \end{bmatrix}$$

$$C = \begin{bmatrix} X^T Y \\ P(X_l) \end{bmatrix}$$

Then we will consider the following analysis of variance table

Source	Degrees of Freedom	Sums of Squares	Mean Square	Expected Mean Square
Total	n_o	Σy_i^2		
Regression - Constraint	$n_R - n_c$	$C^T B$		
Error + Constraint	$n_o + n_c - n_R$	$\Sigma y_i^2 - C^T B$	$\dfrac{\Sigma y_i^2 - C^T B}{n_o + n_c - n_R}$	σ_l^2

where: n_o is the number of observations,

n_R is the number of coefficients in the model equation,

n_c is the number of observations constrained,

σ_ι^2 is the mean variance of the constrained least squares solution.

Next we will examine Figure 3, from which the sums of squares for the constrained least squares fit can be visualized. The calculations are as follows.

(1) The total sums of squares is the summation of the squares of all observations. This is calculated as

$$SS_Y = \Sigma y_i^2 \tag{23}$$

(2) The sum of squares due to regression is the summation of the square of the ordinates to the regression line. This is calculated as

$$SS_R = Y^T X \; [X^T X]^{-1} \; X^T Y \tag{24}$$

(3) The sum of squares due to the error in regression is the summation of the square of the distance from the observation to the regression line. This is calculated as

$$SS_{ER} = \Sigma y_i^2 - SS_R \tag{25}$$

(4) The sum of squares due to the constraint is the summation of the square of the distance between the regression line and the constrained line at each observation. This is calculated as

$$SS_c = SS_R - C^T B \tag{26}$$

(5) The sum of squares due to error plus constraint is the summation of the square of the distance from the observations to the constrained line. This may be calculated as

$$SS_{EC} = \Sigma y_i^2 - C^T B \tag{27}$$

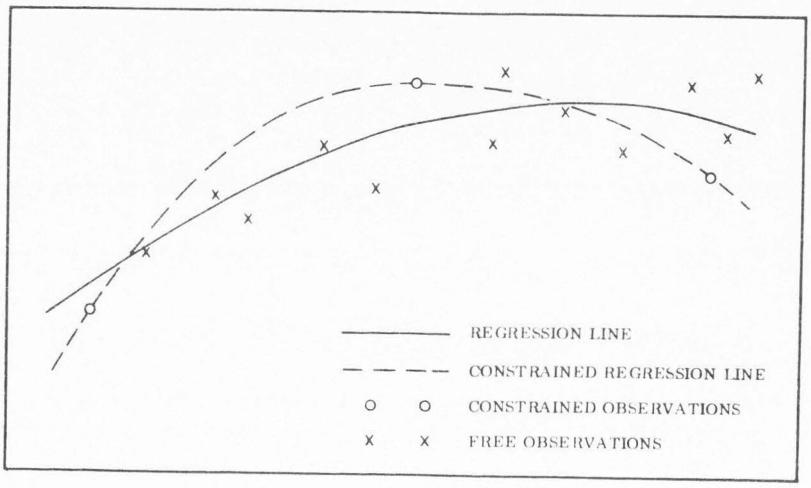

Figure 3. Typical Constrained Least Squares Fit

The mean square for the error plus constraint is

$$\sigma_l^2 = SS_{EC}/(n_o + n_c - n_R) \tag{28}$$

The mean square for the error relative to the regression line is

$$\sigma_R^2 = SS_{ER}/(n_o - n_R) \tag{29}$$

The mean square due to the constraint is

$$\sigma_c^2 = SS_c/n_c \tag{30}$$

Inverting Matrix A by partitioning, we see that the upper left hand corner of the resultant inverse yields the variance-covariance distribution for regression less constraints.

$$= \left[\begin{array}{c|c} [X^TX]^{-1} - [X^TX]^{-1} P^T[PX^TXP^T]^{-1} P[X^TX]^{-1} & - [X^TX]^{-1} P^T[PX^TXP^T]^{-1} \\ \hline - \left[[X^TX]^{-1} P^T[PX^TXP^T]^{-1}\right]^T & PX^TXP^T \end{array}\right] \quad (31)$$

Note that the upper left hand corner of A^{-1} is the difference of two positive definite matrices, and as $\left[X^TX\right]^{-1}$ is the variance-covariance distribution matrix due to regression, then $\left[X^TX\right]^{-1} P^T \left[PX^TXP^T\right]^{-1} P \left[X^TX\right]^{-1}$ is the variance-covariance distribution matrix due to constraint; their weighted sum (Matrix D) is the variance-covariance matrix due to regression plus constraint.

$$D = \sigma_R^2 [X^TX]^{-1} + \sigma_c^2 [X^TX]^{-1} P^T[PX^TXP^T]^{-1} P[X^TX]^{-1} \quad (32)$$

The above derivation for the analysis of variance for one dependent variable will next be applied to calculation of the variance of G at X*.

We will define S_c to be a vector containing the σ_c's for each dependent variable and S_R to be a vector containing the σ_R's. Then, differentiating Eq (8) with respect to L_i, we note that

$$\frac{\partial G(X^*)}{\partial L_i} = - \lambda_i \quad (33)$$

but at X*, $L_i = f_i(X^*)$; and we generalize that

$$\frac{\partial G(X^*)}{\partial f_i} = - \lambda_i \quad (34)$$

Using this relationship, we may compute the mean variance of the derived coefficients of $\hat{H}$. We define $\hat{\lambda}_o = 1.0$ and defining E to be the correlation matrix of $Y^T Y$, the mean variances of the coefficient of H due to constraint and regression are

$$\sigma^2_{Gc} = [S_c]^T [\hat{\lambda}]^T [E] [\hat{\lambda}] [S_c] \tag{35}$$

$$\sigma^2_{GR} = [S_R]^T [\hat{\lambda}]^T [E] [\hat{\lambda}] [S_R] \tag{36}$$

and the variance-covariance matrix for the coefficients of the function $\hat{H}$ is

$$\begin{aligned} D_G = {} & \sigma^2_{GR} [X^T X]^{-1} \\ & + \sigma^2_{Gc} [X^T X]^{-1} P^T [P X^T X P^T]^{-1} P [X^T X]^{-1} \end{aligned} \tag{37}$$

we now expand G in a Taylor's series about the means of x_j's

$$\begin{aligned} \Delta G = G - \bar{G} = {} & \sum_{j=1}^{j=n} \frac{\partial G}{\partial x_j} (x_j - \bar{x}_j) \\ & + \frac{1}{2!} \sum_{j=1}^{j=n} \sum_{k=1}^{k=n} \frac{\partial^2 G}{\partial x_j \, x_k} (x_j - \bar{x}_j) (x_k - \bar{x}_k) \end{aligned} \tag{38}$$

As all the first derivatives of G go to zero at X*, the first summation is identically zero at the solution point.

Thus, when substituting the partial derivatives of $\hat{H}$ (X*) into Eq (38), ΔG becomes

$$\Delta G = \sum_{j=1}^{j=n} c_{jj} (x^*_j - \bar{x}_j)^2 \tag{39}$$

Expanding Eq (39) as a Taylor series with respect to the remaining coefficients, and squaring, we obtain the variance equation:

$$\sigma^2_{G^*} = \sum_{j=1}^{j=n} \sigma^2_{c_{jj}} (x^*_j - \bar{x}_j)^4 + 2 \sum_{j=1}^{j=n-1} \sum_{k=j+1}^{k=n} \rho_{jj,kk} \sigma_{c_{jj}} \sigma_{c_{kk}} (x^*_j - x_j)^2 (x^*_k - \bar{x}_k)^2 \quad (40)$$

where the variance of the coefficients may be found in the matrix D_G.

Using an assumption of normality with $\sigma^2_{G^*}$, confidence limits on G may be established using the t distribution having $n_o + n_c - n_R$ degrees of freedom.

The confidence limits established on the probable error of the extreme solution may be used to determine if the problem should be re-evaluated using the final solution point as the center of a new array.

The sample applied problem had a probable error of 0.02 ft/sec with 95 percent confidence, based upon a set of estimating functions which converged to within a Δ Z of 1 ft/sec. This establishes an error upper bound of 1.02 ft/sec on ideal velocity for iteration No. 10 of the sample problem.

CONCLUSIONS

Model estimation programming has been successfully used to determine optimum design criteria for rocket propulsion systems. The method is original and it may be applied as a solution to the general programming problem.

The method has been programmed for a digital computer and used for the past 18 months. Because of its initial success and ease of application, it is being incorporated into an ever-broadening spectrum of computer programs at the Wasatch Division of Thiokol Chemical Corporation.

REFERENCES

1. Wilde, D. J.: Optimum Seeking Methods. New York, Prentice-Hall, 1964. pp 6-7.

2. Hadley, G.: Nonlinear and Dynamic Programming. Reading, Massachusetts, Addison-Weslay Publishing Co., Inc., 1964, p 5.

3. Hadley: op cit pp 69-72.

4. O'Brien, J. R.: "Propulsion Optimization of a Length Restricted Booster," ARS Journal, Vol 32, September 1962, p 1410.

5. Hadley: op cit p 205.

6. Grogan, G. C., and Powers, S. A.: "An Engineering Approach to Missile/ Rocket Optimization Problems," Sixth Symposium on Ballistic Missile and Aerospace Technology.

7. Zeleznik, F. J., and Sanford, G.: Simultaneous Least Squares Approximation of a Function and Its First Integral with Application to Thermodynamic Data; NASA Technical Note D-767.

8. Saaty, T. L., and Bram, J.: Nonlinear Mathematics. New York City, McGraw-Hill Book Company, 1964.

VITA

Richard Kay Brimhall

Candidate for the Degree of

Master of Science

Thesis: Design Optimization Using Model Estimation Programming

Major Field: Applied Statistics

Biographical Information:

Personal Data: Born at Provo, Utah, February 22, 1928, son of Enos and Amy Kay Brimhall.

Education: Attended elementary and high schools in Provo, Utah; undergraduate studies at Brigham Young University, Provo, Utah, and University of Utah, Salt Lake City; received the Bachelor of Science degree from University of Utah, with a major in Civil Engineering in 1954.

Professional Experience: 1958 to date, preliminary design and system analysis with Thiokol Chemical Corporation, Wasatch Division, Brigham City, Utah, and Rocket Operations Center, Ogden, Utah; currently Associate Scientist in the Post Boost Propulsion Department of the Wasatch Division; 1954 to 1958, Highway Designer, Idaho Department of Highways.