OPTIMIZATION OF SIMULATION EXPERIMENTS

R. E. Taylor

J. W. Schmidt

V. Chachra

Computing Center
and
Department of Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Abstract

The basic objectives of this paper are two-fold. The first objective is to illustrate the use of three multivariable optimization techniques as they are applied in an interactive fashion to the optimization of simulation experiments. The second and more important objective is to present the rationale behind the termination criterion for simulation experiments which is applicable to virtually any multivariable optimization procedure. The termination criterion is statistically based and includes cost factors prevalent for running the simulation as well as the potential savings from continued application of the search. The optimization techniques to be considered in the paper are

1. The sequential one factor-at-a-time technique as proposed by Friedman and Savage,
2. The pattern search method of Hooke and Jeeves, and
3. The successive quadratic approximation technique of Schmidt and Taylor.

It is shown that the termination criteria based upon economic and statistical considerations is most effective for simulation experiments.

Introduction

Digital simulation techniques for discrete systems have progressed rapidly over the past decade. The present development in digital simulation seems to be following two basic avenues. The first area of development is in special purpose computer languages for discrete systems simulation. The second area of development is in the statistical methodology related to the design of simulation experiments and analysis of results. In this paper we shall investigate a third possible avenue of development for discrete systems simulation, namely the use of multivariate optimization techniques for simulation experiments.

Often in performing the simulation analysis of a given system the objective is simply to obtain a measure of system effectiveness for some prescribed values of the decision variables. However, more frequently, the objective is to obtain the specific values of decision variables which will optimize the system effectiveness function. When this is in fact the objective, the problem can be addressed by a body of "multivariable optimization techniques". Indeed these techniques are not new. They have been in existence for many years and have been applied widely to problems of a deterministic nature [4], [18], [19].

The basic objectives of this paper are two fold: The first objective is to illustrate the application of three of these techniques to the optimization of simulation models. The second and more important objective is to present the rationale behind a termination criterion for simulation experiments which is applicable to virtually any multivariable optimization procedure. The termination criterion is statistically based and includes the cost factors prevalent in conducting the simulation analysis as well as potential saving from continued application of the search. The termination criterion will be shown to be effective for these stochastic problems.

The optimization techniques are employed in an interactive manner with the simulation model. The operation is such that particular values of the decision variables are specified by the optimization program to the simulation program. A measure of system effectiveness is determined through simulation which is returned to the optimization program. Based upon that value of the effectiveness function new values of the decision variables are determined and the process is repeated. At some point in this process a termination criterion will be met and the procedure will terminate. This facet will be discussed later.

The optimization techniques to be considered in this paper are:

1. The sequential one-factor-at-a-time technique as proposed by Friedman and Savage (7).
2. The pattern search method of Hooke and Jeeves (10).
3. Successive quadratic approximations of

Schmidt and Taylor (16).

Each of these techniques shall be discussed in detail in a later section of this paper. The search routines will be discussed in terms of a minimization problem. The model to which the techniques were applied is a stochastic inventory system which shall also be discussed in some detail in later sections.

Multivariate Search Procedures

All multivariable search procedures have essentially two basic objectives: (1) to obtain an improved value of the effectiveness function; (2) to provide information useful for locating future experiments where desirable values are likely to be found. The logical organization of a search procedure is such as to accomplish the aforementioned objectives through a three phase operation. The first phase sets the stage by making the initial observation(s) of the effectiveness function. From this initial phase can be determined the general direction of the search. The second phase of the search is characterized by rapid movement toward the optimal. During this phase the effectiveness function is examined through selective manipulation of the decision variables. The final phase of the search is perhaps the most important. This is known as the termination phase and the termination criterion plays a critical part in the overall procedure.

In general, the first phase of any search procedure is designed to "get things underway". For most practical examples this phase consists of the experimenter "arbitrarily" establishing the stating point. Many experimental statistical designs have been created to aid in this process. However, for the procedures discussed herein the stating point is chosen arbitrarily and to some extent the results to be derived from any of these procedures are dependent upon a "lucky" choice for the beginning point. If the experimenter fortunately chooses initial values of decision variables which are close to the optimum levels, money will be saved in achieving a relative optimum. If, on the other hand, luck is not with the experimenter and he selects initial levels which are far from the optimal values, then it likely will cost him more to achieve a relative optimum.

Once the initial experiment has been accomplished the information gained from that may be used to assist future experiments. The procedure of the particular search technique is then applied in an algorithmic fashion. The search procedures discussed herein all operate on the function in a systematic fashion, varying the decision variables in some prescribed manner. This phase of the overall operation is likely to consume the bulk of the activity of the search. As a result of this phase, the effectiveness function should be significantly improved. Later sections of this paper will describe in detail this phase of the operation.

The final phase of the search procedure is called the termination phase and specifies the conditions under which the search procedure will terminate. This phase is of great interest and is considered at length in this paper.

There are several characteristics of search procedures which will be mentioned here for purposes of description. They will not be explored in depth but should be taken into account when considering what technique to apply. These characteristics are listed below.

1. Total number of simulation replications required to obtain an optimum.
2. Ability to move on the response surface in several directions.
3. Ability to vary step length.
4. Ability to deal successfully with a large number of decision variables.
5. Termination criterion.

The Sequential One-Factor-at-a-Time Method

Sectioning or the one-at-a-time method proposed by Friedman and Savage (7) is one of the simplest optimum seeking techniques available and may be applied to functions of any number of decision variables. Suppose $y(x_1,x_2,\ldots,x_n)$ is a cost function to be minimized, where x_i, $i=1,2,\ldots,n$, are the decision variables. To apply the method of sectioning, the analyst fixes the values of the last n-1 variables and varies the first until a minimum, or at least near minimum, is found. Let x_1^o be the minimizing value of x_1 with associated cost $y(x_1^o,x_2,\ldots,x_n)$. The value of x_1 is now fixed at x_1^o, and x_2 is varied until its optimal value is determined, x_2^o. This procedure is repeated for all n decision variables. The entire process is repeated until values of the decision variables are found such that further change in any one of the variables will result in an increase in the value of the objective function.

The sectioning search may be effected in several ways. However, the initial step is always the same. All but one of the decision variables are given fixed values. Let these variables be $x_2,\ldots,x_n$. The initial value of the remaining variable, x_1, must now be set and the measure of effectiveness, $y(x_1,\ldots,x_n)$, evaluated. The initial search over x_1 usually involves changing x_1 in rather large increments. Let δ_{ij} be the jth increment chosen for the ith decision variable and let m be the number of increments for each variable, $j=1,2,\ldots,m$. Choosing δ_{11} relatively large allows the search to rapidly locate the general region of the optimum value of x_1, x_1^*, given the fixed values of the remaining variables. Let us arbitrarily assume that in searching over any decision variable, we first increase the value of the variable and if this does not prove fruitful we then decrease its value. Therefore, the first step in the search moves us to the point $(x_1+\delta_{11},x_2,\ldots,x_n)$. If $y(x_1+\delta_{11},x_2,\ldots,x_n) <$ $y(x_1,\ldots,x_n)$, we must continue to increase x_1, next examining the measure of effectiveness at $(x_1+2\delta_{11},x_2,\ldots,x_n)$. This procedure is continued until a point $(x_1+M\delta_{11},x_2,\ldots,x_n)$ is found such that $y(x_1+M\delta_{11},x_2,\ldots,x_n) > y(x_1+(M-1)\delta_{11},x_2,\ldots$

$.,x_n)$. If the objective function is convex, x_1^* lies between $x_1+(M-2)\delta_{11}$ and $x_1+M\delta_{11}$.

If $y(x_1+\delta_{11},x_2,\ldots,x_n) > y(x_1,\ldots,x_n)$, a further increase in x_1 would not be warranted if the objective function is convex. Therefore, the next point evaluated would be $(x_1-\delta_{11},x_2,\ldots,x_n)$. If $y(x_1-\delta_{11},x_2,\ldots,x_n) > y(x_1,\ldots,x_n)$, then x_1^* is such that $x_1-\delta_{11} < x_1^* < x_1+\delta_{11}$. If $y(x_1-\delta_{11},x_2,\ldots,x_n) < y(x_1,\ldots,x_n)$, x_1 is further reduced until a point $(x_1-M\delta_{11},x_2,\ldots,x_n)$ is found such that $y(x_1-M\delta_{11},x_2,\ldots,x_n) > y(x_1-(M-1)\delta_{11},x_2,\ldots,x_n)$, in which case x_1^* is such that $x_1-M\delta_{11} < x_1^* < x_1-(M-2)\delta_{11}$.

Ignoring boundary constraints, the result of the initial search over x_1 is an interval of width $2\delta_{11}$, the center of which, x_1^o, is the best estimate of x_1^* thus far. At this point the analyst may choose to continue the search over x_1, keeping the remaining decision variables fixed at their previously established values. To accomplish this, the analyst chooses a new increment for x_1, δ_{12}, which is less than the initial increment. The starting point for this search is the center point of the interval about x_1^* which was obtained in the initial search, x_1^o. The procedure described for the initial search of x_1 is then repeated until a new value of x_1^o is derived. The entire process is repeated over and over again until x_1^o is bracketed by a sufficiently small interval. When the search over x_1 terminates, the search over x_2 begins, fixing x_1 at the last value of x_1^o derived and holding $x_3,\ldots,x_n$ at their initial values. The procedure for the search over x_2 is identical to that for x_1. After all n variables have been searched over once, the search returns to x_1 and starts the whole process over again. The search terminates when for every i

$$y(x_1^o,x_2^o,\ldots,x_i^o\pm\delta_{im},\ldots,x_n^o) > y(x_1^o,x_2^o,\ldots,x_i^o,\ldots,x_n^o) \qquad (1)$$

When the initial search over x_1 terminates, the analyst may choose to search over the remaining variables before refining the search over x_1. If this is the case, x_1 is fixed at the initial value of x_1^o, and the search over x_2 is conducted in increments δ_{21}. This process is repeated for all n variables. Here the search returns to x_1, again searching in increments δ_{11}. The search increment for any variable is not reduced until a point $(x_1^o,x_2^o,\ldots,x_n^o)$ is found such that for every i

$$y(x_1^o,x_2^o,\ldots,x_i^o\pm\delta_{i1},\ldots,x_n^o) > y(x_1^o,x_2^o,\ldots,x_i^o,\ldots,x_n^o) \qquad (2)$$

When this condition is achieved, the increments on all variables are reduced to δ_{i2}, $i=1,2,\ldots,n$, and the search over all decision variables is repeated until the termination criterion given is satisfied.

The Pattern Search Method

The philosophy underlying the pattern search technique is based upon the hopeful conjecture that any adjustments of the decision variables which have improved the effectiveness function during early experiments will be worth trying again. The technique begins from the starting point by moving in small steps. The steps grow with repeated success. Failure at any step length indicates that shorter steps are in order. If a change in direction is required, the technique will begin over again with a new pattern. The method is a ridge following technique and a pattern of moves can succeed only if it lies along a straight ridge. In the area of the optimal the steps become very small to avoid overlooking any promising direction. As before $y(\underline{x})$ is the value of the objective function evaluated at the point $\underline{x}$, previously defined as $(x_1, x_2, \ldots, x_n)$. The technique seeks an optimal in a series of cycles. One cycle differs from another basically in the step length employed for the decision variables.

In visualizing what is meant by a "pattern", it is helpful to think of an arrow, its base at one end and its head at the other. A cycle begins at a base point $\underline{b}_1$. At the beginning of a given cycle a step width δ_i is determined for each decision variable. Let $\underline{\delta}_i$ be the vector whose ith component is δ_i, the rest being zero. After evaluating $y(\underline{b}_1)$, $y(\underline{b}_1+\underline{\delta}_1)$ is evaluated. If the new point, $\underline{b}_1 + \underline{\delta}_1$, is better than the base point, this point is called the temporary head $\underline{t}_{11}$, where the first subscript indicates the pattern number under construction and the second subscript indicates the variable number most recently perturbed. If $\underline{b}_1 + \underline{\delta}_1$ is not as good as $\underline{b}_1$, $y(\underline{b}_1-\underline{\delta}_1)$ is evaluated. If this point is better than the base point it is denoted as the temporary head; otherwise $\underline{b}_1$ is designated as the temporary head. This process is repeated for each of the decision variables following the rule that a jth temporary head $(\underline{t}_{ij})$ is obtained from the preceding one, $\underline{t}_{i,j-1}$, as follows:

$$\underline{t}_{ij} = \begin{cases} \underline{t}_{i,j-1} + \underline{\delta}_j & \text{if } y(\underline{t}_{i,j-1} + \underline{\delta}_j) < y(\underline{t}_{i,j-1}) \\ \underline{t}_{i,j-1} - \underline{\delta}_j & \text{if } y(\underline{t}_{i,j-1} - \underline{\delta}_j) < y(\underline{t}_{i,j-1}) \\ \underline{t}_{i,j-1} & \text{if } y(\underline{t}_{i,j-1}) < \\ & \min[y(\underline{t}_{i,j-1} + \underline{\delta}_j),\ y(\underline{t}_{i,j-1} - \underline{\delta}_j)] \end{cases} \quad (3)$$

Equation 3 covers all variables $j (1 \le j \le n)$ if the convention is adopted that

$$\underline{t}_{10} \equiv \underline{b}_1 \ .$$

When all decision variables have been perturbed, the last temporary head point, $\underline{t}_{1n}$ is designated as the second base point $\underline{b}_2$, i.e.,

$$\underline{t}_{1n} \equiv \underline{b}_2 \ .$$

The original base point in the cycle $\underline{b}_1$ and the newly determined base point, $\underline{b}_2$, establish the first pattern which is the arrow joining $\underline{b}_1$ to $\underline{b}_2$.

At this point in the procedure an acceleration step is initiated to establish the next temporary heading. Under the philosophy that if a similar exploration is conducted from $\underline{b}_2$

the results are likely to be the same, the local perturbations are ignored and the search is extended to a new temporary head $\underline{t}_{20}$ for the second pattern based at $\underline{b}_2$. The initial temporary head is given by

$$\begin{aligned} \underline{t}_{20} &\equiv \underline{b}_1 + 2(\underline{b}_2 - \underline{b}_1) \\ &= \underline{b}_2 + \underline{b}_2 - \underline{b}_1 \\ &= 2\underline{b}_2 - \underline{b}_1 \quad . \end{aligned} \tag{4}$$

In other words, the arrow (representing the direction of the pattern) is extended from $\underline{b}_1$ to $\underline{b}_2$, immediately doubling its length. In line with terminology previously used, the double subscript on the temporary head $\underline{t}_{20}$ indicates the initiation of the second pattern with no local explorations yet performed. A local exploration is now carried out about $\underline{t}_{20}$ to correct the tenative second pattern, if necessary. The logical equations governing establishment of new temporary head $\underline{t}_{21}, \underline{t}_{22}, \ldots, \underline{t}_{2n}$ will be similar to Equation 3, the only difference being that the first subscript will be 2 instead of 1. If, after all variables have been perturbed, the last temporary head $\underline{t}_{2n}$ is better than $\underline{b}_2$ it is designated as the third base point $\underline{b}_3$.

As before, a new temporary head $\underline{t}_{30}$ is established by extrapolating from $\underline{b}_2$ through $\underline{b}_3$, i.e.,

$$\underline{t}_{30} = 2\underline{b}_3 - \underline{b}_2 \ . \tag{5}$$

Repeated success in a given direction causes the pattern to grow and as long as this procedure improves the objective function it is continued.

If, however, the attempt to establish a new temporary head is unsuccessful the pattern is destroyed and perturbation of the independent variables is begun at the current δ values about the last successful base. If these perturbations are successful the pattern will again begin to grow and accelerate. If, on the other hand, perturbations about the last successful base are unsuccessful then the cycle is complete. A new cycle is begun by reducing the step size (elements of $\underline{\delta}_1$), and initiating perturbations about the base with the new step size. The termination criterion for the pattern search is normally couched in terms of step size used for perturbations. When the minimum step size is reached the technique is terminated.

Figure 1 illustrates the operation of the pattern search technique for a two variable case. In this example the search begins by proceeding in the positive direction for both decision variables. As the search proceeds successfully during cycle one the pattern continues to grow. At temporary heading $\underline{t}_{40}$ the search falters and is unable to make further improvements at the existing step width. At this point the step width is decreased and the second cycle begins. Note that during the second cycle (indicated by primes) the direction of the search completely changes. At point $\underline{t}'_{50}$ the search is unable to find further improvements. In that the step

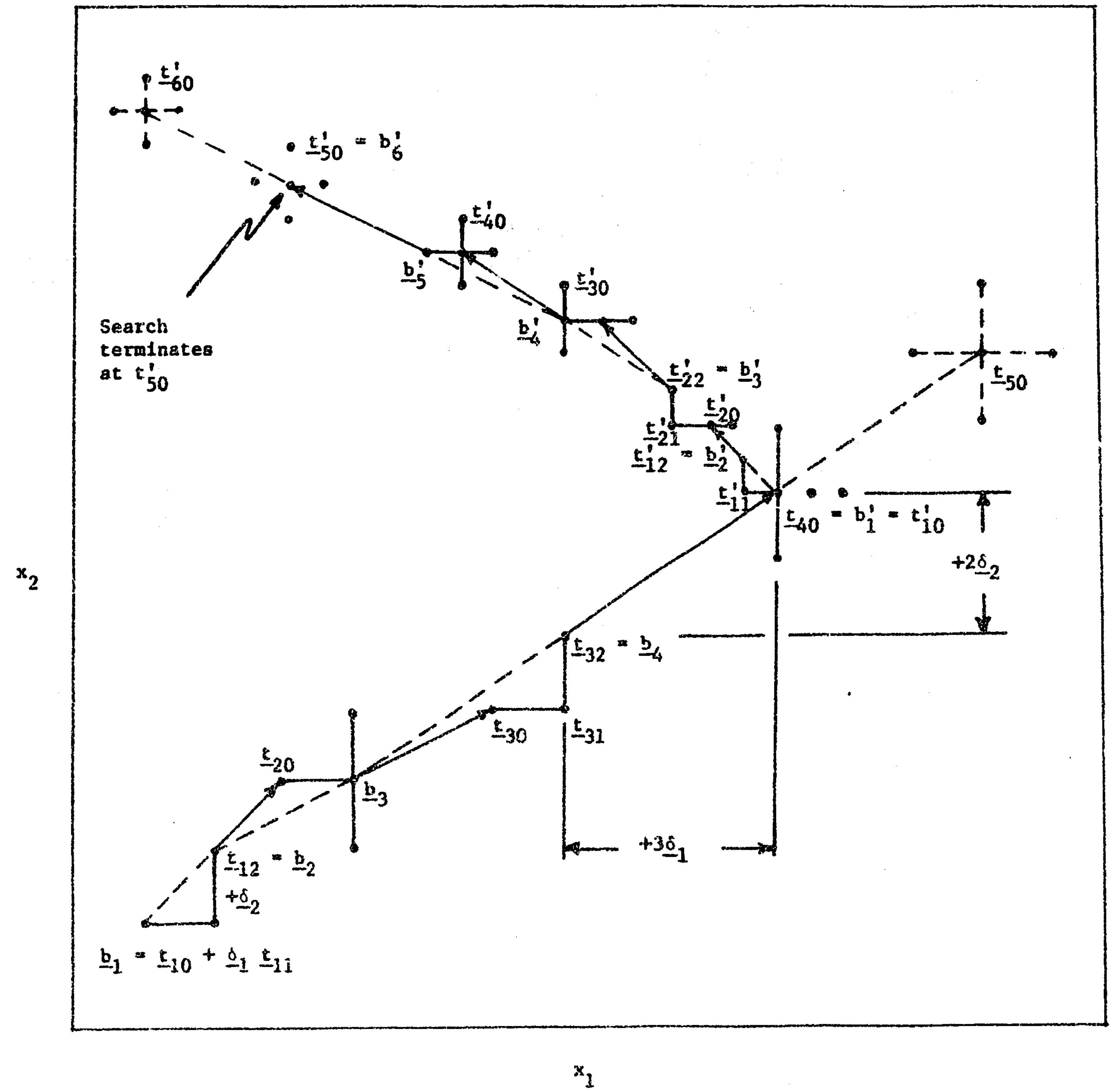

FIGURE 1: Operation of Pattern Search

width has reached its minimum value the procedure terminates. An excellent discussion of this technique can be found in Wilde and Beightler (19).

Successive Quadratic Approximation Method

The search by successive quadratic approximation is based upon the assumption that the objective function can be roughly approximated by a quadratic equation. The reliability of the approximation increases as the region of the optimal to which the approximation applies is reduced. Let $y(x_1,x_2,\ldots,x_n)$ be the objective function and $(x_1,x_2,\ldots,x_n)$ the decision variables. The approximating function, $\hat{y}(x_1,x_2,\ldots,x_n)$, can be expressed by

$$\hat{y}(x_1,x_2,\ldots,x_n) = b_o + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} b_{i+n} x_i^2 . \tag{6}$$

Let m be the number of coefficients in the approximating expression. Therefore

$$m = 2n + 1 .$$

The approximating function given in Equation 6 may be augmented by the addition of terms such as $x_i x_j$ to improve the approximation. However, the authors have found Equation 6 satisfactory in most cases. The constants, b_i, which specify $\hat{y}(x_1,x_2,\ldots,x_n)$ are developed through the method of least squares. Therefore, $y(x_1,x_2,\ldots,x_n)$ must be evaluated at $k \geq m$ points $(x_{1j},x_{2j},\ldots,x_{nj})$, $j=1,2,\ldots,k$. In the context of this paper $y(x_1,x_2,\ldots,x_n)$ is evaluated through simulation although an appropriate mathematical model could be used for this purpose if it were available.

Having fit the approximating function to k points in the solution space, $\hat{y}(x_1,x_2,\ldots,x_n)$ is optimized through the classical methods of calculus. That is

$$\frac{\partial \hat{y}}{\partial x_i} = b_i + 2b_{i+n} x_i = 0 \tag{7}$$

and

$$x_i^* = -b_i/2b_{i+n} \tag{8}$$

The point $(x_1^*,x_2^*,\ldots,x_n^*)$ represents the initial estimate of the optimum for $y(x_1,x_2,\ldots,x_n)$ and is the next point at which $y(x_1,x_2,\ldots,x_n)$ is evaluated. Once $y(x_1,x_2,\ldots,x_n)$ has been evaluated at $(x_1^*,x_2^*,\ldots,x_n^*)$, the point $(x_{1j},x_{2j},\ldots,x_{nj})$ for which $y(x_1,x_2,\ldots,x_n)$ is least optimal, $j=1,2,\ldots,k$, is dropped from further consideration. Let the least optimal point be denoted by $(x_1',x_2',\ldots,x_n')$. Therefore the number of points in the analysis is still k, but $(x_1',x_2',\ldots,x_n')$ is replaced by $(x_1^*,x_2^*,\ldots,x_n^*)$. Again applying the method of least squares, $\hat{y}(x_1,x_2,\ldots,x_n)$ is fit to the new set of k points and the entire procedure is repeated.

As the search progresses the region of investigation of the solution space will generally contract about the optimal point, although this general contraction may be accompanied by periodic expansions. This variation is

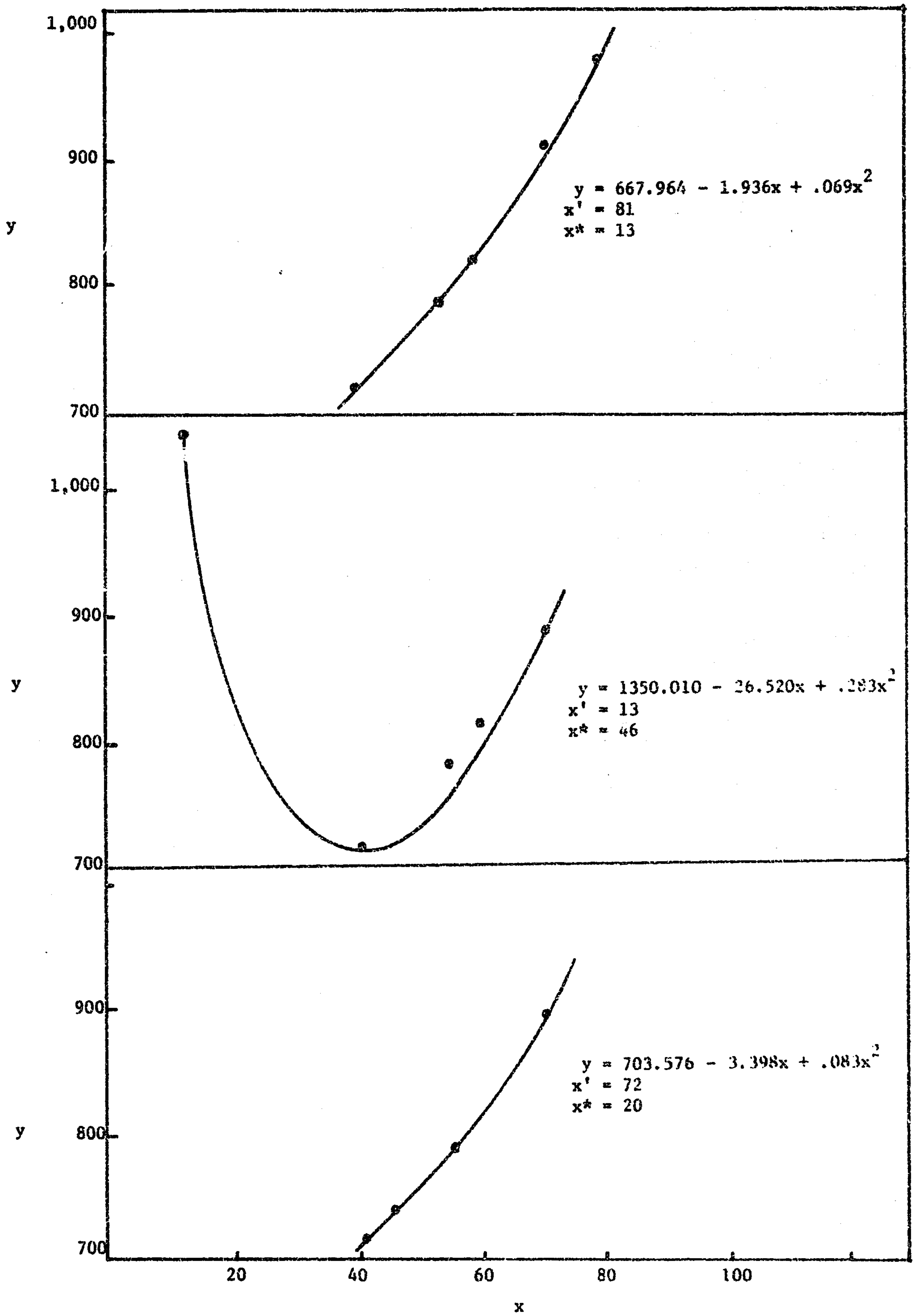

FIGURE 2: Application of Successive Quadratic Approximation

illustrated in Figure 2 where y is a function of one variable, x, and the objective of the search is minimization of y. Three successive iterations of the search are shown in which the region of the search varies from $41 \leq x \leq 81$ to $13 \leq x \leq 72$ to $41 \leq x \leq 72$.

The analyst may adopt a termination criterion of his own choice in using the search by successive quadratic approximation. For example, termination may be effected by specifying a fixed number of iterations. Another alternative is to terminate the search whenever the region of investigation is reduced to a sufficiently small neighborhood. However, these criteria are most effective when $y(x_1,x_2,\ldots,x_n)$ can be expressed in mathematical form.

Termination of the Search

When the system model to be optimized is a mathematical model, the search procedure usually terminates either after a fixed number of iterations or when the step size in the exploratory segment of the search has been reduced to a predefined minimum, although other termination criteria may also be used. The purpose of the application of a search procedure is to identify a point at or near the optimum for the mathematical model. To insure this, the termination criteria is usually defined in such a manner that the search continues well beyond the identification of an adequate approximation to the true optimum. Thus, there is normally wasted computer time resulting from excessive iterations. However, the cost of these extra iterations may not be expensive, since many mathematical models can be evaluated rapidly on a digital computer.

When the system to be optimized is modeled through simulation, the cost of evaluation of the model can be expensive. Thus excessive iteration of the search procedure may result in a situation where more is spent identifying the optimum or near optimum than was saved by finding the optimum. The cost of simulation arises from two sources. Let $(x_1,x_2,\ldots,x_n)$ be a vector of decision variables representing a point at which the system is to be evaluated in the course of the search and let $y(x_1,x_2,\ldots,x_n)$ be the corresponding expected cost of operation of the system. Since $y(x_1,x_2,\ldots,x_n)$ is to be evaluated through simulation, the value of $y(x_1,x_2,\ldots,x_n)$ can only be estimated. Let $\hat{y}$ be the estimate of $y(x_1,x_2,\ldots,x_n)$ obtained by one replication of the simulation. Then

$$\hat{y} = y(x_1,x_2,\ldots,x_n) + \varepsilon \tag{9}$$

where ε is a random variable representing the error due to simulation with mean zero and variance σ^2. One replicate of the simulation at $(x_1,x_2,\ldots,x_n)$ alone may be expensive. In addition, depending upon the value of σ^2, one replicate may provide a poor estimate of $y(x_1,x_2,\ldots,x_n)$. To improve the estimate of $y(x_1,x_2,\ldots,x_n)$, we may replicate the simulation N times at $(x_1,x_2,\ldots,x_n)$. Let $\hat{y}_1,\hat{y}_2,\ldots,\hat{y}_N$ be the simulated values of the cost of operation of the system for replications $1,2,\ldots,N$. Then

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \qquad (10)$$

has mean $y(x_1,x_2,\ldots,x_n)$ and variance $\frac{\sigma^2}{N}$.

Let C_r be the cost of one replicate of the simulation, k the number of iterations of the search until termination, and r_i the number of replicates at the ith point evaluated in the search. Then the total cost of executing the search is:

$$\text{Cost of Search} = C_r \sum_{i=1}^{k} r_i \,. \qquad (11)$$

Let $\bar{y}_1$ be the estimated cost of the system at the first point evaluated in the search and $\bar{y}_o$ the minimum estimated cost found in the course of the search. Then the savings in system cost achieved by the search is $\bar{y}_1 - \bar{y}_o$, and an attempt should be made to design the search such that

$$\bar{y}_1 - \bar{y}_o > C_r \sum_{i=1}^{k} r_i \qquad (12)$$

Of course, there is no guarantee, prior to execution of the search, that the expression in Equation 12 will be satisfied. For example, if the starting point for the search is close to the optimum, large savings as a result of the search may not be possible. Therefore the termination criterion should be designed such that a condition of this type will be detected quickly and the search terminated.

Let $\bar{y}_1,\bar{y}_2,\ldots,\bar{y}_k$ be the costs of operation of the system as estimated through $r_1,r_2,\ldots,r_k$ replications of the simulation for the first k points evaluated in the course of the simulation. Let

$$z_1 = \bar{y}_1$$

and define $z_2,z_3,\ldots,z_p$, $p \leq k$ as a subset of $\bar{y}_1,\ldots,\bar{y}_k$ such that

$$z_j = \bar{y}_\ell$$

where ℓ is the smallest i for which

$$\bar{y}_i < z_{j-1} \,.$$

The definition for the z's given above holds for all optimization procedures which have a single point in solution at any time. In optimization procedures where more than one point is in solution at a given time, as in the sequential quadratic approximation method, a slightly different definition is necessary. If m is the number of points in solution at a given time, then define S_i as the set of m points in solution at the ith iteration. The set S_1 will consist of the first m points that are evaluated. Also define

$$z_1 = \bar{y}_1$$

then,

$$z_j = \bar{y}_\ell$$

where ℓ is the smallest index i for which

$$\bar{y}_i < \max\,[\bar{y} \mid \bar{y} \in S_{j-1}]$$

and S_j is the set of m points obtained from S_{j-1} by replacing the worst point in S_{j-1} by z_j.

To illustrate consider the sequence of values of $\overline{y}_i$ shown below for the first 15 iterations of a search. As shown, z_j is the value of the objective function, as estimated through simulation, at the point in the search at which the jth improvement in the estimated value of the objective function was observed.

Iteration i	Simulated Cost $\overline{y}_i$	j	z_j
1	1000	1	1000
2	1268		
3	1342		
4	987	2	987
5	1014		
6	824	3	824
7	915		
8	927		
9	887		
10	831		
11	807	4	807
12	801	5	801
13	835		
14	763	6	763
15	771		

At the jth improvement the loss resulting from the search up to that point is calculated and given by

$$L_j = z_j - z_1 + R \tag{13}$$

where R is the cost of the search up to and including the jth improvement. R is given by Equation 11 where k is the iteration at which the jth improvement occurred. A simple straight line of the form

$$L = b_o + b_1 x, \tag{14}$$

is then fit to the last $M \leq j$ improvements by the method of least squares. If the slope, b_1, of the straight line given in Equation 14 is significantly less than zero then there is reason to believe that continuation of the search may yield further savings. However, if the slope is greater than or equal to zero, then it is likely that the search should be discontinued in the sense that the maximum savings has already been achieved.

Since L_j is a random variable, the slope, b_1, is a random variable. To determine whether or not the slope is significantly less than zero, a t-test is conducted each time an improvement, z_j, is detected. If,

$$\frac{b_1}{S_{b_1}} < t_\alpha \left[\sum_{t=j-M+1}^{j} r_t \right] \tag{15}$$

where

r_t = the number of replications at the t^{th} improvement

then, the hypothesis

H_o: $b \geq 0$

is rejected, and the search continues. Otherwise the search is terminated.

To implement the termination criteria it is necessary to specify α and determine r_t and M. After each iteration of the search the number of replications at the next point is evaluated and the number of points, M, to which the straight line will be fit, if the t-test is conducted, are determined. However, these calculations require specification of a breakeven rate of return, ρ_b, on investment, a desirable rate of return, ρ_d, and the β error corresponding to the desirable rate of return. The logic incorporated into the search routine then calculates r_t and M such that the equations

$$P(\text{search continues}|b_1 = 0) = \alpha \tag{16}$$

$$P(\text{search continues}|b_1 = \rho_b) = .5 \tag{17}$$

$$P(\text{search continues}|b_1 = \rho_d) = 1-\beta \tag{18}$$

are satisfied as nearly as possible. In any case the minimum values of M and r_t are never less than 2. However, the user may specify a minimum greater than 2 for either M or r_t or both.

The Model to Be Optimized

The function whose expected value is to be minimized is

$$y = 5. \sum_{i=1}^{5} \left[\frac{A_i B_i}{x_i} + \frac{C_i x_i}{2} \left(1 - \frac{A_i}{D_i}\right)\right] + \varepsilon \tag{19}$$

where

i	A_i	B_i	C_i	D_i
1	100	10	1	1000
2	200	20	4	1000
3	300	40	3	1000
4	400	100	5	1000
5	500	50	8	2000

and ε is distributed uniformly on the interval [-25, 25]. The minimum value of E(y) is approximately 7300 at (47, 50, 107, 163, 91). Each procedure began at (500, 500, 500, 500, 500). The value of y at this point is approximately 19820.70. The response surface for this model is a well behaved surface with a rather large "flat" region about the optimal. A sensitivity analysis performed about the optimal shows little response to changes in decision variables for a rather large area. The response surface is flat enough that random error can easily mask out true differences in response thereby possibly causing a premature termination of a search procedure.

Method of Operation

The model discussed in the previous paragraph was coded as a FORTRAN IV subprogram. Each optimization technique was coded as a FORTRAN IV main line program. The macro logic of the overall solution procedure is shown in Figure 3. The general method of operation is that the optimization program specifies values of the decision variables which are passed to the simulation model. The number of replications required at the next point is then determined. This determination is based upon desired confidence levels and the sample variance as discussed previously. After the simulation is complete the subprogram returns a particular value of the effectiveness function. At this time a statistical test is performed to determine if a significant improvement in the objective function has been achieved. Based upon a synthesis of this information, the search program then calculates a new set of decision variables which are passed to the simulation model, and the entire process is repeated. This procedure

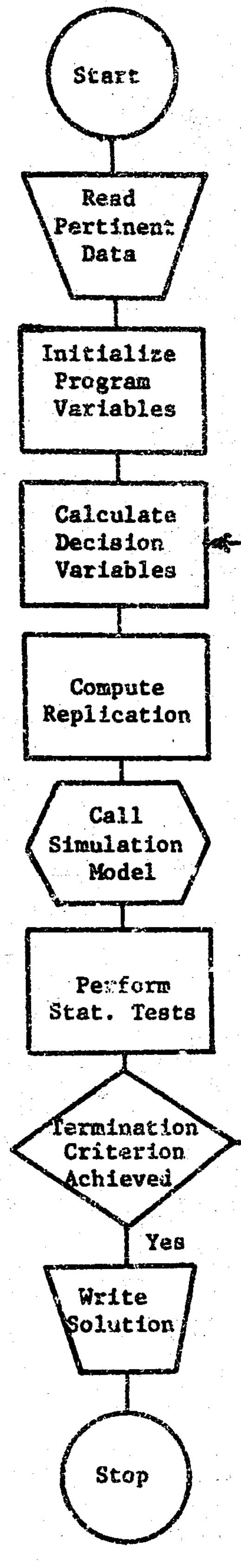

FIGURE 3: Macrologic of Optimization Procedure

continues until the termination criterion is achieved at which time the process terminates.

Simulation Results

The simulation results obtained for the three optimization techniques are summarized in Tables 1-3. Values of $\alpha = 0.10$ and $\beta = 0.20$ were used for the statistical routines. The cost of replication was assumed to be \$2.00. The results show that all three optimization techniques came fairly close to the actual optimal of 7300. This indicates that search procedures can successfully be employed to optimize simulation experiments. The optimal point, total number of replications and the simulated cost for each of the optimization techniques is presented in Table 4. No attempt is made to compare their relative performance.

The results also indicate that the termination criterion based on cost, α and β errors and the rates of return is effective. The number of iterations for the three search procedures ranges from 32 to 275 indicating that termination based only on the number of iterations would be grossly inadequate. Variations in the evaluation of the objective function are inherent in simulation experiments. Depending upon the magnitude of this variation, a step in the right direction may appear to be otherwise. A termination based only on the improvement between successive iterations, when encountering the above situation would terminate the search prematurely. To illustrate, termination under

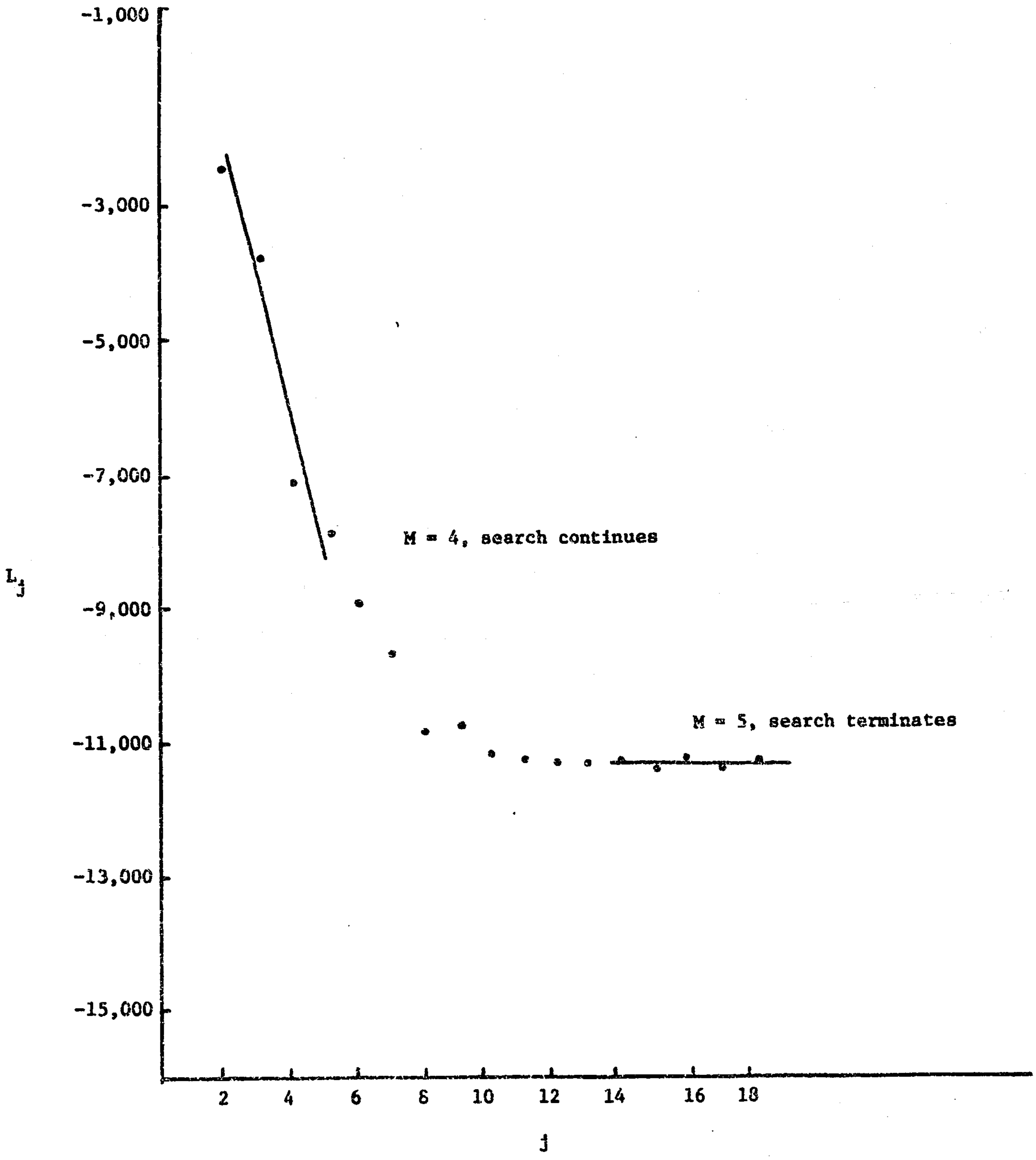

FIGURE 4: Losses L_j from Table 3 for a search where termination occurred at the 18th improvement

Iter #	Simulated Cost	Cumulative # of Replications	Cumulative Replication Cost	Improvement Point Number (j)	Value At Improved Point (z_j)	Loss L_j
1	19819.45	4	8.00	1	19819.45	-
2	20034.90	8	16.00	-	-	-
3	19594.81	12	24.00	2	19594.81	-200.64
4	19369.10	16	32.00	3	19369.10	-418.35
5	19164.61	21	42.00	4	19164.61	-612.84
6	18968.97	25	50.00	5	18968.97	-800.48
7	23701.50	28	56.00	-	-	-
8	19750.59	31	62.00	-	-	-
9	18175.44	34	68.00	6	18175.44	-1576.01
.						
.						
.						
29	7633.54	99	198.00	20	7633.54	-11988.46
42	7601.88	138	276.00	21	7601.88	-11942.12
45	7391.50	144	288.00	22	7391.50	-12140.50
50	7337.99	169	338.00	23	7337.99	-12143.46
73	7324.22	238	476.00	24	7324.22	-12019.23
81	7319.56	262	524.00	25	7319.56	-11975.89

TABLE 1: Simulation Results for One at a Time Search Applied to a Five Variable Inventory Problem

Iter #	Simulated Cost	Cumulative # of Replications	Cumulative Replication Cost	Improvement Point Number (j)	Value At Improved Point (z_j)	Loss L_j
1	19820.92	5	10.00	1	19820.92	-
2	19635.42	9	18.00	2	19635.42	-167.50
3	19010.94	13	26.00	3	19010.94	-783.97
4	18606.25	17	34.00	4	18606.25	-1180.67
5	18089.67	21	42.00	5	18089.67	-1689.24
6	16943.89	26	52.00	6	16943.89	-2825.01
7	14138.26	29	58.00	7	14138.26	-5624.66
8	13948.89	36	72.00	8	13948.89	-5800.01
9	13339.78	38	76.00	9	13339.78	-6405.13
10	12957.38	43	86.00	10	12957.38	-6777.53
11	12546.43	45	90.00	11	12546.43	-7184.49
12	11436.08	47	94.00	12	11436.08	-8290.84
13	7911.42	50	100.00	13	7911.42	-11809.50
14	7911.53	53	106.00	-	-	-
15	8074.42	56	112.00	-	-	-
.						
.						
21	7626.95	74	148.00	14	7626.95	-12045.96
.						
32	7596.78	129	258.00	15	7596.78	-11966.13

TABLE 2: Simulation Results for Pattern Search Applied to a Five Variable Inventory Problem

Iter #	Simulated Cost	Cumulative # of Replications	Cumulative Replication Cost	Improvement Point Number (j)	Value At Improved Point (z_j)	Loss L_j
1	19820.00	4	8.00	1	19820.00	–
2	17372.57	8	16.00	2	17372.57	-2431.43
3	25635.70	12	24.00	–	–	–
5	16057.61	20	40.00	3	16057.61	-3722.39
.						
.						
.						
223	8193.09	456	912.00	10	8193.09	-10714.91
228	7726.22	466	932.00	11	7726.22	-11161.78
230	7635.86	470	940.00	12	7635.86	-11244.14
237	7566.05	484	968.00	13	7566.05	-11285.95
244	7525.02	498	996.00	14	7525.02	-11298.98
250	7520.51	510	1020.00	15	7520.51	-11279.49
251	7377.21	512	1024.00	16	7377.21	-11418.79
265	7373.00	540	1080.00	17	7373.00	-11367.00
269	7338.58	548	1096.00	18	7338.58	-11385.00
275	7337.26	560	1120.00	19	7337.26	-11362.74

TABLE 3: Simulation Results for Sequential Quadratic Approximation Method Applied to a Five Variable Inventory Problem

Optimization Technique	Optimal Point					Total # of Replic.	Simulated Cost
	x_1	x_2	x_3	x_4	x_5		
One-at-a-Time	50.00	50.00	100.00	162.50	87.50	262	7320.11
Pattern Search	100.00	100.00	100.00	180.00	100.00	129	7596.78
Sequential Quadratic Approximation	58.24	47.25	120.99	175.19	92.67	560	7337.26
Theoretical Values	47	50	107	163	91		7300

TABLE 4: Summary of Simulation Results for Three Optimization Techniques Applied to a Five Variable Inventory Problem

this criterion would have occurred at iteration number 42 in Table 1. The termination criterion defined in this paper effectively overcomes the situation described above. Figure 4 illustrates the operation of the termination criterion. The data from the "Losses" column in Table 3 is used to plot Figure 4.

Conclusions

The results of this study demonstrate that search procedures may be effectively used in the optimization of simulation experiments. More importantly, a termination criterion based on cost of replications, α and β errors, minimum rate of return and desired rate of return is proposed. It is found that the termination criterion based on these economic and statistical considerations is effective for simulation experiments. This termination criterion may be used for any search procedure applied to stochastic systems.

Bibliography

1. Box, G. E. P., "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples", Biometrics, 10, 1954.

2. Box, G. E. P., and Wilson, K. B., "On the Experimental Attainment of Optimum Conditions", Journal of the Royal Statistical Society, Vol. 13, No. 1, 1951.

3. Box, M. J., "A Comparison of Several Current Optimization Methods, and the Use of Transformations in Constrained Problems", Computer Journal, Vol. 9, No. 1, May 1966.

4. Brooks, Samuel H., "A Comparison of Maximum-Seeking Methods", Operations Research Journal, Vol. 7, 1959.

5. Draper, Norman R., "Ridge Analysis of Response Surfaces", Technometrics, Vol. 5, No. 4, November 1963.

6. Fletcher, R., and Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization", Computer Journal, Vol. 6, No. 2, July 1963.

7. Friedman, Milton, and Savage, L. J., Chapter 13, "Planning Experiments Seeking Maxima", Techniques of Statistical Analysis, McGraw-Hill, New York, N. Y., 1947.

8. Graybill, Franklin A., An Introduction to Linear Statistical Models, Vol. 1, McGraw-Hill, New York, N. Y., 1961.

9. Gue, Ronald L., and Thomas, Michael E., Mathematical Methods in Operations Research, Macmillan, London, 1968.

10. Hooke, R., and Jeeves, T. A., "Direct Search Solution of Numerical and Statistical Problems", J. Assn. Comp. Mach., Vol. 8, No. 2, April 1961.

11. Kiefer, J., and Wolfowitz, J., "Stochastic Estimation of the Maximum of a Regression Function", Annals of Mathematical Statistics, 22, 1952.

12. Montgomery, Douglas Carter, "Some Techniques for Determining Machine Center Capacities in a Job-Shop", Ph.D. Dissertation, Virginia Polytechnic Institute and State University, 1969.

13. Myers, R. H., Response Surface Methodology, Allyn and Bacon, Inc., Boston, Massachusetts, 1971.

14. Sasser, W. Earl, Burdick, Donald S., Graham, Daniel A., and Naylor, Thomas H., "The Application of Sequential Sampling to Simulation: An Example Inventory Model", Communications of the ACM, 1970.

15. Schmidt, J. W., and Taylor, R. E., Simulation and Analysis of Industrial Systems, Richard D. Irwin, Homewood, Illinois, 1970.

16. Schmidt, J. W., and Taylor, R. E., "System Optimization Through Simulation", Simulation, Vol. 18, No. 2, February 1972.

17. Smith, James Ross, "The Method of Successive Quadratic Approximations With Application to Simulation", Ph.D. Dissertation, Virginia Polytechnic Institute and State University, 1971.

18. Spendley, W., Hext, G. R., and Himsworth, F. R., "Sequential Application of Simplex Designs in Optimization and Evolutionary Operation", Technometrics, November 1962.

19. Wilde, Douglass J., and Beightler, Charles, S., Foundations of Optimization, Prentice-Hall, Englewood Cliffs, N. J., 1967.

20. Wine, R. Lowell, Statistics for Scientists and Engineers, Prentice-Hall, Englewood Cliffs, N. J., 1964.