

Useful Search Techniques to Save Research Time

T. L. Thompson and R. M. Peart
ASSOC. MEMBER ASAE MEMBER ASAE

AGRICULTURAL engineers are always searching for an optimum. As research and design techniques become more sophisticated, and with inexpensive computing time more readily available, methods of searching for this optimum can rely more on precise mathematical expressions and less on opinions and guesses. The search techniques presented here are used when it is impossible or impractical to solve directly for the optimum value. In fact, in many cases the function to be optimized is unknown, as in an experimental problem where, for instance, the environmental humidity for maximum livestock performance is desired. A problem where it may be impractical to solve directly is one that is described by a complex equation or by a series of equations and tabular data which may be empirical. These search techniques are valuable for such problems that cannot be optimized by other techniques, such as setting the derivative equal to zero or by linear programming.

The only requirements are that a value of the function can be determined for any given set of variables and that the function is "well-behaved" (no unbounded or multiple peaks, etc.).

Search and Research in Design

While these techniques are particularly valuable in research, they are also useful to the design engineer. Research and design are not two separate activities of engineers. Every good design involves research, whether it be only a thorough literature search or a long-term research and development program.

Design engineers are finding increased value in mathematical simulation techniques. When the performance of a machine or process can be simulated with a fair degree of accuracy, simulated testing programs take only minutes, and computing costs are practically negligible compared to physical testing costs. Real experiments cannot be eliminated safely, of course, but their length and cost can be greatly reduced. Some of the research techniques de-

scribed here have been used by the senior author in a research program aimed at optimizing the performance of various types of grain driers. A simulation model consisting of a thin-layer drying equation, a mathematical representation of the psychrometric chart, equilibrium relative humidity data, and heat-balance equations was developed to predict the continuous performance of a given drier. With the simulation model, the computer determined the final moisture content while operating with a particular set of operating conditions—air temperature, airflow rate, grain-flow rate, bed depth, and initial moisture content.

Before using mathematical-search techniques, it was difficult to predict a set of operating conditions that resulted in the desired final moisture content. With the methods presented here for finding the zero of a function, it was possible to determine systematically a set of drying conditions which resulted in the desired final moisture with only four or five computer trials. For this determination, one of the variables was varied (holding the others constant) until the desired result was obtained. These search methods made it easy to determine how each variable affected the solution of the overall problem.

The drying methods were then optimized by using the multidimensional search techniques presented here.

METHOD FOR DETERMINING ZEROS OF FUNCTIONS

The zero (or root) of a function is that value of the independent variable x such that y , the dependent variable, is equal to zero. The function $f(x) - \bar{y}$, where \bar{y} is a desired value of y , is also a function of x . Thus these methods can be used to find x for any desired value of y .

The methods presented here can be used to find the zeros of unknown functions. Method I is readily explained by a graphical illustration and can be used on convex or concave functions. Method II is a mathematical procedure that is especially adaptable for use in a digital computer and can be used with any single-valued continuous function. For curved functions, these methods will converge faster than linear interpolation since they account for the curvature of the unknown function.

Method I. Zeros of Convex and Concave Functions

This method was presented by Bell-

man and Dreyfus (1)** and developed further by the senior author. The method utilizes the properties of a convex (concave) curve, that is, a straight line joining any two points on a convex (concave) curve is above (below) the curve, and the extension of this line in either direction is below (above) the curve.

The problem is to reduce the maximum length of interval which includes the desired zero, given that the unknown function is monotonically decreasing (increasing), continuous, and convex (concave) in the interval $[a, b]$ and that $f(a) > 0$ and $f(b) < 0$. The only stipulation on this function is that it can be evaluated for any value of $f(x)$ in the interval. The interval is sequentially reduced until the investigator is satisfied the resulting $f(x)$ value is sufficiently close to zero.

Procedure for Graphically Finding the Zero of an Unknown Convex Monotonically Decreasing Function†

1 Refer to Fig. 1. Draw a straight line from $f(a)$ to $f(b)$. Call the intercept of this line and the x -axis point W . The zero, x^* , is in the interval $[a, W]$. Let ϵ represent the maximum deviation from zero for which the experimenter will accept $y = f(x)$ as being sufficiently close to zero ($|y| < \epsilon$).

2 Arbitrarily select a value of x in the remaining interval (a good choice is at midpoint) and evaluate $f(x)$.

3 If $f(x) > \epsilon$ go to step 4.
 $-\epsilon \leq f(x) \leq \epsilon$ x is the desired value.

$f(x) < -\epsilon$ go to step 5.

4 Refer to Fig. 2. Draw a straight line from $f(a)$ through $f(x)$. Call the intercept of this line with the x -axis point S . Draw a straight line from $f(x)$ to $f(b)$. Call the intercept of this line with the x -axis point W' . The zero, x^* , is in the interval $[S, W']$. Redefine the point x as point a and go to step 2‡.

5 Refer to Fig. 3. Draw a straight line from $f(a)$ to $f(x)$. Call the intercept of this line with the x -axis W' . Draw a straight line through $f(x)$ to $f(b)$. Call the intercept of this line with the x -axis point S' . Let point S be the larger of a and S' . The zero, x^* , is in the interval $[S, W']$. Redefine the point x as point b and go to step 2.‡

† The operations can be described mechanically by using similar triangles.

‡ $f(S)$ and $f(W')$ were not evaluated, therefore, S and W' cannot be used as boundary points.

Fig. 2 and 3 show that this method is very powerful in reducing the maxi-

Paper No. 66-512 presented at the Annual Meeting of the American Society of Agricultural Engineers at Amherst, Mass., June 1966, on a program arranged by the Education and Research Division.

The authors—T. L. THOMPSON and R. M. PEART—are associate professor of agricultural engineering, University of Nebraska, Lincoln; and professor of agricultural engineering, Purdue University. Approved for publication as paper No. 3331 of the journal series of the Purdue University Agricultural Experiment Station.

** Numbers in parentheses refer to the appended references.

imum interval which includes the zero x^* .

Method I can be used with four types of curves—convex or concave and either monotonically increasing or decreasing. However, in order to use this method systematically in the computer, different sets of equations have to be used for each curve type. An alternate method is to transform (reflect about the axis) the different curve types to a single type. However, the curve type has to be known.

Method II. Zeros of Unknown Functions Using an Interpolating Polynomial

This method was developed by the senior author to provide a general method for finding the zero of any single-valued continuous function. The method uses Aitken's algorithm for an interpolating polynomial (2) and is more adaptive to mathematical analysis than Method I, especially in a digital computer.

The method systematically predicts where the zero of the unknown curve is located based on all of the previous observations. In general, after $n+1$ experiments, the prediction is equivalent to solving for the n th order equation through the $n+1$ observed points, and the zero of this equation is the predicted zero of the unknown function.

Procedure for Finding the Zero of Any Single-Valued Continuous Function

Step

1 Arbitrarily select two values of the independent variable (Call them x_1 and x_2) in the vicinity of the desired x and evaluate $f(x_1)$ and $f(x_2)$. Let ϵ represent the maximum deviation from zero for which the experimenter will accept $y = f(x)$ as being sufficiently close to zero.

2 Reindex and arrange the x and y observations in order of increasing y values in a table as shown below. The table appears as it would be after n observations.

X					Y
x_1	x_{11}				y_1
x_2	x_{21}	x_{22}			y_2
x_3	x_{31}	x_{32}	x_{33}		y_3
.	.				.
.	.				.
x_i	x_{i1}	x_{i2}	x_{i3}	...	y_i
.	.				.
.	.				.
x_n	x_{n1}	x_{n2}	x_{n3}	...	y_n
				x_{ni}	
				x_{nn}	

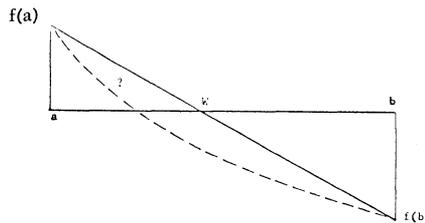


FIG. 1 Initial conditions for Method I.

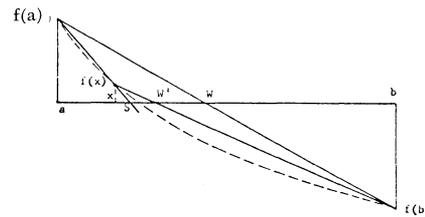


FIG. 2 Method I graphical procedure if $f(x) > 0$.

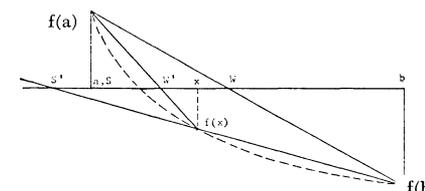


FIG. 3 Method I graphical procedure if $f(x) > 0$.

Let $x_{i1} = x_i, i=1,2,3, \dots, n$
 Each column is calculated from the x column immediately to its left. Calculate

$$X_{ij} = \frac{(x_{j-1, j-1})(y_i) - (x_{i, j-1})(y_{j-1})}{y_i - y_{j-1}}$$

for $j=2,3, \dots, n$
 $i=j, j+1, \dots, n$

x_{nn} is the best estimate of the desired x .

3 Evaluate $y = f(x_{nn})$. If $-\epsilon \leq y \leq \epsilon$, x_{nn} is the desired answer. If not, repeat step 2 using $y = f(x_{nn})$ as another observation.

2	x					y
	0	0				-8
	7.059	7.059	18.824			-5
	14.622	14.622	8.355	14.065		6
	15.000	15.000	7.059	14.622	12.951	9

An example problem and solution is presented below to illustrate the procedure for Method II. The function evaluations were arbitrarily selected by the authors for this illustration.

Example problem:

Refer to Fig. 4. The equivalent curves are drawn for illustration.

Step

1 Let $x_1 = 0, x_2 = 15, \epsilon = 1$, then $y_1 = -8, y_2 = 9$.

2	x			y
	0	0		-8
	15	15	7.059	9

$$x_{22} = \frac{0(9) - 15(-8)}{9 - (-8)} = 7.059$$

This step is equivalent to solving for the intercept of curve 1 and the x axis.

Curve 1 is a straight line through the two points.

3 $y = f(7.059) = -5$
 y is not in the range ± 1 .

2	x				y
	0	0			-8
	7.059	7.059	18.824		-5
	15.000	15.000	7.059	14.622	9

$$x_{22} = \frac{0(-5) - 7.059(-8)}{-5 - (-8)} = 18.824$$

This step is equivalent to solving for the intercept of curve 2 and the x axis.

Curve 2 is a second-order curve through the three points.

3 $y = f(14.622) = 6$
 y is not in the range of ± 1 .

This step is equivalent to solving for the intercept of curve 3 and the x axis.

Curve 3 is a third order curve through the four points.

3 $y = f(12.951) = 4$
 y is not in the range ± 1 .

2	x					y
	0	0				-8
	7.059	7.059	18.824			-5
	14.622	14.622	8.355	14.065		6
	15.000	15.000	7.059	14.622	12.951	9

This step is equivalent to solving for the intercept of curve 4 and the x axis. Curve 4 is a fourth order curve through the four points.

3 $y = f(12.951) = 4$
 y is not in the range ± 1 .

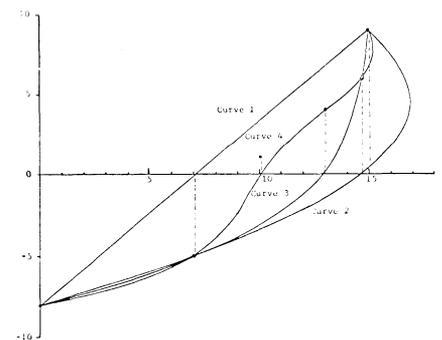


FIG. 4 Graphical solution for Method II. TRANSACTIONS OF THE ASAE • 1968

2	x					y
	0					-8
	7.059	7.059	18.824			-5
	12.951	12.951	8.634	13.163		4
	14.622	14.622	8.355	14.065	11.359	6
	15.000	15.000	7.059	14.622	11.996	9

This step is equivalent to solving for the intercept of curve 4 and the x-axis.

Curve 4 is a fourth order curve through the five points.

$$3 \quad y = f(10.085) = 1$$

y is in the allowable range; therefore, the best estimate of the zero, x^* , is 10.085.

UNIMODAL SEARCH TECHNIQUES

Unimodal search techniques can be used to find the optimal value of an unknown function when it can be assumed that the surface under consideration has only one peak or that there are not any false peaks.

A function is unimodal if, from every point on the surface, a strictly rising path exists which leads to the optimal value or peak. In multidimensional searches, this path does not have to be straight.

Single variable search is important to the understanding and use of multidimensional search methods. One-dimensional searches are used in some of the multidimensional searches.

It is the purpose of this paper to present workable methods and illustrations for each of these procedures. Detailed developments of the methods are presented by Wilde (3).

Single Variable Search

The problem is to find the maximum of an unknown unimodal function, given that the maximum is defined between two limits, i.e., $a \leq x^* \leq b$. A function is unimodal if $f(x)$ increases as x increases from a to x^* and $f(x)$ increases as x approaches x^* from b .

The interval of the x axis where the optimum is known to exist is called the "interval of uncertainty." Before any evaluations are made, the length of the interval of uncertainty is $(b-a)$. After the second and every following evaluation, a portion of this interval can be eliminated from consideration. When maximizing, if $f(x_2) > f(x_1)$ (Fig. 5), the region from a to x_1 can be elim-

inated since $f(x)$ is unimodal. Likewise, if $f(x_1) > f(x_2)$ (Fig. 6), the region from x_2 to b can be eliminated.

Simultaneous Search

Simultaneous search is used when all of the experiments or evaluations, $f(x)$, are performed at the same time. The procedure consists of equally spacing the n evaluations along the interval of uncertainty. This procedure does not take advantage of the unimodality of the function and is inefficient. After the n experiments are performed, and assuming the function is unimodal, the investigator can reduce the interval of uncertainty to $\left[\frac{2(b-a)}{n-1} \right]$, or the length of the two intervals adjacent to the best $f(x)$ value. Referring to Fig. 7, and assuming the function is unimodal, it can be asserted that the optimum lies between x_{k-1} and x_{k+1} where $f(x_k)$ was the best $f(x)$ evaluation.

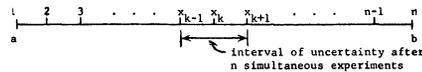


FIG. 7 Simultaneous search with n experiments.

Sequential Search

Sequential search techniques use the results from previous experiments to determine the location of the next experiment. The methods presented consist of various procedures for locating the next experiment in the interval of uncertainty. The purpose is to reduce the interval with the fewest number of experiments. The rate of reduction of the interval of uncertainty does not depend on the outcome of any experiment. The methods minimize the maximum interval of uncertainty.

Let ϵ (epsilon) represent the least separation between two experiments for which a difference between $f(x_1)$ and $f(x_2)$ can be detected, i.e., $f(x) \neq f(x \pm \epsilon)$. Experimental and computational errors should be considered when ϵ is determined for an experiment.

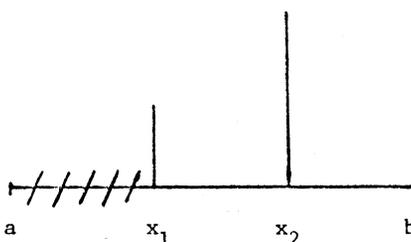


FIG. 5 Remaining interval of uncertainty when $f(x_2) > f(x_1)$.

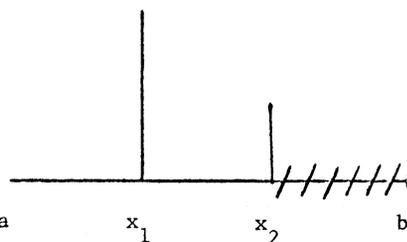


FIG. 6 Remaining interval of uncertainty when $f(x_1) > f(x_2)$.

A *Dichotomous search* consists of placing a pair of experiments in the center of the interval of uncertainty. The two experiments are placed a distance ϵ apart and part of the interval is eliminated from consideration based on the outcome of the two experiments. The procedure is repeated until the interval of uncertainty is sufficiently small.

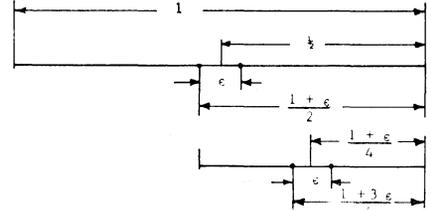


FIG. 8 Dichotomous search with four experiments.

In general, after n (an even number) experiments, the interval of uncertainty is equal to a fraction $[2^{-(n/2)} + (1 - 2^{-(n/2)}) \epsilon]$ of the original interval.

A *Fibonacci search* is based on the value of ϵ and the number of experiments that are to be performed. A Fibonacci series is described by Fibonacci numbers defined as follows: $F_0 = F_1 = 1$, $F_k = F_{k-1} + F_{k-2}$, $k = 2, 3, 4, \dots$. The Fibonacci series F_0, F_1, F_2, \dots is 1, 1, 2, 3, 5, 8, 13, 21, or each number is the sum of the two numbers preceding it. The Fibonacci search method consists of placing the first experiment

a fractional distance $\left[\frac{F_{n-1} + (-1)^n \epsilon}{F_n} \right]$ from one end of the interval of uncertainty, and the second experiment the same distance from the other end, where n is the number of experiments that are to be performed. The function $f(x)$ is evaluated for the two experiments, and part of the interval is eliminated. The third (and each consecutive) experiment is placed symmetrical to the one in the remaining interval, $f(x)$ is evaluated, and part of the interval is eliminated. The procedure is repeated until n experiments have been performed. A simple method for this symmetrical placement is to use the following equation: $x = a + b - x_n$, where a and b are the lower and upper limits of the interval of uncertainty, and x_n is the x value that resulted in the best $f(x)$. After the n sequential experiments have been performed, the interval of uncertainty is a fraction $\left[\frac{1 + F_{n-2} \epsilon}{F_n} \right]$ of the original interval.

Following is an example problem and the solution: Use the Fibonacci search technique to optimize the unimodal function shown in Fig. 9. (Normally the function is not known but is given here for illustrative purposes). Use the figure to evaluate the function at each

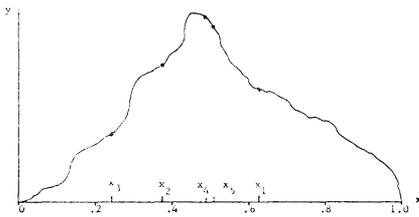


FIG. 9 Arbitrary function for Fibonacci search example problem.

of the x 's. Perform 5 experiments and let $\epsilon = 0.02$.

$$\begin{aligned} \text{Using the formula } x_1 &= \frac{F_4 + (-1)^n \epsilon}{F_5} \\ &= \frac{5 + (-1)^5 (0.02)}{8} \\ &= 4.98/8 = 0.6225 \end{aligned}$$

for the location of the first experiment and $x = a + b - x_b$ for each following experiment, the table below shows the sequence of events for this optimization.

Exp.	a	x_b	b	x	$f(x) ? f(x_b)$	Therefore,
1	0		1.000	0.6225		$x_b = 0.6225$
2	0	0.6225	1.000	0.3775	$f(x) > f(x_b)$	$x_b = 0.3775, b = 0.6225$
3	0	0.3775	0.6225	0.2450	$f(x) < f(x_b)$	$a = 0.2450$
4	0.2450	0.3775	0.6225	0.4900	$f(x) > f(x_b)$	$a = 0.3775, x_b = 0.4900$
5	0.3775	0.4900	0.6225	0.5100	$f(x) < f(x_b)$	$b = 0.5100$

Thus the optimum is in the interval [0.3775, 0.5100]. The length of the remain-interval of uncertainty is the same as the theoretical interval

$$\frac{1 + F_3 \epsilon}{F_5} = \frac{1 + 3(0.02)}{8} = 0.1325.$$

The *Golden Section Search* technique is very similar to the Fibonacci search, but does not depend on the number of experiments or the ϵ value. In fact, this method is the same as the Fibonacci search, except the first experiment is

$$\text{placed a fraction } \frac{2}{1 + \sqrt{5}} = 0.6180340$$

of the distance from one end of the interval of uncertainty, and the experimenter can perform as many experiments as he desires. The interval remaining after n experiments is a fraction $\left[\frac{2}{1 + \sqrt{5}} \right]^{n-1}$ of the original interval of uncertainty.

The golden section is known as a division of a segment into two unequal parts so that the ratio of the whole to the larger is equal to the ratio of the larger to the smaller.

Lattice Search is used when only a certain number of discrete x values are possible. An example is how many salesmen should a company assign to a territory to maximize the company's profit. Since it is impossible to get a fraction of a salesman, only a discrete number of salesmen is possible. Assum-

ing that k discrete x values are arranged so that the resulting function is unimodal, the search consists of locating the first two experiments at the numbered x values corresponding to the two largest Fibonacci numbers (1, 1, 2, 3, 5, 8, 13, 21, ...) included in the k possible x 's which are numbered 1 through k . The next step consists of adding dummy variables to the sequence of x values in order to have a total of $F_n - 1$ values, where F_n is the first Fibonacci number greater than k . After the first two experiments are performed, a portion of the interval is eliminated as in the previous search methods, and the next experiment is performed symmetrical to the one in the remaining interval. The procedure is repeated until the best value of the function is obtained with the two adjacent values of the function evaluated.

Example Problem and Solution

If a manufacturing company can assign up to ten salesmen to a territory, how many should they assign to maximize the company's profit? Assume that the profit function is unimodal. The solution consists of adding two dummy salesmen to the ten possible so that there are $F_6 - 1 = 12$ possibilities.

Next, the company should evaluate their profit with $F_4 = 5$ and $F_5 = 8$ salesmen.

If the company finds that it is more profitable to have 8 salesmen than five, the next evaluation should be made with 10 salesmen (symmetrical to 8 in the interval 6 to 12). One possible sequence of making further evaluations is shown in Fig. 10.

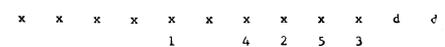


FIG. 10 One possible sequence of events for a lattice search with ten possibilities.

To illustrate the power of these various one-dimensional-search techniques, the following table shows the fraction of the original interval remaining after twenty experiments have been performed:

FRACTION OF ORIGINAL INTERVAL REMAINING AFTER TWENTY EXPERIMENTS	
Simultaneous search	1/ 10
Sequential search	
Dichotomous search	1/ 1,024
Fibonacci search	1/10,946
Golden search	1/ 9,349
Lattice search*	1/17,710

* With 20 experiments, one value can be selected from up to 17,710 possible values, assuming the values are arranged so that the function is unimodal.

For most one-dimensional searches, the Golden section is preferred by the authors because of its power and simplicity. While the number of experiments to perform can be calculated for the Fibonacci and the Golden section methods, the Golden section method does not require this calculation and does not depend on the ϵ value. However, there is a slight reduction in power over the Fibonacci search.

Multivariable Search

The transition from one-dimensional to multivariable search encounters many difficulties. With unknown functions, the gradient and tangent plane at any point cannot be obtained directly, but have to be approximated by performing a series of experiments in the neighborhood of the point in question. With one or two independent variables, it is relatively easy to visualize and illustrate methods of seeking an optimum, but with more variables it is impossible to operate with functions except in a purely mathematical sense. The number of possible solutions and difficulties in optimization increase very rapidly as the number of variables increase.

For each one of the search techniques presented, the method is briefly described and a step-by-step procedure for performing a multidimensional search is given. Each method is illustrated by graphically showing how the method optimizes an ellipsoidal contour in two dimensions. The advantages and disadvantages of the method and an estimate of the number of experiments necessary are given.

All of the search techniques are illustrated with the same example problem. For methods of illustration, the function of the example problem will be known. With a known function, the tangent plane at any point and the maximum along any line in space can be evaluated by using first derivatives. In practice, this simplification will generally not be possible, but is used here to illustrate how the methods perform with an ellipsoidal functional. The example problem used does not necessarily illustrate all of the characteristics of the methods, but is used only to illustrate how the method is performed.

A linear approximation to the tangent plane and thus the gradient line and contour tangent can be obtained at any point $\vec{x}_0 = (x_1, x_2, \dots, x_k)$ by the following procedure, where k is the number of independent variables,

- 1 Evaluate $y(\vec{x}_0)$
- 2 Let $\vec{x}_i = (x_1, x_2, \dots, x_{i-1}, x_i + d_i, x_{i+1}, \dots, x_k)$
Evaluate $y(\vec{x}_i)$
Let $\Delta y_i = y(\vec{x}_i) - y(\vec{x}_0)$ and $m_i = \Delta y_i / d_i$
for $i = 1, 2, \dots, k$

A linear approximation of the tangent plane at \bar{x}_0 is

$$\Delta y = \sum_{j=1}^k m_j \Delta x_j, \text{ where } \Delta x_j = x_j - x_{0j} \text{ and } \Delta y = y - y_0.$$

where x_j is the new value of variable j and y is the resulting height of the tangent plane.

The contour tangent plane is determined by setting Δy equal to zero. The gradient line at point x_0 can be expressed parametrically as

$$\bar{x} = (x_1 + \lambda m_1, x_2 + \lambda m_2, \dots, x_k + \lambda m_k).$$

Thus, with this representation of the gradient line, the value of each x variable is determined by the value of parameter λ .

In practice, the experimenter should further investigate the neighborhood of the optimum point after the search techniques presented here are used. A second-order approximation of the optimum should be made to see if a better value is nearby or if the "optimum" is a saddle point. A detailed method for making a second-order approximation of the surface is presented by Wilde (3).

Contour Tangent Elimination

As previously illustrated, after $k + 1$ experiments have been performed in the neighborhood of a point, the tangent plane at that point can be approximated. The contour tangent at any point x_0 is the hyperplane such that $\Delta y = 0$ (if the response surface is represented by the tangent plane, the value of the function remains the same as at point x_0). The contour tangent elimination method consists of eliminating the areas where $\Delta y < 0$ from the experimental region.

Following is the step-by-step procedure for the contour tangent elimination method:

- 1 Define the feasible region.
- 2 Select an arbitrary point x , preferably in the center of the feasible region, and approximate the tangent plane Δy at this point.
- 3 Eliminate the region defined by $\Delta y < 0$ from the feasible region.
- 4 Return to step 2 and repeat until the remaining feasible region defines the optimum to the experimenter's satisfaction.

Example Problem and Solution

Optimize the function shown in Fig. 11 using the contour tangent elimination method. Do not use the plotted contours to select new arbitrary points, as the functional is usually unknown. At the starting point x_1 , the tangent plane is evaluated, and the region below the contour tangent plane eliminated from consideration. Another arbitrary point x_2 is selected and the re-

gion below its contour tangent plane is eliminated. The same procedure is performed for points x_3, x_4, \dots until the optimum value is obtained.

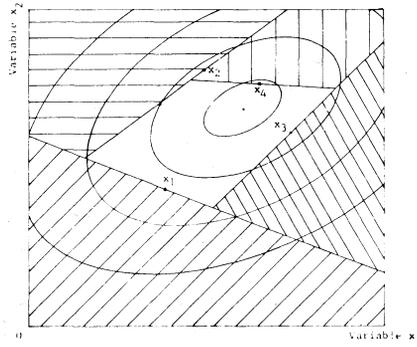


FIG. 11 An illustration of the contour tangent elimination method.

Advantages of contour tangent elimination method:

- 1 The method is simple and is fairly straightforward.
- 2 If the centroid of the feasible region is used as the new point, one-half of the feasible region is eliminated with each iteration.

Disadvantages of contour tangent elimination method:

1 The method guarantees that the optimum will be found only if the function is strongly unimodal[‡].

2 With each iteration, it becomes more difficult to find new arbitrary starting points and increasingly difficult to locate the centroid of the feasible region. With two variables, the feasible region can be plotted and the centroid found quite easily, but with more variables it becomes almost impossible.

3 The experimenter has many constraints and often he does not know how to evaluate them.

Number of experiments for contour tangent elimination method:

$k + 1$ experiments for each iteration, where k is the number of independent variables.

Gradient Method

The "gradient method" or "method of steepest ascent" consists mainly of mathematically climbing the response surface by moving up the surface using the steepest slope. This is very similar to an explorer walking to the top of a mountain when he cannot see the top. He continues to climb by walking up the steepest slope until he reaches the top. If the response surface is unimodal, the gradient method will eventually reach the summit of the surface.

Step-by-step procedure for the gradient method:

- 1 Select an arbitrary starting point \bar{x} .
- 2 Approximate the gradient line and represent it in the form $(x_1 +$

[‡] A function is strongly unimodal if a straight line running to the optimum, x , from any point in the experimental region is a rising path.

$\lambda m_1, x_2 + \lambda m_2, \dots, x_k + \lambda m_k)$ and perform a one-dimensional search maximizing y as λ varies. The methods for approximating the gradient line and performing one-dimensional searches are described above.

- 3 Set $\bar{x} = (x_1 + \lambda^* m_1, x_2 + \lambda^* m_2, \dots, x_k + \lambda^* m_k)$, where λ^* is the optimal λ as determined by the one-dimensional search. Return to step 2 and repeat until the maximum is obtained.

Example Problem and Solution:

Optimize the function shown in Fig. 12 using the gradient method. The gradient line at point p_0 is approximated and the high point (p_2) along this line is determined by a one-dimensional search; p_3 is located at the high point of the gradient line from p_2 ; p_4 is located at the high point of the gradient line from p_3 , and so forth, until the optimum point is obtained.

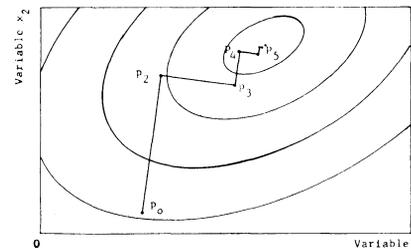


FIG. 12 The gradient method illustrated.

Advantages of gradient method:

- 1 The method can be used on any unimodal function.
- 2 The method will work in the presence of experimental error.
- 3 Gradient methods will inherently stay away from saddle points.

Disadvantages of gradient method:

- 1 Convergence depends on choice of scales; ascent methods will eventually find the peak, and if the scales are selected wisely, convergence can be rapid. (For best choice of scales, make the contours as nearly spherical as possible.)

Number of Experiments for Gradient Method:

$k + 1$ experiments and one one-dimensional search for each iteration.

Sectioning or the Method of Sectional Search

This simple scheme consists of altering only one independent variable at a time and maximizing using a one-dimensional search until the criterion ceases to improve, changing to the next variable, and repeating until no further improvement is obtained. When all of the variables have been used, the process is repeated again until no further improvement is made between subsequent searches of all of the variables.

Step-by-Step Procedure for Sectioning Method:

- 1 Select an arbitrary starting point \vec{x}_0 .
- 2 Maximize the function by allowing one independent variable to vary while holding all of the other variables constant. Use a one-dimensional search to find the maximum.
- 3 Set this variable at the value that resulted in the best value, and allow the next variable to vary. Maximize the function by the one-dimensional search technique. Repeat until all of the variables have been used.
- 4 Go to step 2 above and repeat the sequence until no further improvement is obtained.

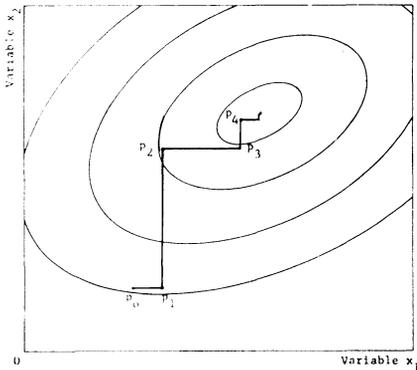


FIG. 13 Method of sectional search illustrated.

Example Problem and Solution:

Optimize the function shown in Fig. 13 using the Method of Sectional Search. From the arbitrary starting point p_0 , the function is maximized over all values of x_1 , while holding x_2 constant. x_1 is set equal to the value that resulted in the best function value, and the function is maximized over variable x_2 . The procedure is continued by varying x_1 , then x_2 , x_1 , x_2 , and so forth until no further improvement is obtained.

Advantages of sectional search:

- 1 The method is simple and easy to perform.
- 2 Performance does not depend on choice of scales.

Disadvantages of sectional search:

- 1 In some cases, the method will not reach the maximum, even when the contours are convex (strong interaction between some of the variables).
- 2 The method will not guarantee convergence to the optimal solution. Therefore, the method is not suitable unless the experimenter knows in advance that ridges are absent.

Number of experiments for sectional search:

k , one-dimensional searches for each iteration.

Pattern Search Technique

The Pattern Search Technique is an optimizing method which has good ridge-following characteristics and requires only very simple calculations to perform. The technique is based on the hopeful conjecture that any set of moves which have been successful during early experiments will be worth trying again. This method uses a set of simple rules and the experience from past observations to direct its next moves.

The search technique automatically adjusts the length and direction of its next excursion, depending on its success in past adventures. The pattern search technique operates by starting at an arbitrary starting point and then climbing up the response surface in the direction that gives improvement, extrapolating in this direction, and then repeating and making corrections in its direction as it goes.

Step-by-Step Procedure for the Pattern Search Technique:

- 1 Select a step size d_i for each independent variable x_i , ($i = 1, 2, \dots, k$)

- 2 Select an arbitrary starting point, p_0

Set $j = 1$, $p = p_0$ and evaluate $y(p)$.

- 3 For any i , find $y(p + d_i e_i)$, where e_i is the i th unit vector in the cartesian basis.

If $y(p + d_i e_i) > y(p)$, set $p = p + d_i e_i$

If $y(p + d_i e_i) \leq y(p)$, evaluate $y(p - d_i e_i)$

If $y(p - d_i e_i) > y(p)$, set $p = p - d_i e_i$

If $y(p - d_i e_i) \leq y(p)$, go on to the next variable.

Repeat this step for all of the variables: $i = 1, 2, 3, \dots, k$.

- 4 Set $p_j = p$ and $p = 2 p_j - p_{j-1}$
- If $y(p_j) > y(p_{j-1})$, set $j = j + 1$, evaluate $y(p)$ and repeat step 3 again.

If $y(p_j) \leq y(p_{j-1})$, set $p_j = p_{j-1} = p$ and reduce the step size d_i for each variable. Evaluate $y(p)$ and repeat step 3 again.

Continue until no more improvement is obtained with a preselected minimum step size for each variable.

Example Problem and Solution

Optimize the function shown in Fig. 14 using the pattern search technique. From the arbitrary starting point p_0 , steps are made in the x_1 and x_2 directions, which indicate an improvement to point p_1 . From point p_1 , an extrapolation of length $p_1 - p_0$ is made in the direction of improvement, and steps are made again to determine point p_2 . From point p_2 , an extrapolation is made, but no improvement is obtained; therefore, the step size is reduced ($\frac{1}{2}$)

and steps are made in the x_1 and x_2 directions and so forth until the optimum is obtained.

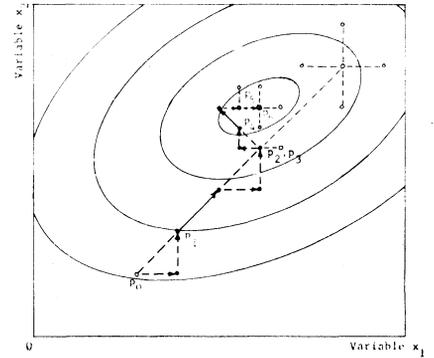


FIG. 14 Pattern search technique illustrated.

Advantages of the Pattern Search Technique:

- 1 The method is powerful on straight ridges.
- 2 Method is simple and requires very simple calculations.
- 3 Performance does not depend on the choice of scale.

Disadvantages of the Pattern Search Technique:

- 1 Performance is sensitive to the step size and the speed at which the grid is reduced to resolve the ridge.
- 2 Pattern search measurements are always taken in directions parallel to the coordinate axes, and for certain contours it could miss the ridge entirely.

Number of Experiments for Pattern Search Technique:

The method required $k + 1$ successful searches, but not more than $2k + 1$ experiments for each iteration.

Partan or the Method of Parallel Tangents

Partan is a method of exactly finding the optimum of concentric ellipsoidal contours in a fixed small number of experiments. The master strategy is based on certain global properties of ellipsoids. Although the method was designed for ellipsoidal functions, it is very powerful on nonellipsoidal contours. It has certain ridge-following properties that make it very attractive, especially when the ridges are straight. Partan can climb like the gradient method, but it accelerates its search in the direction the gradient method is headed, thus increasing its effectiveness for straight ridges and elliptical contours. There are three versions of partan (steep ascent, general, and scale invariant), with each version having a slightly different method of selecting the direction of the next move. All three of the methods are essentially

the same, especially in the number of experiments that have to be performed, but the steep ascent method has the advantage that if the scales of k of the variables are the same, the method will converge in $2(k-1)$ less one-dimensional searches.

Step-by-step procedure for accelerated ascent partan:

- 1 Locate point p_2 at the high point of the gradient line from any arbitrary starting point p_0 .
- 2 Locate point p_3 at the high point of the gradient line from point p_2 .
- 3 Locate p_4 at the high point along the unique line that passes from point p_0 through point p_3 .
- 4 Locate p_5 at the high point of the gradient line from point p_4 .

In general, for even-numbered i step i

Locate point p_{i+1} at the high point of the gradient line from point p_i , step $i+1$

Locate point p_{i+2} at the high point along the unique line that passes from point p_{i-2} through point p_{i+1} .

If the surface consists of concentric ellipsoidal contours, continue for $2k-1$ steps (the optimal solution will be obtained) where k is the number of independent variables. If any of the axes of the ellipsoidal system are equal, the method will reach the optimum sooner.

If the contours are nonellipsoidal, continue until the neighborhood of the optimum is reached or until no further improvement in the solution is obtained.

Example problem and solution:

Maximize the function shown in Fig. 15 using the method of parallel tangents. Point p_0 , p_2 , and p_3 are the same and are determined by the same procedure used for the gradient method example. From point p_3 , a one-dimen-

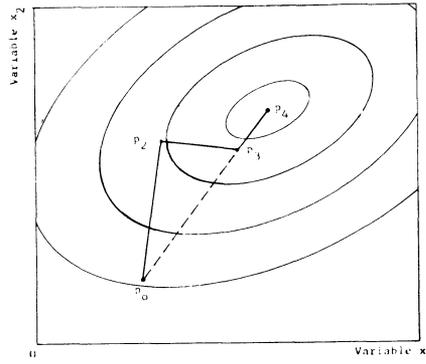


FIG. 15 Method of parallel tangents illustrated.

sional search is performed along the line from point p_0 through point p_3 . Point p_4 , the high point on this line, is the optimum value because the method of parallel tangents locates the optimum in $2k-1 = 3$ one-dimensional searches when the surface has concentric ellipsoidal contours.

Advantages of Partan Search Technique:

- 1 The method finds the optimum in a fixed small number of experiments when applied to concentric ellipsoidal contours.
- 2 Even though the method was designed for ellipsoidal contours, it has all the advantages of the gradient method and has powerful ridge-following characteristics that make it attractive even for non-ellipsoidal contours.
- 3 It can be made invariant to scale changes.
- 4 It is effective in finding and following straight ridges.

Disadvantages of Partan Search Technique:

- 1 The method requires a relatively large amount of stored information.

2 The method involves calculations that are far from simple.

Number of experiments for Partan Search Technique:

k tangent plane measurements ($k + 1$ experiments/tangent plane) and $2k - 1$ one-dimensional searches, where k is the number of independent variables.

For most multidimensional searches, the gradient method is preferred by the authors because it can be used on any unimodal function and will work in the presence of experimental error. While pattern search and the method of sectional search are simple and easy to perform, they could, for certain contours, miss a ridge entirely because their measurements are always taken in directions parallel to the co-ordinates axes. Partan is a very powerful search method, but requires large amounts of stored information and calculations that are far from simple.

In the event that an experimenter does not know enough about a response surface to assume that it is unimodal, the authors suggest that unimodal search techniques be used several times with different starting points to check for unimodality. If all of the explorations lead to the same summit, it appears that the surface may be unimodal. If not, more starting points should be used to determine the number of false peaks. If there appear to be many peaks, unimodal search techniques are not applicable.

References

- 1 Bellman, R. E. and Dreyfus, S. E. Applied dynamic programming. Princeton University Press, Princeton, N.J., pp. 156-160, 1962.
- 2 Henrici, Peter. Elements of numerical analysis. John Wiley & Sons, Inc., New York, pp. 206-207, 1964.
- 3 Wilde, Douglas J. Optimum seeking methods. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.