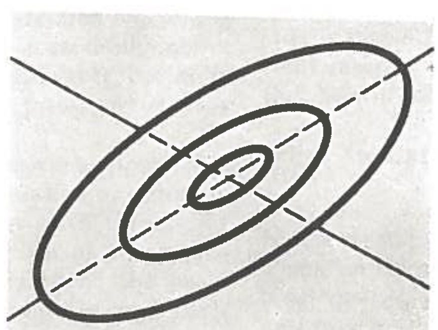
A REPRINT FROM
CHEMICAL ENGINEERING

BOA
1962

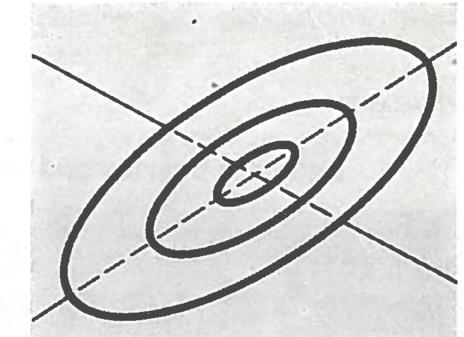
MODERN MATHEMATICAL TOOLS FOR

Optimization

ARNOLD H. BOAS



Reprinted from CHEMICAL ENGINEERING



MODERN MATHEMATICAL TOOLS FOR OPTIMIZATION

Part 1

What Optimization Is All About

There is a wide gap between the mathematical know-how of many practicing chemical engineers and the mathematical tools needed to cope with modern concepts of optimization. This new series will attempt to bridge this gap by introducing the mathematics of these important concepts, emphasizing the practical approach.

ARNOLD H. BOAS, *Socony Mobil Oil Co., Inc.*

Most problems in chemical engineering design or plant operation have at least several, and possibly an infinite, number of solutions.

Selecting the "best" answer to a problem out of the multiplicity of potential solutions is certainly not a new concept to chemical engineers. However, optimum answers are based very often on intuition or past experience.

Now, some sophisticated mathematical tools and techniques are available for the calculation of optimum conditions. These methods, together with the use of modern computers, can minimize much of the guesswork and conjecture that usually surround the problem of choosing the best set of variables in multivariable problems. These techniques, of course, still require skillful use of past experience and even some intuition before they can produce practical information.

There are a number of reasons why engineers are becoming greatly interested in optimization. An important one is that intensive competition in the chemical process industries makes it more necessary than ever that equipment and systems operate at peak performance. Even marginal savings can be extremely vital in this competitive environment.

Watch for Part 2—Lagrange Multipliers, in Jan. 7, 1963, issue. Other parts in future issues will discuss the Fibonacci scan, steepest ascent method, direct search, and a summary of advanced techniques.

Recent interest in this subject is shown by the numerous papers that have appeared in the literature. Some of the latest publications include those of Box,[9] Butcher,[11] Taborek,[26] Wilde,[32] N. Y. U. Symposium,[22] and the AIChE Lecture Series.[1]

However, mathematical techniques required to cope with the problems of optimization may not be too familiar to many practicing chemical engineers. The objective of this series is to present some of these mathematical tools in a comprehensible fashion to engineers who may not feel at home in the area of sophisticated mathematical jargon. Theoretical development will be subservient to practical solutions of problems. References will be noted, so that those readers who are more inclined to the theoretical approach will have ample source material upon which to draw.

It is hoped that at the conclusion of the series sufficient interest will have been instilled in the reader so that he will pursue further reading on the subject.

What to Optimize?

As Box[8] points out, the subject of what to optimize is the first major question that must be answered. Then, an "objective function" is selected, usually related to cost, yield, purity or some other criterion for optimization. This function may be expressed in the form of an equation such as:

$$Y = a + bP + cT + dP^2 + eT^2 + fPT$$

where

Y = yield of a certain product in a chemical reaction.
P = pressure in the reactor.
T = temperature in the reactor.
a, b, c, d, e, f = coefficients in the equation.

Coefficients are usually determined from a series of laboratory or pilot-plant experiments to represent dependence of yield on pressure and temperature. These parameters may vary due to catalyst activity, ambient conditions, feed conditions, etc.

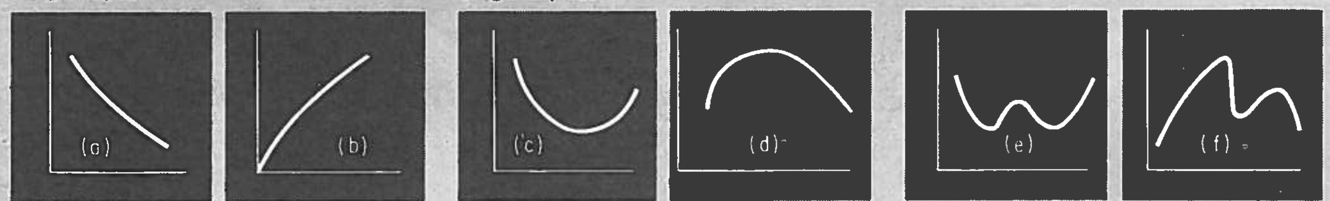
Now the problem is to select P and T so that Y will

Important terms and functions used in optimization techniques—Fig. 1

Monotonic function—A function that is continuously increasing or decreasing (Figs. 1a, b).

Unimodal function—A function with one peak, either maximum or minimum (Figs. 1c, d).

Bimodal function—A function having two maxima or minima (Figs. 1e, f).



State—A detailed description of all factors that may alter the future of the system. A change in state is a way of expressing variation of the system with respect to time. As an example: for a fixed volume process, assume temperature and pressure completely define the system. If the system varies with time, then a description of the temperature and pressure at any time defines the state of the system.

Stochastic process—Webster gives the following definition for the word stochastic: "conjectural; given to, or skillful in, conjecturing." A stochastic (random) process has the characteristic of uncertainty of the output for a given input. Principles of statistics are usually applied and the probable output is considered, based on past history. Very often, techniques of regression analysis are used to establish relations among stochastic variables.

As an example: the yield of product from a certain catalytic reactor may vary with catalyst activity, which varies with time. This relation may not be known quantitatively nor can it be expressed by an analytical expression. The most probable yield would have to be considered at any particular time.

Deterministic process — A process where output is uniquely determined by a given input. As an example: using the reactor for the stochastic case, catalyst activity may be known quite accurately and, therefore, there is no uncertainty.

Adaptive process—A process where output is determined by variable current conditions of input and current operating conditions (see Ref. 8).

Constraint—A restriction that exists among the variables. This may be in the form of an equality or inequality. As an example: temperature in a reactor may have to be below a set maximum value, the mole fractions of components in solution must be positive, etc.

Normalized variables—In the sense that they will be used for search techniques, these are variables that have a lower limit of 0.0 and an upper limit of 1.0. Ordinary variables may be "normalized" by using the maximum and minimum values of the ordinary variables.

Example:

Let x = ordinary variable

y = normalized variable

$= (x - x_{min})/R$

x_{max} = maximum value of x

x_{min} = minimum value of x

R = range of ordinary variable

$= x_{max} - x_{min}$

For the temperature range from 100 F. to 400 F., let us express 200 F. as a normalized variable:

R = 400 F. − 100 F. = 300 F.

x_{min} = 100 F.

$y = (200 - 100)/300 = 0.333$

Evolutionary operation—A system of continuous optimization where small changes in process variables are imposed on an operating plant. These changes are not large enough to disrupt the plant but are of sufficient magnitude to cause a measurable effect on the output (see Ref. 18).

Dynamic programming—A technique that reduces the problem of optimizing N stages simultaneously to one of solving N one-stage problems (see Refs. 3 and 30).

Calculus of variations—A technique that has as its objective the maximization (or minimization) of definite integrals (see Refs. 4, 5, 12, 17, 29).

Pontryagin's maximum principle—A powerful mathematical tool that has been gaining in popularity in recent years. The technique is a modification of the classical calculus of variations (see Ref. 24).

Linear programming—A technique of planning activities in an optimum manner based on a linear mathematical model. This tool finds application in allocation problems, product distribution, meeting product specifications (see Ref. 2, 11a).

be a maximum. The graphical representation of Y vs. P and T may take the form shown in Fig. 2. The variables shown are "free" in that no limits have been placed on them. In reality, for reasons dictated by process or safety, constraints may be imposed on some of the variables. Suppose that P may not exceed P_2 for reasons of safety, and that T has a lower limit of T_1 determined from reaction kinetics. These limits are shown in Fig. 3 where the white area represents the region of interest. Surfaces generated by second-degree equations in two and three variables are discussed by Box.[6]

Care and effort should be taken in defining the objective function as accurately as possible. All the optimization techniques that will be discussed use this function (it need not be in a mathematical form) as a starting point.

Assuming the function has been defined, we can now discuss how to optimize. The various methods of optimization may be classified as follows:

- Analytical.
- Case study.
- Search.

Analytical Method Uses Calculus

The classical method of calculus is used in the analytical method when the first derivative of a differentiable function is set equal to zero.

In the case of a multivariable function, partial derivatives are taken with respect to each of the n variables, and a set of n simultaneous equations is obtained. Where constraints are imposed, Lagrange Multipliers are introduced, one for each constraint, so the number of independent equations and the number of unknowns are identical.

To use the analytical method, a function must be in a mathematical form and must be differentiable. This is usually a serious limitation since many design calculations are of the iterative type, solved by trial-and-error or some numerical method. Another drawback to the analytical method is that only one solution is obtained. There is no indication of the direction or path required to reach this solution.

Case Study Method: Step by Step

Sometimes called "the method of last resort," case studies may be the only method applicable. As the name implies, this technique consists of evaluating various solutions to the problem.

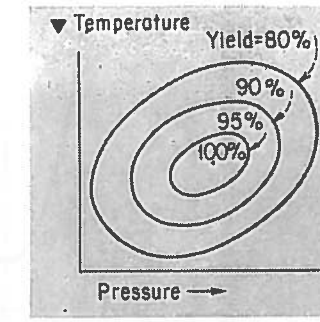
Case after case is tried and results are presented graphically or in tabular form. A study of these graphs or tables yields the "best" combination of variables.

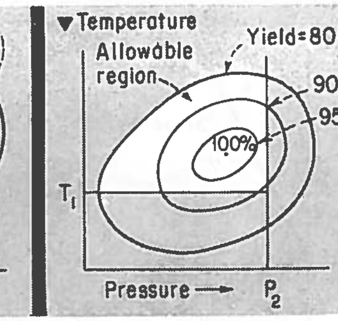
Although the case-study method is well suited for computer use, there are some difficulties in presenting the results for the multidimensional or multivariable problem. However, it is a simple technique and provides many solutions to the problem in contrast to the analytical method that gives one set of values as *the* solution.

Search Methods: "Homing In"

The following table lists some of the more recent search methods that have been described.

Steepest ascent (Box & Wilson)[7] 1951
Gradient search (Zelnik et al.)[33] 1962
Direct search (Hooke and Jeeves)[16] 1961
Gradient-free search (Wilde)[31] 1962
  (a) Lattice methods (b) Contour tangent
Gradient method (Roberts and Lyvers)[25] 1961
  (a) Hemstitching (b) Riding the constraint
Ridge analysis (Lester)[20] 1961
  (Hoerl)[15] 1959
Parallel tangents (PARTAN) (Shah et al.)[26] 1961
  (Buehler et al.)[10] . 1961
Nonlinear digital optimizing program (Mugele)[21] . 1962
Conjugate gradient (Hestenes and Stiefel)[14] 1952
Gradient projection (Rosen)[23a] 1960, 1961





▲ Plots of yield vs. temperature and pressure show maximum yield—Fig. 2

▲ Constraints (minimum temperature, maximum pressure) limit results—Fig. 3

Search methods have certain characteristics in common. A base point is assumed known (one set of conditions that is a solution to the problem, but not necessarily the optimum solution). Then, the method must select the next set of values for the variables and evaluate the objective function once again, hoping that this time, and each successive time, the solution will be closer to the optimum solution.

Each search method has this objective in common but each selects the next set of values in a different manner. A judicious choice can eliminate many wasteful calculations in "homing in" on the optimum.

One requirement in all of these methods is that the dependent variable be unimodal (one peak). However, various techniques are available for scanning for alternative optima.

Basic Concepts of Optimization

There are some basic concepts of maxima, minima, tangents and normals that are important for an understanding of optimization. See Fig. 1 for terminology.

A basic knowledge of the differential calculus is assumed. If the reader is not confident in handling readily differentiable functions, a number of references are included.[9, 19, 23]

Maxima and Minima

The slope of a curve is given by a first derivative. Thus,

$$y = 2x + 1$$

is an equation of a straight line whose slope may be found by taking the first derivative of y with respect to x (see Fig. 4),

$$dy/dx = 2$$

Now, consider the following equation of a parabola (see Fig. 5),

$$y = x^2 + 2$$

The slope at any point on this curve is

$$dy/dx = 2x$$

This slope is continually changing since it is a function of x. For a straight line, the slope remains constant regardless of the value of x.

Change in slope may be represented by the second derivative of y with respect to x.

For the straight line,

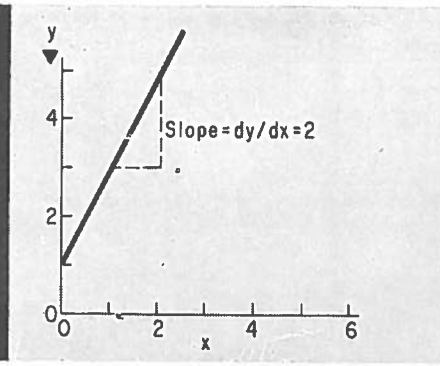
$d^2y/dx^2 = 0$ (slope is constant)

For the parabola,

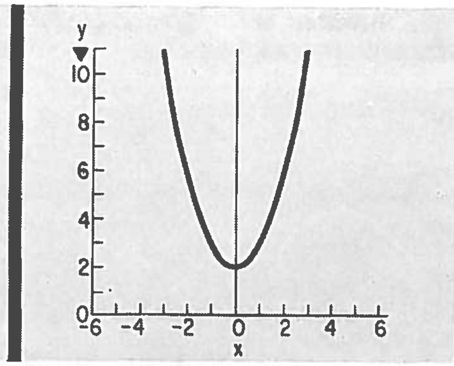
$d^2y/dx^2 = 2$ (slope is changing by 2 units per unit change in x)

These are all differentiable functions. With a discontinuity or kink, the function is not differentiable (see Figs. 6 and 7). In these figures, x_d represents the value of x at the point where the derivatives are nonexistent.

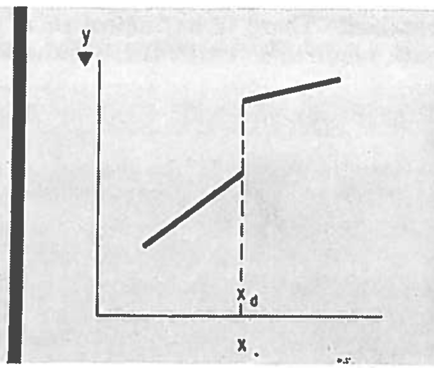
A distinction should be made between local and



Slope of a curve given by the first derivative—Fig. 4



Slope of a parabola is continually changing — Fig. 5



Derivative is not defined for discontinuous function—Fig. 6

global extreme conditions. A function may have an extreme value—a maximum at one value of x—and yet this may not be the true maximum value of the function of x. Fig. 8 indicates such a situation where a local maximum exists at $x = x_1$ but the true maximum (global) in the range of x_a to x_b occurs at $x = x_a$. One method sometimes used in searching for a global extreme is to assume a new starting point and search for a peak once again. If the same peak is reached time and time again, the investigator can have confidence that this is the true peak, although he has no assurance that his objective has been achieved.

Thus, a local maximum (or minimum) will be defined as a point having lower (or higher) adjacent values on either side. A global maximum (or minimum) will be defined as the highest (or lowest) point in the area under consideration.

Our discussion to follow will be limited to *local* extremes.

Functions of One Variable

Rules about extremes of a function of one variable may be summarized as follows:

Rule 1—Extremes of $f(x)$ can occur only where $dy/dx = 0$ or where dy/dx is nonexistent. This rule is a necessary but not sufficient condition, since other conditions must be met.

Rule 2—If, at the point determined from Rule 1, certain derivatives vanish, then the next derivative, which does not vanish, is examined for sign. Say that all the derivatives up to the nth derivative vanish

$$dy/dx = d^2y/dx^2 = d^3y/dx^3 \ldots d^ny/dx^n = 0$$

then the next derivative $d^{n+1}y/dx^{n+1}$ is either positive or negative. If n is an even value, there is a point of inflection. If n is odd, the next derivative (the $n + 1$st) is examined; if it is negative, a maximum exists; if it is positive, a minimum exists.

Rule 3—This rule must be used where dy/dx does not exist (e.g., at a discontinuity). The neighborhood of the critical point (point determined from Rule 1) must be explored. The first derivative, dy/dx, is investigated in this area as x increases in value through the critical point. Sign of dy/dx must be noted:

If the sign goes from plus to minus, a maximum exists; if sign goes from minus to plus, a minimum exists; if sign does not change, there is no extreme.

To illustrate, find the extremes of:

$$y = x^3 - 9x^2 + 24x$$

The first derivative is set equal to zero

$$dy/dx = 3x^2 - 18x + 24 = 0$$

Solving the quadratic equation for x,

$$(x - 4)(x - 2) = 0$$

from which

$$x = 4$$
$$x = 2$$

therefore, extremes may exist at these values of x.

From Rule 2,

$$d^ny/dx^n = 0$$

so that $n = 1$ in this problem

$$d^{n+1}y/dx^{n+1} = d^2y/dx^2$$

The second derivative must now be investigated for sign:

$$d^2y/dx^2 = 6x - 18$$

At $x = 4$

$$d^2y/dx^2 = 6(4) - 18 = 6$$

At $x = 2$

$$d^2y/dx^2 = 6(2) - 18 = -6$$

Following Rule 2, $n = 1$ and is, therefore, odd. The second derivative is positive for $x = 4$, thus a minimum exists at this point; the second derivative is negative for $x = 2$, so a maximum exists at this point.

Fig. 9 illustrates these results graphically. Note that maximum and minimum points are local extremes and not global.

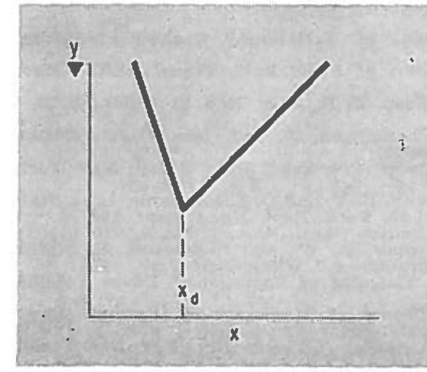
As another example, find the extreme of:

$$y = 5 + x^{2/3}$$

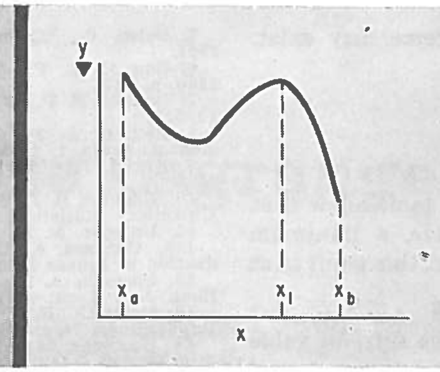
Here, the first derivative is

$$dy/dx = (2/3)x^{-1/3}$$

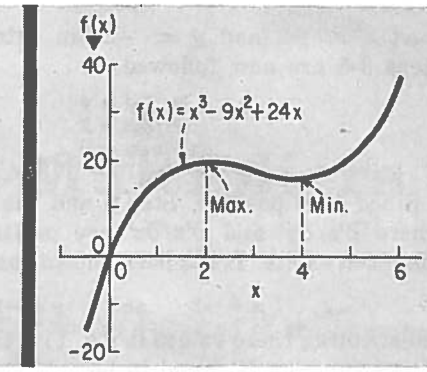
This derivative does not exist at $x = 0$. Using Rule



Derivative does not exist at a kink—Fig. 7

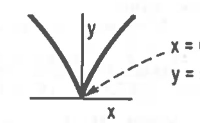


Local and global maxima exist for functions—Fig. 8



Maximum and minimum are local extremes—Fig. 9

3, as x increases through $x = 0$, dy/dx goes from minus to plus. There is a minimum at $x = 0$, $y = 5$.



Functions of Two Variables

Suppose that z is a function of two variables, x and y, and we want to search for a maximum or minimum. The first question to be answered is: Does an extreme exist at all? Without delving into a mathematical proof, the steps will be stated below. Proofs are available in many references.[1, 9, 13, 19, 23, 27]

Step 1—Evaluate the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$.

Step 2—If $\partial z/\partial x = 0$ and $\partial z/\partial y = 0$, then an extreme may exist* and Steps 3 through 5 are to be followed. If either $\partial z/\partial x$ or $\partial z/\partial y$ is not zero, then there is neither a maximum nor a minimum.

Step 3—Evaluate $\partial^2 z/\partial x^2$, $\partial^2 z/\partial y^2$ and $\partial^2 z/(\partial x\, \partial y)$.

Step 4—Evaluate the term M where

$$M = (\partial^2 z/\partial x^2)(\partial^2 z/\partial y^2) - (\partial^2 z/\partial x\, \partial y)^2$$

Step 5—See the table (*right*) to determine whether an extreme condition exists. This constitutes the sufficient condition for an extreme.

* This is a necessary condition for an extreme condition.

For the case of more than two independent variables, the problem becomes much more complex and will not be discussed at this point. This case has been covered by many authors.[1, 13, 23]

M		Optimum
Positive	$\partial^2 z/\partial y^2$ and $\partial^2 z/\partial x^2$ are positive....	Minimum
	$\partial^2 z/\partial y^2$ and $\partial^2 z/\partial x^2$ are negative...	Maximum
Negative	....	Saddle point
Zero	....	Undecided

Given the equation

$$z = x^2 + y^2 + 4x + 4y \qquad (1)$$

Determine: (a) whether an extreme exists (b) if so, the extreme value of this function and the values of x and y at the extreme condition.

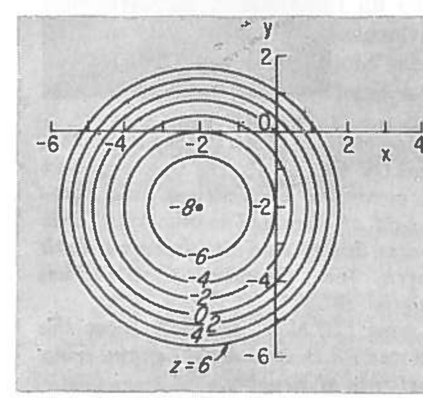
Following the steps in the recommended procedure:

$$\partial z/\partial x = 2x + 4 \qquad (2)$$
$$\partial z/\partial y = 2y + 4 \qquad (3)$$

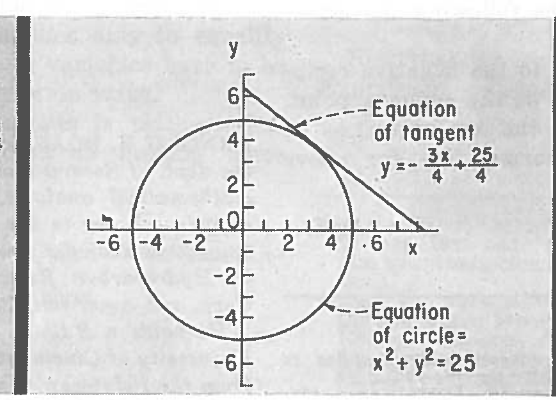
For an extreme condition to exist, Eqs. (2) and (3) are set equal to zero

$$2x + 4 = 0, x = -2$$

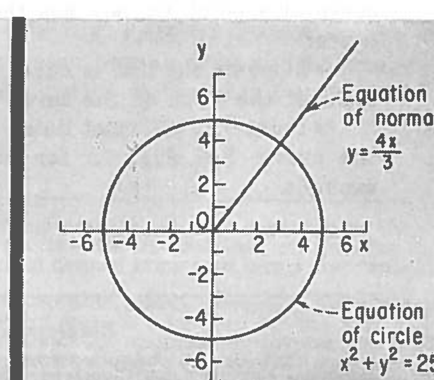
and $\quad 2y + 4 = 0, y = -2$, Now Step 2:



Optimum point shows up in plot of equation—Fig. 10



Tangent calculations are important in optimization—Fig. 11



Normal-line calculations can be graphed—Fig. 12

At $x = -2$ and $y = -2$, an extreme may exist. Steps 3-5 are now followed:

$$3.\quad \partial^2 z/\partial x^2 = 2 \qquad \partial^2 z/\partial y^2 = 2 \qquad \partial^2 z/\partial x\,\partial y = 0$$
$$4.\quad M = (2)(2) - 0 = 4$$

Since M is positive, Step 5 and the table show that where $\partial^2 z/\partial y^2$ and $\partial^2 z/\partial x^2$ are positive, a minimum condition exists. It has been shown that this point is at

$$x = -2 \quad \text{and} \quad y = -2$$

Substituting these values in Eq. (1), the extreme value of the function is found to be

$$z = x^2 + y^2 + 4x + 4y$$
$$= (-2)^2 + (-2)^2 + 4(-2) + 4(-2) = -8$$

Fig. 10 depicts Eq. (1) graphically for various values of z from 6 to -8.

Tangents and Normals to Curves

Tangent lines may be used to approximate a curve over a given interval. This is known as "linearization" and can be applied to reduce nonlinear problems to linear.

The equation of a tangent line to a curve at a point (x_1, y_1) has the following characteristics:

- Slope of the tangent line is identical to that of the curve at the common point (x_1, y_1).
- The point (x_1, y_1) must lie on the line as well as on the curve.

As an example, find the equation of the tangent line to the circle $x^2 + y^2 = 25$ at the point $x = 3$, $y = 4$. Differentiating the equation produces

$$dy/dx = -x/y$$

This is the slope at any point on the circle. Now, at the point, $x = 3$ and $y = 4$

$$dy/dx = -3/4$$

Therefore, the slope, m, of the tangent line at the point is $-3/4$. Since the slope and one point on the line are known, the equation of the line may be calculated from

$$(y - y_1) = m(x - x_1)$$
$$y = -(3x/4) + 25/4 \text{. See Fig. 11.}$$

Equation of a line normal (perpendicular) to a curve at a point (x_1, y_1) has the following characteristics:

- Slope of the line is equal to the negative reciprocal of the slope of the curve at the common point.
- Point (x_1, y_1) must lie on the line as well as on the curve. See Fig. 12 for normal line for above example.

References

1. AIChE Special Lecture Series, "Applied Mathematics in Chemical Engineering," Cleveland, Ohio, May 6, 1961.
2. Baumol, W. J., "Economic Theory and Operations Analysis," Prentice-Hall, Englewood Cliffs, N. J., 1961.
3. Bellman, R., "Dynamic Programming," Princeton University Press, Princeton, N. J., 1957.
4. Bliss, G. A., "Calculus of Variations," University of Chicago Press, Chicago, 1925.
5. Bolza, O., "Calculus of Variations," Chelsea Publishing, 1961.
6. Box, G. E. P., *Jour. of Basic Eng., Trans. ASME*, Mar., 1960, p. 113.
7. Box, G. E. P., Wilson, K. B., *Jour. Roy. Stat. Soc.*, B **13**, 1 (1951).
8. Box, G. E. P., Chanmugan, J., *Ind. Eng. Chem. Fundamentals* **1**, No. 1, 2 (1962).
9. Brand, L., "Advanced Calculus," John Wiley, New York, 1958.
10. Buehler, R. J., Shah, B. V. and O. Kempthorne, Iowa State University Statistical Lab. Tech. Rept. No. 3, Sept. 1961.
11. Butcher, K. L., *British Chem. Eng.*, **7**, No. 7, 510 (1962).

11a. Charnes, A., Cooper, W. W., and Henderson, A., "Introduction to Linear Programming," Wiley, New York, 1953.

12. Forsyth, A. R., "Calculus of Variations," Dover Publications, New York, 1961.
13. Hancock, H., "Theory of Maxima and Minima," Dover Publications, New York, 1961.
14. Hestenes, M. R., Stiefel, E., *Jour. Res. Nat. Bur. Standards* **49**, 409 (1952).
15. Hoerl, A. E., *Chem. Eng. Prog.* **55**, No. 11, 69 (1959).
16. Hooke, R., Jeeves, T. A., *Jour. A. C. M.* **8**, No. 2, 212 (1961).
17. Kimball, W. S., "Calculus of Variations," Butterworth Scientific Publications, London, 1952.
18. Koehler, T., *Chem. Eng. Prog.*, Oct. 1959, p. 76.
19. Korn, G. A., Korn, T. M., "Mathematical Handbook For Scientists and Engineers," McGraw-Hill, New York, 1961.
20. Lester, J. T., *British Chem. Eng.* **6**, No. 12, 851 (1961).
21. Mugele, R. A., "A Nonlinear Digital Optimizing Program for Process Control Systems," Proc. Western Joint Computer Conference, Spring 1962.
22. "Optimization Techniques in Chemical Engineering," N.Y.U. Symposium, May, 1960.
23. Osgood, W. F., "Advanced Calculus," Macmillan, New York, 1929.
24. Rozonoer, L. I., "L. S. Pontryagin's Maximum Principle in the Theory of Optimal Systems," *Automation and Remote Control*, **20**, Oct., Nov., Dec. 1959, pp. 1288-1302, 1405-1421, 1517-1532.
25. Roberts, S. M., Lyvers, H. I., *Ind. Eng. Chem.* **53**, 877 (1961).

25a. Rosen, J. B., *Jour. Soc. Ind. Appl. Math.*, **8**, No. 1,181 (1960) and **9**, No. 4,514 (1961).

26. Shah, B. V., Buehler, R. J. and O. Kempthorne, Iowa State Univ. Statistical Lab. Tech. Rept. No. 2, Apr. 1961, revised Aug. 1962.
27. Sokolnikoff, I. S., Redheffer, R. M., "Mathematics of Physics and Modern Engineering," McGraw-Hill, New York, 1958.
28. Taborek, J. J., *British Chem. Eng.* **5**, No. 11 784 (1960).
29. Weinstock, R., "Calculus of Variations," McGraw-Hill, New York, 1948.
30. Westbrook, G. T., *Hydrocarbon Proc. & Pet. Refiner* **40**, No. 9, 201 (1961).
31. Wilde, D. J., "Gradient-Free Search For an Optimum," presented at the 46th National Meeting of AIChE, Los Angeles, Calif., Feb. 5, 1962.
32. Wilde, D. J., "Advances in Chemical Engineering," Vol. III, "Optimization Methods," p 273-331, Academic Press, 1962.
33. Zellnik, H. E., Sondak, N. E. and Davis, R. S., *Chem. Eng. Prog.* **58**, No. 8, 35 (1962).

Meet the Author

ARNOLD H. BOAS *is associate engineer in the engineering dept. of Socony Mobil Oil Co., N. Y. He is engaged in mathematical analysis, computer applications and optimization theory in the field of chemical engineering. His experience includes process design and development work at Hydrocarbon Research, Inc., Chemical Construction Corp. and American Cyanamid Co.*

He holds a B.Ch.E. from C.C.N.Y., an M.S. from the University of Cincinnati and Ph.D. in chemical engineering from the Polytechnic Institute of Brooklyn.

Dr. Boas is a member of Tau Beta Pi, Sigma Xi, Phi Lambda Upsilon, AIChE, and is a PE in N. Y. State.

Part 2

How to Use Lagrange Multipliers

A very powerful and popular technique—use of Lagrange Multipliers—permits solving for optimum conditions when these conditions are subject to design, operation, cost, or other restrictions.

ARNOLD H. BOAS, *Socony Mobil Oil Co., Inc.**

We pointed out in Part 1 of this series that the analytical method of optimization involves setting the derivatives of functions equal to zero. One extremely powerful analytical technique—use of Lagrange Multipliers—is applied when there are equality constraints or restrictions on the variables (e.g., purity must equal a certain value, flow rate must equal a design value).

The object of optimization studies is to determine values of the independent variables that will maximize (or minimize) some objective function (e.g., yield as a function of temperature and pressure). A Lagrange Expression is developed first. Then, values of the independent variables that optimize this expression are determined, subject to the given constraints of the problem. The objective function will also be optimized by these same values of the independent variables. Very often, it is easier to optimize the Lagrange Expression than the objective function itself, considering the constraints involved.

There are two important rules in applying this technique: (1) the number of Lagrange Multipliers to be introduced must be equal to the number of constraining equations, (2) the Lagrange Expression must be equal to the objective function plus† the product of the Lagrange Multiplier and constraint. This constraint must be in the form of an equation set equal to zero. The discussion that follows applies *only* to equality constraints, *not* situations where variables have to be less than (or greater than) a certain value.

Let's now express these concepts in mathematical nomenclature and then proceed to develop some Lagrange Expressions:

$u(x,y,\ldots)$ = objective function.
$v_1(x,y,\ldots) = 0$ and $v_2(x,y,\ldots) = 0$, etc. are the constraining equations.
$\lambda_1, \lambda_2, \ldots$ etc. are the Lagrange Multipliers.
$w(x,y,\ldots)$ = Lagrange Expression.
$x, y, \ldots$ are independent variables.

* To meet your author, see *Chem. Eng*, Dec. 10, 1962, p. 152.
† Some authors use a minus sign here.

Then, according to the definition of the Lagrange Expression

$$w(x,y,\ldots) = u(x,y,\ldots) + \lambda_1 v_1(x,y,\ldots) + \lambda_2 v_2(x,y,\ldots) \quad (1)$$

Problem 1

Form the Lagrange Expression for the problem of finding values of x and y that lie on a circle of radius 5 and that maximize the function xy.

Solution—The objective function is xy, so

$$u(x,y) = xy$$

The constraining equation is the equation of a circle whose radius is 5, or

$$x^2 + y^2 = 25$$
$$v(x,y) = x^2 + y^2 - 25$$

The Lagrange Multiplier is λ. There is only one multiplier because there is only one constraining equation.

The Lagrange Expression becomes, according to Eq. (1):

$$w(x,y) = xy + \lambda(x^2 + y^2 - 25)$$

Problem 2

Form the Lagrange Expression for finding values of x, y and z that minimize the function $x + 2y^2 + z^2$, subject to the constraint that $x + y + z = 1$.

Solution—This problem involves three independent variables x, y and z. The objective function is $x + 2y^2 + z^2$,

$$u(x,y,z) = x + 2y^2 + z^2$$

Constraining equation is $x + y + z = 1$, so

$$v(x,y,z) = x + y + z - 1$$

Lagrange Multiplier is λ and the Lagrange Expression is

$$w(x,y,z) = x + 2y^2 + z^2 + \lambda(x + y + z - 1)$$

Now that we have seen how to form the Lagrange Expression, we shall proceed to the next step. It can

Part 1—"What Optimization Is All About," appeared in the Dec. 10, 1962 issue, pp. 147-152. It contained an introduction to optimization and defined important terms and functions.
Watch for Part 3—Univariable search methods, including the important Fibonacci scan, is scheduled for the Feb. 4, 1963 issue. Search methods for optimization involve selecting a set of variables and testing to see whether they give an answer nearer the optimum. Based on this answer, another set of variables is chosen and so on.

be shown that the partial derivatives of the Lagrange Expression with respect to each independent variable (including the Lagrange Multipliers) must be equal to zero for an extreme to exist. Thus N simultaneous equations are obtained, one for each of the N independent variables, and it is possible to solve for maximum or minimum conditions. Since each constraining equation introduces one additional equation, one Lagrange Multiplier for each equation is introduced to compensate for this. Some examples will now be considered.

Problem 3

Solve Problem 1 for maximum conditions.

Solution—The Lagrange Expression is:

$$w(x,y) = xy + \lambda(x^2 + y^2 - 25)$$

Taking partial derivatives with respect to each independent variable and setting the resulting equations equal to zero,

$$\partial w/\partial x = y + 2\lambda x = 0 \qquad (2)$$

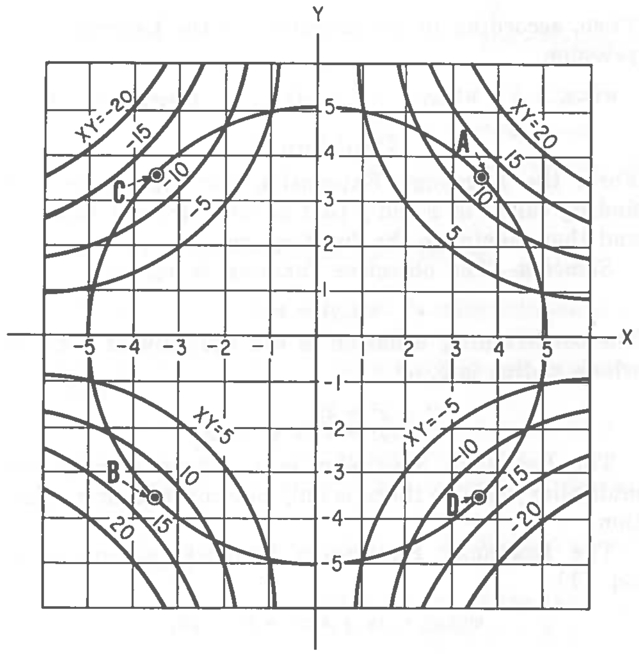
$$\partial w/\partial y = x + 2\lambda y = 0 \qquad (3)$$

$$\partial w/\partial \lambda = x^2 + y^2 - 25 = 0 \qquad (4)$$

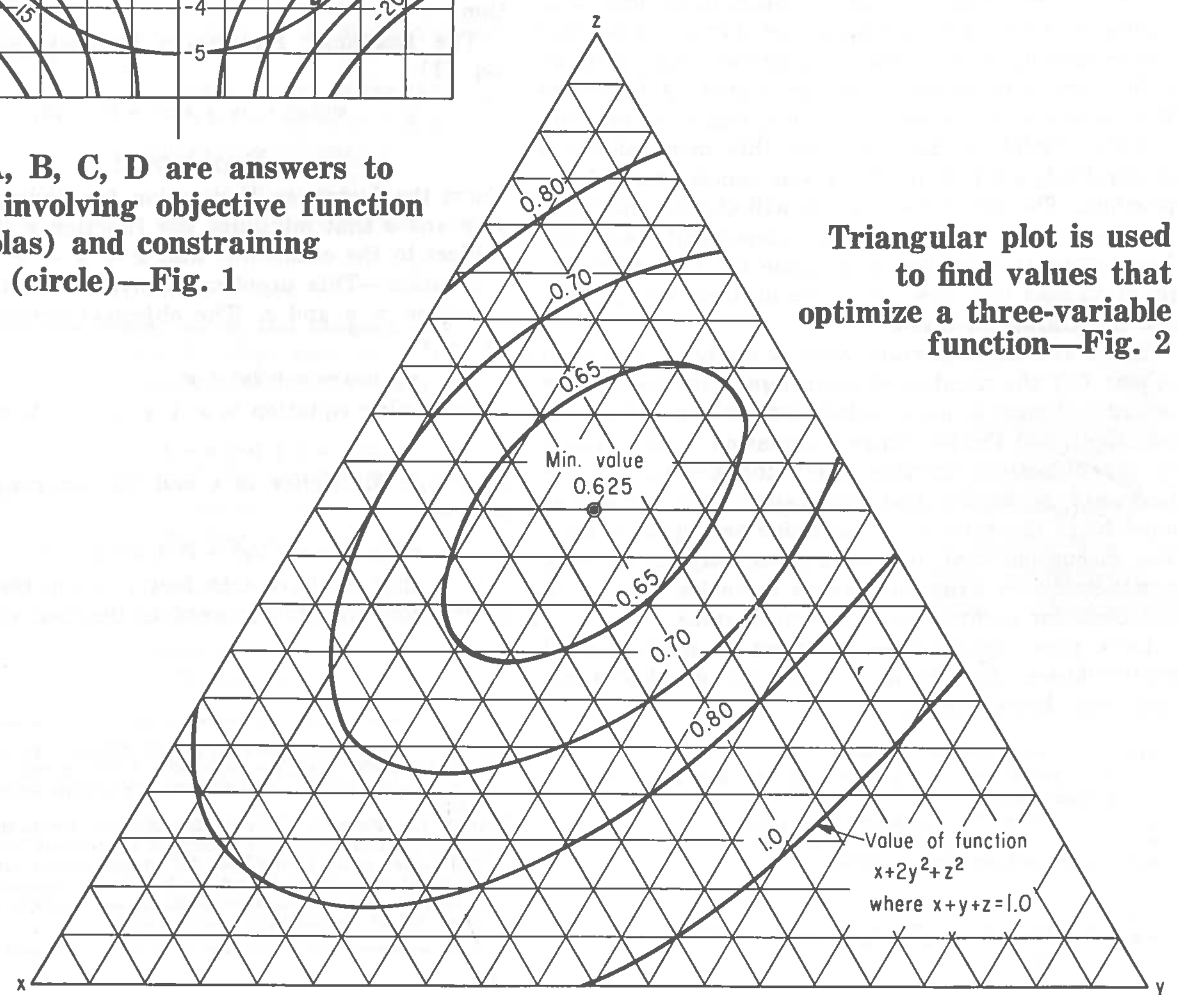
Solving these three equations simultaneously,

$\lambda = 0.50$	$\lambda = -0.50$
$x = \pm 3.54$	$x = \pm 3.54$
$y = \mp 3.54$	$y = \pm 3.54$

Let's designate these four solutions by the following points:



Points A, B, C, D are answers to problem involving objective function (hyperbolas) and constraining equation (circle)—Fig. 1



Triangular plot is used to find values that optimize a three-variable function—Fig. 2

Point A	$x = +3.54$	$y = +3.54$
Point B	$x = -3.54$	$x = -3.54$
Point C	$x = -3.54$	$y = +3.54$
Point D	$x = +3.54$	$y = -3.54$

Points A and B are maximum points, C and D are minimum points. The value of the function at the extreme conditions is ± 12.50. Fig. 1 depicts a plot of the circle (constraining condition), the hyperbolas (objective function) and solutions to the problem, points A, B, C and D.

Problem 4

Solve Problem 2 for minimum conditions.

Solution—The Lagrange Expression has been found to be

$$w(x,y,z) = x + 2y^2 + z^2 + \lambda(x + y + z - 1)$$

Taking partial derivatives with respect to each independent variable and setting the resulting equations equal to zero,

$$\partial w/\partial x = 1 + \lambda = 0 \qquad (5)$$

$$\partial w/\partial y = 4y + \lambda = 0 \qquad (6)$$

$$\partial w/\partial z = 2z + \lambda = 0 \qquad (7)$$

$$\partial w/\partial \lambda = x + y + z - 1 = 0 \qquad (8)$$

Solving these four equations simultaneously,

$$\lambda = -1$$
$$x = 0.25$$
$$y = 0.25$$
$$z = 0.50$$

These are the values of x, y and z whose sum is 1.0 and that minimize the function $x + 2y^2 + z^2$. Value of the function at the extreme condition is

$$x + 2y^2 + z^2 = (0.25) + 2(0.25)^2 + (0.50)^2 = 0.625$$

A triangular plot has been used to represent values of the objective function. Values of 0.65, 0.70, 0.80 and 1.0 are shown in Fig. 2. A point on the plot necessarily satisfies the constraint condition that $x + y + z = 1$. And the minimum point at $x = 0.25$, $y = 0.25$ and $z = 0.50$ is indicated.

Problem 5

Let us assume that the yield, Y, of a chemical reaction is related to temperature, T, and pressure, P, by the following second-order response equation:

$$Y = a_1 + b_1T + c_1P + d_1T^2 + e_1P^2 + f_1TP \qquad (9)$$

where a_1, b_1, c_1, etc., are constants that have been previously determined. Let us, also, assume that the purity, Q, is related to T and P by a similar equation:

$$Q = a_2 + b_2T + c_2P + d_2T^2 + e_2P^2 + f_2TP \qquad (10)$$

If we imagine that Q is fixed at some value, then Eq. (10) represents a constraint. The problem is to find the values of T and P that maximize the yield, Y, subject to the constraining equation.

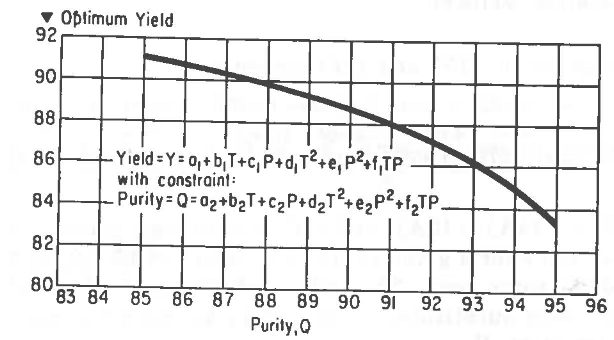
Solution—The objective function is

$$u(T,P) = Y = a_1 + b_1T + c_1T^2 + d_1T^2 + e_1P^2 + f_1TP \qquad (11)$$

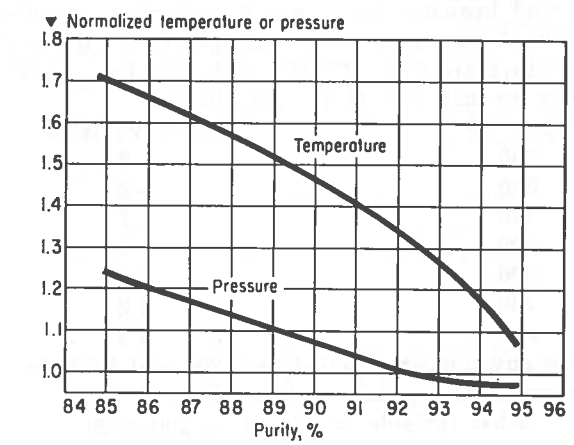
Constraining equation is

$$v(T,P) = a_2 + b_2T + c_2P + d_2T^2 + e_2P^2 + f_2TP - Q \qquad (12)$$

Lagrange Multiplier is λ and the Lagrange Expression becomes



Optimum yield is subject to product-purity restrictions—Fig. 3



Temperature and pressure are a function of optimum yield—Fig. 4

$$w(T,P) = a_1 + b_1T + c_1P + d_1T^2 + e_1P^2 + f_1TP + \lambda(a_2 + b_2T + c_2P + d_2T^2 + e_2P^2 + f_2TP - Q) \qquad (13)$$

Taking partial derivatives with respect to T, P and λ, and setting the resulting equations equal to zero,

$$\partial w/\partial T = b_1 + 2d_1T + f_1P + \lambda(b_2 + 2d_2T + f_2P) = 0 \qquad (14)$$

$$\partial w/\partial P = c_1 + 2e_1P + f_1T + \lambda(c_2 + 2e_2P + f_2T) = 0 \qquad (15)$$

$$\partial w/\partial \lambda = a_2 + b_2T + c_2P + d_2T^2 + e_2P^2 + f_2TP - Q = 0 \qquad (16)$$

Eqs. (14), (15) and (16) are nonlinear simultaneous equations. Standard numerical techniques can be used to solve these equations.[1, 4]

The numerical example of Umland and Smith[5] will now be used to illustrate the method.

Eqs. (9) and (10) are given as:

$$Y = 55.84 + 26.65T + 7.31P - 6.96T^2 - 3.03P^2 + 2.69TP \qquad (9A)$$

$$Q = 85.72 + 8.59T + 21.85P - 5.18T^2 - 9.20P^2 - 6.26TP \qquad (10A)$$

Values of the constants are thus:

$a_1 = 55.84$	$a_2 = 85.72$
$b_1 = 26.65$	$b_2 = 8.59$
$c_1 = 7.31$	$c_2 = 21.85$
$d_1 = -6.96$	$d_2 = -5.18$
$e_1 = -3.03$	$e_2 = -9.20$
$f_1 = 2.69$	$f_2 = -6.26$

Eqs. (14), (15) and (16) become:

$$26.65-13.92T+2.69P+\lambda(8.59-10.36T-6.26P)=0 \quad (14A)$$
$$7.31-6.06P+2.69T+\lambda(21.85-18.40P-6.26T)=0 \quad (15A)$$
$$85.72+8.59T+21.85P-5.18T^2-9.20P^2-6.26TP=Q \quad (16A)$$

Eqs. (14A), (15A), and (16A) have been solved for T, P and λ for a given Q. Here Q values of 85, 90, 92.5 and 95 were used. The values of T and P obtained were then substituted in Eq. (9A) to solve for maximum yield, Y.

Fig. 3 shows the optimum yield vs. specified purity, Q. For example, with a constraint of $Q = 92.5$, the optimum yield is 86.7. And Fig. 4 indicates that for this case $P = 1.00$, $T = 1.32$. The variables of temperature and pressure have been transformed or coded for ease of calculation. For example, if the levels of temperature are 200, 300, 400, 500, 600, 700 and 800 F., then we may code these variables as:

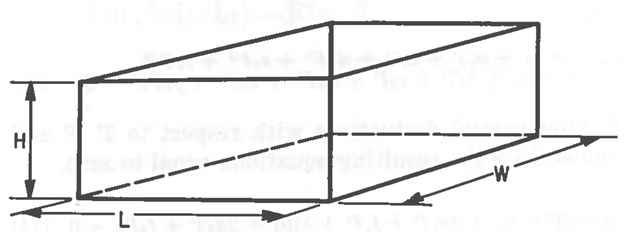
Temp. °F.	Coded Variable
200	−3
300	−2
400	−1
500	0
600	+1
700	+2
800	+3

Then any temperature between 200 and 800F. may be expressed as a coded variable. Thus:

coded variable = (Temp. − 500)/100

Problem 6

Find the dimensions of an open rectangular tank of 1,000-cu. ft. capacity to give the minimum area.



Solution—Consider the general solution to the problem. Call the volume, V

$$V = LWH$$

The objective function is the area to be minimized and it is equal to

$$u(L,W,H) = 2HW + LW + 2HL$$

And the constraining equation is

$$v(L,W,H) = LWH - V$$

The Lagrange Multiplier is λ and the Lagrange Expression is

$$w(L,W,H) = 2HW + LW + 2HL + \lambda(LWH - V)$$

Taking partial derivatives with respect to L, W, H and λ, and setting the resulting equations equal to zero,

$$\partial w/\partial L = W + 2H + \lambda WH = 0 \quad (17)$$
$$\partial w/\partial W = 2H + L + \lambda LH = 0 \quad (18)$$
$$\partial w/\partial H = 2W + 2L + \lambda LW = 0 \quad (19)$$
$$\partial w/\partial \lambda = LWH - V = 0 \quad (20)$$

From Eq. (17),

$$\lambda = \frac{-(W + 2H)}{WH} \quad (21)$$

Substituting in Eq. (18), and solving for W,

$$W = L \quad (22)$$

From Eq. (19), it is found that

$$\lambda = -4/W \quad (23)$$

From Eq. (17),

$$W = 2H \quad (24)$$

therefore,

Volume $= V = LWH = (W)(W)(W/2) = W^3/2 = 1{,}000$ cu. ft

from which

$W = 12.6$ ft.
$L = 12.6$ ft.
$H = 6.3$ ft.

The only constraints considered in this discussion were equality constraints. Very often, inequality constraints are presented where the independent variables must be less than (or greater than) a certain value. This problem is much more complex and will not be discussed in an introductory paper such as this.

The reader is referred to Dorn[2], who considers the required conditions for extrema of nonlinear functions, subject to linear constraints (equalities and inequalities), and Kuhn and Tucker[3] who have generalized the concept of Lagrange Multipliers to include inequality constraints.

References

1. Booth, A. D., "Numerical Methods," Butterworths Scientific Publications, London, England, 1955.
2. Dorn, W. S., *Operations Research* 9, 95 (1961).
3. Kuhn, H. W., Tucker, A. W., "Nonlinear Programming," Proc. 2nd Berkeley Symposium on Math Stat. and Prob., pp. 481-492, U. of Calif. Press, Berkeley, 1951.
4. Scarborough, J. B., "Numerical Mathematical Analysis," 4th Ed., John Hopkins Press, Baltimore, Md., 1958.
5. Umland, A. W., Smith, W. N., *Technometrics* 1, No. 3,289 (1959).

Key Concepts for This Article

For indexing details, see Chem. Eng., Jan. 7, 1963, p. 73 (Reprint No. 222). Words in bold are role indicators; numbers correspond to AIChE system.

Active Concept (8)	Passive Concept (9)	Means/Methods (10)
Optimization	Objective functions	Lagrange's method
Solving	Equations, constraining	Calculus
	Models, mathematical	Mathematics

Key Concepts for Indexing Part 1 of This Series

Part 1 appeared in the Dec. 10, 1962, issue, pp. 147-152

Active Concept (8)	Passive Concept (9)	Means/Methods (10)
Optimization	Objective functions	Analytical methods
Solving	Equations, constraining	Definitions
	Models, mathematical	Case-study methods
		Search methods
		Linearizing
		Calculus
		Mathematics

Part 3

How Search Methods Locate Optimum In Univariable Problems

Search or "hill-climbing" techniques for locating an optimum are useful when no mathematical model is available and data must be collected from experiments. These techniques can eliminate wasteful experimentation and calculations.

ARNOLD H. BOAS, *Socony Mobil Oil Co., Inc.**

In Part 2 of this series, we discussed an optimization technique that is useful when the relationships between variables can be expressed in some mathematical form. But in many cases such mathematical expressions are not available. To obtain a value of the objective function (i.e., the expression to be optimized), a "run" has to be made. This might involve an experimental determination, a computer calculation, an iterative procedure using graphs and tables, etc. Search methods are very useful for handling this type of problem.

There are numerous search techniques reported in the literature (see Part 1, p. 149). All, however, are based on the same principle: a base point is known (one solution to the problem but not necessarily the optimum solution). Then, each search technique selects the next set of values for the variables and tests to see whether the set gives an answer nearer the optimum. Based on this answer, another group of variables are chosen. Judicious choice of these variables can eliminate many wasteful experiments or calculations.

Basically, at the start of a search procedure we know there is a maximum or minimum but we do not know the value of, say, x at the optimum. Suppose only a limited number of experiments can be run to narrow the range in which we know the optimum lies. How should we go about selecting variables for these experiments so that we "home-in" on the optimum most efficiently?

Many practical problems of this type involve more than one independent variable. But to simplify descriptions of some of the important search methods, we are going to assume only one independent variable and negligible random experimental error. This will lead into Part 4 of this series, in which the multivariable problem will be discussed.

Also, the reasonable assumption of unimodality (one peak) will be made: the value of the dependent variable decreases (increases) as the value of the independent variable changes in either direction from the maximum (minimum).

Some of the more common search procedures have been described by Wilde[8] and will be reviewed here.

Part 1—"What Optimization Is All About," appeared in the Dec. 10, 1962, issue, pp. 147-152. It contained an introduction to optimization and defined terms and functions.

Part 2—"How to Use Lagrange Multipliers," appeared in the Jan. 7, 1963, issue, pp. 95-98. It described a useful technique for optimization of mathematical expressions.

Watch for Part 4—Search methods for handling the more complicated problem—multivariable situations—will be described in the next installment.

Principle of Minimax

A useful criterion for an efficient search procedure is the width of the interval within which the optimum point *must* lie, based on the assumptions of the problem. The conservative approach is to assume that "whatever *can* go wrong *will* go wrong in the search procedure." With this disadvantage, we are still in a position to make some conclusive statement about the optimum. We shall consider each search procedure in light of this and obtain the maximum value for the width of the interval. Then, the "best procedure" will be taken to mean that procedure which gives the minimum value of this maximum width; hence, the name minimax.

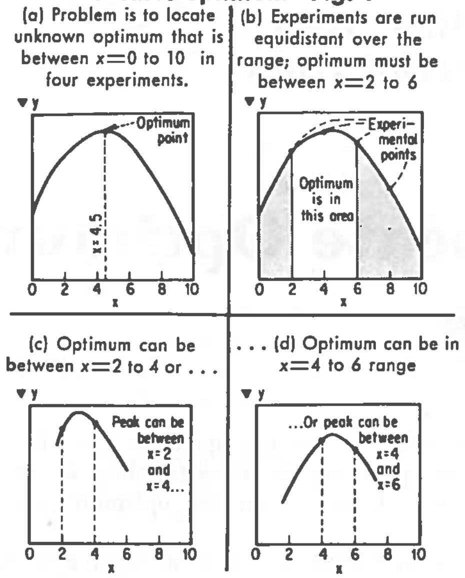
Optimization Problem: Uniform Search

Let's assume that we want to optimize yield, y, in a reactor and that only four experiments are allowed due to certain plant conditions. Also, the unimodal function can be represented as shown in Fig. 1a, where the peak is at 4.5. This maximum is what we are trying to find.

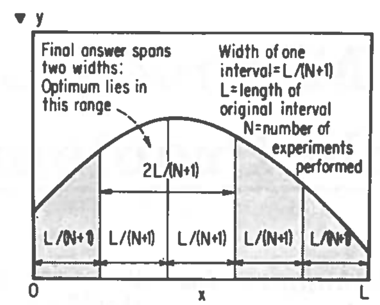
The most obvious way to start is to place the four experiments equidistant over the interval, i.e., at 2, 4, 6 and 8. From Fig. 1b, we see that the value of y at $x = 4$ is higher than the value of y at $x = 2$. Since, we are dealing with a unimodal function, the optimum

* To meet your author, see *Chem. Eng.*, Dec. 10, 1962, p. 152.

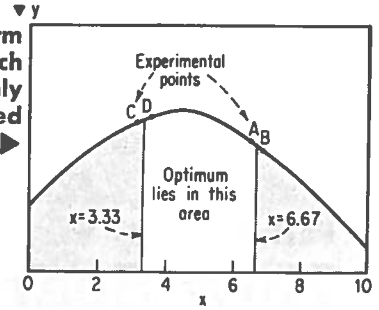
How uniform-search method locates optimum—Fig. 1



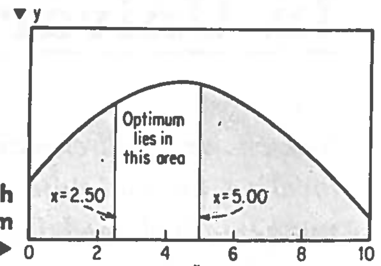
Uniform search divides range of x into equal intervals—Fig. 2



Uniform dichotomous search requires evenly spaced paired experiments—Fig 3



Sequential dichotomous search uses information from previous experiments—Fig. 4



value of y cannot possibly lie between $x = 0$ and $x = 2$. By similar reasoning, the area between $x = 8$ and $x = 10$ can be eliminated as well as $x = 6$ to 8. The area remaining, as seen in Fig. 1b, is the area between $x = 2$ and $x = 6$.

Can anything further be said about the location of the optimum point? Because this is a simple problem, we could draw a curve through the three remaining points and predict the optimum. However, based on the minimax principle, we can state that the optimum point lies between $x = 2$ and $x = 6$. The fact that y at $x = 4$ is higher than y at $x = 6$ presents two possibilities. Figs. 1c and 1d indicate cases where the peak can lie either in the range of $x = 2$ to 4 or $x = 4$ to 6 even though y is higher at $x = 4$ than at $x = 6$.

The procedure just outlined can be described mathematically for the general case of the uniform search. Let:

L = length of the original interval.

F = fraction of original interval within which the optimum lies after performing N experiments.

N = number of experiments performed.

The N experiments divide the entire region into $(N + 1)$ intervals. Width of each interval is $L/(N + 1)$. The optimum can then be specified over the width of two of these intervals, or $2L/(N + 1)$. Therefore,

$$F = \frac{2L}{(N+1)}\left(\frac{1}{L}\right) = \frac{2}{(N+1)}$$

See Fig. 2.

For the specific example illustrated, $N = 4$ and

$$F = 2/5 = 0.40$$

Uniform Dichotomous Search

In the uniform dichotomous search procedure, experiments are performed in pairs. The pairs of experiments are spaced evenly over the entire interval. For the problem under consideration, two experiments* are performed around $x = 6.67$ and two around $x = 3.33$. From Fig. 3, we see that point A is higher than point B and, therefore, the region from $x = 6.67$ to $x = 10$ is eliminated. Point D is higher than point C and the region from $x = 0$ to $x = 3.33$ is also eliminated. Therefore, the optimum lies in the area under the curve bounded by $x = 3.33$ to $x = 6.67$.

Mathematically, this technique can be described as follows: The N experiments divide the region into $(N/2) + 1$ intervals of width $L/[(N/2) + 1]$. The optimum is located over the width of one interval, i.e., $L/[(N/2) + 1]$. Therefore,

$$F = \left(\frac{L}{(N/2)+1}\right)\left(\frac{1}{L}\right) = \frac{2}{N+2}$$

In the specific example considered, $N = 4$ and

$$F = 2/6 = 0.333$$

Sequential Dichotomous Search

Thus far, search methods were considered where all the experiments had to be planned in advance.

A sequential search is one where the investigator takes advantage of the information available from the previous experiments before performing the next one.

A sequential dichotomous search involves running two experiments near the middle of the region so half the area can be eliminated in one fell swoop. In our example, two experiments would be performed around $x = 5$. These would be done on either side of $x = 5$ so that the measurement of y would determine on which side of $x = 5$ the optimum value of y was located. Once half the region has been eliminated (in this example, the region between $x = 5$ and $x = 10$), another pair of experiments is performed near the middle of the remaining region, i.e., $x = 2.5$, so that half of this region can be eliminated. The area remaining will lie between $x = 2.5$ and $x = 5.0$. See Fig. 4.

Thus each pair of experiments bisects the previous interval. At the end of the first pair of experiments, the remaining interval is one-half of the original interval; after the next pair, it is one-quarter and after the $(N/2)$ pair of experiments the remaining interval is $1/(2)^{N/2}$ of the original interval.

F is equal to

$$F = 1/(2)^{N/2}$$

For the case where $N = 4$,

$$F = 1/4 = 0.250$$

Fibonacci Search Technique

A more efficient sequential technique is the Fibonacci search. It has been shown that this is the optimal search routine to follow for the case of one variable and where the assumption of unimodality prevails.[1,2]

Although the name Fibonacci may not be too familiar to some, Fibonacci numbers are far from new. The theory of these numbers goes back to the days of Leonardo of Pisa, also known as Fibonacci.† The original derivation of Fibonacci numbers is presented in "Liber Abacci" (a book about the abacus), written in 1202 (a second edition of this appeared in 1228).

Consider the following numerical sequence:

$$x_n = x_{n-1} + x_{n-2} \quad (n \geqq 2) \tag{1}$$

Note that the term x_n is the sum of the two previous terms in the sequence. When the first two terms are each set equal to 1, the sequence is known as a Fibonacci sequence and the terms in the sequence are known as Fibonacci numbers. The table lists the first 21 Fibonacci numbers.

It might be of interest to develop the expression for the general term in the Fibonacci sequence.

Derivation of General Term

Let us assume that a solution of Eq. (1) has the following form:

$$x_n = k^n \tag{2}$$

Substituting this value for x_n in Eq. (1),

$$k^n = k^{n-1} + k^{n-2} \tag{3}$$

Dividing through by k^{n-2},

$$k^2 = k + 1 \tag{4}$$

Eq. (4) is a quadratic equation in k. Solving for the two values of k,

$$k_1 = \frac{1 + (5)^{1/2}}{2}; \quad k_2 = \frac{1 - (5)^{1/2}}{2}$$

The general solution of Eq. (1) has the following form:

$$x_n = Ak_1^n + Bk_2^n \tag{5}$$

where A and B are arbitrary constants that must be determined from two given values of x_n. These values are $x_0 = 1$ and $x_1 = 1$. Substituting $x_0 = 1$ at $n = 0$ in Eq. (5), the following is obtained:

$$A + B = 1 \tag{6}$$

Substituting $x_1 = 1$ at $n = 1$ gives

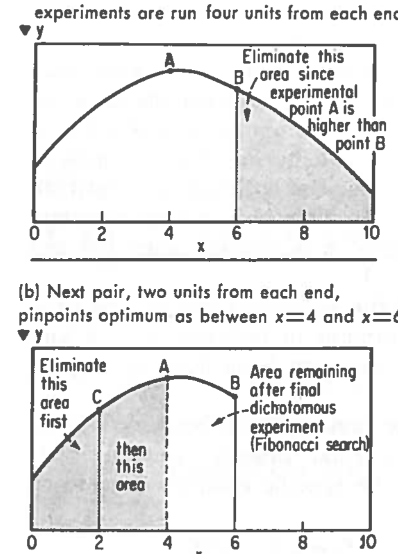
$$1 = Ak_1 + Bk_2$$

Fibonacci number is sum of two previous numbers

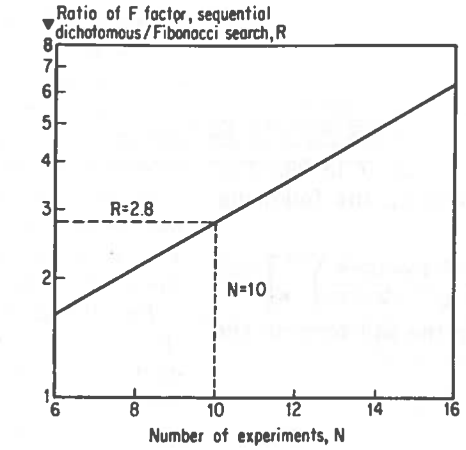
No. of Experiments, N	Fibonacci No. F_N	No. of Experiments, N	Fibonacci No. F_N
0....	1	11....	144
1....	1	12....	233
2....	2	13....	377
3....	3	14....	610
4....	5	15....	987
5....	8	16....	1,597
6....	13	17....	2,584
7....	21	18....	4,181
8....	34	19....	6,765
9....	55	20....	10,946
10....	89		

Fibonacci search: an efficient optimization technique—Fig. 5

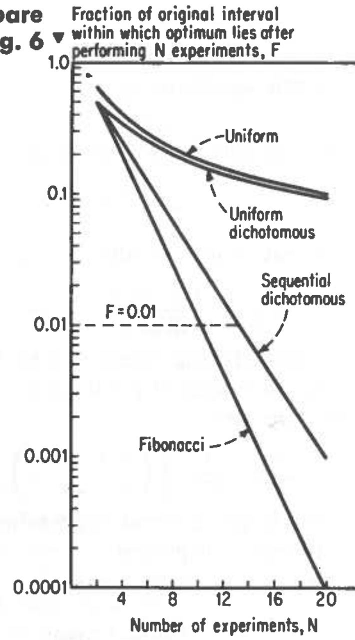
(a) Method uses Fibonacci numbers. Two experiments are run four units from each end.



Comparing Fibonacci vs. sequential dichotomous—Fig. 7



How search techniques compare in efficiency—Fig. 6



* The difference in values of x in the paired experiments has been assumed small enough so as not to influence the final answer.

† An abbreviation of filius Bonacci.

but, k_1 and k_2 have been solved previously; therefore,

$$1 = \frac{A[1 + (5)^{1/2}]}{2} + \frac{B[1 - (5)^{1/2}]}{2} \quad (7)$$

Solving Eqs. (6) and (7) simultaneously,

$$A = \frac{1 + (5)^{1/2}}{2(5)^{1/2}} \quad \text{and} \quad B = -\frac{1 - (5)^{1/2}}{2(5)^{1/2}}$$

Substituting these values for A and B in Eq. (5) together with the values for k_1 and k_2, the following is obtained:

$$x_n = \frac{1}{(5)^{1/2}}\left[\left(\frac{1 + (5)^{1/2}}{2}\right)^{n+1} - \left(\frac{1 - (5)^{1/2}}{2}\right)^{n+1}\right] \quad (8)$$

This is the general expression for the nth term in the Fibonacci sequence.

Since $1/(5)^{1/2} = 0.4472$

$$\frac{1 + (5)^{1/2}}{2} = 1.6180$$

$$\frac{1 - (5)^{1/2}}{2} = -0.6180$$

$$x_n = 0.4472\,[(1.6180)^{n+1} - (-0.6180)^{n+1}] \quad (9)$$

As n increases in value, the last term becomes negligible and the series is approximated by

$$x_n = 0.4472(1.6180)^{n+1} \quad (10)$$

How to Use Fibonacci Search

Now that we have discussed Fibonacci numbers and some of their properties, we are in a position to investigate the Fibonacci search.

A pair of experiments are run equidistant from each end of the interval. This distance, d_1, is determined from the following expression:

$$d_1 = \left(\frac{F_{N-2}}{F_N}\right)L \quad (11)$$

where F_{N-2} is the $N-2$ Fibonacci number
F_N is the Nth Fibonacci number
N is the number of experiments
L is the length of the interval

For the problem that we have been considering,

$$\begin{aligned} N &= 4 \\ L &= 10 \\ F_{N-2} &= F_2 = 2 \\ F_N &= F_4 = 5 \\ d_1 &= (2/5)(10) = 4 \end{aligned}$$

Therefore, the first two experiments are run 4 units from each end, i.e., at $x = 4$ and $x = 6$.

From these two results, we see that point A is higher than point B and hence we eliminate the area from $x = 6$ to $x = 10$. (See Fig. 5a.) The area remaining is from $x = 0$ to $x = 6$ and this becomes the new value of L in Eq. (11). The next value of d (i.e., d_2) is obtained by substituting $N - 1$ for N in Eq. (11):

$$\begin{aligned} d_2 &= (F_{N-3}/F_{N-1})L \\ &= (F_1/F_3)L = (1/3)(6) = 2 \end{aligned}$$

Therefore, the next pair should be run 2 units from each end or at $x = 2$ and $x = 4$; but one of these experiments has already been run at $x = 4$. Hence, only one additional experiment at $x = 2$ is required.

It will always turn out that one of the previous experiments is a Fibonacci experiment for the next run. We see from Fig. 5b that point A is higher than point C and we can, therefore, eliminate the area between $x = 0$ and $x = 2$. The remaining area is between $x = 2$ and $x = 6$. One more experiment is run around $x = 4$ to determine whether the optimum lies between $x = 2$ to 4 or $x = 4$ to 6. This is a dichotomous experiment in conjunction with the experiment that has already been run at $x = 4$.

At the conclusion of the Fibonacci search, we have narrowed down the optimum to between $x = 4$ and $x = 6$ which is better than we have done by any of the other methods.

For the general case, using the Fibonacci search, the fraction of the original interval remaining is equivalent to $1/F_N$. In the specific example considered here,

$$F = 1/F_N = 1/F_4 = 1/5 = 0.200$$

Comparison of Methods

Let us present graphically the methods discussed and the expression for the term F, as shown in Fig. 6. The advantage of the Fibonacci search is obvious. This chart provides a way of determining the number of experiments to perform to obtain a certain F value. For example, if it is desired to narrow down the interval to 1% of the original interval, so that $F = 0.01$, it is seen from the graph that 11 Fibonacci experiments would have to be made, compared with 14 for the sequential dichotomous search. It should be pointed out that all the curves were plotted as if the functions were continuous; the dichotomous search procedures are done in pairs and, therefore, the odd numbers do not have any meaning here.

If the approximate form of the Fibonacci equation is used, i.e. Eq. (10), and the ratio of F factors for the sequential dichotomous to the Fibonacci search is defined by R, then:

$$R = [(0.4472)(1.6180)^{N+1}]/2^{N/2}$$

$$\begin{aligned} R &= (0.4472)(1.6180)(1.6180/2^{0.5})^N \\ &= (0.7236)(1.144)^N \end{aligned}$$

R is plotted vs. N in Fig. 7.

For $N = 10$ experiments, the Fibonacci search is 2.8 times as effective as the sequential dichotomous search; this means that the final interval in the latter search will be 2.8 times that in the former.

References

1. Johnson, Selmer M., "Best Exploration for Maximum Is Fibonaccian," U.S.A.F. Project Rand Research Memo. RM-1590, Nov. 18, 1955.
2. Kiefer, J., *Proc. of the American Mathematical Society*, 4, No. 3, 502 (1953.)
3. Wilde, D. J., "Advances in Chemical Engineering," Vol. III, pp. 273-331, Academic Press, New York, 1962.

Key Concepts for This Article

For indexing details, see Chem. Eng., Jan. 7, 1963, p. 71 (Reprint No. 222). Words in bold are role indicators; numbers correspond to AIChE system.

Active Concept (8)	**Passive Concept (9)**	**Means/Methods (10)**
Optimization	Objective functions	Search methods
Explorations	Variable, dependent	Experimenting

Part 4

Optimizing Multivariable Functions

Many search methods are available for optimizing multivariable situations. Here is a review of some of the most useful techniques.

ARNOLD H. BOAS, *Socony Mobil Oil Co., Inc.**

In Part 3 of this series, we examined optimization methods that are useful where there is only one independent variable. However, the real problems of the engineering world are rarely so simple. Much more common—and much more complex—is the multivariable optimization problem.

The object here is to acquaint the reader with several multivariable search methods, along with their unique features and applications. These methods are useful either when the function to be optimized is known mathematically or when you are planning laboratory runs in order to optimize a function experimentally.

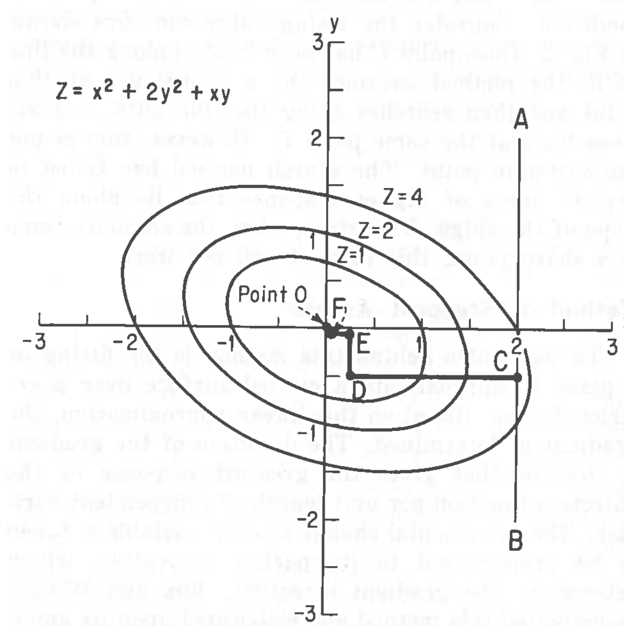
For most multivariable optimization problems, there is no single "right" approach. But, by getting acquainted with the various search methods and having a feel for the problem, individuals may develop a preference for one method over another for a particular type of problem.

Changing One Variable at a Time

The first method that might occur to one is the one-at-a-time technique. Friedman and Savage[4] described this method, which involves keeping all variables constant except one, and varying this to obtain an improvement in the objective function (the expression to be optimized), e.g., a higher value in the case of a maximization problem. This technique works efficiently when searches are conducted along axes parallel to the axes of the contour surfaces. If this is not the case, the search proceeds toward the optimum less efficiently. An example of this is indicated in Fig. 1.

To demonstrate the method mathematically, consider the problem of finding the minimum value of the function $z = x^2 + 2y^2 + xy$. By taking partial derivatives, we see that: $\partial z/\partial x = 2x + y$ and $\partial z/\partial y = 4y + x$. By keeping y constant and varying x, we see that z is related to the constant value of y; similarly, by keeping x constant and varying y, the value of z depends on the constant value of x. This indicates interaction between the variables. When there is no interaction, this method works most efficiently.

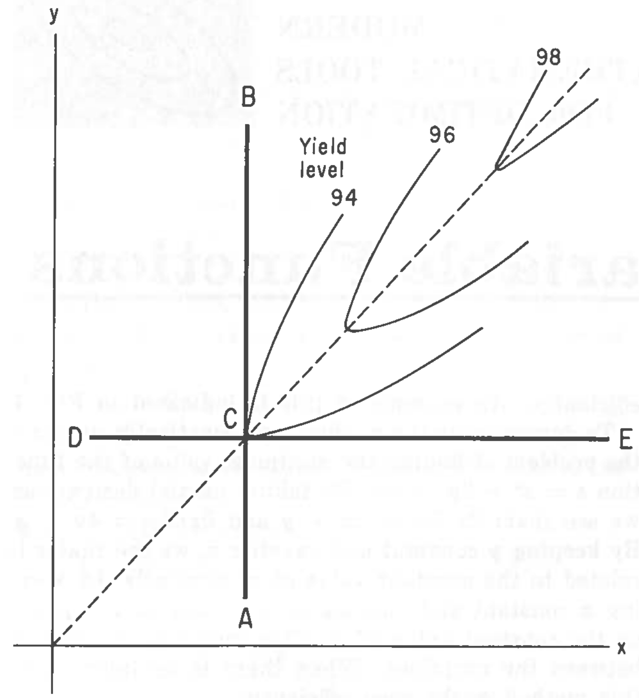
Referring to Fig. 1, an initial search is started with a constant value of $x = 2$. Note that along any one line of search, the problem reduces to a univariable search problem and we may use one of the methods described in Part 3 (Feb. 4). A search along the line ACB takes us to point C, which is the lowest value of the objective function along this line. The value of y at this point is -0.50. A new search is started along the line of constant y until point D is reached, which gives the lowest value of the objective function along the line CD. At this point, $x = 0.25$. The search continues to points E, F, etc. The minimum value of the function is zero at point 0, at which $x = y = 0$. The



One-at-a-time method locates minimum point—Fig. 1

* To meet the author, see *Chem. Eng.*, Dec. 10, 1962, p. 152.

Earlier articles in this series were: "What Optimization Is All About," Dec. 10, 1962; "How to Use Lagrange Multipliers," Jan. 7, 1963; "How Search Methods Locate Optimum in Univariable Problems," Feb. 4, 1963. Other important optimization techniques will be described in the next (and final) installment.



One-at-a-time method can fail to locate the optimum—Fig. 2

following table summarizes the results of the search:

x	y
2.000	−0.500
0.250	−0.067
0.031	−0.008
0.004	−0.001
etc.	etc.

This search can take place via mathematical analysis or by means of manipulation of experimental variables in a pilot plant. It should be mentioned that there are cases where this method will not locate the optimum condition. Consider the rising ridge contours shown in Fig. 2. Once point C has been located along the line ACB, the method assumes the y coordinate at this point and then searches along the line DCE and arrives back at the same point C. However, this is not the optimum point. The search method has failed to explore areas of higher response that lie along the slope of the ridge. Therefore, when the contours come to a sharp point, this method will not work.

Method of Steepest Ascent

The basic idea behind this method is the fitting of a plane to approximate a curved surface over a restricted area. Based on this linear approximation, the gradient is determined. The direction of the gradient is the one that gives the greatest response of the objective function per unit length of independent variable. The incremental change in each variable is taken to be proportional to its partial derivative, which determines the gradient direction. Box and Wilson[1] investigated this method and elaborated upon its applications. (Frequently, the method is called the Box-Wilson Method.) When dealing with a *minimum*, it is called the method of steepest *descent*. Let us work out an example to see how this method is utilized:

Example—Using the same objective function as Fig. 1, i.e., $z = x^2 + 2y^2 + xy$, let us start arbitrarily at the point $x = 2$ and $y = 2$ (point M in Fig. 3). The gradient direction is calculated from the partial derivatives: $\partial z/\partial x = 2x + y$ and $\partial z/\partial y = 4y + x$. At the starting point, these become $\partial z/\partial x = 2(2) + 2 = 6$; and $\partial z/\partial y = 4(2) + 2 = 10$.

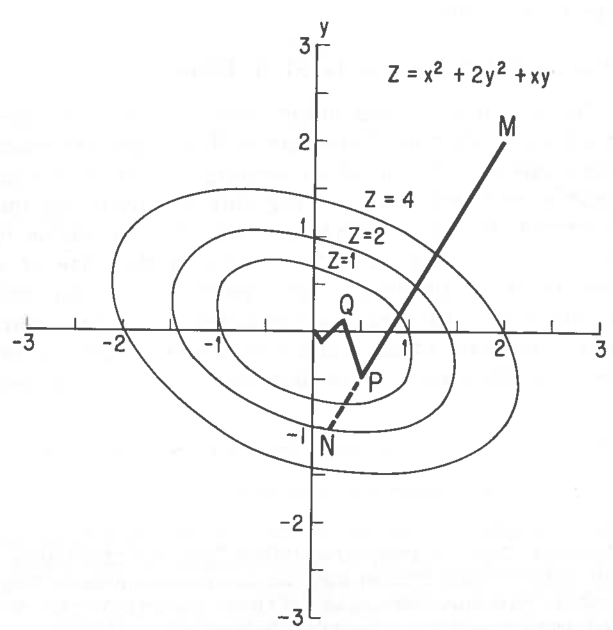
Since both partial derivatives are positive, the objective function varies in the same direction as x and y. Therefore, in a minimization problem, we should *decrease* both x and y to get a lower value of the objective function. The ratio of this decrease is taken to be the ratio of the partial derivatives, i.e., decrease in x divided by the decrease in $y = 6/10 = 0.60$. If the objective function is not known analytically, it is possible to estimate experimentally the partial derivative of z with respect to x by making small incremental changes in x (holding y constant) and noting the corresponding change in z.

The decrease in y is arbitrarily assumed to be 0.50 for the first step. The corresponding decrease in x is 0.60(0.50) = 0.30. Therefore, at the start:

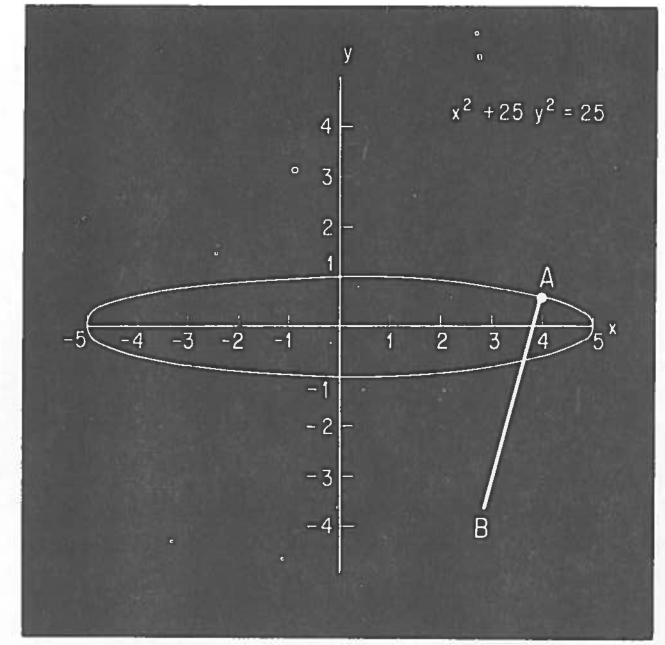
$$x_0 = 2.00 \qquad y_0 = 2.00 \qquad z_0 = 16.00$$
$$x_1 = x_0 - 0.30 = 2.00 - 0.30 = 1.70$$
$$y_1 = y_0 - 0.50 = 2.00 - 0.50 = 1.50$$
$$z_1 = (1.70)^2 + 2(1.50)^2 + (1.70)(1.50) = 9.94$$

As long as the objective function is decreasing, we shall continue along this gradient line. (We could, of course, calculate a new gradient direction each time, if we chose.) Continuing the calculations:

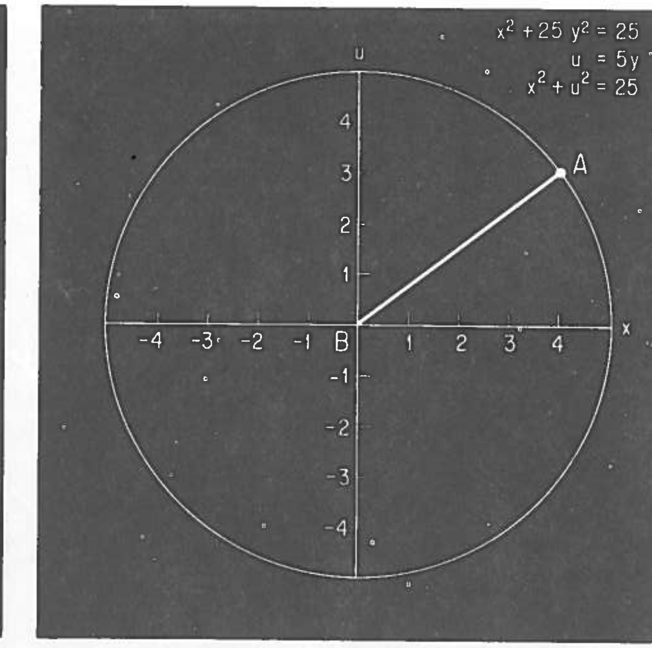
$$x_2 = 1.40 \qquad y_2 = 1.00 \qquad z_2 = 5.36$$
$$x_3 = 1.10 \qquad y_3 = 0.50 \qquad z_3 = 2.26$$
$$x_4 = 0.80 \qquad y_4 = 0.00 \qquad z_4 = 0.64$$
$$x_5 = 0.50 \qquad y_5 = -0.50 \qquad z_5 = 0.50$$
$$x_6 = 0.20 \qquad y_6 = -1.00 \qquad z_6 = 1.84$$



Path of steepest descent leads to minimum point—Fig. 3



If wrong scales are picked, steepest descent can miss optimum—Fig. 4a



By changing scales, steepest descent now leads to minimum point—Fig. 4b

Line MN on Fig. 3 represents the line of steepest descent from the starting point, M. Point N corresponds to x_6, y_6. We note from the figures above that we have obviously gone too far since the objective function is no longer decreasing. We go back to the last successful point, i.e., point x_5, y_5 (point P in Fig. 3), and calculate a new gradient direction as follows: Since $x = 0.50$ and $y = -0.50$ at point P,

$$\frac{\partial z}{\partial x} = 2x + y = 1.00 - 0.50 = 0.50$$

$$\frac{\partial z}{\partial y} = 4y + x = -2.00 + 0.50 = -1.50$$

We notice that $\partial z/\partial x$ is positive as before, which calls for a decrease in x; however, $\partial z/\partial y$ is negative and, therefore, z changes in the opposite direction from y. In order to obtain a decrease in the objective function, y must be *increased*. We are now closer to the optimum and shall arbitrarily reduce the step size in y from 0.50 to 0.30. Since the ratio of step changes in x to y is 1:3 (as seen from the partial derivatives), the change in x will be 0.10. Therefore, a new step 6 will be calculated:

$$x_6 = x_5 - 0.10 = 0.50 - 0.10 = 0.40$$
$$y_6 = y_5 + 0.30 = -0.50 + 0.30 = -0.20$$
$$z_6 = (0.40)^2 + 2(-0.20)^2 + (0.40)(-0.20) = 0.16$$

Continuing along this new gradient line:

$$x_7 = 0.30 \qquad y_7 = 0.10 \qquad z_7 = 0.14$$
$$x_8 = 0.20 \qquad y_8 = 0.40 \qquad z_8 = 0.44$$

At this point, we see that we have gone too far again since z_6 is greater than z_7. The direction is changed once more as indicated in Fig. 3 (point Q). This process is continued until we finally approach the optimum condition.

This step-by-step mathematical operation can be duplicated in the laboratory by first determining the partial derivatives experimentally as noted above, then using this relation to determine the changes that should be made in x and y. Moves are made along gradient lines until the best experimental value of z is found.

Limitations of Steepest Ascent

A few comments about the method of steepest ascent are now in order. The efficiency of the method is related to the choice of scales used. Surfaces with spherical contours give the fastest convergent rates; the closer the contours are to being spherical, the better the convergence. There are many directions of steepest ascent, each depending upon the choice of scales of the independent variables. We want a good response of the objective function relative to a given change in distance of the independent variables. This change in distance, however, is related to the choice of scales, e.g., one inch on a plot may represent 10° F. or 500 psi. or 1 million Btu./hr.

In Fig. 4a there is a plot of the function $z = x^2 + 25y^2$, with $z = 25$.

At point A, $x = 4$, $y = 0.6$, the direction of steepest descent is indicated by line AB. But this does not point to the origin, which is the optimum. However, by a change in scale (i.e., let $u = 5y$ so that $z = x^2 + u^2$) we note the line of steepest descent now does point to the optimum, as indicated in Fig. 4b.

Another shortcoming of the method is the extrapolation. By moving along the gradient line, we have assumed an extrapolation of the partial derivatives. But the shape of the response surface is usually changing as the search continues along a given path; in other words, the assumption that a plane represents the surface may no longer hold.

Still another shortcoming of this method (and most others to be discussed) is the failure to locate the global peak. The method only searches for local extremes. One alternative is to start over again at some other point and explore the surface once more to check for unimodality (one peak).

Attempts have been made to speed the convergence rate of the method of steepest ascent. One of these is the acceleration technique of Forsythe and Motzkin[8]. This method is shown in Fig. 5. Point B is found to give the best value for the objective function along any line ABC. By the method of steepest ascent, the direction BDE is located and point D is found to give the best value along this line. The method of steepest ascent once more locates direction FDG, and point H is found to give the best response along this path.

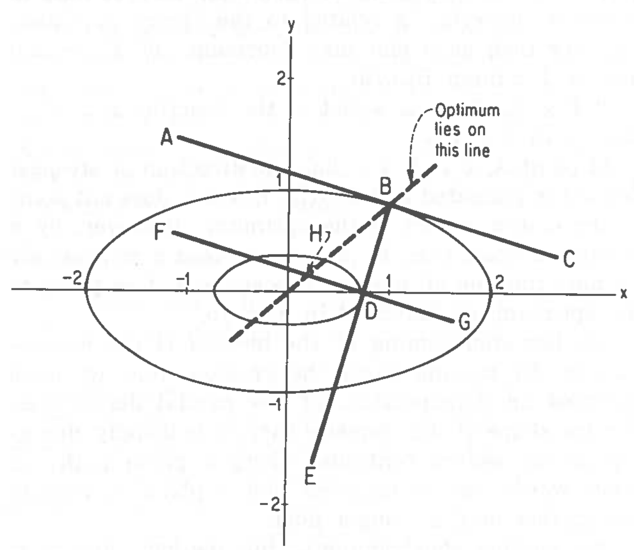
Now, by the acceleration technique, a line through points B and H will pass through the optimum value (at the origin in this case). For a quadratic function, this is exact. Further iterations are required for the nonquadratic surfaces. The theorem holds for two independent dimensions, although Finkel[9] reports it is fairly successful for more than two independent variables.

Direct Search

The "direct search" method of Hooke and Jeeves[7] is a sequential type of search where each solution is compared with the optimum up to that time. A strategy, based on previous results, is established to determine the values of the independent variables for the next trial. This method is devoid of any classical techniques.

No satisfactory rules for the success of direct search have been set forth, and the method can fail. However, different starting points can be used as a check.

A particular strategy must be established for each problem. One type of strategy is known as the "pattern search." This type first makes an exploratory move to study the behavior of the objective function.



Acceleration technique locates line on which optimum lies—Fig. 5

This behavior is of a qualitative nature and merely serves as a guide to the direction of the search. We are only interested in the success or failure of a move at this point. A simple exploratory move would be to change one independent variable at a time.

The second move in the pattern search is the "pattern move." This move uses the information obtained in the exploratory move and actually optimizes the function by moving in the indicated direction. Each pattern move is then followed by a sequence of exploratory moves from the last base point. The entire direct search procedure may be considered as a search from base point to base point. This technique combines some of the features of the univariable and steepest ascent methods.

Approximate solutions are obtained at every trial, improving the function each time. At any given stage of the search, the "best" solution is available up to that point, unlike other methods that tend to overshoot and then retrace their steps. This technique is suitable for electronic computers and circumvents some of the shortcomings of the method of steepest ascent. The authors[7] have used it in curve-fitting problems, solving integral equations, maximizing or minimizing functions with or without restrictions on the independent variables, and solving systems of equations.

Handling of Constraints

Two novel methods for handling constraints on functions have been introduced by Roberts and Lyvers.[10] These are known as "hemstitching" and "riding the constraint." They can be used in conjunction with the optimizing methods mentioned previously.

In hemstitching, a base point is chosen that lies within the constraints, and the search is started according to the method of steepest ascent. After the next point is calculated, it is checked to see whether any of the constraints have been violated. If not, the search continues. When a constraint is violated, the gradient is calculated with respect to the *constraint* rather than with respect to the objective function. Let us work out an example to illustrate this technique:

Example—Find the minimum value of the function $x^2 + y$, subject to the inequality constraint that $x + y \geqq 4$. Use the method of hemstitching to handle the constraint. Start at the point $x = 3$ and $y = 1$.

Solution—The objective function to be minimized is $z = x^2 + y$. Let us first find the partial derivatives of z with respect to x and y: $\partial z/\partial x = 2x$ and $\partial z/\partial y = 1$. Therefore, the ratio of the change in y to the change in x is $1/2x$. At the starting point, this becomes, $1/(2)(3) = 1/6$. Arbitrarily, the change in x will be taken as 0.50, so that the change in y will be 0.08.

The initial points will be denoted by $x_1 = 3.00$, $y_1 = 1.00$. Then:

$$x_2 = x_1 - 0.50 = 3.00 - 0.50 = 2.50$$
$$y_2 = y_1 - 0.08 = 1.00 - 0.08 = 0.92$$

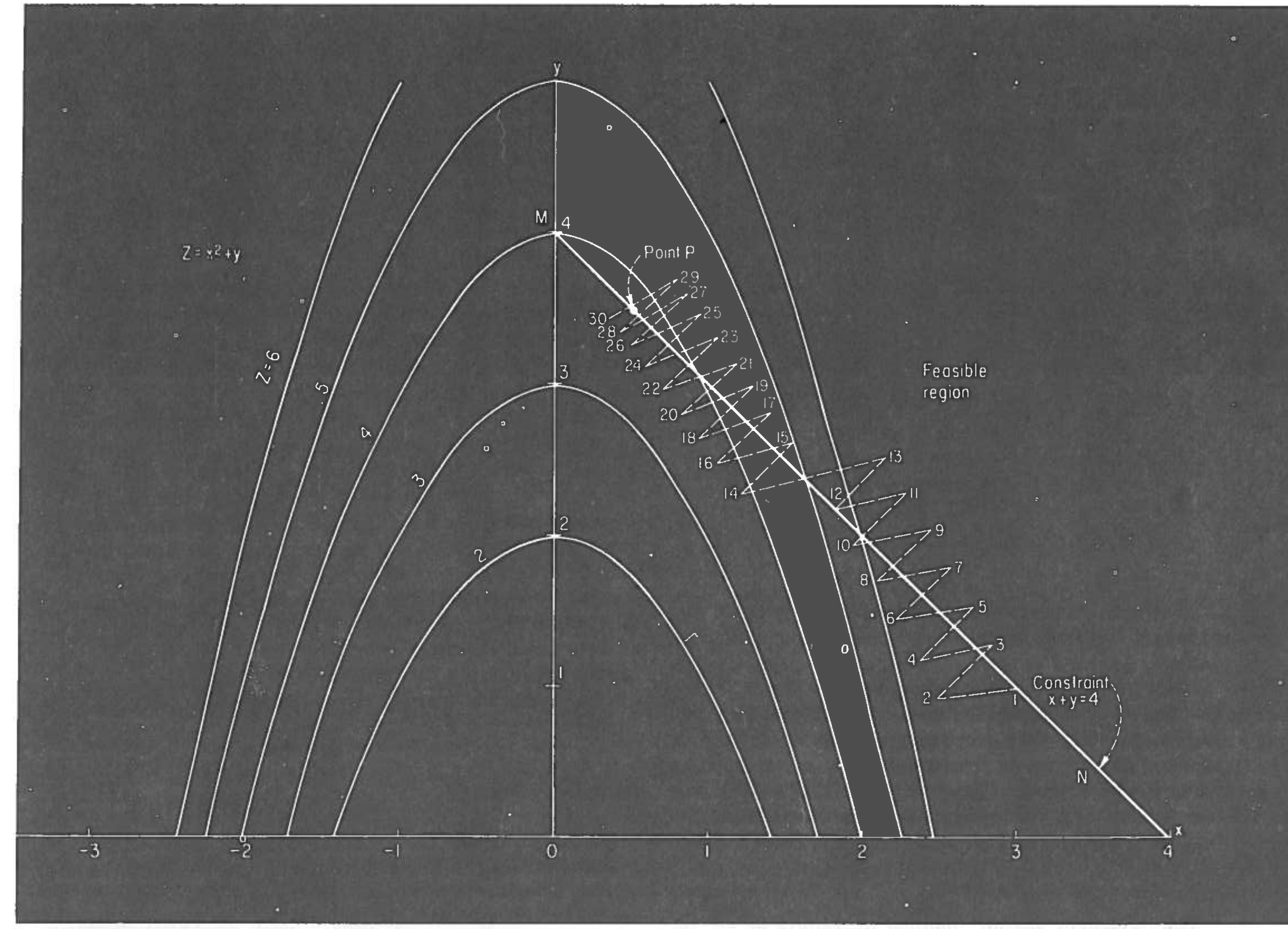
Now, $x + y = 2.50 + 0.92 = 3.42$. This violates the constraint that $x + y \geqq 4$. Let us denote the constraint expression by g so that: $g = x + y$. Taking partial derivatives of g (the constraint expression) with respect to x and y: $\partial g/\partial x = 1$, and $\partial g/\partial y = 1$. The ratio of the partial derivatives is equal, so that equal increments are taken in x and y. Arbitrarily, this will be taken as 0.35. Since the sum of x and y was too low, it is obvious that their values must be increased in order to satisfy the inequality constraint. Therefore:

$$x_3 = x_2 + 0.35 = 2.50 + 0.35 = 2.85$$
$$y_3 = y_2 + 0.35 = 0.92 + 0.35 = 1.27$$

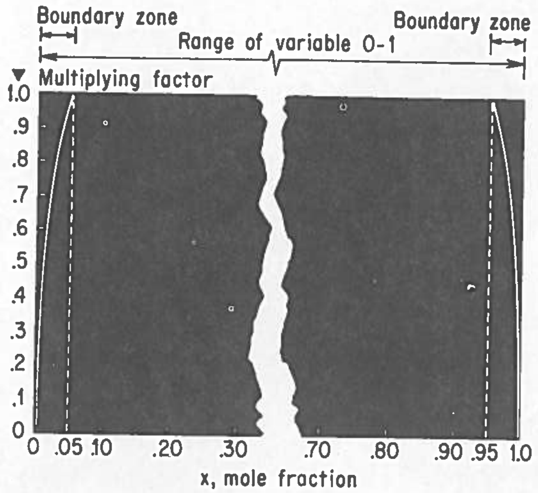
Fig. 6 illustrates this method. The parabolas represent the contours of the objective function. The line MN represents the inequality constraint. Only values of x and y that lie either on or to the right of this line are possible solutions to the problem. The starting point is point 1 and the next point is point 2. This is seen to violate the constraint and point 3 puts us back into the feasible region. The solution to the problem is $x = 0.50$ and $y = 3.5$ ($z = 3.75$ denoted by point P). The crossing of the constraint boundary represents a form of mathematical "hemstitching." The reasoning behind this concept is that the fastest way to move out of the ineligible region is to move orthogonal to the constraint on the objective function.

Another method for handling this type of problem is "riding the constraint." Once a constraint is violated, the point is put back on the constraint line, and values of the independent variables are taken so that the points lie on the constraint. When more than one constraint exists, each time a new constraint is violated the method requires that the points be switched over to the new constraint. The new variables are picked so that they lie on the new constraint; i.e., the one that has been recently violated. The assumption made in this method is that the optimum solution lies on a constraint; the hemstitching technique does not make this assumption.

Still another method for handling constraints was introduced by Rosenbrock[12]. The objective function is modified by means of multiplying factors. Whenever one of the variables violates a constraint, the multiplying factor is zero (i.e., the objective function is multiplied by zero and, hence, is equal to zero). When the variable is within the feasible region, the factor is 1.0 and the objective function assumes its full value. However, when the value of the variable falls within a prescribed "boundary zone," the multiplying factor



Hemstitching method aids optimization where a constraint exists—Fig. 6



Use of multiplying factors is another way to handle constraints—Fig. 7

is assumed to behave parabolically from 0 to 1; the objective function, therefore, varies from 0 to its full value.

Let us look at Fig. 7. The variable x denotes mole fraction and will have the limits of 0 and 1. The width of the boundary zone is arbitrarily defined as 0.05 so that the lower boundary zone is from 0 to 0.05 and the upper boundary zone is from 0.95 to 1.0. Note the way the multiplying factor varies parabolically in the boundary zone at each end of the figure.

Gradient Search Method

The "gradient search" method of Zellnik et al[17] includes a random saddle-point check as well as a random scan for alternate optimums. The advantage of this method over others is claimed to be its efficient scan for saddle points and alternate optimum values.

In all search procedures, the problem of local versus global extremes exists. This method considers a random scan more efficient than a systematic search for alternate optimums. Where N is the number of dimensions, 8N evaluations are made of the objective function. When N is large, a more efficient way to search for alternate optimums would be to start the search over again from a different point.

The Lattice Method

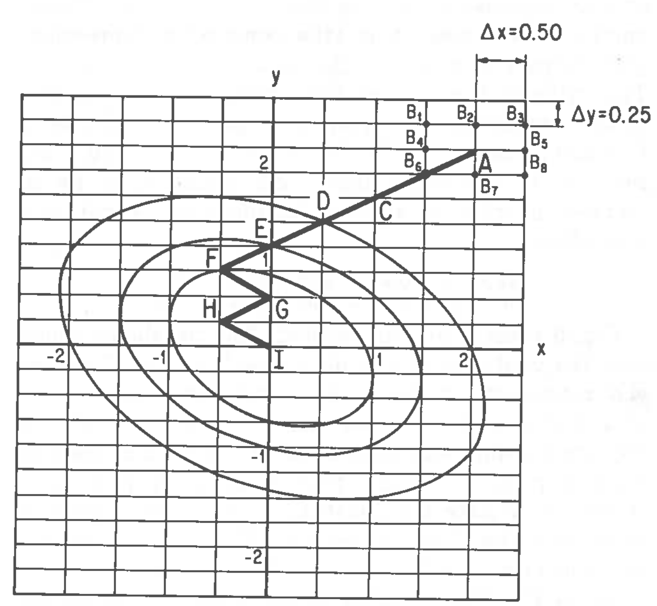
Wilde[18] introduced "gradient free" search methods to circumvent some of the shortcomings of the method of steepest ascent. He notes that the major failing of the method is that the gradient direction is related to the choice of scales. Two other shortcomings are (1) the need for extrapolating along the line of steepest ascent (i.e., assuming that the gradient does not change along the original path) and (2) failure to indicate the step size (i.e., the location of the next block of experiments). Two types of search techniques developed by Wilde to circumvent some of these shortcomings are the "lattice method" and the "contour tangent method."

Lattice method uses the *maximum allowable changes* in the independent variables. In an experimental determination of the objective function, we may not wish to change any one variable too drastically. For example, suppose that the variable x should not change by more than 0.50 units nor the variable y by more than 0.25 units. A grid is constructed where the allowable points are points on the grid. In Fig. 8, such a grid has been constructed; the contour lines are values of the objective function.

Assume that the search starts at point A. There are then eight possible values for the next point, using the maximum allowable changes in the variables, points B_1 through B_8. The direction is determined by noting the sign of the partial derivative. At point B_6, the objective function is found to give a favorable response, with a decrease in both x and y. Therefore, point B_6 becomes the second point in the search. This one point of the eight gives the most favorable response in the function. Partial derivatives are evaluated at B_6, and point C is arrived at by the same procedure. The search continues to point F where it is found that x should be *increased* and y decreased; this takes us to point G. Points H and I (the optimum) are obtained similarly. This technique is called the "lattice search method" because successive experiments are made on the lattice-like grid of consecutively applied constraints.

Tangent Methods

Let us illustrate the "contour tangent method" by examining Fig. 9. Point A is the initial point of the search. A tangent to the objective function contour is drawn at this point. This line divides the entire area into two sections. By assuming strong unimodality, Wilde[18] shows that area I (DEGF) can be eliminated.



Lattice method allows largest possible changes in variables—Fig. 8

(The decision as to which area to eliminate is determined by the partial derivatives.)

The solution must now lie in the remaining area EJLG. A point is selected inside this region. Four points are considered by the author,[18] the midpoint, minimax, center of volume and centroid. One of these is selected as the next point. The easiest to calculate is the midpoint, which is merely all the variables taken at the midpoints between their extreme values. Point B is the next one chosen on Fig. 9. A contour tangent is drawn through B, and area GEIH (Area II) is now eliminated. Similarly, point C is located, a contour tangent drawn, and area III (IJKM) is eliminated. The final area has been narrowed down to KMHL.

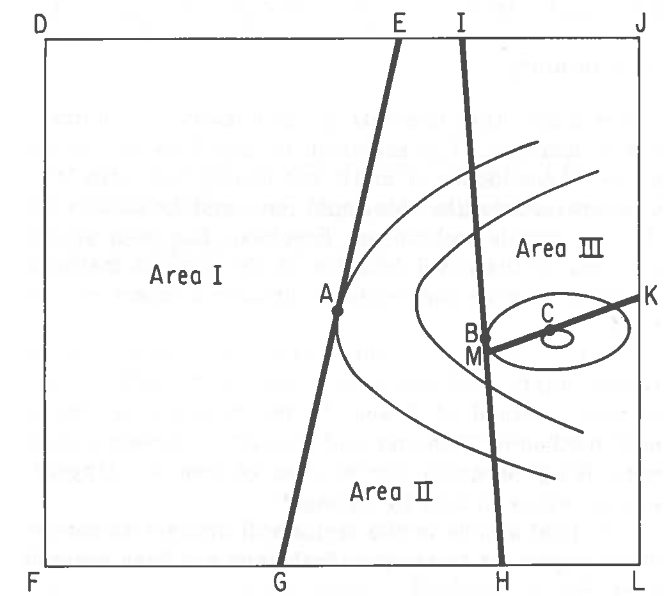
One fundamental requirement in this method is that the dependent variable be strongly unimodal (this requirement, however, applies to the method of steepest ascent as well).

Shah, Buehler and Kempthorne[14] introduced the method of parallel tangents (partan) as a means of locating the extreme value of a function. This method is applicable when the function is known analytically, as well as when it is not but is measurable (e.g., yield of a particular product in a chemical reaction).

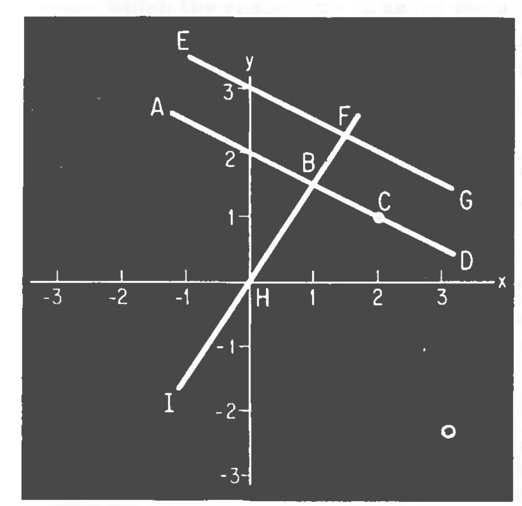
Based on the acceleration technique of Forsythe and Motzkin[8] already described, partan has been developed using the method of steepest ascent. Referring to Fig. 5, the steps are as follows:

1. Starting at point B, determine the line of steepest ascent (line BDE).
2. Find the highest point on this line (point D).
3. Determine the line of steepest ascent from point D (line FG).
4. Find the high point on this line (point H).
5. Draw a line through points B and H, and search for the high point on this line (point at the origin is the optimum).

Another variation of partan does not use steepest ascents at all. By referring to Fig. 5 once again, we see that line AC is parallel to line FG. This fact is now used to develop the variation of partan that does

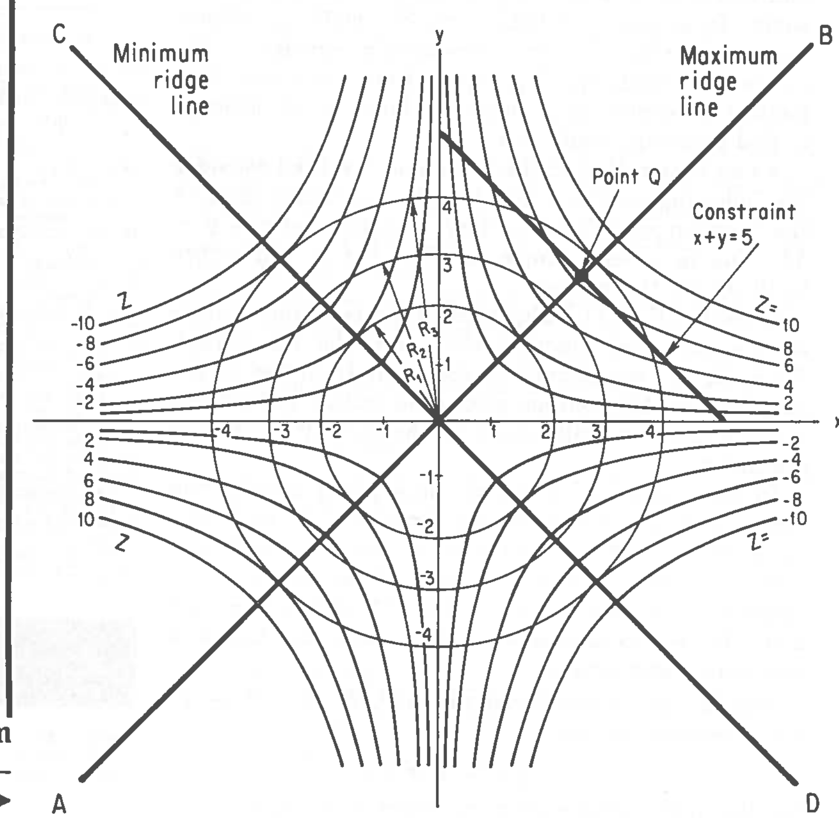


▲Contour tangents narrow down the search area—Fig. 9



▲Parallel tangent method avoids use of steepest ascent —Fig. 10

Ridge analysis finds optimum on complex surfaces— Fig. 11▶



not use steepest ascent (Fig. 10). The steps are as follows:

1. Start at any point, C.
2. Draw a line through point C arbitrarily (line AD).
3. Find the high point on this line (point B).
4. Draw another line parallel to line AD (line EG).
5. Find the high point on this line (point F).
6. Join the two high points to determine a new line of search (line FI).
7. Search along this line until the peak is found (point H).

The contours have intentionally been omitted from the diagram to show that these need not be known for this particular application.

Thus, without the use of gradients, we have a method of locating the peak for elliptical contours. Ordinary steepest ascent methods do not necessarily give finite convergence, whereas partan, which does use steepest ascent, gives convergence regardless of the choice of scales. However, scales should be chosen so as to obtain as nearly circular contours on the function as possible.

The partan method has been extended to the multivariable case where one is dealing with tangent planes rather than tangent lines. With more than three independent variables, we enter the realm of hyperplanes and hyperspace, which makes the solution more complex.

Ridge Analysis

With ridge analysis,[8, 9] it is possible to characterize multivariable optimization problems in two dimensions. It is apparent that in multivariable problems, response surfaces become increasingly complex as the number of variables increases. Ridge analysis is a method of exploring complex surfaces in an attempt to find optimum conditions.

As an illustration of the technique, we shall consider the following problem: What is the maximum value of the function xy, subject to the constraint that $x + y = 5$? This is a very simple problem but it will suffice to illustrate the method.

Consider Fig. 11. The hyperbolas represent values of the objective function that is to be maximized. $R_1, R_2, R_3 \ldots$ are defined as the radii from the origin to points on the contour plot. The radius sweeps out a circle and the peak values of the objective function are noted.

In this particular problem, these peaks take on the same numerical values (except for sign). As each new value of R is chosen and new circles are swept out, different peaks will be obtained. These peaks lie on ridge lines and are shown in Fig. 11 by the lines AB and CD. AB is a maximum ridge line and CD is a minimum ridge line.

For the two-dimensional problem, $R = \sqrt{x^2 + y^2}$, which generalizes to:

$$R = \sqrt{u^2 + v^2 + w^2 + x^2 + y^2 \ldots}$$

for the multivariable problem where $u, v, w, x, y \ldots$ are the independent variables. The geometrical interpretation is quite complex for the multivariable case. But a two-dimensional plot of z vs. R can thus be used even for the multidimensional optimization problem.

For the particular problem above, there is a constraint of $x + y = 5$. This is plotted on Fig. 11, and the intersection of this line with the maximum ridge line AB gives the peak conditions subject to the constraint of the problem. The solution is indicated at point Q, where $x = y = 2.5$ and $z = 6.25$.

In Summary

There are other methods of optimization too numerous to mention. The exclusion of any does not imply a lack of confidence or merit but merely indicates that a review article like this could not possibly include all the worthwhile techniques. Emphasis has been placed on some of the novel features of the various methods in order to give the reader a broader concept of the field.

Some of the more recent methods that the interested reader might find instructive are the "Gradient Projection" method of Rosen[11, 12], the "Conjugate Gradient" method of Hestenes and Stiefel[5], nonlinear digital optimizing program for process control by Mugele[9], and a review article by Spang.[15]

The final article in the series will attempt to review some important techniques that have not been covered thus far.

References

1. Box, G. E. P., Wilson, K. B., *Jour. Roy. Stat. Soc.*, B 13, 1 (1951).
2. Finkel, R. W., Paper 71, Preprints of Papers Presented at 14th National Meeting of Assn. of Computing Machinery (1959).
3. Forsythe, G. E. and Motzkin, T. S., Bull. Am. Math. Soc., 57 183 (1951).
4. Friedman, M. and Savage, L. S. "Selected Techniques of Statistical Analysis," McGraw-Hill, New York, 1947.
5. Hestenes, M. R., Stiefel, E., *Jour. Res. Nat. Bur. Standards, 49* 409 (1952).
6. Hoerl, A. E., *Chem. Eng. Prog., 55,* No. 11, 69 (1959).
7. Hooke, R., Jeeves, T. A., *Jour. Assn. Computing Machinery,* 8, No. 2, 212 (1961).
8. Lester, J. T., *Brit. Chem. Eng. 6,* No. 12, 851 (1961).
9. Mugele, R. A., "A Nonlinear Digital Optimizing Program for Process Control Systems," Proc. Western Joint Computer Conference, Spring 1962.
10. Roberts, S. M., Lyvers, H. I., *Ind. Eng. Chem., 53,* 877 (1961).
11. Rosen, J. B., *S.I.A.M. Jour. 8,* 181-217 (1960).
12. Rosen, J. B., *S.I.A.M. Jour. 9,* 514-532 (1962).
13. Rosenbrock, H. H., *Comput. Jour. 3* 175 (1960).
14. Shah, B. V., Buehler, R. J., Kempthorne, O., Iowa State U. of Science and Technology, Tech. Report No. 2, April 1961, Revised August 1962.
15. Spang, H. A., *S.I.A.M. Review, 4,* No. 4 343 (1962).
16. Wilde, D. J.—"Gradient-Free Search For An Optimum"—Presented at AIChE National Meeting, Los Angeles, Calif., Feb. 5, 1962.
17. Zellnik, H. E., Sondak, N. E., and Davis, R. S., *Chem. Eng. Prog., 58,* No. 8 35 (1962).

Key Concepts for This Article

For indexing details, see Chem. Eng., Jan. 7, 1963, p. 73 (Reprint No. 222). Words in bold are role indicators; numbers correspond to AIChE system.

Active (8)	**Passive (9)**	**Means/Methods (10)**
Optimization	Objective functions	Search methods
Explorations	Variable, dependent	Experimenting
	Multivariable functions	

Part 5

Optimization Via Linear and Dynamic Programming

Here are two basic optimization tools that are useful for solving many common types of engineering problems.

ARNOLD H. BOAS, *Socony Mobil Oil Co., Inc.*

The first four articles in this series have covered some of the major mathematical tools for optimization. However, there are two important methods—linear programming and dynamic programming—that have not been discussed thus far.

Each of these techniques has been extensively described in the literature and no attempt will be made here to delve into the theory. Rather, we will solve some elementary problems via these methods and thus give the reader a feel for the types of problems that can be handled and the form that the solution takes. Mention will be made of many literature references upon which the reader can draw to get a deeper understanding of any particular aspect of these methods.

Let us first look at the concept of linear programming.

Linear Programming

In the context of this article, the word "programming" will indicate the planning of activities with the goal of optimization in mind, e.g., determination of optimum product mix. When the objective function (the function to be optimized) and its constraints are linear, we speak of the optimization process as "linear programming."

The mechanics of linear programming can best be illustrated by a simple problem.

Problem—A chemical plant is planning to produce three products: A units of product 1, B units of product 2, and C units of product 3, having net profits of 10, 4 and 1 per unit, respectively. Due to process conditions, the following constraints must be met:

$$A + B \leqq 5 \text{ and } 2A + B + C \leqq 20.$$

The problem is to find the values of A, B, and C so that profit will be at a maximum.

Solution—The total profit, P, is the objective function to be maximized and can be written: $P = 10A + 4B + C$.

Two new non-negative variables, S_1 and S_2 (called slack or dummy variables), are introduced into the problem. These convert the inequality constraints to equality constraints. Thus:

$$A + B + S_1 = 5 \qquad (1)$$

$$2A + B + C + S_2 = 20 \qquad (2)$$

The method for proceeding now depends on the basic theorem of linear programming, which may be stated as follows: in the optimal solution of any linear programming problem, the total number of non-zero variables (ordinary and slack) is exactly equal to the number of constraints. A solution that satisfies this requirement is known as a *basic* solution to the optimization problem.

In our problem, there are two inequality constraints and hence the optimum solution will contain two non-zero variables. In other words, of the five variables (A, B, C, S_1, S_2), only two of these will not be equal to zero. An additional physical limitation on the problem is that none of the production quantities can be negative; i.e., $A \geqq 0$, $B \geqq 0$, $C \geqq 0$.

The first basic solution that we might look at is: $A = B = C = 0$; $S_1 = 5$; $S_2 = 20$. This solution satisfies the constraints of the problem but is not the optimum solution since profit is zero in this case. But this is only one of many possible basic solutions. The procedure now is to examine other basic solutions until the optimum is found.

For example, suppose that A and B are assumed to be the non-zero variables; then $C = S_1 = S_2 = 0$. From Eq. 1, $A + B = 5$, and from Eq. 2, $2A + B = 20$. Solving these equations simultaneously, $A = 15$, $B = -10$. This is not a feasible solution since B violates the constraint that $B \geqq 0$.

The following table lists the solutions found by taking the variables two at a time and keeping the others equal to zero.

Earlier articles in this series were: "What Optimization Is All About," Dec. 10, 1962; "How to Use Lagrange Multipliers," Jan. 7, 1963; "How Search Methods Locate Optimum in Univariable Problems," Feb. 4, 1963; "Optimizing Multivariable Functions," Mar. 4, 1963.

Table I

A	B	C	S_1	S_2	P
15	−10	0	0	0	N.F.*
5	0	10	0	0	60
0	5	15	0	0	35
10	0	0	−5	0	N.F.
5	0	0	0	10	50
0	20	0	−15	0	N.F.
0	5	0	0	15	20
0	0	20	5	0	20
0	0	0	5	20	0

* Not feasible due to violation of non-negativity constraint.

This table shows that the optimum solution is:

$$A = 5, \quad B = 0, \quad C = 10, \quad S_1 = S_2 = 0, \quad P = 60.$$

In this simple problem, it was possible to investigate all basic solutions and select the optimum. However, when the number of combinations becomes very large, this is not a practical procedure, and more efficient and systematic procedures must be used. The most widely employed technique is the *simplex method* of Dantzig,[6] which has been described by Charnes et al.[3]

Dynamic Programming

Dynamic programming is an optimization technique that is useful for multistage problems. Bellman[7] introduced this concept and expressed the principle of optimality as follows: "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

This technique relies upon decision-making at each stage rather than trying to solve the entire N-stage optimization problem simultaneously. Bellman goes into the mathematics quite deeply but, in line with the theme of this series, the method will be described here mainly in terms of its application to a specific problem.

Multistage problems are quite common in the field of chemical engineering, and thus dynamic programming finds many applications. Some of the more common types of application are in: catalyst regeneration and replacement, feed allocation, multistage unit operations.

Let us illustrate the method by solving a simple problem.

Problem—Consider the heat-exchanger train shown in Fig. 1. The following nomenclature will be used in setting up the problem:

W = rate of fluid flow, lb./hr.
C_p = specific heat, Btu./(lb.)(°F)

In addition, the following terms apply to the heat-exchanger train:

Exchanger	Over-all Heat Transfer Coefficient, Btu./(Hr.)(Sq.Ft.)(°F.)	Area Required, Sq.Ft.	Duty, Btu./Hr.
E-1	$U_1 = 120$	A_1	Q_1
E-2	$U_2 = 80$	A_2	Q_2
E-3	$U_3 = 40$	A_3	Q_3

For this example, all $W(C_p)$ values will be assumed to equal 100,000.

The problem is to select the temperatures T_1 and T_2 so that the total area $(A_1 + A_2 + A_3)$ is a minimum.

Solution—Let us start at exchanger E-3. The temperatures of the two streams in and out of this exchanger are:

$$T_2 \rightarrow 500$$
$$T_3 \leftarrow 600$$

Since the $W(C_p)$ terms are the same for each stream, by heat balance: $600 - T_3 = 500 - T_2$, or $T_3 - T_2 = 100$, which is the mean temperature difference (MTD) for heat transfer.

Now: $Q_3 = U_3A_3(MTD) = (40)(A_3)(100)$
But $Q_3 = WC_p(500 - T_2) = 100{,}000(500 - T_2)$.

Equating the two expressions and solving for A_3:

$$A_3 = 12{,}500 - 25T_2 \quad (3)$$

Let us now consider exchanger E-2. The temperatures in and out of this second heat exchanger are:

$$T_1 \rightarrow T_2$$
$$T_4 \leftarrow 400$$

Since $T_2 - T_1 = 400 - T_4$
$T_4 - T_1 = 400 - T_2 = MTD$
$Q_2 = U_2A_2(MTD) = (80)(A_2)(400 - T_2)$
But $Q_2 = WC_p(T_2 - T_1) = 100{,}000(T_2 - T_1)$
Therefore: $A_2 = 1250(T_2 - T_1)/(400 - T_2)$ (4)

At exchanger E-2, a decision must be made. Either T_2 or T_4 must be set for a fixed T_1 in order to determine the area, A_2. Let us arbitrarily fix T_1 at 100 F. and find the best T_2 to use. Table II shows the results.

This is a one-dimensional optimization problem. We are only looking for the best T_2 to go with $T_1 = 100$. For any T_2, A_2 is calculated from Eq. 4 and A_3 from Eq. 3. The best T_2 is seen to be $T_2 = 277$ F.

This type of one-dimensional optimization is repeated for various values of T_1 and the results are shown in Table III. This table, a summary of a two-stage optimization, shows the best way to design E-2 and E-3 for any T_1. It is not concerned with E-1 or any upstream conditions other than T_1.

Let us now consider exchanger E-1. The temperatures in and out are:

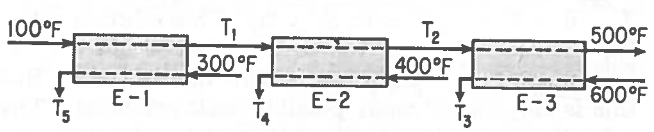
$$100 \rightarrow T_1$$
$$T_5 \leftarrow 300$$

Since: $T_1 - 100 = 300 - T_5$
$300 - T_1 = T_5 - 100 = MTD$
$Q_1 = U_1A_1(MTD) = (120)(A_1)(300 - T_1)$
But $Q_1 = WC_p(T_1 - 100) = 100{,}000(T_1 - 100)$
Therefore: $A_1 = 833(T_1 - 100)/(300 - T_1)$ (5)

Once T_1 is fixed, A_1 can be calculated from Eq. 5. By referring to Table III, we know the best way to design E-2 and E-3 for this T_1. Therefore, we know the best way to design all three stages of the heat-exchanger train for any given T_1.

We now calculate A_1 for each chosen T_1. Table IV summarizes these results and enables us to select the

Dynamic programming can find minimum exchanger area required—Fig. 1



Locating best T_2 and areas for a given T_1—Table II

$T_1 = 100$°F

T_2, °F	A_2, Sq.Ft.	A_3, Sq.Ft.	A_2+A_3, Sq.Ft.
100	0	10,000	10,000
150	250	8,750	9,000
200	625	7,500	8,125
250	1,250	6,250	7,500
270	1,635	5,750	7,385
275	1,750	5,625	7,375
277	1,799	5,575	7,374
280	1,875	5,500	7,375
285	2,011	5,375	7,386
300	2,500	5,000	7,500
350	6,250	3,750	10,000
400	∞	2,500	∞

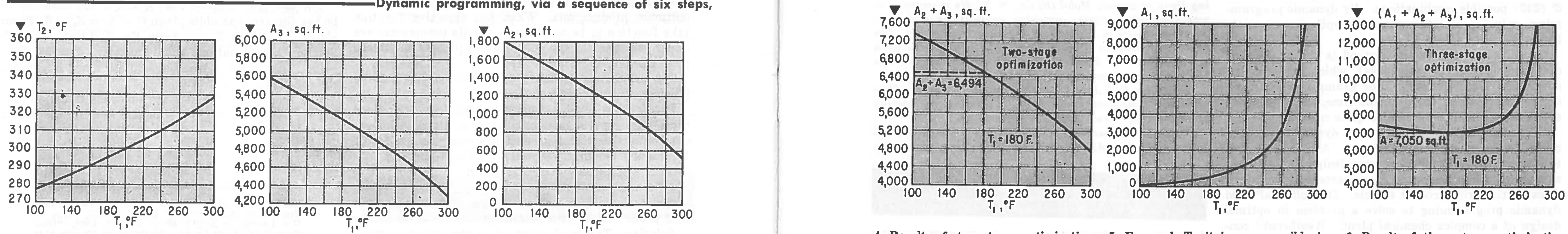
For each T_1 there is a best T_2, A_2 and A_3—Table III

T_1, °F	T_2, °F	A_2, Sq.Ft.	A_3, Sq.Ft.	A_2+A_3, Sq.Ft.
100	277	1,799	5,575	7,374
150	288	1,540	5,300	6,840
160	290	1,477	5,250	6,727
170	293	1,437	5,175	6,612
180	295	1,369	5,125	6,494
190	298	1,324	5,050	6,374
200	300	1,250	5,000	6,250
210	302	1,173	4,950	6,123
220	305	1,118	4,875	5,993
230	308	1,060	4,800	5,860
240	311	997	4,725	5,722
250	313	905	4,675	5,580
260	316	833	4,600	5,433
270	319	756	4,525	5,281
280	322	673	4,450	5,123
290	326	608	4,350	4,958
300	329	511	4,275	4,786

Each T_1 also determines an A_1—Table IV

A_1, Sq.Ft.	$A_1+A_2+A_3$, Sq.Ft.
0	7,374
278	7,118
357	7,084
449	7,061
556	7,050
682	7,056
833	7,083
1,019	7,142
1,250	7,243
1,548	7,408
1,944	7,666
2,500	8,080
3,333	8,766
4,722	10,003
7,500	12,623
15,833	20,791
∞	∞

Dynamic programming, via a sequence of six steps, determines optimum heat exchanger area—Fig. 2



1. Selection of arbitrary T_1 values enables calculation of best corresponding values for the temperature T_2.

2. Fixing of T_1 also determines a corresponding area, A_3, for heat exchanger E-3.

3. Area required for exchanger E-2 can also be calculated for each assumed value of the temperature T_1.

4. Result of two-stage optimization shows sum of areas A_2 and A_3 for each T_1 (final optimum shown).

5. For each T_1, it is now possible to calculate the area required for the heat exchanger E-1.

6. Result of three-stage optimization shows which T_1 gives minimum total exchanger area, solving the problem.

best T_1 and T_2 for the over-all optimum. This is found to be: $T_1 = 180$ *F.*, $T_2 = 295$ *F.*, which gives a total area $(A_1 + A_2 + A_3)$ of 7,050 sq. ft.

If we assume 12 values of T_2 for each T_1 (as in Table II), then we have considered 12 values of T_2 for each of 17 values of T_1 (Table III), and have therefore selected the best case from among 204 (i.e., 12 × 17) possibilities.

Although it was possible to work out this problem entirely with tables, the corresponding graphical solution is shown in Fig. 2. The reader will note that the tabular solution gives the optimum temperature only within ten degrees.

Pros and Cons of Dynamic Programming

A few comments about dynamic programming seem appropriate at this point. Note that dynamic programming does not indicate which particular optimization technique to use at any one stage. This is left up to the investigator.

The problem of constraints usually hinders most optimization techniques. In dynamic programming, however, the constraints are actually helpful because they limit the range to be investigated. In the heat exchanger problem just considered, for example, T_1 is limited to 100-300 F., and hence the values to be studied must be restricted to this range. Discontinuities in the objective function can be handled by dynamic programming because no analytical function need be used; tables and curves are quite adequate.

In general, where there are N stages and k decisions to be made at each stage, the over-all optimization problem involves k^N possible answers. For a five-stage problem with three decisions at each stage, this means 3^5 (243) possible combinations. By dynamic programming, only 5 × 3 (15) one-stage optimization problems have to be solved.

One disadvantage of the method is the problem of dimensionality (many variables), which arises in the optimization at each step. For example, when many decisions are to be made at one time, the single-step optimization itself may become quite complex.

Over-all, however, the method of dynamic programming has found many applications. Aris[6] has applied this technique to chemical reactor design as well as to extraction, while Rudd and Blum[10] extended this work to extraction with product recycle. Dranoff et al[8] used dynamic programming to solve a problem in optimal design of a complex chemical plant. Westbrook[11] considered problems in heat transfer, while Mitten and Nemhauser[9] have illustrated the use of dynamic programming in processes containing branches and loops.

In Conclusion

Several optimization methods that have not been included in this series include random methods, the critical path method, analog methods, and applications of the calculus of variations.

The state of the art in optimization was summed up quite well by C. Storey[12] in his comments regarding the symposium on process optimization held in London in 1962. He noted that we are in a transition state in the field of optimization and that theory is outpacing application. We must wait, he said, for more applications before we can determine which method applies to which problem.

If interest in optimization has been stimulated in the neophyte—if some techniques have been made clearer or some new methods supplied to the reader acquainted with the field—then this series has achieved its objective.

References

Linear Programming

1. Baumol, W. J., "Economic Theory and Operations Analysis," Prentice-Hall, Englewood Cliffs, N. J., 1961.
2. Charnes, A., Cooper, W. W., Henderson, A., "An Introduction to Linear Programming," Wiley, New York, 1953.
3. Dantzig, G. B., "Activity Analysis of Production and Allocation," Chap. 21, Wiley, New York, 1951.
4. Gass, S. I., "Linear Programming: Methods and Applications," McGraw-Hill, New York, 1958.
5. Vajda, S., "Theory of Games and Linear Programming," Wiley, New York, 1956.

Dynamic Programming

6. Aris, R., Rudd, D. F., Amundson, N. R., *Chem. Eng. Sci.*, **12**, 88 (1960).
7. Bellman, R., "Dynamic Programming," Princeton University Press, Princeton, N. J., 1957.
8. Dranoff, J. S., et al., *Operations Research*, **9**, 388 (1961).
9. Mitten, L. G., Nemhauser, G. L., *Chem. Eng. Progr.*, **59**, No. 1, 53 (1963).
10. Rudd, D. F., Blum, E. D., *Chem. Eng. Sci.*, **17**, 277 (1962).
11. Westbrook, G. T., *Hydro Proc. & Pet. Ref.*, **40**, No. 9, 201 (1961).

General

12. Storey, C., *Brit. Chem. Eng.*, **8**, No. 1 (1963).

Meet the Author

ARNOLD H. BOAS *is associate engineer in the Engineering Dept. of Socony Mobil Oil Co., N. Y. He is engaged in mathematical analysis, computer applications and optimization theory in the field of chemical engineering. His experience includes process design and development work at Hydrocarbon Research, Inc., Chemical Construction Corp. and American Cyanamid Co.*

He holds a B.Ch.E. from CCNY, an M.S. from the University of Cincinnati and a Ph.D. in chemical engineering from the Polytechnic Institute of Brooklyn.

Dr. Boas is a member of Tau Beta Pi, Sigma Xi, Phi Lambda Upsilon, AIChE, and is a PE in New York state.

Key Concepts for This Article

For indexing details, see Chem. Eng., Jan. 7, 1963, p. 73 (Reprint No. 222). Words in bold are role indicators; numbers correspond to AIChE system.

Active (8)	Passive (9)	Means/Methods (10)
Optimization	Objective functions	Programming
Explorations	Heat exchangers	Linear
	Variable, dependent	Dynamic

best T_1 and T_2 for the over-all optimum. This is found to be: T_1 = 180 F., T_2 = 295 F., which gives a total area ($A_1 + A_2 + A_3$) of 7,050 sq. ft.

If we assume 12 values of T_2 for each T_1 (as in Table II), then we have considered 12 values of T_2 for each of 17 values of T_1 (Table III), and have therefore selected the best case from among 204 (i.e., 12 × 17) possibilities.

Although it was possible to work out this problem entirely with tables, the corresponding graphical solution is shown in Fig. 2. The reader will note that the tabular solution gives the optimum temperature only within ten degrees.

Pros and Cons of Dynamic Programming

A few comments about dynamic programming seem appropriate at this point. Note that dynamic programming does not indicate which particular optimization technique to use at any one stage. This is left up to the investigator.

The problem of constraints usually hinders most optimization techniques. In dynamic programming, however, the constraints are actually helpful because they limit the range to be investigated. In the heat exchanger problem just considered, for example, T_1 is limited to 100-300 F., and hence the values to be studied must be restricted to this range. Discontinuities in the objective function can be handled by dynamic programming because no analytical function need be used; tables and curves are quite adequate.

In general, where there are N stages and k decisions to be made at each stage, the over-all optimization problem involves k^N possible answers. For a five-stage problem with three decisions at each stage, this means 3^5 (243) possible combinations. By dynamic programming, only 5 × 3 (15) one-stage optimization problems have to be solved.

One disadvantage of the method is the problem of dimensionality (many variables), which arises in the optimization at each step. For example, when many decisions are to be made at one time, the single-step optimization itself may become quite complex.

Over-all, however, the method of dynamic programming has found many applications. Aris[6] has applied this technique to chemical reactor design as well as to extraction, while Rudd and Blum[10] extended this work to extraction with product recycle. Dranoff et al[8] used dynamic programming to solve a problem in optimal design of a complex chemical plant. Westbrook[11] considered problems in heat transfer, while Mitten and Nemhauser[9] have illustrated the use of dynamic programming in processes containing branches and loops.

In Conclusion

Several optimization methods that have not been included in this series include random methods, the critical path method, analog methods, and applications of the calculus of variations.

The state of the art in optimization was summed up quite well by C. Storey[12] in his comments regarding the symposium on process optimization held in London in 1962. He noted that we are in a transition state in the field of optimization and that theory is outpacing application. We must wait, he said, for more applications before we can determine which method applies to which problem.

If interest in optimization has been stimulated in the neophyte—if some techniques have been made clearer or some new methods supplied to the reader acquainted with the field—then this series has achieved its objective.

References

Linear Programming

1. Baumol, W. J., "Economic Theory and Operations Analysis," Prentice-Hall, Englewood Cliffs, N. J., 1961.
2. Charnes, A., Cooper, W. W., Henderson, A., "An Introduction to Linear Programming," Wiley, New York, 1953.
3. Dantzig, G. B., "Activity Analysis of Production and Allocation," Chap. 21, Wiley, New York, 1951.
4. Gass, S. I., "Linear Programming: Methods and Applications," McGraw-Hill, New York, 1958.
5. Vajda, S., "Theory of Games and Linear Programming," Wiley, New York, 1956.

Dynamic Programming

6. Aris, R., Rudd, D. F., Amundson, N. R., *Chem. Eng. Sci.*, 12, 88 (1960).
7. Bellman, R., "Dynamic Programming," Princeton University Press, Princeton, N. J., 1957.
8. Dranoff, J. S., et al., *Operations Research*, 9, 388 (1961).
9. Mitten, L. G., Nemhauser, G. L., *Chem. Eng. Progr.*, 59, No. 1, 53 (1963).
10. Rudd, D. F., Blum, E. D., *Chem. Eng. Sci.*, 17, 277 (1962).
11. Westbrook, G. T., *Hydro Proc. & Pet. Ref.*, 40, No. 9, 201 (1961).

General

12. Storey, C., *Brit. Chem. Eng.*, 8, No. 1 (1963).

Meet the Author

ARNOLD H. BOAS *is associate engineer in the Engineering Dept. of Socony Mobil Oil Co., N. Y. He is engaged in mathematical analysis, computer applications and optimization theory in the field of chemical engineering. His experience includes process design and development work at Hydrocarbon Research, Inc., Chemical Construction Corp. and American Cyanamid Co.*

He holds a B.Ch.E. from CCNY, an M.S. from the University of Cincinnati and a Ph.D. in chemical engineering from the Polytechnic Institute of Brooklyn.

Dr. Boas is a member of Tau Beta Pi, Sigma Xi, Phi Lambda Upsilon, AIChE, and is a PE in New York state.

Key Concepts for This Article

For indexing details, see Chem. Eng., Jan. 7, 1963, p. 73 (Reprint No. 222). Words in bold are role indicators; numbers correspond to AIChE system.

Active (8)	**Passive (9)**	**Means/Methods (10)**
Optimization	Objective functions	Programming
Explorations	Heat exchangers	Linear
	Variable, dependent	Dynamic

A REPRINT FROM

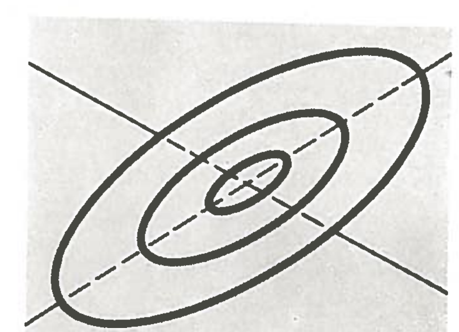
CHEMICAL ENGINEERING

BOA
1962

MODERN MATHEMATICAL TOOLS FOR

Optimization

ARNOLD H. BOAS



Reprinted from CHEMICAL ENGINEERING