METHOD OF LAGRANGE MULTIPLIERS AND THE KUHN-TUCKER CONDITIONS

613-8301

by

PRAMOD KUMAR GUPTA

B.S., BIRLA INSTITUTE OF TECHNOLOGY, INDIA, 1969

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1973

Approved by:

Ching-Lai Hwang

Major Professor

LD
2668
R4
1973
G86
c 2
Document

TABLE OF CONTENTS

0. INTRODUCTION

The general programming problem consists of determining values of n variables $x_1, \ldots, x_n$ which optimize the function, $S = f(x_1, \ldots, x_n)$, and satisfy m constraints given by $g_j(x_1, \ldots, x_n) \{\leq, =, \geq\} b_j$, $j = 1, \ldots, m$. A particular case of the programming problem is a linear programming problem which seeks to determine non-negative values of n variables, $x_i \geq 0$, $i = 1, \ldots, n$, which optimize the linear function, $S = \sum_i c_i x_i$, and satisfy the m linear constraints, $\sum_i a_{ji} x_i \{\leq, =, \geq\} b_j$, $j = 1, \ldots, m$. All programming problems that are not linear are called nonlinear programming problem. Although the class of nonlinear programming problems which has been studied most extensively is that where the constraints are linear and the objective function is nonlinear, the other classes of nonlinear programming problems are that the constraints are nonlinear and the objective function is linear, and that both the constraints and the objective function are nonlinear.

A number of algorithms have been proposed for the solution of the general nonlinear programming problem. However, only a few have been demonstrated to be effective when applied to large-scale nonlinear programming problems, and none of the algorithms has proved to be superior that it can be classified as a universal algorithm for general nonlinear programming problems (Himmelblau, 1972). No general method exists to solve nonlinear programming problems in the sense that the simplex algorithm exists to solve linear programming problems. We may recall that linear programming problems have the following properties (Hadley, 1964).

(1) The set of feasible solutions which satisfies the constraints and the non-negativity restrictions is a convex set. This convex set has a finite number of corners which are referred to as extreme points. (2) The set of all n variables (x_1, ..., x_n), which yield a specific value of the objective function is a hyper-plane, and the hyperplanes corresponding to different values of the objective function are parallel (3) A local maximum or minimum is also the global maximum or minimum of the objective function over the set of feasible solutions. (4) If the optimal value of the objective function is bounded, the optimal solution will be one of the extreme points of the convex set of the feasible solution. Furthermore, if we start the search of the optimal solution at any extreme point of the convex set of feasible solutions, we will reach the optimal extreme point in a series of steps such that at each step we moves only to an adjacent extreme point. No efficient superior algorithm for nonlinear programming problems exists because contrary to linear programming problems for any given nonlinear programming problems, some or all of these features which characterize linear programming problems may be violated.

When the general nonlinear programming problem has that (1) no inequalities appear in the constraints, (2) there are no non-negativity or discreteness restrictions on the variables, (3) the number of equality constraints, m, is less than the number of variables, n, that is, $m < n$, and (4) the objective function, $f(x)$, and the functions in the equality constraints, $g_j(x)$, $j = 1, \ldots, m$ are continuous and possess partial derivatives at least through second order, the problem can be solved by the method of Lagrange multiplier. This classical method is of use mainly in theoretical analyses, and is not, in general, well suited for numerical calculation. The method of Lagrange multipliers can be generalized to handle problems involving inequality constraints and non-negative variables. The necessary conditions for optimizing these problems are the Kuhn-Tucker conditions. Although the theory related to the method of the Lagrange multipliers and the Kuhn-Tucker conditions is not directly concerned with computational techniques, it has been of fundamental importance in developing a numerical procedures for solving nonlinear programming problems, for example, quadratic programming problem.

The method of Lagrange multipliers and the Kuhn-Tucker conditions have been studied extensively by many investigators. In 1951, Kuhn and Tucker published an important paper "Nonlinear programming" dealing with necessary and sufficient conditions for optimal solutions to programming problems, which laid the foundations for a great deal of later work in nonlinear programming. These conditions are known as the Kuhn-Tucker conditions in their honour. They introduced the concept of constrained qualification and rejects certain stationary points as possible optima. Generalization of their theoretical work by other authors appeared later in different books [Arrow, Hurwicz, and Uzawa (1958), Mangasarian (1969)] and papers. Samuelson (1955) described the necessary and sufficient conditions for local optimum of a nonlinear objective function subject to nonlinear equality constraints. He derived these conditions from Taylor's series expansion for single and multi equality constraints. Tucker (1957) presented the use of Lagrange multiplier technique for minimizing a convex objective function subject to linear constraints. He then considered the special case of quadratic programming problems and showed that it is possible to solve them by linear computations. Again Wolfe (1959) introduced, based on the Kuhn-Tucker conditions, the modified simplex method for quadratic programming problems which is widely used and is simple to apply. Some other authors also derived the necessary and sufficient conditions for different cases of nonlinear programming problem from the Kuhn-Tucker conditions: the necessary and sufficient conditions for maximizing an objective function subject to inequality constraints less than zero was given by Dorfman (1958); those for the minimization of an objective function subject to equality and inequality constraints and non-negative of variables by Vajda (1961) and Wilde (1962); those for optimizing a function subject to inequality constriants by Dorn (1963).

Teichroew (1964) summarized all these cases in the table form in his book. Some other authors such as Beveridge and Schechter (1970), Macmillan (1970) and Zangwill (1970) have also described the Lagrange multiplier technique and the Kuhn-Tucker conditions in their books. Beveridge and Schechter (1970) has given a good explaination of the meaning of the sign of Lagrange multipliers which was origianlly covered by Kuhn-Tucker (1951). Hadley (1964), in his book describes the duel problem associated with optimizing an objective function subject to the constraints and the definitions of saddle point.

This report is based upon and an extension of Dr. C. L. Hwang's lecture notes of past two years. The purpose of this report is to give a self-contained, brief and illustrative description of the classical optimization procedures for solving nonlinear programming problems.

Firstly the Lagrangian multiplier technique is presented in Sections 1 through 4 for optimizing the nonlinear objective functions subject to equality constraints. The necessary and sufficient conditions are developed from Taylor's series expansion and numerical examples are solved to illustrate the technique.

The Kuhn-Tucker necessary conditions for optimizing an objective function subject to inequality constraints are presented in Sections 5 and 6, and are illustrated by a numerical example in Section 7. The Wolfe's simplex method for optimal solution of quadratic programming problems which is based on the Kuhn-Tucker conditions is presented in Section 8.

The concept of convexity and concavity of functions, and of convex set, which is basis for determining whether a local optimum is also a global optimal, is presented in Section 9. Section 10 contains the necessary and sufficient conditions for a global optimum.

Finally the dual problem and saddle point and the necessary and sufficient conditions for saddle points are presented in Sections 11 and 12.

1. METHOD OF LAGRANGE MULTIPLIER

A general nonlinear programming problem is usually composed of finding a set of x which minimize (or maximize)

$$S = f(x_1, x_2, \ldots, x_n) \tag{1-1}$$

subject to m equality constraints

$$g_j(x_1, x_2, \ldots, x_n) = 0, \qquad j = 1, \ldots, m \tag{1-2}$$

The optimal solution of the problem can be obtained by the method of Lagrange multiplier.

Let us consider a two-dimensional problem first. The necessary conditions for a point (x_1, x_2) to be the extreme value of

$$S = f(x_1, x_2) \tag{1-3}$$

subject to the equality constraint

$$g(x_1, x_2) = 0 \tag{1-4}$$

are obtained by defining an unconstrained function (Lagrangian)

$$L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda g(x_1, x_2) \tag{1-5}$$

where λ is the Lagrange multiplier.

Now by setting the partial derivatives of $L(x_1, x_2, \lambda)$ with respect to x_1, x_2, and λ equal to zero, we obtain the necessary conditions of optimality as follows

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \tag{1-6}$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0 \tag{1-7}$$

$$-\frac{\partial L}{\partial \lambda} = g(x_1, x_2) = 0 \tag{1-8}$$

By solving the above equations for x_1, x_2, and λ we obtain the stationary point or extremum for the function $f(x_1, x_2)$.

We may generalize the method of Lagrange multiplier by considering the objective function given by equation (1-1) subject to m equality constraints given by equation (1-2). The optimal solution can be obtained by setting the partial derivatives of the Lagrangian function

$$L(x, \lambda) = f(x_1, \ldots, x_n) - \sum_{j=1}^{m} \lambda_j \, g_j(x_1, \ldots, x_n) \tag{1-9}$$

with respect to $x_1, \ldots, x_n$ and $\lambda_1, \ldots, \lambda_m$ equal to zero.

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \qquad i = 1, \ldots, n \tag{1-10}$$

$$-\frac{\partial L}{\partial \lambda_j} = g_j(x_1, \ldots, x_n) = 0, \qquad j = 1, \ldots, m \tag{1-11}$$

The points obtained by solving the above equations will be the stationary points.

2. GEOMETRIC INTERPRETATION OF THE METHOD OF LAGRANGE MULTIPLIER

Figure 1 shows a contour lines of a function $f(x_1, x_2)$ and a constraint $g(x_1, x_2) = 0$.

The extreme value of f, which satisfies the constraint $g(x_1, x_2) = 0$, is the one at which a contour line of the function $f(x_1, x_2)$ touches $g(x_1, x_2) = 0$. At this point, say $(\underline{x}_1, \underline{x}_2)$, the tangents of the two curves must have the same slope. Now the slope of the tangent is given by the ratio of the two partial derivatives of a function. Therefore at $(\underline{x}_1, \underline{x}_2)$, the slope of the tangent is given by,

$$\frac{f_{x_1}}{f_{x_2}} = \frac{g_{x_1}}{g_{x_2}} \tag{2-1}$$

where $f_{x_1} = \frac{\partial f}{\partial x_1}$, $f_{x_2} = \frac{\partial f}{\partial x_2}$, $g_{x_1} = \frac{\partial g}{\partial x_1}$, and $g_{x_2} = \frac{\partial g}{\partial x_2}$.

Equation (2-1) can be written as

$$\frac{f_{x_1}}{g_{x_1}} = \frac{f_{x_2}}{g_{x_2}} \tag{2-2}$$

and let this common ratio be denoted by a undetermined constant λ, known as Lagrange multiplier.

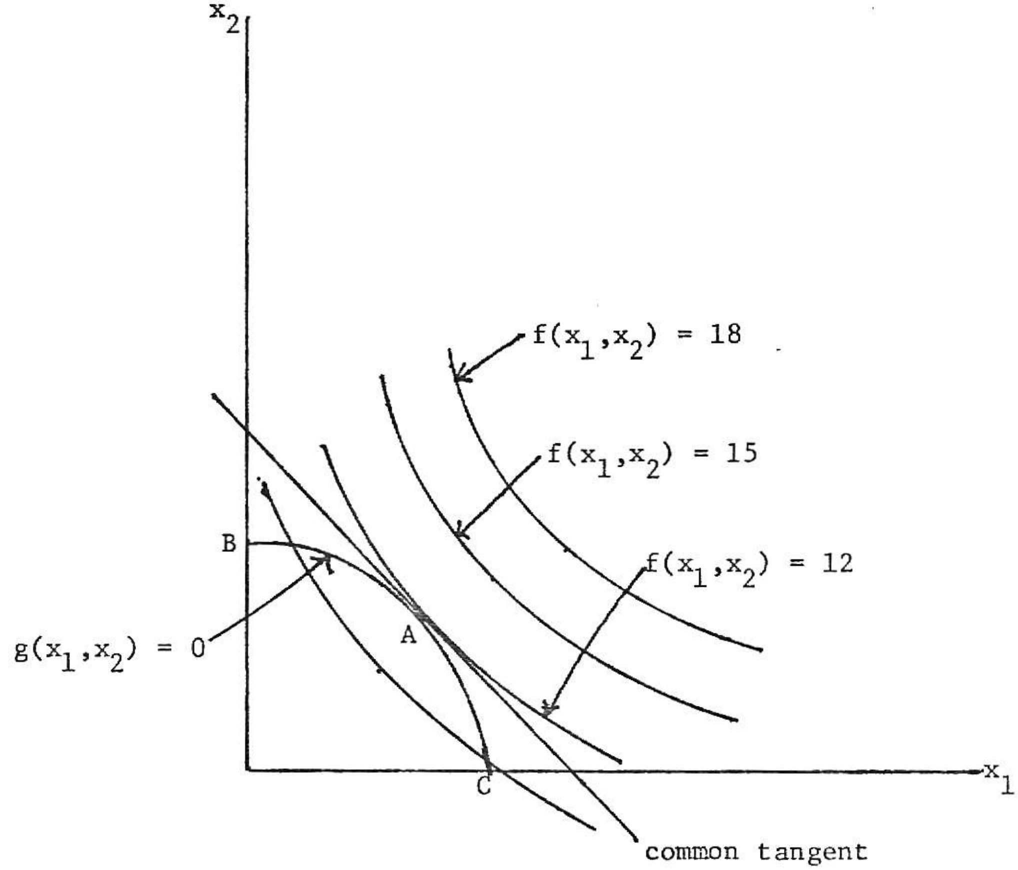


Fig. 1 Contour lines for a function $f(x_1,x_2)$ subject to a constraint $g(x_1,x_2) = 0$

THIS BOOK CONTAINS NUMEROUS PAGES WITH THE ORIGINAL PRINTING ON THE PAGE BEING CROOKED.

THIS IS THE BEST IMAGE AVAILABLE.

Then we can write equation (2-2) as

$$\frac{f_{x_1}}{g_{x_1}} = \frac{f_{x_2}}{g_{x_2}} = \lambda \tag{2-3}$$

or

$$f_{x_1} - \lambda g_{x_1} = 0 \tag{2-4}$$

and

$$f_{x_2} - \lambda g_{x_2} = 0 \tag{2-5}$$

which are the same equations as equations (1-6) and (1-7). Solutions of the above equations give the stationary points.

Figure 1 shows that the stationary point (point A) obtained is a maximum point for the function given by equation (1-3) subject to the constraint, given by equation (1-4). This information is not supplied to us by equations (2-4) and (2-5). Moreover the minimum point for the function f subject to the constraint, $g = 0$, will be either point B or point C, which is also not provided by the method of Lagrange multiplier. Thus the equations (2-4) and (2-5) give us only the stationary point and hence are the necessary conditions only. For knowing exactly whether this stationary point is a local maximum, or local minimum or neither, we will have to check the sufficient conditions discussed in the next section.

Note that it has been assumed that both f and g are differentiable in the neighbourhood of $(\underline{x}_1, \underline{x}_2)$ and that f_{x_2} and g_{x_2} are not zero at this point.

3. NECESSARY AND SUFFICIENT CONDITIONS OF OPTIMALITY

A stationary point, obtained by the method of Lagrange multiplier, may be either a local maximum point, a local minimum point or a saddle point. The necessary and sufficient conditions for a stationary point to be a maximum or minimum can also be developed by the expansion of Taylor's series for the function $f(x_1, x_2)$ as follows

$$f(x_1 + \Delta x_1, x_2 + \Delta x_2) =$$

$$= f(x_1, x_2) + df + \frac{1}{2!} d^2f + \frac{1}{3} d^3f + \ldots \qquad (3\text{-}1)$$

The necessary condition of optimality is that at an extreme point df is zero. Since x_1 and x_2 must satisfy the relation

$$g(x_1, x_2) = 0$$

the differential dg must be zero, that is,

$$\frac{\partial g}{\partial x_1} dx_1 + \frac{\partial g}{\partial x_2} dx_2 = 0 \qquad (3\text{-}2)$$

or

$$dx_1 = - \left(\frac{\partial g}{\partial x_2} / \frac{\partial g}{\partial x_1} \right) dx_2 \qquad (3\text{-}3)$$

if $\frac{\partial g}{\partial x_1} \neq 0$.

Substituting the value of dx_1 given by equation (3-3) into the differential

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 = 0 \qquad (3\text{-}4)$$

gives

$$df = \left[\left(-\frac{\partial f}{\partial x_1}\ \frac{\partial g}{\partial x_2} \Big/ \frac{\partial g}{\partial x_1}\right) + \frac{\partial f}{\partial x_2}\right] dx_2 = 0 \tag{3-5}$$

Since dx_2 can not be zero, a necessary condition for df to be zero in equation (3-5) is

$$\left(-\frac{\partial f}{\partial x_1}\ \frac{\partial g}{\partial x_2} \Big/ \frac{\partial g}{\partial x_1}\right) + \frac{\partial f}{\partial x_2} = 0$$

or

$$\frac{\partial f}{\partial x_1} \Big/ \frac{\partial g}{\partial x_1} = \frac{\partial f}{\partial x_2} \Big/ \frac{\partial g}{\partial x_2} \tag{3-6}$$

Let this common ratio be denoted by λ,

$$\frac{\partial f}{\partial x_1} \Big/ \frac{\partial g}{\partial x_1} = \frac{\partial f}{\partial x_2} \Big/ \frac{\partial g}{\partial x_2} = \lambda \tag{3-7}$$

we have

$$\frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \tag{3-8}$$

$$\frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0 \tag{3-9}$$

which are the same equations given by equations (1-6) and (1-7) for finding the stationary points.

The sufficient conditions for a stationary point to be a maximum, the term d^2f in Taylor's series must be negative. Assuming that x_1 and x_2 are not independent, the second differential of the function, f, is

$$d^2f = d(f_{x_1}\, dx_1 + f_{x_2}\, dx_2) \tag{3-10}$$

where

$$f_{x_1} = \frac{\partial f}{\partial x_1} \tag{3-11}$$

and

$$f_{x_2} = \frac{\partial f}{\partial x_2} \tag{3-12}$$

therefore, we have

$$d^2f = d(f_{x_1})\, dx_1 + f_{x_1}\, d(dx_1)$$

$$+ d(f_{x_2})\, dx_2 + f_{x_2}\, d(dx_2) \tag{3-13}$$

$$= (f_{x_1x_1}\, dx_1 + f_{x_1x_2}\, dx_2)\, dx_1 + f_{x_1}\, d^2x_1$$

$$+ (f_{x_1x_2}\, dx_1 + f_{x_2x_2}\, dx_2)\, dx_2 + f_{x_2}\, d^2x_2 \tag{3-14}$$

$$= f_{x_1x_1}\, (dx_1)^2 + 2f_{x_1x_2}\, dx_1dx_2 +$$

$$f_{x_2x_2}\, (dx_2)^2 + f_{x_1}\, d^2x_1 + f_{x_2}\, d^2x_2 \tag{3-15}$$

where

$$f_{x_ix_j} = \frac{\partial^2 f}{\partial x_i\, \partial x_j}\ , \quad i,\ j = 1,\ 2$$

Similarly from equation (3-2), we obtain

$$d^2g = g_{x_1x_1}(dx_1)^2 + 2g_{x_1x_2}\,dx_1dx_2 + g_{x_2x_2}(dx_2)^2 +$$

$$g_{x_1}d^2x_1 + g_{x_2}d^2x_2 = 0 \tag{3-16}$$

Now solving equation (3-16) for d^2x_1, substituting the value of d^2x_1 so obtained in equation (3-15), and collecting the terms gives

$$d^2f = \left(f_{x_1x_1} - \frac{f_{x_1}}{g_{x_1}}g_{x_1x_1}\right)(dx_1)^2 + 2\left(f_{x_1x_2} - \frac{f_{x_1}}{g_{x_1}}g_{x_1x_2}\right)dx_1\,dx_2 +$$

$$\left(f_{x_2x_2} - \frac{f_{x_1}}{g_{x_1}}g_{x_2x_2}\right)(dx_2)^2 + \left(f_{x_2} - \frac{f_{x_1}}{g_{x_1}}g_{x_2}\right)d^2x_2 \tag{3-17}$$

In the above equation, the last term is zero because the factor in the brackets is df (see equation (3-5)). Substituting the value of dx_1 from equation (3-3) and using λ as defined previously, we have

$$\lambda = \frac{f_{x_1}}{g_{x_1}} = \frac{f_{x_2}}{g_{x_2}} \tag{3-18}$$

and

$$d^2f = -\frac{dx_2^2}{dx_1^2}\Delta_3 \tag{3-19}$$

where

$$\Delta_3 = - \left\{ g_{x_2}^{\ 2} \left[f_{x_1 x_1} - \lambda g_{x_1 x_1} \right] - 2 g_{x_1} g_{x_2} \left[f_{x_1 x_2} - \lambda g_{x_1 x_2} \right] + g_{x_1}^{\ 2} \left[f_{x_2 x_2} - \lambda g_{x_2 x_2} \right] \right\} \tag{3-20}$$

Thus the function $f(x_1, x_2)$ will have a maximum (or minimum) subject to the constraint

$$g(x_1, x_2) = 0$$

if Δ_3 is positive (or negative).

If at a point both the first and second derivative vanish, then we must examine higher order terms in order to develope sufficient conditions for optimality. In other words if Δ_3 is zero, then higher differentials of the function $f(x_1, x_2)$ should be examined to develope sufficient conditions.

The above development is only for two variables with one equality constraint. The development can be generalized to the case of multivariables with more than one constraints, however, the algebra becomes tedious and the results are not practical. The results are presented by Samuelson [1955] and summarized in Teichroew [1964].

4. NUMERICAL EXAMPLES

Example 1: Find a rectangle with maximum area among all the rectangles subject to the constraint of having constant perimeter.

If x_1 is the length of one side of rectangle and x_2 is the length of other side of rectangle, then the problems is to maximize the function

$$f(x_1, x_2) = x_1 x_2 \tag{4-1}$$

subject to

$$2x_1 + 2x_2 = c \tag{4-2}$$

where c is a constant.

We will solve this example firstly by eliminating the constraint and using classical calculus method and then by the method of Lagrange multiplier.

METHOD 1: CLASSICAL DIFFERENTIAL CALCULUS METHOD

In this method, variable x_2 will be eliminated by solving equation (4-2) as follows

$$x_2 = \frac{1}{2}(c - 2x_1) \tag{4-3}$$

Substituting the value of x_2 in equation (4-1), we obtain,

$$f(x_1) = \frac{x_1}{2} c - x_1^2$$

where x_1 needs to be non-negative. Now, setting the first derivative of $f(x_1)$ with respect to x_1 equal to zero, which is the necessary condition for a stationary point, we have

$$\frac{df}{dx_1} = \frac{c}{2} - 2x_1 = 0 \tag{4-4}$$

or

$$x_1 = \frac{c}{4} \tag{4-5}$$

Putting the value of x_1 in equation (4-3) yields

$$x_2 = \frac{c}{4} \tag{4-6}$$

and therefore $f = \frac{c^2}{16}$, which gives the optimum value of the function.

To check whether the stationary point is a maximum or not we have to check the sign of the second derivative of the function with respect to x_1. If it is negative, it means that the stationary point is the maximum point and if it is positive, it means that the stationary point is the minimum point. Since

$$\frac{d^2f}{dx_1^2} = -2 < 0 \tag{4-7}$$

is negative, the stationary point is a maximum. Therefore the rectangle which has the maximum area for a given perimeter will be the one which has both sides equal or

$$x_1 = x_2$$

METHOD 2 - LAGRANGIAN MULTIPLIER METHOD

The above example will now be solved by the method of Lagrangian multiplier. Again, we have to maximize the function

$$f(x_1, x_2) = x_1x_2$$

subject to the constraint

$$g(x_1, x_2) = 2x_1 + 2x_2 - c = 0$$

The Lagrangian function for the above problem will be

$$L(x_1, x_2, \lambda) = x_1 x_2 - \lambda(2x_1 + 2x_2 - c) \tag{4-8}$$

Differentiating the Lagrangian function with respect to x_1, x_2 and λ gives

$$\frac{\partial L}{\partial x_1} = x_2 - 2\lambda = 0 \tag{4-9}$$

$$\frac{\partial L}{\partial x_2} = x_1 - 2\lambda = 0 \tag{4-10}$$

$$-\frac{\partial L}{\partial \lambda} = 2x_1 + 2x_2 - c = 0 \tag{4-11}$$

Solving equation (4-9) for x_2 gives

$$x_2 = 2\lambda \tag{4-12}$$

Solving equation (4-10) for x_1 gives,

$$x_1 = 2\lambda \tag{4-13}$$

Substituting the values of x_2 and x_1 from equations (4-12) and (4-13) in equation (4-11) gives

$$8\lambda = c$$

or

$$\lambda = c/8$$

Substituting the value of λ in equations (4-9) and (4-10) gives,

$$x_1 = c/4$$

$$x_2 = c/4$$

which gives the same result as obtained earlier by the method of classical calculus.

Again a stationary point, obtained by the method of Lagrange multiplier, may be either a local maximum point, a local minimum or neither. To find whether the point obtained above is a maximum or minimum, we will have to check the sufficient conditions discussed in Section 3, where the value of Δ_3 is calculated as

$$\Delta_3 = -\left\{g_{x_2}^{\ 2}\,[f_{x_1x_1} - \lambda g_{x_1x_1}] - 2g_{x_1}g_{x_2}\,[f_{x_1x_2} - \lambda g_{x_1x_2}] + g_{x_1}^{\ 2}\,[f_{x_2x_2} - \lambda g_{x_2x_2}]\right\}$$

For the stationary point to be a maximum, the value of Δ_3 should be positive. Now the various variables in the above equation for this example are

$$f_{x_1} = x_2, \quad f_{x_1x_1} = 0, \quad g_{x_1} = 2, \quad g_{x_1x_1} = 0,$$

$$f_{x_2} = x_1, \quad f_{x_2x_2} = 0, \quad g_{x_2} = 2, \quad g_{x_2x_2} = 0,$$

$$f_{x_1x_2} = 1, \quad g_{x_1x_2} = 0$$

By substituting the above values in the equation for Δ_3, we have

$\Delta_3 = 8$

which is positive and so the stationary point is a maximum point when $x_1 = x_2 = c/4$.

Thus, we see that the results obtained by the method of Lagrange multiplier are just the same as that obtained by the method of classical calculus. But the method of Lagrange multiplier has the advantage that it supplies the value of λ also alongwith the values of x_1 and x_2. Moreover when the number of variable and constraints increases more than two, it becomes difficult to eliminate the variables in the method of classical calculus.

Example 2: Find the dimensions of a closed cylindrical tank of given volume V which has the minimum surface area A.

Solution:

Let us define

R = the radius of the cylinder

H = the hight of the cylinder.

Then the problem is to minimize

$$f = A(R,H) = 2\pi R^2 + 2\pi RH$$

subject to the constraint

$$g(R,H) = \pi R^2 H - V = 0$$

where V is given.

The Lagrangian function is

$$L(R,H,\lambda) = (2\pi R^2 + 2\pi RH) - \lambda\ (\pi R^2 H - V)$$

Differentiating the function with respect to R and H and equating to zero yields

$$\frac{\partial L}{\partial R} = 4\pi R + 2\pi H - 2\lambda\pi RH = 0$$

$$\frac{\partial L}{\partial H} = 2\pi R - \pi\lambda R^2 = 0$$

Combining these equations with the constraint equation

$$g(R,H) = -\frac{\partial L}{\partial \lambda} = \pi R^2 H - V = 0$$

gives us three unknowns with three simultaneous equations. The solutions of these equations give two sets of the stationary points. They are

$$\lambda = \frac{2}{R} \text{ or } \lambda = \frac{4}{H}$$

$$(R,H) = (0,0)$$

and

$$(R,H) = \left(\left(\frac{V}{2\pi}\right)^{1/3},\ 2\left(\frac{V}{2\pi}\right)^{1/3} \right)$$

The first stationary point, $(R,H) = (0,0)$, contradicts the restriction regarding the fixed size of the volume $(V \neq 0)$ except for the trivial case $V = 0$. This example illustrates that the derived stationary point of the Lagrangian function must always be checked to ensure that they do satisfy the system constraints.

The second stationary point does give the minimum area of

$$A = 6\Pi\left(\frac{V}{2\pi}\right)^{2/3}$$

The character of this stationary point can be examined by checking the sufficient condition discussed in Section 3 using equations (3-19) and (3-20). For the stationary point to be a minimum, the value of Δ_3 given by equation (3-20) should be negative. The values of terms in equation (3-20) for this example are

$$f_{x_1} = f_R = 4\pi R + 2\pi H, \qquad f_{x_2} = f_H = 2\pi R$$

$$g_{x_1} = g_R = 2\pi RH, \qquad g_{x_2} = g_H = \pi R^2$$

$$f_{x_1x_1} = f_{RR} = 4\pi, \qquad f_{x_1x_2} = f_{RH} = 2\pi$$

$$f_{x_2x_2} = f_{HH} = 0, \qquad g_{x_1x_1} = g_{RR} = 2\pi H$$

$$g_{x_1x_2} = g_{RH} = 2\pi R, \qquad g_{x_2x_2} = g_{HH} = 0$$

Then

$$\Delta_3 = -\ 12\pi^3 \left(\frac{V}{2\pi}\right)^{4/3}$$

which is negative so that the stationary point is a minimum.

The character of the stationary point, however, should not be examined using the Lagrangian function (an unconstrained function). For example, applying Sylvester's theorem to the Lagrangian function at this stationary point gives

$$\frac{\partial^2 L}{\partial R^2} = 4\pi - 2\pi\lambda H$$

$$= 4\pi - 2\pi\left(+ \frac{4}{H}\right)H$$

$$= - 4\pi$$

$$\frac{\partial^2 L}{\partial R \partial H} = 2\pi - 2\pi\lambda R$$

$$= 2\pi - 2\pi\left(+ \frac{2}{R}\right)R$$

$$= - 2\pi$$

$$\frac{\partial^2 L}{\partial H^2} = 0$$

$$A = h_{11} = \frac{1}{2}\frac{\partial^2 L}{\partial R^2} = - 2\pi$$

$$B = h_{12} = h_{21} = \frac{1}{2}\frac{\partial^2 L}{\partial R \partial H} = -\pi$$

$$C = h_{22} = 0$$

Since

$$A = \left| h_{11} \right| = -\ 2\pi < 0$$

and

$$AC - B^2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} = -\ \pi^2 < 0$$

which represents a saddle point in the Lagrangian with respect R and H at this stationary point. No information is therefore available about the character of the stationary point in f when using only the method of Lagrangian multipliers.

5. THE KUHN-TUCKER CONDITIONS

A general nonlinear programming problem as described in Section 1 can also be composed of finding a set of x which minimizes (or maximizes) a function

$$S = f(x_1, x_2, \ldots, x_n) \tag{5-1}$$

subject to m inequality constrains

$$g_j(x_1, x_2, \ldots, x_n) \leq b_j, \qquad j = 1, \ldots, m \tag{5-2}$$

To obtain the optimal solution for the above problem, certain conditions must be satisfied at the optimal point known as the Kuhn-Tucker conditions.

In this section, we will firstly develope these conditions to maximize (or mimimize) an objective function of two variables

$$S = f(x_1, x_2) \tag{5-3}$$

subject to one inequality constraint

$$g(x_1, x_2) \leq 0 \tag{5-4}$$

and then generalize them for the objective function given by equation (5-1) subject to the constraints given by equation (5-2).

The inequality constraint given by equation (5-4) can be converted to an equality constraint by adding a slack variable z^2 such as

$$g(x_1, x_2) + z^2 = 0 \tag{5-5}$$

Therefore, the problem becomes maximizing the objective function (equation (5-3)) subject to the equality constraint (equation (5-5)).

This problem can now be solved by the method of Lagrange multiplier as the constraint is an equality. The Lagrangian function is

$$L(x_1, x_2, \lambda, z) = f(x_1, x_2) - \lambda[g(x_1, x_2) + z^2] \tag{5-6}$$

The necessary conditions for stationary points are

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \tag{5-7}$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0 \tag{5-8}$$

$$- \frac{\partial L}{\partial \lambda} = g(x_1, x_2) + z^2 = 0 \tag{5-9}$$

$$\frac{\partial L}{\partial z} = - 2\lambda z = 0 \tag{5-10}$$

Conditions (5-7) through (5-10) should all be satisfied at the optimum of $f(x_1, x_2)$. Equation (5-10) can be satisfied when either $\lambda = 0$ or $z = 0$. If z is equal to zero, then from equation (5-9), $g(x_1, x_2)$ is also equal to zero. If z is not equal to zero, then λ should be zero. Then, it is apparant that both conditions (5-9) and (5-10) state that

$$\lambda g(x_1, x_2) = 0 \tag{5-11}$$

Therefore the necessary conditions for the local optimum are given by equations (5-7), (5-8), (5-11) and the unequality constraint, equation (5-4), itself. These are the Kuhn-Tucker conditions for the optimization of a function subject to inequality constraint.

Here, we may look in the condition given by equation (5-10) again. According to this equation either $\lambda = 0$ or $z = 0$. If $\lambda = 0$, the Lagrangian multiplier vanishes and the stationary point exists only in the interior, that is, $g \leq 0$, $z^2 > 0$. If $\lambda \neq 0$, a non-zero Lagrangian multiplier implies that the solution is on the boundary, since $z = 0$ which implies g also equals zero.

Now, in a summary, the necessary conditions for a point (x_1, x_2) to be a local optimum of a function

$$S = f(x_1, x_2)$$

subject to the constraint

$$g(x_1, x_2) \leq 0$$

are that λ and (x_1, x_2) satisfy the following relations

$$\frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0$$

$$\frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0$$

$$\lambda\, g(x_1, x_2) = 0$$

$$g(x_1, x_2) \leq 0$$

These conditions are also sufficient for a stationary point to be a maximum if f is concave and the constraint is convex and to be a minimum if the function is convex and the constraint is convex. (Refer to later Sections).

6. GENERAL STATEMENT OF THE KUHN-TUCKER CONDITIONS

The Kuhn-Tucker conditions can now be generalized as follows.

A point $(x_1, x_2, \ldots, x_n)$ which optimizes a function

$$S = f(x_1, \ldots, x_n) \tag{6-1}$$

subject to the inequality constraints

$$g_j(x_1, \ldots, x_n) \leq 0, \qquad j = 1, \ldots, m \tag{6-2}$$

exists if there is a set of $\lambda_1, \ldots, \lambda_m$ that satisfies the following set of conditions.

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \qquad i = 1, \ldots, n \tag{6-3}$$

$$\lambda_j g_j = 0, \qquad j = 1, \ldots, m \tag{6-4}$$

$$g_j \leq 0, \qquad j = 1, \ldots, m \tag{6-5}$$

$$\lambda_j \geq 0, \qquad j = 1, \ldots, m \text{ (for maximization)} \tag{6-6a}$$

or

$$\lambda_j \leq 0, \qquad j = 1, \ldots, m \text{ (for minimization)} \tag{6-6b}$$

These conditions are also sufficient for a global minimum if f and g_j, $j = 1, \ldots, m$, are all convex and differentiable functions and for a global maximum if f is concave and g_j, $j = 1, \ldots, m$, are all convex and differentiable functions.

Similarly the necessary conditions for optimization of the function, equation (1), subject to the inequality constraints, equation (6-2), and when all x to be non-negative are

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} \leq 0, \quad i = 1, \ldots, n \text{ (for maximization)} \tag{6-7a}$$

or

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_i}{\partial x_i} \geq 0, \quad i = 1, \ldots, n \text{ (for minimization)} \tag{6-7b}$$

$$\sum_{i=1}^{n} x_i \frac{\partial L}{\partial x_i} = 0, \quad i = 1, \ldots, n \tag{6-8}$$

$$\lambda_j g_j = 0, \quad j = 1, \ldots, m \tag{6-9}$$

$$g_j \leq 0, \quad j = 1, \ldots, m \tag{6-10}$$

$$x_i \geq 0, \quad i = 1, \ldots, n \tag{6-11}$$

$$\lambda_j \geq 0, \quad j = 1, \ldots, m \quad \text{(for maximization)} \tag{6-12a}$$

or

$$\lambda_j \leq 0, \quad j = 1, \ldots, m \quad \text{(for minimization)} \tag{6-12b}$$

The derivation of these conditions given by equations (6-7) through (6-11) is discussed in Section 12.

Equations (6-6a), (6-6b), (6-12a) and (6-12b) are based on the fact that if $\lambda > 0$, the stationary point can not be a minimum, and if $\lambda < 0$, it can not be a maximum [Kuhn-Tucker (1951)]. Note that the sign of λ will be affected by several factors such as the type of optimization problem (whether maximization or minimization), the type of inequality constraints (whether $g_j(x) \leq 0$ or $g_j(x) \geq 0$), and the type of Lagrangian function (whether $L(x,\lambda) = f(x) - \sum_j \lambda_j g_j(x)$ or $L(x,\lambda) = f(x) + \sum_j \lambda_j g_j(x)$). Recall that equations (6-6a), (6-6b), (6-12a) and (6-12b) are based on inequality constraints given by equation (6-2) ($g_j(x) \leq 0$), and the Lagrangian function of the form, $L(x,\lambda) = f(x) - \sum_j \lambda_j g_j(x)$.

If there is only one constraint and there exists a stationary point of Lagrangian, $L(x,\lambda)$ for which $\lambda > 0$. Then this point, x_0, must lie on the boundary, since $\lambda g = 0$ and $\lambda > 0$. Therefore, we have

$$g(x_o) = 0$$

and only displacements which give

$$dg \leq 0$$

are admissible (in feasible region). Suppose that there exists a set of displacements dx_i, $i = 1, \ldots, n$, such that

$$dg = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i}\right)_{x_o} dx_i < 0 \qquad (6\text{-}13)$$

Figure 2 shows that an infinitesimal move away from the boundary stationary point, A, will be in the direction of one of the arrows which results in a decrease in $g(x)$ (a move from boundary into interior). Therefore, each move represents a displacement which could be used in equation (6-13). At this restricted stationary point

$$\frac{\partial L}{\partial x_i} = 0, \qquad i = 1, \ldots, n$$

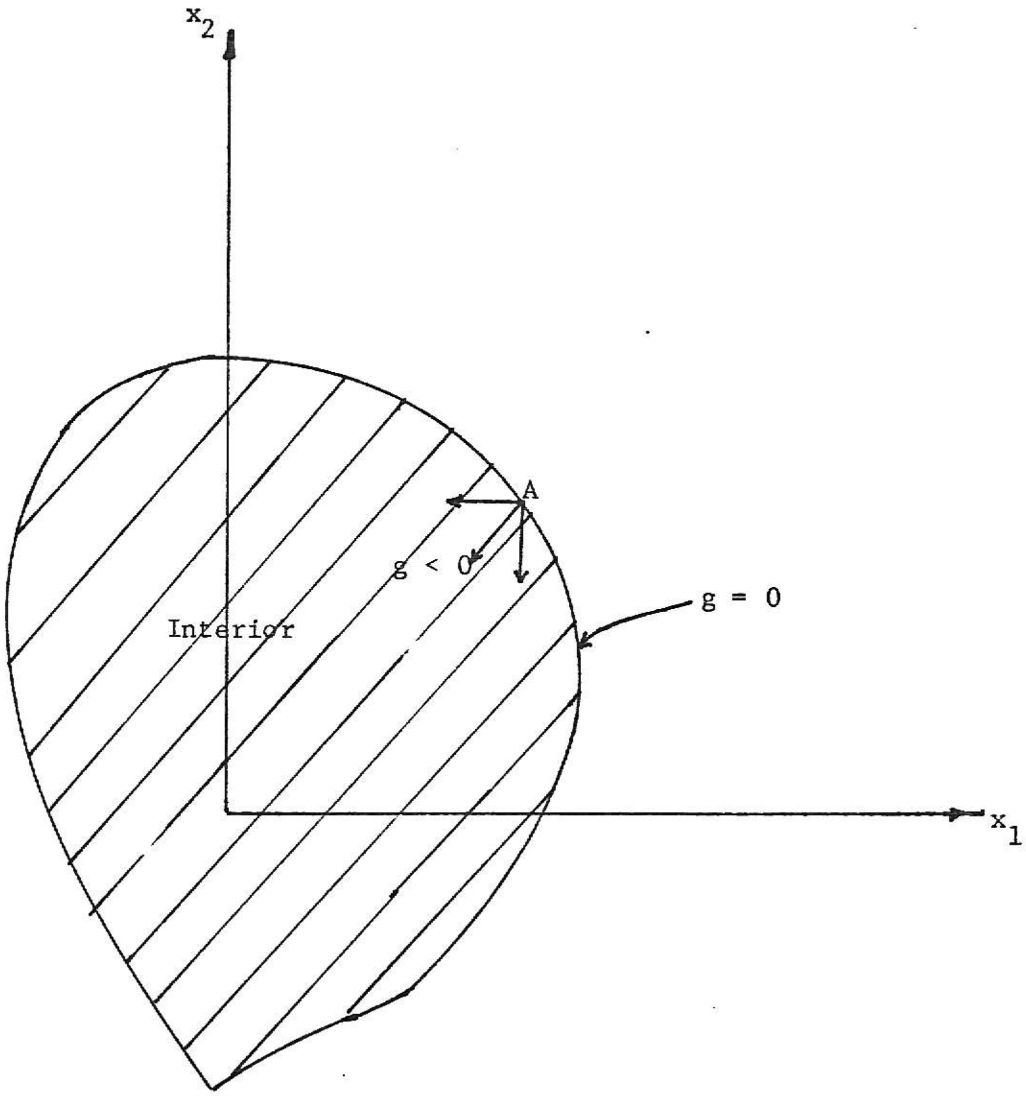


Fig. 2. Accessible regions in two dimensions.

where

$$L(x,\lambda) = f(x) - \lambda g(x)$$

This gives

$$\left(\frac{\partial f}{\partial x_i}\right)_{x_o} = \lambda \left(\frac{\partial g}{\partial x_i}\right)_{x_o}, \qquad i = 1, \ldots, n \tag{6-14}$$

Using equation (6-14), and the summation of the set of displacements yields

$$\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)_{x_o} dx_i = \lambda \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i}\right)_{x_o} dx_i \tag{6-15}$$

or

$$df = \lambda dg$$

If the restricted stationary point is also an accessible stationary point, then

$$df = \lambda dg < 0$$

for positive λ. The accessible stationary point is defined as a stationary point such as point A in Fig. 2 for which there exists at least one set of displacements from a boundary for which the constraining function decreases, that is, equation (6-13) holds. Consequently, an admissible move dx_i, $i = 1, \ldots, n$, into the interior has resulted in a decrease in the objective function, $f(x)$. Then, the stationary point on the boundary can not be a minimum. Therefore, for the accessible restricted point to be a maximum, it is necessary but not sufficient that the Lagrange multiplier must be positive. Similarly, a necessary condition for an accessible stationary point to be a minimum is that the Lagrange multiplier is negative. It is worth noting that these are necessary conditions but they are not sufficient. In a summary the necessary condition on the sign of λ for a single constraint, $g(x) \leq 0$, is given by Table 1.

In general, the above argument can apply to more than one constraint problem, and resulted equations (6-6a), (6-6b), (6-13a) and (6-13b). The detailed discussion and expanded derivation of these conditions are given by Kuhn and Tucker (1951) and Vajda (1961).

Table 1. Necessary condition for λ with a single constraint, $g(x) \leq 0$

Sign of Lagrange multiplier, λ	Stationary point	
	Location	Nature
nonzero, positive	on boundary	Not a minimum
nonzero, negative	on boundary	Not a maximum
zero	on boundary or interior of the feasible region	Either a maximum or minimum or neither

7. AN EXAMPLE

The application of the Kuhn-Tucker conditions for obtaining optimal solutions to nonlinear programming problem is illustrated by the following example.

Maximize

$$f(x_1,x_2) = x_2^2 - x_1 - x_1^2 \tag{7-1}$$

subject to an inequality constraint

$$g(x_1,x_2) = x_1^2 + x_2^2 - 1 \leq 0 \tag{7-2}$$

The Lagrangian is

$$L(x_1,x_2,\lambda) = (x_2^2-x_1-x_1^2) - \lambda(x_1^2+x_2^2-1) \tag{7-3}$$

The Kuhn-Tucker conditions are (see equations (6-3) through (6-6)).

$$\frac{\partial L}{\partial x_1} = -1 - 2x_1 - 2\lambda x_1 = 0 \tag{7-4}$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - 2\lambda x_2 = 0 \tag{7-5}$$

$$\lambda(x_1^2 + x_2^2 - 1) = 0 \tag{7-6}$$

$$x_1^2 + x_2^2 - 1 \leq 0 \tag{7-7}$$

$$\lambda \geq 0 \tag{7-8}$$

The point which will satisfy all the above conditions will be the stationary point. As a start considering first equation (7-8), the value of λ can either be = 0 or > 0. When $\lambda = 0$, equations (7-4) and (7-5) can be solved for the values of x_1 and x_2 and we obtain $x_1 = -0.5$, and $x_2 = 0$. This gives the point B on Figure 3. The value of the objective function at point B is $f(-0.5, 0) = 0.25$.

Next, if $\lambda > 0$, equations (7-4), (7-5) and (7-6) can be rewritten as follows

$$1 + 2x_1 + 2\lambda x_1 = 0 \tag{7-9}$$

$$2x_2\,(1-\lambda) = 0 \tag{7-10}$$

$$x_1^2 + x_2^2 - 1 = 0 \tag{7-11}$$

From equation (7-10), it can be concluded that either $x_2 = 0$ or $(1-\lambda) = 0$. Considering first $x_2 = 0$, we will try to find the stationary point which will satisfy all the necessary conditions. Therefore substituting $x_2 = 0$ into equation (7-11), we obtain

$$x_1 = \pm 1 \tag{7-12}$$

Equation (7-12) gives two values of x_1. Substituting these values into equation (7-9), the values of λ come out to be -3/2 and -1/2, which are non-positive. This solution violates the condition (7-8), which states that $\lambda \geq 0$, and will be neglected. Therefore, $x_2 = 0$ can not be a possible case for the stationary point.

Now considering the case $1-\lambda = 0$, we obtain

$$\lambda = 1$$

Substituting this value of λ into equation (7-9) gives

$$1 + 2x_1 + 2x_1 = 0 \tag{7-13}$$

or

$$x_1 = -.25$$

Substituting the value of x_1 into equation (7-11) gives

$$x_2 = \pm .968 \tag{7-14}$$

This gives points C and D on Figure 3. The value of the objective function at these points is $f(x_1, x_2)_{C\&D} = 1.125$.

The above results, along with the objective function and the constraint are plotted in Figure 3, from which it is clear that among the three stationary points given by the Kuhn-Tucker conditions only C and D are global maximum points and B is a saddle point. This indicates that the Kuhn-Tucker conditions are only necessary conditions for optimality and are not sufficient. There are certain conditions of sufficiency (described in Section 10), which, if are satisfied, ensures the stationary point given by the Kuhn-Tucker conditions to be global optimum. For the present problem, the sufficient conditions for global maximum are that the objective function should be concave and the constraint should be convex. In this case, the objective function is a saddl because its Hessian matrix is neither a positive definite nor a negative definite (refer Section 9 for definitions of positive and negative definite) and the constraint is a convex function as its Hessian matrix is a positive definite. Therefore, the sufficient conditions are not satisfied in this case and the stationary points given by the Kuhn-Tucker conditions may

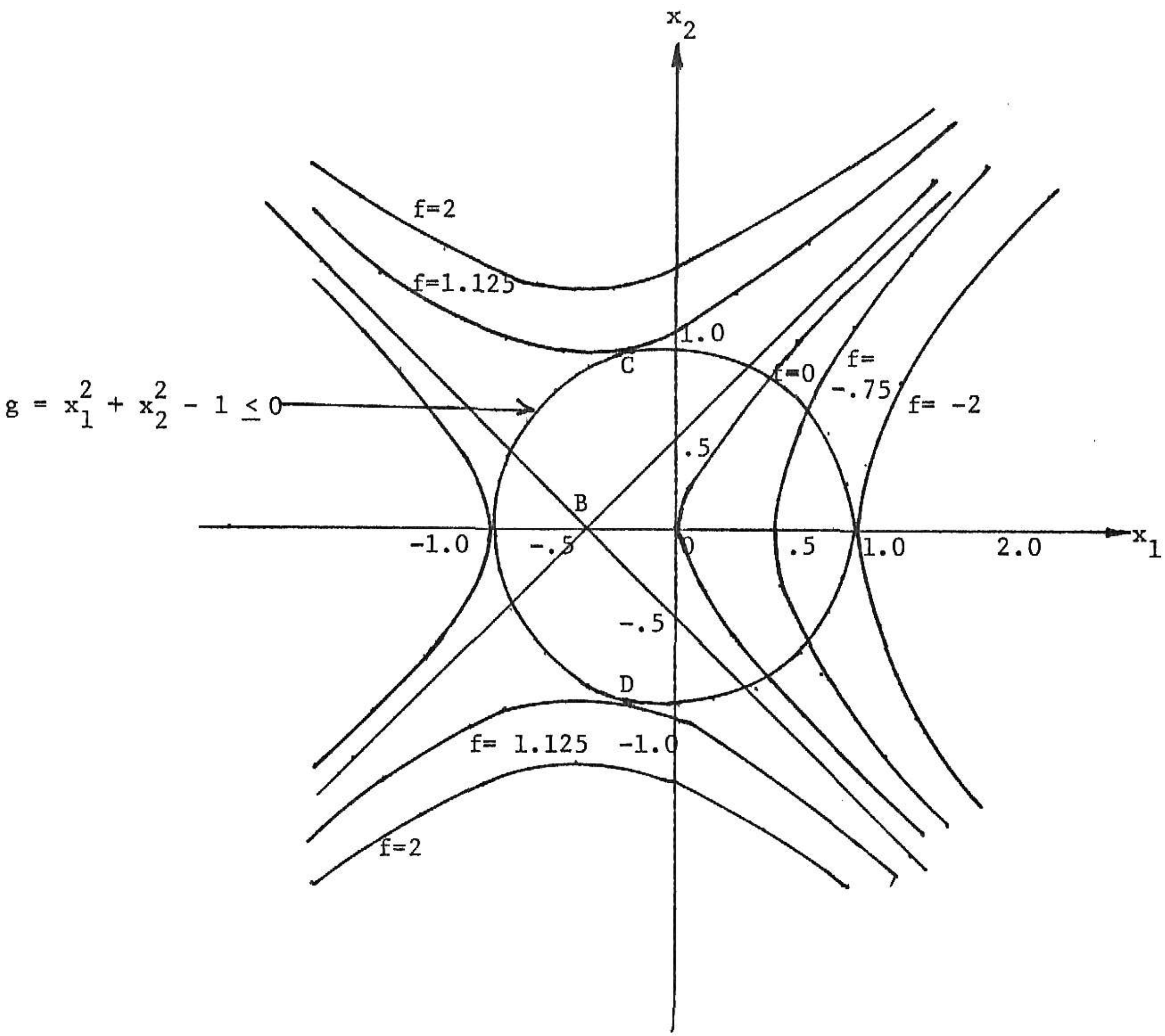


Fig. 3 Contour lines of the objective function, $f(x_1,x_2) = x_2^2-x_1-x_1^2$, and the constraint, $g(x_1,x_2) = x_1^2+x_2^2-1 \leq 0$, for the example.

or may not be global maximum. This is also obvious from the three stationary points given by the Kuhn-Tucker conditions, from which only two points (C and D) are global maximum and point B is a saddle point.

8. QUADRATIC PROGRAMMING

Quadratic programming deals with the problems of optimizing a quadratic objective function subject to linear constraints. A quadratic programming problem can be written as

Maximize (or minimize)

$$f(x) = \sum_{i=1}^{n} c_i x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} q_{ik} x_i x_k$$

$$= c^T x + x^T Q x \qquad (8\text{-}1)$$

subject to the linear constraints

$$g_j(x) = \sum_{i=1}^{n} a_{ji} x_i \leq b_j, \qquad j = 1,2, \ldots, m \qquad (8\text{-}2)$$

and

$$x_i \geq 0, \qquad i = 1,2, \ldots, n$$

where $q_{ik} = q_{ki}$ are given constants and elements of Q matrix.

There are several solution procedures for the special case of the quadratic programming problem based on the concavity (for maximization) and convexity (for minimization) of the objective function. The Wolfe simplex method [1959] for quadratic programming is presented in this section since it has been widely used and it is based on the Kuhn-Tucker conditions.

The Kuhn-Tucker conditions for the quadratic programming problem are given as follows:

$$\frac{\partial L(x,y)}{\partial x_i} = \frac{\partial}{\partial x_i}\left(f(x) - \sum_{j=1}^{m} y_{n+j}g_j(x)\right) \le 0 \qquad (8\text{-}3)$$

or

$$\frac{\partial L}{\partial x_i} = c_i + \sum_{k=1}^{n} q_{ik}x_k - \sum_{j=1}^{m} a_{ij}y_{n+j} \le 0, \qquad i = 1,2, \ldots, n \qquad (8\text{-}4)$$

The inequality in the above equation can be changed to equality by adding y_i, slack variables, that is

$$\frac{\partial L}{\partial x_i} = c_i + \sum_{k=1}^{n} q_{ik}x_k - \sum_{j=1}^{m} a_{ij}y_{n+j} + y_i = 0, \; i = 1, 2, \ldots, n \qquad (8\text{-}5)$$

The other Kuhn-Tucker conditions are

$$-\frac{\partial L(x,y)}{\partial y_{n+j}} = \sum_{i=1}^{n} a_{ji}x_i + x_{n+j} = b_j, \qquad j = 1,2, \ldots, m \qquad (8\text{-}6)$$

$$x_i \ge 0, \qquad i = 1,2, \ldots, n+m \qquad (8\text{-}7)$$

$$y_i \ge 0, \qquad i = 1,2, \ldots, n+m \qquad (8\text{-}8)$$

$$x_i y_i = 0, \qquad i = 1,2, \ldots, n+m \qquad (8\text{-}9)$$

In the above equations y_{n+j} $(j = 1, \ldots, m)$ are the Lagrange multipliers and y_i $(i = 1, \ldots, n)$ and x_{n+j} $(j = 1, \ldots, m)$ are slack variables. Thus $(x_1, x_2, \ldots, x_n)$ will give the stationary points if there exists values of $x_{n+1}, \ldots, x_{n+m}$; $y_1, \ldots, y_{n+m}$ such that $(x_1, \ldots, x_{n+m},$

$y_1, \ldots, y_{n+m}$) satisfies all conditions given by equations (8-5) through (8-9). The problem now reduces to *finding a feasible solution to these* conditions which are linear programming constraints involving 2(n+m) variables except the last restriction given by equation (8-9) (Refer to equations (6-8) and (6-9)). The condition $x_i y_i = 0$ in equation (8-9) says that it is not possible for both x_i and y_i to be basic variables when considering basic feasible solutions. Therefore, the problem reduces to finding an initial basic feasible solution to any linear programming problem having these normal constraints (equations (8-5) through (8-8)), subject to this additional restriction (equation (8-9)) on the identity of the basic variables.

The standard linear programming trick, when there is not an obvious basic feasible solution is to introduce artificial variables which are eventually forced to equal zero. Let z_i, $i = 1,2, \ldots, n$ be these artificial variables, where the only restriction on them is

$$z_i \geq 0, \qquad\qquad i = 1,2, \ldots, n$$

Therefore, equation (8-5) is changed to

$$-\sum_{k=1}^{n} q_{ik} x_k + \sum_{j=1}^{m} a_{ij}\, y_{n+j} - y_i + c_i\, z_i = c_i, \qquad i = 1,2, \ldots, n$$

Now a feasible solution to this artificial problem is feasible for the real problem if and only if $z_i = 0$ for $i = 1,2, \ldots, n$. Therefore, $\sum_{i=1}^{n} z_i$ should be decreased to zero in order to obtain the desired feasible solution.

Therefore, the solution of the following problem along with a modification of the simplex method stated below will give the optimal solution.

Minimize

$$\sum_{i=1}^{n} z_i$$

subject to

$$-\sum_{k=1}^{n} q_{ik}\, x_k + \sum_{j=1}^{m} a_{ij}\, y_{n+j} - y_i + c_i\, z_i = c_i, \qquad i = 1, 2, \ldots, n$$

$$\sum_{i=1}^{n} a_{ji}\, x_i + x_{n+j} = b_j, \qquad j = 1, 2, \ldots, m$$

$$x_i \geq 0, \qquad i = 1, 2, \ldots, n+m$$

$$y_i \geq 0, \qquad i = 1, 2, \ldots, n+m$$

$$z_i \geq 0, \qquad i = 1, 2, \ldots, n.$$

The modification is that y_i is not permitted to become a basic variable whenever x_i is already a basic variable, and vice-versa for $i = 1, 2, \ldots, n+m$. This ensures that $x_i y_i = 0$ for each value of i. When the optimal solution $(\underline{x}_1, \ldots, \underline{x}_{n+m}; \underline{y}_1, \ldots, \underline{y}_{n+m}; z_1 = 0, \ldots, z_n = 0)$ is obtained to this problem, $(\underline{x}_1, \ldots, \underline{x}_n)$ is the optimal solution to the quadratic programming problem.

The stationary point obtained above will be global optimum if the sufficient conditions described in section 10 are satisfied. Since in this case constraints are linear, it forms a convex set. Therefore, it will depend on the convexity (for minimization) or concavity (for maximization) of the quadratic objective function that the point is global optimal.

The solution procedures shall be illustrated by the following example.

EXAMPLE 1

$$\text{Maximize} \quad f(x) = -(x_1-3)^2 - (x_2-3)^2$$

$$= 6x_1 + 6x_2 - x_1^2 - x_2^2 - 18$$

subject to

$$x_1 + x_2 \leq 7$$

$$-x_1 + 3x_2 \leq 9$$

$$3x_1 - x_2 \leq 9$$

$$x_1 \;\&\; x_2 \geq 0$$

Using slack variables, x_3, x_4, and x_5, the inequality constraints become equality constraints as

$$x_1 + x_2 + x_3 - 7 = 0$$

$$-x_1 + 3x_2 + x_4 - 9 = 0$$

$$3x_1 - x_2 + x_5 - 9 = 0$$

The Lagrangian function is

$$L = 6x_1 + 6x_2 - x_1^2 - x_2^2 - 18 - y_3\ (x_1 + x_2 + x_3 - 7)$$

$$-y_4\ (-x_1 + 3x_2 + x_4 - 9) - y_5\ (3x_1 - x_2 + x_5 - 9)$$

The Kuhn-Tucker conditions are

$$\frac{\partial L}{\partial x_1} = 6 - 2x_1 - y_3 + y_4 - 3y_5 \leq 0$$

$$\frac{\partial L}{\partial x_2} = 6 - 2x_2 - y_3 - 3y_4 + y_5 \leq 0$$

$$-\frac{\partial L}{\partial y_3} = x_1 + x_2 + x_3 - 7 = 0$$

$$-\frac{\partial L}{\partial y_4} = -x_1 + 3x_2 + x_4 - 9 = 0$$

$$-\frac{\partial L}{\partial y_5} = 3x_1 - x_2 + x_5 - 9 = 0$$

$$x_i \geq 0, \qquad i = 1, \ldots, 5$$

$$y_j \geq 0, \qquad j = 3, 4, 5$$

$$x_i y_i = 0$$

The first and second inequality conditions above ($\frac{\partial L}{\partial x_1} \leq 0$ and $\frac{\partial L}{\partial x_2} \leq 0$) can be changed to equality condition by introducing slack variables y_1 and y_2 as

$$2x_1 + y_3 - y_4 + 3y_5 - y_1 = 6$$

$$2x_2 + y_3 + 3y_4 - y_5 - y_2 = 6$$

The problem to be solved by revised linear programming is:

Minimize

$$z_1 + z_2$$

subject to

$$2x_1 + y_3 - y_4 + 3y_5 - y_1 + 6z_1 = 6$$

$$2x_2 + y_3 + 3y_4 - y_5 - y_2 + 6z_2 = 6$$

$$x_1 + x_2 + x_3 = 7$$

$$-x_1 + 3x_2 + x_4 = 9$$

$$3x_1 - x_2 + x_5 = 9$$

$$x_i \geq 0, \qquad i = 1, \ldots, 5$$

$$y_i \geq 0, \qquad i = 1, \ldots, 5$$

$$z_i \geq 0, \qquad i = 1, 2$$

$$x_i y_i = 0, \qquad i = 1, \ldots, 5$$

The above problem has been solved by the modified simplex method (satisfying $x_i y_i = 0$) as well as by an I.B.M. L.P. subroutine. The solution by the modified simplex method is shown on Table 2 and the optimal value of the variables are as follows

$x_1 = 3$

$x_2 = 3$

$x_3 = 1$

$x_4 = 3$

$x_5 = 3$

all other variables are zero.

The above values of the x_1 and x_2 gives a value of the objective function as 0. Moreover this solution gives us a value of $x_i y_i = 0$, $(i = 1,2)$ as the value of the variables y_1 and $y_2 = 0$.

The results from the IBM L.P. subroutine solution gives a value of the variables as follows

$x_1 = 4$

$x_2 = 3$

$x_4 = 4$

$y_4 = 2$

$y_2 = 6$

All other variables are zero.

This solution yields a value of $x_2 y_2 = 18$, which is in direct conflict of the constraint, $x_i y_i = 0$. The maximum value of the

Table 2. The modified simplex method solution

C_i	basis	b_i	x^o_1	x^o_2	x^o_3	x^o_4	x^o_5	y^o_3	y^o_4	y^o_5	y^o_1	y^o_2	z^1_1	z^1_2
1	z_1	1	$\frac{1}{3}$	0	0	0	0	$\frac{1}{6}$	$-\frac{1}{6}$	$\frac{1}{2}$	$-\frac{1}{6}$	0	1	0
1	z_2	1	0	$\frac{1}{3}$	0	0	0	$\frac{1}{6}$	$\frac{1}{2}$	$-\frac{1}{6}$	0	$-\frac{1}{6}$	0	1
0	x_3	7	1	1	1	0	0	0	0	0	0	0	0	0
0	x_4	9	-1	3	0	1	0	0	0	0	0	0	0	0
0	x_5	9	3	-1	0	0	1	0	0	0	0	0	0	0
	z=2		$\frac{1}{3}$	$\frac{1}{3}$	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{6}$	0	0
0	x_1	3	1	0	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	0	3	0
1	z_1	1	0	$\frac{1}{3}$	0	0	0	$\frac{1}{6}$	$\frac{1}{2}$	$-\frac{1}{6}$	0	$-\frac{1}{6}$	0	1
0	x_3	4	0	1	1	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	0	-3	0
0	x_4	12	0	3	0	1	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	0	3	0
0	x_5	0	0	-1	0	0	1	$-\frac{3}{2}$	$\frac{3}{2}$	$-\frac{9}{2}$	$\frac{3}{2}$	0	9	0
	z=1		0	$\frac{1}{3}$	0	0	0	$\frac{1}{6}$	$\frac{1}{2}$	$-\frac{1}{6}$	0	$-\frac{1}{6}$	-1	0
0	x_1	3	1	0	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	0	3	0
0	x_2	3	0	1	0	0	0	$\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	3
0	x_3	1	0	0	1	0	0	-1	-1	2	$\frac{1}{2}$	$\frac{1}{2}$	-3	-3
0	x_4	3	0	0	0	1	0	-1	-4	0	$-\frac{1}{2}$	$\frac{3}{2}$	3	-9
0	x_5	3	0	0	0	0	1	-1	3	-5	$\frac{3}{2}$	$-\frac{1}{2}$	9	0
	z=0		0	0	0	0	0	0	0	0	0	0	-1	-1

The artificial vectors could be eliminated or transfered from the basic solution.

$x_1 = 3$

$x_2 = 3$

slacks $\begin{cases} x_3 = 1 \\ x_4 = 3 \\ x_5 = 3 \end{cases}$

objective function comes out to be -1, which is less than the value of the objective function given by the modified simplex method.

These constraints are plotted in Figure 4 and indicates the position of the maximum point from both the methods. Although both the points are in the feasible region of the original quadratic programming problem, since the basic simplex method in the IBM L.P. subroutine does not satisfy $x_i y_i = 0$, therefore, the solution is not the optimal one.

Since the quadratic objective function is negative-definite (by judging from the Hessian matrix), the function is concave. Therefore, the maximum point found by the modified simplex method is a global maximum point.

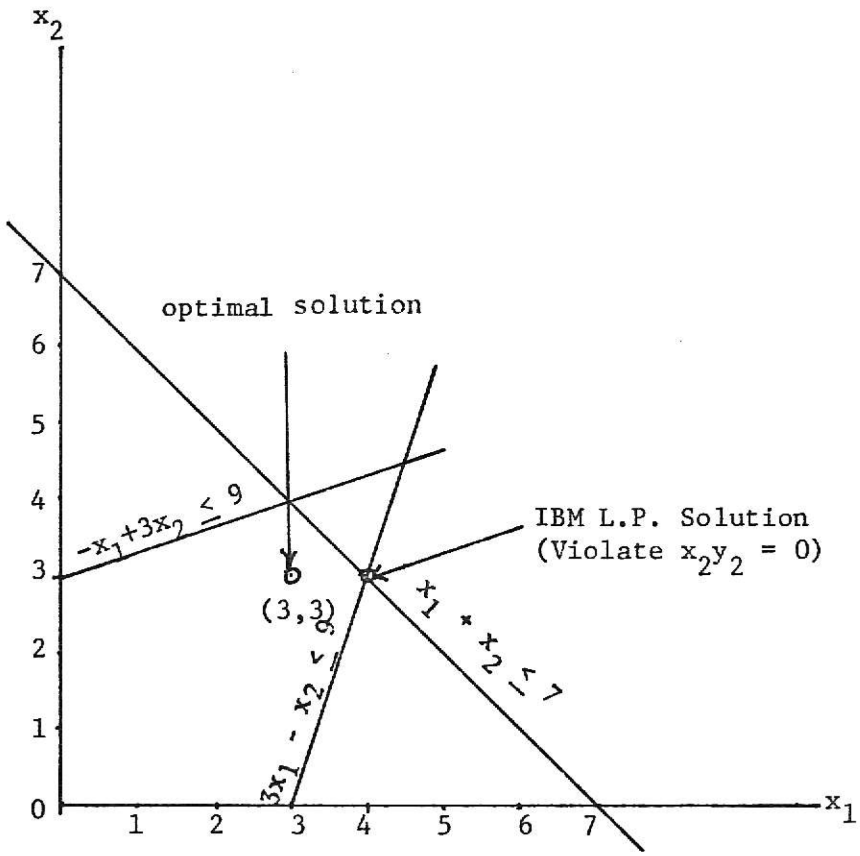


Fig. 4 Linear programming solution of the quadratic programming problem.

9. CONVEXITY AND CONCAVITY OF FUNCTIONS

In this section, we will describe the procedures to recognize whether a function is convex or concave. The concept of convexity and concavity assists in determining whether a local optimum is also a global optimum or not.

9.1 CONVEX FUNCTION

A function of one variable

$$S = f(x)$$

is called convex over the domain R if for any two points x' and x" ε R, and for any value of x defined by

$$x = (1-\lambda)\ x' + \lambda x'', \qquad 0 \leq \lambda \leq 1$$

The inequality

$$f(x) \leq (1-\lambda)\ f(x') + \lambda f(x'') \tag{9-1}$$

must be satisfied. In the above expression

$$\lambda = \frac{x - x'}{x'' - x'}$$

The function is strictly convex if for $x' \neq x''$, the sign of equation (9-1) may be replaced by a inequality sign only. Figure 5 geometrically illustrates a strictly convex function of one independent variable. As is clear from the figure that within the range $x' \leq x \leq x''$, a convex function can not have any value larger than the values of the function obtained by linear interpolation between f(x') and f(x"). In other words for a convex function a line segment joining any two points on the surface will always lie on or above the surface.

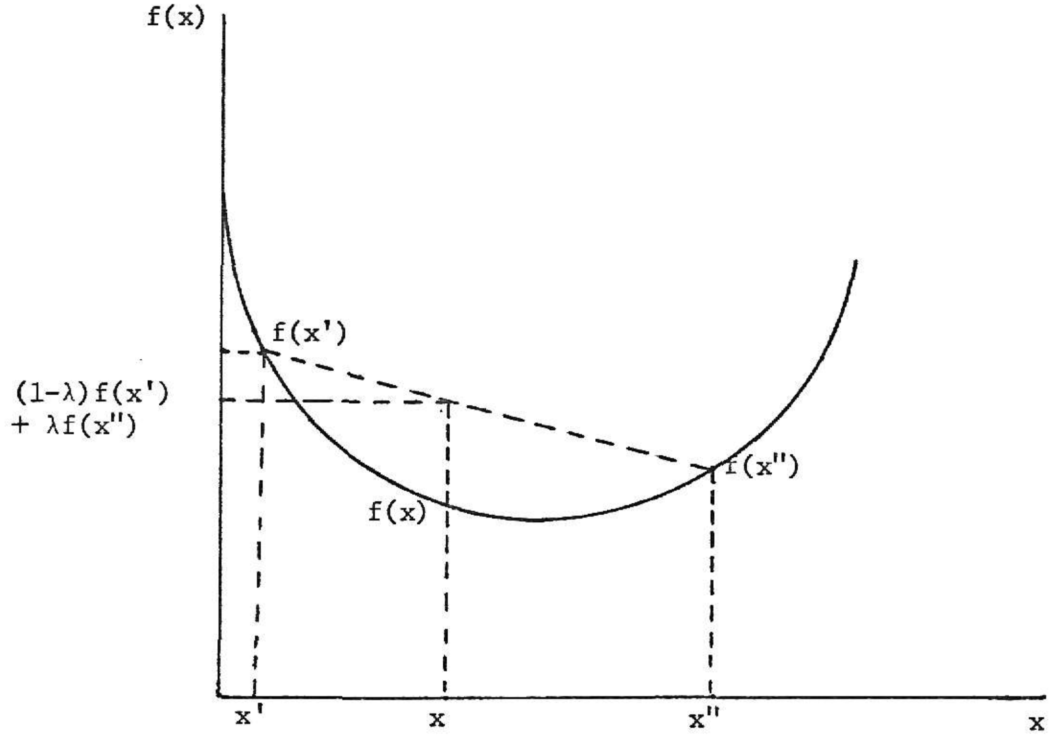


Fig. 5 A convex function of one variable.

The above definition for a convex function of one variable can be extended to the function of multivariables. Thus a function

$$S = f(x_1, \ldots, x_n)$$

is called convex if for any pair of sets $(x_1', \ldots, x_n')$ and $(x_1'', \ldots, x_n'')$ and for any λ such that

$$\lambda = \frac{(x_1, \ldots, x_n) - (x_1', \ldots, x_n')}{(x_1'', \ldots, x_n'') - (x_1', \ldots, x_n')}$$

we have

$$f((1-\lambda)x_1' + \lambda x_1'', \ldots, (1-\lambda)x_n' + \lambda x_n'')$$

$$\leq (1-\lambda)\ f(x_1', \ldots, x_n') + \lambda f(x_1'', \ldots, x_n'') \qquad (9\text{-}2)$$

9.2 CONCAVE FUNCTION

A function of one variable

$$S = f(x)$$

is called concave over the domain R if for any two points x' and $x'' \in$ R, and for any value of x defined by

$$x = (1-\lambda)x' + \lambda x'', \qquad 0 \leq \lambda \leq 1$$

the inequality

$$f(x) \geq (1-\lambda)f(x') + \lambda f(x'') \qquad (9\text{-}3)$$

must be satisfied. In the above expression

$$\lambda = \frac{x - x'}{x'' - x'}$$

Figure 6 geometrically illustrate a strictly concave function of one variable. It is clear from the figure that, a concave function can not have any value smaller than the values of the function obtained by linear interpolation between $f(x')$ and $f(x'')$.

The above definition for a concave function of one variable can be extended to the function of multivariables. Thus a function

$$S = f(x_1, \ldots, x_n)$$

is called concave if for any pair of sets $(x_1', \ldots, x_n')$ and $(x_1'', \ldots, x_n'')$ and any λ as defined below

$$\lambda = \frac{(x_1, \ldots, x_n) - (x_1', \ldots, x_n')}{(x_1'', \ldots, x_n'') - (x_1', \ldots, x_n')}$$

we have

$$f((1-\lambda)\ x_1' + \lambda x_1'', \ldots, (1-\lambda)x_n' + \lambda x_n'') \geq$$

$$(1-\lambda)\ f(x_1', \ldots, x_n') + \lambda\ f(x_1'', \ldots, x_n'') \qquad (9\text{-}4)$$

A function of two variable can be drawn graphically in three dimensions by plotting the function on vertical axis ($f(x_1,x_2)$ axis) and variables on horizontal axes (x_1 and x_2 axes) as shown in Figures 7 and 8. It can also be drawn in two dimensions (on $x_1 - x_2$ plane) as contour lines of function as shown again in Figures 7 and 8. A contour line is a set of points for which objective function has a constant value. If we see the contour lines of a convex function as shown in Figure 7, we will find that the value of the function increases as we go outward from center. Similarly if we see the

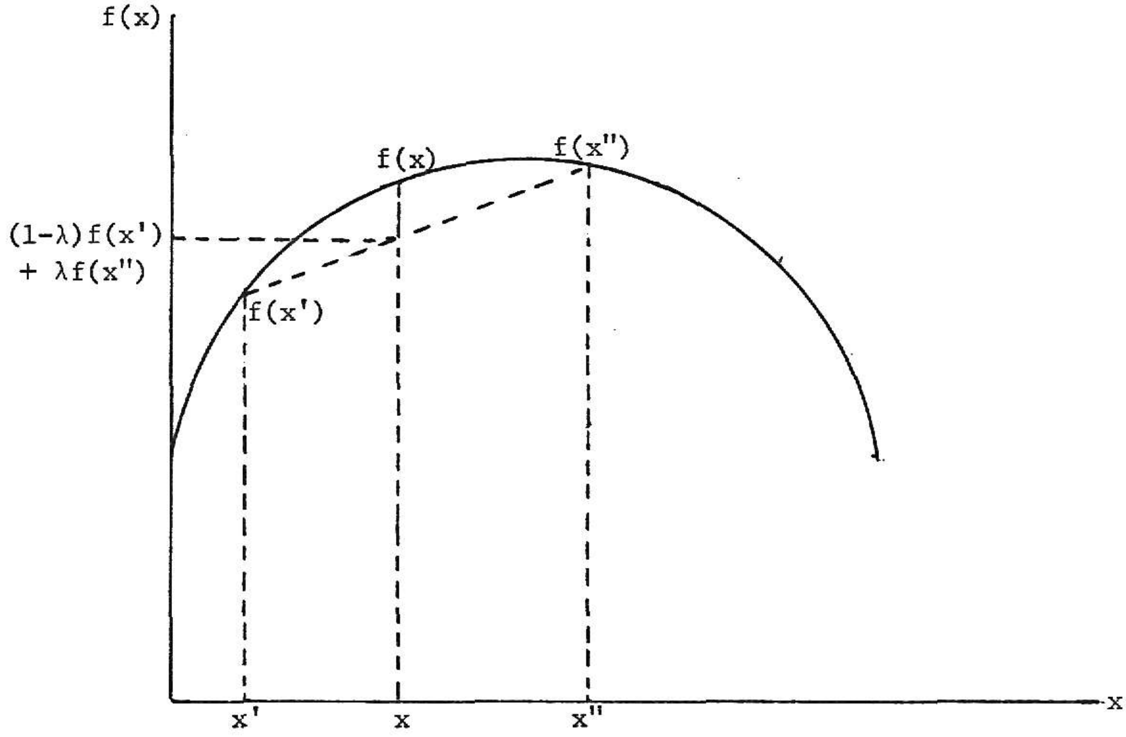


Fig. 6 A concave function of one variable.

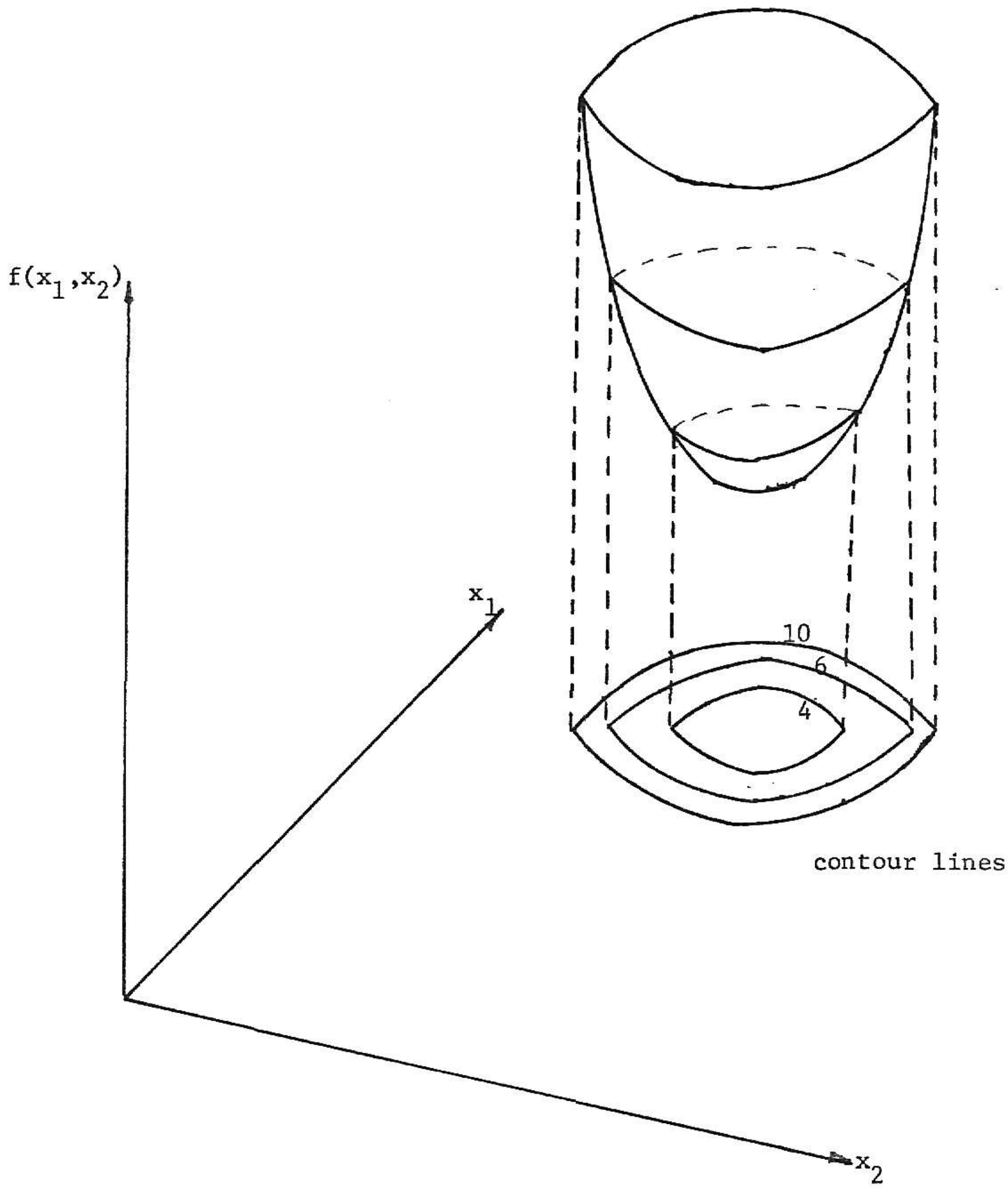


Fig. 7 Convex function of two variables

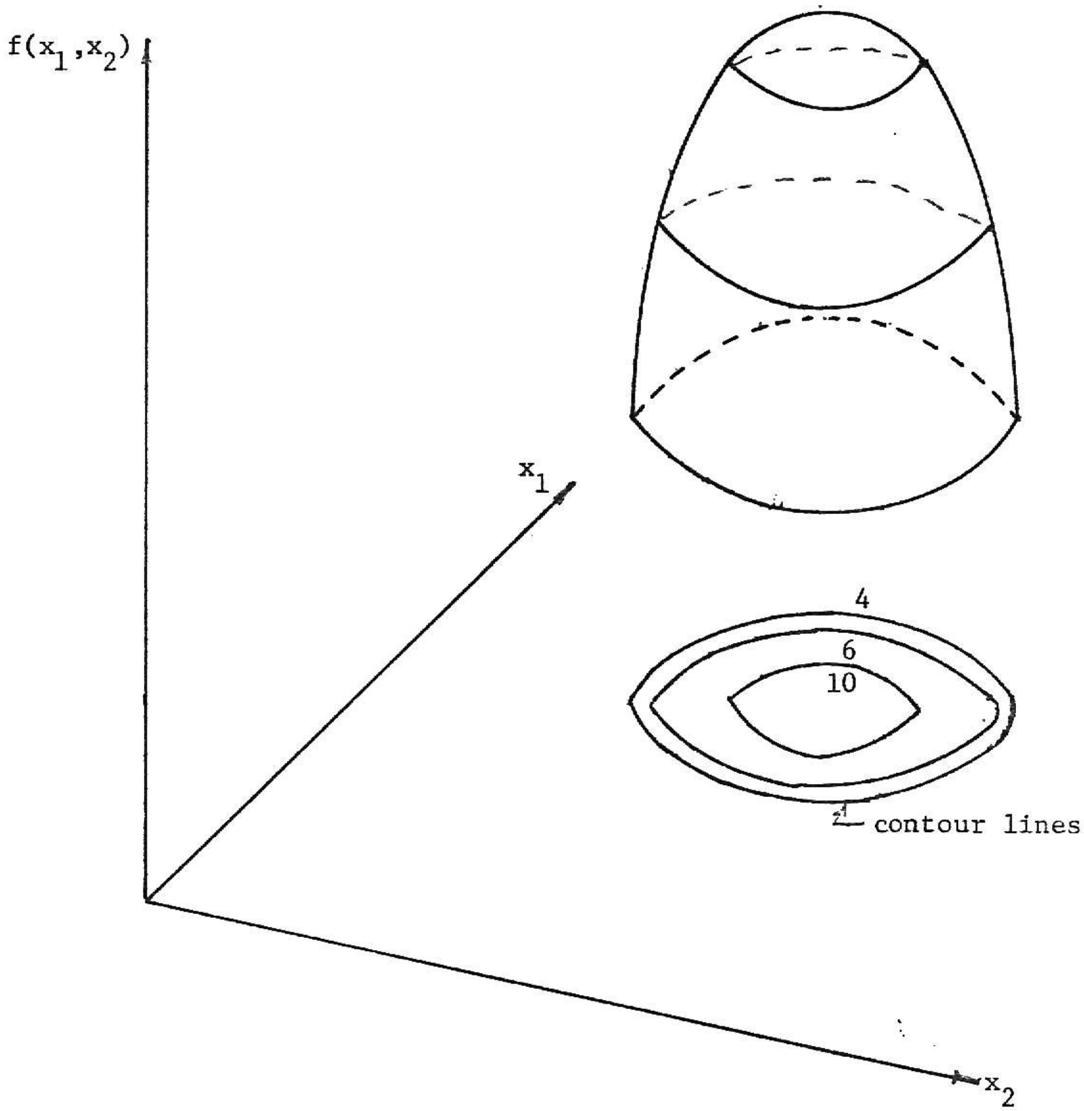


Fig. 8 Concave function of two variables.

contour lines of a concave function as shown in Figure 8, we will find that the value of the function decreases as we go to outward from center.

The functions of multivariables are hard to sketch on a paper but have properties analogous to the functions of one or two variables.

From the above discussion it is clear that a concave function is the reverse of a convex function and vice-versa. Thus a function f(x) is strictly concave if - f(x) is strictly convex.

9.3 PROPERTIES OF CONVEX FUNCTIONS AND CONCAVE FUNCTIONS

A differentiable convex function has following properties.

(a) $f(x'') - f(x') \geq \nabla^T f(x')(x''-x')$ for all x', x'', (9-5)

where $\nabla f(x')$ is the gradient at point x' and is denoted by

$$\nabla f(x') = \begin{bmatrix} \frac{\partial f(x')}{\partial x_1} \\ \vdots \\ \frac{\partial f(x')}{\partial x_n} \end{bmatrix}$$

and $\nabla^T f(x')$ is the transpose of $\nabla f(x')$.

(b) The Hessian matrix or the matrix of the second partial derivatives of f(x) with respect to x is a positive definite (or a positive semidefinite) for all x if f(x) is strictly convex (or convex).

(c) Over the domain R, f(x) has only one extremum.

When f(x) is a function of a single variable, equation (9-5) can be written as

$$\frac{df(x')}{dx} \leq \frac{f(x'') - f(x')}{x'' - x'} , \qquad x' < x''$$

This says that the slope of f(x) at x' is less than, or equal to, the slope of the secant through the points (x', f(x')) and (x'' , f(x'')). This is geometrically shown in Figure 9.

Similarly a differntiable concave function has following properties.

(a) $f(x'') - f(x') \leq \nabla^T f(x') (x''-x')$ for all x', x''

where $\nabla f(x')$ is the gradient of the function and $\nabla^T f(x')$ is the transpose of $\nabla f(x')$.

(b) The Hessian metrix is a negative definite (or a negative semidefinite) for all x if f(x) is strictly concave (or concave).

(c) Over the domain R, f(x) has only one extremum.

9.4 HESSIAN MATRIX AND PRINCIPAL MINORS

A function of two variables $f(x_1, x_2)$ can be expanded in the Taylor series about the point (a,b), as below.

$$f(a + \Delta x_1, b + \Delta x_2) - f(a, b)$$

$$= \left.\frac{\partial f}{\partial x_1}\right|_{a,b} \Delta x_1 + \left.\frac{\partial f}{\partial x_2}\right|_{a,b} \Delta x_2$$

$$+ \frac{1}{2} \left.\frac{\partial^2 f}{\partial x_1^2}\right|_{a,b} (\Delta x_1)^2 + \left.\frac{\partial^2 f}{\partial x_1 \, \partial x_2}\right|_{a,b} (\Delta x_1)(\Delta x_2)$$

$$+ \frac{1}{2} \left.\frac{\partial^2 f}{\partial x_2^2}\right|_{a,b} (\Delta x_2)^2 + \ . \ . \ . \qquad (9\text{-}7)$$

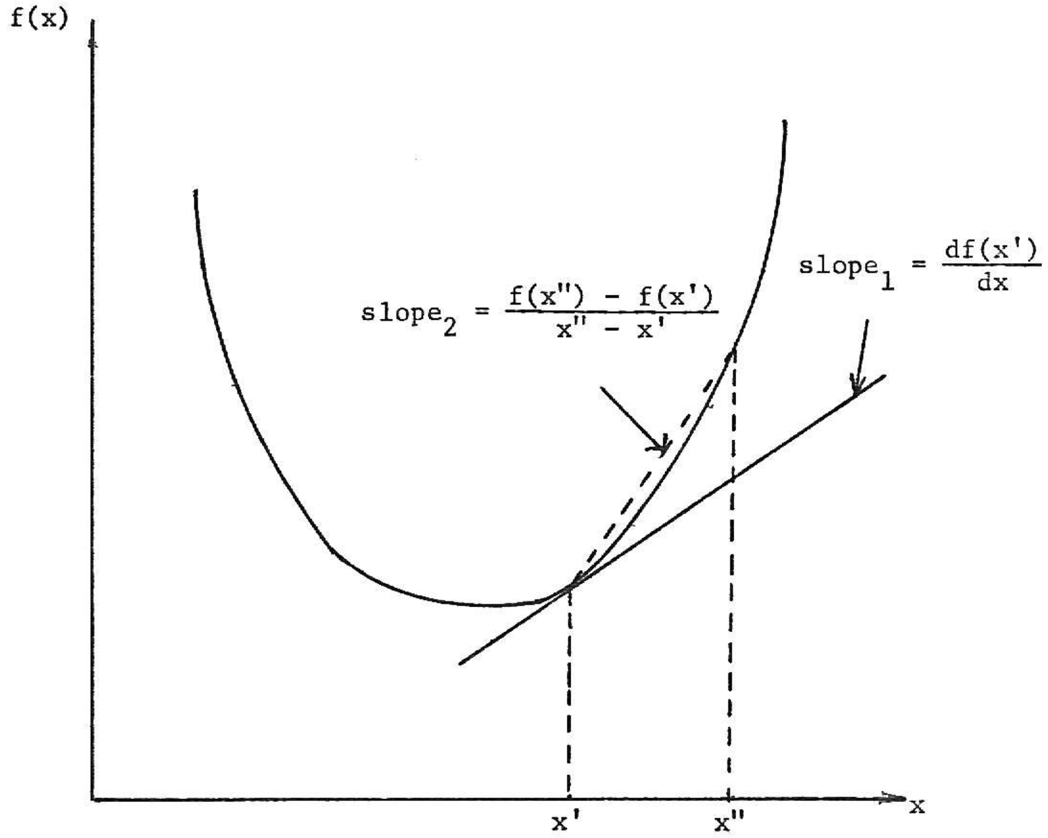


Fig. 9 A property of a convex function of a single variable
slope 1 $\leq$ slope 2

The second order terms in the Taylor series may also be written in the form

$$G = \sum_{j=1}^{2} \sum_{k=1}^{2} h_{jk} \, \Delta x_j \, \Delta x_k \tag{9-8}$$

where

$$h_{jk} = \frac{1}{2} \frac{\partial^2 f}{\partial x_j \partial x_k} \bigg|_{a,b} \tag{9-9}$$

or alternately in vector matrix form as

$$G = (\Delta x)^T \, H(\Delta x) \tag{9-10}$$

where T denotes the transpose of the column matrix, Δx, that is,

$$\Delta x = \begin{pmatrix} \Delta x_1 \\ \\ \Delta x_2 \end{pmatrix}, \; (\Delta x)^T = (\Delta x_1, \Delta x_2), \; H = \begin{pmatrix} h_{11} & h_{12} \\ \\ h_{21} & h_{22} \end{pmatrix} \tag{9-11}$$

The quantity G will be referred to as a quadratic form and the matrix H as the Hessian matrix.

In general, whether a quadratic form has a positive or a negative value depends on the values assigned to Δx_1 and Δx_2. If for all values of Δx_1 and Δx_2, G has only positive values and is zero only if both Δx_1 and Δx_2 are zero, then G is said to be positive definite. Similarly, G is said to be negative definite when G takes on only negative values.

When G takes on a plus or minus sign depending on the values of Δx_1 and Δx_2, the quadratic form is said to be indefinite. A saddle point exists at a stationary point (a, b) where G is indefinite.

If, for all values of Δx_1 and Δx_2, we have $G \geq 0$, then G is said to be <u>positive semidefinite</u>, and if we have $G \leq 0$ for all values of Δx_1 and Δx_2, G is said to be <u>negative semidefinite</u>. For a semidefinite quadratic form, G may be zero for values of Δx_1 and Δx_2 other than zero. Geometrically speaking a semidefinite point is a point of inflection in one direction and a maximum or minimum point in the other.

There are several ways in which we may determine whether or not a given quadratic form is a positive definite quadratic form. One of the simplest ways is to apply <u>Sylvester's theorm</u> [Fan, Erickson, Hwang, 1971], which states that the necessary and sufficient condition for a quadratic form G to be positive definite is that each of the principal minors of the matrix, H, of the quadratic form be greater than zero. That is, if

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}$$

with $h_{12} = h_{21}$, then a necessary and sufficient condition that G be greater than zero is that each of the determinants

$$H_1 = h_{11} \text{ and } H_2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}$$

be greater than zero.

For a negative definite quadratic form, the preceding theorem may be applied by considering the form, minus G, that is, if the sign of every element in the matrix H is changed, then for the new matrix H' each of the determinants H_1' and H_2' should be greater than zero.

The functions which contain more than two decision variables may be treated in a way similar to the case of the two decision variables. A Taylor series expansion of the function

$$S = f(x_1, x_2, \ldots, x_n) \tag{9-12}$$

about a stationary point $a = (a_1, a_2, \ldots, a_n)$ gives

$$f(a_1 + \Delta x_1, a_2 + \Delta x_2, \ldots, a_n + \Delta x_n) - f(a_1, a_2, \ldots, a_n)$$

$$= \sum_{k=1}^{n} \left.\frac{\partial f}{\partial x_k}\right|_a \Delta x_k + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left.\frac{\partial^2 f}{\partial x_j \partial x_k}\right|_a \Delta x_j \Delta x_k + \ldots \tag{9-13}$$

Here the quadratic form is given by

$$G = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left.\frac{\partial^2 f}{\partial x_j \partial x_k}\right|_a \Delta x_j \Delta x_k \tag{9-14}$$

In terms of matrices we may write

$$G = (\Delta x)^T H (\Delta x) \tag{9-15}$$

where T denotes the transpose and where

$$(\Delta x) = \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix}, \quad (\Delta x)^T = (\Delta x_1, \Delta x_2, \ldots, \Delta x_n), \quad H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \\ h_{n1} & h_{ns} & \cdots & h_{nn} \end{pmatrix} \tag{9-16}$$

The elements, h_{jk}, of the matrix H are

$$h_{jk} = \frac{1}{2} \frac{\partial^2 f}{\partial x_j \partial x_k} \bigg|_a \qquad \begin{array}{l} j = 1, 2, \ldots, n \\ k = 1, 2, \ldots, n \end{array} \tag{9-17}$$

Sylvester's theorem may again be used to determine if the quadratic form, G, is positive definite. The necessary and sufficient condition that G be a positive definite quadratic form is that each of the principal minors of H be greater than zero, that is, each of the determinants

$$H_1 = h_{11}, \quad H_2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}, \quad \ldots, \quad H_n = \begin{vmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{vmatrix} \tag{9-18}$$

must be greater than zero.

A necessary and sufficient condition for G to be negative definite quadratic form is that - G be a positive definite quadratic form or that each of the principal minors of the matrix - H be greater than zero according to the conditions given in equation (9-18). If we wish to test H directly for the negative definite case, the necessary and sufficient condition for G to be negative definite is that the signs of H_i be alternately negative and positive as i goes from 1 to n with H_1 being negative, that is, H_i is negative if i is odd and positive if i is even.

9.5 LINEAR FUNCTIONS AND CONVEX SET

Linear functions are both concave and convex functions, but they are neither strictly concave nor strictly convex.

A convex set is defined as a set of points in a n-dimensional space, if for all pairs of two points x', x'' in the set, the straight line segment joining them is also entirely in the set. Thus in a convex set, any point x, where

$$x = (1-\lambda)x' + \lambda x'' \qquad \text{for } 0 \leq \lambda \leq 1$$

is also in the set.

Figure 10 illustrate a convex and a non convex set. It is clear from the figure that a line joining any two points will also lie within the set for a convex set whereas it is not true for a noncovex set.

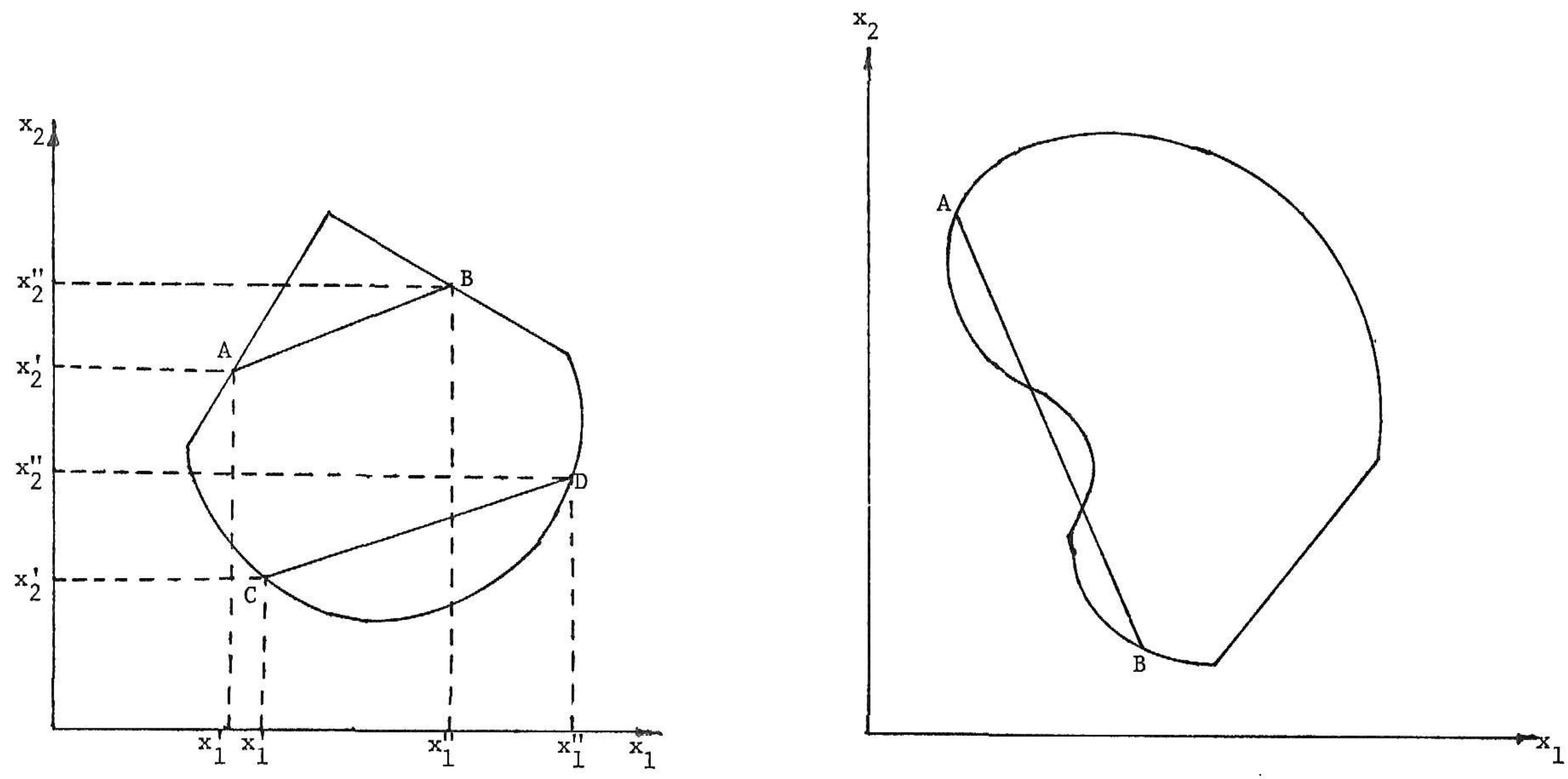
If a function $g(x)$ is a convex function over the non-negative orthant of the n-dimensional (euclidean) spaces, E^n, then, if the set of points V satisfying

$$g(x) \leq b, \qquad x \geq 0$$

is not empty, V is a convex set. Similarly, if $g(x)$ is a concave function over the non-negative orthant of E^n, then if the set of points satisfying

$$g(x) \geq b, \qquad x \geq 0$$

is not empty, V is a convex set. Consequently, since the intersection of convex sets is convex, if the set of points W satisfying



convex set

non-convex set

Fig. 10 A convex set and a non-convex set.

$$g_i(x)\{\leq,\geq\}b_i, \qquad i = 1, \ldots, m$$

$$x \geq 0$$

is not empty, then W is convex when $g_i(x)$ is a convex function over the non-negative orthant of E^n if a $\leq$ sign holds for the ith constraint and $g_i(x)$ is a concave function over the non-negative orthant of E^n if a $\geq$ sign holds for the ith constraint. These statements are proved by Hadley [1964].

10. THE NECESSARY AND SUFFICIENT CONDITIONS FOR A GLOBAL OPTIMUM

The Kuhn-Tucker conditions (the necessary conditions given by equations (6-3) through (6-6) for $-\infty < x < \infty$, and by equations (6-7) through (6-11) for $x \geq 0$) provide the candidates for local minimum or local maximum points. The sufficient conditions for a local minimum (or local maximum) to be a global minimum (or maximum) are that the objective function must be convex (or concave) and the optimal point lies in a closed convex set formed by the constraints in E^n.

A general nonlinear programming problem can be again stated as:

Minimize (or maximize)

$$f(x) \tag{10-1}$$

subject to

$$g_j(x) \{\leq, =, \geq\} 0, \qquad j = 1, 2, \ldots, m \tag{10-2}$$

where

$$-\infty < x_i < \infty \quad , \qquad i = 1, 2, \ldots, n \tag{10-3}$$

or

$$x_i \geq 0 \quad , \qquad i = 1, 2, \ldots, n \tag{10-4}$$

In the published literature, the nonlinear programming problems are formulated in eight cases presented in Table 3 and each case was treated separately [Case I and II by Dorn (1963), Case III by Samuelson (1955), Case V and VII by Vajda (1961) and Dorfman (1958), Case VI by Wilde (1962) and Kuhn-Tucker (1951).]

The necessary conditions (the Kuhn-Tucker conditions) for Case I through Case IV which x can take any value ($-\infty < x < \infty$) are given by equations (6-3) through (6-6), and those for Case V through Case VIII which x takes only non-negative value ($x \geq 0$) are given by equations (6-6) through (6-10).

The sufficient conditions for a local optimal point to be a global optimal point for Case I and for Case V are identical. A local optimal point which

Table 3. Classification of Nonlinear Programming Problems

$-\infty < x_i < +\infty$, $i=1,\ldots,n$	$x_i \geq 0$, $i=1,\ldots,n$
Case I Min (or Max) $f(x)$ Subject to $g_j(x) \leq 0,\ j=1,\ldots,m$	Case V Min (or Max) $f(x)$ Subject to $g_j(x) \leq 0,\ j=1,\ldots,m$
Case II Min (or Max) $f(x)$ Subject to $g_j(x) \geq 0,\ j=1,\ldots,m$	Case VI Min (or Max) $f(x)$ Subject to $g_j(x) \geq 0,\ j=1,\ldots,m$
Case III Min (or Max) $f(x)$ Subject to $g_j(x) = 0,\ j=1,\ldots,m$	Case VII Min (or Max) $f(x)$ Subject to $g_j(x) = 0,\ j=1,\ldots,m$
Case IV Min (or Max) $f(x)$ Subject to $\sum_{i=1}^{n} a_{ji}x_i \leq b_j,\ j=1,\ldots,m$	Case VIII Min (or Max) $f(x)$ Subject to $\sum_{i=1}^{n} a_{ji}x_i \leq b_j,\ j=1,\ldots,m$

satisfies the Kuhn-Tucker conditions will be the global optimal point if the objective function is convex (for minimization) or concave (for maximization), and if the constraints form a closed convex set V in E^n in which the optimal point lies. The convexity or concavity of a function can be identified by Hessian matrix (see Sections 9-3 and 9-4). A set of points V satisfying the constraints

$$g_i(x) \leq 0, \qquad i = 1, 2, \ldots, m \qquad (10\text{-}5)$$

will be a closed convex set if $g_i(x)$, i=1, 2, ..., m are all convex functions.

It is worth recalling that if f(x) is a convex function, the negative of the function, -f(x), is a concave function and vice versa. Also, if $g_i(x) \geq 0$, the negative of function will change the inequality sign, that is, $-g_i(x) \leq 0$. Therefore, a set of point V satisfying the constraints

$$g_i(x) \geq 0, \qquad i = 1, 2, \ldots, m \qquad (10\text{-}6)$$

will be a closed convex set if $g_i(x)$, i = 1, 2, ..., m are all concave functions. Consequently, the sufficient conditions of a optimal point for Case II and Case VI to be a global optimum are that the local optimal point satisfies the Kuhn-Tucker conditions, the objective function is convex (for minimization) or concave (for maximization), and all the function, $g_i(x)$, i = 1, 2, ..., m are concave.

Since the linear constraints $(\sum_i a_{ji} x_i \leq b_j)$ form a closed convex set, the sufficient conditions for a local optimal point to be a global optimum for Cases IV and VIII are that the optimal point satisfies the Kuhn-Tucker conditions, and the objective function is convex (for minimization) and concave (for maximization).

The sufficient conditions for a local optimum to be a global optimum for Cases III and VII which have equality constraints, $g_i(x) = 0$, i = 1, 2, ..., m; will be treated at the next section.

EXAMPLE

An example of Case I will be considered in detail. The problem is:

Minimize

$$f(x) = x_1^2 + x_2^2 - 8x_1 + 16 \tag{10-7}$$

subject to

$$g(x) = x_1^2 - 4x_2 + 2 \leq 0 \tag{10-8}$$

Since the objective function and the constraint both are positive definite, both are convex. Therefore, the objective function and the set formed by the constraint both satisfy the sufficient conditions stated above and the stationary point given by the Kuhn-Tucker conditions should be a global minimum.

The function and constraint are plotted to illustrate that the sufficient conditions given above gives us the correct global minimum. In Figure 11, the function and the constraint are plotted on x_1 and x_2 axis. The object function is plotted as contour lines. In the figure, the feasible region is shown by sectional lines. Since the objective function is convex, it is increasing in value from centre contour line to outward contour line. The stationary point given by the Kuhn-Tucker conditions will be point A, because at this point the objective function just touches the constraint. Point A gives the minimum value of the function subject to the given constraint. As we go towards the centre of the contour lines, the value of the function decreases but all these values will fall outside the feasible region. Similarly if we go away from point A outward from the centre, the value of the function fall in the feasible region but it keeps on increasing. Thus the minimum value of the function, which satisfies the above constraint is at point A. So if the sufficient conditions stated above are satisfied, the stationary point given by the Kuhn-Tucker conditions for the function given by equation (10-7) subject to the constraint given by equation (10-8) will be a global minimum.

The objective function and the constraint are also plotted in three dimensions as shown in Figure 12. The plot of the function is a cone and the plot of the constraint is a paraboloid. The function and the constraint are also projected on $x_1 - x_2$ axis giving the same plot as is shown in Figure 11. From the figure its clear that the point where cone (the plot of function) touches the paraboliod (the plot of constraint) is the optimum point, because that is the only point which minimizes the function and is in the feasible region also.

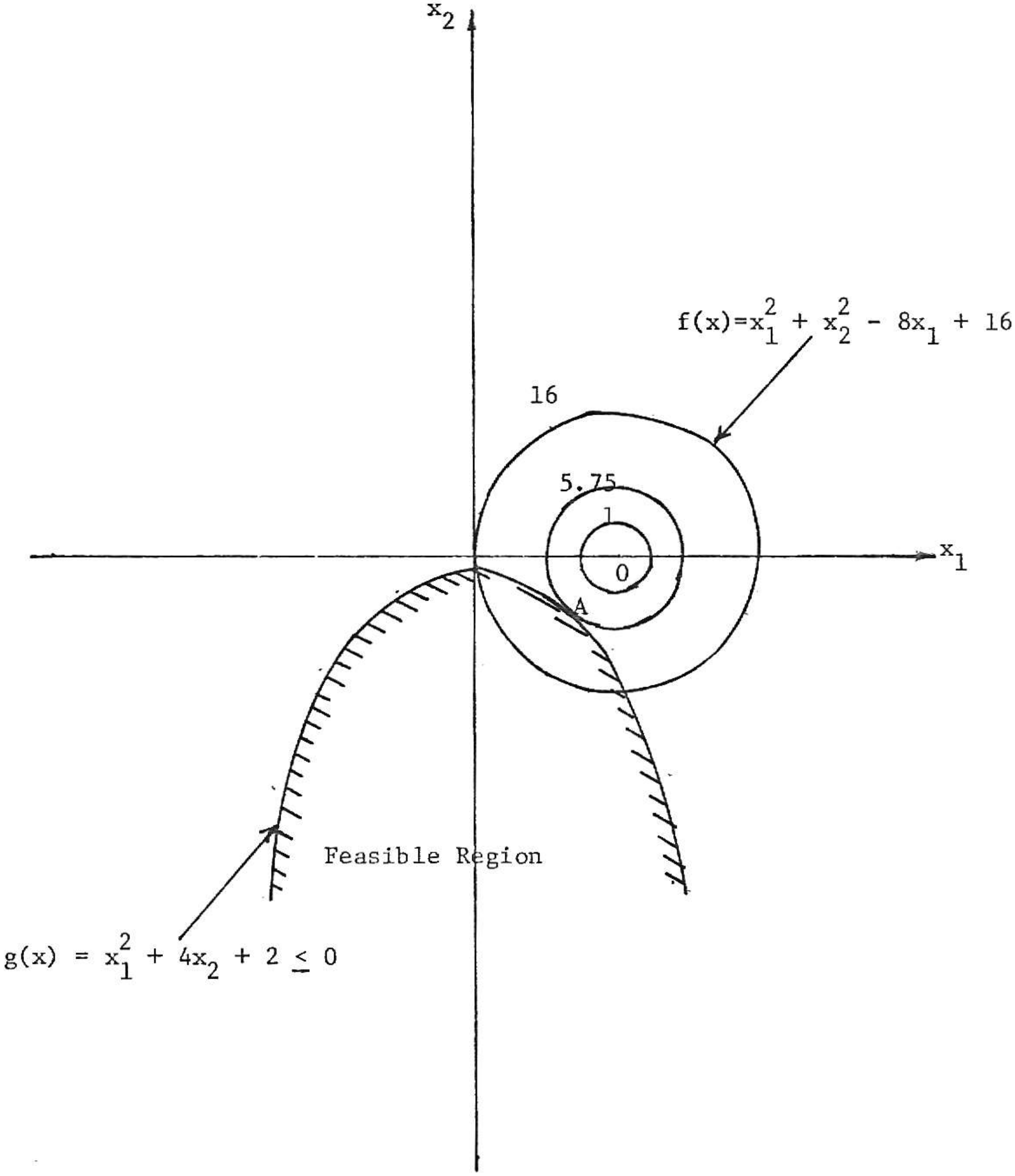


Fig. 11 Plot of the objective function $f(x) = x_1^2 + x_2^2 - 8x_1 + 16$ and the constraint $g(x) = x_1^2 + 4x_2 + 2 \leq 0$ in two dimensions.

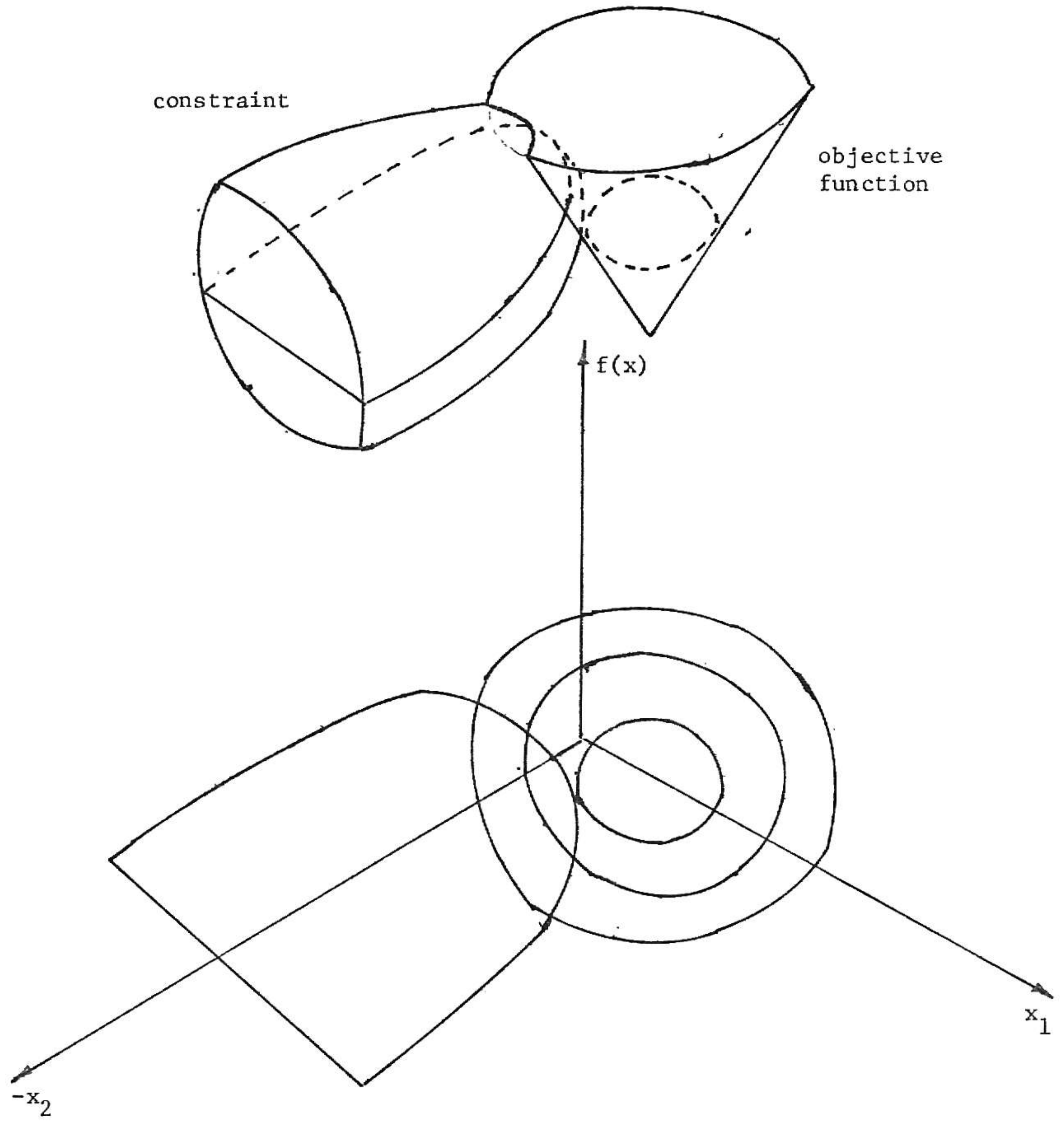


Fig. 12 Plot of the objective function $f(x) = x_1^2+x_2^2-8x_1+16$ and the constraint $g(x) = x_1^2-4x_2+2 \leq 0$ in three dimensions.

11. DUAL PROBLEM AND SADDLE POINT

Let us assume that the objective function

$$S = f(x) \tag{11-1}$$

has a relative maximum at a point x_o for x satisfying the m equality constraints,

$$g_j(x) = b_j, \qquad j = 1, \ldots, m \tag{11-2}$$

Let $\lambda_o^T = (\lambda_1^o, \ldots, \lambda_m^o)$ be a vector containing a set of Lagrange multipliers corresponding to x_o where the Lagrangian is given by

$$L(x,\lambda) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x) \tag{11-3}$$

Then, the Lagrangian, $L(x,\lambda)$, used to have an unconstrained relative maximum at x_o, that is,

$$L(x,\lambda_o) \leq L(x_o,\lambda_o) \tag{11-4}$$

for all x in a certain ε-neighborhood of x_o. Now, let us assume that there exists a δ-neighborhood of λ_o in E^m such that for any λ in this δ-neighborhood, the Lagrangian, $L(x,\lambda)$, has an unconstrained relative maximum with respect to x at a point which satisfies the condition

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_i}{\partial x_i} = 0, \quad i = 1, \ldots, n \tag{11-5}$$

Suppose that equation (11-5) has a unique solution for every λ in the δ-neighborhood of λ_o, then the maximizing point $\hat{x}$ will be a function of λ and can be written as $\hat{x}(\lambda)$. Then, in the δ-neighborhood of λ_o we can write

$$\max_x L(x,\lambda) = L(\hat{x},\lambda) = h(\lambda) \tag{11-6}$$

that is, the maximum of the Lagrangian with respect to x will be a function of λ which is denoted by $h(\lambda)$. It is worth noting that the maximum of the Lagrangian, $L(\hat{x},\lambda)$, at λ_o will be $L(x_o,\lambda_o)$ and equals to $f(x_o)$, that is,

$$h(\lambda_o) = L(x_0,\lambda_o) = f(x_o) \tag{11-7}$$

Next, we will examine whether or not $h(\lambda)$ has a relative maximum or minimum at λ_o. For x satisfying the m equality constraints, equation (11-2), and for a fixed λ, the Lagrangian will be

$$L(x,\lambda) = f(x)$$

Hence

$$\max_x L(x,\lambda) = \max_x f(x) = f(x_o) = h(\lambda_o), \quad x \in Y \tag{11-8}$$

where Y is a set of points x satisfying $g_j(x) = b_j$, $j = 1, \ldots, m$. Since in equation (11-8) x is limited to satisfy the equality constraints given by equation (11-2) whereas in equation (11-6) x is not limited to the constraints, the value of the function $h(\lambda)$ given by equation (11-6) should be greater than or equal to the value of the function $h(\lambda_o)$ given by equation (11-8), that is,

$$h(x) \geq h(\lambda_o) \tag{11-9}$$

and hence $h(\lambda)$ has a relative minimum at λ_o. Since $h(\lambda) = L(x,\lambda)$ when x and λ are related by equation (11-5), we may conclude that $L(x,\lambda)$ has a relative minimum at λ_o with respect to λ subject to the constraints given by equation (11-5) which determine x for each λ.

The problem of minimizing the Lagrangian, $L(x,\lambda)$, given by equation (11-3), with respect to λ subject to the constraints (11-5), is called a dual of the primal problem which seeks to maximize $f(x)$, equation (11-1), subject to the constraints (11-2). These dual problems have the property that in a neighborhood of x_o,

$$\min_{\lambda} L(x,\lambda) = \max_{x} f(x) = f(x_o) = h(\lambda_o) \tag{11-10}$$

subject to the appropriate constraints, equation (11-5) and (11-2) respectively.

From equation (11-6) in the above discussion, we have

$$\max_{x} L(x,\lambda) = h(\lambda)$$

and from equation (11-9), we have

$$\min_{\lambda} h(\lambda) = h(\lambda_o)$$

therefore, we can write

$$f(x_o) = \min_{\lambda} \max_{x} L(x,\lambda) \tag{11-11}$$

and $f(x_o)$ is a solution to a min-max problem. A function $L(x,\lambda)$ is said to have a saddle point at the point (x_o,λ_o), if there exists an

$\varepsilon > 0$ such that for all x, $|x - x_o| < \varepsilon$, and for all λ, $|\lambda-\lambda_o| < \varepsilon$,

$$L(x,\lambda_o) \leq L(x_o, \lambda_o) \leq L(x_o,\lambda) \tag{11-12}$$

In the above discussion, we have shown that if all assumptions made above hold, the Lagrangian function has a saddle point at (x_o,λ_o). However, the saddle point is a degenerate form since the strict equality holds on the right of the following equation.

$$L(x,\lambda_o) \leq L(x_o,\lambda_o) = L(x_o,\lambda)$$

This is deduced from our previous assumption that $L(x,\lambda)$ has a relative maximum with respect to x for any λ in a δ-neighborhood of λ_o, that is

$$L(x,\lambda_o) \leq L(x_o,\lambda_o) = f(x_o)$$

and recall that $L(x_o,\lambda) = f(x_o)$ since $x_o \in Y$.

The material presented here is originally covered in Courant and Hilbert (1953). The above presentation follows closely that of Hadley (1964).

Example

The linear programming (LP) dual problem is a special case of the general nonlinear programming problem discussed above.

Let the primal LP problem be

$$\max_{x} \; c^T x$$

subject to

$$Ax \leq b$$

$$x \geq 0$$

Then, the Lagrangian will be

$$L(x,\lambda) = c^T x - \lambda^T(Ax - b)$$

The dual problem becomes

$$\min_{\lambda} L(x,\lambda) = \min_{\lambda} [c^T x - \lambda^T(Ax - b)]$$

subject to (equation (11-5))

$$\frac{\partial L}{\partial x} = 0 = c^T - \lambda^T A$$

$$\lambda \geq 0$$

However, $c^T - \lambda^T A = 0$ implies

$$c^T x - \lambda^T Ax = 0$$

Therefore, the dual problem reduces to the usual form of the dual LP problem

$$\min_{\lambda} \lambda^T b$$

subject to

$$A^T \lambda = c$$

$$\lambda \geq 0$$

12. NECESSARY AND SUFFICIENT CONDITIONS FOR SADDLE POINTS

In the last section, we explained that a function $L(x,\lambda)$, x being an n - component and λ being an m-component vector, is said to have a saddle point at (x_0, λ_0) if

$$L(x, \lambda_0) \leq L(x_0, \lambda_0) \leq L(x_0, \lambda) \tag{12-1}$$

holds for all x in an ε neighborhoods of x_0 and all λ in an ε neighborhoods of λ_0. This saddle point will also be a global saddle point at (x_0, λ_0) if equation (12-1) holds for all x and λ.

In this section, we will develop conditions which will insure in finding the saddle point for the function $L(x,\lambda)$ and hence the optimum point for the objective function $f(x)$ subject to the constraints.

We will consider three possibilities in which variables x and λ can vary. 1) x,λ are restricted to be non-negative. These will be represented by $x^{(1)}$ and $\lambda^{(1)}$ ($x^{(1)} \geq 0$ and $\lambda^{(1)} \geq 0$). 2) x,λ are restricted to be non-positive. These will be represented by $x^{(2)}$ and $\lambda^{(2)}$ ($x^{(2)} \leq 0$ and $\lambda^{(2)} \leq 0$). 3) x,λ are unrestricted in sign. These will be represented by $x^{(3)}$ and $\lambda^{(3)}$.

Now x and λ in three components can be represented as follows

$$x = [x^{(1)}, x^{(2)}, x^{(3)}]$$

$$\lambda = [\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}]$$

where $x^{(1)}$ has s components, $x^{(2)}$ has t-s components, and $x^{(3)}$ has n-t components. Similarly, $\lambda^{(1)}$ has u components, $\lambda^{(2)}$ has v-u components, and $\lambda^{(3)}$ has m-v components.

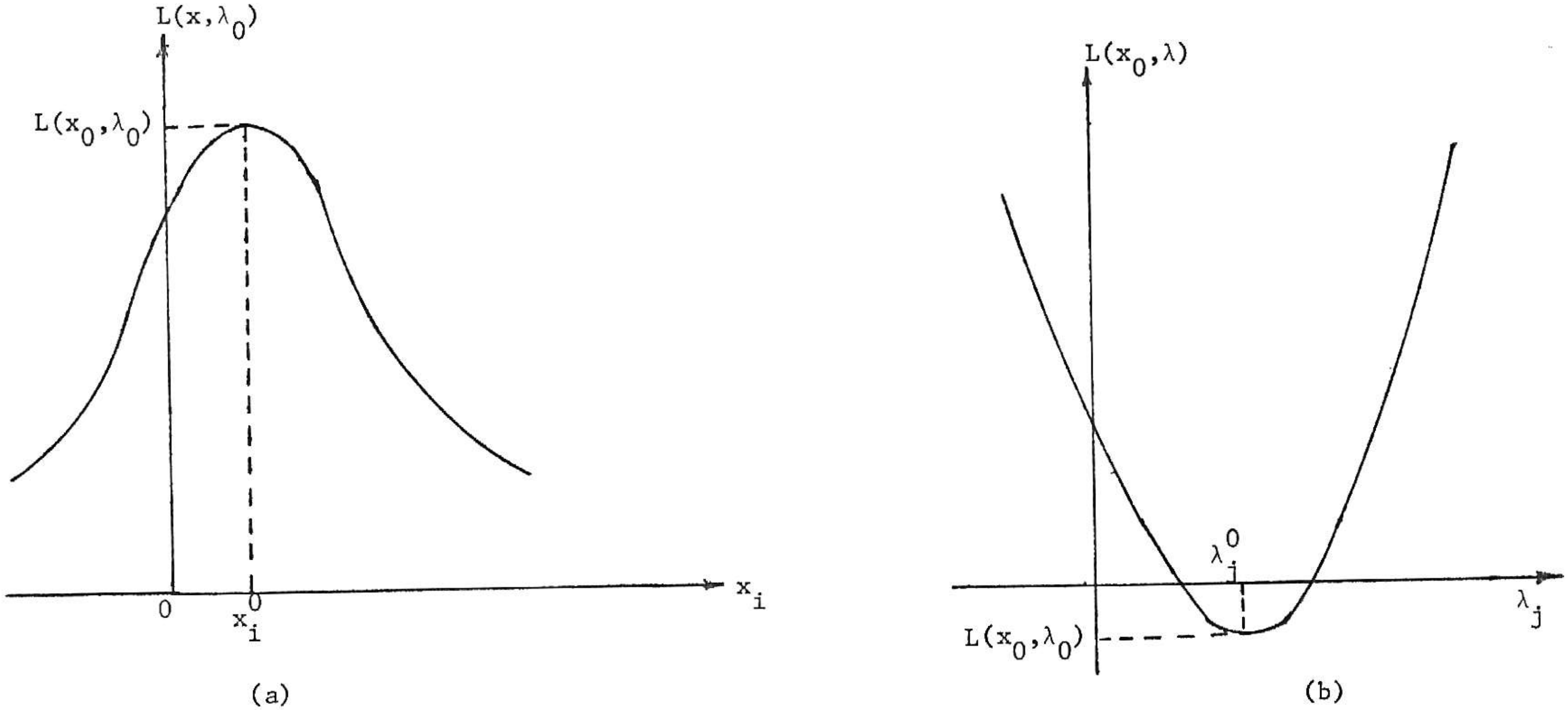


Fig. 13 Saddle point of $L(x,\lambda)$ when x and λ are unrestricted.

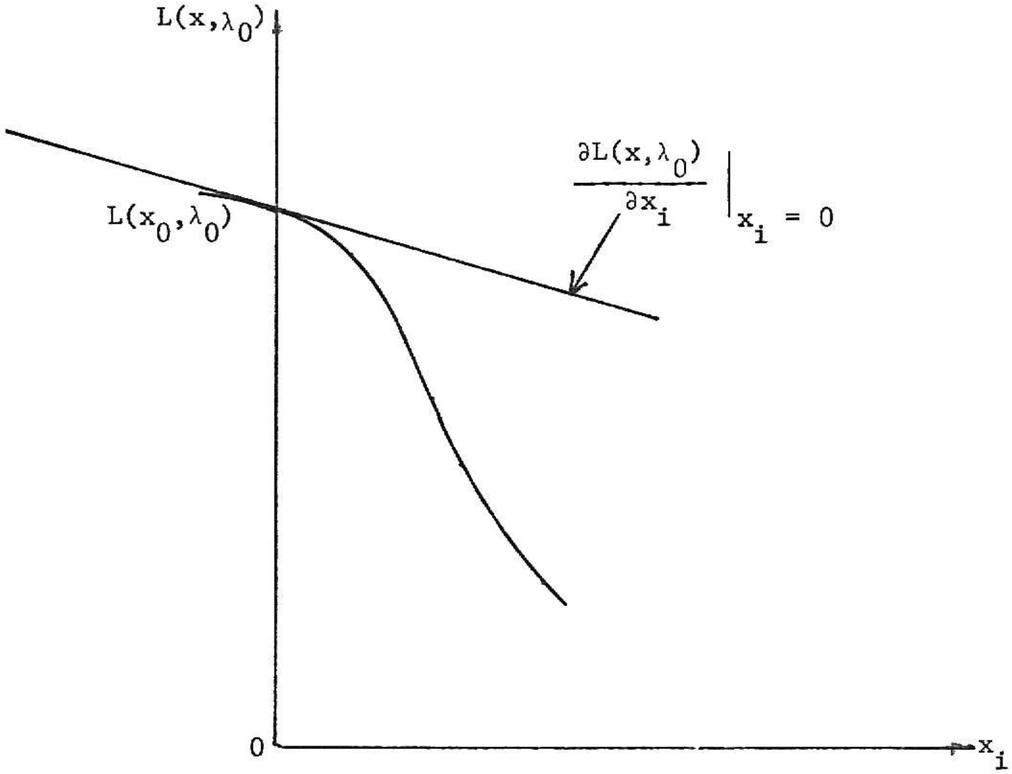


Fig. 14 For maximizing $L(x,\lambda_0)$ and non-negative x_i, $\frac{\partial L}{\partial x_i} \leq 0$ at $x_i = 0$.

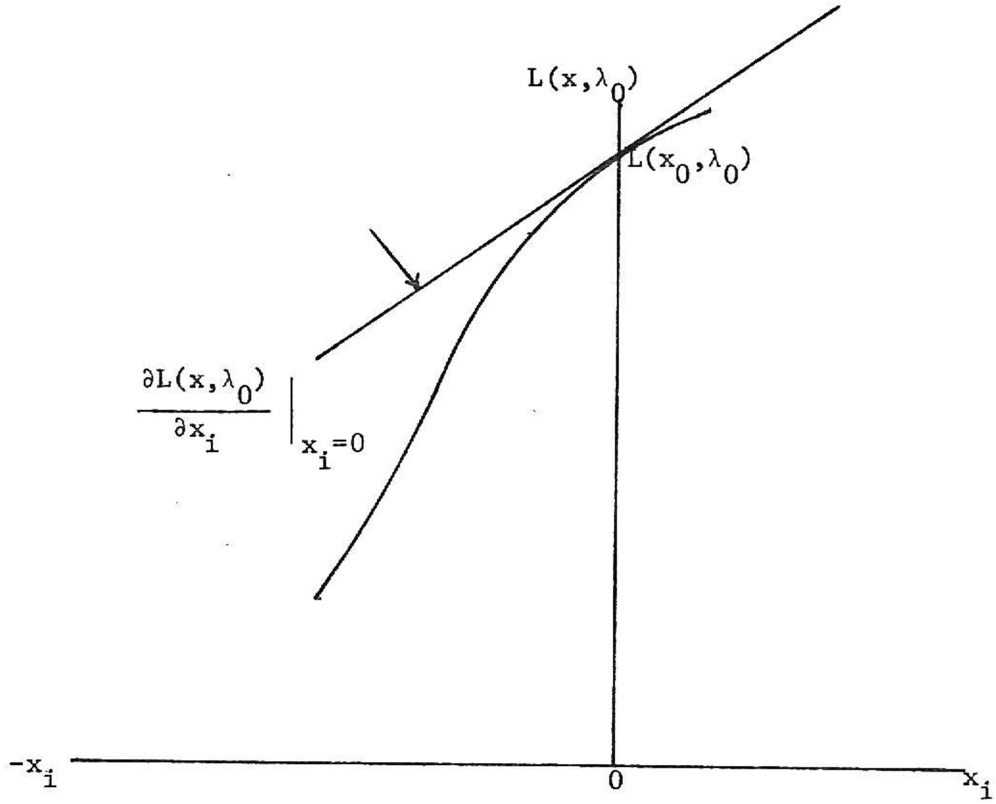


Fig. 15 For maximizing $L(x,\lambda_0)$ and non-positive x_i, $\partial L/\partial x_i \geq 0$ at $x_i = 0$

Let us also denote by W_1, the set of points of x such that the components of x satisfy the above conditions, by W_2, the set of points λ such that the components of λ satisfy the above conditions and by W the set of points $[x,\lambda]$ with $x \varepsilon W_1$ and $\lambda \varepsilon W_2$.

Now the function $L(x,\lambda)$ is said to have a saddle point at (x_0,λ_0) for $(x,\lambda) \varepsilon W$ if $(x_0,\lambda_0) \varepsilon W$ and there exists an $\varepsilon > 0$ such that equation (12-1) holds good for all $x \varepsilon W_1$ in an ε neighborhood of x_0 and all $\lambda \varepsilon W_2$ in an ε neighborhood of λ_0.

We will develop the Kuhn-Tucker conditions for the three cases explained above. Case 3 will be considered first in which the components of x and λ are not restricted in signs. Therefore, if $L(x,\lambda)$ has a saddle point at (x_0,λ_0) for $(x,\lambda) \varepsilon W$, then from the definition of a saddle point, (x_0,λ_0) must satisfy the equations

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) = 0, \qquad i = t + 1, \ldots, n \tag{12-2}$$

$$\frac{\partial}{\partial \lambda_j} L(x_0,\lambda_0) = 0, \qquad j = v + 1, \ldots, m \tag{12-3}$$

Figure 13is a plot of Lagrangian $L(x,\lambda_0)$ as a function of x_i when i has one of the values $t + 1, \ldots, n$, and set $x_k = x_k^0$, $k \neq i$, $\lambda = \lambda_0$, then in a neighborhood of x_i^0, $L(x,\lambda_0)$ must have a maximum point where $\frac{\partial L}{\partial x_i} = 0$ at (x_0,λ_0). Under similar conditions $L(x_0,\lambda)$ as a function of λ_j, where j has one of the values $v+1, \ldots, m$, would pass through a relative minimum where $\frac{\partial L}{\partial \lambda_j} = 0$ at (x_0,λ_0).

It should be noted that if $x_i^0 \neq 0$ for one of the first t components of x, it must again be true that

$$\frac{\partial L(x_0,\lambda_0)}{\partial x_i} = 0 \tag{12-4}$$

and also $\lambda_j^0 \neq 0$ for one of the first v components of λ, it must be true that

$$\frac{\partial L(x_0,\lambda_0)}{\partial \lambda_j} = 0 \tag{12-5}$$

Let us now consider the situation that $x_i^0 = 0$, where i has one of the values 1, ..., t. Assume *first that i comes from the set* 1, ..., s, so that x_i must be non-negative. Then in order that $L(x,\lambda_0) \leq L(x_0,\lambda_0)$ for $(x,\lambda_0) \varepsilon W$, it must true that

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) \leq 0 \tag{12-6}$$

Intuitively if we plot $L(x,\lambda_0)$ as a function of x_i when setting $x_k = x_k^0$, $k \neq i$, $\lambda = \lambda_0$, we must obtain something like the curve shown in Fig. 14 Similarly if i belongs to the set s+1, ..., t (x_i must be non-positive) and $x_i^0 = 0$, then in order that $L(x,\lambda_0) \leq L(x_0,\lambda_0)$ for $(x,\lambda_0) \varepsilon W$, it must be true that

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) \geq 0 \tag{12-7}$$

Figure 15 illustrates this situation.

Similarly, if $\lambda_j^0 = 0$, it must be true that

$$\frac{\partial L(x_0,\lambda_0)}{\partial \lambda_j} \geq 0, \quad j = 1, \ldots, u \tag{12-8}$$

$$\frac{\partial L(x_0,\lambda_0)}{\partial \lambda_j} \leq 0, \qquad j = u+1, \ldots, v \tag{12-9}$$

In the above discussion we have shown that it must be true that either $\partial L(x_0,\lambda_0)/\partial x_i = 0$ or $x_i^0 = 0$, and either $\partial L(x_0,\lambda_0)/\partial \lambda_j = 0$ or $\lambda_j^0 = 0$. Hence, if $L(x,\lambda)$ has a saddle point at (x_0,λ_0) for $(x,\lambda) \varepsilon W$, then (x_0,λ_0) must satisfy

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) \leq 0, \qquad i = 1, \ldots, s \tag{12-10}$$

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) \geq 0, \qquad i = s + 1, \ldots, t \tag{12-11}$$

$$\frac{\partial}{\partial x_i} L(x_0,\lambda_0) = 0, \qquad i = t + 1, \ldots, n \tag{12-12}$$

where

$$x_i^0 \geq 0,\ i = 1, \ldots, s; \quad x_i^0 \leq 0, \quad i = s + 1, \ldots, t;$$

$$x_i^0 \text{ unrestricted}, \quad i = t + 1, \ldots, n \tag{12-13}$$

The above equations can be summarized as follows

$$x_i^0 \frac{\partial}{\partial x_i} L(x_0,\lambda_0) = 0, \qquad i = 1, \ldots, n \tag{12-14}$$

Similarly

$$\frac{\partial}{\partial \lambda_j} L(x_0,\lambda_0) \geq 0, \qquad j = 1, \ldots, u \tag{12-15}$$

$$\frac{\partial}{\partial \lambda_j} L(x_0,\lambda_0) \leq 0, \qquad j = u+1, \ldots, v \tag{12-16}$$

$$\frac{\partial}{\partial \lambda_j} L(x_0,\lambda_0) = 0, \qquad j = v+1, \ldots, m \tag{12-17}$$

where

$$\lambda_j^0 \geq 0, \quad j = 1, \ldots, u; \qquad \lambda_j^0 \leq 0, \quad j = u+1, \ldots, v;$$

$$\lambda_j^0 \text{ unrestricted}, \quad j = v+1, \ldots, m \tag{12-18}$$

The above equations can be summarized as follows

$$\lambda_j^0 \frac{\partial}{\partial \lambda_j} L(x_0,\lambda_0) = 0, \qquad j = 1, \ldots, m \tag{12-19}$$

The above equations, equations (12-10) through (12-19), represent a set of necessary conditions which (x_0,λ_0) must satisfy if $L(x,\lambda)$ has a saddle point at (x_0,λ_0) for $(x,\lambda) \in W$.

Now we will develop the sufficient conditions which must be satisfied to ensure that $L(x,\lambda)$ has a saddle point at (x_0,λ_0) for $(x,\lambda) \in W$. Let us define the gradient of $L(x,\lambda)$ with respect to x evaluated at (x_0,λ_0) to be

$$\nabla_x L(x_0,\lambda_0) = \left[\frac{\partial}{\partial x_1} L(x_0,\lambda_0), \ldots, \frac{\partial}{\partial x_n} L(x_0,\lambda_0)\right] \tag{12-20}$$

and the gradient of $L(x,\lambda)$ with respect to λ evaluated at (x_0,λ_0) to be

$$\nabla_\lambda L(x_0,\lambda_0) = \left[\frac{\partial}{\partial \lambda_1} L(x_0,\lambda_0), \ldots, \frac{\partial}{\partial \lambda_m} L(x_0,\lambda_0)\right] \tag{12-21}$$

Then on summing equation (12-14) over i, we have

$$\nabla_x L(x_0,\lambda_0)x_0 = 0 \tag{12-22}$$

Similarly, on summing equation (12-19) over j we obtain

$$\nabla_\lambda L(x_0,\lambda_0)\lambda_0 = 0 \tag{12-23}$$

Now the sufficient conditions can be defined as follows. Let (x_0,λ_0) be a point satisfying necessary conditions (12-10) through (12-19). Then if there exists an ε neighborhood about (x_0,λ_0) such that for points $(x,\lambda_0) \in W$ in this neighborhood

$$L(x,\lambda_0) \leq L(x_0,\lambda_0) + \nabla_x L(x_0,\lambda_0)\ (x-x_0) \tag{12-24}$$

and for points $(x_0,\lambda) \in W$ in this neighborhood

$$L(x_0,\lambda) \geq L(x_0,\lambda_0) + \nabla_\lambda L(x_0,\lambda_0)\ (\lambda-\lambda_0) \tag{12-25}$$

then $L(x,\lambda)$ has a saddle point at (x_0,λ_0) for $(x,\lambda) \in W$. If equation (12-24) and (12-25) hold for all $x \in W_1$, and $\lambda \in W_2$, it follows that $L(x,\lambda)$ has a global saddle point at (x_0,λ_0) for $(x,\lambda) \in W$.

The sufficient conditions can be proved as follows. Since equation (12-22) holds at (x_0,λ_0)

$$\nabla_x L(x_0,\lambda_0)(x-x_0) = \nabla_x L(x_0,\lambda_0)x$$

However, if $x \in W_1$, then

$$x_i \geq 0, \quad i = 1, \ldots, s; \quad x_i \leq 0, \quad i = s+1, \ldots, t$$

and equations (12-10) and (12-11) hold. Then

$$\nabla_x L(x_0,\lambda_0)x \leq 0$$

Consequently, whenever equation (12-24) holds for $x \varepsilon W$, $L(x,\lambda_0) \leq L(x_0,\lambda_0)$. Similarly, since equations (12-23), (12-15) and (12-16) hold at (x_0,λ_0), we obtain

$$\nabla_\lambda L(x_0,\lambda_0)(\lambda-\lambda_0) = \nabla_\lambda L(x_0,\lambda_0)\, \lambda \geq 0$$

Therefore, for any $\lambda \varepsilon W_2$ such that equation (12-25) holds, $L(x_0,\lambda_0) \leq L(x_0,\lambda)$. Combination of these results we can conclude that for $x \varepsilon W_1$ and $\lambda \varepsilon W_2$ such that equations (12-24) and (12-25) hold, $L(x,\lambda_0) \leq L(x_0,\lambda_0) \leq L(x_0,\lambda)$, provided that (x_0,λ_0) satisfies equation (12-10) though (12-19). Since equations (12-24) and (12-25) hold either in an ε- neighborhood of (x_0,λ_0)(or every where), it follows that $L(x,\lambda)$ has a saddle point (or a global saddle point) at (x_0,λ_0) for $(x,\lambda) \varepsilon W$.

We can also deduce one more interesting relation from the fact that if $L(x,\lambda_0)$ is a concave function of x then equation (12-24) holds good and similarly if $L(x_0,\lambda)$ is a convex *function of* λ then equation (12-25) holds. Thus another equivalent sufficient condition that $L(x,\lambda)$ has a saddle point at (x_0,λ_0) is that $L(x_0,\lambda_0)$ satisfied equations (12-10) through (12-19) and that for all $x \varepsilon W_1$ in an ε neighborhood of x_0, $L(x,\lambda_0)$ is a concave function of x, and for $\lambda \varepsilon W_2$ in an ε neighborhood of x_0, $L(x_0,\lambda)$ is a convex function of λ. If $L(x,\lambda_0)$ is a concave function of x for all $x \varepsilon W_1$ and $L(x_0,\lambda)$ is a convex function of λ for all $\lambda \varepsilon W_2$, then $L(x,\lambda)$ has a global saddle point at (x_0,λ_0) for $(x,\lambda) \varepsilon W$.

REFERENCES

1. Beveridge, S. G., and Schechter, R. S. (1970), Optimization-Theory and Practice, McGraw-Hill, New York, pp. 244-287.

2. Dorfman, R., Samuelson, P. A. and Solow, R. M. (1958), Linear Programming and Economic Analysis, McGraw-Hill, New York.

3. Dorn, W. S. (1963), Non-Linear Programming - A Survey, Management Science, vol. 9, pp. 171-208.

4. Fan, L. T., Erickson, L. E. and Hwang, C. L. (1971), Methods of Optimization, Vol. 2 Equality Constraints and Optimization, Institute for Systems Design and Optimization, Kansas State University.

5. Fan, L. T., Erickson, L. E. and Hwang, C. L. (1971), Methods of Optimization, Vol. 1 Unconstrained Optimization and the Differential Calculus, Institute for Systems Design and Optimization, Kansas State University.

6. Hadley, G. (1964), Nonlinear and Dynamic Programming, Addison-Wesley, Reading, Massachusetts, Chapters 1,3 and 6.

7. Hiller, F. S. and Liberman, G. J. (1967), Introduction to Operations Research, Holden-Day Inc., pp. 574-580.

8. Himmelblau, D. M. (1972), Applied Nonlinear Programming, McGraw-Hill, New York.

9. Kuhn, H. W. and Tucker, A. W. (1951), Non-Linear Programming, Proceedings of the Second Berkeley Symposium on Mathematical Statics and Probability, J. Neyman, Ed., University of California Press, Berkeley, California, pp. 481-492.

10. McMillan, C. (1970), Mathematical Programming, John Wiley & Sons, New York.

11. Mangasarian, O. L. (1969), Nonlinear Programming, McGraw Hill, New York.

12. Samuelson, P. A. (1955), Foundations of Economic Analysis, Harvard University Press, Cambridge, Mass.

13. Teichroew, D. (1964), An Introduction to Management Science, John Wiley & Sons, New York, pp. 295-312, pp. 551-568.

14. Tucker, A. W. (1957), Linear and Nonlinear Programming, Operations Research, Vol. 5, pp. 244-257.

15. Vajda, S. (1961), Mathematical Programming, Addison-Wesley, Reading, Mass, Chapter 12.

16. Wilde, D. J. (1962), Differential Calculus in Nonlinear Programming, Operations Research, 10, pp. 764-773.

17. Wolfe, P. (1959), The Simplex Method for Quadratic Programming, Econometrica, 27, pp. 382-398.

18. Zangwill, W. I. (1969), Nonlinear Programming - A Unified Approach, Prentice Hall, Englewood Cliffs, New Jersey, Chapters 2 and 3.

13. CONCLUDING REMARKS

A self-contained, brief and illustrative description of the classical optimization procedures, the method of Lagrange multipliers and the Kuhn-Tucker conditions, for solving nonlinear programming problems has been presented.

It is usually difficult, if not impossible, to derive the optimal solution for a large scale nonlinear programming problem directly from the Kuhn-Tucker conditions. However, they do provide valuable clues as to the identity of the optimal solution, and they also permit checking whether a proposed solution may be optimal. It is not necessarily true that every point which is a solution to the Kuhn-Tucker conditions will be a point at which the objective function takes on a relative maximum or minimum for all X satisfies the constraints. However, every point at which the objective function does take on a relative maximum or minimum for X satisfies the constraints must be a solution to the Kuhn-Tucker conditions. There are also many valuable indirect applications of the Kuhn-Tucker conditions. One example of them is quadratic programming.

ACKNOWLEDGEMENTS

The author thanks Dr. C. L. Hwang for his valuable guidance and encouragement in the preparation of this report. The author also thanks Dr. F. A. Tillman and Dr. L. T. Fan for serving on the advisory committee of his graduate study.

METHOD OF LAGRANGE MULTIPLIERS AND

THE KUHN-TUCKER CONDITIONS

by

PRAMOD KUMAR GUPTA

B.S., BIRLA INSTITUTE OF TECHNOLOGY, INDIA, 1969

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1973

ABSTRACT

The purpose of this report is to give a self-contained, brief and illustrative description of the classical optimization procedures, the method of Lagrange multipliers and the Kuhn-Tucker conditions, for solving nonlinear programming problems.

Firstly the Lagrangian multiplier technique is presented for optimizing the nonlinear objective functions subject to equality constraints. The necessary and sufficient conditions are developed from Taylor's series expansion and numerical examples are solved to illustrate the technique.

The Kuhn-Tucker necessary conditions for optimizing an objective function subject to inequality constraints are then presented and are illustrated by a numerical example. The Wolfe's simplex method for optimal solution of quadratic programming problems which is based on the Kuhn-Tucker conditions is also presented.

The concept of convexity and concavity of functions, and of convex set, which is basis for determining whether a local optimum is also a global optimal, and then the necessary and sufficient conditions for a global optimum are included.

Finally the dual problem and saddle point and the necessary and sufficient conditions for saddle points are presented.