FUNDAMENTAL METHODS OF MATHEMATICAL ECONOMICS

Third Edition

Alpha C. Chiang

Professor of Economics
The University of Connecticut

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá Hamburg
Johannesburg London Madrid Mexico Montreal New Delhi
Panama Paris São Paulo Singapore Sydney Tokyo Toronto

This book was set in Times Roman by Science Typographers, Inc.
The editors were Patricia A. Mitchell and Gail Gavert;
the production supervisor was Leroy A. Young.
The cover was designed by Carla Bauer.
Halliday Lithograph Corporation was printer and binder.

FUNDAMENTAL METHODS OF MATHEMATICAL ECONOMICS

1 2 3 4 5 6 7 8 9 0 H A L H A L 8 9 8 7 6 5 4

ISBN 0-07-010813-7

Library of Congress Cataloging in Publication Data

Chiang, Alpha C., date
Fundamental methods of mathematical economics.

Bibliography: p.
Includes index.
1. Economics, Mathematical. I. Title.
HB135.C47 1984 330'.01'51 83-19609
ISBN 0-07-010813-7

CHAPTER
TWENTY-ONE
NONLINEAR PROGRAMMING

Linear programming, as discussed in the last two chapters, is in a very real sense an improvement over the classical optimization framework, since constraints may now enter into the problem as inequalities, and, accordingly, we can also explicitly introduce nonnegativity restrictions into the problem. However, the necessity of confining the objective function and the constraints to the linear mold can sometimes be a significant drawback. As a further improvement, therefore, we would welcome an optimization framework that can handle nonlinear objective functions as well as nonlinear inequality constraints. Such a framework is found in *nonlinear programming.*

21.1 THE NATURE OF NONLINEAR PROGRAMMING

The maximization problem of nonlinear programming has the following general format:

$$\begin{aligned}
&\text{Maximize} && \pi = f(x_1, x_2, \ldots, x_n) \\
&\text{subject to} && g^1(x_1, x_2, \ldots, x_n) \leq r_1 \\
(21.1)\qquad & && g^2(x_1, x_2, \ldots, x_n) \leq r_2 \\
& && \cdots\cdots\cdots\cdots\cdots \\
& && g^m(x_1, x_2, \ldots, x_n) \leq r_m \\
&\text{and} && x_j \geq 0 \qquad (j = 1, 2, \ldots, n)
\end{aligned}$$

Denoting the n-tuple of choice variables by the unsubscripted symbol x, we can write (21.1) more succinctly as

$$
\begin{array}{lll}
& \text{Maximize} & \pi = f(x) \\
(21.1') & \text{subject to} & g^i(x) \le r_i \qquad (i = 1, 2, \ldots, m) \\
& \text{and} & x \ge 0
\end{array}
$$

Similarly, the minimization problem can be written in the form of

$$
\begin{array}{lll}
& \text{Minimize} & C = f(x_1, x_2, \ldots, x_n) \\
& \text{subject to} & g^1(x_1, x_2, \ldots, x_n) \ge r_1 \\
(21.2) & & g^2(x_1, x_2, \ldots, x_n) \ge r_2 \\
& & \cdots\cdots\cdots\cdots\cdots \\
& & g^m(x_1, x_2, \ldots, x_n) \ge r_m \\
& \text{and} & x_j \ge 0 \qquad (j = 1, 2, \ldots, n)
\end{array}
$$

or, more succinctly,

$$
\begin{array}{lll}
& \text{Minimize} & C = f(x) \\
(21.2') & \text{subject to} & g^i(x) \ge r_i \qquad (i = 1, 2, \ldots, m) \\
& \text{and} & x \ge 0
\end{array}
$$

From the above, it is clear that a nonlinear program, like a linear one, also consists of three ingredients—an objective function, a set of m constraints, and a set of nonnegativity restrictions on the n choice variables. As in linear programming, m can be either greater than, equal to, or less than n. All the functions in the problem, $f(x)$ and $g^i(x)$, are assumed to be differentiable. Note, also, that in order to be consistent with our earlier discussion of linear programming, we are again adopting only the $\le$ type of constraints in the maximization problem and the $\ge$ type in the minimization problem. In case the reverse type of inequality occurs, it can readily be converted by multiplying through by -1.

The formulations in (21.1) and (21.2) are manifestly the most general optimization problems encountered so far in this volume. As such, they encompass all the previously discussed optimization problems as special cases. Linear programs are obviously special cases. If we take the objective function by itself, we have the free-extremum problem. And, finally, adding to the objective function the constraint section in its strict-equality version and with $m < n$, we obtain the classical problem of constrained optimization.*

* Actually, the connection between nonlinear programming and classical optimization is even closer than has been indicated. For it is possible, if desired, to transform a nonlinear-programming problem into a classical one by taking the following two steps: (1) introduce m dummy variables to convert the inequality constraints into equality ones; and (2) regard each of the n choice variables and each of the m dummy variables as the square of a new, artificial variable, to ensure nonnegativity. (See Exercise 21.1-6.)

Nonlinearities in Economics

Nonlinearities can arise in various ways. In the production problem in linear programming, the per-unit gross profit of each product was assumed to be a constant. But it can very well be a decreasing function of the output level, either because a larger output tends to depress the market price (average revenue), or because increased production tends to raise the average variable cost of the product. If so, the linear objective function $\pi = c_1 x + \cdots + c_n x_n$ must be replaced by a nonlinear version, such as $\pi = c_1(x_1)x_1 + \cdots + c_n(x_n)x_n$, where $c_j(x_j)$ denotes a decreasing function of the variable x_j.

Similarly, in the constraint section, it may happen that the input requirement for resource i in the production of product j decreases with the output level of product j. For instance, the later units of production may conceivably be processable at greater speed than the earlier ones, so that less machine time will be used up by each successive unit of output. This will, of course, undermine the constancy of the coefficient a_{ij}, as assumed in linear programming. It may also happen that the coefficient a_{ij} depends on the output level, not only of product j, but also of another product k. Then there will arise in the constraint section a term which involves the product of the two variables x_j and x_k, and linearity will again be lost.

Whenever the economic circumstances illustrated above are descriptive of the problem at hand, a nonlinear formulation will be more appropriate than a linear one. Unfortunately, many of the convenient features of linear programming will then become unavailable. This fact can be illustrated by some simple nonlinear programs that can be solved graphically.

Graphical Solution

We shall present here three specific examples, each of which will serve to spotlight certain features that distinguish nonlinear programming from linear programming.

Example 1 Minimize $\quad C = (x_1 - 4)^2 + (x_2 - 4)^2$

subject to
$$2x_1 + 3x_2 \geq 6$$
$$-3x_1 - 2x_2 \geq -12$$

and $\quad x_1, x_2 \geq 0$

Because the constraints of this problem are linear, the shape of the feasible region does not differ fundamentally from that of a linear program. Shown as the shaded area in Fig. 21.1a, the feasible region derives its southwestern border from the first constraint, and its northeastern border from the second. Since the objective function is nonlinear, it does not generate a family of parallel, straight isovalue lines. Instead, we get a family of concentric circles, with center at (4, 4) and with each successively smaller circle being associated with a lower value of C. In a

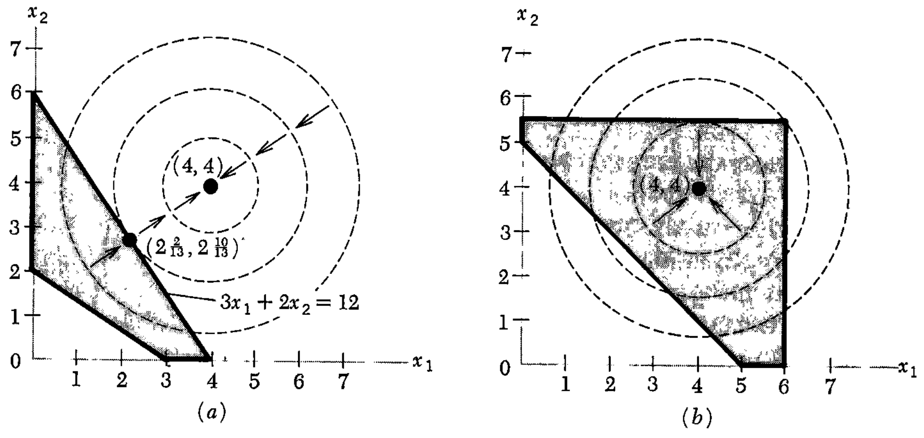

Figure 21.1

free-extremum problem, we would of course choose the point $(x_1, x_2) = (4, 4)$, which yields the minimum value $C = 0$. Being confined to the shaded region, however, the best we can do is to pick the point $(\bar{x}_1, \bar{x}_2) = (2\frac{2}{13}, 2\frac{10}{13})$, where the northeastern border is tangent to one of the circles. The minimized value of C is then $\bar{C} = (2\frac{2}{13} - 4)^2 + (2\frac{10}{13} - 4)^2 = 4\frac{12}{13}$.

The exact value of $\bar{x}_1$ and $\bar{x}_2$ indicated above, while based on the geometric fact of tangency, are found algebraically. First, since the point of tangency in Fig. 21.1*a* lies on the northeastern border, it obviously satisfies the equation

$$3x_1 + 2x_2 = 12 \qquad \text{[from the second constraint]}$$

Next, the circle which is tangent to that border at that point must have the same slope as that border, namely, $-3/2$. Since the slope of the circle is [using the implicit-function rule on the equation $F(x_1, x_2) = (x_1 - 4)^2 + (x_2 - 4)^2 - C = 0$]:

$$\frac{dx_2}{dx_1} = -\frac{\partial F/\partial x_1}{\partial F/\partial x_2} = -\frac{2(x_1 - 4)}{2(x_2 - 4)} = -\frac{x_1 - 4}{x_2 - 4} = -\frac{3}{2}$$

it follows that, by setting this equal to $-3/2$, we can obtain another equation

$$2x_1 - 3x_2 = -4$$

Solved simultaneously, the above pair of linear equations yields the optimal values $(\bar{x}_1, \bar{x}_2) = (2\frac{2}{13}, 2\frac{10}{13})$.

Note that, here, the optimal solution is *not* located at an extreme point of the feasible region, as we would expect in linear programming. Consequently, only *one* constraint is seen to be exactly satisfied, instead of two. Note, also, that whereas to go northeastward towards the point (4, 4) will at first decrease C, to stay on the same course beyond that point will lead to higher values of C instead. Thus we are no longer justified, as under linear programming, in pushing the isovalue curve as far as possible in one single specific direction.

Example 2 Minimize $C = (x_1 - 4)^2 + (x_2 - 4)^2$

subject to

$$x_1 + x_2 \geq 5$$
$$-x_1 \geq -6$$
$$-2x_2 \geq -11$$

and $x_1, x_2 \geq 0$

The present problem differs from the preceding one only in its constraint section. In view of the linearity in the constraints, the feasible region is again a solid polygon, but its new geographic location in relation to the isovalue circles yields a totally new type of outcome. As Fig. 21.1*b* shows, the free-minimum solution point (4, 4) is now contained in the interior of the feasible set, so the constrained optimal solution is also found in that point, with $\overline{C} = 0$. In this example, therefore, the optimal solution does not even lie on the boundary of the feasible region, and, consequently, *none* of the constraints is exactly satisfied at the optimal solution. In contrast to linear programming, it is now no longer possible to narrow down our field of choice to the set of extreme points of the feasible region.

Example 3 Maximize $\pi = 2x_1 + x_2$

subject to

$$-x_1^2 + 4x_1 - x_2 \leq 0$$
$$2x_1 + 3x_2 \leq 12$$

and $x_1, x_2 \geq 0$

In this example, nonlinearity enters through the first constraint. Rewriting the latter in the form of $x_2 \geq -x_1^2 + 4x_1$, where the right-side expression is a quadratic function of x_1, we see that this constraint requires us to pick only the points lying on or above the parabola shown in Fig. 21.2. The second constraint, on the other hand, instructs us to stay on or below a negatively sloped straight line. All told, therefore, the feasible region consists of two disjoint parts, F_1 and F_2. Hence, in this case, the feasible region is not even a convex set!

From the linear objective function, we get a family of linear isovalue curves. As far as F_1 is concerned, point P yields the highest π value, but since F_2 is also feasible, point P qualifies only as a local optimum, not a global one. In fact, any point in F_2 is a better choice than point P. This serves to illustrate that, when the feasible set is not convex, the sufficient conditions of the globality theorem (Sec. 19.3) fail to be satisfied, and a local optimum is therefore not necessarily a global one as well.

To summarize, nonlinear programming differs from linear programming in at least the following five respects, some of which are closely related to each other: (1) The field of choice extends over the entire feasible region, not merely the set of its extreme points; (2) the number of exactly satisfied constraints (and nonnegativity restrictions) may not be equal to the number of choice variables; (3)

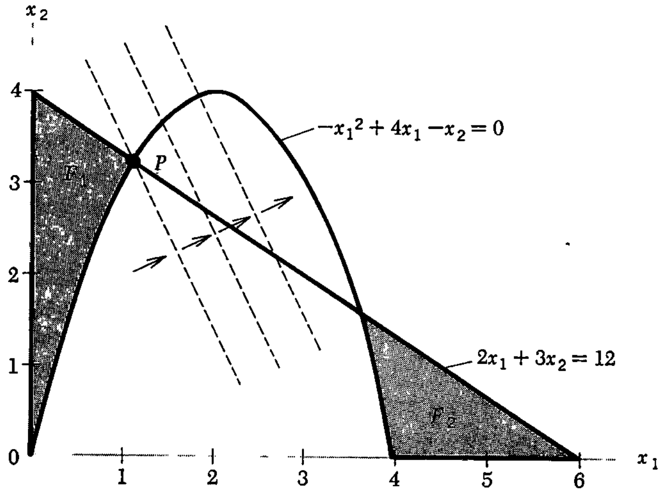

Figure 21.2

adherence to a uniform direction of movement may not lead to continually increasing (or decreasing) values of the objective function; (4) the feasible region may not be a convex set; and (5) a local optimum may not be a global optimum. As a result of these differences, solution methods appropriate for linear programming become largely inapplicable in the nonlinear framework, and new methods become necessary. In this chapter, however, our attention will primarily be focused, not on solution algorithms (which tend to be involved and specialized), but on certain analytical results (necessary conditions and sufficient conditions) that provide the *qualitative* characterizations of an optimal solution, rather than the *quantitative* solution itself.

EXERCISE 21.1

Solve the following nonlinear programs graphically; in each case, give the specific values of $\bar{x}_1$ and $\bar{x}_2$:

1 Minimize $\quad C = x_1^2 + x_2^2$
subject to $\quad x_1 x_2 \geq 25$
and $\quad x_1, x_2 \geq 0$

2 Maximize $\quad \pi = x_1^2 + (x_2 - 2)^2$
subject to $\quad 5x_1 + 3x_2 \leq 15$
and $\quad x_1, x_2 \geq 0$

3 Minimize $\quad C = x_1 + x_2$
subject to $\quad x_1^2 + x_2 \geq 9$
$\quad - x_1 x_2 \geq -8$
and $\quad x_1, x_2 \geq 0$

4 A firm has the linear demand function $x_1 = a - bP_1$ for its first product and $x_2 = c - dP_2$ for the second product. If the average variable costs for the two products are, respectively, $V_1 = m + x_1$ and $V_2 = n + x_2^2$, find its total-gross-profit objective function.

5 The objective function $C = (x_1 - 4)^2 + (x_2 - 4)^2$ in Examples 1 and 2 in the text generates a family of isovalue concentric circles in the x_1x_2 plane. If we introduce a third dimension, C, perpendicular to the x_1x_2 plane, what kind of surface will the objective function yield?

6 Transform the nonlinear program in Exercise 21.1-1 into an equivalent classical constrained-optimization problem by (1) using a dummy variable s, and (2) expressing x_1, x_2, and s as the squares of three other variables, u, v, and w, respectively. Solve this problem in the classical manner, and compare your solution with the graphical solution obtained earlier.

21.2 KUHN-TUCKER CONDITIONS

In the classical optimization problem, with no explicit restrictions on the signs of the choice variables, and with no inequalities in the constraints, the first-order condition for a relative or local extremum is simply that the first partial derivatives of the (smooth) objective function with respect to all the choice variables and the Lagrange multipliers be zero. In nonlinear programming, there exists a similar type of first-order condition, known as the *Kuhn-Tucker conditions.** As we shall see, however, while the classical first-order condition is always necessary, the Kuhn-Tucker conditions cannot be accorded the status of necessary conditions unless a certain proviso is satisfied. On the other hand, under certain specific circumstances, the Kuhn-Tucker conditions turn out to be *sufficient conditions*, or even *necessary-and-sufficient* conditions as well.

Since the Kuhn-Tucker conditions are the single most important analytical result in nonlinear programming, it is essential to have a proper understanding of those conditions as well as their implications. For the sake of expository convenience, we shall develop these conditions in two steps.

Effect of Nonnegativity Restrictions

As the first step, consider a problem with nonnegativity restrictions, but with no other constraints. Taking the single-variable case, in particular, we have:

$$\text{(21.3)} \qquad \begin{array}{ll} \text{Maximize} & \pi = f(x_1) \\ \text{subject to} & x_1 \geq 0 \end{array}$$

where the function f is assumed to be differentiable. In view of the restriction

* H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, California, 1951, pp. 481–492.

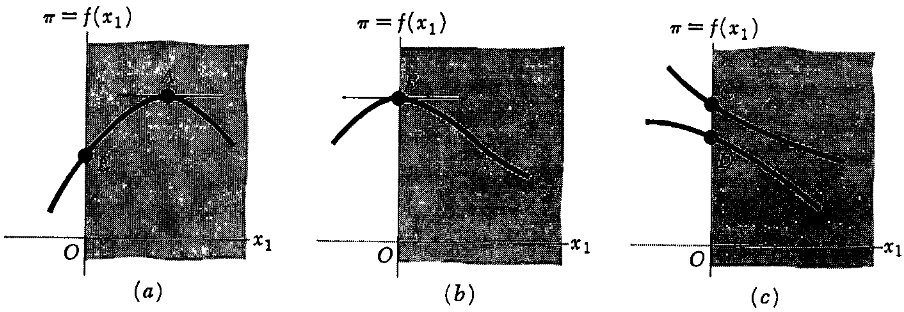

Figure 21.3

$x_1 \geq 0$, three possible situations may arise. First, if a local maximum of π occurs in the interior of the shaded feasible region in Fig. 21.3, such as at point A in diagram a, then we have an *interior solution*. The first-order condition in this case is $d\pi/dx_1 = f'(x_1) = 0$, same as in the classical problem. Second, as illustrated by point B in diagram b, a local maximum can also occur on the vertical axis, where $x_1 = 0$. Even in this second case, where we have a *boundary solution*, the first-order condition $f'(x_1) = 0$ nevertheless remains valid. However, as a third possibility, a local maximum may in the present context take the position of point C or point D in diagram c, because to qualify as a local maximum in the problem (21.3), the candidate point merely has to be higher than the neighboring points *within* the feasible region. In view of this last possibility, the maximum point in a problem like (21.3) can be characterized, not only by the equation $f'(x_1) = 0$, but also by the inequality $f'(x_1) < 0$. Note, on the other hand, that the opposite inequality, $f'(x_1) > 0$, can safely be ruled out, for at a point where the curve is upward-sloping, we can never have a maximum, even if that point is located on the vertical axis, such as point E in diagram a.

The upshot of the above discussion is that, in order for a value of x_1 to give a local maximum of π in the problem (21.3), it must satisfy one of the following three conditions:

$$\begin{cases} f'(x_1) = 0 \quad \text{and} \quad x_1 > 0 \quad \text{[point } A\text{]} & (21.4) \\ f'(x_1) = 0 \quad \text{and} \quad x_1 = 0 \quad \text{[point } B\text{]} & (21.5) \\ f'(x_1) < 0 \quad \text{and} \quad x_1 = 0 \quad \text{[points } C \text{ and } D\text{]} & (21.6) \end{cases}$$

Actually, these three conditions can be consolidated into a single statement:

$$f'(x_1) \leq 0 \qquad x_1 \geq 0 \qquad \text{and} \qquad x_1 f'(x_1) = 0 \tag{21.7}$$

The first inequality in (21.7) is a summary of the information regarding $f'(x_1)$ enumerated in (21.4) through (21.6). The second inequality is a similar summary for x_1; in fact, it merely reiterates the nonnegativity restriction of the problem. And, as the third part of (21.7), we have an equation which expresses an important feature common to (21.4), (21.5), as well as (21.6), namely that, of the two quantities x_1 and $f'(x_1)$, *at least one* must take a zero value, so that the product of the two must be zero. Taken together, the three parts of (21.7) constitute the first-order necessary condition for a local maximum in a problem where the choice variable must be nonnegative. But going a step further, we can also take them to be necessary for a *global* maximum. This is because a global maximum must also be a local maximum and, as such, must also satisfy the necessary condition for a local maximum.

When the problem contains n choice variables:

$$\text{(21.8)}\qquad \begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, \ldots, x_n) \\ \text{subject to} & x_j \geq 0 \qquad (j = 1, 2, \ldots, n) \end{array}$$

the classical first-order condition $f_1 = f_2 = \cdots = f_n = 0$ must be similarly modified. To do this, we can apply the same type of reasoning underlying (21.7) to each choice variable, x_j, taken by itself. Graphically, this amounts to viewing the horizontal axis in Fig. 21.3 as representing each x_j in turn. The required modification of the first-order condition then readily suggests itself:

$$\text{(21.9)}\qquad f_j \leq 0 \qquad x_j \geq 0 \qquad \text{and} \qquad x_j f_j = 0 \qquad (j = 1, 2, \ldots, n)$$

where f_j is the partial derivative $\partial\pi/\partial x_j$.

Effect of Inequality Constraints

With this background, we now proceed to the second step, and try to include inequality constraints as well. For simplicity, let us first deal with a problem with three variables ($n = 3$) and two constraints ($m = 2$):

$$\text{(21.10)}\qquad \begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, x_3) \\ \text{subject to} & g^1(x_1, x_2, x_3) \leq r_1 \\ & g^2(x_1, x_2, x_3) \leq r_2 \\ \text{and} & x_1, x_2, x_3 \geq 0 \end{array}$$

which, with the help of two dummy variables s_1 and s_2, can be transformed into the equivalent form

$$\text{(21.10')}\qquad \begin{array}{ll} \text{Maximize} & \pi = f(x_1, x_2, x_3) \\ \text{subject to} & g^1(x_1, x_2, x_3) + s_1 = r_1 \\ & g^2(x_1, x_2, x_3) + s_2 = r_2 \\ \text{and} & x_1, x_2, x_3, s_1, s_2 \geq 0 \end{array}$$

If the nonnegativity restrictions are absent, we may, in line with the classical approach, form the Lagrangian function (denoting the Lagrange multiplier here by y rather than λ):

(21.11)
$$\boxed{Z^* = f(x_1, x_2, x_3) + y_1[r_1 - g^1(x_1, x_2, x_3) - s_1] + y_2[r_2 - g^2(x_1, x_2, x_3) - s_2]}$$

and write the first-order condition as

$$\frac{\partial Z^*}{\partial x_1} = \frac{\partial Z^*}{\partial x_2} = \frac{\partial Z^*}{\partial x_3} = \frac{\partial Z^*}{\partial s_1} = \frac{\partial Z^*}{\partial s_2} = \frac{\partial Z^*}{\partial y_1} = \frac{\partial Z^*}{\partial y_2} = 0$$

But since the x_j and s_i variables do have to be nonnegative, the first-order condition on those variables should be modified in accordance with (21.9). Consequently, we obtain the following set of conditions instead:

(21.12)
$$\begin{cases} \dfrac{\partial Z^*}{\partial x_j} \le 0 & x_j \ge 0 \quad \text{and} \quad x_j \dfrac{\partial Z^*}{\partial x_j} = 0 \\ \dfrac{\partial Z^*}{\partial s_i} \le 0 & s_i \ge 0 \quad \text{and} \quad s_i \dfrac{\partial Z^*}{\partial s_i} = 0 \\ \dfrac{\partial Z^*}{\partial y_i} = 0 & \begin{pmatrix} i = 1, 2 \\ j = 1, 2, 3 \end{pmatrix} \end{cases}$$

slackness already included

Note that the derivatives $\partial Z^*/\partial y_i$ are still to be set strictly equal to zero. (Why?)

Each line of (21.12) relates to a different type of variable. But we can consolidate the last two lines and, in the process, eliminate the dummy variables s_i from the first-order condition. Inasmuch as $\partial Z^*/\partial s_i = -y_i$, the second line tells us that we must have $-y_i \le 0$, $s_i \ge 0$ and, $-s_i y_i = 0$, or, equivalently,

(21.13)
$$s_i \ge 0 \qquad y_i \ge 0 \qquad \text{and} \qquad y_i s_i = 0$$

But the third line—a restatement of the constraints in (21.10′)—means that $s_i = r_i - g^i(x_1, x_2, x_3)$. By substituting the latter into (21.13), therefore, we can combine the second and third lines of (21.12) into:

$$r_i - g^i(x_1, x_2, x_3) \ge 0 \qquad y_i \ge 0 \qquad \text{and} \qquad y_i[r_i - g^i(x_1, x_2, x_3)] = 0$$

This enables us to express the first-order condition (21.12) in an equivalent form *without* the dummy variables. Using the symbol g_j^i to denote $\partial g^i/\partial x_j$, we now write

(21.14)
$$\begin{cases} \dfrac{\partial Z^*}{\partial x_j} = f_j - (y_1 g_j^1 + y_2 g_j^2) \le 0 & x_j \ge 0 \quad \text{and} \quad x_j \dfrac{\partial Z^*}{\partial x_j} = 0 \\ r_i - g^i(x_1, x_2, x_3) \ge 0 & y_i \ge 0 \quad \text{and} \quad y_i[r_i - g^i(x_1, x_2, x_3)] = 0 \end{cases}$$

These, then, are the Kuhn-Tucker conditions for the problem (21.10), or, more accurately, one version of the Kuhn-Tucker conditions, expressed in terms of the Lagrangian function Z^* in (21.11).

Now that we know the results, though, it is possible to obtain the same set of conditions more directly by using a different Lagrangian function. Given the problem (21.10), let us ignore the nonnegativity restrictions as well as the inequality signs in the constraints and write the purely classical type of Lagrangian function, Z:

$$Z = f(x_1, x_2, x_3) + y_1[r_1 - g^1(x_1, x_2, x_3)] + y_2[r_2 - g^2(x_1, x_2, x_3)] \tag{21.15}$$

Then let us (1) set the partial derivatives $\partial Z/\partial x_j \le 0$, but $\partial Z/\partial y_i \ge 0$, (2) impose nonnegativity restrictions on x_j and y_i, and (3) require complementary slackness to prevail between each variable and the partial derivative of Z with respect to that variable, that is, require their product to vanish. Since the results of these steps, namely,

$$\begin{aligned} &\frac{\partial Z}{\partial x_j} = f_j - (y_1 g_j^1 + y_2 g_j^2) \le 0 \qquad x_j \ge 0 \qquad \text{and} \qquad x_j \frac{\partial Z}{\partial x_j} = 0 \\ &\frac{\partial Z}{\partial y_i} = r_i - g^i(x_1, x_2, x_3) \ge 0 \qquad y_i \ge 0 \qquad \text{and} \qquad y_i \frac{\partial Z}{\partial y_i} = 0 \end{aligned} \tag{21.16}$$

are identical with (21.14), the Kuhn-Tucker conditions are expressible also in terms of the Lagrangian function Z (as against Z^*). Note that, by switching from Z^* to Z, we can not only arrive at the Kuhn-Tucker conditions more directly, but also identify the expression $r_i - g^i(x_1, x_2, x_3)$—which was left nameless in (21.14)—as the partial derivative $\partial Z/\partial y_i$. In the subsequent discussion, therefore, we shall only use the (21.16) version of the Kuhn-Tucker conditions, based on the Lagrangian function Z.

Interpretation of the Kuhn-Tucker Conditions

Parts of the Kuhn-Tucker conditions (21.16) are merely a restatement of certain aspects of the given problem. Thus the conditions $x_j \ge 0$ merely repeat the nonnegativity restrictions, and the conditions $\partial Z/\partial y_i \ge 0$ merely reiterate the constraints. To include these in (21.16), however, has the important advantage of revealing more clearly the remarkable symmetry between the two types of variables, x_j (choice variables) and y_i (Lagrange multipliers). To each variable in each category, there corresponds a marginal condition—$\partial Z/\partial x_j \le 0$ or $\partial Z/\partial y_i \ge 0$—to be satisfied by the optimal solution. Each of the variables must be nonnegative as well. And, finally, each variable is characterized by complementary slackness in relation to a particular partial derivative of the Lagrangian function Z. This means that, for each x_j, we must find in the optimal solution that *either* the marginal condition holds as an equality, as in the classical context, *or* the choice variable in question must take a zero value, *or* both. Analogously, for each y_i, we must find in the optimal solution that *either* the marginal condition holds as an equality—meaning that the ith constraint is exactly satisfied—*or* the Lagrange multiplier vanishes, *or* both.

An even more explicit interpretation is possible, when we look at the expanded expressions for $\partial Z/\partial x_j$ and $\partial Z/\partial y_i$ in (21.16). Assume the problem to be the familiar production problem. Then we have

$f_j \equiv$ the marginal gross profit of the jth product

$y_i \equiv$ the shadow price of the ith resource

$g_j^i \equiv$ the amount of the ith resource used up in producing the marginal unit of the jth product

$y_i g_j^i \equiv$ the marginal imputed cost of the ith resource incurred in producing a unit of the jth product

$\sum_i y_i g_j^i \equiv$ the aggregate marginal imputed cost of the jth product

Thus the marginal condition

$$\frac{\partial Z}{\partial x_j} = f_j - \sum_i y_i g_j^i \leq 0$$

requires that the marginal gross profit of the jth product be no greater than its aggregate marginal imputed cost; i.e., no *under*imputation is permitted. The complementary-slackness condition then means that, if the optimal solution calls for the active production of the jth product ($\bar{x}_j > 0$), the marginal gross profit must be exactly equal to the aggregate marginal imputed cost ($\partial Z/\partial \bar{x}_j = 0$), as would be the situation in the classical optimization problem. If, on the other hand, the marginal gross profit optimally falls short of the aggregate imputed cost ($\partial Z/\partial \bar{x}_j < 0$), entailing *excess* imputation, then that product must not be produced ($\bar{x}_j = 0$).* This latter situation is something that can never occur in the classical context, for if the marginal gross profit is less than the marginal imputed cost, then the output should in that framework be reduced all the way to the level where the marginal condition is satisfied as an equality. What causes the situation of $\partial Z/\partial \bar{x}_j < 0$ to qualify as an optimal one here, is the explicit specification of nonnegativity in the present framework. For then the most we can do in the way of output reduction is to lower production to the level $\bar{x}_j = 0$, and if we still find $\partial Z/\partial \bar{x}_j < 0$ at the zero output, we stop there anyway.

As for the remaining conditions, which relate to the variables y_i, their meanings are even easier to perceive. First of all, the marginal condition $\partial Z/\partial y_i \geq 0$ merely requires the firm to stay within the capacity limitation of every resource in the plant. The complementary-slackness condition then stipulates that, if the ith resource is not fully used in the optimal solution ($\partial Z/\partial \bar{y}_i > 0$), the shadow price of that resource—which is never allowed to be negative—must be set equal to zero ($\bar{y}_i = 0$). On the other hand, if a resource has a positive shadow

* Remember that, given the equation $ab = 0$, where a and b are real numbers, we can legitimately infer that $a \neq 0$ implies $b = 0$, but it is not true that $a = 0$ implies $b \neq 0$, since $b = 0$ is also consistent with $a = 0$.

price in the optimal solution ($\bar{y}_i > 0$), then it is perforce a fully utilized resource ($\partial Z/\partial \bar{y}_i = 0$). These, of course, are nothing but the implications of Duality Theorem II of the preceding chapter.

It is also possible, of course, to take the Lagrange-multiplier value $\bar{y}_i$ to be a measure of how the optimal value of the objective function reacts to a slight relaxation of the ith constraint (see Sec. 12.2). In that light, complementary slackness would mean that, if the ith constraint is optimally not binding ($\partial Z/\partial \bar{y}_i > 0$), then relaxing that particular constraint will not affect the optimal value of the gross profit ($\bar{y}_i = 0$)—just as loosening a belt which is not constricting one's waist to begin with will not produce any greater comfort. If, on the other hand, a slight relaxation of the ith constraint (increasing the endowment of the ith resource) does increase the gross profit ($\bar{y}_i > 0$), then that resource constraint must in fact be binding in the optimal solution ($\partial Z/\partial \bar{y}_i = 0$).

The n-Variable m-Constraint Case

The above discussion can be generalized in a straightforward manner to the maximization problem given in (21.1) or (21.1′). Since there are now n choice variables and m constraints, the Lagrangian function Z will appear in the more general form

$$(21.17) \qquad Z = f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} y_i[r_i - g^i(x_1, x_2, \ldots, x_n)]$$

And the Kuhn-Tucker conditions will simply be

$$(21.18)\begin{cases} \dfrac{\partial Z}{\partial x_j} \le 0 \qquad x_j \ge 0 \qquad \text{and} \qquad x_j \dfrac{\partial Z}{\partial x_j} = 0 \\[2ex] \dfrac{\partial Z}{\partial y_i} \ge 0 \qquad y_i \ge 0 \qquad \text{and} \qquad y_i \dfrac{\partial Z}{\partial y_i} = 0 \end{cases} \qquad \begin{matrix} [\text{maximization}] \\ \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix} \end{matrix}$$

Here, in order to avoid a cluttered appearance, we have not written out the expanded expressions for the partial derivatives $\partial Z/\partial x_j$ and $\partial Z/\partial y_i$. But you are urged to write them out for a more detailed view of the Kuhn-Tucker conditions, similar to what was given in (21.16). Note that, aside from the change in the dimension of the problem, the Kuhn-Tucker conditions remain entirely the same as in the three-variable, two-constraint case discussed before. The interpretation of these conditions should naturally also remain the same.

What if the problem is one of *minimization*, as in (21.2) or (21.2′)? One way of handling it is to convert the problem into a maximization problem and then apply (21.18). To minimize C is equivalent to *maximizing* $-C$, so such a conversion is always feasible. But we must, of course, also reverse the constraint inequalities by multiplying every constraint through by -1. Instead of going through the conversion process, however, we may—again using the Lagrangian function Z as defined in (21.17)—directly apply the minimization version of the Kuhn-Tucker conditions as follows:

$$(21.19)\quad \begin{cases} \dfrac{\partial Z}{\partial x_j} \geq 0 \qquad x_j \geq 0 \qquad \text{and} \qquad x_j \dfrac{\partial Z}{\partial x_j} = 0 \\[2ex] \dfrac{\partial Z}{\partial y_i} \leq 0 \qquad y_i \geq 0 \qquad \text{and} \qquad y_i \dfrac{\partial Z}{\partial y_i} = 0 \end{cases} \qquad \begin{matrix} \text{[minimization]} \\ \hline \begin{pmatrix} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{pmatrix} \end{matrix}$$

This you should compare with (21.18).

Reading (21.18) and (21.19) horizontally (*row*wise), we see that the Kuhn-Tucker conditions for both maximization and minimization problems consist of a set of conditions relating to the choice variables x_j (first row), and another set relating to the Lagrange multipliers y_i (second row). Reading them vertically (*column*wise), on the other hand, we note that, for each x_j and y_i, there is a marginal condition (first column), a nonnegativity restriction (second column), and a complementary-slackness condition (third column). In any given problem, the marginal conditions pertaining to the choice variables always differ, as a group, from the marginal conditions for the Lagrange multipliers in the sense of inequality they take. Also, the marginal conditions for a maximization problem always differ, as a group, from those of a minimization problem in the sense of inequality they take.

Subject to a proviso to be explained in the next section, the Kuhn-Tucker maximum conditions (21.18) and minimum conditions (21.19) are necessary conditions for a local maximum and a local minimum, respectively. But since a global maximum (minimum) must also be a local maximum (minimum), the Kuhn-Tucker conditions can also be taken as necessary conditions for a global maximum (minimum), subject to the same proviso.

An Example

Let us check whether the Kuhn-Tucker conditions are satisfied by the solution in Example 1 of Sec. 21.1, as illustrated in Fig. 21.1*a*. The Lagrangian function for this problem is

$$Z = (x_1 - 4)^2 + (x_2 - 4)^2 + y_1(6 - 2x_1 - 3x_2) + y_2(-12 + 3x_1 + 2x_2)$$

Since the problem is one of minimization, the appropriate conditions are (21.19), which include the four marginal conditions

$$\frac{\partial Z}{\partial x_1} = 2(x_1 - 4) - 2y_1 + 3y_2 \geq 0$$

$$\frac{\partial Z}{\partial x_2} = 2(x_2 - 4) - 3y_1 + 2y_2 \geq 0$$

$$\frac{\partial Z}{\partial y_1} = 6 - 2x_1 - 3x_2 \leq 0$$

$$\frac{\partial Z}{\partial y_2} = -12 + 3x_1 + 2x_2 \leq 0$$

plus the nonnegativity and complementary-slackness conditions. The question is: Can we find nonnegative values $\bar{y}_1$ and $\bar{y}_2$ which, together with the optimal values $\bar{x}_1 = 2\frac{2}{13} = \frac{28}{13}$ and $\bar{x}_2 = 2\frac{10}{13} = \frac{36}{13}$, will satisfy all those conditions?

Given that $\bar{x}_1$ and $\bar{x}_2$ are both nonzero, complementary slackness dictates that $\partial Z/\partial x_1 = 0$ and $\partial Z/\partial x_2 = 0$. Thus, after substituting the $\bar{x}_1$ and $\bar{x}_2$ values into the first two marginal conditions, the latter become two equations

$$-2y_1 + 3y_2 = \tfrac{48}{13}$$

$$-3y_1 + 2y_2 = \tfrac{32}{13}$$

with solution $\bar{y}_1 = 0$, and $\bar{y}_2 = \frac{16}{13} = 1\frac{3}{13}$, which are nonnegative, as required. Since these values, together with $\bar{x}_1$ and $\bar{x}_2$, imply that $\partial Z/\partial \bar{x}_1 = 0$, $\partial Z/\partial \bar{x}_2 = 0$, $\partial Z/\partial \bar{y}_1 < 0$, and $\partial Z/\partial \bar{y}_2 = 0$, which satisfy the marginal inequalities as well as the complementary-slackness conditions, all the Kuhn-Tucker minimum conditions are satisfied.

EXERCISE 21.2

1 Draw a set of diagrams similar to Fig. 21.3 for the minimization case, and deduce a set of necessary conditions for a local minimum corresponding to (21.4) through (21.6). Then condense these conditions into a single statement similar to (21.7).

2 (*a*) Show that, in (21.18), instead of writing

$$y_i \frac{\partial Z}{\partial y_i} = 0 \qquad (i = 1, \ldots, m)$$

as a set of m separate conditions, it is sufficient to write a single equation in the form of

$$\sum_{i=1}^{m} y_i \frac{\partial Z}{\partial y_i} = 0$$

(*b*) Can we do the same for the set of conditions

$$x_j \frac{\partial Z}{\partial x_j} = 0 \qquad (j = 1, \ldots, n)$$

3 Based on the reasoning used in the preceding problem, which set (or sets) of conditions in (21.19) can be condensed into a single equation?

4 Given the minimization problem (21.2), and using the Lagrangian function (21.17), take the derivatives $\partial Z/\partial x_j$ and $\partial Z/\partial y_i$ and write out the expanded version of the Kuhn-Tucker minimum conditions (21.19).

5 Convert the minimization problem (21.2) into a maximization problem, formulate the Lagrangian function, take the derivatives with respect to x_j and y_i, and apply the Kuhn-Tucker maximum conditions (21.18). Are the results consistent with those obtained in the preceding problem?

6 Check the applicability of the Kuhn-Tucker conditions to Example 2 of Sec. 21.1 as follows:

(*a*) Write the Lagrangian function and the Kuhn-Tucker conditions.

(*b*) From the solution given in Fig. 21.1*b*, find the optimal values of $\partial Z/\partial y_i$ $(i = 1, 2, 3)$. What can we then conclude about $\bar{y}_i$?

(*c*) Now find the optimal values of $\partial Z/\partial x_1$ and $\partial Z/\partial x_2$.

(*d*) Are all the Kuhn-Tucker conditions satisfied?

21.3 THE CONSTRAINT QUALIFICATION

In Sec. 21.2, it was pointed out that the Kuhn-Tucker conditions are necessary conditions *only if* a particular proviso is satisfied. That proviso, called the *constraint qualification*, imposes a certain restriction on the constraint functions of a nonlinear program, for the specific purpose of ruling out certain irregularities on the boundary of the feasible set, that would invalidate the Kuhn-Tucker conditions should the optimal solution occur there.

Irregularities at Boundary Points

Let us first illustrate the nature of such irregularities by means of some concrete examples.

Example 1 Maximize $\pi = x_1$

subject to $x_2 - (1 - x_1)^3 \le 0$

and $x_1, x_2 \ge 0$

As shown in Fig. 21.4, the feasible region is the set of points that lie in the first quadrant on or below the curve $x_2 = (1 - x_1)^3$. Since the objective function directs us to maximize x_1, the optimal solution is the point $(1, 0)$. But this solution fails to satisfy the Kuhn-Tucker maximum conditions. To check this, we first write the Lagrangian function

$$Z = x_1 + y_1\left[-x_2 + (1 - x_1)^3\right]$$

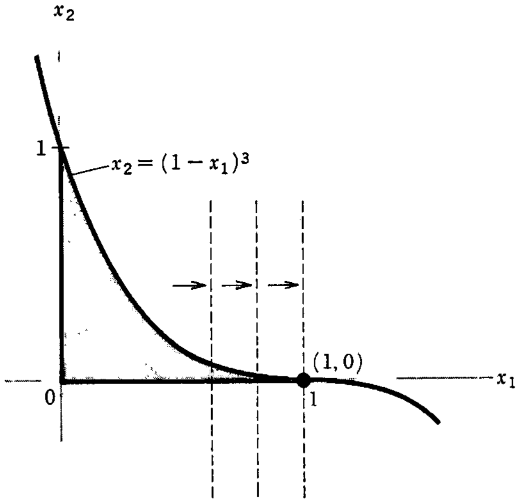

Figure 21.4

As the first marginal condition, we should then have

$$\frac{\partial Z}{\partial x_1} = 1 - 3y_1(1 - x_1)^2 \leq 0$$

In fact, since $\bar{x}_1 = 1$ is positive, complementary slackness requires that this derivative vanish when evaluated at the point (1, 0). However, the actual value we get happens to be $\partial Z/\partial \bar{x}_1 = 1$, thus violating the above marginal condition.

The reason for this anomaly is that the optimal solution, (1, 0), occurs in this example at an outward-pointing *cusp*, which constitutes one type of irregularity that can invalidate the Kuhn-Tucker conditions at a boundary optimal solution. A cusp is a sharp point formed when a curve takes a sudden reversal in direction, such that the slope of the curve on one side of the point is the same as the slope of the curve on the other side of the point. Here, the boundary of the feasible region at first follows the constraint curve, but when the point (1, 0) is reached, it takes an abrupt turn westward and follows the trail of the horizontal axis thereafter. Since the slopes of both the curved side and the horizontal side of the boundary are zero at the point (1, 0), that point is a cusp.

Cusps are the most frequently cited culprit for the failure of the Kuhn-Tucker conditions, but the truth is that the presence of a cusp is neither necessary nor sufficient to cause those conditions to fail at an optimal solution. The following two examples will confirm this.

Example 2 To the problem of the preceding example, let us add a new constraint

$$2x_1 + x_2 \leq 2$$

whose border, $x_2 = 2 - 2x_1$, plots as a straight line with slope -2 which passes through the optimal point in Fig. 21.4. Clearly, the feasible region remains the same as before, and so does the optimal solution at the cusp. But if we write the new Lagrangian function

$$Z = x_1 + y_1\left[-x_2 + (1 - x_1)^3\right] + y_2[2 - 2x_1 - x_2]$$

and the marginal conditions

$$\frac{\partial Z}{\partial x_1} = 1 - 3y_1(1 - x_1)^2 - 2y_2 \le 0$$

$$\frac{\partial Z}{\partial x_2} = -y_1 - y_2 \le 0$$

$$\frac{\partial Z}{\partial y_1} = -x_2 + (1 - x_1)^3 \ge 0$$

$$\frac{\partial Z}{\partial y_2} = 2 - 2x_1 - x_2 \ge 0$$

it turns out that the values $\bar{x}_1 = 1$, $\bar{x}_2 = 0$, $\bar{y}_1 = 1$, and $\bar{y}_2 = \frac{1}{2}$ do satisfy the above four inequalities, as well as the nonnegativity and complementary-slackness conditions. As a matter of fact, $\bar{y}_1$ can be assigned any nonnegative value (not just 1), and all the conditions can still be satisfied—which goes to show that the optimal value of a Lagrange multiplier is not necessarily unique. More importantly, however, this example shows that the Kuhn-Tucker conditions can remain valid despite the cusp.

Example 3 The feasible region of the problem

$$\begin{array}{ll} \text{Maximize} & \pi = x_2 - x_1^2 \\ \text{subject to} & -(10 - x_1^2 - x_2)^3 \le 0 \\ & -x_1 \le -2 \\ \text{and} & x_1, x_2 \ge 0 \end{array}$$

as shown in Fig. 21.5, contains no cusp anywhere. Yet, at the optimal solution, (2, 6), the Kuhn-Tucker conditions nonetheless fail to hold. For, with the Lagrangian function

$$Z = x_2 - x_1^2 + y_1(10 - x_1^2 - x_2)^3 + y_2(-2 + x_1)$$

the second marginal condition would require that

$$\frac{\partial Z}{\partial x_2} = 1 - 3y_1(10 - x_1^2 - x_2)^2 \le 0$$

Indeed, since $\bar{x}_2$ is positive, this derivative should vanish when evaluated at the point (2, 6). But actually we get $\partial Z/\partial \bar{x}_2 = 1$, regardless of the value assigned to y_1. Thus the Kuhn-Tucker conditions can fail even in the absence of a cusp—nay, even when the feasible region is a convex set as in Fig. 21.5. The fundamental reason why cusps are neither necessary nor sufficient for the failure of the Kuhn-Tucker conditions is that the irregularities referred to above relate, not to the shape of the feasible region per se, but to the forms of the constraint functions themselves.

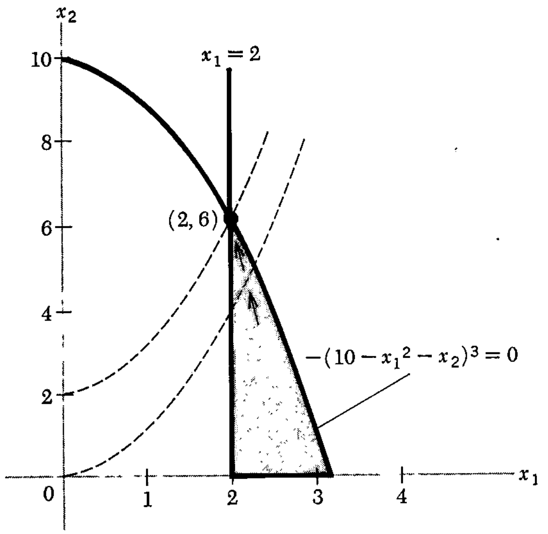

Figure 21.5

The Constraint Qualification

Boundary irregularities—cusp or no cusp—will not occur if a certain constraint qualification is satisfied.

To explain this, let $\bar{x} \equiv (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ be a boundary point of the feasible region and a possible candidate for a solution, and let $dx \equiv (dx_1, dx_2, \ldots, dx_n)$ represent a particular direction of movement from the said boundary point. The direction-of-movement interpretation of the vector dx is perfectly in line with our earlier interpretation of a vector as a directed line segment (an arrow), but here, the point of departure is the point $\bar{x}$ instead of the point of origin, and so the vector dx is *not* in the nature of a radius vector. We shall now impose two requirements on the vector dx. First, if the jth choice variable has a zero value at the point $\bar{x}$, then we shall only permit a nonnegative change on the x_j axis, that is,

$$dx_j \geq 0 \qquad \text{if} \qquad \bar{x}_j = 0 \tag{21.20}$$

Second, if the ith constraint is exactly satisfied at the point $\bar{x}$, then we shall only allow values of $dx_1, \ldots, dx_n$ such that the value of the constraint function $g^i(\bar{x})$ will not increase (for a maximization problem) or will not decrease (for a minimization problem), that is,

$$dg^i(\bar{x}) = g_1^i\, dx_1 + g_2^i\, dx_2 + \cdots + g_n^i\, dx_n \begin{cases} \leq 0 \text{ (maximization)} \\ \geq 0 \text{ (minimization)} \end{cases} \qquad \text{if} \qquad g^i(\bar{x}) = r_i \tag{21.21}$$

where all the partial derivatives g_j^i are to be evaluated at $\bar{x}$. If a vector dx satisfies (21.20) and (21.21), we shall refer to it as a *test vector*. Finally, if there exists a differentiable arc that (1) emanates from the point $\bar{x}$, (2) is contained entirely in the feasible region, and (3) is tangent to a given test vector, we shall call it a *qualifying arc* for that test vector. With this background, the constraint qualification can be stated simply as follows:

> The constraint qualification is satisfied if, for any point $\bar{x}$ on the boundary of the feasible region, there exists a qualifying arc for every test vector dx.

Example 4 We shall show that the optimal point $(1, 0)$ of Example 1 in Fig. 21.4, which fails the Kuhn-Tucker conditions, also fails the constraint qualification. At that point, $\bar{x}_2 = 0$, thus the test vectors must satisfy

$$dx_2 \geq 0 \qquad \text{[by (21.20)]}$$

Moreover, since the (only) constraint, $g^1 = x_2 - (1 - x_1)^3 \leq 0$, is exactly satisfied at $(1, 0)$, we must let

$$g_1^1\, dx_1 + g_2^1\, dx_2 = 3(1 - \bar{x}_1)^2\, dx_1 + dx_2 = dx_2 \leq 0 \qquad \text{[by (21.21)]}$$

These two requirements together imply that we must let $dx_2 = 0$. In contrast, we are free to choose dx_1. Thus, for instance, the vector $(dx_1, dx_2) = (2, 0)$ is an acceptable test vector, as is $(dx_1, dx_2) = (-1, 0)$. The latter test vector would plot in Fig. 21.4 as an arrow starting from $(1, 0)$ and pointing in the due-west direction (not drawn), and it is clearly possible to draw a qualifying arc for it. (The curved boundary of the feasible region itself can serve as a qualifying arc.) On the other hand, the test vector $(dx_1, dx_2) = (2, 0)$ would plot as an arrow starting from $(1, 0)$ and pointing in the due-east direction (not drawn). Since there is no way to draw a smooth arc tangent to this vector and lying entirely within the feasible region, no qualifying arcs exist for it. Hence the optimal solution point $(1, 0)$ violates the constraint qualification.

Example 5 Referring to Example 2 above, let us now illustrate that, after an additional constraint $2x_1 + x_2 \leq 2$ is added to Fig. 21.4, the point $(1, 0)$ will satisfy the constraint qualification, thereby revalidating the Kuhn-Tucker conditions.

As in Example 4, we have to require $dx_2 \geq 0$ (because $\bar{x}_2 = 0$) and $dx_2 \leq 0$ (because the first constraint is exactly satisfied); thus, $dx_2 = 0$. But the second constraint is also exactly satisfied, thereby requiring

$$g_1^2\, dx_1 + g_2^2\, dx_2 = 2\, dx_1 + dx_2 = 2\, dx_1 \leq 0 \qquad \text{[by (21.21)]}$$

With nonpositive dx_1 and zero dx_2, the only admissible test vectors—aside from the null vector itself—are those pointing in the due-west direction in Fig. 21.4 from $(1, 0)$. All of these lie along the horizontal axis in the feasible region, and it is certainly possible to draw a qualifying arc for each test vector. Hence, this time the constraint qualification indeed is satisfied.

Linear Constraints

Earlier, in Example 3, it was demonstrated that the convexity of the feasible set does not guarantee the validity of the Kuhn-Tucker conditions as necessary conditions. However, if the feasible region is a convex set formed by *linear* constraints only, then the constraint qualification will invariably be met, and the Kuhn-Tucker conditions will always hold at an optimal solution. This being the case, we need never worry about boundary irregularities when dealing with a nonlinear program with linear constraints, or, as a special case, a linear program per se.

Example 6 Let us illustrate the linear-constraint result in the two-variable two-constraint framework. For a maximization problem, the linear constraints can be written as

$$a_{11}x_1 + a_{12}x_2 \le r_1$$
$$a_{21}x_1 + a_{22}x_2 \le r_2$$

where we shall take all the parameters to be positive. Then, as indicated in Fig. 21.6, the first constraint border will have a slope of $-a_{11}/a_{12} < 0$, and the second, a slope of $-a_{21}/a_{22} < 0$. The boundary points of the shaded feasible region fall into the following five types: (1) the point of origin, where the two axes intersect, (2) points that lie on one axis segment, such as J and S, (3) points at the intersection of one axis and one constraint border, namely, K and R, (4) points lying on a single constraint border, such as L and N, and (5) the point of intersection of the two constraints, M. We may briefly examine each type in turn with reference to the satisfaction of the constraint qualification.

1. At the origin, no constraint is exactly satisfied, so we may ignore (21.21). But since $x_1 = x_2 = 0$, we must choose test vectors with $dx_1 \ge 0$ and $dx_2 \ge 0$, by (21.20). Hence all test vectors from the origin must point in the due-east, due-north, or northeast directions, as depicted in Fig. 21.6. These vectors all

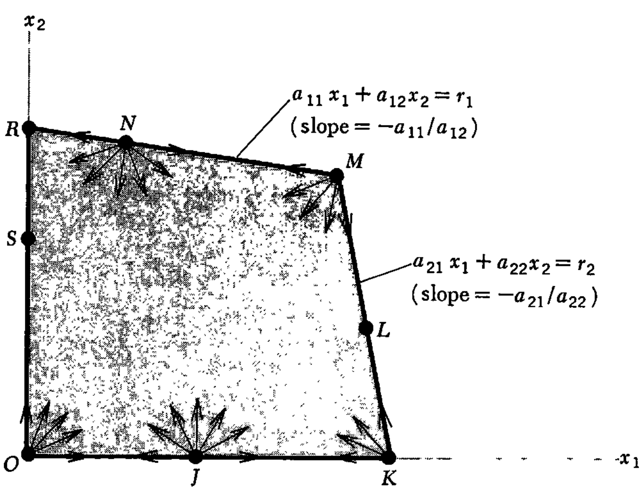

Figure 21.6

happen to fall within the feasible set, and a qualifying arc clearly can be found for each.

2. At a point like J, we can again ignore (21.21). The fact that $x_2 = 0$ means that we must choose $dx_2 \geq 0$, but our choice of dx_1 is free. Hence all vectors would be acceptable except those pointing southward ($dx_2 < 0$). Again all such vectors fall within the feasible region, and there exists a qualifying arc for each. The analysis of point S is similar.
3. At points K and R, both (21.20) and (21.21) must be considered. Specifically, at K, we have to choose $dx_2 \geq 0$ since $x_2 = 0$, so that we must rule out all southward arrows. The second constraint being exactly satisfied, moreover, the test vectors for point K must satisfy

   (21.22) $$g_1^2\, dx_1 + g_2^2\, dx_2 = a_{21}\, dx_1 + a_{22}\, dx_2 \leq 0$$

   Since at K we also have $a_{21}x_1 + a_{22}x_2 = r_2$ (second constraint border), however, we may add this equality to (20.22) and modify the restriction on the test vectors to the form

   (21.22′) $$a_{21}(x_1 + dx_1) + a_{22}(x_2 + dx_2) \leq r_2$$

   Interpreting $(x_j + dx_j)$ to be the new value of x_j attained at the arrowhead of a test vector, we may construe (21.22′) to mean that all test vectors must have their arrowheads located on or below the second constraint border. Consequently, all these vectors must again fall within the feasible region, and a qualifying arc can be found for each. The analysis of point R is analogous.
4. At points such as L and N, neither variable is zero and (21.20) can be ignored. However, for point N, (21.21) dictates that

   (21.23) $$g_1^1\, dx_1 + g_2^1\, dx_2 = a_{11}\, dx_1 + a_{12}\, dx_2 \leq 0$$

   Since point N satisfies $a_{11}x_1 + a_{12}x_2 = r_1$ (first constraint border), we may add this equality to (21.23) and write

   (21.23′) $$a_{11}(x_1 + dx_1) + a_{12}(x_2 + dx_2) \leq r_1$$

   This would require the test vectors to have arrowheads located on or below the first constraint border in Fig. 21.6. Thus we obtain essentially the same kind of result encountered in the other cases. The analysis of point L is analogous.
5. At point M, we may again disregard (21.20), but this time (21.21) requires all test vectors to satisfy both (21.22) and (21.23). Since we may modify the latter conditions to the forms in (21.22′) and (21.23′), all test vectors must now have their arrowheads located on or below the first as well as the second constraint borders. The result thus again duplicates those of the previous cases.

In this example, it so happens that, for every type of boundary point considered, the test vectors all lie within the feasible region. While this locational feature makes the qualifying arcs easy to find, it is by no means a prerequisite for their existence. In a problem with a nonlinear constraint border, in particular, the constraint border itself may serve as a qualifying arc for some test vector that lies outside of the feasible region. An example of this can be found in one of the problems in Exercise 21.3.

EXERCISE 21.3

1 Check whether the solution point $(\bar{x}_1, \bar{x}_2) = (2, 6)$ in Example 3 satisfies the constraint qualification.

2 Maximize $\pi = x_1$
subject to $x_1^2 + x_2^2 \le 1$
and $x_1, x_2 \ge 0$
Solve graphically and check whether the optimal-solution point satisfies (a) the constraint qualification, and (b) the Kuhn-Tucker maximum conditions.

3 Minimize $C = x_1$
subject to $x_1^2 - x_2 \ge 0$
and $x_1, x_2 \ge 0$
Solve graphically. Does the optimal solution occur at a cusp? Check whether the optimal solution satisfies (a) the constraint qualification, and (b) the Kuhn-Tucker minimum conditions.

4 Minimize $C = 2x_1 + x_2$
subject to $x_1^2 - 4x_1 + x_2 \ge 0$
$-2x_1 - 3x_2 \ge -12$
and $x_1, x_2 \ge 0$
Solve graphically for the global minimum, and check whether the optimal solution satisfies (a) the constraint qualification, and (b) the Kuhn-Tucker conditions. (*Hint:* The feasible region is identical with that depicted in Fig. 21.2.)

5 Minimize $C = x_1$
subject to $-x_2 - (1 - x_1)^3 \ge 0$
and $x_1, x_2 \ge 0$
Show that (a) the optimal solution $(\bar{x}_1, \bar{x}_2) = (1, 0)$ does not satisfy the Kuhn-Tucker conditions, but (b) by introducing a new multiplier $y_0 \ge 0$, and modifying the Lagrangian function (21.17) to the form

$$Z_0 = y_0 f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{m} y_i \left[r_i - g^i(x_1, x_2, \ldots, x_n) \right]$$

the Kuhn-Tucker conditions can be satisfied at $(1, 0)$. (*Note:* The Kuhn-Tucker conditions on the multipliers extend only to $y_1, \ldots, y_m$, but not to y_0.)

21.4 KUHN-TUCKER SUFFICIENCY THEOREM: CONCAVE PROGRAMMING

Our discussion hitherto has been concerned with *necessary* conditions for a maximum or minimum in nonlinear programming. Necessary conditions are useful as a screening device for rejecting unqualified candidates for optimal solution. In nonlinear programming, for instance, any interior point in the feasible region that fails the Kuhn-Tucker conditions cannot possibly be an optimal solution. Similarly, a boundary point that satisfies the constraint qualification but fails the Kuhn-Tucker conditions can safely be ruled out as an optimal solution. If a point x^0 does meet the necessary conditions, however, we should not automatically conclude that x^0 constitutes an optimal solution, because certain unqualified candidates may nonetheless pass the screening test, the way an inflection point can satisfy the first-order condition $dy/dx = 0$ in the simplest optimization problem. In other words, with the necessary conditions as a "fishing net," we may catch genuine optimal solutions as well as spurious ones.

To have a *sufficient* condition is a different story, for if a point $\bar{x}$ satisfies a sufficient condition for a maximum, then that point must maximize the objective function. In this sense, sufficient conditions provide a more definitive type of test. But they also happen to have a shortcoming of their own, namely, a sufficient condition as such may not be necessary, so that a genuine optimal solution may nonetheless fail to satisfy the sufficient condition. In other words, with a sufficient condition as a "fishing net," we may very well fail to catch a true optimal solution.

The most gratifying situation is of course one where we possess a necessary-and-sufficient condition. All optimal solutions can be caught by means of such a condition, and yet we do not run the risk of admitting any unqualified candidates.

In this section, we shall show that, under certain circumstances, the Kuhn-Tucker conditions can be regarded as *sufficient* conditions for an extremum, or they may even emerge as *necessary-and-sufficient* conditions.

The Kuhn-Tucker Sufficiency Theorem

In classical optimization problems, the sufficient conditions for maximum and minimum are traditionally expressed in terms of the signs of second-order derivatives or differentials. As we have shown in Sec. 11.5, however, these second-order conditions are closely related to the concepts of concavity and convexity of the objective function. Here, in nonlinear programming, the sufficient conditions will be stated directly in terms of concavity and convexity. And, in fact, these concepts will be applied not only to the objective function $f(x)$, but to the constraint functions $g^i(x)$ as well.

For the *maximization* problem, Kuhn and Tucker offer the following statement of sufficient conditions (sufficiency theorem):

Given the nonlinear program

$$\begin{array}{lll} \text{Maximize} & \pi = f(x) & \\ \text{subject to} & g^i(x) \leq r_i & (i = 1, 2, \ldots, m) \\ \text{and} & x \geq 0 & \end{array}$$

if the following conditions are satisfied:

(a) the objective function $f(x)$ is differentiable and *concave* in the nonnegative orthant (*b*) each constraint function $g^i(x)$ is differentiable and *convex* in the nonnegative orthant

(*c*) the point $\bar{x}$ satisfies the Kuhn-Tucker maximum conditions

then $\bar{x}$ gives a global maximum of $\pi = f(x)$.

Note that, in this theorem, the constraint qualification is nowhere mentioned. This is because we have already assumed, in condition (*c*), that the Kuhn-Tucker conditions are satisfied at $\bar{x}$ and, consequently, the question of the constraint qualification is no longer an issue.

As it stands, the above theorem indicates that conditions (*a*), (*b*), and (*c*) are sufficient to establish $\bar{x}$ to be an optimal solution. Looking at it differently, however, we may also interpret it to mean that, given (*a*) and (*b*), then the Kuhn-Tucker maximum conditions are sufficient for a maximum. In the preceding section, we learned that the Kuhn-Tucker conditions, though not necessary per se, become necessary when the constraint qualification is satisfied. Combining this information with the sufficiency theorem, we may now state that, if the constraint qualification is satisfied and if conditions (*a*) and (*b*) are realized, then the Kuhn-Tucker maximum conditions will be *necessary-and-sufficient* for a maximum. This would be the case, for instance, when all the constraints are linear inequalities, which is sufficient for satisfying the constraint qualification. Later, we shall introduce another set of circumstances that will guarantee the satisfaction of the constraint qualification, even if the $g^i(x)$ functions are not all linear.

The maximization problem dealt with in the sufficiency theorem above is often referred to as *concave programming*. This name arises because Kuhn and Tucker adopt the $\geq$ inequality instead of the $\leq$ inequality in every constraint, so that condition (*b*) would require the $g^i(x)$ functions to be *all concave*, like the $f(x)$ function. But we have modified the formulation in order to achieve consistency with our earlier discussion of linear programming. Though different in form, the two formulations are of course equivalent in substance.

As stated above, the sufficiency theorem deals only with maximization problems. But adaptation to *minimization* problems is by no means difficult. Aside from the appropriate changes in the theorem to reflect the reversal of the problem itself, all we have to do is to interchange the two words *concave* and *convex* in conditions (*a*) and (*b*) and to use the Kuhn-Tucker *minimum* conditions in condition (*c*). (See Exercise 21.4-1.)

Since the Kuhn-Tucker sufficiency theorem is relatively easy to prove, we shall reproduce the proof below.

Proof of the Sufficiency Theorem

For the maximization problem, the Lagrangian function can be expressed as

$$(21.24) \qquad Z = f(x) + \sum_{i=1}^{m} \bar{y}_i\left[r_i - g^i(x)\right] \qquad [\text{cf. } (21.17)]$$

where we have assigned the specific values $\bar{y}_i$ to the Lagrange multipliers, thereby making Z a function of the x_j variables alone. In line with conditions (a) and (b) of the sufficiency theorem, let us postulate $f(x)$ to be concave and each $g^i(x)$ to be convex, which makes each $-g^i(x)$ concave. Then Z, being a sum of concave functions, must also be concave in x. According to (11.24′), the concavity of Z implies that

$$(21.25) \qquad Z(x) \le Z(\bar{x}) + \sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j}(x_j - \bar{x}_j)$$

where $\bar{x}$ is some specific point in the domain, and $\partial Z/\partial \bar{x}_j$ is the partial derivative $\partial Z/\partial x_j$ evaluated at $\bar{x}$. In particular, let us select as $\bar{x}$ and $\bar{y}$ those values of choice variables and Lagrange multipliers that satisfy the Kuhn-Tucker maximum conditions, in line with condition (c) of the sufficiency theorem. Then the Σ expression in (21.25) can be shown to be nonpositive, so that its deletion will not upset the inequality, thereby enabling us to infer that $Z(x) \le Z(\bar{x})$.

To see this, decompose the Σ expression into two terms:

$$T_1 = \sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j} x_j \qquad \text{and} \qquad T_2 = -\sum_{j=1}^{n} \frac{\partial Z}{\partial \bar{x}_j} \bar{x}_j$$

By virtue of complementary slackness at point $\bar{x}$, T_2 must vanish. As to T_1, which involves x_j rather than $\bar{x}_j$, so that complementary slackness does not apply, we can only be sure that $\partial Z/\partial \bar{x}_j \le 0$ (marginal condition) and $x_j \ge 0$ (model specification) for every j. Thus T_1 is nonpositive. Adding up the two terms, we find the Σ expression nonpositive, and therefore we can conclude that

$$(21.26) \qquad Z(x) \le Z(\bar{x})$$

or, by reference to (21.24),

$$(21.26') \qquad f(x) + \sum_{i=1}^{m} \bar{y}_i\left[r_i - g^i(x)\right] \le f(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i\left[r_i - g^i(\bar{x})\right]$$

If we can show that the two Σ expressions in (21.26′) can be deleted without upsetting the inequality, then we can conclude that $f(x) \le f(\bar{x})$, which will establish the claim that $\bar{x}$ maximizes the objective function $f(x)$. This indeed can be done. The Σ expression on the left is necessarily nonnegative because, for each i, we have $\bar{y}_i \ge 0$ (nonnegativity) and $r_i - g^i(x) \ge 0$ (constraint specification). In contrast, the Σ expression on the right must vanish, because $[r_i - g^i(\bar{x})]$ is nothing but $\partial Z/\partial \bar{y}_i$, so that complementary slackness applies. Consequently, $f(x) \le f(\bar{x})$, and $\bar{x}$ is indeed the optimal solution.

The maximum value $\bar{\pi} = f(\bar{x})$ is a *global* maximum. One way of showing this is that the validity of the inequality $f(x) \le f(\bar{x})$ does not hinge on x being located in any restricted neighborhood of $\bar{x}$. More formally, however, we may have recourse to the globality theorem introduced in Sec. 19.3. Our objective function is assumed to be differentiable and concave. As to the constraints, since each $g^i(x)$ is convex, the earlier result in (11.27)—when properly generalized to the n-dimensional case—would indicate that the set

$$S_i^{\le} \equiv \{x \mid g^i(x) \le r_i\}$$

must be a convex set, in fact a closed convex set. Moreover, the feasible region is simply the intersection of the closed convex sets $S_i^{\le}$, $(i = 1, 2, \ldots, m)$, so the feasible set is also a closed convex set. Consequently, the globality theorem is applicable, and any local maximum must be a global maximum. Note that, if the objective function $f(x)$ is *strictly* concave, then the global maximum will be unique.

In the above, by starting from the postulated conditions that (a) the $f(x)$ function is differentiable and concave, (b) each $g^i(x)$ function is differentiable and convex, and (c) point $\bar{x}$ satisfies the Kuhn-Tucker maximum conditions, we deduced that $\bar{x}$ does maximize π. The sufficiency theorem is therefore validated.

Saddle Point

The result in (21.26)—derived on the basis of a specifically chosen y—means that, given $\bar{y}$, $\bar{x}$ *maximizes* the Lagrangian function Z among all admissible x. As it turns out, it is also true that, given $\bar{x}$, $\bar{y}$ *minimizes* Z among all admissible y. That is, we have in fact

$$Z(x, \bar{y}) \le Z(\bar{x}, \bar{y}) \le Z(\bar{x}, y) \tag{21.27}$$

Since this situation is akin to the one depicted in Fig. 11.3a, the point $(\bar{x}, \bar{y})$ is called a *saddle point*, and $Z(\bar{x}, \bar{y})$ a *saddle value* of the Lagrangian function.

It may be recalled that the first inequality in (21.27), which is merely a restatement of (21.26) with the explicit mention that $\bar{y}$ is given, follows from the fact that Z is *concave* in the variables x, given $\bar{y}$. We can also show that the second inequality in (21.27) is tied to the fact that Z is *convex* in the variables y, given $\bar{x}$. Setting $x = \bar{x}$ in the Lagrangian function, we have

$$Z = f(\bar{x}) + \sum_{i=1}^{m} y_i\left[r_i - g^i(\bar{x})\right]$$

Since r_i and $g^i(\bar{x})$ represent constants, Z is clearly linear in the variables y_i, and thus convex in y_i. With reference again to (11.24′), the convexity of Z in y_i carries the implication that

$$Z(y) \ge Z(\bar{y}) + \sum_{i=1}^{m} \frac{\partial Z}{\partial \bar{y}_i}(y_i - \bar{y}_i) \qquad \text{[cf. (21.25)]} \tag{21.28}$$

where we may take $\bar{y}$ to be the y value satisfying the Kuhn-Tucker maximum conditions. Then, by the same type of analysis of (21.25), we can show that the Σ expression in (21.28) must be nonnegative, so that its deletion does not affect the validity of the inequality. Hence, we get $Z(\bar{y}) \leq Z(y)$ or, more completely, $Z(\bar{x}, \bar{y}) \leq Z(\bar{x}, y)$. In view of this discussion, we may characterize a saddle point $(\bar{x}, \bar{y})$ as a point where the given function $Z(x, y)$ is concave with respect to one set of independent variables (here, x), but convex with respect to another (here, y).

When applied to *linear* programming, the saddle-point notion is especially interesting, because it leads directly to the concept of duality. In general, a pair of primal and dual linear programs may be written in vector notation as follows:

Maximize	$\pi = c'x$	Minimize	$\pi^* = r'y$
subject to	$Ax \leq r$	subject to	$A'y \geq c$
and	$x \geq 0$	and	$y \geq 0$

The Lagrangian function (21.17) for each of these emerges in the form

$$Z = c'x + y'(r - Ax) = c'x + y'r - y'Ax \qquad \text{[primal]}$$

and

$$Z' = r'y + x'(c - A'y) = r'y + x'c - x'A'y \qquad \text{[dual]}$$

Note that Z and Z' are transposes of each other. Since Z and Z' are both (1×1), they obviously represent an identical scalar. Thus we are actually maximizing the identical Lagrangian function with respect to the x_j variables (primal program), while minimizing it with respect to the y_i variables (dual program). When the optimal solutions $\bar{x}$ and $\bar{y}$ are attained in these two programs, they must accordingly satisfy the inequality (21.27). This goes to show that the concept of duality and the notion of saddle point amount to the same thing.

By complementary slackness, the $y'(r - Ax)$ component of Z above must vanish in the optimal solution [cf. (21.14), the last equation]. Similarly, the $x'(c - A'y)$ component of Z' must vanish. These facts give us the essence of Duality Theorem II of Sec. 20.1. Going a step further, by subtracting these vanishing components from Z and Z', respectively, we find in the optimal solution that $\bar{Z} = c'\bar{x} = \bar{\pi}$ and $\bar{Z}' = r'\bar{y} = \bar{\pi}^*$. The fact that $\bar{Z} = \bar{Z}'$ then implies $\bar{\pi} = \bar{\pi}^*$, which is the essence of Duality Theorem I. It thus appears that the duality feature of linear programming reveals itself most clearly when the problem is viewed in the context of nonlinear programming.

EXERCISE 21.4

1 Given: Minimize $C = F(x)$
subject to $G^i(x) \geq r_i \quad (i = 1, 2, \ldots, m)$
and $x \geq 0$

(*a*) Convert it into a maximization problem.

(*b*) What in the present problem are the equivalents of the f and g^i functions in the Kuhn-Tucker sufficiency theorem?

(*c*) Hence, what concavity-convexity conditions should be placed on the F and G^i functions to make the sufficient conditions for a maximum applicable here?

(*d*) On the basis of the above, how would you state the Kuhn-Tucker sufficient conditions for a *minimum*?

2 Demonstrate that, when the sufficient conditions for a minimum are satisfied by the minimization program given in the preceding problem, the globality theorem will again be applicable, so that the minimum attained will be global.

3 In (21.27), does $Z(\bar{x}, \bar{y})$ constitute a unique saddle value? How would you modify (21.27) to describe a unique saddle value?

4 Is the Kuhn-Tucker sufficiency theorem applicable to:

(*a*) Maximize $\pi = x_1$
subject to $x_1^2 + x_2^2 \le 1$
and $x_1, x_2 \ge 0$

(*b*) Minimize $C = (x_1 - 3)^2 + (x_2 - 4)^2$
subject to $x_1 + x_2 \ge 4$
and $x_1, x_2 \ge 0$

(*c*) Minimize $C = 2x_1 + x_2$
subject to $x_1^2 - 4x_1 + x_2 \ge 0$
and $x_1, x_2 \ge 0$

21.5 ARROW-ENTHOVEN SUFFICIENCY THEOREM: QUASICONCAVE PROGRAMMING

To apply the Kuhn-Tucker sufficiency theorem, certain concavity-convexity specifications must be met. These constitute quite stringent requirements. In another sufficiency theorem—the Arrow-Enthoven sufficiency theorem*—these specifications are relaxed to the extent of requiring only *quasiconcavity* and *quasiconvexity* in the objective and constraint functions. With the requirements thus weakened, the scope of applicability of the sufficient conditions is correspondingly widened.

In the original formulation of the Arrow-Enthoven paper, with a maximization problem and with constraints in the $\ge$ form, the $f(x)$ and $g^i(x)$ functions must uniformly be quasiconcave in order for their theorem to be applicable. This gives rise to the name *quasiconcave programming*. In our discussion here, however, we shall again use the $\le$ inequality in the constraints of a maximization problem and the $\ge$ inequality in the minimization problem.

* Kenneth J. Arrow and Alain C. Enthoven, "Quasi-concave Programming," *Econometrica*, October, 1961, pp. 779–800.

The Arrow-Enthoven Sufficiency Theorem

The theorem is as follows:

Given the nonlinear program

Maximize	$\pi = f(x)$	
subject to	$g^i(x) \le r_i$	$(i = 1, 2, \ldots, m)$
and	$x \ge 0$	

if the following conditions are satisfied:

(a) the objective function $f(x)$ is differentiable and *quasiconcave* in the nonnegative orthant

(b) each constraint function $g^i(x)$ is differentiable and *quasiconvex* in the nonnegative orthant

(c) the point $\bar{x}$ satisfies the Kuhn-Tucker maximum conditions

(d) any *one* of the following is satisfied:

- (d-i) $f_j(\bar{x}) < 0$ for at least one variable x_j
- (d-ii) $f_j(\bar{x}) > 0$ for some variable x_j that can take on a positive value without violating the constraints
- (d-iii) the n derivatives $f_j(\bar{x})$ are not all zero, and the function $f(x)$ is twice differentiable in the neighborhood of $\bar{x}$ [i.e., all the second-order partial derivatives of $f(x)$ exist at $\bar{x}$]
- (d-iv) the function $f(x)$ is concave

then $\bar{x}$ gives a global maximum of $\pi = f(x)$.

Since the proof of this theorem is somewhat involved, we shall omit it here. However, we do want to call your attention to a few important features of this theorem. For one thing, while Arrow and Enthoven have succeeded in weakening the concavity-convexity specifications to their quasiconcavity-quasiconvexity counterparts, they find it necessary to append a new requirement, (d). Note, though, that only *one* of the four alternatives listed under (d) is required to form a complete set of sufficient conditions. In effect, therefore, the above theorem contains as many as *four* different sets of sufficient conditions for a maximum. In the case of (d-iv), with $f(x)$ concave, it would appear that the Arrow-Enthoven sufficiency theorem becomes identical with the Kuhn-Tucker sufficiency theorem. But this is not true. Inasmuch as Arrow and Enthoven only require the constraint functions $g^i(x)$ to be *quasi*convex, their sufficient conditions are still weaker.

As stated, the theorem lumps together the conditions (a) through (d) as a set of sufficient conditions. But it is also possible to interpret it to mean that, when (a), (b), and (d) are satisfied, then the Kuhn-Tucker maximum conditions become sufficient conditions for a maximum. Furthermore, if the constraint qualification is also satisfied, then the Kuhn-Tucker conditions will become necessary-and-sufficient for a maximum.

Like the Kuhn-Tucker theorem, the Arrow-Enthoven theorem can be adapted with ease to the *minimization* framework. Aside from the obvious changes that are needed to reverse the direction of optimization, we simply have to interchange the words *quasiconcave* and *quasiconvex* in conditions (a) and (b), replace the Kuhn-Tucker maximum conditions by the minimum conditions, reverse the inequalities in (d-i) and (d-ii), and change the word *concave* to *convex* in (d-iv).

A Constraint-Qualification Test

It was earlier mentioned that if all constraint functions are linear, then the constraint qualification is satisfied. In case the $g^i(x)$ functions are nonlinear, the following test offered by Arrow and Enthoven may prove useful in determining whether the constraint qualification is satisfied:

For a maximization problem, if

(a) every constraint function $g^i(x)$ is differentiable and quasiconvex

(b) there exists a point x^0 in the nonnegative orthant such that all the constraints are satisfied as strict inequalities at x^0

(c) *one* of the following is true:

- (c-i) every $g^i(x)$ function is convex
- (c-ii) the partial derivatives of every $g^i(x)$ are not all zero when evaluated at every point x in the feasible region

then the constraint qualification is satisfied.

Again, this test can be adapted to the minimization problem with ease. To do so, just change the word *quasiconvex* to *quasiconcave* in condition (a), and change the word *convex* to *concave* in (c-i). The application of this test will be illustrated later.

EXERCISE 21.5

1 Justify the changes suggested in the text for the Arrow-Enthoven sufficiency theorem and the constraint-qualification test when these are applied to a *minimization* problem.

2 Is the constraint-qualification test in the nature of a set of necessary conditions, or sufficient conditions?

3 Which of the following functions are mathematically acceptable as the objective function of a *maximization* problem which qualifies for the application of the Arrow-Enthoven sufficiency theorem?

(a) $f(x) = x^3 - 2x$ (b) $f(x_1, x_2) = 6x_1 - 9x_2$ (c) $f(x_1, x_2) = x_2 - \ln x_1$

(*Note:* See Exercise 12.4-4.)

4 Is the constraint qualification satisfied, given that the constraints of a *maximization* problem are:

(*a*) $x_1^2 + (x_2 - 5)^2 \leq 4$ and $5x_1 + x_2 \leq 10$

(*b*) $x_1 + x_2 \leq 8$ and $-x_1 x_2 \leq -8$ (*Note:* $-x_1 x_2$ is not convex.)

21.6 ECONOMIC APPLICATIONS

We are now ready to show some simple economic applications of the two sufficiency theorems.

Utility Maximization Revisited

In our previous encounter with consumer theory in Sec. 12.5, the problem was stated in the form of: "Maximize $U = U(x_1, x_2)$, subject to $P_1 x_1 + P_2 x_2 = B$." The constraint appears as an equality, and the choice variables are not specifically restricted to be nonnegative. It is now possible to restate the problem in a more realistic format. Taking the n-variable case, we can write it as a nonlinear program

$$\begin{array}{lll} & \text{Maximize} & U = U(x_1, \ldots, x_n) \\ (21.29) & \text{subject to} & P_1 x_1 + \cdots + P_n x_n \leq B \\ & \text{and} & x_1, \ldots, x_n \geq 0 \end{array}$$

where all the prices are taken to be exogenous. In this new formulation, the consumer is allowed the option of spending less than the budget amount B. Also, the fact that the amount purchased of any commodity cannot be negative is explicitly stated in the problem.

Since the (only) constraint function $g^1(x) = P_1 x_1 + \cdots + P_n x_n$ is linear, the constraint qualification is satisfied, and the Kuhn-Tucker maximum conditions are necessary-and-sufficient if the other conditions cited in either of the two sufficiency theorems are met. To apply the Kuhn-Tucker theorem, we would want $U(x)$ to be differentiable and concave, and $g^1(x)$ to be differentiable and convex. The latter is automatically satisfied since $g^1(x)$ is linear, but the former has to be taken care of by explicit model specification. In checking the Arrow-Enthoven theorem, however, we note that it is only needed to specify $U(x)$ to be differentiable and quasiconcave—a considerably less stringent requirement—provided, of course, that condition (d) can be duly satisfied. If, for instance, there exists some commodity j that lies within the reach of the consumer's budget and yields a positive marginal utility always (nonsatiation), then we will automatically have $U_j(\bar{x}) > 0$ and condition (*d-ii*) will be duly met. But there are of course other ways to satisfy condition (d) so as to make the Kuhn-Tucker maximum conditions necessary-and-sufficient.

The quasiconcavity specification is perfectly in line with our earlier discussion of the classical consumer theory in Sec. 12.5. Recall, however, that there—in a two-commodity model with an equality constraint—the $U(x)$ function was specified to be *strictly* quasiconcave. The purpose of the strictness assumption was to rule out any horizontal platforms on the bell-shaped utility surface in the x_1x_2U space, so that each isovalue set (here, indifference set) will emerge as a thin curve rather than a wide zone. For only then would it make sense to speak of a point-of-tangency optimal solution in the classical tradition. Things are different when the budget constraint is an inequality instead. Since what we are seeking is no longer a point of tangency, but an optimal point within a feasible region, it is now acceptable to have horizontal platforms on the utility surface, and there is thus no need to specify *strict* quasiconcavity.

In Sec. 12.5, the $U(x)$ function was also assumed to be an *increasing* function of x_1 and x_2. Economically, this implies nonsatiation in both commodities. Graphically, this restricts the utility surface to the ascending portion of a bell only, so that the indifference curves will be the usual downward-sloping, convex type rather than the round or oval loops illustrated in Fig 21.7. The rationale for this assumption is again to be found in the tangency-solution feature of the classical framework. For although a tangency point like H would indeed represent the best choice under the circumstances (because ascent to the peak of the utility surface, via a movement toward point A in the base plane, is precluded by the budget), tangency points on the far side of the peak, such as J, would clearly be nonoptimal and should categorically be excluded. Turning to the nonlinear-programming framework, we see that the assumption of nonsatiation can also be of service—to satisfy condition (d-ii). However, since condition (d) in the Arrow-Enthoven sufficiency theorem can be taken care of in other ways, it is not absolutely necessary to assume that $U(x)$ is increasing.

The Kuhn-Tucker conditions for this problem are easy to write and interpret. These will therefore be left to you as an exercise.

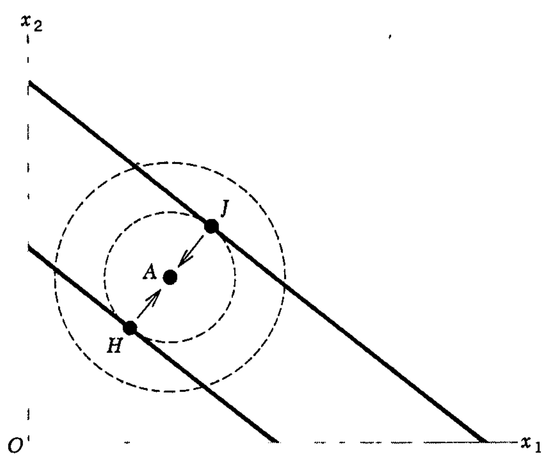

Figure 21.7

Least-Cost Combination Revisited

Let the production function of a firm be of the Cobb-Douglas type: $Q = K^\alpha L^\beta$, where $0 < \alpha, \beta < 1$. Also let $P_K > 0$ and $P_L > 0$ (the prices of capital service and labor service) be exogenously determined. Then the least-cost combination problem can be stated as the nonlinear program

$$\begin{array}{lll} & \text{Minimize} & C = P_K K + P_L L \\ (21.30) & \text{subject to} & K^\alpha L^\beta \geq Q_0 \qquad (Q_0 > 0) \\ & \text{and} & K, L \geq 0 \end{array}$$

This formulation differs from the one in Sec. 12.7 in that it permits the firm to produce more than the requisite quantity Q_0 and that explicit recognition is given to the fact that the amount used of any input cannot be negative. Indeed, given that Q_0 is positive, the inputs K and L must both be positive in the Cobb-Douglas case.

Because the objective function is linear, it is automatically convex as well as quasiconvex. The (only) constraint function, $K^\alpha L^\beta$, is quasiconcave for positive K and L regardless of whether $\alpha + \beta \gtreqless 1$, that is, regardless of whether returns to scale are increasing, constant, or decreasing [see Sec. 12.4, Example 5, and the discussion of (12.52)]. However, this function is *not concave* if $\alpha + \beta > 1$ (increasing returns to scale). In the latter case, each successive unit increment in output will require less than proportionate increase in the two outputs. Geometrically, the successive isoquants for unit increments in output will become more and more tightly spaced as we move away from the point of origin on any given ray in the KL plane. Accordingly, the projection of such a ray onto the production surface will produce a curve showing increasingly more rapid ascent, so that the surface cannot be concave.

In view of this, the Kuhn-Tucker sufficiency theorem becomes inapplicable when there are increasing returns to scale. In contrast, the Arrow-Enthoven theorem is applicable even then. For a minimization problem, the latter theorem specifies a quasiconvex objective function and quasiconcave constraint function(s). These specifications are indeed met in the present problem. In addition, condition (*d-i*) is satisfied, because for a minimization problem we are supposed to have $f_j(\bar{x}) > 0$ for some choice variable x_j, and we do have here $\partial C/\partial K = P_K > 0$ and $\partial C/\partial L = P_L > 0$ for all values of K and L, including $\bar{K}$ and $\bar{L}$.

Going a step further, we may also ascertain that the constraint function $g^1 = K^\alpha L^\beta$ passes the constraint-qualification test. First, g^1 is a quasiconcave function. Second, there certainly exists a nonnegative ordered pair (K, L) that satisfies the constraint as a strict inequality. Third, since every point in the feasible region is characterized by $K > 0$ and $L > 0$ (otherwise Q_0 cannot be positive), the partial derivatives $\partial g^1/\partial K = \alpha K^{\alpha-1}L^\beta$ and $\partial g^1/\partial L = \beta K^\alpha L^{\beta-1}$ are both positive everywhere in the feasible region, thus satisfying condition (*c-ii*) of the test. Consequently, with the constraint qualification satisfied, the Kuhn-Tucker minimum conditions can be taken as necessary-and-sufficient according to the Arrow-Enthoven theorem.

The Sales-Maximizing Firm

In the standard analysis of a firm, the objective of profit maximization is usually assumed. However, when the firm in question is a corporation in which ownership and management are separate, it may very well be rational for the management to pursue the alternative goal of maximizing the sales (revenue).* The total revenue is often taken as an important indicator of the competitive position of the firm within the industry. Moreover, increases in sales revenue are often taken to be a sign of managerial success. It is even conceivable that the remuneration of the management depends directly on this particular performance index. Thus sales maximization appears to be a plausible alternative objective in the corporate setup, provided that, to avoid possible stockholder discontent, the management always sees to it that the profit level does not fall below a certain prescribed minimum, say, π_0, which is below the maximum profit associated with the MR = MC condition.

If so, the problem of the management is to maximize $R = R(Q)$, subject to $\pi = R(Q) - C(Q) \geq \pi_0$, or

$$\begin{array}{lll} \text{Maximize} & R = R(Q) & \\ \text{subject to} & C(Q) - R(Q) \leq -\pi_0 & (\pi_0 > 0) \\ \text{and} & Q \geq 0 & \end{array}$$

As long as $R(Q)$ is differentiable and concave and $C(Q)$ is differentiable and convex—which would imply that the constraint function $C(Q) - R(Q)$ is also differentiable and convex—the Kuhn-Tucker sufficiency theorem can be applied. It is interesting to note that we cannot hope to attain greater generality in this case by relaxing the concavity-convexity assumptions to a quasiconcave $R(Q)$ and a quasiconvex $C(Q)$. Under these weaker assumptions, the constraint function $C(Q) - R(Q)$ is the sum of two quasiconvex functions, but there is no guarantee that it itself is also quasiconvex. Hence the Arrow-Enthoven sufficiency theorem does not automatically become applicable, unless, of course, the constraint function is rendered quasiconvex via a direct, new assumption.

The case of a concave $R(Q)$ function and a convex $C(Q)$ function is illustrated in Fig. 21.8. The two curves in diagram *a* are drawn on the assumption that $R(0) = 0$ and $C(0) > 0$. The curve in diagram *b*, representing the constraint function, is the negative—the mirror image with reference to the horizontal axis—of the profit curve. While any output level in the open interval (Q_1, Q_6) yields a

* William J. Baumol, "On the Theory of Oligopoly," *Econometrica*, August, 1958, pp. 187–198. See also Baumol's *Business Behavior, Value and Growth*, revised edition, Harcourt, Brace, & World, Inc., New York, 1967.

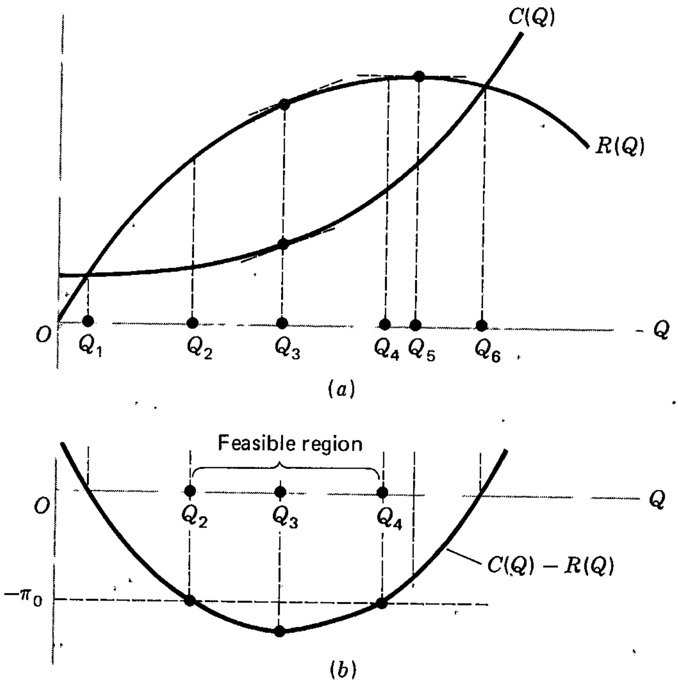

Figure 21.8

positive profit, only the Q levels in the closed interval $[Q_2, Q_4]$ involve a profit no less than π_0 and satisfy the profit constraint. Thus we can identify the interval $[Q_2, Q_4]$ as the feasible region. As may be expected, the feasible region includes Q_3, the output that maximizes the profit.

For this problem, with the Lagrangian function

$$Z = R(Q) + y[-\pi_0 - C(Q) + R(Q)]$$

the Kuhn-Tucker conditions consist of the marginal conditions

$$(21.31)\qquad \begin{aligned} \frac{\partial Z}{\partial Q} &= R'(Q) - yC'(Q) + yR'(Q) \le 0 \\ \frac{\partial Z}{\partial y} &= -\pi_0 - C(Q) + R(Q) \ge 0 \end{aligned}$$

plus the nonnegativity and complementary-slackness conditions. Since $R(0) = 0$ and $C(0) > 0$, a zero output would yield $\partial Z/\partial y = -\pi_0 - C(0) < 0$, which violates the second marginal condition. We must therefore take $Q > 0$ instead—a requirement which is perfectly consistent with the fact that the zero output level lies outside the feasible region. The positivity of Q implies, by complementary slackness, that $\partial Z/\partial Q = 0$; that is, the first weak inequality in (21.31) must be satisfied as an equality. Solving the latter, we then obtain the (constrained) sales-maximizing output rule

$$(21.32) \qquad R'(Q) = \frac{y}{1+y} C'(Q)$$

In this result, the value of y can be either zero or positive. If $y = 0$, the rule reduces to $R'(Q) = 0$, and the firm will push its output all the way to the level where the marginal revenue vanishes—Q_5 in Fig. 21.8. This would be sales maximization in its purest form, because the firm would in that case proceed to the very peak of the total-revenue curve. But such extreme behavior is unfeasible under our assumptions, because Q_5 also lies outside the feasible region. Accordingly, we must take $y > 0$. By complementary slackness, this then implies that $\partial Z/\partial y = 0$, which in turn implies that the profit constraint is to be satisfied as an equality, with the firm attempting to earn just π_0, the minimum profit required.

With a positive y, the sales-maximizing output rule (21.32) indicates that

$$R'(Q) < C'(Q) \qquad \left[\text{since } \frac{y}{1+y} < 1\right]$$

and this would generally yield a higher output level than the profit-maximizing rule $R'(Q) = C'(Q)$. In Fig. 21.8, the profit-maximizing output is Q_3. The output level that satisfies rule (21.32) and yields the exact amount of the minimum required profit is Q_4, which indeed exceeds Q_3.

Solving a Nonlinear Program via the Kuhn-Tucker Conditions

Our discussion of the Kuhn-Tucker conditions has thus far been concentrated on their *analytical* roles (in necessary conditions and sufficient conditions), but these conditions can also play a *computational* role if they happen to be necessary-and-sufficient,* and if the number of choice variables is reasonably small. We shall illustrate the procedure involved with a numerical example of a sales-maximizing firm.

Let the firm have revenue and cost functions

$$R = 32Q - Q^2 \quad [\text{concave}] \qquad \text{and} \qquad C = Q^2 + 8Q + 4 \quad [\text{convex}]$$

and let the minimum profit be $\pi_0 = 18$. Under these circumstances, the Kuhn-Tucker conditions indeed are necessary-and-sufficient. On the basis of (21.31), the two marginal conditions will now take the specific form of

$$\frac{\partial Z}{\partial Q} = 32 - 2Q - y(4Q - 24) \le 0$$

$$\frac{\partial Z}{\partial y} = -2Q^2 + 24Q - 22 \ge 0$$

The value of Q must be either positive or zero. Trying $Q = 0$ first, we immediately

* If they are only sufficient, but not necessary, then an optimal solution may fail the Kuhn-Tucker conditions, so the latter will become ineffective as a "fishing net" to catch such a solution.

encounter a contradiction in the second marginal condition. Thus we must let $Q > 0$ instead. If so, we may infer from complementary slackness that $\partial Z/\partial Q - 0$, which gives us an equation in two variables.

The value of y in that equation must be either positive or zero. Trying $y = 0$ and solving for Q, we get $Q = 16$. Although acceptable in itself, this value of Q would imply $\partial Z/\partial y = -150$, in violation of the second marginal condition (the profit constraint). Thus we must let $y > 0$ instead. But then it follows that $\partial Z/\partial y = 0$ by complementary slackness. Solving this new equation; we obtain the two roots $Q_1 = 11$ and $Q_2 = 1$. But since only the first root is consistent with the condition $\partial Z/\partial Q = 0$, the sales-maximizing output is $\bar{Q} = 11$. In contrast, the $R'(Q) = C'(Q)$ rule would lead to a lower figure, $Q = 6$, as the profit-maximizing output.

From the above, it should be clear that the computational procedure involved is essentially one of trial and error. The basic idea is first to try a zero value for each choice variable. Setting a variable equal to zero always simplifies the marginal conditions by causing certain terms to drop out. If appropriate nonnegative values of Lagrange multipliers can then be found that satisfy all the marginal inequalities, the zero solution will be optimal. If, on the other hand, the zero solution violates some of the inequalities, then we must let one or more choice variables be positive. For every positive choice variable, we may, by complementary slackness, convert a weak inequality marginal condition into a strict equality. Properly solved, such an equality will lead us either to a solution, or to a contradiction that would then compel us to try something else. If a solution exists, such trials will eventually enable us to uncover it.

Note that, in a two-variable problem, we must try out no less than four (2^2) possible combinations of choice-variable signs: $(0,0)$, $(0,+)$, $(+,0)$, and $(+,+)$. A three-variable problem will present as many as either (2^3) possibilities to check, and the complexity will grow rapidly as the number of choice variables increases. This method is therefore unsuitable for high-dimensional nonlinear programs without the help of a computer.

EXERCISE 21.6

1 Write out the Kuhn-Tucker conditions for the problem (21.29).

2 Write out the Kuhn-Tucker conditions for the problem (21.30).

3 Let the total revenue R of a sales-maximizing firm depend on output Q and advertising expenditure A, and let its total cost consist of production cost C and advertising cost A.

(a) Reformulate the nonlinear program given in the text.

(b) What restrictions must be placed on the R and C functions to make the Kuhn-Tucker sufficiency theorem applicable?

(c) What restrictions are needed to make the Arrow-Enthoven sufficiency theorem applicable?

4 Minimize $\quad C = x_1^2 + x_2^2$
subject to $\quad x_1 + x_2 \geq 2$
and $\quad x_1, x_2 \geq 0$

(*a*) Is the Kuhn-Tucker sufficiency theorem applicable to this problem? Are the Kuhn-Tucker minimum conditions necessary-and-sufficient?

(*b*) Write out the Kuhn-Tucker conditions and use these to seek out the optimal solution by trial and error. What are the values of $\bar{x}_1$ and $\bar{x}_2$?

5 (*a*) Formulate a nonlinear program that answers the question: What is the shortest distance from the point of origin (0, 0) to any point lying on or above the straight line passing through the points (0, 2) and (2, 0). Denote distance by d. (*Hint:* Use Pythagoras' theorem.)

(*b*) How does the present program differ from the one given in the preceding problem? Would it be possible to take the solution $(\bar{x}_1, \bar{x}_2)$ of the latter as the solution of the present problem? Why?

(*c*) What is the shortest distance, $\bar{d}$?

21.7 LIMITATIONS OF MATHEMATICAL PROGRAMMING

Because mathematical programming is capable of handling inequality constraints, it has considerably broadened the purview of our optimization discussion. Apart from its obvious applicability to practical problems of industrial and business management, it also enables economists to see the theory of consumption, production and resource allocation in a new light.

However, like other types of methods we have discussed, mathematical programming is not without limitations of its own. For one thing, in the above discussion we have always assumed the choice variables to be continuous. In actuality, though, one or more of the variables may admit of integer values only —an optimal output of 3.75 airplanes, e.g., would not quite make practical sense. This complication fortunately is well taken care of by an offshoot of mathematical-programming techniques, known as *integer programming*, which yields only integer solution values.

A more serious limitation—and this is a limitation not only of mathematical programming, but of all the optimization frameworks considered in this volume —lies in the *static* nature of the solution. In writing an optimal solution, say, $(\bar{x}_1, \ldots, \bar{x}_n)$, we are expressing the best choice that can be made of each variable x_j under a set of given circumstances, but since each $\bar{x}_j$ represents a single numerical value, it can only pertain either to a single point of time, or to a period of time during which all the circumstances postulated in the problem experience no changes. In either case, the problem and its solution are static. A *dynamic* optimization problem, in contrast, would ask for the complete *optimal time path* of each choice variable in a given period, not just a single optimal value. To solve such a problem, however, we must first acquire a knowledge of the mathematical methods of *calculus of variations*, *optimal control theory*, and *dynamic programming*. Since these topics can hardly be explained with a proper degree of thoroughness and clarity within a limited number of pages, they are best to be relegated to a separate volume.

Thus, we have now brought another part of the book to a close by pointing out the limitations of the techniques and analyses involved. The purpose of this is not, of course, to discredit the mathematical methods that have just been painstakingly expounded; rather it is to caution you not to attribute to them a degree of omnipotence they do not possess. Indeed, it is an essential part of learning always to have a clear awareness of the limitations of the analytical methods you study, because without it you could become a slave to the techniques rather than their master!