Part 4

Optimization Problems

Chapter 9

Optimization: A Special Variety of Equilibrium Analysis

When we first introduced the term equilibrium in Chap. 3, we made a broad distinction between goal and nongoal equilibrium. In the latter type, exemplified by our study of market and national-income models, the interplay of certain opposing forces in the model—e.g., the forces of demand and supply in the market models and the forces of leakages and injections in the income models—dictates an equilibrium state, if any, in which these opposing forces are just balanced against each other, thus obviating any further tendency to change. The attainment of this type of equilibrium is the outcome of the impersonal balancing of these forces and does not require the conscious effort on the part of anyone to accomplish a specified goal. True, the consuming households behind the forces of demand and the firms behind the forces of supply are each striving for an optimal position under the given circumstances, but as far as the market itself is concerned, no one is aiming at any particular equilibrium price or equilibrium quantity (unless, of course, the government happens to be trying to peg the price). Similarly, in national-income determination, the impersonal balancing of leakages and injections is what brings about an equilibrium state, and no conscious effort at reaching any particular goal (such as an attempt to alter an undesirable income level by means of monetary or fiscal policies) needs to be involved at all.

In the present part of the book, however, our attention will be turned to the study of *goal equilibrium*, in which the equilibrium state is defined as the optimum position for a given economic unit (a household, a business firm, or even an entire economy) and in which the said economic unit will be deliberately striving for attainment of that equilibrium. As a result, in this context—but only in this context—our earlier warning that equilibrium does not imply desirability becomes irrelevant and immaterial. In this part of the book, our primary focus will be on the classical techniques for locating optimum positions—those using differential calculus. More modern developments, known as mathematical programming, will be discussed in Chap. 13.

9.1 Optimum Values and Extreme Values

Economics is essentially a science of choice. When an economic project is to be carried out, such as the production of a specified level of output, there are normally a number of alternative ways of accomplishing it. One (or more) of these alternatives will, however, be more desirable than others from the standpoint of some criterion, and it is the essence of the optimization problem to choose, on the basis of that specified criterion, the best alternative available.

The most common criterion of choice among alternatives in economics is the goal of *maximizing* something (such as maximizing a firm's profit, a consumer's utility, or the rate of growth of a firm or of a country's economy) or of *minimizing* something (such as minimizing the cost of producing a given output). Economically, we may categorize such maximization and minimization problems under the general heading of *optimization*, meaning "the quest for the best." From a purely mathematical point of view, however, the terms *maximum* and *minimum* do not carry with them any connotation of optimality. Therefore, the collective term for maximum and minimum, as mathematical concepts, is the more matter-of-fact designation *extremum*, meaning an extreme value.

In formulating an optimization problem, the first order of business is to delineate an *objective function* in which the dependent variable represents the object of maximization or minimization and in which the set of independent variables indicates the objects whose magnitudes the economic unit in question can pick and choose, with a view to optimizing. We shall therefore refer to the independent variables as *choice variables*.† The essence of the optimization process is simply to find the set of values of the choice variables that will lead us to the desired extremum of the objective function.

For example, a business firm may seek to maximize profit π, that is, to maximize the difference between total revenue R and total cost C. Since, within the framework of a given state of technology and a given market demand for the firm's product, R and C are both functions of the output level Q, it follows that π is also expressible as a function of Q:

$$\pi(Q) = R(Q) - C(Q)$$

This equation constitutes the relevant objective function, with π as the object of maximization and Q as the (only) choice variable. The optimization problem is then that of choosing the level of Q that maximizes π. Note that while the *optimal* level of π is by definition its *maximal* level, the optimal level of the choice variable Q is itself not required to be either a maximum or a minimum.

To cast the problem into a more general mold for further discussion (though still confining ourselves to objective functions of one variable only), let us consider the general function

$$y = f(x)$$

and attempt to develop a procedure for finding the level of x that will maximize or minimize the value of y. It will be assumed in our discussion that the function f is continuously differentiable.

† They can also be called *decision variables*, or *policy variables*.

9.2 Relative Maximum and Minimum: First-Derivative Test

Since the objective function $y = f(x)$ is stated in the general form, there is no restriction as to whether it is linear or nonlinear or whether it is monotonic or contains both increasing and decreasing parts. From among the many possible types of function compatible with the objective-function form discussed in Sec. 9.1, we have selected three specific cases to be depicted in Fig. 9.1. Simple as they may be, the graphs in Fig. 9.1 should give us valuable insight into the problem of locating the maximum or minimum value of the function $y = f(x)$.

Relative versus Absolute Extremum

If the objective function is a constant function, as in Fig. 9.1*a*, all values of the choice variable x will result in the same value of y, and the height of each point on the graph of the function (such as A or B or C) may be considered a maximum or, for that matter, a minimum—or, indeed, neither. In this case, there is in effect no significant choice to be made regarding the value of x for the maximization or minimization of y.

In Fig. 9.1*b*, the function is strictly increasing, and there is no finite maximum if the set of nonnegative real numbers is taken to be its domain. However, we may consider the end point D on the left (the y intercept) as representing a minimum; in fact, it is in this case the *absolute* (or *global*) minimum in the range of the function.

The points E and F in Fig. 9.1*c*, on the other hand, are examples of a *relative* (or *local*) extremum, in the sense that each of these points represents an extremum in the immediate neighborhood of the point only. The fact that point F is a relative minimum is, of course, no guarantee that it is also the global minimum of the function, although this may happen to be the case. Similarly, a relative maximum point such as E may or may not be a global maximum. Note also that a function can very well have several relative extrema, some of which may be maxima while others are minima.

In most economic problems that we shall be dealing with, our primary, if not exclusive, concern will be with extreme values other than end-point values, for with most such problems the domain of the objective function is restricted to be the set of nonnegative real numbers, and thus an end point (on the left) will represent the zero level of the choice variable, which is often of no practical interest. Actually, the type of function most frequently encountered in economic analysis is that shown in Fig. 9.1*c*, or some variant thereof that contains only a single bend in the curve. We shall therefore continue our discussion mainly with reference to the search for *relative* extrema such as points E and F. This will, however, by no means foreclose the knowledge of an absolute maximum if we want it, because an absolute maximum must be either a relative maximum or one of the end points of the

FIGURE 9.1

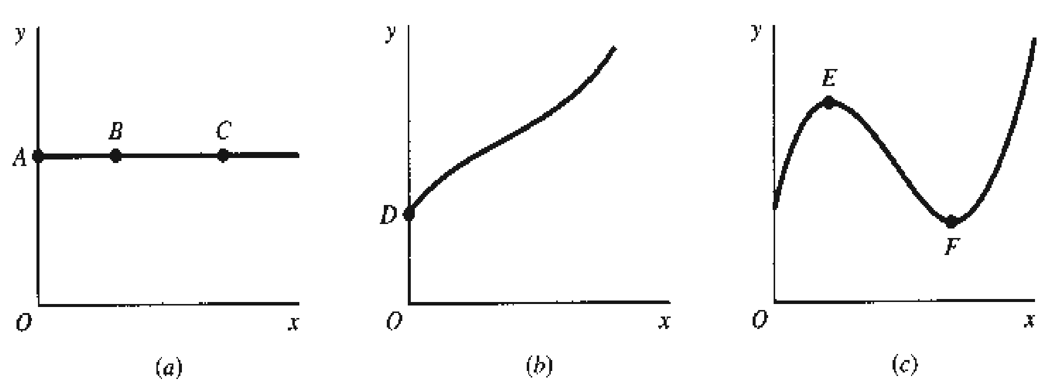
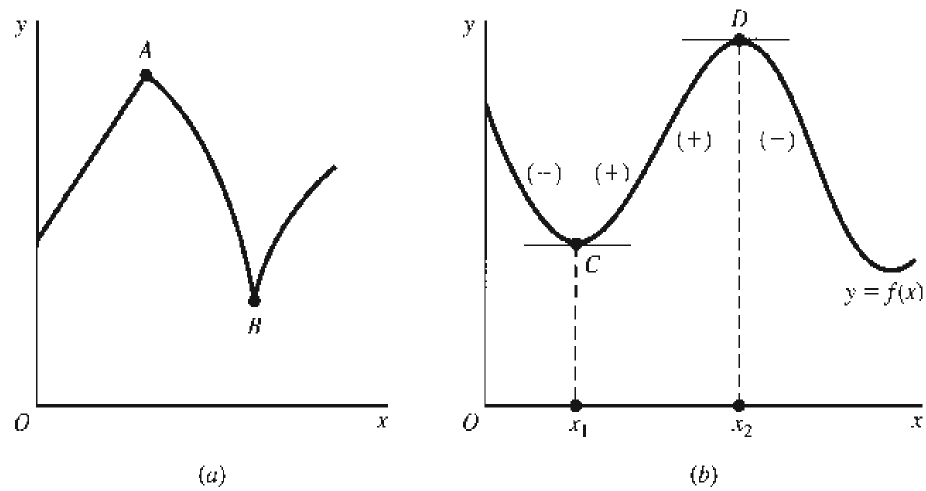


FIGURE 9.2



function. Thus if we know all the relative maxima, it is necessary only to select the largest of these and compare it with the end points in order to determine the absolute maximum. The absolute minimum of a function can be found analogously. Hereafter, the extreme values considered will be *relative* or *local* ones, unless indicated otherwise.

First-Derivative Test

As a matter of terminology, from now on we shall refer to the derivative of a function alternatively as its *first* derivative (short for *first-order* derivative). The reason for this will become apparent shortly.

Given a function $y = f(x)$, the first derivative $f'(x)$ plays a major role in our search for its extreme values. This is due to the fact that, if a relative extremum of the function occurs at $x = x_0$, then either (1) $f'(x_0)$ does not exist, or (2) $f'(x_0) = 0$. The first eventuality is illustrated in Fig. 9.2*a*, where both points A and B depict relative extreme values of y, and yet no derivative is defined at either of these sharp points. Since in the present discussion we are assuming that $y = f(x)$ is continuous and possesses a continuous derivative, however, we are in effect ruling out sharp points. For smooth functions, relative extreme values can occur only where the first derivative has a zero value. This is illustrated by points C and D in Fig. 9.2*b*, both of which represent extreme values, and both of which are characterized by a zero slope—$f'(x_1) = 0$ and $f'(x_2) = 0$. It is also easy to see that when the slope is nonzero we cannot possibly have a relative minimum (the bottom of a valley) or a relative maximum (the peak of a hill). For this reason, we can, in the context of smooth functions, take the condition $f'(x) = 0$ to be a *necessary* condition for a relative extremum (either maximum or minimum).

We must hasten to add, however, that a zero slope, while *necessary*, is *not sufficient* to establish a relative extremum. An example of the case where a zero slope is not associated with an extremum will be presented shortly. By appending a certain proviso to the zero-slope condition, however, we can obtain a decisive test for a relative extremum. This may be stated as follows:

First-derivative test for relative extremum If the first derivative of a function $f(x)$ at $x = x_0$ is $f'(x_0) = 0$, then the value of the function at x_0, $f(x_0)$, will be

a. A relative *maximum* if the derivative $f'(x)$ changes its sign from positive to negative from the immediate left of the point x_0 to its immediate right.

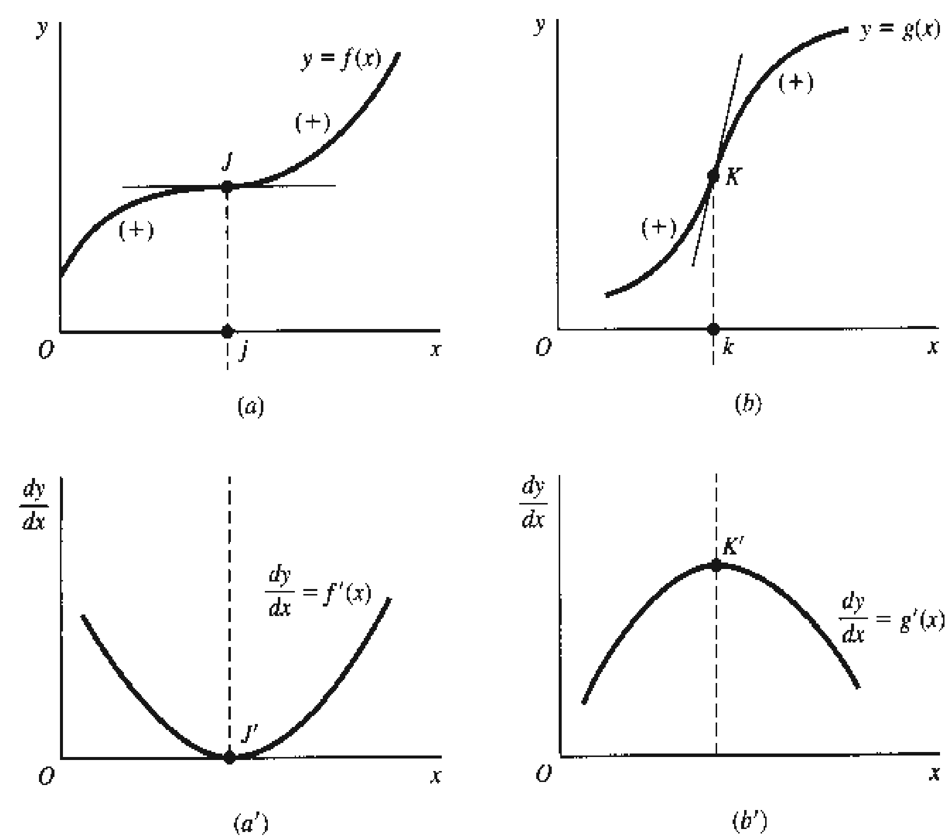
b. A relative *minimum* if $f'(x)$ changes its sign from negative to positive from the immediate left of x_0 to its immediate right.

c. Neither a relative maximum nor a relative minimum if $f'(x)$ has the same sign on both the immediate left and the immediate right of point x_0.

Let us call the value x_0 a *critical value* of x if $f'(x_0) = 0$, and refer to $f(x_0)$ as a *stationary value* of y (or of the function f). The point with coordinates x_0 and $f(x_0)$ can, accordingly, be called a *stationary point*. (The rationale for the word *stationary* should be self-evident—wherever the slope is zero, the point in question is never situated on an upward or downward incline, but is rather at a standstill position.) Then, graphically, the first possibility listed in this test will establish the stationary point as the peak of a hill, such as point D in Fig. 9.2*b*, whereas the second possibility will establish the stationary point as the bottom of a valley, such as point C in the same diagram. Note, however, that in view of the existence of a third possibility, yet to be discussed, we are unable to regard the condition $f'(x) = 0$ as a *sufficient condition* for a relative extremum. But we now see that, *if* the necessary condition $f'(x) = 0$ is satisfied, *then* the change-of-derivative-sign proviso can serve as a *sufficient condition* for a relative maximum or minimum, depending on the direction of the sign change.

Let us now explain the third possibility. In Fig. 9.3*a*, the function f is shown to attain a zero slope at point J (when $x = j$). Even though $f'(j)$ is zero—which makes $f(j)$ a stationary value—the derivative does not change its sign from one side of $x = j$ to the other; therefore, according to the first-derivative test, point J gives neither a maximum nor

FIGURE 9.3



a minimum, as is duly confirmed by the graph of the function. Rather, it exemplifies what is known as an *inflection point.*

The characteristic feature of an inflection point is that, at that point, the derivative (as against the primitive) function reaches an extreme value. Since this extreme value can be either a maximum or a minimum, we have two types of inflection points. In Fig. 9.3a', where we have plotted the derivative $f'(x)$, we see that its value is zero when $x = j$ (see point J') but is positive on both sides of point J'; this makes J' a *minimum* point of the derivative function $f'(x)$.

The other type of inflection point is portrayed in Fig. 9.3b, where the slope of the function $g(x)$ increases till the point k is reached and decreases thereafter. Consequently, the graph of the derivative function $g'(x)$ will assume the shape shown in Fig. 9.3b', where point K' gives a *maximum* value of the derivative function $g'(x)$.[†]

To sum up: A relative extremum must be a stationary value, but a stationary value may be associated with either a relative extremum or an inflection point. To find the relative maximum or minimum of a given function, therefore, the procedure should be first to find the stationary values of the function where the condition $f'(x) = 0$ is satisfied, and then to apply the first-derivative test to determine whether each of the stationary values is a relative maximum, a relative minimum, or neither.

Example 1

Find the relative extrema of the function

$$y = f(x) = x^3 - 12x^2 + 36x + 8$$

First, we find the derivative function to be

$$f'(x) = 3x^2 - 24x + 36$$

To get the critical values, i.e., the values of x satisfying the condition $f'(x) = 0$, we set the quadratic derivative function equal to zero and get the quadratic equation

$$3x^2 - 24x + 36 = 0$$

By factoring the polynomial or by applying the quadratic formula, we then obtain the following pair of roots (solutions):

$$x_1^* = 6 \qquad [\text{at which we have } f'(6) = 0 \text{ and } f(6) = 8]$$
$$x_2^* = 2 \qquad [\text{at which we have } f'(2) = 0 \text{ and } f(2) = 40]$$

Since $f'(6) = f'(2) = 0$, these two values of x are the critical values we desire.

It is easy to verify that, in the immediate neighborhood of $x = 6$, we have $f'(x) < 0$ for $x < 6$, and $f'(x) > 0$ for $x > 6$; thus the value of the function $f(6) = 8$ is a relative minimum. Similarly, since, in the immediate neighborhood of $x = 2$, we find $f'(x) > 0$ for $x < 2$, and $f'(x) < 0$ for $x > 2$, the value of the function $f(2) = 40$ is a relative maximum.

[†] Note that a zero derivative value, while a necessary condition for a relative extremum, is *not* required for an inflection point; for the derivative $g'(x)$ has a positive value at $x = k$, and yet point K is an inflection point.

FIGURE 9.4

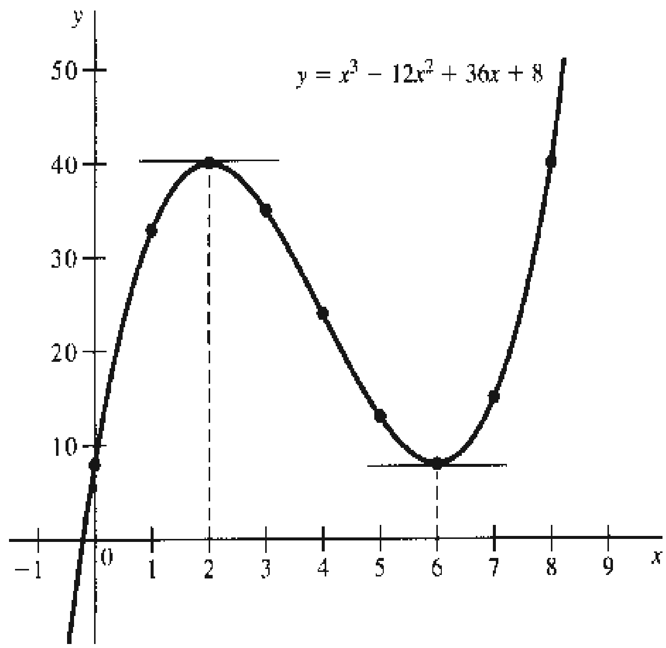


Figure 9.4 shows the graph of the function of this example. Such a graph may be used to verify the location of extreme values obtained through use of the first-derivative test. But, in reality, in most cases "helpfulness" flows in the opposite direction—the mathematically derived extreme values will help in plotting the graph. The accurate plotting of a graph ideally requires knowledge of the value of the function at every point in the domain; but as a matter of actual practice, only a few points in the domain are selected for purposes of plotting, and the rest of the points typically are filled in by interpolation. The pitfall of this practice is that, unless we hit upon the stationary point(s) by coincidence, we shall miss the exact location of the turning point(s) in the curve. Now, with the first-derivative test at our disposal, it becomes possible to locate these turning points precisely.

Example 2 Find the relative extremum of the average-cost function

$$AC = f(Q) = Q^2 - 5Q + 8$$

The derivative here is $f'(Q) = 2Q - 5$, a linear function. Setting $f'(Q)$ equal to zero, we get the linear equation $2Q - 5 = 0$, which has the single root $Q^* = 2.5$. This is the only critical value in this case. To apply the first-derivative test, let us find the values of the derivative at, say, $Q = 2.4$ and $Q = 2.6$, respectively. Since $f'(2.4) = -0.2 < 0$ whereas $f'(2.6) = 0.2 > 0$, we can conclude that the stationary value $AC = f(2.5) = 1.75$ represents a relative minimum. The graph of the function of this example is actually a U-shaped curve, so that the relative minimum already found will also be the absolute minimum. Our knowledge of the exact location of this point should be of great help in plotting the AC curve.

EXERCISE 9.2

1. Find the stationary values of the following (check whether they are relative maxima or minima or inflection points), assuming the domain to be the set of all real numbers:
   (a) $y = -2x^2 + 8x + 7$ (b) $y = 5x^2 + x$ (c) $y = 3x^2 + 3$ (d) $y = 3x^2 - 6x + 2$

2. Find the stationary values of the following (check whether they are relative maxima or minima or inflection points), assuming the domain to be the interval $[0, \infty)$:
   (*a*) $y = x^3 - 3x + 5$
   (*b*) $y = \frac{1}{3}x^3 - x^2 + x + 10$
   (*c*) $y = -x^3 + 4.5x^2 - 6x + 6$
3. Show that the function $y = x + 1/x$ (with $x \neq 0$) has two relative extrema, one a maximum and the other a minimum. Is the "minimum" larger or smaller than the "maximum"? How is this paradoxical result possible?
4. Let $T = \phi(x)$ be a *total* function (e.g., total product or total cost):
   (*a*) Write out the expressions for the *marginal* function M and the *average* function A.
   (*b*) Show that, when A reaches a relative extremum, M and A must have the same value.
   (*c*) What general principle does this suggest for the drawing of a marginal curve and an average curve in the same diagram?
   (*d*) What can you conclude about the elasticity of the total function T at the point where A reaches an extreme value?

9.3 Second and Higher Derivatives

Hitherto we have considered only the first derivative $f'(x)$ of a function $y = f(x)$; now let us introduce the concept of *second derivative* (short for *second-order derivative*), and derivatives of even higher orders. These will enable us to develop alternative criteria for locating the relative extrema of a function.

Derivative of a Derivative

Since the first derivative $f'(x)$ is itself a function of x, it, too, should be differentiable with respect to x, provided that it is continuous and smooth. The result of this differentiation, known as the second derivative of the function f, is denoted by

$f''(x)$ — where the double prime indicates that $f(x)$ has been differentiated with respect to x twice, and where the expression (x) following the double prime suggests that the second derivative is again a function of x

or

$\dfrac{d^2y}{dx^2}$ — where the notation stems from the consideration that the second derivative means, in fact, $\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right)$; hence, the d^2 (read: "d-two") in the numerator and dx^2 (read: "dx squared") in the denominator of this symbol.

If the second derivative $f''(x)$ exists for all x values in the domain, the function $f(x)$ is said to be *twice differentiable;* if, in addition, $f''(x)$ is continuous, the function $f(x)$ is said to be *twice continuously differentiable*. Just as the notation $f \in C^{(1)}$ or $f \in C'$ is often used to indicate that the function f is continuously differentiable, an analogous notation

$$f \in C^{(2)} \qquad \text{or} \qquad f \in C''$$

can be used to signify that f is twice continuously differentiable.