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 Chapter 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

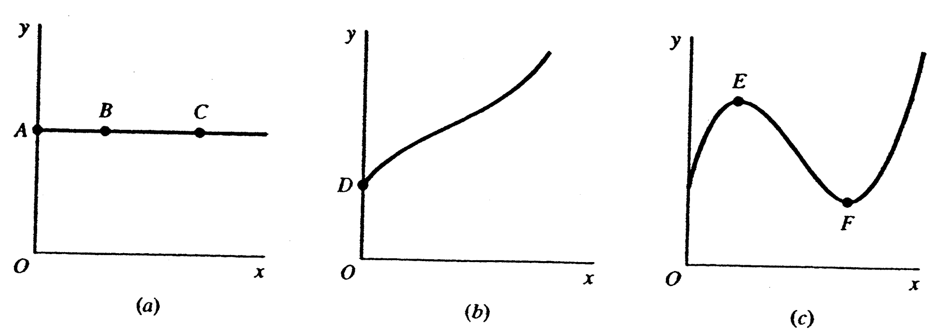
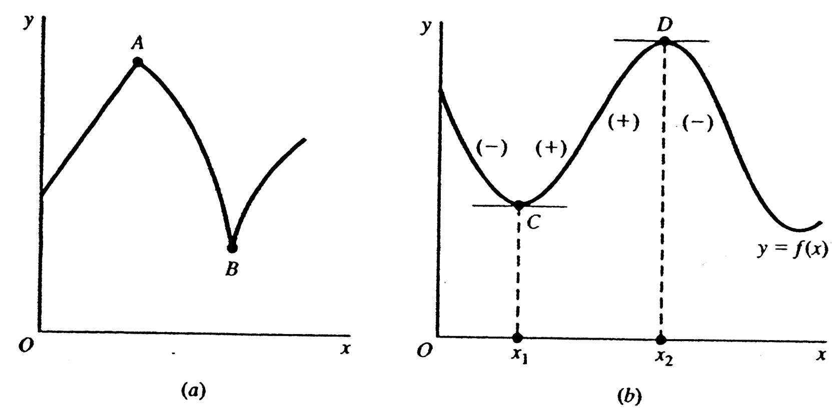


FIGURE 9.2



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.

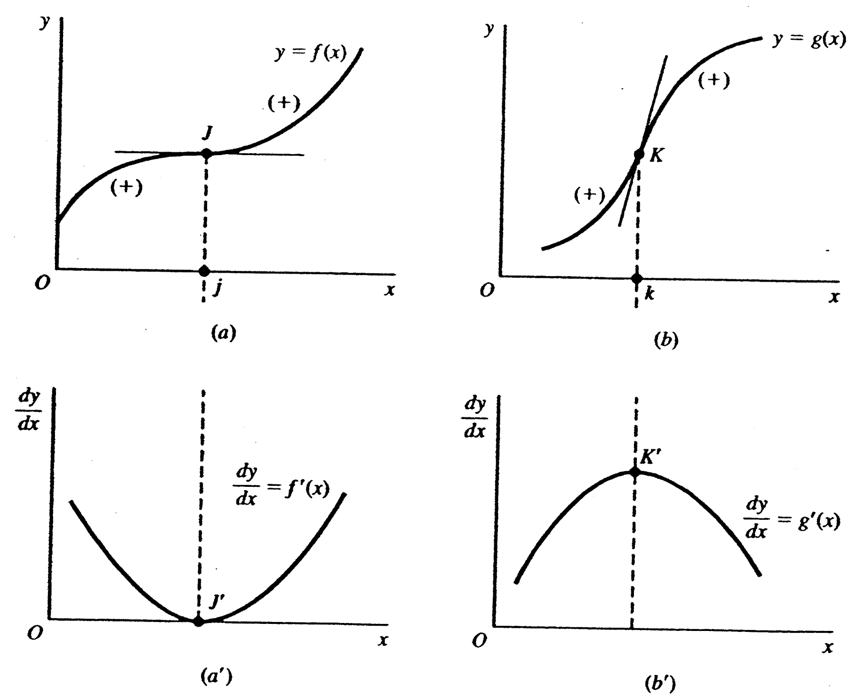
b. A relative *minimum* if $f'(x)$ changes its sign from negative to positive from the immediate left of x_0 to its immediate right.

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

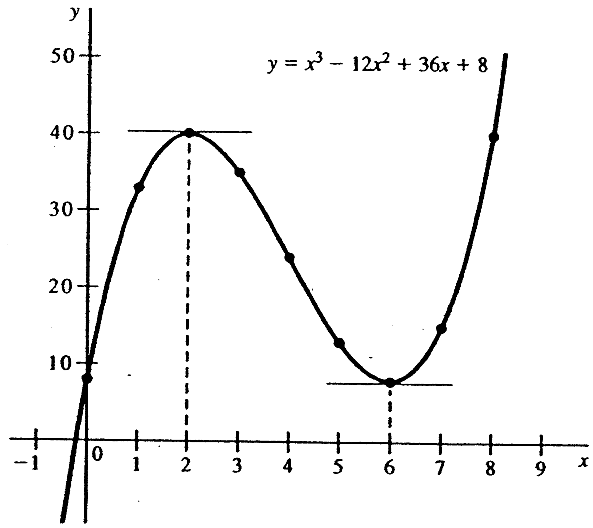


Figure 9.4 shows the graph of the function of this example. Such a graph may be used to verify the location of extreme values obtained through use of the first-derivative test. But, in reality, in most cases "helpfulness" flows in the opposite direction—the mathematically derived extreme values will 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.

As a function of x the second derivative can be differentiated with respect to x again to produce a *third* derivative, which in turn can be the source of a *fourth* derivative, and so on ad infinitum, as long as the differentiability condition is met. These higher-order derivatives are symbolized along the same line as the second derivative:

$$f'''(x), f^{(4)}(x), \ldots, f^{(n)}(x) \qquad \text{[with superscripts enclosed in ( )]}$$

or

$$\frac{d^3y}{dx^3}, \frac{d^4y}{dx^4}, \ldots, \frac{d^ny}{dx^n}$$

The last of these can also be written as $\frac{d^n}{dx^n}y$, where the $\frac{d^n}{dx^n}$ part serves as an operator symbol instructing us to take the nth derivative of (some function) with respect to x.

Almost all the *specific* functions we shall be working with possess continuous derivatives up to any order we desire; i.e., they are continuously differentiable any number of times. Whenever a *general* function is used, such as $f(x)$, we always assume that it has derivatives up to any order we need.

Example 1

Find the first through the fifth derivatives of the function

$$y = f(x) = 4x^4 - x^3 + 17x^2 + 3x - 1$$

The desired derivatives are as follows:

$$\begin{aligned} f'(x) &= 16x^3 - 3x^2 + 34x + 3 \\ f''(x) &= 48x^2 - 6x + 34 \\ f'''(x) &= 96x - 6 \\ f^{(4)}(x) &= 96 \\ f^{(5)}(x) &= 0 \end{aligned}$$

In this particular (polynomial) example, we note that each successive derivative function emerges as a lower-order polynomial—from cubic to quadratic, to linear, to constant. We note also that the fifth derivative, being the derivative of a constant, is equal to zero for all values of x; we could therefore have written it as $f^{(5)}(x) \equiv 0$ as well. The equation $f^{(5)}(x) = 0$ should be carefully distinguished from the equation $f^{(5)}(x_0) = 0$ (zero at x_0 only). Also, understand that the statement $f^{(5)}(x) \equiv 0$ does not mean that the fifth derivative does not exist; it indeed exists, and has the value zero.

Example 2

Find the first four derivatives of the rational function

$$y = g(x) = \frac{x}{1+x} \qquad (x \neq -1)$$

These derivatives can be found either by use of the quotient rule, or, after rewriting the function as $y = x(1+x)^{-1}$, by the product rule:

$$\left.\begin{aligned} g'(x) &= (1+x)^{-2} \\ g''(x) &= -2(1+x)^{-3} \\ g'''(x) &= 6(1+x)^{-4} \\ g^{(4)}(x) &= -24(1+x)^{-5} \end{aligned}\right\} \qquad (x \neq -1)$$

In this case, repeated derivation evidently does not tend to simplify the subsequent derivative expressions.

Note that, like the primitive function $g(x)$, all the successive derivatives obtained are themselves functions of x. Given specific values of x, however, these derivative functions will then take specific values. When $x = 2$, for instance, the second derivative in Example 2 can be evaluated as

$$g''(2) = -2(3)^{-3} = \frac{-2}{27}$$

and similarly for other values of x. It is of the utmost importance to realize that to evaluate this second derivative $g''(x)$ at $x = 2$, as we did, we must first obtain $g''(x)$ from $g'(x)$ and then substitute $x = 2$ into the equation for $g''(x)$. It is *incorrect* to substitute $x = 2$ into $g(x)$ or $g'(x)$ *prior* to the differentiation process leading to $g''(x)$.

Interpretation of the Second Derivative

The derivative function $f'(x)$ measures the rate of change of the function f. By the same token, the second-derivative function f'' is the measure of the rate of change of the first derivative f'; in other words, the second derivative measures the *rate of change* of the *rate of change* of the original function f. To put it differently, with a given infinitesimal increase in the independent variable x from a point $x = x_0$,

$$\left.\begin{matrix} f'(x_0) > 0 \\ f'(x_0) < 0 \end{matrix}\right\} \text{means that the } \textit{value of the function} \text{ tends to } \left\{\begin{matrix} \text{increase} \\ \text{decrease} \end{matrix}\right.$$

whereas, with regard to the second derivative,

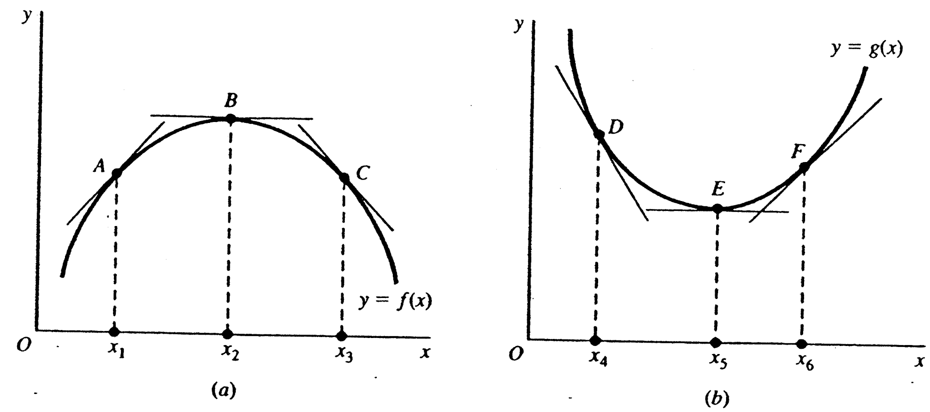
$$\left.\begin{matrix} f''(x_0) > 0 \\ f''(x_0) < 0 \end{matrix}\right\} \text{means that the } \textit{slope of the curve} \text{ tends to } \left\{\begin{matrix} \text{increase} \\ \text{decrease} \end{matrix}\right.$$

Thus a positive first derivative coupled with a positive second derivative at $x = x_0$ implies that the slope of the curve at that point is *positive and increasing*. In other words, the value of the function is increasing at an increasing rate. Likewise, a positive first derivative with a negative second derivative indicates that the slope of the curve is *positive but decreasing*—the value of the function is increasing at a decreasing rate. The case of a negative first derivative can be interpreted analogously, but a warning is in order in this case: When $f'(x_0) < 0$ and $f''(x_0) > 0$, the slope of the curve is *negative and increasing*, but this does *not* mean that the slope is changing, say, from (-10) to (-11); on the contrary, the change should be from (-11), a smaller number, to (-10), a larger number. In other words, the negative slope must tend to be *less* steep as x increases. Lastly, when $f'(x_0) < 0$ and $f''(x_0) < 0$, the slope of the curve must be *negative and decreasing*. This refers to a negative slope that tends to become *steeper* as x increases.

All of this can be further clarified with a graphical explanation. Figure 9.5*a* illustrates a function with $f''(x) < 0$ throughout. Since the slope must steadily decrease as x increases on the graph, we will, when we move from left to right, pass through a point A with a positive slope, then a point B with zero slope, and then a point C with a negative slope. It may happen, of course, that a function with $f''(x) < 0$ is characterized by $f'(x) > 0$ everywhere, and thus plots only as the rising portion of an inverse U-shaped curve, or, with $f'(x) < 0$ everywhere, plots only as the declining portion of that curve.

The opposite case of a function with $f''(x) > 0$ throughout is illustrated in Fig. 9.5*b*. Here, as we pass through points D to E to F, the slope steadily increases and changes from

FIGURE 9.5



negative to zero to positive. Again, we add that a function characterized by $f''(x) > 0$ throughout may, depending on the first-derivative specification, plot only as the declining or the rising portion of a U-shaped curve.

From Fig. 9.5, it is evident that the second derivative $f''(x)$ relates to the *curvature* of a graph; it determines how the curve tends to bend itself. To describe the two types of differing curvatures discussed, we refer to the one in Fig. 9.5*a* as *strictly concave,* and the one in Fig. 9.5*b* as *strictly convex*. And, understandably, a function whose graph is strictly concave (strictly convex) is called a *strictly concave* (*strictly convex*) *function*. The precise geometric characterization of a strictly concave function is as follows. If we pick any pair of points *M* and *N* on its curve and join them by a straight line, the line segment *MN* must lie entirely *below* the curve, except at points *M* and *N*. The characterization of a strictly convex function can be obtained by substituting the word *above* for the word *below* in the last statement. Try this out in Fig. 9.5. If the characterizing condition is relaxed somewhat, so that the line segment *MN* is allowed to lie *either* below the curve, *or* along (coinciding with) the curve, then we will be describing instead a *concave function,* without the adverb *strictly*. Similarly, if the line segment *MN either* lies above, *or* lies along the curve, then the function is *convex,* again without the adverb *strictly*. Note that, since the line segment *MN* may coincide with a (nonstrictly) concave or convex curve, the latter may very well contain a linear segment. In contrast, a *strictly* concave or convex curve can never contain a linear segment anywhere. It follows that while a strictly concave (convex) function is automatically a concave (convex) function, the converse is not true.†

From our earlier discussion of the second derivative, we may now infer that if the second derivative $f''(x)$ is negative for *all* x, then the primitive function $f(x)$ must be a strictly concave function. Similarly, $f(x)$ must be strictly convex, if $f''(x)$ is positive for *all* x. Despite this, it is *not* valid to reverse this inference and say that, if $f(x)$ is strictly concave (strictly convex), then $f''(x)$ must be negative (positive) for all x. This is because, in certain exceptional cases, the second derivative may have a *zero* value at a stationary point on such a curve. An example of this can be found in the function $y = f(x) = x^4$, which plots as a strictly convex curve, but whose derivatives

$$f'(x) = 4x^3 \qquad f''(x) = 12x^2$$

† We shall discuss these concepts further in Sec. 11.5.

indicate that, at the stationary point where $x = 0$, the value of the second derivative is $f''(0) = 0$. Note, however, that at any other point, with $x \neq 0$, the second derivative of this function does have the (expected) positive sign. Aside from the possibility of a zero value at a stationary point, therefore, the second derivative of a strictly concave or convex function may be expected in general to adhere to a single algebraic sign.

For other types of function, the second derivative may take both positive and negative values, depending on the value of x. In Fig. 9.3a and b, for instance, both $f(x)$ and $g(x)$ undergo a sign change in the second derivative at their respective inflection points J and K. According to Fig. 9.3a', the slope of $f'(x)$—that is, the value of $f''(x)$—changes from negative to positive at $x = j$; the exact opposite occurs with the slope of $g'(x)$—that is, the value of $g''(x)$—on the basis of Fig. 9.3b'. Translated into curvature terms, this means that the graph of $f(x)$ turns from strictly concave to strictly convex at point J, whereas the graph of $g(x)$ has the reverse change at point K. Consequently, instead of characterizing an inflection point as a point where the first derivative reaches an extreme value, we may alternatively characterize it as a point where the function undergoes a change in curvature or a change in the sign of its second derivative.

An Application

The two curves in Fig. 9.5 exemplify the graphs of quadratic functions, which may be expressed generally in the form

$$y = ax^2 + bx + c \qquad (a \neq 0)$$

From our discussion of the second derivative, we can now derive a convenient way of determining whether a given quadratic function will have a strictly convex (U-shaped) or a strictly concave (inverse U-shaped) graph.

Since the second derivative of the quadratic function cited is $d^2y/dx^2 = 2a$, this derivative will always have the same algebraic sign as the coefficient a. Recalling that a positive second derivative implies a strictly convex curve, we can infer that a positive coefficient a in the preceding quadratic function gives rise to a U-shaped graph. In contrast, a negative coefficient a leads to a strictly concave curve, shaped like an inverted U.

As intimated at the end of Sec. 9.2, the relative extremum of this function will also prove to be its absolute extremum, because in a quadratic function there can be found only a single valley or peak, evident in a U or inverted U, respectively.

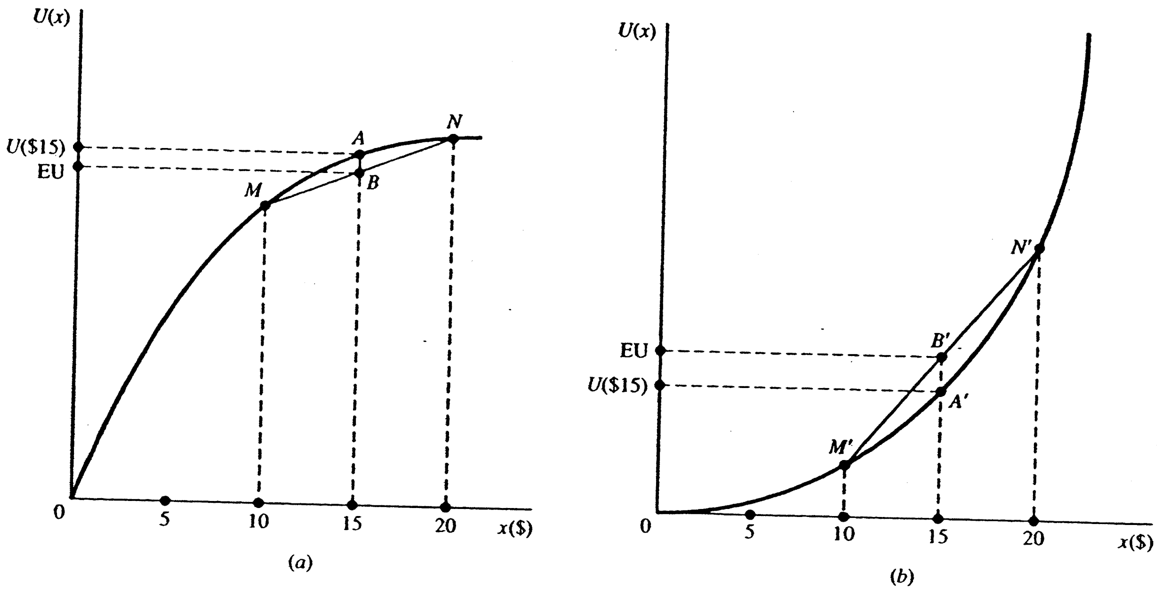
Attitudes toward Risk

The most common application of the concept of marginal utility is to the context of goods consumption. But in another useful application, we consider the marginal utility of *income*, or more to the point of the present discussion, the *payoff* to a betting game, and use this concept to distinguish between different individuals' attitudes toward risk.

Consider the game where, for a fixed sum of money paid in advance (the *cost* of the game), you can throw a die and collect \$10 if an odd number shows up, or \$20 if the number is even. In view of the equal probability of the two outcomes, the mathematically *expected value of payoff* is

$$\text{EV} = 0.5 \times \$10 + 0.5 \times \$20 = \$15$$

FIGURE 9.6



The game is deemed a *fair game,* or *fair bet,* if the cost of the game is exactly \$15. Despite its fairness, playing such a game still involves a risk, for even though the probability distribution of the two possible outcomes is known, the actual result of any individual play is not. Hence, people who are "risk-averse" would consistently decline to play such a game. On the other hand, there are "risk-loving" or "risk-preferring" people who would welcome fair games, or even games with odds set against them (i.e., with the cost of the game exceeding the expected value of payoff).

The explanation for such diverse attitudes toward risk is easily found in the differing utility functions people possess. Assume that a potential player has the strictly concave utility function $U = U(x)$ depicted in Fig. 9.6*a*, where x denotes the payoff, with $U(0) = 0$, $U'(x) > 0$ (positive marginal utility of income or payoff), and $U''(x) < 0$ (diminishing marginal utility) for all x. The economic decision facing this person involves the choice between two courses of action: First, by not playing the game, the person saves the \$15 cost of the game (= EV) and thus enjoys the utility level $U(\$15)$, measured by the height of point A on the curve. Second, by playing, the person has a .5 probability of receiving \$10 and thus enjoying $U(\$10)$ (see point M), plus a .5 probability of receiving \$20 and thus enjoying $U(\$20)$ (see point N). The *expected utility from playing* is, therefore, equal to

$$\text{EU} = 0.5 \times U(\$10) + 0.5 \times U(\$20)$$

which, being the average of the height of M and that of N, is measured by the height of point B, the midpoint on the line segment MN. Since, by the defining property of a strictly concave utility function, line segment MN must lie below arc MN, point B must be lower than point A; that is, EU, the expected utility from playing, falls short of the utility of the cost of the game, and the game should be avoided. For this reason, a strictly concave utility function is associated with risk-averse behavior.

For a risk-loving person, the decision process is analogous, but the opposite choice will be made, because now the relevant utility function is a strictly convex one. In Fig. 9.6*b*,

$U(\$15)$, the utility of keeping the \$15 by not playing the game, is shown by point A' on the curve, and EU, the expected utility from playing, is given by B', the midpoint on the line segment $M'N'$. But this time line segment $M'N'$ lies above arc $M'N'$, and point B' is above point A'. Thus there definitely is a positive incentive to play the game. In contrast to the situation in Fig. 9.6*a*, we can thus associate a strictly convex utility function with risk-loving behavior.

EXERCISE 9.3

1. Find the second and third derivatives of the following functions:
   (*a*) $ax^2 + bx + c$ (*c*) $\dfrac{3x}{1-x}$ $(x \neq 1)$
   (*b*) $7x^4 - 3x - 4$ (*d*) $\dfrac{1+x}{1-x}$ $(x \neq 1)$
2. Which of the following quadratic functions are strictly convex?
   (*a*) $y = 9x^2 - 4x + 8$ (*c*) $u = 9 - 2x^2$
   (*b*) $w = -3x^2 + 39$ (*d*) $v = 8 - 5x + x^2$
3. Draw (*a*) a concave curve which is *not* strictly concave, and (*b*) a curve which qualifies simultaneously as a concave curve and a convex curve.
4. Given the function $y = a - \dfrac{b}{c+x}$ $(a, b, c > 0;\ x \geq 0)$, determine the general shape of its graph by examining (*a*) its first and second derivatives, (*b*) its vertical intercept, and (*c*) the limit of y as x tends to infinity. If this function is to be used as a consumption function, how should the parameters be restricted in order to make it economically sensible?
5. Draw the graph of a function $f(x)$ such that $f'(x) \equiv 0$, and the graph of a function $g(x)$ such that $g'(3) = 0$. Summarize in one sentence the essential difference between $f(x)$ and $g(x)$ in terms of the concept of stationary point.
6. A person who is neither risk-averse nor risk-loving (indifferent toward a fair game) is said to be "risk-neutral."
   (*a*) What kind of utility function would you use to characterize such a person?
   (*b*) Using the die-throwing game detailed in the text, describe the relationship between $U(\$15)$ and EU for the risk-neutral person.

9.4 Second-Derivative Test

Returning to the pair of extreme points B and E in Fig. 9.5 and remembering the newly established relationship between the second derivative and the curvature of a curve, we should be able to see the validity of the following criterion for a relative extremum:

Second-derivative test for relative extremum If the value of the first derivative of a function f 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 second-derivative value at x_0 is $f''(x_0) < 0$.
b. A relative *minimum* if the second-derivative value at x_0 is $f''(x_0) > 0$.

This test is in general more convenient to use than the first-derivative test, because it does not require us to check the derivative sign to both the left and the right of x_0. But it has the drawback that no unequivocal conclusion can be drawn in the event that $f''(x_0) = 0$. For then the stationary value $f(x_0)$ can be *either* a relative maximum, *or* a relative minimum, *or* even an inflectional value.† When the situation of $f''(x_0) = 0$ is encountered, we must either revert to the first-derivative test, or resort to another test, to be developed in Sec. 9.6, that involves the third or even higher derivatives. For most problems in economics, however, the second-derivative test would usually be adequate for determining a relative maximum or minimum.

Example 1

Find the relative extremum of the function

$$y = f(x) = 4x^2 - x$$

The first and second derivatives are

$$f'(x) = 8x - 1 \qquad \text{and} \qquad f''(x) = 8$$

Setting $f'(x)$ equal to zero and solving the resulting equation, we find the (only) critical value to be $x^* = \frac{1}{8}$, which yields the (only) stationary value $f\left(\frac{1}{8}\right) = -\frac{1}{16}$. Because the second derivative is positive (in this case it is indeed positive for any value of x), the extremum is established as a minimum. Further, since the given function plots as a U-shaped curve, the relative minimum is also the absolute minimum.

Example 2

Find the relative extrema of the function

$$y = g(x) = x^3 - 3x^2 + 2$$

The first two derivatives of this function are

$$g'(x) = 3x^2 - 6x \qquad \text{and} \qquad g''(x) = 6x - 6$$

Setting $g'(x)$ equal to zero and solving the resulting quadratic equation, $3x^2 - 6x = 0$, we obtain the critical values $x_1^* = 2$ and $x_2^* = 0$, which in turn yield the two stationary values:

$$g(2) = -2 \qquad [\text{a minimum because } g''(2) = 6 > 0]$$
$$g(0) = 2 \qquad [\text{a maximum because } g''(0) = -6 < 0]$$

Necessary versus Sufficient Conditions

As was the case with the first-derivative test, the zero-slope condition $f'(x) = 0$ plays the role of a *necessary* condition in the second-derivative test. Since this condition is based on the first-order derivative, it is often referred to as the *first-order condition*. Once we find the first-order condition satisfied at $x = x_0$, the negative (positive) sign of $f''(x_0)$ is *sufficient* to establish the stationary value in question as a relative maximum (minimum). These sufficient conditions, which are based on the second-order derivative, are often referred to as *second-order conditions*.

† To see that an inflection point is possible when $f''(x_0) = 0$, let us refer back to Fig. 9.3*a* and 9.3*a*′. Point J in the upper diagram is an inflection point, with $x = j$ as its critical value. Since the $f'(x)$ curve in the lower diagram attains a minimum at $x = j$, the slope of $f'(x)$ [i.e., $f''(x)$] must be zero at the critical value $x = j$. Thus point J illustrates an inflection point occurring when $f''(x_0) = 0$.

To see that a relative extremum is also consistent with $f''(x_0) = 0$, consider the function $y = x^4$. This function plots as a U-shaped curve and has a minimum, $y = 0$, attained at the critical value $x = 0$. Since the second derivative of this function is $f''(x) = 12x^2$, we again obtain a zero value for this derivative at the critical value $x = 0$. Thus this function illustrates a relative extremum occurring when $f''(x_0) = 0$.

TABLE 9.1 Conditions for a Relative Extremum: $y = f(x)$

Condition	Maximum	Minimum
First-order necessary	$f'(x) = 0$	$f'(x) = 0$
Second-order necessary†	$f''(x) \leq 0$	$f''(x) \geq 0$
Second-order sufficient†	$f''(x) < 0$	$f''(x) > 0$

†Applicable only after the first-order necessary condition has been satisfied.

It bears repeating that the first-order condition is *necessary,* but *not sufficient,* for a relative maximum or minimum. (Remember inflection points?) In sharp contrast, the second-order condition that $f''(x)$ be negative (positive) at the critical value x_0 is *sufficient* for a relative maximum (minimum), but it is *not necessary.* [Remember the relative extremum that occurs when $f''(x_0) = 0$?] For this reason, one should carefully guard against the following line of argument: "Since the stationary value $f(x_0)$ is already known to be a minimum, we must have $f''(x_0) > 0$." The reasoning here is faulty because it incorrectly treats the positive sign of $f''(x_0)$ as a necessary condition for $f(x_0)$ to be a minimum.

This is not to say that second-order derivatives can never be used in stating *necessary* conditions for relative extrema. Indeed they can. But care must then be taken to allow for the fact that a relative maximum (minimum) can occur not only when $f''(x_0)$ is negative (positive), but also when $f''(x_0)$ is zero. Consequently, *second-order necessary conditions* must be couched in terms of weak inequalities: for a stationary value $f(x_0)$ to be a relative $\begin{Bmatrix} \text{maximum} \\ \text{minimum} \end{Bmatrix}$, it is necessary that $f''(x_0) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} 0$.

The preceding discussion can be summed up in Table 9.1. All the equations and inequalities in the table are in the nature of conditions (requirements) to be met, rather than descriptive specifications of a given function. In particular, the equation $f'(x) = 0$ does not signify that function f has a zero slope everywhere; rather, it states the stipulation that only those values of x that satisfy this requirement can qualify as critical values.

Conditions for Profit Maximization

We shall now present an economic example of extreme-value problems, i.e., problems of optimization.

One of the first things that a student of economics learns is that, in order to maximize profit, a firm must equate marginal cost and marginal revenue. Let us show the mathematical derivation of this condition. To keep the analysis on a general level, we shall work with the total-revenue function $R = R(Q)$ and total-cost function $C = C(Q)$, both of which are functions of a single variable Q. From these it follows that a profit function (the objective function) may also be formulated in terms of Q (the choice variable):

$$\pi = \pi(Q) = R(Q) - C(Q) \tag{9.1}$$

To find the profit-maximizing output level, we must satisfy the first-order necessary condition for a maximum: $d\pi/dQ = 0$. Accordingly, let us differentiate (9.1) with respect to Q and set the resulting derivative equal to zero: The result is

$$\begin{aligned} \frac{d\pi}{dQ} &\equiv \pi'(Q) = R'(Q) - C'(Q) \\ &= 0 \quad \text{iff} \quad R'(Q) = C'(Q) \end{aligned} \tag{9.2}$$

Thus the *optimum* output (*equilibrium* output) Q^* must satisfy the equation $R'(Q^*) = C'(Q^*)$, or MR = MC. This condition constitutes the first-order condition for profit maximization.

However, the first-order condition may lead to a minimum rather than a maximum; thus we must check the second-order condition next. We can obtain the second derivative by differentiating the first derivative in (9.2) with respect to Q:

$$\frac{d^2\pi}{dQ^2} \equiv \pi''(Q) = R''(Q) - C''(Q)$$
$$\leq 0 \qquad \text{iff} \qquad R''(Q) \leq C''(Q)$$

This last inequality is the second-order necessary condition for maximization. If it is not met, then Q^* cannot possibly maximize profit; in fact, it minimizes profit. If $R''(Q^*) = C''(Q^*)$, then we are unable to reach a definite conclusion. The best scenario is to find $R''(Q^*) < C''(Q^*)$, which satisfies the second-order sufficient condition for a maximum. In that case, we can conclusively take Q^* to be a profit-maximizing output. Economically, this would mean that, if the rate of change of MR is less than the rate of change of MC at the output where MC = MR, then that output will maximize profit.

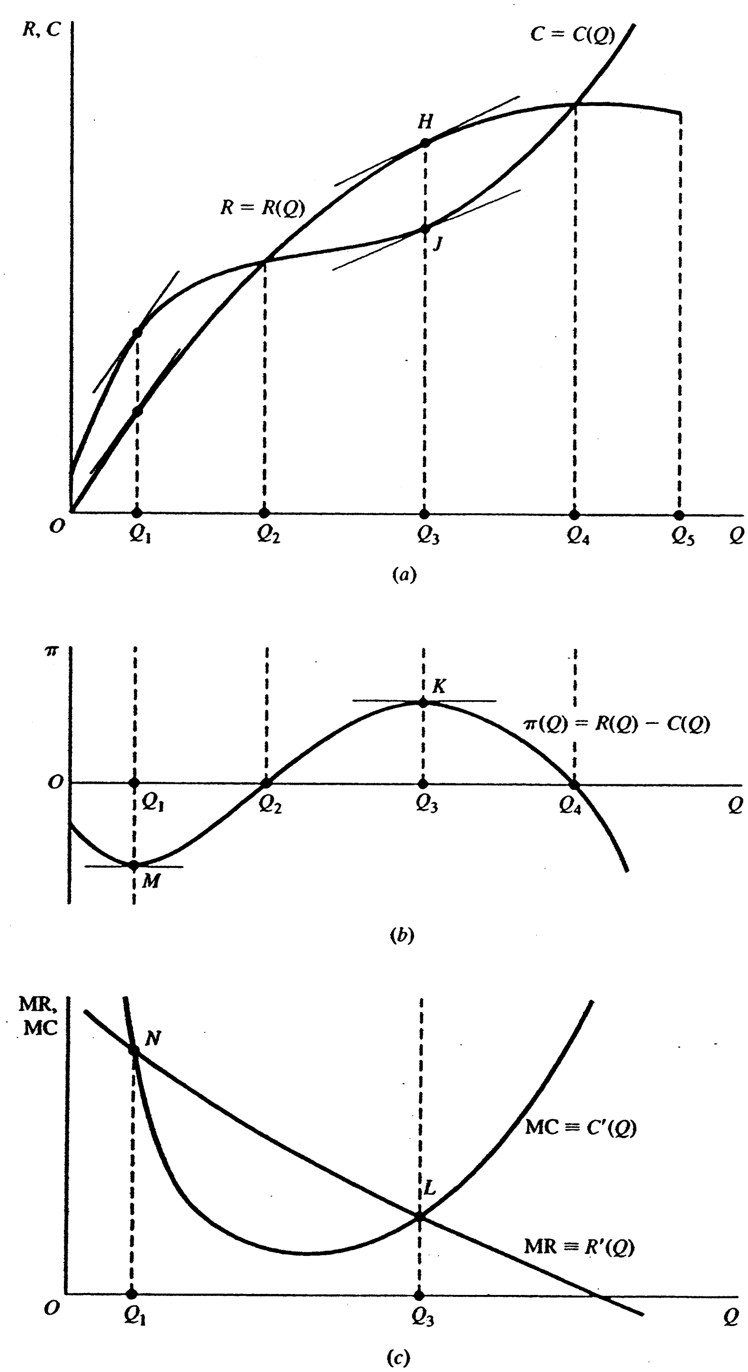
These conditions are illustrated in Fig. 9.7. In Fig. 9.7*a* we have drawn a total-revenue and a total-cost curve, which are seen to intersect twice, at output levels of Q_2 and Q_4. In the open interval (Q_2, Q_4), total revenue R exceeds total cost C, and thus π is positive. But in the intervals $[0, Q_2)$ and $(Q_4, Q_5]$, where Q_5 represents the upper limit of the firm's productive capacity, π is negative. This fact is reflected in Fig. 9.7*b*, where the profit curve—obtained by plotting the vertical distance between the R and C curves for each level of output—lies above the horizontal axis only in the interval (Q_2, Q_4).

When we set $d\pi/dQ = 0$, in line with the first-order condition, it is our intention to locate the peak point K on the profit curve, at output Q_3, where the slope of the curve is zero. However, the relative-minimum point M (output Q_1) will also offer itself as a candidate, because it, too, meets the zero-slope requirement. Below, we shall resort to the second-order condition to eliminate the "wrong" kind of extremum.

The first-order condition $d\pi/dQ = 0$ is equivalent to the condition $R'(Q) = C'(Q)$. In Fig. 9.7*a*, the output level Q_3 satisfies this, because the R and C curves do have the same slope at Q_3 (the tangent lines drawn to the two curves at H and J are parallel to each other). The same is true for output Q_1. Since the equality of the slopes of R and C means the equality of MR and MC, outputs Q_3 and Q_1 must obviously be where the MR and MC curves intersect, as illustrated in Fig. 9.7*c*.

How does the second-order condition enter into the picture? Let us first look at Fig. 9.7*b*. At point K, the second derivative of the π function will (barring the exceptional zero-value case) have a negative value, $\pi''(Q_3) < 0$, because the curve is inverse U-shaped around K; this means that Q_3 will maximize profit. At point M, on the other hand, we would expect that $\pi''(Q_1) > 0$; thus Q_1 provides a relative minimum for π instead. The second-order sufficient condition for a maximum can, of course, be stated alternatively as $R''(Q) < C''(Q)$, that is, that the slope of the MR curve be less than the slope of the MC curve. From Fig. 9.7*c*, it is immediately apparent that output Q_3 satisfies this condition, since the slope of MR is negative while that of MC is positive at point L. But output Q_1 violates this condition because both MC and MR have negative slopes, and that of MR is *numerically smaller* than that of MC at point N, which implies that $R''(Q_1)$ is *greater* than

FIGURE 9.7



$C''(Q_1)$ instead. In fact, therefore, output Q_1 also violates the second-order *necessary* condition for a relative maximum, but satisfies the second-order *sufficient* condition for a relative minimum.

Example 3 Let the $R(Q)$ and $C(Q)$ functions be

$$R(Q) = 1200Q - 2Q^2$$
$$C(Q) = Q^3 - 61.25Q^2 + 1528.5Q + 2000$$

Then the profit function is

$$\pi(Q) = -Q^3 + 59.25Q^2 - 328.5Q - 2000$$

where R, C, and π are all in dollar units and Q is in units of (say) tons per week. This profit function has two critical values, $Q = 3$ and $Q = 36.5$, because

$$\frac{d\pi}{dQ} = -3Q^2 + 118.5Q - 328.5 = 0 \qquad \text{when } Q = \begin{cases} 3 \\ 36.5 \end{cases}$$

But since the second derivative is

$$\frac{d^2\pi}{dQ^2} = -6Q + 118.5 \qquad \begin{cases} > 0 & \text{when } Q = 3 \\ < 0 & \text{when } Q = 36.5 \end{cases}$$

the profit-maximizing output is $Q^* = 36.5$ (tons per week). (The other output minimizes profit.) By substituting Q^* into the profit function, we can find the maximized profit to be $\pi^* = \pi(36.5) = 16{,}318.44$ (dollars per week).

As an alternative approach to the preceding, we can first find the MR and MC functions and then equate the two, i.e., find their intersection. Since

$$R'(Q) = 1200 - 4Q$$
$$C'(Q) = 3Q^2 - 122.5Q + 1528.5$$

equating the two functions will result in a quadratic equation identical with $d\pi/dQ = 0$ which has yielded the two critical values of Q cited previously.

Coefficients of a Cubic Total-Cost Function

In Example 3, a cubic function is used to represent the total-cost function. The traditional total-cost curve $C = C(Q)$, as illustrated in Fig. 9.7*a*, is supposed to contain two wiggles that form a concave segment (decreasing marginal cost) and a subsequent convex segment (increasing marginal cost). Since the graph of a cubic function always contains exactly two wiggles, as illustrated in Fig. 9.4, it should suit that role well. However, Fig. 9.4 immediately alerts us to a problem: the cubic function can possibly produce a downward-sloping segment in its graph, whereas the total-cost function, to make economic sense, should be upward-sloping everywhere (a larger output always entails a higher total cost). If we wish to use a cubic total-cost function such as

$$C = C(Q) = aQ^3 + bQ^2 + cQ + d \tag{9.3}$$

therefore, it is essential to place appropriate restrictions on the parameters so as to prevent the C curve from ever bending downward.

An equivalent way of stating this requirement is that the MC function should be positive throughout, and this can be ensured only if the *absolute minimum* of the MC function turns out to be positive. Differentiating (9.3) with respect to Q, we obtain the MC function

$$MC = C'(Q) = 3aQ^2 + 2bQ + c \tag{9.4}$$

which, because it is a quadratic, plots as a parabola as in Fig. 9.7*c*. In order for the MC curve to stay positive (above the horizontal axis) everywhere, it is necessary that the parabola be U-shaped (otherwise, with an inverse U, the curve is bound to extend itself into the second quadrant). Hence the coefficient of the Q^2 term in (9.4) has to be positive; i.e., we must impose the restriction $a > 0$. This restriction, however, is by no means sufficient, because the minimum value of a U-shaped MC curve—call it $\text{MC}_{\min}$ (a relative minimum which also happens to be an absolute minimum)—may still occur below the horizontal axis. Thus we must next find $\text{MC}_{\min}$ and ascertain the parameter restrictions that would make it positive.

According to our knowledge of relative extremum, the minimum of MC will occur where

$$\frac{d}{dQ}\text{MC} = 6aQ + 2b = 0$$

The output level that satisfies this first-order condition is

$$Q^* = \frac{-2b}{6a} = \frac{-b}{3a}$$

This minimizes (rather than maximizes) MC because the second derivative $d^2(\text{MC})/dQ^2 = 6a$ is assuredly positive in view of the restriction $a > 0$. The knowledge of Q^* now enables us to calculate $\text{MC}_{\min}$, but we may first infer the sign of coefficient b from it. Inasmuch as negative output levels are ruled out, we see that b can never be positive (given $a > 0$). Moreover, since the law of diminishing returns is assumed to set in at a positive output level (that is, MC is assumed to have an initial declining segment), Q^* should be positive (rather than zero). Consequently, we must impose the restriction $b < 0$.

It is a simple matter now to substitute the MC-minimizing output Q^* into (9.4) to find that

$$\text{MC}_{\min} = 3a\left(\frac{-b}{3a}\right)^2 + 2b\frac{-b}{3a} + c = \frac{3ac - b^2}{3a}$$

Thus, to guarantee the positivity of $\text{MC}_{\min}$, we must impose the restriction† $b^2 < 3ac$. This last restriction, we may add, in effect also implies the restriction $c > 0$. (Why?)

The preceding discussion has involved the three parameters a, b, and c. What about the other parameter, d? The answer is that there is need for a restriction on d also, but that has nothing to do with the problem of keeping the MC positive. If we let $Q = 0$ in (9.3), we find

† This restriction may also be obtained by the method of *completing the square*. The MC function can be successively transformed as follows:

$$\begin{aligned}\text{MC} &= 3aQ^2 + 2bQ + c \\ &= \left(3aQ^2 + 2bQ + \frac{b^2}{3a}\right) - \frac{b^2}{3a} + c \\ &= \left(\sqrt{3a}Q + \sqrt{\frac{b^2}{3a}}\right)^2 + \frac{-b^2 + 3ac}{3a}\end{aligned}$$

Since the squared expression can possibly be zero, we must, in order to ensure the positivity of MC, require that $b^2 < 3ac$ on the knowledge that $a > 0$.

that $C(0) = d$. The role of d is thus to determine the vertical intercept of the C curve only, with no bearing on its slope. Since the economic meaning of d is the fixed cost of a firm, the appropriate restriction (in the short-run context) would be $d > 0$.

In sum, the coefficients of the total-cost function (9.3) should be restricted as follows (assuming the short-run context):

$$a, c, d > 0 \qquad b < 0 \qquad b^2 < 3ac \tag{9.5}$$

As you can readily verify, the $C(Q)$ function in Example 3 does satisfy (9.5).

Upward-Sloping Marginal-Revenue Curve

The marginal-revenue curve in Fig. 9.7c is shown to be downward-sloping throughout. This, of course, is how the MR curve is traditionally drawn for a firm under imperfect competition. However, the possibility of the MR curve being partially, or even wholly, upward-sloping can by no means be ruled out a priori.†

Given an average-revenue function $\text{AR} = f(Q)$, the marginal-revenue function can be expressed by

$$\text{MR} = f(Q) + Qf'(Q) \qquad \text{[from (7.7)]}$$

The slope of the MR curve can thus be ascertained from the derivative

$$\frac{d}{dQ}\text{MR} = f'(Q) + f'(Q) + Qf''(Q) = 2f'(Q) + Qf''(Q)$$

As long as the AR curve is downward-sloping (as it would be under imperfect competition), the $2f'(Q)$ term is assuredly negative. But the $Qf''(Q)$ term can be either negative, zero, or positive, depending on the sign of the second derivative of the AR function, i.e., depending on whether the AR curve is strictly concave, linear, or strictly convex. If the AR curve is strictly convex either in its entirety (as illustrated in Fig. 7.2) or along a specific segment, the possibility will exist that the (positive) $Qf''(Q)$ term may dominate the (negative) $2f'(Q)$ term, thereby causing the MR curve to be wholly or partially upward-sloping.

Example 4

Let the average-revenue function be

$$\text{AR} = f(Q) = 8000 - 23Q + 1.1Q^2 - 0.018Q^3$$

As can be verified (see Exercise 9.4-7), this function gives rise to a downward-sloping AR curve, as is appropriate for a firm under imperfect competition. Since

$$\text{MR} = f(Q) + Qf'(Q) = 8000 - 46Q + 3.3Q^2 - 0.072Q^3$$

it follows that the slope of MR is

$$\frac{d}{dQ}\text{MR} = -46 + 6.6Q - 0.216Q^2$$

Because this is a quadratic function and since the coefficient of Q^2 is negative, $d\text{MR}/dQ$ must plot as an inverse-U-shaped curve against Q, such as shown in Fig. 9.5*a*. If a segment of this curve happens to lie above the horizontal axis, the slope of MR will take positive values.

† This point is emphatically brought out in John P. Formby, Stephen Layson, and W. James Smith, "The Law of Demand, Positive Sloping Marginal Revenue, and Multiple Profit Equilibria," *Economic Inquiry*, April 1982, pp. 303–311.

Setting $dMR/dQ = 0$, and applying the quadratic formula, we find the two zeros of the quadratic function to be $Q_1 = 10.76$ and $Q_2 = 19.79$ (approximately). This means that, for values of Q in the open interval (Q_1, Q_2), the dMR/dQ curve does lie above the horizontal axis. Thus the marginal-revenue curve indeed is positively sloped for output levels between Q_1 and Q_2.

The presence of a positively sloped segment on the MR curve has interesting implications. With more bends in its configuration, such an MR curve may produce more than one intersection with the MC curve satisfying the second-order sufficient condition for profit maximization. While all such intersections constitute local optima, however, only one of them is the global optimum that the firm is seeking.

EXERCISE 9.4

1. Find the relative maxima and minima of y by the second-derivative test:
   (*a*) $y = -2x^2 + 8x + 25$ (*c*) $y = \frac{1}{3}x^3 - 3x^2 + 5x + 3$
   (*b*) $y = x^3 + 6x^2 + 9$ (*d*) $y = \dfrac{2x}{1-2x} \quad \left(x \neq \dfrac{1}{2}\right)$
2. Mr. Greenthumb wishes to mark out a rectangular flower bed, using a wall of his house as one side of the rectangle. The other three sides are to be marked by wire netting, of which he has only 64 ft available. What are the length L and width W of the rectangle that would give him the largest possible planting area? How do you make sure that your answer gives the largest, not the smallest area?
3. A firm has the following total-cost and demand functions:
   $$C = \tfrac{1}{3}Q^3 - 7Q^2 + 111Q + 50$$
   $$Q = 100 - P$$
   (*a*) Does the total-cost function satisfy the coefficient restrictions of (9.5)?
   (*b*) Write out the total-revenue function R in terms of Q.
   (*c*) Formulate the total-profit function π in terms of Q.
   (*d*) Find the profit-maximizing level of output Q^*.
   (*e*) What is the maximum profit?
4. If coefficient b in (9.3) were to take a zero value, what would happen to the marginal-cost and total-cost curves?
5. A quadratic profit function $\pi(Q) = hQ^2 + jQ + k$ is to be used to reflect the following assumptions:
   (*a*) If nothing is produced, the profit will be negative (because of fixed costs).
   (*b*) The profit function is strictly concave.
   (*c*) The maximum profit occurs at a positive output level Q^*.
   What parameter restrictions are called for?
6. A purely competitive firm has a single variable input L (labor), with the wage rate W_0 per period. Its fixed inputs cost the firm a total of F dollars per period. The price of the product is P_0.
   (*a*) Write the production function, revenue function, cost function, and profit function of the firm.

(*b*) What is the first-order condition for profit maximization? Give this condition an economic interpretation.

(*c*) What economic circumstances would ensure that profit is maximized rather than minimized?

7. Use the following procedure to verify that the AR curve in Example 4 is negatively sloped:

(*a*) Denote the slope of AR by S. Write an expression for S.

(*b*) Find the maximum value of S, S_{max}, by using the second-derivative test.

(*c*) Then deduce from the value of S_{max} that the AR curve is negatively sloped throughout.

9.5 Maclaurin and Taylor Series

The time has now come for us to develop a test for relative extrema that can apply even when the second derivative turns out to have a zero value at the stationary point. Before we can do that, however, it is first necessary to discuss the so-called expansion of a function $y = f(x)$ into what are known, respectively, as a *Maclaurin series* (expansion around the point $x = 0$) and a *Taylor series* (expansion around any point $x = x_0$).

To *expand* a function $y = f(x)$ around a point x_0 means, in the present context, to transform that function into a *polynomial* form, in which the coefficients of the various terms are expressed in terms of the derivative values $f'(x_0)$, $f''(x_0)$, etc.—all evaluated at the point of expansion x_0. In the Maclaurin series, these will be evaluated at $x = 0$; thus we have $f'(0)$, $f''(0)$, etc., in the coefficients. The result of expansion is a *power series* because, being a polynomial, it consists of a sum of power functions.

Maclaurin Series of a Polynomial Function

Let us consider first the expansion of a *polynomial* function of the nth degree,

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots + a_nx^n \qquad \textbf{(9.6)}$$

into an equivalent nth-degree polynomial where the coefficients (a_0, a_1, etc.) are expressed instead in terms of the derivative values $f'(0)$, $f''(0)$, etc. Since this involves the transformation of one polynomial into another of the same degree, it may seem a sterile and purposeless exercise, but actually it will serve to shed much light on the whole idea of expansion.

Since the power series after expansion will involve the derivatives of various orders of the function f, let us first find these. By successive differentiation of (9.6), we can get the derivatives as follows:

$$\begin{aligned}
f'(x) &= a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \cdots + na_nx^{n-1} \\
f''(x) &= 2a_2 + 3(2)a_3x + 4(3)a_4x^2 + \cdots + n(n-1)a_nx^{n-2} \\
f'''(x) &= 3(2)a_3 + 4(3)(2)a_4x + \cdots + n(n-1)(n-2)a_nx^{n-3} \\
f^{(4)}(x) &= 4(3)(2)a_4 + 5(4)(3)(2)a_5x + \cdots + n(n-1)(n-2)(n-3)a_nx^{n-4} \\
&\vdots \\
f^{(n)}(x) &= n(n-1)(n-2)(n-3)\cdots(3)(2)(1)a_n
\end{aligned}$$

Note that each successive differentiation reduces the number of terms by one—the additive constant in front drops out—until, in the nth derivative, we are left with a single product term (a constant term). These derivatives can be evaluated at various values of x; here we shall evaluate them at $x = 0$, with the result that all terms involving x will drop out. We are then left with the following exceptionally neat derivative values:

$$f'(0) = a_1 \qquad f''(0) = 2a_2 \qquad f'''(0) = 3(2)a_3 \qquad f^{(4)}(0) = 4(3)(2)a_4$$
$$\cdots \qquad f^{(n)}(0) = n(n-1)(n-2)(n-3)\cdots(3)(2)(1)a_n \tag{9.7}$$

If we now adopt a shorthand symbol $n!$ (read: "n factorial"), defined as

$$n! \equiv n(n-1)(n-2)\cdots(3)(2)(1) \qquad (n = \text{a positive integer})$$

so that, for example, $2! = 2 \times 1 = 2$ and $3! = 3 \times 2 \times 1 = 6$, etc. (with $0!$ defined as equal to 1), then the result in (9.7) can be rewritten as

$$a_1 = \frac{f'(0)}{1!} \qquad a_2 = \frac{f''(0)}{2!} \qquad a_3 = \frac{f'''(0)}{3!} \qquad a_4 = \frac{f^{(4)}(0)}{4!} \qquad \cdots \qquad a_n = \frac{f^{(n)}(0)}{n!}$$

Substituting these into (9.6) and utilizing the obvious fact that $f(0) = a_0$, we can now express the given function $f(x)$ as a new, but equivalent, same-degree polynomial in which the coefficients are expressed in terms of derivatives evaluated at $x = 0$:[†]

$$f(x) = \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \qquad \text{[Maclaurin's formula]} \tag{9.8}$$

This new polynomial, called the Maclaurin series of the polynomial function $f(x)$, represents the expansion of the function $f(x)$ around zero ($x = 0$). Note that the point of expansion (here, 0) is simply the value of x that will be used to evaluate $f(x)$ and all its derivatives.

Example 1

Find the Maclaurin series for the function

$$f(x) = 2 + 4x + 3x^2 \tag{9.9}$$

This function has the derivatives

$$\begin{aligned} f'(x) &= 4 + 6x \\ f''(x) &= 6 \end{aligned} \qquad \text{so that} \begin{cases} f'(0) = 4 \\ f''(0) = 6 \end{cases}$$

Thus the Maclaurin series is

$$\begin{aligned} f(x) &= f(0) + f'(0)x + \frac{f''(0)}{2}x^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

The previous line verifies that the Maclaurin series does indeed correctly represent the given function.

[†] Since $0! = 1$ and $1! = 1$, the first two terms on the right of the equals sign in (9.8) can be written more simply as $f(0)$, and $f'(0)x$, respectively. We have included the denominators $0!$ and $1!$ here to call attention to the symmetry among the various terms in the expansion.

Taylor Series of a Polynomial Function

More generally, the polynomial function in (9.6) can be expanded around any point x_0, not necessarily zero. In the interest of simplicity, we shall explain this by means of the specific quadratic function in (9.9) and generalize the result later.

For the purpose of expansion around a specific point x_0, we may first interpret any given value of x as a *deviation* from x_0. More specifically, we shall let $x = x_0 + \delta$, where δ represents the deviation from the value x_0. Upon such interpretation, the given function (9.9) and its derivatives now become

$$\begin{aligned} f(x) &= 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \\ f'(x) &= 4 + 6(x_0 + \delta) \\ f''(x) &= 6 \end{aligned} \tag{9.10}$$

We know that the expression $(x_0 + \delta) = x$ is a variable in the function, but since x_0 in the present context is a *fixed* (chosen) number, only δ can be properly regarded as a variable in (9.10). Consequently, $f(x)$ is in fact a function of δ, say, $g(\delta)$:

$$g(\delta) = 2 + 4(x_0 + \delta) + 3(x_0 + \delta)^2 \qquad [\equiv f(x)]$$

with derivatives

$$\begin{aligned} g'(\delta) &= 4 + 6(x_0 + \delta) \quad && [\equiv f'(x)] \\ g''(\delta) &= 6 && [\equiv f''(x)] \end{aligned}$$

We already know how to expand $g(\delta)$ around zero ($\delta = 0$). According to (9.8), such an expansion will yield the following Maclaurin series:

$$g(\delta) = \frac{g(0)}{0!} + \frac{g'(0)}{1!}\delta + \frac{g''(0)}{2!}\delta^2 \tag{9.11}$$

But since we have let $x = x_0 + \delta$, the fact that $\delta = 0$ implies $x = x_0$; hence, on the basis of the identity $g(\delta) \equiv f(x)$, we can write for the case of $\delta = 0$:

$$g(0) = f(x_0) \qquad g'(0) = f'(x_0) \qquad g''(0) = f''(x_0)$$

Upon substituting these into (9.11), we find the result to represent the expansion of $f(x)$ around the point x_0, because the coefficients now involve the derivatives $f'(x_0)$, $f''(x_0)$, etc., all evaluated at $x = x_0$:

$$f(x)[=g(\delta)] = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 \tag{9.12}$$

You should compare this result—the Taylor polynomial of $f(x)$—with the Maclaurin polynomial of $g(\delta)$ in (9.11).

Since for the specific function under consideration, (9.9), we have

$$f(x_0) = 2 + 4x_0 + 3x_0^2 \qquad f'(x_0) = 4 + 6x_0 \qquad f''(x_0) = 6$$

the Taylor polynomial in (9.12) becomes

$$\begin{aligned} f(x) &= 2 + 4x_0 + 3x_0^2 + (4 + 6x_0)(x - x_0) + \tfrac{6}{2}(x - x_0)^2 \\ &= 2 + 4x + 3x^2 \end{aligned}$$

This verifies that the Taylor polynomial does correctly represent the given function.

The expansion formula in (9.12) can be generalized to apply to the nth-degree polynomial of (9.6). The generalized formula is

$$f(x) = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \qquad \text{[Taylor's formula]} \qquad (9.13)$$

This differs from Maclaurin's formula in (9.8) only in the replacement of zero by x_0 as the point of expansion, and in the replacement of x by the expression $(x - x_0)$. What (9.13) tells us is that, given an nth-degree polynomial $f(x)$, if we let $x = 7$ (say) in the terms on the right of (9.13), select an arbitrary number x_0, then evaluate and add these terms, we will end up exactly with $f(7)$—the value of $f(x)$ at $x = 7$.

Example 2 Taking $x_0 = 3$ as the point of expansion, we can rewrite (9.6) equivalently as

$$f(x) = f(3) + f'(3)(x - 3) + \frac{f''(3)}{2}(x - 3)^2 + \cdots + \frac{f^{(n)}(3)}{n!}(x - 3)^n$$

Expansion of an Arbitrary Function

Heretofore, we have shown how an nth-degree polynomial function can be expressed in another, equivalent, nth-degree polynomial form. As it turns out, it is also possible to express any *arbitrary* function $\phi(x)$—one that is not necessarily a polynomial—in a polynomial form similar to (9.13), provided $\phi(x)$ has finite, continuous derivatives up to the desired order at the expansion point x_0.

According to a mathematical proposition known as *Taylor's theorem*, given an arbitrary function $\phi(x)$, if we know the value of the function at $x = x_0$ [that is, $\phi(x_0)$] and the values of its derivatives at x_0 [that is, $\phi'(x_0)$, $\phi''(x_0)$, etc.], then this function can be expanded around the point x_0 as follows (n = a fixed positive integer arbitrarily chosen):

$$\begin{aligned}\phi(x) &= \left[\frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{\phi^{(n)}(x_0)}{n!}(x - x_0)^n\right] + R_n \\ &\equiv P_n + R_n \qquad \text{[Taylor's formula with remainder]} \qquad (9.14)\end{aligned}$$

where P_n represents the (bracketed) nth-degree polynomial [the first $(n + 1)$ terms on the right], and R_n denotes a *remainder*, to be explained on page 248.[†] The presence of R_n is what distinguishes (9.14) from Taylor's formula (9.13), and for this reason (9.14) is called *Taylor's formula with remainder*. The form of the polynomial P_n and the size of the remainder R_n will depend on the value of n we choose. The larger the n, the more terms there will be in P_n; accordingly, R_n will in general assume a different value for each different n. This fact explains the need for the subscript n in these two symbols. As a memory aid, we can identify n as the order of the highest derivative in P_n. (In the special case of $n = 0$, no derivative will appear in P_n at all.)

[†] The symbol R_n (remainder) is not to be confused with the symbol R^n (n-space).

The appearance of R_n in (9.14) is due to the fact that we are here dealing with an arbitrary function ϕ which cannot always be transformed *exactly* into, but can only be approximated by, the polynomial form shown in (9.13). Therefore, a remainder term is included as a supplement to the P_n part, to represent the discrepancy between $\phi(x)$ and P_n. Thus, P_n constitutes a polynomial approximation to $\phi(x)$, with the term R_n as a measure of the error of approximation. If we choose $n = 1$, for example, we have

$$\phi(x) = [\phi(x_0) + \phi'(x_0)(x - x_0)] + R_1 = P_1 + R_1$$

where P_1 consists of $n + 1 = 2$ terms and constitutes a *linear* approximation to $\phi(x)$. If we choose $n = 2$, a second-power term will appear, so that

$$\phi(x) = \left[\phi(x_0) + \phi'(x_0)(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2\right] + R_2 = P_2 + R_2$$

where P_2, consisting of $n + 1 = 3$ terms, is a *quadratic* approximation to $\phi(x)$. And so forth. The fact that we can create polynomial approximations to any arbitrary function (provided it has finite, continuous derivatives) is of great practical significance. Polynomial functions—even higher-degree ones—are relatively easy to work with, and if they can serve as good approximations to some difficult functions, much convenience is to be gained, as the next two examples will illustrate.

We should point out that the arbitrary function $\phi(x)$ could obviously encompass the nth-degree polynomial of (9.6) as a special case. For this latter case, if the expansion is into another nth-degree polynomial, the result of (9.13) will exactly apply; or in other words, we can use the result in (9.14), with $R_n \equiv 0$. However, if the given nth-degree polynomial $f(x)$ is to be expanded into a polynomial of a *lesser* degree, then the latter can only be considered an approximation to $f(x)$, and a remainder must appear; in that case, the result in (9.14) can be applied with a nonzero remainder. Thus Taylor's formula in the form of (9.14) is perfectly general.

Example 3

Expand the nonpolynomial function

$$\phi(x) = \frac{1}{1+x}$$

around the point $x_0 = 1$, with $n = 4$. We shall need the first four derivatives of $\phi(x)$, which are

$$\phi'(x) = -(1+x)^{-2} \qquad \text{so that} \qquad \phi'(1) = -(2)^{-2} = \frac{-1}{4}$$

$$\phi''(x) = 2(1+x)^{-3} \qquad \phi''(1) = 2(2)^{-3} = \frac{1}{4}$$

$$\phi'''(x) = -6(1+x)^{-4} \qquad \phi'''(1) = -6(2)^{-4} = \frac{-3}{8}$$

$$\phi^{(4)}(x) = 24(1+x)^{-5} \qquad \phi^{(4)}(1) = 24(2)^{-5} = \frac{3}{4}$$

Also, we see that $\phi(1) = \frac{1}{2}$. Thus, setting $x_0 = 1$ in (9.14) and utilizing the obtained derivatives, we obtain the following Taylor series with remainder:

$$\begin{aligned}\phi(x) &= \frac{1}{2} - \frac{1}{4}(x-1) + \frac{1}{8}(x-1)^2 - \frac{1}{16}(x-1)^3 + \frac{1}{32}(x-1)^4 + R_4 \\ &= \frac{31}{32} - \frac{13}{16}x + \frac{1}{2}x^2 - \frac{3}{16}x^3 + \frac{1}{32}x^4 + R_4\end{aligned}$$

It is possible, of course, to choose $x_0 = 0$ as the point of expansion here, too. In that case, with x_0 set equal to zero in (9.14), the expansion will result in a *Maclaurin series with remainder*.

Example 4 Expand the quadratic function

$$\phi(x) = 5 + 2x + x^2$$

around $x_0 = 1$, with $n = 1$. This function is, like (9.9) in Example 1, a second-degree polynomial. But since $n = 1$, our assigned task is to expand it into a *first*-degree polynomial, i.e., to find a linear approximation to the given quadratic function; thus a remainder term is bound to appear. For this reason, $\phi(x)$ should be viewed as an "arbitrary" function for the purpose of this Taylor expansion.

To carry out this expansion, we need only the first derivative $\phi'(x) = 2 + 2x$. Evaluated at $x_0 = 1$, the given function and its derivative yield

$$\phi(x_0) = \phi(1) = 8 \qquad \phi'(x_0) = \phi'(1) = 4$$

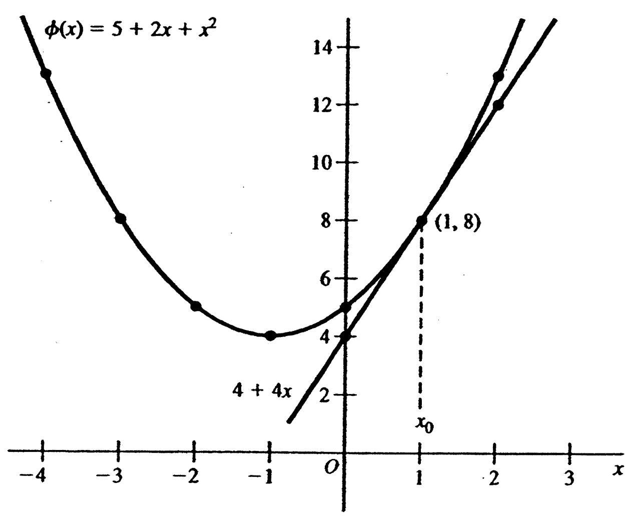
Thus Taylor's formula with remainder gives us

$$\begin{aligned} \phi(x) &= \phi(x_0) + \phi'(x_0)(x - x_0) + R_1 \\ &= 8 + 4(x-1) + R_1 = 4 + 4x + R_1 \end{aligned}$$

where the $(4 + 4x)$ term is a linear approximation and the R_1 term represents the error of approximation.

In Fig. 9.8, $\phi(x)$ plots as a parabola, and its linear approximation as a straight line tangent to the $\phi(x)$ curve at the point (1, 8). The occurrence of the point of tangency at $x = 1$ is not a matter of coincidence; rather, it is the direct consequence of the fact that the point of expansion is set at that particular value of x. This suggests that, when an arbitrary function $\phi(x)$ is approximated by a polynomial, the latter will give the exact value of $\phi(x)$ *at* (and only *at*) the point of expansion, with zero error of approximation ($R_1 = 0$). Elsewhere, R_1 is strictly nonzero and, in fact, shows increasingly larger errors of approximation as we

FIGURE 9.8



try to approximate $\phi(x)$ for x values farther and farther away from the point of expansion x_0. Thus, when attempting to approximate any function $\phi(x)$ by a polynomial, if we are most interested in obtaining an accurate approximation in the neighborhood of a specific value of x, say x_0, then we ought to choose x_0 as the point of expansion.

The construction of Fig. 9.8 is strongly reminiscent of Fig. 8.1. Indeed, both figures are concerned with "approximations." But there is a difference in the scope of approximation. In Fig. 8.1, we attempt to approximate Δy by the differential dy with the help of a tangent line drawn at x_0, a given starting value of x. In Fig. 9.8, on the other hand, we aim more broadly to approximate an entire curve by a particular straight line, i.e., to approximate the height of the curve at any value of x, say, x_1, by the corresponding height of the straight line at x_1. Note that, in both cases, the error of approximation varies with the value of x. In Fig. 8.1, the error (the difference between dy and Δy) gets smaller as Δx gets smaller, or as x gets closer to x_0, at which the tangent line is drawn. In Fig. 9.8, the error (the vertical discrepancy between the straight line and the curve) gets smaller as x approaches x_0, the chosen point of expansion.

Lagrange Form of the Remainder

Now we must comment further on the remainder term. According to the *Lagrange form of the remainder,* we can express R_n as

$$R_n = \frac{\phi^{(n+1)}(p)}{(n+1)!}(x - x_0)^{n+1} \tag{9.15}$$

where p is some number between x (the point where we wish to evaluate the arbitrary function ϕ) and x_0 (the point where we expand the function ϕ). Note that this expression closely resembles the term which should logically follow the last term in P_n in (9.14), except that the derivative involved is here to be evaluated at a point p instead of x_0. Since the point p is, unfortunately, not otherwise specified, this formula does not really enable us to calculate R_n; nevertheless, it does have great analytical significance. Let us therefore illustrate its meaning graphically, although we shall do it only for the simple case of $n = 0$.

When $n = 0$, no derivatives whatever will appear in the polynomial part P_0; therefore (9.14) reduces to

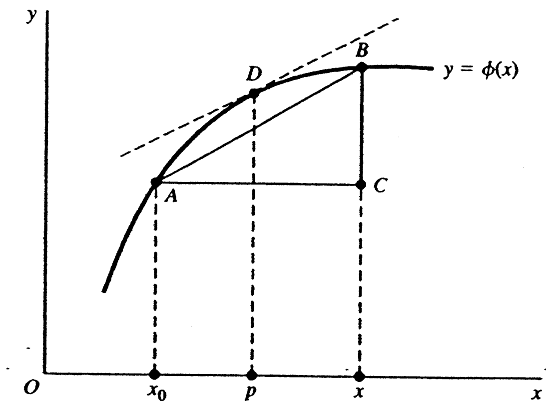
$$\phi(x) = P_0 + R_0 = \phi(x_0) + \phi'(p)(x - x_0)$$

or

$$\phi(x) - \phi(x_0) = \phi'(p)(x - x_0)$$

This result, a simple version of the ***mean-value theorem,*** states that the difference between the value of the function ϕ at x_0 and at any other x value can be expressed as the product of the difference $(x - x_0)$ and the derivative ϕ' evaluated at p (with p being some point between x and x_0). Let us look at Fig. 9.9, where the function $\phi(x)$ is shown as a continuous curve with derivative values defined at all points. Let x_0 be the chosen point of expansion, and let x be *any* point on the horizontal axis. If we try to approximate $\phi(x)$, or distance xB, by $\phi(x_0)$, or distance x_0A, it will involve an error equal to $\phi(x) - \phi(x_0)$, or the distance CB. What the mean-value theorem says is that the error CB—which constitutes the value of the remainder term R_0 in the expansion—can be expressed as $\phi'(p)(x - x_0)$, where p is some point between x and x_0. First we locate, on the curve between points

FIGURE 9.9



A and B, a point D such that the tangent line at D is parallel to line AB; such a point D must exist, since the curve passes from A to B in a continuous and smooth manner. Then, the remainder will be

$$\begin{aligned}
R_0 = CB = \frac{CB}{AC}AC &= (\text{slope of } AB) \cdot AC \\
&= (\text{slope of tangent at } D) \cdot AC \\
&= (\text{slope of curve at } x = p) \cdot AC \\
&= \phi'(p)(x - x_0)
\end{aligned}$$

where the point p is between x and x_0, as required. This demonstrates the rationale of the Lagrange form of the remainder for the case $n = 0$. We can always express R_0 as $\phi'(p)(x - x_0)$ because, even though p cannot be assigned a specific value, we can be sure that such a point exists.

Equation (9.15) provides a way of expressing the remainder term R_n, but it does not eliminate R_n as a source of discrepancy between $\phi(x)$ and the polynomial P_n. However, if it happens that as we increase n (thus raising the degree of the polynomial) indefinitely, we find that

$$R_n \to 0 \text{ as } n \to \infty \qquad \text{so that} \qquad P_n \to \phi(x) \text{ as } n \to \infty$$

then the Taylor series is said to be convergent to $\phi(x)$ at the point of expansion, and the Taylor series can be written as a *convergent infinite series* as follows:

$$\phi(x) = \frac{\phi(x_0)}{0!} + \frac{\phi'(x_0)}{1!}(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots \qquad \textbf{(9.16)}$$

Note that the R_n term is no longer shown; in its place is an ellipsis signifying that the polynomial contains an infinite number of subsequent terms whose mathematical structures follow the pattern indicated by the previous terms. In this (convenient) event, it will be possible to make P_n as accurate an approximation to $\phi(x)$ as we desire by choosing a large enough value for n, that is, by including a large enough number of terms in the polynomial P_n. An important example of this will be discussed in Sec. 10.2.

EXERCISE 9.5

1. Find the value of the following factorial expressions:

   (*a*) 5! (*c*) $\frac{4!}{3!}$ (*e*) $\frac{(n+2)!}{n!}$

   (*b*) 8! (*d*) $\frac{6!}{4!}$

2. Find the first five terms of the Maclaurin series (i.e., choose $n = 4$ and let $x_0 = 0$) for:

   (*a*) $\phi(x) = \frac{1}{1-x}$ (*b*) $\phi(x) = \frac{1-x}{1+x}$

3. Find the Taylor series with $n = 4$ and $x_0 = -2$, for the two functions in Prob. 2.
4. On the basis of Taylor's formula with the Lagrange form of the remainder [see (9.14) and (9.15)], show that at the point of expansion ($x = x_0$) the Taylor series will always give *exactly* the value of the function at that point, $\phi(x_0)$, not merely an approximation.

9.6 *N*th-Derivative Test for Relative Extremum of a Function of One Variable

The expansion of a function into a Taylor (or Maclaurin) series is useful as an approximation device in the circumstance that $R_n \to 0$ as $n \to \infty$, but our present concern is with its application in the development of a general test for a relative extremum.

Taylor Expansion and Relative Extremum

As a preparatory step for that task, let us redefine a relative extremum as follows:

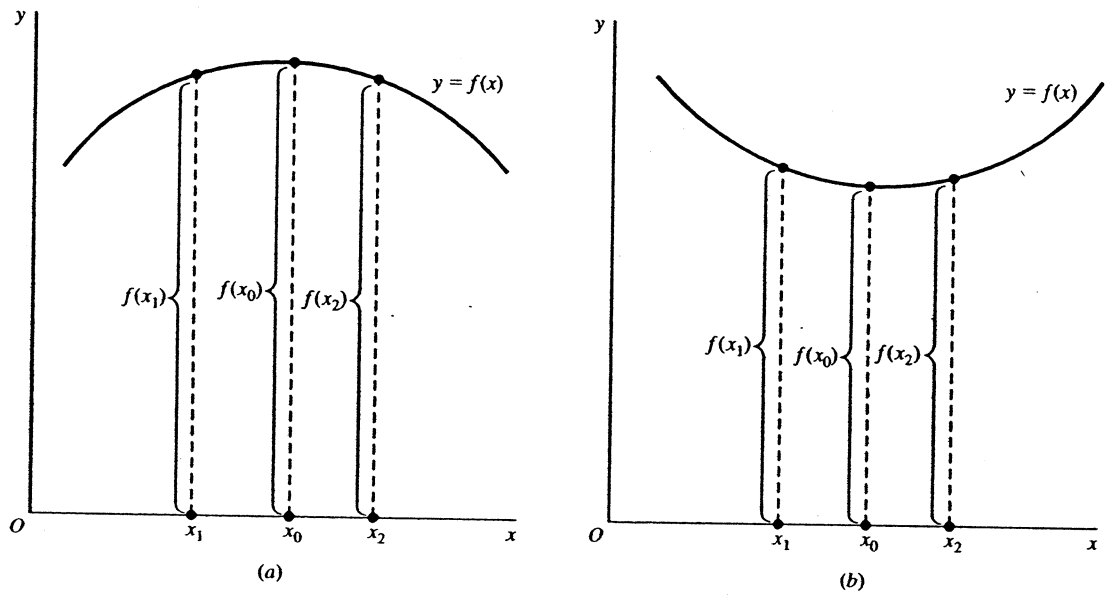
> A function $f(x)$ attains a relative maximum (minimum) value at x_0 if $f(x) - f(x_0)$ is negative (positive) for values of x in the immediate neighborhood of x_0, both to its left and to its right.

This can be made clear by reference to Fig. 9.10, where x_1 is a value of x to the left of x_0, and x_2 is a value of x to the right of x_0. In Fig. 9.10*a*, $f(x_0)$ is a relative maximum; thus $f(x_0)$ exceeds both $f(x_1)$ and $f(x_2)$. In short, $f(x) - f(x_0)$ is negative for any value of x in the immediate neighborhood of x_0. The opposite is true of Fig. 9.10*b*, where $f(x_0)$ is a relative minimum, and thus $f(x) - f(x_0) > 0$.

Assuming $f(x)$ to have finite, continuous derivatives up to the desired order at the point $x = x_0$, the function $f(x)$—not necessarily polynomial—can be expanded around the point x_0 as a Taylor series. On the basis of (9.14) (after duly changing ϕ to f), and using the Lagrange form of the remainder, we can write

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(p)}{(n+1)!}(x - x_0)^{n+1} \qquad (9.17)$$

FIGURE 9.10



If the sign of the expression $f(x) - f(x_0)$ can be determined for values of x to the immediate left and right of x_0, we can readily come to a conclusion as to whether $f(x_0)$ is an extremum, and if so, whether it is a maximum or a minimum. For this, it is necessary to examine the right-hand sum of (9.17). Altogether, there are $(n + 1)$ terms in this sum—n terms from P_n, plus the remainder which is in the $(n + 1)$st degree—and thus the actual number of terms is indefinite, being dependent upon the value of n we choose. However, by properly choosing n, we can always make sure that there will exist only a single term on the right. This will drastically simplify the task of evaluating the sign of $f(x) - f(x_0)$ and ascertaining whether $f(x_0)$ is an extremum, and if so, which kind.

Some Specific Cases

This can be made clearer through some specific illustrations.

Case 1 $\qquad f'(x_0) \neq 0$

If the first derivative at x_0 is nonzero, let us choose $n = 0$, so that the remainder will be in the first degree. Then there will be only $n + 1 = 1$ term on the right side, implying that only the remainder R_0 will be there. That is, we have

$$f(x) - f(x_0) = \frac{f'(p)}{1!}(x - x_0) = f'(p)(x - x_0)$$

where p is some number between x_0 and a value of x in the immediate neighborhood of x_0. Note that p must accordingly be very, very close to x_0.

What is the sign of the expression on the right? Because of the continuity of the derivative, $f'(p)$ will have the same sign as $f'(x_0)$ since, as mentioned before, p is very, very close to x_0. In the present case, $f'(p)$ must be nonzero; in fact, it must be a specific positive or negative number. But what about the $(x - x_0)$ part? When we go from the left of x_0 to its right, x shifts from a magnitude $x_1 < x_0$ to a magnitude $x_2 > x_0$ (see Fig. 9.10). Consequently, the expression $(x - x_0)$ must turn from negative to positive as we move, and $f(x) - f(x_0) = f'(p)(x - x_0)$ must also change sign from the left of x_0 to its right. However, this violates our new definition of a relative extremum; accordingly, there cannot exist a relative extremum at $f(x_0)$ when $f'(x_0) \neq 0$—a fact that is already well known to us.

Case 2 $\qquad f'(x_0) = 0;\ f''(x_0) \neq 0$

In this case, choose $n = 1$, so that the remainder will be in the second degree. Then initially there will be $n + 1 = 2$ terms on the right. But one of these terms will vanish because $f'(x_0) = 0$, and we shall again be left with only one term to evaluate:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{f''(p)}{2!}(x - x_0)^2 \\ &= \tfrac{1}{2} f''(p)(x - x_0)^2 \qquad [\text{because } f'(x_0) = 0] \end{aligned}$$

As before, $f''(p)$ will have the same sign as $f''(x_0)$, a sign that is specified and unvarying, whereas the $(x - x_0)^2$ part, being a square, is invariably positive. Thus the expression $f(x) - f(x_0)$ must take the same sign as $f''(x_0)$ and, according to the earlier definition of relative extremum, will specify

$$\left.\begin{array}{l} \text{A relative maximum of } f(x) \text{ if } f''(x_0) < 0 \\ \text{A relative minimum of } f(x) \text{ if } f''(x_0) > 0 \end{array}\right. \qquad [\text{with } f'(x_0) = 0]$$

You will recognize this as the second-derivative test introduced earlier.

Case 3 $\qquad f'(x_0) = f''(x_0) = 0$, but $f'''(x_0) \neq 0$

Here we are encountering a situation that the second-derivative test is incapable of handling, for $f''(x_0)$ is now zero. With the help of the Taylor series, however, a conclusive result can be established without difficulty.

Let us choose $n = 2$; then three terms will initially appear on the right. But two of these will drop out because $f'(x_0) = f''(x_0) = 0$, so that we again have only one term to evaluate:

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \frac{1}{3!} f'''(p)(x - x_0)^3 \\ &= \frac{1}{6} f'''(p)(x - x_0)^3 \qquad [\text{because } f'(x_0) = 0,\ f''(x_0) = 0] \end{aligned}$$

As previously, the sign of $f'''(p)$ is identical with that of $f'''(x_0)$ because of the continuity of the derivative and because p is very close to x_0. But the $(x - x_0)^3$ part has a varying sign. Specifically, since $(x - x_0)$ is negative to the left of x_0, so also will be $(x - x_0)^3$; yet, to the right of x_0, the $(x - x_0)^3$ part will be positive. Thus there is a change in the sign of $f(x) - f(x_0)$ as we pass through x_0, which violates the definition of a relative extremum. However, we know that x_0 is a critical value $[f'(x_0) = 0]$, and thus it must give an inflection point, inasmuch as it does not give a relative extremum.

Case 4 $\qquad f'(x_0) = f''(x_0) = \cdots = f^{(N-1)}(x_0) = 0$, but $f^{(N)}(x_0) \neq 0$

This is a very general case, and we can therefore derive a general result from it. Note that here all the derivative values are zero until we arrive at the *N*th one.

Analogously to the preceding three cases, the Taylor series for case 4 will reduce to

$$f(x) - f(x_0) = \frac{1}{N!} f^{(N)}(p)(x - x_0)^N$$

Again, $f^{(N)}(p)$ takes the same sign as $f^{(N)}(x_0)$, which is unvarying. The sign of the $(x - x_0)^N$ part, on the other hand, will *vary* if N is *odd* (cf. Cases 1 and 3) and will *remain unchanged* (positive) if N is *even* (cf. Case 2). When N is odd, accordingly, $f(x) - f(x_0)$ will change sign as we pass through the point x_0, thereby violating the definition of a relative extremum (which means that x_0 must give us an inflection point on the curve). But when N is even, $f(x) - f(x_0)$ will not change sign from the left of x_0 to its right, and this will establish the stationary value $f(x_0)$ as a relative maximum or minimum, depending on whether $f^{(N)}(x_0)$ is negative or positive.

*N*th-Derivative Test

At last, then, we may state the following general test.

***N*th-Derivative test for relative extremum of a function of one variable** If the first derivative of a function $f(x)$ at x_0 is $f'(x_0) = 0$ and if the first *nonzero* derivative value at x_0 encountered in successive derivation is that of the *N*th derivative, $f^{(N)}(x_0) \neq 0$, then the stationary value $f(x_0)$ will be

a. A relative *maximum* if N is an even number and $f^{(N)}(x_0) < 0$.
b. A relative *minimum* if N is an even number but $f^{(N)}(x_0) > 0$.
c. An *inflection point* if N is odd.

It should be clear from the preceding statement that the *N*th-derivative test can work if and only if the function $f(x)$ is capable of yielding, sooner or later, a nonzero derivative value at the critical value x_0. While there do exist exceptional functions that fail to satisfy this condition, most of the functions we are likely to encounter will indeed produce nonzero $f^{(N)}(x_0)$ in successive differentiation.† Thus the test should prove serviceable in most instances.

† If $f(x)$ is a constant function, for instance, then obviously $f'(x) = f''(x) = \cdots = 0$, so that no nonzero derivative value can ever be found. This, however, is a trivial case, since a constant function requires no test for extremum anyway. As a nontrivial example, consider the function

$$y = \begin{cases} e^{-1/x^2} & (\text{for } x \neq 0) \\ 0 & (\text{for } x = 0) \end{cases}$$

where the function $y = e^{-1/x^2}$ is an exponential function, yet to be introduced (Chap. 10). By itself, $y = e^{-1/x^2}$ is discontinuous at $x = 0$, because $x = 0$ is not in the domain (division by zero is undefined). However, since $\lim_{x \to 0} y = 0$, we can, by appending the stipulation that $y = 0$ for $x = 0$, fill the gap in the domain and thereby obtain a continuous function. The graph of this function shows that it attains a minimum at $x = 0$. But it turns out that, at $x = 0$, all the derivatives (up to any order) have zero values. Thus we are unable to apply the *N*th-derivative test to confirm the graphically ascertainable fact that the function has a minimum at $x = 0$. For further discussion of this exceptional case, see R. Courant, *Differential and Integral Calculus* (translated by E. J. McShane), Interscience, New York, vol. I, 2d ed., 1937, pp. 196, 197, and 336.

Example 1 Examine the function $y = (7 - x)^4$ for its relative extremum. Since $f'(x) = -4(7 - x)^3$ is zero when $x = 7$, we take $x = 7$ as the critical value for testing, with $y = 0$ as the stationary value of the function. By successive derivation (continued until we encounter a nonzero derivative value at the point $x = 7$), we get

$$f''(x) = 12(7 - x)^2 \quad \text{so that} \quad f''(7) = 0$$
$$f'''(x) = -24(7 - x) \qquad f'''(7) = 0$$
$$f^{(4)}(x) = 24 \qquad f^{(4)}(7) = 24$$

Since 4 is an even number and since $f^{(4)}(7)$ is positive, we conclude that the point (7, 0) represents a relative minimum.

As is easily verified, this function plots as a strictly convex curve. Inasmuch as the second derivative at $x = 7$ is zero (rather than positive), this example serves to illustrate our earlier statement regarding the second derivative and the curvature of a curve (Sec. 9.3) to the effect that, while a positive $f''(x)$ for all x does imply a strictly convex $f(x)$, a strictly convex $f(x)$ does *not* imply a positive $f''(x)$ for all x. More importantly, it also serves to illustrate the fact that, given a strictly convex (strictly concave) curve, the extremum found on that curve must be a minimum (maximum), because such an extremum will *either* satisfy the second-order sufficient condition, *or*, failing that, satisfy another (higher-order) sufficient condition for a minimum (maximum).

EXERCISE 9.6

1. Find the stationary values of the following functions:
   (*a*) $y = x^3$ (*b*) $y = -x^4$ (*c*) $y = x^6 + 5$
   Determine by the *N*th-derivative test whether they represent relative maxima, relative minima, or inflection points.
2. Find the stationary values of the following functions:
   (*a*) $y = (x - 1)^3 + 16$ (*c*) $y = (3 - x)^6 + 7$
   (*b*) $y = (x - 2)^4$ (*d*) $y = (5 - 2x)^4 + 8$
   Use the *N*th-derivative test to determine the exact nature of these stationary values.