CHAPTER

2

Mathematics for Microeconomics

Microeconomic models are constructed using a wide variety of mathematical techniques. In this chapter we provide a brief summary of some of the most important techniques that you will encounter in this book. A major portion of the chapter concerns mathematical procedures for finding the optimal value of some function. Because we will frequently adopt the assumption that an economic actor seeks to maximize or minimize some function, we will encounter these procedures (most of which are based on differential calculus) many times.

After our detailed discussion of the calculus of optimization, we turn to four topics that are covered more briefly. First, we look at a few special types of functions that arise in economic problems. Knowledge of properties of these functions can often be very helpful in solving economic problems. Next, we provide a brief summary of integral calculus. Although integration is used in this book far less frequently than is differentiation, we will nevertheless encounter several situations where we will want to employ integrals to measure areas that are important to economic theory or to add up outcomes that occur over time or across many individuals. One particular use of integration is to examine problems in which the objective is to maximize a stream of outcomes over time. Our third added topic focuses on techniques to be used for such problems in dynamic optimization. Finally, Chapter 2 concludes with a brief summary of mathematical statistics, which will be particularly useful in our study of economic behavior in uncertain situations.

MAXIMIZATION OF A FUNCTION OF ONE VARIABLE

Let's start our study of optimization with a simple example. Suppose that a manager of a firm desires to maximize[1] the profits received from selling a particular good. Suppose also that the profits (π) received depend only on the quantity (q) of the good sold. Mathematically,

$$\pi = f(q). \tag{2.1}$$

Figure 2.1 shows a possible relationship between π and q. Clearly, to achieve maximum profits, the manager should produce output q^*, which yields profits π^*. If a graph such as that of Figure 2.1 were available, this would seem to be a simple matter to be accomplished with a ruler.

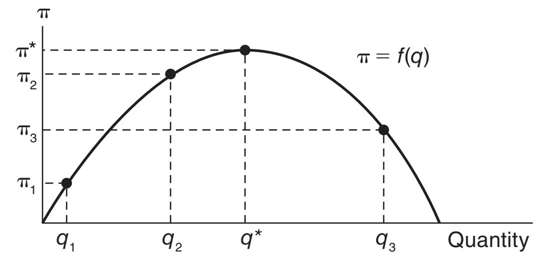
Suppose, however, as is more likely, the manager does not have such an accurate picture of the market. He or she may then try varying q to see where a maximum profit is obtained. For example, by starting at q_1, profits from sales would be π_1. Next, the manager may try output q_2, observing that profits have increased to π_2. The commonsense idea that profits have increased in response to an increase in q can be stated formally as

$$\frac{\pi_2 - \pi_1}{q_2 - q_1} > 0 \quad \text{or} \quad \frac{\Delta\pi}{\Delta q} > 0, \tag{2.2}$$

[1]Here we will generally explore maximization problems. A virtually identical approach would be taken to study minimization problems because maximization of $f(x)$ is equivalent to minimizing $-f(x)$.

FIGURE 2.1 Hypothetical Relationship between Quantity Produced and Profits

If a manager wishes to produce the level of output that maximizes profits, then q^* should be produced. Notice that at q^*, $d\pi/dq = 0$.



where the Δ notation is used to mean "the change in" π or q. As long as $\Delta\pi/\Delta q$ is positive, profits are increasing and the manager will continue to increase output. For increases in output to the right of q^*, however, $\Delta\pi/\Delta q$ will be negative, and the manager will realize that a mistake has been made.

Derivatives

As you probably know, the limit of $\Delta\pi/\Delta q$ for very small changes in q is called the *derivative* of the function, $\pi = f(q)$, and is denoted by $d\pi/dq$ or df/dq or $f'(q)$. More formally, the derivative of a function $\pi = f(q)$ at the point q_1 is defined as

$$\frac{d\pi}{dq} = \frac{df}{dq} = \lim_{h \to 0} \frac{f(q_1 + h) - f(q_1)}{h}. \qquad \textbf{(2.3)}$$

Notice that the value of this ratio obviously depends on the point q_1 that is chosen.

Value of the derivative at a point

A notational convention should be mentioned: Sometimes one wishes to note explicitly the point at which the derivative is to be evaluated. For example, the evaluation of the derivative at the point $q = q_1$ could be denoted by

$$\left.\frac{d\pi}{dq}\right|_{q=q_1}. \qquad \textbf{(2.4)}$$

At other times, one is interested in the value of $d\pi/dq$ for all possible values of q and no explicit mention of a particular point of evaluation is made.

In the example of Figure 2.1,

$$\left.\frac{d\pi}{dq}\right|_{q=q_1} > 0,$$

whereas

$$\left.\frac{d\pi}{dq}\right|_{q=q_3} < 0.$$

What is the value of $d\pi/dq$ at q^*? It would seem to be 0, because the value is positive for values of q less than q^* and negative for values of q greater than q^*. The derivative is the slope of the curve in question; this slope is positive to the left of q^* and negative to the right of q^*. At the point q^*, the slope of $f(q)$ is 0.

First-order condition for a maximum

This result is quite general. For a function of one variable to attain its maximum value at some point, the derivative at that point (if it exists) must be 0. Hence, if a manager could estimate the function $f(q)$ from some sort of real-world data, it would theoretically be possible to find the point where $df/dq = 0$. At this optimal point (say, q^*),

$$\left.\frac{df}{dq}\right|_{q=q^*} = 0. \tag{2.5}$$

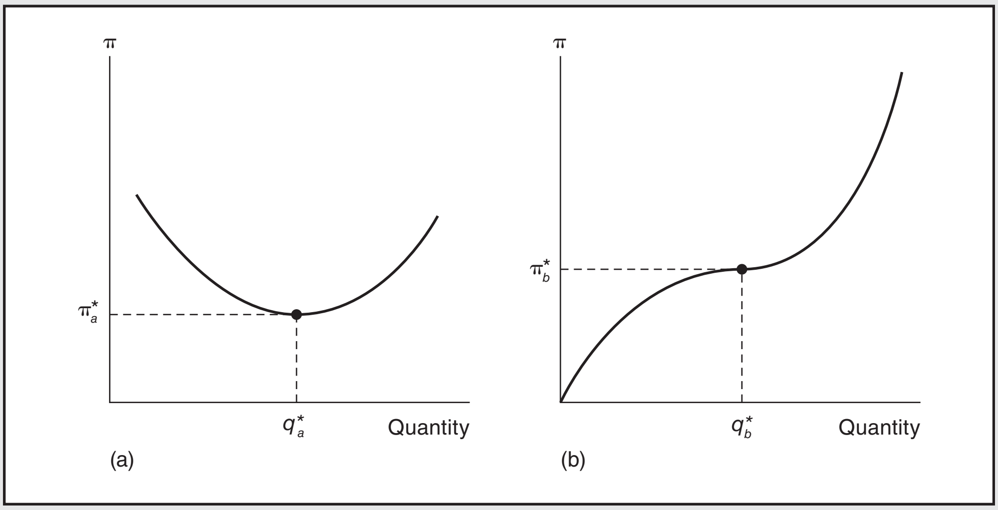
Second-order conditions

An unsuspecting manager could be tricked, however, by a naive application of this first-derivative rule alone. For example, suppose that the profit function looks like that shown in either Figure 2.2a or 2.2b. If the profit function is that shown in Figure 2.2a, the manager, by producing where $d\pi/dq = 0$, will choose point q_a^*. This point in fact yields minimum, not maximum, profits for the manager. Similarly, if the profit function is that shown in Figure 2.2, the manager will choose point q_b^*, which, although it yields a profit greater than that for any output lower than q_b^*, is certainly inferior to any output greater than q_b^*. These situations illustrate the mathematical fact that $d\pi/dq = 0$ is a *necessary* condition for a maximum, but not a *sufficient* condition. To ensure that the chosen point is indeed a maximum point, a second condition must be imposed.

Intuitively, this additional condition is clear: The profit available by producing either a bit more or a bit less than q^* must be smaller than that available from q^*. If this is not true, the manager can do better than q^*. Mathematically, this means that $d\pi/dq$ must be greater

FIGURE 2.2 Two Profit Functions That Give Misleading Results If the First Derivative Rule Is Applied Uncritically

In (a), the application of the first derivative rule would result in point q_a^* being chosen. This point is in fact a point of minimum profits. Similarly, in (b), output level q_b^* would be recommended by the first derivative rule, but this point is inferior to all outputs greater than q_b^*. This demonstrates graphically that finding a point at which the derivative is equal to 0 is a necessary, but not a sufficient, condition for a function to attain its maximum value.



than 0 for $q < q^*$ and must be less than 0 for $q > q^*$. Therefore, at q^*, $d\pi/dq$ must be decreasing. Another way of saying this is that the derivative of $d\pi/dq$ must be negative at q^*.

Second derivatives

The derivative of a derivative is called a *second derivative* and is denoted by

$$\frac{d^2\pi}{dq^2} \quad \text{or} \quad \frac{d^2 f}{dq^2} \quad \text{or} \quad f(q).$$

The additional condition for q^* to represent a (local) maximum is therefore

$$\left.\frac{d^2\pi}{dq^2}\right|_{q=q^*} = \left.f''(q)\right|_{q=q^*} < 0, \tag{2.6}$$

where the notation is again a reminder that this second derivative is to be evaluated at q^*.

Hence, although Equation 2.5 ($d\pi/dq = 0$) is a necessary condition for a maximum, that equation must be combined with Equation 2.6 ($d^2\pi/dq^2 < 0$) to ensure that the point is a local maximum for the function. Equations 2.5 and 2.6 together are therefore sufficient conditions for such a maximum. Of course, it is possible that by a series of trials the manager may be able to decide on q^* by relying on market information rather than on mathematical reasoning (remember Friedman's pool-player analogy). In this book we shall be less interested in how the point is discovered than in its properties and how the point changes when conditions change. A mathematical development will be very helpful in answering these questions.

Rules for finding derivatives

Here are a few familiar rules for taking derivatives. We will use these at many places in this book.

1. If b is a constant, then
$$\frac{db}{dx} = 0.$$
2. If b is a constant, then
$$\frac{d[bf(x)]}{dx} = bf'(x).$$
3. If b is a constant, then
$$\frac{dx^b}{dx} = bx^{b-1}.$$
4. $\dfrac{d\ln x}{dx} = \dfrac{1}{x}$

   where ln signifies the logarithm to the base e ($= 2.71828$).
5. $\dfrac{da^x}{dx} = a^x \ln a$ for any constant a

   A particular case of this rule is $de^x/dx = e^x$.

Now suppose that $f(x)$ and $g(x)$ are two functions of x and that $f'(x)$ and $g'(x)$ exist. Then:

6. $\dfrac{d[f(x) + g(x)]}{dx} = f'(x) + g'(x).$

7. $\dfrac{d[f(x)\cdot g(x)]}{dx} = f(x)g'(x) + f'(x)g(x).$

8. $\dfrac{d[f(x)/g(x)]}{dx} = \dfrac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2},$

   provided that $g(x) \neq 0$.

Finally, if $y = f(x)$ and $x = g(z)$ and if both $f'(x)$ and $g'(z)$ exist, then

9. $\dfrac{dy}{dz} = \dfrac{dy}{dx}\cdot\dfrac{dx}{dz} = \dfrac{df}{dx}\cdot\dfrac{dg}{dz}.$

This result is called the *chain rule*. It provides a convenient way to study how one variable (z) affects another variable (y) solely through its influence on some intermediate variable (x). Some examples are

10. $\dfrac{de^{ax}}{dx} = \dfrac{de^{ax}}{d(ax)}\cdot\dfrac{d(ax)}{dx} = e^{ax}\cdot a = ae^{ax}.$

11. $\dfrac{d[\ln(ax)]}{dx} = \dfrac{d[\ln(ax)]}{d(ax)}\cdot\dfrac{d(ax)}{dx} = \dfrac{1}{ax}\cdot a = \dfrac{1}{x}.$

12. $\dfrac{d[\ln(x^2)]}{dx} = \dfrac{d[\ln(x^2)]}{d(x^2)}\cdot\dfrac{d(x^2)}{dx} = \dfrac{1}{x^2}\cdot 2x = \dfrac{2}{x}.$

FUNCTIONS OF SEVERAL VARIABLES

Economic problems seldom involve functions of only a single variable. Most goals of interest to economic agents depend on several variables, and trade-offs must be made among these variables. For example, the *utility* an individual receives from activities as a consumer depends on the amount of each good consumed. For a firm's *production function*, the amount produced depends on the quantity of labor, capital, and land devoted to production. In these circumstances, this dependence of one variable (y) on a series of other variables ($x_1, x_2, \ldots, x_n$) is denoted by

$$y = f(x_1, x_2, \ldots, x_n). \qquad \textbf{(2.7)}$$

Partial derivatives

We are interested in the point at which y reaches a maximum and in the trade-offs that must be made to reach that point. It is again convenient to picture the agent as changing the variables at his or her disposal (the x's) in order to locate a maximum. Unfortunately, for a function of several variables, the idea of *the* derivative is not well-defined. Just as the steepness of ascent when climbing a mountain depends on which direction you go, so does the slope (or derivative) of the function depend on the direction in which it is taken. Usually, the only directional slopes of interest are those that are obtained by increasing one of the x's while holding all the other variables constant (the analogy for mountain climbing might be to measure slopes only in a north-south or east-west direction). These directional slopes are called *partial derivatives*. The partial derivative of y with respect to (that is, in the direction of) x_1 is denoted by

$$\frac{\partial y}{\partial x_1} \quad \text{or} \quad \frac{\partial f}{\partial x_1} \quad \text{or} \quad f_{x_1} \quad \text{or} \quad f_1.$$

It is understood that in calculating this derivative all of the other x's are held constant. Again it should be emphasized that the numerical value of this slope depends on the value of x_1 and on the (preassigned) values of $x_2, \ldots, x_n$.

EXAMPLE 2.1 Profit Maximization

Suppose that the relationship between profits (π) and quantity produced (q) is given by

$$\pi(q) = 1{,}000q - 5q^2. \tag{2.8}$$

A graph of this function would resemble the parabola shown in Figure 2.1. The value of q that maximizes profits can be found by differentiation:

$$\frac{d\pi}{dq} = 1{,}000 - 10q = 0, \tag{2.9}$$

so

$$q^* = 100. \tag{2.10}$$

At $q = 100$, Equation 2.8 shows that profits are 50,000—the largest value possible. If, for example, the firm opted to produce $q = 50$, profits would be 37,500. At $q = 200$, profits are precisely 0.

That $q = 100$ is a "global" maximum can be shown by noting that the second derivative of the profit function is -10 (see Equation 2.9). Hence, the rate of increase in profits is always decreasing—up to $q = 100$ this rate of increase is still positive, but beyond that point it becomes negative. In this example, $q = 100$ is the only local maximum value for the function π. With more complex functions, however, there may be several such maxima.

QUERY: Suppose that a firm's output (q) is determined by the amount of labor (l) it hires according to the function $q = 2\sqrt{l}$. Suppose also that the firm can hire all of the labor it wants at \$10 per unit and sells its output at \$50 per unit. Profits are therefore a function of l given by $\pi(l) = 100\sqrt{l} - 10l$. How much labor should this firm hire in order to maximize profits, and what will those profits be?

A somewhat more formal definition of the partial derivative is

$$\left.\frac{\partial f}{\partial x_1}\right|_{\bar{x}_2, \ldots, \bar{x}_n} = \lim_{h \to 0} \frac{f(x_1 + h, \bar{x}_2, \ldots, \bar{x}_n) - f(x_1, \bar{x}_2, \ldots, \bar{x}_n)}{h}, \tag{2.11}$$

where the notation is intended to indicate that $x_2, \ldots, x_n$ are all held constant at the preassigned values $\bar{x}_2, \ldots, \bar{x}_n$ so the effect of changing x_1 only can be studied. Partial derivatives with respect to the other variables $(x_2, \ldots, x_n)$ would be calculated in a similar way.

Calculating partial derivatives

It is easy to calculate partial derivatives. The calculation proceeds as for the usual derivative by *treating* $x_2, \ldots, x_n$ *as constants* (which indeed they are in the definition of a partial derivative). Consider the following examples.

1. If $y = f(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$, then

$$\frac{\partial f}{\partial x_1} = f_1 = 2ax_1 + bx_2$$

and

$$\frac{\partial f}{\partial x_2} = f_2 = bx_1 + 2cx_2.$$

Notice that $\partial f/\partial x_1$ is in general a function of both x_1 and x_2 and therefore its value will depend on the particular values assigned to these variables. It also depends on the parameters a, b, and c, which do not change as x_1 and x_2 change.

2. If $y = f(x_1, x_2) = e^{ax_1 + bx_2}$, then

$$\frac{\partial f}{\partial x_1} = f_1 = ae^{ax_1 + bx_2}$$

and

$$\frac{\partial f}{\partial x_2} = f_2 = be^{ax_1 + bx_2}.$$

3. If $y = f(x_1, x_2) = a \ln x_1 + b \ln x_2$, then

$$\frac{\partial f}{\partial x_1} = f_1 = \frac{a}{x_1}$$

and

$$\frac{\partial f}{\partial x_2} = f_2 = \frac{b}{x_2}.$$

Notice here that the treatment of x_2 as a constant in the derivation of $\partial f/\partial x_1$ causes the term $b \ln x_2$ to disappear upon differentiation because it does not change when x_1 changes. In this case, unlike our previous examples, the size of the effect of x_1 on y is independent of the value of x_2. In other cases, the effect of x_1 on y will depend on the level of x_2.

Partial derivatives and the ceteris paribus assumption

In Chapter 1, we described the way in which economists use the ceteris paribus assumption in their models to hold constant a variety of outside influences so the particular relationship being studied can be explored in a simplified setting. Partial derivatives are a precise mathematical way of representing this approach; that is, they show how changes in one variable affect some outcome when other influences are held constant—exactly what economists need for their models. For example, Marshall's demand curve shows the relationship between price (p) and quantity (q) demanded when other factors are held constant. Using partial derivatives, we could represent the slope of this curve by $\partial q/\partial p$ to indicate the ceteris paribus assumptions that are in effect. The fundamental law of demand—that price and quantity move in opposite directions when other factors do not change—is therefore reflected by the mathematical statement "$\partial q/\partial p < 0$." Again, the use of a partial derivative serves as a reminder of the ceteris paribus assumptions that surround the law of demand.

Partial derivatives and units of measurement

In mathematics relatively little attention is paid to how variables are measured. In fact, most often no explicit mention is made of the issue. But the variables used in economics usually refer to real-world magnitudes and therefore we must be concerned with how they are measured. Perhaps the most important consequence of choosing units of measurement is that the partial derivatives often used to summarize economic behavior will reflect these units. For example, if q represents the quantity of gasoline demanded by all U.S. consumers during a given year (measured in billions of gallons) and p represents the price in dollars per gallon, then $\partial q/\partial p$ will measure the change in demand (in billions of gallons per year) for a dollar per gallon change in price. The numerical size of this derivative depends on how q and p are measured. A decision to measure consumption in millions of gallons per year would multiply the size of the derivative by 1,000, whereas a decision to measure price in cents per gallon would reduce it by a factor of 100.

The dependence of the numerical size of partial derivatives on the chosen units of measurement poses problems for economists. Although many economic theories make predictions about the sign (direction) of partial derivatives, any predictions about the numerical magnitude of such derivatives would be contingent on how authors chose to measure their variables. Making comparisons among studies could prove practically impossible, especially given the wide variety of measuring systems in use around the world. For this reason, economists have chosen to adopt a different, unit-free way to measure quantitative impacts.

Elasticity—A general definition

Economists use elasticities to summarize virtually all of the quantitative impacts that are of interest to them. Because such measures focus on the proportional effect of a change in one variable on another, they are unit-free—the units "cancel out" when the elasticity is calculated. Suppose, for example, that y is a function of x and, possibly, other variables. Then the elasticity of y with respect to x (denoted as $e_{y,x}$) is defined as

$$e_{y,x} = \frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{\partial y}{\partial x} \cdot \frac{x}{y}. \tag{2.12}$$

Notice that, no matter how the variables y and x are measured, the units of measurement cancel out because they appear in both a numerator and a denominator. Notice also that, because y and x are positive in most economic situations, the elasticity $e_{y,x}$ and the partial derivative $\partial y/\partial x$ will have the same sign. Hence, theoretical predictions about the direction of certain derivatives will also apply to their related elasticities.

Specific applications of the elasticity concept will be encountered throughout this book. These include ones with which you should be familiar, such as the market price elasticity of demand or supply. But many new concepts that can be expressed most clearly in elasticity terms will also be introduced.

EXAMPLE 2.2 Elasticity and Functional Form

The definition in Equation 2.12 makes clear that elasticity should be evaluated at a specific point on a function. In general the value of this parameter would be expected to vary across different ranges of the function. This observation is most clearly shown in the case where y is a linear function of x of the form

$$y = a + bx + \textbf{other terms}.$$

In this case,

$$e_{y,x} = \frac{\partial y}{\partial x} \cdot \frac{x}{y} = b \cdot \frac{x}{y} = b \cdot \frac{x}{a + bx + \cdots}, \tag{2.13}$$

which makes clear that $e_{y,x}$ is not constant. Hence, for linear functions it is especially important to note the point at which elasticity is to be computed.

If the functional relationship between y and x is of the exponential form

$$y = ax^b$$

then the elasticity is a constant, independent of where it is measured:

$$e_{y,x} = \frac{\partial y}{\partial x} \cdot \frac{x}{y} = abx^{b-1} \cdot \frac{x}{ax^b} = b.$$

A logarithmic transformation of this equation also provides a very convenient alternative definition of elasticity. Because

$$\ln y = \ln a + b \ln x,$$

we have

$$e_{y,x} = b = \frac{\partial \ln y}{\partial \ln x}. \tag{2.14}$$

Hence, elasticities can be calculated through "logarithmic differentiation." As we shall see, this is frequently the easiest way to proceed in making such calculations.

QUERY: Are there any functional forms in addition to the exponential that have a constant elasticity, at least over some range?

Second-order partial derivatives

The partial derivative of a partial derivative is directly analogous to the second derivative of a function of one variable and is called a *second-order partial derivative*. This may be written as

$$\frac{\partial(\partial f/\partial x_i)}{\partial x_j}$$

or more simply as

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = f_{ij}. \tag{2.15}$$

For the examples above:

1. $$\frac{\partial^2 f}{\partial x_1 \partial x_1} = f_{11} = 2a$$
   $$f_{12} = b$$
   $$f_{21} = b$$
   $$f_{22} = 2c.$$

2. $$f_{11} = a^2 e^{ax_1 + bx_2}$$
   $$f_{12} = abe^{ax_1 + bx_2}$$
   $$f_{21} = abe^{ax_1 + bx_2}$$
   $$f_{22} = b^2 e^{ax_1 + bx_2}$$

3. $$f_{11} = \frac{-a}{x_1^2}$$
   $$f_{12} = 0$$
   $$f_{21} = 0$$
   $$f_{22} = \frac{-b}{x_2^2}.$$

Young's theorem

These examples illustrate the mathematical result that, under quite general conditions, the order in which partial differentiation is conducted to evaluate second-order partial derivatives does not matter. That is,

$$f_{ij} = f_{ji} \tag{2.16}$$

for any pair of variables x_i, x_j. This result is sometimes called "Young's theorem." For an intuitive explanation of the theorem, we can return to our mountain-climbing analogy. In this example, the theorem states that the gain in elevation a hiker experiences depends on the directions and distances traveled, but not on the order in which these occur. That is, the gain in altitude is independent of the actual path taken as long as the hiker proceeds from one set of map coordinates to another. He or she may, for example, go one mile north, then one mile east or proceed in the opposite order by first going one mile east, then one mile north. In either case, the gain in elevation is the same since in both cases the hiker is moving from one specific place to another. In later chapters we will make good use of this result because it provides a very convenient way of showing some of the predictions that economic models make about behavior.[2]

Uses of second-order partials

Second-order partial derivatives will play an important role in many of the economic theories that are developed throughout this book. Probably the most important examples relate to the "own" second-order partial, f_{ii}. This function shows how the marginal influence of x_i on y (i.e., $\partial y/\partial x_i$) changes as the value of x_i increases. A negative value for f_{ii} is the mathematical way of indicating the economic idea of diminishing marginal effectiveness. Similarly, the cross-partial f_{ij} indicates how the marginal effectiveness of x_i changes as x_j increases. The sign of this effect could be either positive or negative. Young's theorem indicates that, in general, such cross-effects are symmetric. More generally, the second-order partial derivatives of a function provide information about the curvature of the function. Later in this chapter we will see how such information plays an important role in determining whether various second-order conditions for a maximum are satisfied.

MAXIMIZATION OF FUNCTIONS OF SEVERAL VARIABLES

Using partial derivatives, we can now discuss how to find the maximum value for a function of several variables. To understand the mathematics used in solving this problem, an analogy to the one-variable case is helpful. In this one-variable case, we can picture an agent varying x by a small amount, dx, and observing the change in y, dy. This change is given by

$$dy = f'(x)dx. \tag{2.17}$$

The identity in Equation 2.17 records the fact that the change in y is equal to the change in x times the slope of the function. This formula is equivalent to the *point-slope* formula used for linear equations in basic algebra. As before, the necessary condition for a maximum is that $dy = 0$ for small changes in x around the optimal point. Otherwise, y could be increased by suitable changes in x. But because dx does not necessarily equal 0 in Equation 2.17, $dy = 0$ must imply that at the desired point, $f'(x) = 0$. This is another way of obtaining the first-order condition for a maximum that we already derived.

---

[2]Young's theorem implies that the matrix of the second-order partial derivatives of a function is symmetric. This symmetry offers a number of economic insights. For a brief introduction to the matrix concepts used in economics, see the Extensions to this chapter.

Using this analogy, let's look at the decisions made by an economic agent who must choose the levels of several variables. Suppose that this agent wishes to find a set of x's that will maximize the value of $y = f(x_1, x_2, \ldots, x_n)$. The agent might consider changing only one of the x's, say x_1, while holding all the others constant. The change in y (that is, dy) that would result from this change in x_1 is given by

$$dy = \frac{\partial f}{\partial x_1} dx_1 = f_1 dx_1.$$

This says that the change in y is equal to the change in x_1 times the slope measured in the x_1 direction. Using the mountain analogy again, the gain in altitude a climber heading north would achieve is given by the distance northward traveled times the slope of the mountain measured in a northward direction.

Total differential

If all the x's are varied by a small amount, the total effect on y will be the sum of effects such as that shown above. Therefore the total change in y is defined to be

$$\begin{aligned} dy &= \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n \\ &= f_1 dx_1 + f_2 dx_2 + \cdots + f_n dx_n. \end{aligned} \tag{2.18}$$

This expression is called the *total differential* of f and is directly analogous to the expression for the single-variable case given in Equation 2.17. The equation is intuitively sensible: The total change in y is the sum of changes brought about by varying each of the x's.[3]

First-order condition for a maximum

A necessary condition for a maximum (or a minimum) of the function $f(x_1, x_2, \ldots, x_n)$ is that $dy = 0$ for any combination of small changes in the x's. The only way this can happen is if, at the point being considered,

$$f_1 = f_2 = \cdots = f_n = 0. \tag{2.19}$$

A point where Equations 2.19 hold is called a *critical point*. Equations 2.19 are the necessary conditions for a local maximum. To see this intuitively, note that if one of the partials (say, f_i) were greater (or less) than 0, then y could be increased by increasing (or decreasing) x_i. An economic agent then could find this maximal point by finding the spot where y does not respond to very small movements in any of the x's. This is an extremely important result for economic analysis. It says that any activity (that is, the x's) should be pushed to the point where its "marginal" contribution to the objective (that is, y) is 0. To stop short of that point would fail to maximize y.

---

[3]The total differential in Equation 2.18 can be used to derive the chain rule as it applies to functions of several variables. Suppose that $y = f(x_1, x_2)$ and that $x_1 = g(z)$ and $x_2 = h(z)$. If all of these functions are differentiable, then it is possible to calculate the effects of a change in z on y. The total differential of y is

$$dy = f_1 dx_1 + f_2 dx_2.$$

Dividing this equation by dz gives

$$\frac{dy}{dz} = f_1 \frac{dx_1}{dz} + f_2 \frac{dx_2}{dz} = f_1 \frac{dg}{dz} + f_2 \frac{dh}{dz}.$$

Hence, calculating the effect of z on y requires calculating how z affects both of the determinants of y (that is, x_1 and x_2). If y depends on more than two variables, an analogous result holds. This result acts as a reminder to be rather careful to include all possible effects when calculating derivatives of functions of several variables.

EXAMPLE 2.3 Finding a Maximum

Suppose that y is a function of x_1 and x_2 given by

$$y = -(x_1 - 1)^2 - (x_2 - 2)^2 + 10 \tag{2.20}$$

or

$$y = -x_1^2 + 2x_1 - x_2^2 + 4x_2 + 5.$$

For example, y might represent an individual's health (measured on a scale of 0 to 10), and x_1 and x_2 might be daily dosages of two health-enhancing drugs. We wish to find values for x_1 and x_2 that make y as large as possible. Taking the partial derivatives of y with respect to x_1 and x_2 and applying the necessary conditions given by Equations 2.19 yields

$$\begin{aligned} \frac{\partial y}{\partial x_1} &= -2x_1 + 2 = 0, \\ \frac{\partial y}{\partial x_2} &= -2x_2 + 4 = 0 \end{aligned} \tag{2.21}$$

or

$$\begin{aligned} x_1^* &= 1, \\ x_2^* &= 2. \end{aligned}$$

The function is therefore at a critical point when $x_1 = 1$, $x_2 = 2$. At that point, $y = 10$ is the best health status possible. A bit of experimentation provides convincing evidence that this is the greatest value y can have. For example, if $x_1 = x_2 = 0$, then $y = 5$, or if $x_1 = x_2 = 1$, then $y = 9$. Values of x_1 and x_2 larger than 1 and 2, respectively, reduce y because the negative quadratic terms in Equation 2.20 become large. Consequently, the point found by applying the necessary conditions is in fact a local (and global) maximum.[4]

QUERY: Suppose y took on a fixed value (say, 5). What would the relationship implied between x_1 and x_2 look like? How about for $y = 7$? Or $y = 10$? (These graphs are *contour lines* of the function and will be examined in more detail in several later chapters. See also Problem 2.1.)

Second-order conditions

Again, however, the conditions of Equations 2.19 are not sufficient to ensure a maximum. This can be illustrated by returning to an already overworked analogy: All hilltops are (more or less) flat, but not every flat place is a hilltop. A second-order condition similar to Equation 2.6 is needed to ensure that the point found by applying Equations 2.19 is a local maximum. Intuitively, for a local maximum, y should be decreasing for any small changes in the x's away from the critical point. As in the single-variable case, this necessarily involves looking at the second-order partial derivatives of the function f. These second-order partials must obey certain restrictions (analogous to the restriction that was derived in the single-variable case) if the critical point found by applying Equations 2.19 is to be a local maximum. Later in this chapter we will look at these restrictions.

[4]More formally, the point $x_1 = 1$, $x_2 = 2$ is a global maximum because the function described by Equation 2.20 is concave (see our discussion later in this chapter).

IMPLICIT FUNCTIONS

Although mathematical equations are often written with a "dependent" variable (y) as a function of one or more independent variables (x), this is not the only way to write such a relationship. As a trivial example, the equation

$$y = mx + b \tag{2.22}$$

can also be written as

$$y - mx - b = 0 \tag{2.23}$$

or, even more generally, as

$$f(x, y, m, b) = 0, \tag{2.24}$$

where this functional notation indicates a relationship between x and y that also depends on the slope (m) and intercept (b) parameters of the function, which do not change. Functions written in these forms are sometimes called *implicit functions* because the relationships between the variables and parameters are implicitly present in the equation rather than being explicitly calculated as, say, y as a function of x and the parameters m and b.

Often it is a simple matter to translate from implicit functions to explicit ones. For example, the implicit function

$$x + 2y - 4 = 0 \tag{2.25}$$

can easily be "solved" for x as

$$x = -2y + 4 \tag{2.26}$$

or for y as

$$y = \frac{-x}{2} + 2. \tag{2.27}$$

Derivatives from implicit functions

In many circumstances it is helpful to compute derivatives directly from implicit functions without solving for one of the variables directly. For example, the implicit function $f(x, y) = 0$ has a total differential of $0 = f_x dx + f_y dy$, so

$$\frac{dy}{dx} = -\frac{f_x}{f_y}. \tag{2.28}$$

Hence, the implicit derivative dy/dx can be found as the negative of the ratio of the partial derivatives of the implicit function, providing $f_y \neq 0$.

EXAMPLE 2.4 A Production Possibility Frontier—Again

In Example 1.3 we examined a production possibility frontier for two goods of the form

$$x^2 + 0.25y^2 = 200 \tag{2.29}$$

or, written implicitly,

$$f(x, y) = x^2 + 0.25y^2 - 200 = 0. \tag{2.30}$$

Hence,

(*continued*)

EXAMPLE 2.4 CONTINUED

$$f_x = 2x,$$
$$f_y = 0.5y,$$

and, by Equation 2.28, the opportunity cost trade-off between x and y is

$$\frac{dy}{dx} = \frac{-f_x}{f_y} = \frac{-2x}{0.5y} = \frac{-4x}{y}, \tag{2.31}$$

which is precisely the result we obtained earlier, with considerably less work.

QUERY: Why does the trade-off between x and y here depend only on the ratio of x to y and not on the size of the labor force as reflected by the 200 constant?

Implicit function theorem

It may not always be possible to solve implicit functions of the form $g(x, y) = 0$ for unique explicit functions of the form $y = f(x)$. Mathematicians have analyzed the conditions under which a given implicit function can be solved explicitly with one variable being a function of other variables and various parameters. Although we will not investigate these conditions here, they involve requirements on the various partial derivatives of the function that are sufficient to ensure that there is indeed a unique relationship between the dependent and independent variables.[5] In many economic applications, these derivative conditions are precisely those required to ensure that the second-order conditions for a maximum (or a minimum) hold. Hence, in these cases, we will assert that the *implicit function theorem* holds and that it is therefore possible to solve explicitly for trade-offs among the variables involved.

THE ENVELOPE THEOREM

One major application of the implicit function theorem, which will be used many times in this book, is called the *envelope theorem*; it concerns how the optimal value for a particular function changes when a parameter of the function changes. Because many of the economic problems we will be studying concern the effects of changing a parameter (for example, the effects that changing the market price of a commodity will have on an individual's purchases), this is a type of calculation we will frequently make. The envelope theorem often provides a nice shortcut.

A specific example

Perhaps the easiest way to understand the envelope theorem is through an example. Suppose y is a function of a single variable (x) and a parameter (a) given by

$$y = -x^2 + ax. \tag{2.32}$$

For different values of the parameter a, this function represents a family of inverted parabolas. If a is assigned a specific value, Equation 2.32 is a function of x only, and the value of x that maximizes y can be calculated. For example, if $a = 1$, then $x^* = \frac{1}{2}$ and, for these values of x and a, $y = \frac{1}{4}$ (its maximal value). Similarly, if $a = 2$, then $x^* = 1$ and $y^* = 1$. Hence, an increase

[5] For a detailed discussion of the implicit function theorem in various contexts, see Carl P. Simon and Lawrence Blume, *Mathematics for Economists* (New York: W. W. Norton, 1994), chap. 15.

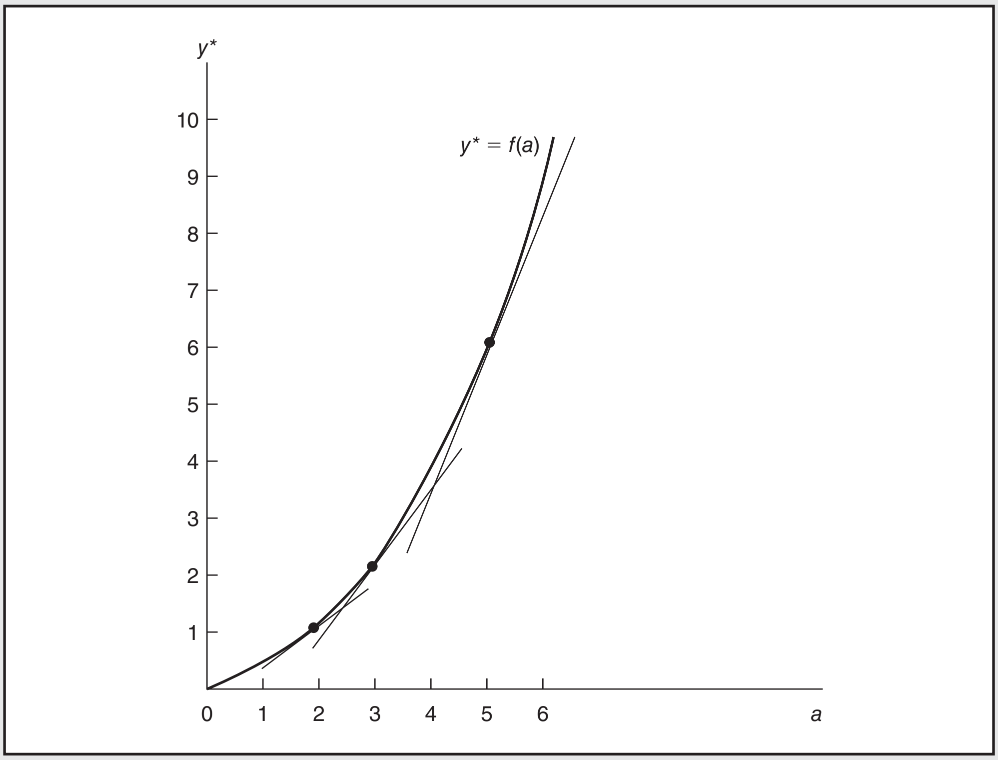
of 1 in the value of the parameter a has increased the maximum value of y by $\frac{3}{4}$. In Table 2.1, integral values of a between 0 and 6 are used to calculate the optimal values for x and the associated values of the objective, y. Notice that as a increases, the maximal value for y also increases. This is also illustrated in Figure 2.3, which shows that the relationship between a and y^* is quadratic. Now we wish to calculate explicitly how y^* changes as the parameter a changes.

TABLE 2.1 Optimal Values of y and x for Alternative Values of a in $y = -x^2 + ax$

Value of a	Value of x^*	Value of y^*
0	0	0
1	$\frac{1}{2}$	$\frac{1}{4}$
2	1	1
3	$\frac{3}{2}$	$\frac{9}{4}$
4	2	4
5	$\frac{5}{2}$	$\frac{25}{4}$
6	3	9

FIGURE 2.3 Illustration of the Envelope Theorem

The envelope theorem states that the slope of the relationship between y^* (the maximum value of y) and the parameter a can be found by calculating the slope of the auxiliary relationship found by substituting the respective optimal values for x into the objective function and calculating $\partial y/\partial a$.



A direct, time-consuming approach

The envelope theorem states that there are two equivalent ways we can make this calculation. First, we can calculate the slope of the function in Figure 2.3 directly. To do so, we must solve Equation 2.32 for the optimal value of x for any value of a:

$$\frac{dy}{dx} = -2x + a = 0;$$

hence,

$$x^* = \frac{a}{2}.$$

Substituting this value of x^* in Equation 2.32 gives

$$\begin{aligned} y^* &= -(x^*)^2 + a(x^*) \\ &= -\left(\frac{a}{2}\right)^2 + a\left(\frac{a}{2}\right) \\ &= -\frac{a^2}{4} + \frac{a^2}{2} = \frac{a^2}{4}, \end{aligned}$$

and this is precisely the relationship shown in Figure 2.3. From the previous equation, it is easy to see that

$$\frac{dy^*}{da} = \frac{2a}{4} = \frac{a}{2} \qquad \textbf{(2.33)}$$

and, for example, at $a = 2$, $dy^*/da = 1$. That is, near $a = 2$ the marginal impact of increasing a is to increase y^* by the same amount. Near $a = 6$, any small increase in a will increase y^* by three times this change. Table 2.1 illustrates this result.

The envelope shortcut

Arriving at this conclusion was a bit complicated. We had to find the optimal value of x for each value of a and then substitute this value for x^* into the equation for y. In more general cases this may be quite burdensome since it requires repeatedly maximizing the objective function. The envelope theorem, providing an alternative approach, states that for small changes in a, dy^*/da can be computed by holding x constant *at its optimal value* and simply calculating $\partial y/\partial a$ from the objective function directly.

Proceeding in this way gives

$$\frac{\partial y}{\partial a} = x, \qquad \textbf{(2.34)}$$

and at x^* we have

$$\frac{\partial y^*}{\partial a} = x^* = \frac{a}{2}. \qquad \textbf{(2.35)}$$

This is precisely the result obtained earlier. The reason that the two approaches yield identical results is illustrated in Figure 2.3. The tangents shown in the figure report values of y for a *fixed* x^*. The tangents' slopes are $\partial y/\partial a$. Clearly, at y^* this slope gives the value we seek.

This result is quite general, and we will use it at several places in this book to simplify our analysis. To summarize, the envelope theorem states that the change in the optimal value of a function with respect to a parameter of that function can be found by partially differentiating the objective function while holding x constant at its optimal value. That is,

$$\frac{dy^*}{da} = \frac{\partial y}{\partial a}\{x = x^*(a)\}, \tag{2.36}$$

where the notation provides a reminder that $\partial y/\partial a$ must be computed at that value of x that is optimal for the specific value of the parameter a being examined.

Many-variable case

An analogous envelope theorem holds for the case where y is a function of several variables. Suppose that y depends on a set of x's $(x_1, \ldots, x_n)$ and on a particular parameter of interest, say, a:

$$y = f(x_1, \ldots, x_n, a). \tag{2.37}$$

Finding an optimal value for y would consist of solving n first-order equations of the form

$$\frac{\partial y}{\partial x_i} = 0 \quad (i = 1, \ldots, n), \tag{2.38}$$

and a solution to this process would yield optimal values for these x's $\left(x_1^*, x_2^*, \ldots, x_n^*\right)$ that would implicitly depend on the parameter a. Assuming the second-order conditions are met, the implicit function theorem would apply in this case and ensure that we could solve each x_i^* as a function of the parameter a:

$$\begin{aligned} x_1^* &= x_1^*(a), \\ x_2^* &= x_2^*(a), \\ &\vdots \\ x_n^* &= x_n^*(a). \end{aligned} \tag{2.39}$$

Substituting these functions into our original objective (Equation 2.37) yields an expression in which the optimal value of y (say, y^*) depends on the parameter a both directly and indirectly through the effect of a on the x^*'s:

$$y^* = f[x_1^*(a), x_2^*(a), \ldots, x_n^*(a), a].$$

Totally differentiating this expression with respect to a yields

$$\frac{dy^*}{da} = \frac{\partial f}{\partial x_1} \cdot \frac{dx_1}{da} + \frac{\partial f}{\partial x_2} \cdot \frac{dx_2}{da} + \cdots + \frac{\partial f}{\partial x_n} \cdot \frac{dx_n}{da} + \frac{\partial f}{\partial a}. \tag{2.40}$$

But, because of the first-order conditions all of these terms except the last are equal to 0 if the x's are at their optimal values. Hence, again we have the envelope result:

$$\frac{dy^*}{da} = \frac{\partial f}{\partial a}, \tag{2.41}$$

where this derivative is to be evaluated at the optimal values for the x's.

EXAMPLE 2.5 The Envelope Theorem: Health Status Revisited

Earlier, in Example 2.3, we examined the maximum values for the health status function

$$y = -(x_1 - 1)^2 - (x_2 - 2)^2 + 10 \tag{2.42}$$

and found that

$$\begin{aligned} x_1^* &= 1, \\ x_2^* &= 2, \end{aligned} \tag{2.43}$$

(continued)

EXAMPLE 2.5 CONTINUED

and

$$y^* = 10.$$

Suppose now we use the arbitrary parameter a instead of the constant 10 in Equation 2.42. Here a might represent a measure of the best possible health for a person, but this value would obviously vary from person to person. Hence,

$$y = f(x_1, x_2, a) = -(x_1 - 1)^2 - (x_2 - 2)^2 + a. \tag{2.44}$$

In this case the optimal values for x_1 and x_2 do not depend on a (they are always $x_1^* = 1$, $x_2^* = 2$), so at those optimal values we have

$$y^* = a \tag{2.45}$$

and

$$\frac{dy^*}{da} = 1. \tag{2.46}$$

People with "naturally better health" will have concomitantly higher values for y^*, providing they choose x_1 and x_2 optimally. But this is precisely what the envelope theorem indicates, because

$$\frac{dy^*}{da} = \frac{\partial f}{\partial a} = 1 \tag{2.47}$$

from Equation 2.44. Increasing the parameter a simply increases the optimal value for y^* by an identical amount (again, assuming the dosages of x_1 and x_2 are correctly chosen).

QUERY: Suppose we focused instead on the optimal dosage for x_1 in Equation 2.42—that is, suppose we used a general parameter, say b, instead of 1. Explain in words and using mathematics why $\partial y^*/\partial b$ would necessarily be 0 in this case.

CONSTRAINED MAXIMIZATION

So far we have focused our attention on finding the maximum value of a function without restricting the choices of the x's available. In most economic problems, however, not all values for the x's are feasible. In many situations, for example, it is required that all the x's be positive. This would be true for the problem faced by the manager choosing output to maximize profits; a negative output would have no meaning. In other instances the x's may be constrained by economic considerations. For example, in choosing the items to consume, an individual is not able to choose any quantities desired. Rather, choices are constrained by the amount of purchasing power available; that is, by this person's budget constraint. Such constraints may lower the maximum value for the function being maximized. Because we are not able to choose freely among all the x's, y may not be as large as it could be. The constraints would be "nonbinding" if we could obtain the same level of y with or without imposing the constraint.

Lagrangian multiplier method

One method for solving constrained maximization problems is the *Lagrangian multiplier method*, which involves a clever mathematical trick that also turns out to have a useful economic interpretation. The rationale of this method is quite simple, although no rigorous presentation will be attempted here.[6] In a prior section, the necessary conditions for a local maximum were discussed. We showed that at the optimal point all the partial derivatives of f must be 0. There are therefore n equations ($f_i = 0$ for $i = 1, \ldots, n$) in n unknowns (the x's). Generally, these equations can be solved for the optimal x's. When the x's are constrained, however, there is at least one additional equation (the constraint) but no additional variables. The set of equations therefore is overdetermined. The Lagrangian technique introduces an additional variable (the Lagrangian multiplier), which not only helps to solve the problem at hand (because there are now $n+1$ equations in $n+1$ unknowns), but also has an interpretation that is useful in a variety of economic circumstances.

The formal problem

More specifically, suppose that we wish to find the values of $x_1, x_2, \ldots, x_n$ that maximize

$$y = f(x_1, x_2, \ldots, x_n), \tag{2.48}$$

subject to a constraint that permits only certain values of the x's to be used. A general way of writing that constraint is

$$g(x_1, x_2, \ldots, x_n) = 0, \tag{2.49}$$

where the function[7] g represents the relationship that must hold among all the x's.

First-order conditions

The Lagrangian multiplier method starts with setting up the expression

$$\mathscr{L} = f(x_1, x_2, \ldots, x_n) + \lambda g(x_1, x_2, \ldots, x_n), \tag{2.50}$$

where λ is an additional variable called the Lagrangian multiplier. Later we will interpret this new variable. First, however, notice that when the constraint holds, $\mathscr{L}$ and f have the same value [because $g(x_1, x_2, \ldots, x_n) = 0$]. Consequently, if we restrict our attention only to values of the x's that satisfy the constraint, finding the constrained maximum value of f is equivalent to finding a critical value of $\mathscr{L}$. Let us proceed then to do so, treating λ also as a variable (in addition to the x's). From Equation 2.50, the conditions for a critical point are:

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial x_1} &= f_1 + \lambda g_1 = 0, \\
\frac{\partial \mathscr{L}}{\partial x_2} &= f_2 + \lambda g_2 = 0, \\
&\vdots \\
\frac{\partial \mathscr{L}}{\partial x_n} &= f_n + \lambda g_n = 0, \\
\frac{\partial \mathscr{L}}{\partial \lambda} &= g(x_1, x_2, \ldots, x_n) = 0.
\end{aligned} \tag{2.51}$$

Equations 2.51 are then the conditions for a critical point for the function $\mathscr{L}$. Notice that there are $n+1$ equations (one for each x and a final one for λ) in $n+1$ unknowns. The equations can generally be solved for $x_1, x_2, \ldots, x_n$, and λ. Such a solution will have two

[6]For a detailed presentation, see A. K. Dixit, *Optimization in Economic Theory*, 2nd ed. (Oxford: Oxford University Press, 1990), chap. 2.

[7]As we pointed out earlier, any function of $x_1, x_2, \ldots, x_n$ can be written in this implicit way. For example, the constraint $x_1 + x_2 = 10$ could be written $10 - x_1 - x_2 = 0$. In later chapters, we will usually follow this procedure in dealing with constraints. Often the constraints we examine will be linear.

properties: (1) the x's will obey the constraint because the last equation in 2.51 imposes that condition; and (2) among all those values of x's that satisfy the constraint, those that also solve Equations 2.51 will make $\mathscr{L}$ (and hence f) as large as possible (assuming second-order conditions are met). The Lagrangian multiplier method therefore provides a way to find a solution to the constrained maximization problem we posed at the outset.[8]

The solution to Equations 2.51 will usually differ from that in the unconstrained case (see Equations 2.19). Rather than proceeding to the point where the marginal contribution of each x is 0, Equations 2.51 require us to stop short because of the constraint. Only if the constraint were ineffective (in which case, as we show below, λ would be 0) would the constrained and unconstrained equations (and their respective solutions) agree. These revised marginal conditions have economic interpretations in many different situations.

Interpretation of the Lagrangian multiplier

So far we have used the Lagrangian multiplier (λ) only as a mathematical "trick" to arrive at the solution we wanted. In fact, that variable also has an important economic interpretation, which will be central to our analysis at many points in this book. To develop this interpretation, rewrite the first n equations of 2.51 as

$$\frac{f_1}{-g_1} = \frac{f_2}{-g_2} = \cdots = \frac{f_n}{-g_n} = \lambda. \tag{2.52}$$

In other words, at the maximum point, the ratio of f_i to g_i is the same for every x_i. The numerators in Equations 2.52 are the marginal contributions of each x to the function f. They show the *marginal benefit* that one more unit of x_i will have for the function that is being maximized (that is, for f).

A complete interpretation of the denominators in Equations 2.52 is probably best left until we encounter these ratios in actual economic applications. There we will see that these usually have a "marginal cost" interpretation. That is, they reflect the added burden on the constraint of using slightly more x_i. As a simple illustration, suppose the constraint required that total spending on x_1 and x_2 be given by a fixed dollar amount, F. Hence, the constraint would be $p_1x_1 + p_2x_2 = F$ (where p_i is the per unit cost of x_i). Using our present terminology, this constraint would be written in implicit form as

$$g(x_1, x_2) = F - p_1x_1 - p_2x_2 = 0. \tag{2.53}$$

In this situation, then,

$$-g_i = p_i \tag{2.54}$$

and the derivative $-g_i$ does indeed reflect the per unit, marginal cost of using x_i. Practically all of the optimization problems we will encounter in later chapters have a similar interpretation for the denominators in Equations 2.52.

Lagrangian multiplier as a benefit-cost ratio

Now we can give Equations 2.52 an intuitive interpretation. They indicate that, at the optimal choices for the x's, the ratio of the marginal benefit of increasing x_i to the marginal cost of increasing x_i should be the same for every x. To see that this is an obvious condition

[8]Strictly speaking, these are the necessary conditions for an interior local maximum. In some economic problems, it is necessary to amend these conditions (in fairly obvious ways) to take account of the possibility that some of the x's may be on the boundary of the region of permissible x's. For example, if all of the x's are required to be nonnegative, it may be that the conditions of Equations 2.51 will not hold exactly, because these may require negative x's. We look at this situation later in this chapter.

for a maximum, suppose that it were not true: Suppose that the "benefit-cost ratio" were higher for x_1 than for x_2. In this case, slightly more x_1 should be used in order to achieve a maximum. Consider using more x_1 but giving up just enough x_2 to keep g (the constraint) constant. Hence, the marginal cost of the additional x_1 used would equal the cost saved by using less x_2. But because the benefit-cost ratio (the amount of benefit per unit of cost) is greater for x_1 than for x_2, the additional benefits from using more x_1 would exceed the loss in benefits from using less x_2. The use of more x_1 and appropriately less x_2 would then increase y because x_1 provides more "bang for your buck." Only if the marginal benefit–marginal cost ratios are equal for all the x's will there be a local maximum, one in which no small changes in the x's can increase the objective. Concrete applications of this basic principle are developed in many places in this book. The result is fundamental for the microeconomic theory of optimizing behavior.

The Lagrangian multiplier (λ) can also be interpreted in light of this discussion. λ is the common benefit-cost ratio for all the x's. That is,

$$\lambda = \frac{\textbf{marginal benefit of } \boldsymbol{x_i}}{\textbf{marginal cost of } \boldsymbol{x_i}} \tag{2.55}$$

for every x_i. If the constraint were relaxed slightly, it would not matter exactly which x is changed (indeed, all the x's could be altered), because, at the margin, each promises the same ratio of benefits to costs. The Lagrangian multiplier then provides a measure of how such an overall relaxation of the constraint would affect the value of y. In essence, λ assigns a "shadow price" to the constraint. A high λ indicates that y could be increased substantially by relaxing the constraint, because each x has a high benefit-cost ratio. A low value of λ, on the other hand, indicates that there is not much to be gained by relaxing the constraint. If the constraint is not binding at all, λ will have a value of 0, thereby indicating that the constraint is not restricting the value of y. In such a case, finding the maximum value of y subject to the constraint would be identical to finding an unconstrained maximum. The shadow price of the constraint is 0. This interpretation of λ can also be shown using the envelope theorem as described later in this chapter.[9]

Duality

This discussion shows that there is a clear relationship between the problem of maximizing a function subject to constraints and the problem of assigning values to constraints. This reflects what is called the mathematical principle of "duality": Any constrained maximization problem has an associated dual problem in constrained *minimization* that focuses attention on the constraints in the original (primal) problem. For example, to jump a bit ahead of our story, economists assume that individuals maximize their utility, subject to a budget constraint. This is the consumer's primal problem. The dual problem for the consumer is to minimize the expenditure needed to achieve a given level of utility. Or, a firm's primal problem may be to minimize the total cost of inputs used to produce a given level of output, whereas the dual problem is to maximize output for a given cost of inputs purchased. Many similar examples will be developed in later chapters. Each illustrates that there are always two ways to look at any constrained optimization problem. Sometimes taking a frontal attack by analyzing the primal problem can lead to greater insights. In other instances, the "back door" approach of examining the dual problem may be more instructive. Whichever route is taken, the results will generally, though not always, be identical, so the choice made will mainly be a matter of convenience.

[9]The discussion in the text concerns problems involving a single constraint. In general, one can handle m constraints ($m < n$) by simply introducing m new variables (Lagrangian multipliers) and proceeding in an analogous way to that discussed above.

EXAMPLE 2.6 Constrained Maximization: Health Status Yet Again

Let's return once more to our (perhaps tedious) health maximization problem. As before, the individual's goal is to maximize

$$y = -x_1^2 + 2x_1 - x_2^2 + 4x_2 + 5,$$

but now assume that choices of x_1 and x_2 are constrained by the fact that he or she can only tolerate one drug dose per day. That is,

$$x_1 + x_2 = 1 \tag{2.56}$$

or

$$1 - x_1 - x_2 = 0.$$

Notice that the original optimal point ($x_1 = 1, x_2 = 2$) is no longer attainable because of the constraint on possible dosages: other values must be found. To do so, we first set up the Lagrangian expression:

$$\mathcal{L} = -x_1^2 + 2x_1 - x_2^2 + 4x_2 + 5 + \lambda(1 - x_1 - x_2). \tag{2.57}$$

Differentiation of $\mathcal{L}$ with respect to x_1, x_2, and λ yields the following necessary condition for a constrained maximum:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_1} &= -2x_1 + 2 - \lambda = 0, \\
\frac{\partial \mathcal{L}}{\partial x_2} &= -2x_2 + 4 - \lambda = 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= 1 - x_1 - x_2 = 0.
\end{aligned} \tag{2.58}$$

These equations must now be solved for the optimal values of x_1, x_2, and λ. Using the first and second equations gives

$$-2x_1 + 2 = \lambda = -2x_2 + 4$$

or

$$x_1 = x_2 - 1. \tag{2.59}$$

Substitution of this value for x_1 into the constraint yields the solution:

$$\begin{aligned} x_2 &= 1, \\ x_1 &= 0. \end{aligned} \tag{2.60}$$

In words, if this person can tolerate only one dose of drugs, he or she should opt for taking only the second drug. By using either of the first two equations, it is easy to complete our solution by showing that

$$\lambda = 2. \tag{2.61}$$

This, then, is the solution to the constrained maximum problem. If $x_1 = 0$, $x_2 = 1$, then y takes on the value 8. Constraining the values of x_1 and x_2 to sum to 1 has reduced the maximum value of health status, y, from 10 to 8.

QUERY: Suppose this individual could tolerate two doses per day. Would you expect y to increase? Would increases in tolerance beyond three doses per day have any effect on y?

EXAMPLE 2.7 Optimal Fences and Constrained Maximization

Suppose a farmer had a certain length of fence, P, and wished to enclose the largest possible rectangular area. What shape area should the farmer choose? This is clearly a problem in constrained maximization. To solve it, let x be the length of one side of the rectangle and y be the length of the other side. The problem then is to choose x and y so as to maximize the area of the field (given by $A = x \cdot y$), subject to the constraint that the perimeter is fixed at $P = 2x + 2y$.

Setting up the Lagrangian expression gives

$$\mathscr{L} = x \cdot y + \lambda(P - 2x - 2y), \tag{2.62}$$

where λ is an unknown Lagrangian multiplier. The first-order conditions for a maximum are

$$\begin{aligned} \frac{\partial \mathscr{L}}{\partial x} &= y - 2\lambda = 0, \\ \frac{\partial \mathscr{L}}{\partial y} &= x - 2\lambda = 0, \\ \frac{\partial \mathscr{L}}{\partial \lambda} &= P - 2x - 2y = 0. \end{aligned} \tag{2.63}$$

The three equations in 2.63 must be solved simultaneously for x, y, and λ. The first two equations say that $y/2 = x/2 = \lambda$, showing that x must be equal to y (the field should be square). They also imply that x and y should be chosen so that the ratio of marginal benefits to marginal cost is the same for both variables. The benefit (in terms of area) of one more unit of x is given by y (area is increased by $1 \cdot y$), and the marginal cost (in terms of perimeter) is 2 (the available perimeter is reduced by 2 for each unit that the length of side x is increased). The maximum conditions state that this ratio should be equal for each of the variables.

Since we have shown that $x = y$, we can use the constraint to show that

$$x = y = \frac{P}{4}, \tag{2.64}$$

and, because $y = 2\lambda$,

$$\lambda = \frac{P}{8}. \tag{2.65}$$

Interpretation of the Lagrangian Multiplier. If the farmer were interested in knowing how much more field could be fenced by adding an extra yard of fence, the Lagrangian multiplier suggests that he or she could find out by dividing the present perimeter by 8. Some specific numbers might make this clear. Suppose that the field currently has a perimeter of 400 yards. If the farmer has planned "optimally," the field will be a square with 100 yards ($= P/4$) on a side. The enclosed area will be 10,000 square yards. Suppose now that the perimeter (that is, the available fence) were enlarged by one yard. Equation 2.65 would then "predict" that the total area would be increased by approximately 50 ($= P/8$) square yards. That this is indeed the case can be shown as follows: Because the perimeter is now 401 yards, each side of the square will be $401/4$ yards. The total area of the field is therefore $(401/4)^2$, which, according to the author's calculator, works out to be 10,050.06 square yards. Hence, the "prediction" of a 50-square-yard increase that is provided by the Lagrangian multiplier proves to be remarkably close. As in all constrained maximization problems, here the Lagrangian multiplier provides useful information about the implicit value of the constraint.

(*continued*)

EXAMPLE 2.7 CONTINUED

Duality. The dual of this constrained maximization problem is that for a given area of a rectangular field, the farmer wishes to minimize the fence required to surround it. Mathematically, the problem is to minimize

$$P = 2x + 2y, \tag{2.66}$$

subject to the constraint

$$A = x \cdot y. \tag{2.67}$$

Setting up the Lagrangian expression

$$\mathscr{L}^D = 2x + 2y + \lambda^D(A - x \cdot y) \tag{2.68}$$

(where the D denotes the dual concept) yields the following first-order conditions for a minimum:

$$\begin{aligned}
\frac{\partial \mathscr{L}^D}{\partial x} &= 2 - \lambda^D \cdot y = 0,\\
\frac{\partial \mathscr{L}^D}{\partial y} &= 2 - \lambda^D \cdot x = 0,\\
\frac{\partial \mathscr{L}^D}{\partial \lambda^D} &= A - x \cdot y = 0.
\end{aligned} \tag{2.69}$$

Solving these equations as before yields the result

$$x = y = \sqrt{A}. \tag{2.70}$$

Again, the field should be square if the length of fence is to be minimized. The value of the Lagrangian multiplier in this problem is

$$\lambda^D = \frac{2}{y} = \frac{2}{x} = \frac{2}{\sqrt{A}}. \tag{2.71}$$

As before, this Lagrangian multiplier indicates the relationship between the objective (minimizing fence) and the constraint (needing to surround the field). If the field were 10,000 square yards, as we saw before, 400 yards of fence would be needed. Increasing the field by one square yard would require about .02 more yards of fence ($= 2/\sqrt{A} = 2/100$). The reader may wish to fire up his or her calculator to show this is indeed the case—a fence 100.005 yards on each side will exactly enclose 10,001 square yards. Here, as in most duality problems, the value of the Lagrangian in the dual is the reciprocal of the value for the Lagrangian in the primal problem. Both provide the same information, although in a somewhat different form.

QUERY: An implicit constraint here is that the farmer's field be rectangular. If this constraint were not imposed, what shape field would enclose maximal area? How would you prove that?

ENVELOPE THEOREM IN CONSTRAINED MAXIMIZATION PROBLEMS

The envelope theorem, which we discussed previously in connection with unconstrained maximization problems, also has important applications in constrained maximization problems. Here we will provide only a brief presentation of the theorem. In later chapters we will look at a number of applications.

Suppose we seek the maximum value of

$$y = f(x_1, \ldots, x_n; a), \tag{2.72}$$

subject to the constraint

$$g(x_1, \ldots, x_n; a) = 0, \tag{2.73}$$

where we have made explicit the dependence of the functions f and g on some parameter a. As we have shown, one way to solve this problem is to set up the Lagrangian expression

$$\mathcal{L} = f(x_1, \ldots, x_n; a) + \lambda g(x_1, \ldots, x_n; a) \tag{2.74}$$

and solve the first-order conditions (see Equations 2.51) for the optimal, constrained values $x_1^*, \ldots, x_n^*$. Alternatively, it can be shown that

$$\frac{dy^*}{da} = \frac{\partial \mathcal{L}}{\partial a}(x_1^*, \ldots, x_n^*; a). \tag{2.75}$$

That is, the change in the maximal value of y that results when the parameter a changes (and all of the x's are recalculated to new optimal values) can be found by partially differentiating the Lagrangian expression (Equation 2.74) and evaluating the resultant partial derivative at the optimal point.[10] Hence, the Lagrangian expression plays the same role in applying the envelope theorem to constrained problems as does the objective function alone in unconstrained problems. As a simple exercise, the reader may wish to show that this result holds for the problem of fencing a rectangular field described in Example 2.7.[11]

INEQUALITY CONSTRAINTS

In some economic problems the constraints need not hold exactly. For example, an individual's budget constraint requires that he or she spend no more than a certain amount per period, but it is at least possible to spend less than this amount. Inequality constraints also arise in the values permitted for some variables in economic problems. Usually, for example, economic variables must be nonnegative (though they can take on the value of zero). In this section we will show how the Lagrangian technique can be adapted to such circumstances. Although we will encounter only a few problems later in the text that require this mathematics, development here will illustrate a few general principles that are quite consistent with economic intuition.

A two-variable example

In order to avoid much cumbersome notation, we will explore inequality constraints only for the simple case involving two choice variables. The results derived are readily generalized. Suppose that we seek to maximize $y = f(x_1, x_2)$ subject to three inequality constraints:

[10]For a more complete discussion of the envelope theorem in constrained maximization problems, see Eugene Silberberg and Wing Suen, *The Structure of Economics: A Mathematical Analysis,* 3rd ed. (Boston: Irwin/McGraw-Hill, 2001), pp. 159–61.

[11]For the primal problem, the perimeter P is the parameter of principal interest. By solving for the optimal values of x and y and substituting into the expression for the area (A) of the field, it is easy to show that $dA/dP = P/8$. Differentiation of the Lagrangian expression (Equation 2.62) yields $\partial\mathcal{L}/\partial P = \lambda$ and, at the optimal values of x and y, $dA/dP = \partial\mathcal{L}/\partial P = \lambda = P/8$. The envelope theorem in this case then offers further proof that the Lagrangian multiplier can be used to assign an implicit value to the constraint.

$$\begin{aligned}
&1.\ g(x_1, x_2) \geq 0; \\
&2.\ x_1 \geq 0; \quad \text{and} \\
&3.\ x_2 \geq 0.
\end{aligned} \tag{2.76}$$

Hence, we are allowing for the possibility that the constraint we introduced before need not hold exactly (a person need not spend all of his or her income) and for the fact that both of the x's must be nonnegative (as in most economic problems).

Slack variables

One way to solve this optimization problem is to introduce three new variables (a, b, and c) that convert the inequality constraints in Equation 2.76 into equalities. To ensure that the inequalities continue to hold, we will square these new variables, ensuring that the resulting values are positive. Using this procedure, the inequality constraints become

$$\begin{aligned}
&1.\ g(x_1, x_2) - a^2 = 0; \\
&2.\ x_1 - b^2 = 0; \quad \text{and} \\
&3.\ x_2 - c^2 = 0.
\end{aligned} \tag{2.77}$$

Any solution that obeys these three equality constraints will also obey the inequality constraints. It will also turn out that the optimal values for a, b, and c will provide several insights into the nature of the solutions to a problem of this type.

Solution by the method of Lagrange

By converting the original problem involving inequalities into one involving equalities, we are now in a position to use Lagrangian methods to solve it. Because there are three constraints, we must introduce three Lagrangian multipliers: λ_1, λ_2, and λ_3. The full Lagrangian expression is

$$\mathscr{L} = f(x_1, x_2) + \lambda_1[g(x_1, x_2) - a^2] + \lambda_2(x_1 - b^2) + \lambda_3(x_2 - c^2). \tag{2.78}$$

We wish to find the values of x_1, x_2, a, b, c, λ_1, λ_2, and λ_3 that constitute a critical point for this expression. This will necessitate eight first-order conditions:

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial x_1} &= f_1 + \lambda_1 g_1 + \lambda_2 = 0, \\
\frac{\partial \mathscr{L}}{\partial x_2} &= f_2 + \lambda_1 g_2 + \lambda_3 = 0, \\
\frac{\partial \mathscr{L}}{\partial a} &= -2a\lambda_1 = 0, \\
\frac{\partial \mathscr{L}}{\partial b} &= -2b\lambda_2 = 0, \\
\frac{\partial \mathscr{L}}{\partial c} &= -2c\lambda_3 = 0, \\
\frac{\partial \mathscr{L}}{\partial \lambda_1} &= g(x_1, x_2) - a^2 = 0, \\
\frac{\partial \mathscr{L}}{\partial \lambda_2} &= x_1 - b^2 = 0, \\
\frac{\partial \mathscr{L}}{\partial \lambda_3} &= x_2 - c^2 = 0,
\end{aligned} \tag{2.79}$$

In many ways these conditions resemble those we derived earlier for the case of a single equality constraint (see Equation 2.51). For example, the final three conditions merely repeat the three revised constraints. This ensures that any solution will obey these conditions. The first two equations also resemble the optimal conditions developed earlier. If λ_2 and λ_3 were 0, the conditions would in fact be identical. But the presence of the additional Lagrangian multipliers in the expressions shows that the customary optimality conditions may not hold exactly here.

Complementary slackness

The three equations involving the variables a, b, and c provide the most important insights into the nature of solutions to problems involving inequality constraints. For example, the third line in Equation 2.79 implies that, in the optimal solution, either λ_1 or a must be 0.[12] In the second case ($a = 0$), the constraint $g(x_1, x_2) = 0$ holds exactly and the calculated value of λ_1 indicates its relative importance to the objective function, f. On the other hand, if $a \neq 0$, then $\lambda_1 = 0$ and this shows that the availability of some slackness in the constraint implies that its value to the objective is 0. In the consumer context, this means that if a person does not spend all his or her income, even more income would do nothing to raise his or her well-being.

Similar complementary slackness relationships also hold for the choice variables x_1 and x_2. For example, the fourth line in Equation 2.79 requires that the optimal solution have either b or λ_2 be 0. If $\lambda_2 = 0$ then the optimal solution has $x_1 > 0$, and this choice variable meets the precise benefit-cost test that $f_1 + \lambda_1 g_1 = 0$. Alternatively, solutions where $b = 0$ have $x_1 = 0$, and also require that $\lambda_2 > 0$. So, such solutions do not involve any use of x_1 because that variable does not meet the benefit-cost test as shown by the first line of Equation 2.79, which implies that $f_1 + \lambda_1 g_1 < 0$. An identical result holds for the choice variable x_2.

These results, which are sometimes called *Kuhn-Tucker conditions* after their discoverers, show that the solutions to optimization problems involving inequality constraints will differ from similar problems involving equality constraints in rather simple ways. Hence, we cannot go far wrong by working primarily with constraints involving equalities and assuming that we can rely on intuition to state what would happen if the problems actually involved inequalities. That is the general approach we will take in this book.[13]

SECOND-ORDER CONDITIONS

So far our discussion of optimization has focused primarily on necessary (first-order) conditions for finding a maximum. That is indeed the practice we will follow throughout much of this book because, as we shall see, most economic problems involve functions for which the second-order conditions for a maximum are also satisfied. In this section we give a brief analysis of the connection between second-order conditions for a maximum and the related curvature conditions that functions must have to ensure that these hold. The economic explanations for these curvature conditions will be discussed throughout the text.

Functions of one variable

First consider the case in which the objective, y, is a function of only a single variable, x. That is,

$$y = f(x). \quad \textbf{(2.80)}$$

[12]We will not examine the degenerate case where both of these variables are 0.

[13]The situation can become much more complex when calculus cannot be relied upon to give a solution, perhaps because some of the functions in a problem are not differentiable. For a discussion, see Avinask K. Dixit, *Optimization in Economic Theory,* 2nd ed. (Oxford: Oxford University Press, 1990).

A necessary condition for this function to attain its maximum value at some point is that

$$\frac{dy}{dx} = f'(x) = 0 \tag{2.81}$$

at that point. To ensure that the point is indeed a maximum, we must have y decreasing for movements away from it. We already know (by Equation 2.81) that for small changes in x, the value of y does not change; what we need to check is whether y is increasing before that "plateau" is reached and declining thereafter. We have already derived an expression for the change in $y(dy)$, which is given by the total differential

$$dy = f'(x)dx. \tag{2.82}$$

What we now require is that dy be decreasing for small increases in the value of x. The differential of Equation 2.82 is given by

$$d(dy) = d^2y = \frac{d[f'(x)dx]}{dx} \cdot dx = f''(x)dx \cdot dx = f''(x)dx^2. \tag{2.83}$$

But

$$d^2y < 0$$

implies that

$$f''(x)dx^2 < 0, \tag{2.84}$$

and since dx^2 must be positive (because anything squared is positive), we have

$$f''(x) < 0 \tag{2.85}$$

as the required second-order condition. In words, this condition requires that the function f have a concave shape at the critical point (contrast Figures 2.1 and 2.2). Similar curvature conditions will be encountered throughout this section.

EXAMPLE 2.8 Profit Maximization Again

In Example 2.1 we considered the problem of finding the maximum of the function

$$\pi = 1{,}000q - 5q^2. \tag{2.86}$$

The first-order condition for a maximum requires

$$\frac{d\pi}{dq} = 1{,}000 - 10q = 0 \tag{2.87}$$

or

$$q^* = 100. \tag{2.88}$$

The second derivative of the function is given by

$$\frac{d^2\pi}{dq^2} = -10 < 0, \tag{2.89}$$

and hence the point $q^* = 100$ obeys the sufficient conditions for a local maximum.

QUERY: Here the second derivative is negative not only at the optimal point; it is always negative. What does that imply about the optimal point? How should the fact that the second derivative is a constant be interpreted?

Functions of two variables

As a second case, we consider y as a function of two independent variables:

$$y = f(x_1, x_2). \tag{2.90}$$

A necessary condition for such a function to attain its maximum value is that its partial derivatives, in both the x_1 and the x_2 directions, be 0. That is,

$$\begin{aligned} \frac{\partial y}{\partial x_1} &= f_1 = \mathbf{0}, \\ \frac{\partial y}{\partial x_2} &= f_2 = \mathbf{0}. \end{aligned} \tag{2.91}$$

A point that satisfies these conditions will be a "flat" spot on the function (a point where $dy = 0$) and therefore will be a candidate for a maximum. To ensure that the point is a local maximum, y must diminish for movements in any direction away from the critical point: In pictorial terms there is only one way to leave a true mountaintop, and that is to go down.

An intuitive argument

Before describing the mathematical properties required of such a point, an intuitive approach may be helpful. If we consider only movements in the x_1 direction, the required condition is clear: The slope in the x_1 direction (that is, the partial derivative f_1) must be diminishing at the critical point. This is a direct application of our discussion of the single-variable case. It shows that, for a maximum, the second partial derivative in the x_1 direction must be negative. An identical argument holds for movements only in the x_2 direction. Hence, both own second partial derivatives (f_{11} and f_{22}) must be negative for a local maximum. In our mountain analogy, if attention is confined only to north-south or east-west movements, the slope of the mountain must be diminishing as we cross its summit—the slope must change from positive to negative.

The particular complexity that arises in the two-variable case involves movements through the optimal point that are not solely in the x_1 or x_2 directions (say, movements from northeast to southwest). In such cases, the second-order partial derivatives do not provide complete information about how the slope is changing near the critical point. Conditions must also be placed on the cross-partial derivative ($f_{12} = f_{21}$) to ensure that dy is decreasing for movements through the critical point in any direction. As we shall see, those conditions amount to requiring that the own second-order partial derivatives be sufficiently negative so as to counterbalance any possible "perverse" cross-partial derivatives that may exist. Intuitively, if the mountain falls away steeply enough in the north-south and east-west directions, relatively minor failures to do so in other directions can be compensated for.

A formal analysis

We now proceed to make these points more formally. What we wish to discover are the conditions that must be placed on the second partial derivatives of the function f to ensure that d^2y is negative for movements in any direction through the critical point. Recall first that the total differential of the function is given by

$$dy = f_1 dx_1 + f_2 dx_2. \tag{2.92}$$

The differential of that function is given by

$$d^2y = (f_{11}dx_1 + f_{12}dx_2)dx_1 + (f_{21}dx_1 + f_{22}dx_2)dx_2 \tag{2.93}$$

or

$$d^2y = f_{11}dx_1^2 + f_{12}dx_2dx_1 + f_{21}dx_1dx_2 + f_{22}dx_2^2. \tag{2.94}$$

Because, by Young's theorem, $f_{12} = f_{21}$, we can arrange terms to get

$$d^2y = f_{11}dx_1^2 + 2f_{12}dx_1dx_2 + f_{22}dx_2^2. \tag{2.95}$$

For Equation 2.95 to be unambiguously negative for any change in the x's (that is, for any choices of dx_1 and dx_2), it is obviously necessary that f_{11} and f_{22} be negative. If, for example, $dx_2 = 0$, then

$$d^2y = f_{11}dx_1^2 \tag{2.96}$$

and $d^2y < 0$ implies

$$f_{11} < 0. \tag{2.97}$$

An identical argument can be made for f_{22} by setting $dx_1 = 0$. If neither dx_1 nor dx_2 is 0, we then must consider the cross partial, f_{12}, in deciding whether or not d^2y is unambiguously negative. Relatively simple algebra can be used to show that the required condition is[14]

$$f_{11}f_{22} - f_{12}^2 > 0. \tag{2.98}$$

Concave functions

Intuitively, what Equation 2.98 requires is that the own second partial derivatives ($f_{11}$and f_{22}) be sufficiently negative so that their product (which is positive) will outweigh any possible perverse effects from the cross-partial derivatives ($f_{12} = f_{21}$). Functions that obey such a condition are called *concave functions.* In three dimensions, such functions resemble inverted teacups (for an illustration, see Example 2.10). This image makes it clear that a flat spot on such a function is indeed a true maximum because the function always slopes downward from such a spot. More generally, concave functions have the property that they always lie below any plane that is tangent to them—the plane defined by the maximum value of the function is simply a special case of this property.

EXAMPLE 2.9 Second-Order Conditions: Health Status for the Last Time

In Example 2.3 we considered the health status function

$$y = f(x_1, x_2) = -x_1^2 + 2x_1 - x_2^2 + 4x_2 + 5. \tag{2.99}$$

The first-order conditions for a maximum are

$$\begin{aligned} f_1 &= -2x_1 + 2 = 0, \\ f_2 &= -2x_2 + 4 = 0 \end{aligned} \tag{2.100}$$

or

$$\begin{aligned} x_1^* &= 1, \\ x_2^* &= 2. \end{aligned} \tag{2.101}$$

---

[14]The proof proceeds by adding and subtracting the term $(f_{12}dx_2)^2/f_{11}$ to Equation 2.95 and factoring. But this approach is only applicable to this special case. A more easily generalized approach that uses matrix algebra recognizes that Equation 2.95 is a "Quadratic Form" in dx_1 and dx_2, and that Equations 2.97 and 2.98 amount to requiring that the Hessian matrix

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$$

be "negative definite." In particular, Equation 2.98 requires that the determinant of this Hessian be positive. For a discussion, see the Extensions to this chapter.

The second-order partial derivatives for Equation 2.99 are

$$\begin{aligned} f_{11} &= -2, \\ f_{22} &= -2, \\ f_{12} &= 0. \end{aligned} \tag{2.102}$$

These derivatives clearly obey Equations 2.97 and 2.98, so both necessary and sufficient conditions for a local maximum are satisfied.[15]

QUERY: Describe the concave shape of the health status function and indicate why it has only a single global maximum value.

Constrained maximization

As another illustration of second-order conditions, consider the problem of choosing x_1 and x_2 to maximize

$$y = f(x_1, x_2), \tag{2.103}$$

subject to the linear constraint

$$c - b_1 x_1 - b_2 x_2 = 0 \tag{2.104}$$

(where c, b_1, b_2 are constant parameters in the problem). This problem is of a type that will be frequently encountered in this book and is a special case of the constrained maximum problems that we examined earlier. There we showed that the first-order conditions for a maximum may be derived by setting up the Lagrangian expression

$$\mathscr{L} = f(x_1, x_2) + \lambda(c - b_1 x_1 - b_2 x_2). \tag{2.105}$$

Partial differentiation with respect to x_1, x_2, and λ yields the familiar results:

$$\begin{aligned} f_1 - \lambda b_1 &= 0, \\ f_2 - \lambda b_2 &= 0, \\ c - b_1 x_1 - b_2 x_2 &= 0. \end{aligned} \tag{2.106}$$

These equations can in general be solved for the optimal values of x_1, x_2, and λ. To ensure that the point derived in that way is a local maximum, we must again examine movements away from the critical points by using the "second" total differential:

$$d^2 y = f_{11} dx_1^2 + 2 f_{12} dx_1 dx_2 + f_{22} dx_2^2. \tag{2.107}$$

In this case, however, not all possible small changes in the x's are permissible. Only those values of x_1 and x_2 that continue to satisfy the constraint can be considered valid alternatives to the critical point. To examine such changes, we must calculate the total differential of the constraint:

$$-b_1 dx_1 - b_2 dx_2 = 0 \tag{2.108}$$

or

$$dx_2 = -\frac{b_1}{b_2} dx_1. \tag{2.109}$$

[15]Notice that Equations 2.102 obey the sufficient conditions not only at the critical point but also for all possible choices of x_1 and x_2. That is, the function is concave. In more complex examples this need not be the case: The second-order conditions need be satisfied only at the critical point for a local maximum to occur.

This equation shows the relative changes in x_1 and x_2 that are allowable in considering movements from the critical point. To proceed further on this problem, we need to use the first-order conditions. The first two of these imply

$$\frac{f_1}{f_2} = \frac{b_1}{b_2}, \tag{2.110}$$

and combining this result with Equation 2.109 yields

$$dx_2 = -\frac{f_1}{f_2}dx_1. \tag{2.111}$$

We now substitute this expression for dx_2 in Equation 2.107 to demonstrate the conditions that must hold for d^2y to be negative:

$$\begin{aligned} d^2y &= f_{11}dx_1^2 + 2f_{12}dx_1\left(-\frac{f_1}{f_2}dx_1\right) + f_{22}\left(-\frac{f_1}{f_2}dx_1\right)^2 \\ &= f_{11}dx_1^2 - 2f_{12}\frac{f_1}{f_2}dx_1^2 + f_{22}\frac{f_1^2}{f_2^2}dx_1^2. \end{aligned} \tag{2.112}$$

Combining terms and putting each over a common denominator gives

$$d^2y = (f_{11}f_2^2 - 2f_{12}f_1f_2 + f_{22}f_1^2)\frac{dx_1^2}{f_2^2}. \tag{2.113}$$

Consequently, for $d^2y < 0$, it must be the case that

$$f_{11}f_2^2 - 2f_{12}f_1f_2 + f_{22}f_1^2 < 0. \tag{2.114}$$

Quasi-concave functions

Although Equation 2.114 appears to be little more than an inordinately complex mass of mathematical symbols, in fact the condition is an important one. It characterizes a set of functions termed *quasi-concave functions*. These functions have the property that the set of all points for which such a function takes on a value greater than any specific constant is a convex set (that is, any two points in the set can be joined by a line contained completely within the set). Many economic models are characterized by such functions and, as we will see in considerable detail in Chapter 3, in these cases the condition for quasi-concavity has a relatively simple economic interpretation. Problems 2.9 and 2.10 examine two specific quasi-concave functions that we will frequently encounter in this book. Example 2.10 shows the relationship between concave and quasi-concave functions.

EXAMPLE 2.10 Concave and Quasi-Concave Functions

The differences between concave and quasi-concave functions can be illustrated with the function[16]

$$y = f(x_1, x_2) = (x_1 \cdot x_2)^k, \tag{2.115}$$

where the x's take on only positive values, and the parameter k can take on a variety of positive values.

[16]This function is a special case of the Cobb-Douglas function. See also Problem 2.10 and the Extensions to this chapter for more details on this function.

No matter what value k takes, this function is quasi-concave. One way to show this is to look at the "level curves" of the function by setting y equal to a specific value, say c. In this case

$$y = c = (x_1 x_2)^k \quad \text{or} \quad x_1 x_2 = c^{1/k} = c'. \tag{2.116}$$

But this is just the equation of a standard rectangular hyperbola. Clearly the set of points for which y takes on values larger than c is convex because it is bounded by this hyperbola.

A more mathematical way to show quasi-concavity would apply Equation 2.114 to this function. Although the algebra of doing this is a bit messy, it may be worth the struggle. The various components of Equation 2.114 are:

$$\begin{aligned} f_1 &= kx_1^{k-1}x_2^k, \\ f_2 &= kx_1^k x_2^{k-1}, \\ f_{11} &= k(k-1)x_1^{k-2}x_2^k, \\ f_{22} &= k(k-1)x_1^k x_2^{k-2}, \\ f_{12} &= k^2 x_1^{k-1}x_2^{k-1}. \end{aligned} \tag{2.117}$$

So,

$$\begin{aligned} f_{11}f_2^2 - 2f_{12}f_1f_2 + f_{22}f_1^2 &= k^3(k-1)x_1^{3k-2}x_2^{3k-2} - 2k^4x_1^{3k-2}x_2^{3k-2} \\ &\quad + k^3(k-1)x_1^{3k-2}x_2^{3k-2} \\ &= 2k^3x_1^{3k-2}x_2^{3k-2}(-1), \end{aligned} \tag{2.118}$$

which is clearly negative, as is required for quasi-concavity.

Whether or not the function f is concave depends on the value of k. If $k < 0.5$ the function is indeed concave. An intuitive way to see this is to consider only points where $x_1 = x_2$. For these points,

$$y = (x_1^2)^k = x_1^{2k}, \tag{2.119}$$

which, for $k < 0.5$, is concave. Alternatively, for $k > 0.5$, this function is convex.

A more definitive proof makes use of the partial derivatives from Equation 2.117. In this case the condition for concavity can be expressed as

$$\begin{aligned} f_{11}f_{22} - f_{12}^2 &= k^2(k-1)^2x_1^{2k-2}x_2^{2k-2} - k^4x_1^{2k-2}x_2^{2k-2} \\ &= x_1^{2k-2}x_2^{2k-2}[k^2(k-1)^2 - k^4] \\ &= x_1^{2k-1}x_2^{2k-1}[k^2(-2k+1)], \end{aligned} \tag{2.120}$$

and this expression is positive (as is required for concavity) for

$$(-2k+1) > 0 \quad \text{or} \quad k < 0.5.$$

On the other hand, the function is convex for $k > 0.5$.

A graphic illustration. Figure 2.4 provides three-dimensional illustrations of three specific examples of this function: for $k = 0.2$, $k = 0.5$, and $k = 1$. Notice that in all three cases the level curves of the function have hyperbolic, convex shapes. That is, for any fixed value of y the functions are quite similar. This shows the quasi-concavity of the function. The primary differences among the functions are illustrated by the way in which the value of y increases as

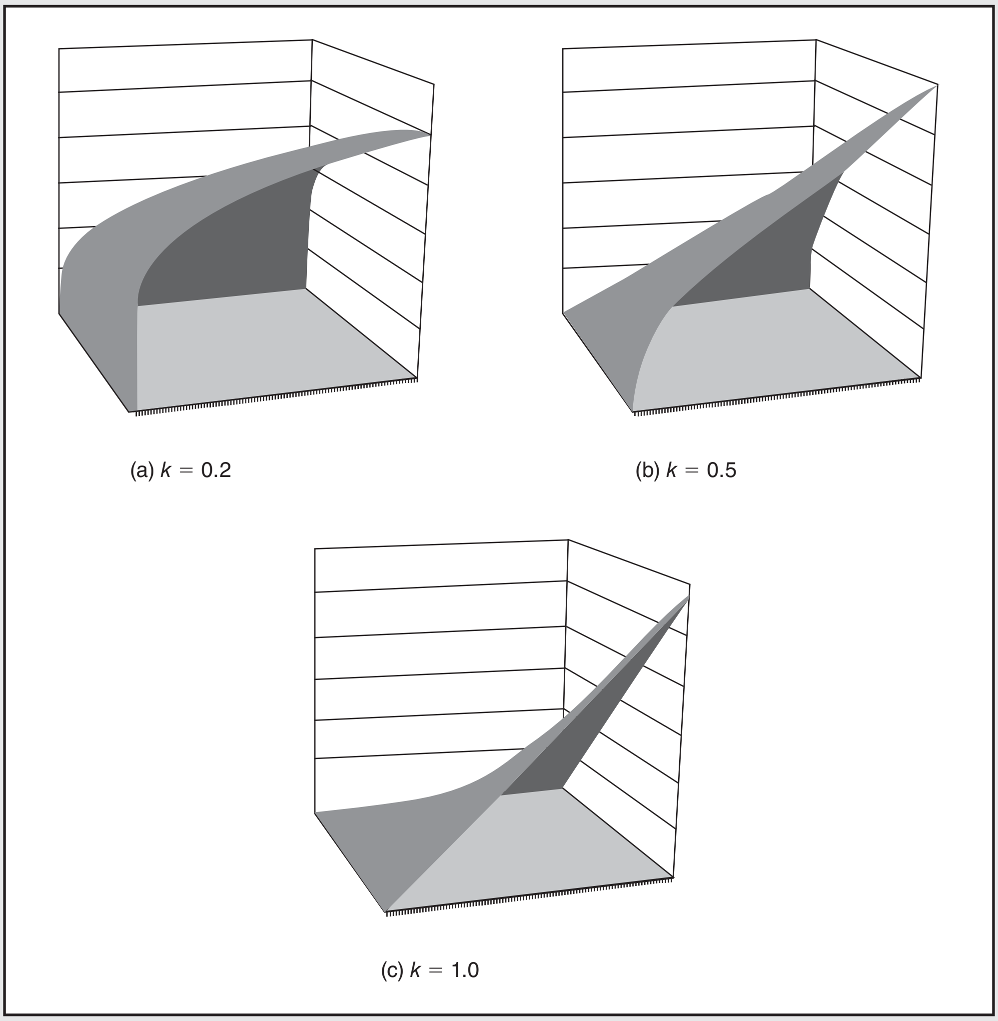
(*continued*)

EXAMPLE 2.10 CONTINUED

both x's increase together. In Figure 2.4a (when $k = 0.2$), the increase in y slows as the x's increase. This gives the function a rounded, teacuplike shape that indicates its concavity. For $k = 0.5$, y appears to increase linearly with increases in both of the x's. This is the borderline between concavity and convexity. Finally, when $k = 1$ (as in Figure 2.4c), simultaneous increases in the values of both of the x's increase y very rapidly. The spine of the function looks convex to reflect such increasing returns.

FIGURE 2.4 Concave and Quasi-Concave Functions

In all three cases these functions are quasi-concave. For a fixed y, their level curves are convex. But only for $k = 0.2$ is the function strictly concave. The case $k = 1.0$ clearly shows nonconcavity because the function is not below its tangent plane.



A careful look at Figure 2.4a suggests that any function that is concave will also be quasi-concave. You are asked to prove that this is indeed the case in Problem 2.8. This example shows that the converse of this statement is not true—quasi-concave functions need not necessarily be concave. Most functions we will encounter in this book will also illustrate this fact; most will be quasi-concave but not necessarily concave.

QUERY: Explain why the functions illustrated both in Figure 2.4a and 2.4c would have maximum values if the x's were subject to a linear constraint, but only the graph in Figure 2.4a would have an unconstrained maximum.

HOMOGENEOUS FUNCTIONS

Many of the functions that arise naturally out of economic theory have additional mathematical properties. One particularly important set of properties relates to how the functions behave when all (or most) of their arguments are increased proportionally. Such situations arise when we ask questions such as what would happen if all prices increased by 10 percent or how would a firm's output change if it doubled all of the inputs that it uses. Thinking about these questions leads naturally to the concept of homogeneous functions. Specifically, a function $f(x_1, x_2, \ldots, x_n)$ is said to be homogeneous of degree k if

$$f(tx_1, tx_2, \ldots, tx_n) = t^k f(x_1, x_2, \ldots, x_n). \tag{2.121}$$

The most important examples of homogeneous functions are those for which $k = 1$ or $k = 0$. In words, when a function is homogeneous of degree one, a doubling of all of its arguments doubles the value of the function itself. For functions that are homogeneous of degree 0, a doubling of all of its arguments leaves the value of the function unchanged. Functions may also be homogeneous for changes in only certain subsets of their arguments—that is, a doubling of some of the x's may double the value of the function if the other arguments of the function are held constant. Usually, however, homogeneity applies to changes in all of the arguments in a function.

Homogeneity and derivatives

If a function is homogeneous of degree k and can be differentiated, the partial derivatives of the function will be homogeneous of degree $k - 1$. A proof of this follows directly from the definition of homogeneity. For example, differentiating Equation 2.121 with respect to its first argument gives

$$\frac{\partial f(tx_1, \ldots, tx_n)}{\partial x_1} \cdot t = t^k \frac{\partial f(x_1, \ldots, x_n)}{\partial x_1}$$

or

$$f_1(tx_1, \ldots, tx_n) = t^{k-1} f_1(x_1, \ldots, x_n), \tag{2.122}$$

which shows that f_1 meets the definition for homogeneity of degree $k - 1$. Because marginal ideas are so prevalent in microeconomic theory, this property shows that some important properties of marginal effects can be inferred from the properties of the underlying function itself.

Euler's theorem

Another useful feature of homogeneous functions can be shown by differentiating the definition for homogeneity with respect to the proportionality factor, t. In this case, we differentiate the right side of Equation 2.121 first:

$$kt^{k-1}f_1(x_1,\ldots,x_n) = x_1 f_1(tx_1,\ldots,tx_n) + \cdots + x_n f_n(tx_1,\ldots,tx_n).$$

If we let $t = 1$, this equation becomes

$$kf(x_1,\ldots,x_n) = x_1 f_1(x_1,\ldots,x_n) + \cdots + x_n f_n(x_1,\ldots,x_n). \tag{2.123}$$

This equation is termed Euler's theorem (after the mathematician who also discovered the constant e) for homogeneous functions. It shows that, for a homogeneous function, there is a definite relationship between the values of the function and the values of its partial derivatives. Several important economic relationships among functions are based on this observation.

Homothetic functions

A homothetic function is one that is formed by taking a monotonic transformation of a homogeneous function.[17] Monotonic transformations, by definition, preserve the order of the relationship between the arguments of a function and the value of that function. If certain sets of x's yield larger values for f, they will also yield larger values for a monotonic transformation of f. Because monotonic transformations may take many forms, however, they would not be expected to preserve an exact mathematical relationship such as that embodied in homogeneous functions. Consider, for example, the function $f(x,y) = x \cdot y$. Clearly this function is homogeneous of degree 2—a doubling of its two arguments will multiply the value of the function by 4. But the monotonic transformation, F, that simply adds 1 to f [that is, $F(f) = f + 1 = xy + 1$] is not homogeneous at all. Hence, except in special cases, homothetic functions do not possess the homogeneity properties of their underlying functions. Homothetic functions do, however, preserve one nice feature of homogeneous functions. This property is that the implicit trade-offs among the variables in a function depend only on the ratios of those variables, not on their absolute values. Here we show this for the simple two-variable, implicit function $f(x,y) = 0$. It will be easier to demonstrate more general cases when we get to the economics of the matter later in this book.

Equation 2.28 showed that the implicit trade-off between x and y for a two-variable function is given by

$$\frac{dy}{dx} = -\frac{f_x}{f_y}.$$

If we assume f is homogeneous of degree k, its partial derivatives will be homogeneous of degree $k - 1$ and the implicit trade-off between x and y is

$$\frac{dy}{dx} = -\frac{t^{k-1}f_x(tx,ty)}{t^{k-1}f_y(tx,ty)} = -\frac{f_x(tx,ty)}{f_y(tx,ty)}. \tag{2.124}$$

Now let $t = 1/y$ and Equation 2.124 becomes

$$\frac{dy}{dx} = -\frac{f_x(x/y,1)}{f_y(x/y,1)}, \tag{2.125}$$

which shows that the trade-off depends only on the ratio of x to y. Now if we apply any monotonic transformation, F (with $F' > 0$), to the original homogeneous function f, we have

[17]Because a limiting case of a monotonic transformation is to leave the function unchanged, all homogeneous functions are also homothetic.

$$\frac{dy}{dx} = -\frac{F'f_x(x/y, 1)}{F'f_y(x/y, 1)} = -\frac{f_x(x/y, 1)}{f_y(x/y, 1)}, \tag{2.126}$$

and this shows both that the trade-off is unaffected by the monotonic transformation and that it remains a function only of the ratio of x to y. In Chapter 3 (and elsewhere) this property will make it very convenient to discuss some theoretical results with simple two-dimensional graphs, for which we need not consider the overall levels of key variables, but only their ratios.

EXAMPLE 2.11 Cardinal and Ordinal Properties

In applied economics it is sometimes important to know the exact numerical relationship among variables. For example, in the study of production, one might wish to know precisely how much extra output would be produced by hiring another worker. This is a question about the "cardinal" (i.e., numerical) properties of the production function. In other cases, one may only care about the order in which various points are ranked. In the theory of utility, for example, we assume that people can rank bundles of goods and will choose the bundle with the highest ranking, but that there are no unique numerical values assigned to these rankings. Mathematically, ordinal properties of functions are preserved by any monotonic transformation because, by definition, a monotonic transformation preserves order. Usually, however, cardinal properties are not preserved by arbitrary monotonic transformations.

These distinctions are illustrated by the functions we examined in Example 2.10. There we studied monotonic transformations of the function

$$f(x_1, x_2) = (x_1 x_2)^k \tag{2.127}$$

by considering various values of the parameter k. We showed that quasi-concavity (an ordinal property) was preserved for all values of k. Hence, when approaching problems that focus on maximizing or minimizing such a function subject to linear constraints we need not worry about precisely which transformation is used. On the other hand, the function in Equation 2.127 is concave (a cardinal property) only for a narrow range of values of k. Many monotonic transformations destroy the concavity of f.

The function in Equation 2.127 also can be used to illustrate the difference between homogeneous and homothetic functions. A proportional increase in the two arguments of f would yield

$$f(tx_1, tx_2) = t^{2k} x_1 x_2 = t^{2k} f(x_1, x_2). \tag{2.128}$$

Hence, the degree of homogeneity for this function depends on k—that is, the degree of homogeneity is not preserved independently of which monotonic transformation is used. Alternatively, the function in Equation 2.127 is homothetic because

$$\frac{dx_2}{dx_1} = -\frac{f_1}{f_2} = -\frac{kx_1^{k-1}x_2^k}{kx_1^k x_2^{k-1}} = -\frac{x_2}{x_1}. \tag{2.129}$$

That is, the trade-off between x_2 and x_1 depends only on the ratio of these two variables and is unaffected by the value of k. Hence, homotheticity is an ordinal property. As we shall see, this property is quite convenient when developing graphical arguments about economic propositions.

QUERY: How would the discussion in this example be changed if we considered monotonic transformations of the form $f(x_1, x_2, k) = x_1 x_2 + k$ for various values of k?

INTEGRATION

Integration is another of the tools of calculus that finds a number of applications in microeconomic theory. The technique is used both to calculate areas that measure various economic outcomes and, more generally, to provide a way of summing up outcomes that occur over time or across individuals. Our treatment of the topic here necessarily must be brief, so readers desiring a more complete background should consult the references at the end of this chapter.

Anti-derivatives

Formally, integration is the inverse of differentiation. When you are asked to calculate the integral of a function, $f(x)$, you are being asked to find a function that has $f(x)$ as its derivative. If we call this "anti-derivative" $F(x)$, this function is supposed to have the property that

$$\frac{dF(x)}{dx} = F'(x) = f(x). \tag{2.130}$$

If such a function exists then we denote it as

$$F(x) = \int f(x)\,dx. \tag{2.131}$$

The precise reason for this rather odd-looking notation will be described in detail later. First, let's look at a few examples. If $f(x) = x$ then

$$F(x) = \int f(x)\,dx = \int x\,dx = \frac{x^2}{2} + C, \tag{2.132}$$

where C is an arbitrary "constant of integration" that disappears upon differentiation. The correctness of this result can be easily verified:

$$F'(x) = \frac{d(x^2/2 + C)}{dx} = x + 0 = x. \tag{2.133}$$

Calculating anti-derivatives

Calculation of anti-derivatives can be extremely simple, or difficult, or agonizing, or impossible, depending on the particular $f(x)$ specified. Here we will look at three simple methods for making such calculations, but, as you might expect, these will not always work.

1. ***Creative guesswork.*** Probably the most common way of finding integrals (anti-derivatives) is to work backwards by asking "what function will yield $f(x)$ as its derivative?" Here are a few obvious examples:

$$\begin{aligned}
F(x) &= \int x^2\,dx = \frac{x^3}{3} + C,\\
F(x) &= \int x^n\,dx = \frac{x^{n+1}}{n+1} + C,\\
F(x) &= \int (ax^2 + bx + c)\,dx = \frac{ax^3}{3} + \frac{bx^2}{2} + cx + C,\\
F(x) &= \int e^x\,dx = e^x + C,\\
F(x) &= \int a^x\,dx = \frac{a^x}{\ln a} + C,\\
F(x) &= \int \left(\frac{1}{x}\right) dx = \ln(|x|) + C,\\
F(x) &= \int (\ln x)\,dx = x\ln x - x + C.
\end{aligned} \tag{2.134}$$

You should use differentiation to check that all of these obey the property that $F'(x) = f(x)$. Notice that in every case the integral includes a constant of integration because anti-derivatives are unique only up to an additive constant which would become zero upon differentiation. For many purposes, the results in Equation 2.134 (or trivial generalizations of them) will be sufficient for our purposes in this book. Nevertheless, here are two more methods that may work when intuition fails.

2. ***Change of variable***. A clever redefinition of variables may sometimes make a function much easier to integrate. For example, it is not at all obvious what the integral of $2x/(1+x^2)$ is. But, if we let $y = 1 + x^2$, then $dy = 2xdx$ and

$$\int \frac{2x}{1+x^2}\, dx = \int \frac{1}{y}\, dy = \ln(|y|) = \ln(|1+x^2|). \tag{2.135}$$

The key to this procedure is in breaking the original function into a term in y and a term in dy. It takes a lot of practice to see patterns for which this will work.

3. ***Integration by parts***. A similar method for finding integrals makes use of the differential expression $duv = udv + vdu$ for any two functions u and v. Integration of this differential yields

$$\int duv = uv = \int u\, dv + \int v\, du \quad \text{or} \quad \int u\, dv = uv - \int v\, du. \tag{2.136}$$

Here the strategy is to define functions u and v in a way that the unknown integral on the left can be calculated by the difference between the two known expressions on the right. For example, it is by no means obvious what the integral of xe^x is. But we can define $u = x$ (so $du = dx$) and $dv = e^x dx$ (so $v = e^x$). Hence we now have

$$\int xe^x\, dx = \int u\, dv = uv - \int v\, du = xe^x - \int e^x\, dx = (x-1)e^x + C. \tag{2.137}$$

Again, only practice can suggest useful patterns in the ways in which u and v can be defined.

Definite integrals

The integrals we have been discussing so far are "indefinite" integrals—they provide only a general function that is the anti-derivative of another function. A somewhat different, though related, approach uses integration to sum up the area under a graph of a function over some defined interval. Figure 2.5 illustrates this process. We wish to know the area under the function $f(x)$ from $x = a$ to $x = b$. One way to do this would be to partition the interval into narrow slivers of $x(\Delta x)$ and sum up the areas of the rectangles shown in the figure. That is:

$$\text{area under } f(x) \approx \sum_i f(x_i)\Delta x_i, \tag{2.138}$$

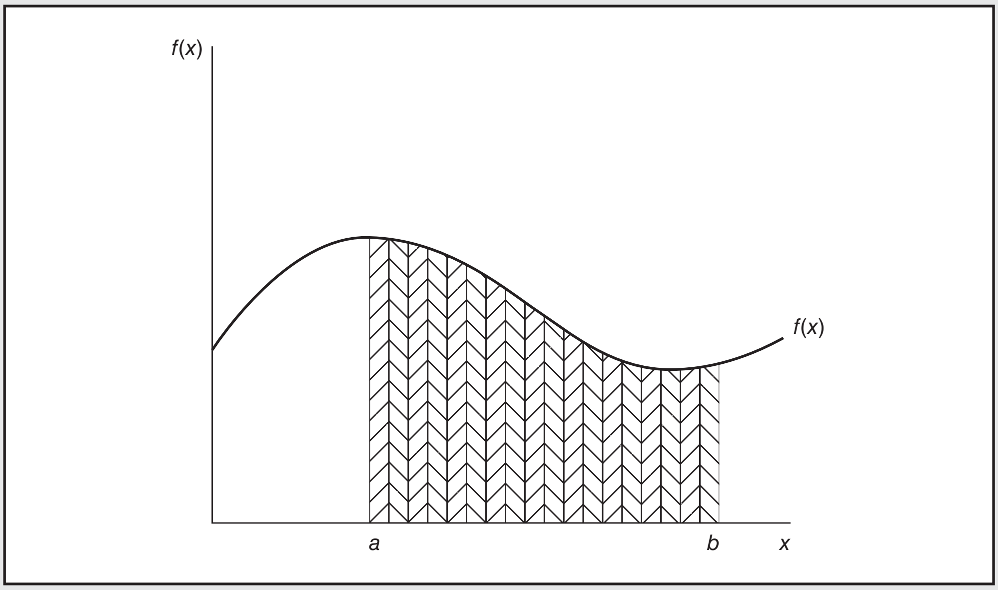
where the notation is intended to indicate that the height of each rectangle is approximated by the value of $f(x)$ for a value of x in the interval. Taking this process to the limit by shrinking the size of the Δx intervals yields an exact measure of the area we want and is denoted by:

$$\text{area under } f(x) = \int_{x=a}^{x=b} f(x)\, dx. \tag{2.139}$$

This then explains the origin of the oddly shaped integral sign—it is a stylized S, indicating "sum." As we shall see, integrating is a very general way of summing the values of a continuous function over some interval.

FIGURE 2.5 Definite Integrals Show the Areas under the Graph of a Function

Definite integrals measure the area under a curve by summing rectangular areas as shown in the graph. The dimension of each rectangle is $f(x)dx$.



Fundamental theorem of calculus

Evaluating the integral in Equation 2.139 is very simple if we know the anti-derivative of $f(x)$, say, $F(x)$. In this case we have

$$\textbf{area under} f(x) = \int_{x=a}^{x=b} f(x)\,dx = F(b) - F(a). \tag{2.140}$$

That is, all we need do is calculate the anti-derivative of $f(x)$ and subtract the value of this function at the lower limit of integration from its value at the upper limit of integration. This result is sometimes termed the "fundamental theorem of calculus" because it directly ties together the two principal tools of calculus, derivatives and integrals. In Example 2.12, we show that this result is much more general than simply a way to measure areas. It can be used to illustrate one of the primary conceptual principles of economics—the distinction between "stocks" and "flows."

EXAMPLE 2.12 Stocks and Flows

The definite integral provides a useful way for summing up any function that is providing a continuous flow over time. For example, suppose that net population increase (births minus deaths) for a country can be approximated by the function $f(t) = 1{,}000e^{0.02t}$. Hence, the net population change is growing at the rate of 2 percent per year—it is 1,000 new people in year 0, 1,020 new people in the first year, 1,041 in the second year, and so forth. Suppose we wish to know how much in total the population will increase within 50 years. This might be a tedious calculation without calculus, but using the fundamental theorem of calculus provides an easy answer:

$$\textbf{increase in population} = \int_{t=0}^{t=50} f(t)\,dt = \int_{t=0}^{t=50} 1{,}000e^{0.02t}\,dt = F(t)\Big|_0^{50}$$

$$= \frac{1{,}000e^{0.02t}}{0.02}\Big|_0^{50} = \frac{1{,}000e}{0.02} - 50{,}000 = 85{,}914 \qquad \textbf{(2.141)}$$

[where the notation $|_a^b$ indicates that the expression is to be evaluated as $F(b) - F(a)$]. Hence, the conclusion is that the population will grow by nearly 86,000 people over the next 50 years. Notice how the fundamental theorem of calculus ties together a "flow" concept, net population increase (which is measured as an amount *per year*), with a "stock" concept, total population (which is measured at a specific date and does not have a time dimension). Note also that the 86,000 calculation refers only to the total *increase* between year zero and year fifty. In order to know the actual total population at any date we would have to add the number of people in the population at year zero. That would be similar to choosing a constant of integration in this specific problem.

Now consider an application with more economic content. Suppose that total costs for a particular firm are given by $C(q) = 0.1q^2 + 500$ (where q represents output during some period). Here the term $0.1q^2$ represents variable costs (costs that vary with output) whereas the 500 figure represents fixed costs. Marginal costs for this production process can be found through differentiation—$MC = dC(q)/dq = 0.2q$—hence, marginal costs are increasing with q and fixed costs drop out upon differentiation. What are the total costs associated with producing, say, $q = 100$? One way to answer this question is to use the total cost function directly: $C(100) = 0.1(100)^2 + 500 = 1{,}500$. An alternative way would be to integrate marginal cost over the range 0 to 100 to get total variable cost:

$$\textbf{variable cost} = \int_{q=0}^{q=100} 0.2q\,dq = 0.1q^2\Big|_0^{100} = 1{,}000 - 0 = 1{,}000, \qquad \textbf{(2.142)}$$

to which we would have to add fixed costs of 500 (the constant of integration in this problem) to get total costs. Of course, this method of arriving at total cost is much more cumbersome than just using the equation for total cost directly. But the derivation does show that total variable cost between any two output levels can be found through integration as the area below the marginal cost curve—a conclusion that we will find useful in some graphical applications.

QUERY: How would you calculate the total variable cost associated with expanding output from 100 to 110? Explain why fixed costs do not enter into this calculation.

Differentiating a definite integral

Occasionally we will wish to differentiate a definite integral—usually in the context of seeking to maximize the value of this integral. Although performing such differentiations can sometimes be rather complex, there are a few rules that should make the process easier.

1. ***Differentiation with respect to the variable of integration***. This is a trick question, but instructive nonetheless. A definite integral has a constant value; hence its derivative is zero. That is:

$$\frac{d\int_a^b f(x)\,dx}{dx} = 0. \qquad \textbf{(2.143)}$$

The summing process required for integration has already been accomplished once we write down a definite integral. It does not matter whether the variable of integration is x or t or anything else. The value of this integrated sum will not change when the variable x changes, no matter what x is (but see rule 3 below).

2. *Differentiation with respect to the upper bound of integration*. Changing the upper bound of integration will obviously change the value of a definite integral. In this case, we must make a distinction between the variable determining the upper bound of integration (say, x) and the variable of integration (say, t). The result then is a simple application of the fundamental theorem of calculus. For example:

$$\frac{d\int_a^x f(t)dt}{dx} = \frac{d[F(x) - F(a)]}{dx} = f(x) - 0 = f(x), \qquad \text{(2.144)}$$

where $F(x)$ is the antiderivative of $f(x)$. By referring back to Figure 2.5 we can see why this conclusion makes sense—we are asking how the value of the definite integral changes if x increases slightly. Obviously, the answer is that the value of the integral increases by the height of $f(x)$ (notice that this value will ultimately depend on the specified value of x).

If the upper bound of integration is a function of x, this result can be generalized using the chain rule:

$$\frac{d\int_a^{g(x)} f(t)\,dt}{d(x)} = \frac{d[F(g(x)) - F(a)]}{dx} = \frac{d[F(g(x))]}{dx} = f\frac{dg(x)}{dx} = fg'(x), \qquad \text{(2.145)}$$

where, again, the specific value for this derivative would depend on the value of x assumed.

Finally, notice that differentiation with respect to a lower bound of integration just changes the sign of this expression:

$$\frac{d\int_{g(x)}^b f(t)\,dt}{dx} = \frac{d[F(b) - F(g(x))]}{dx} = -\frac{dF(g(x))}{dx} = -fg'(x). \qquad \text{(2.146)}$$

3. *Differentiation with respect to another relevant variable*. In some cases we may wish to integrate an expression that is a function of several variables. In general, this can involve multiple integrals and differentiation can become quite complicated. But there is one simple case that should be mentioned. Suppose that we have a function of two variables, $f(x, y)$, and that we wish to integrate this function with respect to the variable x. The specific value for this integral will obviously depend on the value of y and we might even ask how that value changes when y changes. In this case, it is possible to "differentiate through the integral sign" to obtain a result. That is:

$$\frac{d\int_a^b f(x, y)\,dx}{dy} = \int_a^b f_y(x, y)\,dx. \qquad \text{(2.147)}$$

This expression shows that we can first partially differentiate $f(x, y)$ with respect to y before proceeding to compute the value of the definite integral. Of course, the resulting value may still depend on the specific value that is assigned to y, but often it will yield more economic insights than the original problem does. Some further examples of using definite integrals are found in Problem 2.8.

DYNAMIC OPTIMIZATION

Some optimization problems that arise in microeconomics involve multiple periods.[18] We are interested in finding the optimal time path for a variable or set of variables that succeeds in optimizing some goal. For example, an individual may wish to choose a path of lifetime

[18]Throughout this section we treat dynamic optimization problems as occurring over time. In other contexts, the same techniques can be used to solve optimization problems that occur across a continuum of firms or individuals when the optimal choices for one agent affect what is optimal for others.

consumptions that maximizes his or her utility. Or a firm may seek a path for input and output choices that maximizes the present value of all future profits. The particular feature of such problems that makes them difficult is that decisions made in one period affect outcomes in later periods. Hence, one must explicitly take account of this interrelationship in choosing optimal paths. If decisions in one period did not affect later periods, the problem would not have a "dynamic" structure—one could just proceed to optimize decisions in each period without regard for what comes next. Here, however, we wish to explicitly allow for dynamic considerations.

The optimal control problem

Mathematicians and economists have developed many techniques for solving problems in dynamic optimization. The references at the end of this chapter provide broad introductions to these methods. Here, however, we will be concerned with only one such method that has many similarities to the optimization techniques discussed earlier in this chapter—the optimal control problem. The framework of the problem is relatively simple. A decision maker wishes to find the optimal time path for some variable $x(t)$ over a specified time interval $[t_0, t_1]$. Changes in x are governed by a differential equation:

$$\frac{dx(t)}{dt} = g[x(t), c(t), t], \tag{2.148}$$

where the variable $c(t)$ is used to "control" the change in $x(t)$. In each period of time, the decision maker derives value from x and c according to the function $f[x(t), c(t), t]$ and his or her goal to optimize $\int_{t_0}^{t_1} f[x(t), c(t), t]\, dt$. Often this problem will also be subject to "endpoint" constraints on the variable x. These might be written as $x(t_0) = x_0$ and $x(t_1) = x_1$.

Notice how this problem is "dynamic." Any decision about how much to change x this period will affect not only the future value of x, it will also affect future values of the outcome function f. The problem then is how to keep $x(t)$ on its optimal path.

Economic intuition can help to solve this problem. Suppose that we just focused on the function f and chose x and c to maximize it at each instant of time. There are two difficulties with this "myopic" approach. First, we are not really free to "choose" x at any time. Rather, the value of x will be determined by its initial value x_0 and by its history of changes as given by Equation 2.148. A second problem with this myopic approach is that it disregards the dynamic nature of the problem by not asking how this period's decisions affect the future. We need some way to reflect the dynamics of this problem in a single period's decisions. Assigning the correct value (price) to x at each instant of time will do just that. Because this implicit price will have many similarities to the Lagrangian multipliers studied earlier in this chapter, we will call it $\lambda(t)$. The value of x is treated as a function of time because the importance of x can obviously change over time.

The maximum principle

Now let's look at the decision maker's problem at a single point in time. He or she must be concerned with both the current value of the objective function $f[x(t), c(t), t]$ and with the implied change in the value of $x(t)$. Because the current value of $x(t)$ is given by $\lambda(t)x(t)$, the instantaneous rate of change of this value is given by:

$$\frac{d[\lambda(t)x(t)]}{dt} = \lambda(t)\frac{dx(t)}{dt} + x(t)\frac{d\lambda(t)}{dt}, \tag{2.149}$$

and so at any time t a comprehensive measure of the value of concern[19] to the decision maker is

$$H = f[x(t), c(t), t] + \lambda(t)g[x(t), c(t), t] + x(t)\frac{d\lambda(t)}{dt}. \quad \textbf{(2.150)}$$

This comprehensive value represents both the current benefits being received and the instantaneous change in the value of x. Now we can ask what conditions must hold for $x(t)$ and $c(t)$ to optimize this expression.[20] That is:

$$\begin{aligned} \frac{\partial H}{\partial c} &= f_c + \lambda g_c = 0 \quad \text{or} \quad f_c = -\lambda g_c; \\ \frac{\partial H}{\partial x} &= f_x + \lambda g_x + \frac{\partial \lambda(t)}{dt} = 0 \quad \text{or} \quad f_x + \lambda g_x = -\frac{\partial \lambda(t)}{\partial t}. \end{aligned} \quad \textbf{(2.151)}$$

These are then the two optimality conditions for this dynamic problem. They are usually referred to as the "maximum principle." This solution to the optimal control problem was first proposed by the Russian mathematician L. S. Pontryagin and his colleagues in the early 1960s.

Although the logic of the maximum principle can best be illustrated by the economic applications we will encounter later in this book, a brief summary of the intuition behind them may be helpful. The first condition asks about the optimal choice of c. It suggests that, at the margin, the gain from c in terms of the function f must be balanced by the losses from c in terms of the value of its ability to change x. That is, present gains must be weighed against future costs.

The second condition relates to the characteristics that an optimal time path of $x(t)$ should have. It implies that, at the margin, any net gains from more current x (either in terms of f or in terms of the accompanying value of changes in x) must be balanced by changes in the implied value of x itself. That is, the net current gain from more x must be weighed against the declining future value of x.

EXAMPLE 2.13 Allocating a Fixed Supply

As an extremely simple illustration of the maximum principle, assume that someone has inherited 1,000 bottles of wine from a rich uncle. He or she intends to drink these bottles over the next 20 years. How should this be done to maximize the utility from doing so?

Suppose that this person's utility function for wine is given by $u[c(t)] = \ln c(t)$. Hence the utility from wine drinking exhibits diminishing marginal utility ($u' > 0, u'' < 0$). This person's goal is to maximize

$$\int_0^{20} u[c(t)]\, dt = \int_0^{20} \ln c(t)\, dt. \quad \textbf{(2.152)}$$

Let $x(t)$ represent the number of bottles of wine remaining at time t. This series is constrained by $x(0) = 1{,}000$ and $x(20) = 0$. The differential equation determining the evolution of $x(t)$ takes the simple form:[21]

[19]We denote this current value expression by H to suggest its similarity to the Hamiltonian expression used in formal dynamic optimization theory. Usually the Hamiltonian does not have the final term in Equation 2.150, however.

[20]Notice that the variable x is not really a choice variable here—its value is determined by history. Differentiation with respect to x can be regarded as implicitly asking the question: "If $x(t)$ were optimal, what characteristics would it have?"

[21]The simple form of this differential equation (where dx/dt depends only on the value of the control variable, c) means that this problem is identical to one explored using the "calculus of variations" approach to dynamic optimization. In such a case, one can substitute dx/dt into the function f and the first-order conditions for a maximum can be compressed into

$$\frac{dx(t)}{dt} = -c(t). \tag{2.153}$$

That is, each instant's consumption just reduces the stock of remaining bottles. The current value Hamiltonian expression for this problem is

$$H = \ln c(t) + \lambda[-c(t)] + x(t)\frac{d\lambda}{dt}, \tag{2.154}$$

and the first-order conditions for a maximum are

$$\begin{aligned} \frac{\partial H}{\partial c} &= \frac{1}{c} - \lambda = 0, \\ \frac{\partial H}{\partial x} &= \frac{d\lambda}{dt} = 0. \end{aligned} \tag{2.155}$$

The second of these conditions requires that λ (the implicit value of wine) be constant over time. This makes intuitive sense: because consuming a bottle of wine always reduces the available stock by one bottle, any solution where the value of wine differed over time would provide an incentive to change behavior by drinking more wine when it is cheap and less when it is expensive. Combining this second condition for a maximum with the first condition implies that $c(t)$ itself must be constant over time. If $c(t) = k$, the number of bottles remaining at any time will be $x(t) = 1{,}000 - kt$. If $k = 50$, the system will obey the end point constraints $x(0) = 1000$ and $x(20) = 0$. Of course, in this problem you could probably guess that the optimum plan would be to drink the wine at the rate of 50 bottles per year for 20 years because diminishing marginal utility suggests one does not want to drink excessively in any period. The maximum principle confirms this intuition.

More complicated utility. Now let's take a more complicated utility function that may yield more interesting results. Suppose that the utility of consuming wine at any date, t, is given by

$$u[c(t)] = \begin{cases} [c(t)]^{\gamma}/\gamma & \text{if } \gamma \neq 0, \gamma < 1; \\ \ln c(t) & \text{if } \gamma = 0. \end{cases} \tag{2.156}$$

Assume also that the consumer discounts future consumption at the rate δ. Hence this person's goal is to maximize

$$\int_0^{20} u[c(t)]\,dt = \int_0^{20} e^{-\delta t}\frac{[c(t)]^{\gamma}}{\gamma}\,dt \tag{2.157}$$

subject to the following constraints:

$$\begin{aligned} \frac{dx(t)}{dt} &= -c(t), \\ x(0) &= 1{,}000, \\ x(20) &= 0. \end{aligned} \tag{2.158}$$

Setting up the current value Hamiltonian expression yields

$$H = e^{-\delta t}\frac{[c(t)]^{\gamma}}{\gamma} + \lambda(-c) + x(t)\frac{d\lambda(t)}{dt}, \tag{2.159}$$

and the maximum principle requires that

(*continued*)

the single equation $f_x = df_{dx/dt}/dt$, which is termed the "Euler equation." In Chapter 17 we will encounter many Euler equations.

EXAMPLE 2.13 CONTINUED

$$\frac{\partial H}{\partial c} = e^{-\delta t}[c(t)]^{\gamma-1} - \lambda = 0 \quad \text{and}$$
$$\frac{\partial H}{\partial x} = 0 + 0 + \frac{d\lambda}{dt} = 0. \tag{2.160}$$

Hence, we can again conclude that the implicit value of the wine stock (λ) should be constant over time (call this constant k) and that

$$e^{-\delta t}[c(t)]^{\gamma-1} = k \quad \text{or} \quad c(t) = k^{1/(\gamma-1)}e^{\delta t/(\gamma-1)}. \tag{2.161}$$

So, optimal wine consumption should fall over time in order to compensate for the fact that future consumption is being discounted in the consumer's mind. If, for example, we let $\delta = 0.1$ and $\gamma = -1$ ("reasonable" values, as we will show in later chapters), then

$$c(t) = k^{-0.5}e^{-0.05t} \tag{2.162}$$

Now we must do a bit more work in choosing k to satisfy the endpoint constraints. We want

$$\int_0^{20} c(t)\,dt = \int_0^{20} k^{-0.5}e^{-0.05t}\,dt = -20k^{-0.5}e^{-0.05t}\Big|_0^{20}$$
$$= -20k^{-0.5}(e^{-1} - 1) = 12.64k^{-0.5} = 1{,}000. \tag{2.163}$$

Finally, then, we have the optimal consumption plan as

$$c(t) \approx 79e^{-0.05t}. \tag{2.164}$$

This consumption plan requires that wine consumption start out fairly high and decline at a continuous rate of 5 percent per year. Because consumption is continuously declining, we must use integration to calculate wine consumption in any particular year (x) as follows:

$$\text{consumption in year } x \approx \int_{x-1}^{x} c(t)\,dt = \int_{x-1}^{x} 79e^{-0.05t}\,dt = -1{,}580e^{-0.05t}\Big|_{x-1}^{x}$$
$$= 1{,}580(e^{-0.05(x-1)} - e^{-0.05x}). \tag{2.165}$$

If $x = 1$, consumption is about 77 bottles in this first year. Consumption then declines smoothly, ending with about 30 bottles being consumed in the 20th year.

QUERY: Our first illustration was just an example of the second in which $\delta = \gamma = 0$. Explain how alternative values of these parameters will affect the path of optimal wine consumption. Explain your results intuitively (for more on optimal consumption over time, see Chapter 17).

MATHEMATICAL STATISTICS

In recent years microeconomic theory has increasingly focused on issues raised by uncertainty and imperfect information. To understand much of this literature, it is important to have a good background in mathematical statistics. The purpose of this section is, therefore, to summarize a few of the statistical principles that we will encounter at various places in this book.

Random variables and probability density functions

A *random variable* describes (in numerical form) the outcomes from an experiment that is subject to chance. For example, we might flip a coin and observe whether it lands heads or tails. If we call this random variable x, we can denote the possible outcomes ("realizations") of the variable as:

$$x = \begin{cases} 1 & \text{if coin is heads,} \\ 0 & \text{if coin is tails.} \end{cases}$$

Notice that, prior to the flip of the coin, x can be either 1 or 0. Only after the uncertainty is resolved (that is, after the coin is flipped) do we know what the value of x is.[22]

Discrete and continuous random variables

The outcomes from a random experiment may be either a finite number of possibilities or a continuum of possibilities. For example, recording the number that comes up on a single die is a random variable with six outcomes. With two dice, we could either record the sum of the faces (in which case there are 12 outcomes, some of which are more likely than others) or we could record a two-digit number, one for the value of each die (in which case there would be 36 equally likely outcomes). These are examples of discrete random variables.

Alternatively, a continuous random variable may take on any value in a given range of real numbers. For example, we could view the outdoor temperature tomorrow as a continuous variable (assuming temperatures can be measured very finely) ranging from, say, $-50°C$ to $+50°C$. Of course, some of these temperatures would be very unlikely to occur, but in principle the precisely measured temperature could be anywhere between these two bounds. Similarly, we could view tomorrow's percentage change in the value of a particular stock index as taking on all values between -100% and, say, $+1{,}000\%$. Again, of course, percentage changes around 0% would be considerably more likely to occur than would be the extreme values.

Probability density functions

For any random variable, its *probability density function* (PDF) shows the probability that each specific outcome will occur. For a discrete random variable, defining such a function poses no particular difficulties. In the coin flip case, for example, the PDF [denoted by $f(x)$] would be given by

$$\begin{aligned} f(x = 1) &= 0.5, \\ f(x = 0) &= 0.5. \end{aligned} \tag{2.166}$$

For the roll of a single die, the PDF would be:

$$\begin{aligned} f(x = 1) &= 1/6, \\ f(x = 2) &= 1/6, \\ f(x = 3) &= 1/6, \\ f(x = 4) &= 1/6, \\ f(x = 5) &= 1/6, \\ f(x = 6) &= 1/6. \end{aligned} \tag{2.167}$$

[22]Sometimes random variables are denoted by $\tilde{x}$ to make a distinction between variables whose outcome is subject to random chance and (nonrandom) algebraic variables. This notational device can be useful for keeping track of what is random and what is not in a particular problem and we will use it in some cases. When there is no ambiguity, however, we will not employ this special notation.

Notice that in both of these cases the probabilities specified by the PDF sum to 1.0. This is because, by definition, one of the outcomes of the random experiment must occur. More generally, if we denote all of the outcomes for a discrete random variable by x_i for $i = 1, \ldots, n$, then we must have:

$$\sum_{i=1}^{n} f(x_i) = 1. \tag{2.168}$$

For a continuous random variable we must be careful in defining the PDF concept. Because such a random variable takes on a continuum of values, if we were to assign any non-zero value as the probability for a specific outcome (i.e., a temperature of +25.53470°C), we could quickly have sums of probabilities that are infinitely large. Hence, for a continuous random variable we define the PDF $f(x)$ as a function with the property that the probability that x falls in a particular small interval dx is given by the area of $f(x)dx$. Using this convention, the property that the probabilities from a random experiment must sum to 1.0 is stated as follows:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.0. \tag{2.169}$$

A few important PDFs

Most any function will do as a probability density function provided that $f(x) \geq 0$ and the function sums (or integrates) to 1.0. The trick, of course, is to find functions that mirror random experiments that occur in the real world. Here we look at four such functions that we will find useful in various places in this book. Graphs for all four of these functions are shown in Figure 2.6.

1. ***Binomial distribution***. This is the most basic discrete distribution. Usually x is assumed to take on only two values, 1 and 0. The PDF for the binomial is given by:

$$\begin{aligned} &f(x = 1) = p, \\ &f(x = 0) = 1 - p, \\ &\text{where} \quad 0 < p < 1. \end{aligned} \tag{2.170}$$

The coin flip example is obviously a special case of the binomial where $p = 0.5$.

2. ***Uniform distribution***. This is the simplest continuous PDF. It assumes that the possible values of the variable x occur in a defined interval and that each value is equally likely. That is:

$$\begin{aligned} f(x) &= \frac{1}{b-a} \quad &&\text{for } a \leq x \leq b; \\ f(x) &= 0 &&\text{for } x < a \text{ or } x > b. \end{aligned} \tag{2.171}$$

Notice that here the probabilities integrate to 1.0:

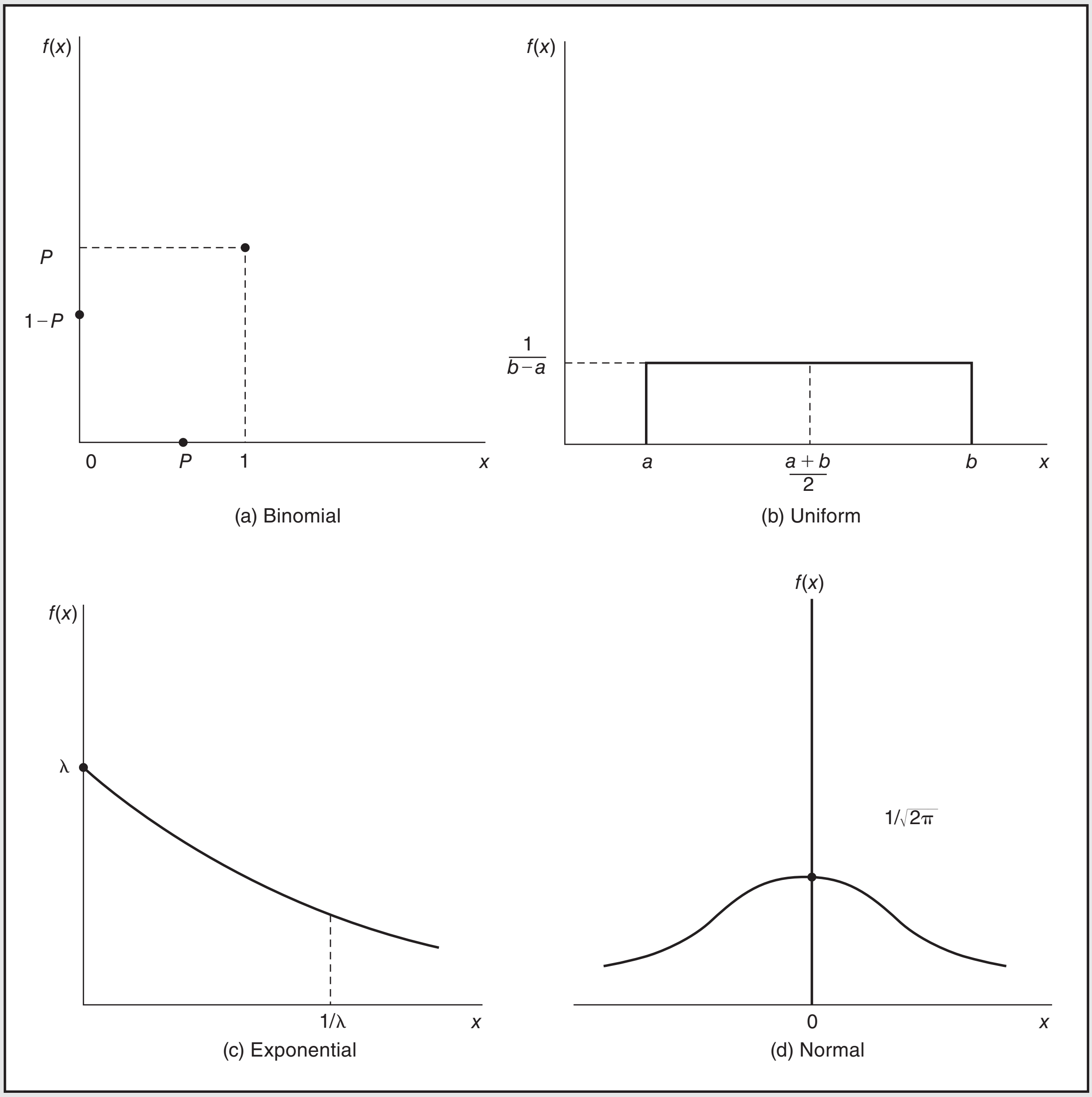
$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{a}^{b} \frac{1}{b-a}\,dx = \left.\frac{x}{b-a}\right|_a^b = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1.0. \tag{2.172}$$

3. ***Exponential distribution***. This is a continuous distribution for which the probabilities decline at a smooth exponential rate as x increases. Formally:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases} \tag{2.173}$$

FIGURE 2.6 Four Common Probability Density Functions

Random variables that have these PDFs are widely used. Each graph indicates the expected value of the PDF shown.



where λ is a positive constant. Again, it is easy to show that this function integrates to 1.0:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{0}^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_{0}^{\infty} = 0 - (-1) = 1.0. \qquad \textbf{(2.174)}$$

4. *Normal distribution*. The Normal (or Gaussian) distribution is the most important in mathematical statistics. It's importance stems largely from the *central limit theorem,* which states that the distribution of any sum of independent random variables will increasingly approximate the Normal distribution as the number of such variables increase. Because sample averages can be regarded as sums of independent random variables, this theorem says that any sample average will have a Normal distribution no matter what the distribution of the population from which the sample is selected. Hence, it may often be appropriate to assume a random variable has a Normal distribution if it can be thought of as some sort of average.

The mathematical form for the Normal PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{2.175}$$

and this is defined for all real values of x. Although the function may look complicated, a few of its properties can be easily described. First, the function is symmetric around zero (because of the x^2 term). Second, the function is asymptotic to zero as x becomes very large or very small. Third, the function reaches its maximal value at $x = 0$. This value is $1/\sqrt{2\pi} \approx 0.4$. Finally, the graph of this function has a general "bell shape"—a shape used throughout the study of statistics. Integration of this function is relatively tricky (though easy in polar coordinates). The presence of the constant $1/\sqrt{2\pi}$ is needed if the function is to integrate to 1.0.

Expected value

The *expected value* of a random variable is the numerical value that the random variable might be expected to have, on average.[23] It is the "center of gravity" of the probability density function. For a discrete random variable that takes on the values $x_1, x_2, \ldots, x_n$, the expected value is defined as

$$E(x) = \sum_{i=1}^{n} x_i f(x_i). \tag{2.176}$$

That is, each outcome is weighted by the probability that it will occur and the result is summed over all possible outcomes. For a continuous random variable, Equation 2.176 is readily generalized as

$$E(x) = \int_{-\infty}^{+\infty} x f(x)\, dx. \tag{2.177}$$

Again, in this integration, each value of x is weighted by the probability that this value will occur.

The concept of expected value can be generalized to include the expected value of any function of a random variable [say, $g(x)$]. In the continuous case, for example, we would write

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x)\, dx. \tag{2.178}$$

[23]The expected value of a random variable is sometimes referred to as the *mean* of that variable. In the study of sampling this can sometimes lead to confusion between the expected value of a random variable and the separate concept of the sample arithmetic average.

As a special case, consider a linear function $y = ax + b$. Then

$$E(y) = E(ax + b) = \int_{-\infty}^{+\infty} (ax + b)f(x)\,dx$$
$$= a\int_{-\infty}^{+\infty} xf(x)\,dx + b\int_{-\infty}^{+\infty} f(x)\,dx = aE(x) + b. \quad \textbf{(2.179)}$$

Sometimes expected values are phrased in terms of the *cumulative distribution function* (CDF) $F(x)$, defined as

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \quad \textbf{(2.180)}$$

That is, $F(x)$ represents the probability that the random variable t is less than or equal to x. With this notation, the expected value of $g(x)$ is defined as

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\,dF(x). \quad \textbf{(2.181)}$$

Because of the fundamental theorem of calculus, Equation 2.181 and Equation 2.178 mean exactly the same thing.

EXAMPLE 2.14 Expected Values of a Few Random Variables

The expected values of each of the random variables with the simple PDFs introduced earlier are easy to calculate. All of these expected values are indicated on the graphs of the functions' PDFs in Figure 2.6.

1. *Binomial*. In this case:

$$E(x) = 1 \cdot f(x = 1) + 0 \cdot f(x = 0) = 1 \cdot p + 0 \cdot (1 - p) = p. \quad \textbf{(2.182)}$$

For the coin flip case (where $p = 0.5$), this says that $E(x) = p = 0.5$—the expected value of this random variable is, as you might have guessed, one half.

2. *Uniform*. For this continuous random variable,

$$E(x) = \int_{a}^{b} \frac{x}{b - a}\,dx = \frac{x^2}{2(b - a)}\bigg|_{a}^{b} = \frac{b^2}{2(b - a)} - \frac{a^2}{2(b - a)} = \frac{b + a}{2}. \quad \textbf{(2.183)}$$

Again, as you might have guessed, the expected value of the uniform distribution is precisely halfway between a and b.

3. *Exponential*. For this case of declining probabilities:

$$E(x) = \int_{0}^{\infty} x\lambda e^{-\lambda x}\,dx = -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\bigg|_{0}^{\infty} = \frac{1}{\lambda}, \quad \textbf{(2.184)}$$

where the integration follows from the integration by parts example shown earlier in this chapter (Equation 2.137). Notice here that the faster the probabilities decline, the lower is the expected value of x. For example, if $\lambda = 0.5$ then $E(x) = 2$, whereas if $\lambda = 0.05$ then $E(x) = 20$.

(*continued*)

EXAMPLE 2.14 CONTINUED

4. ***Normal.*** Because the Normal PDF is symmetric around zero, it seems clear that $E(x) = 0$. A formal proof uses a change of variable integration by letting $u = x^2/2$ $(du = xdx)$:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} xe^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u}\, du = \frac{1}{\sqrt{2\pi}} [-e^{-x^2/2}]\Big|_{-\infty}^{+\infty} = \frac{1}{\sqrt{2\pi}} [0 - 0] = 0. \quad \textbf{(2.185)}$$

Of course, the expected value of a normally distributed random variable (or of any random variable) may be altered by a linear transformation, as shown in Equation 2.179.

QUERY: A linear transformation changes a random variable's expected value in a very predictable way—if $y = ax + b$, then $E(y) = aE(x) + b$. Hence, for this transformation [say, $h(x)$] we have $E[h(x)] = h[E(x)]$. Suppose instead that x were transformed by a concave function, say $g(x)$ with $g' > 0$ and $g'' < 0$. How would $E[g(x)]$ compare to $g[E(x)]$?

Note: This is an illustration of Jensen's inequality, a concept we will pursue in detail in Chapter 7. See also Problem 2.13.

Variance and standard deviation

The expected value of a random variable is a measure of central tendency. On the other hand, the *variance* of a random variable [denoted by σ_x^2 or $\text{Var}(x)$] is a measure of dispersion. Specifically, the variance is defined as the "expected squared deviation" of a random variable from its expected value. Formally:

$$\text{Var}(x) = \sigma_x^2 = E[(x - E(x))^2] = \int_{-\infty}^{+\infty} (x - E(x))^2 f(x)\, dx. \quad \textbf{(2.186)}$$

Somewhat imprecisely, the variance measures the "typical" squared deviation from the central value of a random variable. In making the calculation, deviations from the expected value are squared so that positive and negative deviations from the expected value will both contribute to this measure of dispersion. After the calculation is made, the squaring process can be reversed to yield a measure of dispersion that is in the original units in which the random variable was measured. This square root of the variance is called the "standard deviation" and is denoted as $\sigma_x \left(= \sqrt{\sigma_x^2}\right)$. The wording of the term effectively conveys its meaning: σ_x is indeed the typical ("standard") deviation of a random variable from its expected value.

When a random variable is subject to a linear transformation, its variance and standard deviation will be changed in a fairly obvious way. If $y = ax + b$, then

$$\sigma_y^2 = \int_{-\infty}^{+\infty} [ax + b - E(ax + b)]^2 f(x)\, dx = \int_{-\infty}^{+\infty} a^2[x - E(x)]^2 f(x)\, dx = a^2\sigma_x^2. \quad \textbf{(2.187)}$$

Hence, addition of a constant to a random variable does not change its variance, whereas multiplication by a constant multiplies the variance by the square of the constant. It is clear therefore that multiplying a variable by a constant multiplies its standard deviation by that constant: $\sigma_{ax} = a\sigma_x$.

EXAMPLE 2.15 Variances and Standard Deviations for Simple Random Variables

Knowing the variances and standard deviations of the four simple random variables we have been looking at can sometimes be quite useful in economic applications.

1. *Binomial*. The variance of the binomial can be calculated by applying the definition in its discrete analog:

$$\sigma_x^2 = \sum_{i=1}^{n}(x_i - E(x))^2 f(x_i) = (1-p)^2 \cdot p + (0-p)^2(1-p)$$
$$= (1-p)(p - p^2 + p^2) = p(1-p). \tag{2.188}$$

Hence, $\sigma_x = \sqrt{p(1-p)}$. One implication of this result is that a binomial variable has the largest variance and standard deviation when $p = 0.5$, in which case $\sigma_x^2 = 0.25$ and $\sigma_x = 0.5$. Because of the relatively flat parabolic shape of $p(1-p)$, modest deviations of p from 0.5 do not change this variance substantially.

2. *Uniform*. Calculating the variance of the uniform distribution yields a mildly interesting result:

$$\sigma_x^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\, dx = \left(x - \frac{a+b}{2}\right)^3 \cdot \frac{1}{3(b-a)}\bigg|_a^b$$
$$= \frac{1}{3(b-a)}\left[\frac{(b-a)^3}{8} - \frac{(a-b)^3}{8}\right] = \frac{(b-a)^2}{12}. \tag{2.189}$$

This is one of the few places where the number 12 has any use in mathematics other than in measuring quantities of oranges or doughnuts.

3. *Exponential*. Integrating the variance formula for the exponential is relatively laborious. Fortunately, the result is quite simple; for the exponential, it turns out that $\sigma_x^2 = 1/\lambda^2$ and $\sigma_x = 1/\lambda$. Hence, the mean and standard deviation are the same for the exponential distribution—it is a "one-parameter distribution."

4. *Normal*. In this case also, the integration can be burdensome. But again the result is simple: for the Normal distribution, $\sigma_x^2 = \sigma_x = 1$. Areas below the Normal curve can be readily calculated and tables of these are available in any statistics text. Two useful facts about the Normal PDF are:

$$\int_{-1}^{+1} f(x)\, dx \approx 0.68 \quad \text{and} \quad \int_{-2}^{+2} f(x)\, dx \approx 0.95. \tag{2.190}$$

That is, the probability is about two thirds that a Normal variable will be within ±1 standard deviation of the expected value and "most of the time" (i.e., with probability 0.95) it will be within ±2 standard deviations.

Standardizing the Normal. If the random variable x has a standard Normal PDF, it will have an expected value of 0 and a standard deviation of 1. However, a simple linear transformation can be used to give this random variable any desired expected value (μ) and standard deviation (σ). Consider the transformation $y = \sigma x + \mu$. Now

$$E(y) = \sigma E(x) + \mu = \mu \quad \text{and} \quad \text{Var}(y) = \sigma_y^2 = \sigma^2 \,\text{Var}(x) = \sigma^2. \tag{2.191}$$

Reversing this process can be used to "standardize" any Normally distributed random variable (y) with an arbitrary expected value (μ) and standard deviation (σ) (this is sometimes denoted

(continued)

EXAMPLE 2.15 CONTINUED

as $y \sim N(\mu, \sigma)$) by using $z = (y - \mu)/\sigma$. For example, SAT scores (y) are distributed Normally with an expected value of 500 points and a standard deviation of 100 points (that is, $y \sim N(500, 100)$). Hence, $z = (y - 500)/100$ has a standard Normal distribution with expected value 0 and standard deviation 1. Equation 2.190 shows that approximately 68 percent of all scores lie between 400 and 600 points and 95 percent of all scores lie between 300 and 700 points.

QUERY: Suppose that the random variable x is distributed uniformly along the interval $[0, 12]$. What are the mean and standard deviation of x? What fraction of the x distribution is within ± 1 standard deviation of the mean? What fraction of the distribution is within ± 2 standard deviations of the expected value? Explain why this differs from the fractions computed for the Normal distribution.

Covariance

Some economic problems involve two or more random variables. For example, an investor may consider allocating his or her wealth among several assets the returns on which are taken to be random. Although the concepts of expected value, variance, and so forth carry over more or less directly when looking at a single random variable in such cases, it is also necessary to consider the relationship between the variables to get a complete picture. The concept of covariance is used to quantify this relationship. Before providing a definition, however, we will need to develop some background.

Consider a case with two continuous random variables, x and y. The probability density function for these two variables, denoted by $f(x, y)$, has the property that the probability associated with a set of outcomes in a small area (with dimensions $dxdy$) is given by $f(x, y)dxdy$. To be a proper PDF, it must be the case that:

$$f(x, y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1. \tag{2.192}$$

The single-variable measures we have already introduced can be developed in this two-variable context by "integrating out" the other variable. That is,

$$\begin{aligned} E(x) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y)\, dy\, dx \quad \text{and} \\ \mathrm{Var}(x) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(x)]^2 f(x, y)\, dy\, dx. \end{aligned} \tag{2.193}$$

In this way, the parameters describing the random variable x are measured over all possible outcomes for y after taking into account the likelihood of those various outcomes.

In this context, the *covariance* between x and y seeks to measure the direction of association between the variables. Specifically the covariance between x and y [denoted as $\mathrm{Cov}(x, y)$] is defined as

$$\mathrm{Cov}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(x)][y - E(y)] f(x, y)\, dx\, dy. \tag{2.194}$$

The covariance between two random variables may be positive, negative, or zero. If values of x that are greater than $E(x)$ tend to occur relatively frequently with values of y that are greater than $E(y)$ (and similarly, if low values of x tend to occur together with low values of y), then the covariance will be positive. In this case, values of x and y tend to move in the same direction. Alternatively, if high values of x tend to be associated with low values for y (and vice versa), the covariance will be negative.

Two random variables are defined to be *independent* if the probability of any particular value of, say, x is not affected by the particular value of y that might occur (and vice versa).[24] In mathematical terms, this means that the PDF must have the property that $f(x,y) = g(x)h(y)$—that is, the joint probability density function can be expressed as the product of two single-variable PDFs. If x and y are independent, their covariance will be zero:

$$\begin{aligned}\mathbf{Cov}(x,y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(x)][y-E(y)]g(x)h(y)\,dx\,dy\\ &= \int_{-\infty}^{+\infty}[x-E(x)]g(x)\,dx\cdot\int_{-\infty}^{+\infty}[y-E(y)]h(y)\,dy = 0\cdot 0 = 0.\end{aligned} \tag{2.195}$$

The converse of this statement is not necessarily true, however. A zero covariance does not necessarily imply statistical independence.

Finally, the covariance concept is crucial for understanding the variance of sums or differences of random variables. Although the expected value of a sum of two random variables is (as one might guess) the sum of their expected values:

$$\begin{aligned}E(x+y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\,dx\,dy\\ &= \int_{-\infty}^{+\infty}xf(x,y)\,dy\,dx + \int_{-\infty}^{+\infty}yf(x,y)\,dx\,dy = E(x)+E(y),\end{aligned} \tag{2.196}$$

the relationship for the variance of such a sum is more complicated. Using the definitions we have developed yields

$$\begin{aligned}\mathbf{Var}(x+y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x+y-E(x+y)]^2 f(x,y)\,dx\,dy\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(x)+y-E(y)]^2 f(x,y)\,dx\,dy\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(x)]^2+[y-E(y)]^2+2[x-E(x)][y-E(y)]f(x,y)\,dx\,dy\\ &= \mathbf{Var}(x)+\mathbf{Var}(y)+2\,\mathbf{Cov}(x,y).\end{aligned} \tag{2.197}$$

Hence, if x and y are independent then $\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y)$. The variance of the sum will be greater than the sum of the variances if the two random variables have a positive covariance and will be less than the sum of the variances if they have a negative covariance. Problems 2.13 and 2.14 provide further details on statistical issues that arise in microeconomic theory.

---

[24] A formal definition relies on the concept of conditional probability. The conditional probability of an event B given that A has occurred (written $P(B|A)$ is defined as $P(B|A) = P(A \text{ and } B)/P(A)$; B and A are defined to be independent if $P(B|A) = P(B)$. In this case, $P(A \text{ and } B) = P(A)\cdot P(B)$.

SUMMARY

Despite the formidable appearance of some parts of this chapter, this is not a book on mathematics. Rather, the intention here was to gather together a variety of tools that will be used to develop economic models throughout the remainder of the text. Material in this chapter will then be useful as a handy reference.

One way to summarize the mathematical tools introduced in this chapter is by stressing again the economic lessons that these tools illustrate:

- Using mathematics provides a convenient, shorthand way for economists to develop their models. Implications of various economic assumptions can be studied in a simplified setting through the use of such mathematical tools.
- The mathematical concept of the derivatives of a function is widely used in economic models because economists are often interested in how marginal changes in one variable affect another variable. Partial derivatives are especially useful for this purpose because they are defined to represent such marginal changes when all other factors are held constant.
- The mathematics of optimization is an important tool for the development of models that assume that economic agents rationally pursue some goal. In the unconstrained case, the first-order conditions state that any activity that contributes to the agent's goal should be expanded up to the point at which the marginal contribution of further expansion is zero. In mathematical terms, the first-order condition for an optimum requires that all partial derivatives be zero.
- Most economic optimization problems involve constraints on the choices agents can make. In this case the first-order conditions for a maximum suggest that each activity be operated at a level at which the ratio of the marginal benefit–of the activity to its marginal cost is the same for all activities actually used. This common marginal benefit–marginal cost ratio is also equal to the Lagrangian multiplier, which is often introduced to help solve constrained optimization problems. The Lagrangian multiplier can also be interpreted as the implicit value (or shadow price) of the constraint.
- The implicit function theorem is a useful mathematical device for illustrating the dependence of the choices that result from an optimization problem on the parameters of that problem (for example, market prices). The envelope theorem is useful for examining how these optimal choices change when the problem's parameters (prices) change.
- Some optimization problems may involve constraints that are inequalities rather than equalities. Solutions to these problems often illustrate "complementary slackness." That is, either the constraints hold with equality and their related Lagrangian multipliers are nonzero, or the constraints are strict inequalities and their related Lagrangian multipliers are zero. Again this illustrates how the Lagrangian multiplier implies something about the "importance" of constraints.
- The first-order conditions shown in this chapter are only the necessary conditions for a local maximum or minimum. One must also check second-order conditions that require that certain curvature conditions be met.
- Certain types of functions occur in many economic problems. Quasi-concave functions (those functions for which the level curves form convex sets) obey the second-order conditions of constrained maximum or minimum problems when the constraints are linear. Homothetic functions have the useful property that implicit trade-offs among the variables of the function depend only on the ratios of these variables.
- Integral calculus is often used in economics both as a way of describing areas below graphs and as a way of summing results over time. Techniques that involve various ways of differentiating integrals play an important role in the theory of optimizing behavior.
- Many economic problems are dynamic in that decisions at one date affect decisions and outcomes at later dates. The mathematics for solving such dynamic optimization problems is often a straightforward generalization of Lagrangian methods.
- Concepts from mathematical statistics are often used in studying the economics of uncertainty and information. The most fundamental concept is the notion of a random variable and its associated probability density function. Parameters of this distribution, such as its expected value or its variance, also play important roles in many economic models.

PROBLEMS

2.1

Suppose $U(x, y) = 4x^2 + 3y^2$.

a. Calculate $\partial U/\partial x$, $\partial U/\partial y$.

b. Evaluate these partial derivatives at $x = 1$, $y = 2$.

c. Write the total differential for U.

d. Calculate dy/dx for $dU = 0$—that is, what is the implied trade-off between x and y holding U constant?

e. Show $U = 16$ when $x = 1, y = 2$.

f. In what ratio must x and y change to hold U constant at 16 for movements away from $x = 1$, $y = 2$?

g. More generally, what is the shape of the $U = 16$ contour line for this function? What is the slope of that line?

2.2

Suppose a firm's total revenues depend on the amount produced (q) according to the function

$$\boldsymbol{R = 70q - q^2}.$$

Total costs also depend on q:

$$\boldsymbol{C = q^2 + 30q + 100}.$$

a. What level of output should the firm produce in order to maximize profits $(R - C)$? What will profits be?

b. Show that the second-order conditions for a maximum are satisfied at the output level found in part (a).

c. Does the solution calculated here obey the "marginal revenue equals marginal cost" rule? Explain.

2.3

Suppose that $f(x, y) = xy$. Find the maximum value for f if x and y are constrained to sum to 1. Solve this problem in two ways: by substitution and by using the Lagrangian multiplier method.

2.4

The dual problem to the one described in Problem 2.3 is

$$\begin{array}{ll} \textbf{minimize} & \boldsymbol{x + y} \\ \textbf{subject to} & \boldsymbol{xy = 0.25}. \end{array}$$

Solve this problem using the Lagrangian technique. Then compare the value you get for the Lagrangian multiplier to the value you got in Problem 2.3. Explain the relationship between the two solutions.

2.5

The height of a ball that is thrown straight up with a certain force is a function of the time (t) from which it is released given by $f(t) = -0.5gt^2 + 40t$ (where g is a constant determined by gravity).

a. How does the value of t at which the height of the ball is at a maximum depend on the parameter g?

b. Use your answer to part (a) to describe how maximum height changes as the parameter g changes.

c. Use the envelope theorem to answer part (b) directly.

d. On the Earth $g = 32$, but this value varies somewhat around the globe. If two locations had gravitational constants that differed by 0.1, what would be the difference in the maximum height of a ball tossed in the two places?

2.6

A simple way to model the construction of an oil tanker is to start with a large rectangular sheet of steel that is x feet wide and $3x$ feet long. Now cut a smaller square that is t feet on a side out of each corner of the larger sheet and fold up and weld the sides of the steel sheet to make a traylike structure with no top.

a. Show that the volume of oil that can be held by this tray is given by

$$V = t(x - 2t)(3x - 2t) = 3tx^2 - 8t^2x + 4t^3.$$

b. How should t be chosen so as to maximize V for any given value of x?

c. Is there a value of x that maximizes the volume of oil that can be carried?

d. Suppose that a shipbuilder is constrained to use only 1,000,000 square feet of steel sheet to construct an oil tanker. This constraint can be represented by the equation $3x^2 - 4t^2 = 1{,}000{,}000$ (because the builder can return the cut-out squares for credit). How does the solution to this constrained maximum problem compare to the solutions described in parts (b) and (c)?

2.7

Consider the following constrained maximization problem:

$$\begin{array}{ll} \textbf{maximize} & y = x_1 + 5 \ln x_2 \\ \textbf{subject to} & k - x_1 - x_2 = 0, \end{array}$$

where k is a constant that can be assigned any specific value.

a. Show that if $k = 10$, this problem can be solved as one involving only equality constraints.

b. Show that solving this problem for $k = 4$ requires that $x_1 = -1$.

c. If the x's in this problem must be nonnegative, what is the optimal solution when $k = 4$?

d. What is the solution for this problem when $k = 20$? What do you conclude by comparing this solution to the solution for part (a)?

Note: This problem involves what is called a "quasi-linear function." Such functions provide important examples of some types of behavior in consumer theory—as we shall see.

2.8

Suppose that a firm has a marginal cost function given by $MC(q) = q + 1$.

a. What is this firm's total cost function? Explain why total costs are known only up to a constant of integration, which represents fixed costs.

b. As you may know from an earlier economics course, if a firm takes price (p) as given in its decisions then it will produce that output for which $p = MC(q)$. If the firm follows this profit-maximizing rule, how much will it produce when $p = 15$? Assuming that the firm is just breaking even at this price, what are fixed costs?

c. How much will profits for this firm increase if price increases to 20?

d. Show that, if we continue to assume profit maximization, then this firm's profits can be expressed solely as a function of the price it receives for its output.

e. Show that the increase in profits from $p = 15$ to $p = 20$ can be calculated in two ways: (i) directly from the equation derived in part (d); and (ii) by integrating the inverse marginal cost function $[MC^{-1}(p) = p - 1]$ from $p = 15$ to $p = 20$. Explain this result intuitively using the envelope theorem.

Analytical Problems

2.9 Concave and quasi-concave functions

Show that if $f(x_1, x_2)$ is a concave function then it is also a quasi-concave function. Do this by comparing Equation 2.114 (defining quasi-concavity) to Equation 2.98 (defining concavity). Can you give an intuitive reason for this result? Is the converse of the statement true? Are quasi-concave functions necessarily concave? If not, give a counterexample.

2.10 The Cobb-Douglas function

One of the most important functions we will encounter in this book is the Cobb-Douglas function:

$$y = (x_1)^{\alpha}(x_2)^{\beta},$$

where α and β are positive constants that are each less than 1.

a. Show that this function is quasi-concave using a "brute force" method by applying Equation 2.114.

b. Show that the Cobb-Douglas function is quasi-concave by showing that any contour line of the form $y = c$ (where c is any positive constant) is convex and therefore that the set of points for which $y > c$ is a convex set.

c. Show that if $\alpha + \beta > 1$ then the Cobb-Douglas function is not concave (thereby illustrating again that not all quasi-concave functions are concave).

Note: The Cobb-Douglas function is discussed further in the Extensions to this chapter.

2.11 The power function

Another function we will encounter often in this book is the "power function":

$$y = x^{\delta},$$

where $0 \leq \delta \leq 1$ (at times we will also examine this function for cases where δ can be negative, too, in which case we will use the form $y = x^{\delta}/\delta$ to ensure that the derivatives have the proper sign).

a. Show that this function is concave (and therefore also, by the result of Problem 2.9, quasi-concave). Notice that the $\delta = 1$ is a special case and that the function is "strictly" concave only for $\delta < 1$.

b. Show that the multivariate form of the power function

$$y = f(x_1, x_2) = (x_1)^{\delta} + (x_2)^{\delta}$$

is also concave (and quasi-concave). Explain why, in this case, the fact that $f_{12} = f_{21} = 0$ makes the determination of concavity especially simple.

c. One way to incorporate "scale" effects into the function described in part (b) is to use the monotonic transformation

$$g(x_1, x_2) = y^{\gamma} = [(x_1)^{\delta} + (x_2)^{\delta}]^{\gamma},$$

where γ is a positive constant. Does this transformation preserve the concavity of the function? Is g quasi-concave?

2.12 Taylor approximations

Taylor's theorem shows that any function can be approximated in the vicinity of any convenient point by a series of terms involving the function and its derivatives. Here we look at some applications of the theorem for functions of one and two variables.

a. Any continuous and differentiable function of a single variable, $f(x)$, can be approximated near the point a by the formula

$$f(x) = f(a) + f'(a)(x - a) + 0.5f''(a)(x - a)^2 + \textbf{terms in } f''', f'''', \ldots.$$

Using only the first three of these terms results in a *quadratic* Taylor approximation. Use this approximation together with the definition of concavity given in Equation 2.85 to show that any concave function must lie on or below the tangent to the function at point a.

b. The quadratic Taylor approximation for any function of two variables, $f(x, y)$, near the point (a, b) is given by

$$\begin{aligned} f(x, y) = {} & f(a, b) + f_1(a, b)(x - a) + f_2(a, b)(y - b) \\ & + 0.5[f_{11}(a,b)(x - a)^2 + 2f_{12}(a, b)(x - a)(y - b) + f_{22}(y - b)^2]. \end{aligned}$$

Use this approximation to show that any concave function (as defined by Equation 2.98) must lie on or below its tangent plane at (a, b).

2.13 More on expected value

Because the expected value concept plays an important role in many economic theories, it may be useful to summarize a few more properties of this statistical measure. Throughout this problem, x is assumed to be a continuous random variable with probability density function $f(x)$.

a. (Jensen's inequality) Suppose that $g(x)$ is a concave function. Show that $E[g(x)] \le g[E(x)]$. *Hint:* Construct the tangent to $g(x)$ at the point $E(x)$. This tangent will have the form $c + dx \ge g(x)$ for all values of x and $c + dE(x) = g[E(x)]$ where c and d are constants.

b. Use the procedure from part (a) to show that if $g(x)$ is a convex function then $E[g(x)] \ge g[E(x)]$.

c. Suppose x takes on only nonnegative values—that is, $0 \le x \le \infty$. Use integration by parts to show that

$$E(x) = \int_0^\infty [1 - F(x)]\, dx,$$

where $F(x)$ is the cumulative distribution function for x [that is, $F(x) = \int_0^x f(t)\, dt$].

d. (Markov's inequality) Show that if x takes on only positive values then the following inequality holds:

$$P(x \ge t) \le \frac{E(x)}{t}.$$

Hint: $E(x) = \int_0^\infty xf(x)\, dx = \int_0^t xf(x)\, dx + \int_t^\infty xf(x)\, dx$.

e. Consider the probability density function $f(x) = 2x^{-3}$ for $x \ge 1$.

(1) Show that this is a proper PDF.

(2) Calculate $F(x)$ for this PDF.

(3) Use the results of part (c) to calculate $E(x)$ for this PDF.

(4) Show that Markov's inequality holds for this function.

f. The concept of conditional expected value is useful in some economic problems. We denote the expected value of x conditional on the occurrence of some event, A, as $E(x|A)$. To compute this value we need to know the PDF for x given that A has occurred [denoted by $f(x|A)$]. With this notation, $E(x|A) = \int_{-\infty}^{+\infty} xf(x|A)\,dx$. Perhaps the easiest way to understand these relationships is with an example. Let

$$f(x) = \frac{x^2}{3} \quad \text{for} \quad -1 \le x \le 2.$$

(1) Show that this is a proper PDF.

(2) Calculate $E(x)$.

(3) Calculate the probability that $-1 \le x \le 0$.

(4) Consider the event $0 \le x \le 2$, and call this event A. What is $f(x|A)$?

(5) Calculate $E(x|A)$.

(6) Explain your results intuitively.

2.14 More on variances and covariances

This problem presents a few useful mathematical facts about variances and covariances.

a. Show that $\mathrm{Var}(x) = E(x^2) - [E(x)]^2$.

b. Show that the result in part (a) can be generalized as $\mathrm{Cov}(x, y) = E(xy) - E(x)E(y)$. *Note:* If $\mathrm{Cov}(x, y) = 0$, then $E(xy) = E(x)E(y)$.

c. Show that $\mathrm{Var}(ax \pm by) = a^2\,\mathrm{Var}(x) + b^2\,\mathrm{Var}(y) \pm 2ab\,\mathrm{Cov}(x, y)$.

d. Assume that two independent random variables, x and y, are characterized by $E(x) = E(y)$ and $\mathrm{Var}(x) = \mathrm{Var}(y)$. Show that $E(0.5x + 0.5y) = E(x)$. Then use part (c) to show that $\mathrm{Var}(0.5x + 0.5y) = 0.5\,\mathrm{Var}(x)$. Describe why this fact provides the rationale for diversification of assets.

SUGGESTIONS FOR FURTHER READING

Dadkhan, Kamran. *Foundations of Mathematical and Computational Economics.* Mason, OH: Thomson/South-Western, 2007.

This is a good introduction to many calculus techniques. The book shows how many mathematical questions can be approached using popular software programs such as Matlab or Excel.

Dixit, A. K. *Optimization in Economic Theory,* 2nd ed. New York: Oxford University Press, 1990.

A complete and modern treatment of optimization techniques. Uses relatively advanced analytical methods.

Hoy, Michael, John Livernois, Chris McKenna, Ray Rees, and Thanasis Stengos. *Mathematics for Economists,* 2nd ed. Cambridge, MA: MIT Press, 2001.

A complete introduction to most of the mathematics covered in microeconomics courses. The strength of the book is its presentation of many worked-out examples, most of which are based on micro-economic theory.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory.* New York: Oxford University Press, 1995.

Encyclopedic treatment of mathematical microeconomics. Extensive mathematical appendices cover relatively high-level topics in analysis.

Samuelson, Paul A. *Foundations of Economic Analysis.* Cambridge, MA: Harvard University Press, 1947. Mathematical Appendix A.

A basic reference. Mathematical Appendix A provides an advanced treatment of necessary and sufficient conditions for a maximum.

Silberberg, E., and W. Suen. *The Structure of Economics: A Mathematical Analysis,* 3rd ed. Boston: Irwin/McGraw-Hill, 2001.

A mathematical microeconomics text that stresses the observable predictions of economic theory. The text makes extensive use of the envelope theorem.

Simon, Carl P., and Lawrence Blume. *Mathematics for Economists.* New York: W. W. Norton, 1994.

A very useful text covering most areas of mathematics relevant to economists. Treatment is at a relatively high level. Two topics discussed better here than elsewhere are differential equations and basic point-set topology.

Sydsaeter, K., A. Strom, and P. Berck. *Economists' Mathematical Manual,* 3rd ed. Berlin: Springer-Verlag, 2000.

An indispensable tool for mathematical review. Contains 32 chapters covering most of the mathematical tools that economists use.

Discussions are very brief, so this is not the place to encounter new concepts for the first time.

Taylor, Angus E., and W. Robert Mann. *Advanced Calculus,* 3rd ed. New York: John Wiley, 1983, pp. 183–95.

A comprehensive calculus text with a good discussion of the Lagrangian technique.

Thomas, George B., and Ross L. Finney. *Calculus and Analytic Geometry,* 8th ed. Reading, MA: Addison-Wesley, 1992.

Basic calculus text with excellent coverage of differentiation techniques.

EXTENSIONS

Second-Order Conditions and Matrix Algebra

The second-order conditions described in Chapter 2 can be written in very compact ways by using matrix algebra. In this extension, we look briefly at that notation. We return to this notation at a few other places in the extensions and problems for later chapters.

Matrix algebra background

The extensions presented here assume some general familiarity with matrix algebra. A succinct reminder of these principles might include:

1. An $n \times k$ *matrix*, **A,** is a rectangular array of terms of the form

$$A = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix}.$$

   Here $i = 1, n$; $j = 1, k$. Matrices can be added, subtracted, or multiplied providing their dimensions are conformable.

2. If $n = k$, then **A** is a square matrix. A square matrix is symmetric if $a_{ij} = a_{ji}$. The *identity matrix*, $\mathbf{I}_n$, is an $n + n$ square matrix where $a_{ij} = 1$ if $i = j$ and $a_{ij} = 0$ if $i \neq j$.

3. The **determinant** of a square matrix (denoted by |A|) is a scalar (i.e., a single term) found by suitably multiplying together all of the terms in the matrix. If **A** is 2×2,

$$|\mathbf{A}| = a_{11}a_{22} - a_{21}a_{12}.$$

$$\text{Example: If } A = \begin{bmatrix} 1 & 3 \\ 5 & 2 \end{bmatrix} \text{ then}$$

$$|\mathbf{A}| = 2 - 15 = -13.$$

4. The *inverse* of an $n \times n$ square matrix, **A**, is another $n \times n$ matrix, A^{-1}, such that

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}_n.$$

   Not every square matrix has an inverse. A necessary and sufficient condition for the existence of $\mathbf{A}^{-1}$ is that $|\mathbf{A}| \neq 0$.

5. The *leading principal minors* of an $n \times n$ square matrix **A** are the series of determinants of the first p rows and columns of **A,** where $p = 1, n$. If **A** is 2×2, then the first leading principal minor is a_{11} and the second is $a_{11}a_{22} - a_{21}a_{12}$.

6. An $n \times n$ square matrix, **A,** is *positive definite* if all of its leading principal minors are positive. The matrix is *negative definite* if its principal minors alternate in sign starting with a minus.[1]

7. A particularly useful symmetric matrix is the *Hessian matrix* formed by all of the second-order partial derivatives of a function. If f is a continuous and twice differentiable function of n variables, then its Hessian is given by

$$\mathbf{H}(f) = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & & & \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}.$$

   Using these notational ideas, we can now examine again some of the second-order conditions derived in Chapter 2.

E2.1 Concave and convex functions

A *concave function* is one that is always below (or on) any tangent to it. Alternatively, a *convex function* is always above (or on) any tangent. The concavity or convexity of any function is determined by its second derivative(s). For a function of a single variable, $f(x)$, the requirement is straightforward. Using the Taylor approximation at any point (x_0)

$$f(x_0 + dx) = f(x_0) + f'(x_0)dx + f''(x_0)\frac{dx^2}{2} + \textbf{higher-order terms}.$$

Assuming that the higher-order terms are 0, we have

$$f(x_0 + dx) \leq f(x_0) + f'(x_0)dx$$

if $f''(x_0) \leq 0$ and

$$f(x_0 + dx) \geq f(x_0) + f'(x_0)dx$$

if $f''(x_0) \geq 0$. Because the expressions on the right of these inequalities are in fact the equation of the tangent to the function at x_0, it is clear that the

[1]If some of the determinants in this definition are 0 then the matrix is said to be positive semidefinite or negative semidefinite.

function is (locally) concave if $f''(x_0) \leq 0$ and (locally) convex if $f''(x_0) \geq 0$.

Extending this intuitive idea to many dimensions is cumbersome in terms of functional notation, but relatively simple when matrix algebra is used. Concavity requires that the Hessian matrix be negative definite whereas convexity requires that this matrix be positive definite. As in the single variable case, these conditions amount to requiring that the function move consistently away from any tangent to it no matter what direction is taken.[2]

If $f(x_1, x_2)$ is a function of two variables, the Hessian is given by

$$\mathbf{H} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}.$$

This is negative definite if

$$f_{11} < 0 \quad \text{and} \quad f_{11}f_{22} - f_{21}f_{12} > 0,$$

which is precisely the condition described in Equation 2.98. Generalizations to functions of three or more variables follow the same matrix pattern.

Example 1

For the health status function in Chapter 2 (Equation 2.20), the Hessian is given by

$$\mathbf{H} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix},$$

and the first and second leading principal minors are

$$\mathbf{H}_1 = -2 < 0 \quad \text{and}$$
$$\mathbf{H}_2 = (-2)(-2) - 0 = 4 > 0.$$

Hence, the function is concave.

Example 2

The Cobb-Douglas function $x^a y^b$ where $a, b \in (0, 1)$ is used to illustrate utility functions and production functions in many places in this text. The first- and second-order derivatives of the function are

$$f_x = ax^{a-1}y^b,$$
$$f_y = bx^a y^{b-1},$$
$$f_{xx} = a(a-1)x^{a-2}y^b,$$
$$f_{yy} = b(b-1)x^a y^{b-2}.$$

Hence, the Hessian for this function is

$$\mathbf{H} = \begin{bmatrix} a(a-1)x^{a-2}y^b & abx^{a-1}y^{b-1} \\ abx^{a-1}y^{b-1} & b(b-1)x^a y^{b-2} \end{bmatrix}.$$

The first leading principal minor of this Hessian is

$$\mathbf{H}_1 = a(a-1)x^{a-2}y^b < 0$$

and so the function will be concave, providing

$$\begin{aligned} \mathbf{H}_2 &= a(a-1)(b)(b-1)x^{2a-2}y^{2b-2} - a^2b^2x^{2a-2}y^{2b-2} \\ &= ab(1-a-b)x^{2a-2}y^{2b-2} > 0. \end{aligned}$$

This condition clearly holds if $a + b < 1$. That is, in production function terminology, the function must exhibit diminishing returns to scale to be concave. Geometrically, the function must turn downward as both inputs are increased together.

E2.2 Maximization

As we saw in Chapter 2, the first-order conditions for an unconstrained maximum of a function of many variables requires finding a point at which the partial derivatives are zero. If the function is concave it will be below its tangent plane at this point and therefore the point will be a true maximum.[3] Because the health status function is concave, for example, the first-order conditions for a maximum are also sufficient.

E2.3 Constrained maxima

When the x's in a maximization or minimization problem are subject to constraints, these constraints have to be taken into account in stating second-order conditions. Again, matrix algebra provides a compact (if not very intuitive) way of denoting these conditions. The notation involves adding rows and columns of the Hessian matrix for the unconstrained problem and then checking the properties of this augmented matrix.

Specifically, we wish to maximize

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

subject to the constraint[4]

$$g(x_1, \ldots, x_n) = 0.$$

[2] A proof using the multivariable version of Taylor's approximation is provided in Simon and Blume (1994), chap. 21.

[3] This will be a "local" maximum if the function is concave only in a region, or "global" if the function is concave everywhere.

[4] Here we look only at the case of a single constraint. Generalization to many constraints is conceptually straightforward but notationally complex. For a concise statement see Sydsaeter, Strom, and Berck (2000), p. 93.

We saw in Chapter 2 that the first-order conditions for a maximum are of the form

$$f_i + \lambda g_i = 0,$$

where λ is the Lagrangian multiplier for this problem. Second-order conditions for a maximum are based on the augmented ("bordered") Hessian[5]

$$\mathbf{H}_\mathrm{b} = \begin{bmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & f_{11} & f_{12} & & f_{1n} \\ g_2 & f_{21} & f_{22} & & f_{2n} \\ \vdots & & & & \\ g_n & f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}.$$

For a maximum, $(-1)\mathbf{H}_\mathrm{b}$ must be negative definite—that is, the leading principal minors of $\mathbf{H}_\mathrm{b}$ must follow the pattern $- + - + -$ and so forth, starting with the second such minor.[6]

The second-order conditions for minimum require that $(-1)\mathbf{H}_\mathrm{b}$ be positive definite—that is, all of the leading principal minors of $\mathbf{H}_\mathrm{b}$ (except the first) should be negative.

Example

The Lagrangian for the constrained health status problem (Example 2.6) is

$$\mathcal{L} = -x_1^2 + 2x_1 - x_2^2 + 4x_2 + 5 + \lambda(1 - x_1 - x_2),$$

and the bordered Hessian for this problem is

$$\mathbf{H}_\mathrm{b} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & -2 & 0 \\ -1 & 0 & -2 \end{bmatrix}.$$

The second leading principal minor here is

$$\mathbf{H}_{\mathrm{b}2} = \begin{bmatrix} 0 & -1 \\ -1 & -2 \end{bmatrix} = -1,$$

and the third is

$$\mathbf{H}_{\mathrm{b}3} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & -2 & 0 \\ -1 & 0 & -2 \end{bmatrix}$$
$$= 0 + 0 + 0 - (-2) - 0 - (-2) = 4,$$

so the leading principal minors of the $\mathbf{H}_\mathrm{b}$ have the required pattern and the point

$$x_2 = 1, \quad x_1 = 0,$$

is a constrained maximum.

[5]Notice that, if $g_{ij} = 0$ for all i and j, then $\mathbf{H}_\mathrm{b}$ can be regarded as the simple Hessian associated with the Lagrangian expression given in Equation 2.50, which is a function of the $n+1$ variables $\lambda, x_1, \ldots, x_n$.

[6]Notice that the first leading principal minor of $\mathbf{H}_\mathrm{b}$ is 0.

Example

In the optimal fence problem (Example 2.7), the bordered Hessian is

$$\mathbf{H}_\mathrm{b} = \begin{bmatrix} 0 & -2 & -2 \\ -2 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

and

$$\mathbf{H}_{\mathrm{b}2} = -4,$$
$$\mathbf{H}_{\mathrm{b}3} = 8,$$

so again the leading principal minors have the sign pattern required for a maximum.

E2.4 Quasi-concavity

If the constraint g is linear, then the second-order conditions explored in Extension 2.3 can be related solely to the shape of the function to be optimized, f. In this case the constraint can be written as

$$g(x_1, \ldots, x_n) = c - b_1x_1 - b_2x_2 - \cdots - b_nx_n = 0,$$

and the first-order conditions for a maximum are

$$f_i = \lambda b_i, \quad i = 1, \ldots, n.$$

Using the conditions, it is clear that the bordered Hessian $\mathbf{H}_\mathrm{b}$ and the matrix

$$\mathbf{H}' = \begin{bmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & & f_{1n} \\ f_2 & f_{21} & f_{22} & & f_{2n} \\ f_n & f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$$

have the same leading principal minors except for a (positive) constant of proportionality.[7] The conditions for a maximum of f subject to a linear constraint will be satisfied provided $\mathbf{H}'$ follows the same sign conventions as $\mathbf{H}_\mathrm{b}$—that is, $(-1)\mathbf{H}'$ must be negative definite. A function f for which $\mathbf{H}'$ does follow this pattern is called *quasi-concave*. As we shall see, f has the property that the set of points x for which $f(x) \ge c$ (where c is any constant) is convex. For such a function, the necessary conditions for a maximum are also sufficient.

Example

For the fences problem, $f(x, y) = xy$ and $\mathbf{H}'$ is given by

[7]This can be shown by noting that multiplying a row (or a column) of a matrix by a constant multiplies the determinant by that constant.

$$\mathbf{H}' = \begin{bmatrix} \mathbf{0} & y & x \\ y & \mathbf{0} & \mathbf{1} \\ x & \mathbf{1} & \mathbf{0} \end{bmatrix}.$$

So

$$\mathbf{H}'_2 = -y^2 < \mathbf{0},$$
$$\mathbf{H}'_3 = \mathbf{2}xy > \mathbf{0},$$

and the function is quasi-concave.[8]

Example

More generally, if f is a function of only two variables, then quasi-concavity requires that

$$\mathbf{H}'_2 = -(f_1)^2 < \mathbf{0} \quad \textbf{and}$$
$$\mathbf{H}'_3 = -f_{11}f_2^2 - f_{22}f_1^2 + \mathbf{2}f_1 f_2 f_{12} > \mathbf{0},$$

which is precisely the condition stated in Equation 2.114. Hence, we have a fairly simple way of determining quasi-concavity.

References

Simon, C. P., and L. Blume. *Mathematics for Economists.* New York: W.W. Norton, 1994.

Sydsaeter, R., A. Strom, and P. Berck. *Economists' Mathematical Manual*, 3rd ed. Berlin: Springer-Verlag, 2000.

[8]Since $f(x, y) = xy$ is a form of a Cobb-Douglas function that is not concave, this shows that not every quasi-concave function is concave. Notice that a monotonic function of f (such as $f^{1/3}$) would be concave, however.