Chapter 15

Continuous Time: First-Order Differential Equations

In the Domar growth model, we have solved a simple differential equation by direct integration. For more complicated differential equations, there are various established methods of solution. Even in the latter cases, however, the fundamental idea underlying the methods of solution is still the techniques of integral calculus. For this reason, the solution to a differential equation is often referred to as the *integral* of that equation.

Only *first-order* differential equations will be discussed in the present chapter. In this context, the word *order* refers to the highest order of the derivatives (or differentials) appearing in the differential equation; thus a first-order differential equation can contain only the first derivative, say, dy/dt.

15.1 First-Order Linear Differential Equations with Constant Coefficient and Constant Term

The first derivative dy/dt is the only one that can appear in a first-order differential equation, but it may enter in various powers: dy/dt, $(dy/dt)^2$, or $(dy/dt)^3$. The highest power attained by the derivative in the equation is referred to as the *degree* of the differential equation. In case the derivative dy/dt appears only in the first degree, and so does the dependent variable y, and furthermore, no product of the form $y(dy/dt)$ occurs, then the equation is said to be *linear*. Thus a first-order linear differential equation will generally take the form†

$$\frac{dy}{dt} + u(t)y = w(t) \qquad (15.1)$$

† Note that the derivative term dy/dt in (15.1) has a unit coefficient. This is not to imply that it can never actually have a coefficient other than one, but when such a coefficient appears, we can always "normalize" the equation by dividing each term by the said coefficient. For this reason, the form given in (15.1) may nonetheless be regarded as a *general* representation.

where u and w are two functions of t, as is y. In contrast to dy/dt and y, however, no restriction whatsoever is placed on the independent variable t. Thus the functions u and w may very well represent such expressions as t^2 and e^t or some more complicated functions of t; on the other hand, u and w may also be constants.

This last point leads us to a further classification. When the function u (the coefficient of the dependent variable y) is a constant, and when the function w is a constant additive term, (15.1) reduces to the special case of a first-order linear differential equation with *constant coefficient and constant term.* In this section, we shall deal only with this simple variety of differential equations.

The Homogeneous Case

If u and w are constant functions and if w happens to be identically zero, (15.1) will become

$$\frac{dy}{dt} + ay = 0 \tag{15.2}$$

where a is some constant. This differential equation is said to be *homogeneous* on account of the zero constant term (compare with homogeneous-equation systems). The defining characteristic of a homogeneous equation is that when all the variables (here, dy/dt and y) are multiplied by a given constant, the equation remains valid. This characteristic holds if the constant term is zero, but will be lost if the constant term is not zero.

Equation (15.2) can be written alternatively as

$$\frac{1}{y}\frac{dy}{dt} = -a \tag{15.2'}$$

But you will recognize that the differential equation (14.16) we met in the Domar model is precisely of this form. Therefore, by analogy, we should be able to write the solution of (15.2) or (15.2') *immediately* as follows:

$$y(t) = Ae^{-at} \qquad [\textit{general}\text{ solution}] \tag{15.3}$$

or

$$y(t) = y(0)e^{-at} \qquad [\textit{definite}\text{ solution}] \tag{15.3'}$$

In (15.3), there appears an arbitrary constant A, therefore it is a *general solution*. When any particular value is substituted for A, the solution becomes a *particular solution* of (15.2). There is an infinite number of particular solutions, one for each possible value of A, including the value $y(0)$. This latter value, however, has a special significance: $y(0)$ is the only value that can make the solution satisfy the initial condition. Since this represents the result of definitizing the arbitrary constant, we shall refer to (15.3') as the *definite solution* of the differential equation (15.2) or (15.2').

You should observe two things about the solution of a differential equation: (1) the solution is not a numerical value, but rather a function $y(t)$—a time path if t symbolizes time; and (2) the solution $y(t)$ is free of any derivative or differential expressions, so that as soon as a specific value of t is substituted into it, a corresponding value of y can be calculated directly.

The Nonhomogeneous Case

When a nonzero constant takes the place of the zero in (15.2), we have a *nonhomogeneous* linear differential equation

$$\frac{dy}{dt} + ay = b \tag{15.4}$$

The solution of this equation will consist of the sum of two terms, one of which is called the *complementary function* (which we shall denote by y_c), and the other known as the *particular integral* (to be denoted by y_p). As will be shown, each of these has a significant economic interpretation. Here, we shall present only the method of solution; its rationale will become clear later.

Even though our objective is to solve the *non*homogeneous equation (15.4), frequently we shall have to refer to its homogeneous version, as shown in (15.2). For convenient reference, we call the latter the *reduced equation* of (15.4). The nonhomogeneous equation (15.4) itself can accordingly be referred to as the *complete equation*. It turns out that the complementary function y_c is nothing but the general solution of the reduced equation, whereas the particular integral y_p is simply *any* particular solution of the complete equation.

Our discussion of the homogeneous case has already given us the general solution of the reduced equation, and we may therefore write

$$y_c = Ae^{-at} \qquad \text{[by (15.3)]}$$

What about the particular integral? Since the particular integral is *any* particular solution of the complete equation, we can first try the simplest possible type of solution, namely, y being some constant ($y = k$). If y is a constant, then it follows that $dy/dt = 0$, and (15.4) will become $ay = b$, with the solution $y = b/a$. Therefore, the constant solution will work as long as $a \neq 0$. In that case, we have

$$y_p = \frac{b}{a} \qquad (a \neq 0)$$

The sum of the complementary function and the particular integral then constitutes the general solution of the complete equation (15.4):

$$y(t) = y_c + y_p = Ae^{-at} + \frac{b}{a} \qquad \text{[general solution, case of } a \neq 0] \qquad \textbf{(15.5)}$$

What makes this a general solution is the presence of the arbitrary constant A. We may, of course, definitize this constant by means of an initial condition. Let us say that y takes the value $y(0)$ when $t = 0$. Then, by setting $t = 0$ in (15.5), we find that

$$y(0) = A + \frac{b}{a} \qquad \text{and} \qquad A = y(0) - \frac{b}{a}$$

Thus we can rewrite (15.5) into

$$y(t) = \left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a} \qquad \text{[definite solution, case of } a \neq 0] \qquad \textbf{(15.5')}$$

It should be noted that the use of the initial condition to definitize the arbitrary constant is—and should be—undertaken as the *final* step, after we have found the general solution to the complete equation. Since the values of both y_c and y_p are related to the value of $y(0)$, both of these must be taken into account in definitizing the constant A.

Example 1 Solve the equation $dy/dt + 2y = 6$, with the initial condition $y(0) = 10$. Here, we have $a = 2$ and $b = 6$; thus, by (15.5'), the solution is

$$y(t) = (10 - 3)e^{-2t} + 3 = 7e^{-2t} + 3$$

Example 2 Solve the equation $dy/dt + 4y = 0$, with the initial condition $y(0) = 1$. Since $a = 4$ and $b = 0$, we have

$$y(t) = (1 - 0)e^{-4t} + 0 = e^{-4t}$$

The same answer could have been obtained from (15.3'), the formula for the homogeneous case. The homogeneous equation (15.2) is merely a special case of the nonhomogeneous equation (15.4) when $b = 0$. Consequently, the formula (15.3') is also a special case of formula (15.5') under the circumstance that $b = 0$.

What if $a = 0$, so that the solution in (15.5') is undefined? In that case, the differential equation is of the extremely simple form

$$\frac{dy}{dt} = b \tag{15.6}$$

By straight integration, its general solution can be readily found to be

$$y(t) = bt + c \tag{15.7}$$

where c is an arbitrary constant. The two component terms in (15.7) can, in fact, again be identified as the complementary function and the particular integral of the given differential equation, respectively. Since $a = 0$, the complementary function can be expressed simply as

$$y_c = Ae^{-at} = Ae^{0} = A \qquad (A = \text{an arbitrary constant})$$

As to the particular integral, the fact that the constant solution $y = k$ fails to work in the present case of $a = 0$ suggests that we should try instead a *nonconstant* solution. Let us consider the simplest possible type of the latter, namely, $y = kt$. If $y = kt$, then $dy/dt = k$, and the complete equation (15.6) will reduce to $k = b$, so that we may write

$$y_p = bt \qquad (a = 0)$$

Our new trial solution indeed works! The general solution of (15.6) is therefore

$$y(t) = y_c + y_p = A + bt \qquad [\text{general solution, case of } a = 0] \tag{15.7'}$$

which is identical with the result in (15.7), because c and A are but alternative notations for an arbitrary constant. Note, however, that in the present case, y_c is a constant whereas y_p is a function of time—the exact opposite of the situation in (15.5).

By definitizing the arbitrary constant, we find the definite solution to be

$$y(t) = y(0) + bt \qquad [\text{definite solution, case of } a = 0] \tag{15.7''}$$

Example 3 Solve the equation $dy/dt = 2$, with the initial condition $y(0) = 5$. The solution is, by (15.7''),

$$y(t) = 5 + 2t$$

Verification of the Solution

It is true of all solutions of differential equations that their validity can always be checked by differentiation.

If we try that on the solution (15.5'), we can obtain the derivative

$$\frac{dy}{dt} = -a\left[y(0) - \frac{b}{a}\right]e^{-at}$$

When this expression for dy/dt and the expression for $y(t)$ as shown in (15.5') are substituted into the left side of the differential equation (15.4), that side should reduce exactly to the value of the constant term b on the right side of (15.4) if the solution is correct. Performing this substitution, we indeed find that

$$-a\left[y(0) - \frac{b}{a}\right]e^{-at} + a\left\{\left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a}\right\} = b$$

Thus our solution is correct, provided it also satisfies the initial condition. To check the latter, let us set $t = 0$ in the solution (15.5'). Since the result

$$y(0) = \left[y(0) - \frac{b}{a}\right] + \frac{b}{a} = y(0)$$

is an identity, the initial condition is indeed satisfied.

It is recommended that, as a final step in the process of solving a differential equation, you make it a habit to check the validity of your answer by making sure (1) that the derivative of the time path $y(t)$ is consistent with the given differential equation and (2) that the definite solution satisfies the initial condition.

EXERCISE 15.1

1. Find y_c, y_p, the general solution, and the definite solution, given:
   (a) $\frac{dy}{dt} + 4y = 12$; $y(0) = 2$
   (b) $\frac{dy}{dt} - 2y = 0$; $y(0) = 9$
   (c) $\frac{dy}{dt} + 10y = 15$; $y(0) = 0$
   (d) $2\frac{dy}{dt} + 4y = 6$; $y(0) = 1\frac{1}{2}$
2. Check the validity of your answers to Prob. 1.
3. Find the solution of each of the following by using an appropriate formula developed in the text:
   (a) $\frac{dy}{dt} + y = 4$; $y(0) = 0$
   (b) $\frac{dy}{dt} = 23$; $y(0) = 1$
   (c) $\frac{dy}{dt} - 5y = 0$; $y(0) = 6$
   (d) $\frac{dy}{dt} + 3y = 2$; $y(0) = 4$
   (e) $\frac{dy}{dt} - 7y = 7$; $y(0) = 7$
   (f) $3\frac{dy}{dt} + 6y = 5$; $y(0) = 0$
4. Check the validity of your answers to Prob. 3.

15.2 Dynamics of Market Price

In the (macro) Domar growth model, we found an application of the *homogeneous* case of linear differential equations of the first order. To illustrate the *nonhomogeneous* case, let us present a (micro) dynamic model of the market.

The Framework

Suppose that, for a particular commodity, the demand and supply functions are as follows:

$$Q_d = \alpha - \beta P \qquad (\alpha, \beta > 0)$$
$$Q_s = -\gamma + \delta P \qquad (\gamma, \delta > 0) \tag{15.8}$$

Then, according to (3.4), the equilibrium price should be[1]

$$P^* = \frac{\alpha + \gamma}{\beta + \delta} \qquad (= \text{some positive constant}) \tag{15.9}$$

If it happens that the initial price $P(0)$ is precisely at the level of P^*, the market will clearly be in equilibrium already, and no dynamic analysis will be needed. In the more interesting case of $P(0) \neq P^*$, however, P^* is attainable (if ever) only after a due process of adjustment, during which not only will price change over time but Q_d and Q_s, being functions of P, must change over time as well. In this light, then, the price and quantity variables can *all* be taken to be *functions of time*.

Our dynamic question is this: Given sufficient time for the adjustment process to work itself out, does it tend to bring price to the equilibrium level P^*? That is, does the time path $P(t)$ tend to converge to P^*, as $t \to \infty$?

The Time Path

To answer this question, we must first find the time path $P(t)$. But that, in turn, requires a specific pattern of price change to be prescribed first. In general, price changes are governed by the relative strength of the demand and supply forces in the market. Let us assume, for the sake of simplicity, that the rate of price change (with respect to time) at any moment is always directly proportional to the *excess demand* $(Q_d - Q_s)$ prevailing at that moment. Such a pattern of change can be expressed symbolically as

$$\frac{dP}{dt} = j(Q_d - Q_s) \qquad (j > 0) \tag{15.10}$$

where j represents a (constant) *adjustment coefficient*. With this pattern of change, we can have $dP/dt = 0$ if and only if $Q_d = Q_s$. In this connection, it may be instructive to note two senses of the term *equilibrium price*: the intertemporal sense (P being constant over time) and the market-clearing sense (the equilibrium price being one that equates Q_d and Q_s). In the present model, the two senses happen to coincide with each other, but this may not be true of all models.

By virtue of the demand and supply functions in (15.8), we can express (15.10) specifically in the form

$$\frac{dP}{dt} = j(\alpha - \beta P + \gamma - \delta P) = j(\alpha + \gamma) - j(\beta + \delta)P$$

or

$$\frac{dP}{dt} + j(\beta + \delta)P = j(\alpha + \gamma) \tag{15.10'}$$

[1] We have switched from the symbols (a, b, c, d) of (3.4) to $(\alpha, \beta, \gamma, \delta)$ here to avoid any possible confusion with the use of a and b as parameters in the differential equation (15.4) which we shall presently apply to the market model.

Since this is precisely in the form of the differential equation (15.4), and since the coefficient of P is nonzero, we can apply the solution formula (15.5′) and write the solution—the time path of price—as

$$P(t) = \left[P(0) - \frac{\alpha+\gamma}{\beta+\delta}\right]e^{-j(\beta+\delta)t} + \frac{\alpha+\gamma}{\beta+\delta}$$
$$= [P(0) - P^*]e^{-kt} + P^* \qquad [\text{by (15.9); } k \equiv j(\beta+\delta)] \qquad (15.11)$$

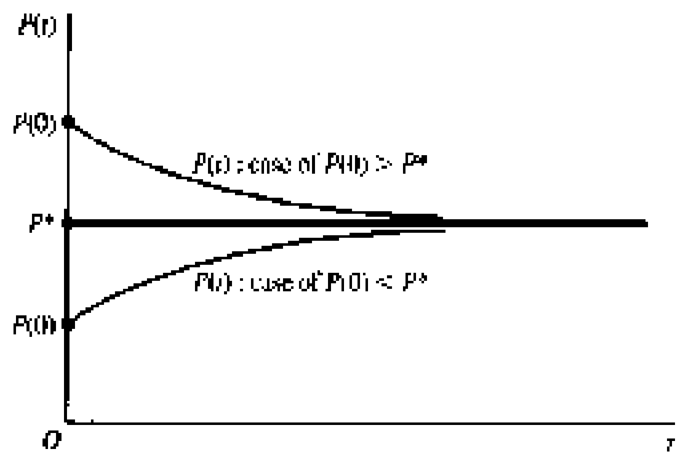
The Dynamic Stability of Equilibrium

In the end, the question originally posed, namely, whether $P(t) \to P^*$ as $t \to \infty$, amounts to the question of whether the first term on the right of (15.11) will tend to zero as $t \to \infty$. Since $P(0)$ and P^* are both constant, the key factor will be the exponential expression e^{-kt}. In view of the fact that $k > 0$, that expression does tend to zero as $t \to \infty$. Consequently, on the assumptions of our model, the time path will indeed lead the price toward the equilibrium position. In a situation of this sort, where the time path of the relevant variable $P(t)$ *converges* to the level P^*—interpreted here in its role as the intertemporal (rather than market-clearing) equilibrium—the equilibrium is said to be *dynamically stable*.

The concept of dynamic stability is an important one. Let us examine it further by a more detailed analysis of (15.11). Depending on the relative magnitudes of $P(0)$ and P^*, the solution (15.11) really encompasses three possible cases. The first is $P(0) = P^*$, which implies $P(t) = P^*$. In that event, the time path of price can be drawn as the horizontal straight line in Fig. 15.1. As mentioned earlier, the attainment of equilibrium is in this case a fait accompli. Second, we may have $P(0) > P^*$. In this case, the first term on the right of (15.11) is positive, but it will decrease as the increase in t lowers the value of e^{-kt}. Thus the time path will approach the equilibrium level P^* from above, as illustrated by the top curve in Fig. 15.1. Third, in the opposite case of $P(0) < P^*$, the equilibrium level P^* will be approached from below, as illustrated by the bottom curve in the same figure. In general, to have dynamic stability, the *deviation* of the time path from equilibrium must either be identically zero (as in case 1) or steadily decrease with time (as in cases 2 and 3).

A comparison of (15.11) with (15.5′) tells us that the P^* term, the counterpart of b/a, is nothing but the particular integral y_p, whereas the exponential term is the (definitized) complementary function y_c. Thus, we now have an economic interpretation for y_c and y_p: y_p represents the *intertemporal equilibrium level* of the relevant variable, and y_c is the *deviation from equilibrium*. Dynamic stability requires the asymptotic vanishing of the complementary function as t becomes infinite.

FIGURE 15.1



In this model, the particular integral is a constant, so we have a *stationary equilibrium* in the intertemporal sense, represented by P^*. If the particular integral is nonconstant, as in (15.7'), on the other hand, we may interpret it as a *moving equilibrium*.

An Alternative Use of the Model

What we have done in the preceding is to analyze the dynamic stability of equilibrium (the convergence of the time path), given certain sign specifications for the parameters. An alternative type of inquiry is: In order to ensure dynamic stability, what specific restrictions must be imposed upon the parameters?

The answer to that is contained in the solution (15.11). If we allow $P(0) \neq P^*$, we see that the first (y_c) term in (15.11) will tend to zero as $t \to \infty$ if and only if $k > 0$—that is, if and only if

$$j(\beta + \delta) > 0$$

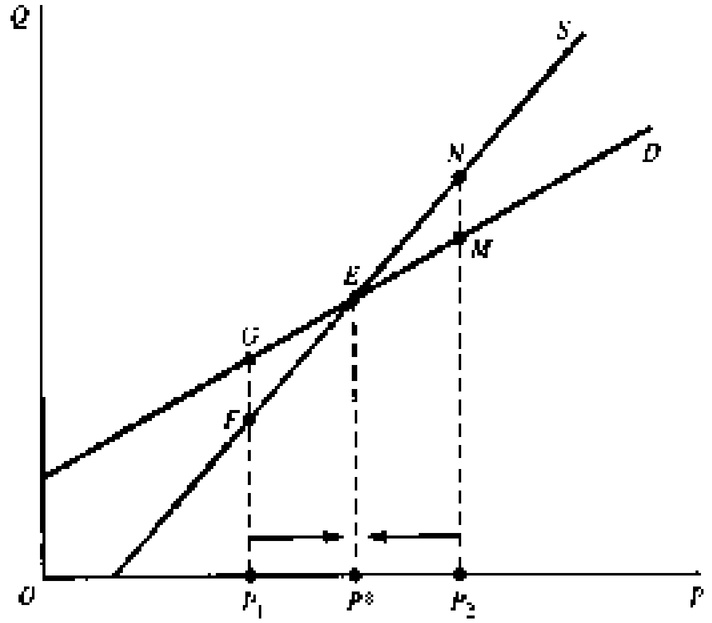
Thus, we can take this last inequality as the required restriction on the parameters j (the adjustment coefficient of price), β (the negative of the slope of the demand curve, plotted with Q on the *vertical* axis), and δ (the slope of the supply curve, plotted similarly).

In case the price adjustment is of the "normal" type, with $j > 0$, so that excess demand drives price up rather than down, then this restriction becomes merely $(\beta + \delta) > 0$ or, equivalently,

$$\delta > -\beta$$

To have dynamic stability in that event, the slope of the supply must exceed the slope of the demand. When both demand and supply are normally sloped ($-\beta < 0$, $\delta > 0$), as in (15.8), this requirement is obviously met. But even if one of the curves is sloped "perversely," the condition may still be fulfilled, such as when $\delta = 1$ and $-\beta = 1/2$ (positively sloped demand). The latter situation is illustrated in Fig. 15.2, where the equilibrium price P^* is, as usual, determined by the point of intersection of the two curves. If the initial price happens to be at P_1, then Q_d (distance P_1G) will exceed Q_s (distance P_1F), and the excess demand (FG) will drive price up. On the other hand, if price is initially at P_2, then

FIGURE 15.2



there will be a *negative* excess demand *MN*, which will drive the price down. As the two arrows in the figure show, therefore, the price adjustment in this case will be *toward* the equilibrium, no matter which side of P^* we start from. We should emphasize, however, that while these arrows can display the direction, they are incapable of indicating the magnitude of change. Thus Fig. 15.2 is basically static, not dynamic, in nature, and can serve only to illustrate, not to replace, the dynamic analysis presented.

EXERCISE 15.2

1. If both the demand and supply in Fig. 15.2 are negatively sloped instead, which curve should be steeper in order to have dynamic stability? Does your answer conform to the criterion $\delta > -\beta$?
2. Show that (15.10') can be rewritten as $dP/dt + k(P - P^*) = 0$. If we let $P - P^* \equiv \Delta$ (signifying deviation), so that $d\Delta/dt = dP/dt$, the differential equation can be further rewritten as
$$\frac{d\Delta}{dt} + k\Delta = 0$$
Find the time path $\Delta(t)$, and discuss the condition for dynamic stability.
3. The dynamic market model discussed in this section is closely patterned after the static one in Sec. 3.2. What specific new feature is responsible for transforming the static model into a dynamic one?
4. Let the demand and supply be
$$Q_d = \alpha - \beta P + \sigma \frac{dP}{dt} \qquad Q_s = -\gamma + \delta P \qquad (\alpha, \beta, \gamma, \delta > 0)$$
(*a*) Assuming that the rate of change of price over time is directly proportional to the excess demand, find the time path $P(t)$ (general solution).
(*b*) What is the intertemporal equilibrium price? What is the market-clearing equilibrium price?
(*c*) What restriction on the parameter σ would ensure dynamic stability?
5. Let the demand and supply be
$$Q_d = \alpha - \beta P - \eta \frac{dP}{dt} \qquad Q_s = \delta P \qquad (\alpha, \beta, \eta, \delta > 0)$$
(*a*) Assuming that the market is cleared at every point of time, find the time path $P(t)$ (general solution).
(*b*) Does this market have a dynamically stable intertemporal equilibrium price?
(*c*) The assumption of the present model that $Q_d = Q_s$ for all t is identical with that of the static market model in Sec. 3.2. Nevertheless, we still have a dynamic model here. How come?

15.3 Variable Coefficient and Variable Term

In the more general case of a first-order linear differential equation

$$\frac{dy}{dt} + u(t)y = w(t) \tag{15.12}$$

$u(t)$ and $w(t)$ represent a variable coefficient and a variable term, respectively. How do we find the time path $y(t)$ in this case?

The Homogeneous Case

For the homogeneous case, where $w(t) = 0$, the solution is still easy to obtain. Since the differential equation is in the form

$$\frac{dy}{dt} + u(t)y = 0 \qquad \text{or} \qquad \frac{1}{y}\frac{dy}{dt} = -u(t) \tag{15.13}$$

we have, by integrating both sides in turn with respect to t,

$$\text{Left side} = \int \frac{1}{y}\frac{dy}{dt}\,dt = \int \frac{dy}{y} = \ln y + c \qquad (\text{assuming } y > 0)$$

$$\text{Right side} = \int -u(t)\,dt = -\int u(t)\,dt$$

In the latter, the integration process cannot be carried further because $u(t)$ has not been given a specific form; thus we have to settle for just a general integral expression. When the two sides are equated, the result is

$$\ln y = -c - \int u(t)\,dt$$

Then the desired y path can be obtained by taking the antilog of $\ln y$:

$$y(t) = e^{\ln y} = e^{-c}e^{-\int u(t)\,dt} = Ae^{-\int u(t)\,dt} \qquad \text{where } A \equiv e^{-c} \tag{15.14}$$

This is the general solution of the differential equation (15.13).

To highlight the variable nature of the coefficient $u(t)$, we have so far explicitly written out the argument t. For notational simplicity, however, we shall from here on omit the argument and shorten $u(t)$ to u.

As compared with the general solution (15.3) for the constant-coefficient case, the only modification in (15.14) is the replacement of the e^{-at} expression by the more complicated expression $e^{-\int u\,dt}$. The rationale behind this change can be better understood if we interpret the at term in e^{-at} as an integral: $\int a\,dt = at$ (plus a constant which can be absorbed into the A term, since e raised to a constant power is again a constant). In this light, the difference between the two general solutions in fact turns into a similarity. For in both cases we are taking the coefficient of the y term in the differential equation—a constant term a in one case, and a variable term u in the other—and integrating that with respect to t, and then taking the negative of the resulting integral as the exponent of e.

Once the general solution is obtained, it is a relatively simple matter to get the definite solution with the help of an appropriate initial condition.

Example 1

Find the general solution of the equation $\frac{dy}{dt} + 3t^2y = 0$. Here we have $u = 3t^2$, and $\int u\,dt = \int 3t^2\,dt = t^3 + c$. Therefore, by (15.14), we may write the solution as

$$y(t) = Ae^{-(t^3+c)} = Ae^{-t^3}e^{-c} = Be^{-t^3} \qquad \text{where } B \equiv Ae^{-c}$$

Observe that if we had omitted the constant of integration c, we would have lost no information, because then we would have obtained $y(t) = Ae^{-t^3}$, which is really the identical solution since A and B both represent arbitrary constants. In other words, the expression e^{-c}, where the constant c makes its only appearance, can always be subsumed under the other constant A.

The Nonhomogeneous Case

For the nonhomogeneous case, where $w(t) \neq 0$, the solution is not as easy to obtain. We shall try to find that solution via the concept of exact differential equations, to be discussed in Sec. 15.4. It does no harm, however, to state the result here first: Given the differential equation (15.12), the general solution is

$$y(t) = e^{-\int u\,dt}\left(A + \int we^{\int u\,dt}\,dt\right) \tag{15.15}$$

where A is an arbitrary constant that can be definitized if we have an appropriate initial condition.

It is of interest that this general solution, like the solution in the constant-coefficient constant-term case, again consists of two additive components. Furthermore, one of these two, $Ae^{-\int u\,dt}$, is nothing but the general solution of the reduced (homogeneous) equation, derived earlier in (15.14), and is therefore in the nature of a complementary function.

Example 2

Find the general solution of the equation $\frac{dy}{dt} + 2ty = t$. Here we have

$$u = 2t \qquad w = t \qquad \text{and} \qquad \int u\,dt = t^2 + k \qquad (k \text{ arbitrary})$$

Thus, by (15.15), we have

$$\begin{aligned} y(t) &= e^{-(t^2+k)}\left(A + \int te^{t^2+k}\,dt\right) \\ &= e^{-t^2}e^{-k}\left(A + e^k\int te^{t^2}\,dt\right) \\ &= Ae^{-k}e^{-t^2} + e^{-t^2}\left(\frac{1}{2}e^{t^2} + c\right) \qquad [e^{-k}e^k = 1] \\ &= (Ae^{-k} + c)e^{-t^2} + \frac{1}{2} \\ &= Be^{-t^2} + \frac{1}{2} \qquad \text{where } B \equiv Ae^{-k} + c \text{ is arbitrary} \end{aligned}$$

The validity of this solution can again be checked by differentiation.

It is interesting to note that, in this example, we could again have omitted the constant of integration k, as well as the constant of integration c, without affecting the final outcome. This is because both k and c may be subsumed under the arbitrary constant B in the final solution. You are urged to try out the simpler process of applying (15.15) without using the constants k and c, and verify that the same solution will emerge.

Example 3

Solve the equation $\frac{dy}{dt} + 4ty = 4t$. This time we shall omit the constants of integration. Since

$$u = 4t \qquad w = 4t \qquad \text{and} \qquad \int u\,dt = 2t^2 \qquad [\text{constant omitted}]$$

the general solution is, by (15.15),

$$\begin{aligned} y(t) &= e^{-2t^2}\left(A + \int 4te^{2t^2}\,dt\right) = e^{-2t^2}\left(A + e^{2t^2}\right) \qquad [\text{constant omitted}] \\ &= Ae^{-2t^2} + 1 \end{aligned}$$

As may be expected, the omission of the constants of integration serves to simplify the procedure substantially.

The differential equation $\frac{dy}{dt} + uy = w$ in (15.12) is more general than the equation $\frac{dy}{dt} + ay = b$ in (15.4), since u and w are not necessarily constant, as are a and b. Accordingly, solution formula (15.15) is also more general than solution formula (15.5). In fact, when we set $u = a$ and $w = b$, (15.15) should reduce to (15.5). This is indeed the case. For when we have

$$u = a \qquad w = b \qquad \text{and} \qquad \int u\,dt = at \qquad \text{[constant omitted]}$$

then (15.15) becomes

$$y(t) = e^{-at}\left(A + \int be^{at}\,dt\right) = e^{-at}\left(A + \frac{b}{a}e^{at}\right) \qquad \text{[constant omitted]}$$

$$= Ae^{-at} + \frac{b}{a}$$

which is identical with (15.5).

EXERCISE 15.3

Solve the following first-order linear differential equations; if an initial condition is given, definitize the arbitrary constant:

1. $\frac{dy}{dt} + 5y = 15$
2. $\frac{dy}{dt} + 2ty = 0$
3. $\frac{dy}{dt} + 2ty = t;\ y(0) = \frac{3}{2}$
4. $\frac{dy}{dt} + t^2 y = 5t^2;\ y(0) = 6$
5. $2\frac{dy}{dt} + 12y + 2e^t = 0;\ y(0) = \frac{6}{7}$
6. $\frac{dy}{dt} + y = t$

15.4 Exact Differential Equations

We shall now introduce the concept of exact differential equations and use the solution method pertaining thereto to obtain the solution formula (15.15) previously cited for the differential equation (15.12). Even though our immediate purpose is to use it to solve a *linear* differential equation, an exact differential equation can be either linear or nonlinear by itself.

Exact Differential Equations

Given a function of two variables $F(y, t)$, its total differential is

$$dF(y,t) = \frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial t}\,dt$$

When this differential is set equal to zero, the resulting equation

$$\frac{\partial F}{\partial y}\,dy + \frac{\partial F}{\partial t}\,dt = 0$$

is known as an *exact differential equation*, because its left side is exactly the differential of the function $F(y, t)$. For instance, given

$$F(y, t) = y^2 t + k \qquad (k \text{ a constant})$$

the total differential is

$$dF = 2yt\,dy + y^2\,dt$$

thus the differential equation

$$2yt\,dy + y^2\,dt = 0 \qquad \text{or} \qquad \frac{dy}{dt} + \frac{y^2}{2yt} = 0 \tag{15.16}$$

is exact.

In general, a differential equation

$$M\,dy + N\,dt = 0 \tag{15.17}$$

is exact if and only if there exists a function $F(y, t)$ such that $M = \partial F/\partial y$ and $N = \partial F/\partial t$. By Young's theorem, which states that $\partial^2 F/\partial t\,\partial y = \partial^2 F/\partial y\,\partial t$, however, we can also state that (15.17) is exact if and only if

$$\frac{\partial M}{\partial t} = \frac{\partial N}{\partial y} \tag{15.18}$$

This last equation gives us a simple test for the exactness of a differential equation. Applied to (15.16), where $M = 2yt$ and $N = y^2$, this test yields $\partial M/\partial t = 2y = \partial N/\partial y$; thus the exactness of the said differential equation is duly verified.

Note that no restrictions have been placed on the terms M and N with regard to the manner in which the variable y occurs. Thus an exact differential equation may very well be *nonlinear* (in y). Nevertheless, it will always be of the first order and the first degree.

Being exact, the differential equation merely says

$$dF(y, t) = 0$$

Thus its general solution should clearly be in the form

$$F(y, t) = c$$

To solve an exact differential equation is basically, therefore, to search for the (primitive) function $F(y, t)$ and then set it equal to an arbitrary constant. Let us outline a method of finding this for the equation $M\,dy + N\,dt = 0$.

Method of Solution

To begin with, since $M = \partial F/\partial y$, the function F must contain the integral of M with respect to the variable y; hence we can write out a preliminary result—in a yet indeterminate form—as follows:

$$F(y, t) = \int M\,dy + \psi(t) \tag{15.19}$$

Here M, a *partial* derivative, is to be integrated with respect to y only; that is, t is to be treated as a constant in the integration process, just as it was treated as a constant in the partial differentiation of $F(y, t)$ that resulted in $M = \partial F/\partial y$.[†] Since, in differentiating $F(y, t)$ partially with respect to y, any additive term containing only the variable t and/or some constants (but with no y) would drop out, we must now take care to reinstate such terms in the integration process. This explains why we have introduced in (15.19) a general term $\psi(t)$, which, though not exactly the same as a constant of integration, has a precisely identical role to play as the latter. It is relatively easy to get $\int M\,dy$; but how do we pin down the exact form of this $\psi(t)$ term?

The trick is to utilize the fact that $N = \partial F/\partial t$. But the procedure is best explained with the help of specific examples.

Example 1 Solve the exact differential equation

$$2yt\,dy + y^2\,dt = 0 \qquad \text{[reproduced from (15.16)]}$$

In this equation, we have

$$M = 2yt \qquad \text{and} \qquad N = y^2$$

STEP i By (15.19), we can first write the preliminary result

$$F(y, t) = \int 2yt\,dy + \psi(t) = y^2t + \psi(t)$$

Note that we have omitted the constant of integration, because it can automatically be merged into the expression $\psi(t)$.

STEP ii If we differentiate the result from Step i partially with respect to t, we can obtain

$$\frac{\partial F}{\partial t} = y^2 + \psi'(t)$$

But since $N = \partial F/\partial t$, we can equate $N = y^2$ and $\partial F/\partial t = y^2 + \psi'(t)$, to get

$$\psi'(t) = 0$$

STEP iii Integration of the last result gives us

$$\psi(t) = \int \psi'(t)\,dt = \int 0\,dt = k$$

and now we have a specific form of $\psi(t)$. It happens in the present case that $\psi(t)$ is simply a constant; more generally, it can be a nonconstant function of t.

STEP iv The results of Steps i and iii can be combined to yield

$$F(y, t) = y^2t + k$$

The solution of the exact differential equation should then be $F(y, t) = c$. But since the constant k can be merged into c, we may write the solution simply as

$$y^2t = c \qquad \text{or} \qquad y(t) = ct^{-1/2}$$

where c is arbitrary.

[†] Some writers employ the operator symbol $\int(\cdots)\,\partial y$ to emphasize that the integration is with respect to y only. We shall still use the symbol $\int(\cdots)\,dy$ here, since there is little possibility of confusion.

Example 2 Solve the equation $(t + 2y)\,dy + (y + 3t^2)\,dt = 0$. First let us check whether this is an exact differential equation. Setting $M = t + 2y$ and $N = y + 3t^2$, we find that $\partial M/\partial t = 1 = \partial N/\partial y$. Thus the equation passes the exactness test. To find its solution, we again follow the procedure outlined in Example 1.

STEP i Apply (15.19) and write

$$F(y, t) = \int (t + 2y)\,dy + \psi(t) = yt + y^2 + \psi(t) \qquad \text{[constant merged into } \psi(t)]$$

STEP ii Differentiate this result with respect to t, to get

$$\frac{\partial F}{\partial t} = y + \psi'(t)$$

Then, equating this to $N = y + 3t^2$, we find that

$$\psi'(t) = 3t^2$$

STEP iii Integrate this last result to get

$$\psi(t) = \int 3t^2\,dt = t^3 \qquad \text{[constant may be omitted]}$$

STEP iv Combine the results of Steps i and iii to get the complete form of the function $F(y, t)$:

$$F(y, t) = yt + y^2 + t^3$$

which implies that the solution of the given differential equation is

$$yt + y^2 + t^3 = c$$

You should verify that setting the total differential of this equation equal to zero will indeed produce the given differential equation.

This four-step procedure can be used to solve any exact differential equation. Interestingly, it may even be applicable when the given equation is *not* exact. To see this, however, we must first introduce the concept of integrating factor.

Integrating Factor

Sometimes an inexact differential equation can be made exact by multiplying every term of the equation by a particular common factor. Such a factor is called an *integrating factor*.

Example 3 The differential equation

$$2t\,dy + y\,dt = 0$$

is not exact, because it does not satisfy (15.18):

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial t}(2t) = 2 \neq \frac{\partial N}{\partial y} = \frac{\partial}{\partial y}(y) = 1$$

However, if we multiply each term by y, the given equation will turn into (15.16), which has been established to be exact. Thus y is an integrating factor for the differential equation in the present example.

When an integrating factor can be found for an inexact differential equation, it is always possible to render it exact, and then the four-step solution procedure can be readily put to use.

Solution of First-Order Linear Differential Equations

The general first-order linear differential equation

$$\frac{dy}{dt} + uy = w$$

which, in the format of (15.17), can be expressed as

$$dy + (uy - w)\,dt = 0 \tag{15.20}$$

has the integrating factor

$$e^{\int u\,dt} \equiv \exp\left(\int u\,dt\right)$$

This integrating factor, whose form is by no means intuitively obvious, can be "discovered" as follows. Let I be the (yet unknown) integrating factor. Multiplication of (15.20) through by I should convert it into an exact differential equation

$$\underbrace{I}_{M}\,dy + \underbrace{I(uy - w)}_{N}\,dt = 0 \tag{15.20'}$$

The exactness test dictates that $\partial M/\partial t = \partial N/\partial y$. Visual inspection of the M and N expressions suggests that, since M consists of I only, and since u and w are functions of t alone, the exactness test will reduce to a very simple condition if I is also a function of t alone. For then the test $\partial M/\partial t = \partial N/\partial y$ becomes

$$\frac{dI}{dt} = Iu \qquad \text{or} \qquad \frac{dI/dt}{I} = u$$

Thus the special form $I = I(t)$ can indeed work, provided it has a rate of growth equal to u, or more explicitly, $u(t)$. Accordingly, $I(t)$ should take the specific form

$$I(t) = Ae^{\int u\,dt} \qquad \text{[cf. (15.13) and (15.14)]}$$

As can be easily verified, however, the constant A can be set equal to 1 without affecting the ability of $I(t)$ to meet the exactness test. Thus we can use the simpler form $e^{\int u\,dt}$ as the integrating factor.

Substitution of this integrating factor into (15.20′) yields the exact differential equation

$$e^{\int u\,dt}\,dy + e^{\int u\,dt}(uy - w)\,dt = 0 \tag{15.20''}$$

which can then be solved by the four-step procedure.

STEP i First, we apply (15.19) to obtain

$$F(y, t) = \int e^{\int u\,dt}\,dy + \psi(t) = ye^{\int u\,dt} + \psi(t)$$

The result of integration emerges in this simple form because the integrand is independent of the variable y.

STEP ii Next, we differentiate the result from Step i with respect to t, to get

$$\frac{\partial F}{\partial t} = yue^{\int u\,dt} + \psi'(t) \qquad \text{[chain rule]}$$

And, since this can be equated to $N = e^{\int u\,dt}(uy - w)$, we have

$$\psi'(t) = -we^{\int u\,dt}$$

STEP iii Straight integration now yields

$$\psi(t) = -\int we^{\int u\,dt}\,dt$$

Inasmuch as the functions $u = u(t)$ and $w = w(t)$ have not been given specific forms, nothing further can be done about this integral, and we must be contented with this rather general expression for $\psi(t)$.

STEP iv Substituting this $\psi(t)$ expression into the result of Step i, we find that

$$F(y, t) = ye^{\int u\,dt} - \int we^{\int u\,dt}\,dt$$

So the general solution of the exact differential equation (15.20')—and of the equivalent, though inexact, first-order linear differential equation (15.20)—is

$$ye^{\int u\,dt} - \int we^{\int u\,dt}\,dt = c$$

Upon rearrangement and substitution of the (arbitrary constant) symbol c by A, this can be written as

$$y(t) = e^{-\int u\,dt}\left(A + \int we^{\int u\,dt}\,dt\right) \qquad (15.21)$$

which is exactly the result given earlier in (15.15).

EXERCISE 15.4

1. Verify that each of the following differential equations is exact, and solve by the four-step procedure:
   (a) $2yt^3\,dy + 3y^2t^2\,dt = 0$
   (b) $3y^2t\,dy + (y^3 + 2t)\,dt = 0$
   (c) $t(1 + 2y)\,dy + y(1 + y)\,dt = 0$
   (d) $\frac{dy}{dt} + \frac{2y^4t + 3t^2}{4y^3t^2} = 0$ [*Hint:* First convert to the form of (15.17).]
2. Are the following differential equations exact? If not, try t, y, and y^2 as possible integrating factors.
   (a) $2(t^3 + 1)\,dy + 3yt^2\,dt = 0$
   (b) $4y^3t\,dy + (2y^4 + 3t)\,dt = 0$
3. By applying the four-step procedure to the general exact differential equation $M\,dy + N\,dt = 0$, derive the following formula for the general solution of an exact differential equation:

$$\int M\,dy + \int N\,dt - \int\left(\frac{\partial}{\partial t}\int M\,dy\right)dt = c$$

15.5 Nonlinear Differential Equations of the First Order and First Degree

In a *linear* differential equation, we restrict to the *first degree* not only the derivative dy/dt, but also the dependent variable y, and we do not allow the product $y(dy/dt)$ to appear. When y appears in a power higher than one, the equation becomes *nonlinear* even if it only contains the derivative dy/dt in the first degree. In general, an equation in the form

$$f(y, t)\,dy + g(y, t)\,dt = 0 \tag{15.22}$$

or

$$\frac{dy}{dt} = h(y, t) \tag{15.22'}$$

where there is no restriction on the powers of y and t, constitutes a first-order first-degree nonlinear differential equation because dy/dt is a first-order derivative in the first power. Certain varieties of such equations can be solved with relative ease by more or less routine procedures. We shall briefly discuss three cases.

Exact Differential Equations

The first is the now-familiar case of exact differential equations. As was pointed out earlier, the y variable can appear in an exact equation in a high power, as in (15.16)—$2yt\,dy + y^2\,dt = 0$—which you should compare with (15.22). True, the cancellation of the common factor y from both terms on the left will reduce the equation to a linear form, but the exactness property will be lost in that event. As an *exact* differential equation, therefore, it must be regarded as nonlinear.

Since the solution method for exact differential equations has already been discussed, no further comment is necessary here.

Separable Variables

The differential equation in (15.22)

$$f(y, t)\,dy + g(y, t)\,dt = 0$$

may happen to possess the convenient property that the function f is in the variable y alone, while the function g involves only the variable t, so that the equation reduces to the special form

$$f(y)\,dy + g(t)\,dt = 0 \tag{15.23}$$

In such an event, the variables are said to be *separable*, because the terms involving y—consolidated into $f(y)$—can be mathematically separated from the terms involving t, which are collected under $g(t)$. To solve this special type of equation, only simple integration techniques are required.

Example 1 Solve the equation $3y^2\,dy - t\,dt = 0$. First let us rewrite the equation as

$$3y^2\,dy = t\,dt$$

Integrating the two sides (each of which is a differential) and equating the results, we get

$$\int 3y^2\,dy = \int t\,dt \qquad \text{or} \qquad y^3 + c_1 = \frac{1}{2}t^2 + c_2$$

Thus the general solution can be written as

$$y^3 = \frac{1}{2}t^2 + c \qquad \text{or} \qquad y(t) = \left(\frac{1}{2}t^2 + c\right)^{1/3}$$

The notable point here is that the integration of each term is performed with respect to a different variable; it is this which makes the separable-variable equation comparatively easy to handle.

Example 2 Solve the equation $2t\,dy + y\,dt = 0$. At first glance, this differential equation does not seem to belong in this spot, because it fails to conform to the general form of (15.23). To be specific, the coefficients of dy and dt are seen to involve the "wrong" variables. However, a simple transformation—dividing through by $2yt$ ($\neq 0$)—will reduce the equation to the separable-variable form

$$\frac{1}{y}\,dy + \frac{1}{2t}\,dt = 0$$

From our experience with Example 1, we can work toward the solution (without first transposing a term) as follows:[1]

$$\int \frac{1}{y}\,dy + \int \frac{1}{2t}\,dt = c$$

so

$$\ln y + \frac{1}{2}\ln t = c \qquad \text{or} \qquad \ln(yt^{1/2}) = c$$

Thus the solution is

$$yt^{1/2} = e^c = k \qquad \text{or} \qquad y(t) = kt^{-1/2}$$

where k is an arbitrary constant, as are the symbols c and A employed elsewhere.

Note that, instead of solving the equation in Example 2 as we did, we could also have transformed it first into an exact differential equation (by the integrating factor y) and then solved it as such. The solution, already given in Example 1 of Sec. 15.4, must of course be identical with the one just obtained by separation of variables. The point is that a given differential equation can often be solvable in more than one way, and therefore one may have a choice of the method to be used. In other cases, a differential equation that is not amenable to a particular method may nonetheless become so after an appropriate transformation.

Equations Reducible to the Linear Form

If the differential equation $dy/dt = h(y, t)$ happens to take the specific nonlinear form

$$\frac{dy}{dt} + Ry = Ty^m \tag{15.24}$$

where R and T are two functions of t, and m is any number other than 0 and 1 (what if $m = 0$ or $m = 1$?), then the equation—referred to as a *Bernoulli equation*—can always be reduced to a linear differential equation and be solved as such.

[1] In the integration result, we should, strictly speaking, have written $\ln|y|$ and $\frac{1}{2}\ln|t|$. If y and t can be assumed to be positive, as is appropriate in the majority of economic contexts, then the result given in the text will occur.

The reduction procedure is relatively simple. First, we can divide (15.24) by y^m, to get

$$y^{-m}\frac{dy}{dt} + Ry^{1-m} = T$$

If we adopt a shorthand variable z as follows:

$$z = y^{1-m} \qquad \left[\text{so that } \frac{dz}{dt} = \frac{dz}{dy}\frac{dy}{dt} = (1-m)y^{-m}\frac{dy}{dt}\right]$$

then the preceding equation can be written as

$$\frac{1}{1-m}\frac{dz}{dt} + Rz = T$$

Moreover, after multiplying through by $(1-m)\,dt$ and rearranging, we can transform the equation into

$$dz + [(1-m)Rz - (1-m)T]\,dt = 0 \qquad \textbf{(15.24')}$$

This is seen to be a first-order linear differential equation of the form (15.20), in which the variable z has taken the place of y.

Clearly, we can apply formula (15.21) to find its solution $z(t)$. Then, as a final step, we can translate z back to y by reverse substitution.

Example 3 Solve the equation $dy/dt + ty = 3ty^2$. This is a Bernoulli equation, with $m = 2$ (giving us $z = y^{1-m} = y^{-1}$), $R = t$, and $T = 3t$. Thus, by (15.24'), we can write the linearized differential equation as

$$dz + (-tz + 3t)\,dt = 0$$

By applying formula (15.21), the solution can be found to be

$$z(t) = A\exp\left(\tfrac{1}{2}t^2\right) + 3$$

(As an exercise, trace out the steps leading to this solution.)

Since our primary interest lies in the solution $y(t)$ rather than $z(t)$, we must perform a reverse transformation using the equation $z = y^{-1}$, or $y = z^{-1}$. By taking the reciprocal of $z(t)$, therefore, we get

$$y(t) = \frac{1}{A\exp\left(\tfrac{1}{2}t^2\right) + 3}$$

as the desired solution. This is a *general solution*, because an arbitrary constant A is present.

Example 4 Solve the equation $dy/dt + (1/t)y = y^3$. Here, we have $m = 3$ (thus $z = y^{-2}$), $R = 1/t$, and $T = 1$; thus the equation can be linearized into the form

$$dz + \left(\frac{-2}{t}z + 2\right)dt = 0$$

As you can verify, by the use of formula (15.21), the solution of this differential equation is

$$z(t) = At^2 + 2t$$

It then follows, by the reverse transformation $y = z^{-1/2}$, that the general solution in the original variable is to be written as

$$y(t) = (At^2 + 2t)^{-1/2}$$

As an exercise, check the validity of the solutions of these last two examples by differentiation.

EXERCISE 15.5

1. Determine, for each of the following, (1) whether the variables are separable and (2) whether the equation is linear or else can be linearized:

   (a) $2t\,dy + 2y\,dt = 0$ (c) $\frac{dy}{dt} = -\frac{t}{y}$

   (b) $\frac{y}{y+t}\,dy + \frac{2t}{y+t}\,dt = 0$ (d) $\frac{dy}{dt} = 3y^2t$
2. Solve (a) and (b) in Prob. 1 by separation of variables, taking y and t to be positive. Check your answers by differentiation.
3. Solve (c) in Prob. 1 as a separable-variable equation and, also, as a Bernoulli equation.
4. Solve (d) in Prob. 1 as a separable-variable equation and, also, as a Bernoulli equation.
5. Verify the correctness of the intermediate solution $z(t) = At^2 + 2t$ in Example 4 by showing that its derivative dz/dt is consistent with the linearized differential equation.

15.6 The Qualitative-Graphic Approach

The several cases of nonlinear differential equations previously discussed (exact differential equations, separable-variable equations, and Bernoulli equations) have all been solved *quantitatively*. That is, we have in every case sought and found a time path $y(t)$ which, for each value of t, tells the specific corresponding value of the variable y.

At times, we may not be able to find a quantitative solution from a given differential equation. Yet, in such cases, it may nonetheless be possible to ascertain the *qualitative* properties of the time path—primarily, whether $y(t)$ converges—by directly observing the differential equation itself or by analyzing its graph. Even when quantitative solutions are available, moreover, we may still employ the techniques of qualitative analysis if the qualitative aspect of the time path is our principal or exclusive concern.

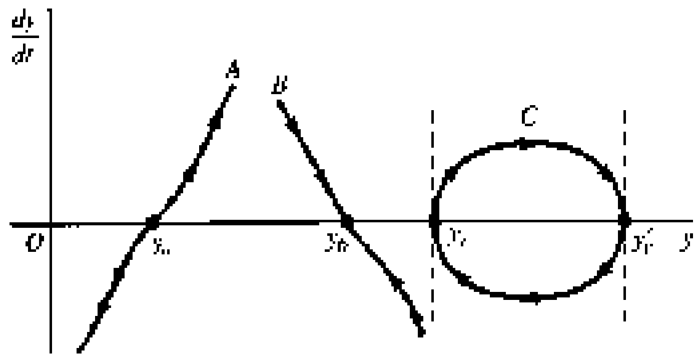
The Phase Diagram

Given a first-order differential equation in the general form

$$\frac{dy}{dt} = f(y)$$

either linear or nonlinear in the variable y, we can plot dy/dt against y as in Fig. 15.3. Such a geometric representation, feasible whenever dy/dt is a function of y alone, is called a *phase diagram*, and the graph representing the function f, a *phase line*. (A differential equation of this form—in which the time variable t does not appear as a separate argument of

FIGURE 15.3



the function f—is said to be an *autonomous* differential equation.) Once a phase line is known, its configuration will impart significant qualitative information regarding the time path $y(t)$. The clue to this lies in the following two general remarks:

1. Anywhere *above* the horizontal axis (where $dy/dt > 0$), y must be increasing over time and, as far as the y axis is concerned, must be moving from left to right. By analogous reasoning, any point *below* the horizontal axis must be associated with a leftward movement in the variable y, because the negativity of dy/dt means that y decreases over time. These directional tendencies explain why the arrowheads on the illustrative phase lines in Fig. 15.3 are drawn as they are. Above the horizontal axis, the arrows are uniformly pointed toward the right—toward the northeast or southeast or due east, as the case may be. The opposite is true below the y axis. Moreover, these results are independent of the algebraic sign of y; even if phase line A (or any other) is transplanted to the left of the vertical axis, the direction of the arrows will not be affected.
2. An equilibrium level of y—in the intertemporal sense of the term—if it exists, can occur only on the horizontal axis, where $dy/dt = 0$ (y stationary over time). To find an equilibrium, therefore, it is necessary only to consider the intersection of the phase line with the y axis.[†] To test the dynamic stability of equilibrium, on the other hand, we should also check whether, regardless of the initial position of y, the phase line will always guide it toward the equilibrium position at the said intersection.

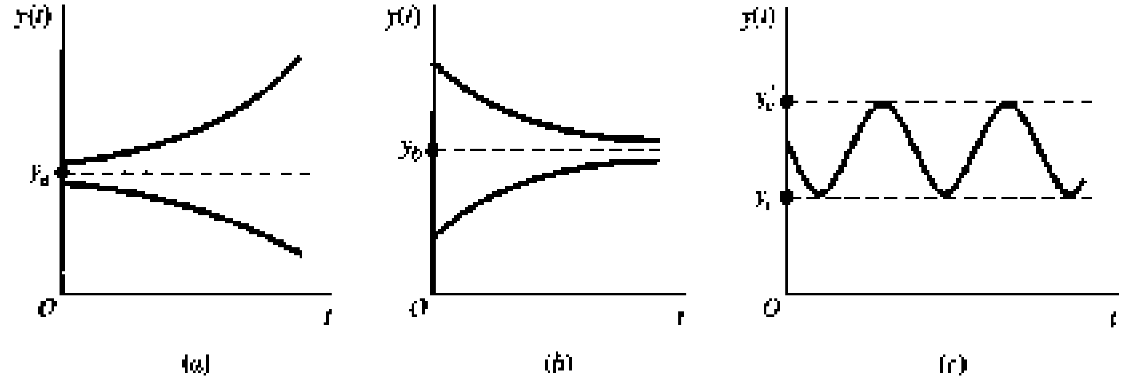
Types of Time Path

On the basis of the preceding general remarks, we may observe three different types of time path from the illustrative phase lines in Fig. 15.3.

Phase line A has an equilibrium at point y_a; but *above* as well as *below* that point, the arrowheads consistently lead away from equilibrium. Thus, although equilibrium can be attained if it happens that $y(0) = y_a$, the more usual case of $y(0) \neq y_a$ will result in y being ever-increasing [if $y(0) > y_a$] or ever-decreasing [if $y(0) < y_a$]. Besides, in this case the deviation of y from y_a tends to grow at an increasing pace because, as we follow the arrowheads on the phase line, we deviate farther from the y axis, thereby encountering ever-increasing numerical values of dy/dt as well. The time path $y(t)$ implied by phase line A can therefore be represented by the curves shown in Fig. 15.4a, where y is plotted against t (rather than dy/dt against y). The equilibrium y_a is dynamically unstable.

[†] However, not all intersections represent equilibrium positions. We shall see this when we discuss phase line C in Fig. 15.3.

FIGURE 15.4



In contrast, phase line B implies a stable equilibrium at y_b. If $y(0) = y_b$, equilibrium prevails at once. But the important feature of phase line B is that, even if $y(0) \neq y_b$, the movement along the phase line will guide y toward the level of y_b. The time path $y(t)$ corresponding to this type of phase line should therefore be of the form shown in Fig. 15.4*b*, which is reminiscent of the dynamic market model.

The preceding discussion suggests that, in general, it is the slope of the phase line at its intersection point which holds the key to the dynamic stability of equilibrium or the convergence of the time path. A (finite) *positive* slope, such as at point y_a, makes for dynamic *instability*; whereas a (finite) *negative* slope, such as at y_b, implies dynamic *stability*.

This generalization can help us to draw qualitative inferences about given differential equations without even plotting their phase lines. Take the linear differential equation in (15.4), for instance:

$$\frac{dy}{dt} + ay = b \qquad \text{or} \qquad \frac{dy}{dt} = -ay + b$$

Since the phase line will obviously have the (constant) slope $-a$, here assumed nonzero, we may immediately infer (without drawing the line) that

$$a \gtrless 0 \quad \Leftrightarrow \quad y(t) \begin{Bmatrix} \text{converges to} \\ \text{diverges from} \end{Bmatrix} \text{equilibrium}$$

As we may expect, this result coincides perfectly with what the quantitative solution of this equation tells us:

$$y(t) = \left[y(0) - \frac{b}{a}\right]e^{-at} + \frac{b}{a} \qquad [\text{from (15.5')}]$$

We have learned that, starting from a nonequilibrium position, the convergence of $y(t)$ hinges on the prospect that $e^{-at} \to 0$ as $t \to \infty$. This can happen if and only if $a > 0$; if $a < 0$, then $e^{-at} \to \infty$ as $t \to \infty$, and $y(t)$ cannot converge. Thus, our conclusion is one and the same, whether it is arrived at quantitatively or qualitatively.

It remains to discuss phase line C, which, being a closed loop sitting across the horizontal axis, does not qualify as a *function* but shows instead a *relation* between dy/dt and y.[†] The interesting new element that emerges in this case is the possibility of a periodically fluctuating time path. The way that phase line C is drawn, we shall find y fluctuating between the two values y_c and y_c' in a perpetual motion. In order to generate the periodic

[†] This can arise from a second-degree differential equation $(dy/dt)^2 = f(y)$.

fluctuation, the loop must, of course, straddle the horizontal axis in such a manner that dy/dt can alternately be positive and negative. Besides, at the two intersection points y_c and y_c', the phase line should have an infinite slope; otherwise the intersection will resemble either y_a or y_b, neither of which permits a continual flow of arrowheads. The type of time path $y(t)$ corresponding to this looped phase line is illustrated in Fig. 15.4c. Note that, whenever $y(t)$ hits the upper bound y_c' or the lower bound y_c, we have $dy/dt = 0$ (local extrema); but these values certainly do not represent equilibrium values of y. In terms of Fig. 15.3, this means that not all intersections between a phase line and the y axis are equilibrium positions.

In sum, for the study of the dynamic stability of equilibrium (or the convergence of the time path), one has the alternative either of finding the time path itself or else of simply drawing the inference from its phase line. We shall illustrate the application of the latter approach with the Solow growth model. Henceforth, we shall denote the intertemporal equilibrium value of y by $\bar{y}$, as distinct from y^*.

EXERCISE 15.6

1. Plot the phase line for each of the following, and discuss its qualitative implications:
   (a) $\frac{dy}{dt} = y - 7$ (c) $\frac{dy}{dt} = 4 - \frac{y}{2}$
   (b) $\frac{dy}{dt} = 1 - 5y$ (d) $\frac{dy}{dt} = 9y - 11$
2. Plot the phase line for each of the following and interpret:
   (a) $\frac{dy}{dt} = (y+1)^2 - 16 \qquad (y \geq 0)$
   (b) $\frac{dy}{dt} = \frac{1}{2}y - y^2 \qquad (y \geq 0)$
3. Given $dy/dt = (y-3)(y-5) = y^2 - 8y + 15$:
   (a) Deduce that there are two possible equilibrium levels of y, one at $y = 3$ and the other at $y = 5$.
   (b) Find the sign of $\frac{d}{dy}\left(\frac{dy}{dt}\right)$ at $y = 3$ and $y = 5$, respectively. What can you infer from these?

15.7 Solow Growth Model

The growth model of Professor Robert Solow,[1] a Nobel laureate, is purported to show, among other things, that the razor's-edge growth path of the Domar model is primarily a result of the particular production-function assumption adopted therein and that, under alternative circumstances, the need for delicate balancing may not arise.

The Framework

In the Domar model, output is explicitly stated as a function of capital alone: $\kappa = \rho K$ (the productive capacity, or potential output, is a constant multiple of the stock of capital). The

[1] Robert M. Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, February 1956, pp. 65–94.

absence of a labor input in the production function carries the implication that labor is always combined with capital in a *fixed* proportion, so that it is feasible to consider explicitly only one of these factors of production. Solow, in contrast, seeks to analyze the case where capital and labor can be combined in *varying* proportions. Thus his production function appears in the form

$$Q = f(K, L) \qquad (K, L > 0)$$

where Q is output (net of depreciation), K is capital, and L is labor—all being used in the *macro* sense. It is assumed that f_K and f_L are positive (positive marginal products), and f_{KK} and f_{LL} are negative (diminishing returns to each input). Furthermore, the production function f is taken to be linearly homogeneous (constant returns to scale). Consequently, it is possible to write

$$Q = Lf\left(\frac{K}{L}, 1\right) = L\phi(k) \qquad \text{where } k \equiv \frac{K}{L} \tag{15.25}$$

In view of the assumed signs of f_K and f_{KK}, the newly introduced ϕ function (which, be it noted, has only a single argument, k) must be characterized by a positive first derivative and a negative second derivative. To verify this claim, we first recall from (12.49) that

$$f_K \equiv \text{MPP}_K = \phi'(k)$$

hence $f_K > 0$ automatically means $\phi'(k) > 0$. Then, since

$$f_{KK} = \frac{\partial}{\partial K}\phi'(k) = \frac{d\phi'(k)}{dk}\frac{\partial k}{\partial K} = \phi''(k)\frac{1}{L} \qquad \text{[see (12.48)]}$$

the assumption $f_{KK} < 0$ leads directly to the result $\phi''(k) < 0$. Thus the ϕ function—which, according to (12.46), gives the APP_L for every capital–labor ratio—is one that increases with k at a decreasing rate.

Given that Q depends on K and L, it is necessary now to stipulate how the latter two variables themselves are determined. Solow's assumptions are:

$$\dot{K}\left(\equiv \frac{dK}{dt}\right) = sQ \qquad \text{[constant proportion of } Q \text{ is invested]} \tag{15.26}$$

$$\frac{\dot{L}}{L}\left(\equiv \frac{dL/dt}{L}\right) = \lambda \qquad (\lambda > 0) \qquad \text{[labor force grows exponentially]} \tag{15.27}$$

The symbol s represents a (constant) marginal propensity to save, and λ, a (constant) rate of growth of labor. Note the dynamic nature of these assumptions; they specify not how the *levels* of K and L are determined, but how their *rates of change* are.

Equations (15.25) through (15.27) constitute a complete model. To solve this model, we shall first condense it into a single equation in one variable. To begin with, substitute (15.25) into (15.26) to get

$$\dot{K} = sL\phi(k) \tag{15.28}$$

Since $k = K/L$, and $K = kL$, however, we can obtain another expression for $\dot{K}$ by differentiating the latter identity:

$$\begin{aligned} \dot{K} &= L\dot{k} + k\dot{L} \qquad &\text{[product rule]} \\ &= L\dot{k} + k\lambda L \qquad &\text{[by (15.27)]} \end{aligned} \tag{15.29}$$

When (15.29) is equated to (15.28) and the common factor L eliminated, the result emerges that

$$\dot{k} = s\phi(k) - \lambda k \tag{15.30}$$

This equation—a differential equation in the variable k, with two parameters s and λ—is the fundamental equation of the Solow growth model.

A Qualitative-Graphic Analysis

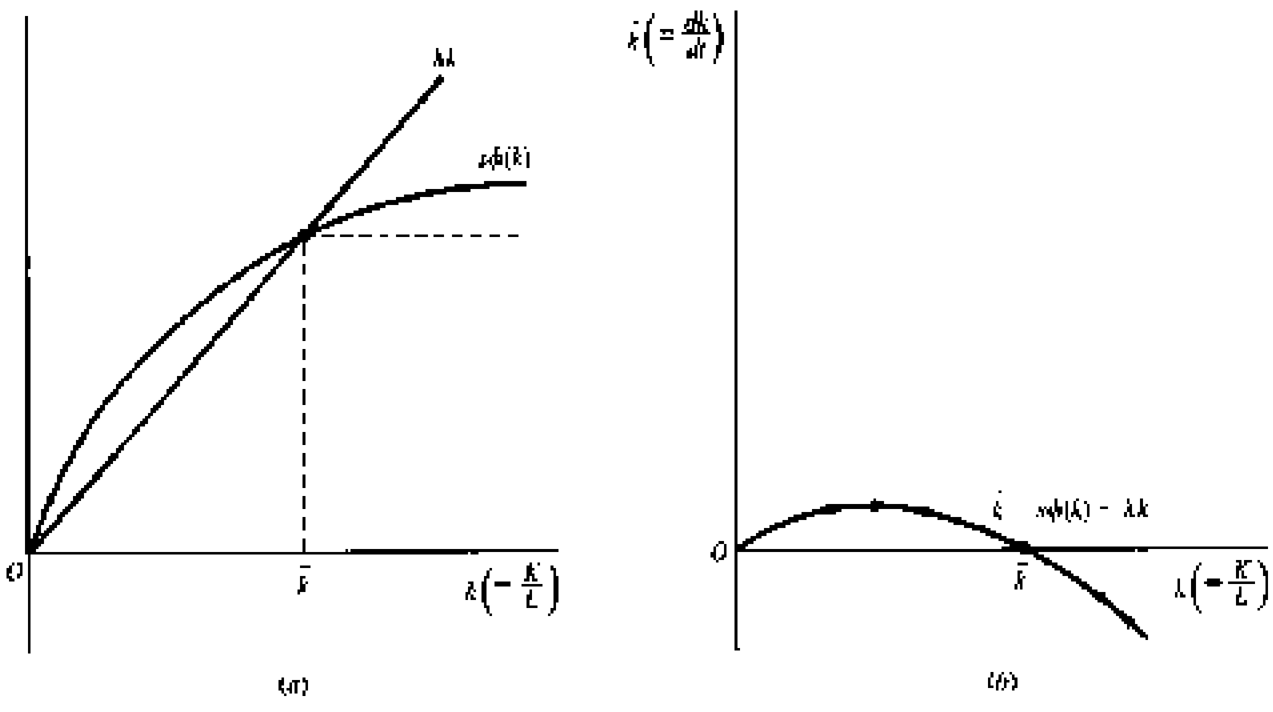
Because (15.30) is stated in a general-function form, no specific quantitative solution is available. Nevertheless, we can analyze it qualitatively. To this end, we should plot a phase line, with $\dot{k}$ on the vertical axis and k on the horizontal.

Since (15.30) contains two terms on the right, however, let us first plot these as two separate curves. The λk term, a linear function of k, will obviously show up in Fig. 15.5*a* as a straight line, with a zero vertical intercept and a slope equal to λ. The $s\phi(k)$ term, on the other hand, plots as a curve that increases at a decreasing rate, like $\phi(k)$, since $s\phi(k)$ is merely a constant fraction of the $\phi(k)$ curve. If we consider K to be an indispensable factor of production, we must start the $s\phi(k)$ curve from the point of origin; this is because if $K = 0$ and thus $k = 0$, Q must also be zero, as will be $\phi(k)$ and $s\phi(k)$. The way the curve is actually drawn also reflects the implicit assumption that there exists a set of k values for which $s\phi(k)$ exceeds λk, so that the two curves intersect at some positive value of k, namely $\bar{k}$.

Based upon these two curves, the value of $\dot{k}$ for each value of k can be measured by the vertical distance between the two curves. Plotting the values of $\dot{k}$ against k, as in Fig. 15.5*b*, will then yield the phase line we need. Note that, since the two curves in Fig. 15.5*a* intersect when the capital–labor ratio is $\bar{k}$, the phase line in Fig. 15.5*b* must cross the horizontal axis at $\bar{k}$. This marks $\bar{k}$ as the intertemporal equilibrium capital–labor ratio.

Inasmuch as the phase line has a negative slope at $\bar{k}$, the equilibrium is readily identified as a stable one; given any (positive) initial value of k, the dynamic movement of the model

FIGURE 15.5



must lead us convergently to the equilibrium level $\bar{k}$. The significant point is that once this equilibrium is attained—and thus the capital–labor ratio is (by definition) unvarying over time—capital must thereafter grow apace with labor, at the identical rate λ. This will imply, in turn, that net investment must grow at the rate λ (see Exercise 15.7-2). Note, however, that the word *must* is used here not in the sense of requirement, but with the implication of automaticity. Thus, what the Solow model serves to show is that, given a rate of growth of labor λ, the economy by itself, and without the delicate balancing à la Domar, can eventually reach a state of steady growth in which investment will grow at the rate λ, the same as K and L. Moreover, in order to satisfy (15.25), Q must grow at the same rate as well because $\phi(k)$ is a constant when the capital–labor ratio remains unvarying at the level $\bar{k}$. Such a situation, in which the relevant variables all grow at an identical rate, is called a *steady state*—a generalization of the concept of *stationary state* (in which the relevant variables all remain constant, or in other words all grow at the zero rate).

Note that, in the preceding analysis, the production function is assumed for convenience to be invariant over time. If the state of technology is allowed to improve, on the other hand, the production function will have to be duly modified. For instance, it may be written instead in the form

$$Q = T(t)f(K, L) \qquad \left(\frac{dT}{dt} > 0\right)$$

where T, some measure of technology, is an increasing function of time. Because of the increasing multiplicative term $T(t)$, a fixed amount of K and L will turn out a larger output at a future date than at present. In this event, the $s\phi(k)$ curve in Fig. 15.5 will be subject to a secular upward shift, resulting in successively higher intersections with the λk ray and also in larger values of $\bar{k}$. With technological improvement, therefore, it will become possible, in a succession of steady states, to have a larger and larger amount of capital equipment available to each representative worker in the economy, with a concomitant rise in productivity.

A Quantitative Illustration

The preceding analysis had to be qualitative, owing to the presence of a general function $\phi(k)$ in the model. But if we specify the production function to be a linearly homogeneous Cobb-Douglas function, for instance, then a quantitative solution can be found as well.

Let us write the production function as

$$Q = K^\alpha L^{1-\alpha} = L\left(\frac{K}{L}\right)^\alpha = Lk^\alpha$$

so that $\phi(k) = k^\alpha$. Then (15.30) becomes

$$\dot{k} = sk^\alpha - \lambda k \qquad \text{or} \qquad \dot{k} + \lambda k = sk^\alpha$$

which is a Bernoulli equation in the variable k [see (15.24)], with $R = \lambda$, $T = s$, and $m = \alpha$. Letting $z = k^{1-\alpha}$, we obtain its linearized version

$$dz + [(1-\alpha)\lambda z - (1-\alpha)s]\,dt = 0$$

or

$$\frac{dz}{dt} + \underbrace{(1-\alpha)\lambda}_{a} z = \underbrace{(1-\alpha)s}_{b}$$

This is a linear differential equation with a constant coefficient a and a constant term b. Thus, by formula (15.5′), we have

$$z(t) = \left[z(0) - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

The substitution of $z = k^{1-\alpha}$ will then yield the final solution

$$k^{1-\alpha} = \left[k(0)^{1-\alpha} - \frac{s}{\lambda}\right]e^{-(1-\alpha)\lambda t} + \frac{s}{\lambda}$$

where $k(0)$ is the initial value of the capital–labor ratio k.

This solution is what determines the time path of k. Recalling that $(1 - \alpha)$ and λ are both positive, we see that as $t \to \infty$ the exponential expression will approach zero; consequently,

$$k^{1-\alpha} \to \frac{s}{\lambda} \qquad \text{or} \qquad k \to \left(\frac{s}{\lambda}\right)^{1/(1-\alpha)} \qquad \text{as } t \to \infty$$

Therefore, the capital–labor ratio will approach a constant as its equilibrium value. This equilibrium or steady-state value, $(s/\lambda)^{1/(1-\alpha)}$, varies directly with the propensity to save s, and inversely with the rate of growth of labor λ.

EXERCISE 15.7

1. Divide (15.30) through by k, and interpret the resulting equation in terms of the growth rates of k, K, and L.
2. Show that, if capital is growing at the rate λ (that is, $K = Ae^{\lambda t}$), net investment I must also be growing at the rate λ.
3. The original input variables of the Solow model are K and L, but the fundamental equation (15.30) focuses on the capital-labor ratio k instead. What assumption(s) in the model is(are) responsible for (and make possible) this shift of focus? Explain.
4. Draw a phase diagram for each of the following, and discuss the qualitative aspects of the time path $y(t)$:

   (a) $\dot{y} = 3 - y - \ln y$ (b) $\dot{y} = e^y - (y + 2)$

Chapter 16

Higher-Order Differential Equations

In Chap. 15, we discussed the methods of solving a *first-order* differential equation, one in which there appears no derivative (or differential) of orders higher than 1. At times, however, the specification of a model may involve the second derivative or a derivative of an even higher order. We may, for instance, be given a function describing "the rate of change of the rate of change" of the income variable Y, say,

$$\frac{d^2Y}{dt^2} = kY$$

from which we are supposed to find the time path of Y. In this event, the given function constitutes a *second-order* differential equation, and the task of finding the time path $Y(t)$ is that of *solving* the second-order differential equation. The present chapter is concerned with the methods of solution and the economic applications of such higher-order differential equations, but we shall confine our discussion to the *linear* case only.

A simple variety of linear differential equations of order n is of the following form:

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_{n-1}\frac{dy}{dt} + a_n y = b \qquad (16.1)$$

or, in an alternative notation,

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_n y = b \qquad (16.1')$$

This equation is of *order n*, because the nth derivative (the first term on the left) is the highest derivative present. It is *linear*, since all the derivatives, as well as the dependent variable y, appear only in the first degree, and moreover, no product term occurs in which y and any of its derivatives are multiplied together. You will note, in addition, that this differential equation is characterized by *constant coefficients* (the a's) and a *constant term* (b). The constancy of the coefficients is an assumption we shall retain throughout this chapter. The constant term b, on the other hand, is adopted here as a first approach; later, in Sec. 16.5, we shall drop it in favor of a variable term.

16.1 Second-Order Linear Differential Equations with Constant Coefficients and Constant Term

For pedagogic reasons, let us first discuss the method of solution for the *second-order* case ($n = 2$). The relevant differential equation is then the simple one

$$y''(t) + a_1 y'(t) + a_2 y = b \qquad (16.2)$$

where a_1, a_2, and b are all constants. If the term b is identically zero, we have a *homogeneous* equation, but if b is a nonzero constant, the equation is *nonhomogeneous*. Our discussion will proceed on the assumption that (16.2) is nonhomogeneous; in solving the nonhomogeneous version of (16.2), the solution of the homogeneous version will emerge automatically as a by-product.

In this connection, we recall a proposition introduced in Sec. 15.1 which is equally applicable here: If y_c is the *complementary function*, i.e., the general solution (containing arbitrary constants) of the reduced equation of (16.2) and if y_p is the *particular integral*, i.e., any particular solution (containing no arbitrary constants) of the complete equation (16.2), then $y(t) = y_c + y_p$ will be the general solution of the complete equation. As was explained previously, the y_p component provides us with the equilibrium value of the variable y in the intertemporal sense of the term, whereas the y_c component reveals, for each point of time, the deviation of the time path $y(t)$ from the equilibrium.

The Particular Integral

For the case of constant coefficients and constant term, the particular integral is relatively easy to find. Since the particular integral can be *any* solution of (16.2), i.e., any value of y that satisfies this nonhomogeneous equation, we should always try the simplest possible type: namely, y = a constant. If y = a constant, it follows that

$$y'(t) = y''(t) = 0$$

so that (16.2) in effect becomes $a_2 y = b$, with the solution $y = b/a_2$. Thus, the desired particular integral is

$$y_p = \frac{b}{a_2} \qquad (\text{case of } a_2 \neq 0) \qquad (16.3)$$

Since the process of finding the value of y_p involves the condition $y'(t) = 0$, the rationale for considering that value as an intertemporal equilibrium becomes self-evident.

Example 1 Find the particular integral of the equation

$$y''(t) + y'(t) - 2y = -10$$

The relevant coefficients here are $a_2 = -2$ and $b = -10$. Therefore, the particular integral is $y_p = -10/(-2) = 5$.

What if $a_2 = 0$—so that the expression b/a_2 is not defined? In such a situation, since the constant solution for y_p fails to work, we must try some *nonconstant* form of solution. Taking the simplest possibility, we may try $y = kt$. Since $a_2 = 0$, the differential equation is now

$$y''(t) + a_1 y'(t) = b$$

but if $y = kt$, which implies $y'(t) = k$ and $y''(t) = 0$, this equation reduces to $a_1 k = b$. This determines the value of k as b/a_1, thereby giving us the particular integral

$$y_p = \frac{b}{a_1}t \qquad (\text{case of } a_2 = 0;\ a_1 \neq 0) \tag{16.3'}$$

Inasmuch as y_p is in this case a nonconstant function of time, we shall regard it as a moving equilibrium.

Example 2 Find the y_p of the equation $y''(t) + y'(t) = -10$. Here, we have $a_2 = 0$, $a_1 = 1$, and $b = -10$. Thus, by (16.3'), we can write

$$y_p = -10t$$

If it happens that a_1 is also zero, then the solution form of $y = kt$ will also break down, because the expression bt/a_1 will now be undefined. We ought, then, to try a solution of the form $y = kt^2$. With $a_1 = a_2 = 0$, the differential equation now reduces to the extremely simple form

$$y''(t) = b$$

and if $y = kt^2$, which implies $y'(t) = 2kt$ and $y''(t) = 2k$, the differential equation can be written as $2k = b$. Thus, we find $k = b/2$, and the particular integral is

$$y_p = \frac{b}{2}t^2 \qquad (\text{case of } a_1 = a_2 = 0) \tag{16.3''}$$

The equilibrium represented by this particular integral is again a moving equilibrium.

Example 3 Find the y_p of the equation $y''(t) = -10$. Since the coefficients are $a_1 = a_2 = 0$ and $b = -10$, formula (16.3'') is applicable. The desired answer is $y_p = -5t^2$.

The Complementary Function

The complementary function of (16.2) is defined to be the general solution of its reduced (homogeneous) equation

$$y''(t) + a_1 y'(t) + a_2 y = 0 \tag{16.4}$$

This is why we stated that the solution of a homogeneous equation will always be a *by-product* in the process of solving a complete equation.

Even though we have never tackled such an equation before, our experience with the complementary function of the first-order differential equations can supply us with a useful hint. From the solutions (15.3), (15.3'), (15.5), and (15.5'), it is clear that exponential expressions of the form Ae^{rt} figure very prominently in the complementary functions of first-order differential equations with constant coefficients. Then why not try a solution of the form $y = Ae^{rt}$ in the second-order equation, too?

If we adopt the trial solution $y = Ae^{rt}$, we must also accept

$$y'(t) = rAe^{rt} \qquad \text{and} \qquad y''(t) = r^2 Ae^{rt}$$

as the derivatives of y. On the basis of these expressions for y, $y'(t)$, and $y''(t)$, the reduced differential equation (16.4) can be transformed into

$$Ae^{rt}(r^2 + a_1 r + a_2) = 0 \tag{16.4'}$$

As long as we choose those values of A and r that satisfy (16.4'), the trial solution $y = Ae^{rt}$ should work. Since e^{rt} can never be zero, we must either let $A = 0$ or see to it that r satisfies the equation

$$r^2 + a_1 r + a_2 = 0 \tag{16.4''}$$

Since the value of the (arbitrary) constant A is to be definitized by use of the initial conditions of the problem, however, we cannot simply set $A = 0$ at will. Therefore, it is essential to look for values of r that satisfy (16.4'').

Equation (16.4'') is known as the *characteristic equation* (or *auxiliary equation*) of the homogeneous equation (16.4), or of the complete equation (16.2). Because it is a quadratic equation in r, it yields two roots (solutions), referred to in the present context as *characteristic roots*, as follows:[1]

$$r_1, r_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \tag{16.5}$$

These two roots bear a simple but interesting relationship to each other, which can serve as a convenient means of checking our calculation: The *sum* of the two roots is always equal to $-a_1$, and their *product* is always equal to a_2. The proof of this statement is straightforward:

$$\begin{aligned} r_1 + r_2 &= \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2} + \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2} = \frac{-2a_1}{2} = -a_1 \\ r_1 r_2 &= \frac{(-a_1)^2 - \left(a_1^2 - 4a_2\right)}{4} = \frac{4a_2}{4} = a_2 \end{aligned} \tag{16.6}$$

The values of these two roots are the only values we may assign to r in the solution $y = Ae^{rt}$. But this means that, in effect, there are *two* solutions which will work, namely,

$$y_1 = A_1 e^{r_1 t} \qquad \text{and} \qquad y_2 = A_2 e^{r_2 t}$$

where A_1 and A_2 are two arbitrary constants, and r_1 and r_2 are the characteristic roots found from (16.5). Since we want only *one* general solution, however, there seems to be one too many. Two alternatives are now open to us: (1) pick either y_1 or y_2 at random, or (2) combine them in some fashion.

The first alternative, though simpler, is unacceptable. There is only one arbitrary constant in y_1 or y_2, but to qualify as a general solution of a *second-order* differential equation, the expression must contain *two* arbitrary constants. This requirement stems from the fact that, in proceeding from a function $y(t)$ to its second derivative $y''(t)$, we "lose" two constants during the two rounds of differentiation; therefore, to revert from a second-order differential equation to the primitive function $y(t)$, two constants should be reinstated. That leaves us only the alternative of combining y_1 and y_2, so as to include both constants

[1] Note that the quadratic equation (16.4'') is in the normalized form; the coefficient of the r^2 term is 1. In applying formula (16.5) to find the characteristic roots of a differential equation, we must first make sure that the characteristic equation is indeed in the normalized form.

A_1 and A_2. As it turns out, we can simply take their *sum*, $y_1 + y_2$, as the general solution of (16.4). Let us demonstrate that, if y_1 and y_2, respectively, satisfy (16.4), then the sum $(y_1 + y_2)$ will also do so. If y_1 and y_2 are indeed solutions of (16.4), then by substituting each of these into (16.4), we must find that the following two equations hold:

$$y_1''(t) + a_1 y_1'(t) + a_2 y_1 = 0$$
$$y_2''(t) + a_1 y_2'(t) + a_2 y_2 = 0$$

By adding these equations, however, we find that

$$\underbrace{[y_1''(t) + y_2''(t)]}_{=\frac{d^2}{dt^2}(y_1+y_2)} + a_1 \underbrace{[y_1'(t) + y_2'(t)]}_{=\frac{d}{dt}(y_1+y_2)} + a_2(y_1 + y_2) = 0$$

Thus, like y_1 or y_2, the sum $(y_1 + y_2)$ satisfies the equation (16.4) as well. Accordingly, the general solution of the homogeneous equation (16.4) or the complementary function of the complete equation (16.2) can, in general, be written as $y_c = y_1 + y_2$.

A more careful examination of the characteristic-root formula (16.5) indicates, however, that as far as the values of r_1 and r_2 are concerned, three possible cases can arise, some of which may necessitate a modification of our result $y_c = y_1 + y_2$.

Case 1 (distinct real roots) When $a_1^2 > 4a_2$, the square root in (16.5) is a real number, and the two roots r_1 and r_2 will take *distinct* real values, because the square root is added to $-a_1$ for r_1, but subtracted from $-a_1$ for r_2. In this case, we can indeed write

$$y_c = y_1 + y_2 = A_1 e^{r_1 t} + A_2 e^{r_2 t} \qquad (r_1 \neq r_2) \qquad \textbf{(16.7)}$$

Because the two roots are distinct, the two exponential expressions must be linearly independent (neither is a multiple of the other); consequently, A_1 and A_2 will always remain as separate entities and provide us with two constants, as required.

Example 4

Solve the differential equation

$$y''(t) + y'(t) - 2y = -10$$

The particular integral of this equation has already been found to be $y_p = 5$, in Example 1. Let us find the complementary function. Since the coefficients of the equation are $a_1 = 1$ and $a_2 = -2$, the characteristic roots are, by (16.5),

$$r_1, r_2 = \frac{-1 \pm \sqrt{1+8}}{2} = \frac{-1 \pm 3}{2} = 1, -2$$

(*Check:* $r_1 + r_2 = -1 = -a_1$; $r_1 r_2 = -2 = a_2$.) Since the roots are distinct real numbers, the complementary function is $y_c = A_1 e^t + A_2 e^{-2t}$. Therefore, the general solution can be written as

$$y(t) = y_c + y_p = A_1 e^t + A_2 e^{-2t} + 5 \qquad (16.8)$$

In order to definitize the constants A_1 and A_2, there is need now for *two* initial conditions. Let these conditions be $y(0) = 12$ and $y'(0) = -2$. That is, when $t = 0$, $y(t)$ and $y'(t)$ are, respectively, 12 and -2. Setting $t = 0$ in (16.8), we find that

$$y(0) = A_1 + A_2 + 5$$

Differentiating (16.8) with respect to t and then setting $t = 0$ in the derivative, we find that

$$y'(t) = A_1 e^t - 2A_2 e^{-2t} \qquad \text{and} \qquad y'(0) = A_1 - 2A_2$$

To satisfy the two initial conditions, therefore, we must set $y(0) = 12$ and $y'(0) = -2$, which results in the following pair of simultaneous equations:

$$A_1 + \ \ A_2 = 7$$
$$A_1 - 2A_2 = -2$$

with solutions $A_1 = 4$ and $A_2 = 3$. Thus the definite solution of the differential equation is

$$y(t) = 4e^t + 3e^{-2t} + 5 \tag{16.8'}$$

As before, we can check the validity of this solution by differentiation. The first and second derivatives of (16.8') are

$$y'(t) = 4e^t - 6e^{-2t} \qquad \text{and} \qquad y''(t) = 4e^t + 12e^{-2t}$$

When these are substituted into the given differential equation along with (16.8'), the result is an identity $-10 = -10$. Thus the solution is correct. As you can easily verify, (16.8') also satisfies both of the initial conditions.

Case 2 (repeated real roots) When the coefficients in the differential equation are such that $a_1^2 = 4a_2$, the square root in (16.5) will vanish, and the two characteristic roots take an identical value:

$$r(= r_1 = r_2) = -\frac{a_1}{2}$$

Such roots are known as *repeated roots*, or *multiple* (here, *double*) *roots*.

If we attempt to write the complementary function as $y_c = y_1 + y_2$, the sum will in this case collapse into a single expression

$$y_c = A_1 e^{rt} + A_2 e^{rt} = (A_1 + A_2)e^{rt} = A_3 e^{rt}$$

leaving us with only one constant. This is not sufficient to lead us from a second-order differential equation back to its primitive function. The only way out is to find another eligible component term for the sum—a term which satisfies (16.4) and yet which is linearly independent of the term $A_3 e^{rt}$, so as to preclude such "collapsing."

An expression that will satisfy these requirements is $A_4 t e^{rt}$. Since the variable t has entered into it multiplicatively, this component term is obviously linearly independent of the $A_3 e^{rt}$ term; thus it will enable us to introduce another constant, A_4. But does $A_4 t e^{rt}$ qualify as a solution of (16.4)? If we try $y = A_4 t e^{rt}$, then, by the product rule, we can find its first and second derivatives to be

$$y'(t) = (rt + 1)A_4 e^{rt} \qquad \text{and} \qquad y''(t) = (r^2 t + 2r)A_4 e^{rt}$$

Substituting these expressions of y, y', and y'' into the left side of (16.4), we get the expression

$$[(r^2 t + 2r) + a_1(rt + 1) + a_2 t]A_4 e^{rt}$$

Inasmuch as, in the present context, we have $a_1^2 = 4a_2$ and $r = -a_1/2$, this last expression vanishes identically and thus is always equal to the right side of (16.4); this shows that A_4te^{rt} does indeed qualify as a solution.

Hence, the complementary function of the double-root case can be written as

$$y_c = A_3e^{rt} + A_4te^{rt} \tag{16.9}$$

Example 5 Solve the differential equation

$$y''(t) + 6y'(t) + 9y = 27$$

Here, the coefficients are $a_1 = 6$ and $a_2 = 9$; since $a_1^2 = 4a_2$, the roots will be repeated. According to formula (16.5), we have $r = -a_1/2 = -3$. Thus, in line with the result in (16.9), the complementary function may be written as

$$y_c = A_3e^{-3t} + A_4te^{-3t}$$

The general solution of the given differential equation is now also readily obtainable. Trying a constant solution for the particular integral, we get $y_p = 3$. It follows that the general solution of the complete equation is

$$y(t) = y_c + y_p = A_3e^{-3t} + A_4te^{-3t} + 3$$

The two arbitrary constants can again be definitized with two initial conditions. Suppose that the initial conditions are $y(0) = 5$ and $y'(0) = -5$. By setting $t = 0$ in the preceding general solution, we should find $y(0) = 5$; that is,

$$y(0) = A_3 + 3 = 5$$

This yields $A_3 = 2$. Next, by differentiating the general solution and then setting $t = 0$ and also $A_3 = 2$, we must have $y'(0) = -5$. That is,

$$y'(t) = -3A_3e^{-3t} - 3A_4te^{-3t} + A_4e^{-3t}$$

and $$y'(0) = -6 + A_4 = -5$$

This yields $A_4 = 1$. Thus we can finally write the definite solution of the given equation as

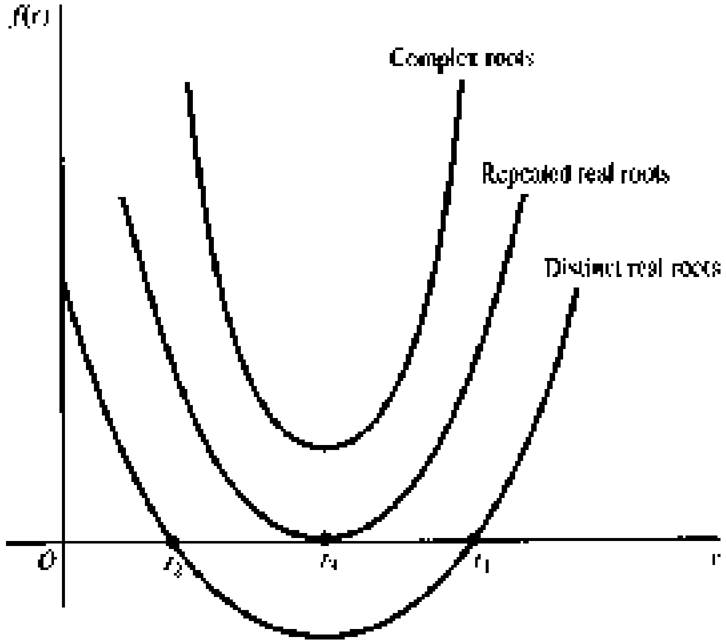
$$y(t) = 2e^{-3t} + te^{-3t} + 3$$

Case 3 (complex roots) There remains a third possibility regarding the relative magnitude of the coefficients a_1 and a_2, namely, $a_1^2 < 4a_2$. When this eventuality occurs, formula (16.5) will involve the square root of a *negative* number, which cannot be handled before we are properly introduced to the concepts of *imaginary* and *complex* numbers. For the time being, therefore, we shall be content with the mere cataloging of this case and shall leave the full discussion of it to Secs. 16.2 and 16.3.

The three cases cited can be illustrated by the three curves in Fig. 16.1, each of which represents a different version of the quadratic function $f(r) = r^2 + a_1r + a_2$. As we learned earlier, when such a function is set equal to zero, the result is a quadratic *equation* $f(r) = 0$, and to solve the latter equation is merely to "find the zeros of the quadratic *function*." Graphically, this means that the roots of the equation are to be found on the horizontal axis, where $f(r) = 0$.

The position of the lowest curve in Fig. 16.1, is such that the curve intersects the horizontal axis twice; thus we can find two distinct roots r_1 and r_2, both of which satisfy the

FIGURE 16.1



quadratic equation $f(r) = 0$ and both of which, of course, are real-valued. Thus the lowest curve illustrates Case 1. Turning to the middle curve, we note that it meets the horizontal axis only once, at r_3. This latter is the only value of r that can satisfy the equation $f(r) = 0$. Therefore, the middle curve illustrates Case 2. Last, we note that the top curve does not meet the horizontal axis at all, and there is thus no real-valued root to the equation $f(r) = 0$. While there exist no real roots in such a case, there are nevertheless two complex numbers that can satisfy the equation, as will be shown in Sec. 16.2.

The Dynamic Stability of Equilibrium

For Cases 1 and 2, the condition for dynamic stability of equilibrium again depends on the algebraic signs of the characteristic roots.

For Case 1, the complementary function (16.7) consists of the two exponential expressions $A_1e^{r_1t}$ and $A_2e^{r_2t}$. The coefficients A_1 and A_2 are arbitrary constants; their values hinge on the initial conditions of the problem. Thus we can be sure of a dynamically stable equilibrium ($y_c \to 0$ as $t \to \infty$), regardless of what the initial conditions happen to be, if and only if the roots r_1 and r_2 are *both* negative. We emphasize the word *both* here, because the condition for dynamic stability does *not* permit even *one* of the roots to be positive or zero. If $r_1 = 2$ and $r_2 = -5$, for instance, it might appear at first glance that the second root, being larger in absolute value, can outweigh the first. In actuality, however, it is the *positive* root that must eventually dominate, because as t increases, e^{2t} will grow increasingly larger, but e^{-5t} will steadily dwindle away.

For Case 2, with repeated roots, the complementary function (16.9) contains not only the familiar e^{rt} expression, but also a multiplicative expression te^{rt}. For the former term to approach zero whatever the initial conditions may be, it is necessary-and-sufficient to have $r < 0$. But would that also ensure the vanishing of te^{rt}? As it turns out, the expression te^{rt} (or, more generally, t^ke^{rt}) possesses the same general type of time path as does e^{rt} ($r \neq 0$). Thus the condition $r < 0$ is indeed necessary-and-sufficient for the entire complementary function to approach zero as $t \to \infty$, yielding a dynamically stable intertemporal equilibrium.

EXERCISE 16.1

1. Find the particular integral of each equation:
   (a) $y''(t) - 2y'(t) + 5y = 2$ (d) $y''(t) + 2y'(t) - y = -4$
   (b) $y''(t) + y'(t) = 7$ (e) $y''(t) = 12$
   (c) $y''(t) + 3y = 9$
2. Find the complementary function of each equation:
   (a) $y''(t) + 3y'(t) - 4y = 12$ (c) $y''(t) - 2y'(t) + y = 3$
   (b) $y''(t) + 6y'(t) + 5y = 10$ (d) $y''(t) + 8y'(t) + 16y = 0$
3. Find the general solution of each differential equation in Prob. 2, and then definitize the solution with the initial conditions $y(0) = 4$ and $y'(0) = 2$.
4. Are the intertemporal equilibriums found in Prob. 3 dynamically stable?
5. Verify that the definite solution in Example 5 indeed (a) satisfies the two initial conditions and (b) has first and second derivatives that conform to the given differential equation.
6. Show that, as $t \to \infty$, the limit of te^{rt} is zero if $r < 0$, but is infinite if $r \geq 0$.

16.2 Complex Numbers and Circular Functions

When the coefficients of a second-order linear differential equation, $y''(t) + a_1 y'(t) + a_2 y = b$, are such that $a_1^2 < 4a_2$, the characteristic-root formula (16.5) would call for taking the square root of a *negative* number. Since the square of any positive or negative real number is invariably positive, whereas the square of zero is zero, only a *nonnegative* real number can ever yield a real-valued square root. Thus, if we confine our attention to the real number system, as we have so far, no characteristic roots are available for this case (Case 3). This fact motivates us to consider numbers outside of the real-number system.

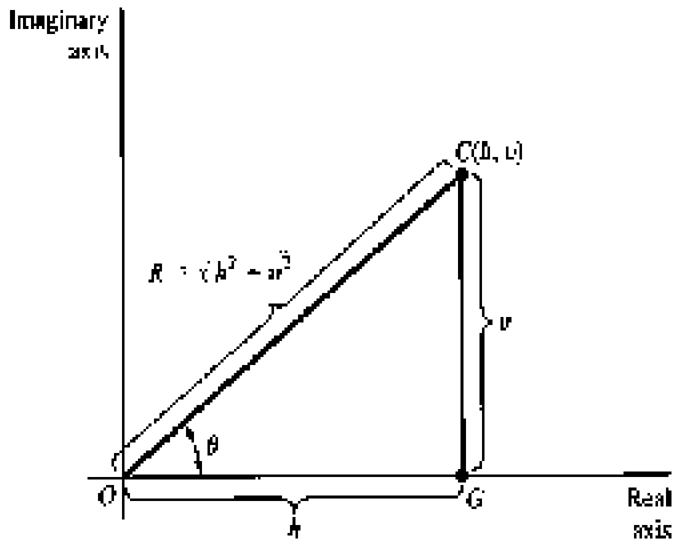
Imaginary and Complex Numbers

Conceptually, it is possible to define a number $i \equiv \sqrt{-1}$, which when squared will equal -1. Because i is the square root of a negative number, it is obviously not real-valued; it is therefore referred to as an *imaginary number*. With it at our disposal, we may write a host of other imaginary numbers, such as $\sqrt{-9} = \sqrt{9}\sqrt{-1} = 3i$ and $\sqrt{-2} = \sqrt{2}i$.

Extending its application a step further, we may construct yet another type of number—one that contains a *real* part as well as an *imaginary* part, such as $(8 + i)$ and $(3 + 5i)$. Known as *complex numbers*, these can be represented generally in the form $(h + vi)$, where h and v are two real numbers.[†] Of course, in case $v = 0$, the complex number will reduce to a real number, whereas if $h = 0$, it will become an imaginary number. Thus the *set of all real numbers* (call it **R**) constitutes a subset of the *set of all complex numbers* (call it **C**). Similarly, the *set of all imaginary numbers* (call it **I**) also constitutes a subset of **C**. That is, $\mathbf{R} \subset \mathbf{C}$, and $\mathbf{I} \subset \mathbf{C}$. Furthermore, since the terms *real* and *imaginary* are mutually exclusive, the sets **R** and **I** must be disjoint; that is $\mathbf{R} \cap \mathbf{I} = \varnothing$.

[†] We employ the symbols h (for horizontal) and v (for vertical) in the general complex-number notation, because we shall presently plot the values of h and v, respectively, on the horizontal and vertical axes of a two-dimensional diagram.

FIGURE 16.2



A complex number $(h + vi)$ can be represented graphically in what is called an *Argand diagram*, as illustrated in Fig. 16.2. By plotting h horizontally on the *real axis* and v vertically on the *imaginary axis*, the number $(h + vi)$ can be specified by the point (h, v), which we have alternatively labeled C. The values of h and v are algebraically signed, of course, so that if $h < 0$, the point C will be to the left of the point of origin; similarly, a negative v will mean a location below the horizontal axis.

Given the values of h and v, we can also calculate the length of the line OC by applying Pythagoras's theorem, which states that the square of the hypotenuse of a right-angled triangle is the sum of the squares of the other two sides. Denoting the length of OC by R (for radius vector), we have

$$R^2 = h^2 + v^2 \qquad \text{and} \qquad R = \sqrt{h^2 + v^2} \tag{16.10}$$

where the square root is always taken to be positive. The value of R is sometimes called the *absolute value*, or *modulus*, of the complex number $(h + vi)$. (Note that changing the signs of h and v will produce no effect on the absolute value of the complex number, R.) Like h and v, then, R is real-valued, but unlike these other values, R is always positive. We shall find the number R to be of great importance in the ensuing discussion.

Complex Roots

Meanwhile, let us return to formula (16.5) and examine the case of complex characteristic roots. When the coefficients of a second-order differential equation are such that $a_1^2 < 4a_2$, the square-root expression in (16.5) can be written as

$$\sqrt{a_1^2 - 4a_2} = \sqrt{4a_2 - a_1^2}\sqrt{-1} = \sqrt{4a_2 - a_1^2}\, i$$

Hence, if we adopt the shorthand

$$h = \frac{-a_1}{2} \qquad \text{and} \qquad v = \frac{\sqrt{4a_2 - a_1^2}}{2}$$

the two roots can be denoted by a pair of *conjugate complex numbers*:

$$r_1, r_2 = h \pm vi$$

These two complex roots are said to be "conjugate" because they always appear together, one being the *sum* of h and vi, and the other being the *difference* between h and vi. Note that they share the same absolute value R.

Example 1 Find the roots of the characteristic equation $r^2 + r + 4 = 0$. Applying the familiar formula, we have

$$r_1, r_2 = \frac{-1 \pm \sqrt{-15}}{2} = \frac{-1 \pm \sqrt{15}\sqrt{-1}}{2} = \frac{-1}{2} \pm \frac{\sqrt{15}}{2}i$$

which constitute a pair of conjugate complex numbers.

As before, we can use (16.6) to check our calculations. If correct, we should have $r_1 + r_2 = -a_1 (= -1)$ and $r_1 r_2 = a_2 (= 4)$. Since we do find

$$r_1 + r_2 = \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2}\right) + \left(\frac{-1}{2} - \frac{\sqrt{15}i}{2}\right) = \frac{-1}{2} + \frac{-1}{2} = -1$$

and

$$r_1 r_2 = \left(\frac{-1}{2} + \frac{\sqrt{15}i}{2}\right)\left(\frac{-1}{2} - \frac{\sqrt{15}i}{2}\right) = \left(\frac{-1}{2}\right)^2 - \left(\frac{\sqrt{15}i}{2}\right)^2 = \frac{1}{4} - \frac{-15}{4} = 4$$

our calculation is indeed validated.

Even in the complex-root case (Case 3), we may express the complementary function of a differential equation according to (16.7); that is,

$$y_c = A_1 e^{(h+vi)t} + A_2 e^{(h-vi)t} = e^{ht}(A_1 e^{vit} + A_2 e^{-vit}) \qquad (16.11)$$

But a new feature has been introduced: the number i now appears in the exponents of the two expressions in parentheses. How do we interpret such imaginary exponential functions?

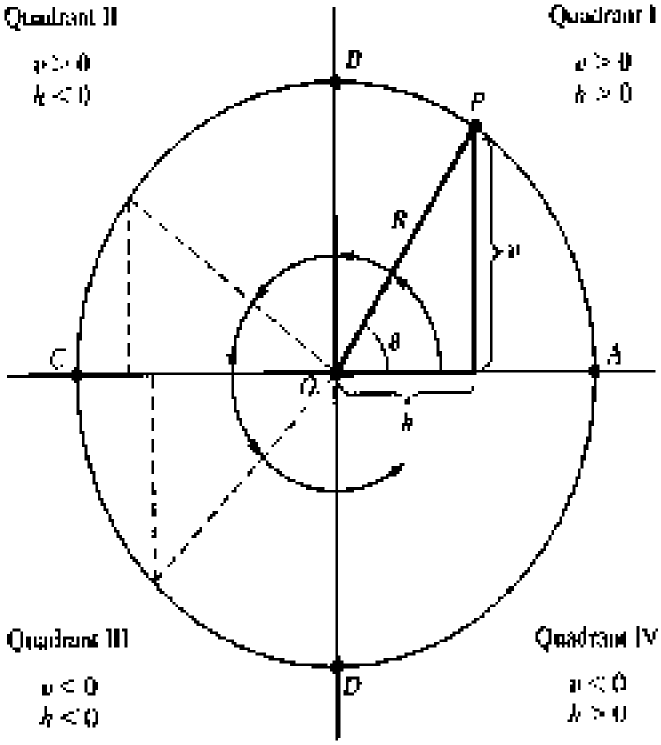
To facilitate their interpretation, it will prove helpful first to transform these expressions into equivalent *circular-function* forms. As we shall presently see, the latter functions characteristically involve periodic fluctuations of a variable. Consequently, the complementary function (16.11), being translatable into circular-function forms, can also be expected to generate a cyclical type of time path.

Circular Functions

Consider a circle with its center at the point of origin and with a radius of length R, as shown in Fig. 16.3. Let the radius, like the hand of a clock, rotate in the counterclockwise direction. Starting from the position OA, it will gradually move into the position OP, followed successively by such positions as OB, OC, and OD; and at the end of a cycle, it will return to OA. Thereafter, the cycle will simply repeat itself.

When in a specific position—say, OP—the clock hand will make a definite angle θ with line OA, and the tip of the hand (P) will determine a vertical distance v and a horizontal distance h. As the angle θ changes during the process of rotation, v and h will vary, although

FIGURE 16.3



R will not. Thus the ratios v/R and h/R must change with θ; that is, these two ratios are both functions of the angle θ. Specifically, v/R and h/R are called, respectively, the *sine* (function) of θ and the *cosine* (function) of θ:

$$\sin\theta \equiv \frac{v}{R} \tag{16.12}$$

$$\cos\theta \equiv \frac{h}{R} \tag{16.13}$$

In view of their connection with a circle, these functions are referred to as *circular functions*. Since they are also associated with a triangle, however, they are alternatively called *trigonometric functions*. Another (and fancier) name for them is *sinusoidal functions*. The sine and cosine functions are not the only circular functions; another frequently encountered one is the *tangent* function, defined as

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{v}{h} \qquad (h \neq 0)$$

Our major concern here, however, will be with the sine and cosine functions.

The independent variable in a circular function is the angle θ, so the mapping involved here is from an *angle* to a *ratio of two distances*. Usually, angles are measured in *degrees* (for example, 30, 45, and 90°); in analytical work, however, it is more convenient to measure angles in *radians* instead. The advantage of the radian measure stems from the fact that, when θ is so measured, the derivatives of circular functions will come out in neater expressions—much as the base e gives us neater derivatives for exponential and logarithmic functions. But just how much is a radian? To explain this, let us return to Fig. 16.3, where we have drawn the point P so that the length of the *arc AP* is exactly equal to the radius R. A *radian* (abbreviated as *rad*) can then be defined as the size of the angle θ

(in Fig. 16.3) formed by such an R-length arc. Since the circumference of the circle has a total length of $2\pi R$ (where $\pi = 3.14159\ldots$), a complete circle must involve an angle of 2π rad altogether. In terms of degrees, however, a complete circle makes an angle of 360°; thus, by equating 360° to 2π rad, we can arrive at the following conversion table:

Degrees	360	270	180	90	45	0
Radians	2π	$\frac{3\pi}{2}$	π	$\frac{\pi}{2}$	$\frac{\pi}{4}$	0

Properties of the Sine and Cosine Functions

Given the length of R, the value of $\sin\theta$ hinges upon the way the value of v changes in response to changes in the angle θ. In the starting position OA, we have $v = 0$. As the clock hand moves counterclockwise, v starts to assume an increasing positive value, culminating in the maximum value of $v = R$ when the hand coincides with OB, that is, when $\theta = \pi/2$ rad ($= 90°$). Further movement will gradually shorten v, until its value becomes zero when the hand is in the position OC, i.e., when $\theta = \pi$ rad ($= 180°$). As the hand enters the third quadrant, v begins to assume negative values; in the position OD, we have $v = -R$. In the fourth quadrant, v is still negative, but it will increase from the value of $-R$ toward the value of $v = 0$, which is attained when the hand returns to OA—that is, when $\theta = 2\pi$ rad ($= 360°$). The cycle then repeats itself.

When these illustrative values of v are substituted into (16.12), we can obtain the results shown in the "$\sin\theta$" row of Table 16.1. For a more complete description of the sine function, however, see the graph in Fig. 16.4a, where the values of $\sin\theta$ are plotted against those of θ (expressed in radians).

The value of $\cos\theta$, in contrast, depends instead upon the way that h changes in response to changes in θ. In the starting position OA, we have $h = R$. Then h gradually shrinks, till $h = 0$ when $\theta = \pi/2$ (position OB). In the second quadrant, h turns negative, and when $\theta = \pi$ (position OC), $h = -R$. The value of h gradually increases from $-R$ to zero in the third quadrant, and when $\theta = 3\pi/2$ (position OD), we find that $h = 0$. In the fourth quadrant, h turns positive again, and when the hand returns to position OA ($\theta = 2\pi$), we again have $h = R$. The cycle then repeats itself.

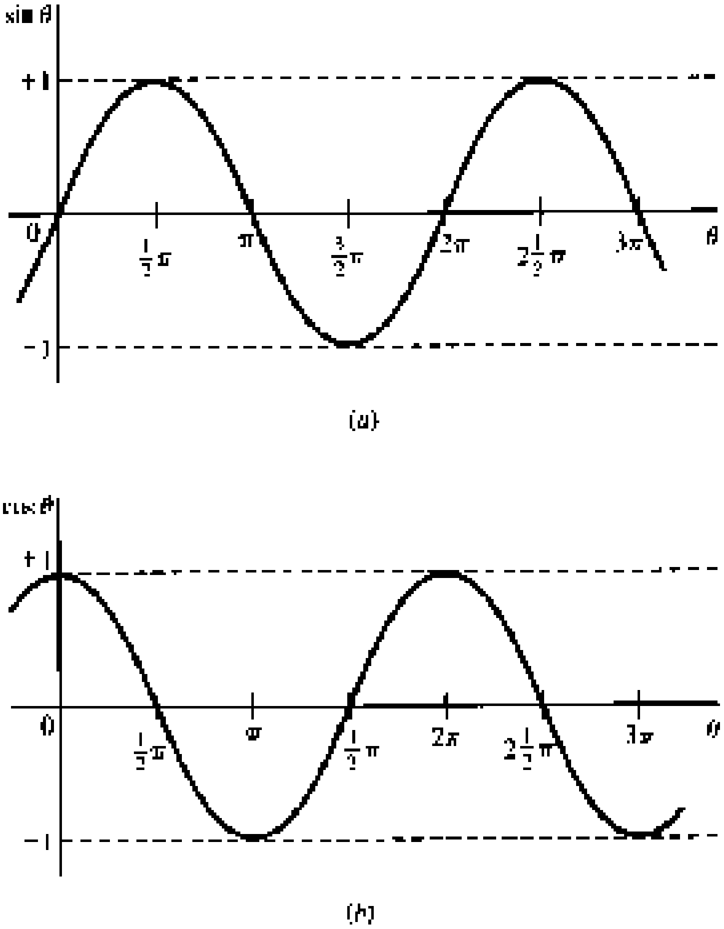
The substitution of these illustrative values of h into (16.13) yields the results in the bottom row of Table 16.1, but Fig. 16.4b gives a more complete depiction of the cosine function.

The $\sin\theta$ and $\cos\theta$ functions share the same domain, namely, the set of all real numbers (radian measures of θ). In this connection, it may be pointed out that a *negative* angle simply refers to the reverse rotation of the clock hand; for instance, a clockwise movement

TABLE 16.1

θ	0	$\frac{1}{2}\pi$	π	$\frac{3}{2}\pi$	2π
$\sin\theta$	0	1	0	-1	0
$\cos\theta$	1	0	-1	0	1

FIGURE 16.4



from OA to OD in Fig. 16.3 generates an angle of $-\pi/2$ rad ($= -90°$). There is also a common range for the two functions, namely, the closed interval $[-1, 1]$. For this reason, the graphs of $\sin\theta$ and $\cos\theta$ are, in Fig. 16.4, confined to a definite horizontal band.

A major distinguishing property of the sine and cosine functions is that both are *periodic;* their values will repeat themselves for every 2π rad (a complete circle) the angle θ travels through. Each function is therefore said to have a *period* of 2π. In view of this periodicity feature, the following equations hold (for any integer n):

$$\sin(\theta + 2n\pi) = \sin\theta \qquad \cos(\theta + 2n\pi) = \cos\theta$$

That is, adding (or subtracting) any integer multiple of 2π to any angle θ will affect neither the value of $\sin\theta$ nor that of $\cos\theta$.

The graphs of the sine and cosine functions indicate a constant range of fluctuation in each period, namely, ± 1. This is sometimes alternatively described by saying that the *amplitude* of fluctuation is 1. By virtue of the identical period and the identical amplitude, we see that the $\cos\theta$ curve, if shifted rightward by $\pi/2$, will be exactly coincident with the $\sin\theta$ curve. These two curves are therefore said to differ only in *phase,* i.e., to differ only in the location of the peak in each period. Symbolically, this fact may be stated by the equation

$$\cos\theta = \sin\left(\theta + \frac{\pi}{2}\right)$$

The sine and cosine functions obey certain identities. Among these, the more frequently used are

$$\begin{aligned}\sin(-\theta) &\equiv -\sin\theta \\ \cos(-\theta) &\equiv \cos\theta\end{aligned} \tag{16.14}$$

$$\sin^2\theta + \cos^2\theta \equiv 1 \qquad [\text{where } \sin^2\theta \equiv (\sin\theta)^2, \text{ etc.}] \tag{16.15}$$

$$\begin{aligned}\sin(\theta_1 \pm \theta_2) &\equiv \sin\theta_1\cos\theta_2 \pm \cos\theta_1\sin\theta_2 \\ \cos(\theta_1 \pm \theta_2) &\equiv \cos\theta_1\cos\theta_2 \mp \sin\theta_1\sin\theta_2\end{aligned} \tag{16.16}$$

The pair of identities (16.14) serves to underscore the fact that the cosine function is symmetrical with respect to the vertical axis (that is, θ and $-\theta$ always yield the same cosine value), while the sine function is not. Shown in (16.15) is the fact that, for any magnitude of θ, the sum of the squares of its sine and cosine is always unity. And the set of identities in (16.16) gives the sine and cosine of the sum and difference of two angles θ_1 and θ_2.

Finally, a word about derivatives. Being continuous and smooth, both $\sin\theta$ and $\cos\theta$ are differentiable. The derivatives, $d(\sin\theta)/d\theta$ and $d(\cos\theta)/d\theta$, are obtainable by taking the limits, respectively, of the difference quotients $\Delta(\sin\theta)/\Delta\theta$ and $\Delta(\cos\theta)/\Delta\theta$ as $\Delta\theta \to 0$. The results, stated here without proof, are

$$\frac{d}{d\theta}\sin\theta = \cos\theta \tag{16.17}$$

$$\frac{d}{d\theta}\cos\theta = -\sin\theta \tag{16.18}$$

It should be emphasized, however, that these derivative formulas are valid only when θ is measured in radians; if measured in degrees, for instance, (16.17) will become $d(\sin\theta)/d\theta = (\pi/180)\cos\theta$ instead. It is for the sake of getting rid of the factor $(\pi/180)$ that radian measures are preferred to degree measures in analytical work.

Example 2 Find the slope of the $\sin\theta$ curve at $\theta = \pi/2$. The slope of the sine curve is given by its derivative ($= \cos\theta$). Thus, at $\theta = \pi/2$, the slope should be $\cos(\pi/2) = 0$. You may refer to Fig. 16.4 for verification of this result.

Example 3 Find the second derivative of $\sin\theta$. From (16.17), we know that the first derivative of $\sin\theta$ is $\cos\theta$, therefore the desired second derivative is

$$\frac{d^2}{d\theta^2}\sin\theta = \frac{d}{d\theta}\cos\theta = -\sin\theta$$

Euler Relations

In Sec. 9.5, it was shown that any function which has finite, continuous derivatives up to the desired order can be expanded into a polynomial function. Moreover, if the remainder term R_n in the resulting Taylor series (expansion at any point x_0) or Maclaurin series (expansion at $x_0 = 0$) happens to approach zero as the number of terms n becomes infinite, the polynomial may be written as an infinite series. We shall now expand the sine and cosine functions and then attempt to show how the imaginary exponential expressions encountered in (16.11) can be transformed into circular functions having equivalent expansions.

For the sine function, write $\phi(\theta) = \sin\theta$; it then follows that $\phi(0) = \sin 0 = 0$. By successive derivation, we can get

$$\left.\begin{aligned} \phi'(\theta) &= \cos\theta \\ \phi''(\theta) &= -\sin\theta \\ \phi'''(\theta) &= -\cos\theta \\ \phi^{(4)}(\theta) &= \sin\theta \\ \phi^{(5)}(\theta) &= \cos\theta \\ \vdots \quad & \quad \vdots \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \phi'(0) &= \cos 0 = 1 \\ \phi''(0) &= -\sin 0 = 0 \\ \phi'''(0) &= -\cos 0 = -1 \\ \phi^{(4)}(0) &= \sin 0 = 0 \\ \phi^{(5)}(0) &= \cos 0 = 1 \\ \vdots \quad & \quad \vdots \end{aligned}\right.$$

When substituted into (9.14), where θ now replaces x, these will give us the following Maclaurin series with remainder:

$$\sin\theta = 0 + \theta + 0 - \frac{\theta^3}{3!} + 0 + \frac{\theta^5}{5!} + \cdots + \frac{\phi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

Now, the expression $\phi^{(n+1)}(p)$ in the last (remainder) term, which represents the $(n+1)$st derivative evaluated at $\theta = p$, can only take the form of $\pm\cos p$ or $\pm\sin p$ and, as such, can only take a value in the interval $[-1, 1]$, regardless of how large n is. On the other hand, $(n+1)!$ will grow rapidly as $n \to \infty$—in fact, much more rapidly than θ^{n+1} as n increases. Hence, the remainder term will approach zero as $n \to \infty$, and we can therefore express the Maclaurin series as an infinite series:

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \qquad \textbf{(16.19)}$$

Similarly, if we write $\psi(\theta) = \cos\theta$, then $\psi(0) = \cos 0 = 1$, and the successive derivatives will be

$$\left.\begin{aligned} \psi'(\theta) &= -\sin\theta \\ \psi''(\theta) &= -\cos\theta \\ \psi'''(\theta) &= \sin\theta \\ \psi^{(4)}(\theta) &= \cos\theta \\ \psi^{(5)}(\theta) &= -\sin\theta \\ \vdots \quad & \quad \vdots \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \psi'(0) &= -\sin 0 = 0 \\ \psi''(0) &= -\cos 0 = -1 \\ \psi'''(0) &= \sin 0 = 0 \\ \psi^{(4)}(0) &= \cos 0 = 1 \\ \psi^{(5)}(0) &= -\sin 0 = 0 \\ \vdots \quad & \quad \vdots \end{aligned}\right.$$

On the basis of these derivatives, we can expand $\cos\theta$ as follows:

$$\cos\theta = 1 + 0 - \frac{\theta^2}{2!} + 0 + \frac{\theta^4}{4!} + \cdots + \frac{\psi^{(n+1)}(p)}{(n+1)!}\theta^{n+1}$$

Since the remainder term will again tend toward zero as $n \to \infty$, the cosine function is also expressible as an infinite series, as follows:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \qquad \textbf{(16.20)}$$

You must have noticed that, with (16.19) and (16.20) at hand, we are now capable of constructing a table of sine and cosine values for all possible values of θ (in radians). However, our immediate interest lies in finding the relationship between imaginary exponential expressions and circular functions. To this end, let us now expand the two exponential expressions $e^{i\theta}$ and $e^{-i\theta}$. The reader will recognize that these are but special cases of the expression e^x, which has previously been shown, in (10.6), to have the expansion

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots$$

Letting $x = i\theta$, therefore, we can immediately obtain

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \cdots \\ &= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned}$$

Similarly, by setting $x = -i\theta$, the following result will emerge:

$$\begin{aligned} e^{-i\theta} &= 1 - i\theta + \frac{(-i\theta)^2}{2!} + \frac{(-i\theta)^3}{3!} + \frac{(-i\theta)^4}{4!} + \frac{(-i\theta)^5}{5!} + \cdots \\ &= 1 - i\theta - \frac{\theta^2}{2!} + \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} - \frac{i\theta^5}{5!} - \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) - i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned}$$

By substituting (16.19) and (16.20) into these two results, the following pair of identities—known as the *Euler relations*—can readily be established:

$$e^{i\theta} \equiv \cos\theta + i\sin\theta \qquad (16.21)$$

$$e^{-i\theta} \equiv \cos\theta - i\sin\theta \qquad (16.21')$$

These will enable us to translate any imaginary exponential function into an equivalent linear combination of sine and cosine functions, and vice versa.

Example 4 Find the value of $e^{i\pi}$. First let us convert this expression into a trigonometric expression. By setting $\theta = \pi$ in (16.21), it is found that $e^{i\pi} = \cos\pi + i\sin\pi$. Since $\cos\pi = -1$ and $\sin\pi = 0$, it follows that $e^{i\pi} = -1$.

Example 5 Show that $e^{-i\pi/2} = -i$. Setting $\theta = \pi/2$ in (16.21'), we have

$$e^{-i\pi/2} = \cos\frac{\pi}{2} - i\sin\frac{\pi}{2} = 0 - i(1) = -i$$

Alternative Representations of Complex Numbers

So far, we have represented a pair of conjugate complex numbers in the general form $(h \pm vi)$. Since h and v refer to the abscissa and ordinate in the Cartesian coordinate system of an Argand diagram, the expression $(h \pm vi)$ represents the *Cartesian form* of a pair of conjugate complex numbers. As a by-product of the discussion of circular functions and Euler relations, we can now express $(h \pm vi)$ in two other ways.

Referring to Fig. 16.2, we see that as soon as h and v are specified, the angle θ and the value of R also become determinate. Since a given θ and a given R can together identify a *unique point in the Argand diagram*, we may employ θ and R to specify the particular pair of complex numbers. By rewriting the definitions of the sine and cosine functions in (16.12) and (16.13) as

$$v = R\sin\theta \quad \text{and} \quad h = R\cos\theta \tag{16.22}$$

the conjugate complex numbers $(h \pm vi)$ can be transformed as follows:

$$h \pm vi = R\cos\theta \pm Ri\sin\theta = R(\cos\theta \pm i\sin\theta)$$

In so doing, we have in effect switched from the Cartesian coordinates of the complex numbers (h and v) to what are called their *polar coordinates* (R and θ). The right-hand expression in the preceding equation, accordingly, exemplifies the *polar form* of a pair of conjugate complex numbers.

Furthermore, in view of the Euler relations, the polar form may also be rewritten into the *exponential form* as follows: $R(\cos\theta \pm i\sin\theta) = Re^{\pm i\theta}$. Hence, we have a total of three alternative representations of the conjugate complex numbers:

$$h \pm vi = R(\cos\theta \pm i\sin\theta) = Re^{\pm i\theta} \tag{16.23}$$

If we are given the values of R and θ, the transformation to h and v is straightforward: we use the two equations in (16.22). What about the reverse transformation? With given values of h and v, no difficulty arises in finding the corresponding value of R, which is equal to $\sqrt{h^2 + v^2}$. But a slight ambiguity arises in regard to θ: the desired value of θ (in radians) is that which satisfies the two conditions $\cos\theta = h/R$ and $\sin\theta = v/R$; but for given values of h and v, θ is not unique! (Why?) Fortunately, the problem is not serious, for by confining our attention to the interval $[0, 2\pi)$ in the domain, the indeterminancy is quickly resolved.

Example 6

Find the Cartesian form of the complex number $5e^{3i\pi/2}$. Here we have $R = 5$ and $\theta = 3\pi/2$; hence, by (16.22) and Table 16.1,

$$h = 5\cos\frac{3\pi}{2} = 0 \qquad \text{and} \qquad v = 5\sin\frac{3\pi}{2} = -5$$

The Cartesian form is thus simply $h - vi = -5i$.

Example 7

Find the polar and exponential forms of $(1 + \sqrt{3}i)$. In this case, we have $h = 1$ and $v = \sqrt{3}$; thus $R = \sqrt{1+3} = 2$. Table 16.1 is of no use in locating the value of θ this time, but Table 16.2, which lists some additional selected values of $\sin\theta$ and $\cos\theta$, will help. Specifically,

TABLE 16.2

θ	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{3\pi}{4}$
$\sin\theta$	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$
$\cos\theta$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}\left(=\frac{\sqrt{2}}{2}\right)$	$\frac{1}{2}$	$\frac{-1}{\sqrt{2}}\left(=\frac{-\sqrt{2}}{2}\right)$

we are seeking the value of θ such that $\cos\theta = h/R = 1/2$ and $\sin\theta = v/R = \sqrt{3}/2$. The value $\theta = \pi/3$ meets the requirements. Thus, according to (16.23), the desired transformation is

$$1 + \sqrt{3}i = 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right) = 2e^{i\pi/3}$$

Before leaving this topic, let us note an important extension of the result in (16.23). Supposing that we have the nth power of a complex number—say, $(h + vi)^n$—how do we write its polar and exponential forms? The exponential form is the easier to derive. Since $h + vi = Re^{i\theta}$, it follows that

$$(h + vi)^n = (Re^{i\theta})^n = R^n e^{in\theta}$$

Similarly, we can write

$$(h - vi)^n = (Re^{-i\theta})^n = R^n e^{-in\theta}$$

Note that the power n has brought about two changes: (1) R now becomes R^n, and (2) θ now becomes $n\theta$. When these two changes are inserted into the polar form in (16.23), we find that

$$(h \pm vi)^n = R^n(\cos n\theta \pm i\sin n\theta) \qquad (16.23')$$

That is,

$$[R(\cos\theta \pm i\sin\theta)]^n = R^n(\cos n\theta \pm i\sin n\theta)$$

Known as *De Moivre's theorem*, this result indicates that, to raise a complex number to the nth power, one must simply modify its polar coordinates by raising R to the nth power and multiplying θ by n.

EXERCISE 16.2

1. Find the roots of the following quadratic equations:
   (a) $r^2 - 3r + 9 = 0$ (c) $2x^2 + x + 8 = 0$
   (b) $r^2 + 2r + 17 = 0$ (d) $2x^2 - x + 1 = 0$
2. (a) How many degrees are there in a radian?
   (b) How many radians are there in a degree?
3. With reference to Fig. 16.3, and by using Pythagoras's theorem, prove that
   (a) $\sin^2\theta + \cos^2\theta \equiv 1$ (b) $\sin\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$
4. By means of the identities (16.14), (16.15), and (16.16), show that:
   (a) $\sin 2\theta \equiv 2\sin\theta\cos\theta$
   (b) $\cos 2\theta \equiv 1 - 2\sin^2\theta$
   (c) $\sin(\theta_1 + \theta_2) + \sin(\theta_1 - \theta_2) \equiv 2\sin\theta_1\cos\theta_2$
   (d) $1 + \tan^2\theta = \frac{1}{\cos^2\theta}$
   (e) $\sin\left(\frac{\pi}{2} - \theta\right) \equiv \cos\theta$ (f) $\cos\left(\frac{\pi}{2} - \theta\right) \equiv \sin\theta$
5. By applying the chain rule:
   (a) Write out the derivative formulas for $\frac{d}{d\theta}\sin f(\theta)$ and $\frac{d}{d\theta}\cos f(\theta)$, where $f(\theta)$ is a function of θ.
   (b) Find the derivatives of $\cos\theta^3$, $\sin(\theta^2 + 3\theta)$, $\cos e^\theta$, and $\sin(1/\theta)$.

6. From the Euler relations, deduce that:

(a) $e^{-i\pi} = -1$ (c) $e^{i\pi/4} = \frac{\sqrt{2}}{2}(1+i)$

(b) $e^{i\pi/3} = \frac{1}{2}(1+\sqrt{3}i)$ (d) $e^{-3i\pi/4} = -\frac{\sqrt{2}}{2}(1+i)$

7. Find the Cartesian form of each complex number:

(a) $2\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)$ (b) $4e^{i\pi/3}$ (c) $\sqrt{2}e^{-i\pi/4}$

8. Find the polar and exponential forms of the following complex numbers:

(a) $\frac{3}{2} + \frac{3\sqrt{3}}{2}i$ (b) $4(\sqrt{3}+i)$

16.3 Analysis of the Complex-Root Case

With the concepts of complex numbers and circular functions at our disposal, we are now prepared to approach the complex-root case (Case 3), referred to in Sec. 16.1. You will recall that the classification of the three cases, according to the nature of the characteristic roots, is concerned only with the complementary function of a differential equation. Thus, we can continue to focus our attention on the reduced equation

$$y''(t) + a_1 y'(t) + a_2 y = 0 \qquad \text{[reproduced from (16.4)]}$$

The Complementary Function

When the values of the coefficients a_1 and a_2 are such that $a_1^2 < 4a_2$, the characteristic roots will be the pair of conjugate complex numbers

$$r_1, r_2 = h \pm vi$$

where

$$h = -\frac{1}{2}a_1 \qquad \text{and} \qquad v = \frac{1}{2}\sqrt{4a_2 - a_1^2}$$

The complementary function, as was already previewed, will thus be in the form

$$y_c = e^{ht}(A_1 e^{vit} + A_2 e^{-vit}) \qquad \text{[reproduced from (16.11)]}$$

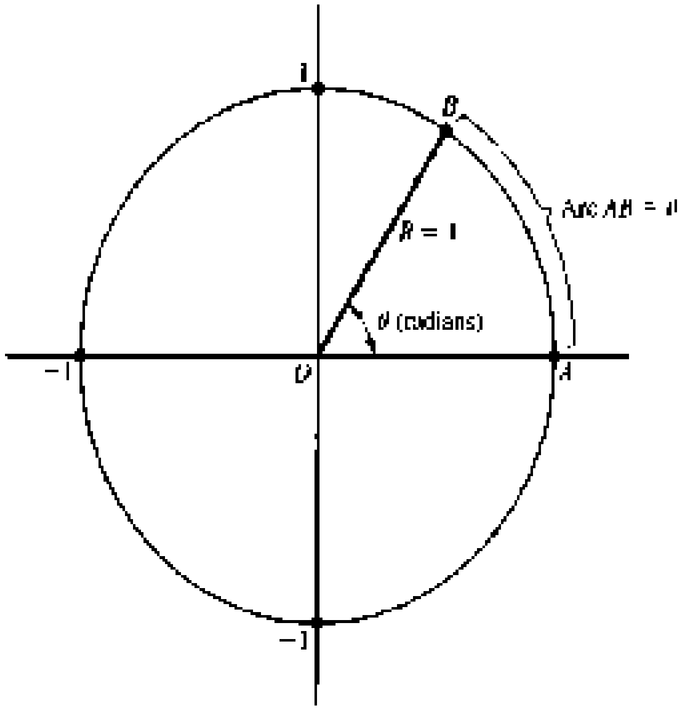
Let us first transform the imaginary exponential expressions in the parentheses into equivalent trigonometric expressions, so that we may interpret the complementary function as a circular function. This may be accomplished by using the Euler relations. Letting $\theta = vt$ in (16.21) and (16.21'), we find that

$$e^{vit} = \cos vt + i\sin vt \quad \text{and} \quad e^{-vit} = \cos vt - i\sin vt$$

From these, it follows that the complementary function in (16.11) can be rewritten as

$$\begin{aligned} y_c &= e^{ht}[A_1(\cos vt + i\sin vt) + A_2(\cos vt - i\sin vt)] \\ &= e^{ht}[(A_1 + A_2)\cos vt + (A_1 - A_2)i\sin vt] \end{aligned} \tag{16.24}$$

FIGURE 16.5



Furthermore, if we employ the shorthand symbols

$$A_5 \equiv A_1 + A_2 \quad \text{and} \quad A_6 \equiv (A_1 - A_2)i$$

it is possible to simplify (16.24) into[1]

$$y_c = e^{ht}(A_5 \cos vt + A_6 \sin vt) \tag{16.24'}$$

where the new arbitrary constants A_5 and A_6 are later to be definitized.

If you are meticulous, you may feel somewhat uneasy about the substitution of θ by vt in the foregoing procedure. The variable θ measures an angle, but vt is a magnitude in units of t (in our context, time). Therefore, how can we make the substitution $\theta = vt$? The answer to this question can best be explained with reference to the *unit circle* (a circle with radius $R = 1$) in Fig. 16.5. True, we have been using θ to designate an angle; but since the angle is measured in radian units, the value of θ is always the ratio of the length of arc AB to the radius R. When $R = 1$, we have specifically

$$\theta \equiv \frac{\text{arc } AB}{R} \equiv \frac{\text{arc } AB}{1} \equiv \text{arc } AB$$

In other words, θ is not only the radian measure of the angle, but also the length of the arc AB, which is a number rather than an angle. If the passing of time is charted on the circumference of the unit circle (counterclockwise), rather than on a straight line as we do in plotting a time series, it really makes no difference whatsoever whether we consider the

[1] The fact that in defining A_6, we include in it the imaginary number i is by no means an attempt to "sweep the dirt under the rug." Because A_6 is an arbitrary constant, it can take an imaginary as well as a real value. Nor is it true that, as defined, A_6 will necessarily turn out to be imaginary. Actually, if A_1 and A_2 are a pair of conjugate complex numbers, say, $m \pm ni$, then A_5 and A_6 will both be real: $A_5 = A_1 + A_2 = (m + ni) + (m - ni) = 2m$, and $A_6 = (A_1 - A_2)i = [(m + ni) - (m - ni)]i = (2ni)i = -2n$.

lapse of time as an increase in the radian measure of the angle θ or as a lengthening of the arc AB. Even if $R \neq 1$, moreover, the same line of reasoning can apply, except that in that case θ will be equal to (arc AB)$/R$ instead; i.e., the angle θ and the arc AB will bear a fixed proportion to each other, instead of being equal. Thus, the substitution $\theta = vt$ is indeed legitimate.

An Example of Solution

Let us find the solution of the differential equation

$$y''(t) + 2y'(t) + 17y = 34$$

with the initial conditions $y(0) = 3$ and $y'(0) = 11$.

Since $a_1 = 2$, $a_2 = 17$, and $b = 34$, we can immediately find the particular integral to be

$$y_p = \frac{b}{a_2} = \frac{34}{17} = 2 \qquad \text{[by (16.3)]}$$

Moreover, since $a_1^2 = 4 < 4a_2 = 68$, the characteristic roots will be the pair of conjugate complex numbers $(h \pm vi)$, where

$$h = -\frac{1}{2}a_1 = -1 \qquad \text{and} \qquad v = \frac{1}{2}\sqrt{4a_2 - a_1^2} = \frac{1}{2}\sqrt{64} = 4$$

Hence, by (16.24′), the complementary function is

$$y_c = e^{-t}(A_5 \cos 4t + A_6 \sin 4t)$$

Combining y_c and y_p, the general solution can be expressed as

$$y(t) = e^{-t}(A_5 \cos 4t + A_6 \sin 4t) + 2$$

To definitize the constants A_5 and A_6, we utilize the two initial conditions. First, by setting $t = 0$ in the general solution, we find that

$$\begin{aligned} y(0) &= e^0(A_5 \cos 0 + A_6 \sin 0) + 2 \\ &= (A_5 + 0) + 2 = A_5 + 2 \qquad [\cos 0 = 1;\ \sin 0 = 0] \end{aligned}$$

By the initial condition $y(0) = 3$, we can thus specify $A_5 = 1$. Next, let us differentiate the general solution with respect to t—using the product rule and the derivative formulas (16.17) and (16.18) while bearing in mind the chain rule [Exercise 16.2-5]—to find $y'(t)$ and then $y'(0)$:

$$y'(t) = -e^{-t}(A_5 \cos 4t + A_6 \sin 4t) + e^{-t}[A_5(-4 \sin 4t) + 4A_6 \cos 4t]$$

so that

$$\begin{aligned} y'(0) &= -(A_5 \cos 0 + A_6 \sin 0) + (-4A_5 \sin 0 + 4A_6 \cos 0) \\ &= -(A_5 + 0) + (0 + 4A_6) = 4A_6 - A_5 \end{aligned}$$

By the second initial condition $y'(0) = 11$, and in view that $A_5 = 1$, it then becomes clear that $A_6 = 3$.[1] The definite solution is, therefore,

$$y(t) = e^{-t}(\cos 4t + 3 \sin 4t) + 2 \qquad \textbf{(16.25)}$$

[1] Note that, here, A_6 indeed turns out to be a real number, even though we have included the imaginary number i in its definition.

As before, the y_p component ($= 2$) can be interpreted as the intertemporal equilibrium level of y, whereas the y_c component represents the deviation from equilibrium. Because of the presence of circular functions in y_c, the time path (16.25) may be expected to exhibit a fluctuating pattern. But what specific pattern will it involve?

The Time Path

We are familiar with the paths of a simple sine or cosine function, as shown in Fig. 16.4. Now we must study the paths of certain variants and combinations of sine and cosine functions so that we can interpret, in general, the complementary function (16.24′)

$$y_c = e^{ht}(A_5 \cos vt + A_6 \sin vt)$$

and, in particular, the y_c component of (16.25).

Let us first examine the term $(A_5 \cos vt)$. By itself, the expression $(\cos vt)$ is a circular function of (vt), with period 2π ($= 6.2832$) and amplitude 1. The period of 2π means that the graph will repeat its configuration every time that (vt) increases by 2π. When t alone is taken as the independent variable, however, repetition will occur every time t increases by $2\pi/v$, so that with reference to t—as is appropriate in dynamic economic analysis—we shall consider the period of $(\cos vt)$ to be $2\pi/v$. (The amplitude, however, remains at 1.) Now, when a multiplicative constant A_5 is attached to $(\cos vt)$, it causes the range of fluctuation to change from ± 1 to $\pm A_5$. Thus the amplitude now becomes A_5, though the period is unaffected by this constant. In short, $(A_5 \cos vt)$ is a cosine function of t, with period $2\pi/v$ and amplitude A_5. By the same token, $(A_6 \sin vt)$ is a sine function of t, with period $2\pi/v$ and amplitude A_6.

There being a common period, the sum $(A_5 \cos vt + A_6 \sin vt)$ will also display a repeating cycle every time t increases by $2\pi/v$. To show this more rigorously, let us note that for given values of A_5 and A_6 we can always find two constants A and ε, such that

$$A_5 = A \cos \varepsilon \qquad \text{and} \qquad A_6 = -A \sin \varepsilon$$

Thus we may express the said sum as

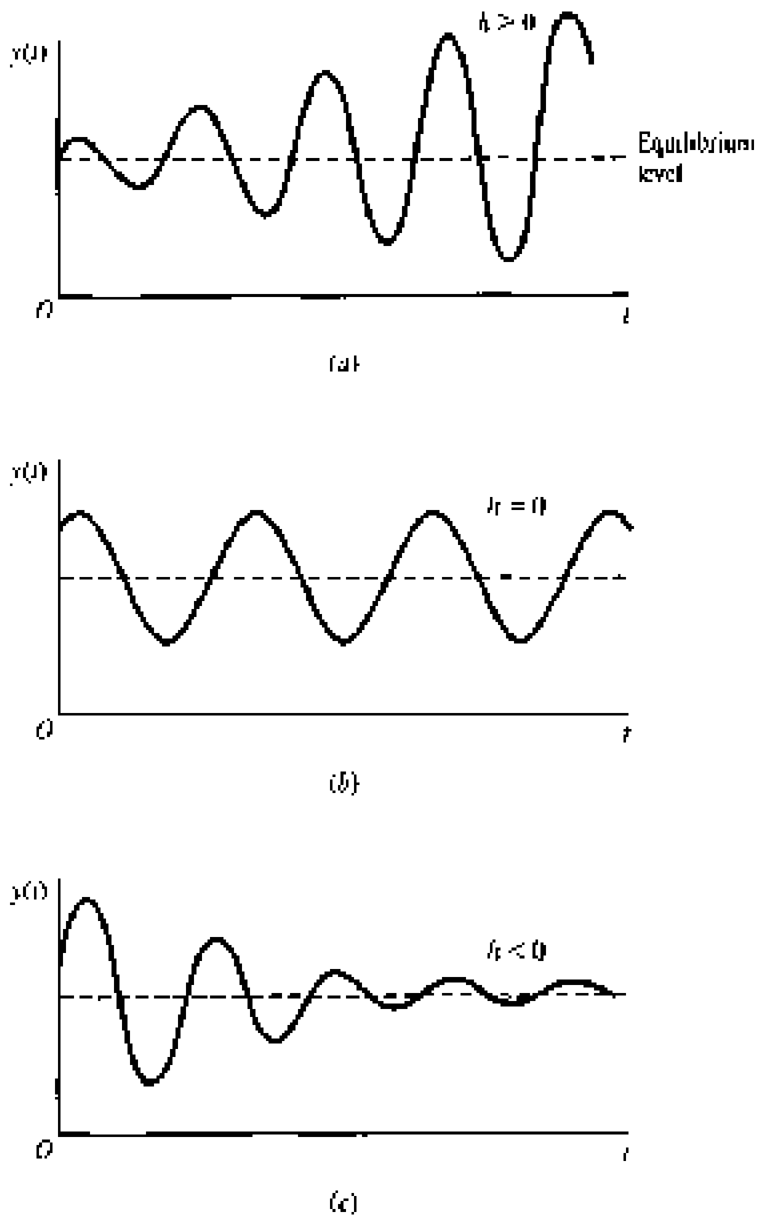
$$\begin{aligned} A_5 \cos vt + A_6 \sin vt &= A \cos \varepsilon \cos vt - A \sin \varepsilon \sin vt \\ &= A(\cos vt \cos \varepsilon - \sin vt \sin \varepsilon) \\ &= A \cos(vt + \varepsilon) \qquad \text{[by (16.16)]} \end{aligned}$$

This is a modified cosine function of t, with amplitude A and period $2\pi/v$, because every time that t increases by $2\pi/v$, $(vt + \varepsilon)$ will increase by 2π, which will complete a cycle on the cosine curve.

Had y_c consisted only of the expression $(A_5 \cos vt + A_6 \sin vt)$, the implication would have been that the time path of y would be a never-ending, constant-amplitude fluctuation around the equilibrium value of y, as represented by y_p. But there is, in fact, also the multiplicative term e^{ht} to consider. This latter term is of major importance, for, as we shall see, it holds the key to the question of whether the time path will converge.

If $h > 0$, the value of e^{ht} will increase continually as t increases. This will produce a magnifying effect on the amplitude of $(A_5 \cos vt + A_6 \sin vt)$ and cause ever-greater deviations from the equilibrium in each successive cycle. As illustrated in Fig. 16.6*a*, the time path will in this case be characterized by *explosive fluctuation*. If $h = 0$, on the other hand,

FIGURE 16.6



then $e^{ht} = 1$, and the complementary function will simply be $(A_5 \cos vt + A_6 \sin vt)$, which has been shown to have a constant amplitude. In this second case, each cycle will display a uniform pattern of deviation from the equilibrium as illustrated by the time path in Fig. 16.6*b*. This is a time path with *uniform fluctuation*. Last, if $h < 0$, the term e^{ht} will continually decrease as t increases, and each successive cycle will have a smaller amplitude than the preceding one, much as the way a ripple dies down. This case is illustrated in Fig. 16.6*c*, where the time path is characterized by *damped fluctuation*. The solution in (16.25), with $h = -1$, exemplifies this last case. It should be clear that only the case of damped fluctuation can produce a *convergent* time path; in the other two cases, the time path is *nonconvergent* or *divergent*.[1]

In all three diagrams of Fig. 16.6, the intertemporal equilibrium is assumed to be stationary. If it is a moving one, the three types of time path depicted will still fluctuate around it, but since a moving equilibrium generally plots as a curve rather than a horizontal straight

[1] We shall use the two words *nonconvergent* and *divergent* interchangeably, although the latter is more strictly applicable to the explosive than to the uniform variety of nonconvergence.

line, the fluctuation will take on the nature of, say, a series of business cycles around a secular trend.

The Dynamic Stability of Equilibrium

The concept of convergence of the time path of a variable is inextricably tied to the concept of dynamic stability of the intertemporal equilibrium of that variable. Specifically, the equilibrium is dynamically stable if, and only if, the time path is convergent. The condition for convergence of the $y(t)$ path, namely, $h < 0$ (Fig. 16.6c), is therefore also the condition for dynamic stability of the intertemporal equilibrium of y.

You will recall that, for Cases 1 and 2 where the characteristic roots are real, the condition for dynamic stability of equilibrium is that every characteristic root be negative. In the present case (Case 3), with complex roots, the condition seems to be more specialized; it stipulates only that the real part (h) of the complex roots ($h \pm vi$) be negative. However, it is possible to unify all three cases and consolidate the seemingly different conditions into a single, generally applicable one. Just interpret any real root r as a complex root whose imaginary part is zero ($v = 0$). Then the condition "the *real* part of every characteristic root be negative" clearly becomes applicable to all three cases and emerges as the only condition we need.

EXERCISE 16.3

Find the y_p and the y_c, the general solution, and the definite solution of each of the following:

1. $y''(t) - 4y'(t) + 8y = 0$; $y(0) = 3$, $y'(0) = 7$
2. $y''(t) + 4y'(t) + 8y = 2$; $y(0) = 2\frac{1}{4}$, $y'(0) = 4$
3. $y''(t) + 3y'(t) + 4y = 12$; $y(0) = 2$, $y'(0) = 2$
4. $y''(t) - 2y'(t) + 10y = 5$; $y(0) = 6$, $y'(0) = 8\frac{1}{2}$
5. $y''(t) + 9y = 3$; $y(0) = 1$, $y'(0) = 3$
6. $2y''(t) - 12y'(t) + 20y = 40$; $y(0) = 4$, $y'(0) = 5$
7. Which of the differential equations in Probs. 1 to 6 yield time paths with (*a*) damped fluctuation; (*b*) uniform fluctuation; (*c*) explosive fluctuation?

16.4 A Market Model with Price Expectations

In the earlier formulation of the dynamic market model, both Q_d and Q_s are taken to be functions of the current price P alone. But sometimes buyers and sellers may base their market behavior not only on the current price but also on the price *trend* prevailing at the time, for the price trend is likely to lead them to certain *expectations* regarding the price level in the future, and these expectations can, in turn, influence their demand and supply decisions.

Price Trend and Price Expectations

In the continuous-time context, the price-trend information is to be found primarily in the two derivatives dP/dt (whether price is rising) and d^2P/dt^2 (whether increasing at an increasing rate). To take the price trend into account, let us now include these derivatives as additional arguments in the demand and supply functions:

$$Q_d = D[P(t), P'(t), P''(t)]$$
$$Q_s = S[P(t), P'(t), P''(t)]$$

If we confine ourselves to the linear version of these functions and simplify the notation for the independent variables to P, P', and P'', we can write

$$\begin{aligned} Q_d &= \alpha - \beta P + mP' + nP'' \qquad (\alpha, \beta > 0) \\ Q_s &= -\gamma + \delta P + uP' + wP'' \qquad (\gamma, \delta > 0) \end{aligned} \tag{16.26}$$

where the parameters α, β, γ, and δ are merely carryovers from the previous market models, but m, n, u, and w are new.

The four new parameters, whose signs have not been restricted, embody the buyers' and sellers' price expectations. If $m > 0$, for instance, a rising price will cause Q_d to increase. This would suggest that buyers expect the rising price to *continue* to rise and, hence, prefer to increase their purchases now, when the price is still relatively low. The opposite sign for m would, on the other hand, signify the expectation of a prompt reversal of the price trend, so the buyers would prefer to cut back current purchases and wait for a lower price to materialize later. The inclusion of the parameter n makes the buyers' behavior depend also on the rate of change of dP/dt. Thus the new parameters m and n inject a substantial element of price speculation into the model. The parameters u and w carry a similar implication on the sellers' side of the picture.

A Simplified Model

For simplicity, we shall assume that only the demand function contains price expectations. Specifically, we let m and n be nonzero, but let $u = w = 0$ in (16.26). Further assume that *the market is cleared at every point of time*. Then we may equate the demand and supply functions to obtain (after normalizing) the differential equation

$$P'' + \frac{m}{n}P' - \frac{\beta + \delta}{n}P = -\frac{\alpha + \gamma}{n} \tag{16.27}$$

This equation is in the form of (16.2) with the following substitutions:

$$y = P \qquad a_1 = \frac{m}{n} \qquad a_2 = -\frac{\beta + \delta}{n} \qquad b = -\frac{\alpha + \gamma}{n}$$

Since this pattern of change of P involves the second derivative P'' as well as the first derivative P', the present model is certainly distinct from the dynamic market model presented in Sec. 15.2.

Note, however, that the present model differs from the previous model in yet another way. In Sec. 15.2, a dynamic adjustment mechanism, $dP/dt = j(Q_d - Q_s)$ is present. Since that equation implies that $dP/dt = 0$ if and only if $Q_d = Q_s$, the intertemporal sense and the market-clearing sense of equilibrium are coincident in that model. In contrast, the present model assumes market clearance at every moment of time. Thus every price attained in the market is an equilibrium price in the market-clearing sense, although it may not qualify as the intertemporal equilibrium price. In other words, the two senses of equilibrium are now disparate. Note, also, that the adjustment mechanism $dP/dt = j(Q_d - Q_s)$, containing a derivative, is what makes the previous market model dynamic

In the present model, with no adjustment mechanism, the dynamic nature of the model emanates instead from the expectation terms mP' and nP''.

The Time Path of Price

The intertemporal equilibrium price of this model—the particular integral P_p (formerly y_p)—is easily found by using (16.3). It is

$$P_p = \frac{b}{a_2} = \frac{\alpha+\gamma}{\beta+\delta}$$

Because this is a (positive) constant, it represents a stationary equilibrium.

As for the complementary function P_c (formerly y_c), there are three possible cases.

Case 1 (distinct real roots)

$$\left(\frac{m}{n}\right)^2 > -4\left(\frac{\beta+\delta}{n}\right)$$

The complementary function of this case is, by (16.7),

$$P_c = A_1e^{r_1t} + A_2e^{r_2t}$$

where

$$r_1, r_2 = \frac{1}{2}\left[-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta+\delta}{n}\right)}\right] \tag{16.28}$$

Accordingly, the general solution is

$$P(t) = P_c + P_p = A_1e^{r_1t} + A_2e^{r_2t} + \frac{\alpha+\gamma}{\beta+\delta} \tag{16.29}$$

Case 2 (double real roots)

$$\left(\frac{m}{n}\right)^2 = -4\left(\frac{\beta+\delta}{n}\right)$$

In this case, the characteristic roots take the single value

$$r = -\frac{m}{2n}$$

thus, by (16.9), the general solution may be written as

$$P(t) = A_3e^{-mt/2n} + A_4te^{-mt/2n} + \frac{\alpha+\gamma}{\beta+\delta} \tag{16.29'}$$

Case 3 (complex roots)

$$\left(\frac{m}{n}\right)^2 < -4\left(\frac{\beta+\delta}{n}\right)$$

In this third and last case, the characteristic roots are the pair of conjugate complex numbers

$$r_1, r_2 = h \pm vi$$

where

$$h = -\frac{m}{2n} \qquad \text{and} \qquad v = \frac{1}{2}\sqrt{-4\left(\frac{\beta+\delta}{n}\right) - \left(\frac{m}{n}\right)^2}$$

Therefore, by (16.24′), we have the general solution

$$P(t) = e^{-mt/2n}(A_5 \cos vt + A_6 \sin vt) + \frac{\alpha+\gamma}{\beta+\delta} \qquad (16.29'')$$

A couple of general conclusions can be deduced from these results. First, if $n > 0$, then $-4(\beta+\delta)/n$ must be negative and hence less than $(m/n)^2$. Hence Cases 2 and 3 can immediately be ruled out. Moreover, with n positive (as are β and δ), the expression under the square-root sign in (16.28) necessarily exceeds $(m/n)^2$, and thus the square root must be greater than $|m/n|$. The $\pm$ sign in (16.28) would then produce one positive root (r_1) and one negative root (r_2). Consequently, the intertemporal equilibrium is dynamically unstable, unless the definitized value of the constant A_1 happens to be zero in (16.29).

Second, if $n < 0$, then all three cases become feasible. Under Case 1, we can be sure that both roots will be negative if m is negative. (Why?) Interestingly, the repeated root of Case 2 will also be negative if m is negative. Moreover, since h, the real part of the complex roots in Case 3, takes the same value as the repeated root r in Case 2, the negativity of m will also guarantee that h is negative. In short, for all three cases, the dynamic stability of equilibrium is ensured when the parameters m and n are both negative.

Example 1

Let the demand and supply functions be

$$Q_d = 42 - 4P - 4P' + P''$$
$$Q_s = -6 + 8P$$

with initial conditions $P(0) = 6$ and $P'(0) = 4$. Assuming market clearance at every point of time, find the time path $P(t)$.

In this example, the parameter values are

$$\alpha = 42 \qquad \beta = 4 \qquad \gamma = 6 \qquad \delta = 8 \qquad m = -4 \qquad n = 1$$

Since n is positive, our previous discussion suggests that only Case 1 can arise, and that the two (real) roots r_1 and r_2 will take opposite signs. Substitution of the parameter values into (16.28) indeed confirms this, for

$$r_1, r_2 = \frac{1}{2}(4 \pm \sqrt{16+48}) = \frac{1}{2}(4 \pm 8) = 6, -2$$

The general solution is, then, by (16.29),

$$P(t) = A_1 e^{6t} + A_2 e^{-2t} + 4$$

By taking the initial conditions into account, moreover, we find that $A_1 = A_2 = 1$, so the definite solution is

$$P(t) = e^{6t} + e^{-2t} + 4$$

In view of the positive root $r_1 = 6$, the intertemporal equilibrium ($\bar{P}_p = 4$) is dynamically unstable.

The preceding solution is found by use of formulas (16.28) and (16.29). Alternatively, we can first equate the given demand and supply functions to obtain the differential equation

$$P'' - 4P' - 12P = -48$$

and then solve this equation as a specific case of (16.2).

Example 2 Given the demand and supply functions

$$Q_d = 40 - 2P - 2P' - P''$$
$$Q_s = -5 + 3P$$

with $P(0) = 12$ and $P'(0) = 1$, find $P(t)$ on the assumption that the market is always cleared.

Here the parameters m and n are both negative. According to our previous general discussion, therefore, the intertemporal equilibrium should be dynamically stable. To find the specific solution, we may first equate Q_d and Q_s to obtain the differential equation (after multiplying through by -1)

$$P'' + 2P' + 5P = 45$$

The intertemporal equilibrium is given by the particular integral

$$\bar{P}_p = \frac{45}{5} = 9$$

From the characteristic equation of the differential equation,

$$r^2 + 2r + 5 = 0$$

we find that the roots are complex:

$$r_1, r_2 = \frac{1}{2}(-2 \pm \sqrt{4 - 20}) = \frac{1}{2}(-2 \pm 4i) = -1 \pm 2i$$

This means that $h = -1$ and $v = 2$, so the general solution is

$$P(t) = e^{-t}(A_5 \cos 2t + A_6 \sin 2t) + 9$$

To definitize the arbitrary constants A_5 and A_6, we set $t = 0$ in the general solution, to get

$$P(0) = e^0(A_5 \cos 0 + A_6 \sin 0) + 9 = A_5 + 9 \qquad [\cos 0 = 1;\ \sin 0 = 0]$$

Moreover, by differentiating the general solution and then setting $t = 0$, we find that

$$P'(t) = -e^{-t}(A_5 \cos 2t + A_6 \sin 2t) + e^{-t}(-2A_5 \sin 2t + 2A_6 \cos 2t)$$

[product rule and chain rule]

and $$P'(0) = -e^0(A_5 \cos 0 + A_6 \sin 0) + e^0(-2A_5 \sin 0 + 2A_6 \cos 0)$$
$$= -(A_5 + 0) + (0 + 2A_6) = -A_5 + 2A_6$$

Thus, by virtue of the initial conditions $P(0) = 12$ and $P'(0) = 1$, we have $A_5 = 3$ and $A_6 = 2$. Consequently, the definite solution is

$$P(t) = e^{-t}(3 \cos 2t + 2 \sin 2t) + 9$$

This time path is obviously one with periodic fluctuation; the period is $2\pi/v = \pi$. That is, there is a complete cycle every time that t increases by $\pi = 3.14159\ldots$. In view of the multiplicative term e^{-t}, the fluctuation is damped. The time path, which starts from the initial price $P(0) = 12$, converges to the intertemporal equilibrium price $P_p = 9$ in a cyclical fashion.

EXERCISE 16.4

1. Let the parameters m, n, u, and w in (16.26) be all nonzero.
   (a) Assuming market clearance at every point of time, write the new differential equation of the model.
   (b) Find the intertemporal equilibrium price.
   (c) Under what circumstances can periodic fluctuation be ruled out?
2. Let the demand and supply functions be as in (16.26), but with $u = w = 0$ as in the text discussion.
   (a) If the market is not always cleared, but adjusts according to
   $$\frac{dP}{dt} = j(Q_d - Q_s) \qquad (j > 0)$$
   write the appropriate new differential equation.
   (b) Find the intertemporal equilibrium price $\bar{P}$ and the market-clearing equilibrium price P^*.
   (c) State the condition for having a fluctuating price path. Can fluctuation occur if $n > 0$?
3. Let the demand and supply be
   $$Q_d = 9 - P + P' + 3P'' \qquad Q_s = -1 + 4P - P' + 5P''$$
   with $P(0) = 4$ and $P'(0) = 4$.
   (a) Find the price path, assuming market clearance at every point of time.
   (b) Is the time path convergent? With fluctuation?

16.5 The Interaction of Inflation and Unemployment

In this section, we illustrate the use of a second-order differential equation with a macro model dealing with the problem of inflation and unemployment.

The Phillips Relation

One of the most widely used concepts in the modern analysis of the problem of inflation and unemployment is the Phillips relation.[1] In its original formulation, this relation depicts an empirically based negative relation between the rate of growth of money wage and the rate of unemployment:

$$w = f(U) \qquad [f'(U) < 0] \tag{16.30}$$

[1] A. W. Phillips, "The Relationship Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," *Economica*, November 1958, pp. 283–299.

where the lowercase letter w denotes the rate of growth of money wage W (i.e., $w \equiv \dot{W}/W$) and U is the rate of unemployment. It thus pertains only to the labor market. Later usage, however, has adapted the Phillips relation into a function that links the *rate of inflation* (instead of w) to the rate of unemployment. This adaptation may be justified by arguing that mark-up pricing is in wide use, so that a positive w, reflecting growing money-wage cost, would necessarily carry inflationary implications. And this makes the rate of inflation, like w, a function of U. The inflationary pressure of a positive w can, however, be offset by an increase in labor productivity, assumed to be exogeneous, and denoted here by T. Specifically, the inflationary effect can materialize only to the extent that money wage grows faster than productivity. Denoting the rate of inflation—that is, the rate of growth of the price level P—by the lowercase letter p, ($p \equiv \dot{P}/P$), we may thus write

$$p = w - T \tag{16.31}$$

Combining (16.30) and (16.31), and adopting the linear version of the function $f(U)$, we then get an adapted Phillips relation

$$p = \alpha - T - \beta U \qquad (\alpha, \beta > 0) \tag{16.32}$$

The Expectations-Augmented Phillips Relation

More recently, economists have preferred to use the *expectations-augmented* version of the Phillips relation

$$w = f(U) + g\pi \qquad (0 < g \leq 1) \tag{16.30'}$$

where π denotes the expected rate of inflation. The underlying idea of (16.30′), as propounded by the Nobel laureate Professor Friedman,[1] is that if an inflationary trend has been in effect long enough, people are apt to form certain inflation expectations which they then attempt to incorporate into their money-wage demands. Thus w should be an increasing function of π. Carried over to (16.32), this idea results in the equation

$$p = \alpha - T - \beta U + g\pi \qquad (0 < g \leq 1) \tag{16.33}$$

With the introduction of a new variable to denote the expected rate of inflation, it becomes necessary to hypothesize how inflation expectations are specifically formed.[2] Here we adopt the *adaptive expectations* hypothesis

$$\frac{d\pi}{dt} = j(p - \pi) \qquad (0 < j \leq 1) \tag{16.34}$$

Note that, rather than explain the absolute magnitude of π, this equation describes instead its pattern of change over time. If the actual rate of inflation p turns out to exceed the expected rate π, the latter, having now been proven to be too low, is revised upward ($d\pi/dt > 0$). Conversely, if p falls short of π, then π is revised in the downward direction. In format, (16.34) closely resembles the adjustment mechanism $dP/dt = j(Q_d - Q_s)$ of

[1] Milton Friedman, "The Role of Monetary Policy," *American Economic Review*, March 1968, pp. 1–17.

[2] This is in contrast to Sec. 16.4, where price expectations were discussed without introducing a new variable to represent the expected price. As a result, the assumptions regarding the formation of expectations were only implicitly embedded in the parameters m, n, u, and w in (16.26).

the market model. But here the driving force behind the adjustment is the discrepancy between the *actual* and *expected* rates of inflation, rather than Q_d and Q_s.

The Feedback from Inflation to Unemployment

It is possible to consider (16.33) and (16.34) as constituting a complete model. Since there are three variables in a two-equation system, however, one of the variables has to be taken as exogenous. If π and p are considered endogenous, for instance, then U must be treated as exogenous. A more satisfying alternative is to introduce a third equation to explain the variable U, so that the model will be richer in behavioral characteristics. More significantly, this will provide us with an opportunity to take into account the feedback effect of inflation on unemployment. Equation (16.33) tells us how U affects p—largely from the supply side of the economy. But p surely can affect U in return. For example, the rate of inflation may influence the consumption-saving decisions of the public, hence also the aggregate demand for domestic production, and the latter will, in turn, affect the rate of unemployment. Even in the conduct of government policies of demand management, the rate of inflation can make a difference in their effectiveness. Depending on the rate of inflation, a given level of money expenditure (fiscal policy) could translate into varying levels of real expenditure, and similarly, a given rate of nominal-money expansion (monetary policy) could mean varying rates of real-money expansion. And these, in turn, would imply differing effects on output and unemployment.

For simplicity, we shall only take into consideration the feedback through the conduct of monetary policy. Denoting the nominal money balance by M and its rate of growth by $m \equiv \dot{M}/M$, let us postulate that[†]

$$\frac{dU}{dt} = -k(m - p) \qquad (k > 0) \tag{16.35}$$

Recalling (10.25), and applying it backward, we see that the expression $(m - p)$ represents the rate of growth of *real* money:

$$m - p = \frac{\dot{M}}{M} - \frac{\dot{P}}{P} = r_M - r_P = r_{(M/P)}$$

Thus (16.35) stipulates that dU/dt is negatively related to the rate of growth of real-money balance. Inasmuch as the variable p now enters into the determination of dU/dt, the model now contains a feedback from inflation to unemployment.

The Time Path of π

Together, (16.33) through (16.35) constitute a closed model in the three variables π, p, and U. By eliminating two of the three variables, however, we can condense the model into a single differential equation in a single variable. Suppose that we let that single variable be π. Then we may first substitute (16.33) into (16.34) to get

$$\frac{d\pi}{dt} = j(\alpha - T - \beta U) - j(1 - g)\pi \tag{16.36}$$

[†] In an earlier discussion, we denoted the money supply by M_s, to distinguish it from the demand for money M_d. Here, we can simply use the unsubscripted letter M, since there is no fear of confusion.

Had this equation contained the expression dU/dt instead of U, we could have substituted (16.35) into (16.36) directly. But as (16.36) stands, we must first deliberately create a dU/dt term by differentiating (16.36) with respect to t, with the result

$$\frac{d^2\pi}{dt^2} = -j\beta\frac{dU}{dt} - j(1-g)\frac{d\pi}{dt} \tag{16.37}$$

Substitution of (16.35) into this then yields

$$\frac{d^2\pi}{dt^2} = j\beta km - j\beta kp - j(1-g)\frac{d\pi}{dt} \tag{16.37'}$$

There is still a p variable to be eliminated. To achieve that, we note that (16.34) implies

$$p = \frac{1}{j}\frac{d\pi}{dt} + \pi \tag{16.38}$$

Using this result in (16.37'), and simplifying, we finally obtain the desired differential equation in the variable π alone:

$$\frac{d^2\pi}{dt^2} + \underbrace{[\beta k + j(1-g)]}_{a_1}\frac{d\pi}{dt} + \underbrace{(j\beta k)}_{a_2}\pi = \underbrace{j\beta km}_{b} \tag{16.37''}$$

The particular integral of this equation is simply

$$\pi_p = \frac{b}{a_2} = m$$

Thus, in this model, the intertemporal equilibrium value of the expected rate of inflation hinges exclusively on the rate of growth of nominal money.

For the complementary function, the two roots are, as before,

$$r_1, r_2 = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right) \tag{16.39}$$

where, as may be noted from (16.37''), both a_1 and a_2 are positive. On a priori grounds, it is not possible to determine whether a_1^2 would exceed, equal, or be less than $4a_2$. Thus all three cases of characteristic roots—distinct real roots, repeated real roots, or complex roots—can conceivably arise. Whichever case presents itself, however, the intertemporal equilibrium will prove dynamically stable in the present model. This can be explained as follows: Suppose, first, that Case 1 prevails, with $a_1^2 > 4a_2$. Then the square root in (16.39) yields a real number. Since a_2 is positive, $\sqrt{a_1^2 - 4a_2}$ is necessarily less than $\sqrt{a_1^2} = a_1$. It follows that r_1 is negative, as is r_2, implying a dynamically stable equilibrium. What if $a_1^2 = 4a_2$ (Case 2)? In that event, the square root is zero, so that $r_1 = r_2 = -a_1/2 < 0$. And the negativity of the repeated roots again implies dynamic stability. Finally, for Case 3, the real part of the complex roots is $h = -a_1/2$. Since this has the same value as the repeated roots under Case 2, the identical conclusion regarding dynamic stability applies.

Although we have only studied the time path of π, the model can certainly yield information on the other variables, too. To find the time path of, say, the U variable, we can *either* start off by condensing the model into a differential equation in U rather than π (see Exercise 16.5-2) *or* deduce the U path from the π path already found (see Example 1).

Example 1

Let the three equations of the model take the specific forms

$$p = \frac{1}{6} - 3U + \pi \tag{16.40}$$

$$\frac{d\pi}{dt} = \frac{3}{4}(p - \pi) \tag{16.41}$$

$$\frac{dU}{dt} = -\frac{1}{2}(m - p) \tag{16.42}$$

Then we have the parameter values $\beta = 3$, $h = 1$, $j = \frac{3}{4}$, and $k = \frac{1}{2}$; thus, with reference to (16.37''), we find

$$a_1 = \beta k + j(1 - g) = \frac{3}{2} \qquad a_2 = j\beta k = \frac{9}{8} \qquad \text{and} \qquad b = j\beta k m = \frac{9}{8}m$$

The particular integral is $b/a_2 = m$. With $a_1^2 < 4a_2$, the characteristic roots are complex:

$$r_1, r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right) = \frac{1}{2}\left(-\frac{3}{2} \pm \frac{3}{2}i\right) = -\frac{3}{4} \pm \frac{3}{4}i$$

That is, $h = -\frac{3}{4}$ and $v = \frac{3}{4}$. Consequently, the general solution for the expected rate of inflation is

$$\pi(t) = e^{-3t/4}\left(A_5 \cos\frac{3}{4}t + A_6 \sin\frac{3}{4}t\right) + m \tag{16.43}$$

which depicts a time path with damped fluctuation around the equilibrium value m.

From this, we can also deduce the time paths for the p and U variables. According to (16.41), p can be expressed in terms of π and $d\pi/dt$ by the equation

$$p = \frac{4}{3}\frac{d\pi}{dt} + \pi$$

The π path in the general solution (16.43) implies the derivative

$$\begin{aligned}\frac{d\pi}{dt} &= -\frac{3}{4}e^{-3t/4}\left(A_5 \cos\frac{3}{4}t + A_6 \sin\frac{3}{4}t\right)\\ &\quad + e^{-3t/4}\left(-\frac{3}{4}A_5 \sin\frac{3}{4}t + \frac{3}{4}A_6 \cos\frac{3}{4}t\right) \qquad \text{[product rule and chain rule]}\end{aligned}$$

Using the solution (16.43) and its derivative, we thus have

$$p(t) = e^{-3t/4}\left(A_6 \cos\frac{3}{4}t - A_5 \sin\frac{3}{4}t\right) + m \tag{16.44}$$

Like the *expected* rate of inflation π, the *actual* rate of inflation p also has a fluctuating time path converging to the equilibrium value m.

As for the U variable, (16.40) tells us that it can be expressed in terms of π and p as follows:

$$U = \frac{1}{3}(\pi - p) + \frac{1}{18}$$

By virtue of the solutions (16.43) and (16.44), therefore, we can write the time path of the rate of unemployment as

$$U(t) = \frac{1}{3}e^{-3t/4}\left[(A_5 - A_6)\cos\frac{3}{4}t + (A_5 + A_6)\sin\frac{3}{4}t\right] + \frac{1}{18} \tag{16.45}$$

This path is, again, one with damped fluctuation, with $\frac{1}{18}$ as $\overline{U}$, the dynamically stable intertemporal equilibrium value of U.

Because the intertemporal equilibrium values of π and p are both equal to the monetary-policy parameter m, the value of m—the rate of growth of nominal money—provides the axis around which the time paths of π and p fluctuate. If a change occurs in m, a new equilibrium value of π and p will immediately replace the old one, and whatever values the π and p variables happen to take at the moment of the monetary-policy change will become the initial values from which the new π and p paths emanate.

In contrast, the intertemporal equilibrium value $\overline{U}$ does not depend on m. According to (16.45), U converges to the constant $\frac{1}{18}$ regardless of the rate of growth of nominal money, and hence regardless of the equilibrium rate of inflation. This constant equilibrium value of U is referred to as the *natural rate of unemployment*. The fact that the natural rate of unemployment is consistent with any equilibrium rate of inflation can be represented in the Up space by a vertical straight line parallel to the p axis. That vertical line relating the equilibrium values of U and p to each other, is known as the *long-run Phillips curve*. The vertical shape of this curve, however, is contingent upon a special parameter value assumed in this example. When that value is altered, as in Exercise 16.5-4, the long-run Phillips curve may no longer be vertical.

EXERCISE 16.5

1. In the inflation-unemployment model, retain (16.33) and (16.34) but delete (16.35) and let U be exogenous instead.
   (*a*) What kind of differential equation will now arise?
   (*b*) How many characteristic roots can you obtain? Is it possible now to have periodic fluctuation in the complementary function?
2. In the text discussion, we condensed the inflation-unemployment model into a differential equation in the variable π. Show that the model can alternatively be condensed into a second-order differential equation in the variable U, with the same a_1 and a_2 coefficients as in (16.37''), but a different constant term $b = kj[\alpha - T - (1 - g)m]$.
3. Let the adaptive expectations hypothesis (16.34) be replaced by the so-called perfect foresight hypothesis $\pi = p$, but retain (16.33) and (16.35).
   (*a*) Derive a differential equation in the variable p.
   (*b*) Derive a differential equation in the variable U.
   (*c*) How do these equations differ fundamentally from the one we obtained under the adaptive expectations hypothesis?
   (*d*) What change in parameter restriction is now necessary to make the new differential equations meaningful?
4. In Example 1, retain (16.41) and (16.42) but replace (16.40) by
   $$p = \frac{1}{6} - 3U + \frac{1}{3}\pi$$
   (*a*) Find $p(t)$, $\pi(t)$, and $U(t)$.
   (*b*) Are the time paths still fluctuating? Still convergent?
   (*c*) What are $\overline{p}$ and $\overline{U}$, the intertemporal equilibrium values of p and U?
   (*d*) Is it still true that $\overline{U}$ is functionally unrelated to $\overline{p}$? If we now link these two equilibrium values to each other in a long-run Phillips curve, can we still get a vertical curve? What assumption in Example 1 is thus crucial for deriving a vertical long-run Phillips curve?

16.6 Differential Equations with a Variable Term

In the differential equations considered in Sec. 16.1,

$$y''(t) + a_1 y'(t) + a_2 y = b$$

the right-hand term b is a constant. What if, instead of b, we have on the right a *variable term*: i.e., some function of t such as bt^2, e^{bt}, or $b \sin t$? The answer is that we must then modify our particular integral y_p. Fortunately, the complementary function is not affected by the presence of a variable term, because y_c deals only with the reduced equation, whose right side is always zero.

Method of Undetermined Coefficients

We shall explain a method of finding y_p, known as the *method of undetermined coefficients*, which is applicable to constant-coefficient variable-term differential equations, as long as the variable term and its successive derivatives together contain only a *finite* number of distinct types of expression (apart from multiplicative constants). The explanation of this method can best be carried out with a concrete illustration.

Example 1 Find the particular integral of

$$y''(t) + 5y'(t) + 3y = 6t^2 - t - 1 \qquad (16.46)$$

By definition, the particular integral is a value of y satisfying the given equation, i.e., a value of y that will make the left side identically equal to the right side regardless of the value of t. Since the left side contains the function $y(t)$ and the derivatives $y'(t)$ and $y''(t)$—whereas the right side contains multiples of the expressions t^2, t, and a constant—we ask: What general function form of $y(t)$, along with its first and second derivatives, will give us the three types of expression t^2, t, and a constant? The obvious answer is a function of the form $B_1t^2 + B_2t + B_3$ (where B_i are coefficients yet to be determined), for if we write the particular integral as

$$y(t) = B_1t^2 + B_2t + B_3$$

we can derive

$$y'(t) = 2B_1t + B_2 \qquad \text{and} \qquad y''(t) = 2B_1 \qquad (16.47)$$

and these three equations are indeed composed of the said types of expression. Substituting these into (16.46) and collecting terms, we get

$$\text{Left side} = (3B_1)t^2 + (10B_1 + 3B_2)t + (2B_1 + 5B_2 + 3B_3)$$

And when this is equated term by term to the right side, we can determine the coefficients B_i as follows:

$$\left.\begin{aligned} 3B_1 &= 6 \\ 10B_1 + 3B_2 &= -1 \\ 2B_1 + 5B_2 + 3B_3 &= -1 \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} B_1 &= 2 \\ B_2 &= -7 \\ B_3 &= 10 \end{aligned}\right.$$

Thus the desired particular integral can be written as

$$y_p = 2t^2 - 7t + 10$$

This method can work only when the number of expression types is finite. (See Exercise 16.6-1.) In general, when this prerequisite is met, the particular integral may be taken as being in the form of a linear combination of all the distinct expression types contained in the given variable term, as well as in all its derivatives. Note, in particular, that a constant expression should be included in the particular integral, if the original variable term or any of its successive derivatives contains a constant term.

Example 2 As a further illustration, let us find the general form for the particular integral suitable for the variable term $(b \sin t)$. Repeated differentiation yields, in this case, the successive derivatives $(b \cos t)$, $(-b \sin t)$, $(-b \cos t)$, $(b \sin t)$, etc., which involve only two distinct types of expression. We may therefore try a particular integral of the form $(B_1 \sin t + B_2 \cos t)$.

A Modification

In certain cases, a complication arises in applying the method. When the coefficient of the y term in the given differential equation is zero, such as in

$$y''(t) + 5y'(t) = 6t^2 - t - 1$$

the previously used trial form for the y_p, namely, $B_1t^2 + B_2t + B_3$, will fail to work. The cause of this failure is that, since the $y(t)$ term is out of the picture and since only derivatives $y'(t)$ and $y''(t)$ as shown in (16.47) will be substituted into the left side, no B_1t^2 term will ever appear on the left to be equated to the $6t^2$ term on the right. The way out of this kind of difficulty is to use instead the trial solution $t(B_1t^2 + B_2t + B_3)$; or if this too fails (e.g., given the equation $y''(t) = 6t^2 - t - 1$), to use $t^2(B_1t^2 + B_2t + B_3)$, and so on.

Indeed, the same trick may be employed in yet another difficult circumstance, as is illustrated in Example 3.

Example 3 Find the particular integral of

$$y''(t) + 3y'(t) - 4y = 2e^{-4t} \tag{16.48}$$

Here, the variable term is in the form of e^{-4t}, but all of its successive derivatives (namely, $-8e^{-4t}$, $32e^{-4t}$, $-128e^{-4t}$, etc.) take the same form as well. If we try the solution

$$y(t) = Be^{-4t} \qquad [\text{with } y'(t) = -4Be^{-4t} \text{ and } y''(t) = 16Be^{-4t}]$$

and substitute these into (16.48), we obtain the inauspicious result that

$$\text{Left side} = (16 - 12 - 4)Be^{-4t} = 0 \tag{16.49}$$

which obviously cannot be equated to the right-side term $2e^{-4t}$.

What causes this to happen is the fact that the exponential coefficient in the variable term (-4) happens to be equal to one of the roots of the characteristic equation of (16.48):

$$r^2 + 3r - 4 = 0 \qquad (\text{roots } r_1, r_2 = 1, -4)$$

The characteristic equation, it will be recalled, is obtained through a process of differentiation;[†] but the expression $(16 - 12 - 4)$ in (16.49) is derived through the same process. Not surprisingly, therefore, $(16 - 12 - 4)$ is merely a specific version of $(r^2 + 3r - 4)$ with r set equal to -4. Since -4 happens to be a characteristic root, the quadratic expression

$$r^2 + 3r - 4 = 16 - 12 - 4$$

must of necessity be identically zero.

[†] See the text discussion leading to (16.4?).

To cope with this situation, let us try instead the solution

$$y(t) = Bte^{-4t}$$

with derivatives

$$y'(t) = (1 - 4t)Be^{-4t} \qquad \text{and} \qquad y''(t) = (-8 + 16t)Be^{-4t}$$

Substituting these into (16.48) will now yield: left side $= -5Be^{-4t}$. When this is equated to the right side, we determine the coefficient to be $B = -2/5$. Consequently, the desired particular integral of (16.48) can be written as

$$y_p = \frac{-2}{5}te^{-4t}$$

EXERCISE 16.6

1. Show that the method of undetermined coefficients is inapplicable to the differential equation $y''(t) - ay'(t) + by = t^{-1}$.
2. Find the particular integral of each of the following equations by the method of undetermined coefficients:
   (a) $y''(t) + 2y'(t) + y = t$
   (b) $y''(t) + 4y'(t) + y = 2t^2$
   (c) $y''(t) + y'(t) + 2y = e^t$
   (d) $y''(t) + y'(t) + 3y = \sin t$

16.7 Higher-Order Linear Differential Equations

The methods of solution introduced in the previous sections are readily extended to an nth-order linear differential equation. With constant coefficients and a constant term, such an equation can be written generally as

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_ny = b \qquad \textbf{(16.50)}$$

Finding the Solution

In this case of constant coefficients and constant term, the presence of the higher derivatives does not materially affect the method of finding the particular integral discussed earlier.

If we try the simplest possible type of solution, $y = k$, we can see that all the derivatives from $y'(t)$ to $y^{(n)}(t)$ will be zero; hence (16.50) will reduce to $a_nk = b$, and we can write

$$y_p = k = \frac{b}{a_n} \qquad (a_n \neq 0) \qquad [\text{cf. } (16.3)]$$

In case $a_n = 0$, however, we must try a solution of the form $y = kt$. Then, since $y'(t) = k$, all the higher derivatives will vanish. (16.50) can be reduced to $a_{n-1}k = b$, thereby yielding the particular integral

$$y_p = kt = \frac{b}{a_{n-1}}t \qquad (a_n = 0;\ a_{n-1} \neq 0) \qquad [\text{cf. } (16.3')]$$

If it happens that $a_n = a_{n-1} = 0$, then this last solution will fail, too; instead, a solution of the form $y = kt^2$ must be tried. Further adaptations of this procedure should be obvious.

As for the complementary function, inclusion of the higher-order derivatives in the differential equation has the effect of raising the degree of the characteristic equation. The complementary function is defined as the general solution of the reduced equation

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_{n-1} y'(t) + a_n y = 0 \qquad \textbf{(16.51)}$$

Trying $y = Ae^{rt}\ (\neq 0)$ as a solution and utilizing the knowledge that this implies $y'(t) = rAe^{rt}, y''(t) = r^2Ae^{rt}, \ldots, y^{(n)}(t) = r^nAe^{rt}$, we can rewrite (16.51) as

$$Ae^{rt}(r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n) = 0$$

This equation is satisfied by any value of r which satisfies the following (nth-degree polynomial) characteristic equation

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0 \qquad \textbf{(16.51')}$$

There will, of course, be n roots to this polynomial, and each of these should be included in the general solution of (16.51). Thus our complementary function should in general be in the form

$$y_c = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \cdots + A_n e^{r_n t} \qquad \left(= \sum_{i=1}^{n} A_i e^{r_i t} \right)$$

As before, however, some modifications must be made in case the n roots are not all real and distinct. First, suppose that there are repeated roots, say, $r_1 = r_2 = r_3$. Then, to avoid "collapsing," we must write the first three terms of the solutions as $A_1 e^{r_1 t} + A_2 t e^{r_1 t} + A_3 t^2 e^{r_1 t}$ [cf. (16.9)]. In case we have $r_4 = r_1$ as well, the fourth term must be altered to $A_4 t^3 e^{r_1 t}$, etc.

Second, suppose that two of the roots are complex, say,

$$r_5, r_6 = h \pm vi$$

then the fifth and sixth terms in the preceding solution should be combined into the following expression:

$$e^{ht}(A_5 \cos vt + A_6 \sin vt) \qquad \text{[cf. (16.24')]}$$

By the same token, if two *distinct* pairs of complex roots are found, there must be two such trigonometric expressions (with a different set of values of h, v, and two arbitrary constants for each).[†] As a further possibility, if there happen to be two pairs of *repeated* complex roots, then we should use e^{ht} as the multiplicative term for one but use te^{ht} for the other. Also, even though h and v have identical values in the repeated complex roots, a different pair of arbitrary constants must now be assigned to each.

Once y_p and y_c are found, the general solution of the complete equation (16.50) follows easily. As before, it is simply the sum of the complementary function and the particular integral: $y(t) = y_c + y_p$. In this general solution, we can count a total of n arbitrary constants. Thus, to definitize the solution, as many as n initial conditions will be required.

[†] It is of interest to note that, inasmuch as complex roots always come in conjugate pairs, we can be sure of having *at least* one real root when the differential equation is of an *odd* order, i.e., when n is an odd number.

Example 1 Find the general solution of

$$y^{(4)}(t) + 6y'''(t) + 14y''(t) + 16y'(t) + 8y = 24$$

The particular integral of this fourth-order equation is simply

$$y_p = \frac{24}{8} = 3$$

Its characteristic equation is, by (16.51'),

$$r^4 + 6r^3 + 14r^2 + 16r + 8 = 0$$

which can be factored into the form

$$(r + 2)(r + 2)(r^2 + 2r + 2) = 0$$

From the first two parenthetical expressions, we can obtain the double roots $r_1 = r_2 = -2$, but the last (quadratic) expression yields the pair of complex roots $r_3, r_4 = -1 \pm i$, with $h = -1$ and $v = 1$. Consequently, the complementary function is

$$y_c = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t}(A_3 \cos t + A_4 \sin t)$$

and the general solution is

$$y(t) = A_1 e^{-2t} + A_2 t e^{-2t} + e^{-t}(A_3 \cos t + A_4 \sin t) + 3$$

The four constants A_1, A_2, A_3, and A_4 can be definitized, of course, if we are given four initial conditions.

Note that all the characteristic roots in this example either are real and negative or are complex and with a negative real part. The time path must therefore be convergent, and the intertemporal equilibrium is dynamically stable.

Convergence and the Routh Theorem

The solution of a high-degree characteristic equation is not always an easy task. For this reason, it should be of tremendous help if we can find a way of ascertaining the convergence or divergence of a time path without having to solve for the characteristic roots. Fortunately, there does exist such a method, which can provide a qualitative (though nongraphic) analysis of a differential equation.

This method is to be found in the *Routh theorem*,[†] which states that:

> The real parts of all of the roots of the nth-degree polynomial equation
>
> $$a_0 r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$
>
> are negative if and only if the first n of the following sequence of determinants
>
> $$|a_1|; \quad \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}; \quad \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}; \quad \ldots$$
>
> all are positive.

In applying this theorem, it should be remembered that $|a_1| \equiv a_1$. Further, it is to be understood that we should take $a_m = 0$ for all $m > n$. For example, given a third-degree

[†] For a discussion of this theorem, and a sketch of its proof, see Paul A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, 1947, pp. 429–435, and the references there cited.

polynomial equation ($n = 3$), we need to examine the signs of the first *three* determinants listed in the Routh theorem; for that purpose, we should set $a_4 = a_5 = 0$.

The relevance of this theorem to the convergence problem should become self-evident when we recall that, in order for the time path $y(t)$ to converge regardless of what the initial conditions happen to be, all the characteristic roots of the differential equation must have negative real parts. Since the characteristic equation (16.51') is an nth-degree polynomial equation, with $a_0 = 1$, the Routh theorem can be of direct help in the testing of convergence. In fact, we note that the coefficients of the characteristic equation (16.51') are wholly identical with those of the given differential equation (16.51), so it is perfectly acceptable to substitute the coefficients of (16.51) directly into the sequence of determinants shown in the Routh theorem for testing, provided that we always take $a_0 = 1$. Inasmuch as the condition cited in the theorem is given on the "if and only if" basis, it obviously constitutes a necessary-and-sufficient condition.

Example 2 Test by the Routh theorem whether the differential equation of Example 1 has a convergent time path. This equation is of the fourth order, so $n = 4$. The coefficient are $a_0 = 1$, $a_1 = 6$, $a_2 = 14$, $a_3 = 16$, $a_4 = 8$, and $a_5 = a_6 = a_7 = 0$. Substituting these into the first four determinants, we find their values to be 6, 68, 800, and 6,400, respectively. Because they are all positive, we can conclude that the time path is convergent.

EXERCISE 16.7

1. Find the particular integral of each of the following:
   (*a*) $y'''(t) + 2y''(t) + y'(t) + 2y = 8$
   (*b*) $y'''(t) + y''(t) + 3y'(t) = 1$
   (*c*) $3y'''(t) + 9y''(t) = 1$
   (*d*) $y^{(4)}(t) + y''(t) = 4$
2. Find the y_p and the y_c (and hence the general solution) of:
   (*a*) $y'''(t) - 2y''(t) - y'(t) + 2y = 4$
   [*Hint:* $r^3 - 2r^2 - r + 2 = (r - 1)(r + 1)(r - 2)$]
   (*b*) $y'''(t) + 7y''(t) + 15y'(t) + 9y = 0$
   [*Hint:* $r^3 + 7r^2 + 15r + 9 = (r - 1)(r^2 + 6r + 9)$]
   (*c*) $y'''(t) + 6y''(t) + 10y'(t) - 8y = 8$
   [*Hint:* $r^3 + 6r^2 + 10r + 8 = (r - 4)(r^2 + 2r + 2)$]
3. On the basis of the signs of the characteristic roots obtained in Prob. 2, analyze the dynamic stability of equilibrium. Then check your answer by the Routh theorem.
4. Without finding their characteristic roots, determine whether the following differential equations will give rise to convergent time paths:
   (*a*) $y'''(t) - 10y''(t) + 27y'(t) - 18y = 3$
   (*b*) $y'''(t) + 11y''(t) + 34y'(t) + 24y = 5$
   (*c*) $y'''(t) + 4y''(t) - 5y'(t) - 2y = -2$
5. Deduce from the Routh theorem that, for the second-order linear differential equation $y''(t) + a_1 y'(t) + a_2 y = b$, the solution path will be convergent regardless of initial conditions if and only if the coefficients a_1 and a_2 are both positive.

Chapter 17

Discrete Time: First-Order Difference Equations

In the continuous-time context, the pattern of change of a variable y is embodied in the derivatives $y'(t)$, $y''(t)$, etc. The time change involved in these is occurring continuously. When time is, instead, taken to be a *discrete* variable, so that the variable t is allowed to take integer values only, the concept of the derivative obviously will no longer be appropriate. Then, as we shall see, the pattern of change of the variable y must be described by so-called differences, rather than by derivatives or differentials, of $y(t)$. Accordingly, the techniques of differential equations will give way to those of *difference equations*.

When we are dealing with discrete time, the value of variable y will change only when the variable t changes from one integer value to the next, such as from $t = 1$ to $t = 2$. Meanwhile, nothing is supposed to happen to y. In this light, it becomes more convenient to interpret the values of t as referring to *periods*—rather than *points*—of time, with $t = 1$ denoting period 1 and $t = 2$ denoting period 2, and so forth. Then we may simply regard y as having one unique value in each time period. In view of this interpretation, the discrete-time version of economic dynamics is often referred to as *period analysis*. It should be emphasized, however, that "period" is being used here not in the calendar sense but in the analytical sense. Hence, a period may involve one extent of calendar time in a particular economic model, but an altogether different one in another. Even in the same model, moreover, each successive period should not necessarily be construed as meaning equal calendar time. In the analytical sense, a period is merely a length of time that elapses before the variable y undergoes a change.

17.1 Discrete Time, Differences, and Difference Equations

The change from continuous time to discrete time produces no effect on the fundamental nature of dynamic analysis, although the formulation of the problem must be altered. Basically, our dynamic problem is still to find a time path from some given pattern of change of a variable y over time. But the pattern of change should now be represented by the difference quotient $\Delta y/\Delta t$, which is the discrete-time counterpart of the derivative dy/dt. Recall, however, that t can now take only integer values; thus, when we are comparing the values of y in two consecutive periods, we must have $\Delta t = 1$. For this reason, the difference quotient $\Delta y/\Delta t$ can be simplified to the expression Δy; this is called the *first difference* of y. The symbol Δ, meaning difference, can accordingly be interpreted as a directive to take the first difference of (y). As such, it constitutes the discrete-time counterpart of the operator symbol d/dt.

The expression Δy can take various values, of course, depending on which two consecutive time periods are involved in the difference-taking (or "differencing"). To avoid ambiguity, let us add a time subscript to y and define the first difference more specifically, as follows

$$\Delta y_t \equiv y_{t+1} - y_t \tag{17.1}$$

where y_t means the value of y in the tth period, and y_{t+1} is its value in the period immediately following the tth period. With this symbology, we may describe the pattern of change of y by an equation such as

$$\Delta y_t = 2 \tag{17.2}$$

or

$$\Delta y_t = -0.1 y_t \tag{17.3}$$

Equations of this type are called *difference equations*. Note the striking resemblance between the last two equations, on the one hand, and the differential equations $dy/dt = 2$ and $dy/dt = -0.1y$ on the other.

Even though difference equations derive their name from difference expressions such as Δy_t, there are alternate equivalent forms of such equations which are completely free of Δ expressions and which are more convenient to use. By virtue of (17.1), we can rewrite (17.2) as

$$y_{t+1} - y_t = 2 \tag{17.2'}$$

or

$$y_{t+1} = y_t + 2 \tag{17.2''}$$

For (17.3), the corresponding alternate equivalent forms are

$$y_{t+1} - 0.9y_t = 0 \tag{17.3'}$$

or

$$y_{t+1} = 0.9y_t \tag{17.3''}$$

The double-prime-numbered versions will prove convenient when we are calculating a y value from a known y value of the preceding period. In later discussions, however, we shall employ mostly the single-prime-numbered versions, i.e., those of (17.2') and (17.3').

It is important to note that the choice of time subscripts in a difference equation is somewhat arbitrary. For instance, without any change in meaning, (17.2') can be rewritten as $y_t - y_{t-1} = 2$, where $(t - 1)$ refers to the period which immediately precedes the tth. Or, we may express it equivalently as $y_{t+2} - y_{t+1} = 2$.

Also, it may be pointed out that, although we have consistently used subscripted y symbols, it is also acceptable to use $y(t)$, $y(t+1)$, and $y(t-1)$ in their stead. In order to avoid using the notation $y(t)$ for both continuous-time and discrete-time cases, however, we shall, in the discussion of period analysis, adhere to the subscript device.

Analogous to differential equations, difference equations can be either linear or nonlinear, homogeneous or nonhomogeneous, and of the first or second (or higher) orders. Take (17.2′) for instance. It can be classified as: (1) linear, for no y term (of any period) is raised to the second (or higher) power or is multiplied by a y term of another period; (2) nonhomogeneous, since the right-hand side (where there is no y term) is nonzero; and (3) of the first order, because there exists only a *first difference* Δy_t, involving a one-period time lag only. (In contrast, a second-order difference equation, to be discussed in Chap. 18, involves a two-period lag and thus entails three y terms: y_{t+2}, y_{t+1}, as well as y_t.)

Actually, (17.2′) can also be characterized as having constant coefficients and a constant term (= 2). Since the constant-coefficient case is the only one we shall consider, this characterization will henceforth be implicitly assumed. Throughout the present chapter, the constant-term feature will also be retained, although a method of dealing with the variable-term case will be discussed in Chap. 18.

Check that the equation (17.3′) is also linear and of the first order; but unlike (17.2′), it is homogeneous.

17.2 Solving a First-Order Difference Equation

In solving a differential equation, our objective was to find a time path $y(t)$. As we know, such a time path is a function of time which is totally free from any derivative (or differential) expressions and which is perfectly consistent with the given differential equation as well as with its initial conditions. The time path we seek from a difference equation is similar in nature. Again, it should be a function of t—a formula defining the values of y in every time period—which is consistent with the given difference equation as well as with its initial conditions. Besides, it must not contain any difference expressions such as Δy_t (or expressions like $y_{t+1} - y_t$).

Solving differential equations is, in the final analysis, a matter of integration. How do we solve a difference equation?

Iterative Method

Before developing a general method of attack, let us first explain a relatively pedestrian method, the *iterative method*—which, though crude, will prove immensely revealing of the essential nature of a so-called solution.

In this chapter we are concerned only with the first-order case; thus the difference equation describes the pattern of change of y between *two* consecutive periods only. Once such a pattern is specified, such as by (17.2″), and once we are given an initial value y_0, it is no problem to find y_1 from the equation. Similarly, once y_1 is found, y_2 will be immediately obtainable, and so forth, by repeated application (iteration) of the pattern of change specified in the difference equation. The results of iteration will then permit us to infer a time path.

Example 1 Find the solution of the difference equation (17.2), assuming an initial value of $y_0 = 15$. To carry out the iterative process, it is more convenient to use the alternative form of the difference equation (17.2''), namely, $y_{t+1} = y_t + 2$, with $y_0 = 15$. From this equation, we can deduce step-by-step that

$$y_1 = y_0 + 2$$
$$y_2 = y_1 + 2 = (y_0 + 2) + 2 = y_0 + 2(2)$$
$$y_3 = y_2 + 2 = [y_0 + 2(2)] + 2 = y_0 + 3(2)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

and, in general, for any period t,

$$y_t = y_0 + t(2) = 15 + 2t \tag{17.4}$$

This last equation indicates the y value of any time period (including the initial period $t = 0$); it therefore constitutes the solution of (17.2).

The process of iteration is crude—it corresponds roughly to solving simple differential equations by straight integration—but it serves to point out clearly the manner in which a time path is generated. In general, the value of y_t will depend in a specified way on the value of y in the immediately preceding period (y_{t-1}); thus a given initial value y_0 will successively lead to $y_1, y_2, \ldots$, via the prescribed pattern of change.

Example 2 Solve the difference equation (17.3); this time, let the initial value be unspecified and denoted simply by y_0. Again it is more convenient to work with the alternative version in (17.3''), namely, $y_{t+1} = 0.9y_t$. By iteration, we have

$$y_1 = 0.9y_0$$
$$y_2 = 0.9y_1 = 0.9(0.9y_0) = (0.9)^2 y_0$$
$$y_3 = 0.9y_2 = 0.9(0.9)^2 y_0 = (0.9)^3 y_0$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

These can be summarized into the solution

$$y_t = (0.9)^t y_0 \tag{17.5}$$

To heighten interest, we can lend some economic content to this example. In the simple multiplier analysis, a single investment expenditure in period 0 will call forth successive rounds of spending, which in turn will bring about varying amounts of income increment in succeeding time periods. Using y to denote *income increment*, we have y_0 = the amount of investment in period 0; but the subsequent income increments will depend on the marginal propensity to consume (MPC). If MPC = 0.9 and if the income of each period is consumed only in the next period, then 90 percent of y_0 will be consumed in period 1, resulting in an income increment in period 1 of $y_1 = 0.9y_0$. By similar reasoning, we can find $y_2 = 0.9y_1$, etc. These, we see, are precisely the results of the iterative process cited previously. In other words, the multiplier process of income generation can be described by a difference equation such as (17.3''), and a solution like (17.5) will tell us what the magnitude of income increment is to be in any time period t.

Example 3 Solve the homogeneous difference equation

$$my_{t+1} - ny_t = 0$$

Upon normalizing and transposing, this may be written as

$$y_{t+1} = \left(\frac{n}{m}\right) y_t$$

which is the same as (17.3″) in Example 2 except for the replacement of 0.9 by n/m. Hence, by analogy, the solution should be

$$y_t = \left(\frac{n}{m}\right)^t y_0$$

Watch the term $\left(\frac{n}{m}\right)^t$. It is through this term that various values of t will lead to their corresponding values of y. It therefore corresponds to the expression e^{rt} in the solutions to differential equations. If we write it more generally as b^t (b for base) and attach the more general multiplicative constant A (instead of y_0), we see that the solution of the general homogeneous difference equation of Example 3 will be in the form

$$y_t = Ab^t$$

We shall find that this expression Ab^t will play the same important role in difference equations as the expression Ae^{rt} did in differential equations.† However, even though both are exponential expressions, the former is to the base b, whereas the latter is to the base e. It stands to reason that, just as the type of the continuous-time path $y(t)$ depends heavily on the value of r, the discrete-time path y_t hinges principally on the value of b.

General Method

By this time, you must have become quite impressed with the various similarities between differential and difference equations. As might be conjectured, the general method of solution presently to be explained will parallel that for differential equations.

Suppose that we are seeking the solution to the first-order difference equation

$$y_{t+1} + ay_t = c \tag{17.6}$$

where a and c are two constants. The general solution will consist of the sum of two components: a *particular solution* y_p, which is *any* solution of the complete nonhomogeneous equation (17.6), and a *complementary function* y_c, which is the general solution of the reduced equation of (17.6):

$$y_{t+1} + ay_t = 0 \tag{17.7}$$

The y_p component again represents the intertemporal equilibrium level of y, and the y_c component, the deviations of the time path from that equilibrium. The sum of y_c and y_p constitutes the *general* solution, because of the presence of an arbitrary constant. As before, in order to definitize the solution, an initial condition is needed.

Let us first deal with the complementary function. Our experience with Example 3 suggests that we may try a solution of the form $y_t = Ab^t$ (with $Ab^t \neq 0$, for otherwise y_t will turn out simply to be a horizontal straight line lying on the t axis); in that case, we also

† You may object to this statement by pointing out that the solution (17.4) in Example 1 does not contain a term in the form of Ab^t. This latter fact, however, arises only because in Example 1 we have $b = n/m = 1/1 = 1$, so that the term Ab^t reduces to a constant.

have $y_{t+1} = Ab^{t+1}$. If these values of y_t and y_{t+1} hold, the homogeneous equation (17.7) will become

$$Ab^{t+1} + aAb^t = 0$$

which, upon canceling the nonzero common factor Ab^t, yields

$$b + a = 0 \qquad \text{or} \qquad b = -a$$

This means that, for the trial solution to work, we must set $b = -a$; then the complementary function should be written as

$$y_c (= Ab^t) = A(-a)^t$$

Now let us search for the particular solution, which has to do with the complete equation (17.6). In this regard, Example 3 is of no help at all, because that example relates only to a homogeneous equation. However, we note that for y_p we can choose *any* solution of (17.6); thus if a trial solution of the simplest form $y_t = k$ (a constant) can work out, no real difficulty will be encountered. Now, if $y_t = k$, then y will maintain the same constant value over time, and we must have $y_{t+1} = k$ also. Substitution of these values into (17.6) yields

$$k + ak = c \qquad \text{and} \qquad k = \frac{c}{1+a}$$

Since this particular k value satisfies the equation, the particular integral can be written as

$$y_p (= k) = \frac{c}{1+a} \qquad (a \neq -1)$$

This being a constant, a stationary equilibrium is indicated in this case.

If it happens that $a = -1$, as in Example 1, however, the particular solution $c/(1+a)$ is not defined, and some other solution of the nonhomogeneous equation (17.6) must be sought. In this event, we employ the now-familiar trick of trying a solution of the form $y_t = kt$. This implies, of course, that $y_{t+1} = k(t+1)$. Substituting these into (17.6), we find

$$k(t+1) + akt = c \qquad \text{and} \qquad k = \frac{c}{t+1+at} = c \qquad [\text{because } a = -1]$$

thus $$y_p (= kt) = ct$$

This form of the particular solution is a nonconstant function of t; it therefore represents a moving equilibrium.

Adding y_c and y_p together, we may now write the general solution in one of the two following forms:

$$y_t = A(-a)^t + \frac{c}{1+a} \qquad [\text{general solution, case of } a \neq -1] \qquad \textbf{(17.8)}$$

$$y_t = A(-a)^t + ct = A + ct \qquad [\text{general solution, case of } a = -1] \qquad \textbf{(17.9)}$$

Neither of these is completely determinate, in view of the arbitrary constant A. To eliminate this arbitrary constant, we resort to the initial condition that $y_t = y_0$ when $t = 0$. Letting $t = 0$ in (17.8), we have

$$y_0 = A + \frac{c}{1+a} \qquad \text{and} \qquad A = y_0 - \frac{c}{1+a}$$

Consequently, the definite version of (17.8) is

$$y_t = \left(y_0 - \frac{c}{1+a}\right)(-a)^t + \frac{c}{1+a} \qquad [\textit{definite solution, case of } a \neq -1] \qquad \textbf{(17.8')}$$

Letting $t = 0$ in (17.9), on the other hand, we find $y_0 = A$, so the definite version of (17.9) is

$$y_t = y_0 + ct \qquad [\text{definite solution, case of } a = -1] \qquad \textbf{(17.9')}$$

If this last result is applied to Example 1, the solution that emerges is exactly the same as the iterative solution (17.4).

You can check the validity of each of these solutions by the following two steps. First, by letting $t = 0$ in (17.8'), see that the latter equation reduces to the identity $y_0 = y_0$, signifying the satisfaction of the initial condition. Second, by substituting the y_t formula (17.8') and a similar y_{t+1} formula—obtained by replacing t with $(t + 1)$ in (17.8')—into (17.6), see that the latter reduces to the identity $c = c$, signifying that the time path is consistent with the given difference equation. The check on the validity of solution (17.9') is analogous.

Example 4 Solve the first-order difference equation

$$y_{t+1} - 5y_t = 1 \qquad \left(y_0 = \tfrac{7}{4}\right)$$

Following the procedure used in deriving (17.8'), we can find y_c by trying a solution $y_t = Ab^t$ (which implies $y_{t+1} = Ab^{t+1}$). Substituting these values into the homogeneous version $y_{t+1} - 5y_t = 0$ and canceling the common factor Ab^t, we get $b = 5$. Thus

$$y_c = A(5)^t$$

To find y_p, try the solution $y_t = k$, which implies $y_{t+1} = k$. Substituting these into the complete difference equation, we find $k = -\frac{1}{4}$. Hence

$$y_p = -\tfrac{1}{4}$$

It follows that the general solution is

$$y_t = y_c + y_p = A(5)^t - \tfrac{1}{4}$$

Letting $t = 0$ here and utilizing the initial condition $y_0 = \frac{7}{4}$, we obtain $A = 2$. Thus the definite solution may finally be written as

$$y_t = 2(5)^t - \tfrac{1}{4}$$

Since the given difference equation of this example is a special case of (17.6), with $a = -5$, $c = 1$, and $y_0 = \frac{7}{4}$, and since (17.8') is the solution "formula" for this type of difference equation, we could have found our solution by inserting the specific parameter values into (17.8'), with the result that

$$y_t = \left(\frac{7}{4} - \frac{1}{1-5}\right)(5)^t + \frac{1}{1-5} = 2(5)^t - \frac{1}{4}$$

which checks perfectly with the earlier answer.

Note that the y_{t+1} term in (17.6) has a unit coefficient. If a given difference equation has a nonunit coefficient for this term, it must be normalized before using the solution formula (17.8').

EXERCISE 17.2

1. Convert the following difference equations into the form of (17.2'):
   (*a*) $\Delta y_t = 7$
   (*b*) $\Delta y_t = 0.3y_t$
   (*c*) $\Delta y_t = 2y_t - 9$
2. Solve the following difference equations by iteration:
   (*a*) $y_{t+1} = y_t - 1$ $\quad (y_0 = 10)$
   (*b*) $y_{t+1} = \alpha y_t$ $\quad (y_0 = \beta)$
   (*c*) $y_{t+1} = \alpha y_t - \beta$ $\quad (y_t = y_0 \text{ when } t = 0)$
3. Rewrite the equations in Prob. 2 in the form of (17.6), and solve by applying formula (17.8') or (17.9'), whichever is appropriate. Do your answers check with those obtained by the iterative method?
4. For each of the following difference equations, use the procedure illustrated in the derivation of (17.8') and (17.9') to find y_c, y_p, and the definite solution:
   (*a*) $y_{t+1} + 3y_t = 4$ $\quad (y_0 = 4)$
   (*b*) $2y_{t+1} - y_t = 6$ $\quad (y_0 = 7)$
   (*c*) $y_{t+1} = 0.2y_t + 4$ $\quad (y_0 = 4)$

17.3 The Dynamic Stability of Equilibrium

In the continuous-time case, the dynamic stability of equilibrium depends on the Ae^{rt} term in the complementary function. In period analysis, the corresponding role is played by the Ab^t term in the complementary function. Since its interpretation is somewhat more complicated than Ae^{rt}, let us try to clarify it before proceeding further.

The Significance of *b*

Whether the equilibrium is dynamically stable is a question of whether or not the complementary function will tend to zero as $t \to \infty$. Basically, we must analyze the path of the term Ab^t as t is increased indefinitely. Obviously, the value of b (the base of this exponential term) is of crucial importance in this regard. Let us first consider its significance alone, by disregarding the coefficient A (by assuming $A = 1$).

For analytical purposes, we can divide the range of possible values of b, $(-\infty, +\infty)$, into seven distinct regions, as set forth in the first two columns of Table 17.1, arranged in descending order of magnitude of b. These regions are also marked off in Fig. 17.1 on a vertical b scale, with the points $+1$, 0, and -1 as the demarcation points. In fact, these latter three points in themselves constitute the regions II, IV, and VI. Regions III and V, on the other hand, correspond to the set of all positive fractions and the set of all negative fractions, respectively. The remaining two regions, I and VII, are where the numerical value of b exceeds unity.

In each region, the exponential expression b^t generates a different type of time path. These are exemplified in Table 17.1 and illustrated in Fig. 17.1. In region I (where $b > 1$), b^t must increase with t at an increasing pace. The general configuration of the time path will therefore assume the shape of the top graph in Fig. 17.1. Note that this graph is shown

TABLE 17.1
A Classification of the Values of b

Region	Value of b		Value of b^t	Value of b^t in Different Time Periods: $t=0$	$t=1$	$t=2$	$t=3$	$t=4\cdots$
I	$b>1$	$(\lvert b\rvert>1)$	e.g., $(2)^t$	1	2	4	8	16
II	$b=1$	$(\lvert b\rvert=1)$	$(1)^t$	1	1	1	1	1
III	$0<b<1$	$(\lvert b\rvert<1)$	e.g., $\left(\frac{1}{2}\right)^t$	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
IV	$b=0$	$(\lvert b\rvert=0)$	$(0)^t$	0	0	0	0	0
V	$-1<b<0$	$(\lvert b\rvert<1)$	e.g., $\left(-\frac{1}{2}\right)^t$	1	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{8}$	$\frac{1}{16}$
VI	$b=-1$	$(\lvert b\rvert=1)$	$(-1)^t$	1	-1	1	-1	1
VII	$b<-1$	$(\lvert b\rvert>1)$	e.g., $(-2)^t$	1	-2	4	-8	16

as a step function rather than as a smooth curve; this is because we are dealing with period analysis. In region II ($b=1$), b^t will remain at unity for all values of t. Its graph will thus be a horizontal straight line. Next, in region III, b^t represents a positive fraction raised to integer powers. As the power is increased, b^t must decrease, though it will always remain positive. The next case, that of $b=0$ in region IV, is quite similar to the case of $b=1$; but here we have $b^t=0$ rather than $b^t=1$, so its graph will coincide with the horizontal axis. However, this case is of peripheral interest only, since we have earlier adopted the assumption that $Ab^t \neq 0$

When we move into the negative regions, an interesting new phenomenon occurs: The value of b^t will *alternate* between positive and negative values from period to period! This fact is clearly brought out in the last three rows of Table 17.1 and in the last three graphs of Fig. 17.1. In region V, where b is a negative fraction, the alternating time path tends to get closer and closer to the horizontal axis (cf. the positive-fraction region, III). In contrast, when $b=-1$ (region VI), a perpetual alternation between $+1$ and -1 results. And finally, when $b<-1$ (region VII), the alternating time path will deviate farther and farther from the horizontal axis.

What is striking is that, whereas the phenomenon of a fluctuating time path cannot possibly arise from a single Ae^{rt} term (the complex-root case of the second-order differential equation requires a *pair* of complex roots), fluctuation can be generated by a single b^t (or Ab^t) term. Note, however, that the character of the fluctuation is somewhat different; unlike the circular-function pattern, the fluctuation depicted in Fig. 17.1 is nonsmooth. For this reason, we shall employ the word *oscillation* to denote the new, nonsmooth type of fluctuation, even though many writers do use the terms fluctuation and oscillation interchangeably.

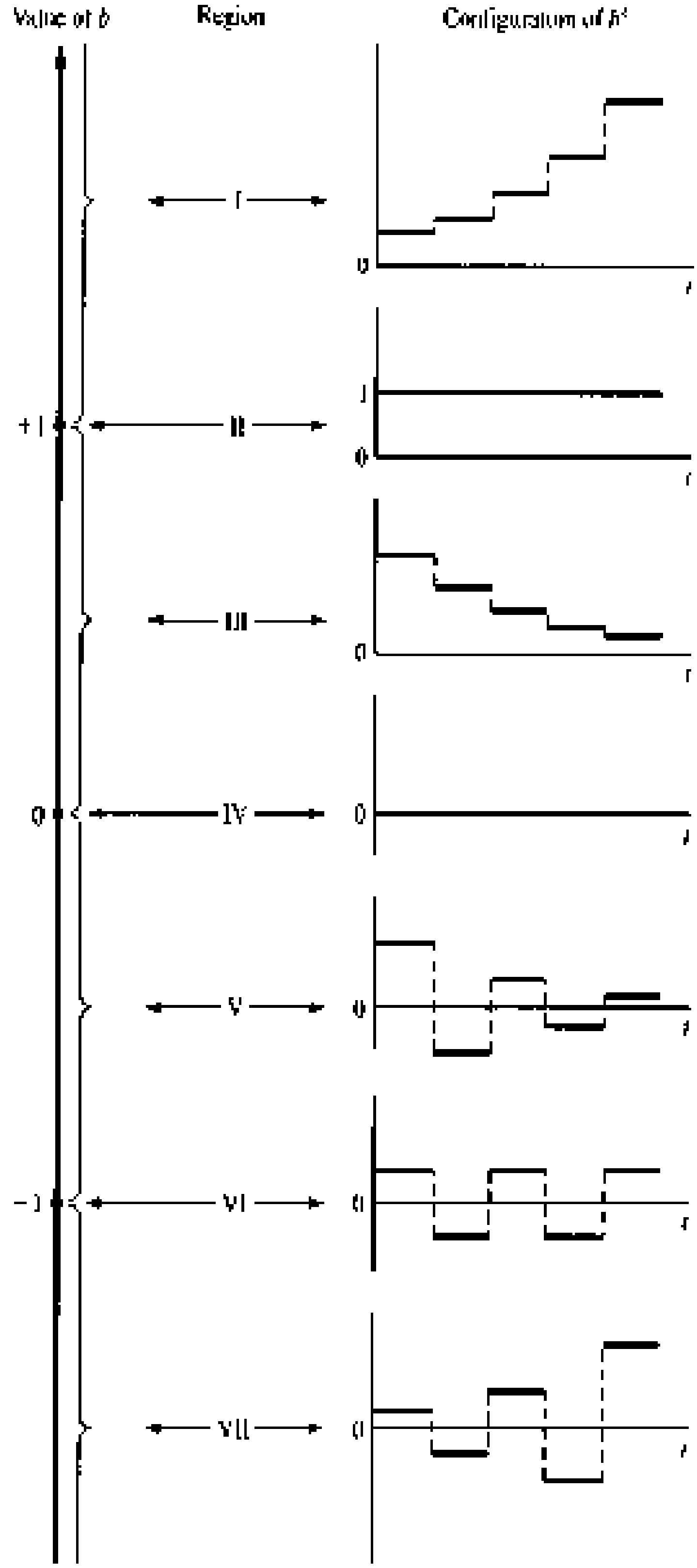
The essence of the preceding discussion can be conveyed in the following general statement: The time path of b^t ($b \neq 0$) will be

$$\left.\begin{matrix}\text{Nonoscillatory}\\ \text{Oscillatory}\end{matrix}\right\} \quad \text{if} \quad \begin{cases} b>0 \\ b<0 \end{cases}$$

$$\left.\begin{matrix}\text{Divergent}\\ \text{Convergent}\end{matrix}\right| \quad \text{if} \quad \begin{cases} |b|>1 \\ |b|<1 \end{cases}$$

It is important to note that, whereas the convergence of the expression e^{rt} depends on the *sign* of r, the convergence of the b^t expression hinges, instead, on the *absolute value* of b.

FIGURE 17.1



The Role of A

So far we have deliberately left out the multiplicative constant A. But its effects—of which there are two—are relatively easy to take into account. First, the *magnitude* of A can serve to "blow up" (if, say, $A = 3$) or "pare down" (if, say, $A = \frac{1}{5}$) the values of b^t. That is, it can produce a *scale effect* without changing the basic configuration of the time path. The *sign* of A, on the other hand, does materially affect the shape of the path because, if b^t is multiplied by $A = -1$, then each time path shown in Fig. 17.1 will be replaced by its own mirror image with reference to the horizontal axis. Thus, a negative A can produce a *mirror effect* as well as a scale effect.

Convergence to Equilibrium

The preceding discussion presents the interpretation of the Ab^t term in the complementary function, which, as we recall, represents the deviations from some intertemporal equilibrium level. If a term (say) $y_p = 5$ is added to the Ab^t term, the time path must be shifted up vertically by a constant value of 5. This will in no way affect the convergence or divergence of the time path, but it will alter the level with reference to which convergence or divergence is gauged. What Fig. 17.1 pictures is the convergence (or lack of it) of the Ab^t expression to zero. When the y_p is included, it becomes a question of the convergence of the time path $y_t = y_c + y_p$ to the equilibrium level y_p.

In this connection, let us add a word of explanation for the special case of $b = 1$ (region III). A time path such as

$$y_t = A(1)^t + y_p = A + y_p$$

gives the impression that it converges, because the multiplicative term $(1)^t = 1$ produces no explosive effect. Observe, however, that y_t will now take the value $(A + y_p)$ rather than the equilibrium value y_p; in fact, it can never reach y_p (unless $A = 0$). As an illustration of this type of situation, we can cite the time path in (17.9), in which a moving equilibrium $y_p = ct$ is involved. This time path is to be considered divergent, not because of the appearance of t in the particular solution but because, with a nonzero A, there will be a constant deviation from the moving equilibrium. Thus, in stipulating the condition for convergence of time path y_t to the equilibrium y_p, we must rule out the case of $b = 1$.

In sum, the solution

$$y_t = Ab^t + y_p$$

is a convergent path if and only if $|b| < 1$.

Example 1 What kind of time path is represented by $y_t = 2(-\frac{4}{5})^t + 9$? Since $b = -\frac{4}{5} < 0$, the time path is oscillatory. But since $|b| = \frac{4}{5} < 1$, the oscillation is damped, and the time path converges to the equilibrium level of 9.

You should exercise care not to confuse $2(-\frac{4}{5})^t$ with $-2(\frac{4}{5})^t$; they represent entirely different time-path configurations.

Example 2 How do you characterize the time path $y_t = 3(2)^t + 4$? Since $b = 2 > 0$, no oscillation will occur. But since $|b| = 2 > 1$, the time path will diverge from the equilibrium level of 4.

EXERCISE 17.3

1. Discuss the nature of the following time paths:

(a) $y_t = 3^t + 1$ (c) $y_t = 5\left(-\frac{1}{10}\right)^t + 3$

(b) $y_t = 2\left(\frac{1}{3}\right)^t$ (d) $y_t = -3\left(\frac{1}{4}\right)^t + 2$

2. What is the nature of the time path obtained from each of the difference equations in Exercise 17.2-4?
3. Find the solutions of the following, and determine whether the time paths are oscillatory and convergent:
   (a) $y_{t+1} - \frac{1}{3}y_t = 6 \qquad (y_0 = 1)$
   (b) $y_{t+1} + 2y_t = 9 \qquad (y_0 = 4)$
   (c) $y_{t+1} + \frac{1}{4}y_t = 5 \qquad (y_0 = 2)$
   (d) $y_{t+1} - y_t = 3 \qquad (y_0 = 5)$

17.4 The Cobweb Model

To illustrate the use of first-order difference equations in economic analysis, we shall cite two variants of the market model for a single commodity. The first variant, known as the *cobweb model*, differs from our earlier market models in that it treats Q_s as a function not of the current price but of the price of the preceding time period.

The Model

Consider a situation in which the producer's output decision must be made one period in advance of the actual sale—such as in agricultural production, where planting must precede by an appreciable length of time the harvesting and sale of the output. Let us assume that the output decision in period t is based on the then-prevailing price P_t. Since this output will not be available for the sale until period $(t+1)$, however, P_t will determine not Q_{st}, but $Q_{s,t+1}$. Thus we now have a "lagged" supply function†

$$Q_{s,t+1} = S(P_t)$$

or, equivalently, by shifting back the time subscripts by one period,

$$Q_{st} = S(P_{t-1})$$

When such a supply function interacts with a demand function of the form

$$Q_{dt} = D(P_t)$$

interesting dynamic price patterns will result.

Taking the linear versions of these (lagged) supply and (unlagged) demand functions, and assuming that in each time period the market price is always set at a level which clears the market, we have a market model with the following three equations:

$$\begin{aligned} Q_{dt} &= Q_{st} \\ Q_{dt} &= \alpha - \beta P_t \qquad (\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_{t-1} \qquad (\gamma, \delta > 0) \end{aligned} \tag{17.10}$$

† We are making the implicit assumption here that the entire output of a period will be placed on the market, with no part of it held in storage. Such an assumption is appropriate when the commodity in question is perishable or when no inventory is ever kept. A model with inventory will be considered in Sec. 17.5.

By substituting the last two equations into the first, however, the model can be reduced to a single first-order difference equation as follows:

$$\beta P_t + \delta P_{t-1} = \alpha + \gamma$$

In order to solve this equation, it is desirable first to normalize it and shift the time subscripts ahead by one period [alter t to $(t+1)$, etc.]. The result,

$$P_{t+1} + \frac{\delta}{\beta} P_t = \frac{\alpha + \gamma}{\beta} \tag{17.11}$$

will then be a replica of (17.6), with the substitutions

$$y = P \qquad a = \frac{\delta}{\beta} \qquad \text{and} \qquad c = \frac{\alpha + \gamma}{\beta}$$

Inasmuch as δ and β are both positive, it follows that $a \neq -1$. Consequently, we can apply formula (17.8′), to get the time path

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} \tag{17.12}$$

where P_0 represents the initial price.

The Cobwebs

Three points may be observed in regard to this time path. In the first place, the expression $(\alpha + \gamma)/(\beta + \delta)$, which constitutes the particular integral of the difference equation, can be taken as the intertemporal equilibrium price of the model:[†]

$$\overline{P} = \frac{\alpha + \gamma}{\beta + \delta}$$

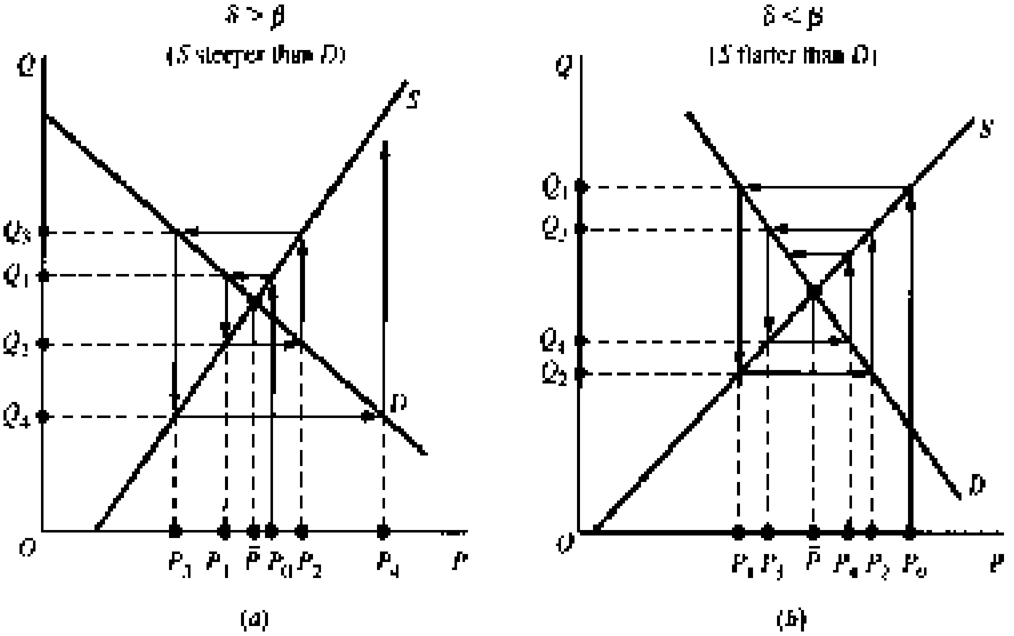
Because this is a constant, it is a stationary equilibrium. Substituting $\overline{P}$ into our solution, we can express the time path P_t alternatively in the form

$$P_t = (P_0 - \overline{P})\left(-\frac{\delta}{\beta}\right)^t + \overline{P} \tag{17.12'}$$

This leads us to the second point, namely, the significance of the expression $(P_0 - \overline{P})$. Since this corresponds to the constant A in the Ab^t term, its sign will bear on the question of whether the time path will commence above or below the equilibrium (mirror effect), whereas its magnitude will decide how far above or below (scale effect). Lastly, there is the expression $(-\delta/\beta)$, which corresponds to the b component of Ab^t. From our model specification that $\beta, \delta > 0$, we can deduce an oscillatory time path. It is this fact which gives rise to the cobweb phenomenon, as we shall presently see. There can, of course, arise *three*

[†] As far as the market-clearing sense of equilibrium is concerned, the price reached in each period is an equilibrium price, because we have assumed that $Q_{dt} = Q_{st}$ for every t.

FIGURE 17.2



possible varieties of oscillation patterns in the model. According to Table 17.1 or Fig. 17.1, the oscillation will be

$$\left.\begin{array}{l}\text{Explosive}\\ \text{Uniform}\\ \text{Damped}\end{array}\right\} \quad \text{if} \quad \delta \gtreqless \beta$$

where the term *uniform oscillation* refers to the type of path in region VI.

In order to visualize the cobwebs, let us depict the model (17.10) in Fig. 17.2. The second equation of (17.10) plots as a downward-sloping linear demand curve, with its slope numerically equal to β. Similarly, a linear supply curve with a slope equal to δ can be drawn from the third equation, if we let the Q axis represent in this instance a *lagged* quantity supplied. The case of $\delta > \beta$ (S steeper than D) and the case of $\delta < \beta$ (S flatter than D) are illustrated in Fig. 17.2*a* and *b*, respectively. In either case, however, the intersection of D and S will yield the intertemporal equilibrium price $\bar{P}$.

When $\delta > \beta$, as in Fig. 17.2*a*, the interaction of demand and supply will produce an explosive oscillation as follows. Given an initial price P_0 (here assumed above $\bar{P}$), we can follow the arrowhead and read off on the S curve that the quantity supplied in the next period (period 1) will be Q_1. In order to clear the market, the quantity demanded in period 1 must also be Q_1, which is possible if and only if price is set at the level of P_1 (see downward arrow). Now, via the S curve, the price P_1 will lead to Q_2 as the quantity supplied in period 2, and to clear the market in the latter period, price must be set at the level of P_2 according to the demand curve. Repeating this reasoning, we can trace out the prices and quantities in subsequent periods by simply following the arrowheads in the diagram, thereby spinning a "cobweb" around the demand and supply curves. By comparing the price levels, $P_0, P_1, P_2, \ldots$, we observe in this case not only an oscillatory pattern of change but also a tendency for price to widen its deviation from $\bar{P}$ as time goes by. With the cobweb being spun from inside out, the time path is divergent and the oscillation explosive.

By way of contrast, in the case of Fig. 17.2*b*, where $\delta < \beta$, the spinning process will create a cobweb which is centripetal. From P_0, if we follow the arrowheads, we shall be led ever closer to the intersection of the demand and supply curves, where $\overline{P}$ is. While still oscillatory, this price path is convergent.

In Fig. 17.2 we have not shown a third possibility, namely, that of $\delta = \beta$. The procedure of graphical analysis involved, however, is perfectly analogous to the other two cases. It is therefore left to you as an exercise.

The preceding discussion has dealt only with the time path of P (that is, P_t); after P_t is found, however, it takes but a short step to get to the time path of Q. The second equation of (17.10) relates Q_{dt} to P_t, so if (17.12) or (17.12') is substituted into the demand equation, the time path of Q_{dt} can be obtained immediately. Moreover, since Q_{dt} must be equal to Q_{st} in each time period (clearance of market), we can simply refer to the time path as Q_t rather than Q_{dt}. On the basis of Fig. 17.2, the rationale of this substitution is easily seen. Each point on the D curve relates a P_t to a Q_t pertaining to the same time period; therefore, the demand function can serve to map the time path of price into the time path of quantity.

You should note that the graphical technique of Fig. 17.2 is applicable even when the D and S curves are nonlinear.

EXERCISE 17.4

1. On the basis of (17.10), find the time path of Q, and analyze the condition for its convergence.
2. Draw a diagram similar to those of Fig. 17.2 to show that, for the case of $\delta = \beta$, the price will oscillate uniformly with neither damping nor explosion.
3. Given demand and supply for the cobweb model as follows, find the intertemporal equilibrium price, and determine whether the equilibrium is stable:
   (*a*) $Q_{dt} = 18 - 3P_t \qquad Q_{st} = -3 + 4P_{t-1}$
   (*b*) $Q_{dt} = 22 - 3P_t \qquad Q_{st} = -2 + P_{t-1}$
   (*c*) $Q_{dt} = 19 - 6P_t \qquad Q_{st} = 6P_{t-1} - 5$
4. In model (17.10), let the $Q_{dt} = Q_{st}$ condition and the demand function remain as they are, but change the supply function to
   $Q_{st} = -\gamma - \delta P_t^*$
   where P_t^* denotes the *expected price* for period t. Furthermore, suppose that sellers have the "adaptive" type of price expectation:[*]
   $P_t^* = P_{t-1}^* + \eta(P_{t-1} - P_{t-1}^*) \qquad (0 < \eta \le 1)$
   where η (the Greek letter eta) is an expectation-adjustment coefficient.
   (*a*) Give an economic interpretation to the preceding equation. In what respects is it similar to, and different from, the adaptive expectations equation (16.34)?
   (*b*) What happens if η takes its maximum value? Can we consider the cobweb model as a special case of the present model?

[*] See Marc Nerlove, "Adaptive Expectations and Cobweb Phenomena," *Quarterly Journal of Economics*, May 1958, pp. 227–240.

(c) Show that the new model can be represented by the first-order difference equation

$$P_{t+1} - \left(1 - \eta - \frac{\eta\delta}{\beta}\right) P_t = \frac{\eta(\alpha + \gamma)}{\beta}$$

(*Hint:* Solve the supply function for P_t^*, and then use the information that $Q_{st} = Q_{dt} = \alpha - \beta P_t$.)

(d) Find the time path of price. Is this path necessarily oscillatory? Can it be oscillatory? Under what circumstances?

(e) Show that the time path P_t, if oscillatory, will converge only if $1 - 2/\eta < -\delta/\beta$. As compared with the cobweb solution (17.12) or (17.12'), does the new model have a wider or narrower range for the stability-inducing values of $-\delta/\beta$?

5. The cobweb model, like the previously encountered dynamic market models, is essentially based on the static market model presented in Sec. 3.2. What economic assumption is the dynamizing agent in the present case? Explain.

17.5 A Market Model with Inventory

In the preceding model, price is assumed to be set in such a way as to clear the current output of every time period. The implication of that assumption is either that the commodity is a perishable which cannot be stocked or that, though it is stockable, no inventory is ever kept. Now we shall construct a model in which sellers do keep an inventory of the commodity.

The Model

Let us assume the following:

1. Both the quantity demanded, Q_{dt}, and the quantity currently produced, Q_{st}, are unlagged linear functions of price P_t.
2. The adjustment of price is effected not through market clearance in every period, but through a process of price-setting by the sellers: At the beginning of each period, the sellers set a price for that period after taking into consideration the inventory situation. If, as a result of the preceding-period price, inventory accumulated, the current-period price is set at a lower level than before, in order to "move" the merchandise; but if inventory decumulated instead, the current price is set higher than before.
3. The price adjustment made from period to period is inversely proportional to the observed change in the inventory (stock).

With these assumptions, we can write the following equations:

$$\begin{aligned} Q_{dt} &= \alpha - \beta P_t & (\alpha, \beta > 0) \\ Q_{st} &= -\gamma + \delta P_t & (\gamma, \delta > 0) \\ P_{t+1} &= P_t - \sigma(Q_{st} - Q_{dt}) & (\sigma > 0) \end{aligned} \qquad (17.13)$$

where σ denotes the *stock-induced-price-adjustment* coefficient. Note that (17.13) is really nothing but the discrete-time counterpart of the market model of Sec. 15.2, although we have now couched the price-adjustment process in terms of *inventory* $(Q_{st} - Q_{dt})$ rather than *excess demand* $(Q_{dt} - Q_{st})$. Nevertheless, the analytical results will turn out to be much different; for one thing, with discrete time, we may encounter the phenomenon of oscillations. Let us derive and analyze the time path P_t.

The Time Path

By substituting the first two equations into the third, the model can be condensed into a single difference equation:

$$P_{t+1} - [1 - \sigma(\beta + \delta)]P_t = \sigma(\alpha + \gamma) \qquad (17.14)$$

and its solution is given by (17.8'):

$$\begin{aligned} P_t &= \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)[1 - \sigma(\beta + \delta)]^t + \frac{\alpha + \gamma}{\beta + \delta} \\ &= (P_0 - \overline{P})[1 - \sigma(\beta + \delta)]^t + \overline{P} \end{aligned} \qquad (17.15)$$

Obviously, therefore, the dynamic stability of the model will hinge on the expression $1 - \sigma(\beta + \delta)$; for convenience, let us refer to this expression as b.

With reference to Table 17.1, we see that, in analyzing the exponential expression b^t, seven distinct regions of b values may be defined. However, since our model specifications $(\sigma, \beta, \delta > 0)$ have effectually ruled out the first two regions, there remain only five possible cases, as listed in Table 17.2. For each of these regions, the b specification of the second column can be translated into an equivalent σ specification, as shown in the third column. For instance, for region III, the b specification is $0 < b < 1$; therefore, we can write

$$0 < 1 - \sigma(\beta + \delta) < 1$$

$$-1 < -\sigma(\beta + \delta) < 0 \qquad \text{[subtracting 1 from all three parts]}$$

$$\text{and} \qquad \frac{1}{\beta + \delta} > \sigma > 0 \qquad \text{[dividing through by } -(\beta + \delta)\text{]}$$

TABLE 17.2
Types of Time Path

Region	Value of $b \equiv 1 - \sigma(\beta + \delta)$	Value of σ	Nature of Time Path P_t
III	$0 < b < 1$	$0 < \sigma < \frac{1}{\beta + \delta}$	Nonoscillatory and convergent
IV	$b = 0$	$\sigma = \frac{1}{\beta + \delta}$	Remaining in equilibrium[†]
V	$-1 < b < 0$	$\frac{1}{\beta + \delta} < \sigma < \frac{2}{\beta + \delta}$	With damped oscillation
VI	$b = -1$	$\sigma = \frac{2}{\beta + \delta}$	With uniform oscillation
VII	$b < -1$	$\sigma > \frac{2}{\beta + \delta}$	With explosive oscillation

[†] The fact that price will be remaining in equilibrium in this case can also be seen directly from (17.14). With $\sigma = 1/(\beta + \delta)$, the coefficient of P_t becomes zero, and (17.14) reduces to $P_{t+1} = \sigma(\alpha + \gamma) = (\alpha + \gamma)/(\beta + \delta) = \overline{P}$.

This last gives us the desired equivalent σ specification for region III. The translation for the other regions may be carried out analogously. Since the type of time path pertaining to each region is already known from Fig. 17.1, the σ specification enables us to tell from given values of σ, β, and δ the general nature of the time path P_t, as outlined in the last column of Table 17.2.

Example 1

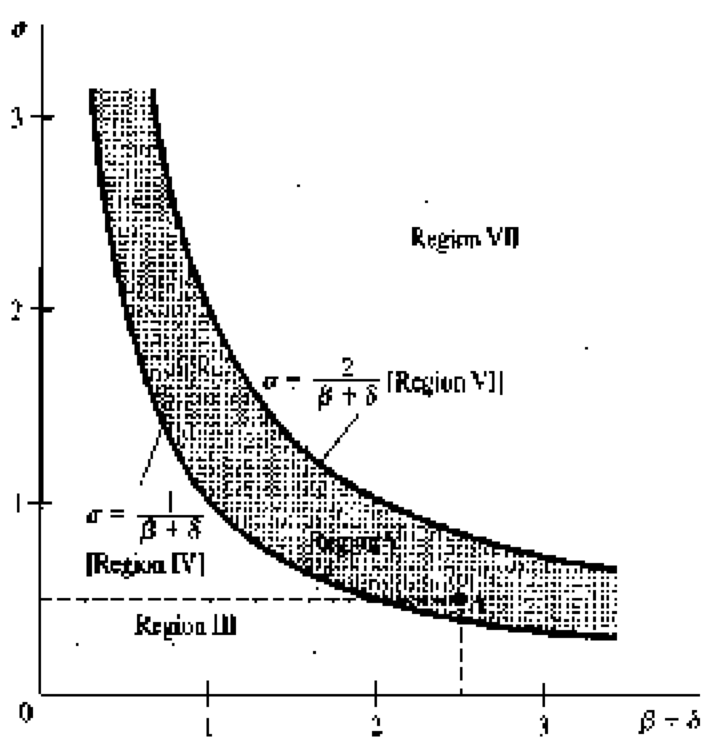
If the sellers in our model always increase (decrease) the price by 10 percent of the amount of the decrease (increase) in inventory, and if the demand curve has a slope of −1 and the supply curve a slope of 15 (both slopes with respect to the price axis), what type of time path P_t will we find?

Here, we have $\sigma = 0.1$, $\beta = 1$, and $\delta = 15$. Since $1/(\beta + \delta) = \frac{1}{16}$ and $2/(\beta + \delta) = \frac{1}{8}$, the value of σ $(= \frac{1}{10})$ lies between the former two values; it is thus a case of region V. The time path P_t will be characterized by damped oscillation.

Graphical Summary of the Results

The substance of Table 17.2, which contains as many as five different possible cases of σ specification, can be made much easier to grasp if the results are presented graphically. Inasmuch as the σ specification involves essentially a comparison of the relative magnitudes of the parameters σ and $(\beta + \delta)$, let us plot σ against $(\beta + \delta)$, as in Fig. 17.3. Note that we need only concern ourselves with the positive quadrant because, by model specification, σ and $(\beta + \delta)$ are both positive. From Table 17.2, it is clear that regions IV and VI are specified by the equations $\sigma = 1/(\beta + \delta)$ and $\sigma = 2/(\beta + \delta)$, respectively. Since each of these plots as a rectangular hyperbola, the two regions are graphically represented by the two hyperbolic curves in Fig. 17.3. Once we have the two hyperbolas, moreover, the other three regions immediately fall into place. Region III, for instance, is merely the set of points lying below the lower hyperbola, where we have σ less than $1/(\beta + \delta)$. Similarly, region V is represented by the set of points falling between the two hyperbolas, whereas all the points located above the higher hyperbola pertain to region VII.

FIGURE 17.3



Example 2 If $\alpha = \frac{1}{2}$, $\beta = 1$, and $\delta = \frac{3}{2}$, will our model (17.13) yield a convergent time path P_t? The given parametric values correspond to point A in Fig. 17.3. Since it falls within region V, the time path is convergent, though oscillatory.

You will note that, in the two models just presented, our analytical results are in each instance stated as a set of alternative possible cases—three types of oscillatory path for the cobwebs, and five types of time path in the inventory model. This richness of analytical results stems, of course, from the parametric formulation of the models. The fact that our result cannot be stated in a single unequivocal answer is, of course, a merit rather than a weakness.

EXERCISE 17.5

1. In solving (17.14), why should formula (17.8') be used instead of (17.9')?
2. On the basis of Table 17.2, check the validity of the translation from the b specification to the α specification for regions IV through VII.
3. If model (17.13) has the following numerical form:

   $$Q_{dt} = 21 - 2P_t$$
   $$Q_{st} = -3 + 6P_t$$
   $$P_{t+1} = P_t - 0.3(Q_{st} - Q_{dt})$$

   find the time path P_t and determine whether it is convergent.
4. Suppose that, in model (17.13), the supply in each period is a fixed quantity, say, $Q_{st} = k$, instead of a function of price. Analyze the behavior of price over time. What restriction should be imposed on k to make the solution economically meaningful?

17.6 Nonlinear Difference Equations—The Qualitative-Graphic Approach

Thus far we have only utilized *linear* difference equations in our models; but the facts of economic life may not always acquiesce to the convenience of linearity. Fortunately, when nonlinearity occurs in the case of first-order difference-equation models, there exists an easy method of analysis that is applicable under fairly general conditions. This method, graphic in nature, closely resembles that of the qualitative analysis of first-order differential equations presented in Sec. 15.6.

Phase Diagram

Nonlinear difference equations in which only the variables y_{t+1} and y_t appear, such as

$$y_{t+1} + y_t^3 = 5 \qquad \text{or} \qquad y_{t+1} + \sin y_t - \ln y_t = 3$$

can be categorically represented by the equation

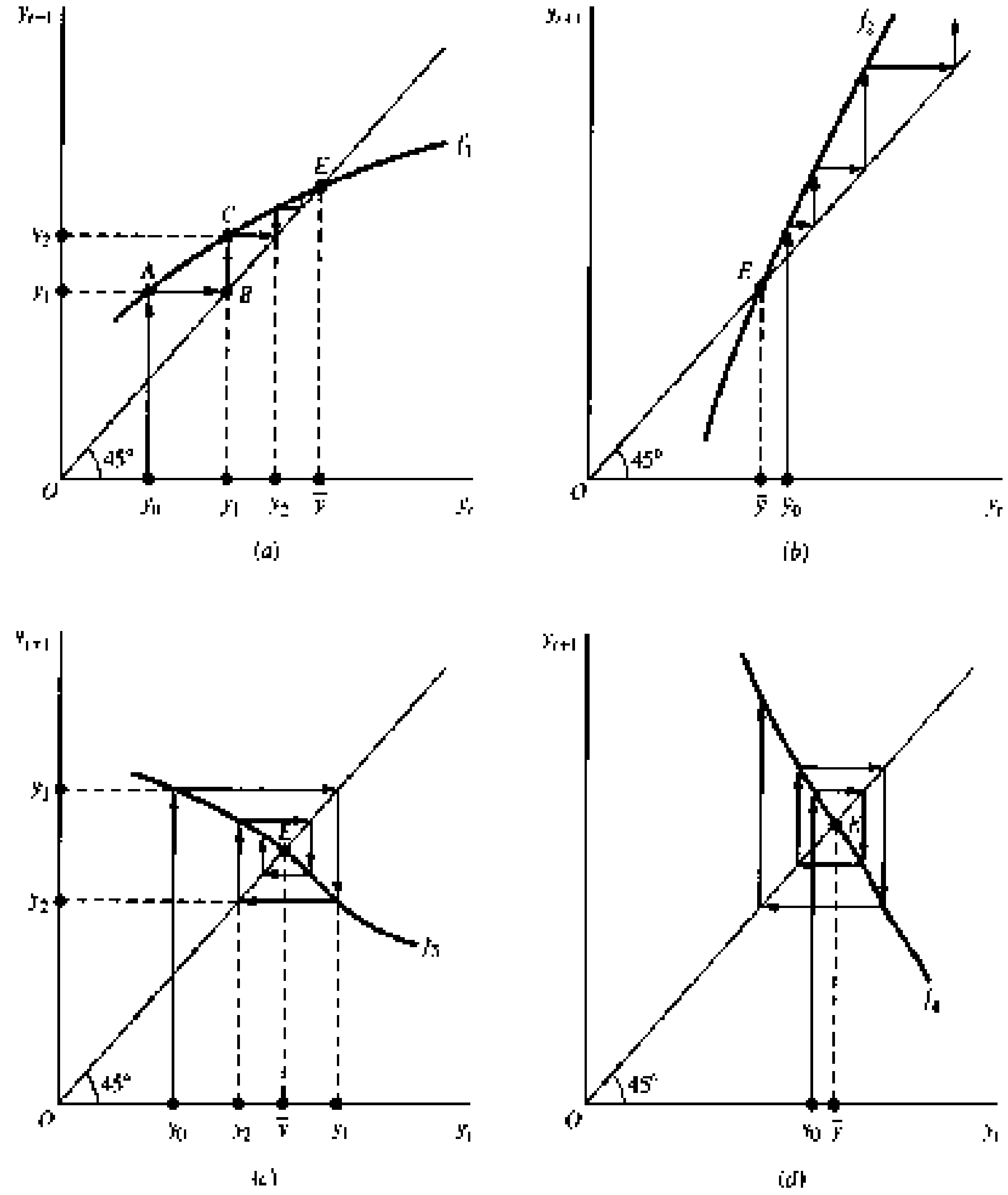
$$y_{t+1} = f(y_t) \tag{17.16}$$

where f can be a function of any degree of complexity, as long as it is a function of y_t alone without t as another argument. When the two variables y_{t+1} and y_t are plotted against each other in a Cartesian coordinate plane, the resulting diagram constitutes a *phase diagram*, and the curve corresponding to f is a *phase line*. From these, it is possible to analyze the time path of the variable by the process of iteration.

The terms phase diagram and phase line are used here in analogy to the differential-equation case; but note one dissimilarity in the construction of the diagram. In the differential-equation case, we plotted dy/dt against y as in Fig. 15.3, so that, in order to be perfectly analogous in the present case, we should have Δy_t on the vertical axis and y_t on the horizontal. This is not impossible to do, but it is much more convenient to place y_{t+1} on the vertical axis instead, as we have done in Fig. 17.4 where the same scale is used on both axes. Note the presence of a 45° line in each diagram of Fig. 17.4; this line will prove to be of great service in carrying out our graphic analysis.

Let us illustrate the procedure involved by means of Fig. 17.4*a*, where we have drawn a phase line (labeled f_1) representing a specific difference equation $y_{t+1} = f_1(y_t)$. If we are

FIGURE 17.4



given an initial value y_0 (plotted on the horizontal axis), by iteration we can trace out all the subsequent values of y as follows. First, since the phase line f_1 maps the initial value y_0 into y_1 according to the equation

$$y_1 = f_1(y_0)$$

we can go straight up from y_0 to the phase line, hit point A, and read its height on the vertical axis as the value of y_1. Next, we seek to map y_1 into y_2 according to the equation

$$y_2 = f_1(y_1)$$

For this purpose, we must first plot y_1 on the horizontal axis—similarly to y_0 during the first mapping. This required transplotting of y_1 from the vertical axis to the horizontal is most easily accomplished by the use of the 45° line, which, having a slope of +1, is the locus of points with identical abscissa and ordinate, such as (2, 2) and (5, 5). Thus, to transplot y_1 from the vertical axis, we can simply go across to the 45° line, hit point B, and then turn straight down to the horizontal axis to locate the point y_1. By repeating this process, we can map y_1 to y_2 via point C on the phase line, and then use the 45° line for transplotting y_2, etc.

Now that the nature of the iteration is clear, we may observe that the desired iteration can be achieved simply by following the arrowheads from y_0 to A (on the phase line), to B (on the 45° line), to C (on the phase line), etc.—always alternating between the two lines—without it ever being necessary to resort to the axes again.

Types of Time Path

The graphic iterations just outlined are, of course, equally applicable to the other three diagrams in Fig. 17.4. Actually, these four diagrams serve to illustrate four basic varieties of phase lines, each implying a different type of time path. The first two phase lines, f_1 and f_2, are characterized by positive slopes, with one slope being less than unity and the other one greater than unity:

$$0 < f_1'(y_t) < 1 \qquad \text{and} \qquad f_2'(y_t) > 1$$

The remaining two, on the other hand, are negatively sloped; specifically, we have

$$-1 < f_3'(y_t) < 0 \qquad \text{and} \qquad f_4'(y_t) < -1$$

In each diagram of Fig. 17.4, the intertemporal equilibrium value of y (namely $\overline{y}$) is located at the intersection of the phase line and the 45° line, which we have labeled E. This is so because the point E on the phase line, being simultaneously a point on the 45° line, will map a y_t into a y_{t+1} of identical value; and when $y_{t+1} = y_t$, by definition y must be in equilibrium intertemporally. Our principal task is to determine whether, given an initial value $y_0 \neq \overline{y}$, the pattern of change implied by the phase line will lead us consistently toward $\overline{y}$ (convergent) or away from it (divergent).

For the phase line f_1, the iterative process leads from y_0 to $\overline{y}$ in a steady path, without oscillation. You can verify that, if y_0 is placed to the right of $\overline{y}$, there will also be a steady movement toward $\overline{y}$, although it will be in the leftward direction. These time paths are convergent to equilibrium, and their general configurations would be of the same type as shown in region III of Fig. 17.1.

Given the phase line f_2, whose slope exceeds unity, however, a divergent time path emerges. From an initial value y_0 greater than $\bar{y}$, the arrowheads lead steadily away from the equilibrium to higher and higher y values. As you can verify, an initial value lower than $\bar{y}$ gives rise to a similar steady divergent movement, though in the opposite direction.

When the phase line is negatively inclined, as in f_3 and f_4, the steady movement gives way to oscillation, and there appears now the phenomenon of *overshooting* the equilibrium mark. In diagram c, y_0 leads to y_1, which exceeds $\bar{y}$, only to be followed by y_2, which falls short of $\bar{y}$, etc. The convergence of the time path will, in such cases, depend on the slope of the phase line being less than 1 in its absolute value. This is the case of the phase line f_3, where the extent of overshooting tends to diminish in successive periods. For the phase line f_4, whose slope exceeds 1 numerically, on the other hand, the opposite tendency prevails, resulting in a divergent time path.

The oscillatory time paths generated by phase lines f_3 and f_4 are reminiscent of the cobwebs in Fig. 17.2. In Fig. 17.4c or d, however, the cobweb is spun around a phase line (which contains a lag) and the 45° line, instead of around a demand curve and a (lagged) supply curve. Here, a 45° line is used as a mechanical aid for transplotting a value of y, whereas in Fig. 17.2, the D curve (which plays a role similar to that of the 45° line in Fig. 17.4) is an integral part of the model itself. Specifically, once Q_{st} is determined on the supply curve, we let the arrowheads hit the D curve for the purpose of finding a price that will "clear the market," as was the rule of the game in the cobweb model. Consequently, there is a basic difference in the labeling of the axes: in Fig. 17.2 there are two entirely different variables, P and Q, but in Fig. 17.4 the axes represent the values of the same variable y in two consecutive periods. Note however, that if we analyze the graph of the difference equation (17.11) which summarizes the cobweb model, rather than the separate demand and supply functions in (17.10), then the resulting diagram will be a phase line such as shown in Fig. 17.4. In other words, there really exist two alternative ways of graphically analyzing the cobweb model, which will yield the identical result.

The basic rule emerging from the preceding consideration of the phase line is that the *algebraic sign* of its slope determines whether there will be *oscillation*, and the *absolute value* of its slope governs the question of *convergence*. If the phase line happens to contain both positively and negatively sloped segments, and if the absolute value of its slope is at some points greater, and elsewhere less, than 1, the time path will naturally become more complicated. However, even in such cases, the graphic-iterative analysis can be employed with equal ease. Of course, an initial value must be given to us before the iteration can be duly started. Indeed, in these more complicated cases, a different initial value can lead to a time path of an altogether different breed (see Exercises 17.6-2 and 17.6-3).

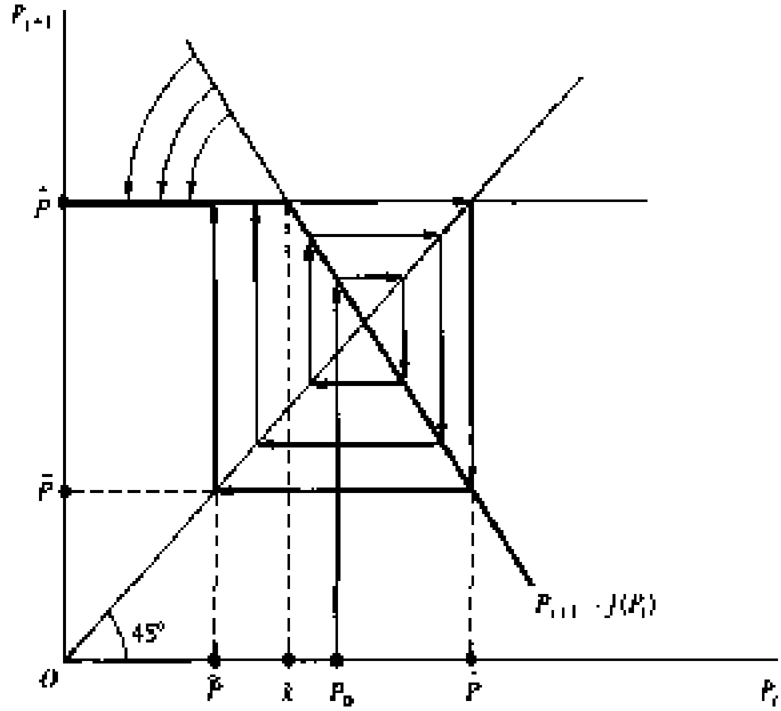
A Market with a Price Ceiling

We shall now cite an economic example of a nonlinear difference equation. In Fig. 17.4, the four nonlinear phase lines all happen to be of the smooth variety; in the present example, we shall show a nonsmooth phase line.

As a point of departure, let us take the *linear* difference equation (17.11) of the cobweb model and rewrite it as

$$P_{t+1} = \frac{\alpha + \gamma}{\beta} - \frac{\delta}{\beta} P_t \qquad \left(\frac{\delta}{\beta} > 0\right) \tag{17.17}$$

FIGURE 17.5



This is in the format of $P_{t+1} = f(P_t)$, with $f'(P_t) = -\delta/\beta < 0$. We have plotted this linear phase line in Fig. 17.5 on the assumption that the slope is greater than 1 in absolute value, implying *explosive* oscillation.

Now let there be imposed a legal price ceiling $\hat{P}$ (read: "P caret" or, less formally, "P hat"). This can be shown in Fig. 17.5 as a horizontal straight line because, irrespective of the level of P_t, P_{t+1} is now forbidden to exceed the level of $\hat{P}$. What this does is to invalidate that part of the phase line lying above $\hat{P}$ or, to view it differently, to bend down the upper part of the phase line to the level of $\hat{P}$, thus resulting in a kinked phase line.[†] In view of the kink, the new (heavy) phase line is not only nonlinear but nonsmooth as well. Like a step function, this kinked line will require more than one equation to express it algebraically:

$$P_{t+1} = \begin{cases} \hat{P} & (\text{for } P_t \leq k) \\ \dfrac{\alpha+\gamma}{\beta} - \dfrac{\delta}{\beta} P_t & (\text{for } P_t > k) \end{cases} \tag{17.17'}$$

where k denotes the value of P_t at the kink.

Assuming an initial price P_0, let us trace out the time path of price iteratively. During the first stage of iteration, when the downward-sloping segment of the phase line is in effect, the explosive oscillatory tendency clearly manifests itself. After a few periods, however, the arrowheads begin to hit the ceiling price, and thereafter the time path will develop into a perpetual cyclical movement between $\hat{P}$ and an effective *price floor* $\tilde{P}$ (read: "P tilde" or, less formally, "P wiggle"). Thus, by virtue of the price ceiling, the intrinsic explosive tendency of the model is effectively contained, and the ever-widening oscillation is now tamed into a uniform oscillation producing a so-called limit cycle.

[†] Strictly speaking, we should also "bend" that part of the phase line lying to the right of the point $\bar{P}$ on the horizontal axis. But it does no harm to leave it as it is, as long as the other end has already been bent, because the transplotting of P_{t+1} to the horizontal axis will carry the upper limit of $\hat{P}$ over to the P_t axis automatically.

What is significant about this result is that, whereas in the case of a linear phase line a uniformly oscillatory path can be produced if and only if the slope of the phase line is -1, now after the introduction of nonlinearity the same analytical result can arise even when the phase line has a slope other than -1. The economic implication of this is of considerable import. If one observes a more or less uniform oscillation in the actual time path of a variable and attempts to explain it by means of a *linear* model, one will be forced to rely on the rather special—and implausible—model specification that the phase-line slope is exactly -1. But if nonlinearity is introduced, in either the smooth or the nonsmooth variety, then a host of more reasonable assumptions can be used, each of which can equally account for the observed feature of uniform oscillation.

EXERCISE 17.6

1. In difference-equation models, the variable t can only take integer values. Does this imply that in the phase diagrams of Fig. 17.4 the variables y_t and y_{t-1} must be considered as discrete variables?
2. As a phase line, use the left half of an inverse U-shaped curve, and let it intersect the 45° line at two points L (left) and R (right).
   (*a*) Is this a case of multiple equilibria?
   (*b*) If the initial value y_0 lies to the left of L, what kind of time path will be obtained?
   (*c*) What if the initial value lies between L and R?
   (*d*) What if the initial value lies to the right of R?
   (*e*) What can you conclude about the dynamic stability of equilibrium at L and at R, respectively?
3. As a phase line, use an inverse U-shaped curve. Let its upward-sloping segment intersect the 45° line at point L, and let its downward-sloping segment intersect the 45° line at point R. Answer the same five questions raised in the Prob. 2. (*Note:* Your answer will depend on the particular way the phase line is drawn; explore various possibilities.)
4. In Fig. 17.5, rescind the legal price ceiling and impose a minimum price P_m instead.
   (*a*) How will the phase line change?
   (*b*) Will it be kinked? Nonlinear?
   (*c*) Will there also develop a uniformly oscillatory movement in price?
5. With reference to (17.17′) and Fig. 17.5, show that the constant k can be expressed as

$$k = \frac{\alpha + \gamma}{\delta} - \frac{\beta}{\delta}\hat{P}$$

Chapter 18

Higher-Order Difference Equations

The economic models in Chap. 17 involve difference equations that relate P_t and P_{t-1} to each other. As the P value in one period can uniquely determine the P value in the next, the time path of P becomes fully determinate once an initial value P_0 is specified. It may happen, however, that the value of an economic variable in period t (say, y_t) depends not only on y_{t-1} but also on y_{t-2}. Such a situation will give rise to a difference equation of the second order.

Strictly speaking, a *second-order difference equation* is one that involves an expression $\Delta^2 y_t$, called the *second difference* of y_t, but contains no differences of order higher than 2. The symbol Δ^2, the discrete-time counterpart of the symbol d^2/dt^2, is an instruction to "take the second difference" as follows:

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_{t+1} - y_t) && \text{[by (17.1)]}\\ &= (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t) && \text{[again by (17.1)]}^{1}\\ &= y_{t+2} - 2y_{t+1} + y_t\end{aligned}$$

Thus a second difference of y_t is transformable into a sum of terms involving a two-period time lag. Since expressions like $\Delta^2 y_t$ and Δy_t are quite cumbersome to work with, we shall simply redefine a second-order difference equation as one involving a two-period time lag in the variable. Similarly, a third-order difference equation is one that involves a three-period time lag, etc.

Let us first concentrate on the method of solving a second-order difference equation, leaving the generalization to higher-order equations in Section 18.4. To keep the scope of discussion manageable, we shall only deal with linear difference equations with constant coefficients in the present chapter. However, both the constant-term and variable-term varieties will be examined.

[1] That is, we first move the subscripts in the $(y_{t+1} - y_t)$ expression forward by one period, to get a new expression $(y_{t+2} - y_{t+1})$, and then we subtract from the latter the original expression. Note that, since the resulting difference may be written as $\Delta y_{t+1} - \Delta y_t$, we may infer the following rule of operation:

$$\Delta(y_{t+1} - y_t) = \Delta y_{t+1} - \Delta y_t$$

This is reminiscent of the rule applicable to the derivative of a sum or difference.

18.1 Second-Order Linear Difference Equations with Constant Coefficients and Constant Term

A simple variety of second-order difference equations takes the form

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \tag{18.1}$$

You will recognize this equation to be linear, nonhomogeneous, and with constant coefficients (a_1, a_2) and constant term c.

Particular Solution

As before, the solution of (18.1) may be expected to have two components: a particular solution y_p representing the intertemporal equilibrium level of y, and a complementary function y_c specifying, for every time period, the deviation from the equilibrium. The particular solution, defined as any solution of the complete equation, can sometimes be found simply by trying a solution of the form $y_t = k$. Substituting this constant value of y into (18.1), we obtain

$$k + a_1 k + a_2 k = c \qquad \text{and} \qquad k = \frac{c}{1 + a_1 + a_2}$$

Thus, so long as $(1 + a_1 + a_2) \neq 0$, the particular integral is

$$y_p (= k) = \frac{c}{1 + a_1 + a_2} \qquad (\text{case of } a_1 + a_2 \neq -1) \tag{18.2}$$

Example 1 Find the particular integral of $y_{t+2} - 3y_{t+1} + 4y_t = 6$. Here we have $a_1 = -3$, $a_2 = 4$, and $c = 6$. Since $a_1 + a_2 \neq -1$, the particular solution can be obtained from (18.2) as follows:

$$y_p = \frac{6}{1 - 3 + 4} = 3$$

In case $a_1 + a_2 = -1$, then the trial solution $y_t = k$ breaks down, and we can try $y_t = kt$ instead. Substituting the latter into (18.1) and bearing in mind that we now have $y_{t+1} = k(t+1)$ and $y_{t+2} = k(t+2)$, we find that

$$k(t+2) + a_1 k(t+1) + a_2 kt = c$$

$$\text{and} \qquad k = \frac{c}{(1 + a_1 + a_2)t + a_1 + 2} = \frac{c}{a_1 + 2} \qquad [\text{since } a_1 + a_2 = -1]$$

Thus we can write the particular solution as

$$y_p (= kt) = \frac{c}{a_1 + 2} t \qquad (\text{case of } a_1 + a_2 = -1;\ a_1 \neq -2) \tag{18.2'}$$

Example 2 Find the particular solution of $y_{t+2} + y_{t+1} - 2y_t = 12$. Here, $a_1 = 1$, $a_2 = -2$, and $c = 12$. Obviously, formula (18.2) is not applicable, but (18.2') is. Thus,

$$y_p = \frac{12}{1 + 2} t = 4t$$

This particular solution represents a moving equilibrium.

If $a_1 + a_2 = -1$, but at the same time $a_1 = -2$ (that is, if $a_1 = -2$ and $a_2 = 1$), then we can adopt a trial solution of the form $y_t = kt^2$, which implies $y_{t-1} = k(t+1)^2$, etc. As you may verify, in this case the particular solution turns out to be

$$y_p = kt^2 = \frac{c}{2}t^2 \qquad \text{(case of } a_1 = -2;\ a_2 = 1) \tag{18.2''}$$

However, since this formula applies only to the unique case of the difference equation $y_{t+2} - 2y_{t+1} + y_t = c$, its usefulness is rather limited.

Complementary Function

To find the complementary function, we must concentrate on the reduced equation

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = 0 \tag{18.3}$$

Our experience with first-order difference equations has taught us that the expression Ab^t plays a prominent role in the general solution of such an equation. Let us therefore try a solution of the form $y_t = Ab^t$, which naturally implies that $y_{t+1} = Ab^{t+1}$, and so on. It is our task now to determine the values of A and b.

Upon substitution of the trial solution into (18.3), the equation becomes

$$Ab^{t+2} + a_1 Ab^{t+1} + a_2 Ab^t = 0$$

or, after canceling the (nonzero) common factor Ab^t,

$$b^2 + a_1 b + a_2 = 0 \tag{18.3'}$$

This quadratic equation—the *characteristic equation* of (18.3) or of (18.1)—which is comparable to (16.4''), possesses the two *characteristic roots*

$$b_1, b_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2} \tag{18.4}$$

each of which is acceptable in the solution Ab^t. In fact, *both* b_1 *and* b_2 should appear in the general solution of the homogeneous difference equation (18.3) because, just as in the case of differential equations, this general solution must consist of two *linearly independent* parts, each with its own multiplicative arbitrary constant.

Three possible situations may be encountered in regard to the characteristic roots, depending on the square-root expression in (18.4). You will find these parallel very closely the analysis of second-order differential equations in Sec. 16.1.

Case 1 (distinct real roots) When $a_1^2 > 4a_2$, the square root in (18.4) is a real number, and b_1 and b_2 are real and distinct. In that event, b_1^t and b_2^t are linearly independent, and the complementary function can simply be written as a linear combination of these expressions; that is,

$$y_c = A_1 b_1^t + A_2 b_2^t \tag{18.5}$$

You should compare this with (16.7).

Example 3

Find the solution of $y_{t+2} + y_{t+1} - 2y_t = 12$. This equation has the coefficients $a_1 = 1$ and $a_2 = -2$; from (18.4), the characteristic roots can be found to be $b_1, b_2 = 1, -2$. Thus, the complementary function is

$$y_c = A_1(1)^t + A_2(-2)^t = A_1 + A_2(-2)^t$$

Since, in Example 2, the particular solution of the given difference equation has already been found to be $y_p = 4t$, we can write the general solution as

$$y_t = y_c + y_p = A_1 + A_2(-2)^t + 4t$$

There are still two arbitrary constants A_1 and A_2 to be definitized; to accomplish this, two initial conditions are necessary. Suppose that we are given $y_0 = 4$ and $y_1 = 5$. Then, since by letting $t = 0$ and $t = 1$ successively in the general solution we find

$$y_0 = A_1 + A_2 \qquad (= 4 \text{ by the first initial condition})$$

$$y_1 = A_1 - 2A_2 + 4 \qquad (= 5 \text{ by the second initial condition})$$

the arbitrary constants can be definitized to $A_1 = 3$ and $A_2 = 1$. The definite solution then can finally be written as

$$y_t = 3 + (-2)^t + 4t$$

Case 2 (repeated real roots) When $a_1^2 = 4a_2$, the square root in (18.4) vanishes, and the characteristic roots are repeated:

$$b(= b_1 = b_2) = -\frac{a_1}{2}$$

Now, if we express the complementary function in the form of (18.5), the two components will collapse into a single term:

$$A_1 b_1^t + A_2 b_2^t = (A_1 + A_2)b^t \equiv A_3 b^t$$

This will not do, because we are now short of one constant.

To supply the missing component—which, we recall, should be linearly independent of the term $A_3 b^t$—the old trick of multiplying b^t by the variable t will again work. The new component term is therefore to take the form $A_4 t b^t$. That this is linearly independent of $A_3 b^t$ should be obvious, for we can never obtain the expression $A_4 t b^t$ by attaching a constant coefficient to $A_3 b^t$. That $A_4 t b^t$ does indeed qualify as a solution of the homogeneous equation (18.3), just as $A_3 b^t$ does, can easily be verified by substituting $y_t = A_4 t b^t$ [and $y_{t+1} = A_4(t+1)b^{t+1}$, etc.] into (18.3)[*] and seeing that the latter will reduce to an identity $0 = 0$.

The complementary function for the repeated-root case is therefore

$$y_c = A_3 b^t + A_4 t b^t \tag{18.6}$$

which you should compare with (16.9).

Example 4 Find the complementary function of $y_{t+2} + 6y_{t+1} + 9y_t = 4$. The coefficients being $a_1 = 6$ and $a_2 = 9$, the characteristic roots are found to be $b_1 = b_2 = -3$. We therefore have

$$y_c = A_3(-3)^t + A_4 t(-3)^t$$

If we proceed a step further, we can easily find $y_p = \frac{1}{4}$, so the general solution of the given difference equation is

$$y_t = A_3(-3)^t + A_4 t(-3)^t + \tfrac{1}{4}$$

Given two initial conditions, A_3 and A_4 can again be assigned definite values.

[*] In this substitution it should be kept in mind that we have in the present case $a_1^2 = 4a_2$ and $b = -a_1/2$.

Case 3 (complex roots) Under the remaining possibility of $a_1^2 < 4a_2$, the characteristic roots are conjugate complex. Specifically, they will be in the form

$$b_1, b_2 = h \pm vi$$

where

$$h = -\frac{a_1}{2} \qquad \text{and} \qquad v = \frac{\sqrt{4a_2 - a_1^2}}{2} \tag{18.7}$$

The complementary function itself thus becomes

$$y_c = A_1 b_1^t + A_2 b_2^t = A_1(h + vi)^t + A_2(h - vi)^t$$

As it stands, y_c is not easily interpreted. But fortunately, thanks to De Moivre's theorem, given in (16.23'), this complementary function can easily be transformed into trigonometric terms, which we have learned to interpret.

According to the said theorem, we can write

$$(h \pm vi)^t = R^t(\cos\theta t \pm i \sin\theta t)$$

where the value of R (always taken to be positive) is, by (16.10),

$$R = \sqrt{h^2 + v^2} = \sqrt{\frac{a_1^2 + 4a_2 - a_1^2}{4}} = \sqrt{a_2} \tag{18.8}$$

and θ is the radian measure of the angle in the interval $[0, 2\pi)$, which satisfies the conditions

$$\cos\theta = \frac{h}{R} = \frac{-a_1}{2\sqrt{a_2}} \qquad \text{and} \qquad \sin\theta = \frac{v}{R} = \sqrt{1 - \frac{a_1^2}{4a_2}} \tag{18.9}$$

Therefore, the complementary function can be transformed as follows:

$$\begin{aligned} y_c &= A_1 R^t(\cos\theta t + i \sin\theta t) + A_2 R^t(\cos\theta t - i \sin\theta t) \\ &= R^t[(A_1 + A_2)\cos\theta t + (A_1 - A_2)i \sin\theta t] \\ &= R^t(A_5 \cos\theta t + A_6 \sin\theta t) \end{aligned} \tag{18.10}$$

where we have adopted the shorthand symbols

$$A_5 \equiv A_1 + A_2 \qquad \text{and} \qquad A_6 \equiv (A_1 - A_2)i$$

The complementary function (18.10) differs from its differential-equation counterpart (16.24') in two important respects. First, the expressions $\cos\theta t$ and $\sin\theta t$ have replaced the previously used $\cos vt$ and $\sin vt$. Second, the multiplicative factor R^t (an exponential with base R) has replaced the natural exponential expression e^{ht}. In short, we have switched from the Cartesian coordinates (h and v) of the complex roots to their polar coordinates (R and θ). The values of R and θ can be determined from (18.8) and (18.9) once h and v become known. It is also possible to calculate R and θ directly from the parameter values a_1 and a_2 via (18.8) and (18.9), provided we first make certain that $a_1^2 < 4a_2$ and that the roots are indeed complex.

Example 5 Find the general solution of $y_{t+2} + \frac{1}{4}y_t = 5$. With coefficients $a_1 = 0$ and $a_2 = \frac{1}{4}$, this constitutes an illustration of the complex-root case of $a_1^2 < 4a_2$. By (18.7), the real and imaginary parts of the roots are $h = 0$ and $v = \frac{1}{2}$. It follows from (18.8) that

$$R = \sqrt{0 + \left(\frac{1}{2}\right)^2} = \frac{1}{2}$$

Since the value of θ is that which can satisfy the two equations

$$\cos\theta = \frac{h}{R} = 0 \qquad \text{and} \qquad \sin\theta = \frac{v}{R} = 1$$

it may be concluded from Table 16.1 that

$$\theta = \frac{\pi}{2}$$

Consequently, the complementary function is

$$y_c = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \sin\frac{\pi}{2}t\right)$$

To find y_p, let us try a constant solution $y_t = k$ in the complete equation. This yields $k = 4$; thus, $y_p = 4$, and the general solution can be written as

$$y_t = \left(\frac{1}{2}\right)^t \left(A_5 \cos\frac{\pi}{2}t + A_6 \sin\frac{\pi}{2}t\right) + 4 \tag{18.11}$$

Example 6 Find the general solution of $y_{t+2} - 4y_{t+1} + 16y_t = 0$. In the first place, the particular solution is easily found to be $y_p = 0$. This means that the general solution $y_t\ (= y_c + y_p)$ will be identical with y_c. To find the latter, we note that the coefficients $a_1 = -4$ and $a_2 = 16$ do produce complex roots. Thus we may substitute the a_1 and a_2 values directly into (18.8) and (18.9) to obtain

$$R = \sqrt{16} = 4$$

$$\cos\theta = \frac{4}{2\cdot 4} = \frac{1}{2} \qquad \text{and} \qquad \sin\theta = \sqrt{1 - \frac{16}{4\cdot 16}} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

The last two equations enable us to find from Table 16.2 that

$$\theta = \frac{\pi}{3}$$

It follows that the complementary function—which also serves as the general solution here—is

$$y_c(= y_t) = 4^t\left(A_5 \cos\frac{\pi}{3}t + A_6 \sin\frac{\pi}{3}t\right) \tag{18.12}$$

The Convergence of the Time Path

As in the case of first-order difference equations, the convergence of the time path y_t hinges solely on whether y_c tends toward zero as $t \to \infty$. What we learned about the various configurations of the expression b^t, in Fig. 17.1, is therefore still applicable, although in the present context we shall have to consider *two* characteristic roots rather than one.

Consider first the case of distinct real roots: $b_1 \neq b_2$. If $|b_1| > 1$ and $|b_2| > 1$, then both component terms in the complementary function (18.5)—$A_1b_1^t$ and $A_2b_2^t$—will be explosive, and thus y_c must be divergent. In the opposite case of $|b_1| < 1$ and $|b_2| < 1$, both terms in y_c will converge toward zero as t is indefinitely increased, as will y_c also. What if $|b_1| > 1$ but $|b_2| < 1$? In this intermediate case, it is evident that the $A_2b_2^t$ term tends to "die down," while the other term tends to deviate farther from zero. It follows that the $A_1b_1^t$ term must eventually dominate the scene and render the path divergent.

Let us call the root with the higher *absolute* value the *dominant root*. Then it appears that it is the dominant root b_1 which really sets the tone of the time path, at least with regard to its ultimate convergence or divergence. Such is indeed the case. We may state, thus, that a *time path will be convergent—whatever the initial conditions may be—if and only if the dominant root is less than* 1 *in absolute value*. You can verify that this statement is valid for the cases where both roots are greater than or less than 1 in absolute value (discussed previously), and where one root has an absolute value of 1 exactly (*not* discussed previously). Note, however, that even though the eventual convergence depends on the dominant root alone, the *non-dominant* root will exert a definite influence on the time path, too, at least in the beginning periods. Therefore, the exact configuration of y_t is still dependent on both roots.

Turning to the repeated-root case, we find the complementary function to consist of the terms A_3b^t and A_4tb^t, as shown in (18.6). The former is already familiar to us, but a word of explanation is still needed for the latter, which involves a multiplicative t. If $|b| > 1$, the b^t term will be explosive, and the multiplicative t will simply serve to intensify the explosiveness as t increases. If $|b| < 1$, on the other hand, the b^t part (which tends to zero as t increases) and the t part will run counter to each other; i.e., the value of t will offset rather than reinforce b^t. Which force will prove the stronger? The answer is that the damping force of b^t will always win over the exploding force of t. For this reason, the basic requirement for convergence in the repeated-root case is still that the root be less than 1 in absolute value.

Example 7

Analyze the convergence of the solutions in Examples 3 and 4. For Example 3, the solution is

$$y_t = 3 + (-2)^t + 4t$$

where the roots are 1 and -2, respectively $[3(1)^t = 3]$, and where there is a moving equilibrium $4t$. The dominant root being -2, the time path is divergent.

For Example 4, where the solution is

$$y_t = A_3(-3)^t + A_4t(-3)^t + \frac{1}{4}$$

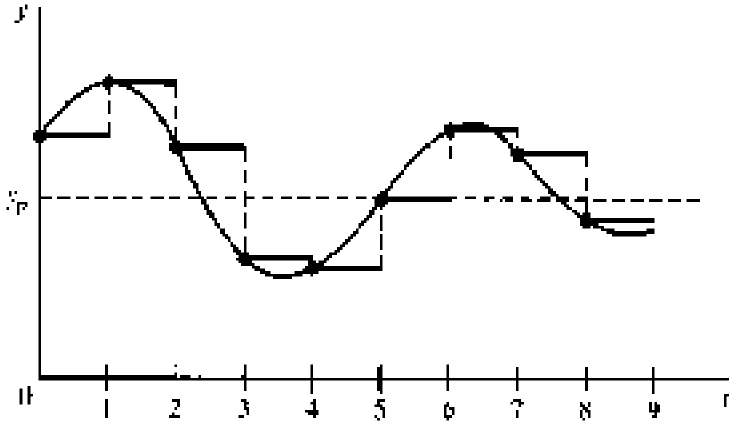
and where $|b| = 3$, we also have divergence

Let us now consider the complex-root case. From the general form of the complementary function in (18.10),

$$y_c = R^t(A_5 \cos\theta t + A_6 \sin\theta t)$$

it is clear that the parenthetical expression, like the one in (16.24'), will produce a fluctuating pattern of a periodic nature. However, since the variable t can only take integer values 0, 1, 2, ... in the present context, we shall catch and utilize only a subset of the points on the graph of a circular function. The y value at each such point will always prevail for a whole period, till the next relevant point is reached. As illustrated in Fig. 18.1, the resulting path is neither the usual oscillatory type (not alternating between values above and below

FIGURE 18.1



y_p in consecutive periods), nor the usual fluctuating type (not smooth); rather, it displays a sort of *stepped* fluctuation. As far as convergence is concerned, though, the decisive factor is really the R^t term, which, like the e^{ht} term in (16.24′), will dictate whether the stepped fluctuation is to be intensified or mitigated as t increases. In the present case, the fluctuation can be gradually narrowed down if and only if $R < 1$. Since R is by definition the absolute value of the conjugate complex roots ($h \pm vi$), the condition for convergence is again that the characteristic roots be less than unity in absolute value.

To summarize: For all three cases of characteristic roots, the time path will converge to a (stationary or moving) intertemporal equilibrium—regardless of what the initial conditions may happen to be—if and only if the absolute value of every root is less than 1.

Example 8 Are the time paths (18.11) and (18.12) convergent? In (18.11) we have $R = \frac{1}{2}$; therefore the time path will converge to the stationary equilibrium (= 4). In (18.12), on the other hand, we have $R = 4$, so the time path will not converge to the equilibrium (= 0).

EXERCISE 18.1

1. Write out the characteristic equation for each of the following, and find the characteristic roots:

   (a) $y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 2$ (c) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 5$

   (b) $y_{t+2} - 4y_{t+1} + 4y_t = 7$ (d) $y_{t+2} - 2y_{t+1} + 3y_t = 4$

2. For each of the difference equations in Prob. 1 state on the basis of its characteristic roots whether the time path involves oscillation or stepped fluctuation, and whether it is explosive.
3. Find the particular solutions of the equations in Prob. 1. Do these represent stationary or moving equilibria?
4. Solve the following difference equations:

   (a) $y_{t+2} + 3y_{t+1} - \frac{7}{4}y_t = 9$ $(y_0 = 6;\ y_1 = 3)$

   (b) $y_{t+2} - 2y_{t+1} + 2y_t = 1$ $(y_0 = 3;\ y_1 = 4)$

   (c) $y_{t+2} - y_{t+1} + \frac{1}{4}y_t = 2$ $(y_0 = 4;\ y_1 = 7)$

5. Analyze the time paths obtained in Prob. 4.

18.2 Samuelson Multiplier-Acceleration Interaction Model

As an illustration of the use of second-order difference equations in economics, let us cite a classic work of Professor Paul Samuelson, the first economist to win the Nobel Prize. We refer to his classic *interaction* model, which seeks to explore the dynamic process of income determination when the acceleration principle is in operation along with the Keynesian multiplier.[†] Among other things, that model serves to demonstrate that the mere interaction of the multiplier and the accelerator is capable of generating cyclical fluctuations endogenously.

The Framework

Suppose that national income Y_t is made up of three component expenditure streams: consumption C_t, investment I_t, and government expenditure G_t. Consumption is envisaged as a function not of current income but of the income of the prior period, Y_{t-1}; for simplicity, it is assumed that C_t is strictly proportional to Y_{t-1}. Investment, which is of the "induced" variety, is a function of the prevailing trend of consumer spending. It is through this induced investment, of course, that the acceleration principle enters into the model. Specifically, we shall assume I_t to bear a fixed ratio to the consumption increment $\Delta C_{t-1} \equiv C_t - C_{t-1}$. The third component, G_t, on the other hand, is taken to be exogenous; in fact, we shall assume it to be a constant and simply denote it by G_0.

These assumptions can be translated into the following set of equations:

$$\begin{aligned} Y_t &= C_t + I_t + G_0 \\ C_t &= \gamma Y_{t-1} \qquad (0 < \gamma < 1) \\ I_t &= \alpha(C_t - C_{t-1}) \qquad (\alpha > 0) \end{aligned} \tag{18.13}$$

where γ (the Greek letter gamma) represents the marginal propensity to consume, and α stands for the accelerator (short for *acceleration coefficient*). Note that, if induced investment is expunged from the model, we are left with a first-order difference equation which embodies the dynamic multiplier process (cf. Example 2 of Sec. 17.2). With induced investment included, however, we have a second-order difference equation that depicts the interaction of the multiplier and the accelerator.

By virtue of the second equation, we can express I_t in terms of income as follows:

$$I_t = \alpha(\gamma Y_{t-1} - \gamma Y_{t-2}) = \alpha\gamma(Y_{t-1} - Y_{t-2})$$

Upon substituting this and the C_t equation into the first equation in (18.13) and rearranging, the model can be condensed into the single equation

$$Y_t - \gamma(1+\alpha)Y_{t-1} + \alpha\gamma Y_{t-2} = G_0$$

or, equivalently (after shifting the subscripts forward by two periods),

$$Y_{t+2} - \gamma(1+\alpha)Y_{t+1} + \alpha\gamma Y_t = G_0 \tag{18.14}$$

Because this is a second-order linear difference equation with constant coefficients and constant term, it can be solved by the method just learned.

[†] Paul A. Samuelson, "Interactions between the Multiplier Analysis and the Principle of Acceleration," *Review of Economic Statistics*, May 1939, pp. 75–78; reprinted in American Economic Association, *Readings in Business Cycle Theory*, Richard D. Irwin, Inc., Homewood, Ill., 1944, pp. 261–269.

The Solution

As the particular solution, we have, by (18.2),

$$Y_p = \frac{G_0}{1-\gamma(1+\alpha)+\alpha\gamma} = \frac{G_0}{1-\gamma}$$

It may be noted that the expression $1/(1-\gamma)$ is merely the multiplier that would prevail in the absence of induced investment. Thus $G_0/(1-\gamma)$—the exogeneous expenditure item times the multiplier—should give us the equilibrium income Y^* in the sense that this income level satisfies the equilibrium condition "national income = total expenditure" [cf. (3.24)]. Being the particular solution of the model, however, it also gives us the intertemporal equilibrium income $\overline{Y}$.

With regard to the complementary function, there are three possible cases. Case 1 ($a_1^2 > 4a_2$), in the present context, is characterized by

$$\gamma^2(1+\alpha)^2 > 4\alpha\gamma \qquad \text{or} \qquad \gamma(1+\alpha)^2 > 4\alpha$$

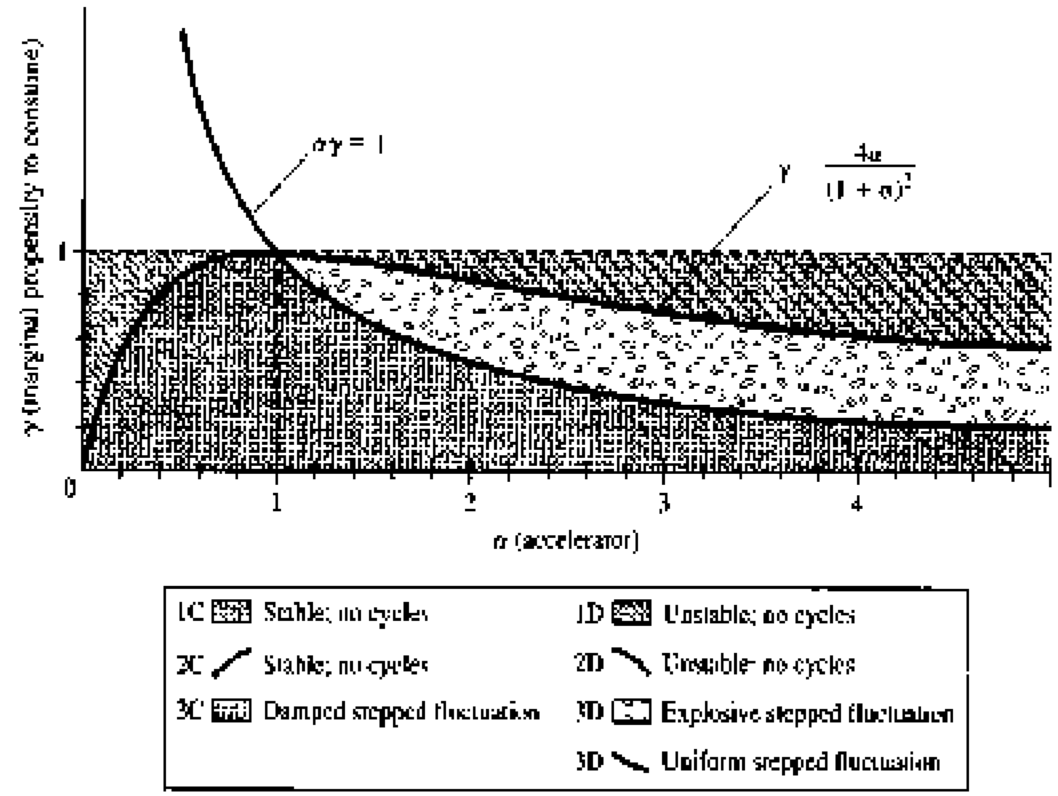
or

$$\gamma > \frac{4\alpha}{(1+\alpha)^2}$$

Similarly, to characterize Cases 2 and 3, we only need to change the > sign in the last inequality to = and <, respectively. In Fig. 18.2, we have drawn the graph of the equation $\gamma = 4\alpha/(1+\alpha)^2$. According to the preceding discussion, the (α, γ) pairs that are located exactly *on* this curve pertain to Case 2. On the other hand, the (α, γ) pairs lying *above* this curve (involving higher γ values) have to do with Case 1, and those lying *below* the curve with Case 3.

This tripartite classification, with its graphical representation in Fig. 18.2, is of interest because it reveals clearly the conditions under which cyclical fluctuations can emerge

FIGURE 18.2



endogenously from the interaction of the multiplier and the accelerator. But this tells nothing about the convergence or divergence of the time path of Y. It remains, therefore, for us to distinguish, under each case, between the *damped* and the *explosive* subcases. We could, of course, take the easy way out by simply illustrating such subcases by citing specific numerical examples. But let us attempt the more rewarding, if also more arduous, task of delineating the general conditions under which convergence and divergence will prevail.

Convergence versus Divergence

The difference equation (18.14) has the characteristic equation

$$b^2 - \gamma(1+\alpha)b + \alpha\gamma = 0$$

which yields the two roots

$$b_1, b_2 = \frac{\gamma(1+\alpha) \pm \sqrt{\gamma^2(1+\alpha)^2 - 4\alpha\gamma}}{2}$$

Since the question of convergence versus divergence depends on the values of b_1 and b_2, and since b_1 and b_2, in turn, depend on the values of the parameters α and γ, the conditions for convergence and divergence should be expressible in terms of the values of α and γ. To do this, we can make use of the fact that—by (16.6)—the two characteristic roots are always related to each other by the following two equations:

$$b_1 + b_2 = \gamma(1+\alpha) \tag{18.15}$$

$$b_1 b_2 = \alpha\gamma \tag{18.15'}$$

One the basis of these two equations, we may observe that

$$\begin{aligned}(1-b_1)(1-b_2) &= 1 - (b_1 + b_2) + b_1 b_2 \\ &= 1 - \gamma(1+\alpha) + \alpha\gamma = 1 - \gamma\end{aligned} \tag{18.16}$$

In view of the model specification that $0 < \gamma < 1$, it becomes necessary to impose on the two roots the condition

$$0 < (1-b_1)(1-b_2) < 1 \tag{18.17}$$

Let us now examine the question of convergence under Case 1, where the roots are real and distinct. Since, by assumption, α and γ are both positive, (18.15′) tells us that $b_1 b_2 > 0$, which implies that b_1 and b_2 possess the same algebraic sign. Furthermore, since $\gamma(1+\alpha) > 0$, (18.15) indicates that both b_1 and b_2 must be positive. Hence, the time path Y_t cannot have oscillations in Case 1.

Even though the signs of b_1 and b_2 are now known, there actually exist under Case 1 as many as five possible combinations of (b_1, b_2) values, each with its own implication regarding the corresponding values for α and γ:

$$\begin{array}{llcl}
(i) & 0 < b_2 < b_1 < 1 & \Rightarrow & 0 < \gamma < 1;\ \alpha\gamma < 1 \\
(ii) & 0 < b_2 < b_1 = 1 & \Rightarrow & \gamma = 1 \\
(iii) & 0 < b_2 < 1 < b_1 & \Rightarrow & \gamma > 1 \\
(iv) & 1 = b_2 < b_1 & \Rightarrow & \gamma = 1 \\
(v) & 1 < b_2 < b_1 & \Rightarrow & 0 < \gamma < 1;\ \alpha\gamma > 1
\end{array}$$

Possibility *i*, where both b_1 and b_2 are positive fractions, duly satisfies condition (18.17) and hence conforms to the model specification $0 < \gamma < 1$. The product of the two roots must also be a positive fraction under this possibility, and this, by (18.15'), implies that $\alpha\gamma < 1$. In contrast, the next three possibilities all violate condition (18.17) and result in inadmissible γ values (see Exercise 18.2-3). Hence they must be ruled out. But Possibility *v* may still be acceptable. With both b_1 and b_2 greater than one, (18.17) may still be satisfied if $(1 - b_1)(1 - b_2) < 1$. But this time we have $\alpha\gamma > 1$ (rather than < 1) from (18.15'). The upshot is that there are only two admissible subcases under Case 1. The first—Possibility *i*—involves fractional roots b_1 and b_2, and therefore yields a convergent time path of Y. The other subcase—Possibility *v*—features roots greater than one, and thus produces a divergent time path. As far as the values of α and γ are concerned, however, the question of convergence and divergence only hinges on whether $\alpha\gamma < 1$ or $\alpha\gamma > 1$. This information is summarized in the top part of Table 18.1, where the convergent subcase is labeled 1C, and the divergent subcase 1D.

The analysis of Case 2, with repeated roots, is similar in nature. The roots are now $b = \gamma(1 + \alpha)/2$, with a positive sign because α and γ are positive. Thus, there is again no oscillation. This time we may classify the value of b into three possibilities only:

$$
\begin{array}{lll}
(vi) & 0 < b < 1 & \Rightarrow \quad \gamma < 1;\ \alpha\gamma < 1 \\
(vii) & b = 1 & \Rightarrow \quad \gamma = 1 \\
(viii) & b > 1 & \Rightarrow \quad \gamma < 1;\ \alpha\gamma > 1
\end{array}
$$

Under Possibility *vi*, $b\,(= b_1 = b_2)$ is a positive fraction; thus the implications regarding α and γ are entirely identical with those of Possibility *i* under Case 1. In an analogous manner, Possibility *viii*, with $b\ (= b_1 = b_2)$ greater than one, can satisfy (18.17) only if $1 < b < 2$; if so, it yields the same results as Possibility *v*. On the other hand, Possibility *vii* violates (18.17) and must be ruled out. Thus there are again only two admissible subcases. The first—Possibility *vi*—yields a convergent time path, whereas the other—Possibility *viii*—gives a divergent one. In terms of α and γ, the convergent and divergent subcases are again associated, respectively, with $\alpha\gamma < 1$ and $\alpha\gamma > 1$. These results are listed in the middle part of Table 18.1, where the two subcases are labeled 2C (convergent) and 2D (divergent).

TABLE 18.1 Cases and Subcases of the Samuelson Model

Case	Subcase	Values of α and γ	Time Path Y_t
1. Distinct real roots			
$\gamma > \frac{4\alpha}{(1+\alpha)^2}$	1C: $0 < b_2 < b_1 < 1$ 1D: $1 < b_2 < b_1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	Nonoscillatory and nonfluctuating
2. Repeated real roots			
$\gamma = \frac{4\alpha}{(1+\alpha)^2}$	2C: $0 < b < 1$ 2D: $b > 1$	$\alpha\gamma < 1$ $\alpha\gamma > 1$	Nonoscillatory and nonfluctuating
3. Complex roots			
$\gamma < \frac{4\alpha}{(1+\alpha)^2}$	3C: $R < 1$ 3D: $R \geq 1$	$\alpha\gamma < 1$ $\alpha\gamma \geq 1$	With stepped fluctuation

Finally, in Case 3, with complex roots, we have stepped fluctuation, and hence endogenous business cycles. In this case, we should look to the absolute value $R = \sqrt{a_2}$ [see (18.8)] for the clue to convergence and divergence, where a_2 is the coefficient of the y_t term in the difference equation (18.1). In the present model, we have $R = \sqrt{\alpha\gamma}$, which gives rise to the following three possibilities:

$$\begin{array}{lccc} (ix) & R < 1 & \Rightarrow & \alpha\gamma < 1 \\ (x) & R = 1 & \Rightarrow & \alpha\gamma = 1 \\ (xi) & R > 1 & \Rightarrow & \alpha\gamma > 1 \end{array}$$

Even though all of these happen to be admissible (see Exercise 18.2-4), only the $R < 1$ possibility entails a convergent time path and qualifies as Subcase 3C in Table 18.1. The other two are thus collectively labeled as Subcase 3D.

In sum, we may conclude from Table 18.1 that a convergent time path can occur if and only if $\alpha\gamma < 1$.

A Graphical Summary

The preceding analysis has resulted in a somewhat complex classification of cases and subcases. It would help to have a visual representation of the classificatory scheme. This is supplied in Fig. 18.2.

The set of all admissible (α, γ) pairs in the model is shown in Fig. 18.2 by the variously shaded rectangular area. Since the values of $\gamma = 0$ and $\gamma = 1$ are excluded, as is the value $\alpha = 0$, the shaded area is a sort of rectangle without sides. We have already graphed the equation $\gamma = 4\alpha/(1+\alpha)^2$ to mark off the three major cases of Table 18.1: The points on that curve pertain to Case 2; the points lying to the north of the curve (representing higher γ values) belong to Case 1; those lying to the south (with lower γ values) are of Case 3. To distinguish between the convergent and divergent subcases, we now add the graph of $\alpha\gamma = 1$ (a rectangular hyperbola) as another demarcation line. The points lying to the north of this rectangular hyperbola satisfy the inequality $\alpha\gamma > 1$, whereas those located below it correspond to $\alpha\gamma < 1$. It is then possible to mark off the subcases easily. Under Case 1, the broken-line shaded region, being below the hyperbola, corresponds to Subcase 1C, but the solid-line shaded region is associated with Subcase 1D. Under Case 2, which relates to the points lying on the curve $\gamma = 4\alpha/(1+\alpha)^2$, Subcase 2C covers the upward-sloping portion of that curve, and Subcase 2D, the downward-sloping portion. Finally, for Case 3, the rectangular hyperbola serves to separate the dot-shaded region (Subcase 3C) from the pebble-shaded region (Subcase 3D). The latter, you should note, also includes the points located *on* the rectangular hyperbola itself, because of the *weak inequality* in the specification $\alpha\gamma \geq 1$.

Since Fig. 18.2 is the repository of *all* the qualitative conclusions of the model, given any ordered pair (α, γ), we can always find the correct subcase graphically by plotting the ordered pair in the diagram.

Example 1 If the accelerator is 0.8 and the marginal propensity to consume is 0.7, what kind of interaction time path will result? The ordered pair (0.8, 0.7) is located in the dot-shaded region, Subcase 3C; thus the time path is characterized by damped stepped fluctuation.

Example 2 What kind of interaction is implied by $\alpha = 2$ and $\gamma = 0.5$? The ordered pair (2, 0.5) lies exactly on the rectangular hyperbola, under Subcase 3D. The time path of Y will again display stepped fluctuation, but it will be neither explosive nor damped. By analogy to the cases of uniform oscillation and uniform fluctuation, we may term this situation as "uniform stepped fluctuation." However, the uniformity feature in this latter case cannot in general be expected to be a perfect one, because, similarly to what was done in Fig. 18.1, we can only accept those points on a sine or cosine curve that correspond to integer values of t, but these values of t may hit an entirely different set of points on the curve in each period of fluctuation.

EXERCISE 18.2

1. By consulting Fig. 18.2, find the subcases to which the following sets of values of α and γ pertain, and describe the interaction time path qualitatively.

   (*a*) $\alpha = 3.5;\ \gamma = 0.8$ (*c*) $\alpha = 0.2;\ \gamma = 0.9$

   (*b*) $\alpha = 2;\ \gamma = 0.7$ (*d*) $\alpha = 1.5;\ \gamma = 0.6$

2. From the values of α and γ given in parts (*a*) and (*c*) of Prob. 1, find the numerical values of the characteristic roots in each instance, and analyze the nature of the time path. Do your results check with those obtained earlier?
3. Verify that Possibilities *ii*, *iii*, and *iv* in Case 1 imply inadmissible values of γ.
4. Show that in Case 3 we can never encounter $\gamma \geq 1$.

18.3 Inflation and Unemployment in Discrete Time

The interaction of inflation and unemployment, discussed earlier in the continuous-time framework, can also be couched in discrete time. Using essentially the same economic assumptions, we shall illustrate in this section how that model can be reformulated as a difference-equation model.

The Model

The earlier continuous-time formulation (Sec. 16.5) consisted of three differential equations:

$$p = \alpha - T - \beta U + g\pi \qquad \text{[expectations-augmented Phillips relation]} \qquad (16.33)$$

$$\frac{d\pi}{dt} = j(p - \pi) \qquad \text{[adaptive expectations]} \qquad (16.34)$$

$$\frac{dU}{dt} = -k(m - p) \qquad \text{[monetary policy]} \qquad (16.35)$$

Three endogenous variables are present: p (actual rate of inflation), π (expected rate of inflation), and U (rate of unemployment). As many as six parameters appear in the model; among these, the parameter m—the rate of growth of nominal money (or, the rate of monetary expansion)—differs from the others in that its magnitude is set as a policy decision.

When cast into the period-analysis mold, the Phillips relation (16.33) simply becomes

$$p_t = \alpha - T - \beta U_t + g\pi_t \qquad (\alpha, \beta > 0;\ 0 < g \le 1) \qquad (18.18)$$

In the adaptive-expectations equation, the derivative must be replaced by a difference expression:

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \qquad (0 < j \le 1) \qquad (18.19)$$

By the same token, the monetary-policy equation should be changed to[†]

$$U_{t+1} - U_t = -k(m - p_{t+1}) \qquad (k > 0) \tag{18.20}$$

These three equations constitute the new version of the inflation-unemployment model.

The Difference Equation in p

As the first step in the analysis of this new model, we again try to condense the model into a single equation in a single variable. Let that variable be p. Accordingly, we shall focus our attention on (18.18). However, since (18.18)—unlike the other two equations—does not by itself describe a pattern of change, it is up to us to create such a pattern. This is accomplished by *differencing* p_t, i.e., by taking the first difference of p_t, according to the definition

$$\Delta p_t \equiv p_{t+1} - p_t$$

Two steps are involved in this. First, we shift the time subscripts in (18.18) forward one period, to get

$$p_{t+1} = \alpha - T - \beta U_{t+1} + g\pi_{t+1} \tag{18.18'}$$

Then we subtract (18.18) from (18.18'), to obtain the first difference of p_t that gives the desired pattern of change:

$$\begin{aligned} p_{t+1} - p_t &= -\beta(U_{t+1} - U_t) + g(\pi_{t+1} - \pi_t) \\ &= \beta k(m - p_{t+1}) + gj(p_t - \pi_t) \qquad \text{[by (18.20) and (18.19)]} \end{aligned} \tag{18.21}$$

Note that, on the second line of (18.21), the patterns of change of the other two variables as given in (18.19) and (18.20) have been incorporated into the pattern of change of the p variable. Thus (18.21) now embodies all the information in the present model.

However, the π_t term is extraneous to the study of p and needs to be eliminated from (18.21). To that end, we make use of the fact that

$$g\pi_t = p_t - (\alpha - T) + \beta U_t \qquad \text{[by (18.18)]} \tag{18.22}$$

Substituting this into (18.21) and collecting terms, we obtain

$$(1 + \beta k)p_{t+1} - [1 - j(1 - g)]p_t + j\beta U_t = \beta k m + j(\alpha - T) \tag{18.23}$$

But there now appears a U_t term to be eliminated. To do that, we difference (18.23) to get a $(U_{t+1} - U_t)$ term and then use (18.20) to eliminate the latter. Only after this rather lengthy process of substitutions, do we get the desired difference equation in the p variable alone, which, when duly normalized, takes the form

$$p_{t+2} - \underbrace{\frac{1 + gj + (1 - j)(1 + \beta k)}{1 + \beta k}}_{a_1} p_{t+1} + \underbrace{\frac{1 - j(1 - g)}{1 + \beta k}}_{a_2} p_t = \underbrace{\frac{j\beta k m}{1 + \beta k}}_{c} \tag{18.24}$$

[†] We have assumed that the change in U_t depends on $(m - p_{t+1})$, the rate of growth of real money in period $(t + 1)$. As an alternative, it is possible to make it depend on the rate of growth of real money in period t, $(m - p_t)$ (see Exercise 18.3-4).

The Time Path of p

The intertemporal equilibrium value of p, given by the particular integral of (18.24), is

$$\overline{p} = \frac{c}{1+a_1+a_2} = \frac{j\beta km}{\beta kj} = m \qquad \text{[by (18.2)]}$$

As in the continuous-time model, therefore, the equilibrium rate of inflation is exactly equal to the rate of monetary expansion.

As to the complementary function, there may arise either distinct real roots (Case 1), or repeated real roots (Case 2), or complex roots (Case 3), depending on the relative magnitudes of a_1^2 and $4a_2$. In the present model,

$$a_1^2 \gtreqless 4a_2 \qquad \text{iff} \qquad [1+gj+(1-j)(1+\beta k)]^2 \gtreqless 4[1-j(1-g)](1+\beta k) \qquad \textbf{(18.25)}$$

If $g=\frac{1}{2}$, $j=\frac{1}{3}$ and $\beta k=5$, for instance, then $a_1^2=(5\frac{1}{6})^2$ whereas $4a_2=20$; thus Case 1 results. But if $g=j=1$, then $a_1^2=4$ while $4a_2=4(1+\beta k)>4$, and we have Case 3 instead. In view of the larger number of parameters in the present model, however, it is not feasible to construct a classificatory graph like Fig. 18.2 in the Samuelson model.

Nevertheless, the analysis of convergence can still proceed along the same line as in Sec. 18.2. Specifically, we recall from (16.6) that the two characteristic roots b_1 and b_2 must satisfy the following two relations:

$$b_1+b_2=-a_1=\frac{1+gj}{1+\beta k}+1-j>0 \qquad \textbf{(18.26)}$$

[see (18.24)]

$$b_1b_2=a_2=\frac{1-j(1-g)}{1+\beta k}\in(0,1) \qquad \textbf{(18.26')}$$

Furthermore, we have in the present model

$$(1-b_1)(1-b_2)=1-(b_1+b_2)+b_1b_2=\frac{\beta jk}{1+\beta k}>0 \qquad \textbf{(18.27)}$$

Now consider Case 1, where the two roots b_1 and b_2 are real and distinct. Since their product b_1b_2 is positive, b_1 and b_2 must take the same sign. Because their sum is positive, moreover, b_1 and b_2 must both be positive, implying that no oscillation can occur. From (18.27), we can infer that neither b_1 nor b_2 can be equal to one; for otherwise $(1-b_1)(1-b_2)$ would be zero, in violation of the indicated inequality. This means that, in terms of the various possibilities of (b_1, b_2) combinations enumerated in the Samuelson model, Possibilities *ii* and *iv* cannot arise here. It is also unacceptable to have one root greater, and the other root less, than one; for otherwise $(1-b_1)(1-b_2)$ would be negative. Thus Possibility *iii* is ruled out as well. It follows that b_1 and b_2 must be *either* both greater than one, *or* both less than one. If $b_1>1$ and $b_2>1$ (Possibility *v*), however, (18.26′) would be violated. Hence the only viable eventuality is Possibility *i*, with b_1 and b_2 both being positive fractions, so that the time path of p is convergent.

The analysis of Case 2 is basically not much different. By practically identical reasoning, we can conclude that the repeated root b can only turn out to be a positive fraction in this model; that is, Possibility *vi* is feasible, but not Possibilities *vii* and *viii*. The time path of p in Case 2 is again nonoscillatory and convergent.

For Case 3, convergence requires that R (the absolute value of the complex roots) be less than one. By (18.8), $R = \sqrt{a_2}$. Inasmuch as a_2 is a positive fraction [see (18.26')], we do have $R < 1$. Thus the time path of p in Case 3 is also convergent, although this time there will be stepped fluctuation.

The Analysis of U

If we wish to analyze instead the time path of the rate of unemployment, we may take (18.20) as the point of departure. To get rid of the p term in that equation, we first substitute (18.18') to get

$$(1+\beta k)U_{t+1} - U_t = k(\alpha - T - m) + kg\pi_{t+1} \qquad \textbf{(18.28)}$$

Next, to prepare for the substitution of the other equation, (18.19), we difference (18.28) to find that

$$(1+\beta k)U_{t+2} - (2+\beta k)U_{t+1} + U_t = kg(\pi_{t+2} - \pi_{t+1}) \qquad \textbf{(18.29)}$$

In view of the presence of a difference expression in π on the right, we can substitute for it a forward-shifted version of the adaptive-expectations equation. The result of this,

$$(1+\beta k)U_{t+2} - (2+\beta k)U_{t+1} + U_t = kgj(p_{t+1} - \pi_{t+1}) \qquad \textbf{(18.30)}$$

is the embodiment of all the information in the model.

However, we must eliminate the p and π variables before a proper difference equation in U will emerge. For this purpose, we note from (18.20) that

$$kp_{t+1} = U_{t+1} - U_t + km \qquad \textbf{(18.31)}$$

Moreover, by multiplying (18.22) through by $(-kj)$ and shifting the time subscripts, we can write

$$\begin{aligned}
-kjg\pi_{t+1} &= -kjp_{t+1} + kj(\alpha - T) - \beta kjU_{t+1} \\
&= -j(U_{t+1} - U_t + km) + kj(\alpha - T) - \beta kjU_{t+1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad [\text{by } (18.31)] \\
&= -j(1+\beta k)U_{t+1} + jU_t + kj(\alpha - T - m) \qquad \textbf{(18.32)}
\end{aligned}$$

These two results express p_{t+1} and π_{t+1} in terms of the U variable and can thus enable us, on substitution into (18.30), to obtain—at long last!—the desired difference equation in the U variable alone:

$$\begin{aligned}
U_{t+2} - \frac{1+gj+(1-j)(1+\beta k)}{1+\beta k}U_{t+1} + \frac{1-j(1-g)}{1+\beta k}U_t \\
= \frac{kj[\alpha - T - (1-g)m]}{1+\beta k} \qquad \textbf{(18.33)}
\end{aligned}$$

It is noteworthy that the two constant coefficients on the left (a_1 and a_2) are identical with those in the difference equation for p [i.e., (18.24)]. As a result, the earlier analysis of the complementary function of the p path should be equally applicable to the present context. But the constant term on the right of (18.33) does differ from that of (18.24). Consequently, the particular solutions in the two situations will be different. This is as it should be, for, coincidence aside, there is no inherent reason to expect the intertemporal equilibrium rate of unemployment to be the same as the equilibrium rate of inflation.

The Long-Run Phillips Relation

It is readily verified that the intertemporal equilibrium rate of unemployment is

$$\overline{U} = \frac{1}{\beta}[\alpha - T - (1 - g)m]$$

But since the equilibrium rate of inflation has been found to be $\overline{p} = m$, we can link $\overline{U}$ to $\overline{p}$ by the equation

$$\overline{U} = \frac{1}{\beta}[\alpha - T - (1 - g)\overline{p}] \qquad (18.34)$$

Because this equation is concerned only with the *equilibrium* rates of unemployment and inflation, it is said to depict the *long-run* Phillips relation.

A special case of (18.34) has received a great deal of attention among economists: the case of $g = 1$. If $g = 1$, the $\overline{p}$ term will have a zero coefficient and thus drop out of the picture. In other words, $\overline{U}$ will become a constant function of $\overline{p}$. In the standard Phillips diagram, where the rate of unemployment is plotted on the horizontal axis, this outcome gives rise to a vertical long-run Phillips curve. The $\overline{U}$ value in this case, referred to as the *natural rate of unemployment*, is then consistent with any equilibrium rate of inflation, with the notable policy implication that, in the long run, there is no trade-off between the twin evils of inflation and unemployment as exists in the short run.

But what if $g < 1$? In that event, the coefficient of $\overline{p}$ in (18.34) will be negative. Then the long-run Phillips curve will turn out to be downward-sloping, thereby still providing a trade-off relation between inflation and unemployment. Whether the long-run Phillips curve is vertical or negatively sloped is, therefore, critically dependent on the value of the g parameter, which, according to the expectations-augmented Phillips relation, measures the extent to which the expected rate of inflation can work its way into the wage structure and the actual rate of inflation. All of this may sound familiar to you. This is because we discussed the topic in Example 1 in Sec. 16.5, and you have also worked on it in Exercise 16.5-4.

EXERCISE 18.3

1. Supply the intermediate steps leading from (18.23) to (18.24).
2. Show that if the model discussed in this section is condensed into a difference equation in the variable π, the result will be the same as (18.24) except for the substitution of π for p.
3. The time paths of p and U in the model discussed in this section have been found to be consistently convergent. Can divergent time paths arise if we drop the assumption that $g \leq 1$? If yes, which divergent "possibilities" in Cases 1, 2, and 3 will now become feasible?
4. Retain equations (18.18) and (18.19), but change (18.20) to

   $U_{t+1} - U_t = -k(m - p_t)$

   (a) Derive a new difference equation in the variable p.

   (b) Does the new difference equation yield a different $\overline{p}$?

   (c) Assume that $j = g = 1$. Find the conditions under which the characteristic roots will fall under Cases 1, 2, and 3, respectively.

   (d) Let $j = g = 1$. Describe the time path of p (including convergence or divergence) when $\beta k = 3$, 4, and 5, respectively.

18.4 Generalizations to Variable-Term and Higher-Order Equations

We are now ready to extend our methods in two directions, to the variable-term case and to difference equations of higher orders.

Variable Term in the Form of cm^t

When the constant term c in (18.1) is replaced by a variable term—some function of t—the only effect will be on the particular solution. (Why?) To find the new particular solution, we can again apply the method of undetermined coefficients. In the differential-equation context (Sec. 16.6), that method requires that the variable term and its successive derivatives together take only a finite number of distinct types of expression, apart from multiplicative constants. Applied to difference equations, the requirement should be amended to read: "the variable term and its successive *differences* must together take only a finite number of distinct expression types, apart from multiplicative constants." Let us illustrate this method by concrete examples, first taking a variable term in the form cm^t, where c and m are constants.

Example 1

Find the particular solution of

$$y_{t+2} + y_{t+1} - 3y_t = 7^t$$

Here, we have $c = 1$ and $m = 7$. First, let us ascertain whether the variable term 7^t yields a finite number of expression types on successive differencing. According to the rule of differencing ($\Delta y_t = y_{t+1} - y_t$), the *first* difference of the term is

$$\Delta 7^t = 7^{t+1} - 7^t = (7-1)7^t = 6(7)^t$$

Similarly, the *second* difference, $\Delta^2(7^t)$, can be expressed as

$$\Delta(\Delta 7^t) = \Delta 6(7^t) = 6(7)^{t+1} - 6(7)^t = 6(7-1)7^t = 36(7)^t$$

Moreover, as can be verified, all successive differences will, like the first and second, be some multiple of 7^t. Since there is only a single expression type, we can try a solution $y_t = B(7)^t$ for the particular solution, where B is an undetermined coefficient.

Substituting the trial solution and its corresponding versions for periods $(t+1)$ and $(t+2)$ into the given difference equation, we obtain

$$B(7)^{t+2} + B(7)^{t+1} - 3B(7)^t = 7^t \qquad \text{or} \qquad B(7^2 + 7 - 3)(7)^t = 7^t$$

Thus,

$$B = \frac{1}{49 + 7 - 3} = \frac{1}{53}$$

and we can write the particular solution as

$$y_p = B(7)^t = \frac{1}{53}(7)^t$$

This may be taken as a moving equilibrium. You can verify the correctness of the solution by substituting it into the difference equation and seeing to it that there will result an identity, $7^t = 7^t$.

The result reached in Example 1 can be easily generalized from the variable term 7^t to that of cm^t. From our experience, we expect all the successive differences of cm^t to take the same form of expression: namely, Bm^t, where B is some multiplicative constant. Hence we can try a solution $y_t = Bm^t$ for the particular solution, when given the difference equation

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = cm^t \tag{18.35}$$

Using the trial solution $y_t = Bm^t$, which implies $y_{t+1} = Bm^{t+1}$, etc., we can rewrite equation (18.35) as

$$Bm^{t+2} + a_1 Bm^{t+1} + a_2 Bm^t = cm^t$$

or

$$B(m^2 + a_1 m + a_2)m^t = cm^t$$

Hence the coefficient B in the trial solution should be

$$B = \frac{c}{m^2 + a_1 m + a_2}$$

and the desired particular solution of (18.35) can be written as

$$y_p = Bm^t = \frac{c}{m^2 + a_1 m + a_2} m^t \qquad (m^2 + a_1 m + a_2 \neq 0) \tag{18.36}$$

Note that the denominator of B is not allowed to be zero. If it happens to be,[†] we must then use the trial solution $y_t = Btm^t$ instead; or, if that too fails, $y_t = Bt^2 m^t$.

Variable Term in the Form of ct^n

Let us now consider variable terms in the form ct^n, where c is any constant, and n is a positive integer.

Example 2

Find the particular solution of

$$y_{t+2} + 5y_{t+1} + 2y_t = t^2$$

The first three differences of t^2 (a special case of ct^n with $c = 1$ and $n = 2$) are found as follows:[‡]

$$\begin{aligned}
\Delta t^2 &= (t+1)^2 - t^2 = 2t + 1 \\
\Delta^2 t^2 &= \Delta(\Delta t^2) = \Delta(2t+1) = \Delta 2t + \Delta 1 \\
&= 2(t+1) - 2t + 0 = 2 \qquad [\Delta \text{ constant} = 0] \\
\Delta^3 t^2 &= \Delta(\Delta^2 t^2) = \Delta 2 = 0
\end{aligned}$$

Since further differencing will only yield zero, there are altogether three distinct types of expression: t^2 (from the variable term itself), t, and a constant (from the successive differences).

Let us therefore try the solution

$$y_t = B_0 + B_1 t + B_2 t^2$$

[†] Analogous to the situation in Example 3 of Sec. 16.6, this eventuality will materialize when the constant m happens to be equal to a characteristic root of the difference equation. The characteristic roots of the difference equation of (18.35) are the values of b that satisfy the equation $b^2 + a_1 b + a_2 = 0$. If one root happens to have the value m, then it must follow that $m^2 + a_1 m + a_2 = 0$.

[‡] These results should be compared with the first three derivatives of t^2:

$$\frac{d}{dt}t^2 = 2t \qquad \frac{d^2}{dt^2}t^2 = 2 \qquad \text{and} \qquad \frac{d^3}{dt^3}t^2 = 0$$

for the particular solution, with undetermined coefficients B_0, B_1, and B_2. Note that this solution implies

$$\begin{aligned} y_{t+1} &= B_0 + B_1(t+1) + B_2(t+1)^2 \\ &= (B_0 + B_1 + B_2) + (B_1 + 2B_2)t + B_2t^2 \\ y_{t+2} &= B_0 + B_1(t+2) + B_2(t+2)^2 \\ &= (B_0 + 2B_1 + 4B_2) + (B_1 + 4B_2)t + B_2t^2 \end{aligned}$$

When these are substituted into the difference equation, we obtain

$$(8B_0 + 7B_1 + 9B_2) + (8B_1 + 14B_2)t + 8B_2t^2 = t^2$$

Equating the two sides term by term, we see that the undetermined coefficients are required to satisfy the following simultaneous equations:

$$\begin{aligned} 8B_0 + 7B_1 + 9B_2 &= 0 \\ 8B_1 + 14B_2 &= 0 \\ 8B_2 &= 1 \end{aligned}$$

Thus, their values must be $B_0 = \frac{13}{256}$, $B_1 = -\frac{7}{32}$, and $B_2 = \frac{1}{8}$, giving us the particular solution

$$y_p = \frac{13}{256} - \frac{7}{32}t + \frac{1}{8}t^2$$

Our experience with the variable term t^2 should enable us to generalize the method to the case of ct^n. In the new trial solution, there should obviously be a term B_nt^n, to correspond to the given variable term. Furthermore, since successive differencing of the term yields the distinct expressions $t^{n-1}, t^{n-2}, \dots, t$, and B_0 (constant), the new trial solution for the case of the variable term ct^n should be written as

$$y_t = B_0 + B_1t + B_2t^2 + \cdots + B_nt^n$$

But the rest of the procedure is entirely the same.

It must be added that such a trial solution may also fail to work. In that event, the trick—already employed on countless other occasions—is again to multiply the original trial solution by a sufficiently high power of t. That is, we can instead try $y_t = t(B_0 + B_1t + B_2t^2 + \cdots + B_nt^n)$, etc.

Higher-Order Linear Difference Equations

The *order* of a difference equation indicates the highest-order difference present in the equation; but it also indicates the maximum number of periods of time lag involved. An nth-order linear difference equation (with constant coefficients and constant term) may thus be written in general as

$$y_{t+n} + a_1y_{t+n-1} + \cdots + a_{n-1}y_{t+1} + a_ny_t = c \qquad (18.37)$$

The method of finding the particular solution of this does not differ in any substantive way. As a starter, we can still try $y_t = k$ (the case of stationary intertemporal equilibrium). Should this fail, we then try $y_t = kt$ or $y_t = kt^2$, etc., in that order.

In the search for the complementary function, however, we shall now be confronted with a characteristic equation which is an nth-degree polynomial equation:

$$b^n + a_1b^{n-1} + \cdots + a_{n-1}b + a_n = 0 \qquad (18.38)$$

There will now be n characteristic roots b_i $(i = 1, 2, \ldots, n)$, all of which should enter into the complementary function thus:

$$y_c = \sum_{i=1}^{n} A_i b_i^t \tag{18.39}$$

provided, of course, that the roots are all real and distinct. In case there are repeated real roots (say, $b_1 = b_2 = b_3$), then the first three terms in the sum in (18.39) must be modified to

$$A_1 b_1^t + A_2 t b_1^t + A_3 t^2 b_1^t \qquad \text{[cf. (18.6)]}$$

Moreover, if there is a pair of conjugate complex roots—say, b_{n-1}, b_n—then the last two terms in the sum in (18.39) are to be combined into the expression

$$R^t(A_{n-1}\cos\theta t + A_n \sin\theta t)$$

A similar expression can also be assigned to any other pair of complex roots. In case of two *repeated* pairs, however, one of the two must be given a multiplicative factor of tR^t instead of R^t.

After y_p and y_c are both found, the general solution of the complete difference equation (18.37) is again obtained by summing; that is,

$$y_t = y_p + y_c$$

But since there will be a total of n arbitrary constants in this solution, no less than n initial conditions will be required to definitize it.

Example 3 Find the general solution of the third-order difference equation

$$y_{t+3} - \frac{7}{8}y_{t+2} + \frac{1}{8}y_{t+1} + \frac{1}{32}y_t = 9$$

By trying the solution $y_t = k$, the particular solution is easily found to be $y_p = 32$. As for the complementary function, since the cubic characteristic equation

$$b^3 - \frac{7}{8}b^2 + \frac{1}{8}b + \frac{1}{32} = 0$$

can be factored into the form

$$\left(b - \frac{1}{2}\right)\left(b - \frac{1}{2}\right)\left(b + \frac{1}{8}\right) = 0$$

the roots are $b_1 = b_2 = \frac{1}{2}$ and $b_3 = -\frac{1}{8}$. This enables us to write

$$y_c = A_1\left(\frac{1}{2}\right)^t + A_2 t\left(\frac{1}{2}\right)^t + A_3\left(-\frac{1}{8}\right)^t$$

Note that the second term contains a multiplicative t; this is due to the presence of repeated roots. The general solution of the given difference equation is then simply the sum of y_c and y_p.

In this example, all three characteristic roots happen to be less than 1 in their absolute values. We can therefore conclude that the solution obtained represents a time path which converges to the stationary equilibrium level 32.

Convergence and the Schur Theorem

When we have a high-order difference equation that is not easily solved, we can nonetheless determine the convergence of the relevant time path qualitatively without having to struggle with its actual quantitative solution. You will recall that the time path can converge if and only if every root of the characteristic equation is less than 1 in absolute value.

In view of this, the following theorem—known as the *Schur theorem*†—becomes directly applicable:

> The roots of the *n*th-degree polynomial equation
>
> $$a_0 b^n + a_1 b^{n-1} + \cdots + a_{n-1} b + a_n = 0$$
>
> will all be less than unity in absolute value if and only if the following n determinants
>
> $$\Delta_1 = \left|\begin{array}{c:c} a_0 & a_n \\ \hdashline a_n & a_0 \end{array}\right| \qquad \Delta_2 = \left|\begin{array}{cc:cc} a_0 & 0 & a_n & a_{n-1} \\ a_1 & a_0 & 0 & a_n \\ \hdashline a_n & 0 & a_0 & a_1 \\ a_{n-1} & a_n & 0 & a_0 \end{array}\right| \qquad \cdots$$
>
> $$\Delta_n = \left|\begin{array}{cccc:cccc} a_0 & 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & 0 & 0 & a_n & \cdots & a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 & 0 & 0 & \cdots & a_n \\ \hdashline a_n & 0 & \cdots & 0 & a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_n & \cdots & 0 & 0 & a_0 & \cdots & a_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & \cdots & a_n & 0 & 0 & \cdots & a_0 \end{array}\right|$$
>
> are all positive.

Note that, since the condition in the theorem is given on the "if and only if" basis, it is a necessary-and-sufficient condition. Thus the Schur theorem is a perfect difference-equation counterpart of the Routh theorem introduced earlier in the differential-equation framework.

The construction of these determinants is based on a simple procedure. This is best explained with the aid of the dashed lines which partition each determinant into four *areas*. Each area of the kth determinant, Δ_k, always consists of a $k \times k$ subdeterminant. The *upper-left* area has a_0 alone in the diagonal, zeros above the diagonal, and progressively larger subscripts for the successive coefficients in each column below the diagonal elements. When we transpose the elements of the upper-left area, we obtain the *lower-right* area. Turning to the *upper-right* area, we now place the a_n coefficient alone in the diagonal, with zeros below the diagonal, and progressively smaller subscripts for the successive coefficients as we go up each column from the diagonal. When the elements of this area are transposed, we get the *lower-left* area.

The application of this theorem is straightforward. Since the coefficients of the characteristic equation are the same as those appearing on the left side of the original difference equation, we can introduce them directly into the determinants cited. Note that, in our context, we always have $a_0 = 1$.

Example 4

Does the time path of the equation $y_{t+2} + 3y_{t+1} + 2y_t = 12$ converge? Here we have $n = 2$, and the coefficients are $a_0 = 1$, $a_1 = 3$, and $a_2 = 2$. Thus we get

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3 < 0$$

† For a discussion of this theorem and its history, see John S. Chipman, *The Theory of Inter-Sectoral Money Flows and Income Formation*, The Johns Hopkins Press, Baltimore, 1951, pp. 119–120.

Since this already violates the convergence condition, there is no need to proceed to Δ_2.

Actually, the characteristic roots of the given difference equation are easily found to be $b_1, b_2 = -1, -2$, which indeed imply a divergent time path.

Example 5 Test the convergence of the path of $y_{t+2} + \frac{1}{6}y_{t+1} - \frac{1}{6}y_t = 2$ by the Schur theorem. Here the coefficients are $a_0 = 1$, $a_1 = \frac{1}{6}$, $a_2 = -\frac{1}{6}$ (with $n = 2$). Thus we have

$$\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 \end{vmatrix} = \frac{35}{36} > 0$$

$$\Delta_2 = \begin{vmatrix} a_0 & 0 & a_2 & a_1 \\ a_1 & a_0 & 0 & a_2 \\ a_2 & 0 & a_0 & a_1 \\ a_1 & a_2 & 0 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 1 & 0 & -\frac{1}{6} \\ -\frac{1}{6} & 0 & 1 & \frac{1}{6} \\ \frac{1}{6} & -\frac{1}{6} & 0 & 1 \end{vmatrix} = \frac{1{,}176}{1{,}296} > 0$$

These do satisfy the necessary-and-sufficient condition for convergence.

EXERCISE 18.4

1. Apply the definition of the "differencing" symbol Δ, to find:
   (a) Δt (b) $\Delta^2 t$ (c) Δt^3
   Compare the results of differencing with those of differentiation.
2. Find the particular solution of each of the following:
   (a) $y_{t+2} + 2y_{t+1} + y_t = 3^t$
   (b) $y_{t+2} - 5y_{t+1} - 6y_t = 2(6)^t$
   (c) $3y_{t+2} + 9y_t = 3(4)^t$
3. Find the particular solutions of:
   (a) $y_{t+2} - 2y_{t+1} + 5y_t = t$
   (b) $y_{t+2} - 2y_{t+1} + 5y_t = 4 + 2t$
   (c) $y_{t+2} + 5y_{t+1} + 2y_t = 18 + 6t + 8t^2$
4. Would you expect that, when the variable term takes the form $m^t + t^n$, the trial solution should be $B(m)^t + (B_0 + B_1 t + \cdots + B_n t^n)$? Why?
5. Find the characteristic roots and the complementary function of:
   (a) $y_{t+3} - \frac{1}{2}y_{t+2} - y_{t+1} + \frac{1}{2}y_t = 0$
   (b) $y_{t+3} - 2y_{t+2} + \frac{5}{4}y_{t+1} - \frac{1}{4}y_t = 1$
   [*Hint:* Try factoring out $(b - \frac{1}{4})$ in both characteristic equations.]
6. Test the convergence of the solutions of the following difference equations by the Schur theorem:
   (a) $y_{t+2} + \frac{1}{2}y_{t+1} - \frac{1}{2}y_t = 3$
   (b) $y_{t+2} - \frac{1}{9}y_t = 1$
7. In the case of a third-order difference equation
   $y_{t+3} + a_1 y_{t+2} + a_2 y_{t+1} + a_3 y_t = c$
   what are the exact forms of the determinants required by the Schur theorem?

Chapter 19

Simultaneous Differential Equations and Difference Equations

Heretofore, our discussion of economic dynamics has been confined to the analysis of a single dynamic (differential or difference) equation. In the present chapter, methods for analyzing a system of simultaneous dynamic equations are introduced. Because this would entail the handling of several variables at the same time, you might anticipate a great deal of new complications. But the truth is that much of what we have already learned about single dynamic equations can be readily extended to systems of simultaneous dynamic equations. For instance, the solution of a dynamic system would still consist of a set of particular integrals or particular solutions (intertemporal equilibrium values of the various variables) and complementary functions (deviations from equilibriums). The complementary functions would still be based on the reduced equations, i.e., the homogeneous versions of the equations in the system. And the dynamic stability of the system would still depend on the signs (if differential equation system) or the absolute values (if difference equation system) of the characteristic roots in the complementary functions. Thus the problem of a dynamic system is only slightly more complicated than that of a single dynamic equation.

19.1 The Genesis of Dynamic Systems

There are two general ways in which a dynamic system can come into being. It may emanate from a given *set* of interacting patterns of change. Or it may be derived from a *single given pattern of change, provided the latter consists of a dynamic equation of the second (or higher) order.*

Interacting Patterns of Change

The most obvious case of a given set of interacting patterns of change is that of a multisector model where each sector, as described by a dynamic equation, impinges on at least one of the other sectors. A dynamic version of the input-output model, for example, could involve n industries whose output changes produce dynamic repercussions on the other industries. Thus it constitutes a dynamic system. Similarly, a dynamic general-equilibrium market model would involve n commodities that are interrelated in their price adjustments. Thus, there is again a dynamic system.

However, interacting patterns of change can be found even in a single-sector model. The various variables in such a model represent, not different sectors or different commodities, but different *aspects* of an economy. Nonetheless, they can affect one another in their dynamic behavior, so as to provide a network of interactions.[†] A concrete example of this has in fact been encountered in Chap. 18. In the inflation-unemployment model, the expected rate of inflation π follows a pattern of change, (18.19), that depends not only on π, but also on the rate of unemployment U (through the actual rate of inflation p). Reciprocally, the pattern of change of U, (18.20), is dependent on π (again through p). Thus the dynamics of π and U must be simultaneously determined. In retrospect, therefore, the inflation-unemployment model could have been treated as a simultaneous-equation dynamic model. And that would have obviated the long sequence of substitutions and eliminations that were undertaken to condense the model into a single equation in one variable. Below, in Sec. 19.4, we shall indeed rework that model, viewed as a dynamic system. Meanwhile, the notion that the same model can be analyzed either as a single equation or as an equation system supplies a natural cue to the discussion of the second way to have a dynamic system.

The Transformation of a High-Order Dynamic Equation

Suppose that we are given an nth-order differential (or difference) equation in *one* variable. Then, as will be shown, it is always possible to transform that equation into a mathematically equivalent system of n simultaneous *first*-order differential (or difference) equations in n variables. In particular, a second-order differential equation can be rewritten as two simultaneous first-order differential equations in two variables.[‡] Thus, even if we happen to start out with only one (high-order) dynamic equation, a dynamic system can nevertheless be derived through the artifice of mathematical transformation. This fact, incidentally, has an important implication: In the ensuing discussion of dynamic systems, we need only be concerned with systems of *first*-order equations, for if a higher-order equation is present, we can always transform it first into a set of first-order equations. This will result in a larger number of equations in the system, but the order will then be lowered to the minimum.

To illustrate the transformation procedure, let us consider the single difference equation

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = c \tag{19.1}$$

If we concoct an artificial new variable x_t, defined by

$$x_t \equiv y_{t+1} \qquad \text{(implying } x_{t+1} \equiv y_{t+2}\text{)}$$

we can then express the original second-order equation by means of *two* first-order (one-period lag) simultaneous equations as follows:

$$\begin{aligned} x_{t+1} \quad + a_1 x_t + a_2 y_t &= c \\ y_{t+1} - \quad x_t \qquad\qquad &= 0 \end{aligned} \tag{19.1'}$$

[†] Note that if we have two dynamic equations in the two variables y_1 and y_2 such that the pattern of change of y_1 depends exclusively on y_1 itself, and similarly for y_2, we really do not have a simultaneous-equation system. Instead, we have merely two separate dynamic equations, each of which can be analyzed by itself, with no requirement of "simultaneity."

[‡] Conversely, two first-order differential (or difference) equations in two variables can be consolidated into a single second-order equation in one variable, as we did in Secs. 16.5 and 18.3.

It is easily seen that, as long as the second equation (which defines the variable x_t) is satisfied, the first is identical with the original given equation. By a similar procedure, and using more artificial variables, we can similarly transform a higher-order single equation into an equivalent system of simultaneous first-order equations. You can verify, for instance, that the third-order equation

$$y_{t+3} + y_{t+2} - 3y_{t+1} + 2y_t = 0 \tag{19.2}$$

can be expressed as

$$\begin{aligned} w_{t+1} \qquad\qquad + w_t - 3x_t + 2y_t &= 0 \\ x_{t+1} \qquad - w_t \qquad\qquad &= 0 \\ y_{t+1} \qquad - x_t \qquad &= 0 \end{aligned} \tag{19.2'}$$

where $x_t \equiv y_{t+1}$ (so that $x_{t+1} \equiv y_{t+2}$) and $w_t \equiv x_{t+1}$ (so that $w_{t+1} \equiv x_{t+2} \equiv y_{t+3}$).

By a perfectly similar procedure, we can also transform an nth-order *differential* equation into a system of n first-order equations. Given the second-order differential equation

$$y''(t) + a_1 y'(t) + a_2 y(t) = 0 \tag{19.3}$$

for instance, we can introduce a new variable $x(t)$, defined by

$$x(t) \equiv y'(t) \qquad [\text{implying } x'(t) \equiv y''(t)]$$

Then (19.3) can be rewritten as the following system of two first-order equations:

$$\begin{aligned} x'(t) \qquad + a_1 x(t) + a_2 y(t) &= 0 \\ y'(t) - \quad x(t) \qquad\qquad &= 0 \end{aligned} \tag{19.3'}$$

where, you may note, the second equation performs the function of defining the newly introduced x variable, as did the second equation in (19.1'). Essentially the same procedure can also be used to transform a higher-order differential equation. The only modification is that we must introduce a correspondingly larger number of new variables.

19.2 Solving Simultaneous Dynamic Equations

The methods for solving simultaneous differential equations and simultaneous difference equations are quite similar. We shall thus discuss them together in this section. For our present purposes, we shall confine the discussion to linear equations with constant coefficients only.

Simultaneous Difference Equations

Suppose that we are given the following system of linear difference equations:

$$\begin{aligned} x_{t+1} \qquad + 6x_t + 9y_t &= 4 \\ y_{t+1} - \quad x_t \qquad &= 0 \end{aligned} \tag{19.4}$$

How do we find the time paths of x and y such that both equations in this system will be satisfied? Essentially, our task is again to seek the particular integrals and complementary functions, and sum these to obtain the desired time paths of the two variables.

Since particular integrals represent intertemporal equilibrium values, let us denote them by $\overline{x}$ and $\overline{y}$. As before, it is advisable first to try constant solutions, namely, $x_{t+1} = x_t = \overline{x}$ and $y_{t+1} = y_t = \overline{y}$. This will indeed work in the present case, for upon substituting these trial solutions into (19.4) we get

$$\left.\begin{aligned} 7\overline{x} + 9\overline{y} &= 4 \\ -\overline{x} + \overline{y} &= 0 \end{aligned}\right\} \quad \Rightarrow \quad \overline{x} = \overline{y} = \frac{1}{4} \tag{19.5}$$

(In case such constant solutions fail to work, however, we must then try solutions of the form $x_t = k_1 t$, $y_t = k_2 t$, etc.)

For the complementary functions, we should, drawing on our previous experience, adopt trial solutions of the form

$$x_t = mb^t \qquad \text{and} \qquad y_t = nb^t \tag{19.6}$$

where m and n are arbitrary constants and the base b represents the characteristic root. It is then automatically implied that

$$x_{t+1} = mb^{t+1} \qquad \text{and} \qquad y_{t+1} = nb^{t+1} \tag{19.7}$$

Note that, to simplify matters, we are employing the same base $b \neq 0$ for both variables, although their coefficients are allowed to differ. It is our aim to find the values of b, m, and n that can make the trial solutions (19.6) satisfy the *reduced* (homogeneous) version of (19.4).

Upon substituting the trial solutions into the reduced version of (19.4) and canceling the common factor $b^t \neq 0$, we obtain the two equations

$$\begin{aligned} (b+6)m + 9n &= 0 \\ -m + bn &= 0 \end{aligned} \tag{19.8}$$

This can be considered as a linear homogeneous-equation system in the two variables m and n—if we are willing to consider b as a parameter for the time being. Because the system (19.8) is homogeneous, it can yield only the trivial solution $m = n = 0$ if its coefficient matrix is nonsingular (see Table 5.1 in Sec. 5.5). In that event, the complementary functions in (19.6) will both be identically zero, signifying that x and y never deviate from their intertemporal equilibrium values. Since that would be an uninteresting special case, we shall try to rule out that trivial solution by requiring the coefficient matrix of the system to be *singular*. That is, we shall require the determinant of that matrix to vanish:

$$\begin{vmatrix} b+6 & 9 \\ -1 & b \end{vmatrix} = b^2 + 6b + 9 = 0 \tag{19.9}$$

From this quadratic equation, we find that $b (= b_1 = b_2) = -3$ is the only value which can prevent m and n from both being zero in (19.8). We shall therefore only use this value of b. Equation (19.9) is called the *characteristic equation*, and its roots the *characteristic roots*, of the given simultaneous difference-equation system.

Once we have a specific value of b, (19.8) gives us the corresponding solution values of m and n. The system being homogeneous, however, there will actually emerge an infinite number of solutions for (m, n), expressible in the form of an equation $m = kn$, where k is a constant. In fact, for each root b_i, there will in general be a distinct equation $m_i = k_i n_i$. Even with repeated roots, with $b_1 = b_2$, we should still use two such equations, $m_1 = k_1 n_1$

and $m_2 = k_2 n_2$ in the complementary functions. Moreover, with repeated roots, we recall from (18.6) that the complementary functions should be written as

$$x_c = m_1(-3)^t + m_2 t(-3)^t$$
$$y_c = n_1(-3)^t + n_2 t(-3)^t$$

The factors of proportionality between m_i and n_i must, of course, satisfy the given equation system (19.4), which mandates that $y_{t+1} = x_t$, i.e.,

$$n_1(-3)^{t+1} + n_2(t+1)(-3)^{t+1} = m_1(-3)^t + m_2 t(-3)^t$$

Dividing through by $(-3)^t$, we get

$$-3n_1 - 3n_2(t+1) = m_1 + m_2 t$$

or, after rearranging,

$$-3(n_1 + n_2) - 3n_2 t = m_1 + m_2 t$$

Equating the terms with t on the two sides of the equals sign, and similarly for the terms without t, we find

$$m_1 = -3(n_1 + n_2) \qquad \text{and} \qquad m_2 = -3n_2$$

If we now write $n_1 = A_3$, $n_2 = A_4$, then it follows that

$$m_1 = -3(A_3 + A_4) \qquad m_2 = -3A_4$$

Thus the complementary functions can be written as

$$\begin{aligned} x_c &= -3(A_3 + A_4)(-3)^t - 3A_4 t(-3)^t \\ &= -3A_3(-3)^t - 3A_4(t+1)(-3)^t \qquad \textbf{(19.10)} \\ y_c &= A_3(-3)^t + A_4 t(-3)^t \end{aligned}$$

where A_3 and A_4 are arbitrary constants. Then the general solution follows easily by combining the particular solutions in (19.5) with the complementary functions just found. All that remains, then, is to definitize the two arbitrary constants A_3 and A_4 with the help of appropriate initial or boundary conditions.

One significant feature of the preceding solution is that, since both time paths have identical b^t expressions in them, they must either both converge or both diverge. This makes sense because, in a model with dynamically interdependent variables, a general intertemporal equilibrium cannot prevail unless no dynamic motion is present anywhere in the system. In the present case, with repeated roots $b = -3$, the time paths of both x and y will display explosive oscillation.

Matrix Notation

In order to bring out the basic parallelism between the methods of solving a single equation and an equation system, the preceding exposition was carried out without the benefit of matrix notation. Let us now see how the latter can be utilized here. Even though it may seem pointless to apply matrix notation to a simple system of only two equations, the possibility of extending that notation to the n-equation case should make it a worthwhile exercise.

First of all, the given system (19.4) may be expressed as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} + \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \tag{19.4'}$$

or, more succinctly, as

$$Iu + Kv = d \tag{19.4''}$$

where I is the 2×2 identity matrix; K is the 2×2 matrix of the coefficients of the x_t and y_t terms; and u, v, and d are column vectors defined as follows:[†]

$$u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} \qquad v = \begin{bmatrix} x_t \\ y_t \end{bmatrix} \qquad d = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

The reader may find one feature puzzling: Since we know $Iu = u$, why not drop the I? The answer is that, even though it seems redundant now, the identity matrix will be needed in subsequent operations, and therefore we shall retain it as in (19.4'').

When we try constant solutions $x_{t+1} = x_t = \overline{x}$ and $y_{t+1} = y_t = \overline{y}$ for the particular solutions, we are in effect setting $u = v = \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$; this will reduce (19.4'') to

$$(I + K) \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = d$$

If the inverse $(I + K)^{-1}$ exists, we can express the particular solutions as

$$\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} = (I + K)^{-1} d \tag{19.5'}$$

This is of course a general formula, for it is valid for any matrix K and vector d as long as $(I + K)^{-1}$ exists. Applied to our numerical example, we have

$$(I + K)^{-1} d = \begin{bmatrix} 7 & 9 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & -\frac{9}{16} \\ \frac{1}{16} & \frac{7}{16} \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

Therefore, $\overline{x} = \overline{y} = \frac{1}{4}$, which checks with (19.5).

Turning to the complementary functions, we see that the trial solutions (19.6) and (19.7) give the u and v vectors the specific forms

$$u = \begin{bmatrix} mb^{t+1} \\ nb^{t+1} \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} \qquad \text{and} \qquad v = \begin{bmatrix} mb^t \\ nb^t \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix} b^t$$

When substituted into the reduced equation $Iu + Kv = 0$, these trial solutions will transform the latter into

$$I \begin{bmatrix} m \\ n \end{bmatrix} b^{t+1} + K \begin{bmatrix} m \\ n \end{bmatrix} b^t = 0$$

[†] The symbol v here denotes a vector. Do not confuse it with the v in the complex-number notation $h \pm vi$, where it represents a scalar.

or, after multiplying through by b^{-t} (a scalar) and factoring,

$$(bI + K)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \qquad \textbf{(19.8')}$$

where 0 is a zero vector. It is from this homogeneous-equation system that we are to find the appropriate values of b, m, and n to be used in the trial solutions in order to make the latter determinate.

To avoid trivial solutions for m and n, it is necessary that

$$|bI + K| = 0 \qquad \textbf{(19.9')}$$

And this is the characteristic equation which will give us the characteristic roots b_i. You can verify that if we substitute

$$bI = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} \qquad \text{and} \qquad K = \begin{bmatrix} 6 & 9 \\ -1 & 0 \end{bmatrix}$$

into this equation, the result will precisely be (19.9), yielding the repeated roots $b = -3$.

In general, each root b_i will elicit from (19.8') a particular set of infinite number of solution values of m and n which are tied to each other by the equation $m_i = k_i n_i$. It is therefore possible to write, for each value of b_i,

$$n_i = A_i \qquad \text{and} \qquad m_i = k_i A_i$$

where A_i are arbitrary constants to be definitized later. When substituted into the trial solutions, these expressions for n_i and m_i along with the values b_i will lead to specific forms of complementary functions. If all roots are distinct real numbers, we may apply (18.5) and write

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \Sigma m_i b_i^t \\ \Sigma n_i b_i^t \end{bmatrix} = \begin{bmatrix} \Sigma k_i A_i b_i^t \\ \Sigma A_i b_i^t \end{bmatrix}$$

With repeated roots, however, we must apply (18.6) instead and, as a result, the complementary functions will contain terms with an extra multiplicative t, such as $m_1 b^t + m_2 t b^t$ (for x_c) and $n_1 b^t + n_2 t b^t$ (for y_c). The factors of proportionality between m_i and n_i are to be determined by the relationship between the variables x and y as stipulated in the given equation system, as illustrated in (19.10) in our numerical example. Finally, in the complex-root case, the complementary functions should be written with (18.10) as their prototype.

Finally, to get the general solution, we can simply form the sum

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$$

Then it remains only to definitize the arbitrary constants A_i.

The extension of this procedure to the n-equation system should be self-evident. When n is large, however, the characteristic equation—an nth-degree polynomial equation—may not be easy to solve quantitatively. In that event, we may again find the Schur theorem to be of help in yielding certain qualitative conclusions about the time paths of the variables in the system. All these variables, we recall, are assigned the same base b in the trial solutions, so they must end up with the same b_i^t expressions in the complementary functions and share the same convergence properties. Thus a single application of the Schur theorem will enable us to determine the convergence or divergence of the time path of every variable in the system.

Simultaneous Differential Equations

The method of solution just described can also be applied to a first-order linear *differential-equation* system. About the only major modification needed is to change the trial solutions to

$$x(t) = me^{rt} \qquad \text{and} \qquad y(t) = ne^{rt} \tag{19.11}$$

which imply that

$$x'(t) = rme^{rt} \qquad \text{and} \qquad y'(t) = rne^{rt} \tag{19.12}$$

In line with our notational convention, the characteristic roots are now denoted by r instead of b.

Suppose that we are given the following equation system:

$$\begin{aligned} x'(t) + 2y'(t) + 2x(t) + 5y(t) &= 77 \\ y'(t) + \quad x(t) + 4y(t) &= 61 \end{aligned} \tag{19.13}$$

First, let us rewrite it in matrix notation as

$$Ju + Mv = g \tag{19.13'}$$

where the matrices are

$$J = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \qquad u = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} \qquad M = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \qquad v = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \qquad g = \begin{bmatrix} 77 \\ 61 \end{bmatrix}$$

Note, that, in view of the appearance of the $2y'(t)$ term in the first equation of (19.13), we have to use the matrix J in place of the identity matrix I, as in (19.4''). Of course, if J is nonsingular (so that J^{-1} exists), then we can in a sense *normalize* (19.13') by premultiplying every term therein by J^{-1}, to get

$$J^{-1}Ju + J^{-1}Mv = J^{-1}g \qquad \text{or} \qquad Iu + Kv = d$$
$$(K \equiv J^{-1}M;\ d \equiv J^{-1}g) \tag{19.13''}$$

This new format is an exact duplicate of (19.4''), although it must be remembered that the vectors u and v have altogether different meanings in the two different contexts. In the ensuing development, we shall adhere to the $Ju + Mv = g$ formulation given in (19.13').

To find the particular integrals, let us try constant solutions $x(t) = \bar{x}$ and $y(t) = \bar{y}$—which imply that $x'(t) = y'(t) = 0$. If these solutions hold, the vectors v and u will become $v = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$ and $u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and (19.13') will reduce to $Mv = g$. Thus the solution for $\bar{x}$ and $\bar{y}$ can be written as

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \bar{v} = M^{-1}g \tag{19.14}$$

which you should compare with (19.5'). In numerical terms, our present problem yields the following particular integrals:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{5}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 77 \\ 61 \end{bmatrix} = \begin{bmatrix} 1 \\ 15 \end{bmatrix}$$

Next, let us look for the complementary functions. Using the trial solutions suggested in (19.11) and (19.12), the vectors u and v become

$$u = \begin{bmatrix} m \\ n \end{bmatrix} re^{rt} \qquad \text{and} \qquad v = \begin{bmatrix} m \\ n \end{bmatrix} e^{rt}$$

Substitution of these into the reduced equation

$$Ju + Mv = 0$$

yields the result

$$J \begin{bmatrix} m \\ n \end{bmatrix} re^{rt} + M \begin{bmatrix} m \\ n \end{bmatrix} e^{rt} = 0$$

or, after multiplying through by the scalar e^{-rt} and factoring,

$$(rJ + M) \begin{bmatrix} m \\ n \end{bmatrix} = 0 \tag{19.15}$$

You should compare this with (19.8'). Since our objective is to find *nontrivial* solutions of m and n (so that our trial solutions will also be nontrivial), it is necessary that

$$|rJ + M| = 0 \tag{19.16}$$

The analog of (19.9'), this last equation—the characteristic equation of the given equation system—will yield the roots r_i that we need. Then, we can find the corresponding (nontrivial) values of m_i and n_i.

In our present example, the characteristic equation is

$$|rJ + M| = \begin{vmatrix} r+2 & 2r+5 \\ 1 & r+4 \end{vmatrix} = r^2 + 4r + 3 = 0 \tag{19.16'}$$

with roots $r_1 = -1$, $r_2 = -3$. Substituting these into (19.15), we get

$$\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = 0 \qquad (\text{for } r_1 = -1)$$

$$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = 0 \qquad (\text{for } r_2 = -3)$$

It follows that $m_1 = -3n_1$ and $m_2 = -n_2$, which we may also express as

$$\begin{aligned} m_1 &= 3A_1 \qquad \text{and} \qquad & m_2 &= A_2 \\ n_1 &= -A_1 & n_2 &= -A_2 \end{aligned}$$

Now that r_i, m_i, and n_i have all been found, the complementary functions can be written as the following linear combinations of exponential expressions:

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} \Sigma m_i e^{r_i t} \\ \Sigma n_i e^{r_i t} \end{bmatrix} \qquad \text{[distinct real roots]}$$

And the general solution will emerge in the form

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$$

In our present example, the solution is

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 3A_1e^{-t} + A_2e^{-3t} + 1 \\ -A_1e^{-t} - A_2e^{-3t} + 15 \end{bmatrix}$$

Moreover, if we are given the initial conditions $x(0) = 6$ and $y(0) = 12$, the arbitrary constants can be found to be $A_1 = 1$ and $A_2 = 2$. These will serve to definitize the preceding solution.

Once more we may observe that, since the e^{rit} expressions are shared by both time paths $x(t)$ and $y(t)$, the latter must either both converge or both diverge. The roots being -1 and -3 in the present case, both time paths converge to their respective equilibria, namely, $\overline{x} = 1$ and $\overline{y} = 15$.

Even though our example consists of a two-equation system only, the method certainly extends to the general n-equation system. When n is large, quantitative solutions may again be difficult, but once the characteristic equation is found, a qualitative analysis will always be possible by resorting to the Routh theorem.

Further Comments on the Characteristic Equation

The term "characteristic equation" has now been encountered in *three* separate contexts: In Sec. 11.3, we spoke of the characteristic equation of a matrix; in Secs. 16.1 and 18.1, the term was applied to a single linear differential equation and difference equation; now, in this section, we have just introduced the characteristic equation of a system of linear difference or differential equations. Is there a connection between the three?

There indeed is, and the connection is a close one. In the first place, given a single equation and an equivalent equation system—as exemplified by the equation (19.1) and the system (19.1'), or the equation (19.3) and the system (19.3')—their characteristic equations must be identical. For illustration, consider the difference equation (19.1), $y_{t+2} + a_1y_{t+1} + a_2y_t = c$. We have earlier learned to write its characteristic equation by directly transplanting its constant coefficients into a quadratic equation:

$$b^2 + a_1b + a_2 = 0$$

What about the equivalent system (19.1')? Taking that system to be in the form of $Iu + Kv = d$, as in (19.4''), we have the matrix $K = \begin{bmatrix} a_1 & a_2 \\ -1 & 0 \end{bmatrix}$. So the characteristic equation is

$$|bI + K| = \begin{vmatrix} b + a_1 & a_2 \\ -1 & b \end{vmatrix} = b^2 + a_1b + a_2 = 0 \qquad \text{[by (19.9')]} \qquad \textbf{(19.17)}$$

which is precisely the same as the one obtained from the single equation as was asserted. Naturally, the same type of result holds also in the differential-equation framework, the only difference being that we would, in accordance with our convention, replace the symbol b by the symbol r in the latter framework.

It is also possible to link the characteristic equation of a difference- (or differential-) equation system to that of a particular square matrix, which we shall call D. Referring to the definition in (11.14), but using the symbol b (instead of r) for the difference-equation framework, we can write the characteristic equation of matrix D as follows:

$$|D - bI| = 0 \tag{19.18}$$

In general, if we multiply every element of the determinant $|D - bI|$ by -1, the value of the determinant will be unchanged if matrix D contains an *even* number of rows (or columns), and will change its sign if D contains an *odd* number of rows. In the present case, however, since $|D - bI|$ is to be set equal to zero, multiplying every element by -1 will not matter, regardless of the dimension of matrix D. But to multiply every element of the determinant $|D - bI|$ by -1 is tantamount to multiplying the matrix $(D - bI)$ by -1 (see Example 6 of Sec. 5.3) before taking its determinant. Thus, (19.18) can be rewritten as

$$|bI - D| = 0 \tag{19.18'}$$

When this is equated to (19.17), it becomes clear that if we pick the matrix $D = -K$, then its characteristic equation will be identical with that of the system (19.1'). This matrix, $-K$, has a special meaning: If we take the *reduced* version of the system, $Iu + Kv = 0$, and express it in the form of $Iu = -Kv$, or simply $u = -Kv$, we see that $-K$ is the matrix that can transform the vector $v = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$ into the vector $u = \begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix}$ in that particular equation.

Again, the same reasoning can be adapted to the differential-equation system (19.3'). However, in the case of a system such as (19.13'), $Ju + Mv = g$, where—unlike in the system (19.3')—the first term is Ju rather than Iu, the characteristic equation is in the form

$$|rJ + M| = 0 \qquad \text{[cf. (19.16')]}$$

For this case, if we wish to find the expression for the matrix D, we must first normalize the equation $Ju + Mv = g$ into the form of (19.13''), and then take $D = -K = -J^{-1}M$.

In sum, given (1) a single difference or differential equation, and (2) an equivalent equation system, from which we can also obtain (3) an appropriate matrix D, if we try to find the characteristic equations of all three of these, the results must be one and the same.

EXERCISE 19.2

1. Verify that the difference-equation system (19.4) is equivalent to the single equation $y_{t+2} + 6y_{t+1} + 9y_t = 4$, which was solved earlier as Example 4 in Sec. 18.1. How do the solutions obtained by the two different methods compare?
2. Show that the characteristic equation of the difference equation (19.2) is identical with that of the equivalent system (19.2').
3. Solve the following two difference-equation systems:

(a)
$$\begin{aligned} x_{t+1} + x_t + 2y_t &= 24 \\ y_{t+1} + 2x_t - 2y_t &= 9 \end{aligned} \qquad (\text{with } x_0 = 10 \text{ and } y_0 = 9)$$

(b)
$$\begin{aligned} x_{t+1} - x_t - \tfrac{1}{3}y_t &= -1 \\ x_{t+1} - y_{t+1} - \tfrac{1}{6}y_t &= 8\tfrac{1}{2} \end{aligned} \qquad (\text{with } x_0 = 5 \text{ and } y_0 = 4)$$

4. Solve the following two differential-equation systems:

   (a) $x'(t) \quad - x(t) - 12y(t) = -60$

   $y'(t) + x(t) + 6y(t) = 36 \qquad$ [with $x(0) = 13$ and $y(0) = 4$]

   (b) $x'(t) \quad - 2x(t) + 3y(t) = 10$

   $y'(t) - x(t) + 2y(t) = 9 \qquad$ [with $x(0) = 8$ and $y(0) = 5$]

5. On the basis of the differential-equation system (19.13), find the matrix D whose characteristic equation is identical with that of the system. Check that the characteristic equations of the two are indeed the same.

19.3 Dynamic Input-Output Models

Our first encounter with input-output analysis was concerned with the question: How much should be produced in each industry so that the input requirements of all industries, as well as the final demand (open system), will be exactly satisfied? The context was static, and the problem was to solve a simultaneous-equation system for the *equilibrium* output levels of all industries. When certain additional economic considerations are incorporated into the model, the input-output system can take on a dynamic character, and there will then result a difference- or differential-equation system of the type discussed in Sec. 19.2.

Three such dynamizing considerations will be considered here. To keep the exposition simple, however, we shall illustrate with two-industry open systems only. Nevertheless, since we shall employ matrix notation, the generalization to the n-industry case should not prove difficult, for it can be accomplished simply by duly changing the dimensions of the matrices involved. For purposes of such generalization, it will prove advisable to denote the variables not by x_t and y_t but by $x_{1,t}$ and $x_{2,t}$, so that we can extend the notation to $x_{n,t}$ when needed. You will recall that, in the input-output context, x_i represents the output (measured in dollars) of the ith industry; the new subscript t will now add a time dimension to it. The input-coefficient symbol a_{ij} will still mean the dollar worth of the ith commodity required in the production of a dollar's worth of the jth commodity, and d_t will again indicate the final demand for the ith commodity.

Time Lag in Production

In a static two-industry open system, the output of industry I should be set at the level of demand as follows:

$$x_1 = a_{11}x_1 + a_{12}x_2 + d_1$$

Now assume a one-period lag in production, so that the amount demanded in period t determines not the current output but the output of period $(t + 1)$. To depict this new situation, we must modify the preceding equation to the form

$$x_{1,t+1} = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} \tag{19.19}$$

Similarly, we can write for industry II:

$$x_{2,t+1} = a_{21}x_{1,t} + a_{22}x_{2,t} + d_{2,t} \tag{19.19'}$$

Thus we now have a system of simultaneous difference equations; this constitutes a dynamic version of the input-output model.

In matrix notation, the system consists of the equation

$$x_{t+1} - Ax_t = d_t \tag{19.20}$$

where $x_{t-1} = \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix}$ $\quad x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}$ $\quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $\quad d_t = \begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}$

Clearly, (19.20) is in the form of (19.4″), with only two exceptions. First, unlike vector u, vector x_{t+1} does not have an identity matrix I as its "coefficient." However, as explained earlier, this really makes no analytical difference. The second, and more substantive, point is that the vector d_t, with a time subscript, implies that the final-demand vector is being viewed as a function of time. If this function is nonconstant, a modification will be required in the method of finding the particular solutions, although the complementary functions will remain unaffected. The following example will illustrate the modified procedure.

Example 1 Given the exponential final-demand vector

$$d_t = \begin{bmatrix} \delta^t \\ \delta^t \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t \qquad (\delta = \text{a positive scalar})$$

find the particular solutions of the dynamic input-output model (19.20). In line with the method of undetermined coefficients introduced in Sec. 18.4, we should try solutions of the form $x_{1,t} = \beta_1\delta^t$ and $x_{2,t} = \beta_2\delta^t$, where β_1 and β_2 are undetermined coefficients. That is, we should try

$$x_t = \begin{bmatrix} \beta_1\delta^t \\ \beta_2\delta^t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t \tag{19.21}$$

which implies[†]

$$x_{t+1} = \begin{bmatrix} \beta_1\delta^{t+1} \\ \beta_2\delta^{t+1} \end{bmatrix} = \begin{bmatrix} \beta_1\delta \\ \beta_2\delta \end{bmatrix} \delta^t = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t$$

If the indicated trial solutions hold, then the system (19.20) will become

$$\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t - \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta^t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \delta^t$$

or, on canceling the common scalar multiplier $\delta^t \neq 0$,

$$\begin{bmatrix} \delta - a_{11} & -a_{12} \\ -a_{21} & \delta - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{19.22}$$

[†] You will note that the vector $\begin{bmatrix} \beta_1\delta \\ \beta_2\delta \end{bmatrix}$ can be rewritten in several equivalent forms:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \delta \quad \text{or} \quad \delta \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{or} \quad \delta \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

We choose the third alternative here because in a subsequent step we shall want to add $\begin{bmatrix} \delta & 0 \\ 0 & \delta \end{bmatrix}$ to another 2×2 matrix. The first two alternative forms will entail problems of dimension conformability.

Assuming the coefficient matrix on the extreme left to be nonsingular, we can readily find β_1 and β_2 (by Cramer's rule) to be

$$\beta_1 = \frac{\delta - a_{22} + a_{12}}{\Delta} \qquad \text{and} \qquad \beta_2 = \frac{\delta - a_{11} + a_{21}}{\Delta} \tag{19.22'}$$

where $\Delta \equiv (\delta - a_{11})(\delta - a_{22}) - a_{12}a_{21}$. Since β_1 and β_2 are now expressed entirely in the known values of the parameters, we only need to insert them into the trial solution (19.21) to get the definite expressions for the particular solutions.

A more general version of the type of final-demand vector discussed here is given in Exercise 19.3-1.

The procedure for finding the complementary functions of (19.20) is no different from that presented in Sec. 19.2. Since the homogeneous version of the equation system is $x_{t+1} - Ax_t = 0$, the characteristic equation should be

$$|bI - A| = \begin{vmatrix} b - a_{11} & -a_{12} \\ -a_{21} & b - a_{22} \end{vmatrix} = 0 \qquad [\text{cf. (19.9')}]$$

From this we can find the characteristic roots b_1 and b_2 and thence proceed to the remaining steps of the solution process.

Excess Demand and Output Adjustment

The model formulation in (19.20) can also arise from a different economic assumption. Consider the situation in which the excess demand for each product always tends to induce an output increment equal to the excess demand. Since the excess demand for the first product in period t amounts to

$$\underbrace{a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t}}_{\text{demanded}} - \underbrace{x_{1,t}}_{\text{supplied}}$$

the output adjustment (increment) $\Delta x_{1,t}$ is to be set exactly equal to that level:

$$\Delta x_{1,t} (\equiv x_{1,t+1} - x_{1,t}) = a_{11}x_{1,t} + a_{12}x_{2,t} + d_{1,t} - x_{1,t}$$

However, if we add $x_{1,t}$ to both sides of this equation, the result will become identical with (19.19). Similarly, our output-adjustment assumption will give an equation the same as (19.19') for the second industry. In short, the same mathematical model can result from altogether different economic assumptions.

So far, the input-output system has been viewed only in the discrete-time framework. For comparison purposes, let us now cast the output-adjustment process in the continuous-time mold.

In the main, this would call for use of the symbol $x_i(t)$ in lieu of $x_{i,t}$, and of the derivative $x_i'(t)$ in lieu of the difference $\Delta x_{i,t}$. With these changes, our output-adjustment assumption will manifest itself in the following pair of differential equations:

$$x_1'(t) = a_{11}x_1(t) + a_{12}x_2(t) + d_1(t) - x_1(t)$$
$$x_2'(t) = a_{21}x_1(t) + a_{22}x_2(t) + d_2(t) - x_2(t)$$

At any instant of time $t = t_0$, the symbol $x_i(t_0)$ tells us the rate of output flow per unit of time (say, per month) that prevails at the said instant, and $d_i(t_0)$ indicates the final demand per month prevailing at that instant. Hence the right-hand sum in each equation indicates the rate of excess demand per month, measured at $t = t_0$. The derivative $x_i'(t_0)$ at the left, on the other hand, represents the rate of output adjustment per month called forth by the excess demand at $t = t_0$. This adjustment will eradicate the excess demand (and bring about equilibrium) in a month's time, but only if both the excess demand and the output adjustment stay unchanged at the current rates. In actuality, the excess demand will vary with time, as will the induced output adjustment, thus resulting in a cat-and-mouse game of chase. The solution of the system, consisting of the time paths of the output x_i, supplies a chronicle of this chase. If the solution is convergent, the cat (output adjustment) will eventually be able to catch the mouse (excess demand), asymptotically (as $t \to \infty$).

After proper rearrangement, this system of differential equations can be written in the format of (19.13′) as follows:

$$Ix' + (I - A)x = d \tag{19.23}$$

where $x' = \begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} \qquad x = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad d = \begin{bmatrix} d_1(t) \\ d_2(t) \end{bmatrix}$

(the prime denoting derivative, not transpose). The complementary functions can be found by the method discussed earlier. In particular, the characteristic roots are to be found from the equation

$$|rI + (I - A)| = \begin{vmatrix} r + 1 - a_{11} & -a_{12} \\ -a_{21} & r + 1 - a_{22} \end{vmatrix} = 0 \qquad \text{[cf. (19.16)]}$$

As for the particular integrals, if the final-demand vector contains nonconstant functions of time $d_1(t)$ and $d_2(t)$ as its elements, a modification will be needed in the method of solution. Let us illustrate with a simple example.

Example 2 Given the final-demand vector

$$d = \begin{bmatrix} \lambda_1 e^{\rho t} \\ \lambda_2 e^{\rho t} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} e^{\rho t}$$

where λ_i and ρ are constants, find the particular integrals of the dynamic model (19.23). Using the method of undetermined coefficients, we can try solutions of the form $x_i(t) = \beta_i e^{\rho t}$, which imply, of course, that $x_i'(t) = \rho\beta_i e^{\rho t}$. In matrix notation, these can be written as

$$x = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \tag{19.24}$$

and $\quad x' = \rho \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} e^{\rho t} \qquad$ [cf. footnote in Example 1]

Upon substituting into (19.23) and canceling the common (nonzero) scalar multiplier $e^{\rho t}$, we obtain

$$\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

or

$$\begin{bmatrix} \rho + 1 - a_{11} & -a_{12} \\ -a_{21} & \rho + 1 - a_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \tag{19.25}$$

If the leftmost matrix is nonsingular, we can apply Cramer's rule and determine the values of the coefficients β_i to be

$$\begin{aligned}\beta_1 &= \frac{\lambda_1(\rho+1-a_{22})+\lambda_2 a_{12}}{\Delta} \\ \beta_2 &= \frac{\lambda_2(\rho+1-a_{11})+\lambda_1 a_{21}}{\Delta}\end{aligned} \tag{19.25'}$$

where $\Delta \equiv (\rho+1-a_{11})(\rho+1-a_{22})-a_{12}a_{21}$. The *undetermined coefficients* having thus been determined, we can introduce these values into the trial solution (19.24) to obtain the desired particular integrals.

Capital Formation

Another economic consideration that can give rise to a dynamic input-output system is capital formation, including the accumulation of inventory.

In the static discussion, we only considered the output level of each product needed to satisfy current demand. The needs for inventory accumulation or capital formation were either ignored, or subsumed under the final-demand vector. To bring capital formation into the open, let us now consider—along with an input-coefficient matrix $A = [a_{ij}]$—a capital-coefficient matrix

$$C = [c_{ij}] = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

where c_{ij} denotes the dollar worth of the ith commodity needed by the jth industry as new capital (either equipment or inventory, depending on the nature of the ith commodity) as a result of an output increment of \$1 in the jth industry. For example, if an increase of \$1 in the output of the soft-drink (jth) industry induces it to add \$2 worth of bottling equipment (ith commodity), then $c_{ij} = 2$. Such a capital coefficient thus reveals a marginal capital-output ratio of sorts, the ratio being limited to one type of capital (the ith commodity) only. Like the input coefficients a_{ij}, the capital coefficients are assumed to be fixed. The idea is for the economy to produce each commodity in such quantity as to satisfy not only the input-requirement demand plus the final demand, but also the capital-requirement demand for it.

If time is *continuous*, output increment is indicated by the derivatives $x_i'(t)$; thus the output of each industry should be set at

$$\begin{aligned} x_1(t) &= a_{11}x_1(t) + a_{12}x_2(t) + c_{11}x_1'(t) + c_{12}x_2'(t) + d_1(t) \\ x_2(t) &= \underbrace{a_{21}x_1(t) + a_{22}x_2(t)}_{\text{input requirement}} + \underbrace{c_{21}x_1'(t) + c_{22}x_2'(t)}_{\text{capital requirement}} + \underbrace{d_2(t)}_{\text{final demand}} \end{aligned}$$

In matrix notation, this is expressible by the equation

$$Ix = Ax + Cx' + d$$

or

$$Cx' + (A - I)x = -d \tag{19.26}$$

If time is *discrete*, the capital requirement in period t will be based on the output increment $x_{i,t} - x_{i,t-1}$ $(\equiv \Delta x_{i,t-1})$; thus the output levels should be set at

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix}}_{\text{input requirement}} + \underbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} x_{1,t} - x_{1,t-1} \\ x_{2,t} - x_{2,t-1} \end{bmatrix}}_{\text{capital requirement}} + \underbrace{\begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix}}_{\text{final demand}}$$

or $$x_t = Ax_t + C(x_t - x_{t-1}) + d_t$$

By shifting the time subscripts forward one period, and collecting terms, however, we can write the equation in the form

$$(I - A - C)x_{t+1} + Cx_t = d_{t+1} \tag{19.27}$$

The differential-equation system (19.26) and the difference-equation system (19.27) can again be solved, of course, by the method of Sec. 19.2. It also goes without saying that these two matrix equations are both extendible to the n-industry case simply by an appropriate redefinition of the matrices and a corresponding change in the dimensions thereof.

In the preceding, we have discussed how a dynamic input-output model can arise from such considerations as time lags and adjustment mechanisms. When similar considerations are applied to general-equilibrium market models, the latter will tend to become dynamic in much the same way. But, since the formulation of such models is analogous in spirit to input-output models, we shall dispense with a formal discussion thereof and merely refer you to the illustrative cases in Exercises 19.3-6 and 19.3-7.

EXERCISE 19.3

1. In Example 1, if the final-demand vector is changed to $d_t = \begin{bmatrix} \lambda_1 \delta^t \\ \lambda_2 \delta^t \end{bmatrix}$, what will the particular solutions be? After finding your answers, show that the answers in Example 1 are merely a special case of these, with $\lambda_1 = \lambda_2 = 1$.
2. (*a*) Show that (19.22) can be written more concisely as
   $(\delta I - A)\beta = u$
   (*b*) Of the five symbols used, which are scalars? Vectors? Matrices?
   (*c*) Write the solution for β in matrix form, assuming $(\delta I - A)$ to be nonsingular.
3. (*a*) Show that (19.25) can be written more concisely as
   $(\rho I + I - A)\beta = \lambda$
   (*b*) Which of the five symbols represent scalars, vectors, and matrices, respectively?
   (*c*) Write the solution for β in matrix form, assuming $(\rho I + I - A)$ to be nonsingular.
4. Given $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ and $d_t = \begin{bmatrix} (\frac{12}{10})^t \\ (\frac{12}{10})^t \end{bmatrix}$ for the discrete-time production-lag input-output model described in (19.20), find (*a*) the particular solutions; (*b*) the complementary functions; and (*c*) the definite time paths, assuming initial outputs $x_{1,0} = \frac{187}{39}$ and $x_{2,0} = \frac{72}{13}$. (Use fractions, not decimals, in all calculations.)

5. Given $A = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} \\ \frac{3}{10} & \frac{2}{10} \end{bmatrix}$ and $d = \begin{bmatrix} e^{t/10} \\ 2e^{t/10} \end{bmatrix}$ for the continuous-time output-adjustment input-output model described in (19.23), find (*a*) the particular integrals; (*b*) the complementary functions; and (*c*) the definite time paths, assuming initial conditions $x_1(0) = \frac{23}{6}$ and $x_2(0) = \frac{25}{6}$. (Use fractions, not decimals, in all calculations.)

6. In an n-commodity market, all Q_{di} and Q_{si} (with $i = 1, 2, \ldots, n$) can be considered as functions of the n prices $P_1, \ldots, P_n$, and so can the excess demand for each commodity $E_i \equiv Q_{di} - Q_{si}$. Assuming linearity, we can write

$$\begin{aligned} E_1 &= a_{10} + a_{11}P_1 + a_{12}P_2 + \cdots + a_{1n}P_n \\ E_2 &= a_{20} + a_{21}P_1 + a_{22}P_2 + \cdots + a_{2n}P_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ E_n &= a_{n0} + a_{n1}P_1 + a_{n2}P_2 + \cdots + a_{nn}P_n \end{aligned}$$

or, in matrix notation,

$$E = a + AP$$

(*a*) What do these last four symbols stand for—scalars, vectors, or matrices? What are their respective dimensions?

(*b*) Consider all prices to be functions of time, and assume that $dP_i/dt = \alpha_i E_i$ $(i = 1, 2, \ldots, n)$. What is the economic interpretation of this last set of equations?

(*c*) Write out the differential equations showing each dP_i/dt to be a linear function of the n prices.

(*d*) Show that, if we let P' denote the $n \times 1$ column vector of the derivatives dP_i/dt, and if we let α denote an $n \times n$ diagonal matrix, with $\alpha_1, \alpha_2, \ldots, \alpha_n$ (in that order) in the principal diagonal and zeros elsewhere, we can write the preceding differential-equation system in matrix notation as $P' - \alpha AP = \alpha a$.

7. For the n-commodity market of Prob. 6, the discrete-time version would consist of a set of difference equations $\Delta P_{i,t} = \alpha_i E_{i,t}$ $(i = 1, 2, \ldots, n)$, where $E_{i,t} = a_{i0} + a_{i1}P_{1,t} + a_{i2}P_{2,t} + \cdots + a_{in}P_{n,t}$.

(*a*) Write out the excess-demand equation system, and show that it can be expressed in matrix notation as $E_t = a + AP_t$.

(*b*) Show that the price adjustment equations can be written as $P_{t+1} - P_t = \alpha E_t$, where α is the $n \times n$ diagonal matrix defined in Prob. 6.

(*c*) Show that the difference-equation system of the present discrete-time model can be expressed in the form $P_{t+1} - (I + \alpha A)P_t = \alpha a$.

19.4 The Inflation-Unemployment Model Once More

Having illustrated the multisector type of dynamic systems with input-output models, we shall now provide an economic example of simultaneous dynamic equations in the one-sector setting. For this purpose, the inflation-unemployment model, already encountered twice before in two different guises, can be called back into service once again.

Simultaneous Differential Equations

In Sec. 16.5 the inflation-unemployment model was presented in the continuous-time framework via the following three equations:

$$p = \alpha - T - \beta U + g\pi \qquad (\alpha, \beta > 0;\ 0 < g \leq 1) \qquad \textbf{(16.33)}$$

$$\frac{d\pi}{dt} = j(p - \pi) \qquad (0 < j \leq 1) \qquad \textbf{(16.34)}$$

$$\frac{dU}{dt} = -k(\mu - p) \qquad (k > 0) \qquad \textbf{(16.35)}$$

except that we have adopted the Greek letter μ here to replace m in (16.35) in order to avoid confusion with our earlier usage of the symbol m in the methodological discussion of Sec. 19.2. In the treatment of this model in Sec. 16.5, since we were not yet equipped then to deal with simultaneous dynamic equations, we approached the problem by condensing the model into a single equation in one variable. That necessitated a quite laborious process of substitutions and eliminations. Now, in view of the coexistence of two given patterns of change in the model for π and U, we shall treat the model as one of two simultaneous differential equations.

When (16.33) is substituted into the other two equations, and the derivatives $d\pi/dt \equiv \pi'(t)$ and $dU/dt \equiv U'(t)$ written more simply as π' and U', the model assumes the form

$$\begin{aligned} \pi' \quad + j(1 - g)\pi + j\beta U &= j(\alpha - T) \\ U' - \qquad kg\pi + k\beta U &= k(\alpha - T - \mu) \end{aligned} \qquad \textbf{(19.28)}$$

or, in matrix notation,

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{J} \begin{bmatrix} \pi' \\ U' \end{bmatrix} + \underbrace{\begin{bmatrix} j(1-g) & j\beta \\ -kg & k\beta \end{bmatrix}}_{M} \begin{bmatrix} \pi \\ U \end{bmatrix} = \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix} \qquad \textbf{(19.28')}$$

From this system, the time paths of π and U can be found simultaneously. Then, if desired, we can derive the p path by using (16.33).

Solution Paths

To find the particular integrals, we can simply set $\pi' = U' = 0$ (to make π and U stationary over time) in (19.28') and solve for π and U. In our earlier discussion, in (19.14), such solutions were obtained through matrix inversion, but Cramer's rule can certainly be used, too. Either way, we can find that

$$\bar{\pi} = \mu \qquad \text{and} \qquad \bar{U} = \frac{1}{\beta}[\alpha - T - (1 - g)\mu] \qquad \textbf{(19.29)}$$

The result that $\bar{\pi} = \mu$ (the equilibrium expected rate of inflation equals the rate of monetary expansion) coincides with that reached in Sec. 16.5. As to the rate of unemployment U, we made no attempt to find its equilibrium level in that section. If we did (on the basis of the differential equation in U given in Exercise 16.5-2), however, the answer would be no different from the $\bar{U}$ solution in (19.29).

Turning to the complementary functions, which are based on the trial solutions me^{rt} and ne^{rt}, we can determine m, n, and r from the reduced matrix equation

$$(rJ + M)\begin{bmatrix} m \\ n \end{bmatrix} = 0 \qquad \text{[from (19.15)]}$$

which, in the present context, takes the form

$$\begin{bmatrix} r + j(1-g) & j\beta \\ -kg & r + k\beta \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{19.30}$$

To avoid trivial solutions for m and n from this homogeneous system, the determinant of the coefficient matrix must be made to vanish; that is, we require

$$|rJ + M| = r^2 + [k\beta + j(1-g)]r + k\beta j = 0 \tag{19.31}$$

This quadratic equation, a specific version of the characteristic equation $r^2 + a_1 r + a_2 = 0$, has coefficients

$$a_1 = k\beta + j(1-g) \qquad \text{and} \qquad a_2 = k\beta j$$

And these, as we would expect, are precisely the a_1 and a_2 values in (16.37″)—a single-equation version of the present model in the variable π. As a result, the previous analysis of the three cases of characteristic roots should apply here with equal validity. Among other conclusions, we may recall that, regardless of whether the roots happen to be real or complex, the real part of each root in the present model turns out to be always negative. Thus the solution paths are always convergent.

Example 1

Find the time paths of π and U, given the parameter values

$$\alpha - T = \frac{1}{6} \qquad \beta = 3 \qquad g = 1 \qquad j = \frac{3}{4} \qquad \text{and} \qquad k = \frac{1}{2}$$

Since these parameter values duplicate those in Example 1 in Sec. 16.5, the results of the present analysis can be readily checked against those of the said section.

First, it is easy to determine that the particular integrals are

$$\overline{\pi} = \mu \qquad \text{and} \qquad \overline{U} = \frac{1}{3}\left(\frac{1}{6}\right) = \frac{1}{18} \qquad \text{[by (19.29)]} \tag{19.32}$$

The characteristic equation being

$$r^2 + \frac{3}{2}r + \frac{9}{8} = 0 \qquad \text{[by (19.31)]}$$

the two roots turn out to be complex:

$$r_1, r_2 = \frac{1}{2}\left(-\frac{3}{2} \pm \sqrt{\frac{9}{4} - \frac{9}{2}}\right) = -\frac{3}{4} \pm \frac{3}{4}i \qquad \left(\text{with } h = -\frac{3}{4} \text{ and } v = \frac{3}{4}\right) \tag{19.33}$$

Substitution of the two roots (along with the parameter values) into (19.30) yields, respectively, the matrix equations

$$\begin{bmatrix} -\frac{3}{4}(1-i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1+i) \end{bmatrix} \begin{bmatrix} m_1 \\ n_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{from } r_1 = -\frac{3}{4} + \frac{3}{4}i\right] \tag{19.34}$$

$$\begin{bmatrix} -\frac{3}{4}(1+i) & \frac{9}{4} \\ -\frac{1}{2} & \frac{3}{4}(1-i) \end{bmatrix} \begin{bmatrix} m_2 \\ n_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \left[\text{from } r_2 = -\frac{3}{4} - \frac{3}{4}i\right] \tag{19.34'}$$

Since r_1 and r_2 are designed—via (19.31)—to make the coefficient matrix singular, each of the preceding two matrix equations actually contains only one independent equation, which can determine only a proportionality relation between the arbitrary constants m_i and n_i. Specifically, we have

$$\frac{1}{3}(1-i)m_1 = n_1 \qquad \text{and} \qquad \frac{1}{3}(1+i)m_2 = n_2$$

The complementary functions can, accordingly, be expressed as

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = \begin{bmatrix} m_1 e^{r_1 t} + m_2 e^{r_2 t} \\ n_1 e^{r_1 t} + n_2 e^{r_2 t} \end{bmatrix}$$

$$= e^{ht}\begin{bmatrix} m_1 e^{vit} + m_2 e^{-vit} \\ n_1 e^{vit} + n_2 e^{-vit} \end{bmatrix} \qquad [\text{by (16.11)}]$$

$$= e^{ht}\begin{bmatrix} (m_1+m_2)\cos vt + (m_1-m_2)i\sin vt \\ (n_1+n_2)\cos vt + (n_1-n_2)i\sin vt \end{bmatrix} \qquad [\text{by (16.24)}]$$

If, for notational simplicity, we define new arbitrary constants

$$A_5 \equiv m_1 + m_2 \quad \text{and} \quad A_6 \equiv (m_1 - m_2)i$$

it then follows that†

$$n_1 + n_2 = \frac{1}{3}(A_5 - A_6) \qquad (n_1 - n_2)i = \frac{1}{3}(A_5 + A_6)$$

So, using these, and incorporating the h and v values of (19.33) into the complementary functions, we end up with

$$\begin{bmatrix} \pi_c \\ U_c \end{bmatrix} = e^{-3t/4}\begin{bmatrix} A_5\cos\frac{3}{4}t + A_6\sin\frac{3}{4}t \\ \frac{1}{3}(A_5 - A_6)\cos\frac{3}{4}t + \frac{1}{3}(A_5+A_6)\sin\frac{3}{4}t \end{bmatrix} \tag{19.35}$$

Finally, by combining the particular integrals in (19.32) with the above complementary functions, we can obtain the solution paths of π and U. As may be expected, these paths are exactly the same as those in (16.43) and (16.45) in Sec. 16.5.

Simultaneous Difference Equations

The simultaneous-equation treatment of the inflation-unemployment model in discrete time is similar in spirit to the preceding continuous-time discussion. We shall thus merely give the highlights.

† This can be seen from the following:

$$n_1 + n_2 = \frac{1}{3}(1-i)m_1 + \frac{1}{3}(1+i)m_2 = \frac{1}{3}[(m_1+m_2) - (m_1 - m_2)i]$$

$$= \frac{1}{3}(A_5 - A_6)$$

$$(n_1 - n_2)i = \left[\frac{1}{3}(1-i)m_1 - \frac{1}{3}(1+i)m_2\right]i = \frac{1}{3}[(m_1 - m_2) - (m_1+m_2)i]i$$

$$= \frac{1}{3}(A_6 + A_5) \qquad [i^2 = -1]$$

The model in question, as given in Sec. 18.3, consists of three equations, two of which describe the patterns of change of π and U, respectively:

$$p_t = \alpha - T - \beta U_t + g\pi_t \tag{18.18}$$

$$\pi_{t+1} - \pi_t = j(p_t - \pi_t) \tag{18.19}$$

$$U_{t+1} - U_t = -k(\mu - p_{t+1}) \tag{18.20}$$

Eliminating p_t and collecting terms, we can rewrite the model as the difference-equation system

$$\underbrace{\begin{bmatrix} 1 & 0 \\ -kg & 1+\beta k \end{bmatrix}}_{J}\begin{bmatrix} \pi_{t-1} \\ U_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} -(1-j+jg) & j\beta \\ 0 & -1 \end{bmatrix}}_{K}\begin{bmatrix} \pi_t \\ U_t \end{bmatrix}$$

$$= \begin{bmatrix} j(\alpha - T) \\ k(\alpha - T - \mu) \end{bmatrix} \tag{19.36}$$

Solution Paths

If stationary equilibriums exist, the particular solutions of (19.36) can be expressed as $\overline{\pi} = \pi_t = \pi_{t+1}$ and $\overline{U} = U_t = U_{t+1}$. Substituting $\overline{\pi}$ and $\overline{U}$ into (19.36), and solving the system (by matrix inversion or Cramer's rule), we obtain

$$\overline{\pi} = \mu \qquad \text{and} \qquad \overline{U} = \frac{1}{\beta}[\alpha - T - (1-g)\mu] \tag{19.37}$$

The $\overline{U}$ value is the same as what was found in Sec. 18.3. Although we did not find $\overline{\pi}$ in the latter section, the information in Exercise 18.3-2 indicates that $\overline{\pi} = \mu$, which agrees with (19.37). In fact, you may note, the results in (19.37) are also identical with the intertemporal equilibrium values obtained in the continuous-time framework in (19.29).

The search for the complementary functions, based this time on the trial solutions mb^t and nb^t, involves the reduced matrix equation

$$(bJ + K)\begin{bmatrix} m \\ n \end{bmatrix} = 0$$

or, in view of (19.36),

$$\begin{bmatrix} b-(1-j+jg) & j\beta \\ -bkg & b(1+\beta k)-1 \end{bmatrix}\begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{19.38}$$

In order to avoid trivial solutions from this homogeneous system, we require

$$|bJ + K| = (1+\beta k)b^2 - [1 + gj + (1-j)(1+\beta k)]b$$
$$+ (1 - j + jg) = 0 \tag{19.39}$$

The normalized version of this quadratic equation is the characteristic equation $b^2 + a_1 b + a_2 = 0$, with the same a_1 and a_2 coefficients as in (18.24) and (18.33) in Sec. 18.3. Consequently, the analysis of the three cases of characteristic roots undertaken in that section should equally apply here.

For each root, b_i, (19.38) supplies us with a specific proportionality relation between the arbitrary constants m_i and n_i, and these enable us to link the arbitrary constants in the complementary function for U to those in the complementary function for π. Then, by combining the complementary functions and the particular solutions, we can get the time paths of π and U.

EXERCISE 19.4

1. Verify (19.29) by using Cramer's rule.
2. Verify that the same proportionality relation between m_1 and n_1 emerges whether we use the first or the second equation in the system (19.34).
3. Find the time paths (general solutions) of π and U, given:

$$p = \frac{1}{6} - 2U + \frac{1}{3}\pi$$
$$\pi' = \frac{1}{4}(p - \pi)$$
$$U' = -\frac{1}{2}(\mu - p)$$

4. Find the time paths (general solutions) of π and U, given:

(a)
$$p_t = \frac{1}{2} - 3U_t + \frac{1}{2}\pi_t$$
$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$
$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

(b)
$$p_t = \frac{1}{4} - 4U_t + \pi_t$$
$$\pi_{t+1} - \pi_t = \frac{1}{4}(p_t - \pi_t)$$
$$U_{t+1} - U_t = -(\mu - p_{t+1})$$

19.5 Two-Variable Phase Diagrams

The preceding sections have dealt with the *quantitative* solutions of *linear* dynamic systems. In the present section, we shall discuss the *qualitative-graphic* (phase-diagram) analysis of a *nonlinear* differential-equation system. More specifically, our attention will be focused on the first-order differential-equation system in two variables, in the general form of

$$x'(t) = f(x, y)$$
$$y'(t) = g(x, y)$$

Note that the time derivatives $x'(t)$ and $y'(t)$ depend only on x and y and that the variable t does not enter into the f and g functions as a separate argument. This feature, which makes the system an *autonomous system*, is a prerequisite for the application of the phase-diagram technique.[†]

The two-variable phase diagram, like the one-variable version in Sec. 15.6, is limited in that it can answer only qualitative questions—those concerning the location and the dynamic stability of the intertemporal equilibrium(s). But, again like the one-variable version, it has the compensating advantages of being able to handle nonlinear systems as comfortably as linear ones and to address problems couched in terms of general functions as readily as those in terms of specific ones.

[†] In the one-variable phase diagram introduced earlier in Sec. 15.6, the equation $dy/dt = f(y)$ is also restricted to be *autonomous*, being forbidden to have the variable t as an explicit argument in the function f.

The Phase Space

When constructing the one-variable phase diagram (Fig. 15.3) for the (autonomous) differential equation $dy/dt = f(y)$, we simply plotted dy/dt against y on the two axes in a two-dimensional phase space. Now that the number of variables is *doubled*, however, how can we manage to meet the apparent need for more axes? The answer, fortunately, is that the 2-space is all we need.

To see why this is feasible, observe that the most crucial task of phase-diagram construction is to determine the direction of movement of the variable(s) over time. It is this information, as embodied in the arrowheads in Fig. 15.3, that enables us to derive the final qualitative inferences. For the drawing of the said arrowheads, only two things are required: (1) a demarcation line—call it the "$dy/dt = 0$" line—that provides the locale for any prospective equilibrium(s) and, more importantly, separates the phase space into two regions, one characterized by $dy/dt > 0$ and the other by $dy/dt < 0$ and (2) a real line on which the increases and decreases of y that are implied by any nonzero values of dy/dt can be indicated. In Fig. 15.3, the demarcation line cited in item 1 is found in the horizontal axis. But that axis actually also serves as the real line cited in item 2. This means that the vertical axis, for dy/dt, can actually be given up without loss, provided we take care to distinguish between the $dy/dt > 0$ region and the $dy/dt < 0$ region—say, by labeling the former with a plus sign, and the latter with a minus sign. This dispensability of one axis is what makes feasible the placement of a two-variable phase diagram in the 2-space. We now need *two* real lines instead of one. But this is automatically taken care of by the standard x and y axes of a two-dimensional diagram. We now also need *two* demarcation lines (or curves), one for $dx/dt = 0$ and the other for $dy/dt = 0$. But these are both graphable in a two-dimensional phase space. And once these are drawn, it would not be difficult to decide which sides of these lines or curves should be marked with plus and minus signs, respectively.

The Demarcation Curves

Given the following autonomous differential-equation system

$$\begin{aligned} x' &= f(x, y) \\ y' &= g(x, y) \end{aligned} \tag{19.40}$$

where x' and y' are short for the time derivatives $x'(t)$ and $y'(t)$, respectively, the two demarcation curves—to be denoted by $x' = 0$ and $y' = 0$—represent the graphs of the two equations

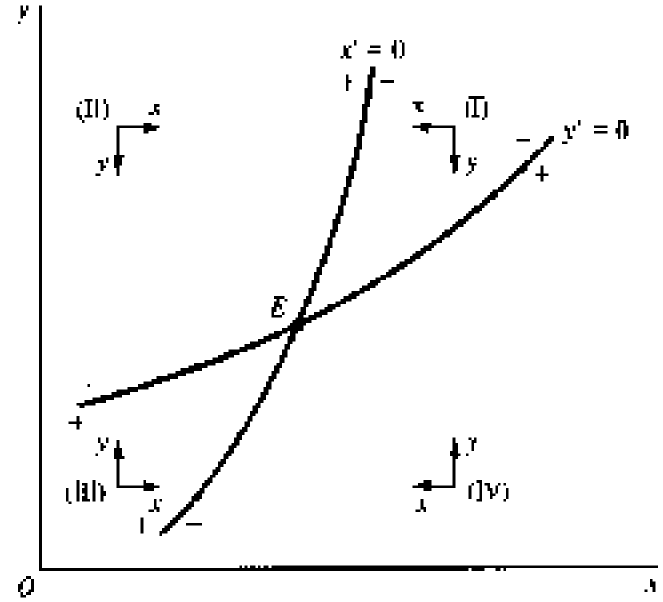
$$f(x, y) = 0 \qquad [x' = 0 \text{ curve}] \tag{19.41}$$

$$g(x, y) = 0 \qquad [y' = 0 \text{ curve}] \tag{19.42}$$

If the specific form of the f function is known, (19.41) can be solved for y in terms of x and the solution plotted in the xy plane as the $x' = 0$ curve. Even if not, however, we can nonetheless resort to the implicit-function rule and ascertain the slope of the $x' = 0$ curve to be

$$\left.\frac{dy}{dx}\right|_{x'=0} = -\frac{\partial f/\partial x}{\partial f/\partial y} = -\frac{f_x}{f_y} \qquad (f_y \neq 0) \tag{19.43}$$

FIGURE 19.1



As long as the signs of the partial derivatives f_x and f_y ($\neq 0$) are known, a qualitative clue to the slope of the $x' = 0$ curve is available from (19.43). By the same token, the slope of the $y' = 0$ curve can be inferred from the derivative

$$\left.\frac{dy}{dx}\right|_{y'=0} = -\frac{g_x}{g_y} \qquad (g_y \neq 0) \tag{19.44}$$

For a more concrete illustration, let us assume that

$$f_x < 0 \qquad f_y > 0 \qquad g_x > 0 \qquad \text{and} \qquad g_y < 0 \tag{19.45}$$

Then both the $x' = 0$ and $y' = 0$ curves will be positively sloped. If we further assume that

$$-\frac{f_x}{f_y} > -\frac{g_x}{g_y} \qquad [x' = 0 \text{ curve steeper than } y' = 0 \text{ curve}]$$

then we may encounter a situation such as that shown in Fig. 19.1. Note that the demarcation lines are now possibly curved. Note, also, that they are now no longer required to coincide with the axes.

The two demarcation curves, intersecting at point E, divide the phase space into four distinct regions, labeled I through IV. Point E, where x and y are both stationary ($x' = y' = 0$), represents the intertemporal equilibrium of the system. At any other point, however, either x or y (or both) would be changing over time, in directions dictated by the signs of the time derivatives x' and y' at that point. In the present instance, we happen to have $x' > 0$ ($x' < 0$) to the left (right) of the $x' = 0$ curve; hence the plus (minus) signs on the left (right) of that curve. These signs are based on the fact that

$$\frac{\partial x'}{\partial x} = f_x < 0 \qquad [\text{by (19.40) and (19.45)}] \tag{19.46}$$

which implies that, as we move continually from west to east in the phase space (as x increases), x' undergoes a steady decrease, so that the sign of x' must pass through three stages, in the order $+$, 0, $-$. Analogously, the derivative

$$\frac{\partial y'}{\partial y} = g_y < 0 \qquad [\text{by (19.40) and (19.45)}] \tag{19.47}$$

implies that, as we move continually from south to north (as y increases), y' steadily decreases, so that the sign of y' must pass through three stages, in the order $+$, 0, $-$. Thus we are led to append the plus signs below, and the minus signs above, the $y' = 0$ curve in Fig. 19.1.

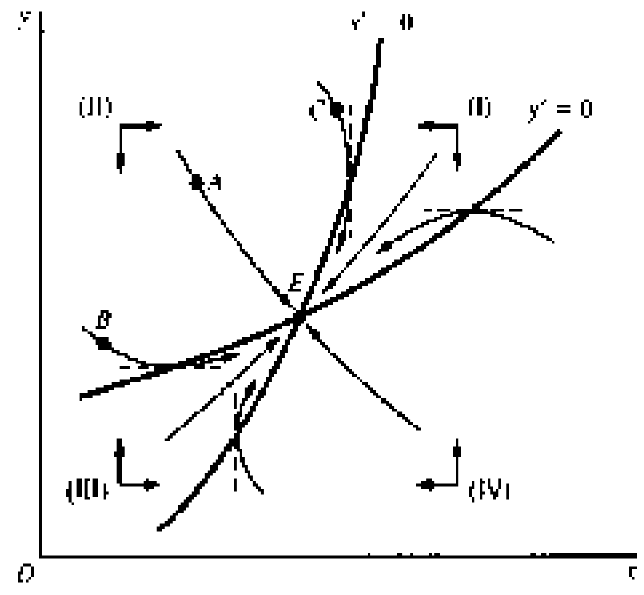
On the basis of these plus and minus signs, a set of directional arrows can now be drawn to indicate the intertemporal movement of x and y. For any point in region I, x' and y' are both negative. Hence x and y must both decrease over time, producing a *westward* movement for x, and a *southward* movement for y. As indicated by the two arrows in region I, given an initial point located in region I, the intertemporal movement must be in the general southwestward direction. The exact opposite is true in region III, where x' and y' are both positive, so that both the x and y variables must increase over time. In contrast, x' and y' have different signs in region II. With x' positive and y' negative, x should move *eastward* and y *southward*. And region IV displays a tendency exactly opposite to region II.

Streamlines

For a better grasp of the implications of the directional arrows, we can sketch a series of *streamlines* in the phase diagram. Also referred to as *phase trajectories* (or *trajectories* for short) or *phase paths*, these streamlines serve to map out the dynamic movement of the system from any conceivable initial point. A few of these are illustrated in Fig. 19.2, which reproduces the $x' = 0$ and $y' = 0$ curves in Fig 19.1. Since every point in the phase space must be located on one streamline or another, there should exist an infinite number of streamlines, all of which conform to the directional requirements imposed by the xy arrows in every region. For depicting the general qualitative character of the phase diagram, however, a few representative streamlines should normally suffice.

Several features may be noted about the streamlines in Fig. 19.2. First, all of them happen to lead toward point E. This makes E a stable (here, globally stable) intertemporal equilibrium. Later, we shall encounter other types of streamline configurations. Second, while some streamlines never venture beyond a single region (such as the one passing through point A), others may cross over from one region into another (such as those passing through B and C). Third, where a streamline crosses over, it must have either an infinite slope (crossing the $x' = 0$ curve) or a zero slope (crossing the $y' = 0$ curve). This is due to the

FIGURE 19.2



fact that, along the $x' = 0$ ($y' = 0$) curve, $x(y)$ is stationary over time, so the streamline must not have any horizontal (vertical) movement while crossing that curve. To ensure that these slope requirements are consistently met, it would be advisable, as soon as the demarcation curves have been put in place, to add a few short *vertical* sketching bars across the $x' = 0$ curve and a few *horizontal* ones across the $y' = 0$ curve, as guidelines for the drawing of the streamlines.† Fourth, and last, although the streamlines do explicitly point out the directions of movement of x and y over time, they provide no specific information regarding velocity and acceleration, because the phase diagram does not allow for an axis for t (time). It is for this reason, of course, that streamlines carry the alternative name of *phase paths*, as opposed to *time* paths. The only observation we can make about velocity is qualitative in nature: As we move along a streamline closer and closer to the $x' = 0$ ($y' = 0$) curve, the velocity of approach in the horizontal (vertical) direction must progressively diminish. This is due to the steady decrease in the absolute value of the derivative $x' \equiv dx/dt$ ($y' \equiv dy/dt$) that occurs as we move toward the demarcation line on which x' (y') takes a zero value.

Types of Equilibrium

Depending on the configurations of the streamlines surrounding a particular intertemporal equilibrium, that equilibrium may fall into one of four categories: (1) nodes, (2) saddle points, (3) foci or focuses, and (4) vortices or vortexes.

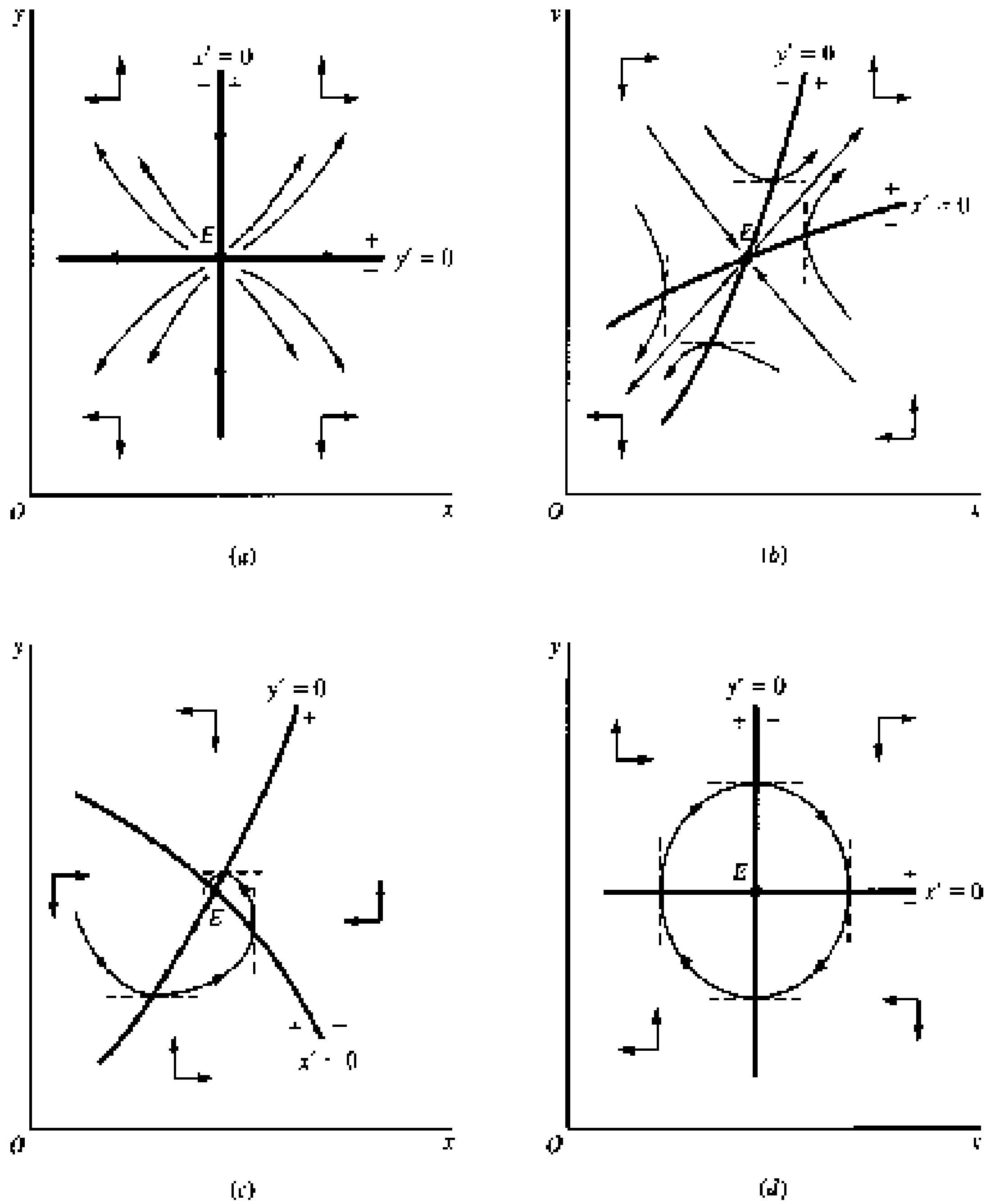
A *node* is an equilibrium such that all the streamlines associated with it either flow noncyclically toward it (*stable node*) or flow noncyclically away from it (*unstable node*). We have already encountered a stable node in Fig. 19.2. An unstable node is shown in Fig. 19.3*a*. Note that in this particular illustration, it happens that the streamlines never cross over from region to region. Also, the $x' = 0$ and $y' = 0$ curves happen to be linear, and, in fact, they themselves serve as streamlines.

A *saddle point* is an equilibrium with a double personality—it is stable in some directions, but unstable in others. More accurately, with reference to the illustration in Fig. 19.3*b*, a saddle point has exactly one pair of streamlines—called the *stable branches* of the saddle point—that flow directly and consistently toward the equilibrium, and exactly one pair of streamlines—the *unstable branches*—that flow directly and consistently away from it. All the other trajectories head toward the saddle point initially but sooner or later turn away from it. This double personality, of course, is what inspired the name "saddle point." Since stability is observed only on the stable branches, which are not reachable as a matter of course, a saddle point is generically classified as an *unstable* equilibrium.

The third type of equilibrium, *focus*, is one characterized by whirling trajectories, all of which either flow cyclically toward it (*stable focus*), or flow cyclically away from it (*unstable focus*). Figure 19.3*c* illustrates a stable focus, with only one streamline explicitly drawn in order to avoid clutter. What causes the whirling motion to occur? The answer lies in the way the $x' = 0$ and $y' = 0$ curves are positioned. In Fig. 19.3*c*, the two demarcation curves are sloped in such a way that they take turns in blockading the streamline flowing in a direction prescribed by a particular set of xy arrows. As a result, the streamline is frequently compelled to cross over from one region into another, tracing out a spiral. Whether we get

† To aid your memory, note that the sketching bars across the $x' = 0$ curve should be *perpendicular* to the x axis. Similarly, the sketching bars across the $y' = 0$ curve should be *perpendicular* to the y axis.

FIGURE 19.3



a stable focus (as is the case here) or an unstable one depends on the relative placement of the two demarcation curves. But in either case, the slope of the streamline at the crossover points must still be either infinite (crossing $x' = 0$) or zero (crossing $y' = 0$).

Finally, we may have a *vortex* (or *center*). This is again an equilibrium with whirling streamlines, but these streamlines now form a family of loops (concentric circles or ovals) orbiting around the equilibrium in a perpetual motion. An example of this is given in Fig. 19.3*d*, where, again, only a single streamline is shown. Inasmuch as this type of equilibrium is unattainable from any initial position away from point E, a vortex is automatically classified as an unstable equilibrium.

All the illustrations in Fig. 19.3 display a unique equilibrium. When sufficient nonlinearity exists, however, the two demarcation curves may intersect more than once, thereby producing multiple equilibria. In that event, a combination of the previously cited types of intertemporal equilibrium may exist in the same phase diagram. Although there will then be more than four regions to contend with, the underlying principle of phase-diagram analysis will remain basically the same.

Inflation and Monetary Rule à la Obst

As an economic illustration of the two-variable phase diagram, we shall present a model due to Professor Obst,† which purports to show the ineffectiveness of the conventional (hence the need for a new) type of countercyclical monetary-policy rule, when an "inflation adjustment mechanism" is at work. Such a model contrasts with our earlier discussion of inflation in that, instead of studying the implications of a *given* rate of monetary expansion, it looks further into the efficacy of two different monetary *rules*, each prescribing a different set of monetary actions to be pursued in the face of various inflationary conditions.

A crucial assumption of the model is the inflation adjustment mechanism

$$\frac{dp}{dt} = h\left(\frac{M_s - M_d}{M_s}\right) = h\left(1 - \frac{M_d}{M_s}\right) \qquad (h > 0) \qquad \textbf{(19.48)}$$

which shows that the effect of an excess supply of money ($M_s > M_d$) is to raise the rate of inflation p, rather than the price level P. The clearance of the money market would thus imply not price stability, but only a stable rate of inflation. To facilitate the analysis, the second equality in (19.48) serves to shift the focus from the excess supply of money to the demand-supply ratio of money, M_d/M_s, which we shall denote by μ. On the assumption that M_d is directly proportional to the nominal national product PQ, we can write

$$\mu \equiv \frac{M_d}{M_s} = \frac{aPQ}{M_s} \qquad (a > 0)$$

The rates of growth of the several variables are then related by

$$\begin{aligned}\frac{d\mu/dt}{\mu} &= \frac{da/dt}{a} + \frac{dP/dt}{P} + \frac{dQ/dt}{Q} - \frac{dM_s/dt}{M_s} \\ & \qquad\qquad\qquad\qquad [\text{by (10.24) and (10.25)}] \\ &\equiv p + q - m \qquad\qquad [a = \text{a constant}] \qquad \textbf{(19.49)}\end{aligned}$$

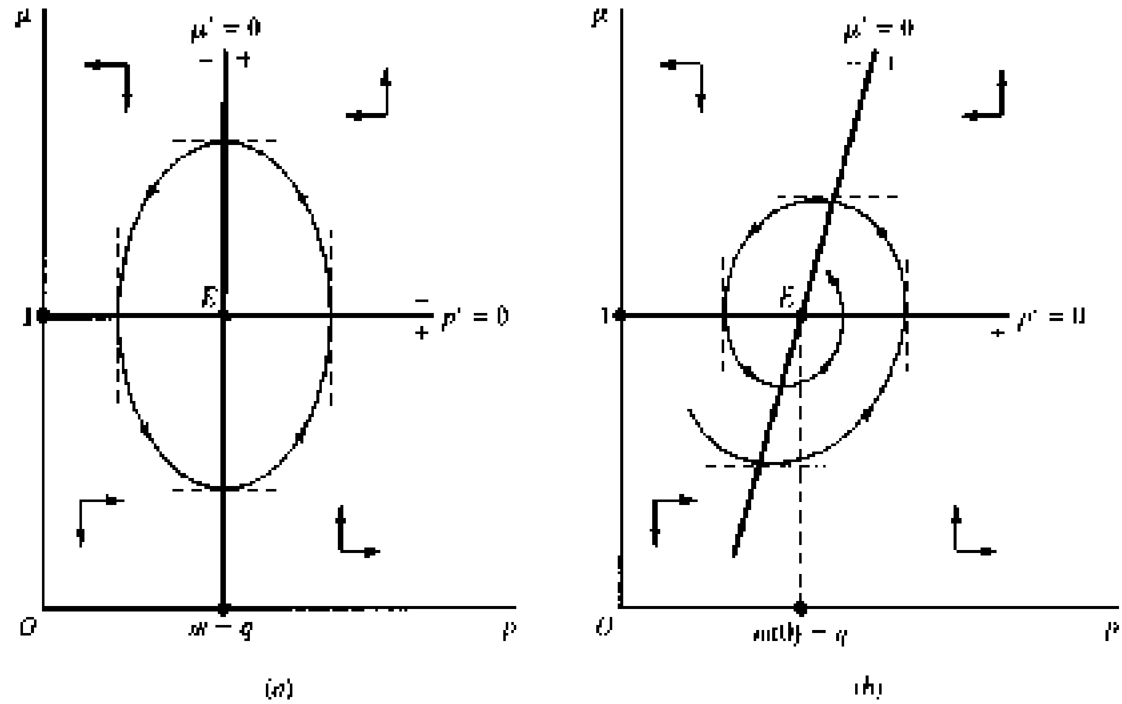
where the lowercase letters p, q, and m denote, respectively, the rate of inflation, the (exogenous) rate of growth of the real national product, and the rate of monetary expansion.

Equations (19.48) and (19.49), a set of two differential equations, can jointly determine the time paths of p and μ, if, for the time being, m is taken to be exogenous. Using the symbols p' and μ' to represent the time derivatives $p'(t)$ and $\mu'(t)$, we can express this system more concisely as

$$\begin{aligned} p' &= h(1-\mu) \\ \mu' &= (p + q - m)\mu \end{aligned} \qquad \textbf{(19.50)}$$

† Norman P. Obst, "Stabilization Policy with an Inflation Adjustment Mechanism," *Quarterly Journal of Economics*, May 1978, pp. 355–359. No phase diagrams are given in the Obst paper, but they can be readily constructed from the model.

FIGURE 19.4



Given that h is positive, we can have $p' = 0$ if and only if $1 - \mu = 0$. Similarly, since μ is always positive, $\mu' = 0$ if and only if $p + q - m = 0$. Thus the $p' = 0$ and $\mu' = 0$ demarcation curves are associated with the equations

$$\mu = 1 \qquad [p' = 0 \text{ curve}] \tag{19.51}$$

$$p = m - q \qquad [\mu' = 0 \text{ curve}] \tag{19.52}$$

As shown in Fig. 19.4a, these plot as a horizontal line and a vertical line, respectively, and yield a unique equilibrium at E. The equilibrium value $\overline{\mu} = 1$ means that in equilibrium M_d and M_s are equal, clearing the money market. The fact that the equilibrium rate of inflation is shown to be positive reflects an implicit assumption that $m > q$.

Since the $p' = 0$ curve corresponds to the $x' = 0$ curve in our previous discussion, it should have vertical sketching bars. And the other curve should have horizontal ones. From (19.50), we find that

$$\frac{\partial p'}{\partial \mu} = -h < 0 \qquad \text{and} \qquad \frac{\partial \mu'}{\partial p} = \mu > 0 \tag{19.53}$$

with the implication that a northward movement across the $p' = 0$ curve passes through the $(+, 0, -)$ sequence of signs for p', and an eastward movement across the $\mu' = 0$ curve, the $(-, 0, +)$ sequence of signs for μ'. Thus we obtain the four sets of directional arrows as drawn, which generate streamlines (only one of which is shown) that orbit counterclockwise around point E. This, of course, makes E a vortex. Unless the economy happens initially to be at E, it is impossible to attain equilibrium. Instead, there will be never-ceasing fluctuation.

The preceding conclusion is, however, the consequence of an *exogenous* rate of monetary expansion. What if we now endogenize m by adopting an anti-inflationary monetary rule? The "conventional" monetary rule would call for gearing the rate of monetary expansion negatively to the rate of inflation:

$$m = m(p) \qquad m'(p) < 0 \qquad [\text{conventional monetary rule}] \tag{19.54}$$

Such a rule would modify the second equation in (19.50) to

$$\mu' = [p + q - m(p)]\mu \tag{19.55}$$

and alter (19.52) to

$$p = m(p) - q \qquad [\mu' = 0 \text{ curve under conventional monetary rule}] \tag{19.56}$$

Given that $m(p)$ is monotonic, there exists only one value of p—say, p_1—that can satisfy this equation. Hence the new $\mu' = 0$ curve must still emerge as a vertical straight line, although with a different horizontal intercept $p_1 = m(p_1) - q$. Moreover, from (19.55), we find that

$$\frac{\partial \mu'}{\partial p} = [1 - m'(p)]\mu > 0 \qquad [\text{by } (19.54)]$$

which is *qualitatively* no different from the derivative in (19.53). It follows that the directional arrows must also remain as they are in Fig. 19.4*a*. In short, we would end up with a vortex as before.

The alternative monetary rule proposed by Obst is to gear m to the *rate of change* (rather than the *level*) of the rate of inflation:

$$m = m(p') \qquad m'(p') < 0 \qquad [\text{alternative monetary rule}] \tag{19.57}$$

Under this rule, (19.55) and (19.56) will become, respectively,

$$\mu' = [p + q - m(p')]\mu \tag{19.58}$$

$$p = m(p') - q \qquad [\mu' = 0 \text{ curve under alternative monetary rule}] \tag{19.59}$$

This time the $\mu' = 0$ curve would become upward-sloping. For, differentiating (19.59) with respect to μ via the chain rule, we have

$$\frac{dp}{d\mu} = m'(p')\frac{dp'}{d\mu} = m'(p')(-h) > 0 \qquad [\text{by } (19.50)]$$

so, by the inverse-function rule, $d\mu/dp$—the slope of the $\mu' = 0$ cruve—is also positive. This new situation is illustrated in Fig. 19.4*b*, where, for simplicity, the $\mu' = 0$ curve is drawn as a straight line, with an arbitrarily assigned slope.† Despite the slope change, the partial derivative

$$\frac{\partial \mu'}{\partial p} = \mu > 0 \qquad [\text{from } (19.58)]$$

is unchanged from (19.53), so the μ arrows should retain their original orientation in Fig. 19.4*a*. The streamlines (only one of which is shown) will now twirl inwardly toward the equilibrium at $\overline{\mu} = 1$ and $\overline{p} = m(0) - q$, where $m(0)$ denotes $m(p')$ evaluated at $p' = 0$. Thus the alternative monetary rule is seen to be capable of converting a vortex into a stable focus, thereby making possible the asymptotic elimination of the perpetual fluctuation in the rate of inflation. Indeed, with a sufficiently flat $\mu' = 0$ curve, it is even possible to turn the vortex into a stable node.

† The slope is inversely proportional to the absolute value of $m'(p')$. The more sensitively the rate of monetary expansion m is made to respond to the rate of change of the rate of inflation p', the flatter the $\mu' = 0$ curve will be in Fig. 19.4*b*.

EXERCISE 19.5

1. Show that the two-variable phase diagram can also be used, if the model consists of a single second-order differential equation, $y''(t) = f(y', y)$, instead of two first-order equations.
2. The plus and minus signs appended to the two sides of the $x' = 0$ and $y' = 0$ curves in Fig. 19.1 are based on the partial derivatives $\partial x'/\partial x$ and $\partial y'/\partial y$, respectively. Can the same conclusions be obtained from the derivatives $\partial x'/\partial y$ and $\partial y'/\partial x$?
3. Using Fig. 19.2, verify that if a streamline does not have an infinite (zero) slope when crossing the $x' = 0$ ($y' = 0$) curve, it will necessarily violate the directional restrictions imposed by the xy arrows.
4. As special cases of the differential-equation system (19.40), assume that
   (a) $f_x = 0 \quad f_y > 0 \quad g_x > 0 \quad$ and $\quad g_y = 0$
   (b) $f_x = 0 \quad f_y < 0 \quad g_x < 0 \quad$ and $\quad g_y = 0$
   For each case, construct an appropriate phase diagram, draw the streamlines, and determine the nature of the equilibrium.
5. (a) Show that it is possible to produce *either* a stable node or a stable focus from the differential-equation system (19.40), if
   $f_x < 0 \quad f_y > 0 \quad g_x < 0 \quad$ and $\quad g_y < 0$
   (b) What special feature(s) in your phase-diagram construction are responsible for the difference in the outcomes (node versus focus)?
6. With reference to the Obst model, verify that if the positively sloped $\mu' = 0$ curve in Fig. 19.4b is made sufficiently flat, the streamlines, although still characterized by crossovers, will converge to the equilibrium in the manner of a node rather than a focus.

19.6 Linearization of a Nonlinear Differential-Equation System

Another qualitative technique of analyzing a *nonlinear* differential-equation system is to draw inferences from the *linear approximation* to that system, to be derived from the Taylor expansion of the given system around its equilibrium.† We learned in Sec. 9.5 that a linear (or even a higher-order polynomial) approximation to an arbitrary function $\phi(x)$ can give us the exact value of $\phi(x)$ at the point of expansion, but will entail progressively larger errors of approximation as we move farther away from the point of expansion. The same is true of the linear approximation to a nonlinear system. At the point of expansion—here, the equilibrium point E—the linear approximation can pinpoint exactly the same equilibrium as the original nonlinear system. And in a sufficiently small neighborhood of E, the linear approximation should have the same general streamline configuration as the original system. As long as we are willing to confine our stability inferences to the immediate neighborhood of the equilibrium, therefore, the linear approximation could serve as an adequate source of information. Such analysis, referred to as *local stability analysis*, can be used

† In the case of multiple equilibria, each equilibrium requires a separate linear approximation.

either by itself, or as a supplement to the phase-diagram analysis. We shall deal with the two-variable case only.

Taylor Expansion and Linearization

Given an arbitrary (successively differentiable) one-variable function $\phi(x)$, the Taylor expansion around a point x_0 gives the series

$$\phi(x) = \phi(x_0) + \phi'(x_0)(x - x_0) + \frac{\phi''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{\phi^{(n)}(x_0)}{n!}(x - x_0)^n + R_n$$

where a polynomial involving various powers of $(x - x_0)$ appears on the right. A similar structure characterizes the Taylor expansion of a function of two variables $f(x, y)$ around any point (x_0, y_0). With two variables in the picture, however, the resulting polynomial would comprise various powers of $(y - y_0)$ as well as $(x - x_0)$—in fact, also the products of these two expressions:

$$\begin{aligned} f(x, y) = f(x_0, y_0) &+ f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &+ \frac{1}{2!}[f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) \\ &+ f_{yy}(x_0, y_0)(y - y_0)^2] + \cdots + R_n \end{aligned} \tag{19.60}$$

Note that the coefficients of the $(x - x_0)$ and $(y - y_0)$ expressions are now the *partial* derivatives of f, all evaluated at the expansion point (x_0, y_0).

From the Taylor series of a function, the linear approximation—or *linearization* for short—is obtained by simply dropping all terms of order higher than one. Thus, for the one-variable case, the linearization is the following linear function of x:

$$\phi(x_0) + \phi'(x_0)(x - x_0)$$

Similarly, the linearization of (19.60) is the following linear function of x and y:

$$f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Besides, by substituting the function symbol g for f in this result, we can also get the corresponding linearization of $g(x, y)$. It follows that, given the nonlinear system

$$\begin{aligned} x' &= f(x, y) \\ y' &= g(x, y) \end{aligned} \tag{19.61}$$

its linearization around the expansion point (x_0, y_0) can be written as

$$\begin{aligned} x' &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ y' &= g(x_0, y_0) + g_x(x_0, y_0)(x - x_0) + g_y(x_0, y_0)(y - y_0) \end{aligned} \tag{19.62}$$

If the specific forms of the functions f and g are known, then $f(x_0, y_0)$, $f_x(x_0, y_0)$, $f_y(x_0, y_0)$ and their counterparts for the g function can all be assigned specific values and the linear system (19.62) solved quantitatively. However, even if the f and g functions are given in *general* forms, qualitative analysis is still possible, provided only that the signs of f_x, f_y, g_x, and g_y are ascertainable.

The Reduced Linearization

For purposes of local stability analysis, the linearization (19.62) can be put into a simpler form. First, since our point of expansion is to be the equilibrium point $(\bar{x}, \bar{y})$, we should replace (x_0, y_0) by $(\bar{x}, \bar{y})$. More substantively, since at the equilibrium point we have $x' = y' = 0$ by definition, it follows that

$$f(\bar{x}, \bar{y}) = g(\bar{x}, \bar{y}) = 0 \qquad \text{[by (19.61)]}$$

so the first term on the right side of each equation in (19.62) can be dropped. Making these changes, then multiplying out the remaining terms on the right of (19.62) and rearranging, we obtain another version of the linearization:

$$\begin{aligned} x' - f_x(\bar{x}, \bar{y})x - f_y(\bar{x}, \bar{y})y &= -f_x(\bar{x}, \bar{y})\bar{x} - f_y(\bar{x}, \bar{y})\bar{y} \\ y' - g_x(\bar{x}, \bar{y})x - g_y(\bar{x}, \bar{y})y &= -g_x(\bar{x}, \bar{y})\bar{x} - g_y(\bar{x}, \bar{y})\bar{y} \end{aligned} \qquad \textbf{(19.63)}$$

Note that, in (19.63), each term on the right of the equals signs represents a constant. We took the trouble to separate out these constant terms so that we can now drop them all, to get to the reduced equations of the linearization. The result, which may be written in matrix notation as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} - \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \textbf{(19.64)}$$

constitutes the *reduced linearization* of (19.61). Inasmuch as qualitative analysis depends exclusively on the knowledge of the characteristic roots, which, in turn, hinge only on the reduced equations of a system, (19.64) is all we need for the desired local stability analysis.

Going a step further, it may be observed that the only distinguishing property of the reduced linearization lies in the matrix of partial derivatives—the Jacobian matrix of the nonlinear system (19.61)—evaluated at the equilibrium $(\bar{x}, \bar{y})$. Hence, in the final analysis, the local stability or instability of the equilibrium is predicated solely on the makeup of the said Jacobian. For notational convenience in the ensuing discussion, we shall denote the Jacobian evaluated at the equilibrium by J_E and its elements by a, b, c, and d:

$$J_E \equiv \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x}, \bar{y})} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \textbf{(19.65)}$$

It will be assumed that the two differential equations are functionally independent. Then we shall always have $|J_E| \neq 0$. (For some cases where $|J_E| = 0$, see Exercise 19.6-4.)

Local Stability Analysis

According to (19.16), and using (19.65), the characteristic equation of the reduced linearization should be

$$\begin{vmatrix} r - a & -b \\ -c & r - d \end{vmatrix} = r^2 - (a + d)r + (ad - bc) = 0$$

It is clear that the characteristic roots depend critically on the expressions $(a + d)$ and $(ad - bc)$. The latter is merely the determinant of the Jacobian in (19.65):

$$ad - bc = |J_E|$$

And the former, representing the *sum* of the principal-diagonal elements of that Jacobian, is called the *trace* of J_E, symbolized by tr J_E:

$$a + d = \text{tr } J_E$$

Accordingly, the characteristic roots can be expressed as

$$r_1, r_2 = \frac{\text{tr } J_E \pm \sqrt{(\text{tr } J_E)^2 - 4|J_E|}}{2}$$

The relative magnitudes of $(\text{tr } J_E)^2$ and $4|J_E|$ will determine whether the two roots are real or complex, that is, whether the time paths of x and y are steady or fluctuating. To check the dynamic stability of equilibrium, on the other hand, we need to ascertain the algebraic signs of the two roots. For that purpose, the following two relationships will prove to be most helpful:

$$r_1 + r_2 = \text{tr } J_E \qquad \text{[cf. (16.5) and (16.6)]} \tag{19.66}$$

$$r_1 r_2 = |J_E| \tag{19.67}$$

Case 1 ($(\text{tr } J_E)^2 > 4|J_E|$) In this case, the roots are real and distinct, and no fluctuation is possible. Hence the equilibrium can be either a node or a saddle point, but never a focus or vortex. In view that $r_1 \neq r_2$, there exist three distinct possibilities of sign combination: both roots negative, both roots positive, and two roots with opposite signs.[1] Taking into account the information in (19.66) and (19.67), these three possibilities are characterized by:

$$\begin{aligned}
(i) \quad & r_1 < 0, r_2 < 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E < 0 \\
(ii) \quad & r_1 > 0, r_2 > 0 \quad \Rightarrow \quad |J_E| > 0; \text{tr } J_E > 0 \\
(iii) \quad & r_1 > 0, r_2 < 0 \quad \Rightarrow \quad |J_E| < 0; \text{tr } J_E \gtreqless 0
\end{aligned}$$

Under Possibility *i*, with both roots negative, both complementary functions x_c and y_c tend to zero as t becomes infinite. The equilibrium is thus a stable node. The opposite is true under Possibility *ii*, which describes an unstable node. In contrast, with two roots of opposite signs, Possibility *iii* yields a saddle point.

To see this last case more clearly, recall that the complementary functions of the two variables under Case 1 take the general form

$$\begin{aligned}
x_c &= A_1 e^{r_1 t} + A_2 e^{r_2 t} \\
y_c &= k_1 A_1 e^{r_1 t} + k_2 A_2 e^{r_2 t}
\end{aligned}$$

where the arbitrary constants A_1 and A_2 are to be determined from the initial conditions. If the initial conditions are such that $A_1 = 0$, the positive root r_1 will drop out of the picture, leaving it to the negative root r_2 to make the equilibrium stable. Such initial conditions pertain to the points located on the stable branches of the saddle point. On the other hand, if the initial conditions are such that $A_2 = 0$, the negative root r_2 will vanish from the scene, leaving it to the positive root r_1 to make the equilibrium unstable. Such initial conditions relate to the points lying on the unstable branches. Inasmuch as all the other initial conditions also involve $A_1 \neq 0$, they must all give rise to divergent complementary functions, too. Thus Possibility *iii* yields a saddle point.

[1] Since we have ruled out $|J_E| = 0$, no root can take a zero value.

Case 2 $(\text{tr } J_E)^2 = 4|J_E|$ As the roots are repeated in this case, only two possibilities of sign combination can arise:

$$(iv) \quad r_1 < 0, r_2 < 0 \quad \Rightarrow \quad |J_E| > 0; \text{ tr } J_E < 0$$
$$(v) \quad r_1 > 0, r_2 > 0 \quad \Rightarrow \quad |J_E| > 0; \text{ tr } J_E > 0$$

These two possibilities are mere duplicates of Possibilities *i* and *ii*. Thus they point to a stable node and an unstable node, respectively.

Case 3 $(\text{tr } J_E)^2 < 4|J_E|$ This time, with complex roots $h \pm vi$, cyclical fluctuation is present, and we must encounter either a focus or a vortex. On the basis of (19.66) and (19.67), we have in the present case

$$\text{tr } J_E = r_1 + r_2 = (h + vi) + (h - vi) = 2h$$
$$|J_E| = r_1 r_2 = (h + vi)(h - vi) = h^2 + v^2$$

Thus tr J_E has to take the same sign as h, whereas $|J_E|$ is invariably positive. Consequently, there are three possible outcomes:

$$(vi) \quad h < 0 \quad \Rightarrow \quad |J_E| > 0; \text{ tr } J_E < 0$$
$$(vii) \quad h > 0 \quad \Rightarrow \quad |J_E| > 0; \text{ tr } J_E > 0$$
$$(viii) \quad h = 0 \quad \Rightarrow \quad |J_E| > 0; \text{ tr } J_E = 0$$

These are associated, respectively, with damped fluctuation, explosive fluctuation, and uniform fluctuation. In other word, Possibility *vi* implies a stable focus; Possibility *vii*, an unstable focus; and Possibilty *viii*, a vortex.

The conclusions from the preceding discussion are summarized in Table 19.1 to facilitate qualitative inferences from the signs of $|J_E|$ and tr J_E. Three features of the table are especially noteworthy. First, a negative $|J_E|$ is exclusively tied to the saddle-point type of equilibrium. This suggests that $|J_E| < 0$ is a necessary-and-sufficient condition for a saddle point. Second, a zero value for tr J_E occurs only under two circumstances—when there is a saddle point or a vortex. These two circumstances are, however, distinguishable from each other by the sign of $|J_E|$. Accordingly, a zero tr J_E coupled with a positive $|J_E|$ is necessary-and-sufficient for a vortex. Third, while a negative sign for tr J_E is necessary for dynamic stability, it is *not* sufficient, on account of the possibility of a saddle point.

TABLE 19.1
Local Stability Analysis of a Two-Variable Nonlinear Differential-Equation System

Case	Sign of $\|J_E\|$	Sign of tr J_E	Type of Equilibrium
1. $(\text{tr } J_E)^2 > 4\|J_E\|$	+	−	Stable node
	+	+	Unstable node
	−	+, 0, −	Saddle point
2. $(\text{tr } J_E)^2 = 4\|J_E\|$	+	−	Stable node
	+	+	Unstable node
3. $(\text{tr } J_E)^2 < 4\|J_E\|$	+	−	Stable focus
	+	+	Unstable focus
	+	0	Vortex

Nevertheless, when a negative tr J_E is accompanied by a positive $|J_E|$, we do have a necessary-and-sufficient condition for dynamic stability.

The discussion leading to the summary in Table 19.1 has been conducted in the context of a linear approximation to a *nonlinear* system. However, the contents of that table are obviously applicable also to the qualitative analysis of a system that is *linear* to begin with. In the latter case, the elements of the Jacobian matrix will be a set of given constants, so there is no need to evaluate them at the equilibrium. Since there is no approximation process involved, the stability inferences will no longer be "local" in nature but will have global validity.

Example 1

Analyze the local stability of the nonlinear system

$$\begin{aligned} x' &= f(x, y) = xy - 2 \\ y' &= g(x, y) = 2x - y \end{aligned} \qquad (x, y \geq 0)$$

First, setting $x' = y' = 0$, and noting the nonnegativity of x and y, we find a single equilibrium E at $(\bar{x}, \bar{y}) = (1, 2)$. Then, by taking the partial derivatives of x' and y', and evaluating them at E, we obtain

$$J_E = \begin{bmatrix} f_x & f_y \\ g_x & g_y \end{bmatrix}_{(\bar{x},\bar{y})} = \begin{bmatrix} y & x \\ 2 & -1 \end{bmatrix}_{(1,2)} = \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}$$

Since $|J_E| = -4$ is negative, we can immediately conclude that the equilibrium is locally a saddle point.

Note that while the first row of the Jacobian matrix originally contains the variables y and x, the second row does not. The reason for the difference is that the second equation in the given system is originally linear, and requires no linearization.

Example 2

Given the nonlinear system

$$\begin{aligned} x' &= x^2 - y \\ y' &= 1 - y \end{aligned}$$

we can, by setting $x' = y' = 0$, find two equilibrium points: $E_1 = (1, 1)$ and $E_2 = (-1, 1)$. Thus we need two separate linearizations. Evaluating the Jacobian $\begin{bmatrix} 2x & -1 \\ 0 & -1 \end{bmatrix}$ at the two equilibriums in turn, we obtain

$$J_{E1} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix} \qquad \text{and} \qquad J_{E2} = \begin{bmatrix} -2 & -1 \\ 0 & -1 \end{bmatrix}$$

The first of these has a negative determinant; thus $E_1 = (1, 1)$ is locally a saddle point. From the second, we find that $|J_{E2}| = 2$ and tr $J_{E2} = -3$. Hence, by Table 19.1, $E_2 = (-1, 1)$ is locally a stable node under Case 1.

Example 3

Does the linear system

$$\begin{aligned} x' &= x - y + 2 \\ y' &= x + y + 4 \end{aligned}$$

possess a stable equilibrium? To answer such a qualitative question, we can simply concentrate on the reduced equations and ignore the constants 2 and 4 altogether. As may be expected from a linear system, the Jacobian $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ has as its elements four constants.

Inasmuch as its determinant and trace are both equal to 2, the equilibrium falls under Case 3 and is an unstable focus. Note that this conclusion is reached without having to solve for the equilibrium. Note, also, that the conclusion is in this case globally valid.

Example 4 Analyze the local stability of the Obst model (19.50),

$$p' = h(1-\mu)$$
$$\mu' = (p+q-m)\mu$$

assuming that the rate of monetary expansion m is exogenous (no monetary rule is followed). According to Fig. 19.4a, the equilibrium of this model occurs at $E = (\bar{p}, \bar{\mu}) = (m-q, 1)$. The Jacobian matrix evaluated at E is

$$J_E = \begin{bmatrix} \dfrac{\partial p'}{\partial p} & \dfrac{\partial p'}{\partial \mu} \\ \dfrac{\partial \mu'}{\partial p} & \dfrac{\partial \mu'}{\partial \mu} \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m \end{bmatrix}_{(m-q,1)} = \begin{bmatrix} 0 & -h \\ 1 & 0 \end{bmatrix}$$

Since $|J_E| = h > 0$, and $\text{tr}\, J_E = 0$, Table 19.1 indicates that the equilibrium is locally a vortex. This conclusion is consistent with that of the phase-diagram analysis in Sec. 19.5.

Example 5 Analyze the local stability of the Obst model, assuming that the alternative monetary rule is as follows:

$$p' = h(1-\mu) \qquad \text{[from (19.50)]}$$
$$\mu' = [p+q-m(p')]\mu \qquad \text{[from (19.58)]}$$

Note that since p' is a function of μ, the function $m(p')$ is in the present model also a function of μ. Thus we have to apply the product rule in finding $\partial\mu'/\partial\mu$. At the equilibrium E, where $p' = \mu' = 0$, we have $\bar{\mu} = 1$ and $\bar{p} = m(0) - q$. The Jacobian evaluated at E is, therefore,

$$J_E = \begin{bmatrix} 0 & -h \\ \mu & p+q-m(p') - m'(p')(-h)\mu \end{bmatrix}_E = \begin{bmatrix} 0 & -h \\ 1 & m'(0)h \end{bmatrix}$$

where $m'(0)$ is negative by (19.57). According to Table 19.1, with $|J_E| = h > 0$ and $\text{tr}\, J_E = m'(0)h < 0$, we can have either a stable focus or a stable node, depending on the relative magnitudes of $(\text{tr}\, J_E)^2$ and $4|J_E|$. To be specific, the larger the absolute value of the derivative $m'(0)$, the larger the absolute value of $\text{tr}\, J_E$ will be and the more likely $(\text{tr}\, J_E)^2$ will exceed $4|J_E|$, to produce a stable node instead of a stable focus. This conclusion is again consistent with what we learned from the phase-diagram analysis.

EXERCISE 19.6

1. Analyze the local stability of each of the following nonlinear systems:

(a) $x' = e^x - 1$
$y' = ye^x$

(b) $x' = x + 2y$
$y' = x^2 + y$

(c) $x' = 1 - e^y$
$y' = 5x - y$

(d) $x' = x^3 + 3x^2y + y$
$y' = x(1 + y^2)$

2. Use Table 19.1 to determine the type of equilibrium a nonlinear system would have locally, given that:

   (a) $f_x = 0$ $f_y > 0$ $g_x > 0$ and $g_y = 0$

   (b) $f_x = 0$ $f_y < 0$ $g_x < 0$ and $g_y = 0$

   (c) $f_x < 0$ $f_y > 0$ $g_x < 0$ and $g_y < 0$

   Are your results consistent with your answers to Exercises 19.5-4 and 19.5-5?

3. Analyze the local stability of the Obst model, assuming that the conventional monetary rule is followed.

4. The following two systems both possess zero-valued Jacobians. Construct a phase diagram for each, and deduce the locations of all the equilibriums that exist:

   (a) $x' = x + y$ (b) $x' = 0$

   $y' = -x - y$ $y' = 0$

Chapter 20

Optimal Control Theory

At the end of Chap. 13, we referred to dynamic optimization as a type of problem we were not ready to tackle because we did not yet have the tools of dynamic analysis such as differential equations. Now that we have acquired such tools, we can finally try a taste of dynamic optimization.

The classical approach to dynamic optimization is called the *calculus of variations*. In the later development of this methodology, however, a more powerful approach known as *optimal control theory* has, for the most part, supplanted the calculus of variations. For this reason, we shall, in this chapter, confine our attention to optimal control theory, explaining its basic nature, introducing the major solution tool called *the maximum principle*, and illustrating its use in some elementary economic models.†

20.1 The Nature of Optimal Control

In static optimization, the task is to find a single value for each choice variable, such that a stated objective function will be maximized or minimized, as the case may be. Such a problem is devoid of a time dimension. In contrast, time enters explicitly and prominently in a dynamic optimization problem. In such a problem, we will always have in mind a planning period, say from an initial time $t = 0$ to a terminal time $t = T$, and try to find the best course of action to take during that entire period. Thus the solution for any variable will take the form of not a single value, but a complete time path.

Suppose the problem is one of profit maximization over a time period. At any point of time t, we have to choose the value of some *control variable*, $u(t)$, which will then affect the value of some *state variable*, $y(t)$, via a so-called *equation of motion*. In turn, $y(t)$ will determine the profit $\pi(t)$. Since our objective is to maximize the profit over the entire period, the objective function should take the form of a definite integral of π from $t = 0$ to $t = T$. To be complete, the problem also specifies the initial value of the state variable y,

† For a more complete treatment of optimal control theory (as well as "calculus of variations"), the student is referred to *Elements of Dynamic Optimization* by Alpha C. Chiang, McGraw-Hill, New York, 1992, now published by Waveland Press, Inc., Prospect Heights, Illinois. This chapter draws heavily from material in this cited book.

$y(0)$, and the terminal value of y, $y(T)$, or alternatively, the range of values that $y(T)$ is allowed to take.

Taking into account the preceding, we can state the simplest problem of optimal control as:

$$\begin{aligned}
&\text{Maximize} && \int_0^T F(t, y, u)\, dt \\
&\text{subject to} && \frac{dy}{dt} \equiv y' = f(t, y, u) \qquad\qquad (20.1)\\
& && y(0) = A \qquad y(T) \text{ free} \\
&\text{and} && u(t) \in U \qquad \text{for all } t \in [0, T]
\end{aligned}$$

The first line of (20.1), the objective function, is an integral whose integrand $F(t, y, u)$ stipulates how the choice of the control variable u at time t, along with the resulting y at time t, determines our object of maximization at t. The second line is the equation of motion for the state variable y. What this equation does is to provide the mechanism whereby our choice of control variable u can be translated into a specific pattern of movement of the state variable y. Normally, the linkage between u and y can be adequately described by a first-order differential equation $y' = f(t, y, u)$. However, if it happens that the pattern of change of the state variable requires a second-order differential equation, then we must transform this equation into a pair of first-order differential equations. In that case an additional state variable will be introduced. Both the integrand F and the equation of motion are assumed to be continuous in all their arguments and possess continuous first-order partial derivatives with respect to the state variable y and the time variable t, but not necessarily the control variable u. In the third line, we indicate that the initial state, the value of y at $t = 0$, is a constant A, but the terminal state $y(T)$ is left unrestricted. Finally, the fourth line indicates that the permissible choices of u are limited to a control region U. It may happen, of course, that $u(t)$ is not restricted.

Illustration: A Simple Macroeconomic Model

Consider an economy that produces output Y using capital K and a fixed amount of labor L, according to the production function

$$Y = Y(K, L)$$

Further, output is used either for consumption C or for investment I. If we ignore the problem of depreciation, then

$$I \equiv \frac{dK}{dt}$$

In other words, investment is the change in capital stock over time. Thus we can also write investment as

$$I = Y - C = Y(K, L) - C = \frac{dK}{dt}$$

which gives us a first-order differential equation in the variable K.

If our objective is to maximize some form of social utility over a fixed planning period, then the problem becomes

$$
\begin{aligned}
&\text{Maximize} && \int_0^T U(C)\,dt \\
&\text{subject to} && \frac{dK}{dt} = Y(K, L) - C \qquad \textbf{(20.2)} \\
&\text{and} && K(0) = K_0 \qquad K(T) = K_T
\end{aligned}
$$

where K_0 and K_T are the initial value and terminal (target) value of K. Note that in (20.2), the terminal state is a fixed value, not left free as in (20.1). Here C serves as the control variable and K is the state variable. The problem is to choose the optimal control path $C(t)$ such that its impact on output Y and capital K, and the repercussions therefrom upon C itself, will together maximize the aggregate utility over the planning period.

Pontryagin's Maximum Principle

The key to optimal control theory is a first-order necessary condition known as the *maximum principle*.[1] The statement of the maximum principle involves an approach that is akin to the Lagrangian function and the Lagrangian multiplier variable. For optimal control problems, these are known as the *Hamiltonian function* and *costate variable*, concepts we will now develop.

The Hamiltonian

In (20.1), there are three variables: time t, the state variable y, and the control variable u. We now introduce a new variable known as the costate variable and denoted by $\lambda(t)$. Like the Lagrange multiplier, the costate variable measures the shadow price of the state variable.

The costate variable is introduced into the optimal control problem via a *Hamiltonian function* (or Hamiltonian, for short). The Hamiltonian is defined as

$$H(t, y, u, \lambda) \equiv F(t, y, u) + \lambda(t) f(t, y, u) \qquad \textbf{(20.3)}$$

where H denotes the Hamiltonian and is a function of four variables: t, y, u, and λ.

The Maximum Principle

The maximum principle—the main tool for solving problems of optimal control—is so named because, as a first-order necessary condition it requires us to choose u so as to maximize the Hamiltonian H *at every point of time.*

Since, aside from the control variable, u, H involves the state variable y and costate variable λ, the statement of the maximum principle also stipulates how y and λ should change over time, via an *equation of motion for the state variable y* (*state equation* for

[1] The term "maximum principle" is attributed to L. S. Pontryagin and his associates, and is often referred to as Pontryagin's maximum principle. See *The Mathematical Theory of Optimal Control Processes* by L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, Interscience, New York, 1962 (translated by K. N. Trirogoff).

short) as well as an *equation of motion for the costate variable* λ (*costate equation* for short). The state equation always comes as part of the problem statement itself, as in the second equation in (20.1). But in the view that (20.3) implies $\partial H/\partial\lambda = f(t, y, u)$, the maximum principle describes the state equation

$$y' = f(t, y, u) \text{ as } y' = \frac{\partial H}{\partial \lambda} \tag{20.4}$$

In contrast, λ does not appear in the problem statement (20.1) and its equation of motion enters into the picture purely as an optimization condition. The costate equation is

$$\lambda' \left(\equiv \frac{d\lambda}{dt}\right) = -\frac{\partial H}{\partial y} \tag{20.5}$$

Note that both equations of motion are stated in terms of the partial derivatives of H, suggesting some symmetry, but there is a negative sign attached to $\partial H/\partial y$ in (20.5).

Equations (20.4) and (20.5) constitute a system of two differential equations. Thus we need two boundary conditions to definitize the two arbitrary constants that will arise in the process of solution. If both the initial state $y(0)$ and the terminal state $y(T)$ are fixed, then these specifications can be used to definitize the constants. But if, as in problem (20.1), the terminal state is not fixed, then something called a *transversality condition* must be included as part of the maximum principle, to fill the gap left by the missing boundary condition.

Summing up the preceding, we can state the various components of the maximum principle for problem (20.1) as follows:

$$\begin{aligned}
&(i) \quad H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) && \text{for all } t \in [0, T] \\
&(ii) \quad y' = \frac{\partial H}{\partial \lambda} && \text{(state equation)} \\
&(iii) \quad \lambda' = -\frac{\partial H}{\partial y} && \text{(costate equation)} \\
&(iv) \quad \lambda(T) = 0 && \text{(transversality condition)}
\end{aligned} \tag{20.6}$$

Condition *i* in (20.6) states that at every time t the value of $u(t)$, the optimal control, must be chosen so as to maximize the value of the Hamiltonian over all admissible values of $u(t)$. In the case where the Hamiltonian is differentiable with respect to u and yields an interior solution, Condition *i* can be replaced by

$$\frac{\partial H}{\partial u} = 0$$

However, if the control region is a closed set, then boundary solutions are possible and $\partial H/\partial u = 0$ may not apply. In fact, the maximum principle does not even require the Hamiltonian to be differentiable with respect to u.

Conditions *ii* and *iii* of the maximum principle, $y' = \partial H/\partial\lambda$ and $\lambda' = -\partial H/\partial y$, give us two equations of motion, referred to as the Hamiltonian system for the given problem. Condition *iv*, $\lambda(T) = 0$, is the transversality condition appropriate for the free-terminal-state problem only.

Example 1

To illustrate the use of the maximum principle, let us first consider a simple noneconomic example—that of finding the shortest path from a given point A to a given straight line. In Fig. 20.1, we have plotted the point A on the vertical axis in the ty plane, and drawn the straight line as a vertical one at $t = T$. Three (out of an infinite number of) admissible paths are shown, each with a different length. The length of any path is the aggregate of small path segments, each of which can be considered as the hypotenuse (not drawn) of a triangle formed by small movements dt and dy. Denoting the hypotenuse by dh, we have, by Pythagoras's theorem,

$$dh^2 = dt^2 + dy^2$$

Dividing both sides by dt^2 and taking the square root yields

$$\frac{dh}{dt} = \left[1 + \left(\frac{dy}{dt}\right)^2\right]^{1/2} = [1 + (y')^2]^{1/2} \tag{20.7}$$

The total length of the path can then be found by integrating (20.7) with respect to t, from $t = 0$ to $t = T$. If we let $y' = u$ be the control variable, (20.7) can be expressed as

$$\frac{dh}{dt} = (1 + u^2)^{1/2} \tag{20.7'}$$

To minimize the integral of (20.7′) is, of course, equivalent to maximizing the *negative* of (20.7′). Thus the shortest-path problem is:

$$\text{Maximize} \quad \int_0^T -(1 + u^2)^{1/2} dt$$

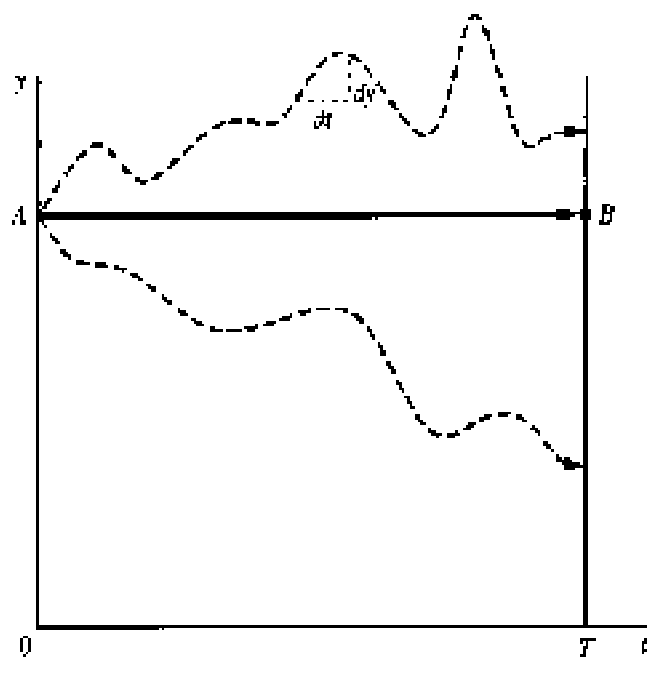
$$\text{subject to} \quad y' = u$$

$$\text{and} \quad y(0) = A \qquad y(T) \text{ free}$$

The Hamiltonian for the problem is, by (20.3),

$$H = -(1 + u^2)^{1/2} + \lambda u$$

FIGURE 20.1



Since H is differentiable in u, and u is unrestricted, the following first-order condition can be used to maximize H:

$$\frac{\partial H}{\partial u} = -\frac{1}{2}(1+u^2)^{-1/2}(2u) + \lambda = 0$$

or $$u(t) = \lambda(1-\lambda^2)^{-1/2}$$

Checking the second-order condition, we find that

$$\frac{\partial^2 H}{\partial u^2} = -(1+u^2)^{-3/2} < 0$$

which verifies that the solution to $u(t)$ does maximize the Hamiltonian. Since $u(t)$ is a function of λ, we need a solution to the costate variable. From the first-order conditions, the equation of motion for the costate variable is

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

since H is independent of y. Thus, λ is a constant. To definitize this constant, we can make use of the transversality condition $\lambda(T) = 0$. Since λ can take only a single value, now known to be zero, we actually have $\lambda(t) = 0$ for all t. Thus we can write

$$\lambda^*(t) = 0 \qquad \text{for all } t \in [0, T]$$

It follows that the optimal control is

$$u^*(t) = \lambda^*[1-(\lambda^*)^2]^{-1/2} = 0$$

Finally, using the equation of motion for the state variable, we see that

$$y' = u = 0$$

or $$y^*(t) = c_0 \qquad \text{(a constant)}$$

Incorporating the initial condition

$$y(0) = A$$

we can conclude that $c_0 = A$, and write

$$y^*(t) = A \qquad \text{for all } t$$

In Fig. 20.1, this path is the line AB. The shortest path is found to be a straight line with a zero slope.

Example 2 Find the optimal control path that will

$$\text{Maximize} \quad \int_0^1 (y - u^2)\,dt$$

$$\text{subject to} \quad y' = u$$

$$\text{and} \quad y(0) = 5 \qquad y(1) \text{ free}$$

This problem is in the format of (20.1), except that u is unrestricted.

The Hamiltonian for this problem,

$$H = y - u^2 + \lambda u$$

is concave in u, and u is unrestricted, so we can maximize H by applying the first-order condition (also sufficient because of concavity of H):

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

which gives us

$$u(t) = \frac{\lambda}{2} \quad \text{or} \quad y' = \frac{\lambda}{2} \tag{20.8}$$

The equation of motion for λ is

$$\lambda' = -\frac{\partial H}{\partial y} = -1 \tag{20.8'}$$

The last two equations constitute the differential-equation system for this problem.

We can first solve for λ by straight integration of (20.8') to get

$$\lambda(t) = c_1 - t \qquad (c_1 \text{ arbitrary})$$

Moreover, by the transversality condition in (20.6), we must have $\lambda(1) = 0$. Setting $t = 1$ in the last equation yields $c_1 = 1$. Thus the optimal costate path is

$$\lambda^*(t) = 1 - t$$

It follows that $y' = \frac{1}{2}(1 - t)$, by (20.8), and by integration,

$$y(t) = \frac{1}{2}t - \frac{1}{4}t^2 + c_2 \qquad (c_2 \text{ arbitrary})$$

The arbitrary constant can be definitized by using the initial condition $y(0) = 5$. Setting $t = 0$ in the preceding equation, we get $5 = y(0) = c_2$. Thus the optimal path for the state variable is

$$y^*(t) = \frac{1}{2}t - \frac{1}{4}t^2 + 5$$

and the corresponding optimal control path is

$$u^*(t) = \frac{1}{2}(1 - t)$$

Example 3 Find the optimal control path that will

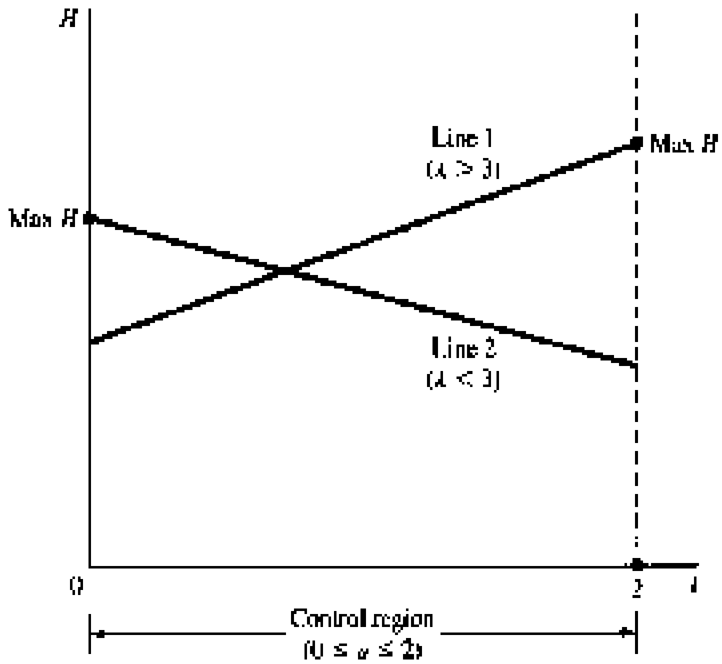
$$\begin{aligned} \text{Maximize} \quad & \int_0^2 (2y - 3u)\, dt \\ \text{subject to} \quad & y' = y + u \\ & y(0) = 4 \qquad y(2) \text{ free} \\ \text{and} \quad & u(t) \in [0, 2] \end{aligned}$$

The fact that the control variable is restricted to the closed set [0, 2] gives rise to the possibility of boundary solutions.

The Hamiltonian function

$$H = 2y - 3u + \lambda(y + u) = (2 + \lambda)y + (\lambda - 3)u$$

FIGURE 20.2



is linear in u. If we plot H against u in the uH plane, we get a straight line with slope $\partial H/\partial u = \lambda - 3$, which is positive if $\lambda > 3$ (Line 1), but negative if $\lambda < 3$ (Line 2), as illustrated in Fig. 20.2. If at any time λ exceeds 3, then the maximum H occurs at the upper boundary of the control region and we must choose $u = 2$. If, on the other hand, λ falls below 3, then in order to maximize H, we must choose $u = 0$. In short, $u^*(t)$ depends on $\lambda(t)$ as follows:

$$u^*(t) = \begin{Bmatrix} 2 \\ 0 \end{Bmatrix} \quad \text{if} \quad \lambda(t) \begin{Bmatrix} > \\ < \end{Bmatrix} 3 \tag{20.9}$$

Thus, it is critical to find $\lambda(t)$. To do this, we start from the costate equation

$$\lambda' = -\frac{\partial H}{\partial y} = -2 - \lambda \qquad \text{or} \qquad \lambda' + \lambda = -2$$

The general solution of this equation is

$$\lambda(t) = Ae^{-t} - 2 \qquad [\text{by (15.5)}]$$

where A is an arbitrary constant. By using the transversality condition $\lambda(T) = \lambda(2) = 0$, we find that $A = 2e^2$. Thus the definite solution for λ is

$$\lambda^*(t) = 2e^{2-t} - 2 \tag{20.10}$$

which is a decreasing function of t, falling steadily from the initial value $\lambda^*(0) = 2e^2 - 2 = 12.778$ to a terminal value $\lambda^*(2) = 2e^0 - 2 = 0$. This means that λ^* must pass through the point $\lambda = 3$ at some critical time τ, when the optimal u has to be switched from $u^* = 2$ to $u^* = 0$.

To find this critical time τ, we set $\lambda^*(\tau) = 3$ in (20.10):

$$3 = \lambda^*(\tau) = 2e^{2-\tau} - 2 \qquad \text{or} \qquad e^{2-\tau} = \frac{5}{2} = 2.5$$

Taking the natural log of both sides, we get

$$\ln e^{2-\tau} = \ln 2.5 \qquad \text{or} \qquad 2 - \tau = \ln 2.5$$

Thus

$$\tau = 2 - \ln 2.5 = 1.084 \qquad \text{(approx.)}$$

and the optimal control turns out to consist of two phases in the time interval [0, 2]:

$$\text{Phase 1: } u^*[0, \tau) = 2 \qquad \text{Phase 2: } u^*[\tau, 2] = 0$$

20.2 Alternative Terminal Conditions

What happens to the maximum principle when the terminal condition is different from the one in (20.1)? In (20.1), we face a vertical terminal line—with a fixed terminal time but unrestricted terminal state as illustrated in Fig. 20.1. The maximum principle for the maximization problem requires that

$$(i) \quad H(t, y, u^*, \lambda) \geq H(t, y, u, \lambda) \qquad \text{for all } t \in [0, T]$$

$$(ii) \quad y' = \frac{\partial H}{\partial \lambda}$$

$$(iii) \quad \lambda' = -\frac{\partial H}{\partial y}$$

with the transversality condition

$$(iv) \quad \lambda(T) = 0$$

With alternative terminal conditions, Conditions *i*, *ii*, and *iii* will remain the same, but Condition *iv* (the transversality condition) must be duly modified.

Fixed Terminal Point

If the terminal point is fixed so that the terminal condition is $y(T) = y_T$ with both T and y_T given, then the terminal condition itself should provide the information to definitize one constant. In this case, no transversality condition is needed.

Horizontal Terminal Line

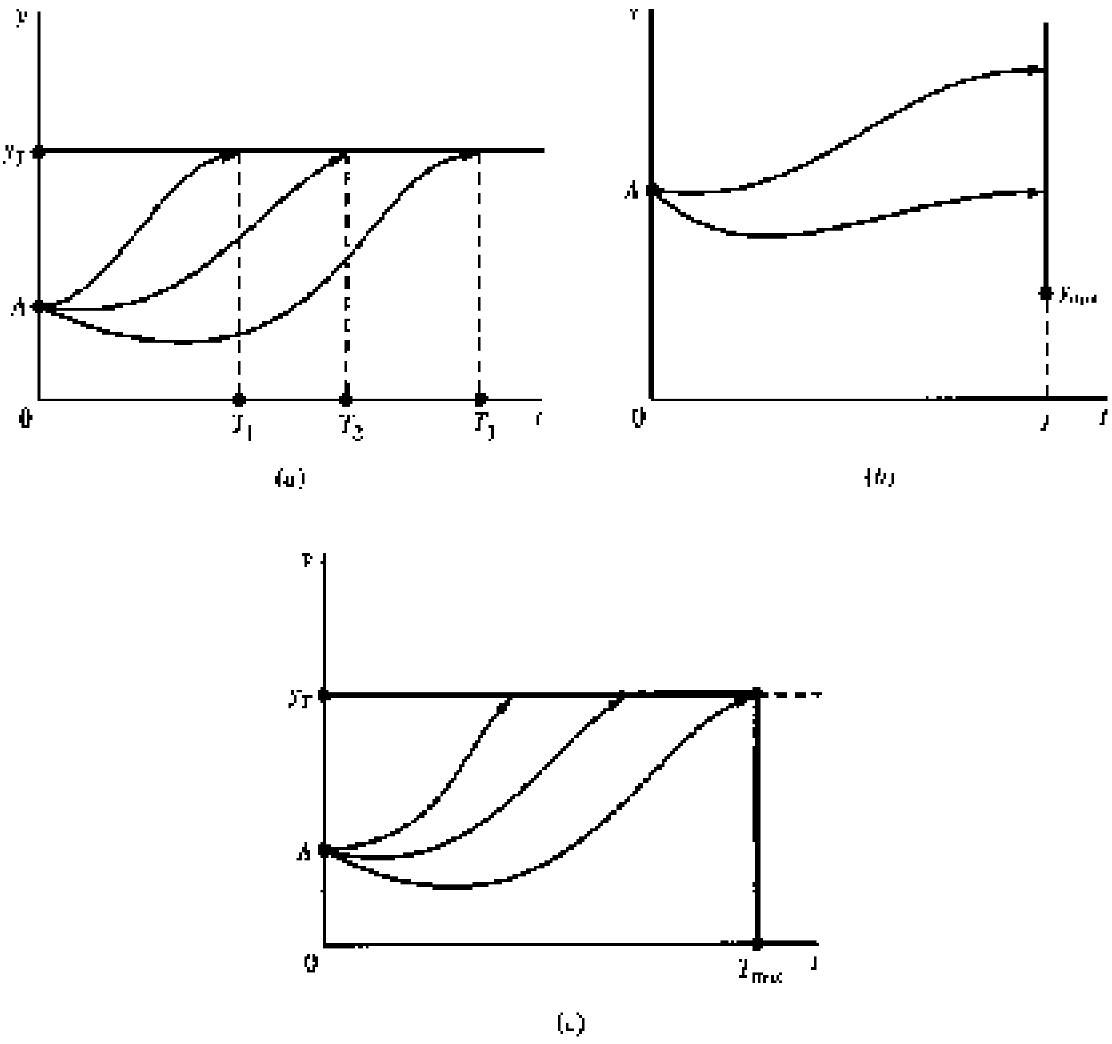
Suppose that the terminal state is fixed at a given target level y_T but the terminal time T is free, so that we have the flexibility to reach the target in a hurry or at a leisurely pace. We then have a horizontal terminal line as illustrated in Fig. 20.3a, which allows us to choose between T_1, T_2, T_3, or other terminal times to reach the target level of y. For this case, the transversality condition is a restriction on the Hamiltonian (rather than the costate variable) at $t = T$:

$$H_{t=T} = 0 \qquad (20.11)$$

Truncated Vertical Terminal Line

If we have a fixed terminal time T, and the terminal state is free but subject to the proviso that $y_T \geq y_{\min}$, where $y_{\min}$ denotes a given minimum permissible level of y, we face a truncated vertical terminal line, as illustrated in Fig. 20.3b.

FIGURE 20.3



The transversality condition for this case can be stated like the complementary-slackness condition found in the Kuhn-Tucker conditions:

$$\lambda(T) \geq 0 \qquad y_T \geq y_{min} \qquad (y_T - y_{min})\lambda(T) = 0 \qquad \textbf{(20.12)}$$

The practical approach for solving this type of problem is to first try $\lambda(T) = 0$ as the transversality condition and test if the resulting y_T^* satisfies the restriction $y_T^* \geq y_{min}$. If so, the problem is solved. If not, then treat the problem as a given terminal point problem with y_{min} as the terminal state.

Truncated Horizontal Terminal Line

When the terminal state is fixed at y_T and the terminal time is free but subject to the restriction $T \leq T_{max}$, where T_{max} denotes the latest permissible time (a deadline) to reach the given y_T, we face a truncated horizontal terminal line as illustrated in Fig. 20.3c. The transversality condition becomes

$$H_{t=T_{max}} \geq 0 \qquad T \leq T_{max} \qquad (T - T_{max})H_{t=T_{max}} = 0 \qquad \textbf{(20.13)}$$

This again appears in the format of the complementary-slackness condition.

The practical approach to solving this type of problem is to try $H_{t=T_{max}} = 0$ first. If the resulting solution value is $T^* \leq T_{max}$, then the problem is solved. If not, then we must take

T_{max} as a fixed terminal time which, together with the given y_T, defines a fixed end point, and solve the problem as a fixed-end-point problem.

Example 1 In the problem

$$\text{Maximize} \quad \int_0^1 (y - u^2)\,dt$$
$$\text{subject to} \quad y' = u$$
$$\text{and} \quad y(0) = 2 \qquad y(1) = \alpha$$

the terminal point is fixed, even though $y(1)$ is assigned a parametric rather than numerical value here.

The Hamiltonian function

$$H = y - u^2 + \lambda u$$

is concave in u, so we can set $\partial H/\partial u = 0$ to maximize H:

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

Thus

$$u = \frac{\lambda}{2}$$

which shows that in order to solve for $u(t)$, we need to solve for $\lambda(t)$ first.

The two equations of motion are

$$y'(= u) = \frac{\lambda}{2}$$
$$\lambda'\left(= -\frac{\partial H}{\partial y}\right) = -1$$

Direct integration of the last equation yields

$$\lambda(t) = c_1 - t \qquad (c_1 \text{ arbitrary})$$

which implies that

$$y' = \frac{1}{2}c_1 - \frac{1}{2}t$$

Again, by direct integration, we find that

$$y(t) = \frac{c_1}{2}t - \frac{1}{4}t^2 + c_2 \qquad (c_2 \text{ arbitrary})$$

To definitize the two arbitrary constants, we make use of the initial condition $y(0) = 2$, and the terminal condition $y(1) = \alpha$. Setting $t = 0$ and $t = 1$, successively, in the preceding equation, we obtain

$$2 = y(0) = c_2 \qquad \alpha = y(1) = \frac{c_1}{2} - \frac{1}{4} + c_2$$

Thus, $c_2 = 2$, and $c_1 = 2\alpha - \frac{7}{2}$.

Therefore, we can write the optimal paths of this problem as:

$$y^*(t) = (a - \frac{7}{4})t - \frac{1}{4}t^2 + 2$$
$$\lambda^*(t) = 2a - \frac{7}{2} - t$$
$$u^*(t) = a - \frac{7}{4} - \frac{1}{2}t$$

Example 2 The problem

$$\text{Maximize} \quad \int_0^T -(t^2 + u^2)\,dt$$
$$\text{subject to} \quad y' = u$$
$$\text{and} \quad y(0)4 \qquad y(T) = 5 \qquad T \text{ free}$$

exemplifies the case of horizontal terminal line where the terminal state is fixed but the time of arrival at the target level of y is unrestricted. In fact, it is one of our tasks to solve for the optimal value of T.

Since the Hamiltonian

$$H = -t^2 - u^2 + \lambda u$$

is concave in u, we can again maximize H by using the first-order condition

$$\frac{\partial H}{\partial u} = -2u + \lambda = 0$$

which gives us

$$u = \frac{\lambda}{2} \tag{20.14}$$

The concavity of H makes it unnecessary to check the second-order condition, but if we wish, it is easy to check that $\partial^2 H/\partial u^2 = -2 < 0$, sufficient for a maximum of H.

The equation of motion for λ is

$$\lambda' = -\frac{\partial H}{\partial y} = 0$$

which implies that λ is a constant. But we cannot yet determine its exact value at this point.

Turning to the equation of motion for y,

$$y' = u = \frac{\lambda}{2} \qquad \text{[by (20.14)]}$$

we can obtain, by direct integration,

$$y(t) = \frac{\lambda}{2}t + c \tag{20.15}$$

Since $y(0) = 4$, we see that $c = 4$. Furthermore, the transversality condition (20.11) requires that

$$H_{t=T} = -T^2 - \frac{\lambda^2}{4} + \frac{\lambda^2}{2} = -T^2 + \frac{\lambda^2}{4} = 0 \qquad \text{[by (20.14)]}$$

Solving the preceding equation for T, and taking the positive square root, we get

$$T = \frac{\lambda}{2} \tag{20.16}$$

Since λ is constant, so is T. We try now to find its exact value.

Applying the terminal-state specification $y(T) = 5$ to (20.15), and recalling that $c = 4$, we get

$$y(T) = \frac{\lambda}{2}T + 4 = 5$$

In view of (20.16), the last equation can be rewritten as $T^2 = 1$. Thus, by taking the square root, we can determine the optimal arrival time to be

$$T^* = 1 \qquad \text{(negative root unacceptable)}$$

From this, we can readily deduce that

$$\begin{aligned} \lambda^*(t) &= 2T^* = 2 && \text{[by (20.16)]} \\ u^*(t) &= \frac{\lambda}{2} = 1 && \text{[by (20.14)]} \\ y^*(t) &= t + 4 && \text{[by (20.15)]} \end{aligned}$$

The last result shows that, in this example, the optimal y path is a straight line going from the given initial point to the horizontal terminal line.

EXERCISE 20.2

Find the optimal paths of the control, state, and costate variables that will

1. Maximize $\displaystyle\int_0^1 (y - u^2)\,dt$

   subject to $y' = u$

   and $y(0) = 2 \qquad y(1)$ free

2. Maximize $\displaystyle\int_0^8 6y\,dt$

   subject to $y' = y + u$

   $y(0) = 10 \qquad y(8)$ free

   and $u(t) \in [0, 2]$

3. Maximize $\displaystyle\int_0^T -(au + bu^2)\,dt$

   subject to $y' = y - u$

   and $y(0) = y_0 \qquad y(t)$ free

4. Maximize $\displaystyle\int_0^T (yu - u^2 - y^2)\,dt$

   subject to $y' = u$

   and $y(0) = y_0 \qquad y(t)$ free

5. Maximize $\int_0^{20} -\frac{1}{2}u^2\,dt$

subject to $y' = u$

and $y(0) = 10 \quad y(20) = 0$

6. Maximize $\int_0^4 3y\,dt$

subject to $y' = y + u$

$y(0) = 5 \quad y(4) \geq 300$

and $0 \leq u(t) \leq 2$

7. Maximize $\int_0^1 -u^2\,dt$

subject to $y' = y + u$

and $y(0) = 1 \quad y(1) = 0$

8. Maximize $\int_1^2 (y + ut - u^2)\,dt$

subject to $y' = u$

and $y(1) = 3 \quad y(2) = 4$

9. Maximize $\int_0^2 (2y - 3u - au^2)\,dt$

subject to $y' = u + y$

and $y(0) = 5 \quad y(2)$ free

20.3 Autonomous Problems

In the general control problem framework, the variable t can enter the objective function and state equation directly. The general specification

$$\text{Maximize} \quad \int_0^T F(t, y, u)\,dt$$
$$\text{subject to} \quad y' = f(t, y, u)$$
$$\text{and} \quad \text{boundary conditions}$$

where t explicitly enters into F and f means the date matters. That is, the value generated by the activity $u(t)$ depends not only on the level, but also on exactly when this activity takes place.

Problems in which t is absent from the objective function and state equation such as

$$\text{Maximize} \quad \int_0^T F(y, u)\,dt$$
$$\text{subject to} \quad y' = f(y, u)$$
$$\text{and} \quad \text{boundary conditions}$$

are called *autonomous problems*. In such problems, since the Hamiltonian

$$H = F(y, u) + \lambda f(y, u)$$

does not contain t as an argument, the equations of motion are easier to solve; moreover, they are amenable to the use of phase-diagram analysis.

In still other cases, in an otherwise autonomous problem, time t enters into the picture as part of the discount factor e^{-rt}, but nowhere else, so that the objective function takes the form of

$$\int_0^T G(y, u)e^{-rt}\, dt$$

Strictly speaking, this problem is nonautonomous. However, it is easy to convert the problem into an autonomous one by employing the so-called *current-value Hamiltonian*, defined as:

$$H_c \equiv He^{rt} = G(y, u) + \mu f(y, u) \tag{20.17}$$

where

$$\mu \equiv \lambda e^{rt} \tag{20.18}$$

is the *current-value Lagrange multiplier*. By focusing on the current (undiscounted) value, we are able to eliminate t from the original Hamiltonian.

Using H_c in lieu of H, we must revise the maximum principle to:

$$\begin{aligned}
&(i)\ H_c(y, u^*, \mu) \ge H_c(y, u, \mu) \qquad \text{for all } t \in [0, T] \\
&(ii)\ y' = \frac{\partial H_c}{\partial \mu} \\
&(iii)\ \mu' = -\frac{\partial H_c}{\partial y} + r\mu \\
&(iv)\ \mu(T) = 0 \qquad \text{(for vertical terminal line)} \\
&\quad\ \text{or } [H_c]_{t=T} = 0 \qquad \text{(for horizontal terminal line)}
\end{aligned} \tag{20.19}$$

20.4 Economic Applications

Lifetime Utility Maximization

Suppose a consumer has the utility function $U(C(t))$, where $C(t)$ is consumption at time t. The consumer's utility function is concave, and has the following properties:

$$U' > 0 \qquad U'' < 0$$

The consumer is also endowed with an initial stock of wealth, or capital, K_0, with income stream derived from the stock of capital according to the following:

$$Y = rK$$

where r is the market rate of interest. The consumer uses the income to purchase C. In addition, the consumer can consume the capital stock. Any income not consumed is added to the capital stock as investment. Thus,

$$K' \equiv I = Y - C = rK - C$$

The consumer's lifetime utility maximization problem is to

$$\text{Maximize} \quad \int_0^T U(C(t))e^{-\delta t}\,dt$$

$$\text{subject to} \quad K' = rK(t) - C(t)$$

$$\text{and} \quad K(0) = K_0 \qquad K(T) \geq 0$$

where δ is the consumer's personal rate of time preference ($\delta \geq 0$). It is assumed that $C(t) > 0$ and $K(t) > 0$ for all t.

The Hamiltonian is

$$H = U(C(t))e^{-\delta t} + \lambda(t)\,[rK(t) - C(t)]$$

where C is the control variable, and K is the state variable. Since $U(C)$ is concave, and the constraint is linear in C, we know that the Hamiltonian is concave and the maximization of H can be achieved by simply setting $\partial H/\partial C = 0$. Thus we have

$$\frac{\partial H}{\partial C} = U'(C)e^{-\delta t} - \lambda = 0 \qquad \textbf{(20.20)}$$

$$K' = rK(t) - C(t) \qquad \textbf{(20.20')}$$

$$\lambda' = -\frac{\partial H}{\partial K} = -r\lambda \qquad \textbf{(20.20'')}$$

Equation (20.20) states that the discounted marginal utility should be equated to the present shadow price of an additional unit of capital. Differentiating (20.20) with respect to t, we get

$$U''(C)C'e^{-\delta t} - \delta U'(C)e^{-\delta t} = \lambda' \qquad \textbf{(20.21)}$$

In view of (20.20) and (20.20″) we have

$$\lambda' = -r\lambda = -rU'(C)e^{-\delta t}$$

which can be substituted into (20.21) to yield

$$U''(C)C'(t)e^{-\delta t} - \delta U'(C)e^{-\delta t} = -rU'(C)\,e^{-\delta t}$$

or, after canceling the common factor $e^{-\delta t}$ and rearranging,

$$\frac{-U''(C(t))}{U'(C(t))}C'(t) = r - \delta$$

Since $U' > 0$ and $U'' < 0$, the sign of the derivative $C'(t)$ has to be the same as $(r - \delta)$. Therefore, if $r > \delta$, the optimal consumption will rise over time; if $r < \delta$, the optimal consumption will decline over time.

Solving (20.20″) directly gives us

$$\lambda(t) = \lambda_0 e^{-rt}$$

where $\lambda_0 > 0$ is the constant of integration. Combining this with (20.20) gives us

$$U'(C(t)) = \lambda e^{\delta t} = \lambda_0 e^{(\delta - r)t}$$

which shows that the marginal utility of consumption will optimally decrease over time if $r > \delta$, but increase over time if $r < \delta$.

Since the terminal condition $K(T) \geq 0$ identifies the present problem as one with a truncated vertical terminal line, the appropriate transversality condition is, by (20.12),

$$\lambda(T) \geq 0 \quad K(T) \geq 0 \quad K(T)\lambda(T) = 0$$

The key condition is the complementary-slackness stipulation, which means that either the capital stock K must be exhausted on the terminal date, or the shadow price of capital λ must fall to zero on the terminal date. By assumption, $U'(C) > 0$, the marginal utility can never be zero. Therefore, the marginal value of capital cannot be zero. This implies that the capital stock should optimally be exhausted by the terminal date T in this model.

Exhaustible Resource

Let $s(t)$ denote a stock of an exhaustible resource and $q(t)$ be the rate of extraction at any time t such that

$$s' = -q$$

The extracted resource produces a final consumer good c such that

$$c = c(q) \qquad \text{where} \qquad c' > 0, c'' < 0 \tag{20.22}$$

The consumption good is the sole argument in the utility function of a representative consumer with the following properties:

$$U = U(c) \qquad \text{where} \qquad U' > 0, U'' < 0 \tag{20.22'}$$

The consumer wishes to maximize the utility function over a given interval $[0, T]$. Since c is a function of q, the rate of extraction, q will serve as the control variable. For simplicity, we ignore the issue of discounting over time. The dynamic problem is then to choose the optimal extraction rate that maximizes the utility function subject only to a nonnegativity constraint on the state variable $s(t)$, the stock of the exhaustible resource. The formulation is

$$\begin{aligned} &\text{Maximize} && \int_0^T U(c(q))\,dt \\ &\text{subject to} && s' = -q \\ &\text{and} && s(0) = s_0 \qquad s(T) \geq 0 \end{aligned} \tag{20.23}$$

where s_0 and T are given.

The Hamiltonian for the problem is

$$H = U(c(q)) - \lambda q$$

Since H is concave in q by model specifications on the $U(c(q))$ function, we can maximize H by setting $\partial H/\partial q = 0$:

$$\frac{\partial H}{\partial q} = U'(c(q))c'(q) - \lambda = 0 \tag{20.24}$$

The concavity of H assures us that (20.24) maximizes H, but we can easily check the second-order condition and confirm that $\partial^2 H/\partial q^2$ is negative.

The maximum principle stipulates that

$$\lambda' = -\frac{\partial H}{\partial s} = 0$$

which implies that

$$\lambda(t) = c_0 \qquad \text{a constant} \tag{20.25}$$

To determine c_0, we turn to the transversality conditions. Since the model specifies $K(T) \geq 0$, it has a truncated vertical terminal line, so (20.12) applies:

$$\lambda(T) \geq 0 \qquad s(T) \geq 0 \qquad s(T)\lambda(T) = 0$$

In practical applications, the initial step is to try $\lambda(T) = 0$, solve for q, and see if the solution will work. Since $\lambda(T)$ is a constant, to try $\lambda(T) = 0$ implies $\lambda(t) = 0$ for all t, and $\partial H/\partial q$ in (20.24) reduces to

$$U'(c)c'(q) = 0$$

which (in principle) can be solved for q. Since t is not an explicit argument of U or c, the solution path for q is constant over time:

$$q^*(t) = q^*$$

Now, we check if q^* satisfies the restriction $s(T) \geq 0$. If q^* is a constant, then the equation of motion

$$s' = -q$$

can be readily integrated, yielding

$$s(t) = -qt + c_1 \qquad [c_1 = \text{constant of integration}]$$

Using the initial condition $s(0) = s_0$ yields a solution for the constant of integration

$$c_1 = s_0$$

and the optimal state path is

$$s(t) = s_0 - q^*t \tag{20.26}$$

Without specifying the functional forms for U and c, no numerical solution can be found for q^*. However, from the transversality conditions, we can conclude that if $s(T) \geq 0$, then q^* as derived in the solution is acceptable. But if $s(T) < 0$ for the given q^*, then the extraction rate is too high and we need to find a different solution. Since the trial solution $\lambda(T) = 0$ failed, we now take the alternative of $\lambda(T) > 0$. Even in this case, though, λ is still a constant by (20.25). And (20.24) can still (in principle) yield a constant, but different, solution value q_2^*. It follows that (20.26) remains valid. But this time, with $\lambda(T) > 0$, the transversality condition (20.12) dictates that $s(T) = 0$, or in view of (20.26),

$$s_0 - q_2^*T = 0$$

Thus we can write the revised (constant) optimal rate of extraction as

$$q_2^* = \frac{s_0}{T}$$

This new solution value should represent a lower extraction rate that would not violate the $s(T) \geq 0$ boundary condition.

EXERCISE 20.4

1. Maximize $\int_0^T (K - aK^2 - I^2)dt \quad (a > 0)$

   subject to $K' = I - \delta K \quad (\delta > 0)$

   and $K(0) = K_0 \quad K(T)$ free

2. Solve the following exhaustible resource problem for the optimal extraction path:

   Maximize $\int_0^T \ln(q)e^{-\delta t}dt$

   subject to $s' = -q$

   and $s(0) = s_0 \quad s(t) \geq 0$

20.5 Infinite Time Horizon

In this section we introduce the problem of dynamic optimization over an infinite planning period. Infinite time horizon models tend to introduce complexities with respect to transversality conditions and optimal time paths that differ from those developed earlier. Rather than address these issues here, we shall illustrate the methodology of such models with a version of the *neoclassical optimal growth model*.

Neoclassical Optimal Growth Model

The standard neoclassical production function expresses output Y as a function of two inputs: labor L and capital K. Its general form is

$$Y = Y(K, L)$$

where $Y(K, L)$ is a linearly homogeneous function with the properties

$$Y_L > 0 \qquad Y_K > 0 \qquad Y_{LL} < 0 \qquad Y_{KK} < 0$$

Rewriting the production function in per capita terms yields

$$y = \phi(k) \qquad \text{with } \phi'(k) > 0 \qquad \text{and} \qquad \phi''(k) < 0$$

where $y = Y/L$ and $k = K/L$. Total output Y is allocated to consumption C or gross investment I. Let δ be the rate of depreciation of the capital stock K. Then net investment or changes to the capital stock can be written as

$$K' = I - \delta K = Y - C - \delta K$$

Denoting per capita consumption as $c \equiv C/L$, we can write as

$$\frac{1}{L}K' = y - c - \delta k \tag{20.27}$$

The right-hand side of (20.27) is in per capita terms, but the left-hand side is not. To unify, we note that

$$K' = \frac{dK}{dt} = \frac{d}{dt}(kL) = k\frac{dL}{dt} + L\frac{dk}{dt} \tag{20.28}$$

If the population growth rate is[4]

$$\frac{dL/dt}{L} = n \qquad \text{so that} \qquad \frac{dL}{dt} = nL$$

then (20.28) becomes

$$K' = knL + Lk' \qquad \text{or} \qquad \frac{1}{L}K' = kn + k'$$

Substituting this into (20.27) transforms the latter into an equation entirely in per capita terms:

$$k' = y - c - (n+\delta)k = \phi(k) - c - (n+\delta)k \tag{20.27'}$$

Let $U(c)$ be the social welfare function (expressed in per capita terms), where

$$U'(c) > 0 \qquad \text{and} \qquad U''(c) < 0$$

and, to eliminate corner solutions, we also assume

$$U'(c) \to \infty \qquad \text{as } c \to 0$$

and

$$U'(c) \to 0 \qquad \text{as } c \to \infty$$

If ρ denotes the social discount rate and the initial population is normalized to one, the objective function can be expressed as

$$V = \int_0^\infty U(c)e^{-\rho t}L_0e^{nt}\,dt = \int_0^\infty U(c)e^{-(\rho-n)t}\,dt$$

$$= \int_0^\infty U(c)e^{-rt}\,dt \qquad \text{where } r = \rho - n$$

In this version of the neoclassical optimal growth model, utility is weighted by a population that grows continuously at a rate of n. However, if $r = \rho - n > 0$, then the model is mathematically no different from one without population weights but with a positive discount rate r.

The optimal growth problem can now be stated as

$$\begin{aligned} &\text{Maximize} && \int_0^\infty U(c)e^{-rt}\,dt \\ &\text{subject to} && k' = \phi(k) - c - (n+\delta)k \\ & && k(0) = k_0 \\ &\text{and} && 0 \le c(t) \le \phi(k) \end{aligned} \tag{20.29}$$

where k is the state variable and c is the control variable.

[4] In this model we assume labor force and population to be one and the same.

The Hamiltonian for the problem is

$$H = U(c)e^{-rt} + \lambda[\phi(k) - c - (n+\delta)k]$$

Since H is concave in c, the maximum of H corresponds to an interior solution in the control region $[0 < c < f(k)]$, and therefore we can find the maximum of H from

$$\frac{\partial H}{\partial c} = U'(c)e^{-rt} - \lambda = 0$$

or

$$U'(c) = \lambda e^{rt} \tag{20.30}$$

The economic interpretation of (20.30) is that, along the optimal path, the marginal utility of per capita consumption should equal the shadow price of capital (λ) weighted by e^{rt}. Checking second-order conditions, we find

$$\frac{\partial^2 H}{\partial c^2} = U''(c)e^{-rt} < 0$$

Therefore, the Hamiltonian is maximized.

From the maximum principle, we have two equations of motion

$$k' = \frac{\partial H}{\partial \lambda} = \phi(k) - c - (n+\delta)k$$

and

$$\lambda' = -\frac{\partial H}{\partial k} = -\lambda[\phi'(k) - (n+\delta)]$$

The two equations of motion combined with the $U'(c) = \lambda e^{rt}$ should in principle define a solution for c, k, λ. However, at this level of generality we are unable to do more than undertake qualitative analysis of the model. Anything more would require specific forms of both the utility and production functions.

The Current-Value Hamiltonian

Since the preceding model is an example of an autonomous problem (t is not a separate argument in the utility function or state equation but appears only in the discount factor), we may use the current-value Hamiltonian written as

$$H_c = He^{rt} = U(c) + \mu\,[\phi(k) - c - (n+\delta)k] \qquad \text{[see (20.17)]}$$

where $\mu = \lambda e^{rt}$.

The maximum principle calls for

$$\frac{\partial H_c}{\partial c} = U'(c) - \mu = 0 \qquad \text{or} \qquad \mu = U'(c) \tag{20.31}$$

$$k' = \frac{\partial H_c}{\partial \mu} = \phi(k) - c - (n+\delta)k \tag{20.31'}$$

$$\begin{aligned}\mu' &= -\frac{\partial H_c}{\partial k} + r\mu = -\mu[\phi'(k) - (n+\delta)] + r\mu \\ &= -\mu[\phi'(k) - (n+\delta+r)]\end{aligned} \tag{20.31''}$$

Equations (20.31') and (20.31'') constitute an autonomous differential equation system. This makes possible a qualitative analysis by phase diagram.

Constructing a Phase Diagram

The variables in the differential equations (20.31′) and (20.31″) are k and μ. Since (20.31) involves a function of c, namely $U'(c)$, rather than the plain c itself, it would be simpler to construct a phase diagram in the kc space rather than the $k\mu$ space. To do this, we shall try to eliminate μ. Since $\mu = U'(c)$, by (20.31), differentiation with respect to t gives us

$$\mu' = U''(c)c'$$

Substituting these expressions for μ and μ' into (20.31″) yields

$$c' = -\frac{U'(c)}{U''(c)}[\phi'(k) - (n + \delta + r)]$$

which is a differential equation in c. We now have the autonomous differential equation system

$$k' = \phi(k) - c - (n + \delta)k \qquad \textbf{(20.31′)}$$

and

$$c' = -\frac{U'(c)}{U''(c)}[\phi'(k) - (n + \delta + r)] \qquad \textbf{(20.32)}$$

To construct the phase diagram in the kc space, we first draw the $k' = 0$ and $c' = 0$ curves which are defined by

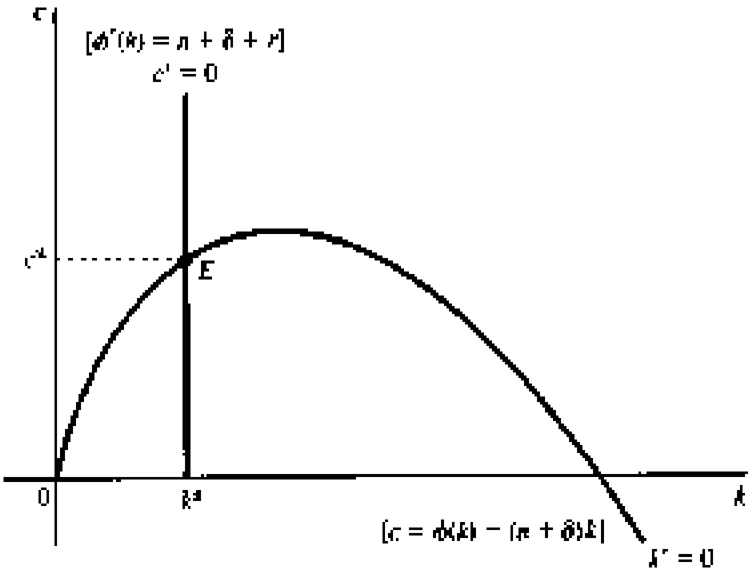
$$c = \phi(k) - (n + \delta)k \qquad (k' = 0) \qquad \textbf{(20.33)}$$

and

$$\phi'(k) = n + \delta + r \qquad (c' = 0) \qquad \textbf{(20.34)}$$

These two curves are illustrated in Fig. 20.4. The equation for the $k' = 0$ curve, (20.33), has the same structure as the fundamental equation of the Solow growth model, (15.30). Thus the $k' = 0$ curve has the same general shape as the one in Fig. 15.5b. The $c' = 0$ curve, on the other hand, plots as a vertical line because given the model specifications $\phi'(k) > 0$ and $\phi''(k) < 0$, $\phi(k)$ is associated with an upward-sloping concave curve, with a different slope at every point on the curve, so that only a unique value of k can satisfy (20.34). The intersection of the two curves at point E determines the intertemporal equilibrium values of

FIGURE 20.4



k and c, because at point E, neither k nor c will change in value over time, resulting in a *steady state*. We could label these values as $\bar{k}$ and $\bar{c}$ for intertemporal equilibrium values, but we shall label them as k^* and c^* instead, because they also represent the equilibrium values for optimal growth.

Analyzing the Phase Diagram

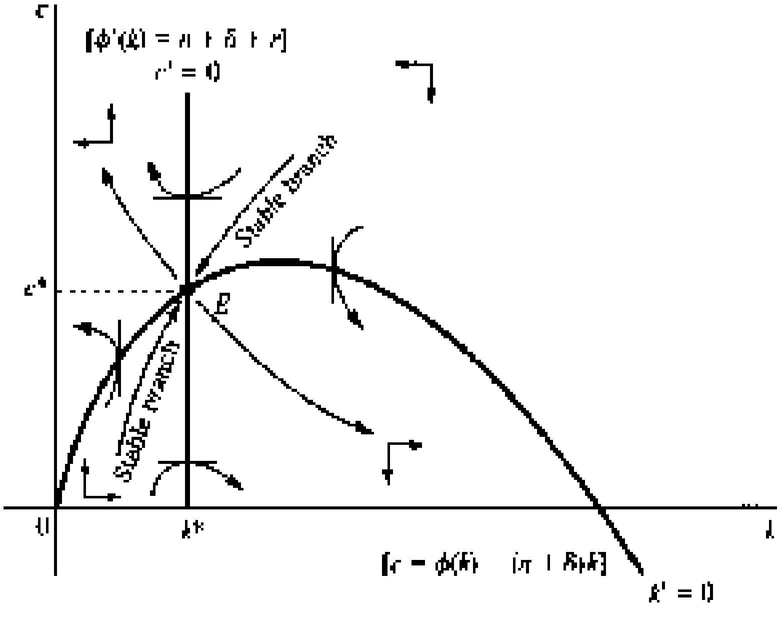
The intersection point E in Fig. 20.4 gives us a unique steady state. But what happens if we are initially at some point other than E? Returning to our system of first-order differential equations (20.31′) and (20.32), we can deduce that

$$\frac{\partial k'}{\partial c} = -1 < 0 \qquad \text{and} \qquad \frac{\partial c'}{\partial k} = -\frac{U'(c)}{U''(c)}\phi''(k) < 0$$

Since $\partial k'/\partial c < 0$, all the points below the $k' = 0$ curve are characterized by $k' > 0$ and all the points above the curve by $k' < 0$. Similarly, since $\partial c'/\partial k < 0$, all the points to the left of the $c' = 0$ line are characterized by $c' > 0$ and all the points to the right of the line by $c' < 0$. Thus the $k' = 0$ curve and the $c' = 0$ line divide the phase space into four regions, each with its own distinct pairing of signs of c' and k'. These are reflected in Fig. 20.5 by the right-angled directional arrows in each region.

The streamlines that follow the directional arrows in each region tell us that the steady state at point E is a saddle point. If we have an initial point that lies on one of the two stable branches of the saddle point, the dynamics of the system will lead us to point E. But any initial point that does not lie on a stable branch will make us either skirt around point E, never reaching it, or move steadily away from it. If we follow the streamlines of the latter instances, we will eventually (as $t \to \infty$) end up either with $k = 0$ (exhaustion of capital) or $c = 0$ (per capita consumption dwindling to zero)—both of which are economically unacceptable. Thus, the only viable alternative is to choose a (k, c) pair so as to locate our economy on a stable branch—a "yellow brick road," so to speak—that will take us to the steady state at E. We have not explicitly talked about the transversality condition, but if we had, it would have guided us to the steady state at E, where the per capita consumption can be maintained at a constant level ever after.

FIGURE 20.5



20.6 Limitations of Dynamic Analysis

The static analysis presented in Part 2 of this volume dealt only with the question of what the equilibrium position will be under certain given conditions of a model. The major query was: What values of the variables, *if attained*, will tend to perpetuate themselves? But the *attainability* of the equilibrium position was taken for granted. When we proceeded to the realm of comparative statics, in Part 3, the central question shifted to a more interesting problem: How will the equilibrium position shift in response to a certain change in a parameter? But the attainability aspect was again brushed aside. It was not until we reached the dynamic analysis in Part 5 that we looked the question of attainability squarely in the eye. Here we specifically ask: If initially we are away from an equilibrium position—say, because of a recent disequilibrating parameter change—will the various forces in the model tend to steer us toward the new equilibrium position? Furthermore, in a dynamic analysis, we also learn the particular character of the path (whether steady, fluctuating, or oscillatory) the variable will follow on its way to the equilibrium (if at all). The significance of dynamic analysis should therefore be self-evident.

However, in concluding its discussion, we should also take cognizance of the limitations of dynamic analysis. For one thing, to make the analysis manageable, dynamic models are often formulated in terms of linear equations. While simplicity may thereby be gained, the assumption of linearity will in many cases entail a considerable sacrifice of realism. Since a time path which is germane to a linear model may not always approximate that of a nonlinear counterpart, as we have seen in the price-ceiling example in Sec. 17.6, care must be exercised in the interpretation and application of the results of linear dynamic models. In this connection, however, the qualitative-graphic approach may perform an extremely valuable service, because under quite general conditions it can enable us to incorporate nonlinearity into a model without adding undue complexity to the analysis.

Another shortcoming usually found in dynamic economic models is the use of constant coefficients in differential or difference equations. Inasmuch as the primary role of the coefficients is to specify the parameters of the model, the constancy of coefficients—again assumed for the sake of mathematical manageability—essentially serves to "freeze" the economic environment of the problem under investigation. In other words, it means that the endogenous adjustment of the model is being studied in a sort of economic vacuum, such that no exogenous factors are allowed to intrude. In certain cases, of course, this problem may not be too serious, because many economic parameters do tend to stay relatively constant over long periods of time. And in some other cases, we may be able to undertake a comparative-dynamic type of analysis, to see how the time path of a variable will be affected by a change in certain parameters. Nevertheless, when we are interpreting a time path that extends into the distant future, we should always be careful not to be overconfident about the validity of the path in its more remote stretches, if simplifying assumptions of constancy have been made.

You realize, of course, that to point out its limitations as we have done here is by no means intended to disparage dynamic analysis as such. Indeed, it will be recalled that each type of analysis hitherto presented has been shown to have its own brand of limitations. As long as it is duly interpreted and properly applied, therefore, dynamic analysis—like any other type of analysis—can play an important part in the study of economic phenomena. In particular, the techniques of dynamic analysis have enabled us to extend the study of optimization into the realm of dynamic optimization in this chapter, in which the solution we seek is no longer a static optimum state, but an entire optimal time path.

The Greek Alphabet

Α	α	alpha
Β	β	beta
Γ	γ	gamma
Δ	δ	delta
Ε	ϵ	epsilon
Ζ	ζ	zeta
Η	η	eta
Θ	θ	theta
Ι	ι	iota
Κ	κ	kappa
Λ	λ	lambda
Μ	μ	mu
Ν	ν	nu
Ξ	ξ	xi
Ο	o	omicron
Π	π	pi
Ρ	ρ	rho
Σ	σ	sigma
Τ	τ	tau
Υ	υ	upsilon
Φ	ϕ (or φ)	phi
Χ	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Mathematical Symbols

1. Sets

$a \in S$	a is an element of (belongs to) set S
$b \notin S$	b is not an element of set S
$S \subset T$	set S is a subset of (is contained in) set T
$T \supset S$	set T includes set S
$A \cup B$	the union of set A and set B
$A \cap B$	the intersection of set A and set B
$\tilde{S}$	the complement of set S
$\{\ \}$ or $\varnothing$	the null set (empty set)
$\{a, b, c\}$	the set with elements a, b, and c
$\{x \mid x$ has property $P\}$	the set of all objects with property P
$\min\{a, b, c\}$	the smallest element of the specified set
R	the set of all real numbers
R^2	the two-dimensional real space
R^n	the n-dimensional real space
(x, y)	ordered pair
(x, y, z)	ordered triple
(a, b)	open interval from a to b
$[a, b]$	closed interval from a to b

2. Matrices and Determinants

A' or A^T	the transpose of matrix A
A^{-1}	the inverse of matrix A
$\|A\|$	the determinant of matrix A
$\|J\|$	Jacobian determinant
$\|H\|$	Hessian determinant
$\|\overline{H}\|$	bordered Hessian determinant
$r(A)$	the rank of matrix A
$\text{tr}A$	the trace of A

0	null matrix (zero matrix)
$u \cdot v$	the inner product (dot product) of vectors u and v
$u'v$	the scalar product of two vectors

3. Calculus

Given $y = f(x)$, a function of a single variable x:

$\lim_{x \to \infty} f(x)$	the limit of $f(x)$ as x approaches infinity
dy	the first differential of y
d^2y	the second differential of y
$\frac{dy}{dx}$ or $f'(x)$	the first derivative of the function $y = f(x)$
$\left.\frac{dy}{dx}\right\|_{x=x_0}$ or $f'(x_0)$	the first derivative evaluated at $x = x_0$
$\frac{d^2y}{dx^2}$ or $f''(x)$	the second derivative of $y = f(x)$
$\frac{d^ny}{dx^n}$ or $f^{(n)}(x)$	the nth derivative of $y = f(x)$
$\int f(x)\,dx$	indefinite integral of $f(x)$
$\int_a^b f(x)\,dx$	definite integral of $f(x)$ from $x = a$ to $x = b$

Given the function $y = f(x_1, x_2, \ldots, x_n)$:

$\frac{\partial y}{\partial x_i}$ or f_i	the partial derivative of f with respect to x_i
$\nabla f \equiv \text{grad } f$	the gradient of f
$\frac{dy}{dx_i}$	the total derivative of f with respect to x_i
$\frac{\S y}{\S x_i}$	the partial total derivative of f with respect to x_i

4. Differential and Difference Equations

$\dot{y} \equiv \frac{dy}{dt}$	the time derivative of y
Δy_t	the first difference of y_t
$\Delta^2 y_t$	the second difference of y_t
y_p	particular integral
y_c	complementary function

5. Others

$\sum_{i=1}^{n} x_i$	the sum of x_i as i ranges from 1 to n

$p \Rightarrow q$	p only if q (p implies q)
$p \Leftarrow q$	p if q (p is implied by q)
$p \Leftrightarrow q$	p if and only if q
iff	if and only if
$\|m\|$	the absolute value of the number m
$n!$	n factorial $\equiv n(n-1)(n-2)\cdots(3)(2)(1)$
$\log_b x$	the logarithm of x to base b
$\log_e x$ or $\ln x$	the natural logarithm of x (to base e)
e	the base of natural logarithms and natural exponential functions
$\sin\theta$	sine function of θ
$\cos\theta$	cosine function of θ
R_n	the remainder term when the Taylor series involves an nth-degree polynomial

A Short Reading List

Abadie, J. (ed.): *Nonlinear Programming,* North-Holland Publishing Company, Amsterdam, 1967. (A collection of papers on certain theoretical and computational aspects of nonlinear programming; Chapter 2, by Abadie, deals with the Kuhn-Tucker theorem in relation to the constraint qualification.)

Allen, R. G. D.: *Mathematical Analysis for Economists,* Macmillan & Co., Ltd., London, 1938. (A clear exposition of differential and integral calculus; determinants are discussed, but not matrices; no set theory, and no mathematical programming.)

______: *Mathematical Economics,* 2d ed., St. Martin's Press, Inc., New York, 1959. (Discusses a legion of mathematical economic models; explains linear differential and difference equations and matrix algebra.)

Almon, C.: *Matrix Methods in Economics,* Addison-Wesley Publishing Company, Inc., Reading, Mass., 1967. (Matrix methods are discussed in relation to linear-equation systems, input-output models, linear programming, and nonlinear programming. Characteristic roots and characteristic vectors are also covered.)

Baldani, J., J. Bradfield, and R. Turner: *Mathematical Economics,* The Dryden Press, Orlando, 1996.

Baumol, W. J.: *Economic Dynamics: An Introduction,* 3d ed., The Macmillan Company, New York, 1970. (Part IV gives a lucid explanation of simple difference equations; Part V treats simultaneous difference equations; differential equations are only briefly discussed.)

Braun, M.: *Differential Equations and Their Applications: An Introduction to Applied Mathematics,* 4th ed., Springer-Verlag, Inc., New York, 1993. (Contains interesting applications of differential equations, such as the detection of art forgeries, the spread of epidemics, the arms race, and the disposal of nuclear waste.)

Burmeister, E., and A. R. Dobell: *Mathematical Theories of Economic Growth,* The Macmillan Company, New York, 1970. (A thorough exposition of growth models of varying degrees of complexity.)

Chiang, Alpha C.: *Elements of Dynamic Optimization,* McGraw-Hill Book Company, 1992, now published by Waveland Press, Inc., Prospect Heights, Ill.

Clark, Colin W.: *Mathematical Bioeconomics: The Optimal Management of Renewable Resources,* 2nd ed., John Wiley & Sons, Inc., Toronto, 1990. (A thorough explanation of optimal control theory and its use in both renewable and nonrenewable resources.)

Coddington, E. A., and N. Levinson: *Theory of Ordinary Differential Equations*, McGraw-Hill Book Company, New York, 1955. (A basic mathematical text on differential equations.)

Courant, R.: *Differential and Integral Calculus* (trans. E. J. McShane), Interscience Publishers, Inc., New York, vol. I, 2d ed., 1937, vol. II, 1936. (A classic treatise on calculus.)

———, and F. John: *Introduction to Calculus and Analysis*, Interscience Publishers, Inc., New York, vol. I, 1965, vol. II, 1974. (An updated version of the preceding title.)

Dorfman, R., P. A. Samuelson, and R. M. Solow: *Linear Programming and Economic Analysis*, McGraw-Hill Book Company, New York, 1958. (A detailed treatment of linear programming, game theory, and input-output analysis.)

Franklin, J.: *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*, Springer-Verlag, Inc., New York, 1980. (A delightful presentation of mathematical programming.)

Frisch, R.: *Maxima and Minima: Theory and Economic Applications* (in collaboration with A. Nataf), Rand McNally & Company, Chicago, Ill., 1966. (A thorough treatment of extremum problems, done primarily in the classical tradition.)

Goldberg, S.: *Introduction to Difference Equations*, John Wiley & Sons, Inc., New York, 1958. (With economic applications.)

Hadley, G.: *Linear Algebra*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1961. (Covers matrices, determinants, convex sets, etc.)

———: *Linear Programming*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962. (A clearly written, mathematically oriented exposition.)

———: *Nonlinear and Dynamic Programming*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1964. (Covers nonlinear programming, stochastic programming, integer programming, and dynamic programming; computational aspects are emphasized.)

Halmos, P. R.: *Naive Set Theory*, D. Van Nostrand Company, Inc., Princeton, N.J., 1960. (An informal and hence readable introduction to the basics of set theory.)

Hands, D. Wade: *Introductory Mathematical Economics*, 2nd ed., Oxford University Press, New York, 2004.

Henderson, J. M., and R. E. Quandt: *Microeconomic Theory: A Mathematical Approach*, 3d ed., McGraw-Hill Book Company, New York, 1980. (A comprehensive mathematical treatment of microeconomic topics.)

Hoy, M., J. Livernois, C. McKenna, R. Rees, and T. Stengos: *Mathematics for Economics*, 2nd ed., The MIT Press, Cambridge, Mass. 2001.

Intriligator, M. D.: *Mathematical Optimization and Economic Theory*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1971. (A thorough discussion of optimization methods, including the classical techniques, linear and nonlinear programming, and dynamic optimization; also applications to the theories of the consumer and the firm, general equilibrium and welfare economics, and theories of growth.)

Kemeny, J. G., J. L. Snell, and G. L. Thompson: *Introduction to Finite Mathematics*, 3d ed., Prentice Hall, Inc., Englewood Cliffs, N.J., 1974. (Covers such topics as sets, matrices, probability, and linear programming.)

Klein, Michael W.: *Mathematical Methods for Economics*, 2nd ed., Addison-Wesley Publishing Company, Inc., Reading, Mass. 2002.

Koo, D.: *Elements of Optimization: With Applications in Economics and Business*, Springer-Verlag, Inc., New York, 1977. (Clear discussion of classical optimization methods, mathematical programming as well as optimal control theory.)

Koopmans, T. C. (ed.): *Activity Analysis of Production and Allocation*, John Wiley & Sons, Inc., New York, 1951, reprinted by Yale University Press, 1972. (Contains a number of important papers on linear programming and activity analysis.)

———: *Three Essays on the State of Economic Science*, McGraw-Hill Book Company, New York, 1957. (The first essay contains a good exposition of convex sets; the third essay discusses the interaction of *tools* and *problems* in economics.)

Lambert, Peter J., *Advanced Mathematics for Economists: Static and Dynamic Optimization*, Blackwell Publishers, New York, 1985.

Leontief, W. W.: *The Structure of American Economy, 1919–1939*, 2d ed., Oxford University Press, Fair Lawn, N.J., 1951. (The pioneering work in input-output analysis.)

Samuelson, P. A.: *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Mass., 1947. (A classic in mathematical economics, but very difficult to read.)

Silberberg, Eugene, and Wing Suen: *The Structure of Economics: A Mathematical Analysis*, 3rd ed., McGraw-Hill Book Company, New York, 2001. (Primarily a microeconomic focus, this book has a strong discussion of the envelope theorem and a wide variety of applications.)

Sydsæter, Knut, and Peter Hammond: *Essential Mathematics for Economic Analysis*, Prentice Hall, Inc., London, 2002.

Takayama, A.: *Mathematical Economics*, 2nd ed., The Dryden Press, Hinsdale, Ill., 1985. (Gives a comprehensive treatment of economic theory in mathematical terms, with concentration on two specific topics: competitive equilibrium and economic growth.)

Thomas, G. B., and R. L. Finney: *Calculus and Analytic Geometry*, 9th ed., Addison-Wesley Publishing Company, Inc., Reading, Mass., 1996. (A clearly written introduction to calculus.)

Answers to Selected Exercises

Exercise 2.3

1. (*a*) $\{x \mid x > 34\}$
3. (*a*) $\{2, 4, 6, 7\}$ (*c*) $\{2, 6\}$ (*e*) $\{2\}$
8. There are 16 subsets.
9. *Hint:* Distinguish between the two symbols $\notin$ and $\not\subset$.

Exercise 2.4

1. (*a*) $\{(3, a), (3, b), (6, a), (6, b), (9, a), (9, b)\}$
3. No.
5. Range $= \{y \mid 8 \leq y \leq 32\}$

Exercise 2.5

2. (*a*) and (*b*) differ in the sign of the slope measure; (*a*) and (*c*) differ in the vertical intercept.
4. When negative values are permissible, quadrant III has to be used too.
5. (*a*) x^{19}
6. (*a*) x^6

Exercise 3.2

1. $P^* = 2\frac{3}{11}$, and $Q^* = 14\frac{2}{11}$
3. *Note:* In 2(*a*), $c = 10$ (not 6).
5. *Hint:* $b + d = 0$ implies $d = -b$.

Exercise 3.3

1. (*a*) $x_1^* = 5$, and $x_2^* = 3$
3. (*a*) $(x - 6)(x + 1)(x - 3) = 0$, or $x^3 - 8x^2 + 9x + 18 = 0$
5. (*a*) -1, 2, and 3 (*c*) -1, $\frac{1}{2}$, and $-\frac{1}{4}$

Exercise 3.4

3. $P_1^* = 3\frac{6}{17}$ $P_2^* = 3\frac{8}{17}$ $Q_1^* = 11\frac{7}{17}$ $Q_2^* = 8\frac{7}{17}$

Exercise 3.5

1. (*b*) $Y^* = (a - bd + I_0 + G_0)/[1 - b(1 - t)]$

$T^* = [d(1 - b) + t(a + I_0 + G_0)]/[1 - b(1 - t)]$

$C^* = [a - bd + b(1 - t)(I_0 + G_0)]/[1 - b(1 - t)]$

3. *Hint*: After substituting the last two equations into the first, consider the resulting equation as a quadratic equation in the variable $w \equiv Y^{1/2}$. Only one root is acceptable, $w_1^* = 11$, giving $Y^* = 121$ and $C^* = 91$. The other root leads to a negative C^*.

Exercise 4.1

1. The elements in the (column) vector of constants are: $0, a, -c$.

Exercise 4.2

1. (*a*) $\begin{bmatrix} 7 & 3 \\ 9 & 7 \end{bmatrix}$ (*c*) $\begin{bmatrix} 21 & -3 \\ 18 & 27 \end{bmatrix}$

3. In this special case, AB happens to be equal to $BA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

4. (*b*) $\underset{(2\times 2)}{\begin{bmatrix} 49 & 3 \\ 4 & 3 \end{bmatrix}}$ (*c*) $\underset{(2\times 1)}{\begin{bmatrix} 3x + 5y \\ 4x + 2y - 7z \end{bmatrix}}$

6. (*a*) $x_2 + x_3 + x_4 + x_5$ (*c*) $b(x_1 + x_2 + x_3 + x_4)$

7. (*b*) $\sum_{i=2}^{4} a_i(x_{i+1} + i)$ (*d*) *Hint*: $x^0 = 1$ for $x \neq 0$

Exercise 4.3

1. (*a*) $uv' = \begin{bmatrix} 15 & 5 & -5 \\ 3 & 1 & -1 \\ 9 & 3 & -3 \end{bmatrix}$ (*c*) $xx' = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix}$

(*e*) $u'v = 13$ (*g*) $u'u = 35$

3. (*a*) $\sum_{i=1}^{n} P_iQ_i$ (*b*) $P \cdot Q$ or $P'Q$ or $Q'P$

5. (*a*) $2v = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$ (*c*) $u - v = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$

7. (*a*) $d = \sqrt{27}$

9. (*c*) $d(v, 0) = (v \cdot v)^{1/2}$

Exercise 4.4

1. (*a*) $\begin{bmatrix} 5 & 17 \\ 11 & 17 \end{bmatrix}$

2. No; it should be $A - B = -B + A$.
4. (*a*) $k(A + B) = k[a_{ij} + b_{ij}] = [ka_{ij} + kb_{ij}] = [ka_{ij}] + [kb_{ij}] = k[a_{ij}] + k[b_{ij}] = kA + kB$ (Can you justify each step?)

Exercise 4.5

1. (*a*) $AI_3 = \begin{bmatrix} -1 & 5 & 7 \\ 0 & -2 & 4 \end{bmatrix}$ (*c*) $I_2x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
3. (*a*) 5×3 (*c*) 2×1
4. *Hint:* Multiply the given diagonal matrix by itself, and examine the resulting product matrix for conditions for idempotency.

Exercise 4.6

1. $A' = \begin{bmatrix} 0 & -1 \\ 4 & 3 \end{bmatrix}$ and $B' = \begin{bmatrix} 3 & 0 \\ -8 & 1 \end{bmatrix}$
3. *Hint:* Define $D \equiv AB$, and apply (4.11).
5. *Hint:* Define $D = AB$, and apply (4.14).

Exercise 5.1

1. (*a*) (5.2) (*c*) (5.3) (*e*) (5.3)
3. (*a*) Yes. (*d*) No.
5. (*a*) $r(A) = 3$; A is nonsingular. (*b*) $r(B) = 2$; B is singular.

Exercise 5.2

1. (*a*) -6 (*c*) 0 (*e*) $3abc - a^3 - b^3 - c^3$
3. $|M_b| = \begin{vmatrix} d & f \\ g & i \end{vmatrix}$ $|C_b| = -\begin{vmatrix} d & f \\ g & i \end{vmatrix}$
4. (*a*) *Hint:* Expand by the third column.
5. 20 (not -20)

Exercise 5.3

3. (*a*) Property IV. (*b*) Property III (applied to both rows).
4. (*a*) Singular. (*c*) Singular.
5. (*a*) Rank < 3 (*c*) Rank < 3
7. A is nonsingular because $|A| = 1 - b \neq 0$.

Exercise 5.4

1. $\sum_{i=1}^{4} a_{i3} |C_{i2}|$ $\sum_{j=1}^{4} a_{2j} |C_{4j}|$
3. (*a*) Interchange the two diagonal elements of A; multiply the two off-diagonal elements of A by -1.
   (*b*) Divide by $|A|$.

4. (*a*) $E^{-1} = \frac{1}{20}\begin{bmatrix} 3 & 2 & -3 \\ -7 & 2 & 7 \\ -6 & -4 & 26 \end{bmatrix}$ (*c*) $G^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

Exercise 5.5

1. (*a*) $x_1^* = 4$, and $x_2^* = 3$ (*c*) $x_1^* = 2$, and $x_2^* = 1$
2. (*a*) $A^{-1} = \frac{1}{7}\begin{bmatrix} 1 & 2 \\ -2 & 3 \end{bmatrix}$; $x^* = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ (*c*) $A^{-1} = \frac{1}{15}\begin{bmatrix} 1 & 7 \\ -1 & 8 \end{bmatrix}$; $x^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$
3. (*a*) $x_1^* = 2, x_2^* = 0, x_3^* = 1$ (*c*) $x^* = 0, y^* = 3, z^* = 4$
4. *Hint:* Apply (5.8) and (5.13).

Exercise 5.6

1. (*a*) $A^{-1} = \frac{1}{1-b+bt}\begin{bmatrix} 1 & 1 & -b \\ b(1-t) & 1 & -b \\ t & t & 1-b \end{bmatrix}$

$$\begin{bmatrix} Y^* \\ C^* \\ T^* \end{bmatrix} = \frac{1}{1-b+bt}\begin{bmatrix} I_0 + G_0 + a - bd \\ b(1-t)(I_0+G_0) + a - bd \\ t(I_0+G_0) + at + d(1-b) \end{bmatrix}$$

(*b*) $|A| = 1 - b + bt$ $|A_1| = I_0 + G_0 - bd + a$

$|A_2| = a - bd + b(1-t)(I_0 + G_0)$ $|A_3| = d(1-b) + t(a + I_0 + G_0)$

Exercise 5.7

1. $x_1^* = 69.53$, $x_2^* = 57.03$, and $x_3^* = 42.58$
3. (*a*) $A = \begin{bmatrix} 0.10 & 0.50 \\ 0.60 & 0 \end{bmatrix}$; the matrix equation is $\begin{bmatrix} 0.90 & -0.50 \\ -0.60 & 1.00 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1{,}000 \\ 2{,}000 \end{bmatrix}$.

   (*c*) $x_1^* = 3{,}333\frac{1}{3}$, and $x_2^* = 4{,}000$
4. Element 0.33: 33¢ of Commodity II is needed as input for producing \$1 of Commodity I.

Exercise 6.2

1. (*a*) $\Delta y/\Delta x = 8x + 4\Delta x$ (*b*) $dy/dx = 8x$ (*c*) $f'(3) = 24$, $f'(4) = 32$
3. (*a*) $\Delta y/\Delta x = 5$; a constant function.

Exercise 6.4

1. Left-side limit = right-side limit = 15; the limit is 15.
3. (*a*) 5 (*b*) 5

Exercise 6.5

1. (*a*) $-3/4 < x$ (*c*) $x < 1/2$
3. (*a*) $-7 < x < 5$ (*c*) $-4 \le x \le 1$

Exercise 6.6

1. (*a*) 7 (*c*) 17
3. (*a*) $2\frac{1}{2}$ (*c*) 2

Exercise 6.7

2. (*a*) $N^2 - 5N - 2$ (*b*) Yes. (*c*) Yes.

3. (*a*) $(N + 2)/(N^2 + 2)$ (*b*) Yes. (*c*) Continuous in the domain.

6. Yes; each function is continuous and smooth.

Exercise 7.1

1. (*a*) $dy/dx = 12x^{11}$ (*c*) $dy/dx = 35x^4$ (*e*) $dw/du = -2u^{-1/2}$

3. (*a*) $f'(x) = 18$; $f'(1) = f'(2) = 18$

(*c*) $f'(x) = 10x^{-3}$; $f'(1) = 10$, $f'(2) = 1\frac{1}{4}$

Exercise 7.2

1. $\text{VC} = Q^3 - 5Q^2 + 12Q$; $\dfrac{d\text{VC}}{dQ} = 3Q^2 - 10Q + 12$ is the MC function.

3. (*a*) $3(27x^2 + 6x - 2)$ (*c*) $12x(x + 1)$ (*e*) $-x(9x + 14)$

4. (*b*) $\text{MR} = 60 - 6Q$

7. (*a*) $(x^2 - 3)/x^2$ (*c*) $30/(x + 5)^2$

8. (*a*) a (*c*) $-a/(ax + b)^2$

Exercise 7.3

1. $-2x[3(5 - x^2)^2 + 2]$

3. (*a*) $18x(3x^2 - 13)^2$ (*c*) $5a(ax + b)^4$

5. $x = \frac{1}{7}y - 3$, $dx/dy = \frac{1}{7}$

Exercise 7.4

1. (*a*) $\partial y/\partial x_1 = 6x_1^2 - 22x_1x_2$, and $\partial y/\partial x_2 = -11x_1^2 + 6x_2$

(*c*) $\partial y/\partial x_1 = 2(x_2 - 2)$, and $\partial y/\partial x_2 = 2x_1 + 3$

3. (*a*) 12 (*c*) 10/9

5. (*a*) $U_1 = 2(x_1 + 2)(x_2 + 3)^3$, and $U_2 = 3(x_1 + 2)^2(x_2 + 3)^2$

Exercise 7.5

1. $\partial Q^*/\partial a = d/(b + d) > 0$ $\partial Q^*/\partial b = -d(a + c)/(b + d)^2 < 0$

$\partial Q^*/\partial c = -b/(b + d) < 0$ $\partial Q^*/\partial d = b(a + c)/(b + d)^2 > 0$

2. $\partial Y^*/\partial I_0 = \partial Y^*/\partial \alpha = 1/(1 - \beta + \beta\delta) > 0$

Exercise 7.6

1. (*a*) $|J| = 0$; the functions are dependent.

(b) $|J| = -20x_2$; the functions are independent.

Exercise 8.1

1. (*a*) $dy = -3(x^2 + 1)\,dx$ (*c*) $dy = [(1 - x^2)/(x^2 + 1)^2]\,dx$

3. (*a*) $dC/dY = b$, and $C/Y = (a + bY)/Y$

Exercise 8.2

2. (*a*) $dz = (6x + y)\,dx + (x - 6y^2)\,dy$

3. (a) $dy = [x_2/(x_1 + x_2)^2]\,dx_1 - [x_1/(x_1 + x_2)^2]\,dx_2$
4. $\varepsilon_{QP} = 2bP^2/(a + bP^2 + R^{1/2})$
6. $\varepsilon_{XP} = -2/(Y_f^{1/2}P^2 + 1)$

Exercise 8.3

3. (a) $dy = 3[(2x_2 - 1)(x_3 + 5)\,dx_1 + 2x_1(x_3 + 5)\,dx_2 + x_1(2x_2 - 1)\,dx_3]$
4. *Hint:* Apply the definitions of differential and total differential.

Exercise 8.4

1. (a) $dz/dy = x + 10y + 6y^2 = 28y + 9y^2$
   (c) $dz/dy = -15x + 3y = 108y - 30$
3. $dQ/dt = [a\alpha A/K + b\beta A/L + A'(t)]K^{\alpha}L^{\beta}$
4. (b) $\S W/\S u = 10uf_1 + f_2 \quad \S W/\S v = 3f_1 - 12v^2 f_2$

Exercise 8.5

5. (a) Defined; $dy/dx = -(3x^2 - 4xy + 3y^2)/(-2x^2 + 6xy) = -9/8$
   (b) Defined: $dy/dx = -(4x + 4y)/(4x - 4y^3) = 2/13$
7. The condition $F_y \neq 0$ is violated at (0, 0).
8. The product of partial derivatives is equal to -1.

Exercise 8.6

1. (c) $(dY^*/dG_0) = 1/(S' + T' - I') > 0$
3. $(\partial P^*/\partial Y_0) = D_{Y_0}/(S_{P^*} - D_{P^*}) > 0 \quad (\partial Q^*/\partial Y_0) = D_{Y_0}S_{P^*}/(S_{P^*} - D_{P^*}) > 0$
   $(\partial P^*/\partial T_0) = -S_{T_0}/(S_{P^*} - D_{P^*}) > 0 \quad (\partial Q^*/\partial T_0) = -S_{T_0}D_{P^*}/(S_{P^*} - D_{P^*}) < 0$

Exercise 9.2

1. (a) When $x = 2$, $y = 15$ (a relative maximum).
   (c) When $x = 0$, $y = 3$ (a relative minimum).
2. (a) The critical value $x = -1$ lies outside the domain; the critical value $x = 1$ leads to $y = 3$ (a relative minimum).
4. (d) The elasticity is one.

Exercise 9.3

1. (a) $f''(x) = 2a$, $f'''(x) = 0$ (c) $f''(x) = 6(1 - x)^{-3}$, $f'''(x) = 18(1 - x)^{-4}$
3. (b) A straight line.
5. Every point on $f(x)$ is a stationary point, but the only stationary point on $g(x)$ we know of is at $x = 3$.

Exercise 9.4

1. (a) $f(2) = 33$ is a maximum.
   (c) $f(1) = 5\frac{1}{3}$ is a maximum; $f(5) = -5\frac{1}{3}$ is a minimum.
2. *Hint:* First write an area function A in terms of one variable (either L or W) alone.
3. (d) $Q^* = 11$ (e) Maximum profit $= 111\frac{1}{3}$

5. (*a*) $k < 0$ (*b*) $h < 0$ (*c*) $j > 0$

7. (*b*) S is maximized at the output level 20.37 (approximately).

Exercise 9.5

1. (*a*) 120 (*c*) 4 (*e*) $(n+2)(n+1)$

2. (*a*) $1 + x + x^2 + x^3 + x^4$

3. (*b*) $-63 - 98x - 62x^2 - 18x^3 - 2x^4 + R_4$

Exercise 9.6

1. (*a*) $f(0) = 0$ is an inflection point. (*c*) $f(0) = 5$ is a relative minimum.

2. (*b*) $f(2) = 0$ is a relative minimum.

Exercise 10.1

1. (*a*) Yes. (*b*) Yes.

3. (*a*) $5e^{5t}$ (*c*) $-12e^{-2t}$

5. (*a*) The curve with $a = -1$ is the mirror image of the curve with $a = 1$ with reference to the horizontal axis.

Exercise 10.2

1. (*a*) 7.388 (*b*) 1.649

2. (*c*) $1 + 2x + \frac{1}{2!}(2x)^2 + \frac{1}{3!}(2x)^3 + \cdots$

3. (*a*) $\$70e^{0.12}$ (*b*) $\$690e^{0.10}$

Exercise 10.3

1. (*a*) 4 (*c*) 4

2. (*a*) 7 (*c*) -3 (*e*) 6

3. (*a*) 26 (*c*) $\ln 3 - \ln B$ (*f*) 3

Exercise 10.4

1. The requirement prevents the function from degenerating into a constant function.

3. *Hint:* Take log to base b.

4. (*a*) $y = e^{(3\ln 8)t}$ or $y = e^{6.2383t}$ (*c*) $y = 5e^{(\ln 5)t}$ or $y = 5e^{1.6094t}$

5. (*a*) $t = (\ln y)/(\ln 7)$ or $t = 0.5139 \ln y$
   (*c*) $t = 3\ln(9y)/(\ln 15)$ or $t = 1.1078 \ln(9y)$

6. (*a*) $r = \ln 1.05$ (*c*) $r = 2\ln 1.03$

Exercise 10.5

1. (*a*) $2e^{2t+4}$ (*c*) $2te^{t^2+1}$ (*e*) $(2ax+b)e^{ax^2+bx+c}$

3. (*a*) $5/t$ (*c*) $1/(t+19)$ (*e*) $1/[x(1+x)]$

5. *Hint:* Use (10.21), and apply the chain rule.

7. (*a*) $3(8-x^2)/[(x+2)^2(x+4)^2]$

Exercise 10.6

1. $t^* = 1/r^2$
2. $d^2A/dt^2 = -A(\ln 2)/4\sqrt{t^3} < 0$

Exercise 10.7

1. (a) $2/t$ (c) $\ln b$ (e) $1/t - \ln 3$
3. $r_y = kr_x$
7. $|\varepsilon_d| = n$
11. $r_Q = \varepsilon_{QK} r_K + \varepsilon_{QL} r_L$

Exercise 11.2

1. $z^* = 3$ is a minimum.
3. $z^* = c$, which is a minimum in case (a), a maximum in case (b), and a saddle point in case (c).
5. (a) Any pair (x, y) other than $(2, 3)$ yields a positive z value.
   (b) Yes. (c) No. (d) Yes ($d^2z = 0$).

Exercise 11.3

1. (a) $q = 4u^2 + 4uv + 3v^2$ (c) $q = 5x^2 + 6xy$
3. (a) Positive definite. (c) Neither.
5. (a) Positive definite. (c) Negative definite. (e) Positive definite.
6. (a) $r_1, r_2 = \frac{1}{2}(7 \pm \sqrt{17})$; $u'Du$ is positive definite.
   (c) $r_1, r_2 = \frac{1}{2}(5 \pm \sqrt{61})$, $u'Fu$ is indefinite.
7. $v_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$, $v_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$

Exercise 11.4

1. $z^* = 0$ (minimum)
3. $z^* = -11/40$ (minimum)
4. $z^* = 2 - e$ (minimum), attained at $(x^*, y^*, w^*) = (0, 0, 1)$
5. (b) *Hint:* See (11.16).
6. (a) $r_1 = 2$ $r_2 = 4 + \sqrt{6}$ $r_3 = 4 - \sqrt{6}$

Exercise 11.5

1. (a) Strictly convex. (c) Strictly convex.
2. (a) Strictly concave. (c) Neither.
3. No.
5. (a) Disk. (b) Yes.
7. (a) Convex combination, with $\theta = 0.5$. (b) Convex combination, with $\theta = 0.2$.

Exercise 11.6

1. (*a*) No. (*b*) $Q_1^* = P_{10}/4$ and $Q_2^* = P_{20}/4$
3. $|\varepsilon_{d1}| = 1\frac{5}{8}$ $|\varepsilon_{d2}| = 1\frac{1}{3}$ $|\varepsilon_{d3}| = 1\frac{1}{2}$
5. (*a*) $\pi = P_0 Q(a, b)(1 + \frac{1}{2} i_0)^{-2} - P_{a0} a - P_{b0} b$

Exercise 11.7

1. $(\partial a^*/\partial P_{a0}) = P_0 Q_{bb} e^{-rt}/|J| < 0$ $(\partial b^*/\partial P_{a0}) = -P_0 Q_{ab} e^{-rt}/|J| < 0$
2. (*a*) Four. (*b*) $(\partial a^*/\partial P_0) = (Q_b Q_{ab} - Q_a Q_{bb}) P_0 (1 + i_0)^{-2}/|J| > 0$
   (*c*) $(\partial a^*/\partial i_0) = (Q_a Q_{bb} - Q_b Q_{ab}) P_0^2 (1 + i_0)^{-3}/|J| < 0$

Exercise 12.2

1. (*a*) $z^* = 1/2$, attained when $\lambda^* = 1/2$, $x^* = 1$, and $y^* = 1/2$
   (*c*) $z^* = -19$, attained when $\lambda^* = -4$, $x^* = 1$, and $y^* = 5$
4. $Z_\lambda = -G(x, y) = 0$ $Z_x = f_x - \lambda G_x = 0$ $Z_y = f_y - \lambda G_y = 0$
5. *Hint:* Distinguish between identical equality and conditional equality.

Exercise 12.3

1. (*a*) $|\overline{H}| = 4$; z^* is a maximum. (*c*) $|\overline{H}| = -2$; z^* is a minimum.

Exercise 12.4

2. (*a*) Quasiconcave, but not strictly so. (*c*) Strictly quasiconcave.
4. (*a*) Neither. (*c*) Quasiconvex, but not quasiconcave.
5. *Hint:* Review Sec. 9.4.
7. *Hint:* Use either (12.21) or (12.25').

Exercise 12.5

1. (*b*) $\lambda^* = 3$, $x^* = 16$, $y^* = 11$ (*c*) $|\overline{H}| = 48$; condition is satisfied.
3. $(\partial x^*/\partial B) = 1/2P_x > 0$ $(\partial x^*/\partial P_x) = -(B + P_y)/2P_x^2 < 0$
   $(\partial x^*/\partial P_y) = 1/2P_x > 0$ etc.
5. Not valid.
7. No to both (*a*) and (*b*)—see (12.32) and (12.33').

Exercise 12.6

1. (*a*) Homogeneous of degree one. (*c*) Not homogeneous.
   (*e*) Homogeneous of degree two.
4. They are true.
7. (*a*) Homogeneous of degree $a + b + c$.
8. (*a*) $j^2 Q = g(jK, jL)$ (*b*) *Hint:* Let $j = 1/L$.
   (*d*) Homogeneous of degree one in K and L.

Exercise 12.7

1. (*a*) 1 : 2 : 3 (*b*) 1 : 4 : 9
2. *Hint:* Review Figs. 8.2 and 8.3.
4. *Hint:* This is a total derivative.
6. (*a*) Downward-sloping straight lines. (*b*) $\sigma \to \infty$ as $\rho \to -1$
8. (*a*) 7 (*c*) $\ln 5 - 1$

Exercise 13.1

3. The conditions $x_j(\partial Z/\partial x_j) = 0$ and the conditions $\lambda_i(\partial Z/\partial \lambda_i) = 0$ can be condensed.
5. Consistent.

Exercise 13.2

1. No qualifying arc can be found for a test vector such as $(dx_1, dx_2) = (1, 0)$.
3. $(x_1^*, x_2^*) = (0, 0)$ is a cusp. The constraint qualification is satisfied (all test vectors are horizontal and pointing eastward); the Kuhn-Tucker conditions are satisfied, too.
4. All the conditions can be satisfied by choosing $y_0^* = 0$ and $y_1^* \geq 0$.

Exercise 13.4

2. (*a*) Yes. (*b*) Yes. (*c*) No.
4. (*a*) Yes. (*b*) Yes.

Exercise 14.2

1. (*a*) $-8x^{-2} + c$, $(x \neq 0)$ (*c*) $\frac{1}{6}x^6 - \frac{3}{2}x^2 + c$
2. (*a*) $13e^x + c$ (*c*) $5e^x - 3x^{-1} + c$, $(x \neq 0)$
3. (*a*) $3\ln|x| + c$, $(x \neq 0)$ (*c*) $\ln(x^2 + 3) + c$
4. (*a*) $\frac{2}{3}(x+1)^{3/2}(x+3) - \frac{4}{15}(x+1)^{5/2} + c$

Exercise 14.3

1. (*a*) $4\frac{1}{3}$ (*b*) $3\frac{1}{4}$ (*e*) $2\left(\frac{a}{3} + c\right)$
2. (*a*) $\frac{1}{2}(e^{-2} - e^{-4})$ (*c*) $e^2\left(\frac{1}{2}e^4 - \frac{1}{2}e^2 + e - 1\right)$
3. (*b*) Underestimate. (*e*) $f(x)$ is Riemann integrable.

Exercise 14.4

1. None.
2. (*a*), (*c*), (*d*) and (*e*).
3. (*a*), (*c*) and (*d*) convergent; (*e*) divergent.

Exercise 14.5

1. (*a*) $R(Q) = 14Q^2 - \frac{10}{3}e^{0.3Q} + \frac{10}{3}$ (*b*) $R(Q) = 10Q/(1+Q)$
3. (*a*) $K(t) = 9t^{4/3} + 25$
5. (*a*) 29,000

Exercise 14.6

1. Capital alone is considered. Since labor is normally necessary for production as well, the underlying assumption is that K and L are always used in a fixed proportion.
3. *Hint:* Use (6.8).
4. *Hint:* $\ln u - \ln v = \ln \frac{u}{v}$

Exercise 15.1

1. (*a*) $y(t) = -e^{-4t} + 3$ (*c*) $y(t) = \frac{3}{2}(1 - e^{-10t})$
3. (*a*) $y(t) = 4(1 - e^{-t})$ (*c*) $y(t) = 6e^{5t}$ (*e*) $y(t) = 8e^{7t} - 1$

Exercise 15.2

1. The D curve should be steeper.
3. The price adjustment mechanism generates a differential equation.
5. (*a*) $P(t) = A \exp\left(-\frac{\beta + \delta}{\eta} t\right) + \frac{\alpha}{\beta + \delta}$ (*b*) Yes.

Exercise 15.3

1. $y(t) = Ae^{-5t} + 3$
3. $y(t) = e^{-t^2} + \frac{1}{2}$
5. $y(t) = e^{-6t} - \frac{1}{7}e^{t}$
6. *Hint:* Review Sec. 14.2, Example 17.

Exercise 15.4

1. (*a*) $y(t) = (c/t^3)^{1/2}$ (*c*) $yt + y^2 t = c$

Exercise 15.5

1. (*a*) Separable; linear when written as $\frac{dy}{dt} + \frac{1}{t}y = 0$

   (*c*) Separable; reducible to a Bernoulli equation.
3. $y(t) = (A - t^2)^{1/2}$

Exercise 15.6

1. (*a*) Upward-sloping phase line; dynamically unstable equilibrium.

   (*c*) Downward-sloping phase line; dynamically stable equilibrium.
3. The sign of the derivative measures the slope of the phase line.

Exercise 15.7

1. $r_k = r_K - r_L$ [*cf.* (10.25)]
4. (*a*) Plot $(3 - y)$ and $\ln y$ as two separate curves, and then subtract. A single equilibrium exists (at a y value between 1 and 3) and is dynamically stable.

Exercise 16.1

1. (*a*) $y_p = 2/5$ (*c*) $y_p = 3$ (*e*) $y_p = 6t^2$
3. (*a*) $y(t) = 6e^t + e^{-4t} - 3$ (*c*) $y(t) = e^t + te^t + 3$
6. *Hint:* Apply L'Hôpital's rule.

Exercise 16.2

1. (*a*) $\frac{3}{2} \pm \frac{1}{2}\sqrt{3}i$ (*c*) $-\frac{1}{4} \pm \frac{3}{4}\sqrt{7}i$
3. (*b*) *Hint.* When $\theta = \pi/4$, line OP is a 45° line.
5. (*a*) $\frac{d}{d\theta}\sin f(\theta) = f'(\theta)\cos f(\theta)$ (*b*) $\frac{d}{d\theta}\cos\theta^3 = -3\theta^2\sin\theta^3$
7. (*a*) $\sqrt{3} + i$ (*c*) $1 - i$

Exercise 16.3

1. $y(t) = e^{2t}(3\cos 2t + \frac{1}{2}\sin 2t)$
3. $y(t) = e^{-3t/2}\left(-\cos\frac{\sqrt{7}}{2}t + \frac{\sqrt{7}}{7}\sin\frac{\sqrt{7}}{2}t\right) + 3$
5. $y(t) = \frac{2}{3}\cos 3t + \sin 3t + \frac{1}{3}$

Exercise 16.4

1. (*a*) $P'' + \frac{m - u}{n - w}P' - \frac{\beta + \delta}{n - w} = -\frac{\alpha + \gamma}{n - w} \quad (n \neq w)$ (*b*) $\overline{P}_p = \frac{\alpha + \gamma}{\beta + \delta}$
3. (*a*) $P(t) = e^{t/2}(2\cos\frac{3}{2}t + 2\sin\frac{3}{2}t) + 2$

Exercise 16.5

1. (*a*) $\frac{d\pi}{dt} + j(1 - g) = j(\alpha - T - \beta U)$

   (*b*) No complex roots; no fluctuation.
3. (*c*) Both are first-order differential equations. (*d*) $g \neq 1$
4. (*a*) $\pi(t) = e^{-t/4}\left(A_5\cos\frac{\sqrt{2}}{4}t + A_6\sin\frac{\sqrt{2}}{4}t\right) + m$ (*c*) $\overline{P} = m$; $\overline{U} = \frac{1}{18} - \frac{2}{9}m$

Exercise 16.6

2. (*a*) $y_p = t - 2$ (*c*) $y_p = \frac{1}{4}e^t$

Exercise 16.7

1. (*a*) $y_p = 4$ (*c*) $y_p = \frac{1}{18}t^2$
3. (*a*) Divergent. (*c*) Convergent.

Exercise 17.2

1. (*a*) $y_{t+1} = y_t + 7$ (*c*) $y_{t+1} = 3y_t - 9$
3. (*a*) $y_t = 10 + t$ (*c*) $y_t = y_0\alpha^t - \beta(1 + \alpha + \alpha^2 + \cdots + \alpha^{t-1})$

Exercise 17.3

1. (*a*) Nonoscillatory; divergent. (*c*) Oscillatory; convergent.
3. (*a*) $y_t = -8(1/3)^t + 9$ (*c*) $y_t = -2(-1/4)^t + 4$

Exercise 17.4

1. $Q_t = \alpha - \beta(P_0 - \overline{P})(-\delta/\beta)^t - \beta\overline{P}$
3. (*a*) $\overline{P} = 3$; explosive oscillation. (*c*) $\overline{P} = 2$; uniform oscillation.
5. The lag in the supply function.

Exercise 17.5

1. $a = -1$
3. $P_t = (P_0 - 3)(-1.4)^t + 3$, with explosive oscillation.

Exercise 17.6

1. No.
2. (*b*) Nonoscillatory, explosive downward movement.
   (*d*) Damped, steady downward movement toward R.
4. (*a*) At first downward-sloping, then becoming horizontal.

Exercise 18.1

1. (*a*) $\frac{1}{2} \pm \frac{1}{2}i$ (*c*) $\frac{1}{2}, -1$
3. (*a*) 4 (stationary) (*c*) 5 (stationary)
4. (*b*) $y_t = \sqrt{2}^t\left(2\cos\frac{\pi}{4}t + \sin\frac{\pi}{4}t\right) + 1$

Exercise 18.2

1. (*a*) Subcase 1D. (*c*) Subcase 1C.
3. *Hint:* Use (18.16).

Exercise 18.3

3. Possibilities *v*, *viii*, *x*, and *xi* will become feasible.
4. (*a*) $p_{t+2} - [2 - j(1-g) - \beta k]p_{t+1} + [1 - j(1-g) - \beta k(1-j)]p_t = j\beta km$
   (*c*) $\beta k \gtreqless 4$

Exercise 18.4

1. (*a*) 1 (*c*) $3t^2 + 3t + 1$
3. (*a*) $y_p = \frac{1}{4}t$ (*c*) $y_p = 2 - t + t^2$
5. (*a*) $1/2, -1$ and 1

Exercise 19.2

2. $b^3 + b^2 - 3b + 2 = 0$

3. (*a*) $x_t = -(3)^t + 4(-2)^t + 7 \qquad y_t = 2(3)^t + 2(-2)^t + 5$

4. (*a*) $x(t) = 4e^{-2t} - 3e^{-3t} + 12 \qquad y(t) = -e^{-2t} + e^{-3t} + 4$

Exercise 19.3

2. (*c*) $\beta = (\delta I - A)^{-1}u$

3. (*c*) $\beta = (\rho I + I - A)^{-1}\lambda$

5. (*c*) $x_1(t) = 4e^{-4t/10} + 2e^{-11t/10} + \frac{17}{6}e^{t/10}; \quad x_2(t) = 3e^{-4t/10} - 2e^{-11t/10} + \frac{19}{6}e^{t/10}$

Exercise 19.4

4. (*a*)
$$\begin{bmatrix} \pi_t \\ U_t \end{bmatrix} = \begin{bmatrix} A_1 \\ \dfrac{23 - \sqrt{193}}{48} A_1 \end{bmatrix} \left(\frac{33 + \sqrt{193}}{64} \right)^t + \begin{bmatrix} A_2 \\ \dfrac{23 + \sqrt{193}}{48} A_2 \end{bmatrix} \left(\frac{33 - \sqrt{193}}{64} \right)^t + \begin{bmatrix} \mu \\ \dfrac{1}{6}(1 - \mu) \end{bmatrix}$$

Exercise 19.5

1. The single equation can be rewritten as two first-order equations.
2. Yes.
4. (*a*) Saddle point.

Exercise 19.6

1. (*a*) $|J_E| = 1$ and $\text{tr } J_E = 2$; locally unstable node.
   (*c*) $|J_E| = 5$ and $\text{tr } J_E = -1$; locally stable focus
2. (*a*) Locally a saddle point. (*c*) Locally stable node or stable focus.
4. (*a*) The $x' = 0$ and $y' = 0$ curves coincide, and provide a lineful of equilibrium points.

Exercise 20.2

1. $\lambda^* = 1 - t \qquad u^* = \dfrac{1 - t}{2} \qquad y^* = \dfrac{t}{2} - \dfrac{t^2}{4} + 2$

6. $\lambda^*(t) = 3e^{4-t} - 3 \qquad u^*(t) = 2 \qquad y^*(t) = 7e^t - 2$

Exercise 20.4

1. $\lambda^* = \delta/(\delta^2 + \alpha) \qquad K^* = 1/2(\delta^2 + \alpha)$