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[†]

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

[†] 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[†]

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

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

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

[†] 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

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

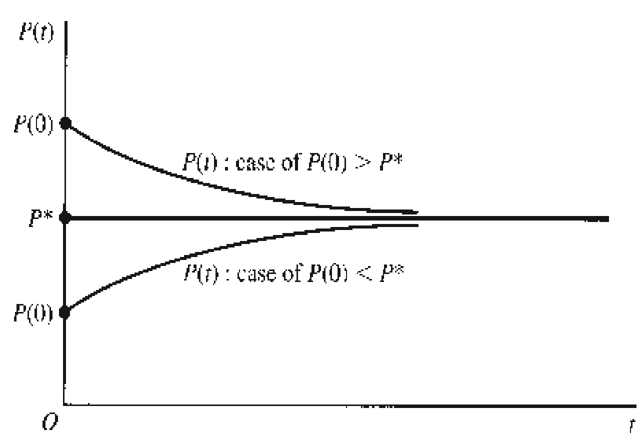
The Dynamic Stability of Equilibrium

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

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

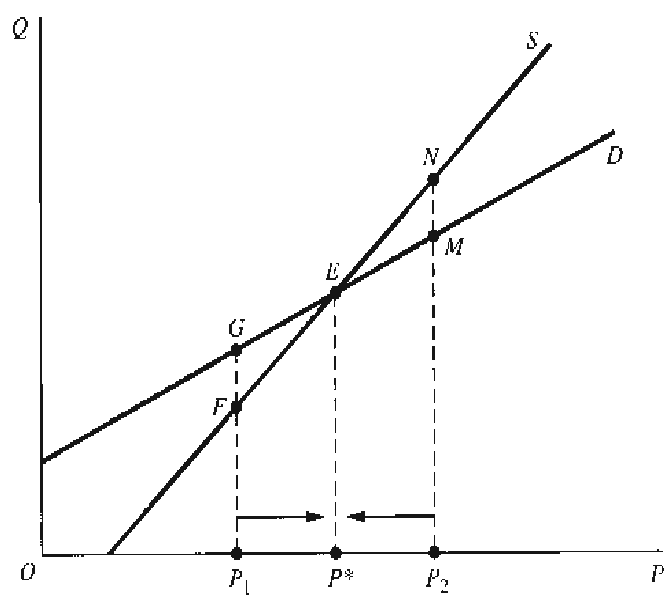
Thus, we can take this last inequality as the required restriction on the parameters j (the adjustment coefficient of price), β (the negative of the slope of the demand curve, plotted with Q on the *vertical* axis), and δ (the slope of the supply curve, plotted similarly).

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. $\frac{dy}{dt} + t^2y = 5t^2;\ y(0) = 6$
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

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

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:[†]

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

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.

[†] 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

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

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) $\dfrac{dy}{dt} = -\dfrac{t}{y}$

   (b) $\dfrac{y}{y+t}\,dy + \dfrac{2t}{y+t}\,dt = 0$ (d) $\dfrac{dy}{dt} = 3y^2t$

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.

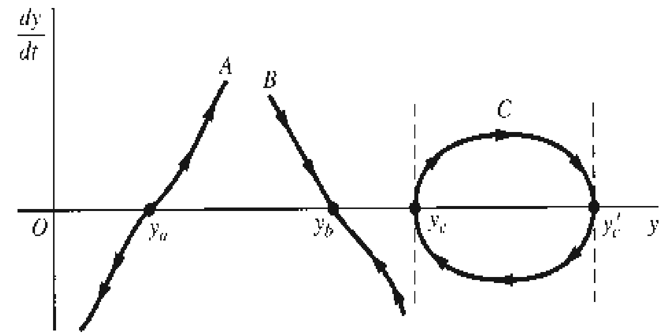
The Phase Diagram

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.

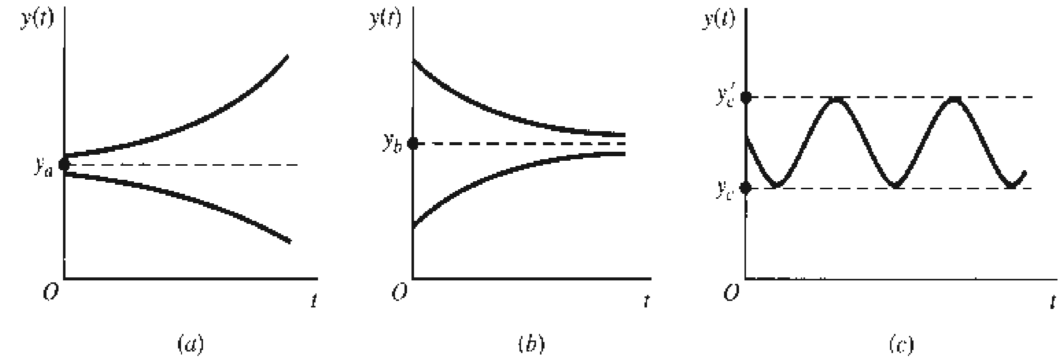
Types of Time Path

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) $\dfrac{dy}{dt} = y - 7$ (c) $\dfrac{dy}{dt} = 4 - \dfrac{y}{2}$
   (b) $\dfrac{dy}{dt} = 1 - 5y$ (d) $\dfrac{dy}{dt} = 9y - 11$
2. Plot the phase line for each of the following and interpret:
   (a) $\dfrac{dy}{dt} = (y+1)^2 - 16 \quad (y \geq 0)$
   (b) $\dfrac{dy}{dt} = \dfrac{1}{2}y - y^2 \quad (y \geq 0)$
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 $\dfrac{d}{dy}\left(\dfrac{dy}{dt}\right)$ 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,[†] 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

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

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

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

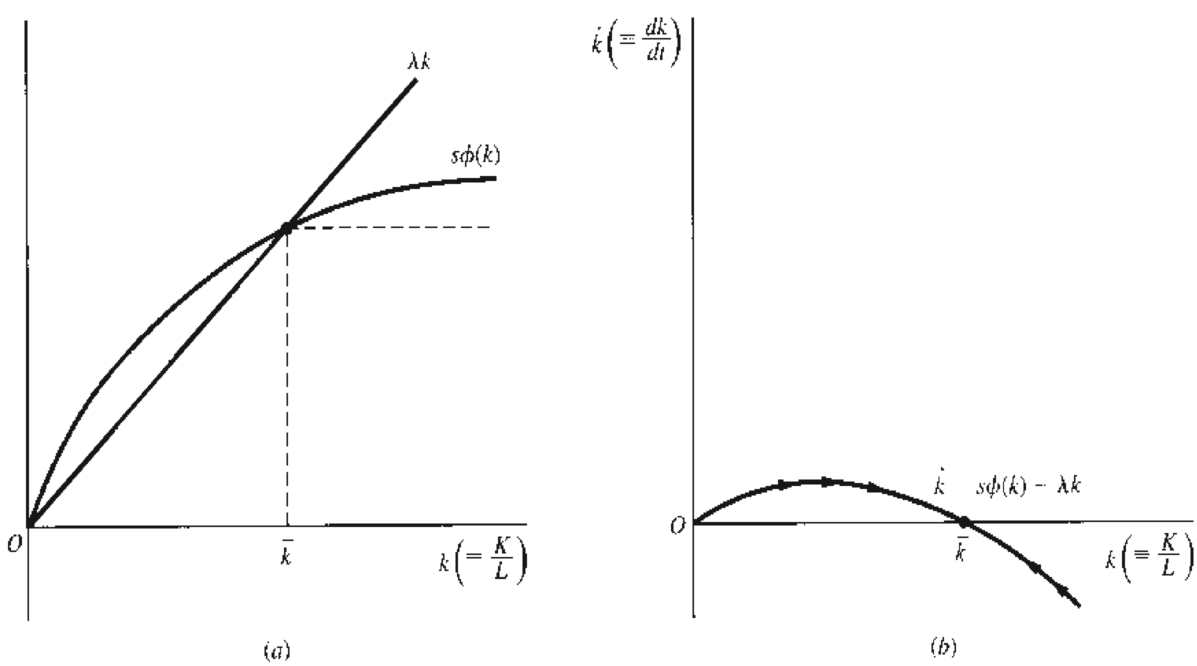
Because (15.30) is stated in a general-function form, no specific quantitative solution is available. Nevertheless, we can analyze it qualitatively. To this end, we should plot a phase line, with $\dot{k}$ on the vertical axis and k on the horizontal.

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:

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

or, in an alternative notation,

$$y^{(n)}(t) + a_1y^{(n-1)}(t) + \cdots + a_{n-1}y'(t) + a_ny = b \qquad \textbf{(16.1')}$$

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

$$y''(t) + a_1 y'(t) + a_2 y = b \tag{16.2}$$

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

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

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:[†]

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

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

[†] 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

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

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

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

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

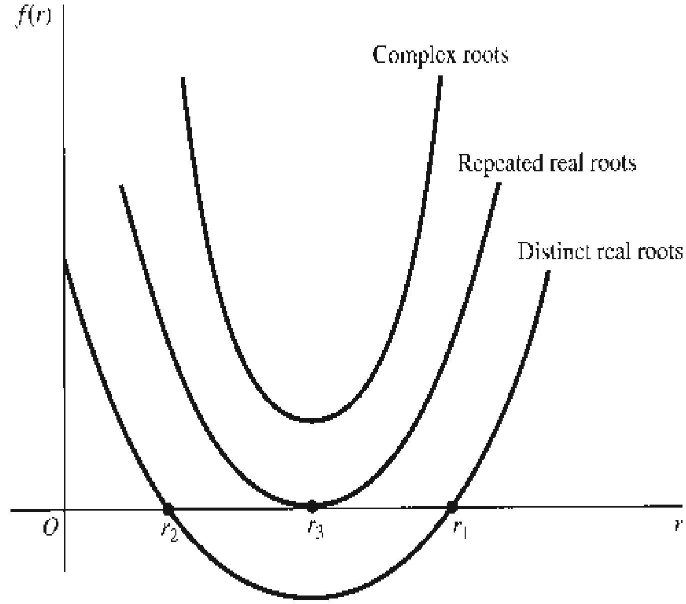
$$y(t) = 2e^{-3t} + te^{-3t} + 3$$

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) $y''(t) - 2y'(t) + 5y = 2$
   (b) $y''(t) + y'(t) = 7$
   (c) $y''(t) + 3y = 9$
   (d) $y''(t) + 2y'(t) - y = -4$
   (e) $y''(t) = 12$
2. Find the complementary function of each equation:
   (a) $y''(t) + 3y'(t) - 4y = 12$
   (b) $y''(t) + 6y'(t) + 5y = 10$
   (c) $y''(t) - 2y'(t) + y = 3$
   (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, $y''(t) + a_1y'(t) + a_2y = b$, 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.

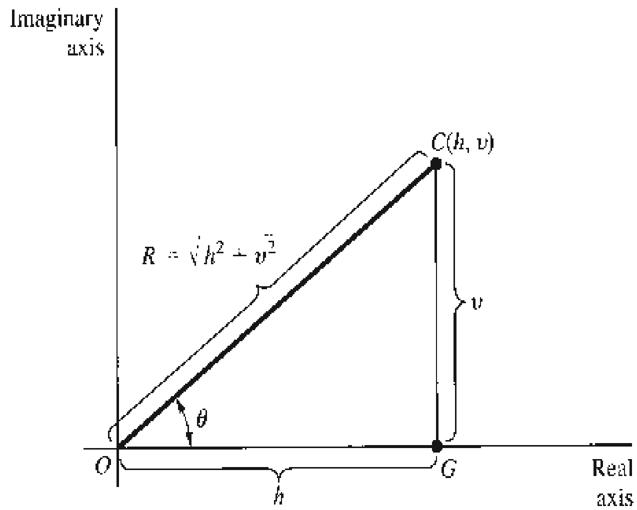
Imaginary and Complex Numbers

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,

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

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?

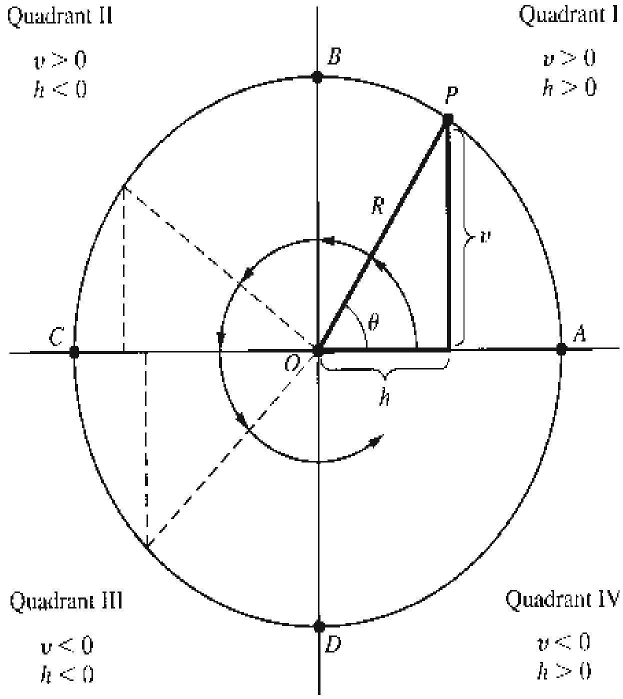
To facilitate their interpretation, it will prove helpful first to transform these expressions into equivalent *circular-function* forms. As we shall presently see, the latter functions characteristically involve periodic fluctuations of a variable. Consequently, the complementary function (16.11), being 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. 16.4*a*, 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.

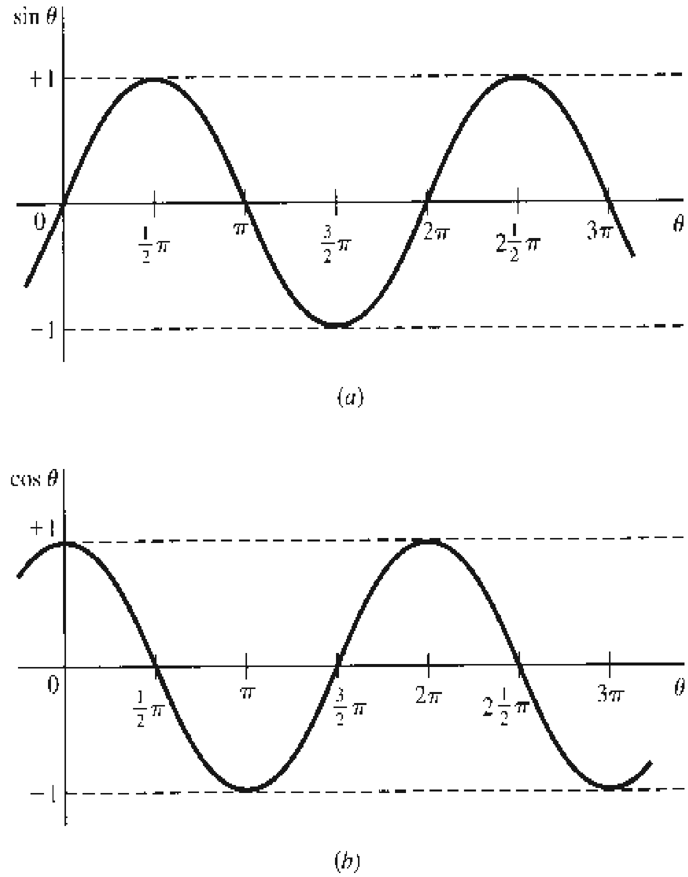
The substitution of these illustrative values of h into (16.13) yields the results in the bottom row of Table 16.1, but Fig. 16.4*b* gives a more complete depiction of the cosine function.

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 ($= -90^\circ$). 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

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

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

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

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

$$\frac{d}{d\theta}\sin\theta = \cos\theta \qquad (16.17)$$

$$\frac{d}{d\theta}\cos\theta = -\sin\theta \qquad (16.18)$$

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:

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \tag{16.19}$$

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:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots \tag{16.20}$$

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:

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

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

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*) $1 + \tan^2\theta \equiv \dfrac{1}{\cos^2\theta}$
   (*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

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

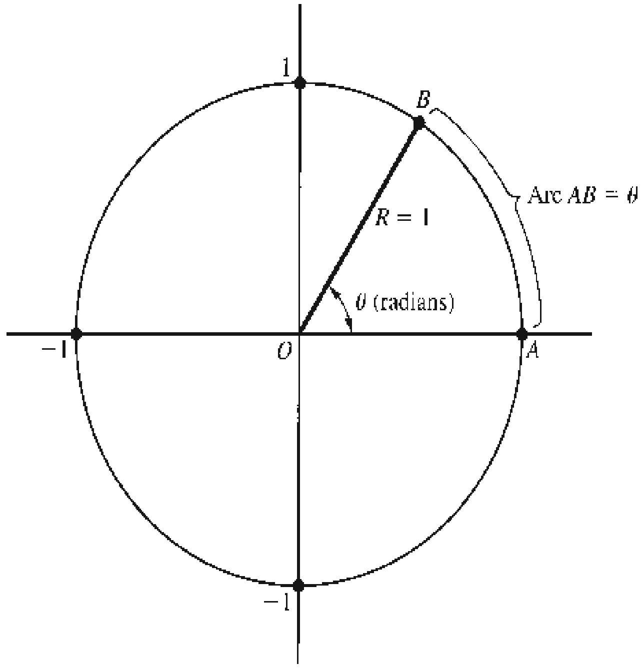
Let us first transform the imaginary exponential expressions in the parentheses into equivalent trigonometric expressions, so that we may interpret the complementary function as a circular function. This may be accomplished by using the Euler relations. Letting $\theta = vt$ in (16.21) and (16.21′),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

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

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

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

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†

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

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

† 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 $A_6 = 3$.[†] The definite solution is, therefore,

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

[†] 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

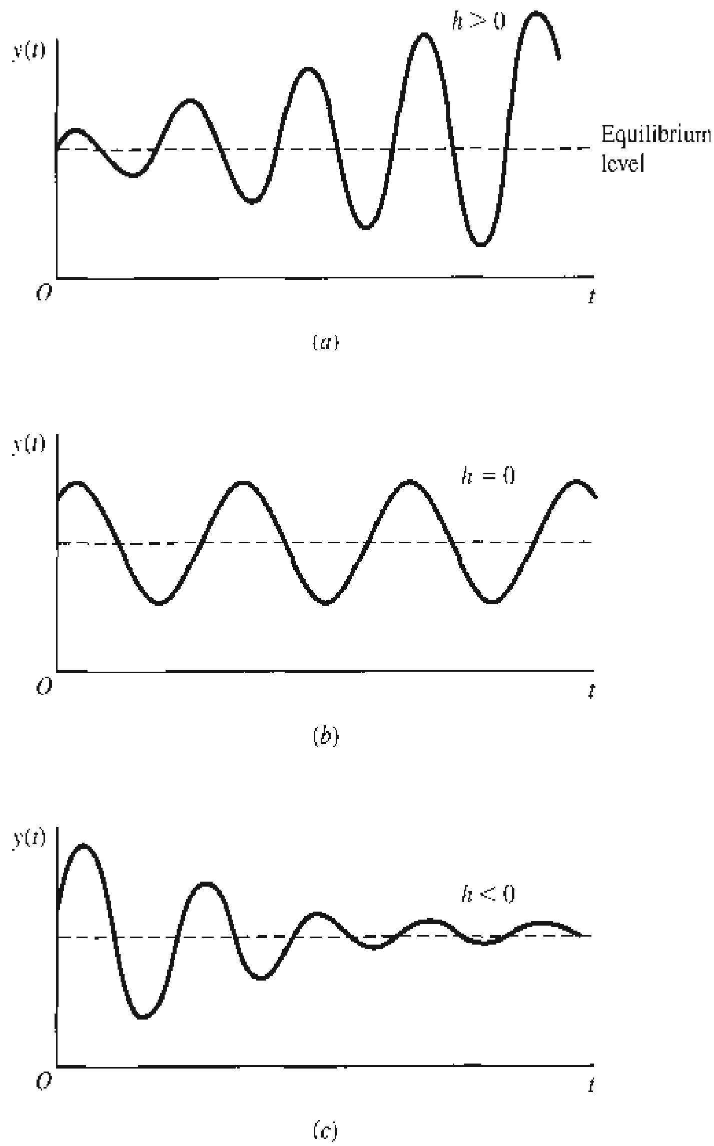
$$\begin{aligned} A_5 \cos vt + A_6 \sin vt &= A \cos \varepsilon \cos vt - A \sin \varepsilon \sin vt \\ &= A(\cos vt \cos \varepsilon - \sin vt \sin \varepsilon) \\ &= A \cos(vt + \varepsilon) \qquad \text{[by (16.16)]} \end{aligned}$$

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*.†

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

† 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. 16.6*c*), 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