2

First-order Differential Equations

In this chapter, methods will be given for solving first-order differential equations. Remember that first-order means that the first derivative of the unknown function is the highest derivative appearing in the equation. While this implies that the most general first-order differential equation has the form $F(t,x,x') = 0$ for some function F, in this chapter we will assume that the equation can be solved explicitly for x'. This means that our first-order differential equation can always be put in the form:

$$x' = f(t,x), \tag{2.1}$$

where f denotes an arbitrary function of two variables. To see why such an assumption makes sense, suppose the differential equation is

$$(x'(t))^2 + 4x'(t) + 3x(t) = t.$$

It would be messy, but not impossible, to use the quadratic formula to extract two differential equations of the form $x' = f(t,x)$ from this quadratic equation. However, one could also imagine equations where solving for $x'(t)$ is not even possible, and in such a case, some of our methods may not be applicable.

The material in this chapter will cover several analytic methods for solving first-order differential equations, each requiring the function f in (2.1) to have a special form. Two different graphical methods are also described; one for the general equation (2.1), and a more specific method for an autonomous equation where f is a function depending only on x. Numerical methods for first-order equations are introduced, and theoretical issues of existence and uniqueness of solutions are discussed. In the examples you will be presented with some real-world problems in applied mathematics. Acquiring a solid understanding of first-order equations will lay the groundwork for everything that follows.

2.1 Separable First-order Equations

The first analytic method we will consider applies to first-order equations that can be written in the form

$$dx/dt = g(t)h(x); \tag{2.2}$$

that is, when the function $f(t, x)$ can be factored into a product of a function of t times a function of x. Such a differential equation is called **separable**.

Example 2.1.1. *Determine which of the following first-order differential equations are separable. Hint: try to factor the right-hand side if the equation does not initially appear to be separable.*

(a) $x' = xt + 2x$

(b) $x' = x + \cos(t)$

(c) $x' = xt^2 + t^2 - tx - t$

(d) $x' = x^2 + x + 3$

Solution. Equation (a) can be factored into $x' = x(t + 2)$, so it is separable with $g(t) = t + 2$ and $h(x) = x$. Equation (b) is not separable as $x + \cos(t)$ cannot be factored into a function of t multiplied by a function of x. With a bit of work, equation (c) can be factored as $x' = (t^2 - t)(x + 1)$, so it is separable with $g(t) = (t^2 - t)$ and $h(x) = (x + 1)$. Equation (d) can be written as $x' = (1)(x^2 + x + 3)$, with $g(t) = 1$ and $h(x) = x^2 + x + 3$, so it is separable. ■

Special Cases of $x' = g(t)h(x)$.

(1) If $h(x) = 1$, the separable equation $x' = g(t)$ is just an integration problem, and the solution is

$$x = \int g(t)dt;$$

that is, x is just the **indefinite integral** of the function $g(t)$. Remember that this means that x can be *any* function $G(t)$ such that $G'(t) = g(t)$, and this introduces an arbitrary constant into the solution. As an example, the solution of $x' = t + 1$ is

$$x(t) = \int (t + 1)dt = \frac{t^2}{2} + t + C.$$

Even in this simple case the solution is an infinite one-parameter family of functions.

(2) If $g(t) = 1$, the separable equation $x' = h(x)$ is called an **autonomous** first-order differential equation. Unless $h(x)$ is a constant, it is no longer possible to solve the equation by simple integration, and the method given below must be used. Autonomous first-order differential equations are important and will be investigated more thoroughly in Section 2.7. In the above examples, only equation (d) is autonomous. The other three contain functions of t (other than the unknown function $x(t)$) on the right-hand side.

You may have already been shown, in calculus, an easy method for solving separable equations. Consider the following example.

Example 2.1.2. *Solve the differential equation* $\frac{dx}{dt} = -tx^2$.

Solution. Split dx/dt into two pieces dx and dt, and do a bit of algebra to write

$$-\frac{dx}{x^2} = tdt.$$

Integrate each side with respect to its own variable to obtain

$$\int \left(-\frac{1}{x^2}\right) dx = \int tdt \Longrightarrow \frac{1}{x} = \frac{t^2}{2} + C, \tag{2.3}$$

where the arbitrary constants on each side have been collected on the right. Solve this equation for x to obtain the one-parameter family of solutions

$$x(t) = \frac{1}{t^2/2 + C}.$$

■

You should check that the function $x(t)$ does satisfy the differential equation for any value of the constant C. It appears that this method works, but splitting $\frac{dx}{dt}$ into two pieces is not a mathematically condoned operation; therefore, a justification of the method needs to be given.

If an equation is separable, and $x'(t)$ is written as dx/dt, both sides of the equation $dx/dt = g(t)h(x)$ can be divided by $h(x)$, and the equation becomes

$$\frac{1}{h(x(t))}\frac{dx}{dt} = g(t). \tag{2.4}$$

The two sides of (2.4) will be identical if, and only if, their indefinite integrals are the same up to an additive constant; that is,

$$\int \frac{1}{h(x(t))}\left(\frac{dx}{dt}\right) dt = \int g(t)dt + C.$$

The method of simple substitution can be applied to the integral on the left. If we substitute $u = x(t)$, then $du = (dx/dt)dt$, and the equation becomes

$$\int [1/h(u)]du = \int g(t)dt + C. \tag{2.5}$$

Now let $H(u)$ be any function such that $H'(u) = 1/h(u)$ and $G(t)$ any function with $G'(t) = g(t)$. Then (2.5) implies that

$$H(u) + C_1 = G(t) + C_2 \Longrightarrow H(u) = G(t) + C,$$

where C is the constant $C_2 - C_1$.

Replacing u again by $x(t)$:

$$H(x(t)) = G(t) + C. \tag{2.6}$$

Check carefully that the expression $H(x) = G(t) + C$ in (2.6) is exactly the same as the solution obtained in (2.3) in the above example. It is an **implicit solution** of (2.2); that is, it defines a relation between the unknown function x and its independent variable t. If it can be solved explicitly for x as a function of t, the result is called an **explicit solution** of the differential equation. As expected, the integration produces an infinite one-parameter family of solutions.

Everything that has been said so far justifies the following step-by-step procedure for solving separable equations.

To solve a separable first-order differential equation $x'(t) = g(t)h(x)$:

- Write the equation in the form $dx/dt = g(t)h(x)$.
- Multiply both sides by dt, divide by $h(x)$, and integrate, to put the equation in the form
$$\int [1/h(x)]dx = \int g(t)dt.$$
- Find any function $H(x)$ such that $H'(x) = 1/h(x)$ and any function $G(t)$ such that $G'(t) = g(t)$.
- Write the solution as $H(x) = G(t) + C$.
- If possible, solve the equation from the previous step explicitly for x, as a function of t.

The next example shows how this method works.

Example 2.1.3. *Solve the separable differential equation* $x' = t(x + 1)$.

Solution. First write the equation in the form
$$dx/dt = t(x + 1). \tag{2.7}$$
Separate the variables (including dx and dt) so only the variable x appears on the left and t on the right:
$$\frac{1}{(x+1)}dx = tdt.$$
Integration of each side with respect to its own variable of integration leads to the implicit solution
$$\ln|x + 1| = t^2/2 + C,$$
which can be solved explicitly for x by applying the exponential function:
$$e^{\ln|x+1|} = |x + 1| = e^{(t^2/2+C)} = e^C e^{(t^2/2)}.$$
If the positive constant e^C is replaced by a nonzero constant A that can be either positive or negative, the absolute value signs can be dropped, to give
$$x + 1 = Ae^{(t^2/2)},$$
where $A = \pm e^C$. Then the explicit solution is
$$x(t) = Ae^{(t^2/2)} - 1. \tag{2.8}$$
■

The formula for $x(t)$ is the one-parameter family of curves that we expect to get as the solution of a first-order differential equation, and notice that the parameter A was introduced in the step where the equation was integrated.

As we saw in Section 1.2, a **particular solution** to a first-order differential equation is a solution in which there are no arbitrary constants. It will be shown (in Section 2.4) that, in general, to obtain a particular solution of a first-order differential equation it is necessary and sufficient to give one initial condition of the form $x(t_0) = x_0$.

Again from Section 1.2, we call a one-parameter family of solutions of a first-order differential equation, containing a single constant of integration, a **general solution** if it contains every solution of the equation. The analytic solution of a separable equation, found by the method just described, will contain all solutions of the equation with the possible exception of **constant solutions**. These are solutions of the form $x(t) \equiv C$ which make the right-hand side of the differential equation $x' = f(t, x)$ identically equal to 0. Note that with $x \equiv C$ the differential equation is satisfied, because the derivative of a constant function is also zero; that is, $\frac{d}{dt}(C) \equiv 0 \equiv f(t, C)$.

Referring back to Example 2.1.3, the function $x \equiv -1$ is the only constant solution of the differential equation $x' = t(x+1)$. In this case it is given by the solution formula (2.8) when the constant $A = 0$; therefore, (2.8) is the general solution of (2.7) if we allow the value $A = 0$.

When solving a separable equation it is wise to find all constant solutions first, since they may be lost when the equation is divided by $h(x)$.

Example 2.1.4. *Solve the initial-value problem* $x' = t/x$, $x(0) = 1$.

Solution. First note that this differential equation has no constant solutions; that is, there are no constant values for x that make $t/x \equiv 0$. Write the equation as $dx/dt = t/x$. Then by multiplying by dt and x, and integrating,

$$xdx = tdt \Longrightarrow \int xdx = \int tdt \Longrightarrow x^2/2 = t^2/2 + C.$$

The expression $x^2/2 = t^2/2 + C$ is an implicit solution and yields two explicit solutions

$$x(t) = \pm\sqrt{t^2 + 2C}.$$

We can satisfy the initial condition by substituting $t = 0$ into the general solution and setting $x(0) = 1$:

$$x(0) = \pm\sqrt{0 + 2C} = 1.$$

This implies that C must be 1/2 and the sign of the square root must be taken to be positive. Now the unique solution to the initial-value problem, $x(t) = \sqrt{t^2 + 1}$, is completely determined. ■

The following two applications show how separable differential equations and initial-value problems can arise in real-world situations.

2.1.1 Application 1: Population Growth.

One of the simplest differential equations arises in the study of the growth of biological populations. Consider a population with size $P(t)$ at time t. If it is assumed that the population has a constant birth rate α and constant death rate β per unit of time, then an equation for the rate of growth of the population is

$$dP/dt = \alpha P(t) - \beta P(t) = (\alpha - \beta)P(t) = rP(t), \tag{2.9}$$

where r is called the *net growth rate* of the population. This is a separable differential equation with general solution (Check it!):

$$P(t) = Ke^{rt},$$

where K is the arbitrary constant of integration. The initial value is frequently given as the size of the population at time $t = 0$. Then $P(0) = Ke^{r0} = K$, and the particular solution of this initial-value problem is $P(t) = P(0)e^{rt}$. This means that, t units of time after the initial time, the population will have grown exponentially (or decreased exponentially if $\beta > \alpha$). Populations do not grow exponentially forever, and biologists usually use more complicated equations of growth to take this into account.

One assumption that can be made is that as the population P increases, its growth rate decreases, due to the effects of crowding, intra-species competition, etc. The simplest way to decrease the growth rate as P increases is to assume that the growth rate is linear in P; that is, replace r in (2.9) by $R = r - \gamma P(t)$. Then

$$dP/dt = (r - \gamma P(t))P(t) = rP(t)\left(1 - \frac{\gamma}{r}P(t)\right) = rP(t)(1 - P(t)/N),$$

where we have defined a new constant $N = r/\gamma$. The equation

$$dP/dt = rP(t)(1 - P(t)/N) \tag{2.10}$$

is called the **logistic growth equation**.

Notice that the rate of growth dP/dt goes to 0 as $P(t) \to N$. This limiting value of the population, N, is called the **carrying capacity** of the ecosystem in which the population lives. The parameter r, which now gives the approximate rate of growth when the population is small, is called the **intrinsic growth rate** of P.

The logistic growth equation (2.10) is an autonomous differential equation and therefore is separable, but the expression $dP/[P(1 - P/N)]$ has to be integrated using partial fractions (or with the use of computer algebra). In either case, using our technique for separable equations, we have

$$\frac{dP}{[P(1 - P/N)]} = rdt \Longrightarrow \int \frac{1}{P(1 - P/N)}dP = \int rdt. \tag{2.11}$$

To compute the integral on the left, we can use partial fractions to write

$$\frac{1}{P(1 - P/N)} \equiv \frac{1}{P} - \frac{1}{P - N}.$$

You should check this last equality carefully (there is a review of partial fraction expansions in Section 2 of Chapter 6).

Integration of (2.11), using the partial fraction expression, now results in

$$\ln|P| - \ln|P - N| = rt + K.$$

To solve for $P(t)$, apply the exponential function to both sides and use the properties of the exponential and logarithmic functions to write

$$\begin{aligned}
e^{\ln|P| - \ln|P-N|} &= e^{rt+K} \Longrightarrow \\
\frac{P}{P - N} &= K_1 e^{rt}, \text{where } K_1 = \pm e^K \Longrightarrow \\
P &= PK_1 e^{rt} - NK_1 e^{rt} \Longrightarrow \\
P - PK_1 e^{rt} &= -NK_1 e^{rt} \Longrightarrow \\
P &= \frac{-NK_1 e^{rt}}{1 - K_1 e^{rt}} = \frac{NK_1 e^{rt}}{K_1 e^{rt} - 1} = \frac{N}{1 - Ce^{-rt}}
\end{aligned} \tag{2.12}$$

where $C = \frac{1}{K_1}$. Note that the differential equation $P' = rP(1 - P/N)$ has two constant solutions $P \equiv 0$ and $P \equiv N$. Using the value $C = 0$ in the solution (2.12) gives $P = N$, but no finite value of C makes this solution identically 0. To have a general solution, we must add the solution $P \equiv 0$ to the formula in (2.12).

In Section 2.4 the interval of existence of the solutions of the logistic population equation will be carefully examined and it will be shown that solutions with $P(0)$ between 0 and N exist for all t. Solutions with $P(0) > N$ exist for all $t > 0$ but have a vertical asymptote at a negative value of t. Solutions with $P(0) < 0$ tend to $-\infty$ at a positive value of t, but these are not physically realizable as populations.

2.1.2 Application 2: Newton's Law of Cooling.

Newton's Law of Cooling is a well-known law of physics that states that if a small body of temperature T is placed in a room with constant air temperature A, the rate of change of the temperature T is directly proportional to the temperature difference $A - T$. This law can be expressed in the form of a differential equation:

$$T'(t) = k(A - T(t)),$$

where $T(t)$ is the temperature of the small body at time t, A is the surrounding (ambient) air temperature, and k is a positive constant that depends on the physical properties of the small body. The only constant solution of the equation is $T(t) \equiv A$, which says that if the body is initially at the ambient temperature, it will remain there.

The equation can be seen to be separable and can be solved by writing

$$dT/dt = k(A - T) \Longrightarrow \int \frac{dT}{A - T} = \int k dt \Longrightarrow -\ln|A - T| = kt + C \Longrightarrow$$
$$|A - T| = e^{-(kt+C)} \Longrightarrow A - T(t) = \alpha e^{-kt},$$

where $\alpha = \pm e^{-C}$ can be any positive or negative real number. The explicit solution is

$$T(t) = A - \alpha e^{-kt}. \tag{2.13}$$

The constant solution $T \equiv A$ is obtained from the formula by letting α have the value zero. The long-term behavior is very easy to determine here, since if $k > 0$, then $T(t) \to A$ as $t \to \infty$. Thus the temperature of the small body tends to the constant room temperature, which makes good sense physically.

Consider the following very practical example that uses Newton's Law of Cooling.

Example 2.1.5. *A cup of coffee, initially at temperature* $T(0) = 210°$*, is placed in a room in which the temperature is* $70°$*. If the temperature of the coffee after five minutes has dropped to* $185°$*, at what time will the coffee reach a nice drinkable temperature of* $160°$*?*

Solution. If we assume the cup of coffee cools according to Newton's Law of Cooling, the general solution given by (2.13) with $A = 70$, can be used to write

$$T(t) = 70 - \alpha e^{-kt}.$$

Using the initial condition, we can find the value of α:

$$T(0) = 70 - \alpha e^{-0k} = 70 - \alpha = 210 \Longrightarrow \alpha = -140.$$

The temperature function can now be written as $T(t) = 70 + 140e^{-kt}$. To find the value of the parameter k, use the given value $T(5)$:

$$T(5) = 70 + 140e^{-5k} = 185 \Longrightarrow e^{-5k} = \frac{115}{140} \Longrightarrow$$
$$k = -\frac{1}{5}\ln(\frac{115}{140}) \approx 0.0393.$$

The value for k completely determines the temperature function; that is,

$$T(t) = 70 + 140e^{-0.0393t}$$

for all $t > 0$. Now the answer to the original question can be found by solving the equation $T(\hat{t}) = 160$ for $\hat{t}$. The approximate value for $\hat{t}$ is 11.2 minutes. ■

In the last example, if the value of the physical parameter k had been known beforehand, only one value of the temperature would have been required to determine the function $T(t)$ exactly. In this problem, the value of the parameter k had to be determined experimentally from the given data, thus necessitating the temperature to be read at two different times. This sort of thing is even more likely to occur in problems that come from nonphysical sciences, where parameters are usually not known physical constants and must be experimentally determined from the data provided.

Exercises 2.1. *In problems 1–4, determine whether the equation is separable.*

1. $x' + 2x = e^{-t}$
2. $x' + 2x = 1$
3. $x' = \frac{x+1}{t+1}$
4. $x' = \frac{\sin t}{\cos x}$

Put equations 5–14 into the form $x'(t) = g(t)h(x)$, and solve by the method of separation of variables.

5. $x' = \frac{x}{t}$
6. $x' = \frac{t}{x}$
7. $x' = x + 5$
8. $x' = 3x - 2$
9. $x' = x\cos(t)$
10. $x' = (1 + t)(2 + x)$
11. $xx' = 1 + 2t + 3t^2$
12. $x' = (t + 1)(\cos(x))^2$
13. $x' = t + tx^2$
14. $x' = 2 - tx^2 - t + 2x^2$ (*Hint: factor.*)

In 15–20, solve the initial-value problem.

15. $y' = y + 1, y(0) = 2$
16. $y' = ty, y(0) = 3$
17. $x' = x\cos(t), \; x(0) = 1$
18. $x' = (1 + t)(2 + x), \; x(0) = -1$

19. $x' = (t+1)(\cos(x))^2,\ x(0) = 1$

20. $P' = 2P(1-P),\ P(0) = 1/2$

21. (*Newton's Law of Cooling*) *A cold Pepsi is taken out of a* 40° *refrigerator and placed on a picnic table. Five minutes later the Pepsi has warmed up to* 50°. *If the outside temperature remains constant at* 90°, *what will be the temperature of the Pepsi after it has been on the table for twenty minutes? What happens to the temperature of the Pepsi over the long term?*

22. (*Newton's Law of Cooling*) *Disclaimer: The following problem is known not to be a very good physical example of Newton's Law of Cooling, since the thermal conductivity of a corpse is hard to measure; in spite of this, body temperature is often used to estimate time of death.*

    At 7 AM *one morning detectives find a murder victim in a closed meat locker. The temperature of the victim measures* 88°. *Assume the meat locker is always kept at* 40°, *and at the time of death the victim's temperature was* 98.6°. *When the body is finally removed at 8* AM, *its temperature is* 86°.

    (a) *When did the murder occur?*
    (b) *How big an error in the time of death would result if the live body temperature was known only to be between* 98.2° *and* 101.4°*?*

23. (*Orthogonal Curves*) *In the exercises for Section* 1.3 *the family of curves orthogonal to the family* $\bar{y} = cx^3$ *was shown to satisfy the differential equation* $y' = -\frac{x}{3y}$. *Use the method for separable equations to solve this equation, and find a formula for the orthogonal family.*

24. (*Orthogonal Curves*) *Show that the family of curves orthogonal to the family* $\bar{y} = ce^x$ *satisfies the equation* $y' = -\frac{1}{y}$ (*you may want to refer back to the model labeled MATHEMATICS in Section* 1.3). *Solve this separable equation and plot three curves from both families. If you make the scales identical on both axes, the curves should appear perpendicular at their points of intersection.*

25. (*Population Growth*) *In this problem you are asked to compare two different ways of modeling a population; either by a simple exponential growth equation, or by a logistic growth model.*

Table 2.1. Census

t	year	pop (millions)
0	1800	5.2
0.5	1850	23.2
1.0	1900	76.2
1.5	1950	151.3
2.0	2000	281.4

One population for which reasonable data is available is the population of the United States. The Census Table above gives census data for the population (*in millions*) *from 1800 to 2000 in 50 year intervals. If the logistic function* $P(t) = \frac{N}{1+Ce^{-rt}}$ *is fit exactly to the three data points for 1800, 1900, and 2000, the values for the parameters are* $N = 331.82$, $C = 62.811$, *and* $r = 2.93$, *and the equation becomes*

$$P(t) = \frac{331.82}{1 + 62.811e^{-2.93t}}.$$

Note that t *is measured in hundreds of years, with* $t = 0$ *denoting the year 1800.*

(a) *Plot (preferably using a computer) an accurate graph of* $P(t)$ *and mark the five data points on the graph (three of them should lie* exactly *on the curve).*
(b) *Where does the point* $(t, P(t)) = (1.5, 151.3)$ *lie, relative to the curve? Can you think of a reason why—famine, disease, war?*
(c) *What does the logistic model predict for* $P(2.1)$*; that is, the population in 2010? (The census data gives 308.7 million.)*
(d) *What does the model predict for the population in 2100? Does this seem reasonable?*
(e) *Now fit a simple exponential model* $p(t) = ce^{rt}$ *to the same data, using the two points for the years 1900 and 1950 to evaluate the parameters* r *and* c*. What does this model predict for the population in 2100? Does this seem more or less reasonable than the result in (d)?*
Note: If you use $t = 0$ *for 1900 and* $t = 1$ *for 1950, then the population in 2100 is* $p(4)$.

COMPUTER PROBLEMS. *Use your computer algebra system to solve the equation in each of the odd-numbered exercises 5–13 above. The Maple or Mathematica instructions you need are given at the end of the exercises in Section* 1.2. *You can use the answers in Appendix* A *to check your computer results.*

2.2 Graphical Methods, the Slope Field

For any first-order differential equation

$$x' = f(t, x), \tag{2.14}$$

whether or not it can be solved by some analytic method, it is possible to obtain a large amount of graphical information about the general behavior of the solution curves from the differential equation itself. In Section 2.4 you will see that if the function $f(t, x)$ is everywhere continuous in both variables t and x and has a continuous derivative with respect to x, the family of solutions of (2.14) forms a set of nonintersecting curves that fill the entire (t, x)-plane. In this section we will see how to use the slope function $f(t, x)$ to sketch a field of tangent vectors (called a **slope field**) that show graphically how the solution curves (also called **trajectories**) flow through the plane. This can all be done even when it is impossible to find an exact formula for $x(t)$.

Figure 2.1 shows a slope field for the spruce-budworm equation, which has the general form

$$x' = rx\left(1 - \frac{x}{N}\right) - \frac{ax^2}{b^2 + x^2}.$$

This equation models the growth of a population of pests that attack fir trees. It is essentially a logistic growth equation with an added term that models the effect of predation on the pests, primarily by birds.

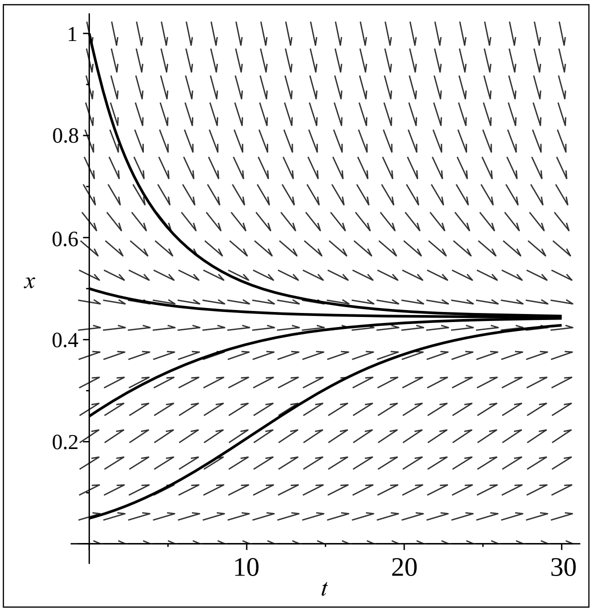


Figure 2.1. Slope field for the equation $x' = 0.2x(1-x) - 0.3x^2/(1+x^2)$

You should try to solve this equation with your CAS. It may print out some long unintelligible formula, but basically it is not able to solve the equation in terms of elementary functions, such as polynomials, exponentials, sines and cosines, etc. Notice, however, how easy it is to see how the family of solutions behaves. In Figure 2.1, Maple was used to draw the slope field and four numerically computed solutions (much more will be said about numerical solutions in Section 2.6).

To understand how a slope field is drawn, let $x(t)$ be a solution of (2.14). Then if $x(t)$ passes through some point $(\bar{t}, \bar{x})$ in the plane, the differential equation states that the graph of $x(t)$ at that point must have slope $f(\bar{t}, \bar{x})$. Using just the function $f(t, x)$, a slope field can be drawn by choosing an arbitrary set of points (t_i, x_i), and through each of them drawing a short line with slope $f(t_i, x_i)$; that is, a line that is tangent to the solution curve passing through that point.

Definition 2.1. *A **slope field** for a first-order differential equation $x' = f(t, x)$ is a field of short lines (or arrows) of slope $f(t_i, x_i)$ drawn through each point (t_i, x_i) in some chosen grid of points in the (t, x)-plane.*

The slope lines are often drawn as arrows, all of the same length, pointing in the positive t-direction. They may be drawn with center at the point (t_i, x_i), or alternatively, with the tail of the arrow at (t_i, x_i).

Four of the five models described in Section 1.3 are illustrated by slope fields, and this would be a good time to look back at them to see how much information they provide. The next example will show you exactly how a slope field is constructed.

Example 2.2.1. *Sketch a slope field for the differential equation*

$$x' = x + t.$$

Solution. We will arbitrarily choose a grid of points with integer coordinates in the region $-3 \le t \le 3, -3 \le x \le 3$ (see Figure 2.2). At each grid point (t_i, x_i), the slope line (which in Figure 2.2 is an arrow centered at the point) will have slope $f(t_i, x_i) = x_i + t_i$. For example, the arrow at $(-1, 2)$ has slope $2 + (-1) = 1$ and the arrow at $(2, -2)$ has slope $(-2) + 2 = 0$. We can put the slopes at all the integer grid points into a table:

Slopes at integer pairs (t, x) for the equation $x' = x + t$

	3	0	1	2	3	4	5	6
	2	−1	0	1	2	3	4	5
	1	−2	−1	0	1	2	3	4
x	0	−3	−2	−1	0	1	2	3
	−1	−4	−3	−2	−1	0	1	2
	−2	−5	−4	−3	−2	−1	0	1
	−3	−6	−5	−4	−3	−2	−1	0
		−3	−2	−1	0	1	2	3
					t			

The slope marks are plotted in Figure 2.2. A solution curve passing through the point $(1, 0)$ has been sketched in as well, by drawing it so that at each point it is tangent to the slope line at that point. In order for this to work, it has to be assumed that the direction of the slope lines changes continuously in both the t and x directions; that is, the function f should at least be a continuous function of both of its variables.

Just by looking at Figure 2.2, certain conclusions can be drawn. It appears that all solutions lying above the line of slope -1 through the point $(0, -1)$ tend to ∞, and solutions lying below that line tend to $-\infty$. What we cannot determine from the slope field is whether the solutions exist for all t, or have a vertical asymptote at a finite value of t.

One might also hazard a guess from looking at the slope field that the straight line $x = -t - 1$ is a solution of the differential equation. To check, let $x(t) = -t - 1$. Then $x'(t) = -1$ and substituting into $x' = x + t$,

$$x(t) + t = (-t - 1) + t = -1 \equiv x'(t),$$

which verifies that it is a solution. ■

Using a slope field is hardly a precise solution method, but it quickly gives a picture of how the entire family of solutions behaves. Therefore, in a sense, it provides a picture of the *general solution* of the differential equation. The finer the grid, the more information one has to work with.

Figure 2.3 shows a slope field for the logistic growth equation $P' = 0.5P\left(1 - \frac{P}{4}\right)$. We have seen such a picture before, but the point to be made here is that since the equation is autonomous, the slope function P' depends only on P. This means that the slopes along any horizontal line $P =$ constant will *all be the same*, so a slope field for an autonomous equation can be drawn very quickly. Again, it is impossible to tell from the graph whether solutions exist for all t; however, we know from our work with separable equations in Section 2.1 that if $P(0) > N = 4$ or $P(0) < 0$ the corresponding solution has a vertical asymptote.

If it is necessary to sketch a complicated slope field by hand, there is a more efficient way of choosing the grid of points at which the slope lines or arrows are to be

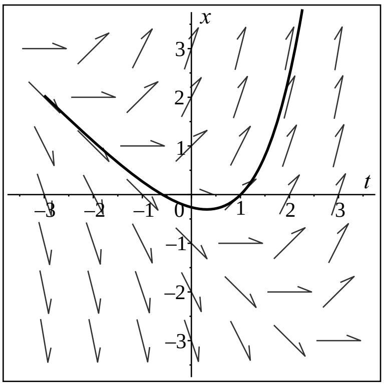


Figure 2.2. Slope field, $x' = x + t$

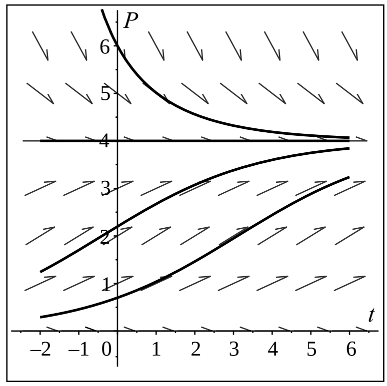


Figure 2.3. Slope field, $P' = \frac{1}{2}P\left(1 - \frac{P}{4}\right)$

drawn. Also, in some cases this method yields information about the long-term behavior of the solutions that is not obvious from a rectangular grid of slopes (as we see in the next example). Consider the right-hand side of a differential equation $x' = f(t, x)$. The equation

$$f(t, x) = m, \quad \text{for } m \text{ any real number,}$$

defines a curve, or set of curves, in the (t, x)-plane along which all the slope vectors must have the same slope m. Such a curve is called an **isocline**, or curve of equal slopes, for the differential equation (2.14). If an isocline for slope m is sketched in the plane, slope lines all of slope m can be drawn along it very quickly.

Example 2.2.2. *Sketch a slope field for $x' = x^2 - t$ by using isoclines.*

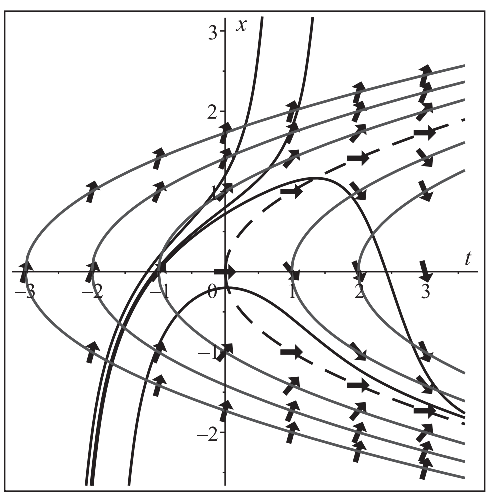


Figure 2.4. Slope field for $x' = x^2 - t$; the parabolic curves are isoclines

Solution. The isoclines are the curves having equation $x^2 - t = m$. They are parabolas, rotated by 90^0, and the isoclines for slope $m = -2, -1, 0, 1, 2,$ and 3 have been sketched in Figure 2.4. Be sure to note that, in general, isoclines are **not** solution curves. Along each of the isoclines, slope lines with appropriate slope have been drawn at equal intervals. It is then possible to sketch some approximate solution curves. There appear to be two types of solutions to this differential equation; those that increase monotonically with t, and those that approach the lower branch of the parabola $x^2 = t$ (and hence ultimately approach $-\infty$). It can be proved analytically that there is a unique solution separating the two families of curves. From Figure 2.4 it can be estimated that the initial condition that separates the different long-term behaviors is around $x(0) = 0.7$. This differential equation cannot be solved analytically in terms of elementary functions, but we will see that it is possible to approximate this special solution, as closely as desired, with the numerical methods described in Section 2.6. ■

2.2.1 Using Graphical Methods to Visualize Solutions.

Drawing slope fields and approximate solution curves is something that computers do very nicely. Almost any computer program that can draw a slope field for a first-order differential equation will also have the facility for drawing approximate solution curves through given initial points. This makes it possible to visualize graphically how the behavior of solutions changes when the parameters in an equation are varied. Be sure to take note of the fact that this can always be done without having an analytic solution of the differential equation.

A simple example of this can be seen in the case of a modified logistic growth equation, one with a harvesting term. This is the differential equation

$$P' = rP\left(1 - \frac{P}{N}\right) - H, \tag{2.15}$$

described in the previous section, where we have subtracted a constant H, representing the number of individuals being removed from the population (i.e., harvested) per unit of time. The example will demonstrate the power of a graphical method. The equation happens to have an analytic solution, but slope fields can be obtained, and studied, even when the equation cannot be solved analytically.

Example 2.2.3. *Consider a population of deer living in some woods. The deer population is assumed to be growing according to the harvested logistic equation*

$$\frac{dP}{dt} = 0.4P\left(1 - \frac{P}{100}\right) - H, \tag{2.16}$$

with $P(t)$ equal to the number of deer at time t years after the initial time. We know from previous work that if no deer are removed (that is, $H = 0$) the population will approach the carrying capacity of 100 *deer as t increases. The residents living near the wooded area would like to cut down the deer population by allowing hunters to "harvest" H deer each year, and wonder how this would affect the population. Figure* 2.5 *shows a Maple plot of the slope field with selected solution curves of* (2.16) *for three different values of the harvesting parameter H. Use the slope fields to describe the effect of increasing H.*

Solution. When $H = 0$, we have already seen that the population tends to the carrying capacity $P = 100$ for any positive initial condition. This can be seen in the left-hand graph. Even with an initial population of 2 it predicts a deer population of 100 after about 20 years.

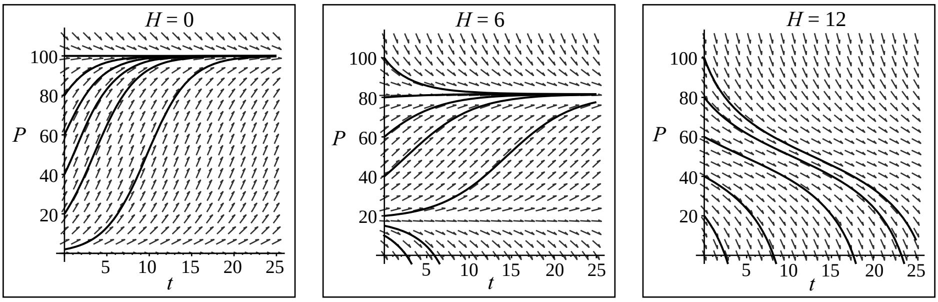


Figure 2.5. Deer population with $H = 0$, $H = 6$, and $H = 12$

If $H = 6$, meaning that 6 deer are removed from the herd each year, it is seen that the limiting population is reduced to about 80. Furthermore, if there are fewer than 20 deer initially, the herd appears to die out.

The right-hand graph with $H = 12$ is significantly different from the other two. It now appears that no matter how many deer are present initially, the herd will ultimately go extinct. This radical change in the behavior of the solutions, as the parameter H crosses a certain value, is called a **bifurcation** of the system modeled by the differential equation. The value of H at which the change occurs is called a **bifurcation value** of the parameter.

We will have a lot more to say about this when we study autonomous first-order equations in Section 2.7. At this point you are just seeing a graphical description of the change; for some equations, including this one, it will turn out to be possible to find the exact bifurcation value of a parameter by an analytic method. ■

Exercises 2.2. *For the differential equations in 1–5, make a hand sketch of a slope field in the region $-3 \leq t \leq 3, -3 \leq x \leq 3$, using an integer grid.*

1. $x' = x + t/2$

2. $x' = -xt/(1 + t^2)$

3. $x' = x(1 - x/2)$

4. $x' = 1 + t/2$

5. $x' = t/x$ *(Note: at any point where $x = 0$ the slopes will be vertical.)*

For the differential equations in 6–9, sketch a slope field in the region $-3 \leq t \leq 3, -3 \leq x \leq 3$ using the method of isoclines.

6. $x' = x + t$

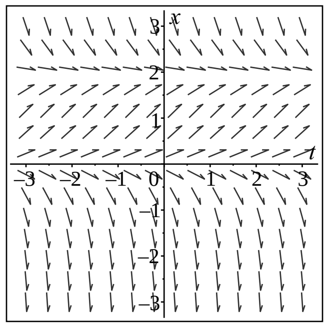
7. $x' = x^3 + t$

8. $x' = x^2$

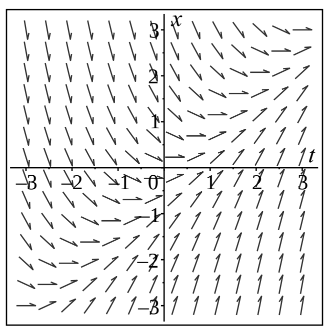
9. $x' = t^2 - x$

In Exercises 10–12, you are given a differential equation. The corresponding slope field has also been drawn. Sketch enough solution curves in the slope field to be able to describe the behavior of the family of solutions in the entire (t, x)-plane. As part of your description, explain how the long-term behavior depends on the initial condition $x(0)$. Be as specific as possible.

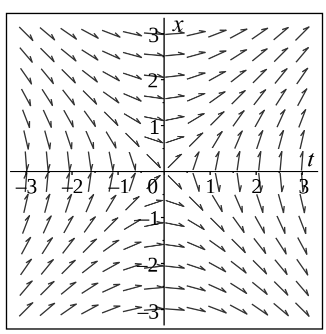
10. $x' = 2x(1 - x/2)$



11. $x' + x = t$



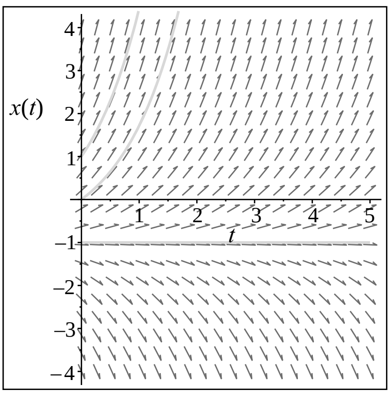
12. $x' = t/x$

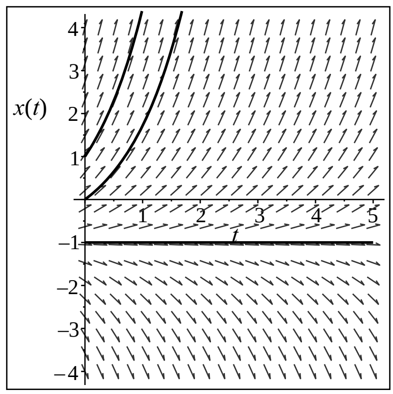


COMPUTER PROBLEMS. *You will need two new instructions in Maple (or in Mathematica) to produce slope fields and sample solutions for first-order differential equations. In Maple the first instruction,* `with(DEtools):`, *loads the necessary routines. Then the instruction* `DEplot();` *can be used to draw the slope field. The format of the second instruction is*

```
DEplot({deq},[x(t)],t=t0..t1,
[[x(0)=x0],[x(0)=x1],...]],x=a..b);
```

You can also add options at the end of the DEplot instruction. Executing the instruction `?DEplot;` *will show you what is available. It is probably best to specify the option* `stepsize=0.01` *to make sure the solution curves are drawn accurately.*





As an example, the two instructions

```
with(DEtools):
DEplot({x'(t)=1+x(t)},[x(t)],t=0..5,
[[x(0)=0],[x(0)=-1],[x(0)=1]],x=-4..4);
```

produced the slope field and three solution curves shown in the left-hand graph above. The graph on the right was done by adding two options; `color=blue` *to make the arrows blue and* `linecolor=black` *to make the curves black.*
The equivalent instructions in Mathematica are `<<VectorFieldPlots`:` *to load the routines, and*

```
VectorFieldPlot[{1,deq}{t,t0,t1},{x,a,b},
Axes->True,Frame->True]
```

to plot the slope field. Again, it would be a good idea to look at the help page for other options. The option `ScaleFunction->(1&)` *will ensure that the arrows are all of equal length, which makes the graph easier to read. The integer* 1 *in front of the differential equation implies that you are solving a first-order equation.*

You need to get used to these instructions, because they are going to receive heavy use all through the text.
Use your own computer algebra system or graphing calculator to create a detailed slope-field for equations 13–15.

13. $x' = x + t/2$ *(same as 1 above)*

14. $x' = x^2(1 - x)$

15. $x' = 1 + t/2$ *(same as 4 above)*

2.3 Linear First-order Differential Equations

Methods for solving linear differential equations have been around for a long time. While it is not correct to say that all linear equations can be solved analytically and all nonlinear equations cannot, it is close to the truth. It is so close, in fact, that before computers were readily available (circa 1955) engineers spent much of their time "linearizing" their problems so they could be solved.

A first-order differential equation is **linear** in $x(t)$ if it can be written in the form

$$a_1(t)x'(t) + a_0(t)x(t) = b(t),$$

where the functions a_0, a_1, and b are arbitrary functions of t. Since we are assuming that our first-order equations can be solved explicitly for $x'(t)$, the function $a_1(t)$ must be nonzero on some interval of t so that we can divide by it and write the equation as

$$x'(t) + \frac{a_0(t)}{a_1(t)}x(t) = \frac{b(t)}{a_1(t)}.$$

This leads to the following definition.

Definition 2.2. *A first-order* ***linear differential equation in standard form*** *is an equation that can be written as*

$$x'(t) + p(t)x(t) = q(t), \tag{2.17}$$

for some functions p and q. If $q(t) \equiv 0$, it is called a ***homogeneous linear equation****.*

What linear really means is that the unknown function x and its derivative can only appear in the equation multiplied by functions of t. You cannot have terms like $x^2, x^3, e^x, \ln(x), \sin(x)$, etc. However, the coefficient functions p and q can be arbitrary functions of t.

Example 2.3.1. *For each equation (a)–(d), determine if it is linear with respect to the unknown variables x and x'. If it is linear, state what $p(t)$ and $q(t)$ are. If it is not linear, explain why.*

(a) $x' = -3x + \sin(t)$

(b) $x' = x\sin(t) + t^3$

(c) $4x' + e^x t = 0$

(d) $xx' = 2t$

Solution. Equation (a) is linear: write it as $x' + 3x = \sin(t)$, so that $p(t) = 3$ and $q(t) = \sin(t)$. Equation (b) is linear: write it as $x' - x\sin(t) = t^3$, so that $p(t) = -\sin(t)$ and $q(t) = t^3$. Equation (c) is nonlinear: the term e^x is not a linear term in x. Equation (d) is also nonlinear; trying to put it into standard form would result in a term of the form $\frac{1}{x}$ which is not a linear function of x. ■

Observe that if a linear first-order equation is homogeneous, then it is also separable; that is, an equation of the form $x' + p(t)x = 0$ can also be written in the form $\frac{dx}{dt} = -p(t)x \equiv g(t)h(x)$, with $h(x) = x$ and $g(t) \equiv -p(t)$. This means that our method for solving separable equations can be applied to show that

$$\frac{dx}{dt} = -p(t)x \Longrightarrow \int \frac{dx}{x} = \int(-p(t))dt \Longrightarrow \ln|x| = -\int p(t)dt;$$

therefore, the **general solution of the homogeneous equation** can be written as

$$x(t) = e^{-\int p(t)dt}. \tag{2.18}$$

We will usually choose the simplest function $P(t)$ such that $P'(t) = p(t)$, and write the solution in the form

$$x(t) = \alpha e^{-P(t)}. \tag{2.19}$$

Example 2.3.2. *Solve the homogeneous linear equation $x' + \cos(t)x = 0$.*

Solution. In this case, $p(t) = \cos(t)$ and $\int p(t)dt = \sin(t) + C$. If we take $P(t) = \sin(t)$, the one-parameter family of solutions can be written as

$$x(t) = \alpha e^{-\sin(t)}.$$

To show that the solution is correct for any value of α, compute the derivative of x:

$$x'(t) = \alpha e^{-\sin(t)}\frac{d}{dt}(-\sin(t)) = -\alpha\cos(t)e^{-\sin(t)}.$$

Now substitution of x and x' into the equation results in the identity

$$x' + \cos(t)x = -\alpha\cos(t)e^{-\sin(t)} + \cos(t)\alpha e^{-\sin(t)} \equiv 0,$$

as expected. ■

Solving a nonhomogeneous linear equation is not quite as simple, as the next example illustrates.

Example 2.3.3. *Solve the equation*

$$x' + 2x = te^{-2t}. \tag{2.20}$$

Solution. This time we have no way of integrating the left-hand side of the equation, since we do not know what the function $x(t)$ is. The trick here is to multiply the entire equation by a *positive* function that will make the left side easy to integrate. Note that this will not change the set of solutions. In the case of (2.20) we are going to multiply by the function e^{2t}, and it will be explained below how we picked this particular multiplier.

Making sure to multiply both sides of (2.20) by e^{2t}, we get

$$e^{2t}x' + e^{2t}2x = e^{2t}(te^{-2t}) = t. \tag{2.21}$$

The left side of the equation can be seen to be the exact derivative of the product of the multiplier e^{2t} and the unknown function x; that is,

$$e^{2t}x' + e^{2t}2x \equiv \frac{d}{dt}\left(e^{2t}x\right).$$

This is true for *any* function $x(t)$, since it is just a particular case of the product rule for differentiation.

If we now write (2.21) in the form $\frac{d}{dt}\left(e^{2t}x(t)\right) = t$ and integrate, we get the implicit solution

$$\frac{d}{dt}\left(e^{2t}x(t)\right) = t \Longrightarrow \int \frac{d}{dt}\left(e^{2t}x(t)\right)dt = \int t dt \Longrightarrow e^{2t}x(t) = \frac{t^2}{2} + C.$$

Solving for x results in the explicit solution

$$x(t) = e^{-2t}\left(\frac{t^2}{2}\right) + Ce^{-2t}.$$

Check it! ■

The only question that remains, in the case of an arbitrary linear differential equation, is how to find a function $\mu(t)$ such that multiplying the equation by μ makes it integrable. Such a function μ is called an **integrating factor** for the differential equation.

The left side of (2.17), after multiplication by a function $\mu(t)$, is

$$\mu[x' + px] = \mu x' + (\underline{\mu p})x. \tag{2.22}$$

If $\mu p \equiv \mu'$, the right-hand side of (2.22) is exactly equal to the derivative of the product μx; that is, $\mu x' + \underline{\mu'} x \equiv \frac{d}{dt}(\mu x)$ by the product rule for differentiation, and this holds for any functions μ and x.

We already know how to find a function μ that satisfies $\mu p \equiv \mu'$. This is just a homogeneous linear equation $\mu' - p(t)\mu = 0$, and we have already shown that its solution is

$$\mu(t) = e^{\int p(t)dt}. \tag{2.23}$$

Be very careful of plus and minus signs here. We had to write the condition $\mu' = \mu p$ as $\mu' - p\mu = 0$ to put it into standard form. Then the general solution of the homogeneous equation (2.18) gave us another minus sign.

Using the integrating factor μ in (2.23), it is now possible to solve the equation

$$\mu[x' + px] = \mu q$$

by integrating both sides with respect to t:

$$\int \mu[x' + px]dt = \int \mu q dt \Longrightarrow \int \frac{d}{dt}(\mu x)dt = \int \mu q dt$$

$$\Longrightarrow \mu x = \int \mu q dt + C.$$

Dividing the final expression by the nonzero function μ gives the explicit solution

$$x(t) = \frac{1}{\mu(t)}\left[\int \mu(t)q(t)dt + C\right]. \tag{2.24}$$

Although the equation (2.24) is important theoretically as the **general solution of the first-order linear differential equation**, you do not need to memorize it to solve a simple linear equation. To do that, just follow the steps used in the derivation. This can be written in the form of a 5-step procedure.

To solve a linear first-order differential equation:

- Put the equation in standard form,
$$x'(t) + p(t)x(t) = q(t).$$
Be sure to divide through by the coefficient of $x'(t)$ if it is not already equal to 1.
- Find the simplest possible antiderivative $P(t)$ of the coefficient function $p(t)$.
- Let $\mu(t) = e^{P(t)}$, and multiply both sides of the equation in step one by μ (do not forget to multiply $q(t)$ by μ).
- Integrate both sides of the resulting equation with respect to t (be sure to add a constant of integration on one side). If you have done the first three steps correctly, the integral of the left-hand side is just $\mu(t)x(t)$.
- Divide both sides of the equation by $\mu(t)$ to obtain the explicit solution $x(t)$.

The easiest way to see how all of this works is by example.

Example 2.3.4. *Solve the equation* $x' = x + te^t$.

Solution. This is easily seen to be a linear equation with standard form

$$x' - x = te^t. \tag{2.25}$$

The coefficient $p(t)$ of x is -1, and $\int p(t)dt = \int(-1)dt = -t + C$ implies that we can use the integrating factor $\mu(t) = e^{-t}$. Now multiply both sides of the equation by μ:

$$e^{-t}x' - e^{-t}x = e^{-t}(te^t) = t.$$

Using the product rule for differentiation, the left side is the derivative of $\mu x \equiv e^{-t}x$; therefore, the equation can be written as

$$\frac{d}{dt}(e^{-t}x) = t.$$

It is now possible to integrate both sides of the equation with respect to t:

$$\int \frac{d}{dt}(e^{-t}x)dt = \int tdt \Longrightarrow e^{-t}x = t^2/2 + C$$

and solve explicitly for $x(t)$:

$$x(t) = e^t(t^2/2 + C).$$

This is the general solution of $x' = x + te^t$, and you should check it by differentiating $x(t)$ and substituting x and x' into the differential equation. ■

It can be seen that the solution of a first-order linear differential equation always contains a single constant that can be used to satisfy one initial condition. The next example shows how this is done.

Example 2.3.5. *Solve the initial-value problem*

$$tx' + x = \cos(t), \quad x(1) = 2. \tag{2.26}$$

Also, determine the behavior of the solution as $t \to \infty$, both by looking at the formula for the solution and by using a slope field.

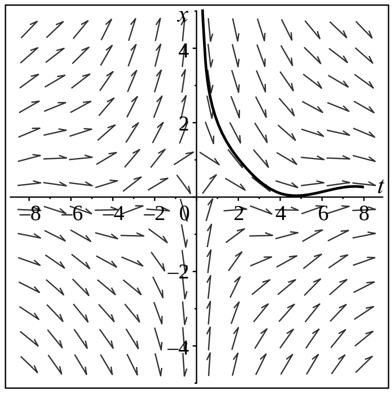


Figure 2.6. $tx' + x = \cos(t)$

Solution. To put (2.26) into standard form, it must be divided by t:

$$x' + (1/t)x = \cos(t)/t,$$

where it must now be assumed that $t \neq 0$. The coefficient of x in the standard form of the equation is $p(t) = 1/t$ and a simple antiderivative is $P(t) = \ln(t)$; therefore, the integrating factor we will use is $\mu(t) = e^{\ln(t)} = t$. Remember, we can always choose the simplest possible antiderivative of p. If the differential equation is multiplied by $\mu(t) = t$ and integrated,

$$\int \left[tx'(t) + x(t)\right]dt \equiv \int \frac{d}{dt}\left[tx(t)\right]dt = \int \cos(t)dt.$$

This implies that $tx(t) = \sin(t) + C$, and solving explicitly for x,

$$x(t) = \left[\sin(t) + C\right]/t, \quad t \neq 0.$$

To satisfy the initial condition $x(1) = 2$, substitute $t = 1$ and $x = 2$ into the equation:

$$x(1) = \sin(1) + C = 2,$$

and solve for C. Then $C = 2 - \sin(1)$, and the solution to the initial-value problem is $x(t) = [\sin(t) + 2 - \sin(1)]/t$. From the exact solution we can see that $x \to 0$ as $t \to \infty$. At $t = 0$, $x(t)$ has a vertical asymptote. A slope field for this equation is shown in Figure 2.6, with the solution through the initial point $(1, 2)$ drawn in. From the slope field it appears that all solutions with initial conditions given at $t > 0$ approach zero as $t \to \infty$. The analytic solution can be used to verify this. ■

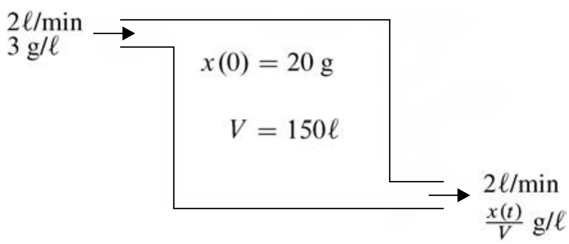
2.3.1 Application: Single-compartment Mixing Problem. We will end this section on linear equations with a very important application that comes up in several applied areas. We will see that they can be as diverse as pollution control and medical dosing. In this application we consider a problem called a **one-compartment mixing problem**. At this point we are only able to solve a mixing problem involving a single compartment in which mixing takes place. By the end of Chapter 4 we will be able to treat mixing problems with any number of compartments.

Basically, the problem consists of finding a formula for the amount of some "pollutant" in a container, into which the pollutant is entering at a fixed rate and also flowing out at a fixed rate. The general rule used to model this situation is conservation of mass:

rate of change of pollutant per unit time = rate in − rate out.

If we denote by $x(t)$ the amount of pollutant in the container at time t, its rate of change per unit time is given by $\frac{dx}{dt}$; therefore, the problem leads to a differential equation for the function $x(t)$.

Example 2.3.6. *Consider a fish tank that initially contains* 150 *liters of water with* 20 *grams of salt (pollutant) dissolved in it. The salt concentration in the tank needs to be increased from* 20/150 *grams per liter to* 1 *gram per liter to accommodate a new species of fish. Water containing* 3 *grams of salt per liter is allowed to run into the tank at a rate of* 2 *liters per minute. The thoroughly stirred mixture in the tank is also allowed to drain out at the same rate of* 2 *liters per minute. Find a differential equation for the amount of salt* $x(t)$ *in the tank at time* t*. Use the initial condition* $x(0) = 20$ *to find the time it will take to increase the salt concentration in the tank to* 1 *gram per liter.*



Solution. We are assuming that the mixture in the tank is thoroughly stirred, so that the salt concentration $x(t)/\text{vol}$ is instantaneously the same throughout the tank. Let $x(t)$ denote the number of grams of salt in the tank at time t. Then the rate at which salt enters or leaves the tank is measured in units of grams/minute, so

$$\frac{dx}{dt} = \text{rate in} - \text{rate out} = (2\ell/\text{min})(3\text{g}/\ell) - (2\ell/\text{min})\left(\frac{x(t)}{\text{vol}}\text{g}/\ell\right)$$

and the differential equation we are looking for is

$$\frac{dx}{dt} = 6 - 2\frac{x(t)}{150} \equiv 6 - \frac{x(t)}{75}. \tag{2.27}$$

This differential equation is both separable and linear, and we will arbitrarily choose to solve it by the method for separable equations:

$$\frac{dx}{dt} = \frac{450 - x}{75} \Longrightarrow \int \frac{dx}{450 - x} = \int \frac{dt}{75} \Longrightarrow -\ln(|450 - x|) = \frac{t}{75} + C.$$

Exponentiating both sides,

$$\frac{1}{450 - x} = Ke^{t/75} \implies x = 450 - (1/K)e^{-t/75}, \quad \text{where } K = \pm e^C.$$

Using the initial condition $x(0) = 20$ to find the value of $1/K$,

$$x(0) = 450 - 1/K = 20 \implies 1/K = 430$$

and the solution of the IVP is

$$x(t) = 450 - 430e^{-t/75}. \tag{2.28}$$

To find the time when the concentration of salt in the tank reaches 1 g/ℓ, set

$$1 = x(t)/\text{vol} = (450 - 430e^{-t/75})/150$$

and solve for t. This gives $t \approx 27$ minutes as the time it takes to increase the concentration to the required value. ■

It can be seen from (2.28) that as $t \to \infty$, the solution $x(t)$ approaches 450 as we would expect. This is the value of x for which the concentration $x(t)$/vol in the tank is the same as the concentration of the solution constantly flowing in.

If the flow rates in and out of the container are not the same, the volume will change over time, and the differential equation will no longer be autonomous, and therefore no longer separable. In the next example it will be shown that we have the tools to obtain an analytic solution in this case.

Example 2.3.7. *In the previous example it took about 27 minutes for the solution to reach the desired concentration. One way to speed up the process would be to let the salt water flow in at a faster rate. Assume that the input flow is increased to* 2.5ℓ*/min, but that the output flow cannot be increased. If the maximum amount of water the container can hold is* 160 *liters, will the salt concentration reach the desired level before the container overflows?*

Solution. The differential equation now becomes

$$x'(t) = (2.5\ \ell/\text{min})(3\ \text{g}/\ell) - (2\ \ell/\text{min})\left(\frac{x(t)}{\text{vol}}\ \text{g}/\ell\right).$$

Since the rate of water flowing in is 2.5ℓ/min and the rate out is still 2ℓ/min, the volume of fluid in the tank will increase by 0.5 liter each minute; therefore, the volume at time t is $V(t) = 150 + \frac{1}{2}t$.

The differential equation $x' = 7.5 - \frac{2x}{150+\frac{1}{2}t}$ is not separable, but it is linear and its standard form is

$$x' + \left(\frac{4}{300 + t}\right)x = 7.5.$$

The integrating factor is $e^{\int \frac{4}{300+t} dt}$, and choosing the simplest antiderivative we can take $\mu(t) = e^{4(\ln(300+t))} = (300+t)^4$. Multiplying by μ and integrating gives

$$\begin{aligned}(300+t)^4 x' + 4(300+t)^3 x &= 7.5(300+t)^4 \\ &\Longrightarrow \int \left((300+t)^4 x' + 4\left((300+t)^3 x\right)\right) dt \\ &\equiv \int \frac{d}{dt}\left((300+t)^4 x\right) dt = \int 7.5(300+t)^4 dt;\end{aligned}$$

therefore,

$$(300+t)^4 x = \frac{7.5(300+t)^5}{5} + C$$

and

$$x(t) = 1.5(300+t) + \frac{C}{(300+t)^4}.$$

Using the initial condition $x(0) = 20$,

$$20 = 1.5(300) + \frac{C}{(300)^4}$$

and the value of C is $-430(300)^4$. To make the concentration in the tank equal $1\text{g}/\ell$, set

$$1 = x(t)/V(t) = \frac{x(t)}{150 + t/2} = \left(\frac{2}{300+t}\right) x(t) = 3.0 - \frac{860(300)^4}{(300+t)^5},$$

and solve for t. This gives $t \approx 22.4$ minutes, but $V(22.4) = 150 + 0.5(22.4) = 161.2$ liters; therefore, the tank will overflow just before the solution reaches the desired concentration. ■

Exercises 2.3. *In Exercises 1–6, state whether the equation is linear, separable, both, or neither.*

1. $x' = tx$
2. $x' = \cos(t)x$
3. $x' = t^2 \sin(x)$
4. $x' = t + x$
5. $y' = 1/(1 + e^{-y})$
6. $y' = e^t y + e^t$

In Exercises 7–10, use the method described in this section to find an analytic solution to the equation and describe what happens to the solution as $t \to \infty$.

7. $x' + 2x = e^{-2t} \sin(t)$
8. $x' = 2x + 1$ *This equation is also separable.*
9. $tx' - 2x = 1 + t$
10. $tx' = 2x + t^2$

In Exercises 11–14, solve the initial-value problems.

11. $x' = -x + e^{2t}, \quad x(0) = 1$

12. $x' + 2x = e^{-2t}\cos(t), \quad x(0) = -1$

13. $tx' + x = 3t^2 - t, \quad x(1) = 0$

14. $x' + 2tx = 3t, \quad x(0) = 4$

15. *In Section* 1.3*, the velocity $v(t)$ of a free-falling body was shown to satisfy the differential equation*

$$mv' = mg - kv^p,$$

where m is the mass of the body in kilograms, g is the gravitational constant, and p is a real constant.

   (a) *For what value(s) of p is the equation linear in $v(t)$?*
   (b) *Solve the equation with $p = 1$, $\frac{k}{m} = 0.6$, and $g = 9.8m/sec^2$.*
   (c) *What is the value of the terminal velocity; that is, $\lim_{t\to\infty} v(t)$? Convert the velocity from meters per second to miles per hour.*
   (d) *Does the terminal velocity depend on the initial velocity $v(0)$? Explain.*

16. *In Exercises* 1.3 *(under ENGINEERING), the RL circuit problem*

$$i'(t) + 3i(t) = 10\sin(t)$$

was given, and you were asked to find the value of $i(0)$ that makes the solution be $i(t) = 3\sin(t) - \cos(t)$. Use the method described in this section to solve the equation, and determine the correct value of $i(0)$. (The following integral formula may help: $\int e^{at}\sin(bt)dt = e^{at}\left(\frac{a\sin(bt)-b\cos(bt)}{a^2+b^2}\right)$.) Use your CAS to draw a slope field for the equation and plot the solution using the initial condition you found.

The next two problems refer to the mixing problem in subsection 2.3.1.

17. *Continuing Examples* 2.3.6 *and* 2.3.7*, assume now that instead of increasing the input flow as in Example* 2.3.7*, we decrease the input flow to* $1.5\ell/min$ *(the output flow is not changed). Will the salt concentration reach the desired level before the container empties?*

18. *Continuing Example* 2.3.7*, find the maximum input flow rate such that the tank will just reach its capacity of* 160 *liters when the salt concentration reaches* 1 *gram per liter. How many minutes does it take? Your answer should be exact to the nearest second. (This one is hard!)*

2.4 Existence and Uniqueness of Solutions

In this section we will consider the question of whether a given initial-value problem

$$x' = f(t,x), \quad x(t_0) = x_0 \tag{2.29}$$

has a solution passing through a given initial point (t_0, x_0), and, if it does, whether there can be more than one such solution through that point. These are the questions of existence and uniqueness, and they become very important when one is solving applied problems.

There is a basic theorem, proved in advanced courses on differential equations, which can be applied to the initial-value problem (2.29).

Theorem 2.1. (Existence and Uniqueness): *Given the differential equation* $x' = f(t,x)$*, if* f *is defined and continuous everywhere inside a rectangle* $\mathbf{R} = \{(t,x) \mid a \leq t \leq b, c \leq x \leq d\}$ *in the* (t,x)*-plane, containing the point* (t_0, x_0) *in its interior, then* there exists *a solution* $x(t)$ *passing through the point* (t_0, x_0)*, and this solution is continuous on an interval* $t_0 - \varepsilon < t < t_0 + \varepsilon$ *for some* $\varepsilon > 0$*. If* $\frac{\partial f}{\partial x}$ *is continuous in* $\mathbf{R}$*, there is* exactly one *such solution; that is, the solution is unique.*

Comment 1. Even if the function f is differentiable, the theorem does not imply that the solution exists for all t. For example, it may have a vertical asymptote close to $t = t_0$.

Comment 2. To prove the uniqueness part of Theorem 2.1 it is not necessary for the function f to be differentiable. It only needs to satisfy what is called a **Lipschitz condition**.

Definition 2.3. *A function* $f(t,x)$ *defined on a bounded rectangle* $\mathbf{R}$ *is said to satisfy a* **Lipschitz condition** *on* $\mathbf{R}$ *if for any two points* (t, x_1), (t, x_2) *in* $\mathbf{R}$ *there exists a positive constant* M *such that*

$$|f(t,x_2) - f(t,x_1)| \leq M|x_2 - x_1|.$$

Note that if the derivative $\frac{\partial f}{\partial x}$ *does* exist and is continuous in $\mathbf{R}$, then we can use the maximum value of $|\frac{\partial f}{\partial x}|$ in $\mathbf{R}$ as a Lipschitz constant M, since the Mean Value Theorem implies that the slope of the secant $\frac{f(t,x_2)-f(t,x_1)}{x_2-x_1}$ is equal to $\frac{\partial f}{\partial x}$ at some intermediate point between x_1 and x_2. One of the exercises at the end of this section will ask you to find a Lipschitz constant for a nondifferentiable slope function. This may convince you that if f is differentiable it is a much easier condition to use.

The proof of Theorem 2.1 requires advanced analysis, but the idea is straightforward. First notice that any solution $x(t)$ of the IVP (2.29) is also a solution of the integral equation

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds. \tag{2.30}$$

To see this, simply differentiate both sides of (2.30) using the Fundamental Theorem of Calculus on the right. Therefore, solving (2.29) is equivalent to solving (2.30).

To solve (2.30), start by letting $x_0(t)$ be the constant $x_0(t) \equiv x_0$. Then create a sequence of functions $\{x_n(t)\}_{n=1}^{\infty}$ as follows:

$$x_1(t) = x_0 + \int_{t_0}^{t} f(s, x_0(s))ds,$$

$$x_2(t) = x_0 + \int_{t_0}^{t} f(s, x_1(s))ds,$$

$$\ldots, x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, x_n(s))ds, \ldots.$$

In order to show that $x(t) = lim_{n\to\infty} x_n(t)$ is the solution of our IVP we would need to be able to claim that

$$lim_{n\to\infty} \int_{t_0}^{t} f(s, x_n(s))ds = \int_{t_0}^{t} f(s, lim_{n\to\infty} x_n(s))ds.$$

This statement requires a theorem called the monotone convergence theorem for integrals, and this is where the hard analysis comes in.

The following example illustrates how such a successive approximation scheme might work. Be warned that this is a very simple equation, and in general the iteration procedure does not give a constructive method for producing a solution.

Example 2.4.1. *Show that the procedure described above leads to the correct solution of the IVP*

$$x' = x, \; x(0) = 1.$$

Solution. This IVP is equivalent to the integral equation $x(t) = 1 + \int_0^t x(s)ds$. Let $x_0(t) \equiv 1$. Then the next three functions in the sequence are:

$$x_1(t) = 1 + \int_0^t x_0(s)ds = 1 + \int_0^t 1ds = 1 + t;$$

$$x_2(t) = 1 + \int_0^t x_1(s)ds = 1 + \int_0^t (1 + s)ds = 1 + t + t^2/2;$$

$$x_3(t) = 1 + \int_0^t x_2(s)ds = 1 + \int_0^t (1 + s + s^2/2)ds = 1 + t + t^2/2 + t^3/(3 \cdot 2);$$

and it can be seen that $x_n(t)$ is the nth Taylor polynomial for the function e^t. As $n \to \infty$ we know that this approaches the function e^t which *is* the solution of the IVP. ■

Theorem 2.1 does, however, give us a very useful piece of information. In some of the slope fields pictured in Section 2.2, it was noted that solutions appear to form a space-filling set of curves. We are now able to test that assertion, using the following lemma.

Lemma 2.1. *Let $x' = f(t, x)$ be a first-order differential equation with f and $\frac{\partial f}{\partial x}$ both continuous for all values of t and x in some region* **R** *in the plane. Then inside the region* **R** *the solution curves of the differential equation will form a nonintersecting space-filling family of curves.*

Proof. Theorem 2.1 guarantees that there exists a unique solution through any point in **R** since f and $\frac{\partial f}{\partial x}$ are both continuous; this means that every point in the region has a solution curve passing through it. Furthermore, two solutions cannot intersect anywhere inside **R**, since if they did, there would be a point with two or more solutions through it and this would contradict uniqueness guaranteed by the theorem. □

We first apply these results to the logistic growth equation.

Example 2.4.2. *What does Theorem* 2.1 *tell us about solutions of the equation*

$$x' = x(1 - x)? \tag{2.31}$$

Solution. The functions $f(t,x) = x(1-x) = x - x^2$ and $\frac{\partial f}{\partial x} = 1 - 2x$ are both continuous everywhere in the (t,x)-plane; therefore, there exists a unique solution through any point (t_0, x_0). Lemma 2.1 tells us that the set of solutions fills the entire (t,x)-plane and that two different solution curves never intersect.

Since $x(t) \equiv 0$ and $x(t) \equiv 1$ are both constant solutions of the differential equation, the theorem also implies that any solution with $0 < x(0) < 1$ must remain bounded between 0 and 1 for all t for which it exists. This is easily seen, since a solution starting inside the strip $0 < x < 1$ cannot have a point of intersection with either of the bounding solutions, hence it can never exit the strip.

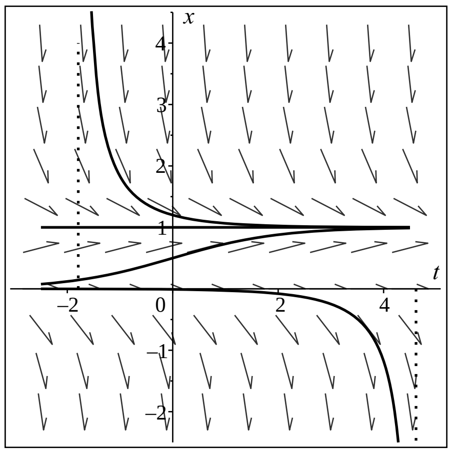


Figure 2.7. Solutions of $x' = x(1-x)$

Be very careful however not to assume that every solution must exist for all values of t. Theorem 2.1 does not tell us how big the t-interval of existence is for a solution of $x' = x(1-x)$. We solved this equation in Section 2.1.1 (it is separable). The general solution of $x' = x(1-x)$ is $x(t) = \frac{1}{1+Ce^{-t}}$, where $C = \frac{1}{x(0)} - 1$ if $x(0) \neq 0$. If $0 < x(0) < 1$ the constant C is positive, and therefore the denominator of the solution is never equal to zero. This means that the solution exists for all t. (This could also have been determined from the fact that solutions starting in this strip can never leave.) If $x(0) > 1$ or $x(0) < 0$, the denominator of $x(t)$ will have a zero at $t = \ln(-C) = \ln(1 - \frac{1}{x(0)})$ and $x(t)$ will have a vertical asymptote there. Figure 2.7 shows the solutions through $x(0) = -0.01, 0.5, 1.0$, and 1.2. The vertical asymptote for the solution with $x(0) = -0.01$ is at $t = \ln(101) \approx 4.6$, and for the solution with $x(0) = 1.2$ the asymptote is at $t = \ln(1/6) \approx -1.8$. ■

A second example will show what happens when the equation is not well-behaved everywhere.

Example 2.4.3. *Does the IVP* $tx' + x = \cos(t), \quad x(1) = 2$, *have a unique solution?*

Solution. The function $f(t,x) = (-x + \cos(t))/t$ is continuous everywhere except where $t = 0$. As shown in Figure 2.8, a rectangle that contains no points of discontinuity of f can be drawn around the initial point $(t_0, x_0) = (1, 2)$. This implies that

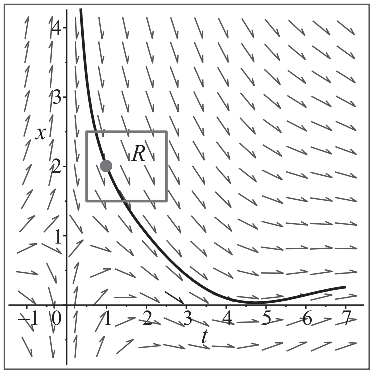


Figure 2.8. Solutions of $tx' + x = \cos(t)$

there exists at least one solution of the IVP. To check its uniqueness, it is necessary to compute

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left(-\frac{1}{t}x + \frac{\cos(t)}{t}\right) = -\frac{1}{t}.$$

This function is also continuous wherever $t \neq 0$; therefore, the IVP has a unique solution (shown in Figure 2.8). We found this solution in Example 2.3.5 to be $x(t) = (\sin(t) + 2 - \sin(1))/t$, so the solution of our IVP turns out to only be defined for $t > 0$, and has a vertical asymptote at $t = 0$. ■

The next example will show what can happen when the slope function does not satisfy all of the hypotheses of Theorem 2.1. Figure 2.9 shows a slope field for the

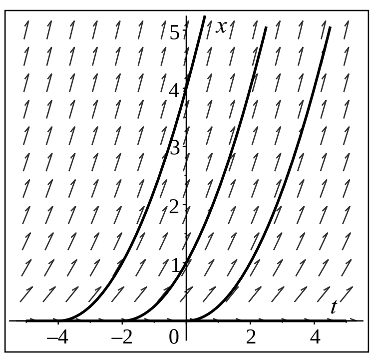


Figure 2.9. Slope field for the equation $x' = \sqrt{x}$

equation $x' = \sqrt{x}$. No arrows appear in the lower half of the plane. This is because the slope function $\sqrt{x}$ is not defined for negative x. It can also be seen that this is an autonomous differential equation, and slopes along any horizontal line (for constant x) will all be the same.

Example 2.4.4. *Does the IVP* $x' = \sqrt{x}, \quad x(0) = 0$ *have a solution? Is it unique?*

Solution. The function $f(t,x) = \sqrt{x}$ is defined at $x = 0$, but we cannot enclose the initial point $(0,0)$ *inside* a rectangle containing only points with $x \geq 0$. When $x < 0$, $\sqrt{x}$ is not even defined; therefore, f does not satisfy the continuity hypothesis at the initial point $(0,0)$, and Theorem 2.1 gives us *no* information about solutions of this IVP.

The differential equation $x' = \sqrt{x}$ is separable and can be solved by integration, as follows:

$$\frac{dx}{dt} = \sqrt{x} \Longrightarrow \frac{dx}{\sqrt{x}} = dt \Longrightarrow \int \frac{dx}{\sqrt{x}} \equiv \int x^{-1/2}dx = \int dt$$

$$\Longrightarrow 2x^{1/2} = t + C \Longrightarrow x(t) = \left(\frac{t+C}{2}\right)^2.$$

To satisfy the initial condition $x(0) = (\frac{0+C}{2})^2 = 0$, we must have $C = 0$; therefore, $x(t) = \frac{t^2}{4}$ is a solution of the IVP. However, $x(t) \equiv 0$ is also a solution; therefore, in this case the solution of the IVP is definitely *not* unique. This should not be surprising, since Theorem 2.1 requires the function f to have a continuous derivative with respect to x, and $\frac{\partial f}{\partial x} = \frac{1}{2\sqrt{x}}$ is not even defined at $x = 0$. ■

Back in Section 1.1 we showed in Example 1.1.2 that one solution of the differential equation $x' = -t/x$ is $x(t) = \sqrt{1-t^2}$. The question was raised as to whether the equation has any solutions that exist outside of the interval $-1 \leq t \leq 1$. We are now able to answer that question.

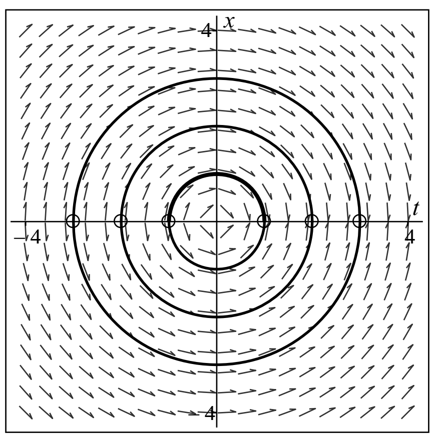


Figure 2.10. Solutions of $x' = -t/x$

The slope function $f(t,x) = -t/x$ and its derivative $\frac{\partial f}{\partial x} = t/x^2$ are both continuous everywhere except on the t-axis, where $x = 0$. This means that the upper half-plane, and also the lower half-plane, will be filled with nonintersecting solution curves.

The differential equation $x' = -t/x$ is separable and can be solved by writing

$$\frac{dx}{dt} = -\frac{t}{x} \Longrightarrow \int xdx = \int (-t)dt \Longrightarrow \frac{x^2}{2} = -\frac{t^2}{2} + C \Longrightarrow x^2 + t^2 = 2C \equiv \alpha^2.$$

This implicit solution produces a family of circles in the (t, x)-plane, but each circle gives us the solutions $x(t) = \pm\sqrt{\alpha^2 - t^2}$ of two initial-value problems; one satisfying $x(0) = \alpha$ and the other satisfying $x(0) = -\alpha$. What this means is that every point on the t-axis is a point at which two solution curves meet. Any point not on the t-axis will have a unique solution passing through it. This is illustrated in the slope field shown in Figure 2.10. Note very carefully the direction of the vectors in the slope field. What is happening to the slope of a solution curve as x approaches zero?

Exercises 2.4. *Exercises 1–6 test your understanding of Theorem* 2.1 *and Lemma* 2.1.

1. *Use Theorem* 2.1 *to show that the differential equation* $x' = \frac{x^2}{1+t^2}$ *has a unique solution through every initial point* (t_0, x_0). *Can solution curves ever intersect?*

2. *Use Theorem* 2.1 *to prove that the solution of an initial-value problem for the equation* $x' = \frac{x}{1+t^2}$, *with* $x(0) > 0$, *can never become negative. Hint: First find a constant solution of the differential equation for some constant* C.

3. *Does the equation* $x' = x^2 - t$ *have a unique solution through every initial point* (t_0, x_0)? *Can solution curves ever intersect for this differential equation? If so, where?*

4. *Does Theorem* 2.1 *imply that the solution of* $x' = x^2 - t$, $x(0) = 1.0$, *is defined for all* t? *What can you say about the size of the interval for which it exists?*

5. *Draw a slope field for the equation* $x' = x^{2/3}$. *Can solution curves ever intersect? If so, where? (Note that this is a separable equation.)*

6. *Consider the differential equation* $x' = t/x$.

   (a) *Use Theorem* 2.1 *to prove that there is a unique solution through the initial point* $x(1) = 1/2$.

   (b) *Show that for* $t > 0$, $x(t) = t$ *is a solution of* $x' = t/x$.

   (c) *Use* (b) *to show that the solution in* (a) *is defined for all* $t > 1$.

   (d) *Solve the initial-value problem in* (a) *analytically (see Example* 2.1.4*). Find the exact* t*-interval on which the solution is defined. Sketch the solution in the slope field in Figure* 2.11.

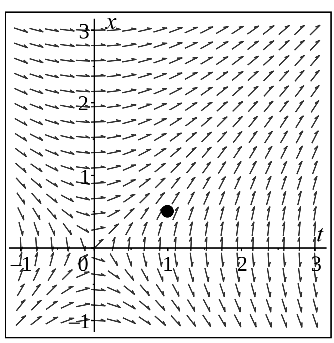


Figure 2.11. Slope field for $x' = t/x$

Student Project. *The Student Project at the end of Chapter* 1 *compared several functions $S(z)$ which varied from* 0 *to* 1 *as z went from $-\infty$ to $+\infty$. These functions, called response functions, are used to model the growth of activity in a collection of nerve cells. One of the simplest functions of this type is the piecewise linear function*

$$S(z) = \begin{cases} 0 & \text{if } z < -1/2, \\ z + 1/2 & \text{if } -1/2 \le z \le 1/2, \\ 1 & \text{if } z > 1/2. \end{cases} \tag{2.32}$$

A simple neural model using the function S is the equation

$$x'(t) = f(t, x) = -x(t) + S\big(20(x(t) - 0.5 - 0.4\cos(2\pi t))\big).$$

To study the behavior of solutions of this equation in the square $B = \{(t, x) \mid 0 \le t \le 1, 0 \le x \le 1\}$ we need to know if there exist unique solutions through any given initial point in B.

1. *Show that the derivative $\frac{\partial f}{\partial x}$ is NOT a continuous function in B.*
2. *Show that f does satisfy a Lipschitz condition in B and find a Lipschitz constant M.*
3. *Plot five solutions of $x' = f(t, x)$ in B. If you are using Maple, the function $S(z)$ can be defined by the command:* `S:=z->piecewise(z<-0.5,0,z<0.5,z+0.5,1.0)`. *Plot solutions through the initial points $(t_0, x_0) = (0, 0.1), (0, 0.3), (0, 0.5), (0, 0.7), (0, 0.9)$.*

2.5 More Analytic Methods for Nonlinear First-order Equations

Two additional methods for solving first-order differential equations are described in this section. Together with the previous methods we have looked at, our list is still far from exhaustive. If there is any analytic method for solving a given differential equation, your computer algebra system will probably know how to apply it. Discovering and programming methods for finding analytic solutions of differential equations is one of the things that keeps the programmers for Maple, *Mathematica*, etc. busy; however, understanding how these methods work will enable you to be an intelligent user of these programs.

2.5.1 Exact Differential Equations. In this section we use the variables x and y instead of t and x since this is the way exact equations are usually written in textbooks, as functions of two space variables. However, everything remains true if x is replaced by t and y is replaced by x.

Let F be a differentiable function of two independent variables x and y. The expression

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy$$

is called the *total differential* of the function F. From calculus it is known that if dF is identically equal to zero in some region of the (x, y)-plane, then the function $F(x, y)$ must be a constant in that region. This can be used to find analytic solutions for a certain type of first-order differential equation.

Given a first-order differential equation written in the form

$$g(x,y)dx + h(x,y)dy = 0, \tag{2.33}$$

suppose we can show that there exists a differentiable function $F(x,y)$ such that $\frac{\partial F}{\partial x} = g(x,y)$ and $\frac{\partial F}{\partial y} = h(x,y)$. (This is not going to be true in general.) If it is true, (2.33) states that the total differential of the function F is zero, and therefore we have found an implicit solution of (2.33), namely $F(x,y) = C$.

Example 2.5.1. *Solve the differential equation*

$$y' = -x/y. \tag{2.34}$$

This is the same equation, namely $x' = -t/x$, that we looked at at the end of Section 2.4*, with the variables renamed. A slope field for* (2.34) *is pictured in Figure* 2.10.

Solution. We first write (2.34) in the form $dy/dx = -x/y$ and then expand it, by assuming that the differentials dx and dy can be separated, as

$$-xdx = ydy \quad \text{or} \quad xdx + ydy = 0.$$

The function $F(x,y) = \frac{x^2+y^2}{2}$ satisfies $\frac{\partial F}{\partial x} = x$ and $\frac{\partial F}{\partial y} = y$. (Check it!) This means that F is a function such that $\frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy \equiv 0$; therefore, $F(x,y) = \frac{x^2+y^2}{2} \equiv C$ is an implicit solution of (2.34). ■

This is exactly the same implicit solution we found for $x' = -t/x$ using the method for separable equations.

Definition 2.4. *A first-order differential equation $y' = f(x,y)$ is called* ***exact*** *if it can be written in the form*

$$M(x,y)dx + N(x,y)dy = 0 \tag{2.35}$$

where $M(x,y) = \frac{\partial F}{\partial x}$ and $N(x,y) = \frac{\partial F}{\partial y}$ for some differentiable function $F(x,y)$.

In the calculus of several variables it is shown that if F is a twice continuously differentiable function of two variables, then the second-order mixed partial derivatives

$$\frac{\partial^2 F}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 F}{\partial y \partial x}$$

are equal. If $M = \frac{\partial F}{\partial x}$ and $N = \frac{\partial F}{\partial y}$, then

$$\frac{\partial M}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) \equiv \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) = \frac{\partial N}{\partial x}.$$

It is also true, but slightly more difficult to prove, that if $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, then there exists a function $F(x,y)$ with $\frac{\partial F}{\partial x} = M$ and $\frac{\partial F}{\partial y} = N$; therefore, a simple way to test (2.35) for exactness is to check that the functions M and N satisfy $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$.

Example 2.5.2. *Determine whether or not the following differential equations are exact.*

(i) $(x^2 + y)dx + (x - \sin(y))\,dy = 0$

(ii) $(x^2y)dx + (x^3/3 + 4y^2 + 1)dy = 0$

(iii) $(x + y)dx - (x - y)dy = 0$

Solution. In the first equation, $M = x^2 + y$ and $N = x - \sin(y)$. The partial derivatives are

$$\frac{\partial M}{\partial y} = \frac{\partial}{\partial y}(x^2 + y) = 1 \quad \text{and} \quad \frac{\partial N}{\partial x} = \frac{\partial}{\partial x}(x - \sin(y)) = 1;$$

therefore, the equation is exact.

In the second equation, $M = x^2y$ and $N = (x^3/3 + 4y^2 + 1)$. The partial derivatives are

$$\frac{\partial M}{\partial y} = \frac{\partial}{\partial y}(x^2y) = x^2 \quad \text{and} \quad \frac{\partial N}{\partial x} = \frac{\partial}{\partial x}(x^3/3 + 4y^2 + 1) = x^2;$$

therefore, the equation is exact.

In the third equation, $M = x + y$ and $N = -(x - y)$. The partial derivatives are

$$\frac{\partial M}{\partial y} = \frac{\partial}{\partial y}(x + y) = 1 \quad \text{and} \quad \frac{\partial N}{\partial x} = \frac{\partial}{\partial x}(-x + y) = -1.$$

Since the partial derivatives are not the same, equation (iii) is not exact. ■

To solve an exact equation $M(x, y)dx + N(x, y)dy = 0$: Make sure that $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$, and then use the following three steps.

(1) Set $\frac{\partial F}{\partial x} = M(x, y)$ and integrate once with respect to x to get

$$F(x, y) = \int M(x, y)dx + Q(y).$$

Note that $Q(y)$ represents the "constant" of integration with respect to x, and the integration is done as though x is the only variable, and y is a parameter.

(2) Differentiate the function F found in step (1), partially with respect to y, and set the result equal to $N(x, y)$:

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\left(\int M(x, y)dx\right) + Q'(y) \equiv N(x, y).$$

(3) If you have done steps (1) and (2) correctly, the equation resulting from step (2) will define $Q'(y)$ as a function of y only. Antidifferentiate $Q'(y)$ to obtain $Q(y)$. The function F from step (1), with this value of $Q(y)$, will provide an implicit solution $F(x, y) = C$ of the given exact differential equation.

The examples below demonstrate the use of this method.

Example 2.5.3. *Solve the differential equation* $(x^2 + y)dx + (x - \sin(y))\,dy = 0$.

Solution. We showed in Example 2.5.2 that $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x} = 1$; therefore, this equation is exact. As suggested in step (1), we set $\frac{\partial F}{\partial x} = M = x^2 + y$. Then

$$F = \int (x^2 + y)dx = x^3/3 + yx + Q(y).$$

When integrating partially with respect to x, an arbitrary function $Q(y)$ acts as the constant of integration since $\frac{\partial}{\partial x}(Q(y)) \equiv 0$.

Step (2) says to differentiate this version of F partially with respect to y and set the result equal to N:

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\left(x^3/3 + yx + Q(y)\right) = 0 + x + Q'(y) \equiv N = x - \sin(y).$$

Therefore, the function Q must satisfy $Q'(y) = -\sin(y)$, and integration gives $Q(y) = \cos(y)$. It is not necessary to add a constant to $Q(y)$ since the function F will be set equal to an arbitrary constant.

Substituting $Q(y)$ into $F = x^3/3 + yx + Q(y)$ we have the complete function $F(x, y) = x^3/3+yx+\cos(y)$; therefore, an implicit solution of the differential equation is given by

$$x^3/3 + yx + \cos(y) = C.$$

In this case it is not possible to solve for $y(x)$ explicitly.

Alternatively, we could have started by setting $\frac{\partial F}{\partial y} = N = x - \sin(y)$ and integrated with respect to y to obtain $F = \int(x - \sin(y))dy = xy + \cos(y) + P(x)$. Then $\frac{\partial F}{\partial x} = y + P'(x) \equiv M = x^2 + y$ implies $P'(x) = x^2$ and $P(x) = x^3/3$. This results in the same function $F(x, y) = xy + \cos(y) + x^3/3$. Sometimes one of the two methods turns out to be much easier than the other. ■

Example 2.5.4. *Solve the initial-value problem*

$$y' = f(x, y) = \frac{-2xy}{1 + x^2 + 3y^2}, \quad y(0) = 1. \tag{2.36}$$

Solution. The differential equation (2.36) is neither separable nor linear. It is easy to see that $f(x, y)$ is defined and continuous for all x and y, and so is its partial derivative with respect to y; therefore, the Existence and Uniqueness Theorem tells us that there is a unique solution through any initial point and solutions cannot intersect in the (x, y)-plane. Note that $y \equiv 0$ is a constant solution of (2.36). A slope field for this equation is shown in Figure 2.12.

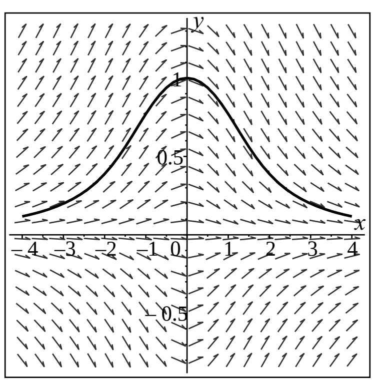


Figure 2.12. Slope field for $y' = \frac{-2xy}{1+x^2+3y^2}$

To find an algebraic solution of the IVP, we first set

$$\frac{dy}{dx} = \frac{-2xy}{1 + x^2 + 3y^2},$$

and then write the equation in the form

$$M(x,y)dx + N(x,y)dy = (2xy)dx + (1 + x^2 + 3y^2)dy = 0.$$

Noting that $\frac{\partial M}{\partial y} = 2x = \frac{\partial N}{\partial x}$, we see that the method for solving exact equations can be used.

Now, following the three-step method described above,

(1) Let $F(x,y) = \int M(x,y)dx + Q(y) = \int (2xy)dx + Q(y) = x^2y + Q(y)$.

(2) Differentiating the function found in (1) with respect to y,

$$\frac{\partial F}{\partial y} = x^2 + Q'(y) \equiv N(x,y) = 1 + x^2 + 3y^2$$

and the equation for $Q'(y) \equiv \frac{dQ}{dy}$ is $Q'(y) = 1 + 3y^2$.

(3) Integrating $Q'(y) = 1 + 3y^2$ gives $Q(y) = y + y^3$.

We have now found that $F(x,y)$ is the function $x^2y + y + y^3$ and therefore

$$x^2y + y + y^3 = C \tag{2.37}$$

is an implicit solution of (2.36). In this example, we can use the initial condition $y(0) = 1$ to find the constant C. Letting $x = 0$ and $y = 1$ in (2.37) gives $(0)(1)+1+(1)^3 = 2 = C$, and the unique solution to (2.36) is $x^2y + y + y^3 = 2$.

We can go one step further and find an explicit formula for the solution y. There is a formula for roots of cubic polynomials (see the CRC tables) which is messy, but can sometimes be used to produce a reasonable result. In this case it provides the following explicit solution:

$$y(x) = \left(\sqrt{1 + \left(\frac{x^2+1}{3}\right)^3} + 1\right)^{1/3} - \left(\sqrt{1 + \left(\frac{x^2+1}{3}\right)^3} - 1\right)^{1/3}. \tag{2.38}$$

This solution curve is shown in the slope field in Figure 2.12. At this point, it would be interesting to see what your computer algebra system gives as a solution to (2.36). It may not be written in exactly the same form as (2.38), but you can check that it is the same function by graphing both functions on the same set of axes. ■

The method for solving exact equations can be extended by also allowing for multiplication by an integrating factor. This technique can be used to make an equation of the form $P(x,y)dx + Q(x,y)dy = 0$ into an exact equation. An excellent discussion of this method can be found in the book by Polking, Boggess, and Arnold[1].

[1]J. Polking, A. Boggess, and D. Arnold, *Differential Equations with Boundary Value Problems, 2nd ed.*, Pearson, Prentice Hall, 2005.

2.5.2 Bernoulli Equations.

An equation of the form

$$x' = p(t)x + q(t)x^n, \quad n \neq 0, 1 \tag{2.39}$$

is called a **Bernoulli equation**. If $n = 0$ or $n = 1$ the equation is easily seen to be linear, and our method for solving linear equations can be used. For any other value of n, the substitution of a new dependent variable $v(t) = (x(t))^{1-n}$ turns the equation into a linear equation in v. To see this, differentiate $v(t)$ by the chain rule:

$$v' = \frac{d}{dt}(x^{1-n}) = (1-n)x^{-n}x'.$$

Multiplying equation (2.39) by x^{-n},

$$x^{-n}x' = p(t)x^{1-n} + q(t) = p(t)v + q(t),$$

and therefore

$$v' = (1-n)(p(t)v + q(t)) = (1-n)p(t)v + (1-n)q(t), \tag{2.40}$$

which is a linear differential equation for v.

Example 2.5.5. *As a first example we will show that the logistic growth equation*

$$P' = rP(1 - P/N) \tag{2.41}$$

is a Bernoulli equation for any values of the parameters r and N.

Solution. We can rewrite (2.41) in the form

$$P' = rP - \frac{r}{N}P^2$$

and this has the form of a Bernoulli equation with $n = 2$, $p(t) = r$, and $q(t) = -r/N$. Letting $v = P^{1-n} = P^{-1} = 1/P$ and using (2.40), the equation for v is

$$v' = (1-2)p(t)v + (1-2)q(t) = -rv + \frac{r}{N}.$$

The linear equation $v' + rv = \frac{r}{N}$ can be solved by multiplying by the integrating factor $\mu = e^{rt}$, which gives

$$e^{rt}v' + re^{rt}v = \frac{r}{N}e^{rt} \Longrightarrow \frac{d}{dt}(e^{rt}v) = \frac{r}{N}e^{rt},$$

and integrating,

$$e^{rt}v = \frac{r}{N} \cdot \frac{e^{rt}}{r} + C.$$

If both sides are multiplied by e^{-rt},

$$v = 1/N + Ce^{-rt}.$$

To find P, we substitute back into $v = 1/P$ to write

$$P = 1/v = \frac{1}{1/N + Ce^{-rt}} = \frac{N}{1 + NCe^{-rt}},$$

which is equivalent to the solution we found in Section 2.1 by separation of variables. If we let $C_1 = -NC$ the two forms are the same. ■

Another Bernoulli equation is considered in the next example.

Example 2.5.6. *Solve the initial-value problem*

$$y' = -2y + e^t y^3, \quad y(0) = 1. \tag{2.42}$$

Solution. This is a Bernoulli equation with $n = 3$, $p(t) = -2$, and $q(t) = e^t$. If we make the substitution $v = y^{1-n} = y^{-2}$, then the equation for v is

$$v' = (1-3)\cdot(-2)\cdot v + (1-3)\cdot e^t = 4v - 2e^t.$$

Solving for v by our method for linear equations, with $\mu = e^{-4t}$,

$$e^{-4t}v' - 4e^{-4t}v = -2e^t \cdot e^{-4t} = -2e^{-3t} \Longrightarrow \int \frac{d}{dt}(e^{-4t}v)dt = \int(-2e^{-3t})dt$$

$$\Longrightarrow e^{-4t}v = (-2e^{-3t})/(-3) + C \Longrightarrow v(t) = \frac{2}{3}e^t + Ce^{4t}.$$

To find y, note that $v = y^{-2}$ implies that $y = \pm\sqrt{1/v}$. Therefore,

$$y(t) = \pm\left(\frac{2}{3}e^t + Ce^{4t}\right)^{-\frac{1}{2}}. \tag{2.43}$$

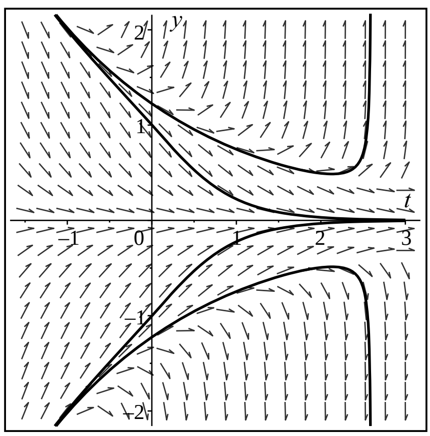


Figure 2.13. Slope field for $y' = -2y + e^t y^3$

To satisfy the initial condition $y(0) = 1$, set $\pm(\frac{2}{3} + C)^{-\frac{1}{2}} = 1$. We must use the plus sign, and then $C = \frac{1}{3}$. The unique solution of the initial value problem (2.42) is

$$y(t) = 1/\sqrt{\frac{2}{3}e^t + \frac{1}{3}e^{4t}} = \sqrt{\frac{3}{e^t(2 + e^{3t})}},$$

and it is defined for all real t. ■

It is important in Example 2.5.6 to note that $y = 0$ is a constant solution of the differential equation (2.42), but it is not given by the "general" formula (2.43). This means that when using Bernoulli's Method it is necessary to check for constant solutions first.

The Uniqueness and Existence Theorem can be used to show that (2.42) has unique solutions through every point of the (t, y)-plane. It is interesting to speculate whether the positive solutions will tend to 0 or to $+\infty$ as $t \to \infty$. One might suppose that the term $e^t y^3$ in the slope function would drive solutions to $+\infty$. However, the slope field shown in Figure 2.13, as well as the algebraic solution, suggests that there is a positive value of $y(0)$ below which all positive solutions tend to 0 and above which they tend to $+\infty$, possibly at a vertical asymptote. Exercise 21. below asks you to find this value of $y(0)$.

2.5.3 Using Symmetries of the Slope Field.

In Figure 2.13 a certain type of symmetry is evident in the slope field for the equation in Example 2.5.6. If we write the equation as $y' = f(t, y) = -2y + e^t y^3$, we can see analytically that the slope function f satisfies

$$f(t, -y) = -2(-y) + e^t(-y)^3 = -(-2y + e^t y^3) = -f(t, y);$$

that is, the slope function is symmetric about the t-axis.

Any first-order differential equation $y' = f(t, y)$ with f satisfying $f(t, -y) \equiv -f(t, y)$ has the property that if $y(t)$ is any solution of the equation, then $-y(t)$ is also a solution. To prove this, assume $y' = f(t, y)$, $f(t, -y) \equiv -f(t, y)$, and let $w(t) = -y(t)$. Then

$$w' = -y' = -f(t, y) = -f(t, -w) \equiv -(-f(t, w)) = f(t, w),$$

where the next-to-last equality uses the symmetry property of the slope field. Having shown that $w' = f(t, w)$, we can conclude that the function $w \equiv -y$ is a solution of the same differential equation satisfied by y. This symmetry of the solution curves about the t-axis can be clearly seen in Figure 2.13.

Exercises 2.5. *Determine whether or not equations 1–6 are exact.*

1. $(x + y)dx + xdy = 0$
2. $(2x + y)dx + (x - y)dy = 0$
3. $\sin(y)dx + x\cos(y)dy = 0$
4. $ye^x dx + xe^y dy = 0$
5. $2xydx + (x^2 + y^2)dy = 0$
6. $(x^2 + xy + 3y^2)dx + (y^2 + xy + 3x^2)dy = 0$

Find a general solution for equations 7–10.

7. $(x + y)dx + (x + 1)dy = 0$
8. $(2 + y)dx + (x - 3)dy = 0$
9. $(y + \sin(y))dx + (1 + x + x\cos(y))dy = 0$
10. $(xy^2 + 2y)dx + (x^2 y + 2x + 1)dy = 0$

For equations 11–14, show that the equation is exact, and solve the IVP:

11. $(1 + xy)dx + \frac{1}{2}x^2 dy = 0,\ y(1) = 1$
12. $(x + y)dx + (x + 1)dy = 0,\ y(0) = 2$
13. $\sin(y)dx + (x\cos(y))dy = 0,\ y(1) = \frac{\pi}{2}$

14. $ye^x dx + (2 + e^x)dy = 0,\ y(0) = 1$

Solve the following Bernoulli equations.

15. $P' = 2P(1 - P/4)$
16. $P' = P(2 - 5P)$
17. $y' = -y + e^t y^2,\ y(0) = 1$
18. $y' = 2y + e^{3t} y^2,\ y(0) = -1$
19. $y' + y = ty^3$
20. $y' + 2y = -4ty^3$
21. *For the Bernoulli equation $y' = -2y + e^t y^3$ in Example 2.5.6, find the positive solution that separates solutions tending to zero from those having a vertical asymptote at a positive value of t. What is the initial value $y(0)$ for this solution?*
22. *In the PHYSICS example in Chapter* 1.3, *the equation for the velocity of a skydiver was given as*

$$v'(t) = g - \frac{k}{m}(v(t))^p.$$

Assume now that $p = 2$, so the term representing the friction due to the air is proportional to the square of the velocity.

(a) *Define the constants $\alpha = \frac{k}{m}$ and $b = \sqrt{\frac{mg}{k}}$. Show that the function $x(t) = v(t) + b$ satisfies the Bernoulli equation*

$$x'(t) = (2b\alpha)x(t) - \alpha(x(t))^2.$$

(b) *Solve this Bernoulli equation for $x(t)$, and show that the general solution is*

$$v(t) \equiv x(t) - b = \frac{b(1 - 2bCe^{-2b\alpha t})}{1 + 2bCe^{-2b\alpha t}}.$$

(c) *Find the terminal velocity $\bar{v} = \lim_{t\to\infty}(v(t))$. Is it equal to $\left(\frac{mg}{k}\right)^{\frac{1}{p}}$ as stated in Chapter* 1.3?

2.6 Numerical Methods

In this section you will be shown how solutions to initial-value problems

$$x'(t) = f(t, x), \quad x(t_0) = x_0, \tag{2.44}$$

can be numerically approximated. The idea is not to make you an expert in numerical analysis, but to provide you with enough information so you will know when an approximate solution is accurate to the number of significant digits required by your work.

A numerical method for solving (2.44) produces a set of discrete points

$$(t_0, x_0), (t_1, x_1), \ldots, (t_N, x_N)$$

where $t_j = t_0 + j\Delta t$ and x_j is an approximation to the value of the solution at t_j; that is, $x_j \approx x(t_j)$. Once this list of points is obtained, it is usually plotted by connecting the points by straight-line segments; or, if you are using a sophisticated computer package, a curve fitting routine may be used to fit a smooth curve through the points.

For the three methods described here, it will be seen that the accuracy of the method depends directly on the step size Δt. For a given method, the smaller the value of Δt, the more accurate the approximation as long as the word size of the computer is large enough to avoid round-off error. However, it will also be shown that the error depends significantly on the method being used to compute the approximation.

If the slope function f satisfies the conditions of the Existence and Uniqueness Theorem 2.1, it is always possible to compute a numerical approximation to the solution. In theory, if f can be differentiated enough times, $x(t)$ could be approximated by the first $N+1$ terms in its Taylor series at $t = t_0$; that is, by

$$x(t_0 + h) = x(t_0) + x'(t_0)h + \frac{x''(t_0)h^2}{2!} + \cdots + \frac{x^{(N)}(t_0)h^N}{N!} + E_N(h), \tag{2.45}$$

where the error $E_N(h)$ in this approximation is known from calculus to be

$$E_N(h) = \frac{x^{(N+1)}(\xi)h^{N+1}}{(N+1)!} \tag{2.46}$$

for ξ some value of t between t_0 and $t_0 + h$.

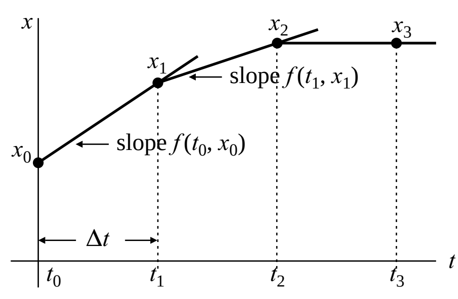
In general, this is not a good procedure to use because, as most calculus students already know, it may take a very large number of terms in the series to approximate $x(t)$ accurately for t very far from t_0. Furthermore, for each different differential equation the formula $f(t, x)$ for x' changes and the derivatives $x'', x''', \ldots$ would all have to be recalculated. These can be very complicated calculations; for example, the chain rule gives $x''(t) = \frac{d}{dt}(x'(t)) = \frac{d}{dt}(f(t, x(t))) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}x'(t)$, and higher derivatives become successively more complex. The numerical methods described in this section avoid these problems by starting at $t = t_0$, with a *small* Δt. The value $x(t_0 + \Delta t)$ is approximated by a *small* number of terms in the Taylor series, and then the process is repeated at $t_1 = t_0 + \Delta t, t_2 = t_0 + 2\Delta t, \ldots$. There will be a small error at each step, due to truncating the series, and these errors can accumulate.

2.6.1 Euler's Method.

Euler's Method is one of the oldest and simplest numerical methods for obtaining approximate solutions to the initial value problem (2.44). It uses only the first two terms of the Taylor series; that is, on each interval $[t_j, t_{j+1}]$ it approximates $x(t)$ by the *tangent line approximation* at the left end of the interval. This approximation is

$$x(t_j + \Delta t) \approx x(t_j) + x'(t_j)\Delta t \equiv x(t_j) + f\left(t_j, x(t_j)\right)\Delta t.$$

Notice that *no derivatives* of f have to be evaluated, only the function f itself.



Algorithm for Euler's Method
Given $x' = f(t,x),\ x(t_0) = x_0$, to find approximate values of $x(t)$ on the interval $t_0 \le t \le t_{max}$:
(1) Choose a small stepsize $\Delta t = \frac{t_{max}-t_0}{N}$, with N a positive integer.
(2) For $j = 0, 1, \ldots, N-1$ compute

$$x_{j+1} = x_j + f(t_j, x_j)\Delta t$$
$$t_{j+1} = t_j + \Delta t.$$

(3) Plot the points $(t_j, x_j),\ j = 0, 1, \ldots, N$. If straight lines are drawn between consecutive points, this results in a piecewise linear approximation to the solution $x(t)$ on the interval $[t_0, t_{max}]$.

Example 2.6.1. *Approximate the solution of* $x' = t - x,\ x(0) = 1$, *on the interval* $[0, 2]$.

Solution. As a first try we arbitrarily let $\Delta t = \frac{2-0}{N}$ with N chosen to be 4. Then $\Delta t = 0.5$. It helps to make a table, as shown below. The values $t_j = t_0 + j\Delta t$ are all determined once Δt is chosen.

j	t_j	x_j	$f(t_j, x_j) = t_j - x_j$	$x_{j+1} = x_j + f(t_j, x_j)\Delta t$
0	0	1.0	-1.0	$1.0 + (-1.0)(0.5) = 0.5$
1	0.5	0.5	0	$0.5 + 0(0.5) = 0.5$
2	1.0	0.5	0.5	$0.5 + 0.5(0.5) = 0.75$
3	1.5	0.75	0.75	$0.75 + (0.75)(0.5) = 1.125$
4	2.0	1.125		

■

Comments.

- The value of x_{j+1} at the end of each row is used as the value of x_j in the following row. The initial conditions give the values of t_0 and x_0 in the first line.

- The function $f(t_j, x_j)$ is the slope function and depends on the differential equation being solved.

Using the analytic method for solving linear differential equations, the exact solution of $x' = t - x$, $x(0) = 1$ is $x(t) = t - 1 + 2e^{-t}$. Check it! This gives the exact value $x(2) = 2 - 1 + 2e^{-2} \approx 1.27067$. The absolute error in our numerically computed value of $x(2)$ in the abovetable is $|1.27067 - 1.125| \approx 0.14567$.

If the calculations are redone with $N = 8$, that is, with $\Delta t = 0.25$, the corresponding table is

j	t_j	x_j	$f(t_j, x_j) = t_j - x_j$	$x_{j+1} = x_j + f(t_j, x_j)\Delta t$
0	0	1.0	−1.0	0.75
1	0.25	0.75	−0.50	0.625
2	0.50	0.625	−0.125	0.59375
3	0.75	0.59375	0.15625	0.6328125
4	1.00	0.6328125	0.3671875	0.72460938
5	1.25	0.72460938	0.52539063	0.85595703
6	1.50	0.85595703	0.64404297	1.01696777
7	1.75	1.01696777	0.73303223	1.20022583
8	2.00	1.20022583		

The error in the approximate value for $x(2)$ is now $|1.27067 - 1.200226| \approx 0.0704$. By cutting the step size Δt in half the error has been cut approximately in half, from 0.146 to 0.070. Figure 2.14 shows the two approximate solutions plotted with the exact solution.

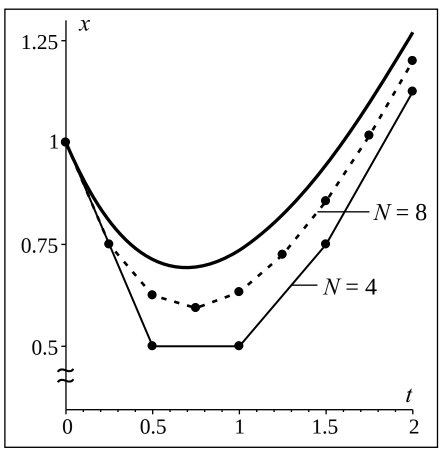


Figure 2.14. Euler approximations to $x' = t - x,\ x(0) = 1$, with $N = 4, 8$

Error and the Order of a Numerical Method. Two different types of error can occur when solving a differential equation numerically. The first is called **round-off error**, and is due to rounding or chopping the result after each step in the procedure, in order to store it in a computer with fixed word size. For example, if $x(0) = \frac{1}{3}$, and the computer running the program has a word size of 4 decimal digits, then it would probably store $x(0)$ as 0.3333, already introducing an error greater than 0.00003 into the calculation. Fortunately, modern computers can store a lot more digits than this so it is usually not a problem, unless the procedure requires a huge number of steps.

The type of error you really need to be concerned about is caused by the approximations made by whatever method you are using. This is called **truncation error**, and in the case of Euler's Method it is caused by dropping all but the first two terms in the Taylor series for $x(t+\Delta t)$. From (2.46), it is seen that the maximum error made by using the approximation

$$x(t+\Delta t) \approx x(t) + x'(t)\Delta t = x(t) + f(t, x(t))\Delta t$$

is given by $E_1(\Delta t) = x''(\xi)(\Delta t)^2/2$ where ξ is some value of t between t and $t+\Delta t$. If we know that $|x''(t)| \leq M$ on the entire interval $[t_0, t_{max}]$, then we can say that the error E made in a single step (called the **local truncation error**) is bounded by

$$E \leq \frac{M}{2}(\Delta t)^2.$$

Numerical analysts refer to this as an **error of order 2 in** Δt, and usually write it as $\mathcal{O}((\Delta t)^2)$.

Definition 2.5. *An error of order n, denoted by* $\mathcal{O}((\Delta t)^n)$*, is an error which approaches a constant times* $(\Delta t)^n$ *as* $\Delta t \to 0$.

Since Euler's Method takes N steps to go from t_0 to t_{max}, and $N = \frac{t_{max}-t_0}{\Delta t}$, it might seem that the total error in the calculation would be bounded by N times the local error E. The product NE is a constant times Δt, hence is $\mathcal{O}(\Delta t)$. Unfortunately this is not quite correct. After the first step we are using the approximation x_1 instead of the true value of $x(t_1)$. When we write $x_2 = x_1 + \Delta t f(t_1, x_1)$ we not only have the truncation error to worry about, but also the error in x_1, and these errors can propagate.

In fact it can be shown that the total error increases exponentially over the total number of steps. An excellent discussion of this is given in the book by Hubbard and West[2], in which it is shown that the total error over the N steps is bounded by

$$E(t_{max}) \leq \frac{C}{K}\left(e^{K|t_{max}-t_0|} - 1\right)\Delta t$$

for some constant C, where K can be taken to be the maximum value of $|\frac{\partial f}{\partial x}|$ on a rectangle containing the entire solution. For a fixed length interval $|t_{max} - t_0|$, this will still be just a constant times Δt, but if the interval of integration is increased, the error can grow exponentially. This should warn you that using Euler's Method over very long time intervals may produce large errors even if Δt is very small.

Based on the preceding discussion, we say that Euler's Method is a *first-order* numerical method. What this means, practically, is that if an equation is solved using Euler's Method with a given Δt, and then again with $\frac{\Delta t}{2}$, the error in the second calculation should be about $\frac{1}{2}$ of the error in the first.

---

[2]J. H. Hubbard and B. H. West, *Differential Equations, A Dynamical Systems Approach, Part I*, Springer Verlag, 1990.

An nth-order numerical method is one whose error is $\mathcal{O}((\Delta t)^n)$; that is, the error is equal to a constant times $(\Delta t)^n$, in the limit as $\Delta t \to 0$. If an equation is solved by an nth-order method with a given Δt, and then again with $\frac{\Delta t}{2}$, the error in the second calculation should be about $(\frac{1}{2})^n$ times the error in the first. This is what makes higher-order methods so accurate for small values of Δt.

We will see that higher-order methods are needed in real-world applications, but Euler's Method is by far the easiest one to understand.

2.6.2 Improved Euler Method. A second-order numerical method called the Improved Euler Method, or Heun's Method, uses two values of the slope function in each interval $[t_j, t_{j+1}]$. The slope $m_0 = f(t_j, x_j)$ at the left end of the interval is computed first. Then the Euler approximation $\bar{x}_{j+1} = x_j + m_0\Delta t$ is used to calculate an approximation m_1 to the slope at the right-hand end of the t-interval. The average slope $m = \frac{m_0+m_1}{2}$ is used to compute $x_{j+1} = x_j + m\Delta t$ which is the improved approximation of $x(t_{j+1})$. See Figure 2.15.

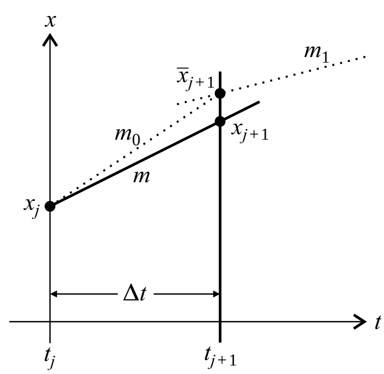


Figure 2.15. Improved Euler Method

This results in the formula

$$x_{j+1} = x_j + \frac{\Delta t}{2}(m_0 + m_1) = x_j + \frac{\Delta t}{2}\left[f(t_j, x_j) + f\left(t_j + \Delta t, x_j + \Delta t f(t_j, x_j)\right)\right].$$

With some difficulty, it can be shown that this formula for x_{j+1} agrees with the Taylor series for $x(t_j + \Delta t)$ in its first *three* terms. This means that the truncation error at each step is $\mathcal{O}\left((\Delta t)^3\right)$; therefore, using the same argument as before, the accumulated error in the Improved Euler Method, over the interval $[t_0, t_{max}]$, is $\mathcal{O}\left((\Delta t)^2\right)$. Hence, this is a **second-order method**, and cutting the step size in half reduces the error by approximately $\frac{1}{4} = (\frac{1}{2})^2$.

Algorithm for the Improved Euler Method

Replace Step (2) in the Euler Method by the following:

(2′) For $j = 0, 1, \dots, N-1$ compute

$$\begin{aligned} m_0 &= f(t_j, x_j) \\ \tilde{x}_{j+1} &= x_j + m_0 \Delta t \\ t_{j+1} &= t_j + \Delta t \\ m_1 &= f(t_{j+1}, \tilde{x}_{j+1}) \\ m &= \frac{m_0 + m_1}{2} \\ x_{j+1} &= x_j + m \Delta t. \end{aligned}$$

Example 2.6.2. *Use the Improved Euler Method, with* $\Delta t = 0.5$*, to approximate the solution of* $x' = t - x$, $x(0) = 1$*, on the interval* $[0, 2]$.

j	t_j	x_j	$m_0 = t_j - x_j$	$\tilde{x}_{j+1} = x_j + m_0 \Delta t$	$m_1 = t_{j+1} - \tilde{x}_{j+1}$	$m = \frac{m_0+m_1}{2}$	x_{j+1}
0	0	1.0	−1.0	0.5	0	−0.5	0.75
1	0.5	0.75	−0.25	0.625	0.375	0.0625	0.78125
2	1.0	0.78125	0.21875	0.89063	0.60938	0.41406	0.98828
3	1.5	0.98828	0.51172	1.24414	0.75586	0.63379	1.30518
4	2.0	1.30518					

Solution. The absolute error in $x(2)$ is approximately $|1.30518 - 1.27067| = 0.03451$, which is even less than the error we found using Euler's Method with twice as many steps. ■

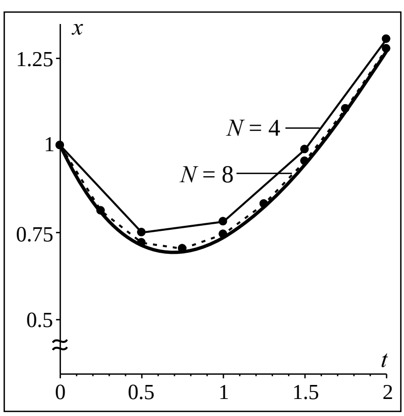


Figure 2.16. Improved Euler approximations to $x' = t - x$, $x(0) = 1$, with $N = 4, 8$

With $N = 8$, the Improved Euler Method gives an approximation of 1.27756 for $x(2)$, which is in error by approximately 0.00689. Figure 2.16 shows the two approximate solutions with $N = 4$ and $N = 8$. Comparing these with Figure 2.14 shows the distinct improvement, and also the fact that the solutions now converge to the exact solution from above.

By halving the step size, the error has been reduced to less than a quarter of its size, which is what we would expect with a second-order method. The Improved Euler Method requires two evaluations of $f(t, x)$ for each step, but again no derivatives of f have to be calculated.

2.6.3 Fourth-order Runge-Kutta Method. A lot of work was done in the first half of the twentieth century to develop methods that are far more accurate. Most computer algebra systems use what are called Runge-Kutta Methods, developed by two German mathematicians. One of the major motivations behind these methods was to find a way to avoid having to differentiate the slope function. The fourth-order Runge-Kutta Method, for example, uses four evaluations of the slope function on each t-interval to produce an approximation which agrees with the first five terms of the Taylor series, and therefore has an accumulated error $\mathcal{O}\left((\Delta t)^4\right)$ on the interval $[t_0, t_{max}]$.

There are many different Runge-Kutta formulas for any given order, but the algorithm below is the one that has gained acceptance because of its simplicity. Although it is tedious to compute by hand, it is easy to program for a calculator or computer. If a differential equation needs to be solved within a Maple or *Mathematica* procedure, for example, the algorithm given below can be used exactly as written.

Algorithm for the fourth-order Runge-Kutta Method
Replace Step (2) in Euler's Method by the following:
(2*) For $j = 0, 1, \ldots, N-1$ compute

$$\begin{aligned}
m_1 &= f(t_j, x_j) \\
m_2 &= f(t_j + \frac{\Delta t}{2}, x_j + m_1\frac{\Delta t}{2}) \\
m_3 &= f(t_j + \frac{\Delta t}{2}, x_j + m_2\frac{\Delta t}{2}) \\
m_4 &= f(t_j + \Delta t, x_j + m_3\Delta t) \\
x_{j+1} &= x_j + \frac{\Delta t}{6}(m_1 + 2m_2 + 2m_3 + m_4) \\
t_{j+1} &= t_j + \Delta t.
\end{aligned}$$

For the IVP $x' = t - x,\ \ x(0) = 1$, four steps of the fourth-order Runge-Kutta Method with $\Delta t = 0.5$ result in a value of $x(2) \approx 1.27110$. The absolute error in this value is $|1.27067 - 1.27110| = 0.00043$, which is much smaller than the error in either Euler's Method or the Improved Euler Method, as expected. If Δt is halved, the error in $x(2)$ will be approximately $\frac{1}{16} = (\frac{1}{2})^4$ times its original size. This means that halving the step size Δt will result in an answer with at least one more significant decimal digit. The following table shows the values of $x(2)$, using the fourth-order Runge-Kutta method with $\Delta t = 0.25, 0.125, 0.0625$, and 0.03125.

Δt	$x(2)$
0.25	1.2706 9228
0.125	1.2706 7178
0.0625	1.2706 7063
0.03125	1.2706 7057

When solving real-world problems, it is often the case that no exact solution is possible; it then becomes critical to be able to estimate the accuracy of a numerical solution. Based on the preceding discussion, if one is using the fourth-order Runge-Kutta Method, it is reasonable to compare two solutions having step sizes Δt and $\Delta t/2$, and use the position of the digit in which they differ to estimate how many significant digits are correct.

Example 2.6.3. *In Example 2.2.2 we drew a slope field for the differential equation $x' = x^2 - t$, and remarked that no simple analytic formula for the solution exists. The slope field and four numerically computed trajectories for this equation are seen in Figure 2.17. It appears that there may be a unique solution that separates solutions that tend to infinity as t increases from those that enter the region inside the parabola $x(t) = \pm\sqrt{t}$ and then tend asymptotically to $-\sqrt{t}$ as $t \to \infty$. Notice that once a solution crosses the upper branch of the parabola, the slope function $f(t,x) = x^2 - t$ becomes negative, implying that the solution must go down.*

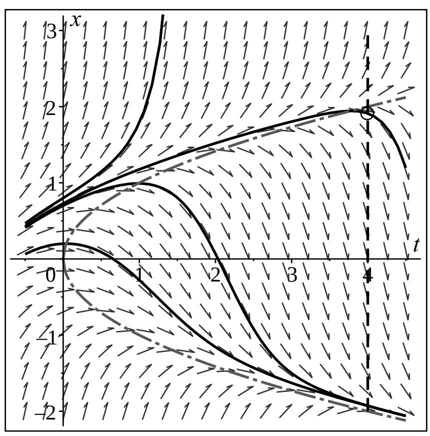


Figure 2.17. Solutions of $x' = x^2 - t$, $x(0) = 0.2, 0.7, 0.72901, 0.8$

One way to obtain an estimate for the initial value $\hat{x}_0$ for the special solution separating the two families is to pick initial conditions $(K, \sqrt{K})$ far out on the parabola (that is, with K large) and then integrate the solution back from $t = K$ to $t = 0$. Most numerical differential equation solvers allow you to enter a negative value for Δt, or alternatively to specify a range of t from 0 *to K and enter the initial condition as $x(K) = \sqrt{K}$. This was done with a very accurate Maple routine, and resulted in the following values:*

Each of the initial values in the table must be slightly below $\hat{x}_0$ since these three solution curves have already met the parabola at $x(K) = \sqrt{K}$, and therefore are destined to go down as t increases. It is easy to test the accuracy of this initial value by using initial

K	$x(0)$
4	0.7290109
9	0.7290112
16	0.7290112

conditions $x(0) = 0.7290112 \pm \varepsilon$ *to see if the solutions ultimately go in different directions. This would be a good way to test the accuracy of your differential equation solver. In most applications an accuracy of* 7 *decimal places is much more than is needed.*

You should now feel fairly confident in your ability to estimate the accuracy of the solution of a given IVP, even when the equation cannot be solved analytically. The table below shows the value of $x(4)$ *obtained by using the simple fourth-order Runge-Kutta method described above to solve the IVP*

$$x' = x^2 - t, \quad x(0) = 0.72901.$$

To include the possibility of round-off error, two calculations were made, one using 8 *significant digits in the calculations (comparable to having a word size of eight decimal digits), and one made with* 16 *digits. This is possible to do in Maple by using a variable called* `Digits`. *The table below shows the computed values of* $x(4)$ *for step sizes* $\Delta t = 0.125, 0.125/2, 0.125/2^2, \ldots, 0.125/2^{10}$, *for the two word sizes.*

Δt	$x(4)$, Digits=8	$x(4)$, Digits= 16
0.125	2.0645 996	2.0649 8090
0.0625	1.9151 634	1.9164 0489
0.03125	1.9036 732	**1.9**072 4317
0.015625	1.9099 448	**1.906**6 6497
0.0078125	1.9131 016	**1.9066** 2862
0.00390625	1.9417 954	**1.9066 2**634

In the right-hand column the values of $x(4)$ *are gaining one more significant digit each time the step size* Δt *is divided by two; and* $x(4) = 1.9066$ *can be assumed accurate to* 5 *significant digits. This is what we would expect to happen with our fourth-order method, and as long as the number of steps needed, relative to the significant digits used in the calculation is reasonable, this is what you will see.*

What you need to realize is that if you are working on a machine that keeps only a small number of significant digits in its calculations, then if Δt *is very small, or a huge number of steps are needed, bad things can happen. In the center column, with an* 8*-digit word size, the value of* $x(4)$ *never really converged to its true value. If* Δt *was made even smaller, the answer would soon become "garbage," just due to the round-off error at each step.*

Conclusion. *For reasonable problems where only three or four significant digits are needed in the answer, a fourth-order method is fine; and all you need to do is run it with a reasonably small step size, and then compare the answer to what you get with half that step size. You may need to repeat this until you see the final value agree to the number of significant digits you require. If an agreement does not occur, then you may need to increase the word size of the computer or calculator you are using, or you may need to* *find a different type of equation solver that is geared to your particular problem. There are many more sophisticated methods for numerically solving differential equations. To find out about these you should consult a book on numerical analysis.*

Exercises 2.6. *Problems 1 and 2 ask you to compute tables for the Euler Method and Modified Euler Method by hand, for the IVP $x' = t - x$, $x(0) = 1$. To make these a reasonable length, you are going to find values of $x(1)$, instead of $x(2)$ (as in the Examples). The exact solution is $x(t) = t - 1 + 2e^{-t}$, which gives $x(1) \approx 0.735759$.*

1. *For the IVP $x' = t - x, \quad x(0) = 1$,*
   (a) *Do 8 steps of Euler's Method, with $\Delta t = 0.125$, to find an approximation to $x(1)$.*
   (b) *Complete the following table (values are from Example 2.6.1):*

Δt	$x(1)$	error in $x(1)$
0.5	0.50000	0.235759
0.25	0.632813	0.102946
0.125	?	?

   (c) *Is the error in $x(1)$ what you would expect from a first-order method? Explain.*

2. *For the IVP $x' = t - x$, $x(0) = 1$,*
   (a) *Do 4 steps of the Modified Euler Method, with $\Delta t = 0.25$, to find an approximation to $x(1)$. Then do 8 steps with $\Delta t = 0.125$.*
   (b) *Complete the following table (values are from Example 2.6.2):*

Δt	$x(1)$	error in $x(1)$
0.5	0.781250	0.045491
0.25	?	?
0.125	?	?

   (c) *Are the errors in $x(1)$ what you would expect from a second-order method? Explain.*

3. *Use whatever technology you have available to solve (numerically) the IVP*
$$x' = t - x, \quad x(0) = 1.$$
Find out how to set the parameters and step size to obtain a value of $x(2)$ accurate to 6 decimal places. Explain exactly what you had to do.

4. *Use the same settings as in Exercise 3 to solve the IVP*
$$x' = x^2 - t, \quad x(0) = 0.729011.$$
Decrease the step size at least once to make sure your answer is exact to 5 decimal places.
   (a) *Does the solution intersect the parabola $t = x^2$?*

(b) *If it does, at what value of* t*?*

(c) *Does this seem reasonable? Explain. (Refer to Figure* 2.17.*)*

Problem 5.b *demonstrates an inherent danger in using approximate solutions.*

5. (a) *Compute the Euler Method approximation to the solution of* $x' = x^2,\ x(0) = 1,$ *on the interval* $[0, 1.2]$ *with* $\Delta t = 0.2$.

   (b) *Solve the differential equation exactly (it is separable), and explain what you find. What is happening in (a)?*

6. *Use your numerical equation solver to find the value of* $x(0.8)$ *for the solution of* $x' = x^2$, $x(0) = 1$. *Compare this with the exact solution obtained in problem* 5.b.

7. *Using a computer algebra system, such as Maple, see what it gives for* $x(1)$ *and* $x(1.2)$ *for the IVP* $x' = x^2$, $x(0) = 1$. *Explain what happens.*

For each problem 8–13, use an appropriate analytic method to solve the IVP. Then use your own numerical equation solver to solve the problem, and compare the value of the numerical solution at $t = 5$ *with the analytic solution at* $t = 5$. *Determine the exact error in your numerical result.*

8. $x' = \frac{t+2}{x},\quad x(0) = 2$

9. $x' + \frac{1}{5}x = e^{-0.2t}\sin(t),\quad x(0) = -1$

10. $x' = \frac{1-t/2}{2+x},\quad x(0) = -1$

11. $tx' = 2x + t^2,\quad x(1) = 1$

12. $\sin(y)dt + t\cos(y)dy = 0,\quad y(1) = \pi/2$ *(Exact equation)*

13. $y' = -y + e^t y^2,\quad y(0) = 1$ *(Bernoulli equation)*

COMPUTER PROBLEMS. In Section 1.2, the Maple instruction

```
dsolve({ODE,initcond},y(t))
```

was introduced to produce the exact solution of a solvable differential equation. If the equation is not solvable, this same instruction can be used to produce a numerical solution. The example below will produce a numerical solution of the IVP $y' = t/y$, $y(0) = 1$, which has the exact solution $y(t) = \sqrt{t^2 + 1}$, but will be used to illustrate the method.

In Maple, enter the instruction

```
sol:=dsolve({y'(t)=t/y(t),y(0)=1},type=numeric);
```

This returns a procedure for computing numerical values of the solution.

`sol(1.5);` produces the output `[t=1.5,y(t)=1.80277548...]`

To pick off the value of $y(1.5)$, use

`ans:=op(2,op(2,sol(1.5)));` This will be just the number 1.80277548... .

The corresponding instructions in *Mathematica* are

```
approxsol = NDSolve[{y'[t]==t/y[t],y[0]==1},y[t],{t,0,2}]
```

`ya[t_]=y[t]/.First[approxsol]` returns a list `ya` containing approximate values of the solution on the interval $[0, 2]$. To obtain a value at a specific point, `ya[1.5]` returns the value 1.80278, found by interpolation using the list of approximate values of y on the interval $[0, 2]$.

2.7 Autonomous Equations, the Phase Line

A first-order differential equation $x' = f(t, x)$ is called **autonomous** if the slope function f depends only on x, and not explicitly on t. In other words, the rate of change of x does not depend on time, but only on the current state of the dependent variable x. The equations $x' = x^2$ and $P' = rP(1 - P)$ are autonomous; the equations $x' = x + t$ and $x' = x^2 - t$ are not. An autonomous first-order equation can be written in the form

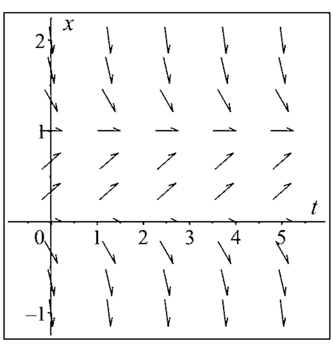
$$x' \equiv \frac{dx}{dt} = f(x). \tag{2.47}$$

An autonomous first-order equation is always separable.

Solutions of autonomous first-order equations have very limited types of behavior. Suppose that $f(r) = 0$ for some real number r. Then the constant function $x(t) \equiv r$ is a solution of (2.47), since both $f(x)$ and x' are identically zero for all values of t.

Definition 2.6. *A constant function $x(t) \equiv r$, such that $f(r) = 0$, is called an **equilibrium solution** of* (2.47).

Since the slopes depend only on x, a slope field (as defined in Section 2.2) for an autonomous equation $x' = f(x)$ is completely determined once slopes along any vertical line $t = \bar{t}$ are plotted. In fact, if $f(x)$ is defined and continuous for all x, the behavior of solutions of the equation can be determined from the slope lines along the x-axis. This leads to the construction of what is called a **phase line** for the differential equation (2.47).



Slope field for $x' = 2x(1 - x)$

phase line

To draw a phase line for the equation $x' = f(x)$:

- Find all real numbers $r_1 < r_2 < \cdots$ such that $f(r_i) = 0$, and label these values on a vertical x-axis. These points represent the equilibrium solutions. We assume first that the function f is equal to zero at only finitely many values of x; if not, a phase line can always be drawn using a finite interval on the x-axis.
- For each interval $(-\infty, r_1), (r_1, r_2), \ldots$ pick any value $x = \bar{x}$ in the interval and determine whether $f(\bar{x})$ is positive or negative. Draw an arrow on the axis, in the given interval, pointing up if $f(\bar{x})$ is positive and down if $f(\bar{x})$ is negative. Note that if f is a continuous function, it will have constant sign between any two zeros.

A solution with initial value satisfying $r_i < x(0) < r_{i+1}$ is monotonically increasing if the arrow points up or monotonically decreasing if the arrow points down. If $f(x)$ satisfies the conditions of the Existence and Uniqueness Theorem 2.1, then the solution must remain bounded between the two equilibrium values, since they are solutions and solutions cannot intersect.

Example 2.7.1. *Draw a phase line for the differential equation*

$$x' = f(x) = 0.5(x - 1)(x + 2).$$

Solution. This equation has two equilibrium solutions: $x_1 \equiv 1$ and $x_2 \equiv -2$. These are shown plotted on a phase line in Figure 2.18. When $x = 0$, $f(x) = 0.5(-1)(2) = -1 < 0$ so an arrow is drawn in the downward direction in the interval $(-2, 1)$. In the intervals $(-\infty, -2)$ and $(1, \infty)$, $f(x) > 0$ so the arrows point up. Now consider a solution with initial condition less than -2, for example $x(0) = -3$. Its slope will always be positive, but it can never cross the equilibrium solution $x(t) \equiv -2$ (do you see why?); therefore, it must increase monotonically and approach -2 as a horizontal asymptote, as $t \to \infty$. Similarly, a solution with $-2 < x(0) < 1$ must be monotonically decreasing and bounded between -2 and 1 for all t; therefore, it must approach -2 asymptotically as $t \to \infty$. If $x(0) > 1$, the solution is monotonically increasing. Whether it exists for all t or has a vertical asymptote at some positive value of t cannot be determined geometrically.

In Figure 2.19 solutions have been drawn for certain initial values at $t_0 = 0$, but the picture would look exactly the same if the same initial conditions were specified at any value of t_0. ■

The phase line contains almost all of the information needed to construct the graphs of solutions shown in Figure 2.19. It does not contain information on how fast the curves approach their asymptotes, or where the curves have inflection points, however. This information, which does depend on t, is lost in going to the phase line representation, but note that we did not need to solve the differential equation analytically in order to draw the phase line.

Solutions below the lowest equilibrium solution r_1 and above the highest equilibrium solution r_N may either be defined for all t or become infinite at a finite value of t. The logistic equation that was graphed in Figure 2.7 is an example of an autonomous equation. In that example, the solutions below $x \equiv 0$ and above $x \equiv 1$ were both

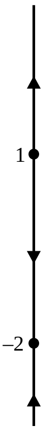


Figure 2.18. Phase line for $x' = 0.5(x-1)(x+2)$

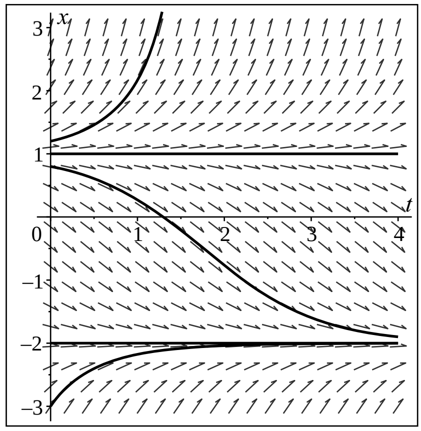


Figure 2.19. Solutions of $x' = 0.5(x-1)(x+2)$

shown to have vertical asymptotes, but in order to do this we had to be able to obtain an analytic solution of the equation.

If the slope function f has infinitely many zeros, for example $f(x) = \sin(x)$, which is zero at $x = 0, \pm\pi, \pm 2\pi, \ldots$, the phase line will only be able to show the behavior around a finite number of these.

2.7.1 Stability—Sinks, Sources, and Nodes.

In Figure 2.19 it can be seen that if solutions start initially close enough to the equilibrium solution $x \equiv -2$, they will tend toward it as $t \to \infty$. An equilibrium solution of this type (with arrows pointing toward it from both sides on the phase line) is called a **sink** and is said to be a *stable equilibrium*. On the other hand, solutions starting close to $x = 1$ all tend to move away from this solution as t increases. An equilibrium solution of this type (with arrows pointing away from it on both sides) is called a **source**. It is said to be an *unstable equilibrium*. If the arrows on the phase line point toward an equilibrium on one side and away from it on the other side, the equilibrium is called a **node**. It is *semi-stable* in the sense that if a solution starts on one side of the equilibrium it will tend toward it, and on the other side it will tend away as $t \to \infty$.

Example 2.7.2. *Draw a phase line for the equation*

$$x' = x(x+1)^2(x-3) = f(x)$$

and label each equilibrium point as a sink, source, or node.

Solution. The equilibrium solutions are the zeros of $f(x)$, namely $x = 0$, -1, and 3. These are shown plotted on the phase line in between Figures 2.20 and 2.21. To determine the direction of the arrows, it helps to draw a graph of the slope function $f(x) = x(x+1)^2(x-3)$. This is shown in Figure 2.20. Be careful not to confuse the graph of $f(x)$ with graphs of the solution curves $x(t)$ shown in Figure 2.21. The graph of $f(x)$ is only used to determine whether the arrow between two equilibria points up or down. It can be seen that $f(x)$ is positive between all pairs of equilibrium points except 0 and 3; therefore, all of the arrows point up except the arrow between 0 and 3.

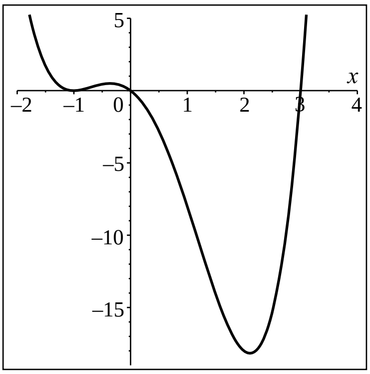
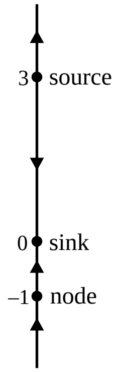


Figure 2.20. Slope function $f(x)$



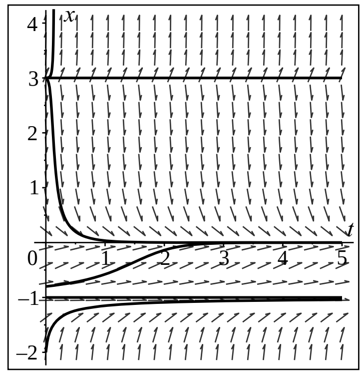


Figure 2.21. $x' = x(x+1)^2(x-3)$

Once the arrows are drawn, it is easy to see that -1 is a node, 0 is a sink, and 3 is a source. Some solutions of this equation are shown in Figure 2.21. ∎

Bifurcation in Equations with Parameters. If an autonomous differential equation contains a parameter α, so that

$$x' = f(x, \alpha),$$

then the phase line will change as α varies. Sometimes no significant change will occur, but there can be values of α at which a small change in the parameter can produce a large change in the qualitative structure of the vector field.

Definition 2.7. *If a small smooth change to the value of a parameter causes a sudden qualitative change in the behavior of the solutions of a differential equation, we say that a* **bifurcation** *has occurred. The value of the parameter at which the change occurs is called a* **bifurcation value** *of the parameter.*

According to Wikipedia, the name "bifurcation" was first introduced in a paper by Henri Poincaré in 1885, a fairly recent event in mathematical terms. For a simple first-order autonomous equation, the only way a bifurcation can occur is if the number or type of one or more of the equilibrium solutions changes.

Example 2.7.3. *Show that the differential equation*

$$x' = f(x, \alpha) = \alpha x - x^3$$

bifurcates when the parameter α passes through the value 0.

Solution. For any $\alpha \leq 0$, $f(x, \alpha) = x(\alpha - x^2) = 0$ has only the solution $x = 0$, but for $\alpha > 0$ there are three equilibrium points $x = 0, \pm\sqrt{\alpha}$. Since the number of equilibria changes as α passes through the value 0, the number 0 is a bifurcation value of α. The direction of the arrows on the phase lines can be determined from the two graphs in Figure 2.22 showing the slope functions $f(x, \alpha)$ for $\alpha = -1$ and $\alpha = +1$, and a characteristic phase line for each case. ∎

If a collection of phase lines is sketched for values close to the bifurcation value, we get what is called a **bifurcation diagram**. By convention, stable equilibria (sinks) are connected by solid curves and unstable equilibria (sources) by dotted curves. These curves show the values of x at which the equilibria occur for values of α between those for which phase lines are drawn. A bifurcation diagram for the equation in Example 2.7.3 is shown in Figure 2.23.

If you have ever seen a pitchfork, it should be clear from the picture in Figure 2.23 why this type of bifurcation is referred to as a **pitchfork bifurcation**. It occurs when a sink turns into a source surrounded by two new sinks.

2.7.1.1 Bifurcation of the Harvested Logistic Equation. In Section 2.2 an example was given (Example 2.2.3) of a harvested logistic growth problem for a herd of deer. Figure 2.5 showed three slope fields for the equation

$$P' = 0.4P\left(1 - \frac{P}{100}\right) - H \tag{2.48}$$

for three values of the harvesting parameter H. It was suggested by the slope fields that something strange happened to the behavior of the solution curves between the values $H = 6$ and $H = 12$. It is now possible to completely describe this phenomenon in terms of a bifurcation caused by the change in the parameter H.

For any fixed values of r, N, and H, the equilibrium solutions of the harvested logistic equation

$$P' = f(P,H) = rP\left(1 - \frac{P}{N}\right) - H$$

are the constant values of P where $f(P,H) = 0$; that is, they are the zeros of the quadratic polynomial

$$f(P,H) = -\frac{r}{N}P^2 + rP - H = -\frac{r}{N}\left(P^2 - NP + \frac{NH}{r}\right).$$

Using the quadratic formula, the zeros of $f(P,H)$ are at

$$P = \frac{N \pm \sqrt{N^2 - \frac{4NH}{r}}}{2} = \frac{N}{2} \pm \frac{N}{2}\sqrt{1 - \left(\frac{4}{rN}\right)H}.$$

Figure 2.24 shows the parabolic slope function $f(P,H)$ for several values of H. When $H = 0$, there are two equilibria, at $P = 0$ and $P = N$. The vertex of the parabola is at

$$f\left(\frac{N}{2}, 0\right) = r\frac{N}{2}\left(1 - \frac{(N/2)}{N}\right) = \frac{rN}{4}.$$

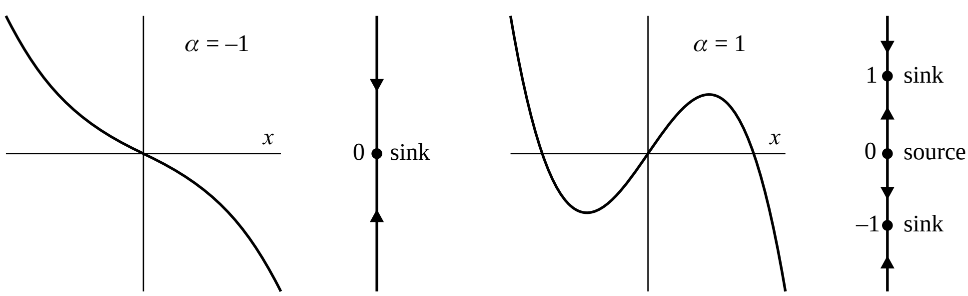


Figure 2.22. $f(x)$ and phase line for $x' = f(x) = \alpha x - x^3$, $\alpha = -1, 1$

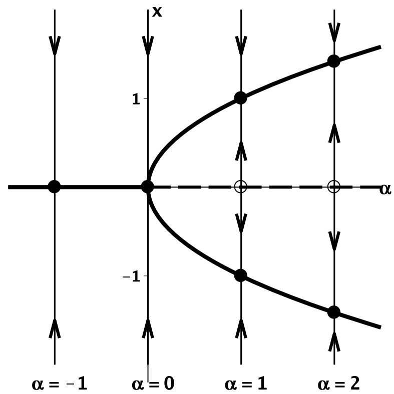


Figure 2.23. Bifurcation diagram for $x' = \alpha x - x^3$

As H increases, the parabola moves down due to the term $-H$, and the two roots get closer together. At the bifurcation value $H^* = \frac{rN}{4}$ there is a single root at $P = \frac{N}{2}$. For $H > \frac{rN}{4}$, there are no roots. Using the graph we can see how the arrows should go on phase lines for values of $H < \frac{rN}{4}, H = H^* = \frac{rN}{4}$, and $H > \frac{rN}{4}$. A characteristic phase line for each case is also shown in Figure 2.24.

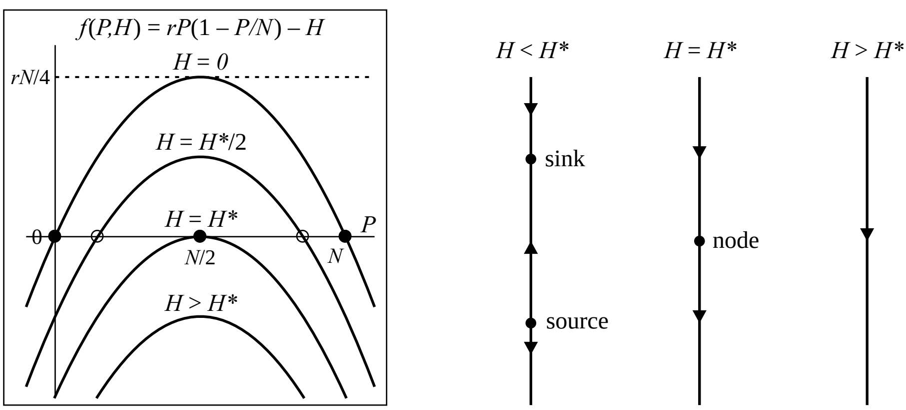


Figure 2.24. Phase lines for the harvested logistic equation $P' = rP(1 - P/N) - H$

For the problem in Example 2.2.3, the value of the carrying capacity was $N = 100$ and the intrinsic growth rate was $r = 0.4$. This gives a bifurcation value for the harvesting parameter $H^* = \frac{rN}{4} = (0.4)(100)/4 = 10$. You should now be able to explain exactly what is happening to the solutions shown in the slope fields in Figure 2.25.

In the left-hand graph $H = 9 < H^*$, and there are two equilibrium solutions; a stable solution with $P(0) \approx 65.8$ and an unstable solution with $P(0) \approx 34.2$. These values are given by the formula $P = \frac{N}{2} \pm \frac{N}{2}\sqrt{1 - \left(\frac{4}{rN}\right)H}$ for the roots of the quadratic. If the size of the herd is initially 66 or greater, it will decrease, but not go below the

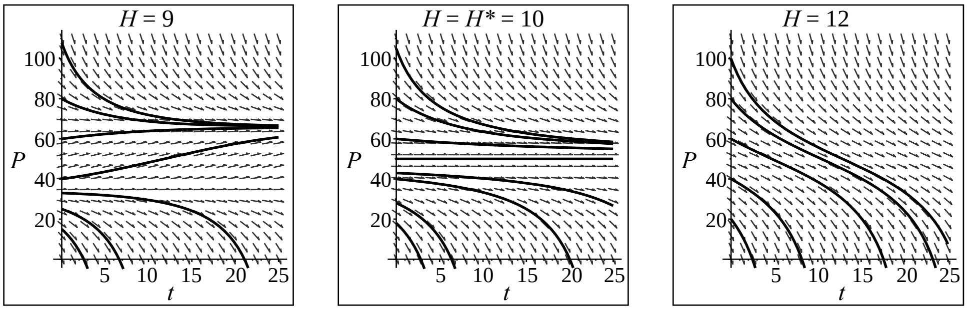


Figure 2.25. Deer population with $H = 9$, $H = H^* = 10$, and $H = 12$

value 65.8. If it is initially between 34 and 65 deer, it will increase to 65. If the herd initially has fewer than 34 deer, it will gradually go extinct.

In the middle graph, ten deer are being removed each year, and the equilibrium point is a node. If there are more than 50 deer initially, the herd will gradually decrease and stabilize at a size of 50, but if there are initially fewer than 50 deer, it will go extinct over time. In the right-hand graph we can see that if more than 10 deer are removed each year, the herd will go extinct no matter how large it is initially. The bifurcation value of the harvesting parameter is called the **critical harvesting rate**. The harvested logistic equation first appeared in a 1975 paper by Fred Brauer and David Sánchez[3], and has been written about by several authors since then.

Exercises 2.7. *In problems 1–6, draw a phase line for the autonomous differential equation and label each equilibrium point as a sink, source, or node.*

1. $x' = x(1 - x/4)$

2. $x' = x^2 - 1$

3. $x' = x^2$

4. $x' = (x^2 - 1)(x + 2)^2$

5. *(spruce-budworm growth equation)* $dP/dt = P(1 - P/5) - \dfrac{0.7P^2}{(0.05)^2 + P^2}$

6. $x' = \sin(x)$ *Hint: plot* $f(x) = \sin(x)$ *to determine the direction of the arrows.*

7. *The logistic growth equation* $x' = rx\left(1 - \frac{x}{N}\right)$*, with* N *and* r *strictly positive constants, is an autonomous equation. Find all equilibrium solutions, draw a phase line, and label each equilibrium with its type. If* $x(0) > 0$*, what must happen to the population as* $t \to \infty$*?*

8. *Some populations are assumed to grow according to an equation of the form*

$$y' = ry(y - A)(B - y),$$

[3] F. Brauer and D. A. Sánchez, Constant rate population harvesting: equilibrium and stability, *Theor. Population Biology* **8** (1975), 12–30

where r, A, and B are all positive constants, and $B > A$. Find all equilibrium solutions, draw a phase line, and label each equilibrium with its type. Determine what happens to the population as $t \to \infty$ if

(a) $0 < y(0) < A$,
(b) $A < y(0) < B$,
(c) $y(0) > B$.

9. *In Section 1.3 the equation for the velocity of a free falling skydiver was assumed to be the autonomous equation*

$$v'(t) = g - \frac{k}{m}(v(t))^p,$$

where p, g, k, and m are all strictly positive constants.

(a) *Find all equilibrium solutions, draw a phase line, and label each equilibrium with its type.*
(b) *What value must $v(t)$ approach as $t \to \infty$? (Remember that this is what we called the* terminal *velocity.)*

10. *The RL-circuit equation $Li'(t) + Ri(t) = E(t)$ was described in Chapter 1, Section 3. The inductance L and resistance R are positive constants. If the voltage $E(t)$ is also a constant, this is an autonomous equation. Draw a phase line for this equation and determine the limiting value of the current $i(t)$ as $t \to \infty$.*

11. *A population of fish in a lake satisfies the growth equation:*

$$dx/dt = f(x, h) = 0.5x(4 - x) - h,$$

where $x(t)$ is thousands of fish in the lake at time t (in years), and h is thousands of fish harvested per year.

(a) *If the harvesting term h is zero, how many fish will the lake support (i.e., what is its carrying capacity)?*
(b) *If the harvesting term is constant at $h = 1$, find all equilibrium solutions and draw a phase line. Label each equilibrium as a sink, source, or node.*
(c) *If $h = 1$ and the initial condition is $x(0) = 0.5$, what happens to the solution as $t \to \infty$? Explain this in terms that a biologist might use.*
(d) *Sketch phase lines for $h = 0, 0.5, 1.0, 1.5, 2.0, 2.5$ and place them side by side as in Figure 2.23 to form a bifurcation diagram.*
(e) *What is the bifurcation point $h = h^*$ for this problem?*

STUDENT PROJECT: Single Neuron Equation. Biologists are currently applying mathematical modeling to a wide range of problems. One such problem involves the function of nerve cells (neurons) in the brain. These cells generate electrical impulses that move along their axons, affecting the activity of neighboring cells. Even for a single neuron, a complete model using what is currently known about its firing mechanism may take several equations. Since the human brain contains on the order of 10^{10} neurons, an exact model is not feasible. In this application you will be looking at a very simplified model (see Figure 2.26) for an isolated population of neurons all having similar properties.

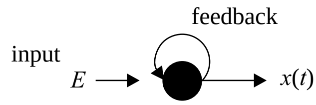


Figure 2.26. Input and output of the neuron population

Let $x(t)$ denote the percent of neurons firing at time t, normalized to be between 0 (low activity) and 1 (high activity). A simple model representing the change of activity level in the population is given by the differential equation

$$\frac{dx}{dt} = -x + S_a(x + E - \theta) \tag{2.49}$$

where E is the level of input activity coming from cells external to the population, θ is a common threshold level for cells in the set, and S_a is a response function that models the change in activity level due to a given input. We will use a standard "sigmoidal" response function

$$S_a(z) = \frac{1}{1 + e^{-az}}.$$

The nonlinear function S_a can be seen to increase monotonically from 0 to 1 as z increases from $-\infty$ to ∞. It is called a sigmoidal function because it has a sort of stylized S-shape. You may remember that solutions of the logistic growth equation had this same shape.

(1) Find a formula for the derivative of $S_a(z)$, and show that it satisfies the identity

$$S_a'(z) \equiv aS_a(z)(1 - S_a(z)).$$

(2) Draw a graph of $S_a(z)$ for $a = 3, 10$, and 20. Where is the slope a maximum? Is it the same in each case? Explain how the graph of $S_a(z - \theta)$ differs from the graph of $S_a(z)$.

The differential equation (2.49) for x is autonomous and therefore it can be analyzed by drawing a phase line. Assume $a = 10$, and the incoming activity E is constant at $E = 0.2$. Equation (2.49) now becomes

$$\frac{dx}{dt} = -x + \frac{1}{1 + e^{-10(x+0.2-\theta)}}. \tag{2.50}$$

Because the value of S_a is always between 0 and 1, if $x > 1$ the slope $x' = -x + S_a(x + 0.2 - \theta)$ is negative and if $x < 0$ it is positive. This means that any equilibrium solutions, that is, values of x where $\frac{dx}{dt} \equiv 0$, must lie between 0 and 1. It also implies that the arrows on the phase line are always pointing down above $x = 1$ and up below $x = 0$.

Figure 2.27 shows graphs of $y = x$ (the dashed line) and the response function $y = 1/(1 + e^{-10(x+0.2-\theta)})$ for $\theta = 0.4, 0.7$, and 1.0. It can be seen that for small θ there will be one equilibrium solution near $x = 1$ and for large θ there will be one equilibrium solution near $x = 0$. This seems reasonable since a high threshold means it takes a lot of input to produce a great amount of activity. For θ in a middle range, however, there can be three equilibrium solutions.

(3) Draw phase lines for equation (2.50) with $\theta = 0.4, 0.5, \ldots, 0.9, 1.0$. Label each equilibrium point as a sink, source, or node. You will need a numerical equation

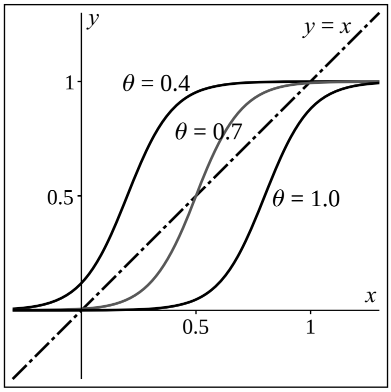


Figure 2.27. $y = 1/(1 + e^{-10(x+0.2-\theta)})$ for $\theta = 0.4, 0.7, 1.0$

solver to find the equilibrium values; that is, the values of x where $x = 1/(1 + e^{-10(x+0.2-\theta)})$. As a check, the three equilibria for $\theta = 0.7$ are $x_1 \approx 0.007188$, $x_2 = 0.5$, and $x_3 \approx 0.992812$.

(4) Make a bifurcation diagram, as in Figure 2.23, by putting the seven phase lines from (3) in a row and joining the equilibrium points.

(5) From the bifurcation diagram, estimate the two bifurcation values of θ where the number of equilibrium points changes from one to three, and then from three back to one.

(6) Find the two bifurcation values of θ analytically. You will need to solve two simultaneous equations obtained by using the fact that at a bifurcation value of θ the curves $y = 1/(1 + e^{-10(x+0.2-\theta)})$ and $y = x$ have a point of tangency. At this point the y-values of the curves are equal and their slopes are also equal.

(7) Use your computer algebra system to draw a slope field for (2.49) with $a = 10$, $E = 0.2$, and $\theta = 0.7$. Let t vary from 0 to 20. Use initial values $x(0) = 0.1, 0.3, 0.5, 0.7$, and 0.9, and describe what happens to the activity as $t \to \infty$.

(8) Redo problem (7) with periodic input $E = E(t) = 0.2(1 + \sin(t))$. With this time-varying input, the equation is no longer autonomous. Explain carefully how the activity differs from that described in (7). Do you think there is a periodic solution separating the two types of solutions in the periodic case? This is an interesting problem, and it might be a good time to look at the paper "Qualitative tools for studying periodic solutions and bifurcations as applied to the periodically harvested logistic equation", by Diego Benardete, V. W. Noonburg, and B. Pollina, *Amer. Math. Monthly*, vol. 115, 202–219 (2008). It discusses, in a very readable way, how one goes about determining the answer to such a question.