FIRST-ORDER ORDINARY DIFFERENTIAL EQUATIONS I:
Introduction and Analytic Methods

David Levermore
Department of Mathematics
University of Maryland

4 September 2012

Because the presentation of this material in lecture will differ from that in the book, I felt that notes that closely follow the lecture presentation might be appreciated.

Contents

1. Introduction: Classification and Overview

1.1. Classification.

We begin by giving the definition of a differential equation.

Definition 1.1. *A* differential equation *is an algebraic relation involving derivatives of one or more unknown functions with respect to one or more independent variables, and possibly either the unknown functions themselves or their independent variables, that holds at every point where all of the functions appearing in it are defined.*

For example, an unknown function $p(t)$ might satisfy the relation

$$\frac{\mathrm{d}p}{\mathrm{d}t} = 5p\,. \tag{1.1}$$

This is a differential equation because it algebraically relates the derivative of the unknown function p to itself. It does not involve the independent variable t. This relation should hold every point t where $p(t)$ and its derivative are defined.

Similarly, unknown functions $u(x, y)$ and $v(x, y)$ might satisfy the relation

$$\partial_x u + \partial_y v = 0\,, \tag{1.2}$$

where $\partial_x u$ and $\partial_y v$ denote partial derivatives. This is a differential equation because it algebraically relates some partial derivatives of the unknown functions u and v to each other. It does not involve the values of either u or v, nor does it involve either of the independent variables, x or y. This relation should hold every point (x, y) where $u(x, y)$, $v(x, y)$ and their partial derivatives appearing in (1.2) are defined.

Here are other examples of differential equations that involve derivatives of a single unkown function:

$$\begin{aligned}
&\text{(a)}\quad \frac{\mathrm{d}v}{\mathrm{d}t} = 9.8 - .05v^2\,, && \text{(b)}\quad \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \sin(\theta) = 0\,,\\
&\text{(c)}\quad \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + x^2 + y^2 = -1\,, && \text{(d)}\quad \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 + 4y^2 = 1\,,\\
&\text{(e)}\quad \frac{\mathrm{d}^2u}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}u}{\mathrm{d}r} + u = 0\,, && \text{(f)}\quad \frac{\mathrm{d}^3u}{\mathrm{d}x^3} + u\frac{\mathrm{d}u}{\mathrm{d}x} + \frac{\mathrm{d}u}{\mathrm{d}x} = 0\,,\\
&\text{(g)}\quad \frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 9x = \cos(2t)\,, && \text{(h)}\quad \frac{\mathrm{d}^2r}{\mathrm{d}t^2} = \frac{1}{r^{13}} - \frac{1}{r^7}\,,\\
&\text{(i)}\quad \frac{\mathrm{d}^2A}{\mathrm{d}x^2} + xA = 0\,, && \text{(j)}\quad W^3\frac{\mathrm{d}W}{\mathrm{d}z} + \tfrac{1}{6}z = 0\,,\\
&\text{(k)}\quad \partial_{tt}h = \partial_{xx}h + \partial_{yy}h\,, && \text{(l)}\quad \partial_{xx}\phi + \partial_{yy}\phi + \partial_{zz}\phi = 0\,,\\
&\text{(m)}\quad \partial_t u + u\partial_x u = \partial_{xx}u\,, && \text{(n)}\quad \partial_t T = \partial_x\big(T^4\partial_x T\big)\,.
\end{aligned} \tag{1.3}$$

In all of these examples except k and l the unknown function itself also appears in the equation. In examples d, e, g, i, and j the independent variable also appears in the equation.

Remark. All of the above examples except one arise in applications. That one is also the only example that has no real-valued solution. Can you spot this odd example?

The subject of differential equations is too vast to be studied in one course. In order to describe the kinds of differential equations that we will study in this course, we need to introduce some ways in which differential equations are classified.

Definition 1.2. *A differential equation is called an* ordinary differential equation (ODE) *if it invloves derivatives with respect to only one independent variable. Otherwise, it is called a* partial differential equation (PDE).

Example (1.1) is an ordinary differential equation. Example (1.2) is a partial differential equation. Of the examples in (1.3):

a – j	are ordinary differential equations ;
k – n	are partial differential equations .

Definition 1.3. *The* order *of a differential equation is the order of the highest derivative that appears in it. An* n^{th}-order differential equation *is one whose order is* n.

Examples (1.1) and (1.2) are first-order differential equations. Of the examples in (1.3):

a, c, d, j	are first-order differential equations ;
b, e, g, h, i, k, l, m, n	are second-order differential equations ;
f	is a third-order differential equation .

Definition 1.4. *A differential equation is said to be* linear *if it can be expressed so that each side of the equation is a sum of terms, each of which is either*

- *a derivative of an unknown function times a factor that is independent of the unknown functions,*
- *an unknown function times a factor that is independent of the unknown fuctions,*
- *or entirely independent of the unknown fuctions.*

Otherwise it is said to be nonlinear.

Examples (1.1) and (1.2) are linear differential equations. Of the examples in (1.3):

e, g, i, k, l	are linear differential equations ;
a – d, f, h, j, m, n	are nonlinear differential equations .

Remark. Linear differential equations are important because much more can be said about them than for general nonlinear differential equations.

Remark. Every n^{th}-order, linear ordinary differential equation for a single unknown function $y(t)$ can be brought into the form

$$p_0(t)\frac{\mathrm{d}^n y}{\mathrm{d}t^n} + p_1(t)\frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + p_{n-1}(t)\frac{\mathrm{d}y}{\mathrm{d}t} + p_n(t)y = r(t)\,, \tag{1.4}$$

where $p_0(t) \neq 0$ and $p_0(t)$, $p_1(t)$, $\cdots$, $p_n(t)$, and $r(t)$ are independant of the unknown function y. Notice that the left-hand side of this equation is a sum of terms, each of which is y or one of its derivatives times a factor $p_k(t)$ that may or may not depend on t, while the right-hand side is simply $r(t)$, which may or may not depend on t. Of the examples in (1.3), e, g, i are linear ordinary differential equations, all of which are second-order and in the form (1.4).

In applications one is often faced with a *system* of coupled differential equations — typically a system of m differential equations for m unknown functions. For example, two unknown functions $p(t)$ and $q(t)$ might satisfy the system

$$\frac{\mathrm{d}p}{\mathrm{d}t} = (6 - 2p - q)\,p\,, \qquad \frac{\mathrm{d}q}{\mathrm{d}t} = (8 - 4p - q)\,q\,. \tag{1.5}$$

Similarly, two unknown functions $u(x, y)$ and $v(x, y)$ might satisfy the system

$$\partial_x u = \partial_y v\,, \qquad \partial_y u + \partial_x v = 0\,. \tag{1.6}$$

The concepts of order, linear, and nonlinear extend to systems.

Definition 1.5. *The* order *of a system of differential equations is the order of the highest derivatrive that appears in the entire system. A system is called* linear *if every equation in it is linear. Otherwise, it is called* nonlinear*.*

Example (1.5) is a first-order, nonlinear system of ordinary differential equations, while (1.6) is a first-order, linear system of partial differential equations.

Remark. Systems that arise in applications can be extremely large. Systems of over 10^{10} ordinary differential equations are being solved numerically every day in order to approximate solutions of systems of partial differential equations.

1.2. **Course Overview.** Differential equations arise in mathematics, physics, chemistry, biology, medicine, pharmacology, communications, electronics, finance, economics, areospace, meteorology, climatology, oil recovery, hydrology, ecology, combustion, image processing, animation, and in many other fields. Partial differential equations are at the heart of most of these applications. You need to know something about ordinary differential equations before you study partial differential equations.

This course will serve as your introduction to ordinary differential equations. More specifically, we will study four classes of ordinary differential equations. We illustrate these four classes below denoting the independent variable by t.

I. We will begin with single first-order ODEs that can be brought into the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t, y)\,. \tag{1.7}$$

You may have seen some of this material in your calculus courses.

II. We will next study single n^{th}-order, linear ODEs that can be brought into the form

$$\frac{\mathrm{d}^n y}{\mathrm{d}t^n} + a_1(t)\frac{\mathrm{d}^{n-1} y}{\mathrm{d}t^{n-1}} + \cdots + a_{n-1}(t)\frac{\mathrm{d}y}{\mathrm{d}t} + a_n(t)y = f(t)\,. \tag{1.8}$$

This material is the heart of the course. We will see many new techniques. Many students find this the most difficult part of the course.

III. We will then turn towards systems of n first-order, linear ODEs that can brought into the form

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}t} &= a_{11}(t)y_1 + a_{12}(t)y_2 + \cdots + a_{1n}(t)y_n + f_1(t)\,,\\ \frac{\mathrm{d}y_2}{\mathrm{d}t} &= a_{21}(t)y_1 + a_{22}(t)y_2 + \cdots + a_{2n}(t)y_n + f_2(t)\,,\\ &\vdots\\ \frac{\mathrm{d}y_n}{\mathrm{d}t} &= a_{n1}(t)y_1 + a_{n2}(t)y_2 + \cdots + a_{nn}(t)y_n + f_n(t)\,.\end{aligned} \tag{1.9}$$

This material builds upon the material covered in part II.

IV. Finally, we will study systems of two first-order, nonlinear ODEs that can brought into the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,y)\,, \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = g(x,y)\,. \tag{1.10}$$

This material builds upon the material in parts I and III.

We will study each of these four classes from four perspectives: analytical methods, graphical methods, numerical methods, and in the setting of word problems. This is far from a complete treatment of the subject. It will however prepare you to learn more about ordinary differential equations or to learn about partial differential equations.

2. First-Order Equations: General and Explicit

2.1. **First-Order Equations.** We now begin our study of first-order ordinary differential equations that involve a single real-valued unknown function $y(t)$. These can always be brought into the form

$$F\left(t, y, \frac{\mathrm{d}y}{\mathrm{d}t}\right) = 0\,.$$

If we try to solve this equation for $\mathrm{d}y/\mathrm{d}t$ in terms of t and y then there might be no solutions or many solutions. For example, equation (c) of (1.3) clearly has no (real) solutions because the sum of nonnegative terms cannot add to -1. On the other hand, equation (d) will be satisfied if either

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \sqrt{1-4y^2}\,, \qquad \text{or} \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\sqrt{1-4y^2}\,.$$

To avoid such complications, we restrict ourselves to equations that are already in the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y)\,. \tag{2.1}$$

Examples (1.1) and (a) of (1.3) are already in this form. Example (j) of (1.3) can easily be brought into this form. And as we saw above, example (d) of (1.3) can be reduced to two equations in this form.

It is important to understand what is meant by a solution of (2.1),

Definition 2.1. *If $Y(t)$ is a function defined for every t in an interval (a,b) then we say $Y(t)$ is a* solution *of (2.1) over (a,b) if*

$$\begin{aligned} &\textit{(i)} \quad Y'(t) \textit{ is defined for every } t \textit{ in } (a,b)\,,\\ &\textit{(ii)} \quad f(t,Y(t)) \textit{ is defined for every } t \textit{ in } (a,b)\,,\\ &\textit{(iii)} \quad Y'(t) = f(t,Y(t)) \textit{ for every } t \textit{ in } (a,b)\,. \end{aligned} \tag{2.2}$$

Remark. One can recast condition (i) as "the function Y is differentiable over (a,b)." This definition is very natural in that it simply states that (i) the thing on left-hand side of the equation makes sense, (ii) the thing on right-hand side of the equation makes sense, and (iii) the two things are equal. This classical notion of solution will suit our needs now. Later we will see situations arise in which we will want to broaden this notion of solution.

Exercise. Consider the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{t}{y}\,.$$

Check that $y = \sqrt{4-t^2}$ is a solution of this equation over $(-2,2)$.

We want to address the following five basic questions about solutions.

- When does (2.1) have solutions?
- Under what conditions is a solution unique?
- How can we find analytic expressions for solutions?
- How can we visualize solutions?
- How can we approximate solutions?

We will focus on the last three questions. They address practical skills that can be applied when faced with a differential equation. The first two questions will be viewed through the lens of the last three. They are important because differential equations that arise in applications are supposed to model or predict something. If an equation either does not have solutions or has more than one solution then it fails to meet this objective. Moreover, in those situations the methods by which we will address the last three questions can give misleading results. Therefore we will study the first two questions with an eye towards avoiding such pitfalls. Rather than addressing these questions for a general $f(t, y)$ in (2.1), we will start by treating special forms $f(t, y)$ of increasing complexity.

2.2. **Review of Explicit Equations.** It is simplest to treat (2.1) when the deriviative is given as an explicit function of t. These so-called *explicit* equations have the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t) \,. \tag{2.3}$$

This case is covered in calculus courses, so we only review it here.

2.2.1. *Recipe for Explicit Equations.* You should recall from your study of calculus that a differentiable function F is said to be a *primitive* or *antiderivative* of f if $F' = f$. We thereby see that $y = Y(t)$ is a solution of (2.3) if and only if Y is a primitive of f. You should also recall from calculus that if you know one primitive F of f then any other primitive Y of f must have the form $Y(t) = F(t) + c$ for some constant c. We thereby see that if (2.3) has one solution then it has a family of solutions given by the indefinite integral of f — namely, by

$$y = \int f(t)\,\mathrm{d}t = F(t) + c \,, \quad \text{where } F'(t) = f(t) \text{ and } c \text{ is any constant} \,. \tag{2.4}$$

Moreover, there are no other solutions of (2.3). Therefore the family (2.4) is called a *general solution* of the differential equation (2.3).

2.2.2. *Initial-Value Problems for Explicit Equations.* In order to pick a unique solution from the family (2.4) one must impose an additional condition that determines c. We do this by imposing a so-called *initial condition* of the form

$$y(t_I) = y_I \,,$$

where t_I is called the *initial time* or *initial point*, while y_I is called the *initial value* or *initial datum*. The combination of equation (2.3) with the above initial condition is the so-called *initial-value problem* given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t) \,, \qquad y(t_I) = y_I \,. \tag{2.5}$$

If f has a primitive F then by (2.4) every solution of the differential equation in the initial-value problem (2.5) has the form $y = F(t) + c$ for some constant c. By imposing the initial condition in (2.5) upon this family we see that

$$F(t_I) + c = y_I \,,$$

which implies that $c = y_I - F(t_I)$. Therefore the unique solution of initial-value problem (2.5) is given by

$$y = y_I + F(t) - F(t_I) \,. \tag{2.6}$$

The above arguments show that the problems of finding either a general solution of (2.3) or the unique solution of the initial-value problem (2.5) reduce to the problem of finding a primitive F of f. Given such an F, a general solution of (2.3) is given by (2.4) while the unique solition of initial-value problem (2.5) is given by (2.6). These arguments however do not insure that such a primitive exists. Of course, for sufficiently simple f we can find a primitive analytically.

Example. Find a general solution to the differential equation

$$\frac{\mathrm{d}w}{\mathrm{d}x} = 6x^2 + 1\,.$$

Solution. By (2.4) a general solution is

$$w = \int \big(6x^2 + 1\big)\,\mathrm{d}x = 2x^3 + x + c\,.$$

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}w}{\mathrm{d}x} = 6x^2 + 1\,, \qquad w(1) = 5\,.$$

Solution. The previous example shows the solution has the form $w = 2x^3 + x + c$ for some constant c. Imposing the initial condition gives $2 \cdot 1^3 + 1 + c = 5$, which implies $c = 2$. Hence, the solution is $w = 2x^3 + x + 2$.

Alternative Solution. By (2.6) with $x_I = 1$, $w_I = 5$, and the primitive $F(x) = 2x^3 + x$ we find

$$\begin{aligned} w &= w_I + F(x) - F(x_I) = 5 + F(x) - F(1) \\ &= 5 + \big(2x^3 + x\big) - \big(2 \cdot 1^3 + 1\big) = 2x^3 + x + 2\,. \end{aligned}$$

Remark. As the solutions to the previous example illustrate, when solving an initial-value problem it is often easier to first find a general solution and then evaluate the c from the initial condition rather than to directly apply formula (2.6). With that approach you do not have to memorize formula (2.6).

2.2.3. *Theory for Explicit Equations.* Finally, even when we cannot find a primitive analytically, we can show that a solution exists by appealing to the Second Fundamental Theorem of Calculus. It states that if f is continuous over an interval (a, b) then for every t_I in (a, b) one has

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{t_I}^{t} f(s)\,\mathrm{d}s = f(t)\,.$$

In other words, f has a primitive over (a, b) that can be expressed as a definite integral. Here s is the "dummy" variable of integration in the above definite integral. If t_I is in (a, b) then the First Fundamental Theorem of Calculus implies that formula (2.6) can be expressed as

$$y = y_I + \int_{t_I}^{t} f(s)\,\mathrm{d}s\,. \tag{2.7}$$

This shows that if f is continuous over an interval (a, b) that contains t_I then the initial-value problem (2.5) has a unique solution over (a, b), which is given by formula (2.7). This formula can be approximated by numerical quadrature for any such f.

It is natural to ask if there is a largest time interval over which a solution exists.

Definition 2.2. *The largest time interval over which a solution (in the sense of Definition 2.1) exists is called its* interval of definition *or* interval of existence*.*

For explicit equations one can usually identify the interval of definition for the solution of the initial-value problem (2.5) by simply looking at $f(t)$. Specifically, if $Y(t)$ is the solution of the initial value problem (2.5) then its interval of definition will be (t_L, t_R) whenever:

- $f(t)$ is continuous over (t_L, t_R),
- the initial time t_I is in (t_L, t_R),
- $f(t)$ is not defined at both $t = t_L$ and $t = t_R$.

This is because the first two bullets along with the formula (2.7) imply that the interval of definition will be at least (t_L, t_R), while the last two bullets along with Definition 2.1 of solution imply that the interval of definition can be no bigger than (t_L, t_R). This argument works when $t_L = -\infty$ or $t_R = \infty$.

3. First-Order Equations: Linear

The simplest class of first-order equations to treat beyond the explicit ones is that of linear equations. It was remarked in (1.4) that every linear first-order ODE for a single unknown function $y(t)$ can be brought into the form

$$p(t)\frac{\mathrm{d}y}{\mathrm{d}t} + q(t)y = r(t)\,, \tag{3.1}$$

where $p(t)$, $q(t)$, and $r(t)$ are given functions of t such that $p(t) \neq 0$ for those t over which the equation is considered. The functions $p(t)$ and $q(t)$ are called *coefficients* while the function $r(t)$ is called the *forcing* or *driving*. Equation (3.1) is called *homogeneous* when the forcing is absent (i.e. when $r(t) = 0$ for all t), and is called *nonhomogeneous* otherwise.

3.1. Linear Normal Form.

Because $p(t) \neq 0$ for those t over which equation (3.1) is being considered, we can divide by $p(t)$ to bring equation (3.1) into its so-called *normal* or *standard* form

$$\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y = f(t)\,, \tag{3.2}$$

where

$$a(t) = \frac{q(t)}{p(t)}\,, \qquad f(t) = \frac{r(t)}{p(t)}\,.$$

Equation (3.2) becomes explicit if the coefficient is absent (i.e. if $a(t) = 0$ for all t). Equation (3.2) is called *homogeneous* if the forcing is absent (i.e. if $f(t) = 0$ for all t), and is called *nonhomogeneous* otherwise.

Remark. The linear normal form (3.2) can be put into the general form (2.1) by simply solving for the derivative of y as

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t) - a(t)y\,.$$

The derivative of the unknown function y thereby is given as a linear function of y whose coefficients are functions of t. We could have chosen this as the normal form for first-order linear equations, but picked (3.2) instead because it is the restriction to first-order of the normal form for higher-order linear equations that we will use later.

The normal form (3.2) will be the starting point for all of the methods and theory we will develop for linear equations. Therefore you should get into the habit of putting every linear equation into its normal form. Because you may not be given a linear equation in form (3.1), you can often discover it is linear by trying to put it into the normal form (3.2).

Example. Consider the equation

$$e^t\frac{\mathrm{d}z}{\mathrm{d}t} + t^2 = \frac{2t+z}{1+t^2}\,.$$

Show that this equation is linear and put it into the normal form (3.2).

Solution. By grouping all the terms involving either the z or its derivative on the left-hand side, while grouping all the other terms on the right-hand side, we obtain

$$e^t\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{1}{1+t^2}\,z = \frac{2t}{1+t^2} - t^2\,.$$

This can be transformed into the normal form (3.2) by multiplying both sides by e^{-t}.

3.2. **Recipe for Homogeneous Linear Equations.** When a linear equation is homogeneous, the problem of finding an analytic solution reduces to that of finding one primitive. You should first put the equation into its normal form (3.2), which becomes simply

$$\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y = 0\,. \tag{3.3}$$

We will show that a general solution of this equation is given by

$$y = e^{-A(t)}c\,, \qquad \text{where } A'(t) = a(t) \text{ and } c \text{ is any constant}\,. \tag{3.4}$$

Hence, for homogenous linear equations the recipe for solution only requires finding the primitive of $a(t)$. This means that for simple enough $a(t)$ you should be able to write down a general solution immediately. We illustrate this with a few examples.

Example. Find a general solution of the equation

$$\frac{\mathrm{d}p}{\mathrm{d}t} = 5p\,.$$

Solution. After the equation is put into normal form, we see that $a(t) = -5$. We can set $A(t) = -5t$. A general solution is then

$$p = e^{5t}c\,, \qquad \text{where } c \text{ is an arbitrary constant}\,.$$

Example. Find a general solution of the equation

$$\frac{\mathrm{d}w}{\mathrm{d}t} + t^2 w = 0\,.$$

Solution. The equation is already in normal form. Because $a(t) = t^2$, we can set $A(t) = \frac{1}{3}t^3$. A general solution is then

$$w = e^{-\frac{1}{3}t^3}c\,, \qquad \text{where } c \text{ is an arbitrary constant}\,.$$

Example. Find a general solution of the equation

$$(1+t^2)\frac{\mathrm{d}z}{\mathrm{d}t} + 4tz = 0\,.$$

Solution. First put the equation into its normal form

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{4t}{1+t^2}\,z = 0\,.$$

Because $a(t) = (4t)/(1+t^2)$, we can set $A(t) = 2\log(1+t^2)$. A general solution is then

$$z = e^{-2\log(1+t^2)}c = \frac{c}{(1+t^2)^2}\,, \qquad \text{where } c \text{ is an arbitrary constant}\,.$$

We will see that recipe (3.4) is a special case of the recipe for solutions of nonhomogeneous linear equations. Therefore we will justify it when we justify that more general recipe.

3.3. **Recipe for Nonhomogeneous Linear Equations.** When a linear equation is nonhomogeneous, the problem of finding an analytic solution reduces to that of finding two primitives. We begin by putting the equation into the normal form (3.2), which is

$$\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y = f(t)\,. \tag{3.5}$$

Below we will show that this is equivalent to the so-called *integrating factor form*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(e^{A(t)}y\Big) = e^{A(t)}f(t)\,, \qquad \text{where } A'(t) = a(t)\,. \tag{3.6}$$

This is an explicit equation for the derivative of $e^{A(t)}y$ that can be integrated to obtain

$$e^{A(t)}y = \int e^{A(t)}f(t)\,\mathrm{d}t = B(t)+c\,, \qquad \text{where } B'(t) = e^{A(t)}f(t) \text{ and } c \text{ is any constant}\,. \tag{3.7}$$

Therefore a general solution of (3.5) is given by the family

$$y = e^{-A(t)}B(t) + e^{-A(t)}c\,. \tag{3.8}$$

Notice that because $B'(t) = e^{A(t)}f(t)$, for a homogeneous equation we can set $B(t) = 0$. In that case this recipe recovers the recipe for homogeneous linear equations given by (3.4).

The key to understanding recipe (3.8) is to understand the equivalence of the normal form (3.5) and integrating factor form (3.6). This equivalence follows from the fact that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(e^{A(t)}y\Big) = e^{A(t)}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}t}\Big(e^{A(t)}\Big)y = e^{A(t)}\frac{\mathrm{d}y}{\mathrm{d}t} + e^{A(t)}A'(t)y = e^{A(t)}\left(\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y\right).$$

This calculation shows that equation (3.6) is simply equation (3.5) multiplied by $e^{A(t)}$. Because the factor $e^{A(t)}$ is always positive, the equations are equivalent. We call $e^{A(t)}$ an *integrating factor* of equation (3.5) because after multiplying both sides of (3.5) by $e^{A(t)}$ the left-hand side can be written as the derivative of $e^{A(t)}y$. An integrating factor thereby allows us to reduce the linear equation (3.5) to the explicit equation (3.6).

Remark. The integrating factor $e^{A(t)}$ is just the reciprocal of a solution given by (3.4) to the associated homogeneous problem. The general solution (3.8) thereby has the form

$$y = Y_P(t) + Y_H(t)\,,$$

where $Y_P(t) = e^{-A(t)}B(t)$ is the particular solution of (3.5) obtained by setting $c = 0$ in (3.8) while $Y_H(t) = e^{-A(t)}c$ is the general solution given by (3.4) of the associated homogeneous problem. General solutions of higher-order nonhomogeneous linear equations share this structure.

Rather than using formula (3.8), you should find general solutions of first-order linear ordinary differential equations by simply retracing the steps by which (3.8) was derived. We illustrate this approach with the following examples.

Example. Find a general solution to

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -3x + e^{2t}\,.$$

Solution. First bring the equation into the normal form

$$\frac{\mathrm{d}x}{\mathrm{d}t} + 3x = e^{2t}\,.$$

An integrating factor is $e^{A(t)}$ where $A'(t) = 3$. By setting $A(t) = 3t$, we then bring the equation into the integrating factor form

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(e^{3t}x\Big) = e^{3t}e^{2t} = e^{5t}\,.$$

By integrating both sides of this equation we obtain

$$e^{3t}x = \int e^{5t}\,\mathrm{d}t = \tfrac{1}{5}e^{5t} + c\,.$$

Therefore a general solution is given by

$$x = \tfrac{1}{5}e^{2t} + e^{-3t}c\,.$$

Example. Find a general solution to

$$(1+t^2)\frac{\mathrm{d}z}{\mathrm{d}t} + 4tz = \frac{1}{\big(1+t^2\big)^2}\,.$$

Solution. First bring the equation into the normal form

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{4t}{1+t^2}z = \frac{1}{\big(1+t^2\big)^3}\,.$$

An integrating factor is $e^{A(t)}$ where $A'(t) = 4t/(1+t^2)$. By setting $A(t) = 2\log(1+t^2)$, we see that

$$e^{A(t)} = e^{2\log(1+t^2)} = \Big(e^{\log(1+t^2)}\Big)^2 = \big(1+t^2\big)^2\,.$$

We then bring the differential equation into the integrating factor form

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big((1+t^2)^2 z\Big) = \big(1+t^2\big)^2\frac{1}{\big(1+t^2\big)^3} = \frac{1}{1+t^2}\,.$$

By integrating both sides of this equation we obtain

$$(1+t^2)^2 z = \int \frac{1}{1+t^2}\,\mathrm{d}t = \tan^{-1}(t) + c\,.$$

Therefore a general solution is given by

$$z = \frac{\tan^{-1}(t)}{\big(1+t^2\big)^2} + \frac{c}{\big(1+t^2\big)^2}\,.$$

3.4. **Linear Initial-Value Problems.** In order to pick a unique solution from the family (3.8) one must impose an additional condition that determines c. We do this by again imposing an *initial condition* of the form

$$y(t_I) = y_I\,,$$

where t_I is called the *initial time* or *initial point* while y_I is called the *initial value* or *initial datum*. The combination of the differential equation (3.2) with this initial condition is

$$\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y = f(t)\,, \qquad y(t_I) = y_I\,. \tag{3.9}$$

This is a so-called *initial-value problem.* By imposing the initial condition upon the family (3.8) we see that

$$y(t_I) = e^{-A(t_I)}B(t_I) + e^{-A(t_I)}c = y_I\,,$$

whereby $c = e^{A(t_I)}y_I - B(t_I)$. Therefore if the primitives $A(t)$ and $B(t)$ exist then the unique solution of initial-value problem (3.9) is given by

$$y = e^{-A(t)+A(t_I)}y_I + e^{-A(t)}\big(B(t) - B(t_I)\big)\,. \tag{3.10}$$

Rather than using formula (3.10), you should solve initial-value problems by simply retracing the steps by which it was derived — namely, by first finding a general solution as we did in the previous section, and then by imposing the initial condition to evaluate c. Formula (3.10) shows us that this approach will always yield the solution. We illustrate this with the following examples.

Example. Solve the initial-value problem

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -3x + e^{2t}\,, \qquad x(0) = 2\,.$$

Solution. We showed above that a general solution of the differential equation is

$$x = \tfrac{1}{5}e^{2t} + e^{-3t}c\,.$$

By imposing the initial condition we find that

$$x(0) = \tfrac{1}{5}e^{0} + e^{0}c = \tfrac{1}{5} + c = 2\,,$$

whereby $c = \frac{9}{5}$. Therefore the solution of the initial-value problem is

$$x = \tfrac{1}{5}e^{2t} + \tfrac{9}{5}e^{-3t}\,.$$

Example. Solve the initial-value problem

$$(1+t^2)\frac{\mathrm{d}z}{\mathrm{d}t} + 4tz = \frac{1}{\big(1+t^2\big)^2}\,, \qquad z(1) = \pi\,.$$

Solution. We showed above that a general solution of the differential equation is

$$z = \frac{\tan^{-1}(t)}{\big(1+t^2\big)^2} + \frac{c}{\big(1+t^2\big)^2}\,.$$

By imposing the initial condition we find that

$$z(1) = \frac{\tan^{-1}(1)}{\big(1+1^2\big)^2} + \frac{c}{\big(1+1^2\big)^2} = \frac{\frac{\pi}{4}}{2^2} + \frac{c}{2^2} = \frac{\pi}{16} + \frac{c}{4} = \pi\,,$$

whereby $c = \frac{15}{4}\pi$. Therefore the solution of the initial-value problem is

$$z = \frac{\tan^{-1}(t)}{\big(1+t^2\big)^2} + \frac{\frac{15}{4}\pi}{\big(1+t^2\big)^2}\,.$$

3.5. **Theory for Linear Equations.** Even when we cannot find primitives $A(t)$ and $B(t)$ analytically, we can show that a solution exists whenever $a(t)$ and $f(t)$ are continuous over an interval (t_L, t_R) that contains the initial time t_I. In that case we can appeal to the Second Fundamental Theorem of Calculus to express $A(t)$ and $B(t)$ as the definite integrals

$$A(t) = \int_{t_I}^{t} a(r)\,\mathrm{d}r\,, \qquad B(t) = \int_{t_I}^{t} e^{A(s)} f(s)\,\mathrm{d}s\,.$$

For this choice of $A(t)$ and $B(t)$ we have $A(t_I) = B(t_I) = 0$, whereby formula (3.10) becomes

$$y = e^{-A(t)} y_I + e^{-A(t)} B(t) = e^{-A(t)} y_I + \int_{t_I}^{t} e^{-A(t)+A(s)} f(s)\,\mathrm{d}s\,. \tag{3.11}$$

The First Fundamental Theorem of Calculus implies that

$$A(t) - A(s) = \int_{s}^{t} a(r)\,\mathrm{d}r\,,$$

whereby formula (3.11) can be expressed as

$$y = \exp\!\left(-\int_{t_I}^{t} a(r)\,\mathrm{d}r\right) y_I + \int_{t_I}^{t} \exp\!\left(-\int_{s}^{t} a(r)\,\mathrm{d}r\right) f(s)\,\mathrm{d}s\,. \tag{3.12}$$

This shows that if a and f are continuous over an interval (t_L, t_R) that contains t_I then the initial-value problem (3.9) has a unique solution over (t_L, t_R) given by formula (3.12). We state this result as a theorem.

Theorem 3.1. *Let $a(t)$ and $f(t)$ be functions defined over the open interval (t_L, t_R) that are also continuous over (t_L, t_R).*

Then for every inital time t_I in (t_L, t_R), and every initial value y_I there exists a unique solution $y = Y(t)$ to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} + a(t)y = f(t)\,, \qquad y(t_I) = y_I\,, \tag{3.13}$$

that is defined over (t_L, t_R).

Moreover, this solution is continuously differentiable and is determined by formula (3.12).

For linear equations one can usually identify the interval of definition for the solution of the initial-value problem (3.9) by simply looking at $a(t)$ and $f(t)$. Specifically, if $Y(t)$ is the solution of the initial value problem (3.9) then its interval of definition will be (t_L, t_R) whenever:

- the coefficient $a(t)$ and forcing $f(t)$ are continuous over (t_L, t_R),
- the initial time t_I is in (t_L, t_R),
- either the coefficient $a(t)$ or the forcing $f(t)$ is not defined at both $t = t_L$ and $t = t_R$.

This is because the first two bullets along with the formula (3.12) imply that the interval of definition will be at least (t_L, t_R), while the last two bullets along with Definition 2.1 of solution imply that the interval of definition can be no bigger than (t_L, t_R) because the equation is not defined at both $t = t_L$ and $t = t_R$. This argument can be applied when either $t_L = -\infty$ or $t_R = \infty$.

Example. Give the interval of definition for the solution of the initial-value problem

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \cot(t)\, z = \frac{1}{\log(t^2)}\,, \qquad z(4) = 3\,.$$

Solution. The coefficient $\cot(t)$ is not defined at $t = n\pi$ where n is any integer, and is continuous everywhere else. The forcing $1/\log(t^2)$ is not defined at $t = 0$ and $t = 1$, and is continuous everywhere else. Therefore the interval of definition is $(\pi, 2\pi)$ because: both $\cot(t)$ and $1/\log(t^2)$ are continuous over this interval; the initial time is $t = 4$, which is in this interval; $\cot(t)$ is not defined at $t = \pi$ and $t = 2\pi$.

Example. Give the interval of definition for the solution of the initial-value problem

$$\frac{\mathrm{d}z}{\mathrm{d}t} + \cot(t)\, z = \frac{1}{\log(t^2)}\,, \qquad z(2) = 3\,.$$

Solution. The interval of definition is $(1, \pi)$ because: both $\cot(t)$ and $1/\log(t^2)$ are continuous over this interval; the initial time is $t = 2$, which is in this interval; $\cot(t)$ is not defined at $t = \pi$ while $1/\log(t^2)$ is not defined at $t = 1$.

Remark. If $y = Y(t)$ is a solution of (3.8) whose interval of definition is (t_L, t_R) then this does not mean that $Y(t)$ will become undefined at either $t = t_L$ or $t = t_R$ when those endpoints are finite. For example, $y = t^4$ solves the initial-value problem

$$t\,\frac{\mathrm{d}y}{\mathrm{d}t} - 4y = 0\,, \qquad y(1) = 1\,,$$

and is defined for every t. However, the interval of definition is just $(0, \infty)$ because the initial time is $t = 1$ and normal form of the equation is

$$\frac{\mathrm{d}y}{\mathrm{d}t} - \frac{4}{t}\,y = 0\,,$$

the coefficient of which is undefined at $t = 0$.

Remark. It is natural to ask why we do not extend our definition of solutions so that $y = t^4$ is considered a solution of the initial-value problem

$$t\,\frac{\mathrm{d}y}{\mathrm{d}t} - 4y = 0\,, \qquad y(1) = 1\,,$$

for every t. For example, we might say that $y = Y(t)$ is a solution provided it is differentiable and satisfies the above equation rather than its normal form. However by this definition the function

$$Y(t) = \begin{cases} t^4 & \text{for } t \geq 0 \\ ct^4 & \text{for } t < 0 \end{cases}$$

also solves the initial-value problem for any c. This shows that because the equation breaks down at $t = 0$, there are many ways to extend the solution $y = t^4$ to $t < 0$. We avoid such complications by requiring the normal form of the equation to be defined.

4. First-Order Equations: Separable

The next simplest class of first-order equations to treat after linear ones is that of so-called separable equations. These have the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)\, g(y)\,. \tag{4.1}$$

Here the derivative of y with respect to t is given as a function of t times a function of y. Equation (4.1) becomes explicit when $g(y)$ is a constant, so we will exclude that case.

4.1. **Recipe for Autonomous Equations.**

When $f(t)$ is a constant equation (4.1) is said to be autonomous. In that case $f(t)$ can be absorbed into $g(y)$ and (4.1) becomes

$$\frac{\mathrm{d}y}{\mathrm{d}t} = g(y)\,. \tag{4.2}$$

The word "autonomous" has Greek roots and means "self governing". Equation (4.2) is called autonomous because it depends only on y. Autonomous equations arise naturally in applications because often the laws governing dynamics are the same at every time.

If $g(y_o) = 0$ at some point y_o then it is clear that $y(t) = y_o$ is a solution of (4.2) that is defined for every t. Because this solution does not depend on t it is called a *stationary solution*. Every zero of g is also called a *stationary point* because it yields such a stationary solution. These points are also sometimes called either *equilibrium points* or *critical points.*

Example. Consider the autonomous equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 4y - y^3\,.$$

Because $4y - y^3 = y(2-y)(2+y)$, we see that $y = 0$, $y - 2$, and $y = -2$ are stationary points of this equation.

Now let us consider how to find the nonstationary solutions of (4.2). Just as we did for the linear case, we will reduce the autonomous case to the explicit case. The trick to doing this is to consider t to be a function of y. This trick works over intervals over which the solution $y(t)$ is a strictly monotonic function of t. This will be the case over intervals where $g(y(t))$ is never zero — i.e. over intervals where the solution $y(t)$ does not hit a stationary point. In that case, by the chain rule we have

$$\frac{\mathrm{d}t}{\mathrm{d}y} = \frac{1}{\dfrac{\mathrm{d}y}{\mathrm{d}t}} = \frac{1}{g(y)}\,.$$

This is an explicit equation for the derivative of t with respect to y. It can be integrated to obtain

$$t = \int \frac{\mathrm{d}y}{g(y)} = G(y) + c\,, \qquad \text{where } G'(y) = \frac{1}{g(y)} \text{ and } c \text{ is any constant}\,. \tag{4.3}$$

Equation (4.3) is called an *implicit solution* of equation (4.2) because if we solve equation (4.3) for y as a differentiable function of t then the result will be a solution of (4.2) wherever $g(y) \neq 0$. Indeed, suppose that $Y(t)$ is differentiable and satisfies

$$t = G\big(Y(t)\big) + c\,.$$

Upon differentiating both sides of this equation with respect to t we see that

$$1 = G'\big(Y(t)\big)\,\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \frac{1}{g\big(Y(t)\big)}\,\frac{\mathrm{d}Y(t)}{\mathrm{d}t}\,, \qquad \text{wherever } g\big(Y(t)\big) \neq 0\,.$$

It follows that $y = Y(t)$ satifies (4.2).

Being able to solve (4.3) for y means finding an inverse function of G — namely, a function G^{-1} with property that

$$G^{-1}\big(G(y)\big) = y \quad \text{for every } y \text{ in some interval within the domain of } G\,.$$

For every such an inverse function, a family of explicit solutions to (4.2) is then given by

$$y = G^{-1}\big(t - c\big)\,. \tag{4.4}$$

As we will see in examples below, such a solution may not exist for every value of $t - c$, or there may be more than one solution.

Remark. This recipe will fail to yield a family of explicit solutions to (4.2) if either we are unable to find an expression for the primitive $G(y)$ in order to obtain (4.3), or we are unable find an explicit inverse function of $G(y)$ in order to obtain (4.4) from (4.3).

Example. Find all solutions to the equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2\,.$$

Solution. This equation has the stationary solution $y = 0$. Nonstationary solutions are given implicitly by

$$t = \int \frac{\mathrm{d}y}{y^2} = -\frac{1}{y} + c\,.$$

We can solve for y explicitly to find the family of solutions

$$y = \frac{1}{c - t}\,.$$

Notice that this solution is not defined at $t = c$. Therefore it is really two solutions — one for $t < c$ and one for $t > c$. Because

$$\lim_{t \to c^-} \frac{1}{c - t} = +\infty\,, \qquad \lim_{t \to c^+} \frac{1}{c - t} = -\infty\,,$$

we say that the solution for $t < c$ "blows-up" to ∞ as $t \to c^-$ while the solution for $t > c$ "blows-up" to $-\infty$ as $t \to c^+$. (Here $t \to c^-$ means that t approaches c from below while $t \to c^+$ means that t approaches c from above.) Also notice that none of these solutions ever hits the stationary solution $y = 0$.

Remark. The above example already shows two important differences between nonlinear and linear equations. For one, it shows that solutions of nonlinear equations can "blow-up" even when the right-hand side of the equation is continuous everywhere. This is in marked contrast with linear equations where the solution will not blow-up if the coefficient and forcing are continuous everywhere. Moreover, it shows that for solutions of nonlinear equations the interval of definition cannot be read off from the equation. This also is in marked contrast with linear equations where we can read off the interval of definition from the coefficient and the forcing.

Example. Find all solutions to the equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2y}\,.$$

Solution. This equation has no stationary solutions. Nonstationary solutions are given implicitly by

$$t = \int 2y\,\mathrm{d}y = y^2 + c\,.$$

We can solve for y explicitly to find two families of solutions

$$y = \pm\sqrt{t-c}\,.$$

Notice that these functions are not defined for $t < c$ and are not differentiable at $t = c$. Therefore they are solutions only when $t > c$. One family gives positive solutions while the other gives negative solutions. It is clear from the right-hand side of the equation that no solution can take the value 0. The derivatives of these solutions are given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2y} = \pm\frac{1}{2\sqrt{t-c}}\,.$$

Notice that these derivatives "blow-up" as $t \to c^+$ while the solutions themselves approach the forbidden value 0.

Remark. The above example shows two more important differences between nonlinear and linear equations. For one, it shows that sometimes we need more than one family of nonstationary solutions to give a general solution. This is in marked contrast with linear equations where a general solution is always given by the single family (3.8). It also shows that the derivative of a solution can "blow-up" at an endpoint of its interval of definition even though the solution does not. The phenomenon does not happen for linear equations.

The next example shows that finding explicit nonstationary solutions by recipe (4.4) can get complicated even when $g(y)$ is a fairly simple polynomial. In fact, it will show that even the simpler task of finding the implicit relation (4.3) satisfied by nonstationary solutions sometimes requires you to recall basic techniques of integration.

Example. Find a general solution of

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 4y - y^3\,.$$

Solution. Because $4y - y^3 = y(2-y)(2+y)$, we see that $y = 0$, $y - 2$, and $y = -2$ are stationary solutions of this equation. Nonstationary solutions are given implicitly by

$$t = \int \frac{\mathrm{d}y}{4y - y^3}\,.$$

A partial fraction decomposition yields the identity

$$\frac{1}{4y-y^3} = \frac{1}{y(2-y)(2+y)} = \frac{\frac{1}{4}}{y} + \frac{\frac{1}{8}}{2-y} - \frac{\frac{1}{8}}{2+y}\,.$$

(You should be able to write down such partial fraction identities directly.) Nonstationary solutions therefore are given implicitly by

$$\begin{aligned} t &= \int \frac{\frac{1}{4}}{y} + \frac{\frac{1}{8}}{2-y} - \frac{\frac{1}{8}}{2+y}\,\mathrm{d}y = \tfrac{1}{4}\log(|y|) - \tfrac{1}{8}\log(|2-y|) - \tfrac{1}{8}\log(|2+y|) + c \\ &= \tfrac{1}{8}\log(y^2) - \tfrac{1}{8}\log(|2-y|) - \tfrac{1}{8}\log(|2+y|) + c = \tfrac{1}{8}\log\left(\frac{y^2}{|4-y^2|}\right) + c \end{aligned}$$

where c is an arbitrary constant. Upon solving this for y^2 we find

$$8(t-c) = \log\left(\frac{y^2}{|4-y^2|}\right),$$

which leads to

$$\frac{y^2}{|4-y^2|} = e^{8(t-c)}.$$

This can then be broken down into two cases.

First, if $y^2 < 4$ then

$$\frac{y^2}{4-y^2} = e^{8(t-c)},$$

which implies that

$$y^2 = \frac{4}{1+e^{-8(t-c)}},$$

and finally that

$$y = \pm\sqrt{\frac{4}{1+e^{-8(t-c)}}}.$$

These solutions exist for every time t. They vanish as $t \to -\infty$ and approach ± 2 as $t \to \infty$.

On the other hand, if $y^2 > 4$ then

$$\frac{y^2}{y^2-4} = e^{8(t-c)},$$

which implies that

$$y^2 = \frac{4}{1-e^{-8(t-c)}}.$$

So long as the denominator is positive, we find the solutions

$$y = \pm\sqrt{\frac{4}{1-e^{-8(t-c)}}}.$$

The denominator is positive if and only if $t > c$. Therefore these solutions exist for every time $t > c$. They diverge (blow-up) to $\pm\infty$ as $t \to c^+$ and approach ± 2 as $t \to \infty$.

Therefore we have found the three stationary solutions $y = 0$, $y = 2$, and $y = -2$, and the four families of nonstationary solutions

$$
\begin{aligned}
y &= -\sqrt{\frac{4}{1 - e^{-8(t-c)}}} && \text{when } -\infty < y < -2 \text{ and } t > c\,,\\
y &= -\sqrt{\frac{4}{1 + e^{-8(t-c)}}} && \text{when } -2 < y < 0\,,\\
y &= +\sqrt{\frac{4}{1 + e^{-8(t-c)}}} && \text{when } 0 < y < 2\,,\\
y &= +\sqrt{\frac{4}{1 - e^{-8(t-c)}}} && \text{when } 2 < y < \infty \text{ and } t > c\,,
\end{aligned}
$$

where c is an arbitrary constant. Notice that none of these nonstationary solutions ever hits one of the stationary solutions. This list includes every solution of the equation. □

4.2. **Recipe for Separable Equations.** Now let us consider the general separable equation (4.1), which has the form

$$
\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)\,g(y)\,. \tag{4.5}
$$

The method for solving separable equations is a slight modification of the one for solving autonomous equations. Indeed, the difficulties that arise are the same ones that arose in the autonomous case. The difference will be that the recipe for implicitly solving separable equations will require us to find two primitives, while recipe (4.3) for implicitly solving autonomous equations requires us to find only one primitive.

If $g(y_o) = 0$ at some point y_o then it is clear that $y(t) = y_o$ is a solution of (4.5) that is defined for every t. As before, this solution is called a *stationary solution*. Every zero of g is called a *stationary point* because it yields such a stationary solution.

To find the nonstationary solutions of equation (4.5) we first put the equation into its so-called *separated differential form*

$$
\frac{1}{g(y)}\,\mathrm{d}y = f(t)\,\mathrm{d}t\,.
$$

Then integrate both sides of this equation to obtain

$$
\int \frac{1}{g(y)}\,\mathrm{d}y = \int f(t)\,\mathrm{d}t\,.
$$

This is equivalent to

$$
F(t) = G(y) + c\,, \tag{4.6}
$$

where F and G satisfy

$$
F'(t) = f(t)\,, \quad G'(y) = \frac{1}{g(y)}\,, \quad \text{and } c \text{ is any constant}\,.
$$

Equation (4.6) is called an *implicit* solution of (4.5).

We call (4.6) an implicit solution of (4.5) because if we solve equation (4.6) for y as a differentiable function of t then the result will be a solution of (4.5) whenever $g(y) \neq 0$. Indeed, suppose that $Y(t)$ is differentiable and satisfies

$$F(t) = G\big(Y(t)\big) + c\,.$$

Upon differentiating both sides of this equation with respect to t we see that

$$f(t) = G'\big(Y(t)\big)\,\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = \frac{1}{g\big(Y(t)\big)}\,\frac{\mathrm{d}Y(t)}{\mathrm{d}t}\,, \qquad \text{wherever } g\big(Y(t)\big) \neq 0\,.$$

It follows that $y = Y(t)$ satifies (4.5).

Being able to solve (4.6) for y means finding an inverse function of G — namely, a function G^{-1} with property that

$$G^{-1}\big(G(y)\big) = y \quad \text{for every } y \text{ in some interval within the domain of } G\,.$$

For every such an inverse function, a family of explicit solutions to (4.5) is then given by

$$y = G^{-1}\big(F(t) - c\big)\,. \tag{4.7}$$

As we will see in examples below, such a solution may not exist for every value of $F(t) - c$, or there may be more than one solution.

Remark. This recipe will fail to yield a family of explicit solutions to (4.5) if either we are unable to find expressions for the primitives $F(t)$ and $G(y)$ in order to obtain (4.6), or we are unable find an explicit inverse function of $G(y)$ in order obtain (4.7) from (4.6).

Example. Give a general solution of

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{y}\,.$$

Solution. This equation is separable. Its right-hand side is undefined when $y = 0$. It has no stationary solutions. Its separated differential form is

$$2y\,\mathrm{d}y = -2x\,\mathrm{d}x\,.$$

(The factors of 2 make our primitives nicer.) Its solutions are given implicitly by

$$y^2 = -x^2 + c\,, \qquad \text{for any constant } c\,.$$

This can be solved explicitly to obain the families

$$y = \pm\sqrt{c - x^2}\,, \qquad \text{for any constant } c\,.$$

These are solutions provided $c - x^2 > 0$, which implies that $c > 0$. For every $c > 0$ each of these families gives a solution whose interval of definition is $(-\sqrt{c}, \sqrt{c})$.

Example. Find all solutions of

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{3x + xz^2}{z + x^2 z}\,.$$

Solution. This equation is separable. Its right-hand side is undefined when $z = 0$. It has no stationary solutions. Its separated differential form is

$$\frac{2z}{3 + z^2}\,\mathrm{d}z = \frac{2x}{1 + x^2}\,\mathrm{d}x\,.$$

(The factors of 2 make our primitives nicer.) Then because

$$F(x) = \int \frac{2x}{1+x^2}\,\mathrm{d}x = \log(1+x^2) + c_1\,, \quad G(z) = \int \frac{2z}{3+z^2}\,\mathrm{d}z = \log(3+z^2) + c_2\,,$$

its solutions are given implicitly by

$$\log(3+z^2) = \log(1+x^2) - c\,, \qquad \text{for any constant } c\,.$$

By exponentiating both sides of this equation we obtain

$$3 + z^2 = (1+x^2)e^{-c}\,,$$

which yields the families

$$z = \pm\sqrt{(1+x^2)e^{-c} - 3}\,, \qquad \text{for any constant } c\,.$$

These are solutions provided $(1+x^2)e^{-c} - 3 > 0$, or equivalently, provided

$$x^2 > 3e^c - 1\,.$$

For $c < -\log(3)$ each of the above families gives a single solution whose interval of definition is $(-\infty, \infty)$. For $c \geq -\log(3)$ each of the families gives two solutions, one whose interval of definition is $(-\infty, -\sqrt{3e^c - 1})$, and one whose interval of definition is $(\sqrt{3e^c - 1}, \infty)$.

4.3. **Separable Initial-Value Problems.** In order to pick a unique solution from among all the solutions we impose an additional condition. As for linear equations, we do this by imposing an *initial condition* of the form $y(t_I) = y_I$, where t_I is called the *initial time* while y_I is called the *initial value*. Differential equation (4.5) combined with this initial condition is the *initial-value problem*

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)\,g(y)\,, \qquad y(t_I) = y_I\,. \tag{4.8}$$

There are two possibilities: either $g(y_I) = 0$ or $g(y_I) \neq 0$. If $g(y_I) = 0$ then it is clear that a solution of (4.8) is the stationary solution $y(t) = y_I$, which is defined for every t. On the other hand, if $g(y_I) \neq 0$ then we can try to use recipe (4.6) to obtain an implicit relation $F(t) = G(y) + c$. The initial condition then implies that

$$F(t_I) = G(y_I) + c\,,$$

whereby $c = F(t_I) - G(y_I)$. The solution of the initial-value problem (4.8) thereby satisfies

$$F(t) = G(y) + F(t_I) - G(y_I)\,.$$

To find the explicit solution, we must solve this equation for y as a function of t. There may be more than one solution of this equation. If so, be sure to take the one that satisfies the initial condition. This means we have to find the inverse function of G that recovers y_I — namely a function G^{-1} with the property that

$$G^{-1}\big(G(y)\big) = y \quad \text{for every } y \text{ in an interval within the domain of } G \text{ that contains } y_I\,.$$

In particular, G^{-1} must satisfy $G^{-1}\big(G(y_I)\big) = y_I$. There is a unique inverse function with this property. The solution of the initial-value problem (4.8) is then given by

$$y = G^{-1}\big(G(y_I) + F(t) - F(t_I)\big)\,.$$

This will be valid for all times t in some open interval that contains the initial time t_I. The largest such interval is the interval of definition for the solution.

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y^2\,, \qquad y(0) = y_I\,.$$

Identify its interval of definition.

Solution. We see that $y = 0$ is the only stationary point of this equation. Hence, if $y_I = 0$ then $y(t) = 0$ is a stationary solution whose interval of definition is $(-\infty, \infty)$. So let us suppose that $y_I \neq 0$. The solution is given implicitly by

$$t = \int \frac{\mathrm{d}y}{y^2} = -\frac{1}{y} + c\,.$$

The initial condition then implies that

$$0 = -\frac{1}{y_I} + c\,,$$

whereby $c = 1/y_I$. The solution therefore is given implicitly by

$$t = -\frac{1}{y} + \frac{1}{y_I}\,.$$

This may be solved for y explicitly by first noticing that

$$\frac{1}{y} = \frac{1}{y_I} - t = \frac{1 - y_I t}{y_I}\,,$$

and then taking the reciporical of each side to obtain

$$y = \frac{y_I}{1 - y_I t}\,.$$

It is clear from this formula that the solution ceases to exist when $t = 1/y_I$. Therefore if $y_I > 0$ then the interval of definition of the solution is $(-\infty, 1/y_I)$ while if $y_I < 0$ then the interval of definition of the solution is $(1/y_I, \infty)$. Notice that both of these intervals contain the initial time $t = 0$. Finally, notice that our explicit solution recovers the stationary solution when $y_I = 0$ even though it was derived assuming that $y_I \neq 0$.

You might think that the blow-up seen in the last example had something to do with the fact that the stationary point $y = 0$ was bad for our recipe. However, the next example shows that blow-up happens even when there are no stationary points.

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 1 + y^2\,, \qquad y(0) = y_I\,.$$

Give its interval of definition.

Solution. Because $1 + y^2 > 0$, we see there are no stationary solutions of this equation. Solutions are given implicitly by

$$t = \int \frac{\mathrm{d}y}{1 + y^2} = \tan^{-1}(y) + c\,.$$

The initial condition then implies that $0 = \tan^{-1}(y_I) + c$, whereby

$$c = -\tan^{-1}(y_I)\,.$$

Here we adopt the usual convention that $-\frac{\pi}{2} < \tan^{-1}(y_I) < \frac{\pi}{2}$. The solution therefore is given implicitly by

$$t = \tan^{-1}(y) - \tan^{-1}(y_I)\,.$$

This may be solved for y explicitly by first noticing that

$$\tan^{-1}(y) = t + \tan^{-1}(y_I)\,,$$

and then taking the tangent of both sides to obtain

$$y = \tan\bigl(t + \tan^{-1}(y_I)\bigr)\,.$$

Because tan becomes undefined at $-\frac{\pi}{2}$ and $\frac{\pi}{2}$, the interval of definition for this solution is $(-\frac{\pi}{2} - \tan^{-1}(y_I), \frac{\pi}{2} - \tan^{-1}(y_I))$. Notice that this interval contains the initial time $t = 0$ because $-\frac{\pi}{2} < \tan^{-1}(y_I) < \frac{\pi}{2}$.

The next example shows another way solutions of nonlinear equations can break down.

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2y}\,, \qquad y(0) = y_I \neq 0\,.$$

Give its interval of definition.

Solution. Notice that because the right-hand side of the differential equation does not make sense when $y = 0$, no solution may take that value. In particuluar, we require that $y_I \neq 0$. Because $1/(2y) \neq 0$, we see there are no stationary solutions of this equation. Solutions therefore are given implicitly by

$$t = \int 2y\,\mathrm{d}y = y^2 + c\,.$$

The initial condition then implies that

$$0 = y_I^2 + c\,,$$

whereby $c = -y_I^2$. The solution therefore is given implicitly by

$$t = y^2 - y_I^2\,.$$

This may be solved for y explicitly by first noticing that

$$y^2 = t + y_I^2\,,$$

and then taking the square root of both sides to obtain

$$y = \pm\sqrt{t + y_I^2}\,.$$

We must then choose the sign of the square root so that the solution agrees with the initial data — i.e. positive when $y_I > 0$ and negative when $y_I < 0$. In either case we obtain

$$y = \mathrm{sign}(y_I)\sqrt{t + y_I^2}\,,$$

where we define

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{if } x > 0\,, \\ 0 & \text{if } x = 0\,, \\ -1 & \text{if } x < 0\,. \end{cases}$$

Finally, notice that to keep the argument of the square root positive we must require that $t > -y_I^2$. The interval of definition for this solution therefore is $(-y_I^2, \infty)$.

Remark. In the above example the solution does not blow-up as t approaches $-y_I^2$. Indeed, the solution $y(t)$ approaches 0 as t approaches $-y_I^2$. However, the derivative of the solution is given by

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2y} = \frac{\operatorname{sign}(y_I)}{2\sqrt{t + y_I^2}}\,,$$

which does blow-up as t approaches $-y_I^2$. This happens because the solution approaches the value 0 where the right-hand side of the equation is not defined.

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{x}{y}\,, \qquad y(3) = -4\,.$$

Give its interval of definition.

Solution. This equation is separable. Its right-hand side is undefined when $y = 0$. It has no stationary solutions. Earlier we showed that its solutions are given implicitly by

$$y^2 = -x^2 - c\,, \qquad \text{for some constant } c\,.$$

The initial condition then implies that $(-4)^2 = -3^2 - c$, from which we solve to find that $c = -(-4)^2 - 3^2 = -16 - 9 = -25$. It follows that

$$y = -\sqrt{25 - x^2}\,,$$

where the negative square root is needed to satisfy the initial condition. This is a solution when $25 - x^2 > 0$, so its interval of definition is $(-5, 5)$.

Example. Find the solution to the initial-value problem

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{3x + xz^2}{z + x^2 z}\,, \qquad z(1) = -3\,.$$

Give its interval of definition.

Solution. This equation is separable. Its right-hand side is undefined when $z = 0$. It has no stationary solutions. Earlier we showed that its solution is given implicitly by

$$\log(3 + z^2) = \log(1 + x^2) - c\,, \qquad \text{for some constant } c\,.$$

The initial condition then implies that

$$\log\big(3 + (-3)^2\big) = \log\big(1 + 1^2\big) - c\,,$$

which gives $c = \log(2) - \log(12) = -\log(6)$. It follows that

$$\log(3 + z^2) = \log(1 + x^2) + \log(6) = \log\big((1 + x^2)\,6\big) = \log(6 + 6x^2)\,,$$

whereby $3 + z^2 = 6 + 6x^2$. Upon solving this for z we obtain

$$z = -\sqrt{3 + 6x^2}\,,$$

where the negative square root is needed to satisfy the initial condition. This is a solution when $3 + 6x^2 > 0$, so its interval of definition is $(-\infty, \infty)$.

4.4. **Analysis of Implicit Solutions.** Even when we cannot obtain explicit solutions by this recipe, often we can use the implicit solution to describe how all solutions behave and to determine their intervals of definition. We begin with an autonomous example.

Example. Solve the general initial-value problem

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{e^x}{1+x}, \qquad x(0) = x_I \neq 1 .$$

Solution. The differential equation is autonomous. Its right-hand side is undefined when $x = -1$. It has no stationary solutions. Its separated differential form is

$$(1+x)e^{-x}\,\mathrm{d}x = \mathrm{d}t .$$

One integration-by-parts yields

$$\int (1+x)e^{-x}\,\mathrm{d}x = -(1+x)e^{-x} + \int e^{-x}\,\mathrm{d}x = -(2+x)e^{-x} + c ,$$

so that the solutions of the differential equation satisfy

$$t = -(2+x)e^{-x} + c .$$

The initial condition implies $0 = -(2+x_I)e^{-x_I} + c$, whereby

$$c = (2+x_I)e^{-x_I} .$$

The solution of the initial-value problem therefore satisfies

$$t - (2+x_I)e^{-x_I} = -(2+x)e^{-x} .$$

This equation cannot be solved for x to obtain explicit solutions.

Let $G(x) = -(2+x)e^{-x}$, so that the solution of the initial-value problem satisfies

$$t - c = G(x) , \qquad \text{where } c = -G(x_I) .$$

Because $G'(x) = (1+x)e^{-x}$, it can be seen that $G'(x) > 0$ for $x > -1$ while $G'(x) < 0$ for $x < -1$. This sign analysis shows that $G(x)$ has a minimum value of $G(-1) = -e$ while

$$\lim_{x\to-\infty} G(x) = +\infty , \qquad \lim_{x\to+\infty} G(x) = 0 .$$

Because $x = -1$ is forbidden, there will be no solution unless $t - c > G(-1) = -e$.

If we sketch a graph of $G(x) = -(2+x)e^{-x}$ then we can see the following.

- The equation $G(x) = t - c$ has one solution with $x > -1$ provided $-e < t - c < 0$. This gives the solution of the initial-value problem when $x_I > -1$. It will be an increasing function $X(t)$ with interval of definition $(c - e, c)$ such that

$$\lim_{t\to(c-e)^+} X(t) = -1 , \qquad \lim_{t\to c^-} X(t) = \infty .$$

- The equation $G(x) = t - c$ has one solution with $x < -1$ provided $-e < t - c < \infty$. This gives the solution of the initial-value problem when $x_I < -1$. It will be a decreasing function $X(t)$ with interval of definition $(c - e, \infty)$ such that

$$\lim_{t\to(c-e)^+} x(t) = -1 , \qquad \lim_{t\to\infty} x(t) = -\infty .$$

A similar analysis can be carried out for separable equations, but more care is required to determine the interval of definition.

Example. Solve the general initial-value problem

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{e^x \cos(t)}{1+x}, \qquad x(t_I) = x_I \neq 1 .$$

Solution. The differential equation is separable. Its right-hand side is undefined when $x = -1$. It has no stationary solutions. Its separated differential form is

$$(1+x)e^{-x}\,\mathrm{d}x = \cos(t)\,\mathrm{d}t .$$

Because

$$\int (1+x)e^{-x}\,\mathrm{d}x = -(2+x)e^{-x} + c_1 , \qquad \int \cos(t)\,\mathrm{d}t = \sin(t) + c_2 ,$$

the solutions of the differential equation satisfy

$$\sin(t) = -(2+x)e^{-x} + c .$$

The initial condition implies $\sin(t_I) = -(2+x_I)e^{-x_I} + c$, whereby

$$c = \sin(t_I) + (2+x_I)e^{-x_I} .$$

The solution of the initial-value problem therefore satisfies

$$\sin(t) - \sin(t_I) - (2+x_I)e^{-x_I} = -(2+x)e^{-x} .$$

This equation cannot be solved for x to obtain explicit solutions.

Let $G(x) = -(2+x)e^{-x}$, so that the solution of the initial-value problem satisfies

$$\sin(t) - c = G(x) , \qquad \text{where } c = \sin(t_I) - G(x_I) .$$

This is the same $G(x)$ we analyzed in the previous example. From the graph of $G(x)$ produced by that analysis we conclude the following.

- The equation $G(x) = \sin(t) - c$ has one solution with $x > -1$ when $-e < \sin(t) - c < 0$. This solves the initial-value problem when $x_I > -1$. There are three possibilities.
  - If $-1 < c \leq 1$ then the solution $X(t) > -1$ exists for so long as $\sin(t) < c$. Its interval of definition is $\big(-\pi - \sin^{-1}(c), \sin^{-1}(c)\big)$, where $\sin^{-1}(c)$ is chosen so the interval contains t_I. We have $X(t) \to \infty$ as t approaches either endpoint.
  - If $1 < c < e - 1$ then the solution $X(t) > -1$ exists for all time and is oscillatory because $-e < \sin(t) - c < 0$ for every t. Its interval of definition is $(-\infty, \infty)$.
  - If $e - 1 \leq c < e + 1$ then the solution $X(t) > -1$ exists for so long as $c - e < \sin(t)$. Its interval of definition is $\big(\sin^{-1}(c-e), \pi - \sin^{-1}(c-e)\big)$, where $\sin^{-1}(c-e)$ is chosen so the interval contains t_I. We have $X(t) \to -1$ as t approaches either endpoint.
- The equation $G(x) = \sin(t) - c$ has one solution with $x < -1$ when $-e < \sin(t) - c$. This solves the initial-value problem when $x_I < -1$. There are two possibilites.
  - If $c < e - 1$ then the solution $X(t) < -1$ exists for all time and is oscillatory because $-e < \sin(t) - c$ for every t. Its interval of definition is $(-\infty, \infty)$.
  - If $e - 1 \leq c < e + 1$ then the solution $X(t) < -1$ exists for so long as $c - e < \sin(t)$. Its interval of definition is $\big(\sin^{-1}(c-e), \pi - \sin^{-1}(c-e)\big)$, where $\sin^{-1}(c-e)$ is chosen so this interval contains t_I. We have $X(t) \to -1$ as t approaches either endpoint.

4.5. **Theory for Separable Equations.** Up until now we have mentioned that you must be careful to check that the nonstationary solutions obtained from recipe (4.7) do not hit any of the stationary solutions, but we have not said why this leads to trouble. The next example illustrates the difficulty that arises.

Example. Find all solutions to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = 3y^{\frac{2}{3}}\,, \qquad y(0) = 0\,.$$

Solution. We see that $y = 0$ is a stationary point of this equation. Therefore $y(t) = 0$ is a stationary solution whose interval of definition is $(-\infty, \infty)$. However, let us carry out our recipe for nonstationary solutions to see where it leads. These solutions are given implicitly by

$$t = \int \frac{\mathrm{d}y}{3y^{\frac{2}{3}}} = \int \tfrac{1}{3} y^{-\frac{2}{3}}\,\mathrm{d}y = y^{\frac{1}{3}} + c\,.$$

Upon solving this for y we find $y = (t-c)^3$ where c is an arbitrary constant. Notice that each of these solutions hits the stationary point when $t = c$. The initial condition then implies that

$$0 = (0 - c)^3 = -c^3\,,$$

whereby $c = 0$. We thereby have found two solutions of the initial-value problem: $y(t) = 0$ and $y(t) = t^3$.

In fact, as we will now show, there are many more solutions of the initial-value problem. Let a and b be any two numbers such that $a \leq 0 \leq b$ and define $y(t)$ by

$$y(t) = \begin{cases} (t-a)^3 & \text{for } t < a\,, \\ 0 & \text{for } a \leq t \leq b\,, \\ (t-b)^3 & \text{for } b < t\,. \end{cases}$$

You will understand this function better if you graph it. It is clearly a differentiable function with

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) = \begin{cases} 3(t-a)^2 & \text{for } t < a\,, \\ 0 & \text{for } a \leq t \leq b\,, \\ 3(t-b)^2 & \text{for } b < t\,, \end{cases}$$

whereby it clearly satisfies the initial-value problem. Its interval of definition is $(-\infty, \infty)$. When $a = b = 0$ this reduces to $y(t) = t^3$.

Similarly, for every $a \leq 0$ we can construct the solution

$$y(t) = \begin{cases} (t-a)^3 & \text{for } t < a\,, \\ 0 & \text{for } a \leq t\,, \end{cases}$$

while for every $b \geq 0$ we can construct the solution

$$y(t) = \begin{cases} 0 & \text{for } t \leq b\,, \\ (t-b)^3 & \text{for } b < t\,. \end{cases}$$

The interval of definition for each of these solutions is also $(-\infty, \infty)$.

Remark. The above example shows a very important difference between nonlinear and linear equations. Specifically, it shows that for nonlinear equations an initial-value problem may not have a unique solution.

The nonuniqueness seen in the previous example arises because $g(y) = 3y^{\frac{2}{3}}$ does not behave nicely at the stationary point $y = 0$. It is clear that g is continuous at 0, but because $g'(y) = 2y^{-\frac{1}{3}}$ we see that g is not differentiable at 0. The following fact states the differentiablity of g is enough to ensure that the solution of the initial-value problem exists and is unique.

Theorem 4.1. *Let $f(t)$ and $g(y)$ be functions defined over the open intervals (t_L, t_R) and (y_L, y_R) respecively such that*

- *f is continuous over (t_L, t_R),*
- *g is continuous over (y_L, y_R),*
- *g is differentiable at each of its zeros in (y_L, y_R).*

Then for every inital time t_I in (t_L, t_R), and every initial value y_I in (y_L, y_R) there exists a unique solution $y = Y(t)$ to the initial-value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)\, g(y)\,, \qquad y(t_I) = y_I\,, \tag{4.9}$$

that is defined over every time interval (a, b) such that

- *t_I is in (a, b),*
- *(a, b) is contained within (t_L, t_R),*
- *$Y(t)$ remains within (y_L, y_R) while t is in (a, b).*

Moreover, this solution is continuously differentiable and is determined by our recipe. This means either $g(y_I) = 0$ and $Y(t) = y_I$ is a stationary solution, or $g(y_I) \neq 0$ and $Y(t)$ is a nonstationary solution that satisfies

$$G\big(Y(t)\big) = F(t)\,, \qquad Y(t_I) = y_I\,,$$

where $F(t)$ is defined for every t in (t_L, t_R) by the definite integral

$$F(t) = \int_{t_I}^{t} f(s)\,\mathrm{d}s\,,$$

while $G(y)$ is defined by the definite integral

$$G(y) = \int_{y_I}^{y} \frac{1}{g(x)}\,\mathrm{d}x\,,$$

whenever the point y is in (y_L, y_R) and neither y nor any point between y and y_I is a stationary point of g.

In particular, if f is continuous over $(-\infty, \infty)$ while g is differentiable over $(-\infty, \infty)$ then the initial-value problem (4.1) has a unique solution that either exists for all time or "blows up" in a finite time. Moreover, this solution is continuously differentiable and is determined by our recipe. We saw this "blow up" behavior in the examples above with $g(y) = y^2$ and $g(y) = 1 + y^2$. Indeed, it can be seen whenever $g(y)$ is a polynomial of degree two or more.

Remark. The above theorem implies that if the initial point y_I lies between two stationary points within (y_L, y_R) then the solution $Y(t)$ exists for all t in (t_L, t_R). This is because the uniqueness assertion implies $Y(t)$ cannot cross any stationary point, and therefore is trapped within (y_L, y_R). In particular, if g is differentiable over $(-\infty, \infty)$ then the only solutions that might "blows up" in a finite time are those that are not trapped above and below by stationary points.

Example. If $g(y) = y^2$ then the only stationary point is $y = 0$. Because $g(y) > 0$ when $y \neq 0$ we see that every nonstationary solution $Y(t)$ will be an increasing function of t. This fact is verified by the formula we derived earlier,

$$Y(t) = \frac{y_I}{1 - y_I t} .$$

When $y_I > 0$ the interval of definition is $(-\infty, 1/y_I)$ and we see that $Y(t) \to +\infty$ as $t \to 1/y_I$ while $Y(t) \to 0$ as $t \to -\infty$. In this case the solution is trapped below as $t \to -\infty$ by the stationary point $y = 0$. Similarly, when $y_I < 0$ the interval of definition is $(1/y_I, \infty)$ and we see that $Y(t) \to -\infty$ as $t \to 1/y_I$ while $Y(t) \to 0$ as $t \to \infty$. In this case the solution is trapped above as $t \to \infty$ by the stationary point $y = 0$.

Remark. Even in such cases where we cannot find an explicit inverse function of $G(y)$ we often can determine the interval of definition of the solution directly from the equation

$$G(y) = F(t) , \quad \text{where} \quad F(t) = \int_{t_I}^{t} f(s)\,\mathrm{d}s , \quad G(y) = \int_{y_I}^{y} \frac{1}{g(x)}\,\mathrm{d}x .$$

For example, if $g(y) > 0$ over $[y_I, \infty)$ then $G(y)$ will be increasing over $[y_I, \infty)$ and the solution $Y(t)$ will be defined over the largest interval (a, b) such that t_I is in (a, b), (a, b) is contained within (t_L, t_R), and

$$F(t) < \lim_{y \to +\infty} G(y) .$$

If the above limit is finite and equal to $F(b)$ then the solution "blows up" as $t \to b^-$.

Similarly, if $g(y) > 0$ over $(-\infty, y_I]$ then $G(y)$ will be increasing over $(-\infty, y_I]$ and the solution $Y(t)$ will be defined over the largest interval (a, b) such that t_I is in (a, b), (a, b) is contained within (t_L, t_R), and

$$\lim_{y \to -\infty} G(y) < F(t) .$$

If the above limit is finite and equal to $F(b)$ then the solution "blows down" as $t \to b^-$.

The same kind of analysis can be carried out to determine the times at which a solution approaches a forbidden value. Two such examples were given in the previous section.