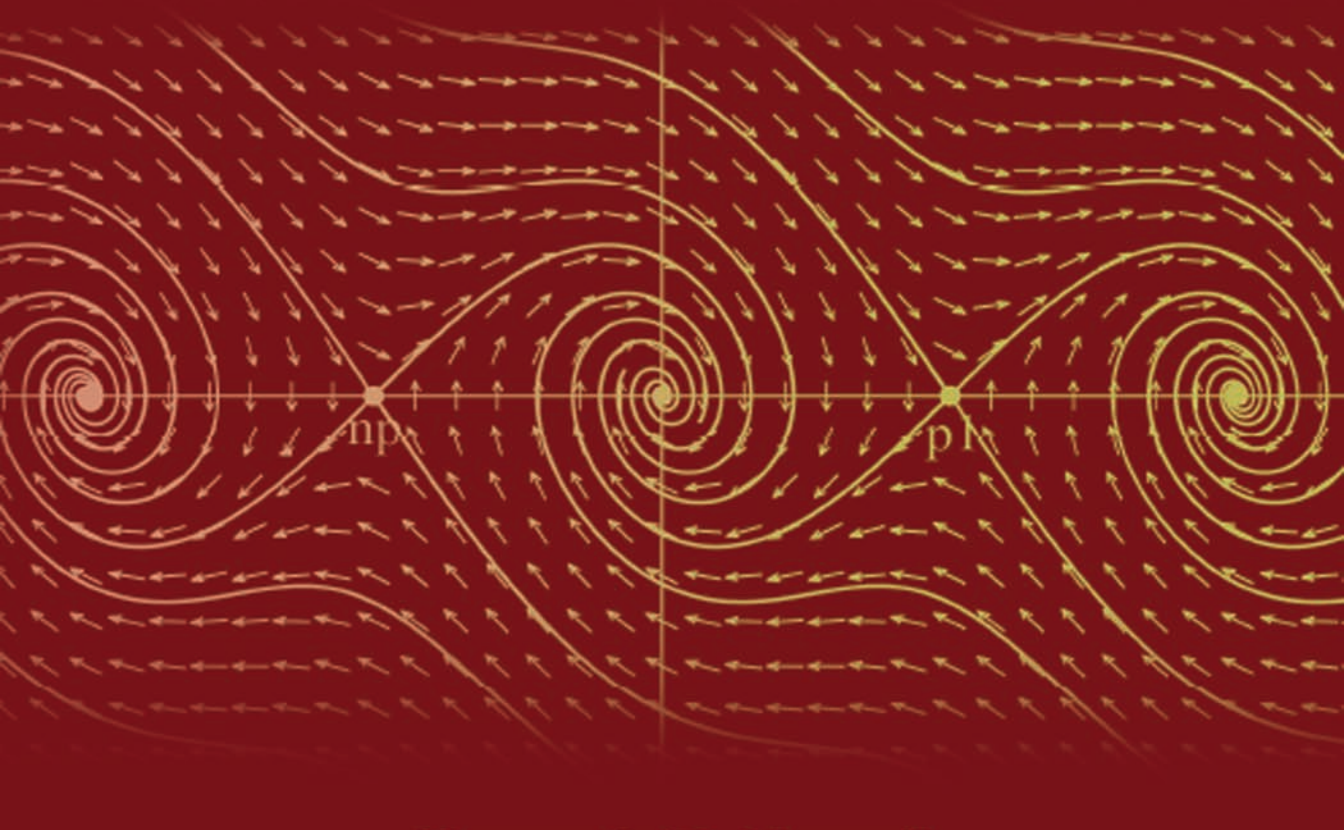
TEXTBOOKS IN MATHEMATICS

ORDINARY DIFFERENTIAL EQUATIONS

An Introduction to the Fundamentals

SECOND EDITION



Kenneth B. Howell

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

ORDINARY DIFFERENTIAL EQUATIONS

An Introduction to the Fundamentals

Second Edition

Textbooks in Mathematics

Series editors:

Al Boggess and Ken Rosen

CRYPTOGRAPHY: THEORY AND PRACTICE, FOURTH EDITION
Douglas R. Stinson and Maura B. Paterson

GRAPH THEORY AND ITS APPLICATIONS, THIRD EDITION
Jonathan L. Gross, Jay Yellen and Mark Anderson

COMPLEX VARIABLES: A PHYSICAL APPROACH WITH APPLICATIONS, SECOND EDITION
Steven G. Krantz

GAME THEORY: A MODELING APPROACH
Richard Alan Gillman and David Housman

FORMAL METHODS IN COMPUTER SCIENCE
Jiacun Wang and William Tepfenhart

SPHERICAL GEOMETRY AND ITS APPLICATIONS
Marshall A. Whittlesey

AN INTRODUCTION TO MATHEMATICAL PROOFS
Nicholas A. Loehr

COMPUTATIONAL PARTIAL DIFFERENTIAL EQUATIONS USING MATLAB®
Jichun Li and Yi-Tung Chen

AN ELEMENTARY TRANSITION TO ABSTRACT MATHEMATICS
Gove Effinger and Gary L. Mullen

MATHEMATICAL MODELING WITH EXCEL, SECOND EDITION
Brian Albright and William P. Fox

PRINCIPLES OF FOURIER ANALYSIS, SECOND EDITION
Kenneth B. Howell

https://www.crcpress.com/Textbooks-in-Mathematics/book-series/CANDHTEXBOOMTH

TEXTBOOKS IN MATHEMATICS

ORDINARY DIFFERENTIAL EQUATIONS

An Introduction to the Fundamentals

Second Edition

Kenneth B. Howell
University of Alabama in Huntsville, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Second edition published 2020
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

[First edition published by CRC Press 2015]

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 978-1-138-60583-1 (hbk)
ISBN: 978-0-429-34742-9 (ebk)

Typeset in NimbusRomNo9L
by Author

Contents

Preface
(With Important Information for the Reader)

This textbook reflects my personal views on how an "ideal" introductory course in ordinary differential equations should be taught, tempered by such practical constraints as the time available, and the level and interests of the students. It also reflects my beliefs that a good text should both be a useful resource beyond the course, and back up its claims with solid and clear proofs (even if some of those proofs are ignored by most readers). Moreover, I hope, it reflects the fact that such a book should be written to engage those normally taking such a course; namely, students who are reasonably acquainted with the differentiation and integration of functions of one variable, but who might not yet be experts and may, on occasion, need to return to their elementary calculus texts for review. Most of these students are not mathematicians and probably have no desire to become professional mathematicians. Still, most are interested in fields of study in which a fundamental understanding of the mathematics and applications of differential equations is valuable.

Do note that this text only assumes the student reader is acquainted with single-variable calculus. It is not assumed that the reader has had courses in multi-variable calculus or linear algebra. If you, the reader, have had (or are taking) these and other more advanced courses, then great. You can delve into a few more topics. In particular, if you've had a course in linear algebra or real analysis, then you can (and should) be on the lookout for points where the theory from those more advanced courses can be applied to simplify some of the discussion.

Of course, while I wrote this text for the students, the needs of the instructors were kept firmly in mind. After all, this is the text my colleagues and I have been using for the last several years.

Whether you are a student, instructor or just a casual reader, there are a number of things you should be aware of before starting the first chapter:

1. *Extra material*: There is more material in this text than can be reasonably covered in a "standard" one-semester introductory course. In part, this is to provide the material for a variety of "standard" courses which may or may not cover such topics as Laplace transforms, series solutions, systems and numerical methods. Beyond that, though, there are expanded discussions of topics normally covered, as well as topics rarely covered, but which are still elementary enough and potentially useful enough to merit discussion. There are also proofs that are not simple and illuminating enough to be included in the basic exposition, but should still be there to keep the author honest and to serve as a reference for others. Because of this extra material, there is an appendix, *Author's Guide to Using This Text*, with advice on which sections must be covered, which are optional, and which are best avoided by the first-time reader. It also contains a few opinionated comments.

2. *Computer math packages*: At several points in the text, the use of a "computer math package" is advised or, in exercises, required. By a "computer math package", I mean one of those powerful software packages such as *Maple* or *Mathematica* that can do symbolic calculations, graphing and so forth. Unfortunately, software changes over time, new products emerge, and companies providing this software can be bought and sold. In addition, you may be able to find other computational resources on the Internet (but be aware that websites can be much more fickle and untrustworthy than major software providers). For these reasons, details on using such software are not included in this text. You will have to figure that out yourself (it's not that hard). I will tell you this: Extensive use of *Maple* was made in preparing this text. In fact, most of the graphs were generated in *Maple* and then cleaned up using commercial graphics software.

   On the subject of computer math packages: Please become reasonably proficient in at least one. If you are reading this, you are probably working in or will be working in a field in which this sort of knowledge is invaluable. But don't think this software can replace a basic understanding of the mathematics you are using. Even a simple calculator is useless to someone who doesn't understand just what $+$ and $\times$ mean. Mindlessly using this software can lead to serious and costly mistakes (as discussed in Section 10.3).

3. *Additional chapters*: By the way, I do not consider this text complete. Additional chapters on systems of differential equations and boundary-value problems are being written for a possible follow-up text. As these chapters become written (and rewritten), they will become available at the website for this text (see below).

4. *Text website*: While this edition remains in publication, I intend to maintain a website for this text containing at least the following:

   - A lengthy solution manual
   - The aforementioned chapters extending the material in this text
   - A list of known errata discovered since the book's publication

   At the time I was writing this, the text's website was at http://howellkb.uah.edu/DEtext/. With luck, that will still be the website's location when you need it. Unfortunately, I cannot guarantee that my university will not change its website policies and conventions, forcing you to search for the current location of the text's website. If you must search for this site, I would suggest starting with the website of the Department of Mathematical Sciences of the University of Alabama in Huntsville.

Those acquainted with the first edition may wonder how this, the second edition, differs from the first. The answer: Not much — known typos have been corrected, some of the discussion has been cleaned up, and many more exercises have been added. In fact, two new "chapters" consist of nothing more than review exercises for Parts II and III. Aside from that, two chapters on numerical methods (extending the original discussion on the Euler method) have been added.

Finally, I must thank the many students and fellow faculty who have used earlier versions of this text and have provided the feedback that I have found invaluable in preparing this edition. Those comments are very much appreciated. And, if you, the reader, should find any errors or would like to make any suggestions or comments regarding this text, please let me know. That, too, would be very much appreciated.

Dr. Kenneth B. Howell
(howellkb@uah.edu)

Part I

The Basics

1

The Starting Point: Basic Concepts and Terminology

Let us begin our study of "differential equations" with a few basic questions — questions that any beginner should ask:

What are "differential equations"?

What can we do with them? Solve them? If so, what do we solve for? And how?

and, of course,

What good are they, anyway?

In this chapter, we will try to answer these questions (along with a few you would not yet think to ask), at least well enough to begin our studies. With luck we will even raise a few questions that cannot be answered now, but which will justify continuing our study. In the process, we will also introduce and examine some of the basic concepts, terminology and notation that will be used throughout this book.

1.1 Differential Equations: Basic Definitions and Classifications

A *differential equation* is an equation involving some function of interest along with a few of its derivatives. Typically, the function is unknown, and the challenge is to determine what that function could possibly be.

Differential equations can be classified either as "ordinary" or as "partial". An *ordinary differential equation* is a differential equation in which the function in question is a function of only one variable. Hence, its derivatives are the "ordinary" derivatives encountered early in calculus. For the most part, these will be the sort of equations we'll be examining in this text. For example,

$$\frac{dy}{dx} = 4x^3$$

$$\frac{dy}{dx} + \frac{4}{x}y = x^2$$

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - 3y = 65\cos(2x)$$

$$4x^2 \frac{d^2y}{dx^2} + 4x\frac{dy}{dx} + [4x^2 - 1]y = 0$$

and

$$\frac{d^4y}{dx^4} = 81y$$

are some differential equations that we will later deal with. In each, y denotes a function that is given by some, yet unknown, formula of x. Of course, there is nothing sacred about our choice of symbols. We will use whatever symbols are convenient for the variables and functions, especially if the problem comes from an application and the symbols help remind us of what they denote (such as when we use t for a measurement of time).[1]

A *partial differential equation* is a differential equation in which the function of interest depends on two or more variables. Consequently, the derivatives of this function are the partial derivatives developed in the later part of most calculus courses.[2] Because the methods for studying partial differential equations often involve solving ordinary differential equations, it is wise to first become reasonably adept at dealing with ordinary differential equations before tackling partial differential equations.

As already noted, this text is mainly concerned with ordinary differential equations. So let us agree that, unless otherwise indicated, the phrase "differential equation" in this text means "ordinary differential equation". If you wish to further simplify the phrasing to "DE" or even to something like "Diffy-Q", go ahead. This author, however, will not be so informal.

Differential equations are also classified by their "order". The *order* of a differential equation is simply the order of the highest order derivative explicitly appearing in the equation. The equations

$$\frac{dy}{dx} = 4x^3 \quad , \quad \frac{dy}{dx} + \frac{4}{x}y = x^2 \quad \text{and} \quad y\frac{dy}{dx} = -9.8x$$

are all first-order equations. So is

$$\frac{dy}{dx} + 3y^2 = y\left(\frac{dy}{dx}\right)^4 \quad ,$$

despite the appearance of the higher powers — dy/dx is still the highest *order* derivative in this equation, even if it is multiplied by itself a few times.

The equations

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - 3y = 65\cos(2x) \quad \text{and} \quad 4x^2\frac{d^2y}{dx^2} + 4x\frac{dy}{dx} + [4x^2 - 1]y = 0$$

are second-order equations, while

$$\frac{d^3y}{dx^3} = e^{4x} \quad \text{and} \quad \frac{d^3y}{dx^3} - \frac{d^2y}{dx^2} + \frac{dy}{dx} - y = x^2$$

are third-order equations.

?▶ *Exercise 1.1:* *What is the order of each of the following equations?*

$$\frac{d^2y}{dx^2} + 3\frac{dy}{dx} - 7y = \sin(x)$$

[1] On occasion, we may write "$y = y(x)$" to explicitly indicate that, in some expression, y denotes a function given by some formula of x with $y(x)$ denoting that "formula of x". More often, it will simply be understood that y is a function given by some formula of whatever variable appears in our expressions.

[2] A brief introduction to partial derivatives is given in Section 3.7 for those who are interested and haven't yet seen partial derivatives.

$$\frac{d^5 y}{dx^5} - \cos(x)\frac{d^3 y}{dx^3} = y^2$$

$$\frac{d^5 y}{dx^5} - \cos(x)\frac{d^3 y}{dx^3} = y^6$$

$$\frac{d^{42} y}{dx^{42}} = \left(\frac{d^3 y}{dx^3}\right)^2 \quad .$$

In practice, higher-order differential equations are usually more difficult to solve than lower-order equations. This, of course, is not an absolute rule. There are some very difficult first-order equations, as well as some very easily solved twenty-seventh-order equations.

Solutions: The Basic Notions*

Any function that satisfies a given differential equation is called a *solution* to that differential equation. "Satisfies the equation", means that, if you plug the function into the differential equation and compute the derivatives, then the result is an equation that is true no matter what real value we replace the variable with. And if that resulting equation is not true for some real values of the variable, then that function is not a solution to that differential equation.

!►*Example 1.1:* *Consider the differential equation*

$$\frac{dy}{dx} - 3y = 0 \quad .$$

If, in this differential equation, we let $y(x) = e^{3x}$ *(i.e., if we replace* y *with* e^{3x} *), we get*

$$\frac{d}{dx}\left[e^{3x}\right] - 3e^{3x} = 0$$

$$\hookrightarrow \qquad 3e^{3x} - 3e^{3x} = 0$$

$$\hookrightarrow \qquad 0 = 0 \quad ,$$

which certainly is true for every real value of x *. So* $y(x) = e^{3x}$ *is a solution to our differential equation.*

On the other hand, if we let $y(x) = x^3$ *in this differential equation, we get*

$$\frac{d}{dx}\left[x^3\right] - 3x^3 = 0$$

$$\hookrightarrow \qquad 3x^2 - 3x^3 = 0$$

$$\hookrightarrow \qquad 3x^2(1 - x) = 0 \quad ,$$

which is true only if $x = 0$ *or* $x = 1$ *. But our interest is not in finding values of* x *that make the equation true; our interest is in finding* functions of x *(i.e.,* $y(x)$ *) that make the equation true* for all *values of* x *. So* $y(x) = x^3$ *is not a solution to our differential equation. (And it makes no sense, whatsoever, to refer to either* $x = 0$ *or* $x = 1$ *as solutions, here.)*

---

* Warning: The discussion of "solutions" here is rather incomplete so that we can get to the basic, intuitive concepts quickly. We will refine our notion of "solutions" in Section 1.3 starting on page 14.

Typically, a differential equation will have many different solutions. Any formula (or set of formulas) that describes all possible solutions is called a *general solution* to the equation.

!►Example 1.2: *Consider the differential equation*

$$\frac{dy}{dx} = 6x \quad .$$

All possible solutions can be obtained by just taking the indefinite integral of both sides,

$$\int \frac{dy}{dx}\,dx = \int 6x\,dx$$

$$\hookrightarrow \qquad y(x) + c_1 = 3x^2 + c_2$$

$$\hookrightarrow \qquad y(x) = 3x^2 + c_2 - c_1$$

where c_1 and c_2 are arbitrary constants. Since the difference of two arbitrary constants is just another arbitrary constant, we can replace the above $c_2 - c_1$ with a single arbitrary constant c and rewrite our last equation more succinctly as

$$y(x) = 3x^2 + c \quad .$$

This formula for y describes all possible solutions to our original differential equation — it is a general solution *to the differential equation in this example. To obtain an individual solution to our differential equation, just replace c with any particular number. For example, respectively letting $c = 1$, $c = -3$, and $c = 827$ yield the following three solutions to our differential equation:*

$$3x^2 + 1 \quad , \quad 3x^2 - 3 \quad \text{and} \quad 3x^2 + 827 \quad .$$

As just illustrated, general solutions typically involve arbitrary constants. In many applications, we will find that the values of these constants are not truly arbitrary but are fixed by additional conditions imposed on the possible solutions (so, in these applications at least, it would be more accurate to refer to the "arbitrary" constants in the general solutions of the differential equations as "yet undetermined" constants).

Normally, when given a differential equation and no additional conditions, we will want to determine all possible solutions to the given differential equation. Hence, "solving a differential equation" often means "finding a general solution" to that differential equation. That will be the default meaning of the phrase "solving a differential equation" in this text. Notice, however, that the resulting "solution" is not a single function that satisfies the differential equation (which is what we originally defined "a solution" to be), but is a formula describing all possible functions satisfying the differential equation (i.e., a "general solution"). Such ambiguity often arises in everyday language, and we'll just have to live with it. Simply remember that, in practice, the phrase "a solution to a differential equation" can refer either to

any single function that satisfies the differential equation,

or

any formula describing all the possible solutions (more correctly called a general solution).

In practice, it is usually clear from the context just which meaning of the word "solution" is being used. On occasions where it might not be clear, or when we wish to be very precise, it is standard to call any single function satisfying the given differential equation a *particular solution*. So, in the last example, the formulas

$$3x^2 + 1 \quad , \quad 3x^2 - 3 \quad \text{and} \quad 3x^2 + 827$$

describe particular solutions to

$$\frac{dy}{dx} = 6x \quad .$$

Initial-Value Problems

One set of auxiliary conditions that often arises in applications is a set of "initial values" for the desired solution. This is a specification of the values of the desired solution and some of its derivatives at a single point. To be precise, an *N^{th}-order set of initial values* for a solution y consists of an assignment of values to

$$y(x_0) \quad , \quad y'(x_0) \quad , \quad y''(x_0) \quad , \quad y'''(x_0) \quad , \quad \ldots \quad \text{and} \quad y^{(N-1)}(x_0)$$

where x_0 is some fixed number (in practice, x_0 is often 0) and N is some nonnegative integer.[3] Note that there are exactly N values being assigned and that the highest derivative in this set is of order $N - 1$.

We will find that N^{th}-order sets of initial values are especially appropriate for N^{th}-order differential equations. Accordingly, the term *N^{th}-order initial-value problem* will always mean a problem consisting of

1. an N^{th}-order differential equation, and
2. an N^{th}-order set of initial values.

For example,

$$\frac{dy}{dx} - 3y = 0 \qquad \text{with} \quad y(0) = 4$$

is a first-order initial-value problem. " ${}^{dy}/_{dx} - 3y = 0$ " is the first-order differential equation, and " $y(0) = 4$ " is the first-order set of initial values. On the other hand, the third-order differential equation

$$\frac{d^3y}{dx^3} + \frac{dy}{dx} = 0$$

along with the third-order set of initial conditions

$$y(1) = 3 \quad , \quad y'(1) = -4 \quad \text{and} \quad y''(1) = 10$$

makes up a third-order initial-value problem.

A *solution* to an initial-value problem is a solution to the differential equation that also satisfies the given initial values. The usual approach to solving such a problem is to first find the general solution to the differential equation (via any of the methods we'll develop later), and then determine the values of the 'arbitrary' constants in the general solution so that the resulting function also satisfies each of the given initial values.

---

[3] Remember, if $y = y(x)$, then

$$y' = \frac{dy}{dx} \quad , \quad y'' = \frac{d^2y}{dx^2} \quad , \quad y''' = \frac{d^3y}{dx^3} \quad , \quad \ldots \quad \text{and} \quad y^{(k)} = \frac{d^ky}{dx^k} \quad .$$

We will use the 'prime' notation for derivatives when the ${}^{d}/_{dx}$ notation becomes cumbersome.

!►Example 1.3: *Consider the initial-value problem*

$$\frac{dy}{dx} = 6x \qquad \text{with} \quad y(1) = 8 \quad .$$

From Example 1.2, we know that the general solution to the above differential equation is

$$y(x) = 3x^2 + c$$

where c *is an arbitrary constant. Combining this formula for* y *with the requirement that* $y(1) = 8$*, we have*

$$8 = y(1) = 3 \cdot 1^2 + c = 3 + c \quad ,$$

which, in turn, requires that

$$c = 8 - 3 = 5 \quad .$$

So the solution to the initial-value problem is given by

$$y(x) = 3x^2 + c \qquad \text{with} \quad c = 5 \quad ;$$

that is,

$$y(x) = 3x^2 + 5 \quad .$$

By the way, the terms "initial values", "initial conditions", and "initial data" are essentially synonymous and, in practice, are used interchangeably.

1.2 Why Care About Differential Equations? Some Illustrative Examples

Perhaps the main reason to study differential equations is that they naturally arise when we attempt to mathematically describe "real-world" processes that vary with, say, time or position. Let us look at one well-known process: the falling of some object towards the earth. To illustrate some of the issues involved, we'll develop two different sets of mathematical descriptions for this process.

By the way, any collection of equations and formulas describing some process is called a *(mathematical) model* of the process, and the process of developing a mathematical model is called, unsurprisingly, *modeling*.

The Situation to Be Modeled:

Let us concern ourselves with the vertical position and motion of an object dropped from a plane at a height of 1,000 meters. Since it's just being dropped, we may assume its initial downward velocity is 0 meters per second. The precise nature of the object — whether it's a falling marble, a frozen duck (live, unfrozen ducks don't usually fall) or some other familiar falling object — is not important at this time. Visualize it as you will.

The first two things one should do when developing a model is to sketch the process (if possible) and to assign symbols to quantities that may be relevant. A crude sketch of the process is in Figure 1.1 (I've sketched the object as a ball since a ball is easy to sketch). Following ancient traditions, let's make the following symbolic assignments:

m = the mass (in grams) of the object

t = time (in seconds) since the object was dropped

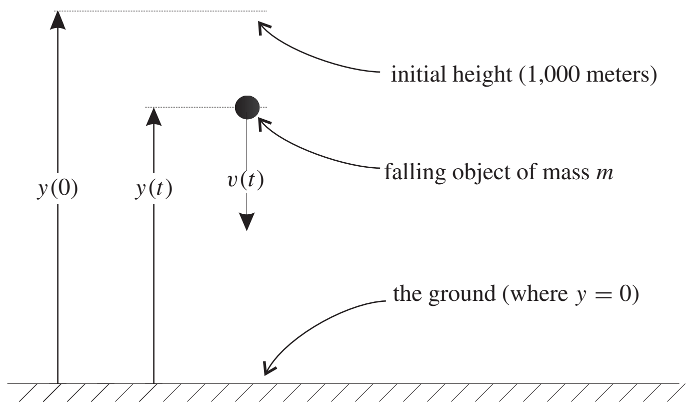


Figure 1.1: Rough sketch of a falling object of mass m.

$y(t)$ = vertical distance (in meters) between the object and the ground at time t

$v(t)$ = vertical velocity (in meters/second) of the object at time t

$a(t)$ = vertical acceleration (in meters/second^2) of the object at time t

Where convenient, we will use y, v and a as shorthand for $y(t)$, $v(t)$ and $a(t)$. Remember that, by the definition of velocity and acceleration,

$$v = \frac{dy}{dt} \qquad \text{and} \qquad a = \frac{dv}{dt} = \frac{d^2y}{dt^2} \quad .$$

From our assumptions regarding the object's position and velocity at the instant it was dropped, we have that

$$y(0) = 1{,}000 \qquad \text{and} \qquad \left.\frac{dy}{dt}\right|_{t=0} = v(0) = 0 \quad . \tag{1.1}$$

These will be our initial values. (Notice how appropriate it is to call these the "initial values" — $y(0)$ and $v(0)$ are, indeed, the initial position and velocity of the object.)

As time goes on, we expect the object to be falling faster and faster downwards, so we expect both the position and velocity to vary with time. Precisely how these quantities vary with time might be something we don't yet know. However, from Newton's laws, we do know

$$F = ma$$

where F is the sum of the (vertically acting) forces on the object. Replacing a with either the corresponding derivative of velocity or position, this equation becomes

$$F = m\frac{dv}{dt} \tag{1.2}$$

or, equivalently,

$$F = m\frac{d^2y}{dt^2} \quad . \tag{1.2$'$}$$

If we can adequately describe the forces acting on the falling object (i.e., the F), then the velocity, $v(t)$, and vertical position, $y(t)$, can be found by solving the above differential equations, subject to the initial conditions in line (1.1).

The Simplest Falling Object Model

The Earth's gravity is the most obvious force acting on our falling object. Checking a convenient physics text, we find that the force of the Earth's gravity acting on an object of mass m is given by

$$F_{\text{grav}} = -gm \qquad \text{where} \qquad g = 9.8 \quad \left(\text{meters/second}^2\right) \quad .$$

Of course, the value for g is an approximation and assumes that the object is not too far above the Earth's surface. It also assumes that we've chosen "up" to be the positive direction (hence the negative sign).

For this model, let us suppose the Earth's gravity, F_{grav}, is the only significant force involved. Assuming this (and keeping in mind that we are measuring distance in meters and time in seconds), we have

$$F = F_{\text{grav}} = -9.8m$$

in the "$F = ma$" equation. In particular, equation (1.2′) becomes

$$-9.8m = m\frac{d^2y}{dt^2} \quad .$$

The mass conveniently divides out, leaving us with

$$\frac{d^2y}{dt^2} = -9.8 \quad .$$

Taking the indefinite integral with respect to t of both sides of this equation yields

$$\int \frac{d^2y}{dt^2}\,dt = \int -9.8\,dt$$

$$\hookrightarrow \qquad \int \frac{d}{dt}\left(\frac{dy}{dt}\right)dt = \int -9.8\,dt$$

$$\hookrightarrow \qquad \frac{dy}{dt} + c_1 = -9.8t + c_2$$

$$\hookrightarrow \qquad \frac{dy}{dt} = -9.8t + c$$

where c_1 and c_2 are the arbitrary constants of integration and $c = c_2 - c_1$. This gives us our formula for ${}^{dy}/_{dt}$ up to an unknown constant c. But recall that the initial velocity is zero,

$$\left.\frac{dy}{dt}\right|_{t=0} = v(0) = 0 \quad .$$

On the other hand, setting t equal to zero in the formula just derived for ${}^{dy}/_{dt}$ yields

$$\left.\frac{dy}{dt}\right|_{t=0} = -9.8 \cdot 0 + c \quad .$$

Combining these two expressions for $y'(0)$ yields

$$0 = \left.\frac{dy}{dt}\right|_{t=0} = -9.8 \cdot 0 + c \quad .$$

Thus, $c = 0$ and our formula for ${}^{dy}/_{dt}$ reduces to

$$\frac{dy}{dt} = -9.8t \quad .$$

Again, we have a differential equation that is easily solved by simple integration,

$$\int \frac{dy}{dt}\,dt = \int -9.8t\,dt$$

$$\hookrightarrow \qquad y(t) + C_1 = -9.8\left[\tfrac{1}{2}t^2\right] + C_2$$

$$\hookrightarrow \qquad y(t) = -4.9t^2 + C$$

where, again, C_1 and C_2 are the arbitrary constants of integration and $C = C_2 - C_1$.[4] Combining this last equation with the initial condition for $y(t)$ (from line (1.1)), we get

$$1{,}000 = y(0) = -4.9 \cdot 0^2 + C \quad .$$

Thus, $C = 1{,}000$ and the vertical position (in meters) at time t is given by

$$y(t) = -4.9t^2 + 1{,}000 \quad .$$

A Better Falling Object Model

The above model does not take into account the resistance of the air to the falling object — a very important force if the object is relatively light or has a parachute. Let us add this force to our model. That is, for our "$F = ma$" equation, we'll use

$$F = F_{\text{grav}} + F_{\text{air}}$$

where F_{grav} is the force of gravity discussed above, and F_{air} is the force due to the air resistance acting on this particular falling body.

Part of our problem now is to determine a good way of describing F_{air} in terms relevant to our problem. To do that, let us list a few basic properties of air resistance that should be obvious to anyone who has stuck their hand out of a car window:

1. The force of air resistance does not depend on the position of the object, only on the relative velocity between it and the surrounding air. So, for us, F_{air} will just be a function of v, $F_{\text{air}} = F_{\text{air}}(v)$. (This assumes, of course, that the air is still — no up- or downdrafts — and that the density of the air remains fairly constant throughout the distance this object falls.)

2. This force is zero when the object is not moving, and its magnitude increases as the speed increases (remember, speed is the magnitude of the velocity). Hence, $F_{\text{air}}(v) = 0$ when $v = 0$, and $|F_{\text{air}}(v)|$ gets bigger as $|v|$ gets bigger.

3. Air resistance acts *against* the direction of motion. This means that the direction of the force of air resistance is opposite to the direction of motion. Thus, the sign of $F_{\text{air}}(v)$ will be opposite that of v.

While there are many formulas for $F_{\text{air}}(v)$ that would satisfy the above conditions, common sense suggests that we first use the simplest. That would be

$$F_{\text{air}}(v) = -\gamma v$$

[4] Note that slightly different symbols are being used to denote the different constants. This is highly recommended to prevent confusion when (and if) we ever review our computations.

where γ is some positive value. The actual value of γ will depend on such parameters as the object's size, shape, and orientation, as well as the density of the air through which the object is moving. For any given object, this value could be determined by experiment (with the aid of the equations we will soon derive).

?► **Exercise 1.2:** *Convince yourself that*

a: *this formula for* $F_{\text{air}}(v)$ *does satisfy the above three conditions, and*

b: *no simpler formula would work.*

We are now ready to derive the appropriate differential equations for our improved model of a falling object. The total force is given by

$$F = F_{\text{grav}} + F_{\text{air}} = -9.8m - \gamma v \quad .$$

Since this formula explicitly involves v instead of ${}^{dy}/_{dt}$, let us use the equation (1.2) version of "$F = ma$" from page 9,

$$F = m\frac{dv}{dt} \quad .$$

Combining the last two equations,

$$m\frac{dv}{dt} = F = -9.8m - \gamma v \quad .$$

Cutting out the middle and dividing through by the mass gives the slightly simpler equation

$$\frac{dv}{dt} = -9.8 - \kappa v \qquad \text{where} \quad \kappa = \frac{\gamma}{m} \quad . \tag{1.3}$$

Remember that γ, m and, hence, κ are positive constants, while $v = v(t)$ is a yet unknown function that satisfies the initial condition $v(0) = 0$. After solving this initial-value problem for $v(t)$, we could then find the corresponding formula for height at time t, $y(t)$, by solving the simple initial-value problem

$$\frac{dy}{dt} = v(t) \qquad \text{with} \quad y(0) = 1{,}000 \quad .$$

Unfortunately, we cannot solve equation (1.3) by simply integrating both sides with respect to t,

$$\int \frac{dv}{dt}\,dt = \int [-9.8 - \kappa v]\,dt \quad .$$

The first integral is not a problem. By the relation between derivatives and integrals, we still have

$$\int \frac{dv}{dt}\,dt = v(t) + c_1$$

where c_1 is an arbitrary constant. It's the other side that is the problem. Since κ is a constant, but $v = v(t)$ is an unknown function of t, the best we can do with the right-hand side is

$$\int [-9.8 - \kappa v]\,dt = -\int 9.8\,dt - \kappa \int v(t)\,dt = -9.8t + c_2 - \kappa \int v(t)\,dt \quad .$$

Again, c_2 is an arbitrary constant. However, since $v(t)$ is an unknown function, its integral is simply another unknown function of t. Thus, letting $c = c_2 - c_1$ and "integrating the equation" simply gives us the rather unhelpful formula

$$v(t) = -9.8t + c - (\kappa \cdot \textit{some unknown function of } t) \quad .$$

Fortunately, this is a text on differential equations, and methods for solving equations such as equation (1.3) will be discussed in Chapters 4 and 5. But there's no need to rush things. The main goal here is just to see how differential equations arise in applications. Of course, now that we have equation (1.3), we also have a good reason to continue on and learn how to solve it.

By the way, if we replace v in equation (1.3) with ${}^{dy}/_{dt}$, we get the second-order differential equation

$$\frac{d^2y}{dt^2} = -9.8 - \kappa\frac{dy}{dt} \quad .$$

This can be integrated, yielding

$$\frac{dy}{dt} = -9.8t - \kappa y + c$$

where c is an arbitrary constant. Again, this is a first-order differential equation that we cannot solve until we delve more deeply into the various methods for solving these equations. And if, in this last equation, we again use the fact that $v = {}^{dy}/_{dt}$, all we get is

$$v = -9.8t - \kappa y + c \tag{1.4}$$

which is another not-very-helpful equation relating the unknown functions $v(t)$ and $y(t)$.[5]

Summary of Our Models and the Related Initial Value Problems

For the first model of a falling body, we had the second-order differential equation

$$\frac{d^2y}{dt^2} = -9.8 \quad .$$

along with the initial conditions

$$y(0) = 1{,}000 \qquad \text{and} \qquad y'(0) = 0 \quad .$$

In other words, we had a second-order initial-value problem. This problem, as we saw, was rather easy to solve.

For the second model, we still had the initial conditions

$$y(0) = 1{,}000 \qquad \text{and} \qquad y'(0) = 0 \quad ,$$

but we found it a little more convenient to write the differential equation as

$$\frac{dv}{dt} = -9.8 - \kappa v \qquad \text{where} \qquad \frac{dy}{dt} = v$$

and κ was some positive constant. There are a couple of ways we can view this collection of equations. First of all, we could simply replace the v with ${}^{dy}/_{dt}$ and say we have the second-order initial problem

$$\frac{d^2y}{dt^2} = -9.8 - \kappa\frac{dy}{dt}$$

with

$$y(0) = 1,000 \qquad \text{and} \qquad y'(0) = 0 \quad .$$

[5] Well, not completely useless — see Exercise 1.10 b on page 20.

Alternatively, we could (as was actually suggested) view the problem as two successive first-order problems:

$$\frac{dv}{dt} = -9.8 - \kappa v \qquad \text{with} \quad v(0) = 0 \quad ,$$

followed by

$$\frac{dy}{dt} = v(t) \qquad \text{with} \quad y(0) = 1{,}000 \quad .$$

The first of these two problems can be solved using methods we'll develop later. And once we have the solution, $v(t)$, to that, the second can easily be solved by integration.

Though, ultimately, the two ways of viewing our second model are equivalent, there are advantages to the second. It is conceptually simple, and it makes it a little easier to use solution methods that will be developed relatively early in this text. It also leads us to finding $v(t)$ before even considering $y(t)$. Moreover, it is probably the velocity of landing, not the height of landing, that most concerns a falling person with (or without) a parachute. Indeed, if we are lucky, the solution to the first, $v(t)$, may tell us everything we are interested in, and we won't have to deal with the initial-value problem for y at all.

Finally, it should be mentioned that, together, the two equations

$$\frac{dv}{dt} = -9.8 - \kappa v \qquad \text{and} \qquad \frac{dy}{dt} = v$$

form a "system of differential equations". That is, they comprise a set of differential equations involving unknown functions that are related to each other. This is an especially simple system since it can be solved by successively solving the individual equations in the system. Much more complicated systems can arise that are not so easily solved, especially when modeling physical systems consisting of many components, each of which can be modeled by a differential equation involving several different functions (as in, say, a complex electronic circuit). Dealing with these sorts of systems will have to wait until we've become reasonably adept at dealing with individual differential equations.

1.3 More on Solutions

Intervals of Interest

When discussing a differential equation and its solutions, we should include a specification of an interval (of nonzero length) over which the solution(s) is (are) to be valid. The choice of this interval, which we may call the *interval of solution*, the *interval of the solution's validity*, or, simply, the *interval of interest*, may be based on the problem leading to the differential equation, on mathematical considerations, or, to a certain extent, on the whim of the person presenting the differential equation. One thing we will insist on, in this text at least, is that **solutions must be continuous over this interval**.

!► *Example 1.4:* *Consider the equation*

$$\frac{dy}{dx} = \frac{1}{(x-1)^2} \quad .$$

Integrating this gives

$$y(x) = c - \frac{1}{x-1}$$

where c is an arbitrary constant. No matter what value c is, however, this function cannot be continuous over any interval containing $x = 1$ because $(x-1)^{-1}$ "blows up" at $x = 1$. So we will only claim that our solutions are valid over intervals that do not include $x = 1$. In particular, we have valid (continuous) solutions to this differential equation over the intervals $[0, 1)$, $(-\infty, 1)$, $(1, \infty)$, *and* $(2, 5)$; *but not over* $(0, 2)$ *or* $(0, 1]$ *or* $(-\infty, \infty)$.

Why should we make such an issue of continuity? Well consider, if a function is not continuous at a point, then its derivatives do not exist at that point — and without the derivatives existing, how can we claim that the function satisfies a particular differential equation?

Another reason for requiring continuity is that the differential equations most people are interested in are models for "real-world" phenomena, and real-world phenomena are normally continuous processes while they occur — the temperature of an object does not instantaneously jump by fifty degrees nor does the position of an object instantaneously change by three kilometers. If the solutions do "blow up" at some point, then

1. some of the assumptions made in developing the model are probably not valid, or
2. a catastrophic event is occurring in our process at that point, or
3. both.

Whatever is the case, it would be foolish to use the solution derived to predict what is happening beyond the point where "things blow up". That should certainly be considered a point where the validity of the solution ends.

Sometimes, it's not the mathematics that restricts the interval of interest, but the problem leading to the differential equation. Consider the simplest falling object model discussed earlier. There we had an object start falling from an airplane at $t = 0$ from a height of 1,000 meters. Solving the corresponding initial-value problem, we obtained

$$y(t) = -4.9t^2 + 1{,}000$$

as the formula for the height above the earth at time t. Admittedly, this satisfies the differential equation for all t, but, in fact, it only gives the height of the object from the time it starts falling, $t = 0$, to the time it hits the ground, T_{hit}.[6] So the above formula for $y(t)$ is a valid description of the position of the object only for $0 \leq t \leq T_{\text{hit}}$; that is, $[0, T_{\text{hit}}]$ is the interval of interest for this problem. Any use of this formula to predict the position of the object at a time outside the interval $[0, T_{\text{hit}}]$ is just plain foolish.

In practice, the interval of interest is often not explicitly given. This may be because the interval is implicitly described in the problem, or because determining this interval is part of the problem (e.g., determining where the solutions must "blow up"). It may also be because the person giving the differential equation is lazy or doesn't care what interval is used because the issue at hand is to find formulas that hold independently of the interval of interest.

In this text, if no interval of interest is given or hinted at, assume it to be any interval that makes sense. Often, this will be the entire real line, $(-\infty, \infty)$.

Solutions Over Intervals

In introducing the concept of the "interval of interest", we have implicitly refined our notion of "a (particular) solution to a differential equation". Let us make that refinement explicit: A *solution* to a differential equation over an interval of interest is a function that is both continuous and satisfies the differential equation over the given interval.

[6] T_{hit} is computed in Exercise 1.9 on page 19.

Recall that the *domain* of a function is the set of all numbers that can be plugged into the function. Naturally, if a function is a solution to a differential equation over some interval, then that function's domain must include that interval.[7]

Since we've refined our definition of particular solutions, we should make the corresponding refinement to our definition of a general solution. A *general solution* to a differential equation over an interval of interest is a formula or set of formulas describing all possible particular solutions over that interval.

Describing Particular Solutions

Let us get somewhat technical for a moment. Suppose we have a solution y to some differential equation over some interval of interest. Remember, we've defined y to be a "function". If you look up the basic definition of "function" in your calculus text, you'll find that, strictly speaking, y is a mapping of one set of numbers (the domain of y) onto another set of numbers (the range of y). This means that, for each value x in the function's domain, y assigns a corresponding number which we usually denote $y(x)$ and call "the value of y at x". If we are lucky, the function y is described by some formula, say, x^2. That is, the value of $y(x)$ can be determined for each x in the domain by the equation

$$y(x) = x^2 \quad .$$

Strictly speaking, the function y, its value at x (i.e., $y(x)$), and any formula describing how to compute $y(x)$ are different things. In everyday usage, however, the fine distinctions between these concepts are often ignored, and we say things like

$$\textit{consider the function } x^2 \qquad \text{or} \qquad \textit{consider } y = x^2$$

instead of the more correct statement

$$\textit{consider the function } y \textit{ where } y(x) = x^2 \textit{ for each } x \textit{ in the domain of } y \,.$$

For our purposes, "everyday usage" will usually suffice, and we won't worry that much about the differences between y, $y(x)$, and a formula describing y. This will save ink and paper, simplify the English, and, frankly, make it easier to follow many of our computations.

In particular, when we seek a particular solution to a differential equation, we will usually be quite happy to find a convenient formula describing the solution. We will then probably mildly abuse terminology by referring to that formula as "the solution". Please keep in mind that, in fact, any such formula is just one *description* of the solution — a very useful description since it tells you how to compute $y(x)$ for every x in the interval of interest. But other formulas can also describe the same function. For example, you can easily verify that

$$x^2 \qquad , \qquad (x+3)(x-3)+9 \qquad \text{and} \qquad \int_{t=0}^{x} 2t \, dt$$

are all formulas describing the same function on the real line.

There will also be differential equations for which we simply cannot find a convenient formula describing the desired solution (or solutions). On those occasions we will have to find some alternative way to describe our solutions. Some of these will involve using the differential equations to sketch approximations to the graphs of their solutions. Other alternative descriptions will involve formulas that approximate the solutions and allow us to generate lists of values approximating a solution at different points. These alternative descriptions may not be as convenient or as accurate as explicit formulas for the solutions, but they will provide usable information about the solutions.

[7] In theory, it makes sense to restrict the domain of a solution to the interval of interest so that irrelevant questions regarding the behavior of the function off the interval have no chance of arising. At this point of our studies, let us just be sure that a function serving as a solution makes sense at least over whatever interval we have interest in.

Additional Exercises

1.3. *For each differential equation given below, three choices for a possible solution* $y = y(x)$ *are given. Determine whether each choice is or is not a solution to the given differential equation. (In each case, assume the interval of interest is the entire real line* $(-\infty, \infty)$ *.)*

a. $\dfrac{dy}{dx} = 3y$

i. $y(x) = e^{3x}$ **ii.** $y(x) = x^3$ **iii.** $y(x) = \sin(3x)$

b. $x\dfrac{dy}{dx} = 3y$

i. $y(x) = e^{3x}$ **ii.** $y(x) = x^3$ **iii.** $y(x) = \sin(3x)$

c. $\dfrac{d^2y}{dx^2} = 9y$

i. $y(x) = e^{3x}$ **ii.** $y(x) = x^3$ **iii.** $y(x) = \sin(3x)$

d. $\dfrac{d^2y}{dx^2} = -9y$

i. $y(x) = e^{3x}$ **ii.** $y(x) = x^3$ **iii.** $y(x) = \sin(3x)$

e. $x\dfrac{dy}{dx} - 2y = 6x^4$

i. $y(x) = x^4$ **ii.** $y(x) = 3x^4$ **iii.** $y(x) = 3x^4 + 5x^2$

f. $\dfrac{d^2y}{dx^2} - 2x\dfrac{dy}{dx} - 2y = 0$

i. $y(x) = \sin(x)$ **ii.** $y(x) = x^3$ **iii.** $y(x) = e^{x^2}$

g. $\dfrac{d^2y}{dx^2} + 4y = 12x$

i. $y(x) = \sin(2x)$ **ii.** $y(x) = 3x$ **iii.** $y(x) = \sin(2x) + 3x$

h. $\dfrac{d^2y}{dx^2} - 6\dfrac{dy}{dx} + 9y = 0$

i. $y(x) = e^{3x}$ **ii.** $y(x) = xe^{3x}$ **iii.** $y(x) = 7e^{3x} - 4xe^{3x}$

1.4. *For each initial-value problem given below, three choices for a possible solution* $y = y(x)$ *are given. Determine whether each choice is or is not a solution to the given initial-value problem.*

a. $\dfrac{dy}{dx} = 4y$ *with* $y(0) = 5$

i. $y(x) = e^{4x}$ **ii.** $y(x) = 5e^{4x}$ **iii.** $y(x) = e^{4x} + 1$

b. $x\dfrac{dy}{dx} = 2y$ *with* $y(2) = 20$

i. $y(x) = x^2$ **ii.** $y(x) = 10x$ **iii.** $y(x) = 5x^2$

c. $\dfrac{d^2y}{dx^2} - 9y = 0$ *with* $y(0) = 1$ *and* $y'(0) = 9$

i. $y(x) = 2e^{3x} - e^{-3x}$ **ii.** $y(x) = e^{3x}$ **iii.** $y(x) = e^{3x} + 1$

d. $x^2\dfrac{d^2y}{dx^2} - 4x\dfrac{dy}{dx} + 6y = 36x^6$ *with* $y(1) = 1$ *and* $y'(1) = 12$

i. $y(x) = 10x^3 - 9x^2$ **ii.** $y(x) = 3x^6 - 2x^2$ **iii.** $y(x) = 3x^6 - 2x^3$

1.5. *For the following, let*

$$y(x) = \sqrt{x^2 + c}$$

where c *is an arbitrary constant.*

a. *Verify that this* y *is a solution to*

$$\frac{dy}{dx} = \frac{x}{y}$$

no matter what value c *is.*

b. *What value should* c *be so that the above* y *satisfies the initial condition*

i. $y(0) = 3$? **ii.** $y(2) = 3$?

c. *Using your results for the above, give a solution to each of the following initial-value problems:*

i. $\dfrac{dy}{dx} = \dfrac{x}{y}$ *with* $y(0) = 3$

ii. $\dfrac{dy}{dx} = \dfrac{x}{y}$ *with* $y(2) = 3$

1.6. *For the following, let*

$$y(x) = Ae^{x^2} - 3$$

where A *is an arbitrary constant.*

a. *Verify that this* y *is a solution to*

$$\frac{dy}{dx} - 2xy = 6x$$

no matter what value A *is.*

b. *In fact, it can be verified (using methods that will be developed later) that the above formula for* y *is a general solution to the above differential equation. Using this fact, finish solving each of the following initial-value problems:*

i. $\dfrac{dy}{dx} - 2xy = 6x$ *with* $y(0) = 1$

ii. $\dfrac{dy}{dx} - 2xy = 6x$ *with* $y(1) = 0$

1.7. *For the following, let*

$$y(x) = A\cos(2x) + B\sin(2x)$$

where A *and* B *are arbitrary constants.*

a. *Verify that this* y *is a solution to*

$$\frac{d^2y}{dx^2} + 4y = 0$$

no matter what values A *and* B *are.*

b. *Again, it can be verified that the above formula for* y *is a general solution to the above differential equation. Using this fact, finish solving each of the following initial-value problems:*

i. $\frac{d^2y}{dx^2} + 4y = 0$ *with* $y(0) = 3$ *and* $y'(0) = 8$

ii. $\frac{d^2y}{dx^2} + 4y = 0$ *with* $y(0) = 0$ *and* $y'(0) = 1$

1.8. *It was stated (on page 7) that "N^{th}-order sets of initial values are especially appropriate for N^{th}-order differential equations." The following problems illustrate one reason this is true. In particular, they demonstrate that, if* y *satisfies some N^{th}-order initial-value problem, then it automatically satisfies particular higher-order sets of initial values. Because of this, specifying the initial values for* $y^{(m)}$ *with* $m \geq N$ *is unnecessary and may even lead to problems with no solutions.*

a. *Assume* y *satisfies the first-order initial-value problem*

$$\frac{dy}{dx} = 3xy + x^2 \qquad \textit{with} \qquad y(1) = 2 \quad .$$

i. *Using the differential equation along with the given value for* $y(1)$, *determine what value* $y'(1)$ *must be.*

ii. *Is it possible to have a solution to*

$$\frac{dy}{dx} = 3xy + x^2$$

that also satisfies both $y(1) = 2$ *and* $y'(1) = 4$ *? (Give a reason.)*

iii. *Differentiate the given differential equation to obtain a second-order differential equation. Using the equation obtained along with the now known values for* $y(1)$ *and* $y'(1)$, *find the value of* $y''(1)$.

iv. *Can we continue and find* $y'''(1)$, $y^{(4)}(1)$, $\ldots$*?*

b. *Assume* y *satisfies the second-order initial-value problem*

$$\frac{d^2y}{dx^2} + 4\frac{dy}{dx} - 8y = 0 \qquad \textit{with} \qquad y(0) = 3 \quad \textit{and} \quad y'(0) = 5 \quad .$$

i. *Find the value of* $y''(0)$ *and of* $y'''(0)$

ii. *Is it possible to have a solution to*

$$\frac{d^2y}{dx^2} + 4\frac{dy}{dx} - 8y = 0$$

that also satisfies all of the following:

$$y(0) = 3 \quad , \quad y'(0) = 5 \quad \textit{and} \quad y'''(0) = 0 \quad ?$$

1.9. *Consider the simplest model we developed for a falling object (see page 10). In that, we derived*

$$y(t) = -4.9t^2 + 1{,}000$$

as the formula for the height y *above ground of some falling object at time* t.

a. *Find* T_{hit}, *the time the object hits the ground.*

b. *What is the velocity of the object when it hits the ground?*

c. *Suppose that, instead of being dropped at $t = 0$, the object is tossed up with an initial velocity of 2 meters per second. If this is the only change to our problem, then:*

i. *How does the corresponding initial-value problem change?*

ii. *What is the solution $y(t)$ to this initial value problem?*

iii. *What is the velocity of the object when it hits the ground?*

1.10. *Consider the "better" falling object model (see page 11), in which we derived the differential equation*

$$\frac{dv}{dt} = -9.8 - \kappa v \tag{1.5}$$

for the velocity. In this, κ is some positive constant used to describe the air resistance felt by the falling object.

a. *This differential equation was derived assuming the air was still. What differential equation would we have derived if, instead, we had assumed there was a steady updraft of 2 meters per second?*

b. *Recall that, from equation (1.5) we derived the equation*

$$v = -9.8t - \kappa y + c$$

relating the velocity v to the distance above ground y and the time t (see page 13). In the following, you will show that it, along with experimental data, can be used to determine the value of κ.

i. *Determine the value of the constant of integration, c, in the above equation using the given initial values (i.e., $y(0) = 1{,}000$ and $v(0) = 0$).*

ii. *Suppose that, in an experiment, the object was found to hit the ground at $t = T_{hit}$ with a speed of $v = v_{hit}$. Use this, along with the above equation, to find κ in terms of T_{hit} and v_{hit}.*

1.11. *For the following, let*

$$y(x) = Ax + Bx \ln|x|$$

where A and B are arbitrary constants.

a. *Verify that this y is a solution to*

$$x^2 \frac{d^2y}{dx^2} - x\frac{dy}{dx} + y = 0 \qquad \text{on the intervals } (0, \infty) \text{ and } (-\infty, 0) \quad ,$$

no matter what values A and B are.

b. *Again, we will later be able to show that the above formula for y is a general solution for the above differential equation. Given this, find the solution to the above differential equation satisfying $y(1) = 3$ and $y'(1) = 8$.*

c. *Why should your answer to **1.11 b** not be considered a valid solution to*

$$x^2 \frac{d^2y}{dx^2} - x\frac{dy}{dx} + y = 0$$

over the entire real line, $(-\infty, \infty)$?

2

Integration and Differential Equations

Often, when attempting to solve a differential equation, we are naturally led to computing one or more integrals — after all, integration is the inverse of differentiation. Indeed, we have already solved one simple second-order differential equation by repeated integration (the one arising in the simplest falling object model, starting on page 10). Let us now briefly consider the general case where integration is immediately applicable, and also consider some practical aspects of using both the indefinite integral and the definite integral.

2.1 Directly-Integrable Equations

We will say that a given first-order differential equation is *directly integrable* if (and only if) it can be (re)written as

$$\frac{dy}{dx} = f(x) \tag{2.1}$$

where $f(x)$ is some known function of just x (no y's). More generally, any N^{th}-order differential equation will be said to be *directly integrable* if and only if it can be (re)written as

$$\frac{d^N y}{dx^N} = f(x) \tag{2.1$'$}$$

where, again, $f(x)$ is some known function of just x (no y's or derivatives of y).

!►*Example 2.1:* *Consider the equation*

$$x^2 \frac{dy}{dx} - 4x = 6 \quad . \tag{2.2}$$

Solving this equation for the derivative:

$$x^2 \frac{dy}{dx} = 4x + 6$$

$$\hookrightarrow \qquad \frac{dy}{dx} = \frac{4x+6}{x^2} \quad .$$

Since the right-hand side of the last equation depends only on x *, we do have*

$$\frac{dy}{dx} = f(x) \qquad \left(\text{with } f(x) = \frac{4x+6}{x^2}\right) \quad .$$

So equation (2.2) is directly integrable.

!►Example 2.2: *Consider the equation*

$$x^2 \frac{dy}{dx} - 4xy = 6 \quad . \tag{2.3}$$

Solving this equation for the derivative:

$$x^2 \frac{dy}{dx} = 4xy + 6$$

$$\hookrightarrow \qquad \frac{dy}{dx} = \frac{4xy + 6}{x^2} \quad .$$

Here, the right-hand side of the last equation depends on both x and y, not just x. So equation (2.3) is not *directly integrable.*

Solving a directly-integrable equation is easy. First solve for the derivative to get the equation into form (2.1) or (2.1′), then integrate both sides as many times as needed to eliminate the derivatives, and, finally, do whatever simplification seems appropriate.

!►Example 2.3: *Again, consider*

$$x^2 \frac{dy}{dx} - 4x = 6 \quad . \tag{2.4}$$

In Example 2.1, we saw that it is directly integrable and can be rewritten as

$$\frac{dy}{dx} = \frac{4x + 6}{x^2} \quad .$$

Integrating both sides of this equation with respect to x (and doing a little algebra):

$$\int \frac{dy}{dx}\,dx = \int \frac{4x + 6}{x^2}\,dx \tag{2.5}$$

$$\begin{aligned} \hookrightarrow \qquad y(x) + c_1 &= \int \left[\frac{4}{x} + \frac{6}{x^2}\right] dx \\ &= 4\int x^{-1}\,dx + 6\int x^{-2}\,dx \\ &= 4\ln|x| + c_2 - 6x^{-1} + c_3 \end{aligned}$$

where c_1, c_2 and c_3 are arbitrary constants. Rearranging things slightly and letting $c = c_2 + c_3 - c_1$, this last equation simplifies to

$$y(x) = 4\ln|x| - 6x^{-1} + c \quad . \tag{2.6}$$

This is our general solution to differential equation (2.4). Since both $\ln|x|$ and x^{-1} are discontinuous just at $x = 0$, the solution can be valid over any interval not containing $x = 0$.

?►Exercise 2.1: *Consider the differential equation in Example 2.2 and explain why the y, which is an unknown function of x, makes it impossible to completely integrate both sides of*

$$\frac{dy}{dx} = \frac{4xy + 6}{x^2}$$

with respect to x.

2.2 On Using Indefinite Integrals

This is a good point to observe that, whenever we take the indefinite integrals of both sides of an equation, we obtain a bunch of arbitrary constants — c_1, c_2, ... (one constant for each integral) — that can be combined into a single arbitrary constant c. In the future, rather than note all the arbitrary constants that arise and how they combine into a single arbitrary constant c that is added to the right-hand side in the end, let us agree to simply add that c at the end. Let's not explicitly note all the intermediate arbitrary constants. If, for example, we had agreed to this before doing the last example, then we could have replaced all that material from equation (2.5) to equation (2.6) with

$$\begin{aligned} \int \frac{dy}{dx}\,dx &= \int \frac{4x+6}{x^2}\,dx \\ \hookrightarrow \qquad y(x) &= \int \left[\frac{4}{x} + \frac{6}{x^2}\right] dx \\ &= 4\int x^{-1}\,dx + 6\int x^{-2}\,dx \\ &= 4\ln|x| - 6x^{-1} + c \quad . \end{aligned}$$

This should simplify our computations a little.

This convention of "implicitly combining all the arbitrary constants" also allows us to write

$$y(x) = \int \frac{dy}{dx}\,dx \tag{2.7}$$

instead of

$$y(x) + \text{some arbitrary constant} = \int \frac{dy}{dx}\,dx \quad .$$

By our new convention, that "some arbitrary constant" is still in equation (2.7) — it's just been moved to the right-hand side of the equation and combined with the constants arising from the integral there.

Finally, like you, this author will get tired of repeatedly saying "where c is an arbitrary constant" when it is obvious that the c (or the c_1 or the A or …) that just appeared in the previous line is, indeed, some arbitrary constant. So let us not feel compelled to constantly repeat the obvious, and agree that, when a new symbol suddenly appears in the computation of an indefinite integral, then, yes, that is an arbitrary constant. Remember, though, to use different symbols for the different constants that arise when integrating a function already involving an arbitrary constant.

!►*Example 2.4:* *Consider solving*

$$\frac{d^2y}{dx^2} = 18x^2 \quad . \tag{2.8}$$

Clearly, this is directly integrable and will require two integrations. The first integration yields

$$\frac{dy}{dx} = \int \frac{d^2y}{dx^2}\,dx = \int 18x^2\,dx = \frac{18}{3}x^3 + c_1 \quad .$$

Cutting out the middle leaves

$$\frac{dy}{dx} = 6x^3 + c_1 \quad .$$

Integrating this, we have

$$y(x) = \int \frac{dy}{dx}\,dx = \int \left[6x^3 + c_1\right] dx = \frac{6}{4}x^4 + c_1 x + c_2 \quad .$$

So the general solution to equation (2.8) is

$$y(x) = \frac{3}{2}x^4 + c_1 x + c_2 \quad .$$

In practice, rather than use the same letter with different subscripts for different arbitrary constants (as we did in the above example), you might just want to use different letters, say, writing

$$y(x) = \frac{3}{2}x^4 + ax + b$$

instead of

$$y(x) = \frac{3}{2}x^4 + c_1 x + c_2 \quad .$$

This sometimes prevents dumb mistakes due to bad handwriting.

2.3 On Using Definite Integrals

Basic Ideas

We have been using the *indefinite* integral to recover $y(x)$ from ${}^{dy}/_{dx}$ via the relation

$$\int \frac{dy}{dx}\,dx = y(x) + c \quad .$$

Here, c is some constant (which we've agreed to automatically combine with other constants from other integrals).

We could just about as easily have used the corresponding *definite* integral relation

$$\int_a^x \frac{dy}{ds}\,ds = y(x) - y(a) \tag{2.9}$$

to recover $y(x)$ from its derivative. Note that, here, we've used s instead of x to denote the variable of integration. This prevents the confusion that can arise when using the same symbol for both the variable of integration *and* the upper limit in the integral. The lower limit, a, can be chosen to be any convenient value. In particular, if we are also dealing with initial values, then it makes sense to set a equal to the point at which the initial values are given. That way (as we will soon see) we will obtain a general solution in which the undetermined constant is simply the initial value.

Aside from getting it into the form

$$\frac{dy}{dx} = f(x) \quad ,$$

there are two simple steps that should be taken before using the definite integral to solve a first-order, directly-integrable differential equation:

1. Pick a convenient value for the lower limit of integration a. In particular, if the value of $y(x_0)$ is given for some point x_0, set $a = x_0$.

2. Rewrite the differential equation with s denoting the variable instead of x (i.e., replace x with s),

$$\frac{dy}{ds} = f(s) \quad . \tag{2.10}$$

After that, simply integrate both sides of equation (2.10) with respect to s from a to x:

$$\int_a^x \frac{dy}{ds}\,ds = \int_a^x f(s)\,ds$$

$$\hookrightarrow \qquad y(x) - y(a) = \int_a^x f(s)\,ds \quad .$$

Then solve for $y(x)$ by adding $y(a)$ to both sides,

$$y(x) = \int_a^x f(s)\,ds + y(a) \quad . \tag{2.11}$$

This is a general solution to the given differential equation. It should be noted that the integral here is a definite integral. Its evaluation does not lead to any arbitrary constants. However, the value of $y(a)$, until specified, can be anything; so $y(a)$ is the "arbitrary constant" in this general solution.

!►Example 2.5: *Consider solving the initial-value problem*

$$\frac{dy}{dx} = 3x^2 \qquad \text{with} \qquad y(2) = 12 \quad .$$

Since we know the value of $y(2)$, we will use 2 as the lower limit for our integrals. Rewriting the differential equation with s replacing x gives

$$\frac{dy}{ds} = 3s^2 \quad .$$

Integrating this with respect to s from 2 to x:

$$\int_2^x \frac{dy}{ds}\,ds = \int_2^x 3s^2\,ds$$

$$\hookrightarrow \qquad y(x) - y(2) = s^3\Big|_2^x = x^3 - 2^3 \quad .$$

Solving for $y(x)$ (and computing 2^3) then gives us

$$y(x) = x^3 - 8 + y(2) \quad .$$

This is a general solution to our differential equation. To find the particular solution that also satisfies $y(2) = 12$, as desired, we simply replace the $y(2)$ in the general solution with its given value,

$$\begin{aligned} y(x) &= x^3 - 8 + y(2) \\ &= x^3 - 8 + 12 = x^3 + 4 \quad . \end{aligned}$$

Of course, rather than go through the procedure just outlined to solve

$$\frac{dy}{dx} = f(x) \quad ,$$

we could, after determining a and $f(s)$, just plug these into equation (2.11),

$$y(x) = \int_a^x f(s)\,ds + y(a) \quad ,$$

and compute the integral. That is, after all, what we derived for any choice of f.

Advantages of Using Definite Integrals

By using definite integrals instead of indefinite integrals, we avoid dealing with arbitrary constants and end up with expressions explicitly involving initial values. This is sometimes convenient.

A much more important advantage of using definite integrals is that they result in concrete, computable formulas even when the corresponding indefinite integrals can*not* be evaluated. Let us look at a classic example.

!► *Example 2.6:* *Consider solving the initial-value problem*

$$\frac{dy}{dx} = e^{-x^2} \qquad \text{with} \qquad y(0) = 0 \quad .$$

In particular, determine the value of $y(x)$ when $x = 10$.

Using indefinite integrals yields

$$y(x) = \int \frac{dy}{dx}\,dx = \int e^{-x^2}\,dx \quad .$$

Unfortunately, this integral was not one you learned to evaluate in calculus.[1] *And if you check the tables, you will discover that no one else has discovered a usable formula for this integral. Consequently, the above formula for $y(x)$ is not very usable. Heck, we can't even isolate an arbitrary constant or see how the solution depends on the initial value.*

On the other hand, using definite integrals, we get

$$\int_0^x \frac{dy}{ds}\,ds = \int_0^x e^{-s^2}\,ds$$

$$\hookrightarrow \qquad y(x) - y(0) = \int_0^x e^{-s^2}\,ds$$

$$\hookrightarrow \qquad y(x) = \int_0^x e^{-s^2}\,ds + y(0) \quad .$$

This last formula explicitly describes how $y(x)$ depends on the initial value $y(0)$. Since we are assuming $y(0) = 0$, this reduces to

$$y(x) = \int_0^x e^{-s^2}\,ds \quad .$$

We still cannot find a computable formula for this integral, but, if we choose a specific value for x, say, $x = 10$, this expression becomes

$$y(10) = \int_0^{10} e^{-s^2}\,ds \quad .$$

The value of this integral can be very accurately approximated using any of a number of numerical integration methods such as the trapezoidal rule or Simpson's rule. In practice, of course, we'll just use the numerical integration command in our favorite computer math package (Maple, Mathematica, etc.). Using any such package, you will find that

$$y(10) = \int_0^{10} e^{-s^2}\,ds \approx 0.886 \quad .$$

[1] Well, you could expand e^{-x^2} in a Taylor series and integrate the series.

In one sense,

$$y(x) = \int f(x)\,dx \tag{2.12}$$

and

$$y(x) = \int_a^x f(s)\,ds + y(a) \tag{2.13}$$

are completely equivalent mathematical expressions. In practice, either can be used just about as easily, *provided* a reasonable formula for the indefinite integral in (2.12) can be found. If no such formula can be found, however, then expression (2.13) is much more useful because it can still be used, along with a numerical integration routine, to evaluate $y(x)$ for specific values of x. Indeed, one can compute $y(x)$ for a large number of values of x, plot each of these values of $y(x)$ against x, and thereby construct a very accurate approximation of the graph of y.

There are other ways to approximate solutions to differential equations, and we will discuss some of them. However, if you can express your solution in terms of definite integrals — even if the integral must be computed approximately — then it is usually best to do so. The other approximation methods for differential equations are typically more difficult to implement, and more likely to result in poor approximations.

Important "Named" Definite Integrals with Variable Limits

You should be familiar with a number of "named" functions (such as the natural logarithm and the arctangent) that can be given by definite integrals. For the two examples just cited,

$$\ln(x) = \int_1^x \frac{1}{s}\,ds \qquad \text{for} \quad x > 0$$

and

$$\arctan(x) = \int_0^x \frac{1}{1+s^2}\,ds \quad .$$

While $\ln(x)$ and $\arctan(x)$ can be defined independently of these integrals, their alternative definitions do not provide us with particularly useful ways to compute these functions by hand (unless x is something special, such as 1). Indeed, if you need the value of $\ln(x)$ or $\arctan(x)$ for, say, $x = 18$, then you are most likely to "compute" these values by having your calculator or computer or published tables[2] tell you the (approximate) value of $\ln(18)$ or $\arctan(18)$. Thus, for computational purposes, we might as well just view $\ln(x)$ and $\arctan(x)$ as names for the above integrals, and be glad that their values can easily be looked up electronically or in published tables.

It turns out that other integrals arise often enough in applications that workers dealing with these applications have decided to "name" these integrals, and to have their values tabulated. Two noteworthy "named integrals" are:

- The *error function*, denoted by erf and given by

$$\operatorname{erf}(x) = \int_0^x \frac{2}{\sqrt{\pi}} e^{-s^2}\,ds \quad .$$

- The *sine-integral function*, denoted by Si and given by[3]

$$\operatorname{Si}(x) = \int_0^x \frac{\sin(s)}{s}\,ds \quad .$$

---

[2] if you are an old-timer

[3] This integral is clearly mis-named since it is not the integral of the sine. In fact, the function being integrated, $\sin(x)/x$, is often called the "sinc" function (pronounced "sink"), so Si should really be called the "sinc-integral function". But nobody does.

Both of these are considered to be well-known functions, at least among certain groups of mathematicians, scientists and engineers. They (the functions, not the people) can be found in published tables and standard mathematical software (such as Maple or Mathematica) alongside such better-known functions as the natural logarithm and the trigonometric functions. Moreover, using tables or software, the value of $\operatorname{erf}(x)$ and $\operatorname{Si}(x)$ for any real value of x can be accurately computed just as easily as can the value of $\arctan(x)$. For these reasons, and because "$\operatorname{erf}(x)$" and "$\operatorname{Si}(x)$" take up less space than the integrals they represent, we will often follow the lead of others and use these function names instead of writing out the integrals.

!►*Example 2.7:* *In Example 2.6, above, we saw that the solution to*

$$\frac{dy}{dx} = e^{-x^2} \qquad \textit{with} \qquad y(0) = 0$$

is

$$y(x) = \int_0^x e^{-s^2}\, ds \quad .$$

Since this integral is the same as the integral for the error function with $^2/_{\sqrt{\pi}}$ *divided out, we can also express our answer as*

$$y(x) = \frac{\sqrt{\pi}}{2} \operatorname{erf}(x) \quad .$$

2.4 Integrals of Piecewise-Defined Functions

Computing the Integrals

Be aware that the functions appearing in differential equations can be piecewise defined, as in

$$\frac{dy}{dx} = f(x) \qquad \text{where} \quad f(x) = \begin{cases} x^2 & \text{if} \quad x < 2 \\ 1 & \text{if} \quad 2 \le x \end{cases} \quad .$$

Indeed, two such functions occur often enough that they have their own names: the *step function*, given by

$$\operatorname{step}(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad 0 \le x \end{cases} \quad ,$$

and the *ramp function*, given by

$$\operatorname{ramp}(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \le x \end{cases} \quad .$$

The reasons for these names should be obvious from their graphs (see Figure 2.1)

Such functions regularly arise when we attempt to model things reacting to discontinuous influences. For example, if $y(t)$ is the amount of energy produced up to time t by some light-sensitive device, and the rate at which this energy is produced depends proportionally on the intensity of the light received by the device, then

$$\frac{dy}{dt} = \operatorname{step}(t)$$

models the energy production of this device when it's kept in the dark until a light bulb (of unit intensity) is suddenly switched on at $t = 0$.

Computing the integrals of such functions is simply a matter of computing the integrals of the various "pieces", and then putting the integrated pieces together appropriately. While this can be done using indefinite integrals, each such integral introduces a new constant of integration which must then be related to the others to ensure that the final result describes a continuous function with a minimum number of arbitrary constants. On the other hand, the intelligent use of a definite integral eliminates the extra bookkeeping arising from an excessive number of "arbitrary" constant, and also ensures that the result is a continuous function (as required for solutions). Let's illustrate this by solving the differential equation given at the start of this section.

!► *Example 2.8 (using definite integrals):* *We seek a general solution to*

$$\frac{dy}{dx} = f(x) \qquad \text{where} \quad f(x) = \begin{cases} x^2 & \text{if } \ x < 2 \\ 1 & \text{if } \ 2 \le x \end{cases} \quad .$$

Taking the definite integral (starting, for no good reason, at 0*), we have*

$$y(x) = \int_0^x f(s)\,ds + y(0) \qquad \text{where} \quad f(s) = \begin{cases} s^2 & \text{if } \ s < 2 \\ 1 & \text{if } \ 2 \le s \end{cases} \quad .$$

Now, if $x \le 2$*, then* $f(s) = s^2$ *for every value of* s *in the interval* $(0, x)$*. So, when* $x \le 2$*,*

$$\int_0^x f(s)\,ds = \int_0^x s^2\,ds = \left.\frac{1}{3}s^3\right|_{s=0}^{x} = \frac{1}{3}x^3 \quad .$$

(Notice that this integral is valid for $x = 2$ *even though the formula used for* $f(s)$*,* s^2*, was only valid for* $s < 2$*.)*

On the other hand, if $2 < x$*, we must break the integral into two pieces, the one over* $(0, 2)$ *and the one over* $(2, x)$*:*

$$\begin{aligned}\int_0^x f(s)\,ds &= \int_0^2 f(s)\,ds + \int_2^x f(s)\,ds \\ &= \int_0^2 s^2\,ds + \int_2^x 1\,ds \\ &= \left.\frac{1}{3}s^3\right|_{s=0}^{2} + \left. s\right|_{s=2}^{x} \\ &= \left[\frac{1}{3}\cdot 2^3 - 0\right] + [x - 2] = x + \frac{2}{3} \quad .\end{aligned}$$

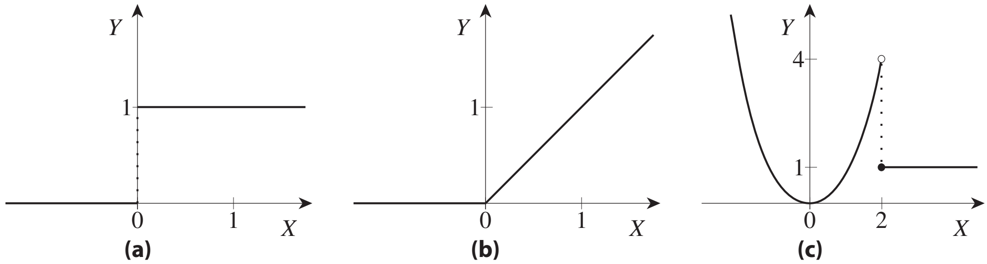


Figure 2.1: Three piecewise defined functions: **(a)** the step function, **(b)** the ramp function, **(c)** $f(x)$ from Example 2.8.

Thus, our general solution is

$$y(x) = \int_0^x f(s)\,ds + y(0) = \begin{cases} \frac{1}{3}x^3 + y(0) & \text{if} \quad x \le 2 \\ x + \frac{2}{3} + y(0) & \text{if} \quad 2 < x \end{cases} .$$

Keep in mind that solutions to differential equations are required to be continuous. After checking the above formulas, it should be obvious that the $y(x)$ obtained in the last example is continuous everywhere except, possibly, at $x = 2$. With a little work we could also verify that, in fact, we also have continuity at $x = 2$. But we won't bother because, in the next subsection, it will be seen that solutions so obtained via definite integration are guaranteed to be continuous, provided the discontinuities in the function being integrated are not too bad.

In practice, a given piecewise defined function may have more than two "pieces", and the differential equation may have order higher than one. For example, you may be called upon to solve

$$\frac{d^2y}{dx^2} = f(x) \qquad \text{where} \quad f(x) = \begin{cases} 0 & \text{if} \quad x < 1 \\ 1 & \text{if} \quad 1 \le x < 2 \\ 0 & \text{if} \quad 2 \le x \end{cases}$$

or even something involving infinitely many pieces, such as

$$\frac{d^4y}{dx^4} = \text{stair}(x) \qquad \text{where} \quad \text{stair}(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad 0 \le x < 1 \\ 2 & \text{if} \quad 2 \le x < 3 \\ 3 & \text{if} \quad 3 \le x < 4 \\ \vdots & \end{cases} . \tag{2.14}$$

The method illustrated in the last example can still be applied; you just have more integrals to keep track of.

Continuity of the Integrals

Theorem 2.1
Let f be a function on an interval (α, β) and let a be a point in that interval. Suppose, further, that f is continuous at all but, at most, a finite number of points in (α, β), and that, at each such point x_0 of discontinuity, the left- and right-hand limits

$$\lim_{x \to x_0^-} f(x) \qquad \textit{and} \qquad \lim_{x \to x_0^+} f(x)$$

exist (and are finite).[4] *Then the function given by*

$$g(x) = \int_a^x f(s)\,ds$$

is continuous on (α, β).

[4] Such discontinuities are said to be *finite-jump* discontinuities.

PROOF: First of all, note that the two requirements placed on f ensure

$$g(x) = \int_a^x f(s)\,ds$$

is well defined for any x in (α, β) using any of the definitions for the integral found in most calculus texts (check this out yourself, using the definition in your calculus text). They also prevent $f(x)$ from "blowing up" on any closed subinterval $[\alpha', \beta']$ of (α, β). Thus, for each such closed subinterval $[\alpha', \beta']$, there is a corresponding finite constant M such that[5]

$$|f(s)| \leq M \qquad \text{whenever} \quad \alpha' \leq s \leq \beta' \quad .$$

Now, to verify the claimed continuity of g, we must show that

$$\lim_{x \to x_0} g(x) = g(x_0) \tag{2.15}$$

for any x_0 in (α, β). But by the definition of g and well-known properties of integration,

$$\begin{aligned} \lim_{x \to x_0} g(x) &= \lim_{x \to x_0} \int_a^x f(s)\,ds \\ &= \lim_{x \to x_0} \left[\int_a^{x_0} f(s)\,ds + \int_{x_0}^x f(s)\,ds \right] \\ &= \lim_{x \to x_0} \left[g(x_0) + \int_{x_0}^x f(s)\,ds \right] = g(x_0) + \lim_{x \to x_0} \int_{x_0}^x f(s)\,ds \quad . \end{aligned}$$

So, to show equation (2.15) holds, it suffices to confirm that

$$\lim_{x \to x_0} \int_{x_0}^x f(s)\,ds = 0 \quad ,$$

which, in turn, is equivalent to confirming that

$$\lim_{x \to x_0^+} \left| \int_{x_0}^x f(s)\,ds \right| = 0 \qquad \text{and} \qquad \lim_{x \to x_0^-} \left| \int_{x_0}^x f(s)\,ds \right| = 0 \quad . \tag{2.16}$$

To do this, pick any two finite values α' and β' satisfying $\alpha < \alpha' < x_0 < \beta' < \beta$. As noted, there is some finite constant $M \geq |f(s)|$ on $[\alpha', \beta']$. So, if $x_0 \leq x \leq \beta'$,

$$0 \leq \left| \int_{x_0}^x f(s)\,ds \right| \leq \int_{x_0}^x |f(s)|\,ds \leq \int_{x_0}^x M\,ds = M[x - x_0] \quad .$$

Similarly, if $\alpha' < x < x_0$, then

$$\begin{aligned} 0 \leq \left| \int_{x_0}^x f(s)\,ds \right| = \left| -\int_x^{x_0} f(s)\,ds \right| &= \left| \int_x^{x_0} f(s)\,ds \right| \\ &\leq \int_x^{x_0} |f(s)|\,ds \leq \int_x^{x_0} M\,ds = M[x_0 - x] \quad . \end{aligned}$$

Hence,

$$0 \leq \lim_{x \to x_0^+} \left| \int_{x_0}^x f(s)\,ds \right| \leq \lim_{x \to x_0^+} M[x - x_0] = M[x_0 - x_0] = 0$$

and

$$0 \leq \lim_{x \to x_0^-} \left| \int_{x_0}^x f(s)\,ds \right| \leq \lim_{x \to x_0^-} M[x_0 - x] = M[x_0 - x_0] = 0 \quad ,$$

which, of course, means that equation set (2.16) holds.

---

[5] The constant M can be the maximum value of $|f(s)|$ on $[\alpha', \beta']$, provided that maximum exists. It may change if either endpoint α' or β' is changed.

Additional Exercises

2.2. *Determine whether each of the following differential equations is or is not directly integrable:*

a. $\dfrac{dy}{dx} = 3 - \sin(x)$

b. $\dfrac{dy}{dx} = 3 - \sin(y)$

c. $\dfrac{dy}{dx} + 4y = e^{2x}$

d. $x\dfrac{dy}{dx} = \arcsin(x^2)$

e. $y\dfrac{dy}{dx} = 2x$

f. $\dfrac{d^2y}{dx^2} = \dfrac{x+1}{x-1}$

g. $x^2\dfrac{d^2y}{dx^2} = 1$

h. $y^2\dfrac{d^2y}{dx^2} = 8x^2$

i. $\dfrac{d^2y}{dx^2} + 3\dfrac{dy}{dx} + 8y = e^{-x^2}$

j. $x^2\dfrac{d^2y}{dx^2} + 3x\dfrac{dy}{dx} = 0$

2.3. *Find a general solution for each of the following directly integrable equations. (Use indefinite integrals on these.)*

a. $\dfrac{dy}{dx} = 4x^3$

b. $\dfrac{dy}{dx} = 20e^{-4x}$

c. $x\dfrac{dy}{dx} + \sqrt{x} = 2$

d. $\sqrt{x+4}\,\dfrac{dy}{dx} = 1$

e. $\dfrac{dy}{dx} = x\cos\left(x^2\right)$

f. $\dfrac{dy}{dx} = x\cos(x)$

g. $x = \left(x^2 - 9\right)\dfrac{dy}{dx}$

h. $1 = \left(x^2 - 9\right)\dfrac{dy}{dx}$

i. $1 = x^2 - 9\dfrac{dy}{dx}$

j. $\dfrac{d^2y}{dx^2} = \sin(2x)$

k. $\dfrac{d^2y}{dx^2} - 3 = x$

l. $\dfrac{d^4y}{dx^4} = 1$

2.4. *Solve each of the following initial-value problems (using the indefinite integral). Also, state the largest interval over which the solution is valid (i.e., the maximal possible interval of interest).*

a. $\dfrac{dy}{dx} = 4x + 10e^{2x}$ *with* $y(0) = 4$

b. $\sqrt[3]{x+6}\,\dfrac{dy}{dx} = 1$ *with* $y(2) = 10$

c. $\dfrac{dy}{dx} = \dfrac{x-1}{x+1}$ *with* $y(0) = 8$

d. $x\dfrac{dy}{dx} + 2 = \sqrt{x}$ *with* $y(1) = 6$

e. $\cos(x)\dfrac{dy}{dx} - \sin(x) = 0$ *with* $y(0) = 3$

f. $\left(x^2+1\right)\dfrac{dy}{dx} = 1$ *with* $y(0) = 3$

g. $x \dfrac{d^2y}{dx^2} + 2 = \sqrt{x}$ *with* $y(1) = 8$ *and* $y'(1) = 6$

2.5 a. *Using definite integrals (as in Example 2.5 on page 25), find the general solution to*

$$\frac{dy}{dx} = \sin\left(\frac{x}{2}\right)$$

with $y(0)$ *acting as the arbitrary constant.*

b. *Using the formula just found for* $y(x)$ *:*

i. *Find* $y(\pi)$ *when* $y(0) = 0$.

ii. *Find* $y(\pi)$ *when* $y(0) = 3$.

iii. *Find* $y(2\pi)$ *when* $y(0) = 3$.

2.6 a. *Using definite integrals (as in Example 2.5 on page 25), find the general solution to*

$$\frac{dy}{dx} = 3\sqrt{x+3}$$

with $y(1)$ *acting as the arbitrary constant.*

b. *Using the formula just found for* $y(x)$ *:*

i. *Find* $y(6)$ *when* $y(1) = 16$.

ii. *Find* $y(6)$ *when* $y(1) = 20$.

iii. *Find* $y(-2)$ *when* $y(1) = 0$.

2.7. *Using definite integrals (as in Example 2.5 on page 25), find the solution to each of the following initial-value problems. (In some cases, you may want to use the error function or the sine-integral function.)*

a. $\dfrac{dy}{dx} = x\,e^{-x^2}$ *with* $y(0) = 3$

b. $\dfrac{dy}{dx} = \dfrac{x}{\sqrt{x^2+5}}$ *with* $y(2) = 7$

c. $\dfrac{dy}{dx} = \dfrac{1}{x^2+1}$ *with* $y(1) = 0$

d. $\dfrac{dy}{dx} = e^{-9x^2}$ *with* $y(0) = 1$

e. $x\dfrac{dy}{dx} = \sin(x)$ *with* $y(0) = 4$

f. $x\dfrac{dy}{dx} = \sin\left(x^2\right)$ *with* $y(0) = 0$

2.8. *Using an appropriate computer math package (such as Maple or Mathematica), graph each of the following over the interval* $0 \leq x \leq 10$ *:*

a. *the error function,* $\operatorname{erf}(x)$.

b. *the sine integral function,* $\operatorname{Si}(x)$.

c. *the solution to*

$$\frac{dy}{dx} = \ln\left|2 + x^2\sin(x)\right| \quad \textit{with} \quad y(0) = 0 \quad .$$

2.9. *Each of the following differential equations involves a function that is (or can be) piecewise defined. Sketch the graph of each of these piecewise defined functions, and find the general solution of each differential equation. If an initial value is also given, then also solve the given initial-value problem:*

a. $\dfrac{dy}{dx} = \operatorname{step}(x)$ *with* $y(0) = 0$ *and* $\operatorname{step}(x)$ *as defined on page 28*

b. $\dfrac{dy}{dx} = f(x)$ *with* $y(0) = 2$ *and* $f(x) = \begin{cases} 0 & \textit{if} \quad x < 1 \\ 1 & \textit{if} \quad 1 \leq x \end{cases}$

c. $\dfrac{dy}{dx} = f(x)$ *with* $y(0) = 0$ *and* $f(x) = \begin{cases} 0 & \textit{if} \quad x < 1 \\ 1 & \textit{if} \quad 1 \le x < 2 \\ 0 & \textit{if} \quad 2 \le x \end{cases}$

d. $\dfrac{dy}{dx} = |x - 2|$

e. $\dfrac{dy}{dx} = \text{stair}(x)$ *for* $x < 4$ *with* $y(0) = 0$ *and* $\text{stair}(x)$ *as defined on page 30*

Part II

First-Order Equations

3

Some Basics about First-Order Equations

For the next few chapters, our attention will be focused on first-order differential equations. We will discover that these equations can often be solved using methods developed directly from the tools of elementary calculus. And even when these equations cannot be explicitly solved, we will still be able to use fundamental concepts from elementary calculus to obtain good approximations to the desired solutions.

But first, let us discuss a few basic ideas that will be relevant throughout our discussion of first-order differential equations.

3.1 Algebraically Solving for the Derivative

Here are some of the first-order differential equations that we have seen or will see in the next few chapters:

$$x^2 \frac{dy}{dx} - 4x = 6 \quad ,$$

$$\frac{dy}{dx} - x^2 y^2 = x^2 \quad ,$$

$$\frac{dy}{dx} + 4xy = 2xy^2$$

and

$$x \frac{dy}{dx} + 4y = x^3 \quad .$$

One thing we can do with each of these equations is to algebraically solve for the derivative. Doing this with the first equation:

$$x^2 \frac{dy}{dx} - 4x = 6$$

$$\hookrightarrow \qquad x^2 \frac{dy}{dx} = 6 + 4x$$

$$\hookrightarrow \qquad \frac{dy}{dx} = \frac{4x + 6}{x^2} \quad .$$

For the second equation:

$$\frac{dy}{dx} - x^2 y^2 = x^2$$

$$\hookrightarrow \qquad \frac{dy}{dx} = x^2 + x^2 y^2 \quad .$$

Solving for the derivative is often a good first step towards solving a first-order differential equation. For example, the first equation above is directly integrable — solving for the derivative yielded

$$\frac{dy}{dx} = \frac{4x+6}{x^2} \quad ,$$

and $y(x)$ can now be found by simply integrating both sides with respect to x.

Even when the equation is not directly integrable and we get

$$\frac{dy}{dx} = \text{“a formula of both } x \text{ and } y\text{”} \quad ,$$

— as in our second equation above,

$$\frac{dy}{dx} = x^2 + x^2y^2$$

— that formula on the right can still give us useful information about the possible solutions and can help us determine which method is appropriate for obtaining the general solution. Observe, for example, that the right-hand side of the last equation can be factored into a formula of x and a formula of y,

$$\frac{dy}{dx} = x^2\left(1+y^2\right) \quad .$$

In the next chapter, we will find that this means the equation is “separable” and can be solved by a procedure developed for just such equations.

For convenience, let us say that a first-order differential equation is in *derivative formula form* if it is written as

$$\frac{dy}{dx} = F(x,y) \tag{3.1}$$

where $F(x,y)$ is some (known) formula of x and/or y. Remember, to convert a given first-order differential equation to derivative form, simply use a little algebra to solve the differential equation for the derivative.

?►Exercise 3.1: *Verify that the derivative formula forms of*

$$\frac{dy}{dx} + 4y = 3y^3 \qquad \textit{and} \qquad x\frac{dy}{dx} + 4xy = 2y^2$$

are

$$\frac{dy}{dx} = 3y^3 - 4y \qquad \textit{and} \qquad \frac{dy}{dx} = \frac{2y^2 - 4xy}{x} \quad ,$$

respectively.

Keep in mind that the right side of equation (3.1), $F(x,y)$, need not always be a formula of *both* x and y. As we saw in an example above, the equation might be directly integrable. In this case, the right side of the above derivative formula form reduces to some $f(x)$, a formula involving only x,

$$\frac{dy}{dx} = f(x) \quad .$$

Alternatively, the right side may end up being a formula involving only y, $F(x,y) = g(y)$. We have a word for such differential equations; that word is “autonomous”. That is, an *autonomous* first-order differential equation is a differential equation that can be written as

$$\frac{dy}{dx} = g(y)$$

where $g(y)$ is some formula involving y but not x. The first equation in the last exercise is an example of an autonomous differential equation. Autonomous equations arise fairly often in applications, and the fact that dy/dx is given by a formula of just y will make an autonomous equation easier to graphically analyze in Chapter 9. But, as we'll see in the next chapter, they are just special cases of "separable" equations, and can be solved using the methods that will be developed there.

You should also be aware that the derivative formula form is not the only way we will attempt to rewrite our first-order differential equations. Frankly, much of the theory for first-order differential equations involves determining how a given differential equation can be rewritten so that we can cleverly apply tricks from calculus to further reduce the equation to something that can be easily integrated. We've already seen this with directly-integrable differential equations (for which the "derivative formula" form is ideal). In the next few chapters, we will see this with other equations for which other forms are useful.

By the way, there are first-order differential equations that cannot be put in derivative formula form. Consider

$$\frac{dy}{dx} + \sin\left(\frac{dy}{dx}\right) = x \quad .$$

It can be safely said that solving this equation for dy/dx is beyond the algebraic skills of most mortals. Fortunately, first-order differential equations that cannot be rewritten in the derivative formula form rarely arise in real-world applications.

3.2 Constant (or Equilibrium) Solutions

There is one type of particular solution that is easily determined for many first-order differential equations using elementary algebra: the "constant" solution.

A *constant solution* to a given differential equation is simply a constant function that satisfies that differential equation. Remember, y is a *constant function* if its value, $y(x)$, is some fixed constant for all x; that is, for some single number y_0,

$$y(x) = y_0 \qquad \text{for all} \quad x \quad .$$

Such solutions are also sometimes called *equilibrium solutions*. In an application involving some process that can vary with x, these solutions describe situations in which the process does not vary with x. This often means that all the factors influencing the process are "balancing out", leaving the process in a "state of equilibrium". As we will later see, this sometimes means that these solutions — whether called constant or equilibrium — are the most important solutions to a given differential equation.[1]

!►*Example 3.1:* *Consider the differential equation*

$$\frac{dy}{dx} = 2xy^2 - 4xy$$

and the constant function

$$y(x) = 2 \qquad \textit{for all} \quad x \quad .$$

[1] According to mathematical tradition, one only refers to a constant solution as an "equilibrium solution" if the differential equation is autonomous.

Since the derivative of a constant function is zero, plugging in this function, $y = 2$ into

$$\frac{dy}{dx} = 2xy^2 - 4xy$$

gives

$$0 = 2x \cdot 2^2 - 4x \cdot 2 \quad ,$$

which, after a little arithmetic and algebra, reduces further to

$$0 = 0 \quad .$$

Hence, our constant function satisfies our differential equation, and, so, is a constant solution to that differential equation.

On the other hand, plugging the constant function

$$y(x) = 3 \qquad \text{for all} \quad x$$

into

$$\frac{dy}{dx} = 2xy^2 - 4xy$$

gives

$$0 = 2x \cdot 3^2 - 4x \cdot 3 \quad .$$

This only reduces to

$$0 = 6x \quad ,$$

which is not valid for all values of x on any nontrivial interval. Thus, $y = 3$ is not a constant solution to our differential equation.

Admittedly, constant functions are not usually considered particularly exciting. The graph of a constant function,

$$y(x) = y_0 \qquad \text{for all} \quad x$$

is just a horizontal line (at $y = y_0$), and its derivative (as noted in the above example) is zero. But the fact that its derivative is zero is what simplifies the task of finding all possible constant solutions to a given differential equation, especially if the equation is in derivative formula form. After all, if we plug a constant function

$$y(x) = y_0 \qquad \text{for all} \quad x$$

into an equation of the form

$$\frac{dy}{dx} = F(x, y) \quad ,$$

then, since the derivative of a constant is zero, this equation reduces to

$$0 = F(x, y_0) \quad .$$

We can then determine all values y_0 that make $y = y_0$ a constant solution for our differential equation by simply determining every constant y_0 that satisfies

$$F(x, y_0) = 0 \qquad \text{for all} \quad x \quad .$$

!►Example 3.2: *Suppose we have a differential equation that, after a bit of algebra, can be written as*

$$\frac{dy}{dx} = (y - 2x)\left(y^2 - 9\right) \quad .$$

If it has a constant solution,

$$y(x) = y_0 \qquad \text{for all} \quad x \quad ,$$

then, after plugging this simple formula for y into the differential equation (and remembering that the derivative of a constant is zero), we get

$$0 = (y_0 - 2x)\left({y_0}^2 - 9\right) \quad , \tag{3.2}$$

which is possible if and only if either

$$y_0 - 2x = 0 \qquad \text{or} \qquad {y_0}^2 - 9 = 0 \quad .$$

Now,

$$y_0 - 2x = 0 \quad \Longleftrightarrow \quad y_0 = 2x \quad .$$

This gives us a value for y_0 that varies with x, contradicting the original assumption that y_0 was a constant. So this does not lead to any constant solutions (or any other solutions, either!). If there is such a solution, $y = y_0$, it must satisfy the other equation,

$${y_0}^2 - 9 = 0 \quad .$$

But

$${y_0}^2 - 9 = 0 \quad \Longleftrightarrow \quad {y_0}^2 = 9 \quad \Longleftrightarrow \quad y_0 = \pm\sqrt{9} = \pm 3 \quad .$$

So there are exactly two constant values for y_0, 3 and -3, that satisfy equation (3.2). And thus, our differential equation has exactly two constant (or equilibrium) solutions,

$$y(x) = 3 \qquad \text{for all} \quad x$$

and

$$y(x) = -3 \qquad \text{for all} \quad x \quad .$$

Keep in mind that, while the constant solutions to a given differential equation may be important, they rarely are the only solutions. And in practice, the solution to a given initial-value problem will typically *not* be one of the constant solutions. However, as we will see later, one of the constant solutions may tell us something about the long-term behavior of the solution to that particular initial-value problem. That is one of the reasons constant solutions are so important.

You should also realize that many differential equations have no constant solutions. Consider, for example, the directly-integrable differential equation

$$\frac{dy}{dx} = 2x \quad .$$

Integrating this, we get the general solution

$$y(x) = x^2 + c \quad .$$

No matter what value we pick for c, this function varies as x varies. It cannot be a constant.

In fact, it is not hard to see that no directly-integrable differential equation,

$$\frac{dy}{dx} = f(x) \quad ,$$

can have a constant solution (unless $f \equiv 0$). Just consider what you get when you integrate the $f(x)$. (That's why we did not mention such solutions when we discussed directly-integrable equations.)

3.3 On the Existence and Uniqueness of Solutions

Unfortunately, not all problems are solvable, and those that are solvable sometimes have several solutions. This is true in mathematics just as it is true in real life.

Before attempting to solve a problem involving some given differential equation and auxiliary condition (such as an initial value), it would certainly be nice to know that the given differential equation actually has a solution satisfying the given auxiliary condition. This would be especially true if the given differential equation looks difficult and we expect that considerable effort will be required in solving it (effort which would be wasted if that solution did not exist). And even if we can find a solution, we normally would like some assurance that it is the only solution.

The following theorem is the standard theorem quoted in most elementary differential equation texts addressing these issues for fairly general first-order initial-value problems.

Theorem 3.1 (on existence and uniqueness)
Consider a first-order initial-value problem

$$\frac{dy}{dx} = F(x, y) \qquad \text{with} \quad y(x_0) = y_0$$

in which both F and $\partial F/\partial y$ are continuous functions on some open region of the XY–plane containing the point (x_0, y_0).[2] *The initial-value problem then has exactly one solution over some open interval (α, β) containing x_0. Moreover, this solution and its derivative are continuous over that interval.*

This theorem assures us that, if we can write a first-order differential equation in the derivative formula form,

$$\frac{dy}{dx} = F(x, y) \quad ,$$

and that $F(x, y)$ is a 'reasonably well-behaved' formula on some region of interest, then our differential equation has solutions — with luck and skill, we will be able to find them. Moreover, if we can find a solution to this equation that also satisfies some initial value $y(x_0) = y_0$ corresponding to a point at which F is 'reasonably well-behaved', then that solution is unique (i.e., it is the only solution) — there is no need to worry about alternative solutions — at least over some interval (α, β). Just what that interval (α, β) is, however, is not explicitly described in this theorem. It turns out to depend in subtle ways on just how well behaved $F(x, y)$ is. More will be said about this in a few paragraphs.

!►*Example 3.3:* *Consider the initial-value problem*

$$\frac{dy}{dx} - x^2y^2 = x^2 \qquad \text{with} \quad y(0) = 3 \quad .$$

As derived earlier, the derivative formula form for this equation is

$$\frac{dy}{dx} = x^2 + x^2y^2 \quad .$$

So

$$F(x, y) = x^2 + x^2y^2$$

[2] The $\partial F/\partial y$ is a "partial derivative". If you are not acquainted with partial derivatives see the appendix on page 59.

and

$$\frac{\partial F}{\partial y} = \frac{\partial}{\partial y}\left[x^2 + x^2y^2\right] = 0 + x^2 2y = 2x^2 y \quad .$$

It should be clear that these two functions are continuous everywhere on the XY–plane. Hence, we can take the entire plane to be that "open region" in the above theorem, which then assures us that the above initial-value problem has one (and only one) solution valid over some interval (a, b) with $a < 0 < b$. Unfortunately, the theorem doesn't tell us what that solution is nor what that interval (a, b) might be. We will have to wait until we develop a method for solving this differential equation.

The proof of the above theorem is nontrivial and can be safely skipped by most beginning readers. In fact, despite the importance of the above theorem, we will rarely explicitly refer to it in the chapters that follow. The main explicit references will be a "graphical" discussion of the theorem in Chapter 9 using methods developed there,[3] and to note that analogous theorems can be proven for higher-order differential equations. Nonetheless, it is an important theorem whose proof should be included in this text if only to assure you that the author is not making it up. Besides, the basic core of the proof is fairly accessible to most readers and contains some clever and interesting ideas. We will go over that basic core in the next section (Section 3.4), leaving the more challenging details for the section after that (Section 3.5).

Part of the proof will be to identify the interval (α, β) mentioned in the above theorem. In fact, the interval (α, β) can be easily determined if F and ${}^{\partial F}/_{\partial y}$ are sufficiently well behaved. That is what the next theorem gives us. Its proof requires just a few modifications of the proof of the above, and will be briefly discussed after that proof.

Theorem 3.2

Consider a first-order initial-value problem

$$\frac{dy}{dx} = F(x, y) \qquad \text{with} \quad y(x_0) = y_0$$

over an interval (α, β) containing x_0, and with $F = F(x, y)$ being a continuous function on the infinite strip

$$\mathcal{R} = \{(x, y) : \alpha < x < \beta \text{ and } -\infty < y < \infty\} \quad .$$

Further suppose that, on $\mathcal{R}$, the partial derivative ${}^{\partial F}/_{\partial y}$ is continuous and is a function of x only.[4] *Then the initial-value problem has exactly one solution over (α, β). Moreover, this solution and its derivative are continuous on that interval.*

In practice, many of our first-order differential equations will not satisfy the conditions described in the last theorem. So this theorem is of relatively limited value for now. However, it leads to higher-order analogs that will be used in developing the theory needed for important classes of higher-order differential equations. That is why Theorem 3.2 is mentioned here.

[3] which you may find more illuminating than the proof given here

[4] More generally, the theorem remains true if we replace the phrase "a function of x only" with "a bounded function on $\mathcal{R}$". Our future interest, however, will be with the theorem as stated.

3.4 Confirming the Existence of Solutions (Core Ideas)

So let us consider the first-order initial-value problem

$$\frac{dy}{dx} = F(x, y) \qquad \text{with} \quad y(x_0) = y_0 \quad ,$$

assuming that both F and ${}^{\partial F}/_{\partial y}$ are continuous on some open region in the XY–plane containing the point (x_0, y_0). Our goal is to verify that a solution y exists over some interval. (This is the existence claim of Theorem 3.1. The uniqueness claim of that theorem will be left as an exercise using material developed in the next section — see Exercise 3.2 on page 56.)

The gist of our proof consists of three steps:

1. Observe that the initial-value problem is equivalent to a corresponding integral equation.
2. Derive a sequence of functions — ψ_0, ψ_1, ψ_2, ψ_3, $\ldots$ — using a formula inspired by that integral equation.
3. Show that this sequence of functions converges on some interval to a solution y of the original initial-value problem.

The "hard" part of the proof is in the details of the last step. We can skip over these details initially, returning to them in the next section.

Two comments should be made here:

1. The ψ_k's end up being approximations to the solution y, and, in theory at least, the method we are about to describe can be used to find approximate solutions to an initial-value problem. Other methods, however, are often more practical.
2. This method was developed by the French mathematician Emile Picard and is often referred to as the *(Picard's) method of successive approximations* or as *Picard's iterative method* (because of the way the ψ_k's are generated).

To simplify discussion let us assume $x_0 = 0$, so that our initial-value problem is

$$\frac{dy}{dx} = F(x, y) \qquad \text{with} \quad y(0) = y_0 \quad . \tag{3.3}$$

There is no loss of generality here. After all, if $x_0 \neq 0$, we can apply the change of variable $s = x - x_0$ and convert our original problem into problem (3.3) (with x replaced by s).

Converting to an Integral Equation

Suppose $y = y(x)$ is a solution to initial-value problem (3.3) on some interval (α, β) with $\alpha < 0 < \beta$. Renaming x as s, our differential equation becomes

$$\frac{dy}{ds} = F(s, y(s)) \qquad \text{for each } s \text{ in } (\alpha, \beta) \quad .$$

Integrating this from 0 to any x in (α, β) and remembering that $y(0) = y_0$, we get

$$\int_0^x \frac{dy}{ds}\, ds = \int_0^x F(s, y(s))\, ds$$

$$\hookrightarrow \qquad y(x) - y(0) = \int_0^x F(s, y(s))\, ds$$

$$\hookrightarrow \qquad y(x) - y_0 = \int_0^x F(s, y(s))\, ds \quad .$$

That is, y satisfies the integral equation

$$y(x) = y_0 + \int_0^x F(s, y(s))\, ds \qquad \text{whenever} \quad \alpha < x < \beta \quad .$$

On the other hand, if y is any continuous function on (α, β) satisfying this integral equation, then basic calculus tells us that, on this interval, y is differentiable with

$$\frac{dy}{dx} = \frac{d}{dx}\left[y_0 + \int_0^x F(s, y(s))\, ds\right] = 0 + \frac{d}{dx}\int_0^x F(s, y(s))\, ds = F(x, y(x)) \quad .$$

and

$$y(0) = y_0 + \underbrace{\int_0^0 F(s, y(s))\, ds}_{0} = y_0 \quad .$$

Thus, y also satisfies our original initial-value problem.

We should note that, in the above, we implicitly assumed $F(x, y)$ was a reasonably behaved function at each point (x, y) where $\alpha < x < \beta$ and $y = y(x)$. In particular, if F is continuous at each of these points, then this continuity, the continuity of y, and the fact that $y' = F(x, y)$ ensures that y is not only differentiable on (α, β) but that y' is continuous on (α, β).

In summary, we have the following theorem:

Theorem 3.3
Let y be any continuous function on some interval (α, β) containing 0, and assume F is a function of two variables continuous at every (x, y) with $\alpha < x < \beta$ and $y = y(x)$. Then y has a continuous derivative on (α, β) and satisfies the initial-value problem

$$\frac{dy}{dx} = F(x, y) \qquad \textit{with} \quad y(0) = y_0 \quad \textit{on} \quad (\alpha, \beta)$$

if and only if y satisfies the integral equation

$$y(x) = y_0 + \int_0^x F(s, y(s))\, ds \qquad \textit{whenever} \quad \alpha < x < \beta \quad .$$

Generating a Sequence of "Approximate Solutions"

Begin with any continuous function ψ_0. For example, we could simply choose ψ_0 to be the constant function

$$\psi_0(x) = y_0 \qquad \text{for all} \quad x \quad .$$

(Later, we will place some additional restrictions on ψ_0, but the above constant function will still be a valid choice for ψ_0.)

Next, let ψ_1 be the function constructed from ψ_0 by

$$\psi_1(x) = y_0 + \int_0^x F(s, \psi_0(s))\, ds \quad .$$

Then construct ψ_2 from ψ_1 via

$$\psi_2(x) = y_0 + \int_0^x F(s, \psi_1(s))\, ds \quad .$$

Continue the process, defining ψ_3 , ψ_4 , ψ_5 , ... by

$$\psi_3(x) = y_0 + \int_0^x F(s, \psi_2(s))\,ds \quad ,$$

$$\psi_4(x) = y_0 + \int_0^x F(s, \psi_3(s))\,ds \quad ,$$

$$\vdots$$

In general, once ψ_k is defined, we define ψ_{k+1} by

$$\psi_{k+1}(x) = y_0 + \int_0^x F(s, \psi_k(s))\,ds \quad . \tag{3.4}$$

Since we apparently can continue this iterative process forever, we have an infinite sequence of functions

$$\psi_0 \quad , \quad \psi_1 \quad , \quad \psi_2 \quad , \quad \psi_3 \quad , \quad \psi_4 \quad , \quad \ldots \quad .$$

In the future, we may refer to this sequence as the *Picard sequence* (based on ψ_0 and F). Note that, for $k = 1, 2, 3, \ldots$,

$$\psi_k(0) = y_0 + \underbrace{\int_0^0 F(s, \psi_{k-1}(s))\,ds}_{0} = y_0 \quad .$$

So each of these ψ_k's satisfies the initial condition in our initial-value problem. Moreover, since each of these ψ_k's is a constant added to an integral from 0 to x , each of these ψ_k's should be continuous at least over the interval of x's on which the integral is finite.

(Naively) Taking the Limit

Now suppose there is an interval (α, β) containing 0 on which this sequence of ψ_k's converges to some continuous function. Let y denote this function,

$$y(x) = \lim_{k\to\infty} \psi_k(x) \qquad \text{for} \quad \alpha < x < \beta \quad .$$

Now let x be any point in (a, b) . Blithely (and naively) taking the limit of both sides of equation (3.4), we get

$$\begin{aligned} y(x) = \lim_{k\to\infty} \psi_k(x) &= \lim_{k\to\infty} \psi_{k+1}(x) \\ &= \lim_{k\to\infty} \left[y_0 + \int_0^x F(s, \psi_k(s))\,ds \right] \\ &= y_0 + \lim_{k\to\infty} \int_0^x F(s, \psi_k(s))\,ds \\ &= y_0 + \int_0^x \lim_{k\to\infty} F(s, \psi_k(s))\,ds \\ &= y_0 + \int_0^x F(s, y(s))\,ds \quad . \end{aligned}$$

Thus (assuming the above limits are valid) we see that y satisfies the integral equation

$$y(x) = y_0 + \int_0^x F(s, y(s))\,ds \qquad \text{for} \quad a < x < b \quad .$$

As noted in Theorem 3.3, this means the function y is a solution to our original initial-value problem, thus verifying the claimed existence of such a solution.

That was the essence of Picard's method of successive approximations.

3.5 Details in the Proof of Theorem 3.1

Confirming the Existence of Solutions

What Are the Remaining Details?

Before proclaiming that we have rigorously verified the existence of a solution to our initial-value problem via the Picard method, we need to rigorously verify the assumptions made in the last section. If you check carefully, you will see that we still need to rigorously confirm the following three statements concerning the functions $\psi_1 , \psi_2 , \ldots$ generated by the Picard iteration method:

1. There is an interval (α, β) containing 0 such that
$$\lim_{k\to\infty} \psi_k(x)$$
exists for each x in (α, β) .

2. The function given by
$$y(x) = \lim_{k\to\infty} \psi_k(x)$$
is continuous on the interval (α, β) .

3. The above defined function y satisfies
$$y(x) = y_0 + \int_0^x F(s, y(s))\, ds \qquad \text{whenever} \quad \alpha < x < \beta \quad .$$

Confirming these claims under the assumptions in Theorem 3.1 on page 42 will be the main goal of this section.[5]

Some Preliminary Bounds

In carrying out our analysis, we will make use of a number of facts normally discussed in standard introductory calculus courses. For example, we will use without comment that fact that, for any summation,
$$\left|\sum_k c_k\right| \leq \sum_k |c_k| \quad .$$
This is the triangle inequality. Recall, also, that "the absolute value of an integral is less than or equal to the integral of the absolute value". We will need to be a little careful about this because the lower limits on our integrals will not always be less than our upper limits. If $\sigma < \tau$, then we do have
$$\left|\int_\sigma^\tau g(s)\, ds\right| \leq \int_\sigma^\tau |g(s)|\, ds \quad .$$

[5] Some of the analysis in this section can be shortened considerably using tools from advanced real analysis. Since the typical reader is not expected to have yet had a such a course, we will not use those tools. However, if you have had such a course and are acquainted with such terms as "uniform convergence" and "Cauchy sequences", then you should look to see how your more advanced mathematics can shorten the analysis given here.

On the other hand, if $\tau < \sigma$, then

$$\left| \int_\sigma^\tau g(s)\,ds \right| = \left| -\int_\tau^\sigma g(s)\,ds \right| = \left| \int_\tau^\sigma g(s)\,ds \right| \leq \int_\tau^\sigma |g(s)|\,ds \quad .$$

Suppose that, in addition, $|g(s)| \leq M$ for all s in some interval containing σ and τ. Then, if $\sigma < \tau$,

$$\left| \int_\sigma^\tau g(s)\,ds \right| \leq \int_\sigma^\tau |g(s)|\,ds \leq \int_\sigma^\tau M\,ds = M[\tau - \sigma] = M\,|\tau - \sigma| \quad ,$$

while, if $\tau < \sigma$,

$$\left| \int_\sigma^\tau g(s)\,ds \right| \leq \int_\tau^\sigma |g(s)|\,ds \leq \int_\tau^\sigma M\,ds = M[-(\tau - \sigma)] = M\,|\tau - \sigma| \quad .$$

So, in general, we have the following little lemma:

Lemma 3.4
If $|g(s)| \leq M$ for all s in some interval containing σ and τ, then

$$\left| \int_\sigma^\tau g(s)\,ds \right| = M\,|\tau - \sigma| \quad .$$

Other facts from calculus will be used and slightly expanded as needed. These facts will include material on the absolute convergence of summations and the Taylor series for the exponentials.

The next lemma establishes the interval (α, β) mentioned in the existence theorem (Theorem 3.1) along with some function bounds that will be useful in our analysis.

Lemma 3.5
Assume both $F(x, y)$ and ${}^{\partial F}/_{\partial y}$ are continuous on some open region $\mathcal{R}$ in the XY–plane containing the point $(0, y_0)$. Then there are positive constants M and B, a closed interval $[\alpha, \beta]$ and a finite distance ΔY such that all the following hold:

1. $\alpha < 0 < \beta$.

2. *The open region $\mathcal{R}$ contains the closed rectangular region*

$$\mathcal{R}_1 = \{\, (x, y) : \alpha \leq x \leq \beta \text{ and } |y - y_0| \leq \Delta Y \,\} \quad .$$

3. *For each (x, y) in $\mathcal{R}_1$,*

$$|F(x, y)| \leq M \qquad \text{and} \qquad \left| \left. \frac{\partial F}{\partial y} \right|_{(x,y)} \right| \leq B \quad .$$

4. $0 < -\alpha M \leq \Delta Y$ *and* $0 < \beta M \leq \Delta Y$.

5. *If ϕ is a continuous function on (α, β) satisfying*

$$|\phi(x) - y_0| \leq \Delta Y \qquad \text{for} \quad \alpha \leq x \leq \beta \quad ,$$

then

$$\psi(x) = y_0 + \int_0^x F(s, \phi(s))\,ds$$

defines the function ψ on the interval $[\alpha, \beta]$. Moreover, ψ is continuous on $[\alpha, \beta]$ and satisfies

$$|\psi(x) - y_0| \leq \Delta Y \qquad \text{for} \quad \alpha \leq x \leq \beta \quad .$$

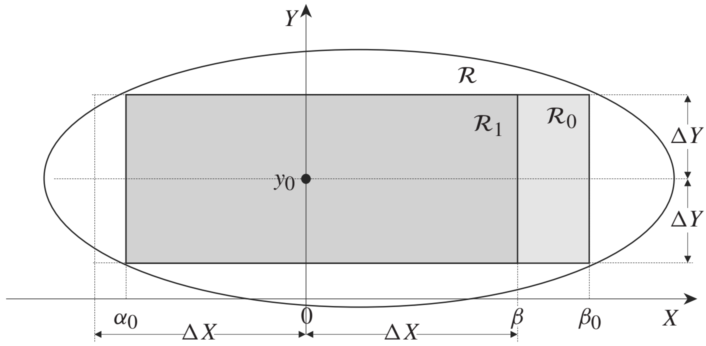


Figure 3.1: Rectangles contained in region $\mathcal{R}$ for the proof of Lemma 3.5 (with $|\alpha_0| < \Delta X$ and $\Delta X < \beta_0$).

PROOF: The goal is to find a rectangle $\mathcal{R}_1$ on which the above holds. We start by noting that, because $\mathcal{R}$ is an open region containing the point $(0, y_0)$, that point is not on the boundary of $\mathcal{R}$, and we can pick a negative value α_0 and two positive values β_0 and ΔY so that the closed rectangular region

$$\mathcal{R}_0 = \{ (x, y) : \alpha_0 \leq x \leq \beta_0 \text{ and } |y - y_0| \leq \Delta Y \}$$

is contained in $\mathcal{R}$, as in Figure 3.1.

Since F and $\partial F/\partial y$ are continuous on $\mathcal{R}$, they (and their absolute values) must be continuous on that portion of $\mathcal{R}$ which is $\mathcal{R}_0$. But recall that a continuous function of one variable on a closed finite interval will always have a maximum value on that interval. Likewise, a continuous function of two variables will always have a maximum value over a closed finite rectangle. Let M and B be, respectively, the maximum values of $|F|$ and $|\partial F/\partial y|$ on $\mathcal{R}_0$. Then, of course,

$$|F(x, y)| \leq M \quad \text{and} \quad \left| \left.\frac{\partial F}{\partial y}\right|_{(x,y)} \right| \leq B \qquad \text{for each } (x, y) \text{ in } \mathcal{R}_0 \quad .$$

Now let us further restrict the possible values of x by first setting

$$\Delta X = \frac{\Delta Y}{M} \qquad \left(\text{so } M = \frac{\Delta Y}{\Delta X}\right) \quad ,$$

and then defining the endpoints of the interval (α, β) by

$$\alpha = \begin{Bmatrix} \alpha_0 & \text{if} & -\Delta X < \alpha_0 \\ -\Delta X & \text{if} & \alpha_0 \leq -\Delta X \end{Bmatrix} \qquad \text{and} \qquad \beta = \begin{Bmatrix} \Delta X & \text{if} & \Delta X < \beta_0 \\ \beta_0 & \text{if} & \beta_0 \leq \Delta X \end{Bmatrix}$$

(again, see Figure 3.1).

By these choices,

$$\alpha_0 \leq \alpha < 0 < \beta \leq \beta_0 \quad ,$$

$$|x| \leq \Delta X \qquad \text{whenever} \quad \alpha \leq x \leq \beta \quad ,$$

$$0 < -\alpha M \leq \Delta X M = \Delta Y \quad ,$$

$$0 < \beta M \leq \Delta X M = \Delta Y \quad ,$$

and the closed rectangle

$$\mathcal{R}_1 \;=\; \{\,(x, y) : \alpha \le x \le \beta \text{ and } |y - y_0| \le \Delta Y\,\}$$

is contained in the closed rectangle $\mathcal{R}_0$, ensuring that

$$|F(x, y)| \;\le\; M \quad\text{and}\quad \left|\left.\frac{\partial F}{\partial y}\right|_{(x,y)}\right| \;\le\; B \qquad \text{for each } (x, y) \text{ in } \mathcal{R}_1 \quad .$$

This takes care of the first four claims of the lemma.

To confirm the lemma's final claim, let ϕ be a continuous function on (α, β) satisfying

$$|\phi(x) - y_0| \;\le\; \Delta Y \qquad \text{for} \quad \alpha \le x \le \beta \quad .$$

Then, $(s, \phi(s))$ is a point in $\mathcal{R}_1$ for each s in the interval $[\alpha, \beta]$. This, in turn, means that $F(s, \phi(s))$ exists and is bounded by M over the interval $[\alpha, \beta]$. Moreover, it is easily verified that the continuity of both F over $\mathcal{R}$ and ϕ over (α, β) ensures that $F(s, \phi(s))$ is a bounded continuous function of s over $[\alpha, \beta]$. Consequently, the integral in

$$\psi(x) \;=\; y_0 \;+\; \int_0^x F(s, \phi(s))\, ds$$

exists (and is finite) for each x in $[\alpha, \beta]$.

To help confirm the claimed continuity of ψ, take any two points x and x_1 in (α, β). Using Lemma 3.4 and the fact that F is bounded by M on $\mathcal{R}_1$, we have that

$$\begin{aligned}
|\psi(x_1) - \psi(x)| \;&=\; \left| y_0 \;+\; \int_0^{x_1} F(s, \phi(s))\, ds \;-\; y_0 \;-\; \int_0^{x} F(s, \phi(s))\, ds \right| \\
&=\; \left| \int_x^{x_1} F(s, \phi(s))\, ds \right| \\
&\le\; M\,|x_1 - x| \quad .
\end{aligned}$$

Hence,

$$\lim_{x \to x_1} |\psi(x_1) - \psi(x)| \;\le\; \lim_{x \to x_1} M\,|x_1 - x| \;=\; M \cdot 0 \;=\; 0 \quad ,$$

which, in turn, means that

$$\lim_{x \to x_1} \psi(x) \;=\; \psi(x_1) \quad ,$$

confirming that ψ is continuous at each x_1 in (α, β). By almost identical arguments, we also have

$$\lim_{x \to \alpha^+} \psi(x) \;=\; \psi(\alpha) \qquad \text{and} \qquad \lim_{x \to \beta^-} \psi(x) \;=\; \psi(\beta) \quad .$$

Altogether, these limits tell us that ψ is continuous on the closed interval $[\alpha, \beta]$.

Finally, let $\alpha \le x \le \beta$. Again using Lemma 3.4 and the boundedness of F, along with the definition of ΔX, we see that

$$|\psi(x) - y_0| \;=\; \left| \int_0^x F(s,\, \phi(s))\, ds \right| \;\le\; M\,|x| \;\le\; M \Delta X \;=\; M \cdot \frac{\Delta Y}{M} \;=\; \Delta Y \quad . \qquad \blacksquare$$

Convergence of the Picard Sequence

Let us now look more closely at the Picard sequence of functions,

$$\psi_0 \; , \; \psi_1 \; , \; \psi_2 \; , \; \psi_3 \; , \; \ldots$$

with ψ_0 being "some continuous function" and

$$\psi_{k+1}(x) \; = \; y_0 \; + \int_0^x F(s, \psi_k(s))\, ds \qquad \text{for} \quad k \; = \; 0, \; 1, \; 2, \; 3, \; \ldots \quad .$$

Remember, F and $\partial F/\partial y$ are continuous on some open region containing the point $(0, y_0)$. This means Lemma 3.5 applies. Let $[\alpha, \beta]$, M, B and ΔY be the interval and constants from that lemma. Let us also now impose an additional restriction on the choice for ψ_0: Let us insist that ψ_0 be any continuous function on $[\alpha, \beta]$ such that

$$|\psi_0(x) - y_0| \; \leq \; \Delta Y \qquad \text{for} \quad \alpha < x < \beta \quad .$$

In particular, we could let ψ_0 be the constant function $\psi_0(x) = y_0$ for all x.

We now want to show that the sequence of ψ_k's converges to a function y on $[\alpha, \beta]$. Our first step in this direction is to observe that, thanks to the additional requirement on ψ_0, Lemma 3.5 can be applied repeatedly to show that ψ_1, ψ_2, ψ_3, $\ldots$ are all well-defined, continuous functions on the interval $[\alpha, \beta]$ with each satisfying

$$|\psi_k(x) - y_0| \; \leq \; \Delta Y \qquad \text{for} \quad \alpha \leq x \leq \beta \quad .$$

Next, we need to establish useful bounds on the sequence

$$|\psi_1(x) - \psi_0(x)| \quad , \quad |\psi_2(x) - \psi_1(x)| \quad , \quad |\psi_3(x) - \psi_2(x)| \quad , \quad \ldots$$

when $\alpha \leq x \leq \beta$. The first is easy:

$$\begin{aligned} |\psi_1(x) - \psi_0(x)| \; &= \; |\psi_1(x) - y_0 - \psi_0(x) + y_0| \\ &= \; |[\psi_1(x) - y_0] \; + \; (-[\psi_0(x) - y_0])| \\ &\leq \; |\psi_1(x) - y_0| \; + \; |\psi_0(x) - y_0| \; \leq \; 2\Delta Y \quad . \end{aligned}$$

To simplify the derivation of useful bounds on the others, let us observe that, if $k \geq 1$,

$$\begin{aligned} |\psi_{k+1}(x) - \psi_k(x)| \; &= \; \left| \left[y_0 \; + \int_0^x F(s, \psi_k(s))\, ds \right] - \left[y_0 \; + \int_0^x F(s, \psi_{k-1}(s))\, ds \right] \right| \\ &= \; \left| \int_0^x [F(s, \psi_k(s)) - F(s, \psi_{k-1}(s))]\, ds \right| \\ &\leq \; \int_0^x |F(s, \psi_k(s)) - F(s, \psi_{k-1}(s))|\, ds \quad . \end{aligned}$$

Now recall that, if f is any continuous and differentiable function on an interval $\mathcal{I}$, and t_1 and t_2 are two points in $\mathcal{I}$, then there is a point τ between t_1 and t_2 such that

$$f(t_2) \; - \; f(t_1) \; = \; f'(\tau)\,[t_2 - t_1] \quad .$$

This was the mean value theorem for derivatives. Consequently, if

$$\left| f'(t) \right| \; \leq \; B \qquad \text{for each } t \text{ in } \mathcal{I} \quad ,$$

then

$$|f(t_2) - f(t_1)| = \left|f'(\tau)\,[t_2 - t_1]\right| = \left|f'(\tau)\right| |t_2 - t_1| \le B\,|t_2 - t_1| \quad .$$

The same holds for partial derivatives. In particular, for each pair of points (x, y_1) and (x, y_2) in the closed rectangle

$$\mathcal{R}_1 = \{ (x, y) : \alpha \le x \le \beta \text{ and } |y - y_0| \le \Delta Y \} \quad ,$$

we have a γ between y_1 and y_2 such that

$$|F(x, y_2) - F(x, y_1)| = \left| \frac{\partial F}{\partial y}\bigg|_{(x,\gamma)} \cdot [y_2 - y_1] \right| \le B\,|y_2 - y_1| \quad .$$

Thus, for $0 \le x \le \beta$ and $k = 1, 2, 3, \ldots$,

$$\begin{aligned} |\psi_{k+1}(x) - \psi_k(x)| &\le \int_0^x |F(s, \psi_k(s)) - F(s, \psi_{k-1}(s))|\,ds \\ &\le \int_0^x B\,|\psi_k(s) - \psi_{k-1}(s)|\,ds \quad . \end{aligned}$$

Repeatedly using this (with $0 \le x \le \beta$), we get

$$\begin{aligned} |\psi_2(x) - \psi_1(x)| &\le \int_0^x B\,|\psi_1(s) - \psi_0(s)|\,ds \\ &\le \int_0^x B \cdot 2\Delta Y\,ds = 2\Delta Y B x \quad , \end{aligned}$$

$$\begin{aligned} |\psi_3(x) - \psi_2(x)| &\le \int_0^x B\,|\psi_2(s) - \psi_1(s)|\,ds \\ &\le \int_0^x B \cdot 2\Delta Y B\,s\,ds = 2\Delta Y \frac{(Bx)^2}{2} \quad , \end{aligned}$$

$$\begin{aligned} |\psi_4(x) - \psi_3(x)| &\le \int_0^x |\psi_3(s) - \psi_2(s)|\,ds \\ &\le \int_0^x B \cdot 2\Delta Y B^2 \frac{s^2}{2}\,ds \le 2\Delta Y \frac{(Bx)^3}{3 \cdot 2} \quad , \end{aligned}$$

$$\begin{aligned} |\psi_5(x) - \psi_4(x)| &\le \int_0^x |\psi_4(s) - \psi_3(s)|\,ds \\ &\le \int_0^x B \cdot 2\Delta Y B^3 \frac{s^3}{3 \cdot 2}\,ds \le 2\Delta Y \frac{(Bx)^4}{4!} \quad , \end{aligned}$$

$$\vdots$$

Continuing, we get

$$|\psi_{k+1}(x) - \psi_k(x)| \le 2\Delta Y \frac{(Bx)^k}{k!} \qquad \text{for} \quad 0 \le x \le \beta \quad \text{and} \quad k = 1, 2, 3, \ldots \quad .$$

Virtually the same arguments give us

$$|\psi_{k+1}(x) - \psi_k(x)| \le 2\Delta Y \frac{(-Bx)^k}{k!} \qquad \text{for} \quad \alpha \le x \le 0 \quad \text{and} \quad k = 1, 2, 3, \ldots \quad .$$

More concisely, for $\alpha \leq x \leq \beta$ and $k = 1, 2, 3, \ldots,$

$$|\psi_{k+1}(x) - \psi_k(x)| \leq 2\Delta Y \frac{(B\,|x|)^k}{k!} \quad . \tag{3.5}$$

At this point it is worth recalling that the Taylor series for e^X is

$$\sum_{k=0}^{\infty} \frac{X^k}{k!}$$

and that this series converges for each real value X. In particular, for any x,

$$2\Delta Y e^{B|x|} = \sum_{k=0}^{\infty} 2\Delta Y \frac{(B\,|x|)^k}{k!} \quad .$$

Now consider the infinite series

$$S(x) = \sum_{k=0}^{\infty} \left[\psi_{k+1}(x) - \psi_k(x)\right] \quad .$$

According to inequality (3.5), the absolute value of each term in this series is bounded by the corresponding term in the Taylor series for $2\Delta Y e^{B|x|}$. The comparison test then tells us that $S(x)$ converges absolutely for each x in $[\alpha, \beta]$. And this means that the limit

$$S(x) = \lim_{N\to\infty} \sum_{k=0}^{N} \left[\psi_{k+1}(x) - \psi_k(x)\right]$$

exists for each x in the interval $[\alpha, \beta]$. But

$$\begin{aligned}\sum_{k=0}^{N} \left[\psi_{k+1}(x) - \psi_k(x)\right] &= [\psi_1(x) - \psi_0(x)] + [\psi_2(x) - \psi_1(x)] + [\psi_3(x) - \psi_2(x)] \\ &\qquad + \cdots + [\psi_N(x) - \psi_{N-1}(x)] + [\psi_{N+1}(x) - \psi_N(x)] \\ &= -\psi_0(x) + \psi_1(x) - \psi_1(x) + \psi_2(x) - \psi_2(x) + \psi_3(x) \\ &\qquad + \cdots - \psi_{N-1}(x) + \psi_N(x) - \psi_N(x) + \psi_{N+1}(x) \quad .\end{aligned}$$

Most of the terms cancel out, leaving us with

$$\sum_{k=0}^{N} \left[\psi_{k+1}(x) - \psi_k(x)\right] = \psi_{N+1}(x) - \psi_0(x) \quad . \tag{3.6}$$

So

$$\lim_{k\to\infty} \psi_k(x) = \lim_{k\to\infty} \left[\psi_0(x) + \sum_{k=0}^{k-1} \left[\psi_{k+1}(x) - \psi_k(x)\right]\right] = \psi_0(x) + S(x) \quad .$$

This shows that the limit

$$y(x) = \lim_{N\to\infty} \psi_N(x)$$

exists for each x in $[\alpha, \beta]$, confirming the first statement we wished to confirm at the beginning of this section (see page 47).

At this point, let us observe that, for $\alpha \leq x \leq \beta$, we have the formulas

$$\psi_N(x) = \psi_0(x) + S(x) = \psi_0(x) + \sum_{k=0}^{N-1} \left[\psi_{k+1}(x) - \psi_k(x)\right] \tag{3.7a}$$

and

$$y(x) = \psi_0(x) + S(x) = \psi_0(x) + \sum_{k=0}^{\infty} \left[\psi_{k+1}(x) - \psi_k(x)\right] \quad . \tag{3.7b}$$

Let us also observe what we get when we combine the above formula for ψ_N with inequality (3.5) and the observations regarding the Taylor series of the exponential:

$$\begin{aligned} |\psi_N(x)| &\leq |\psi_0(x)| + \sum_{k=0}^{N-1} |\psi_{k+1}(x) - \psi_k(x)| \\ &\leq |\psi_0(x)| + \sum_{k=0}^{N} 2\Delta Y \frac{(B\,|x|)^k}{k!} = |\psi_0(x)| + \Delta Y e^{B|x|} \quad . \end{aligned} \tag{3.8a}$$

Likewise

$$|y(x)| \leq |\psi_0(x)| + \Delta Y e^{B|x|} \quad . \tag{3.8b}$$

These observations may later prove useful.

Continuity of the Limit

Now to confirm the continuity of y claimed by the second statement from the beginning of this section. We start by picking any two points x_1 and x in $[\alpha, \beta]$, and any positive integer N , and then observe that, because F is bounded by M ,

$$\begin{aligned} |\psi_N(x_1) - \psi_N(x)| &= \left|\left[y_0 + \int_0^{x_1} F(s, \psi_{N-1}(s))\,ds\right] - \left[y_0 + \int_0^{x} F(s, \psi_{N-1}(s)\,ds\right]\right| \\ &= \left|\int_x^{x_1} F(s, \psi_{N-1}(s))\,ds\right| \\ &\leq M\,|x_1 - x| \quad . \end{aligned}$$

Combined with the definition of y and some basic facts about limits, this gives us

$$|y(x_1) - y(x)| = \lim_{N\to\infty} |\psi_N(x_1) - \psi_N(x)| \leq M\,|x_1 - x| \quad .$$

As demonstrated at the end of the proof of Lemma 3.5, this immediately tells us that y is continuous on $[\alpha, \beta]$.

The Limit as a Solution

Finally, let us verify the third statement made at the beginning of this section, namely that the above defined y satisfies

$$y(x) = y_0 + \int_0^x F(s, y(s))\,ds \qquad \text{whenever} \quad \alpha < x < \beta \quad .$$

This, according to Theorem 3.3 on page 45, is equivalent to showing that y satisfies the differential equation in our initial-value problem over the interval (α, β) .[6]

[6] Yes, we've already shown that y is defined and continuous on $[\alpha, \beta]$, not just (α, β) . However, the derivative of a function is ill-defined at the endpoints of the interval over which it is defined, and that is why we are now limiting x to being in (α, β) .

We start by assuming $\alpha \leq x \leq \beta$. Using equation set (3.7) and inequality (3.5), we see that

$$
\begin{aligned}
|y(x) - \psi_N(x)| &= |[y(x) - \psi_0(x)] - [\psi_N(x) - \psi_0(x)]| \\
&= \left| \sum_{k=0}^{\infty} \left[\psi_{k+1}(x) - \psi_k(x)\right] - \sum_{k=0}^{N-1} \left[\psi_{k+1}(x) - \psi_k(x)\right] \right| \\
&= \left| \sum_{k=N}^{\infty} \left[\psi_{k+1}(x) - \psi_k(x)\right] \right| \\
&\leq \sum_{k=N}^{\infty} |\psi_{k+1}(x) - \psi_k(x)| \\
&\leq \sum_{k=N}^{\infty} 2\Delta Y \frac{(B\,|x|)^k}{k!} \quad .
\end{aligned}
$$

Under the change of index $k = N + n$, this becomes

$$
|y(x) - \psi_N(x)| \leq 2\Delta Y \sum_{n=0}^{\infty} \frac{(B\,|x|)^{N+n}}{(N+n)!} \quad . \tag{3.9}
$$

But

$$
\begin{aligned}
(N+n)! &= \underbrace{(N+n)}_{\geq N}\underbrace{(N+n-1)}_{\geq N-1}\underbrace{(N+n-2)}_{\geq N-2}\cdots\underbrace{(N+n-[N-1])}_{\geq 1}\underbrace{n(n-1)\cdots 2\cdot 1}_{=n!} \\
&\geq N!\,n! \quad .
\end{aligned}
$$

Thus,

$$
\frac{1}{(N+n)!} \leq \frac{1}{N!\,n!}
$$

and

$$
\sum_{n=0}^{\infty} \frac{(B\,|x|)^{N+n}}{(N+n)!} \leq \sum_{n=0}^{\infty} \frac{(B\,|x|)^{N+n}}{N!\,n!} \leq \frac{(B\,|x|)^N}{N!} \sum_{n=0}^{\infty} \frac{(B\,|x|)^n}{n!} = \frac{(B\,|x|)^N}{N!} e^{B|x|} \quad .
$$

Combining this with inequality (3.9) yields

$$
|y(x) - \psi_N(x)| \leq 2\Delta Y \frac{(B\,|x|)^N}{N!} e^{B|x|} \quad .
$$

Consequently,

$$
\begin{aligned}
&\left| \psi_{N+1}(x) - y_0 - \int_0^x F(s, y(s))\,ds \right| \\
&\qquad = \left| \left[y_0 + \int_0^x F(s, \psi_N(s))\,ds \right] - y_0 - \int_0^x F(s, y(s))\,ds \right| \\
&\qquad \leq \int_0^x |F(s, \psi_N(s)) - F(s, y(s))|\,ds \\
&\qquad \leq \int_0^x B\,|\psi_{N-1}(s) - y(s)|\,ds
\end{aligned}
$$

$$\leq \int_0^x B \cdot 2\Delta Y \frac{(B\,|s|)^{N-1}}{(N-1)!} e^{B|s|}\,ds$$
$$= 2\Delta Y \frac{B^N e^{B|x|}}{(N-1)!} \int_0^x |s|^{N-1}\,ds \quad .$$

Computing the last integral leaves us with

$$\left|\psi_{N+1}(x) - y_0 - \int_0^x F(s, y(s))\,ds\right| \leq 2\Delta Y \frac{(B\,|x|)^N}{N!} e^{B|x|} \quad .$$

But, as is well known,

$$\frac{(B\,|x|)^N}{N!} \rightarrow 0 \qquad \text{as} \quad N \rightarrow \infty$$

for any finite value $B\,|x|$. Hence

$$\left|\psi_N(x) - y_0 - \int_0^x F(s, y(s))\,ds\right| \rightarrow 0 \qquad \text{as} \quad N \rightarrow \infty \quad .$$

That is

$$0 = \lim_{N\to\infty} \left[\psi_N(x) - y_0 - \int_0^x F(s, y(s))\,ds\right]$$
$$= \lim_{N\to\infty} \psi_N(x) - y_0 - \int_0^x F(s, y(s))\,ds$$
$$= y(x) - y_0 - \int_0^x F(s, y(s))\,ds \quad ,$$

verifying that

$$y(x) = y_0 + \int_0^x F(s, y(s))\,ds \qquad \text{whenever} \quad \alpha < x < \beta \quad ,$$

as desired.

Where Are We?

Let's stop for a moment and review what we have done. We have just spent several pages rigorously verifying the three statements made at the beginning of this section under the assumptions made in Theorem 3.1 on page 42. By verifying these statements, we've rigorously justified the computations made in the previous section showing that the limit of a Picard sequence is a solution to the initial-value problem in Theorem 3.1. Consequently, we have now rigorously verified the claim in Theorem 3.1 that a solution to the given initial-value problem exists on at least some interval (α, β) .

We now need to show that this y is the only solution on that interval.

The Uniqueness Claim in Theorem 3.1

If you've made it through this section up to this point, then you should have little difficulty in finishing the proof of Theorem 3.1 by doing the following exercises. Do make use of the work we've done in the previous several pages.

?► *Exercise 3.2:* *Consider a first-order initial-value problem*

$$\frac{dy}{dx} = F(x, y) \qquad \textit{with} \quad y(0) = y_0 \quad ,$$

and with both F and $^{\partial F}/_{\partial y}$ being continuous functions on some open region containing the point $(0, y_0)$. Since Lemma 3.5 applies, we can let $[\alpha, \beta]$ be the interval, and M, B and ΔY the positive constants from that lemma. Using this interval and these constants:

a i: *Verify that*

$$0 \leq M|x| \leq \Delta Y \quad \text{for} \quad \alpha \leq x \leq \beta \quad .$$

ii: *Also verify that any solution y to the above initial-value problem satisfies*

$$|y(x) - y_0| \leq M|x| \quad \text{for} \quad a < x < b \quad .$$

Now observe that the last two inequalities yield

$$|y(x) - y_0| \leq M|x| \leq \Delta Y \quad \text{for} \quad \alpha \leq x \leq \beta$$

whenever y is a solution to the above initial-value problem.

b: *For the following, let y_1 and y_2 be any two solutions to the above initial-value problem on (α, β), and let*

$$\psi_0 , \psi_1 , \psi_2 , \psi_3 , \ldots \quad \text{and} \quad \phi_0 , \phi_1 , \phi_2 , \phi_3 , \ldots$$

be the two Picard sequences of functions on (α, β) generated by setting

$$\psi_{k+1}(x) = y_0 + \int_0^x F(s, \psi_k(s))\, ds$$

and

$$\phi_{k+1}(x) = y_0 + \int_0^x F(s, \phi_k(s))\, ds$$

with

$$\psi_0(x) = y_1(x) \quad \text{and} \quad \phi_0(x) = y_2(x) \quad .$$

i: *Using ideas similar to those used above to prove the convergence of the Picard sequence, show that, for each x in (α, β) and each positive integer k,*

$$|\psi_{k+1}(x) - \phi_{k+1}(x)| \leq \int_0^x B\,|\psi_k(s) - \phi_k(s)|\, ds \quad .$$

ii: *Then verify that, for each x in (α, β),*

$$|\psi_0(x) - \phi_0(x)| \leq 2\Delta Y \quad ,$$

and

$$\lim_{k\to\infty} |\psi_{k+1}(x) - \phi_{k+1}(x)| = 0 \quad .$$

(Hint: This is very similar to our showing that $|\psi_{k+1}(x) - \psi_{k+1}(x)| \to 0$ as $k \to \infty$.)

iii: *Verify that, for each x in (α, β) and positive integer k,*

$$\psi_k(x) = y_1(x) \quad \text{and} \quad \phi_k(x) = y_2(x) \quad .$$

iv: *Combine the results of the last two parts to show that*

$$y_1(x) = y_2(x) \quad \text{for} \quad \alpha < x < \beta \quad .$$

The end result of the above set of exercises is that there cannot be two *different* solutions on the interval (α, β) to the initial-value problem. That was the uniqueness claim of Theorem 3.1.

3.6 On Proving Theorem 3.2

We could spend several more enjoyable pages redoing the work in the previous section, but under the assumptions made in Theorem 3.2 on page 43 instead of those in Theorem 3.1. To avoid that, let us briefly discuss how you can modify that work, and, thereby, prove Theorem 3.2.

First of all, recall that much of the initial effort in proving the convergence of the Picard sequence,

$$\psi_0 \ , \ \psi_1 \ , \ \psi_2 \ , \ \psi_3 \ , \ \dots$$

with

$$\psi_{k+1}(x) \ = \ y_0 \ + \int_0^x F(s, \psi_k(s))\, ds \qquad \text{for} \quad k \ = \ 0,\ 1,\ 2,\ 3,\ \dots \quad ,$$

was in showing that there is an interval (α, β) such that, as long as $\alpha \leq s \leq \beta$, then $\psi_k(s)$ is never so large or so small that $(s, \psi_k(s))$ is outside a rectangular region on which F is "well-behaved" (this was the main result of Lemma 3.5 on page 48). However, if (as in Theorem 3.2) $F = F(x, y)$ is a continuous function on the infinite strip

$$\mathcal{R} \ = \ \{ (x, y) : \alpha < x < \beta \text{ and } -\infty < y < \infty \} \quad ,$$

then, for any continuous function ϕ on (α, β), $F(s, \phi(s))$ is a well-defined, continuous function of s over (α, β), and the integral in

$$\psi(x) \ = \ y_0 \ + \int_0^x F(s, \phi(s))\, ds$$

exists (and is finite) whenever $\alpha < x < \beta$. Verifying that ψ is continuous requires a little more thought than was needed in the proof of Lemma 3.5, but is still pretty easy — simply appeal to the continuity of $F(s, \phi(s))$ as a function of s along with the fact that

$$\psi(x_1) - \psi(x) \ = \ \int_x^{x_1} F(s, \phi(s))\, ds$$

to show that

$$\lim_{x \to x_1} \psi(x) \ = \ \psi(x_1) \qquad \text{for each } x_1 \text{ in } (\alpha, \beta) \quad .$$

Consequently, all the functions in the Picard sequence ψ_0, ψ_1, ψ_2, $\dots$ are continuous on (α, β) (provided, of course, that we started with ψ_0 being continuous).

Now choose finite values α_1 and β_1 so that $\alpha < \alpha_1 < 0 < \beta_1 < \beta$; let ΔY be the maximum value of

$$\frac{1}{2} \left| \psi_1(x) - \psi_0(x) \right| \qquad \text{for} \quad \alpha_1 \leq x \leq \beta_1 \quad ,$$

and let $\mathcal{R}_0$ be the infinite strip

$$\mathcal{R}_0 \ = \ \{ (x, y) : \alpha_1 < x < \beta_1 \text{ and } -\infty < y < \infty \} \quad .$$

By the assumptions in the theorem, we know that, on $\mathcal{R}$, the continuous function ${}^{\partial F}/_{\partial y}$ depends only on x. So we can treat it as a continuous function on the closed interval $[\alpha_1, \beta_1]$. But such functions are bounded. Thus, for some positive constant B and every point in $\mathcal{R}_0$,

$$\left| \frac{\partial F}{\partial y} \right| \ \leq \ B \quad .$$

Using this, the bounds on

$$\left| \psi_{k+1}(x) - \psi_k(x) \right| \qquad \text{for} \quad \alpha_1 \leq x \leq \beta_1 \quad \text{and} \quad k = 1,\ 2,\ 3,\ \dots$$

can now be rederived exactly as in the previous section (leading to inequality (3.5) on page 53 and inequality set (3.8) on page 54), and we can then use arguments almost identical to those used in the previous section to show that the Picard sequence converges on (α_1, β_1) to a solution y of the given initial-value problem. The only notable modification is that the bound M used to show the continuity of y must be rederived. For this proof, let M be the maximum value of $F(x, y)$ on the closed rectangle

$$\{ (x, y) : \alpha_1 \leq x \leq \beta_1 \text{ and } |y| \leq H \}$$

where H is the maximum value of

$$|\psi_0(x)| + \Delta Y e^{B|x|} \qquad \text{for} \quad \alpha_1 \leq x \leq \beta_1 \quad .$$

Inequality set (3.8) then tells us that

$$|\psi_k(s)| \leq H \qquad \text{for} \quad \alpha_1 \leq s \leq \beta_1 \quad \text{and} \quad k = 0, 1, 2, 3, \ldots \quad .$$

This, in turn, assures us that

$$|F(s, \psi_k(s))| \leq M \qquad \text{for} \quad \alpha_1 \leq s \leq \beta_1 \quad \text{and} \quad k = 0, 1, 2, 3, \ldots \quad ,$$

which is what we used in the previous section to prove the continuity of y.

Finally, since every point x in the interval (α, β) is also in some such subinterval (α_1, β_1), we must have that the Picard sequence converges at every point x in (α, β), and what it converges to, $y(x)$, is a solution to the given initial-value problem. Straightforward modifications to the arguments outlined in Exercise 3.2 then show that this solution is the only solution.

3.7 Appendix: A Little Multivariable Calculus

There are a few places in our discussions where some knowledge of the calculus of functions of two or more variables (i.e., "multivariable" calculus) is needed. These include the commentary about existence and uniqueness in this chapter (Theorems 3.1 and 3.2), and the use of the multivariable version of the chain rule in Chapter 7. This appendix is a brief introduction to those elements of multivariable calculus that are needed for these discussions. It is for those who have not yet been formally introduced to calculus of several variables, and contains just barely enough to get by.

Functions of Two Variables

At least while we are only concerned with first-order differential equations, the only multivariable calculus we will need involves functions of just two variables, such as

$$f(x, y) = x^2 + x^2 y^2 \quad , \quad g(x, y) = \frac{x^3 + 4y}{x} \quad \text{and} \quad h(x, y) = \sqrt{x^3 + y^2} \quad .$$

These functions will be defined on "regions" of the XY–plane.

Open and Closed Regions

Functions of one variable are typically defined on intervals of the X–axis. For functions of two variables, we must replace the concept of an interval with that of a "region". For our purposes, a *region* (in the XY–plane) refers to the collection of all points enclosed by some curve or set of curves on the plane (with the understanding that this curve or set of curves actually does enclose some collection of points in the plane). If we include the curves with the enclosed points, then we say the region is *closed*; if the curves are all excluded, then we refer to the region as *open*. This corresponds to the distinction between a closed interval $[a, b]$ (which does contain the endpoints), and an open interval (a, b) (which does not contain the endpoints).

!►Example 3.4: *Consider the rectangular region $\mathcal{R}$ whose sides form the rectangle generated from the vertical lines $x = 1$ and $x = 4$ along with the horizontal lines $y = 2$ and $y = 6$. If $\mathcal{R}$ is to be a closed region, then it must include this rectangle; that is,*

$$\mathcal{R} = \{(x, y) : 1 \leq x \leq 4 \text{ and } 2 \leq y \leq 6\} \quad .$$

If $\mathcal{R}$ is to be an open region, then it must exclude this rectangle; that is,

$$\mathcal{R} = \{(x, y) : 1 < x < 4 \text{ and } 2 < y < 6\} \quad .$$

On the other hand, if $\mathcal{R}$ just includes one of its sides, say, its right side,

$$\mathcal{R} = \{(x, y) : 1 < x \leq 4 \text{ and } 2 < y < 6\} \quad ,$$

then it is considered to be neither open or closed.

Limits

The concept of limits for functions of two variables is a natural extension of the concept of limits for functions of one variable.

Given a function $f(x, y)$ of two variables, a point (x_0, y_0) in the plane, and a finite value A, we say that

A is the limit of $f(x, y)$ as (x, y) approaches (x_0, y_0),

equivalently,

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = A \qquad \text{or} \qquad f(x, y) \to A \quad \text{as} \quad (x, y) \to (x_0, y_0) \quad ,$$

if and only if we can make the value of $f(x, y)$ as close (but not necessarily equal) to A as we desire by requiring (x, y) be sufficiently close (but not necessarily equal) to (x_0, y_0). More formally,

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = A$$

if and only if, for every positive value ϵ there is a corresponding positive distance δ_ϵ such that $f(x, y)$ is within ϵ of A whenever (x, y) is within δ_ϵ of (x_0, y_0). That is, (in mathematical shorthand), for each $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that

$$\text{distance from } (x, y) \text{ to } (x_0, y_0) < \delta_\epsilon \quad \Longrightarrow \quad |f(x, y) - A| < \epsilon \quad .$$

The rules for the existence and computation of these limits are straightforward extensions of those for functions of one variable, and need not be discussed in detail here.

!►Example 3.5: *"Obviously", if*

$$f(x, y) = x^2 + x^2y^2 \quad ,$$

then

$$\lim_{(x,y)\to(2,3)} f(x, y) = \lim_{(x,y)\to(2,3)} \left[x^2 + x^2y^2\right] = 2^2 + (2^2)(3^2) = 40 \quad .$$

On the other hand

$$\lim_{(x,y)\to(0,3)} g(x, y)$$

does not exist if

$$g(x, y) = \frac{x^3 + 4y}{x}$$

because $(x, y) \to (0, 3)$ *leads to* ${}^{12}/_{0}$.

Continuity

The only difference between "continuity for a function of one variable" and "continuity for a function of two variables" is the number of variables involved.

Basically, a function $f(x, y)$ is continuous at a point (x_0, y_0) if and only if we can legitimately write

$$\lim_{(x,y)\to(x_0,y_0)} f(x, y) = f(x_0, y_0) \quad .$$

That function is then continuous on a region $\mathcal{R}$ if and only if it is continuous at every point in $\mathcal{R}$. Note that this does require $f(x, y)$ to be defined at every point in the region.

Partial Derivatives

Recall that the derivative of a function of one variable $f = f(t)$ is given by the limit formula

$$\frac{df}{dt} = \lim_{\Delta t\to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

provided the limit exists. The simplest extension of this for a function of two variables $f = f(x, y)$ is the "partial" derivatives with respect to each variable:

1. The *(first) partial derivative with respect to* x is denoted and defined by

$$\frac{\partial f}{\partial x} = \lim_{\Delta x\to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

   provided the limit exists.

2. The *(first) partial derivative with respect to* y is denoted and defined by

$$\frac{\partial f}{\partial y} = \lim_{\Delta y\to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

   provided the limit exists.

Note the notation, ${}^{\partial f}/_{\partial x}$ and ${}^{\partial f}/_{\partial y}$, in which we use ∂ instead of d.[7]

An important thing to observe about the limit formula for ${}^{\partial f}/_{\partial x}$ is that, in essence, x replaces the variable t in the previous formula for ${}^{df}/_{dt}$ while y does not vary. Consequently, to compute ${}^{\partial f}/_{\partial x}$, simply take the derivative of $f(x, y)$ using x as the variable while pretending y is a constant. Likewise, to compute ${}^{\partial f}/_{\partial y}$ simply take the derivative of $f(x, y)$ using y as the variable while pretending x is a constant. As a result, everything already learned about computing ordinary derivatives applies to computing partial derivatives, provided we keep straight which variable is being treated (temporarily) as a constant.

[7] Some authors prefer using such notation as $D_x f$ and f_x instead of ${}^{\partial f}/_{\partial x}$, and $D_y f$ and f_y instead of ${}^{\partial f}/_{\partial y}$.

!►Example 3.6: *Let*

$$f(x, y) = x^2 + x^2y^2 \quad .$$

Then

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left[x^2 + x^2y^2\right] = \frac{\partial}{\partial x}\left[x^2\right] + \frac{\partial}{\partial x}\left[x^2y^2\right] = 2x + 2xy^2 \quad ,$$

while

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\left[x^2 + x^2y^2\right] = \frac{\partial}{\partial y}\left[x^2\right] + \frac{\partial}{\partial y}\left[x^2y^2\right] = 0 + x^2 2y \quad .$$

?►Exercise 3.3: *Let*

$$g(x, y) = x^2y^3 \qquad \textit{and} \qquad h(x, y) = \sin\left(x^2 + y^2\right) \quad .$$

Verify that

$$\frac{\partial g}{\partial x} = 2xy^3 \qquad \textit{and} \qquad \frac{\partial g}{\partial y} = 3x^2y^2 \quad ,$$

while

$$\frac{\partial h}{\partial x} = 2x\cos\left(x^2 + y^2\right) \qquad \textit{and} \qquad \frac{\partial h}{\partial y} = 2y\cos\left(x^2 + y^2\right) \quad .$$

Functions of More than Two Variables

The notation can become a bit more cumbersome, and the pictures even harder to draw, but everything discussed above for functions of two variables naturally extends to functions of three or more variables. For example, we may have a function of three variables $f = f(x, y, z)$ defined on, say, an open box-like region

$$\mathcal{R} = \{ (x, y, z) : x_{\min} < x < x_{\max} \ , \ y_{\min} < y < y_{\max} \text{ and } z_{\min} < z < z_{\max} \}$$

where $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$, $z_{\min}$ and $z_{\max}$ are finite numbers. We will then say that, for any given point (x_0, y_0, z_0) and value A ,

$$\lim_{(x,y,)\to(x_0,y_0,z_0)} f(x, y, z) = A$$

if and only if there is a corresponding positive distance δ_ϵ for every positive value ϵ such that $f(x, y, z)$ is within ϵ of A whenever (x, y, z) is within δ_ϵ of (x_0, y_0, z_0) . We will also say that this function is continuous on $\mathcal{R}$ if and only if we can legitimately write

$$\lim_{(x,y,)\to(x_0,y_0,z_0)} f(x, y, z) = f(x_0, y_0, z_0)$$

for every point (x_0, y_0, z_0) in $\mathcal{R}$. Finally, the three (first) partial derivatives of this function are given by

$$\frac{\partial f}{\partial x} = \lim_{\Delta x\to 0} \frac{f(x + \Delta x, y, z) - f(x, y, z)}{\Delta x} \quad ,$$

$$\frac{\partial f}{\partial y} = \lim_{\Delta y\to 0} \frac{f(x, y + \Delta y, z) - f(x, y, z)}{\Delta y}$$

and

$$\frac{\partial f}{\partial z} = \lim_{\Delta y\to 0} \frac{f(x, y, z + \Delta z) - f(x, y, z)}{\Delta z} \quad ,$$

provided the limits exist. Again, in practice, the partial derivative with respect to any one of the three variables is the derivative obtained by pretending the other variables are constants.

Additional Exercises

3.4. *Rewrite each of the following in derivative formula form, and then find all constant solutions. (In some cases, you may have to use the quadratic formula to find any constant solutions.)*

a. $\dfrac{dy}{dx} + 3xy = 6x$

b. $\sin(x+y) - y\dfrac{dy}{dx} = 0$

c. $\dfrac{dy}{dx} - y^3 = 8$

d. $x^2\dfrac{dy}{dx} + xy^2 = x$

e. $\dfrac{dy}{dx} - y^2 = x$

f. $y^3 - 25y + \dfrac{dy}{dx} = 0$

g. $(x-2)\dfrac{dy}{dx} = y + 3$

h. $(y-2)\dfrac{dy}{dx} = x - 3$

i. $\dfrac{dy}{dx} + 2y - y^2 = -2$

j. $\dfrac{dy}{dx} + (8-x)y - y^2 = -8x$

3.5. *Which of the equations in the above exercise set are autonomous?*

3.6. *Consider the first-order initial-value problem*

$$\frac{dy}{dx} = 2\sqrt{y} \qquad \text{with} \quad y(1) = 0 \quad .$$

a. *Verify that each of the following is a solution on the interval* $(-\infty, \infty)$ *, and graph that solution:*

i. $y(x) = 0 \qquad \text{for} \quad -\infty < x < \infty \,.$

ii. $y(x) = \begin{cases} 0 & \text{if} \quad x < 1 \\ (x-1)^2 & \text{if} \quad 1 \le x \end{cases} \quad .$

iii. $y(x) = \begin{cases} 0 & \text{if} \quad x < 3 \\ (x-3)^2 & \text{if} \quad 3 \le x \end{cases} \quad .$

b. *You've just verified three different functions as being solutions to the above initial-value problem. Why does this not violate Theorem 3.1?*

3.7. *Let* ψ_0 *,* ψ_1 *,* ψ_2 *,* ψ_3 *, ... be the sequence of functions generated by the Picard iterative method (as described in Section 3.4) using the initial-value problem*

$$\frac{dy}{dx} = xy \qquad \text{with} \quad y(0) = 2$$

along with

$$\psi_0(x) = 2 \qquad \text{for all} \quad x \quad .$$

Using the formula for Picard's method (formula (3.4) on page 46), compute the following:

a. $\psi_1(x)$ **b.** $\psi_2(x)$ **c.** $\psi_3(x)$

3.8. *Let* $\psi_0\,,\ \psi_1\,,\ \psi_2\,,\ \psi_3\,,\ \ldots$ *be the sequence of functions generated by the Picard iterative method (as described in Section 3.4) using the initial-value problem*

$$\frac{dy}{dx} = 2x + y^2 \qquad \text{with} \quad y(0) = 3$$

along with

$$\psi_0(x) = 3 \qquad \text{for all} \quad x \quad .$$

Compute the following:

a. $\psi_1(x)$ **b.** $\psi_2(x)$

4

Separable First-Order Equations

As we will see below, the notion of a differential equation being "separable" is a natural generalization of the notion of a first-order differential equation being directly integrable. What's more, a fairly natural modification of the method for solving directly integrable first-order equations gives us the basic approach to solving "separable" differential equations. However, it cannot be said that the theory of separable equations is just a trivial extension of the theory of directly integrable equations. Certain issues can arise that do not arise in solving directly integrable equations. Some of these issues are pertinent to even more general classes of first-order differential equations than those that are just separable, and may play a role later on in this text.

In this chapter we will, of course, learn how to identify and solve separable first-order differential equations. We will also see what sort of issues can arise, examine those issues, and discuss some ways to deal with them. Since many of these issues involve graphing, we will also draw a bunch of pictures.

4.1 Basic Notions

Separability

A first-order differential equation is said to be *separable* if, after solving it for the derivative,

$$\frac{dy}{dx} = F(x, y) \quad ,$$

the right-hand side can then be factored as "a formula of just x" times "a formula of just y",

$$F(x, y) = f(x)g(y) \quad .$$

If this factoring is not possible, the equation is not separable.

More concisely, a first-order differential equation is *separable* if and only if it can be written as

$$\frac{dy}{dx} = f(x)g(y) \tag{4.1}$$

where f and g are known functions.

!►*Example 4.1:* *Consider the differential equation*

$$\frac{dy}{dx} - x^2y^2 = x^2 \quad . \tag{4.2}$$

Solving for the derivative (by adding x^2y^2 to both sides),

$$\frac{dy}{dx} = x^2 + x^2y^2 \quad ,$$

and then factoring out the x^2 on the right-hand side gives

$$\frac{dy}{dx} = x^2\left(1 + y^2\right) \quad ,$$

which is in form

$$\frac{dy}{dx} = f(x)g(y)$$

with

$$f(x) = \underbrace{x^2}_{\text{no } y\text{'s}} \qquad \text{and} \qquad g(y) = \underbrace{\left(1 + y^2\right)}_{\text{no } x\text{'s}} \quad .$$

So equation (4.2) is a separable differential equation.

!▶Example 4.2: *On the other hand, consider*

$$\frac{dy}{dx} - x^2y^2 = 4 \quad . \tag{4.3}$$

Solving for the derivative here yields

$$\frac{dy}{dx} = x^2y^2 + 4 \quad .$$

The right-hand side of this clearly cannot be factored into a function of just x times a function of just y. Thus, equation (4.3) is not separable.

We should (briefly) note that any directly integrable first-order differential equation

$$\frac{dy}{dx} = f(x)$$

can be viewed as also being the separable equation

$$\frac{dy}{dx} = f(x)g(y)$$

with $g(y)$ being the constant 1. Likewise, a first-order autonomous differential equation

$$\frac{dy}{dx} = g(y)$$

can also be viewed as being separable, this time with $f(x)$ being 1. Thus, both directly integrable and autonomous differential equations are all special cases of separable differential equations.

Integrating Separable Equations

As just noted, a directly-integrable equation

$$\frac{dy}{dx} = f(x)$$

can be viewed as the separable equation

$$\frac{dy}{dx} = f(x)g(y) \qquad \text{with} \quad g(y) = 1 \quad .$$

We point this out again because the method used to solve directly-integrable equations (integrating both sides with respect to x) is rather easily adapted to solving separable equations. Let us try to figure out this adaptation using the differential equation from the first example. Then, if we are successful, we can discuss its use more generally.

!►Example 4.3: *Consider the differential equation*

$$\frac{dy}{dx} - x^2 y^2 = x^2 \quad .$$

In Example 4.1, we saw that this is a separable equation, and can be written as

$$\frac{dy}{dx} = x^2 \left(1 + y^2\right) \quad .$$

If we simply try to integrate both sides with respect to x , the right-hand side would become

$$\int x^2 \left(1 + y^2\right) dx \quad .$$

Unfortunately, the y here is really $y(x)$, some unknown formula of x ; so the above is just the integral of some unknown function of x — something we cannot effectively evaluate. To eliminate the y's on the right-hand side, we could, before attempting the integration, divide through by $1 + y^2$, obtaining

$$\frac{1}{1 + y^2} \frac{dy}{dx} = x^2 \quad . \tag{4.4}$$

The right-hand side can now be integrated with respect to x . What about the left-hand side? The integral of that with respect to x is

$$\int \frac{1}{1 + y^2} \frac{dy}{dx} \, dx \quad .$$

Tempting as it is to simply "cancel out the dx's", let's not (at least, not yet). After all, dy/dx is not a fraction; it denotes the derivative $y'(x)$ where $y(x)$ is some unknown formula of x . But y is also shorthand for that same unknown formula $y(x)$. So this integral is more precisely written as

$$\int \frac{1}{1 + [y(x)]^2} y'(x) \, dx \quad .$$

Fortunately, this is just the right form for applying the generic substitution $y = y(x)$ to convert the integral with respect to x to an integral with respect to y . No matter what $y(x)$ might be (so long as it is differentiable), we know

$$\int \underbrace{\frac{1}{1 + [y(x)]^2}}_{\frac{1}{1+y^2}} \underbrace{y'(x) \, dx}_{dy} = \int \frac{1}{1 + y^2} \, dy \quad .$$

Combining all this, we get

$$\int \frac{1}{1 + y^2} \frac{dy}{dx} \, dx = \int \frac{1}{1 + [y(x)]^2} y'(x) \, dx = \int \frac{1}{1 + y^2} \, dy \quad ,$$

which, after cutting out the middle, reduces to

$$\int \frac{1}{1+y^2}\frac{dy}{dx}\,dx = \int \frac{1}{1+y^2}\,dy \quad ,$$

the very equation we would have obtained if we had yielded to temptation and naively "cancelled out the dx*'s".*

Consequently, the equation obtained by integrating both sides of equation (4.4) with respect to x *,*

$$\int \frac{1}{1+y^2}\frac{dy}{dx}\,dx = \int x^2\,dx \quad ,$$

is the same as

$$\int \frac{1}{1+y^2}\,dy = \int x^2\,dx \quad .$$

Doing the indicated integration on both sides then yields

$$\arctan(y) = \frac{1}{3}x^3 + c \quad ,$$

which, in turn, tells us that

$$y = \tan\left(\frac{1}{3}x^3 + c\right) \quad .$$

This is the general solution to our differential equation.

Two generally useful ideas were illustrated in the last example. One is that, whenever we have an integral of the form

$$\int H(y)\frac{dy}{dx}\,dx$$

where y denotes some (differentiable) function of x , then this integral is more properly written as

$$\int H(y(x))\,y'(x)\,dx \quad ,$$

which reduces to

$$\int H(y)\,dy$$

via the substitution $y = y(x)$ (even though we don't yet know what $y(x)$ is). Thus, in general,

$$\int H(y)\frac{dy}{dx}\,dx = \int H(y)\,dy \quad . \tag{4.5}$$

This equation is true whether you derive it rigorously, as we have, or obtain it naively by mechanically canceling out the dx's.[1]

The other idea seen in the example was that, if we divide an equation of the form

$$\frac{dy}{dx} = f(x)g(y)$$

by $g(y)$, then (with the help of equation (4.5)) we can compute the integral with respect to x of each side of the resulting equation,

$$\frac{1}{g(y)}\frac{dy}{dx} = f(x) \quad .$$

This leads us to a *basic procedure for solving separable first-order differential equations*:

[1] One of the reasons our notation is so useful is that naive manipulations of the differentials often do lead to valid equations. Just don't be too naive and cancel out the d's in ${}^{dy}/_{dx}$.

1. Get the differential equation into the form
$$\frac{dy}{dx} = f(x)g(y) \quad .$$
2. Divide through by $g(y)$ to get
$$\frac{1}{g(y)}\frac{dy}{dx} = f(x) \quad .$$
(Note: At this point we've "separated the variables", getting all the y's and derivatives of y on one side, and all the x's on the other.)
3. Integrate both sides with respect to x, making use of the fact that
$$\int \frac{1}{g(y)}\frac{dy}{dx}\,dx = \int \frac{1}{g(y)}\,dy \quad .$$
4. Solve the resulting equation for y.

There are a few issues that can arise in some of these steps, and we will have to slightly refine this procedure to address those issues. Before doing that, though, let us practice with another differential equation for which the above approach can be applied without any difficulty.

!►Example 4.4: *Consider solving the initial-value problem*
$$\frac{dy}{dx} = -\frac{x}{y-3} \qquad \text{with} \quad y(0) = 1 \quad .$$
Here,
$$\frac{dy}{dx} = f(x)g(y) \qquad \text{with} \quad f(x) = -x \quad \text{and} \quad g(y) = \frac{1}{y-3} \quad ,$$
and "dividing through by $g(y)$" is the same as multiplying through by $y-3$. Doing so, and then integrating both sides with respect to x, we get the following:
$$[y-3]\frac{dy}{dx} = -x$$
$$\hookrightarrow \qquad \int [y-3]\frac{dy}{dx}\,dx = -\int x\,dx$$
$$\hookrightarrow \qquad \int [y-3]\,dy = -\int x\,dx$$
$$\hookrightarrow \qquad \tfrac{1}{2}y^2 - 3y = -\tfrac{1}{2}x^2 + c \quad .$$
Though hardly necessary, we can multiply through by 2, obtaining the slightly simpler expression
$$y^2 - 6y = -x^2 + 2c \quad .$$
We are now faced with the less-than-trivial task of solving the last equation for y in terms of x. Since the left-hand side looks something like a quadratic for y, let us rewrite this equation as
$$y^2 - 6y + \left[x^2 - 2c\right] = 0$$
so that we can apply the quadratic formula to solve for y. Applying that venerable formula, we get
$$y = \frac{-(-6) \pm \sqrt{(-6)^2 - 4\left[x^2 - 2c\right]}}{2} = 3 \pm \sqrt{9 - x^2 + 2c} \quad ,$$

which, since $9 + 2c$ *is just another unknown constant, can be written a little more simply as*

$$y = 3 \pm \sqrt{a - x^2} \quad . \tag{4.6}$$

This is the general solution to our differential equation.

Now for the initial-value problem. Combining the general solution just derived with the given initial value at $x = 0$ *yields*

$$1 = y(0) = 3 \pm \sqrt{a - 0^2} = 3 \pm \sqrt{a} \quad .$$

So

$$\pm\sqrt{a} = -2 \quad .$$

This means that $a = 4$ *, and that we must use the* negative *root in formula (4.6) for* y *. Thus, the solution to our initial-value problem is*

$$y = 3 - \sqrt{4 - x^2} \quad .$$

4.2 Constant Solutions
Avoiding Division by Zero

In the above procedure for solving

$$\frac{dy}{dx} = f(x)g(y) \quad ,$$

we divided both sides by $g(y)$. This requires, of course, that $g(y)$ not be zero — which is often *not* the case.

!►Example 4.5: *Consider solving*

$$\frac{dy}{dx} = 2x(y - 5) \quad .$$

As long as $y \neq 5$ *, we can divide through by* $y - 5$ *and follow our basic procedure:*

$$\frac{1}{y-5}\frac{dy}{dx} = 2x$$

$$\hookrightarrow \qquad \int \frac{1}{y-5}\frac{dy}{dx}\,dx = \int 2x\,dx$$

$$\hookrightarrow \qquad \int \frac{1}{y-5}\,dy = \int 2x\,dx$$

$$\hookrightarrow \qquad \ln|y - 5| = x^2 + c$$

$$\hookrightarrow \qquad |y - 5| = e^{x^2+c} = e^{x^2}e^c$$

$$\hookrightarrow \qquad y - 5 = \pm e^{x^2}e^c \quad .$$

So, assuming $y \neq 5$ *, we get*

$$y = 5 \pm e^c e^{x^2} \quad .$$

Notice that, because $e^c \neq 0$ *for every real value* c *, this formula for* y *never gives us* $y = 5$ *for any real choice of* c *and* x *.*

But what about the case where $y = 5$ *?*

Well, suppose $y = 5$ *. To be more specific, let* y *be the constant function*

$$y(x) = 5 \qquad \text{for every} \quad x \quad ,$$

and plug this constant function into our differential equation

$$\frac{dy}{dx} = 2x(y-5) \quad .$$

Recalling (again) that derivatives of constants are zero, we get

$$0 = 2x(5-5) \quad ,$$

which is certainly a true equation. So $y = 5$ *is a solution. In fact, it is one of those "constant" solutions we discussed in the previous chapter.*

Combining all the above, we see that the "general solution" to the given differential equation is actually the set consisting of the solutions

$$y(x) = 5 \qquad \text{and} \qquad y(x) = 5 \pm e^c e^{x^2} \quad .$$

Now consider the general case, where we seek all possible solutions to

$$\frac{dy}{dx} = f(x)g(y) \quad .$$

If y_0 is any single value for which

$$g(y_0) = 0 \quad ,$$

then plugging the corresponding constant function

$$y(x) = y_0 \qquad \text{for all} \quad x$$

into the differential equation gives, after a trivial bit of computation,

$$0 = 0 \quad ,$$

showing that

$$y(x) = y_0 \quad \text{is a constant solution to} \quad \frac{dy}{dx} = f(x)g(y) \quad ,$$

just as we saw (in the above example) that

$$y(x) = 5 \quad \text{is a constant solution to} \quad \frac{dy}{dx} = 2x(y-5) \quad .$$

Conversely, suppose $y = y_0$ is a constant solution to

$$\frac{dy}{dx} = f(x)g(y)$$

(and f is not the zero function). Then the equation is valid with y replaced by the constant y_0 , giving us

$$0 = f(x)g(y_0) \quad ,$$

which, in turn, means that y_0 must be a constant such that

$$g(y_0) = 0 \quad .$$

What all this shows is that our basic method for solving separable equations may miss the constant solutions because those solutions correspond to a division by zero in our basic method.[2]

Because constant solutions are often important in understanding the physical process the differential equation might be modeling, let us be careful to find them. Accordingly, we will insert the following step into our procedure on page 68 for solving separable equations:

- Identify all constant solutions by finding all values y_0, y_1, y_2, … such that

  $$g(y_k) = 0 \quad ,$$

  and then write down

  $$y(x) = y_0 \quad , \quad y(x) = y_1 \quad , \quad y(x) = y_2 \quad , \quad \ldots \quad .$$

  (These are the constant solutions.)

(And we will renumber the other steps as appropriate.)

Sometimes, the formula obtained by our basic procedure for solving can be 'tweaked' to also account for the constant solutions. A standard 'tweak' can be seen by reconsidering the general solution obtained in our last example.

!►Example 4.6: *The general solution obtained in the previous example was the set containing*

$$y(x) = 5 \qquad \text{and} \qquad y(x) = 5 \pm e^c e^{x^2} \quad .$$

If we let $A = \pm e^c$, the second equation reduces to

$$y(x) = 5 + Ae^{x^2} \quad .$$

Remember, though, $A = \pm e^c$ can be any positive or negative number, but cannot be zero (because of the nature of the exponential function). So, by our definition of A, our general solution is

$$y(x) = 5 \tag{4.7a}$$

and

$$y(x) = 5 + Ae^{x^2} \quad \text{where } A \text{ can be any nonzero real number} \quad . \tag{4.7b}$$

However, if we allow A to be zero, then equation (4.7b) reduces to equation (4.7a),

$$y(x) = 5 + 0 \cdot e^{x^2} = 5 \quad ,$$

which means the entire set of possible solutions can be expressed more simply as

$$y(x) = 5 + Ae^{x^2}$$

where A is an arbitrary constant with no restrictions on its possible values.

[2] Because $g(y_0) = 0$ is a 'singular' value for division, many authors refer to constant solutions of separable equations as *singular* solutions.

In the future, we will usually express our general solutions as simply as practical, with the trick of letting

$$A = \pm e^c \quad \text{or} \quad 0$$

often being used without comment. Keep in mind, though, that the sort of tweaking just described is not always possible.

?►Exercise 4.1: *Verify that the general solution to*

$$\frac{dy}{dx} = -y^2$$

is given by the set consisting of

$$y(x) = 0 \qquad \text{and} \qquad y(x) = \frac{1}{x+c} \quad .$$

Is there any way to rewrite these two formulas for $y(x)$ as a single formula using just one arbitrary constant?

The Importance of Constant Solutions

Even if we can use the same general formula to describe all the solutions (constant and otherwise), it is often worthwhile to explicitly identify any constant solutions. To see this, let us now solve the differential equation from Chapter 1 describing a falling object when we take into account air resistance.

!►Example 4.7: *Let $v = v(t)$ be the velocity (in meters per second) at time t of some object of mass m plummeting towards the ground. In Chapter 1, we decided that F_{air}, the force of air resistance acting on the falling body, could be described by*

$$F_{\text{air}} = -\gamma v$$

where γ was some positive constant dependent on the size and shape of the object (and probably determined by experiment). Using this, we obtained the differential equation

$$\frac{dv}{dt} = -9.8 - \kappa v \qquad \text{where} \quad \kappa = \frac{\gamma}{m} \quad .$$

This is a relatively simple separable equation. Assuming v equals a constant v_0 yields

$$0 = -9.8 - \kappa v_0 \quad \Longrightarrow \quad v_0 = -\frac{9.8}{\kappa} = -\frac{9.8m}{\gamma} \quad .$$

So, we have one constant solution,

$$v(t) = v_0 \qquad \text{for all} \quad t$$

where

$$v_0 = -\frac{9.8}{\kappa} = -\frac{9.8m}{\gamma} \quad .$$

For reasons that will soon become clear, v_0 is called the terminal velocity *of the object that is falling.*

To find the other possible solutions, we assume $v \neq v_0$ and proceed:

$$
\begin{aligned}
& & \frac{dv}{dt} &= -9.8 - \kappa v \\
&\hookrightarrow & \frac{1}{9.8+\kappa v}\frac{dv}{dt} &= -1 \\
&\hookrightarrow & \int \frac{1}{9.8+\kappa v}\frac{dv}{dt}\,dt &= -\int 1\,dt \\
&\hookrightarrow & \int \frac{1}{9.8+\kappa v}\,dv &= -\int dt \\
&\hookrightarrow & \frac{1}{\kappa}\ln|9.8+\kappa v| &= -t + c \\
&\hookrightarrow & \ln|9.8+\kappa v| &= -\kappa t + \kappa c \\
&\hookrightarrow & 9.8 + \kappa v &= \pm e^{-\kappa t + \kappa c} \\
&\hookrightarrow & v(t) &= \frac{1}{\kappa}\left[-9.8 \pm e^{\kappa c}e^{-\kappa t}\right] \quad .
\end{aligned}
$$

Since $v_0 = -9.8\kappa^{-1}$, the last equation reduces to

$$v(t) = v_0 + Ae^{-\kappa t} \qquad \text{where} \quad A = \pm\frac{1}{\kappa}e^{\kappa c} \quad .$$

This formula for $v(t)$ yields the constant solution, $v = v_0$, if we allow $A = 0$. Thus, letting A be a completely arbitrary constant, we have that

$$v(t) = v_0 + Ae^{-\kappa t} \tag{4.8a}$$

where

$$v_0 = -\frac{9.8m}{\gamma} \qquad \text{and} \qquad \kappa = \frac{\gamma}{m} \tag{4.8b}$$

describes all possible solutions to the differential equation of interest here. The graphs of some possible solutions (assuming a terminal velocity of -10 meters/second) are sketched in Figure 4.1.

Notice how the constant in the constant solution, v_0, appears in the general solution (equation (4.8a)). More importantly, notice that the exponential term in this solution rapidly goes to zero as t increases, so

$$v(t) = v_0 + Ae^{-\kappa t} \quad \to \quad v(t) = v_0 \qquad \text{as} \quad t \to \infty \quad .$$

This is graphically obvious in Figure 4.1. Consequently, no matter what the initial velocity and initial height were, eventually the velocity of this falling object will be very close to v_0 (provided it doesn't hit the ground first). That is why v_0 is called the terminal velocity. That is also why that constant solution is so important here (and is appropriately also called the equilibrium solution). It accurately predicts the final velocity of any object falling from a sufficiently high height. And if you are that falling object, then that velocity[3] is probably a major concern.

[3] between 120 and 150 miles per hour for a typical human body

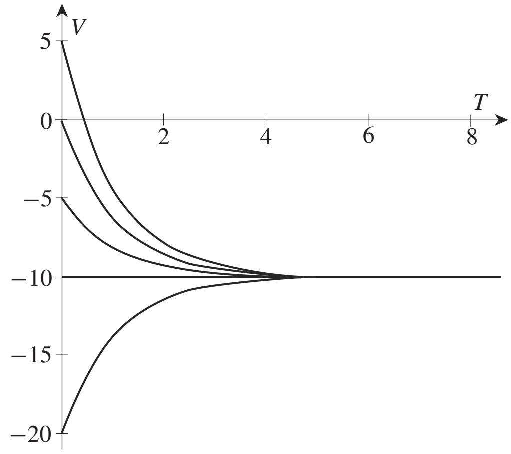


Figure 4.1: Graphs of the velocity of a falling object during the first 8 seconds of its fall assuming a terminal velocity of -10 meters per second. Each graph corresponds to a different initial velocity.

4.3 Explicit Versus Implicit Solutions

Thus far, we have been able to find explicit formulas for all of our solutions; that is, we have been able to carry out the last step in our basic procedure — that of solving the resulting (integrated) equation for y in terms of x — obtaining

$$y = y(x) \qquad \text{where } y(x) \text{ is some formula of } x \text{ (with no } y\text{'s).}$$

For example, as the general solution to

$$\frac{dy}{dx} - x^2y^2 = x^2 \quad ,$$

we obtained (in Example 4.3)

$$y = \underbrace{\tan\left(\frac{1}{3}x^3 + c\right)}_{y(x)} \quad .$$

Unfortunately, this is not always possible.

!▶*Example 4.8:* *Consider*

$$\frac{dy}{dx} = \frac{x+1}{8 + 2\pi \sin(\pi y)} \quad .$$

In this case,

$$g(y) = \frac{1}{8 + 2\pi \sin(\pi y)} \quad ,$$

which can never be zero. So there are no constant solutions, and we can blithely proceed with our procedure. Doing so:

$$\frac{dy}{dx} = \frac{x+1}{8 + 2\pi \sin(\pi y)}$$

$$\hookrightarrow \qquad [8 + 2\pi \sin(\pi y)]\frac{dy}{dx} = x + 1$$

$$\hookrightarrow \qquad \int [8 + 2\pi \sin(\pi y)] \frac{dy}{dx}\, dx = \int x + 1\, dx$$

$$\hookrightarrow \qquad \int [8 + 2\pi \sin(\pi y)]\, dy = \int x + 1\, dx$$

$$\hookrightarrow \qquad 8y - 2\cos(\pi y) = \frac{1}{2}x^2 + x + c \quad .$$

The next step would be to solve the last equation for y in terms of x. But look at that last equation. Can you solve it for y as a formula of x? Neither can anyone else. So we are not able to obtain an explicit formula for y. At best, we can say that $y = y(x)$ satisfies the equation

$$8y - 2\cos(\pi y) = \frac{1}{2}x^2 + x + c \quad .$$

Still, this equation is not without value. It does implicitly describe the possible relations between x and y. In particular, the graphs of this equation can be sketched for different values of c (we'll do this later on in this chapter). These graphs, in turn, give you the graphs you would obtain for $y(x)$ if you could actually find the formula for $y(x)$.

In practice, we must deal with both "explicit" and "implicit" solutions to differential equations. When we have an explicit formula for the solution in terms of the variable, that is, we have something of the form

$$y = y(x) \qquad \text{where } y(x) \text{ is some formula of } x \text{ (with no } y\text{'s)} \quad , \tag{4.9}$$

then we say that we have an *explicit solution* to our differential equation. Technically, it is that "formula of x" in equation (4.9) which is the *explicit solution*. In practice, though, it is common to refer to the entire equation as "an explicit solution". For example, we found that the solution to

$$\frac{dy}{dx} - x^2 y^2 = x^2$$

is explicitly given by

$$y = \tan\left(\frac{1}{3}x^3 + c\right) \quad .$$

Strictly speaking, the explicit solution here is the formula

$$\tan\left(\frac{1}{3}x^3 + c\right) \quad .$$

That, of course, is what is really meant when someone answers the question

$$\textit{What is the explicit solution to} \qquad \frac{dy}{dx} - x^2 y^2 = x^2 \qquad ?$$

with the equation

$$y = \tan\left(\frac{1}{3}x^3 + c\right) \quad .$$

If, on the other hand, we have an equation (other than something like (4.9)) involving the solution and the variable, then that *equation* is called an *implicit solution*. In trying to solve the differential equation in Example 4.8,

$$\frac{dy}{dx} = \frac{x + 1}{8 + 2\pi \sin(\pi y)} \quad ,$$

we derived the equation

$$8y - 2\cos(\pi y) = \frac{1}{2}x^2 + x + c \quad .$$

This equation is an implicit solution for the given differential equation.[4]