

# Some Mathematical Background for Mathematical Epidemiology

Fred Brauer

June 30, 2008



# Contents

<b>Prologue: Mathematical Epidemiology is not an Oxymoron</b>	<b>vii</b>
<b>Preface</b>	<b>ix</b>
<b>I CALCULUS</b>	<b>1</b>
<b>1 Functions and Graphs</b>	<b>3</b>
1.1 Coordinates . . . . .	3
1.2 Functions . . . . .	4
1.2.1 Graphs . . . . .	5
1.2.2 Elementary functions . . . . .	7
1.2.3 Exponential functions and natural logarithms . . . . .	9
1.2.4 Trigonometric functions . . . . .	11
1.3 Some exercises . . . . .	12
<b>2 The Derivative</b>	<b>13</b>
2.1 The meaning of the derivative . . . . .	13
2.2 Limits of Functions . . . . .	15
2.3 Calculation of Derivatives . . . . .	18
2.4 Applications of the Derivative . . . . .	21
2.4.1 Curve sketching . . . . .	22
2.4.2 Maximum - minimum problems . . . . .	23
2.4.3 Optimization . . . . .	23
2.4.4 Exponential growth and decay . . . . .	24
2.5 Local Linearity . . . . .	25
2.5.1 Application: The chain rule . . . . .	26
2.6 Some exercises . . . . .	26
<b>3 The Integral</b>	<b>29</b>
3.1 The Indefinite Integral . . . . .	29
3.2 The Definite Integral . . . . .	33
3.3 The Fundamental Theorem of Calculus . . . . .	34
3.4 Some exercises . . . . .	36

<b>4</b>	<b>Multivariable Calculus</b>	<b>39</b>
4.1	Functions of Two Variables . . . . .	39
4.1.1	Graphic representation of functions . . . . .	40
4.1.2	Linear functions . . . . .	42
4.2	Limits, Continuity, and Partial Derivatives . . . . .	42
4.2.1	Limits and continuity . . . . .	42
4.2.2	Partial derivatives . . . . .	44
4.3	Local Linearity . . . . .	45
4.4	Some exercises . . . . .	48
<b>II</b>	<b>MATRIX ALGEBRA</b>	<b>51</b>
<b>5</b>	<b>Vectors and Matrices</b>	<b>53</b>
5.1	Introduction . . . . .	53
5.2	Vectors and Matrices . . . . .	54
5.3	Systems of Linear Equations . . . . .	58
5.4	The Inverse Matrix . . . . .	59
5.5	Determinants . . . . .	60
5.6	Eigenvalues and Eigenvectors . . . . .	62
5.7	Some exercises . . . . .	66
<b>III</b>	<b>ORDINARY DIFFERENTIAL EQUATIONS</b>	<b>69</b>
<b>6</b>	<b>First Order Equations</b>	<b>71</b>
6.1	Exponential Growth and Decay . . . . .	71
6.1.1	Radioactive decay . . . . .	73
6.2	Solutions and Direction Fields . . . . .	75
6.3	Equations with Variables Separable . . . . .	79
6.4	Some Applications of Separable Equations . . . . .	86
6.4.1	The spread of infectious diseases . . . . .	86
6.4.2	Drug dosage . . . . .	88
6.4.3	Allometry . . . . .	89
6.5	Qualitative Properties . . . . .	90
6.6	Some Qualitative Applications . . . . .	98
6.6.1	Population growth with harvesting . . . . .	98
6.6.2	The spread of infectious diseases . . . . .	101
6.7	Some exercises . . . . .	102
<b>7</b>	<b>Systems</b>	<b>105</b>
7.1	The Phase Plane . . . . .	105
7.2	Linearization at an Equilibrium . . . . .	107
7.3	Linear Systems with Constant Coefficients . . . . .	110
7.4	Stability of Equilibria . . . . .	118
7.5	Some Applications . . . . .	123

7.5.1	Predator-prey systems . . . . .	123
7.5.2	An epidemiological model . . . . .	126
7.6	Some exercises . . . . .	131

## **IV FURTHER TOPICS IN CALCULUS 135**

### **8 Double Integrals 137**

8.1	Double integrals over a rectangle . . . . .	137
8.2	Double integrals over more general regions . . . . .	142
8.3	Some exercises . . . . .	146

### **9 Power Series Expansions 149**

9.1	The Mean Value Theorem . . . . .	149
9.2	Taylor Polynomials . . . . .	152
9.3	Taylor's Theorem . . . . .	157
9.4	The Taylor Series of a Function . . . . .	160
9.5	Convergence of Power Series . . . . .	166
9.6	Some exercises . . . . .	168



# Prologue: Mathematical Epidemiology is not an Oxymoron

What can mathematical modelling contribute to epidemiology? Usually, scientific experiments are designed to obtain information and test hypotheses. The normal process of scientific progress is to observe a phenomenon, hypothesize an explanation, and then devise an experiment to test the hypothesis. A mathematical model is a mathematical description of the situation based on the hypotheses, and the solution of the model gives conclusions which may be compared with experimental results. For example, we might wish to compare two different management strategies for a disease outbreak. However, experiments in epidemiology with controls are often difficult or impossible to design and even if it is possible to arrange an experiment there are serious ethical questions involved in withholding treatment from a control group. Data about a disease outbreak is often obtainable only after the fact from reports of epidemics or of endemic disease levels, but the data may be inaccurate or incomplete. In order to describe the course of a future disease outbreak, formulation of a mathematical model may be the *only* way to compare the effect of different management strategies.

Every model is either too simple to be an accurate description or too complicated to analyze. A mathematician may choose to start with a simple incomplete model to obtain qualitative information, while an epidemiologist may object that the model is too simplistic and omits important aspects of the situation. One way to proceed would be to begin with a simple model and then add more structure to the model to see how much this alters the predicted behaviour.

Mathematical modeling in epidemiology provides understanding of the underlying mechanisms that influence the spread of disease and, in the process, it suggests control strategies. In fact, models may identify behaviours that are unclear in experimental data - often because data are non-reproducible and the number of data points is limited and subject to errors in measurement. For example, one of the fundamental results in mathematical epidemiology is that most mathematical epidemic models, including those that include a high degree of heterogeneity, exhibit “threshold” behavior which in epidemiological terms

can be stated as follows: *If the average number of secondary infections caused by an average infective is less than one a disease will die out, while if it exceeds one the disease will persist.* This broad principle, consistent with observations and quantified via epidemiological models, has been used to estimate the effectiveness of vaccination policies and the likelihood that a disease may be eliminated or eradicated. The possibility of eliminating smallpox worldwide was suggested by a simple mathematical model. Even if it is not possible to verify hypotheses accurately, agreement with hypotheses of a qualitative nature is often valuable. Expressions for the basic reproductive number for HIV in various populations can be used to test the possible effectiveness of vaccines that may provide temporary protection by reducing either HIV-infectiousness or susceptibility to HIV. Models could be used to estimate how widespread a vaccination plan must be to prevent or reduce the spread of HIV.

In the mathematical modeling of disease transmission, as in most other areas of mathematical modeling, there is a trade-off between simple models, which omit most details and are designed only to highlight general qualitative behavior, and detailed models, usually designed for specific situations including short-term quantitative predictions. Detailed models are generally difficult or impossible to solve analytically and hence their usefulness for theoretical purposes is limited, although their strategic value may be high.

There are many different types of models in epidemiology, requiring different mathematical techniques. For an introduction to mathematical epidemiology the central mathematical topics are calculus, ordinary differential equations from a qualitative point of view, matrix algebra, and probability and statistics.

# Preface

These notes are intended for students or workers in epidemiology or public health who may have learned elementary calculus and differential equations, possibly in the distant past and possibly in a form that did not convince them that mathematics could be useful in their chosen field. We have tried to present some basic mathematics to make it possible for “calculus victims” to become “calculus users” in epidemiology.

While the material is aimed at students whose primary interest is not mathematics, it consists of topics in mathematics, not epidemiology. The purpose of the notes is to describe mathematical concepts and techniques that will be useful for review in studying mathematical epidemiology, not to be part of a course in mathematical epidemiology. However, we make no pretense of mathematical rigour. Results are stated but not necessarily proved, and hypotheses are not always stated precisely. The intent is to provide an honest impression for the user of mathematics. However, it is intended only as a review and is *not* meant to be a course for learning this material for the first time. Readers may need to return to their calculus texts for more detailed review where necessary. In addition to descriptions of the mathematics we have included examples from and applications to the biological sciences, especially population biology, and epidemiology. Because the approach to differential equations may be somewhat different from the course taken previously, we have included more detail and more examples here than in the material on calculus.

In many calculus courses the emphasis is on mastering computational techniques. This is true especially in courses aimed at students of the physical sciences or engineering, who are likely to need to use these techniques frequently. On the other hand, for students less likely to need to do many mathematical computations an emphasis on techniques may make the course uninteresting and may not persuade the students that an acquaintance with mathematics may be useful for them in their chosen fields of study. In these notes we try to emphasize ideas and principles over computations, and we try to include enough examples from the biological sciences to persuade epidemiologists that the material is useful for them. We have not, however, included exercises even though working many problems is necessary for real understanding.

We believe that the basics of calculus of functions of one variable, some topics in the differential calculus of functions of several (especially two) variables, the elements of matrix algebra, differential equations of first order, and systems

of two first order differential equations comprise a first level of mathematical understanding for epidemiologists. This material is included in these notes along with some additional topics in calculus such as double integrals and material on power series needed for understanding of generating functions, which appear in probabilistic contexts such as stochastic models and contact networks, but is not needed elsewhere. An understanding of elementary probability, not included here, is also essential. Also, there are topics in mathematical epidemiology that require more advanced mathematics. For example, the study of models with spatial dependence or age structure requires some acquaintance with partial differential equations.

It is important to remember that calculus is not just a collection of techniques. It is also a logical discipline and all statements require proofs. We do not emphasize proofs in these notes, but the reader should be aware that statements should be accepted only if they have been proved rigorously (by someone). Our purpose in these notes is to tell the truth and nothing but the truth, but not necessarily the whole truth. We make statements that have been proved, but usually do not give their proofs.

The first section of these notes contains a brief outline of the major topics in calculus that will be useful in the study of mathematical epidemiology. The second section contains a very brief introduction to matrix algebra. The third section contains an outline of a qualitative approach to ordinary differential equations and two-dimensional systems of ordinary differential equations. This material, together with an introduction to probability and statistics, not included here, should provide a basis for study of mathematical models in epidemiology. The fourth section contains some additional topics in calculus useful in some applications. We hope that these notes may also encourage the study of more advanced mathematical topics that can also be useful.

**Part I**  
**CALCULUS**



# Chapter 1

## Functions and Graphs

### 1.1 Coordinates

In order to represent functions geometrically we make use of *Cartesian coordinates* in the plane. We draw two perpendicular *axes*, normally with the  $x$ -axis horizontal and  $y$ -axis vertical. To each point in the plane we assign a pair of coordinates  $(x, y)$  in the following way. The *abscissa*, or  $x$ -coordinate, measures the horizontal distance (perpendicular to the  $y$ -axis) between the  $y$ -axis and the point, with points to the right of the  $y$ -axis having a positive  $x$ -coordinate and points to the left of the  $y$ -axis having a negative  $x$ -coordinate. The *ordinate*, or  $y$ -coordinate, measures the vertical distance (perpendicular to the  $x$ -axis) between the  $x$ -axis and the point, with points above the  $x$ -axis having a positive  $y$ -coordinate and points below the  $x$ -axis having a negative  $y$ -coordinate. The axes divide the plane into four *quadrants*, called the first through fourth quadrants, depending on the signs of the coordinates.

For each point in the plane there is a unique pair of coordinates which describe the point and for each pair of real numbers there is a unique point in the plane whose coordinates are these numbers (in order). We frequently speak of “the point  $(x, y)$ ” rather than the more precise form “the point whose coordinates are  $(x, y)$ ”; no confusion can arise because of the uniqueness of the correspondence between points and pairs of coordinates.

For example, suppose we count the number of people in a given community who are infected with an epidemic disease on different days. We may plot this information using the number of days since counting began as the  $x$ -coordinate and the number of people infected on that day as the  $y$ -coordinate. This gives a representation as in Figure 1.1, which is actually derived from data for the Great Plague in the village of Eyam in 1666.

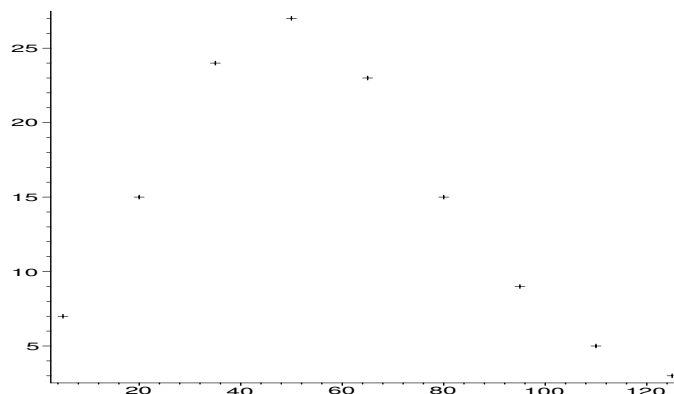


Figure 1.1: Some epidemic data

## 1.2 Functions

We say that  $y$  is a *function* of  $x$  if there is a rule assigning a value of  $y$  to each value of  $x$ . More precisely,  $x$  is a *variable*, called the *independent variable*, which may range over some set of values called the *domain* of the function. To each value in the domain, the function assigns a value of the *dependent variable*  $y$ , and  $y$  then ranges over another set of values called the *range* of the function. A function is given a name such as  $f$ , and we use the notation

$$y = f(x)$$

If  $x_0$  is a value in the domain to which the function  $f$  assigns the value  $y_0$ , we write

$$y_0 = f(x_0)$$

Usually,  $x$ , called the independent variable, and  $y$ , called the dependent variable are both real numbers and the domain and range are both intervals, but this is not required as part of the definition. The specification of a function must include specification of the domain. Sometimes there are obvious restrictions. For example, the function  $f(x) = \sqrt{1 - x^2}$  can be defined only if  $1 - x^2 \geq 0$ , or  $-1 \leq x \leq 1$ . Usually, a function is given by an analytic expression, but not necessarily. For example, a function could be defined by different expressions on different intervals, or by a table of values. However, the definition of a function requires that only one value of  $y$  be given to correspond to each value of  $x$ .

Most of the functions that we will use are *continuous*. A precise definition of continuity requires the concept of *limit* of a function, and we will delay the introduction of this idea until the next section. Intuitively, a function is continuous if its graph contains no breaks, that is, if the graph can be drawn without lifting the pencil from the paper.

While we have spoken of  $y$  as a function of  $x$ , other names are often used for the variables. In many problems the independent variable is time, denoted by  $t$ .

The information contained in Figure 1.1 may be extrapolated to give the graph of a function by drawing straight lines from one point to the next, as in Figure 1.2. This graph is the epidemic curve describing the data. While we

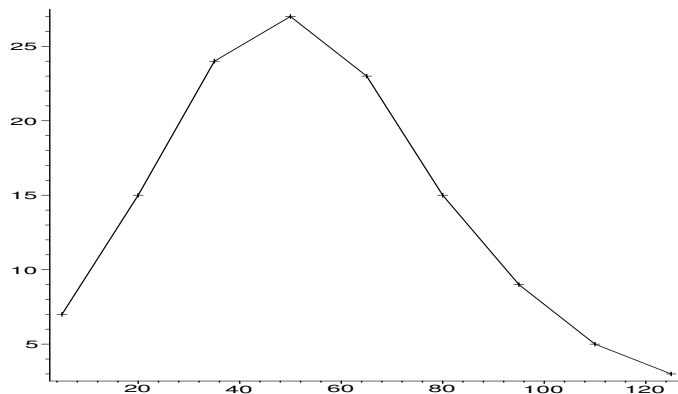


Figure 1.2: An epidemic curve

have chosen to draw this graph as a sequence of line segments, in practice it is more customary to connect the dots by smooth curves.

Mathematically functions may be defined simply as relations between variables, but in other sciences functions are usually defined to describe quantities with a scientific meaning. For example, in economics one sometimes postulates that there is a relation between the price  $p$  of some good and the quantity  $q$  of this good which can be sold at a given price, called the *demand* for the good. This relation is called the *demand relation*. In mechanics, we may describe the position of a moving body as a function of time. In population biology, we may describe the population size at a given time as a function. In epidemiology, we may describe the number of infectives in a population as a function of time. Here, we may count the number of infectives once a day and describe the number of infectives for a given day, but we may also think of the number of infectives as varying continuously over each day.

### 1.2.1 Graphs

The *graph* of a function  $f$  is defined to be the set of all points  $(x, y)$  such that  $x$  is in the domain of the function and  $y = f(x)$ . The graph of a function is a curve, and the requirement that only one value of  $y$  can correspond to any value of  $x$  implies that a vertical line can intersect the graph of a function only once. On occasion, we may be imprecise and speak of a function defined by a relation such as  $y^2 = x$  whose graph intersects every line  $x = x_0$  with  $x_0 > 0$  twice, namely at  $(x_0, \sqrt{x_0})$  and at  $(x_0, -\sqrt{x_0})$ . In such a situation, we should really separate the “function” into two branches  $y = \sqrt{x}$  and  $y = -\sqrt{x}$ , but we may sometimes call it a “multiple-valued function”. It is customary to use

the horizontal axis for the independent variable and the vertical axis for the dependent variable.

**Example 1.** During the course of a disease outbreak, we may report the number of people infected with the disease each day. This amounts to defining a function  $I(t)$  representing the number of people ‘infected with the disease as a function of time. Typically, the graph of this function is similar to the graph shown in Figure 1.1.

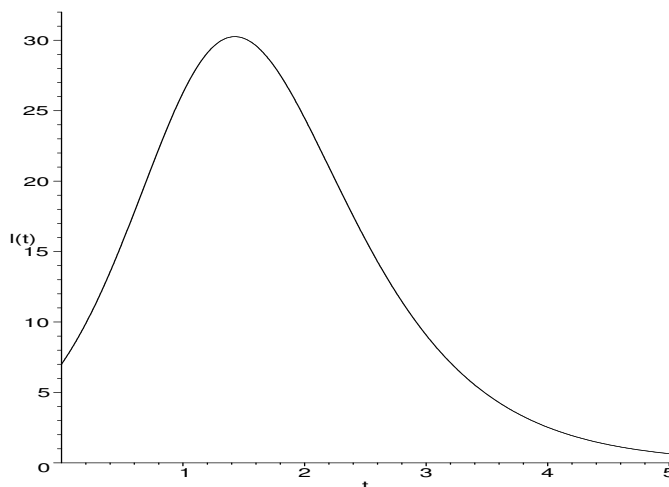


Figure 1.3:  $I$  as a function of  $t$ .

The simplest examples of functions are constant functions,  $y = b$ , and linear functions

$$y = mx + b \quad (1.1)$$

The graph of a constant function is a horizontal line, and the graph of the linear function ((8.1) is a line with slope  $m$ . The slope describes the angle made by the graph of the function with the  $x$ -axis, and it also represents the rate of change of the linear function (8.1). By the rate of change of the function, we mean that a change in  $x$  produces a change  $m$  times as large in  $y$ ; in other words, the change in  $y$  divided by the change in  $x$  is the slope  $m$ . If the slope of the line is positive, the line goes upward to the right, and  $y$  increases when  $x$  increases. If the slope of the line is negative, the line goes downward to the right, and  $y$  decreases when  $x$  increases (Figures 1.4, 1.5). Much of the differential calculus is an attempt to extend these ideas of slope and increasing and decreasing functions to functions which are not linear (Figure 1.6).

In order to graph a linear function (8.1) we need only plot any two points which satisfy the equation (8.1) and then draw the line joining them. One way to draw the graph of a more complicated function would be to plot many points on the graph. Modern technology has created computers and graphing

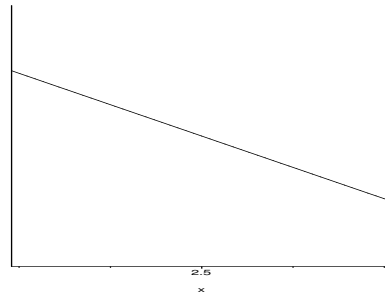
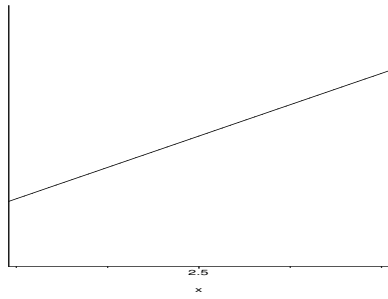


Figure 1.4: A line with positive slope      Figure 1.5: A line with negative slope

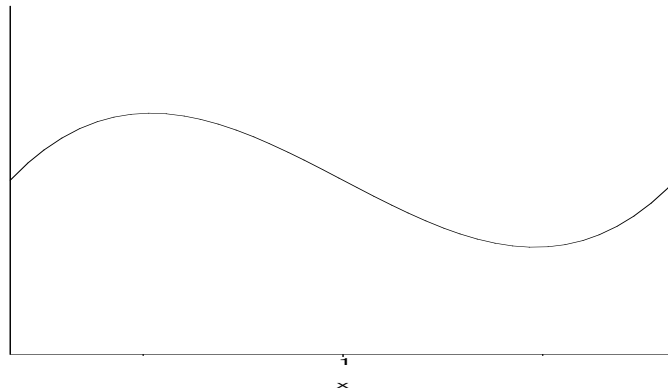


Figure 1.6: A function with varying slope

calculators which will perform this task when given instructions in a language that can be learned readily by calculus students. However, it is important to develop an understanding of the relationships between properties of a function and the nature of its graph, and this can not be acquired without some analysis of the behavior of functions. The ideas of calculus are important in learning about the nature of the graph of a function, and this is one of the areas of applications which we will describe in Section 2.4.

### 1.2.2 Elementary functions

There are some particular functions that we shall use as building blocks for the construction of other functions. Functions which are considered to be basic are called *elementary functions*. However, the choice of a list of elementary functions depends to some extent on the purposes for which they will be used, and different lists may contain different functions.

Functions can be combined in various ways to produce other functions. For example, if  $f(x)$  and  $g(x)$  are two functions, we can form the sum  $f(x) + g(x)$ ,

the product  $cf(x)$  of a constant and a function, the product  $f(x)g(x)$ , and the quotient  $f(x)/g(x)$ .

Perhaps the simplest functions are the powers  $x^n$ , where  $n$  is a positive integer. In particular, if  $n = 0$  the function  $x^n$  is the constant function 1 and if  $n = 1$  the function  $x^n$  is the linear function  $x$ . Functions  $x^n$  with different values of  $n$  may be combined to form *polynomials*; a polynomial of degree  $n$  is a function of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where  $a_0, a_1, \dots, a_n$  are real numbers with  $a_n \neq 0$ , and  $n$  is a positive integer. A *rational function* is a quotient of two polynomials, but note that a rational function can not be defined for values of  $x$  such that the denominator is zero.

**Example 1.** The rate of a chemical reaction is assumed to be governed by the *law of mass action*. This law states that the rate at which a reaction proceeds is proportional to the product of the concentrations of the reactants. Let us consider a reaction in which two substances  $A$  and  $B$  with initial concentrations  $a$  and  $b$  respectively react to form a third substance  $AB$ , with one molecule of each of  $A$  and  $B$  combining to form one molecule of  $AB$ . Let  $x$  denote the concentration of  $AB$  at some time during the reaction. Then the concentrations of  $A$  and  $B$  are  $a - x$  and  $b - x$  respectively. According to the law of mass action, the rate of the reaction is  $k(a - x)(b - x)$ , where  $k$  is some positive constant, called the rate constant. This reaction rate function

$$R(x) = k(a - x)(b - x)$$

is a polynomial function. The rate of reaction depends on the concentration of the product of the reaction. If we wish to find the quantity of product  $x$  at a given time we would have to solve the *differential equation*

$$\frac{dx}{dt} = R(x) = k(a - x)(b - x).$$

The solution of differential equations is a topic to which we will return in Chapter 6.

**Example 2** (The Monod growth function). Suppose the per capita growth rate of an organism depends on the concentration of a nutrient. We let the concentration of the nutrient be  $x$ . If the growth rate has a saturation effect, the per capita growth rate is often described by the *Monod* growth function

$$r(x) = \frac{ax}{A + x}.$$

where  $a$  and  $A$  are positive constants. This function increases from zero when  $x = 0$  and approaches  $a$  when  $x$  becomes very large. The parameter  $A$  represents the value of  $x$  for which the function takes the value  $a/2$ , half its ultimate value as  $x \rightarrow \infty$ . The graph is shown in Figure 1.7.

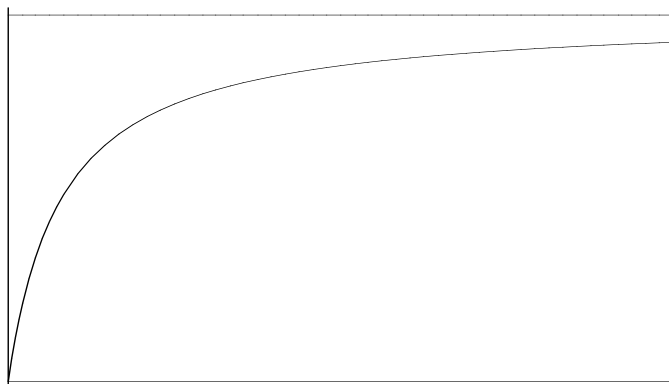


Figure 1.7: A Monod function

Powers obey the *laws of exponents*

$$x^n x^m = x^{n+m}, \quad x^n / x^m = x^{n-m} \quad (n > m), \quad (x^n)^m = x^{nm} \quad (1.2)$$

provided  $n$  and  $m$  are positive integers. We define negative and fractional powers in such a way that these laws remain valid. Then for negative integer powers we must define

$$x^0 = 1, \quad x^{-p} = 1/x^p. \quad (1.3)$$

To define  $x^{1/q}$  when  $q$  is a positive integer, we write  $y = x^{1/q}$  and then require  $y^q = x$ . Note that if  $q$  is an even integer and  $x$  is negative, this is not possible. Thus the function  $x^{1/q}$  with  $q$  an even integer can be defined only for  $x \geq 0$ . We can now define rational powers by

$$x^{p/q} = (x^{1/q})^p \quad (1.4)$$

when  $q$  is a positive integer and  $p$  is an integer, again with the restriction  $x \geq 0$  if  $q$  even. We can now define the function  $x^n$  whenever  $n$  is a rational number, using (8.3) if  $n$  is a negative integer,  $n = -p$  and (8.4) if  $n$  is a rational number  $n = p/q$ . In order to define the function  $x^n$  when  $n$  is an irrational number we will have to make use of the exponential function.

### 1.2.3 Exponential functions and natural logarithms

We can define the function  $a^x$  with  $a$  any positive number of all rational values of  $x$  by the process outlined above for defining  $x^n$  for rational values of  $n$ . We define the function  $a^x$  for all values of  $x$ , including irrational values of  $x$ , “by continuity”. This means that if  $x$  is an irrational number approximated by a sequence of rational numbers  $x_1, x_2, \dots$ , we approximate  $a^x$  by the sequence of numbers  $a^{x_1}, a^{x_2}, \dots$ . If we attempt to draw the graph of the function  $a^x$  by plotting all points  $a^x$  with  $x$  rational, then there are gaps in this sketch for all

irrational  $x$  and our approximation amounts to filling these gaps by drawing a smooth curve through the points which have been plotted.

The *logarithm* to base  $a$  is defined as the *inverse function* of  $a^x$ . By this we mean that if

$$y = a^x$$

then

$$x = \log_a y.$$

In other words, the logarithm of  $y$  to base  $a$  is the exponent to which the base  $a$  must be raised to give  $y$ . We define

$$p = \log_a u, \quad q = \log_a v,$$

so that

$$u = a^p, \quad v = a^q.$$

Then

$$uv = a^p a^q = a^{p+q},$$

so that

$$\log_a(uv) = p + q = \log_a u + \log_a v. \quad (1.5)$$

Also,

$$u/v = a^p/a^q = a^{p-q},$$

so that

$$\log_a(u/v) = p - q = \log_a u - \log_a v. \quad (1.6)$$

In addition,

$$u^n = (a^p)^n = a^{np}$$

so that

$$\log_a(u^n) = np = n \log_a u. \quad (1.7)$$

The rules (8.5), (8.6), (8.7) are the equivalents in terms of logarithms of the laws of exponents (8.2).

For exponential functions and logarithms there is a particularly important choice for the base  $a$ . This is the number  $e = 2.718159\dots$  introduced in the eighteenth century by Euler (1707–1783). It is possible to prove that the expression  $(1 + 1/n)^n$  increases as  $n$  increases, coming ever closer to some number, and this number is defined to be  $e$ . Thus

$$e = \lim_{n \rightarrow \infty} [1 + (1/n)]^n;$$

an equivalent definition obtained by replacing  $n$  by  $1/h$  is

$$e = \lim_{h \rightarrow 0} (1 + h)^{1/h}. \quad (1.8)$$

The base  $e$  is understood in discussing exponential functions unless a different base is specified. Thus by “the” exponential function, we mean the function

$$y = e^x. \quad (1.9)$$

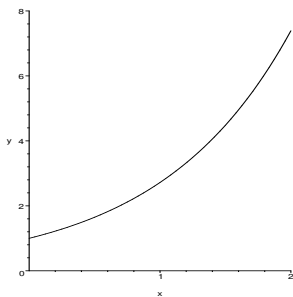


Figure 1.8: The exponential function

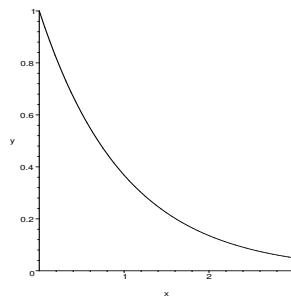


Figure 1.9: The negative exponential function

The graphs of the exponential function  $e^x$  and the negative exponential function  $e^{-x}$  are shown in Figures 1.6 and 1.7.

The *natural logarithm* is the logarithm to base  $e$ , or the inverse of the exponential function, and is denoted by  $\ln$ . In more advanced mathematics the natural logarithm is sometimes denoted by  $\log$  but we will always use  $\ln$  for the natural logarithm.

We observe that

$$x = \ln y; \quad (1.10)$$

is equivalent to (1.9). Note that because  $e^x > 0$  for all  $x$ , (1.9) implies  $y > 0$ , so that the domain of the natural logarithm function  $\ln y$  is  $y > 0$ . From (1.9) and (1.10) with  $x = 1$ , it follows that  $\ln e = 1$ .

### 1.2.4 Trigonometric functions

Many phenomena in the sciences are oscillatory, and one way to describe oscillations is in terms of trigonometric functions. We will not repeat the definitions of the trigonometric functions encountered in precalculus courses, but we remind the student that we will always consider the argument of a trigonometric function to be measured in *radians*, not degrees. The relation between radians and degrees is that  $\pi$  radians is  $180^\circ$ . Thus  $\pi/2$  is  $90^\circ$ ,  $2\pi$  is  $360^\circ$ , etc. The main trigonometric functions which we shall use are  $\sin x$ ,  $\cos x$ , and  $\tan x = \sin x / \cos x$ . The trigonometric functions are *periodic* with period  $2\pi$ , which means that they repeat after  $2\pi$ ,

$$\sin(x + 2\pi) = \sin x, \quad \cos(x + 2\pi) = \cos x.$$

The tangent has period  $\pi$ ,

$$\tan(x + \pi) = \tan x.$$

It is useful to remember that  $\sin 0 = 0$ ,  $\cos 0 = 1$ ,  $\sin \pi/2 = 1$ ,  $\cos \pi/2 = 0$ ,  $\sin \pi = 0$ ,  $\cos \pi = -1$ ,  $\sin 3\pi/2 = -1$ ,  $\cos 3\pi/2 = 0$ ,  $\sin 2\pi = 0$ ,  $\cos 2\pi = 1$ .

Because  $\cos x = 0$  whenever  $x$  is an odd multiple of  $\pi/2$ ,  $\tan x$  is unbounded for these values of  $x$ . The sine and cosine are bounded for all  $x$ ; in fact

$$|\sin x| \leq 1, \quad |\cos x| \leq 1 \quad (-\infty < x < \infty).$$

There are many identities involving trigonometric functions. We list a few of the most important ones.

$$\begin{aligned} \sin^2 x + \cos^2 x &= 1 \\ \sin(x + y) &= \sin x \cos y + \cos x \sin y \\ \sin(x - y) &= \sin x \cos y - \cos x \sin y \\ \cos(x + y) &= \cos x \cos y - \sin x \sin y \\ \cos(x - y) &= \cos x \cos y + \sin x \sin y \\ \sin 2x &= 2 \sin x \cos x \\ \cos 2x &= \cos^2 x - \sin^2 x = 2 \cos^2 x - 1 = 1 - 2 \sin^2 x \end{aligned}$$

### 1.3 Some exercises

1. For what values of  $x$  can the function

$$f(x) = \frac{1}{\sqrt{1-x^2}}$$

be defined?

2. For what values of  $x$  can the function  $f(x) = \sqrt{4 - \sqrt{x}}$  be defined?
3. Sketch the graph of the function  $\ln x$  for  $0 < x < \infty$
4. Is the function

$$f(x) = \begin{cases} -x^2, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

continuous?

5. is it possible to define  $f(1)$  to make the function

$$f(x) = \begin{cases} 1 + x^2, & x < 1 \\ 2 - e^{-x}, & x > 1 \end{cases}$$

continuous?

## Chapter 2

# The Derivative

### 2.1 The meaning of the derivative

The idea of the derivative comes from the attempt to describe the slope of a curve at a point. If the curve is a straight line, its slope may be calculated from the coordinates of any two points on the line. For a curve which is not necessarily a straight line, we may calculate the slope of the line joining any two points on the curve, but this slope will depend on the choice of points. If we fix one of the points on the curve and try to describe the slope of the curve at this point as the slope of the line joining this point to another point on the curve, the slope will depend on the choice of the second point. By the slope of the curve at the chosen point we mean the limit of this slope as the second point approaches the chosen point.

To define the slope of a curve given by a function  $y = f(x)$  at a point  $(x_0, f(x_0))$  on the curve, we choose a second point on the curve  $(x_0+h, f(x_0+h))$ , thinking of  $x_0+h$  as close to  $x_0$ , or  $h$  small. Then the slope of the line joining the two points on the curve, called the *secant line* to the curve is

$$\frac{f(x_0+h) - f(x_0)}{(x_0+h) - x_0} = \frac{f(x_0+h) - f(x_0)}{h}. \quad (2.1)$$

It is important to remember that  $h$  must be different from zero in (2.1). If  $h = 0$ , both numerator and denominator in (2.1) are zero and the expression (1) is *indeterminate*. Of course, taking  $h = 0$  would mean that the two points on the curve used to try to calculate the slope were the same. The slope (2.1) is defined for all  $h \neq 0$ , and in order to describe the slope of the curve  $y = f(x)$  at  $x_0$ , we try to find a number such that the slope  $[f(x_0+h) - f(x_0)]/h$  can be made as close as we wish to this number by taking  $h$  small enough. In practice, this is often accomplished by dividing a factor  $h$  out of both numerator and denominator (legitimate since  $h \neq 0$ ) and examining the result. If there is such a number, we call it the *limit* of the slope (2.1) as  $h \rightarrow 0$ , and this is what we will define to be the slope or *derivative* of the curve  $y = f(x)$  at  $x_0$ , denoted by

$f'(x_0)$ ,

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}. \quad (2.2)$$

**Example 1.** To find the derivative of the function

$$f(x) = x^2$$

at the point  $(x_0, x_0^2)$ , we calculate

$$\begin{aligned} f(x_0 + h) &= (x_0 + h)^2 = x_0^2 + 2x_0h + h^2, \\ f(x_0 + h) - f(x_0) &= 2x_0h + h^2 = h(2x_0 + h) \\ \frac{f(x_0 + h) - f(x_0)}{h} &= 2x_0 + h. \end{aligned}$$

It is clear that as  $h \rightarrow 0$  this quantity approaches  $2x_0$ , and thus we have shown that the derivative of  $x^2$  at  $x_0$  is  $2x_0$ .

The *tangent line* to the curve  $y = f(x)$  at  $(x_0, f(x_0))$  is defined to be the line through this point whose slope is  $f'(x_0)$ , the slope of the curve at that point. Equivalently, the tangent line to a curve at a point is the straight line through the point whose direction is the direction of the curve at the point. Thus the equation of the tangent line is

$$y - f(x_0) = f'(x_0)(x - x_0).$$

Near the point  $(x_0, f(x_0))$  the tangent line approximates the curve. In particular, if the slope  $f'(x_0)$  is positive the curve is rising (going upward to the right) at  $x_0$  and if the slope  $f'(x_0)$  is negative the curve is falling.

The quantity  $f'(x_0)$  is also known as the *derivative* of the function  $f(x)$  at  $x_0$ . Recall that the slope of the graph also represents the rate of change of the function. In many applications, the importance of the derivative lies in its interpretation as the rate of change of the function  $f(x)$  at  $x_0$ . For example, in describing the motion of an object in a line, *velocity* is defined as the rate of change of position; thus if position is known as a function of time the derivative of this function is the velocity. In studying the size of a population as a function of time, the derivative of this function is the rate of change of population size, which may be separated into birth, death, and migration rates. If  $I(t)$  represents the number of infectives in a population in which there is a disease outbreak, then  $I'(t)$  represents the rate of change of the number of infectives. This rate of change can be separated into two parts and can be regarded as the rate of new infections minus the rate at which infectives leave the infective class, either by recovery or by death from disease.

If  $f(x)$  is a function, we may find the derivative  $f'(x_0)$  at every point  $x_0$ . This enables us to define the derivative as a function  $f'(x)$ , namely as the function whose value at  $x_0$  is  $f'(x_0)$ . There is another notation for derivatives in common use; if  $y$  is a function of  $x$  we call the derivative  $dy/dx$ . This quantity  $dy/dx$  is

not a quotient of  $dy$  and  $dx$ ; the expressions  $dx$  and  $dy$  do not have meanings by themselves. However, there are many situations in which the notation is designed so that thinking of  $dy/dx$  as a quotient leads to correct formulae.

## 2.2 Limits of Functions

In describing derivatives, we have assumed that it is always possible to find the necessary limit. This is not necessarily true, and we must discuss some properties of functions in order to give a proper perspective. Here, we are describing how to add rigor to the idea of the derivative and make it mathematically precise.

The first basic concept is that of the limit of a function. Indeed, our definition of the derivative has invoked this idea, when we required  $h$  to approach zero (but did not allow  $h = 0$ ). For a function  $f(x)$ , by the statement

$$\lim_{x \rightarrow a} f(x) = L$$

read as “the limit of  $f(x)$  as  $x$  approaches  $a$  is  $L$ ”, or “ $f(x)$  approaches  $L$  as  $x$  approaches  $a$ ”, we mean that the value of  $f(x)$  is close to  $L$  whenever the value of  $x$  is close to  $a$ . More precisely, we require that the value of  $f(x)$  can be made as close to  $L$  as we care to prescribe by choosing  $x$  close enough to  $a$ . It is important to remember that here  $x$  is not allowed to equal  $a$ ; the value  $f(a)$ , or even whether  $f(a)$  is defined, has nothing to do with the limit of  $f(x)$  as  $x$  approaches  $a$ .

Functions do not necessarily have limits, but if they do the following rules for combinations of functions are true

$$\begin{aligned} \lim_{x \rightarrow a} [f(x) + g(x)] &= \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) \\ \lim_{x \rightarrow a} [cf(x)] &= c \lim_{x \rightarrow a} f(x) \\ \lim_{x \rightarrow a} [f(x)g(x)] &= \left[ \lim_{x \rightarrow a} f(x) \right] \left[ \lim_{x \rightarrow a} g(x) \right] \\ \lim_{x \rightarrow a} [f(x)/g(x)] &= \left[ \lim_{x \rightarrow a} f(x) \right] / \left[ \lim_{x \rightarrow a} g(x) \right] \quad \text{if } \lim_{x \rightarrow a} g(x) \neq 0 \end{aligned} \tag{2.3}$$

Thus an attempt to calculate the limit of a function  $f(x)$  as  $x$  approaches  $a$  always begins by substituting  $x = a$  and trying to calculate  $f(a)$ , as the elementary functions all have the property that  $\lim_{x \rightarrow a} f(x) = f(a)$  and the rules (2.3) say that combinations of elementary functions have the same property unless there is a zero in the denominator.

If substitution of  $x = a$  does give zero in the denominator but not in the numerator, then it is easy to see that the function  $f(x)$  is very large when  $x$  is near  $a$ , and we say that it has an infinite limit,

$$\lim_{x \rightarrow \infty} f(x) = \infty.$$

In this case, the graph of the function  $f(x)$  approaches the line  $x = a$ , and we say that the line  $x = a$  is a *vertical asymptote* of the graph  $y = f(x)$ .

If substitution of  $x = a$  into  $f(x)$  gives zero in both numerator and denominator, then we say that  $f(x)$  is an *indeterminate form* as  $x \rightarrow a$ , and we must perform some algebraic manipulation to find the limit. Often, we can find a factor  $(x - a)$  in both numerator and denominator, and we may cancel this factor *provided*  $x \neq a$  to remove the indeterminacy. This cancellation is legitimate because the limit depends only on values of  $x$  different from  $a$ .

Here is a list of some important limits of particular elementary functions.

$$\begin{aligned} \lim_{x \rightarrow 0} x^n &= \infty \quad (n < 0) \\ \lim_{h \rightarrow 0} (1 + h)^{1/h} &= e \\ \lim_{h \rightarrow 0} (1 + bh)^{1/h} &= e^b \\ \lim_{x \rightarrow 0} e^x - 1/x &= 1 \\ \lim_{x \rightarrow 0} \ell n x &= -\infty \\ \lim_{x \rightarrow 0} x^n \ell n x &= 0 \quad (n > 0) \\ \lim_{x \rightarrow 0} \sin x/x &= 1 \\ \lim_{x \rightarrow 0} (1 - \cos x)/x &= 0 . \end{aligned}$$

The limit of a function as  $x$  becomes very large (either positive or negative) must be defined slightly differently. By the statement  $\lim_{x \rightarrow \infty} f(x) = L$  we mean that the value of  $f(x)$  is close to  $L$  for all sufficiently large  $x$ , or more precisely that we can make the value of  $f(x)$  as close to  $L$  as we care to prescribe by choosing  $x$  large enough. The statement  $\lim_{x \rightarrow -\infty} f(x) = L$  is defined similarly except that now  $x$  is taken “large and negative”, meaning that  $x$  is negative with large absolute value. The limit rules (2.3) are also valid for limits as  $x \rightarrow \infty$  or  $x \rightarrow -\infty$ .

To calculate a limit as  $x \rightarrow \infty$  we begin by trying to “substitute  $x = \infty$ ”, that is, by examining the limit of each term as  $x \rightarrow \infty$ . In practice, this often leads to an indeterminate form of the type “ $\infty/\infty$ ”. In order to evaluate such an indeterminate form we usually divide numerator and denominator by the term which grows most rapidly (“the largest term”) as  $x \rightarrow \infty$  and then try to “substitute  $x = \infty$ ”. We may then obtain a finite quantity, which is the desired limit, or an indeterminate form with zero in the denominator but not in the numerator, which implies that the limit is infinite.

If  $\lim_{x \rightarrow \infty} f(x) = L$ , the graph of the curve  $y = f(x)$  approaches the line  $y = L$  for large  $x$ , and the line  $y = L$  is said to be a *horizontal asymptote* of the curve. Information about horizontal asymptotes is useful in sketching the graph of a function. Observe that functions do not necessarily have horizontal asymptotes, as not all functions have limits as  $x \rightarrow \infty$ . Also, it is possible for a function to have two different horizontal asymptotes, one as  $x \rightarrow \infty$  and another as  $x \rightarrow -\infty$ .

Here is a list of some important limits of particular elementary functions as  $x \rightarrow \infty$ .

$$\begin{aligned} \lim_{x \rightarrow \infty} x^n &= \infty \quad (n > 0) \\ \lim_{x \rightarrow \infty} x^n &= 0 \quad (n < 0) \\ \lim_{h \rightarrow \infty} \left(1 + \frac{b}{x}\right)^x &= e^b \\ \lim_{h \rightarrow \infty} e^{bx} &= \infty \quad (b > 0) \\ \lim_{x \rightarrow \infty} e^{-bx} &= 0 \quad (b > 0) \\ \lim_{x \rightarrow \infty} x^n e^{-bx} &= 0 \quad (b > 0, n \text{ arbitrary}) \\ \lim_{x \rightarrow \infty} \ln x &= \infty \\ \lim_{x \rightarrow \infty} x^{-n} \ln x &= 0 \quad (n > 0) \end{aligned}$$

In Chapter 1 we gave an intuitive description of the property of continuity of a function. Now that we have introduced the idea of limits, we can give a more precise description. A function  $f(x)$  is said to be continuous at a point  $x_0$  if

$$\lim_{x \rightarrow x_0} f(x) = f(x_0). \quad (2.4)$$

This apparently simple statement is actually three separate statements:

- (i) The function  $f(x)$  has a limit as  $x$  approaches  $x_0$ .
- (ii) The function  $f(x)$  is defined at  $x_0$ .
- (iii) The value of the function at  $x_0$  is equal to the limit as  $x$  approaches  $x_0$ .

The statement (2.4) says that for all  $x$  close to  $x_0$  and also for  $x = x_0$ , the value of the function  $f(x)$  is close to  $x_0$ . Thus if  $f(x)$  is continuous at  $x_0$ , the graph of  $y = f(x)$  can not have a “jump” at  $x_0$ . A function  $f(x)$  can fail to be continuous at a point in several ways:

- (i)  $f(x)$  might not be defined at  $x = x_0$  (e.g.  $f(x) = x^2 - 1/(x - 1)$  at  $x = 1$ ).
- (ii)  $f(x)$  might not be bounded near  $x = x_0$  (e.g.  $f(x) = 1/x$  near  $x = 0$ ).
- (iii)  $f(x)$  might be bounded but not have a limit as  $x \rightarrow x_0$  (e.g.  $f(x) = \sin(1/x)$  at  $x = 0$  or  $f(x) = 0$  for  $x < 0$ ,  $f(x) = 1$  for  $x \geq 0$  at  $x = 0$ ).
- (iv)  $f(x)$  might have a limit as  $x \rightarrow x_0$  but  $f(x_0)$  might be “wrong” (e.g.  $f(x) = (x^2 - 1)/(x - 1)$  if  $x \neq 1$ ,  $f(1) = 1$ ).

If a function  $f(x)$  is continuous at every point of an interval, then  $f(x)$  is said to be continuous on the interval. For the most part, the functions with which we deal are continuous at every value for which they are defined. Continuous functions have the following important properties whose proofs (which we omit) depend on some fundamental properties of the real number system:

1. If a function is continuous on a closed bounded interval  $a \leq x \leq b$  it has a *maximum* in the interval, that is, there is a point  $x_0$  in the interval such that  $f(x) \leq f(x_0) = M$  for all  $x, a \leq x \leq b$ .
2. If a function is continuous on a closed bounded interval  $a \leq x \leq b$  it has a *minimum* in the interval, that is, there is a point  $x_0$  in the interval such that  $f(x) \geq f(x_0) = m$  for all  $x, a \leq x \leq b$ .
3. If a function is continuous on a closed bounded interval  $a \leq x \leq b$  then for every number  $c$  between the maximum  $M$  and the minimum  $m$  of the function on the interval there is a point  $\xi$  with  $a \leq \xi \leq b$  such that  $f(\xi) = c$ .

An *increasing* function is a function  $f(x)$  such that  $f(x_1) \leq f(x_2)$  for all  $x_1, x_2$  with  $x_1 < x_2$ . It can be proved that if  $f'(x) \geq 0$  for all  $x$  then the function  $f(x)$  is increasing. A *strictly increasing* function is a function  $f(x)$  such that  $f(x_1) < f(x_2)$  for all  $x_1, x_2$  with  $x_1 < x_2$ . The reader should be warned that some authors use the terms non-decreasing (instead of increasing) and increasing (instead of strictly increasing), and therefore one should be careful to check the terminology on encountering the term "increasing". Decreasing and strictly decreasing functions are defined analogously. A useful theorem states that an increasing function defined on an interval  $a \leq x < \infty$  which is bounded above has a limit as  $x \rightarrow \infty$ . Similarly, a decreasing function defined on an interval  $a \leq x < \infty$  which is bounded below has a limit as  $x \rightarrow \infty$ .

Although in general functions are not necessarily continuous, the functions that we encounter in applications are generally continuous, and also differentiable.

## 2.3 Calculation of Derivatives

It is possible to show that if the derivative  $f'(x_0)$  of a function  $f(x)$  at a point  $x_0$ , as defined by (2.2), exists, then the function  $f(x)$  must be continuous at  $x_0$ . The converse statement is false. A function may be continuous at a point  $x_0$  but fail to have a derivative at  $x_0$ ; for example the function  $f(x) = |x|$  is continuous at  $x = 0$  but does not have a derivative at  $x = 0$  (the quantity

$$\frac{f(x_0 + h) - f(x_0)}{h}$$

is  $+1$  if  $h$  is positive and  $-1$  if  $h$  is negative and thus does not have a limit as  $h \rightarrow 0$ ). It is even possible for a function to be continuous on an interval but to fail to have a derivative at any point of the interval. However, we will avoid such "pathological" functions, and will encounter only functions which have a derivative at every point where they are defined, except possibly for functions like  $|x|$  which obviously fail to have a derivative at certain points.

In order to calculate the derivative of a particular function, we must go back to the definition (2.2). The calculation will require the evaluation of a limit as

$h$  approaches zero. We list some differentiation formulae. For convenience, we use the alternate notation  $dy/dx$  for the derivative, replacing  $y$  by its expression in terms of  $x$ .

$$\begin{aligned}\frac{d}{dx}x^n &= nx^{n-1} \quad (n \text{ rational, either positive or negative, } n \neq 0) \\ \frac{d}{dx}e^x &= e^x \quad (\text{requires } \lim_{h \rightarrow 0} (e^h - 1)/h = 1) \\ \frac{d}{dx}\ln x &= 1/x \quad (\text{requires } \lim_{h \rightarrow 0} (1+h)^{1/h} = e) \\ \frac{d}{dx}\sin x &= \cos x \\ \frac{d}{dx}\cos x &= -\sin x \quad \left( \text{requires } \lim_{h \rightarrow 0} \sin h/h = 1, \lim_{h \rightarrow 0} (1 - \cos h)/h = 0 \right)\end{aligned}\tag{2.5}$$

We will not derive all of these formulae, but we give two examples

**Example 1.** To find the derivative of the function

$$f(x) = e^x$$

at the point  $(x_0, e^{x_0})$ , we calculate

$$\begin{aligned}f(x_0 + h) &= e^{x_0+h} = e^{x_0}e^h, \\ f(x_0 + h) - f(x_0) &= e^{x_0}e^h - e^{x_0} = e^{x_0}[e^h - 1], \\ \frac{f(x_0 + h) - f(x_0)}{h} &= e^{x_0}\frac{e^h - 1}{h}.\end{aligned}$$

Using the relation

$$\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$$

we see that the derivative of  $e^x$  at  $x_0$  is  $e^{x_0}$ .

**Example 2.** To find the derivative of the function

$$f(x) = \sin x$$

at the point  $(x_0, \sin x_0)$ , we calculate

$$\begin{aligned}f(x_0 + h) &= \sin(x_0 + h) = \sin x_0 \cos h + \cos x_0 \sin h, \\ f(x_0 + h) - f(x_0) &= \sin x_0 \cos h + \cos x_0 \sin h - \sin x_0 \\ &= \sin x_0[\cos h - 1] + \cos x_0 \sin h, \\ \frac{f(x_0 + h) - f(x_0)}{h} &= \sin x_0 \frac{\cos h - 1}{h} + \cos x_0 \frac{\sin h}{h}.\end{aligned}$$

Using the relations

$$\lim_{h \rightarrow 0} \frac{\cos h - 1}{h} = 0, \lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$$

we see that the derivative of  $\sin x$  at  $x_0$  is  $\cos x_0$ .

In order to find the derivative of combinations of elementary functions, we may use the following rules. If  $u$  and  $v$  are functions of  $x$  and  $c$  is a constant,

$$\begin{aligned}\frac{d}{dx}(u+v) &= \frac{du}{dx} + \frac{dv}{dx} \\ \frac{d}{dx}(cu) &= c\frac{du}{dx} \\ \frac{d}{dx}(uv) &= u\frac{dv}{dx} + v\frac{du}{dx} \\ \frac{d}{dx}(u/v) &= \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2} \quad (\text{if } v \neq 0)\end{aligned}\tag{2.6}$$

In order to calculate more complicated combinations, there are some more general rules.

**I. Chain rule.** If  $z$  is a function of  $y$  and  $y$  is a function of  $x$ , then  $z$  is a function of  $x$  whose derivative is given by

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}.\tag{2.7}$$

The relation (2.7) is an instance of how the  $dy/dx$  notation is designed to make formulae appear natural. However, cancellation of  $dy$  from numerator and denominator of the right side of (2.7) is not a proof of the validity of (2.7), because  $dx$ ,  $dy$ , and  $dz$  do not have independent meanings and the expressions in (2.7) are not fractions. Another way to express (2.7) which appears more complicated than (2.7) but is sometimes more straightforward is that if  $y$  is a function of  $x$ ,  $y = f(x)$  and  $z$  is a function of  $y$ ,  $z = g(y)$ , then  $z$  is a function of  $x$  called the composite function

$$z = F(x) = g[f(x)]$$

whose derivative is given by

$$F'(x) = g'[f(x)]f'(x).$$

**Example 3.** If  $y = e^{ax}$ , then  $\frac{dy}{dx} = ae^{ax}$ . If  $y = e^{-bx}$ , then  $\frac{dy}{dx} = -be^{-bx}$ .

**Example 4.** If  $y = f(x)$ , then

$$\frac{d \ln y}{dx} = \frac{1}{y} \frac{dy}{dx} = \frac{f'(x)}{f(x)}.$$

**II. Inverse functions.** If  $y$  is a function of  $x$  which can be solved for  $x$  as a function of  $y$ , then

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

(provided  $dy/dx \neq 0$ ). Again, we can state this in functional notation. If  $y$  is a function of  $x$ ,  $y = f(x)$ , which can be inverted, that is, if there is a function  $x = g(y)$  such that

$$x = g[f(x)], \quad y = f[g(y)]$$

then

$$g'(y) = 1/f'(x)$$

(provided  $f'(x) \neq 0$ ).

**Example 5.** If  $y = x^2$ , so that  $\frac{dy}{dx} = 2x$ , then  $\frac{dx}{dy} = 1/2x$ , or

$$\frac{d}{dy}y^{1/2} = \frac{1}{2x} = \frac{1}{2}y^{-1/2}.$$

**III. Implicit differentiation.** If  $y$  is given *implicitly* as a function by a relation between  $x$  and  $y$  of the form  $F(x, y) = 0$ , then it is possible to calculate  $dy/dx$  by differentiating each term in the equation  $F(x, y) = 0$  with respect to  $x$ , remembering to use the chain rule to differentiate terms involving  $y$  with respect to  $x$ .

**Example 1** (Adiabatic expansion of a gas). Under some circumstances a gas expands so rapidly that it does not exchange heat with its surroundings. Such expansion is called *adiabatic*. There are laws of physics which state that in adiabatic expansion the volume  $V$  of gas and the temperature  $T$  are related by an equation

$$TV^{\gamma-1} = C, \tag{2.8}$$

where  $\gamma$  and  $C$  are constants, with  $\gamma \approx 1.4$ . Here  $V$  and  $T$  are functions of the time  $t$ . If we differentiate (2.8) implicitly with respect to  $t$  we obtain

$$\frac{dT}{dt}V^{\gamma-1} + (\gamma - 1)TV^{\gamma-2}\frac{dV}{dt} = 0.$$

This implies

$$\frac{dT}{dt} = -(\gamma - 1)\frac{T}{V}\frac{dV}{dt}.$$

In particular, we note that the time derivatives of  $V$  and  $T$  have opposite sign. Thus if the gas is expanding, the temperature drops.

## 2.4 Applications of the Derivative

The applications that we describe in this section are to mathematical questions. We omit discussion about the mathematical formulation and solution of problems which come from other sciences—questions that may appear to the non-mathematician to have a more legitimate claim to the name application. In this section we shall discuss the use of properties of functions including but not restricted to properties of derivatives in sketching the graph of a given function, in finding where a given function attains its maximum and minimum value, and in describing exponential growth or decay. In the next section we will describe some applications to problems of optimization which arise in other sciences.

### 2.4.1 Curve sketching

If we are given a function  $f(x)$  whose graph we wish to sketch, we may begin by checking for vertical and horizontal asymptotes, as described in the preceding section. Next, we investigate where the function is increasing and where the function is decreasing, remembering that  $f(x)$  is increasing where  $f'(x) > 0$  and decreasing where  $f'(x) < 0$ . We calculate the derivative  $f'(x)$  and find the values of  $x$  for which  $f'(x) = 0$ . These are the places where the graph may switch between increasing and decreasing, and by seeing whether  $f'(x)$  is positive or negative between then we can identify the intervals on which  $f(x)$  is increasing and the intervals on which  $f(x)$  is decreasing. As the graph may “jump between  $-\infty$  and  $+\infty$ ” at a vertical asymptote, as with  $f(x) = 1/x$  at  $x = 0$ , we must remember to watch out for points where  $f'(x)$  does not exist.

**Example 1.** The derivative of the function

$$y = \frac{K}{1 + ce^{-rt}}$$

where  $K, c$ , and  $r$  are positive constants is positive for all  $t$ . It increases from  $K/(1 + c)$  at  $t = 0$  to a limit  $K$  as  $t \rightarrow \infty$ . This function, known as the *logistic* function describes the growth of many quantities with the property that the rate of growth decreases as the quantity increases. Its graph is shown in Figure 2.1.

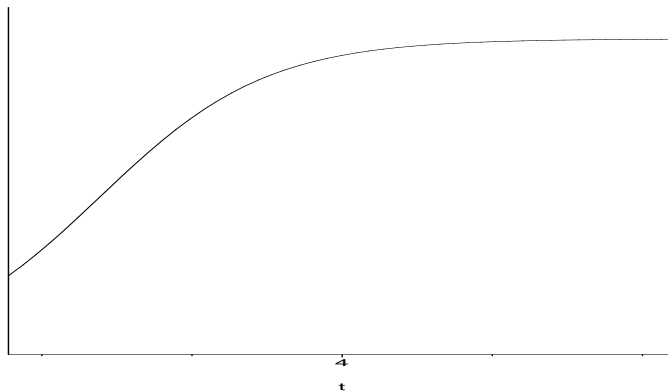


Figure 2.1: A logistic function

A further refinement is obtained from examination of the second derivative  $f''(x)$ . If  $f''(x) > 0$  the curve is said to be *concave upwards* and if  $f''(x) < 0$  the curve is said to be *concave downwards*. The second derivative may also be used to check whether a point at which  $f'(x) = 0$  is a *relative maximum* (if  $f''(x) < 0$ ) or a *relative minimum* (if  $f''(x) > 0$ ). A point at which  $f''(x) = 0$  may indicate a change in the concavity of the function. A point at which the concavity does change is called a *point of inflection*.

With information about horizontal and vertical asymptotes, where the function is increasing, where the function is decreasing, and the concavity, it is possible to sketch the graph. In order to judge the scale, it is wise to plot a few points as well, especially any relative maxima and minima.

### 2.4.2 Maximum - minimum problems

In many applications, the problem to be solved may be formulated as a question of determining where a function  $f(x)$  defined on a given interval attains its maximum or minimum. By maximum, we mean an *absolute* or *global* maximum – a point  $x_0$  such that  $f(x) \leq f(x_0)$  for every  $x$  in the interval. The solution of such a problem depends on the following properties of continuous functions.

1. A function  $f(x)$  which is continuous on a closed bounded interval has a maximum and a minimum.
2. If the maximum or minimum of  $f(x)$  occurs at an interior point  $x_0$  of the interval  $x_0$  where the derivative  $f'(x_0)$  exists, then  $f'(x_0) = 0$ .
3. If  $x_0$  is an interior point of the interval and  $f'(x_0) = 0$ ,  $f''(x_0) > 0$ , then  $x_0$  is a *relative* or *local* minimum (a point such that  $f(x) \geq f(x_0)$  for all  $x$  near  $x_0$ ) and can not be an absolute maximum. Similarly, if  $x_0$  is an interior point and  $f'(x_0) = 0$ ,  $f''(x_0) < 0$ , then  $x_0$  is a relative maximum and can not be an absolute minimum.

In order to find the absolute maximum of a function  $f(x)$  on an interval, we must examine (i) the ends of the interval if they are contained in the interval under consideration, (ii) any points of the interval at which the derivative  $f'(x)$  does not exist, (iii) any points  $x_0$  in the interval such that  $f'(x_0) = 0$ . Thus we calculate the derivative  $f'(x)$  and find all solutions  $x_0$  of the equation  $f'(x_0) = 0$ . Then we calculate the value of the function  $f(x)$  at each such  $x_0$ , at each point where  $f(x)$  does not have a derivative, and at the end points of the interval. The largest of these values is at the maximum and the smallest is at the minimum. We may use the second derivative test at points  $x_0$  where  $f'(x_0) = 0$  to eliminate any relative minima from consideration if we are looking for the maximum. If it is difficult to calculate the second derivative at  $x_0$ , we may calculate the first derivative to the left of  $x_0$  and to the right of  $x_0$  in order to identify relative maxima and relative minima. Finally, we should remember that if the interval under consideration is not closed and bounded then the function  $f(x)$  does not necessarily have a maximum or a minimum. For example, the function  $f(x) = 1/x$  has a minimum but no maximum on the interval  $0 < x \leq 1$ , a maximum but no minimum on the interval  $1 \leq x < \infty$ , and neither a maximum nor a minimum on the interval  $0 < x < \infty$ .

### 2.4.3 Optimization

Many applications of calculus involve optimization of some quantity, and the theory described in Section 2.4.2 provides the tools for such applications. We

will give a few examples to give an idea of the process.

**Example 1.** Suppose that the yield of an agricultural crop depends on the level of nitrogen in the soil, with the yield being given by

$$Y(N) = \frac{N}{N^2 + 1}$$

if  $N \geq 0$ . What level of nitrogen maximizes the yield?

**Solution:** Using the rule for differentiating a quotient of two functions, we have

$$Y'(N) = \frac{(N^2 + 1) - N \cdot 2N}{(N^2 + 1)^2} = \frac{1 - N^2}{(N^2 + 1)^2}$$

which is zero for  $N = \pm 1$ . Only  $N = 1$  is in the interval of definition of the function. Since  $Y'(N) > 0$  if  $0 < N < 1$  and  $Y'(N) < 0$  if  $N > 1$ , this is a relative maximum. Since  $Y(0) = 0$  and  $Y(N) \rightarrow 0$  as  $N \rightarrow \infty$ , the relative maximum at  $N = 1$  is the absolute maximum.

**Example 2.** In Section 1.2, Example 1 we described the rate of a chemical reaction by the function

$$R(x) = k(a - x)(b - x)$$

At what stage of the reaction is this speed the greatest?

**Solution:** The function  $R(x)$  is defined on the interval  $0 \leq x \leq c$ , where  $c = \min(a, b)$ . Its derivative is  $k[2x - (a + b)]$ , which is negative near  $x = 0$  and vanishes for  $x = (a + b)/2 \geq c$ . Since the point at which  $R'(x) = 0$  is outside the interval of definition,  $R(x)$  is a decreasing function and its maximum is taken at  $x = 0$ , at the beginning of the reaction.

#### 2.4.4 Exponential growth and decay

In many situations, the rate of change of a quantity is proportional to the amount. If  $y$  is increasing at a rate proportional to  $y$ , so that  $dy/dt = ay$  for some positive constant  $a$ , then  $y = ce^{at}$  for some constant  $c$ . The *doubling time* is defined to be the time required for  $y$  to double its original value. If  $y = ce^{at}$ , then  $y = c$  when  $t = 0$ , and  $y = 2c$  when  $e^{at} = 2$  or  $t = \ln 2/a$ . Thus the doubling time is  $\ln 2/a$ .

If  $y$  is decreasing at a rate proportional to  $y$ , so that  $dy/dt = -by$  for some positive constant  $b$ , then  $y = ce^{-bt}$  for some constant  $c$ . The *half-life* (a term coming from decay of radioactive elements, which behave in this way) is defined to be the time required for  $y$  to decrease to half its original value. If  $y = ce^{-bt}$ , then  $y = c$  when  $t = 0$  and  $y = c/2$  when  $e^{-bt} = 1/2$ , or  $t = \ln 2/b$ . Thus the half-life is  $\ln 2/b$ .

Exponential growth and decay will be studied further in Section 7.1, where we will show that a function whose derivative is a constant multiple of the function must be an exponential function.

## 2.5 Local Linearity

A fundamental property of differentiable functions is that *locally* (close enough to a starting point) they may be approximated by linear functions.

The definition of the derivative of a function  $f$  of one variable at a point  $x_0$  is

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

Another way to express this definition is to say that

$$\frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0) + \epsilon \quad (2.9)$$

with  $\epsilon$  “small”; more precisely,  $\epsilon \rightarrow 0$  as  $h \rightarrow 0$ . The relation (2.9) may be rewritten

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \epsilon h \quad (2.10)$$

If we let  $x = x_0 + h$ , (2.10) becomes

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + R \quad (2.11)$$

where  $R \rightarrow 0$  “faster than  $h$ ” in the sense that

$$\lim_{h \rightarrow 0} \frac{R}{h} = 0 \quad (2.12)$$

If we think of (2.11) as saying that  $f(x_0) + (x - x_0)f'(x_0)$  approximates  $f(x)$  with an error  $R$ , the relation (2.12) says that the approximation (2.11) is good if  $h$  is small, or  $x$  is close to  $x_0$ . The approximation

$$f(x_0) + (x - x_0)f'(x_0)$$

is a linear function of  $x$ , and is called the *linear approximation* to  $f(x)$  at  $x_0$ . In geometric terms, the curve (straight line)

$$y = f(x_0) + (x - x_0)f'(x_0)$$

is the *tangent line* to the curve  $y = f(x)$  at the point  $(x_0, f(x_0))$ . Because of (2.9), the linear approximation is a good approximation if  $h$  is small, or  $x$  is close to the starting point  $x_0$ . In other words, the function behaves like a linear function *locally* (near  $x_0$ ). For our purposes, the estimate (2.9) for the error in the approximation is sufficient, but more precise estimates involving higher order derivatives are possible. For example, to establish the distinction between a local maximum and a local minimum of a function  $f(x)$  at a point  $x_0$  at which  $f'(x_0) = 0$  in terms of the second derivative requires an error estimate in terms of the second derivative.

**Example 1.** Estimate  $\sqrt{4.1}$  from the linear approximation.

**Solution.** We use the function  $F(x) = x^{1/2}$  with  $x_0 = 4$ ,  $h = 0.1$ . Then we have  $f(x) = f(x_0 + h) = f(4.1) = \sqrt{4.1}$ . Since  $f'(x) = x^{-1/2}/2$ , the linear

approximation is  $f(x_0) + (x - x_0)f'(x_0) = \sqrt{4} + (0.1)(4)^{-1/2}/2 = 2.025$ . [A better way to estimate  $\sqrt{4.1}$  is to use a calculator, obtaining 2.0248457.]

In addition to using the linear approximation to estimate the value of a function at a specific point, we may also use it to approximate the function. Some examples, all for  $x$  near 0, are

$$\begin{aligned} e^x &\approx 1 + x \\ \ln(1 + x) &\approx x \\ \sin x &\approx x \\ (1 + x)^k &\approx 1 + kx \quad (k \text{ any real number}). \end{aligned}$$

### 2.5.1 Application: The chain rule

Suppose  $y$  is a function of  $u$ , say  $y = h(u)$  with  $y_0 = h(u_0)$ , and  $u$  is a function of  $x$ , say  $u = g(x)$  with  $u_0 = g(x_0)$ . Then  $y$  is a function of  $x$ , say  $y = f(x)$  from  $y = h\{g(x)\}$ , with  $y_0 = f(x_0)$ . The chain rule is that

$$f'(x_0) = h'\{g(x_0)\}g'(x_0) \quad (2.13)$$

The colloquial form of (2.13) is

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

The idea behind the proof is that if  $y$  is a linear function of  $u$ ,

$$y - y_0 = a(u - u_0)$$

with  $a = h'(u_0) = h'\{g(x_0)\}$  and  $u$  is a linear function of  $x$ ,

$$u - u_0 = b(x - x_0)$$

with  $b = g'(x_0)$ , then direct substitution gives

$$y - y_0 = a(u - u_0) = ab(x - x_0)$$

and

$$f'(x_0) = ab = h'\{g(x_0)\}g'(x_0)$$

as desired. The full proof of the chain rule is just a matter of going through this calculation and including some bookkeeping on the error terms.

## 2.6 Some exercises

1. Find the slope of the line through the points (2, 4) and (3, 9).
2. Find the slope of the line through the two points on the curve  $y = x^2$  with  $x = 2$  and  $x = 3$ .

3. Find the slope of the line through the two points on the curve  $y = x^2$  with  $x = 2$  and  $x = 2 + h$  ( $h \neq 0$ ).

In each of Exercises 4–9, find the indicated limit.

4.  $\lim_{x \rightarrow 1} \frac{x^2+1}{x+1}$   
5.  $\lim_{x \rightarrow 1} \frac{x^2-1}{x-1}$   
6.  $\lim_{x \rightarrow 1} e^x$   
7.  $\lim_{x \rightarrow \infty} \frac{2x^2+x+1}{x^2-1}$   
8.  $\lim_{x \rightarrow \infty} \frac{x^3+1}{x+1}$   
9.  $\lim_{x \rightarrow \infty} \frac{2x^2+e^{-x}}{x^2-1}$

In each of Exercises 10–12, find the derivative  $\frac{dy}{dx}$ .

10.  $y = x^2$   
11.  $y = x^3 + x^{2/3}$   
12.  $y = (x^2 + 1)^2 e^x$

In each of Exercises 13–14, find the derivative  $\frac{dy}{dx}$  at the indicated point.

13.  $y = x^3 + x$ ,  $x = 2$   
14.  $y = xe^{-x}$ ,  $x = 1$

In each of Exercises 15–17, find the maximum of the given function  $f(x)$  on the given interval.

15.  $f(x) = x - x^2$ ,  $0 \leq x \leq 1$   
16.  $f(x) = 1 - x^2$ ,  $0 \leq x \leq 1$   
17.  $f(x) = xe^{-x}$ ,  $0 \leq x < \infty$

18. Sketch a graph of the function  $y = xe^{-x}$  on  $0 \leq x < \infty$ , indicating relative maxima, relative minima, inflection points, and horizontal asymptotes.  
19. Sketch a graph of the function

$$y = \frac{x}{1+x^2}$$

on  $-\infty < x < \infty$ , indicating relative maxima, relative minima, symmetry, and horizontal asymptotes.

20. If  $y = y_0 e^{-10t}$ , find the value of  $t$  such that  $y = \frac{1}{2}y_0$ .

In each of Exercises 21–24, use the linear approximation with the given  $x_0$  to estimate the value of the given function  $f(x)$  at the given value of  $x$ .

21.  $f(x) = x^4$ ,  $x_0 = 1$ ,  $x = 0.99$

22.  $f(x) = (1 + x)^{1/3}$ ,  $x_0 = 0$ ,  $x = 0.1$

23.  $f(x) = \frac{x}{x-1}$ ,  $x_0 = 2$ ,  $x = 2.2$

24.  $f(x) = (x^2 + 16)^{1/2}$ ,  $x_0 = 3$ ,  $x = 3.2$

25. According to Taylor's theorem, a quadratic approximation to  $f(x)$  is

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2}f''(x_0).$$

(a) If  $x_0$  is a *critical point* of  $f(x)$ , that is, if  $f'(x_0) = 0$ , this approximation becomes

$$f(x) \approx f(x_0) + \frac{(x - x_0)^2}{2}f''(x_0).$$

Show that  $f(x) \geq f(x_0)$  if  $f''(x_0) > 0$ , that is, that  $x_0$  is a relative minimum, and that  $x_0$  is a relative maximum if  $f''(x_0) < 0$ .

## Chapter 3

# The Integral

One interpretation of integration is as the inverse operation to differentiation. This leads to the *indefinite integral* of a function, defined as the collection of functions whose derivative is this function. Every rule for calculating derivatives can be reversed to give a rule for calculating indefinite integrals. In addition, various techniques can be developed for calculating indefinite integrals, but there are relatively simple functions which do not have an indefinite integral that can be expressed in terms of elementary functions.

The *definite integral* of a function over a given interval is defined as a kind of limit of sums, motivated by the idea of the area under a curve. Most applications involve definite integrals. However, the calculation of a definite integral from its definition is extremely complicated. Fortunately, there is a result so important as to be known as the *fundamental theorem of calculus*, which says that a definite integral can be calculated indirectly from the corresponding indefinite integral. This result justifies the effort to develop methods for calculating indefinite integrals.

### 3.1 The Indefinite Integral

If  $f(x)$  is a given function we define its *indefinite integral* (or *antiderivative*), denoted by

$$\int f(x)dx$$

to be the collection of functions of  $x$  whose derivative with respect to  $x$  is  $f(x)$ . Obviously, if  $F(x)$  is one such function then  $F(x) + c$  is another such function for every choice of the constant  $c$ . It is less obvious, but can be proved to be true, that there are no other functions with derivative  $f(x)$ . In other words, two functions having the same function as derivative can differ only by a constant. Thus if we know one function with derivative  $f(x)$ , then we obtain all others by adding on a constant. For this reason, we write formulae for indefinite integrals

like

$$\int x dx = \frac{1}{2}x^2 + c$$

with a constant  $c$ , called the *constant of integration*.

Every formula for calculating the derivative of a given function can be reversed to give a formula for an indefinite integral. Thus the relations (2.5) in Section 2.3 give the integration formulae

$$\begin{aligned} \int x^n dx &= \frac{x^{n+1}}{n+1} + c \quad (n \neq -1) \\ \int e^x dx &= e^x + c \\ \int \frac{1}{x} dx &= \ln x + c \\ \int \cos x dx &= \sin x + c \\ \int \sin x dx &= -\cos x + c \end{aligned} \tag{3.1}$$

The notation  $\int f(x)dx$  means that the result is a function of  $x$ . The expression  $\int f(u)du$  would mean a function of  $u$  whose derivative with respect to  $u$  is  $f(u)$ . The “ $dx$ ” is a marker to tell us what the independent variable is. Thus  $\int f(u)du$  would be a function of  $u$  whose derivative with respect to  $u$  is  $f(u)$ . Then if  $u$  is a function of  $x$ ,  $\int f(u)du$  would be a function of  $x$  whose derivative with respect to  $x$ , by the chain rule, is

$$f\{u(x)\}u'(x).$$

Now we have two different expressions, each of which described a function of  $x$  whose derivative with respect to  $x$  is  $f\{u(x)\}u'(x)$ , namely  $\int f(u)du$  and  $\int f\{u(x)\}u'(x)dx$ . Thus

$$\begin{aligned} \int f(u)du &= \int f\{u(x)\}u'(x)dx \\ &= \int f\{u(x)\}\frac{du}{dx}dx \end{aligned} \tag{3.2}$$

This is the *substitution rule* for indefinite integrals. To apply the substitution rule in evaluating an indefinite integral  $\int g(x)dx$ , we try to break  $g(x)$  into a product of two factors, one of which is the derivative of some function such that the other factor can be expressed in terms of this function. Then we make a substitution, letting  $u$  be this function. For example, in  $\int 2x(x^2 + 1)^{1/2}dx$ ,  $2x$  is the derivative of  $x^2 + 1$ , and if we let  $u = x^2 + 1$ , the integral is

$$\int u^{1/2}\frac{du}{dx}dx = \int u^{1/2}du = \frac{2}{3}u^{3/2} + c = \frac{2}{3}(x^2 + 1)^{3/2} + c.$$

The notation is devised so that if we could cancel  $dx$  from numerator and denominator in (3.2) we would obtain the right formula.

Another situation in which the substitution rule is useful is the evaluation of an integral in which the numerator is the derivative of the denominator. In the integral

$$\int \frac{g'(x)}{g(x)} dx$$

we let  $u = g(x)$  to obtain

$$\int \frac{g'(x)}{g(x)} dx = \int \frac{du}{u} = \ln u + c = \ln[g(x)] + c.$$

However, one must be careful not to write

$$\int \frac{dx}{g(x)} = \ln g(x) + c$$

which is a very common error.

Evaluation of indefinite integrals is a skill developed only with practice. Many integrals can be evaluated with the aid of the right substitution, but there are no firm rules for finding the right substitution. However, trying the wrong substitution does no harm – it may not help, but one can always start over and try something else.

Another useful technique for evaluating indefinite integrals is *integration by parts*. This is just the rule for differentiating a product of two functions turned around. The formula

$$\frac{d}{dx}(uv) = u \frac{dv}{dx} + v \frac{du}{dx}$$

gives

$$\begin{aligned} u \frac{dv}{dx} &= \frac{d}{dx}(uv) - v \frac{du}{dx} \\ \int u \frac{dv}{dx} dx &= \int \frac{d}{dx}(uv) dx - \int v \frac{du}{dx} dx \end{aligned}$$

and since

$$\int \frac{d}{dx}(uv) dx = uv + c$$

we obtain the integration by parts formula

$$\int u \frac{dv}{dx} dx = uv - \int v \frac{du}{dx} dx \quad (3.3)$$

To apply the formula, we try to write the function we are trying to integrate as the product of two factors, one of which ( $dv/dx$ ) is the derivative of a known function and the other of which ( $u$ ) has a reasonably simple derivative. For

example, to evaluate  $\int xe^x dx$ , we take  $dv/dx = e^x$ , so that  $v = e^x$ , and  $u = x$ . Then

$$\begin{aligned}\int xe^x dx &= \int x \frac{d}{dx}(e^x) dx = xe^x - \int e^x dx \\ &= xe^x - e^x + c\end{aligned}$$

As with integration by substitution, the choice of  $u$  and  $v$  is a skill to be developed with practice, and a choice which does not help can be discarded in favour of another attempt.

A table containing the integrals of all elementary functions might appear to be a useful aid to integration. However, there are two problems – not all elementary functions can be integrated in terms of elementary functions, and a complete table of those which can would be too cumbersome to be useful. On the other hand, a table of reasonable size can be very useful, if users are prepared to make substitutions to reduce unknown integrals to those in the table. Here is a brief table of integrals.

1.  $\int x^n dx = x^{n+1}/(n+1) + c \quad (n \neq -1)$
2.  $\int x^{-1} dx = \ell n|x| + c$
3.  $\int dx/(ax+b) = 1/a \ell n|ax+b| + c$
4.  $\int e^{ax} dx = e^{ax}/a + c$
5.  $\int xe^{ax} dx = (ax-1)e^{ax}/a^2 + c$
6.  $\int \ell n x dx = x \ell n x - x + c$
7.  $\int \sin ax dx = -\cos ax/a + c$
8.  $\int \cos ax dx = \sin ax/a + c$
9.  $\int \sin^2 x dx = x/2 - \sin 2x/4 + c$
10.  $\int \cos^2 x dx = x/2 + \sin 2x/4 + c$
11.  $\int dx/(a^2 - x^2) = \ell n|(a+x)/(a-x)|/2a + c$
12.  $\int dx/(x^2 - a^2) = \ell n|(x-a)/(x+a)| + c$

## 3.2 The Definite Integral

The definite integral is motivated by the idea of the area under a curve. Let  $y = f(x) \geq 0$  be a continuous function on the interval  $a \leq x \leq b$ . We begin with the simple idea of the area of a rectangle as the product of its base and its height and we attempt to describe the area under the curve as a sum of rectangles. Because one of the boundaries of the region is not a straight line, we can not do this exactly, but we can approximate. Let us partition the interval  $a \leq x \leq b$  into  $n$  equal subintervals of length  $h = (b-a)/n$  by defining  $x_0 = a, x_1 = a+h, x_2 = a+2h, \dots, x_{n-1} = a+(n-1)h, x_n = a+nh = a+(b-a) = b$ . In each of the subintervals we pick a point, say  $\xi_1$  in  $x_0 \leq x \leq x_1, \xi_2$  in  $x_1 \leq x \leq x_2, \dots, \xi_{n-1}$  in  $x_{n-2} \leq x \leq x_{n-1}, \xi_n$  in  $x_{n-1} \leq x \leq x_n$ . Then the sum

$$f(\xi_1)h + f(\xi_2)h + \dots + f(\xi_n)h \quad (3.4)$$

represents a sum of areas of rectangles, which if  $h$  is small ( $n$  is large) might be expected to be close to our intuitive idea of what the area under the curve should be.

We *define* the area under the curve to be the limit of this sum as  $h \rightarrow 0$  (or  $n \rightarrow \infty$ ), provided this limit exists and is independent of the choice of the points  $\xi_1, \xi_2, \dots, \xi_n$ . It is possible, but usually very cumbersome, to calculate this limit for suitably simple functions  $f(x)$ .

It is possible to define the sum (3.4) whenever the interval  $a \leq x \leq b$  is finite and the function  $f(x)$  is bounded on the interval  $a \leq x \leq b$ . We do not need to require  $f(x) \geq 0$  either; this was done only to give the idea of area under the curve, between the curve and the  $x$ -axis. However, the fact that we can define the sum (3.4) for a large class of intervals and functions does not assure us that the limit of the sum (3.4) as  $h \rightarrow 0$  exists for many functions.

There is a very important theorem stating that if the interval  $a \leq x \leq b$  is finite and if the function  $f(x)$  is *continuous* on the interval  $a \leq x \leq b$ , then the limit of the sum (3.4) does exist, regardless of how the points  $t_1, t_2, \dots, t_n$  are chosen. We say that a function for which this limit exists is *integrable* on the interval  $a \leq x \leq b$  and we call the limiting value of the sum the *definite integral of the function  $f(x)$*  over the interval  $a \leq x \leq b$ , denoted by

$$\int_a^b f(x)dx \quad (3.5)$$

The reason for the notation (3.5) is that we think of  $h$  as  $\Delta x$  and write the sum (3.4) as

$$\sum_{i=1}^n f(\xi_i)\Delta x$$

and think of the integral sign  $\int$  as a “limit” of the summation sign  $\sum$ . The content of the theorem mentioned above is that every continuous function is integrable.

It is important to remember that the “ $x$ ” and “ $dx$ ” in (3.5) are “dummy” variables. The definite integral (3.5) is a number, and if we had thought of  $y$  as a function of a different variable, say  $u$ , we would have obtained exactly the same number for the definite integral, which we would have called  $\int_a^b f(u)du$ . The name of the “variable” in a definite integral is completely irrelevant.

Another extremely important observation is that the definition of the definite integral has absolutely nothing to do with the earlier definition of the indefinite integral. However, the similarity of the notations for definite and indefinite integrals suggests that the two concepts are related. The connection between them is contained in the *fundamental theorem of calculus*.

### 3.3 The Fundamental Theorem of Calculus

In order to discover the relation between the indefinite integral and the definite integral, let us examine the function

$$F(x) = \int_a^x f(u)du. \quad (3.6)$$

Remember that although we have said that the definite integral is a number, (3.6) defines a function because of the variable upper limit of integration  $x$ . For each value of  $x$ , (3.6) defines a number  $F(x)$ , and thus we have a function. Note also that we prefer the form (3.6) to the form

$$\int_a^x f(x)dx$$

to avoid confusion between two different meanings of the same expression  $x$ .

The function  $F(x)$  represents the area under the curve  $y = f(x)$  from  $a$  to  $x$ . In order to find  $F'(x)$ , the derivative of this function, we must calculate

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

Now  $F(x+h) - F(x)$  represents the area under the curve  $y = f(x)$  between  $x$  and  $x+h$ , which is approximately a rectangle with base  $h$  and height  $f(x)$ , thus having area  $hf(x)$ . From this, we conclude that

$$\frac{F(x+h) - F(x)}{h}$$

is approximately  $f(x)$  and thus that

$$F'(x) = f(x) \quad (3.7)$$

This argument can be made rigorous. One part of the fundamental theorem of calculus says that if the function  $f(x)$  is continuous then

$$F(x) = \int_a^x f(u)du \quad (3.8)$$

is an indefinite integral of  $f(x)$ . That is,

$$F'(x) = f(x), \quad \frac{d}{dx} \int_a^x f(u)du = f(x).$$

We may use this fact to calculate the definite integral

$$\int_a^b f(x)dx$$

without having to form a limit of sums. We begin by finding an indefinite integral of  $f(x)$ , say  $G(x)$ . This indefinite integral  $G(x)$  is not necessarily the same as  $F(x)$ , but it differs from  $F(x)$  by a constant. Thus

$$F(x) = G(x) + c \tag{3.9}$$

for some constant  $c$ . The definite integral which we wish to calculate is  $F(b)$ , and the indefinite integral  $F(x)$  has the property that  $F(a) = 0$ . Now we have

$$\int_a^b f(x)dx = F(b) = G(b) + c. \tag{3.10}$$

In order to find  $c$ , we substitute in (3.9) to obtain  $G(a) + c = F(a) = 0$ . Thus  $c = -G(a)$  and from (3.10)

$$\int_a^b f(x)dx = G(b) - G(a).$$

In order to calculate the definite integral

$$\int_a^b f(x)dx$$

we find any indefinite integral of  $f(x)$  and “put on the limits of integration  $a$  and  $b$ ”.

In applications, we are usually faced with the problem of evaluating a definite integral. According to the fundamental theorem of calculus, we can accomplish this by finding an indefinite integral and thus make use of whatever techniques we have learned for calculating indefinite integrals. This is the importance of the fundamental theorem of calculus.

One of the important techniques for the calculation of indefinite integrals is the substitution rule (3.2). We may use a substitution  $u = u(x)$  to replace an indefinite integral with respect to  $x$  by an indefinite integral with respect to  $u$  which we then evaluate, and after this evaluation we return to the original variable  $x$ . For a definite integral, we do not need to return to the original variable but can instead change the limits of integration. The substitution rule for definite integrals is that if

$$u(\alpha) = a, \quad u(\beta) = b$$

then

$$\int_a^b f(u)du = \int_\alpha^\beta f\{u(x)\}u'(x)dx. \quad (3.11)$$

It is not always possible to find an indefinite integral of a given functions. However, there are numerical methods for approximating definite integrals which are easily implemented on a microcomputer, or even in some cases on a hand-held calculator. Such methods are especially useful for functions which are not given by an analytic expression but are instead given by a table of values (e.g., from experimental data).

One way to estimate a definite integral makes use of the idea of the linear approximation.

**Example 2.** Estimate  $\int_0^{0.2} e^{-t^2} dt$ .

**Solution.** We use the function  $F(x) = \int_0^x e^{-t^2} dt$ , for which  $F(0) = 0$  and (by the rule for differentiation of integrals with variable upper limits),  $F'(x) = e^{-x^2}$ . In particular,  $F'(0) = e^0 = 1$ . Thus the linear approximation to  $F(x)$  is  $F(0) + xF'(0) = x$ , and  $\int_0^{0.2} e^{-t^2} dt = F(0.2) \approx 0.2$ . However, there are

numerical methods with much higher accuracy.

### 3.4 Some exercises

In each of Exercises 1–6, evaluate the indicated indefinite integral.

1.  $\int x^2 dx$
2.  $\int x^{-1/3} dx$
3.  $\int e^{-x} dx$
4.  $\int e^{2x} dx$
5.  $\int \frac{2}{x} dx$
6.  $\int \frac{1}{x^2} dx$

In each of Exercises 7–10, evaluate the indicated definite integral.

7.  $\int_2^5 (x^3 + 1) dx$
8.  $\int_1^2 x^{-2/3} dx$
9.  $\int_0^1 e^{-2x} dx$
10.  $\int_1^{10} \frac{dx}{x}$

11. Find the value of the derivative at  $x = 1$  of the function

$$f(x) = \int_0^x e^{-t^2} dt.$$



## Chapter 4

# Multivariable Calculus

### 4.1 Functions of Two Variables

There are many situations in which a quantity depends on more than one variable. Thus we are led to consider functions of two variables, of the form

$$z = f(x, y).$$

More generally, we might consider functions of an arbitrary number of variables. However, as the essential differences between the properties of functions of one variable and functions of more than one variable are displayed for functions of two variables, we shall restrict our attention to functions of two variables in developing the basic theory.

**Example 1.** In regions with severe winter weather, the wind-chill index is often used to describe the apparent severity of the cold. This index is a subjective temperature depending on the actual temperature and the wind velocity. The origin of the wind-chill index is in the work of P. A. Siple and C. F. Passel (1945) in the Antarctic measuring the rate of heat loss from a can of water; later refinements were based on studies of body heat loss and were thus more closely related to subjective reality.

Air temperature (°F)																
–	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40
0	35	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30	-35	-40
5	32	27	22	16	11	6	1	-5	-10	-15	-20	-26	-31	-36	-41	-47
10	22	16	10	4	-3	-9	-15	-21	-27	-33	-40	-46	-52	-58	-64	-70
15	16	9	2	-5	-11	-18	-26	-32	-38	-45	-52	-58	-65	-72	-79	-85
20	11	4	-3	-10	-17	-25	-32	-39	-46	-53	-60	-67	-74	-82	-89	-96
25	8	0	-7	-16	-22	-29	-37	-44	-52	-59	-66	-74	-81	-89	-96	-104
30	5	-2	-10	-18	-25	-33	-41	-48	-56	-63	-71	-79	-86	-94	-102	-109
35	3	-4	-12	-20	-28	-35	-43	-51	-59	-67	-74	-82	-90	-98	-106	-113
40	2	-6	-14	-22	-29	-37	-45	-53	-61	-69	-77	-85	-93	-101	-108	-116
45	1	-1	-15	-23	-31	-39	-47	-55	-62	-70	-78	-86	-94	-102	-110	-118

Table 4.1: Wind chill as function of temperature and wind velocity

However, the wind-chill index is subjective, and is not uniquely defined. It is invariably described by a table, such as the one given in Table 1 rather than by an explicit formula.

To determine a wind-chill index for a given temperature and wind velocity from this table, find the entry in the same column as the temperature along the top of the table and in the same row as the wind velocity down the left side of the table. Thus, for example, if the temperature is  $5^\circ$  and the wind velocity is 20 mph, the wind-chill index is  $-32^\circ$ , meaning that subjectively it would feel as cold as a temperature of  $-32^\circ$  with zero wind velocity.

**Example 2.** Weather maps often give temperature contour lines - curves on the map connecting places with the same temperature. These curves may actually consist of several separate curves. Such a map gives a way of estimating the temperature at a given location.

**Example 3.** Suppose that at time  $t$  the number of bicyclists at milepost  $x$  along a straight road is a function  $u(x, t)$  such that

$$u(x, t) = f(x - ct),$$

where  $f$  is a specified function of one variable. Then  $u(x, 0) = f(x)$  the distribution of bicyclists at time  $t = 0$ , and  $u(x, 1) = f(x - c)$  gives the distribution of bicyclists at time  $t = 1$ . But the graph of  $f(x - c)$  is the same as the graph of  $f(x)$  shifted  $c$  units to the right. Thus  $u(x, t)$  represents a wave with shape  $f(x)$  moving to the right with velocity  $c$ .

### 4.1.1 Graphic representation of functions

A function of one variable  $y = f(x)$  is represented graphically by the set of points  $(x, y)$  in two-dimensional space such that  $y = f(x)$ . Generally, this graph is a curve. Similarly, a function of two variables  $z = f(x, y)$  is represented

graphically by the set of points  $(x, y, z)$  in three-dimensional space such that  $z = f(x, y)$ . Generally, this is a two-dimensional surface. Because our ability to visualize in three dimensions is much weaker than our ability to visualize in two dimensions, we attempt to visualize the surface representing a function  $z = f(x, y)$  by examining its cross sections by planes parallel to the coordinate planes.

The intersection of the surface  $z = f(x, y)$  with a plane  $y = y_0$  parallel to the  $x - z$  plane is a curve  $z = f(x, y_0)$  in this plane. Analytically, we can think of this curve as representing  $z$  as a function of  $x$  for fixed  $y = y_0$ . Similarly, the intersection of the surface  $z = f(x, y)$  with a plane  $x = x_0$  parallel to the  $y - z$  plane is a curve  $z = f(x_0, y)$  in this plane which we can think of as representing  $z$  as a function of  $y$  for fixed  $x = x_0$ .

The intersection of the surface  $z = f(x, y)$  with a horizontal plane  $z = c$  is called a *contour line*. Its projection on the  $x - y$  plane is a curve given implicitly by the equation  $f(x, y) = c$ , called a *level curve* of the function. A level curve may consist of several different component curves. A weather map as described in Example 2 above is in fact a description of the temperature function by its level curves. A topographical survey map is a map which includes level curves giving the altitude as a function of position and providing a way of visualizing a three-dimensional landscape by a plane drawing.

**Example 4.** The level curves of the function  $z = x^2 + y^2$  are the curves  $x^2 + y^2 = c$  in the  $x - y$  plane. If  $c > 0$ , the level curve is a circle with center at the origin; if  $c = 0$  the level “curve” is the point  $(0, 0)$ ; if  $c < 0$ , there is no level curve. This tells us that the graph of the function  $z = x^2 + y^2$  lies above the  $x - y$  plane, touching the  $x - y$  plane at the origin. Another way to describe this information is to say that the function  $z = x^2 + y^2$  has a minimum value of zero, attained at the origin  $x = 0, y = 0$ .

**Example 5.** The level curves of the function  $z = xy$  are the curves  $xy = c$  in the  $x - y$  plane. If  $c > 0$ , the level curve is a hyperbola with branches in the first and third quadrants; if  $c = 0$  the level “curve” is the point  $(0, 0)$ ; if  $c < 0$ , the level curve is a hyperbola with branches in the second and fourth quadrants. Thus the surface describing the function  $z = xy$  goes up from the origin in the first and third quadrants and down from the origin in the second and fourth quadrants. The origin is said to be a *saddle point* of the function.

**Example 6.** Human blood pressure  $p$  may be considered as a function of cardiac output  $x$ , the volume of blood flowing through a person’s heart, and the systems vascular resistance, or SVR, the resistance to blood flowing through veins and arteries. It is sometimes assumed that the functional relation has the form  $p = kxy$  for some constant  $k$ , but a less specific assumption which may be of use for qualitative estimates is that  $\frac{\partial p}{\partial x} > 0, \frac{\partial p}{\partial y} > 0$ . There are medications such as nitroglycerine which lower the SVR, and there are medications such as dopamine which increase cardiac output. For a person with a weak heart it may be considered necessary to increase the cardiac output. However, medication

which accomplishes this will also have the undesirable effect of increasing the blood pressure. Therefore such medication is normally administered along with medication which lowers  $SVR$  to keep the blood pressure constant, or move along a level curve of the function  $p$ . A heart attack causes a drop in cardiac output, thus producing a drop in blood pressure. Medication to decrease  $SVR$  would increase the blood pressure but would not have any effect on cardiac output. If blood pressure is taken as a measure of cardiac output, such medication would appear to be beneficial. However, following a heart attack it is more important to restore a normal cardiac output. Here, the variable  $x$  is the essential and the function  $p$  is secondary in importance.

### 4.1.2 Linear functions

In one variable linear function (represented by straight lines) form a particularly simple class of functions to approximate arbitrary functions near a base point. Geometrically, this idea is expressed by the statement that the tangent line to a curve at a point approximates the curve near that point. For two variables, a *linear function* is a function of the form

$$z = ax + by + c,$$

where  $a$ ,  $b$ , and  $c$  are constants. The surface represented by a linear function is a *plane*. The intersection of the plane  $z = ax + by + c$  with a plane  $y = y_0$  parallel to the  $x - z$  plane has  $z = ax + (by_0 + c)$  and is a straight line with slope  $a$ . Observe that this slope is the same at every point. The intersection of the plane  $z = ax + by + c$  with a plane  $x = x_0$  parallel to the  $y - z$  plane has  $z = by + (ax_0 + c)$  and is a straight line with slope  $b$ . Again, this slope is the same at every point. However, the slope  $a$  in the  $x$ -direction and the slope  $b$  in the  $y$ -direction may be different. On a plane, the slope in a given direction is the same at every point, but the slope, or rate of change of the function, depends on the direction. This will turn out to be an essential difference between functions of one variable and functions of two variables.

## 4.2 Limits, Continuity, and Partial Derivatives

### 4.2.1 Limits and continuity

The definitions of limits and continuity for functions of two variables are the same as the corresponding definitions for functions of one variable provided we think of  $(x, y)$  as a point in the  $xy$ -plane rather than as a pair of numbers. Then a function  $f(x, y)$  of two variables is defined for points  $(x, y)$  in some plane region, which is called the *domain* of the function. We may speak of a limit as  $(x, y)$  approaches  $(x_0, y_0)$ , or as the point  $P$  with coordinates  $(x, y)$  approaches the point  $P_0$  with coordinates  $(x_0, y_0)$ , or as  $x$  approaches  $x_0$  and  $y$  approaches  $y_0$ . The statement

$$\lim_{P \rightarrow P_0} f(x, y) = L$$

means that the value of the function  $f(x, y)$  is close to the number  $L$ , in the sense that the difference of absolute values

$$|f(x, y) - L|$$

is small for every point  $P$  sufficiently close to  $P_0$ . Here  $P$  close to  $P_0$  means that the distance

$$[(x - x_0)^2 + (y - y_0)^2]^{1/2}$$

between  $P$  and  $P_0$  is small or, equivalently, that the absolute values of the differences in coordinates  $|x - x_0|$  and  $|y - y_0|$  are both small.

In the definition of limit, we specify that  $|f(x, y) - L|$  must be small for *all* points  $(x, y)$  close to  $(x_0, y_0)$  (not including  $(x_0, y_0)$ ). Difficulties can arise which are not possible for functions of one variable. Consider, for example, the problem of determining

$$\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}.$$

If the point  $(x, y)$  is on the line  $y = mx$  of slope  $m$  through the origin, then

$$f(x, y) = f(x, mx) = \frac{x(mx)}{x^2 + (mx)^2} = \frac{m}{1 + m^2}.$$

Thus the function is constant on each such line, and there are points arbitrarily close to the origin at which the function has the value  $m/(1 + m^2)$  for every value of  $m$ ,  $-\infty < m < \infty$ . These values  $m/(1 + m^2)$  range from  $-\frac{1}{2}$  to  $\frac{1}{2}$ . The function does not have a limit at the origin even though it appears to be a simple function of  $x$  and  $y$ . The problem is that  $xy/(x^2 + y^2)$  is indeterminate when  $x$  and  $y$  are both zero, even though it is well-defined if either  $x$  or  $y$  is different from zero.

Even if a function has a limit as  $(x, y)$  approaches the origin along every line through the origin, it still may not have a limit at the origin. For example, the function

$$g(x, y) = \frac{x^2 y}{x^4 + y^2}$$

on any line  $y = mx$  has the value

$$g(x, mx) = \frac{x^2(mx)}{x^4 + (mx)^2} = \frac{mx^3}{x^4 + m^2x^2} = \frac{mx}{x^2 + m^2}$$

and this approaches zero as  $x$  approaches zero for every value of  $m$ . However, for a point on the parabola  $y = x^2$ ,

$$g(x, y) = g(x, x^2) = \frac{x^4}{x^4 + x^4} = \frac{1}{2}$$

and thus as  $(x, y)$  approaches  $(0, 0)$  along this parabola  $g(x, y)$  approaches  $1/2$ . Thus this function fails to approach a limit as  $(x, y)$  approaches the origin. For a function  $f(x, y)$  to have a limit  $L$  at a point  $(x_0, y_0)$  the value  $f(x, y)$  must be

close to  $L$  whenever  $(x, y)$  is close to  $(x_0, y_0)$ , independent of the path by which  $(x, y)$  approaches  $(x_0, y_0)$ .

Continuity is defined for functions of two variables in terms of limits, just as for functions of one variable. Thus a function  $f(x, y)$  is *continuous* at the point  $(x_0, y_0)$  if

$$f(x_0, y_0) = \lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y).$$

Continuous functions have the same properties with respect to maximum and minimum values as functions of one variable. In particular, if  $f(x, y)$  is continuous at every point of a set of points  $D$  in the plane which is closed (includes all boundary points) and bounded (is contained in some sufficiently large disc  $(x^2 + y^2)^{1/2} \leq R$ ), then the function  $f(x, y)$  attains an absolute maximum at some point of the domain  $D$ , and also attains an absolute minimum at some point of  $D$ . The determination of maxima and minima by methods involving differentiation is more complicated than for functions of one variable because the concept of derivative is more complicated.

### 4.2.2 Partial derivatives

In describing the rate of change of a function  $f(x, y)$  at a point  $(x_0, y_0)$  we must specify the direction. In general, the rate of change of a function will depend on the direction. We pick two preferred directions – the positive  $x$ -direction and the positive  $y$ -direction and define the *partial derivatives* as the derivatives in these directions, respectively

$$\begin{aligned} \frac{\partial f}{\partial x}(x_0, y_0) &= f_x(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}, \\ \frac{\partial f}{\partial y}(x_0, y_0) &= f_y(x_0, y_0) = \lim_{k \rightarrow 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k} \end{aligned}$$

Both notations,  $\frac{\partial f}{\partial x}$  or  $\frac{\partial f}{\partial y}$ , and  $f_x$  or  $f_y$ , are in common use. Because in the definition of each partial derivative the other variable is held fixed and thus may be viewed as a constant, all the rules for calculation of derivatives of functions of one variable carry over to rules for calculation of partial derivatives. It is possible to define the derivative of a function of two variables in a given direction and to express this directional derivative in terms of the two partial derivatives.

As each partial derivative can be differentiated partially with respect to either variable, we would expect that there are four different second order partial derivatives.  $\frac{\partial^2 f}{\partial x^2} = \frac{\partial f}{\partial x} \left( \frac{\partial f}{\partial x} \right)$  or  $f_{xx}(x, y) = [f_x(x, y)]_x$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) \text{ or } f_{xy}(x, y) = [f_x(x, y)]_y$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial y} \right) \text{ or } f_{yx}(x, y) = [f_y(x, y)]_x$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right) \text{ or } f_{yy}(x, y) = [f_y(x, y)]_y$$

(observe that the order of the two variables is different in the two notations).

However, there are actually only three second order partial derivatives: If the *mixed partial derivatives*  $f_{xy}(x, y)$  and  $f_{yx}(x, y)$  are continuous, then there is a theorem which states that they are identical

$$f_{xy}(x, y) = f_{yx}(x, y).$$

The partial derivative  $f_x(x, y)$  represents the rate of change of  $f$  in the  $x$ -direction. Another interpretation is to consider the function  $f(x, y_0)$ , with  $y$  fixed as  $y_0$ , as a function of one variable; then  $f_x(x, y_0)$  is the derivative of this function, or the slope of the curve given by the intersection of the surface  $z = f(x, y)$  and the plane  $y = y_0$ . The partial derivative  $f_y(x, y)$  has a similar interpretation as the slope of the curve given by the intersection of the surface  $z = f(x, y)$  and the plane  $x = x_0$ .

### 4.3 Local Linearity

A fundamental property of differentiable functions is that *locally* (close enough to a starting point) they may be approximated by linear functions. We have seen this earlier for functions of one variable. Here we will develop the fundamental idea of the linear approximation of a function of two variables, and as an example this idea will be used to study the chain rule.

We would like to estimate

$$f(x_0 + h, y_0 + k) - f(x_0, y_0)$$

for a function  $f$  of two variables. We do this in two stages, first moving from  $(x_0, y_0)$  to  $(x_0 + h, y_0)$ , and then moving from  $(x_0 + h, y_0)$  to  $(x_0 + h, y_0 + k)$  (see Figure 4.1). In each stage only one variable changes, and the rate of change can be measured by a partial derivative.

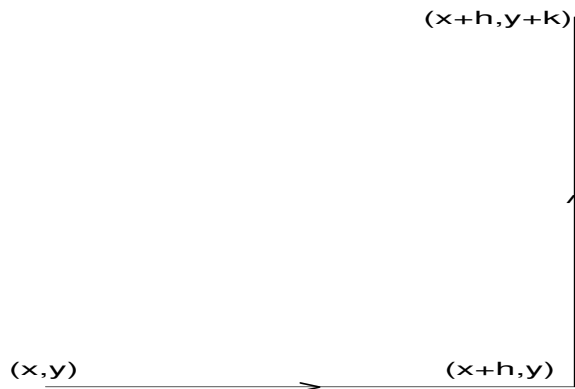


Figure 4.1:

The definition of the partial derivative of a function  $f$  of two variables at a point  $(x_0, y_0)$  is

$$f_x(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}.$$

Another way to express this definition is to say that

$$\frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = f_x(x_0, y_0) + \varepsilon_1 \quad (4.1)$$

with  $\varepsilon_1$ , “small” in the sense that  $\varepsilon_1 \rightarrow 0$  as  $h \rightarrow 0$ . The relation (4.1) may be rewritten

$$f(x_0 + h, y_0) - f(x_0, y_0) = hf_x(x_0, y_0) + \varepsilon_1 h \quad (4.2)$$

To move from  $(x_0 + h, y_0)$  to  $(x_0 + h, y_0 + k)$ , we use the definition of the partial derivative  $f_y$ ,

$$f_y(x_0 + h, y_0) = \lim_{k \rightarrow 0} \frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0)}{k} \quad (4.3)$$

We may write (4.1) as

$$\frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0)}{k} = f_y(x_0 + h, y_0) + \varepsilon_2$$

or

$$f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) = kf_y(x_0 + h, y_0) + \varepsilon_2 k \quad (4.4)$$

Addition of (4.1) and (4.4) gives

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = hf_x(x_0, y_0) + kf_y(x_0 + h, y_0) + \varepsilon_1 h + \varepsilon_2 k \quad (4.5)$$

The relation (4.5) is not the form we want; we wish to replace  $f_y(x_0 + h, y_0)$  by  $f_y(x_0, y_0)$ . If we assume that the partial derivative  $f_y$  is continuous (this is not necessary for the truth of the result, but it simplifies the calculations), we may write

$$f_y(x_0 + h, y_0) = f_y(x_0, y_0) + \varepsilon_3 \quad (4.6)$$

with  $\varepsilon_3 \rightarrow 0$  as  $h \rightarrow 0$ . Now, substitution of (4.6) into (4.5) gives

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = hf_x(x_0, y_0) + kf_y(x_0, y_0) + R \quad (4.7)$$

where

$$R = \varepsilon_1 h + \varepsilon_2 k + \varepsilon_3 k.$$

Then  $R$  is “small” relative to  $h$  and  $k$  in the sense that

$$\lim_{(h,k) \rightarrow (0,0)} \frac{R}{\sqrt{h^2 + k^2}} = 0 \quad (4.8)$$

The quantity  $\sqrt{h^2 + k^2}$  in the denominator in (4.8) is the distance from the point  $(x_0, y_0)$  to the point  $(x_0 + h, y_0 + k)$ . We may let  $x = x_0 + h, y = y_0 + k$  and write (4.7) in the form

$$f(x, y) = f(x_0, y_0) + (x - x_0)f_x(x_0, y_0) + (y - y_0)f_y(x_0, y_0) + R \quad (4.9)$$

Thus the expression

$$f(x_0, y_0) + (x - x_0)f_x(x_0, y_0) + (y - y_0)f_y(x_0, y_0) \quad (4.10)$$

which is linear in  $x$  and  $y$ , approximates  $f(x, y)$  with an error  $R$ . The relation (4.8) says that the approximation is good if  $h$  and  $k$  are small, that is, if the point  $(x, y)$  is close to the point  $(x_0, y_0)$ . The approximation (4.10) is called the *linear approximation* to  $f(x, y)$  at  $(x_0, y_0)$ . Just as for functions of one variable, a function of two variables behaves locally (near  $(x_0, y_0)$ ) like a linear function. The surface

$$z = f(x_0, y_0) + (x - x_0)f_x(x_0, y_0) + (y - y_0)f_y(x_0, y_0),$$

corresponding to the linear approximation is the *tangent plane* to the surface  $z = f(x, y)$  at the point  $(x_0, y_0, f(x_0, y_0))$ .

**Example 1.** Find the linear approximation to the function  $f(x, y) = x^2 + 3y^2$  at the point  $(3, -1)$ .

**Solution.** We calculate  $f_x(x, y) = 2x$  and  $f_y(x, y) = 6y$ . Thus  $f_x(3, -1) = 6$  and  $f_y(3, -1) = -6$ . Since  $f(3, -1) = 12$ , the linear approximation is  $f(3, -1) + f_x(3, -1)(x - 3) + f_y(3, -1)(y + 1) = 12 + 6(x - 3) - 6(y + 1)$ . While this could be simplified algebraically to  $6x - 6y - 12$ , it is often preferable to leave the answer in the form  $12 + 6(x - 3) - 6(y + 1)$ .

**Example 2.** Find the tangent plane to the surface  $z = x^2 + 3y^2$  at the point  $(3, -1)$ .

**Solution.** This is a geometric formulation of exactly the same problem as Example 1. The tangent plane is  $z = 6(x - 3) - 6(y + 1) + 12$  or  $z = 6x - 6y - 12$ .

The local linearity of functions of several variables is the key to many topics in the differential calculus of multivariable functions, including directional derivatives, the chain rule, implicit differentiation, and the total differential. However, we shall not explore these topics here. We do, however, give without proof a useful result on differentiation of integrals that can be proved with the aid of implicit differentiation.

We consider a function defined by an integral of a function of two variables with respect to one of the variables whose upper limit of integration is variable, of the form

$$F(t) = \int_a^t f(t, s) ds.$$

Functions having this form arise in some general models for disease transmission. It can be proved that if  $f(t, s)$  and its partial derivative  $f_t(t, s)$  with respect to  $t$ , are continuous, then  $F(t)$  is differentiable and

$$F'(t) = f(t, t) + \int_a^t f_t(t, s) ds.$$

This result extends the rule for differentiating integrals given by the fundamental theorem of calculus.

#### 4.4 Some exercises

1. For the wind-chill index described by Table 4.1,
  - (a) What wind velocity with temperature  $20^\circ F$  would give the same wind-chill index as a temperature of  $0^\circ F$  and a wind velocity of 10 mph?
  - (b) What temperature with wind velocity 20 mph would give the same wind-chill index as a temperature of  $0^\circ F$  and a wind velocity of 10 mph?
  - (c) Is the effect of increasing wind velocity on the wind-chill index more pronounced on less pronounced as the wind velocity increases?
  - (d) Describe wind velocity as a function of temperature to give the condition that the wind-chill index is equal to  $-20^\circ F$ .
2. If  $f$  is a specified function of one variable, what does the function  $u(x, t) = f(x + ct)$  represent?
3. What are the shapes of the vertical cross sections of the function  $z = x^2 + y^2$ ?
4. What are the level curves of the function  $z = e^{x^2+y^2}$ ?

In each of Exercises 5–8, calculate the partial derivatives of the given function

5.  $f(x, y) = x^2 - xy + y^2$

6.  $f(x, y) = e^{xy}$

7.  $f(x, y) = 4x^{1/4}y^{3/4}$

8.  $f(x, y) = \frac{xy}{x^2+y^2}$

9. For the function  $f(L, K) = bL^\alpha K^\beta$ , verify that

$$Kf_K(L, K) + Lf_L(L, K) = (\alpha + \beta)f(L, K).$$

10. Let  $W$  be the wind-chill index function of air temperature  $T$  and wind velocity  $V$  described by Table 4.1. Estimate  $\frac{\partial W}{\partial T}$  and  $\frac{\partial W}{\partial V}$  when  $T = 10, V = 10$ . In general, what can you say about the sign of  $\frac{\partial W}{\partial T}$  and of  $\frac{\partial W}{\partial V}$ ? What appears to be the value of  $\lim_{V \rightarrow \infty} \frac{\partial W}{\partial V}$ ?

In each of Exercises 11–12, find the linear approximation to the given function at the given point, and the tangent plane to the given surface at the given point.

11.  $z = x^2 - 4y^2$  at  $(2, 1)$
12.  $z = xy$  at  $(3, 0)$
13. For the function  $f(x, y) = e^{xy}$
- (a) Find the linear approximation to  $f(x, y)$  at  $(1, 0)$ .
  - (b) Find the linear approximation to  $f_x(x, y)$  at  $(1, 0)$ .
  - (c) Explain why the partial derivative with respect to  $x$  of the linear approximation to  $f(x, y)$  at  $(1, 0)$  is not the same as the linear approximation to  $f_x(x, y)$  at  $(1, 0)$ .



Part II

**MATRIX ALGEBRA**



## Chapter 5

# Some Properties of Vectors and Matrices

### 5.1 Introduction

A single linear algebraic equation in one unknown has the form

$$ax = b.$$

A system of  $m$  linear equations in  $n$  variables is a system of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m. \end{aligned}$$

The language of vectors and matrices makes it possible to write this system in the simple form  $Ax = b$ , where  $A$  is an  $m \times n$  *matrix* and  $b$  is a *column vector*. We will develop the elementary theory of vectors and matrices to show how to use this language to simplify the analysis of linear systems. This will be useful in the study of systems of differential equations that arise in epidemic models. In models consisting of a system of two differential equations in two unknown functions the matrices and vectors involved will have  $m = n = 2$ , and readers should interpret the results in this chapter with  $m = n = 2$ . We may think of matrix algebra as machinery that will allow us to simplify the language of epidemiological models.

## 5.2 Vectors and Matrices

A real  $n$ -vector is an ordered  $n$ -tuple of real numbers of the form

$$v = [a_1, a_2, \dots, a_n].$$

We also write  $v$  in the form (*column vector*)

$$v = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}. \quad (5.1)$$

In these notes, we use the form (5.1), since this will simplify calculations later. The real numbers will be called *scalars*.

We have the following two vector operations:

1. *Addition*, which is given by the formula

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix};$$

2. *Scalar multiplication*, defined by

$$\lambda \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{bmatrix}.$$

We define the *scalar product* (also called *inner* or *dot* product) of two vectors by

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \dots + a_n b_n.$$

**Example 1.** Consider the vectors

$$v_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Then

$$v_1 + v_2 = \begin{bmatrix} -1 \\ 4 \\ 0 \end{bmatrix}, \quad 3v_1 = \begin{bmatrix} 3 \\ 9 \\ -6 \end{bmatrix},$$

$$v_1 \cdot v_2 = (1)(-2) + (3)(1) + (-2)(2) = -3.$$

It is not hard to see that the operations defined above have the following properties:

$$\begin{aligned} u + v &= v + u \\ (u + v) + w &= u + (v + w) \\ \lambda(u + v) &= \lambda u + \lambda v \\ u \cdot v &= v \cdot u \\ u \cdot (v + w) &= u \cdot v + u \cdot w \\ u \cdot (\lambda v) &= \lambda u \cdot v \end{aligned}$$

An  $m \times n$  matrix is an array of  $mn$  real numbers with  $m$  rows and  $n$  columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \quad (5.2)$$

The matrix  $A$  is also written as  $A = (a_{ij})$ . The *size* of  $A$  is  $m \times n$ . Matrices of the same size can be added, and matrices can be multiplied by scalars, according to the following rules:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix},$$

$$\lambda \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{bmatrix}.$$

**Example 2.** Consider the matrices

$$A = \begin{bmatrix} 2 & -1 & 3 \\ -2 & 5 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & -1 & 3 \\ 4 & 4 & -2 \end{bmatrix}.$$

Then we have

$$\begin{aligned} A + B &= \begin{bmatrix} 2 & -2 & 6 \\ 2 & 9 & -2 \end{bmatrix}, \\ 5A &= \begin{bmatrix} 10 & -5 & 15 \\ -10 & 25 & 0 \end{bmatrix}, \\ -B &= \begin{bmatrix} 0 & 1 & -3 \\ -4 & -4 & 2 \end{bmatrix}. \end{aligned}$$

The matrix operations satisfy the following properties:

$$\begin{aligned} A + B &= B + A \\ A + (B + C) &= (A + B) + C \\ \lambda(A + B) &= \lambda A + \lambda B. \end{aligned}$$

If  $A = (a_{ij})$  is an  $m \times n$  matrix, and  $B = (b_{jk})$  is an  $n \times p$  matrix, then the product  $AB$  is defined as the  $m \times p$  matrix given by  $C = (c_{ik})$  where

$$c_{ik} = \sum_{j=1}^n a_{ij}b_{jk}.$$

That is, the  $(i, k)$ -element of  $C$  is the scalar product of the  $i$ th row of  $A$  and the  $k$ th column of  $B$ . Note that the product of two matrices  $A$  and  $B$  is defined only when the number of columns of  $A$  is equal to the number of rows of  $B$ .

**Example 3.** Let

$$A = \begin{bmatrix} 2 & -1 & 3 \\ -2 & 5 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 4 \\ 0 & 1 & -1 \\ -2 & 5 & 0 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} -2 & 12 & 9 \\ -4 & 7 & -13 \end{bmatrix}.$$

We can check that the matrix product satisfies the following properties

$$\begin{aligned} A(BC) &= (AB)C \\ A(B + C) &= AB + AC \\ A(\lambda B) &= \lambda AB. \end{aligned}$$

It is important to note that the matrix product **is not commutative**, that is, in general  $AB \neq BA$ . Moreover, if the product  $AB$  is defined, in general  $BA$  is not defined.

A matrix is called a *square matrix* if the number of its rows is equal to the number of its columns. The *main diagonal* of an  $n \times n$  square matrix  $A = (a_{ij})$  is the  $n$ -tuple  $(a_{11}, a_{22}, \dots, a_{nn})$ . A square matrix  $D$  is called a *diagonal matrix* if all its elements are zero with the exception of its diagonal:

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

we also write  $D$  as  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ .

**Example 4.** Let

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & -1 \\ -2 & 5 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \text{diag}(2, -1, 5).$$

Then

$$AD = \begin{bmatrix} 4 & 1 & 20 \\ 0 & -1 & -5 \\ -4 & -5 & 0 \end{bmatrix},$$

$$DA = \begin{bmatrix} 4 & -2 & 8 \\ 0 & -1 & 1 \\ -10 & 25 & 0 \end{bmatrix}.$$

The example above suggests the following fact, which is indeed true: If  $A = (v_1 \ v_2 \ \dots \ v_n)$  is a square matrix whose columns are the  $n$ -vectors  $v_1, v_2, \dots, v_n$ ,

and  $B = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$  is a square matrix whose rows are the  $n$ -vectors  $u_1, u_2, \dots, u_n$ ,

then

$$A \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = (\lambda_1 v_1 \ \lambda_2 v_2 \ \dots \ \lambda_n v_n) \quad (5.3a)$$

$$\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)B = \begin{bmatrix} \lambda_1 u_1 \\ \lambda_2 u_2 \\ \vdots \\ \lambda_n u_n \end{bmatrix}. \quad (5.3b)$$

The  $n \times n$  *identity matrix* is the matrix  $I = \text{diag}(1, 1, \dots, 1)$ , and satisfies

$$AI = IA = A \quad (5.4)$$

for all  $n \times n$  matrices  $A$ .

### 5.3 Systems of Linear Equations

A system of  $m$  linear equations in  $n$  variables is a system of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m. \end{aligned}$$

We can write this system in the form  $Ax = b$ , where

$$A = (a_{ij}), \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

The information contained in the above system is also contained in the *augmented matrix*  $A|b$  with  $n+1$  rows and  $m$  columns. The solutions of the system are not changed by adding a multiple of one equation to another equation, multiplication of an equation by a non-zero constant, or interchanging two equations. This statement is equivalent to the statement that the set of solutions is not changed if the augmented matrix is subjected to a sequence of *elementary row operations*. There are three types of elementary row operations, namely

1. addition of a multiple of one row to another row,
2. multiplication of a row by a non-zero constant,
3. interchange of two rows.

The system can be solved by reducing the matrix  $A$  to a triangular matrix using elementary row operations, a method called the Gauss-Jordan method. We demonstrate the method by an example.

**Example 1.** Solve the system

$$\begin{aligned} x - 3y - 5z &= -8 \\ -x + 2y + 4z &= 5 \\ 2x - 5y - 11z &= -9. \end{aligned}$$

We add the first row to the second row, subtract double the first row from the third row and multiply the new second row by  $-1$ . Next we add 3 times the second row to the first row and subtract the second row from the third row. Finally we subtract the third row from the first row, multiply the third row by

$-1/2$ , and subtract the third row from the second row. These operations give the sequence of augmented matrices

$$\begin{aligned} \left( \begin{array}{ccc|c} 1 & -3 & -5 & -8 \\ -1 & 2 & 4 & 5 \\ 2 & -5 & -11 & -9 \end{array} \right) &\sim \left( \begin{array}{ccc|c} 1 & -3 & -5 & -8 \\ 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & 7 \end{array} \right) \\ &\sim \left( \begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & -2 & 4 \end{array} \right) \\ &\sim \left( \begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -2 \end{array} \right). \end{aligned}$$

Then we may read off the solution from the final form as  $x = -3, y = 5, z = -2$ .

## 5.4 The Inverse Matrix

Let  $A$  be a square  $n \times n$  matrix. The inverse of  $A$ , if it exists, is a matrix  $A^{-1}$  such that

$$AA^{-1} = A^{-1}A = I.$$

Not every matrix  $A$  has a matrix. If  $A$  has a matrix, then it is called *invertible*, and its inverse is unique.

**Example 1.** Find the inverse of the matrix

$$A = \begin{bmatrix} 1 & -3 & -5 \\ -1 & 2 & 4 \\ 2 & -5 & -11 \end{bmatrix}.$$

The inverse of  $A$  must be of the form (5.2), so we have to solve the system of 9 linear equations

$$A \cdot (x_{ij}) = I.$$

As in the previous example, we use the Gauss-Jordan method to solve this system of equations. The first and last steps of the algorithm are

$$\left( \begin{array}{ccc|ccc} 1 & -3 & -5 & 1 & 0 & 0 \\ -1 & 2 & 4 & 0 & 1 & 0 \\ 2 & -5 & -11 & 0 & 0 & 1 \end{array} \right) = \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -4 & -1 \\ 0 & 1 & 0 & -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array} \right).$$

Therefore the inverse of  $A$  is the matrix

$$A^{-1} = \begin{bmatrix} -1 & -4 & -1 \\ -\frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

## 5.5 Determinants

The *determinant* of a square matrix is defined inductively by

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} &= a_{11} \det \begin{bmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &\quad - a_{12} \det \begin{bmatrix} a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} + \cdots \\ &\quad + (-1)^{n+1} a_{1n} \det \begin{bmatrix} a_{21} & a_{22} & \cdots & a_{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)} \end{bmatrix}. \end{aligned}$$

**Example 1.** Calculate the determinant of the matrix

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 4 & 1 \\ 0 & -2 & -1 \end{bmatrix}.$$

We have

$$\det A = 2 \det \begin{bmatrix} 4 & 1 \\ -2 & -1 \end{bmatrix} - 0 + \det \begin{bmatrix} 1 & 4 \\ 0 & -2 \end{bmatrix} = 2(-4 + 2) + (-2) = -6.$$

If  $A = (v_1 \dots v_n)$  is a square  $n \times n$  matrix whose columns are the vectors  $v_i$ , then  $\det A$  is an alternating multilinear form on the columns of  $A$ , *i. e.*

$$\det(v_1 \dots \alpha u + \beta v \dots v_n) = \alpha \det(v_1 \dots u \dots v_n) \quad (5.5)$$

$$+ \beta \det(v_1 \dots v \dots v_n)$$

$$\det(v_1 \dots v_i \dots v_j \dots v_n) = -\det(v_1 \dots v_j \dots v_i \dots v_n)$$

From (5.5), we can deduce the following properties of the determinant.

$$\det(v_1 \dots 0 \dots v_n) = 0 \quad (5.6a)$$

$$\det(v_1 \dots u \dots u \dots v_n) = 0 \quad (5.6b)$$

$$\det(v_1 \dots u \dots v + \alpha u \dots v_n) = \det(v_1 \dots u \dots v \dots v_n). \quad (5.6c)$$

The transpose of a matrix  $A = (a_{ij})$  is the matrix  $A^T = (a_{ji})$ , *i. e.* the matrix whose columns are the rows of  $A$ , and whose rows are the columns of  $A$ . If  $A$  is an  $m \times n$  matrix, then  $A^T$  is an  $n \times m$  matrix.

One can show inductively that

$$\det A^T = \det A. \quad (5.7)$$

From (5.7) we conclude that the alternating multilinear properties (5.5-5.6) also hold for the rows of a matrix.

**Example 2.** Calculate the determinant of the matrix

$$A = \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ -2 & 1 & -1 & 3 & 2 \\ 10 & 4 & -2 & 0 & 1 \\ 4 & 6 & -1 & 0 & 3 \end{bmatrix}.$$

We use properties (5.5-5.6) applied to the rows of  $A$  to calculate this determinant.

$$\begin{aligned} \det A &= \det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{bmatrix} = -\det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 4 & 2 & -2 & 0 \\ 0 & 4 & -42 & -30 & 11 \\ 0 & 6 & -17 & -12 & 7 \end{bmatrix} \\ &= -\det \begin{bmatrix} 1 & 0 & 4 & 3 & -1 \\ 0 & 1 & 7 & 9 & 0 \\ 0 & 0 & -26 & -38 & 0 \\ 0 & 0 & -70 & -66 & 11 \\ 0 & 0 & -59 & -66 & 7 \end{bmatrix} = -\det \begin{bmatrix} -26 & -38 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{bmatrix} \\ &= 2 \det \begin{bmatrix} 13 & 19 & 0 \\ -70 & -66 & 11 \\ -59 & -66 & 7 \end{bmatrix} = 2 \left( 13 \det \begin{bmatrix} -66 & 11 \\ -66 & 7 \end{bmatrix} \right. \\ &\quad \left. - 19 \det \begin{bmatrix} -70 & 11 \\ -59 & 7 \end{bmatrix} \right) \\ &= 2(13(-462 + 726) - 19(-490 + 649)) = 822. \end{aligned}$$

**Theorem 1.** *If  $A$  and  $B$  are square matrices of the same size, then*

$$\det(AB) = (\det A)(\det B). \quad (5.7a)$$

If the matrix  $A$  has an inverse  $A^{-1}$ , then the equation (5.7a) implies

$$(\det A)(\det A^{-1}) = 1. \quad (5.8)$$

Thus, from equation (5.8), we can conclude that if  $A$  has an inverse, then  $\det A \neq 0$ . The converse is true and is contained in the following theorem.

**Theorem 2.** Let  $A$  be an  $n \times n$  matrix. Then the following are equivalent

1. The system  $Ax = b$  has a unique solution for each  $n$ -vector  $b$ .
2. The matrix  $A$  is invertible.
3.  $\det A \neq 0$ .

Note that if  $\det A = 0$ , then the system  $Ax = 0$  has nonzero solutions  $x$ . We use this fact in the following section.

## 5.6 Eigenvalues and Eigenvectors

Let  $A$  be a square matrix. We say that  $v \neq 0$  is an *eigenvector* of  $A$  if

$$Av = \lambda v \tag{5.9}$$

for some scalar<sup>1</sup>  $\lambda \in \mathbb{C}$ . The scalar  $\lambda$  is called the *eigenvalue* of  $A$  with respect to  $v$ .<sup>2</sup>

**Example 1.** Consider

$$A = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Then  $v$  is an eigenvector of  $A$  with respect to the eigenvalue  $\lambda = 2$ :

$$Av = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

We now describe an algorithm to find the eigenvalues of a matrix. First, note that if  $v$  is an eigenvector of  $A$  with respect to the eigenvalue  $\lambda$ , then  $\lambda$  and  $v$  are solutions to the equation (5.9) with  $v \neq 0$ . Equation (5.9) can be written in the form

$$(A - \lambda I)v = 0, \tag{5.10}$$

where  $I$  is the identity matrix of the same size as  $A$ . By theorem 2, equation (5.10) has a nonzero solution in  $v$  if and only if

$$\det(A - \lambda I) = 0. \tag{5.11}$$

<sup>1</sup>We now also consider complex numbers as scalars.

<sup>2</sup>We also say that  $v$  is an eigenvector of  $A$  with respect to  $\lambda$ . Note that for each eigenvector of a matrix  $A$  corresponds a unique eigenvalue; however, several eigenvectors may correspond to a single eigenvalue.

Equation (5.11) is called the *characteristic equation* of the matrix  $A$ . Observe that if  $A$  is an  $n \times n$  matrix, then the characteristic equation (5.11) is a polynomial equation in  $\lambda$  of degree  $n$ .

**Example 2.** Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 5 & 8 & -3 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

The characteristic equation of  $A$  is given by

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 5 - \lambda & 8 & -3 \\ 0 & -3 - \lambda & 0 \\ 0 & 0 & 2 - \lambda \end{bmatrix} = (5 - \lambda)(-3 - \lambda)(2 - \lambda).$$

Thus the eigenvalues of  $A$  are  $\lambda_1 = 5$ ,  $\lambda_2 = -3$ , and  $\lambda_3 = 2$ . To calculate the eigenvectors, we solve the equation (5.10) for each  $\lambda_i$ . Using the Gauss-Jordan method we obtain that

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

are the eigenvectors of  $A$  with respect to  $\lambda_1 = 5$ ,  $\lambda_2 = -3$ , and  $\lambda_3 = 2$ .

**Example 3.** Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

The characteristic equation of  $A$  is given by

$$\begin{aligned} 0 &= \det \begin{bmatrix} 1 - \lambda & 1 & 0 \\ 2 & -\lambda & 0 \\ 0 & 0 & 3 - \lambda \end{bmatrix} = (1 - \lambda) \det \begin{bmatrix} -\lambda & 0 \\ 0 & 3 - \lambda \end{bmatrix} - \det \begin{bmatrix} 2 & 0 \\ 0 & 3 - \lambda \end{bmatrix} \\ &= (1 - \lambda)(-\lambda)(3 - \lambda) - 2(3 - \lambda) = -(\lambda - 3)(\lambda + 1)(\lambda - 2). \end{aligned}$$

Thus, the eigenvalues of  $A$  are then  $\lambda_1 = 3$ ,  $\lambda_2 = -1$ , and  $\lambda_3 = 2$ . The eigenvectors of  $A$  with respect to these eigenvalues are, respectively,

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

**Example 4.** Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 3 \\ -4 & 2 \end{bmatrix}.$$

The characteristic equation of  $A$  is the equation

$$0 = \det \begin{bmatrix} 1 - \lambda & 3 \\ -4 & 2 - \lambda \end{bmatrix} = \lambda^2 - 3\lambda + 14.$$

Then the eigenvalues of  $A$  are

$$\lambda_1 = \frac{3}{2} + \frac{\sqrt{47}}{2}i \text{ and } \lambda_2 = \frac{3}{2} - \frac{\sqrt{47}}{2}i.$$

The eigenvectors of  $A$  corresponding to  $\lambda_1$  and  $\lambda_2$  are, respectively,

$$v_1 = \begin{bmatrix} 3 \\ \frac{1}{2} + \frac{\sqrt{47}}{2}i \end{bmatrix}, \quad v_2 = \begin{bmatrix} 3 \\ \frac{1}{2} - \frac{\sqrt{47}}{2}i \end{bmatrix}.$$

Let  $A$  be an  $n \times n$  matrix, and suppose that the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  are all different. If  $v_1, v_2, \dots, v_n$  are the eigenvectors of  $A$  with respect to the  $\lambda_i$ , let

$$\begin{aligned} D &= \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \\ S &= (v_1 \ v_2 \ \dots \ v_n). \end{aligned}$$

The matrix  $D$  is the diagonal matrix whose diagonal elements are the eigenvalues of  $A$ , and  $S$  is the matrix whose columns are the eigenvectors of  $A$ . Hence

$$\begin{aligned} AS &= A(v_1 \ v_2 \ \dots \ v_n) = (Av_1 \ Av_2 \ \dots \ Av_n) = (\lambda_1 v_1 \ \lambda_2 v_2 \ \dots \ \lambda_n v_n) \\ &= (v_1 \ v_2 \ \dots \ v_n) \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = SD, \end{aligned}$$

where we have used (5.3a). One can show that the matrix  $S$  has an inverse. Then, we factor  $A$  in the form

$$A = SDS^{-1}. \tag{5.12}$$

The expression (5.12) is called the *diagonalization* of  $A$ . Not every matrix has a diagonalization. If the matrix  $A$  has a diagonalization, then we say that  $A$  is *diagonalizable*. We have seen, in the case where all the eigenvalues of  $A$  are distinct, that  $A$  is diagonalizable.

**Example 5.** Diagonalize the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

The eigenvalues of  $A$  are  $\lambda_1 = 3, \lambda_2 = 2$ , and  $\lambda_3 = -1$ , and the eigenvectors with respect to these eigenvalues are

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}.$$

Let

$$S = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Then we have  $A = SDS^{-1}$ . Thus, the diagonalization of  $A$  is

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 2/3 & 1/3 & 0 \\ 1/3 & -1/3 & 0 \end{bmatrix}.$$

From (5.3) one can check that

$$(\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n))^k = \text{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k).$$

Thus, if  $A = SDS^{-1}$ ,

$$\begin{aligned} A^k &= (SDS^{-1})(SDS^{-1}) \cdots (SDS^{-1}) \quad (k \text{ factors}) \\ &= SD^k S^{-1}. \end{aligned}$$

If  $p(x) = a_k x^k + \dots + a_1 x + a_0$  is a polynomial and  $A$  is an  $n \times n$  matrix, we define

$$p(A) = a_k A^k + \dots + a_1 A + a_0 I.$$

By the above observations, we see that if  $A$  is diagonalizable, then

$$p(A) = p(SDS^{-1}) = S \cdot p(D) \cdot S^{-1} = S \cdot \text{diag}(p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)) \cdot S^{-1}.$$

**Example 6.** Calculate  $p(A)$  where  $p(x) = x^2 - 3x + 14$ , and

$$A = \begin{bmatrix} 1 & 3 \\ -4 & 2 \end{bmatrix}.$$

The characteristic equation of  $A$  is  $\lambda^2 - 3\lambda + 14 = 0$ . Therefore

$$p(A) = S \cdot p(D) \cdot S^{-1} = S \cdot \text{diag}(p(\lambda_1), p(\lambda_2)) \cdot S^{-1} = S \cdot \text{diag}(0, 0) \cdot S^{-1} = 0.$$

The above example suggests the following fact, which is indeed true: Every diagonalizable matrix satisfies its characteristic equation. In fact, one can show that every matrix satisfies its characteristic equation.

## 5.7 Some exercises

In each of Exercises 1–2, calculate the matrix products  $AB$  and  $BA$ .

1.

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

2.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

In each of Exercises 3–5, find all solutions of the system of equations.

3.

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_2 + x_3 &= -1 \\ x_1 + x_2 &= 1 \end{aligned}$$

4.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 0 \\ x_2 + x_3 &= -1 \\ x_1 + x_2 &= 1 \end{aligned}$$

5.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1 \\ x_2 + x_3 &= -1 \\ x_1 + x_2 &= 1 \end{aligned}$$

In each of Exercises 6–9, calculate the determinant of  $A$ .

6.

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix}$$

7.

$$A = \begin{bmatrix} -1 & 2 & 3 \\ 3 & 2 & 1 \\ -2 & 4 & 6 \end{bmatrix}$$

8.

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 5 & -2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

9.

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

In each of Exercises 10–13, find all eigenvalues and eigenvectors of the matrix  $A$ .

10.

$$A = \begin{bmatrix} -1 & 4 \\ 2 & 1 \end{bmatrix}$$

11.

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}$$

12.

$$A = \begin{bmatrix} 2 & -3 & 3 \\ 4 & -5 & 3 \\ 4 & -4 & 2 \end{bmatrix}$$

13.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

14. Can the matrix

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$$

be diagonalized?



Part III

**ORDINARY  
DIFFERENTIAL  
EQUATIONS**



## Chapter 6

# First Order Ordinary Differential Equations

### 6.1 Exponential Growth and Decay

The rate of change of some quantity is often proportional to the amount of the quantity present. This may be true, for example, of the size of a population with enough resources that its growth is unrestricted, and depends only on an inherent per capita reproductive rate. It can also apply to a decaying population—for example, the mass of a piece of a radioactive substance. In such a case, if  $y(t)$  is the quantity at time  $t$ , then  $y(t)$  satisfies the *differential equation*

$$\frac{dy}{dt} = ay \tag{6.1}$$

where  $a$  is a constant representing the proportional growth or decay rate with  $a$  positive if the quantity is increasing and negative if the quantity is decreasing. It is easy to verify that  $y = ce^{at}$  is a solution of the differential equation (6.1) for every choice of the constant  $c$ . By this we mean that if we substitute the function  $y = ce^{at}$  into the differential equation (6.1) it becomes an identity. If  $y = ce^{at}$ , then  $y' = ace^{at} = ay$ , and this is the necessary verification. Thus the differential equation (6.1) has an infinite family of solutions (one for every choice of the constant  $c$ , including  $c = 0$ ), namely

$$y = ce^{at}. \tag{6.2}$$

In order for a mathematical problem to be a plausible description of a scientific situation, the mathematical problem must have only one solution; if there were multiple solutions we would not know which solution represents the situation. This suggests that the differential equation (6.1) by itself is not enough to specify a description of a physical situation. We must also specify the value of the function  $y$  for some initial time when we may measure the quantity  $y$

and then allow the system to start running. For example, we might impose the additional requirement, called an *initial condition*, that

$$y(0) = y_0. \quad (6.3)$$

A problem consisting of a differential equation together with an initial condition is called an *initial value problem*. We may determine the value of  $c$  for which the solution (6.2) of the differential equation (6.1) also satisfies the initial condition (6.3) by substituting  $t = 0$ ,  $y = y_0$  into the form (6.2). This gives the equation

$$y_0 = c e^0 = c$$

and thus  $c = y_0$ . We now use this value of  $c$  to give the solution of the differential equation (6.1) which also satisfies the initial condition (6.3), namely  $y = y_0 e^{at}$ . In order to show that this is the only solution of the initial value problem (6.1), (6.3), we must show that **every** solution of the differential equation (6.1) is of the form (6.2).

To prove this, suppose that  $y(t)$  is a solution of the differential equation [eqrefde1](#), that is, that  $y'(t) = ay(t)$  for every value of  $t$ . If  $y(t) \neq 0$ , division of this equation by  $y(t)$  gives

$$\frac{y'(t)}{y(t)} = \frac{d}{dt} \ln|y(t)| = a. \quad (6.4)$$

Integration of both sides of (6.4) gives  $\ln|y(t)| = at + k$  for some constant of integration  $k$ . Then

$$|y(t)| = e^{at+k} = e^k e^{at}.$$

Because  $e^{at}$  and  $e^k$  are positive for every value of  $t$ ,  $|y(t)|$  cannot be zero, and thus  $y(t)$  cannot change sign. We may remove the absolute value and conclude that  $y(t)$  is a constant multiple of  $e^{at}$ ,  $y = ce^{at}$ . We note also that if  $y(t)$  is different from zero for one value of  $t$  then  $y(t)$  is different from zero for every value of  $t$ . Thus the division by  $y(t)$  at the beginning of the proof is legitimate unless the solution  $y(t)$  is identically zero. The identically zero function is a solution of the differential equation, as is easily verified by substitution, and it is contained in the family of solutions  $y = ce^{at}$  with  $c = 0$ .

The absolute value which appears in the integration produces some complications which may be avoided if we know that the solution must be non-negative, so that  $|y(t)| = y(t)$ , as is the case in many applications. If we know that a solution  $y(t)$  of the differential equation  $y' = ay$  is positive for all  $t$ , we could replace (6.4) by

$$\frac{d}{dt} \ln y(t) = a$$

and then integrate to obtain  $\ln y(t) = at + k$ ,  $y(t) = e^{at+k} = e^{at} e^k = ce^{at}$ .

The logical argument in the above proof is that if the differential equation has a solution, then that solution must have a certain form. However, it also derives the form and thus serves as a method of determining the solution.

We now have a family of solutions of the differential equation (6.1). In order to determine the member of this family which satisfies a given initial condition, that is, in order to determine the value of the constant  $c$ , we merely substitute the initial condition into the family of solutions. This procedure may be followed in any situation which is described by an initial value problem, including population growth models and radioactive decay.

**Example 1.** Suppose that a given population of protozoa develops according to a simple growth law with a growth rate of 0.6 per member per day, that there are no deaths, and that on day zero the population consists of two members. Find the population size after 10 days.

**Solution.** The population size satisfies the differential equation (6.1) with  $a = 0.6$ , and is therefore given by  $y(t) = ce^{0.6t}$ . Since  $y(0) = 2$ , we substitute  $t = 0$ ,  $y = 2$ , and we obtain  $2 = c$ . Thus the solution which satisfies the initial condition is  $y(t) = 2e^{0.6t}$ , and the population size after 10 days is  $y(10) = 2e^{(0.6)(10)} = 403$  (with population size rounded off to the nearest integer).  $\square$

If we know that a population grows exponentially according to an exponential growth law but do not know the rate of growth we view the solution  $y = ce^{at}$  as containing two parameters which must be determined. This requires knowledge of the population size at two different times to provide two equations which may be solved for these two parameters.

**Example 2.** Suppose that a population which follows an exponential growth law has 50 members at a starting time and 100 members at the end of 10 days. Find the population at the end of 20 days.

**Solution.** The population size at time  $t$  satisfies  $y(t) = ce^{at}$  and  $y(0) = 50$ ,  $y(10) = 100$ . Thus  $y(0) = 50 = ce^0$ ,  $y(10) = ce^{10a} = 100$ . It follows that  $c = 50$  and  $100 = 50e^{10a}$ . We obtain  $e^{10a} = 2$ ,  $a = \frac{\ln 2}{10} = 0.0693$ . Finally, we obtain

$$y(20) = 50 e^{(20)(\ln 2)/10} = 50 e^{2\ln 2} = 50 \cdot 2^2 = 200 \quad \square$$

### 6.1.1 Radioactive decay

Radioactive materials decay because a fraction of their atoms decompose into other substances. If  $y(t)$  represents the mass of a sample of a radioactive substance at time  $t$ , and a fraction  $k$  of its atoms decompose in unit time, then  $y(t+h) - y(t)$  is approximately  $-ky(t)$  and we are led to the differential equation (6.1) with  $a$  replaced by  $-k$ . If it is clear from the nature of the problem that the constant of proportionality must be negative, we will use  $-k$  for the constant of proportionality, giving a differential equation

$$y' = -ky \tag{6.5}$$

with  $k > 0$ .

**Example 3.** The radioactive element strontium 90 has a decay constant  $2.48 \times 10^{-2}$  years<sup>-1</sup>. How long will it take for a quantity of strontium 90 to decrease to half of its original mass?

**Solution.** The mass  $y(t)$  of strontium 90 at time  $t$  satisfies the differential equation (6.7) with  $k = 2.48 \times 10^{-2}$ . If we denote the mass at time  $t = 0$  by  $y_0$ , then  $y(t) = y_0 e^{-(2.48 \times 10^{-2})t}$ . The value of  $t$  for which  $y(t) = y_0/2$  is the solution of

$$\frac{y_0}{2} = y_0 e^{-(2.48 \times 10^{-2})t}.$$

If we divide both sides of this equation by  $y_0$  and then take natural logarithms, we have

$$-(2.48 \times 10^{-2})t = \ln \frac{1}{2} = -\ln 2$$

so that  $t = (\ln 2)/(2.48 \times 10^{-2}) = 27.9$  years.  $\square$

The time required for the mass of a radioactive substance to decrease to half of its starting value is called the **half-life** of the substance. The half-life  $T$  is related to the decay constant  $k$  by the equation

$$T = \frac{\ln 2}{k}$$

because if  $y(t) = y_0 e^{-kt}$  and (by definition)  $y(T) = \frac{y_0}{2}$ , then  $e^{-kT} = \frac{1}{2}$ , so that  $-kT = \ln \frac{1}{2} = -\ln 2$ . For radioactive substances it is common to give the half-life rather than the decay constant.

**Example 4.** Radium 226 is known to have a half-life of 1620 years. Find the length of time required for a sample of radium 226 to be reduced to one fourth of its original size.

**Solution.** The decay constant for radium 226 is  $k = \frac{\ln 2}{1620} = 4.28 \times 10^{-4}$  years<sup>-1</sup>. In terms of  $k$ , the mass of a sample at time  $t$  is  $y_0 e^{-kt}$  if the starting mass is  $y_0$ . The time  $\tau$  at which the mass is  $y_0/4$  is obtained by solving the equation  $y_0/4 = y_0 e^{-k\tau}$  or  $e^{-k\tau} = 1/4$ . Taking natural logarithms we obtain  $-k\tau = \ln 1/4$ , which gives

$$\tau = -\frac{\ln \frac{3}{4}}{k} = \frac{1620(\ln \frac{3}{4})}{\ln 2} = 672 \text{ years. } \square$$

The radioactive element carbon 14 decays to ordinary carbon (carbon 12) with a decay constant  $1.244 \times 10^{-4}$  years<sup>-1</sup>, and thus the half-life of carbon 14 is 5570 years. This has an important application, called carbon dating, for determining the approximate age of fossil materials. The carbon in living matter contains a small proportion of carbon 14 absorbed from the atmosphere. When a plant or animal dies, it no longer absorbs carbon 14 and the proportion of carbon 14 decreases because of radioactive decay. By comparing the proportion of carbon 14 in a fossil with the proportion assumed to have been present before death, it is possible to calculate the time since absorption of carbon 14 ceased.

**Example 5.** Living tissue contains approximately  $6 \times 10^{10}$  atoms of carbon 14 per gram of carbon. A wooden beam in an ancient Egyptian tomb from the First Dynasty contained approximately  $3.33 \times 10^{10}$  atoms of carbon 14 per gram of carbon. How old is the tomb?

**Solution.** The number of atoms of carbon 14 per gram of carbon,  $y(t)$ , is given by  $y(t) = y_0 e^{-kt}$ , with  $y_0 = 6 \times 10^{10}$ ,  $k = 1.244 \times 10^{-4}$ , and  $y(t) = 3.33 \times 10^{10}$  for this particular  $t$  value. Thus the age of the tomb is given by the solution of the equation

$$e^{-(1.244 \times 10^{-4})t} = \frac{3.33 \times 10^{10}}{6 \times 10^{10}} = \frac{3.33}{6},$$

and if we take natural logarithms this reduces to

$$t = -\frac{\ln 3.33 - \ln 6}{1.244 \times 10^{-4}} = 4733 \text{ years. } \square$$

## 6.2 Solutions and Direction Fields

By a differential equation we will mean simply a relation between an unknown function and its derivatives. We will confine ourselves to *ordinary differential equations*, which are differential equations whose unknown function is a function of one variable so that its derivatives are ordinary derivatives. A *partial differential equation* is a differential equation whose unknown function is a function of more than one variable, so that the derivatives involved are partial derivatives. The *order* of a differential equation is the order of the highest-order derivative appearing in the differential equation. In this chapter we shall consider first-order differential equations, relations involving an unknown function  $y(t)$  and its first derivative  $y'(t) = \frac{dy}{dt}$ . The general form of a first-order differential equation is

$$y' = \frac{dy}{dt} = f(t, y), \quad (6.6)$$

with  $f$  a given function of the two variables  $t$  and  $y$ .

By a solution of the differential equation (6.6) we mean a differentiable function  $y$  of  $t$  on some  $t$ -interval  $I$  such that, for every  $t$  in the interval  $I$ ,

$$y'(t) = f(t, y(t)).$$

In other words, differentiating the function  $y(t)$  results in the function  $f(t, y(t))$ . For example, we have seen in the preceding section that, whatever the value of the constant  $c$ , the function  $y = ce^{at}$  is a solution of the differential equation  $y' = ay$  on every  $t$ -interval. We see this by differentiating  $ce^{at}$  to get  $a ce^{at}$ , which we can rewrite as  $ay$ .

To verify whether a given function is a solution of a given differential equation, we need only substitute into the differential equation and check whether it then reduces to an identity.

**Example 1.** Show that the function  $y = \frac{1}{t+1}$  is a solution of the differential equation  $y' = -y^2$ .

**Solution:** For the given function,

$$\frac{dy}{dt} = -\frac{1}{(t+1)^2} = -y^2$$

and this shows that it is indeed a solution.  $\square$

In the same way we can verify that a family of functions satisfies a given differential equation. By a *family of functions* we will mean a function which includes an arbitrary constant, so that each value of the constant defines a distinct function. The family  $ce^{5t}$ , for instance, includes the functions  $e^{5t}$ ,  $-4e^{5t}$ ,  $12e^{5t}$ , and  $\sqrt{3}e^{5t}$ , among others. When we say that a family of functions satisfies a differential equation, we mean that substitution of the family (i.e., the general form) into the differential equation gives an identity satisfied for every choice of the constant.

**Example 2.** Show that for every  $c$  the function  $y = \frac{1}{t+c}$  is a solution of the differential equation  $y' = -y^2$ .

**Solution:** For the given function,

$$\frac{dy}{dt} = -\frac{1}{(t+c)^2} = -y^2,$$

and thus each member of the given family of functions is a solution.  $\square$

In applications we are usually interested in finding not a family of solutions of a differential equation but a solution which satisfies some additional requirement. In the various examples in Section 5.1 the additional requirement was that the solution should have a specified value for a specified value of the independent variable  $t$ . Such a requirement, of the form

$$y(t_0) = y_0 \tag{6.7}$$

is called an *initial condition*, and  $t_0$  is called the *initial time* while  $y_0$  is called the *initial value*. A problem consisting of a differential equation (6.1) together with an initial condition (6.2) is called an *initial value problem*. Geometrically, an initial condition picks out the solution from a family of solutions which passes through the point  $(t_0, y_0)$  in the  $t$ - $y$  plane. Physically, this corresponds to measuring the state of a system at the time  $t_0$  and using the solution of the initial value problem to predict the future behavior of the system.

**Example 4.** Find the solution of the differential equation  $y' = -y^2$  of the form  $y = \frac{1}{t+c}$  which satisfies the initial condition  $y(0) = 1$ .

**Solution:** We substitute the values  $t = 0$ ,  $y = 1$  into the equation  $y = \frac{1}{t+c}$ , and we obtain a condition on  $c$ , namely  $1 = \frac{1}{c}$ , whose solution is  $c = 1$ . The required solution is the function in the given family with  $c = 1$ , namely  $y = \frac{1}{t+1}$ .  $\square$

**Example 5.** Find the solution of the differential equation  $y' = -y^2$  which satisfies the general initial condition  $y(0) = y_0$ , where  $y_0$  is arbitrary.

**Solution:** We substitute the values  $t = 0$ ,  $y = y_0$  into the equation  $y = \frac{1}{t+c}$  and solve the resulting equation  $y_0 = \frac{1}{c}$  for  $c$ , obtaining  $c = \frac{1}{y_0}$  provided  $y_0 \neq 0$ . Thus the solution of the initial value problem is

$$y = \frac{1}{\left(t + \frac{1}{y_0}\right)^2}$$

except if  $y_0 = 0$ . If  $y_0 = 0$ , there is no solution of the initial value problem of the given form; in this case the identically zero function,  $y = 0$ , is a solution. We have now obtained a solution of the initial value problem with arbitrary initial value at  $y = 0$  for the differential equation  $y' = -y^2$ .  $\square$

A family of solutions may arise if we are considering a differential equation with no initial condition imposed, and we will then also be concerned with the question of whether the given family contains all solutions of the differential equation. To answer this question, we will need to make use of a theorem which guarantees that each initial value problem for the given differential equation has exactly one solution. More specifically, if an initial value problem is to be a usable mathematical description of a scientific problem, it must have a solution, for otherwise it would be of no use in predicting behavior. Furthermore, it should have only one solution, for otherwise we would not know which solution describes the system. Thus for applications it is vital that there be a mathematical theory telling us that an initial value problem has exactly one solution. Fortunately, there is a very general theorem which tells us that this is true for the initial value problem (6.6), (6.3) provided the function  $f$  is reasonably smooth. We will state this result and ask the reader to accept it without proof because the proof requires more advanced mathematical knowledge than we have at present.

**EXISTENCE AND UNIQUENESS THEOREM:** If the function  $f(t, y)$  is differentiable with respect to  $y$  in some region of the plane which contains the point  $(t_0, y_0)$ , then the initial value problem consisting of the differential equation  $y' = f(t, y)$  and the initial condition  $y(t_0) = y_0$  has a unique solution which is defined on some  $t$ -interval containing  $t_0$  in its interior.

Even though the function  $f(t, y)$  may be well-behaved in the whole  $t$ - $y$  plane, there is no assurance that a solution will be defined for all  $t$ . As we have seen in Example 1, the solution  $y = \frac{1}{t+1}$  of  $y' = -y^2$ ,  $y(0) = 1$  exists only for  $-1 < t < \infty$ . As we have seen in Example 2, each solution of the family of solutions  $y = \frac{1}{t+c}$  has a different interval of existence. In Example 5 we have shown how to rewrite a family of solutions for a differential equation in terms of an arbitrary initial condition — that is, as a solution of an initial value problem for that differential equation. We have also seen how to identify those initial conditions which cannot be satisfied by a member of the given family. Often there are constant functions which are not members of the given family but

which are solutions and satisfy initial conditions that cannot be satisfied by a member of the family. The existence and uniqueness theorem tells us that if we can find a family of solutions, possibly supplemented by some additional solutions, so that we can find this collection contains a solution corresponding to each possible initial condition, then we have found the set of all solutions of the differential equation.

A differential equation which arises in various applications, including models for population growth and spread of rumors, is the *logistic differential equation*,

$$y' = ry\left(1 - \frac{y}{K}\right) \quad (6.8)$$

containing two parameters  $r$  and  $K$ . The basic idea behind this form is that instead of a constant per capita growth rate  $r$  as in the exponential growth equation it is more realistic to assume that the per capita growth rate decreases as the population size increases. The form  $1 - \frac{y}{K}$  used in the logistic equation is the simplest form for a decreasing per capita growth rate. We may verify that for every constant  $c$  the function

$$y = \frac{K}{1 + ce^{-rt}} \quad (6.9)$$

is a solution of this differential equation. To see this, note that for the given function  $y$ ,

$$y' = \frac{Kcr e^{-rt}}{(1 + ce^{-rt})^2}$$

and

$$1 - \frac{y}{K} = \frac{K - y}{K} = \frac{ce^{-rt}}{(1 + ce^{-rt})}.$$

Thus

$$ry\left(1 - \frac{y}{K}\right) = \frac{Krc e^{-rt}}{(1 + ce^{-rt})^2} = y',$$

and the given function satisfies the logistic differential equation for every choice of  $c$ .

To find the solution which obeys the initial condition  $y(0) = y_0$ , we substitute  $t = 0$ ,  $y = y_0$  into the form (6.9), obtaining  $\frac{K}{1+c} = y_0$  which implies  $c = \frac{K-y_0}{y_0}$  as long as  $y_0 \neq 0$  and gives the solution

$$y = \frac{K}{1 + \left(\frac{K-y_0}{y_0}\right)e^{-rt}} = \frac{Ky_0}{y_0 + (K - y_0)e^{-rt}} \quad (6.10)$$

to the initial value problem with  $y_0 \neq 0$ . Note that the denominator begins at  $K$  (for  $t = 0$ ) and moves toward  $y_0$  as  $t \rightarrow \infty$ . Now suppose that  $K$  represents some physical quantity such that  $K > 0$ . It is easy to see from the form (6.10) that if  $y_0 > 0$ , then the solution  $y(t)$  exists for all  $t > 0$ , and  $\lim_{t \rightarrow \infty} y(t) = K$ . If  $y_0 < 0$ , then this solution does not exist for all  $t > 0$ , because  $y(t) \rightarrow -\infty$  where

the denominator changes sign: as  $y_0 + (K - y_0)e^{-rt} \rightarrow 0$ , or  $t \rightarrow -\log(\frac{-y_0}{K-y_0})$ . If  $y_0 = 0$ , the solution of the initial value problem is not given by (6.10), but is the identically zero function  $y = 0$ . We observe that the family of solutions (6.9) of the logistic differential equation (6.8) includes the constant solution  $y = K$  (with  $c = 0$ ) but not the constant solution  $y = 0$ . The existence and uniqueness theorem shows that, since we have now obtained a solution corresponding to each possible initial condition, we have obtained all solutions of the logistic differential equation.

The geometric interpretation of a solution  $y(t)$  to a differential equation (6.6) is that the curve  $y = y(t)$  has slope  $f(t, y)$  at each point  $(t, y)$  along its length. Thus we might think of approximating the solution curve by piecing together short line segments whose slope at each point  $(t, y)$  is  $f(t, y)$ . To realize this idea, we construct at each point  $(t, y)$  in some region of the plane a short line segment with slope  $f(t, y)$ . The collection of line segments is called the *direction field* of the differential equation (6.6). The direction field can help us to visualize solutions of the differential equation since at each point on its graph a solution curve is tangent to the line segment at that point. We may sketch the solutions of a differential equation by connecting these line segments by smooth curves.

Drawing direction fields by hand is a difficult and time-consuming task. There are computer programs, both self-contained and portions of more elaborate computational systems such as Maple, Matlab, and Mathematica, which can generate direction fields for a differential equation and can also sketch solution curves corresponding to these direction fields. We give some examples here which have been produced by Maple; the reader with access to a facility which is capable of drawing direction fields is urged to reproduce these examples before trying to produce other direction fields.

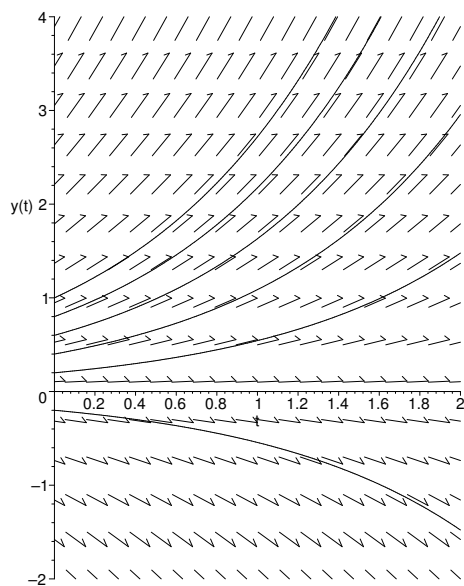
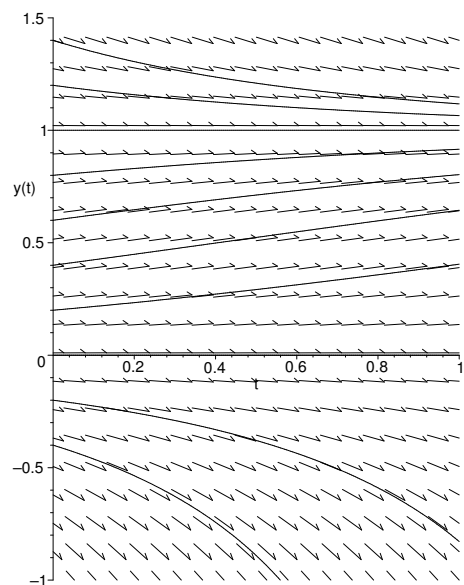
A direction field and some solutions of the differential equation  $y' = y$  are shown in Figure 6.1. The direction field suggests exponential solutions, which we know from Section 6.1 to be correct.

A direction field and some solutions of the differential equation  $y' = y(1 - y)$  are shown in Figure 6.2. The direction field indicates that solutions below the  $t$ -axis are unbounded below while solutions above the  $t$ -axis tend to 1 as  $t \rightarrow \infty$ , consistent with what we have established for this logistic differential equation.

The geometric view of differential equations presented by the direction field will appear again when we examine some qualitative properties in Section 6.6.

## 6.3 Equations with Variables Separable

In this section we shall learn a method for finding solutions of a class of differential equations. The method is based on the method we used in Section 6.1 for solving the differential equation of exponential growth or decay, and is applicable to differential equations with *variables separable*. A differential equation  $y' = f(t, y)$  is called *separable* or is said to have *variables separable* if the function

Figure 6.1: Direction field and solutions for  $y' = y$ Figure 6.2: Direction field and solutions for  $y' = y(1 - y)$ 

$f$  can be expressed in the form

$$f(t, y) = p(t)q(y). \quad (6.11)$$

In particular, if  $f$  is independent of  $t$  and is a function of  $y$  only, the differential equation (6.6) is separable; such an equation is said to be *autonomous*. The reason for the name separable is that the differential equation can then be written as

$$\frac{y'}{q(y)} = p(t)$$

with all the dependence on  $y$  on the left and all the dependence on  $t$  on the right hand side of the equation, provided that  $q(y) \neq 0$ . For example,  $y' = \frac{y}{1+t^2}$  and  $y' = y^2$  are both separable, whereas  $y' = \sin t - 2ty$  is not separable. The examples discussed in Section 6.1 are also separable and the method of solution described for equation (6.1) of Section 6.1 is a special case of the general method of solution to be developed for separable equations. Before explaining this general technique, let us work out some more examples.

**Example 1.** Find all solutions of the differential equation

$$y' = -y^2.$$

**Solution:** We divide the equation by  $y^2$ , permissible if  $y \neq 0$ , to give

$$\frac{1}{y^2} \frac{dy}{dt} = -1.$$

We then integrate both sides of this equation with respect to  $t$ . In order to integrate the left side of the equation we must make the substitution  $y = y(t)$ , where  $y(t)$  is the (as yet unknown) solution. This substitution gives

$$\int \frac{1}{y^2} \frac{dy}{dt} dt = \int \frac{dy}{y^2} = -\frac{1}{y} + c$$

and we obtain

$$-\frac{1}{y} = -t - c \tag{6.12}$$

or  $y = \frac{1}{t+c}$ , with  $c$  a constant of integration. Observe that no matter what value of  $c$  is chosen this solution is never equal to zero. We began by dividing the equation by  $y^2$ , which was legitimate provided  $y \neq 0$ . If  $y = 0$ , we cannot divide, but the constant function  $y = 0$  is a solution. We now have all the solutions of the differential equation, namely the family (6.12) together with the zero solution.  $\square$

**Example 2.** Solve the initial value problem

$$y' = -y^2, \quad y(0) = 1.$$

**Solution:** We begin by using a more colloquial form of separation of variables than the procedure of Example 1. We write the differential equation in the form  $\frac{dy}{dt} = -y^2$  and separate variables by dividing the equation through by  $y^2$  to give

$$-\frac{dy}{y^2} = dt.$$

This form of the equation is meaningless without an interpretation of differentials, but the integrated form

$$-\int \frac{dy}{y^2} = \int dt$$

is meaningful. Carrying out the integration, we obtain  $\frac{1}{y} = t + c$ , where  $c$  is a constant of integration. Since  $y = 1$  for  $t = 0$  we may substitute these values into the solution and obtain  $1 = c$ . Substituting this value of  $c$  we now obtain

$$\frac{1}{y} = t + 1.$$

Solving for  $y$ , we obtain the solution  $y = \frac{1}{t+1}$ , and we may easily verify that this is a solution of the initial value problem, defined for all  $t \geq 0$ .  $\square$

The version of separation of variables used in Example 2 is easier to apply in practice than the more precise version used in Example 1. When we treat the general case we will justify this approach.

**Example 3.** Solve the initial value problem

$$y' = -y^2, \quad y(0) = 0.$$

**Solution:** The procedure used in Example 1 leads to the family of solutions  $y = \frac{1}{t+c}$  for the differential equation  $y' = -y^2$ . When we substitute  $t = 0$ ,  $y = 0$  and attempt to solve for the constant  $c$ , we find that there is no solution. When we divided the differential equation by  $y^2$  we had to assume  $y \neq 0$ , but the constant function  $y \equiv 0$  is also a solution of the differential equation, as may easily be verified. Since this function also satisfies the initial condition, it is the solution of the given initial value problem.  $\square$

Let us now apply the technique of Example 2 to show how to obtain the solution in the general case of a differential equation with variables separable. We will make use of the idea of the definition of the indefinite integral of a function as a new function whose derivative is the derivative of the give function. We first solve the differential equation

$$y' = p(t)q(y) \tag{6.13}$$

where  $p$  is continuous on some interval  $a < t < b$  and  $q$  is continuous on some interval  $c < y < d$ . We solve the differential equation (6.13) by separating variables as in Example 1 dividing by  $q(y)$  and integrating to give

$$\int \frac{dy}{q(y)} = \int p(t)dt + c \tag{6.14}$$

where  $c$  is a constant of integration. We now define

$$Q(y) = \int \frac{dy}{q(y)}, \quad P(t) = \int p(t)dt,$$

by which we mean that  $Q(y)$  is a function of  $y$  whose derivative with respect to  $y$  is  $\frac{1}{q(y)}$  and  $P(t)$  is a function of  $t$  whose derivative with respect to  $t$  is  $p(t)$ . Then (6.14) becomes

$$Q(y) = P(t) + c \quad (6.15)$$

and this equation defines a family of solutions of the differential equation (6.13). To verify that each function  $y(t)$  defined implicitly by (6.15) is indeed a solution of the differential equation (6.13), we differentiate (6.15) implicitly with respect to  $t$ , obtaining

$$\frac{1}{q(y(t))} \frac{dy}{dt} = p(t)$$

on any interval on which  $q(y(t)) \neq 0$ , and this shows that  $y(t)$  satisfies the differential equation (6.13). There may be constant solutions of (6.13) which are not included in the family (6.15). A constant solution to the original equation (6.13) corresponds to a solution of the equation  $q(y) = 0$ .

In order to find the one solution of the differential equation (6.13) which satisfies the initial condition  $y(t_0) = y_0$ , it will help to be more specific in the choice of the indefinite integrals  $Q(y)$  and  $P(t)$ . We let  $Q(y)$  be the indefinite integral of  $\frac{1}{q(y)}$  such that  $Q(y_0) = 0$ , and we let  $P(t)$  be the indefinite integral of  $p(t)$  such that  $P(t_0) = 0$ . Then, because of the fundamental theorem of calculus, we have

$$Q(y) = \int_{y_0}^y \frac{du}{q(u)}, \quad P(t) = \int_{t_0}^t p(s) ds$$

and we may write the solution (6.15) in the form

$$\int_{y_0}^y \frac{du}{q(u)} = \int_{t_0}^t p(s) ds.$$

Now substitution of the initial conditions  $t = t_0$ ,  $y = y_0$  gives  $c = 0$ . Thus the solution of the initial value problem is given implicitly by

$$\int_{y_0}^y \frac{du}{q(u)} = \int_{t_0}^t p(s) ds \quad (6.16)$$

We have now solved the initial value problem in the case  $q(y_0) \neq 0$ . Since  $y(t_0) = y_0$  and the function  $q$  is continuous at  $y_0$ ,  $q(y(t))$  is continuous, and therefore  $q(y(t)) \neq 0$  on some interval containing  $t_0$  (possibly smaller than the original interval  $a < t < b$ ). On this interval,  $y(t)$  is the unique solution of the initial value problem. If  $q(y_0) = 0$  we have the constant function  $y = y_0$  as a solution of the initial value problem in place of the solution given by (6.16).

We shall see in Section 7.5 that the constant solutions of a separable differentiable equation (6.13) play an important role in describing the behavior of all solutions as  $t \rightarrow \infty$ .

The differential equation

$$y' = -ay + b \quad (6.17)$$

where  $a$  and  $b$  are given constants, appears in some of the examples in Section 7.4. In order to solve it, we separate variables, obtaining

$$\int \frac{dy}{-ay + b} = \int dt.$$

Integration gives

$$-\frac{1}{a} \ln(-ay + b) = t + c$$

with  $c$  an arbitrary constant of integration. Algebraic solution now gives

$$\begin{aligned} \ln(-ay + b) &= -a(t + c) \\ -ay + b &= e^{-a(t+c)} = e^{-ac} e^{-at} \\ ay &= b - e^{-ac} e^{-at} \\ y &= \frac{b}{a} - \frac{e^{-ac}}{a} e^{-at} \end{aligned}$$

We now rename the arbitrary constant  $-\frac{e^{-ac}}{a}$  as a new arbitrary constant, which we again call  $c$ , and obtain the family of solutions

$$y = \frac{b}{a} + ce^{-at} \quad (6.18)$$

In our separation of variables we overlooked the constant solution  $y = \frac{b}{a}$ , but this solution is contained in the family (6.18).

In order to find the solution of (6.17) which satisfies the initial condition

$$y(0) = y_0 \quad (6.19)$$

we substitute  $t = 0$ ,  $y = y_0$  into (6.18), obtaining

$$y_0 = \frac{b}{a} + c$$

or  $c = y_0 - \frac{b}{a}$ . This value of  $c$  gives the solution

$$y = \frac{b}{a} + (y_0 - \frac{b}{a})e^{-at} = \frac{b}{a}(1 - e^{-at}) + y_0 e^{-at} \quad (6.20)$$

of the initial value problem.

The solutions (6.18) and (6.20) have been obtained without regard for the sign of the coefficients  $b$  and  $a$ . We will think of  $a$  in (6.18) and (6.20) as positive, and will rewrite the solutions for  $a$  negative by making the replacement  $r = -a$ , thinking of  $r$  as positive. The result of this is that the solutions of the differential equation

$$y' = ry + b \quad (6.21)$$

are given by the family of functions

$$y = -\frac{b}{r} + ce^{rt} \quad (6.22)$$

and the solution of the differential equation (6.21) which satisfies the initial condition (6.19) is

$$y = -\frac{b}{r}(1 - e^{rt}) + y_0 e^{rt} = \frac{b}{r}(e^{rt} - 1) + y_0 e^{rt}. \quad (6.23)$$

In applications, the specific form of the solution of a differential equation is often less important than the behavior of the solution for large values of  $t$ . Because  $e^{-at} \rightarrow 0$  as  $t \rightarrow \infty$  when  $a > 0$ , we see from (6.18) that every solution of (6.17) tends to the limit  $\frac{b}{a}$  as  $t \rightarrow \infty$  if  $a > 0$ . This is an example of *qualitative* information about the behavior of solutions which will be useful in applications. In Section 7.5 we shall examine other qualitative questions—information about the behavior of solutions of a differential equation which may be obtained indirectly rather than by explicit solution.

To conclude this section, we return to the logistic differential equation (6.8) whose solutions were described in Section 7.2. In Section 7.2, we verified that the solutions were given by equation (6.9) but did not show how to obtain these solutions. The differential equation (6.8) is separable, and separation of variables and integration gives

$$\int \frac{K dy}{y(K - y)} = \int r dt$$

provided  $y \neq 0$ ,  $y \neq K$ . In order to evaluate the integral on the left hand side, we use the algebraic relation (which may be obtained by partial fractions)

$$\frac{K}{y(K - y)} = \frac{1}{y} + \frac{1}{K - y}$$

and then rewrite

$$\int \frac{K dy}{y(K - y)} = \int \frac{dy}{y} + \int \frac{dy}{K - y} = \ln|y| - \ln|K - y| = rt + c,$$

so

$$\ln \left| \frac{y}{K - y} \right| = rt + c.$$

We can now exponentiate both sides of the equation to obtain

$$\left| \frac{y}{K - y} \right| = e^{rt+c} = e^{rt} e^c.$$

If we remove the absolute value bars from the left-hand side, we can define a new constant  $C$  on the right-hand side, equal to  $e^{-c}$  if  $0 < y < K$ , and equal to  $-e^{-c}$  otherwise, thus rewriting the right-hand side as  $\frac{1}{C}e^{rt}$ . Finally, we solve for  $y$ , and obtain the family of solutions

$$y = \frac{K}{1 + Ce^{-rt}} \quad (6.24)$$

as in Section 7.2. This family contains all solutions of (6.8) except for the constant solution  $y = 0$ . A qualitative observation is that if  $r > 0$  every positive solution approaches the limit  $K$  as  $t \rightarrow \infty$ . (The only other non-negative solution is the constant solution  $y \equiv 0$ .)

## 6.4 Some Applications of Separable Equations

In this section we describe problems from various topics which lead to separable differential equations as models. While we wish to point out the breadth and variety of examples, our main goal is to encourage some understanding of the modelling process in areas of interest to the reader.

### 6.4.1 The spread of infectious diseases

Consider a population of constant size  $K$  in which an infectious disease is introduced. We divide the population into two classes: susceptibles (not infected) and infectives (infected and contagious). Let  $S(t)$  denote the number of susceptibles and let  $I(t)$  denote the number of infectives, so that  $S(t) + I(t) = K$ . We assume that the disease is spread from infectives to susceptibles through contact. Suppose that an “average” infective makes potentially infective contacts with a constant number  $\beta K$  of individuals in unit time. Then, presumably  $\beta I(t)$  of these individuals are already infected, and the number of new infections caused by an “average” infective in unit time is  $\beta S(t)$ . Thus the total number of new infections in unit time is  $\beta S(t)I(t)$ . We assume also that on recovery infectives have no immunity against re-infection and return to the susceptible class. Finally, we assume that in unit time a fraction  $\gamma$  of the infectives recover, so that the number of recoveries in unit time is  $\gamma I(t)$ . This assumption is equivalent to the assumption that for each  $s \geq 0$  the fraction of the infectives who remain infective for a time interval  $s$  is  $e^{-\gamma s}$ , and that the average length of the infective period is  $1/\gamma$ .

We may formulate a model to describe  $S(t)$  and  $I(t)$  by thinking of a flow rate  $\beta SI$  from the susceptible class to the infective class and a flow rate  $\gamma I$  from the infective class to the susceptible class. As noted with some previous population models,  $S$  and  $I$  should, properly speaking, take on only integer values, to count individuals. However, we will allow them to be real-valued in this model, and consider the model either as an approximation valid for large populations  $K$ , or (if we set  $K = 1$ ) as a representation of *proportions* of the population susceptible and infective. (Models for small populations are usually written as discrete processes, and stochasticity also plays an important role on this scale.)

We now have a pair of differential equations

$$\begin{aligned} S' &= -\beta SI + \gamma I \\ I' &= \beta SI - \gamma I, \end{aligned} \tag{6.25}$$

a model first described by W. O. Kermack and A. G. McKendrick [W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics, Part II, *Proc. Royal Soc. London* 138 (1932), 55–83]. Since  $S(t) + I(t) = K$  for all  $t$ , we may replace  $S$  by  $K - I$  and describe the situation by a single differential equation

$$I' = \beta I(K - I) - \gamma I = (\beta K - \gamma)I - \beta I^2 \tag{6.26}$$

The differential equation (6.26) has the same form as the logistic equation (6.8), with  $r$  replaced by  $\beta K - \gamma$  and  $K$  replaced by  $\left(\frac{\beta K - \gamma}{\beta}\right) = K - \frac{\gamma}{\beta}$ . There is an important difference between (6.26) and (6.8), namely that in our discussion of the solutions of (6.8) we used  $r > 0$ , and in (6.26) it is possible for  $\beta K - \gamma$  to be either positive or negative. If we examine the solution (6.9) of (6.8), we may easily see that if  $r > 0$ , then  $\lim_{t \rightarrow \infty} y(t) = K$  for every solution  $y(t)$  with  $y(0) > 0$ , while if  $r < 0$ , then  $\lim_{t \rightarrow \infty} y(t) = 0$  for every solution  $y(t)$  with  $y(0) > 0$ . If we translate this result to (6.26), we see that if  $\beta K/\gamma < 1$ , then  $\lim_{t \rightarrow \infty} I(t) = 0$ , while if  $\beta K/\gamma > 1$ , then  $\lim_{t \rightarrow \infty} I(t) = K - \frac{\gamma}{\beta} > 0$ .

This result is the famous threshold theorem of Kermack and McKendrick: If the quantity  $\beta K/\gamma$ , called the *basic reproduction number*, is less than 1, then the number of infectives approaches zero, and the number of susceptibles approaches  $K$ . In epidemiological terms this means that the infection dies out. On the other hand, if the basic reproduction number  $\beta K/\gamma$  exceeds 1, then the number of infectives remains positive and tends to a positive limit. In epidemiological terms, this means that the infection remains *endemic*. The quantity  $\beta K/\gamma$  represents the number of secondary infections caused by each infective over the duration of the infection. Since  $\beta K$  is a number of contacts per infective in unit time, the dimensions of  $\beta K$  are  $\text{time}^{-1}$ . Thus the basic reproduction number  $\beta K/\gamma$  is dimensionless, and does not depend on the units used in describing the model.

Another infectious disease model describes an epidemic in which infectives either recover with immunity against reinfection or die of the disease. This differs from the model (6.25) in that the rate  $\gamma I$  of departure from the infective class is now a rate of entry into a third class  $R$  of removed members. This gives the system

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \gamma I \\ R' &= \gamma I, \end{aligned} \tag{6.27}$$

another model first described by W. O. Kermack and A. G. McKendrick [W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics, *Proc. Royal Soc. London* 115 (1927), 700–721]. We may discard the third equation of this system since  $R$  is determined once  $S$  and  $I$  are known. We separate variables in the first equation of (6.27) and rewrite it as

$$\frac{S'}{S} = \beta I,$$

which we may integrate to give

$$\ln S(0) - \ln S(t) = \beta \int_0^t I(s) ds.$$

This equation does not give a solution of (6.27) because  $I$  is not known, but it is a valid relation that yields useful information about solutions when combined

with the result

$$S(0) - S(t) - I(t) = \beta \int_0^t I(s) ds$$

of integrating the equation

$$(S + I)' = -\gamma I$$

obtained by adding the first two equations of (6.27) (see Section 7.5.2).

### 6.4.2 Drug dosage

When a dosage of a drug is administered to a patient, the concentration of the drug in the blood increases. Over time, the concentration decreases as the drug is eliminated from the body. The question we would like to model is how the concentration of drug in the blood changes with repeated doses of the drug.

Experimental evidence indicates that the rate of elimination is proportional to the concentration of drug in the bloodstream. Thus we assume that, if  $C(t)$  is a function representing the concentration of drug in the blood at time  $t$ , its derivative is given by

$$C'(t) = -qC(t).$$

Here  $q$  is a positive constant, called the elimination constant of the drug. We assume also that when a drug is administered, it is absorbed completely immediately. This is certainly at best an approximation to the truth, probably a better approximation for a drug injected directly into the bloodstream than for a drug taken by mouth. Let us assume that the drug is administered regularly at fixed time intervals of length  $T$  with a dose capable of raising the concentration in the blood by  $A$ . Then if we begin with an initial dose  $A$  at time  $t = 0$ ,  $C(t)$  satisfies the initial value problem  $C'(t) = -qC(t)$ ,  $C(0) = A$  until the second dose. Thus for  $0 \leq t \leq T$ , we have  $C(t) = Ae^{-qt}$ , and just before the second dose at time  $T$ ,  $C(T^-) = Ae^{-qT}$ .

We let  $C_i$  be the concentration at the beginning of the  $i$ th interval and  $R_i$  the residual concentration at the end of the  $i$ th interval; we have shown that  $C_1 = A$ ,  $R_1 = Ae^{-qT}$ . After the second dose,  $C_2 = A + Ae^{-qT} = A(1 + e^{-qT})$ . The residual concentration at the end of each interval is the concentration at the beginning of the interval multiplied by  $e^{-qT}$  and the dose at the beginning of the next interval is this residual concentration plus  $A$ . Thus we obtain

$$C_2 = A(1 + e^{-qT}), \quad R_2 = A(1 + e^{-qT})e^{-qT} = A(e^{-qT} + e^{-2qT}).$$

Next we see that

$$C_3 = A + A(e^{-qT} + e^{-2qT}) = A(1 + e^{-qT} + e^{-2qT}), \quad R_3 = A(e^{-qT} + e^{-2qT} + e^{-3qT}).$$

We may give a simpler expression for each of these by using the formula for the sum of a geometric series,

$$1 + r^2 + r^3 + \dots + r^n = \frac{1 - r^{n+1}}{1 - r}$$

to obtain

$$C_3 = A \frac{1 - e^{-3qT}}{1 - e^{-qT}}, \quad R_3 = Ae^{-qT} \frac{1 - e^{-3qT}}{1 - e^{-qT}}.$$

From this we may conjecture (and prove by induction) that for every positive integer  $n$ ,

$$C_n = A \frac{1 - e^{-nqT}}{1 - e^{-qT}}, \quad R_n = Ae^{-qT} \frac{1 - e^{-nqT}}{1 - e^{-qT}}.$$

As  $n \rightarrow \infty$ ,  $e^{-nqT} \rightarrow 0$  because  $e^{-qT} < 1$ . Thus  $C_n$  and  $R_n$  approach limits  $C$  and  $R$  respectively as  $n \rightarrow \infty$ , and

$$C = \frac{A}{1 - e^{-qT}}, \quad R = \frac{Ae^{-qT}}{1 - e^{-qT}}.$$

It is easy to see that both the initial and residual concentration increase with each dose but never exceed the limit values. For example, if a drug with an elimination constant of  $0.1 \text{ hours}^{-1}$  is administered every 8 hours in a dosage of 1 mg/L, the limiting residual concentration is  $\frac{e^{-0.8}}{1 - e^{-0.8}} = 0.816 \text{ mg/L}$ .

If the time interval between doses is short,  $e^{-qT}$  is close to 1, and the ratio  $\frac{R}{A}$  of residual concentration to dose is

$$\frac{e^{-qT}}{1 - e^{-qT}} = \frac{1}{e^{qT} - 1}$$

which is large. On the other hand, if the time interval between doses is long, each  $R_n$  is close to zero and each  $C_n$  is close to  $A$ ; we have a concentration of  $A$  gradually dissipating to almost zero and then a repetition of the same process.

### 6.4.3 Allometry

Let  $x(t)$  and  $y(t)$  be the sizes of two different organs or parts of an individual at time  $t$ . There is considerable empirical evidence to suggest that the relative growth rates  $\frac{1}{x} \frac{dx}{dt}$  and  $\frac{1}{y} \frac{dy}{dt}$  are proportional. This means that there is a constant  $k$ , depending on the nature of the organs, such that

$$\frac{1}{y} \frac{dy}{dt} = k \frac{1}{x} \frac{dx}{dt} \quad (6.28)$$

The relation (6.28) is called the *allometric law*. The single equation (6.28) does not provide enough information to determine  $x$  and  $y$  as functions of  $t$ , but we can eliminate  $t$  and obtain a relation between  $x$  and  $y$ . If we consider  $y$  as a function of  $x$ , then according to the chain rule of calculus,

$$\frac{dy}{dx} = \frac{dy}{dt} \bigg/ \frac{dx}{dt} \quad (6.29)$$

Combining (6.28) and (6.29) we have

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

which is easily solved by separation of variables to give  $\log y = k \ln x + a$ , and finally

$$y = e^{k \ln x} e^a = x^k e^a = c x^k. \quad (6.30)$$

In order to determine the constant  $c$ , we need the values of  $x$  and  $y$  at some starting time.

In experiments, one might measure both  $x$  and  $y$  at various times and then use these measurements to plot  $\ln y$  against  $\ln x$ , that is, to plot the experimental data on logarithmically scaled graph paper. According to the relation (6.30), the graph should be a straight line with slope  $k$  and  $y$ -intercept  $a = \ln c$ . Because of experimental error, the points may not line up perfectly, but it should be possible to draw a line fitting the data reasonably well. One can then measure the slope and  $y$ -intercept of the line to determine the constants  $k$  and  $c$  in the relation (6.30). A more accurate procedure would be to use the method of least squares to estimate the slope and  $y$ -intercept of the line.

## 6.5 Qualitative Properties of Differential Equations

We have seen some instances in which all solutions of a differential equation, or at least all solutions with initial values in some interval, tend to the same limit as  $t \rightarrow \infty$ . Two examples are the (separable) linear differential equation with constant coefficients

$$y' = -ay + b \quad (6.31)$$

for which every solution approaches the limit  $\frac{b}{a}$  as  $t \rightarrow \infty$  provided  $a > 0$  (Section 7.3), and the logistic differential equation (6.8) for which every solution with positive initial value approaches the limit  $K$  as  $t \rightarrow \infty$  (Section 7.3) if  $r > 0$ . In applications we are often particularly interested in the long-term behavior of solutions, especially since many of the models we develop will be complex enough to make finding an explicit solution impractical. In this section, we describe some information of this nature which may be obtained indirectly, without explicit solution of a differential equation. Properties of solutions obtained without actually finding an expression for the solutions are called *qualitative* properties.

We shall study only differential equations which do not depend explicitly on the independent variable  $t$ , of the general form

$$y' = g(y). \quad (6.32)$$

Such differential equations are called autonomous. We will always assume that the function  $g(y)$  is sufficiently smooth that the existence and uniqueness theorem of Section 7.2 is valid, and there is a unique solution of the differential equation (6.32) through each initial point. An autonomous differential equation is always separable, but if the integral  $\int \frac{dy}{g(y)}$  is difficult or impossible to evaluate, solution by separation of variables is impractical. In Section 7.3, when we established the method of separation of variables, we pointed out the possibility

of constant solutions which do not come from separation of variables. It will turn out that examination of these constant solutions is of central importance in the qualitative analysis.

In order to analyze the behavior as  $t \rightarrow \infty$  of the solutions of the differential equation (6.32), and how this is determined from the nature of the constant solutions, we will need to use some properties of autonomous differential equations. We will describe these properties and illustrate them by referring to the logistic equation (6.8), and then we will combine them to give a general procedure for describing the behavior as  $t \rightarrow \infty$  of all solutions of an autonomous differential equation by examining the constant solutions. For simplicity, we shall consider only non-negative values of  $t$ , and we will think of solutions as determined by their initial values for  $t = 0$ .

**Property 1:** If  $\hat{y}$  is a solution of the equation  $g(y) = 0$ , then the constant function  $y = \hat{y}$  is a solution of the differential equation  $y' = g(y)$ . Conversely, if  $y = \hat{y}$  is a constant solution of the differential equation  $y' = g(y)$  then  $\hat{y}$  is a solution of the equation  $g(y) = 0$ .

To establish this property, which we have already used in Section 7.3, we need only observe that because the derivative of a constant function is the zero function, a constant function  $y = \hat{y}$  is a solution of (6.32) if and only if  $g(\hat{y}) = 0$ . A solution  $\hat{y}$  of  $g(y) = 0$  is called an *equilibrium* or *critical point* of the differential equation (6.32), and the corresponding constant solution of (6.32) is called an *equilibrium solution*.

**Example 1.** Find the equilibrium points and equilibrium solutions of the logistic differential equation (6.8).

**Solution:** For the logistic differential equation,  $g(y) = ry(1 - \frac{y}{K})$ , and the solutions of  $g(y) = 0$  are  $y = 0$  and  $y = K$ . Thus 0 and  $K$  are the only equilibria, and  $y \equiv 0, y \equiv K$  are the corresponding equilibrium solutions of the logistic differential equation.  $\square$

The graphs of equilibrium solutions of an autonomous differential equation (6.32) are horizontal lines which separate the  $t - y$  plane into horizontal bands of the form

$$\{(t, y) \mid t \geq 0, y_1 < y < y_2\}$$

where  $y_1$  and  $y_2$  are consecutive equilibria of (6.32), with no equilibrium of (6.32) between  $y_1$  and  $y_2$ . See Figure 6.3 in Example 2 for an illustration.

**Property 2:** The graph of every solution curve of the differential equation  $y' = g(y)$  remains in the same band and is either monotone increasing [ $y'(t) > 0$ ] or monotone decreasing [ $y'(t) < 0$ ] for all  $t \geq 0$ , depending on whether  $g(y) > 0$  or  $g(y) < 0$  respectively in the band.

Suppose that  $y_1$  and  $y_2$  are consecutive equilibria of (6.32) and that  $y_1 < y(0) < y_2$ . If  $g(y)$  is indeed smooth enough to apply the existence and uniqueness

theorem of Section 7.2, then the graph of a solution cannot cross either of the constant solutions which form the boundaries of the band — or else at the crossing there would be a point in the  $t - y$  plane with two solutions passing through it, violating the uniqueness. Therefore the solution must remain in this band for all  $t \geq 0$ . If, for example,  $g(y) > 0$  for  $y_1 < y < y_2$ , then  $y'(t) = g\{y(t)\} > 0$  for  $t \geq 0$ , and the solution  $y(t)$  is monotone increasing. A similar argument shows that if  $g(y) < 0$  for  $y_1 < y < y_2$ , the solution  $y(t)$  is monotone decreasing.

If  $y(0)$  is above the largest equilibrium of (6.32), the band containing the solution is unbounded, and if  $g(y) > 0$  in this band, then the solution  $y(t)$  may be (positively) unbounded and does not necessarily exist for all  $t \geq 0$ . Likewise, if  $y(0)$  is below the smallest equilibrium of (3) and if  $g(y) < 0$  in the band containing the solution, then the solution may be unbounded (negatively) and may fail to exist for all  $t \geq 0$ .

**Example 2.** Describe the bands for the logistic differential equation (6.32) with  $r > 0$ , and find which solutions are increasing.

**Solution:** Since there are two equilibria  $y = 0$  and  $y = K$ , there are three bands to consider (Figure 6.3). If  $y > K$  (Band 1), the function  $g(y) = ry(1 - \frac{y}{K})$  is negative, and therefore solutions  $y(t)$  of (6.8) with  $y(0) > K$  are decreasing for all  $t$ . If  $0 < y < K$  (Band 2), the function  $g(y)$  is positive, and therefore solutions  $y(t)$  with  $0 < y(0) < K$  are increasing for all  $t$ . If  $y < 0$  (Band 3),  $g(y)$  is negative, so that solutions  $y(t)$  with  $y(0) < 0$  are decreasing for all  $t$ .  $\square$

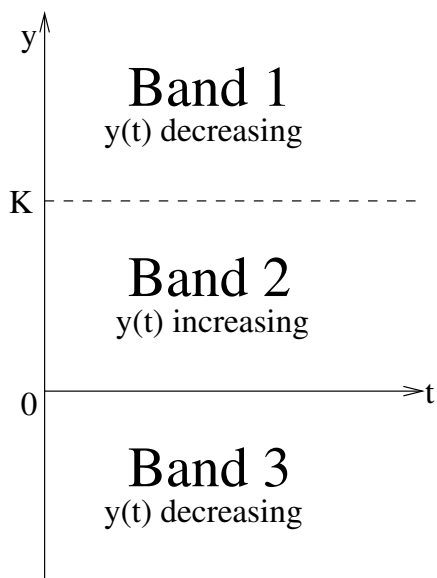


Figure 6.3: The  $t - y$  plane divided into bands by the equilibria of the logistic equation

**Property 3:** Every solution of the differential equation  $y' = g(y)$  which remains bounded for  $0 \leq t < \infty$  approaches a limit as  $t \rightarrow \infty$ .

This property is an immediate consequence of Property 2 and the fact from calculus that a function which is bounded and monotone (either increasing or decreasing) must approach a limit as  $t \rightarrow \infty$ . In order to show which solutions of a given differential equation have a limit, it is necessary to show which solutions remain bounded. A solution may become unbounded positively (i.e.,  $y(t) \rightarrow +\infty$ ) if it is monotone increasing when  $y$  is large. In many applications,  $g(y) < 0$  for large  $y$ , and in such a case every solution remains bounded, because solutions which become large and positive must be decreasing. If, as is also frequently the case in applications,  $y = 0$  is an equilibrium and only non-negative solutions are of interest, solutions cannot cross the line  $y = 0$  and become negatively unbounded (i.e.,  $y(t) \rightarrow -\infty$ ). In many applications,  $y$  stands for a quantity such as the number of members of a population, the mass of a radioactive substance, the quantity of money in an account, or the height of a particle above ground which cannot become negative. In such a situation, only non-negative solutions are significant, and if  $y = 0$  is not an equilibrium but a solution reaches the value zero for some finite  $t$ , we will consider the population system to have collapsed and the population to be zero for all larger  $t$ . If this is the case we need not be concerned with the possibility of solutions becoming negatively unbounded even if  $y = 0$  is not an equilibrium.

**Example 3.** Show that every solution of the logistic differential equation (6.8) with  $y(0) > 0$  approaches a limit as  $t \rightarrow \infty$ .

Solution: For the logistic equation,  $g(y) = ry(1 - \frac{y}{K})$ . As we have remarked in Example 2,  $g(y) < 0$  for  $y > K$ . Therefore every positive solution is bounded and by Property 3 has a limit as  $t \rightarrow \infty$ .  $\square$

We have observed from the form of the solutions of the logistic differential equation that every solution with positive initial value approaches the limit  $K$  as  $t \rightarrow \infty$ . We are now in a position to deduce this information without having to solve the differential equation, by considering the question of what values are possible limits for solutions of an autonomous differential equation.

**Property 4:** The only possible limits as  $t \rightarrow \infty$  of solutions of the differential equation  $y' = g(y)$  are the equilibria of the differential equation.

To see why this property is true, we use the fact from calculus that if a differentiable function tends to a limit as  $t \rightarrow \infty$ , then its derivative must tend to zero. If a solution  $y(t)$  of  $y' = g(y)$  tends to zero, then by the continuity of the function  $g(y)$  we have

$$0 = \lim_{t \rightarrow \infty} y'(t) = \lim_{t \rightarrow \infty} g\{y(t)\} = g\{\lim_{t \rightarrow \infty} y(t)\}.$$

Thus  $\lim_{t \rightarrow \infty} y(t)$  must be a root of the equation  $g(y) = 0$ , and hence an equilibrium of (6.32).

**Example 4.** Show that every solution of the logistic differential equation (6.8) with  $y(0) > 0$  has limit  $K$  as  $t \rightarrow \infty$ .

Solution: We have seen in Example 3 that every solution of (6.8) with  $y(0) > 0$  has a limit, and according to Property 4 a limit must be either  $K$  or 0. However, the limit 0 may be ruled out, as from Example 2 solutions whose initial values are positive and close to zero must be monotone increasing and therefore tend away from zero. Thus every solution of the logistic equation with positive initial value approaches the limit  $K$ . Figure 6.4 illustrates the behavior with  $r = 2$ ,  $K = 2$ .  $\square$

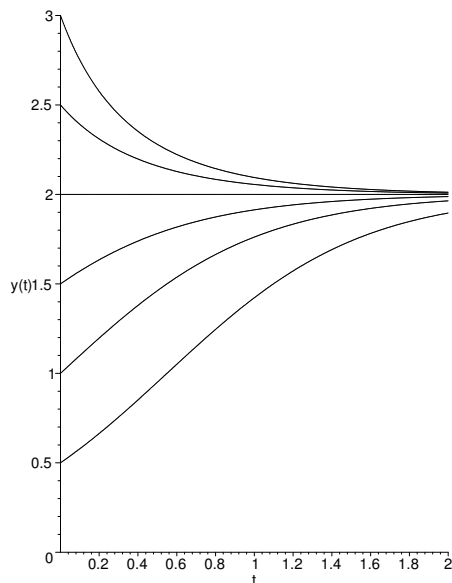


Figure 6.4: Some solutions of the logistic equation (6.8) with  $r = 2$ ,  $K = 2$

In applications, the initial condition usually comes from observations and is subject to experimental error. For a model to be a plausible predictor of what will actually occur, it is important that a small change in the initial value does not produce a large change in the solution. For example, the solution of the logistic differential equation with initial value zero remains zero for all values of  $t$ , but every solution with positive initial value, no matter how small, approaches the limit  $K$  as  $t \rightarrow \infty$ . Because of its extreme sensitivity to changes in the initial value, we do not ascribe practical significance to the equilibrium zero as a limiting value for solutions of the logistic equation.

An equilibrium  $\hat{y}$  of (6.32) such that every solution with initial value sufficiently close to  $\hat{y}$  approaches  $\hat{y}$  as  $t \rightarrow \infty$  is said to be *asymptotically stable*. If there are solutions which start arbitrarily close to an equilibrium but move away from it, then the equilibrium is said to be *unstable*. For the logistic differential equation (6.8) the equilibrium  $y = 0$  is unstable, and the equilibrium  $y = K$  is

asymptotically stable. In applications, unstable equilibria have no significance, because they can be observed only if the initial condition is “just right”.

**Property 5:** An equilibrium  $\hat{y}$  of  $y' = g(y)$  with  $g'(\hat{y}) < 0$  is asymptotically stable; an equilibrium  $\hat{y}$  with  $g'(\hat{y}) > 0$  is unstable.

To establish this property, we note that if  $\hat{y}$  is an equilibrium with  $g'(\hat{y}) < 0$  ( $g(y)$  has a negative slope at  $y = \hat{y}$ ), then  $g(y)$  is positive if  $y < \hat{y}$  and negative if  $y > \hat{y}$  (Figure 6.5(a)). In this case solutions above the equilibrium decrease toward the equilibrium, while solutions below the equilibrium increase toward the equilibrium. This shows that an equilibrium  $\hat{y}$  with  $g'(\hat{y}) < 0$  is asymptotically stable. By a similar argument, if  $g'(\hat{y}) > 0$ , solutions above the equilibrium increase and solutions below the equilibrium decrease, with both moving away from the equilibrium (Figure 6.5(b)).

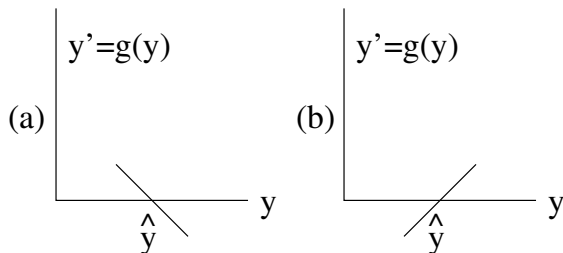


Figure 6.5: Equilibria  $\hat{y}$  with (a)  $g'(\hat{y}) < 0$ , (b)  $g'(\hat{y}) > 0$

**Example 5.** For each equilibrium of the logistic differential equation (6.8), determine whether it is asymptotically stable or unstable.

**Solution:** We have  $g(y) = ry(1 - \frac{y}{K})$ ,  $g'(y) = r - \frac{2ry}{K} = r(1 - \frac{2y}{K})$ . Since  $g'(K) = -r < 0$ , the equilibrium  $y = K$  is asymptotically stable, and since  $g'(0) = r > 0$ , the equilibrium  $y = 0$  is unstable.  $\square$

The properties we have developed make it possible for us to make a complete analysis of the asymptotic behavior, that is, the behavior as  $t \rightarrow \infty$ , of solutions of an autonomous differential equation (6.32) merely by examining the equilibria and the nature of the function  $g(y)$  for large values of  $y$ . We begin by finding all the equilibria (roots of the equation  $g(y) = 0$ ). An equilibrium  $\hat{y}$  with  $g'(\hat{y}) < 0$  is asymptotically stable, and all solutions with initial value in the two bands adjoining this equilibrium tend to it. An equilibrium  $\hat{y}$  with  $g'(\hat{y}) > 0$  is unstable and repels all solutions with initial value in the two bands adjoining it. An equilibrium  $\hat{y}$  with  $g'(\hat{y}) = 0$  must be analyzed more carefully. If  $g(y)$  is negative for values of  $y$  above the largest equilibrium, then no solutions become positively unbounded. If  $g(y)$  is positive for values of  $y$  above the largest equilibrium, then this equilibrium is unstable, and solutions with initial value above this equilibrium become unbounded. If  $g(y)$  is negative for values of  $y$  below the

smallest equilibrium, then this equilibrium is likewise unstable, and solutions with initial value below this equilibrium become negatively unbounded.

A convenient way to display the qualitative behavior of solutions of an autonomous differential equation (6.32) is by drawing the *phase line*. We draw the graph of the function  $g(y)$  and on the  $y$ -axis we may draw arrows to the right where the graph is above the  $y$ -axis and to the left where the graph is below the  $y$ -axis. The reason for doing this is that where the graph is above the axis  $g(y)$  is positive and therefore the solution  $y$  of (6.32) is increasing, while where the graph is below the axis  $g(y)$  is negative and therefore the solution  $y$  of (6.32) is decreasing. The points where the graph crosses the  $y$ -axis are the equilibria, and we can see from the directions of the arrows along the axis which equilibria are asymptotically stable and which are unstable. The phase line is the  $y$ -axis viewed as the line on which the solution curve moves, thinking of  $t$  as a parameter. Thus the solution is described by the motion along the line, whose direction is given by the arrows. The graph of Figure 6.6 describes a situation in which there are asymptotically stable equilibria at  $y = -1$  and  $y = 2$ , and an unstable equilibrium at  $y = 0$ .

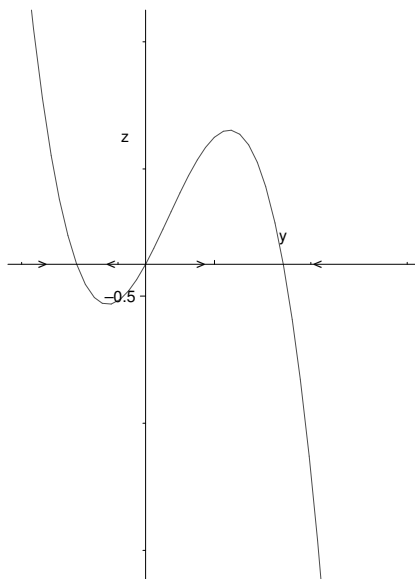


Figure 6.6: A phase line superimposed on the corresponding graph of  $g(y)$

**Example 8.** Describe the asymptotic behavior of solutions of the differential equation (6.31).

**Solution:** Here  $g(y) = -ay + b$ ,  $g'(y) = -a$ . The only equilibrium is  $y = \frac{b}{a}$ . If  $a > 0$ , this equilibrium is asymptotically stable, and  $g(y) < 0$  if  $y$  is large and positive,  $g(y) > 0$  if  $y$  is large and negative. This means that every solution is bounded and approaches the limit  $\frac{b}{a}$ . If  $a < 0$ , however, the equilibrium is

unstable. Further, since  $g(y) > 0$  above the equilibrium and  $g(y) < 0$  below the equilibrium, every solution is unbounded, either positively or negatively.  $\square$

**Example 9.** Describe the asymptotic behavior of solutions with  $y(0) \geq 0$  of the differential equation

$$y' = y(re^{-y} - d), \quad (6.33)$$

where  $r$  and  $d$  are positive constants.

**Solution:** The equilibria of (6.33) are the solutions of  $y(re^{-y} - d) = 0$ . Thus there are two equilibria, namely  $y = 0$  and the solution  $\hat{y}$  of  $re^{-y} = d$ , which is  $\hat{y} = \ln \frac{r}{d}$ . If  $r < d$ ,  $\hat{y} < 0$ , and only the equilibrium  $y = 0$  is of interest. In this case,  $g(y) < 0$  for  $y > 0$ , and solutions with  $y(0) > 0$  decrease to 0 (Figure 6.7).

If  $r > d$ , we define  $K$  to be  $\ln \frac{r}{d}$ , so that the positive equilibrium is  $y = K$ . We may now rewrite the differential equation (6.33) as  $y' = ry(e^{-y} - e^{-K})$ . For the function  $g(y) = ry(e^{-y} - e^{-K})$ , we have  $g'(y) = r(e^{-y} - e^{-K}) - rye^{-y}$  and  $g'(0) = r(1 - e^{-K}) > 0$ , implying that the equilibrium  $y = 0$  is unstable. The equilibrium  $y = K$  is asymptotically stable since  $g'(K) = -rKe^{-K} < 0$ . All positive solutions are bounded, because  $g(y) < 0$  for  $y > K$ . Thus every solution with  $y(0) > 0$  tends to the limit  $K$ , while the solution with initial value zero is the zero function and has limit zero (Figure 6.8).  $\square$

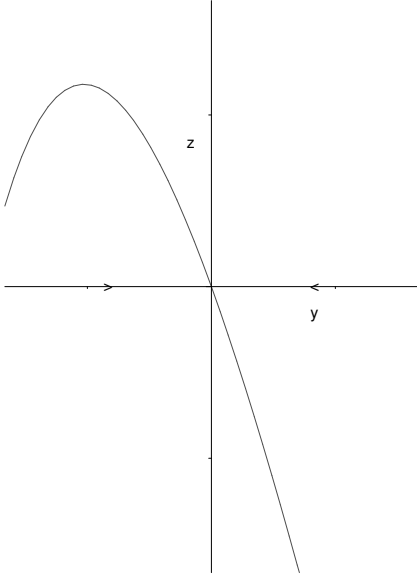


Figure 6.7: A phase line and graph of  $g(y)$  for (6.33), with  $r < d$

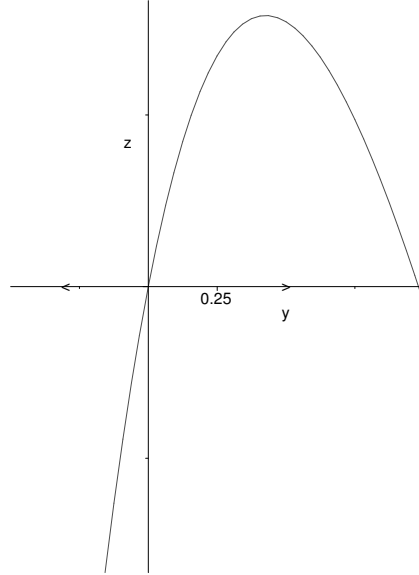


Figure 6.8: A phase line and graph of  $g(y)$  for (6.33), with  $r > d$

We see that the solutions of (6.33), which has been suggested as a population growth model with per capita birth rate  $re^{-y}$  and per capita death rate  $d$ ,

behaves qualitatively in the same manner as solutions of the logistic equation. Some other differential equations which exhibit the same behavior have been proposed as population models. The original formulation of the logistic equation was based on a per capita growth rate which should be a decreasing function of population size, positive for small  $y$  but negative for large  $y$ . It is not difficult to show that every population model of the form (6.32) for which the per capita growth rate is a decreasing function of the population size  $y$ , and which is positive for  $0 < y < K$  and negative for  $y > K$ , has the property that every solution with  $y(0) > 0$  approaches the limit  $K$  as  $t \rightarrow \infty$ .

In many applications, the function  $g(y)$  has the form  $g(y) = yf(y)$ , which has the effect of guaranteeing that  $y = 0$  is an equilibrium. Then  $g'(y) = f(y) + yf'(y)$ . At the equilibrium  $y = 0$ ,  $g'(0) = f(0)$  and at a nonzero equilibrium  $\hat{y}$  with  $f(\hat{y}) = 0$ ,  $g'(\hat{y}) = \hat{y}f'(\hat{y})$ . Thus the equilibrium  $y = 0$  is asymptotically stable if  $f(0) < 0$  and unstable if  $f(0) > 0$ ; a nonzero equilibrium  $\hat{y}$  is asymptotically stable if  $f'(\hat{y}) < 0$  and unstable if  $f'(\hat{y}) > 0$ . We shall restate this result formally as a theorem.

**EQUILIBRIUM STABILITY THEOREM:** An equilibrium  $\hat{y}$  of  $y' = g(y)$  with  $g'(\hat{y}) < 0$  is asymptotically stable; an equilibrium  $\hat{y}$  with  $g'(\hat{y}) > 0$  is unstable. The equilibrium  $y = 0$  of  $y' = yf(y)$  is asymptotically stable if  $f(0) < 0$  and unstable if  $f(0) > 0$ , while a nonzero equilibrium  $\hat{y}$  is asymptotically stable if  $f'(\hat{y}) < 0$  and unstable if  $f'(\hat{y}) > 0$ .

## 6.6 Some Qualitative Applications

The results of the previous section make it possible to obtain information about the behavior of solutions of differential equations without having to solve to find solutions explicitly. This is of importance in many applications not only because it is often much easier to analyze a differential equation qualitatively than to solve it analytically but also because it enables us to describe the behavior of solutions in terms of general properties of the differential equation rather than in terms of the specific functions in the differential equation. For example, in the previous section we mentioned several differential equations used as population models and having the property that every positive solution approaches the same limit. While the logistic differential equation has often been used as a population model, it is only one of many possibilities.

### 6.6.1 Population growth with harvesting

We consider a situation in which a population grows according to a logistic law but members are removed from it at a constant rate in time. For example, the population of deer in Wisconsin is reduced each hunting season by a fixed number determined by the state agency which issues hunting permits. We will assume that  $H$  members of the population are removed in unit time and that the rate of removal is uniform (not seasonal as in deer hunting). The constant

$H$  is called the *harvest rate*, and this type of removal is called *constant-yield harvesting*. Then the population size satisfies a differential equation of the form

$$y' = ry \left(1 - \frac{y}{K}\right) - H, \quad (6.34)$$

where  $r$ ,  $K$ , and  $H$  are positive constants. While this problem can be solved explicitly by separation of variables, the integration must be handled by examining three different cases depending on the values of the constants. It is simpler (and probably more informative) to carry out a qualitative analysis.

The equilibria of (6.34) are the solutions of  $ry - \frac{r}{K}y^2 - H = 0$ , or

$$y^2 - Ky + \frac{HK}{r} = 0 \quad (6.35)$$

These are given by

$$y = \frac{1}{2} \left[ K \pm \sqrt{K^2 - \frac{4HK}{r}} \right]. \quad (6.36)$$

We must distinguish three cases:

- (i)  $K^2 - \frac{4HK}{r} > 0$  or  $H < \frac{rK}{4}$ ,
- (ii)  $K^2 - \frac{4HK}{r} = 0$  or  $H = \frac{rK}{4}$ ,
- (iii)  $K^2 - \frac{4HK}{r} < 0$  or  $H > \frac{rK}{4}$ .

The number and nature of the equilibria are different in the three cases.

In case (i) there are two distinct equilibria given by (6.36), and it is clear that one, which we shall call  $y_1$ , is smaller than  $\frac{K}{2}$  while the other, which we shall call  $y_2$ , is larger than  $\frac{K}{2}$ . If we let  $g(y) = ry - \frac{r}{K}y^2 - H$ , then  $g'(y) = r - \frac{2r}{K}y = \frac{2r}{K}(\frac{K}{2} - y)$ . Thus  $g'(y) > 0$  if  $y < \frac{K}{2}$  and  $g'(y) < 0$  if  $y > \frac{K}{2}$ . We now see from Property 5 of Section 7.5 that an equilibrium  $\hat{y}$  of (6.34) with  $\hat{y} > \frac{K}{2}$  is asymptotically stable while an equilibrium  $\hat{y}$  with  $\hat{y} < \frac{K}{2}$  is unstable. More specifically, in case (i), we see that  $y_1$  is unstable while  $y_2$  is asymptotically stable.

Since  $g(y) < 0$  for  $0 < y < y_1$ , a solution  $y(t)$  with  $0 < y(0) < y_1$  is monotone decreasing and reaches zero in finite time. When this happens, we consider the population to have been wiped out and the model to have collapsed. If  $y(0) > y_1$ , the solution  $y(t)$  approaches the limit  $y_2$  as  $t \rightarrow \infty$ . This is illustrated in Figure 6.9 with  $r = 2$ ,  $K = 2$ ,  $H = \frac{5}{9}$ .

In case (ii) there is a single equilibrium  $\frac{K}{2}$  which is a double root of (2), and  $g'(\frac{K}{2}) = 0$ . The equilibrium stability theorem does not apply here, but we can rewrite  $g(y)$  as

$$g(y) = ry - \frac{r}{K}y^2 - \frac{rK}{4} = -\left(\frac{r}{K}y^2 - ry + \frac{rK}{4}\right) = -\frac{r}{K} \left(y - \frac{K}{2}\right)^2,$$

we see that  $g(y) < 0$  if  $y \neq \frac{K}{2}$ . Thus every solution is monotone decreasing, and solutions starting above  $\frac{K}{2}$  approach the equilibrium  $\frac{K}{2}$ , while solutions starting below  $\frac{K}{2}$  reach zero in finite time. This behavior is illustrated in Figure 6.10 with  $r = 2$ ,  $K = 2$ ,  $H = 1$ .

In case (iii) the roots of (6.35) are complex and there are no equilibria of (6.34). Since  $g(y)$  has no real zeroes and  $g(0) = -H < 0$ ,  $g(y) < 0$  for all  $y$ . Thus every solution is monotone decreasing and reaches zero in finite time; the population is wiped out no matter what the initial population size is, as shown in Figure 6.11 with  $r = 2$ ,  $K = 2$ ,  $H = 1.5$ .

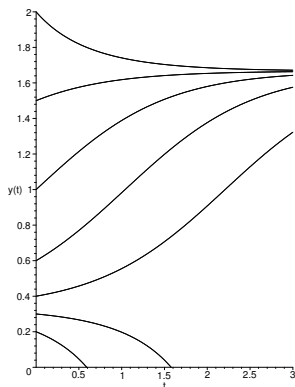


Figure 6.9: Solutions to (6.34),  $H < rK/4$

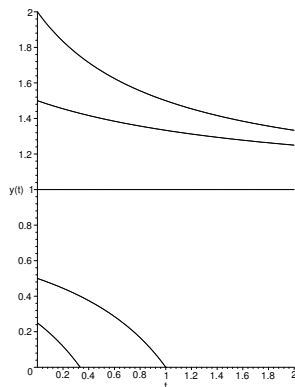


Figure 6.10: Solutions to (6.34),  $H = rK/4$

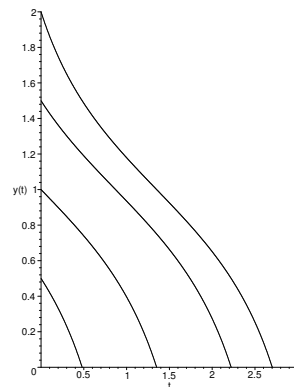


Figure 6.11: Solutions to (6.34),  $H > rK/4$

As a function of  $H$  the limiting population size behaves in the following way for suitably chosen initial population sizes. For  $H = 0$  the limiting population size is  $K$  for every positive initial population size. As  $H$  increases from 0 to  $\frac{rK}{4}$ , the limiting population size  $y_2$  decreases monotonically from  $K$  to  $K/2$ , for all initial conditions  $y(0) > y_1$ . For  $H = \frac{rK}{4}$  the limiting population size is  $\frac{K}{2}$ , provided  $y(0) > \frac{K}{2}$ . However, for  $H > \frac{rK}{4}$ , the population size reaches zero in finite time for every initial population size, meaning that the population is wiped out. Thus the limiting population size decreases continuously (from  $K$  to  $\frac{K}{2}$ ) as  $H$  increases from 0 to  $rK/4$ , and then jumps to zero as  $H$  passes through  $rK/4$ . Such a discontinuity in the limiting behavior of a system is called a (mathematical) *catastrophe*. The corresponding biological catastrophe is the wiping out of the population in finite time as  $H$  increases continuously through a critical value, in this case  $rK/4$ .

For any general population model of the form  $y' = g(y)$  whose per capita growth rate is a monotone decreasing function of population size and positive for small  $y$  but negative for large  $y$ , the response to constant yield harvesting is similar. The process is modelled by the differential equation

$$y' = g(y) - H. \quad (6.37)$$

For such models, the limiting population size is positive if  $H$  is less than a critical harvest rate  $H_c$ , which is the maximum value of the function  $g(y)$ , and jumps to zero as  $H$  passes through  $H_c$ . In the logistic case,  $H_c = \max_y ry \left(1 - \frac{y}{K}\right) = \frac{rK}{4}$ .

### 6.6.2 The spread of infectious diseases

In Section 6.4 we formulated the differential equation

$$I' = (\beta K - \gamma)I - \beta I^2, \quad (6.38)$$

where  $I(t)$  is the number of infective individuals in a population at time  $t$ , as a model for the spread of a communicable disease. As (6.38) is a logistic equation, we were able to use the solution of this equation found in Section 6.3 to describe the asymptotic behavior of solutions.

The qualitative results obtained in Section 6.5 make it possible for us to describe the asymptotic behavior of solutions of (6.38) without solving. The differential equation (6.38) has the form  $I' = g(I)$  with

$$g(I) = (\beta K - \gamma)I - \beta I^2, \quad g'(I) = (\beta K - \gamma) - 2\beta I.$$

There are two equilibria, solutions of  $g(I) = 0$ , given by  $I = 0$  and  $I = K - \frac{\gamma}{\beta}$ . However, if  $\beta K/\gamma < 1$ , the second equilibrium is negative and has no epidemiological meaning. Since  $g'(0) = \beta K - \gamma$ , the equilibrium  $I = 0$  is asymptotically stable if  $\beta K - \gamma < 0$ , i.e., if  $\beta K/\gamma < 1$ , and unstable if  $\beta K/\gamma > 1$ . Since  $g'(K - \frac{\gamma}{\beta}) = -(\beta K - \gamma)$ , the equilibrium  $I = K - \frac{\gamma}{\beta}$  is asymptotically stable if  $\beta K/\gamma > 1$ . Thus there is always a single asymptotically stable equilibrium,  $I = 0$  if  $\beta K/\gamma < 1$  and  $I = K - \frac{\gamma}{\beta}$  if  $\beta K/\gamma > 1$ . The quantity  $\beta K/\gamma$  is called the disease's *basic reproduction number*, usually denoted by  $R_0$ . If  $R_0 < 1$ , every solution of (6.38) approaches the equilibrium  $I = 0$ , called the **disease-free equilibrium**; the disease tends to die out. If  $R_0 > 1$ , every solution of (6.38) approaches the equilibrium  $I = K - \frac{\gamma}{\beta} > 0$ , called the **endemic equilibrium**; the disease persists in the population.

We can interpret the expression for  $R_0$  biologically by recalling from the original discussion in Section 7.4 that an “average” infective makes  $\beta K$  potentially infective contacts in unit time and remains infective an average of  $1/\gamma$  units of time. Therefore  $R_0 = \beta K/\gamma$  represents the average number of secondary infections caused per infected individual before recovery (hence the term basic reproduction number). If each infection is able to replace itself and more ( $R_0 > 1$ ), then we should expect the disease to persist, while if each infection cannot, on average, replace itself before the individual recovers ( $R_0 < 1$ ), then we would say the disease is doing a poor job of reproducing itself and should die out. This kind of insight illustrates the reason why such mathematical models are useful. For this model, one might be able to make this same argument without going through all the details of a qualitative analysis, but not so for the more complicated models which arise in studying the problems of interest to us. *Mathematical* results give rise to *biological* insights, and that is our motivation.

## 6.7 Some exercises

In each of Exercises 1–2, assume that the population size satisfies a simple growth law.

1. Suppose that the birth rate of a given population is 0.36 per member per day with no deaths. If the population size on day zero is 50, what is the population size 10 days later?
2. Suppose a population has 173 members at  $t = 0$  and 262 members at  $t = 10$ . Estimate the population size at  $t = 5$ .
3. If the half-life of a radioactive substance is 30 days, how long would it take until 99 % of the substance decays?
4. In a sample of uranium 238, it is found that 0.02867 % of the mass disintegrates in 10 years. Find the half-life of uranium 238.
5. How old is a fossil in which 85 % of the carbon 14 has disintegrated?
6. \* Show that the solution of the initial value problem  $y' = ay$ ,  $y(t_0) = c$  is  $y(t) = ce^{a(t-t_0)}$ .
7. Show that  $y = (c - t^2)^{-\frac{1}{2}}$  is a solution of the differential equation  $y' = ty^3$  for every choice of the constant  $c$ .
8. Show that  $y = ct$  is a solution of the differential equation  $y' = y/t$  for every choice of the constant  $c$ .
9. Among the family of solutions  $y = (c - t^2)^{-\frac{1}{2}}$  of  $y' = ty^3$ , find the solution such that  $y(0) = 1$ .
10. Find the solution of  $y' = ty^3$  such that  $y(0) = 0$ .
11. Show that the solution of the initial value problem  $y' = ty^3$ ,  $y(0) = -1$  is  $y = -(1 - t^2)^{-\frac{1}{2}}$ .
12. Among the solutions  $y = \frac{1+ce^t}{1-ce^t}$  of the differential equation  $y' = \frac{1}{2}(y^2 - 1)$ , find the ones which satisfy the initial conditions  $y(0) = 1$ ,  $y(0) = 0$ , and  $y(2) = 0$ .
13. Show that if  $\hat{y}$  is a constant such that  $f(t, \hat{y}) = 0$  for all  $t$ , then  $y = \hat{y}$  is a constant solution of  $y' = f(t, y)$ .

In each of Exercises 14–17, find the solution of the given differential equation.

14.  $y' = t^2y$
15.  $y' = \frac{t}{y}$

16.  $y' = -2ty$

17.  $y' = \frac{y+1}{t}$

In each of Exercises 18–20, find the solution of the given differential equation which satisfies the given initial condition.

18.  $y' = t^2y, \quad y(5) = 1$

19.  $y' = \frac{t}{y}, \quad y(0) = 1$

20.  $y' = -2ty, \quad y(0) = y_0$

21. Suppose a population satisfies a logistic model with  $r = 0.4$ ,  $K = 100$ ,  $y(0) = 5$ . Find the population size when  $t = 10$ .

22. Find the limit as  $t \rightarrow \infty$  of the solution of the initial value problem  $y' = -y + 1$ ,  $y(0) = 0$ .

23. Find the limit as  $t \rightarrow \infty$  of the solution of the initial value problem  $y' = -y + 1$ ,  $y(0) = 100$ .

24. Find two solutions of the initial value problem  $y' = 3y^{2/3}$ ,  $y(0) = 0$ .

25. Find all differentiable functions  $f(t)$  such that  $[f(t)]^2 = \int_0^t f(s) ds$  for all  $t \geq 0$ .

26. Find all continuous (not necessarily differentiable) functions  $f(t)$  such that  $[f(t)]^2 = \int_0^t f(s) ds$  for all  $t \geq 0$ .

For each of the differential equations in Exercises 27–32, draw the phase line, find all equilibria and describe the behavior of solutions as  $t \rightarrow \infty$ .

27.  $y' = y$

28.  $y' = -y$

29.  $y' = y^3$

30.  $y' = y(1 - y)(2 - y)$

31.  $y' = -y(1 - y)(2 - y)$

32.  $y' = y^2(y + 1)$

For each of the differential equations in Exercises 33–34, describe the behavior of solutions with  $y(0) > 0$ .

33.  $y' = \frac{ry(K-y)}{K+Ay}$

34.  $y' = ry \left[ 1 - \left( \frac{y}{K} \right)^\theta \right]$



## Chapter 7

# Systems of Ordinary Differential Equations

In the previous chapter we saw how to model and analyze continuously changing quantities using differential equations. In many applications of interest there may be two or more interacting quantities—populations of two or more species, for instance, or parts of a whole, which depend upon each other. When the amount or size of one quantity depends in part on the amount of another, and vice versa, they are said to be *coupled*, and it is not possible or appropriate to model each one separately. In these cases we write models which consist of *systems* of differential equations. In this chapter, we will find that the quantitative and qualitative approaches we used to analyze individual differential equations in Chapter 7 extend in a more or less natural way to cover systems of differential equations. Extending them will require some basic multivariable calculus, principally the use of partial derivatives and the idea of linear approximation.

### 7.1 The Phase Plane

Our purpose is to study *two-dimensional autonomous systems of first-order differential equations* of the general form

$$\begin{aligned}y' &= F(y, z) \\z' &= G(y, z)\end{aligned}\tag{7.1}$$

Usually, it is not possible to solve such a system, by which we mean to find  $y$  and  $z$  as functions of  $t$  so that

$$y'(t) = F(y(t), z(t)), \quad z'(t) = G(y(t), z(t))$$

for all  $t$  in some interval. However, we can often obtain information about the relation between the functions  $y(t)$  and  $z(t)$ . Geometrically, such information is displayed as a curve in the  $y - z$  plane, called the *phase plane* for this system

An *equilibrium* of the system (7.1) is a solution  $(y_\infty, z_\infty)$  of the pair of equations

$$F(y, z) = 0, \quad G(y, z) = 0.$$

Geometrically, an equilibrium is a point in the phase plane. In terms of the system (7.1), an equilibrium gives a constant solution  $y = y_\infty, z = z_\infty$  of the system. This definition is completely analogous to the definition of an equilibrium given for a first-order differential equation in Section 7.5.

The *orbit* of a solution  $y = y(t), z = z(t)$  of the system (7.1) is the curve in the  $y - z$  phase plane consisting of all points  $(y(t), z(t))$  for  $0 \leq t < \infty$ . A closed orbit corresponds to a periodic solution because the orbit must travel repeatedly around the closed orbit as  $t$  increases.

There is a geometric interpretation of orbits which is analogous to the interpretation given for solutions of first-order differential equations in Section 7.2. Just as the curve  $y = y(t)$  has slope  $y' = f(t, y)$  at each point  $(t, y)$  along its length, an orbit of (7.1) (considering  $z$  as an implicit function of  $y$ ) has slope

$$\frac{dz}{dy} = \frac{z'}{y'} = \frac{G(y, z)}{F(y, z)}$$

at each point of the orbit. The *direction field* for a two-dimensional autonomous system is a collection of line segments in the phase plane with this slope at each point  $(y, z)$ , and an orbit must be a curve which is tangent to the direction field at each point of the curve. Computer algebra systems such as Maple and Mathematica may be used to draw the direction field for a given system.

**Example 1.** Describe the orbits of the system

$$y' = z, \quad z' = -y.$$

**Solution:** If we consider  $z$  as a function of  $y$ , we have

$$\frac{dz}{dy} = z'/y' = -\frac{y}{z}.$$

Solution by separation of variables gives

$$\int z \, dz = - \int y \, dy,$$

and integration gives  $\frac{z^2}{2} = -\frac{y^2}{2} + c$ . Thus every orbit is a circle  $y^2 + z^2 = 2c$  with centre at the origin, and every solution is periodic.  $\square$

To find equilibria of a system (7.1), it is helpful to draw the *nullclines*, namely the curves  $F(y, z) = 0$  on which  $y' = 0$ , and  $G(y, z) = 0$  on which  $z' = 0$ . An equilibrium is an intersection of these two curves.

## 7.2 Linearization of a System at an Equilibrium

Sometimes it is possible to find the orbits in the phase plane of a system of differential equations, but it is rarely possible to solve a system of differential equations analytically. For this reason, our study of systems will concentrate on qualitative properties. The linearization of a system of differential equations at an equilibrium is a linear system with constant coefficients, whose solutions approximate the solutions of the original system near the equilibrium. In this section we shall see how to find the linearization of a system. In the next section we shall see how to solve linear systems with constant coefficients, and this will enable us to understand much of the behavior of solutions of a system near an equilibrium.

Let  $(y_\infty, z_\infty)$  be an equilibrium of a system

$$\begin{aligned}y' &= F(y, z) \\z' &= G(y, z)\end{aligned}\tag{7.2}$$

that is, a point in the phase plane such that

$$F(y_\infty, z_\infty) = 0, \quad G(y_\infty, z_\infty) = 0.\tag{7.3}$$

We will assume that the equilibrium is *isolated*, that is, that there is a circle centered around  $(y_\infty, z_\infty)$  which does not contain any other equilibrium. We shift the origin to the equilibrium by letting  $y = y_\infty + u$ ,  $z = z_\infty + v$ , and then make linear approximations to  $F(y_\infty + u, z_\infty + v)$  and  $G(y_\infty + u, z_\infty + v)$ . The difference here between a one-dimensional system and a two-dimensional one is that the linear approximation in two or more dimensions uses partial derivatives. Our approximations are

$$\begin{aligned}F(y_\infty + u, z_\infty + v) &\approx F(y_\infty, z_\infty) + F_y(y_\infty, z_\infty)u + F_z(y_\infty, z_\infty)v \\G(y_\infty + u, z_\infty + v) &\approx G(y_\infty, z_\infty) + G_y(y_\infty, z_\infty)u + G_z(y_\infty, z_\infty)v\end{aligned}\tag{7.4}$$

with error terms  $h_1$  and  $h_2$  respectively which are negligible relative to the linear terms in (7.4) when  $u$  and  $v$  are small (i.e., close to the equilibrium).

The linearization of the system (7.2) at the equilibrium  $(y_\infty, z_\infty)$  is defined to be the linear system with constant coefficients

$$\begin{aligned}u' &= F_y(y_\infty, z_\infty)u + F_z(y_\infty, z_\infty)v \\v' &= G_y(y_\infty, z_\infty)u + G_z(y_\infty, z_\infty)v.\end{aligned}\tag{7.5}$$

To obtain it, we first note that  $y' = u'$ ,  $z' = v'$ , and then substitute (7.4) into (7.2). By (7.3) the constant terms are zero, and for the linearization we neglect the higher-order terms  $h_1$  and  $h_2$ . The *coefficient matrix* of the linear system (7.5) is the matrix of constants

$$\begin{bmatrix} F_y(y_\infty, z_\infty) & F_z(y_\infty, z_\infty) \\ G_y(y_\infty, z_\infty) & G_z(y_\infty, z_\infty) \end{bmatrix}.$$

In population models, this matrix is often called the *community matrix* of the system at equilibrium. Its entries describe the effect of a change in each variable on the growth rates of the two variables.

**Example 1.** Find the linearization at each equilibrium of the system

$$y' = A - \beta yz - \mu y, \quad z' = \beta yz - (\gamma + \mu)z.$$

(This corresponds to the classical Kermack-McKendrick SIR model for a possibly endemic disease, which we shall study in Section 8.4. Here  $y$  corresponds to the number of susceptible, uninfected individuals,  $z$  to the number of infected, infective individuals,  $A$  to the birth rate,  $\mu$  to the death rate and  $\beta$  and  $\gamma$  to the infection and recovery rates, respectively.)

**Solution:** The equilibria are the solutions of  $A = y(\beta z + \mu)$ ,  $z(\beta y) = (\gamma + \mu)z$ . To satisfy the second of these equations, we must have either  $z = 0$  or  $\beta y = \gamma + \mu$ . If  $z = 0$ , the first equation gives  $y = A/\mu$ . If  $z > 0$ , the second equation gives  $\beta y = \gamma + \mu$ . Since

$$\begin{aligned} \frac{\partial}{\partial y} [-\beta yz - \mu y] &= -\beta z - \mu, & \frac{\partial}{\partial z} [-\beta yz] &= -\beta y, \\ \frac{\partial}{\partial y} [\beta yz - (\gamma + \mu)z] &= \beta z, & \frac{\partial}{\partial z} [\beta yz - (\gamma + \mu)z] &= \beta y - (\gamma + \mu), \end{aligned}$$

the linearization at an equilibrium  $(y_\infty, z_\infty)$  is

$$\begin{aligned} u' &= -(\beta z_\infty + \mu)u - \beta y_\infty v, \\ v' &= \beta z_\infty u + (\beta y_\infty - (\gamma + \mu))v. \end{aligned}$$

At the equilibrium  $(A/\mu, 0)$  the linearization is

$$\begin{aligned} u' &= -\mu u - \beta \frac{A}{\mu} v \\ v' &= \beta \frac{A}{\mu} - (\gamma + \mu)v \end{aligned}$$

At the other equilibrium with  $I > 0$  the linearization is

$$\begin{aligned} u' &= -(\beta I_\infty + \mu)u - (\mu + \alpha)v \\ v' &= \beta I_\infty u \quad \square \end{aligned}$$

An equilibrium of the system (7.2) with the property that every orbit with initial value sufficiently close to the equilibrium remains close to the equilibrium for all  $t \geq 0$ , and approaches the equilibrium as  $t \rightarrow \infty$ , is said to be *locally asymptotically stable*. An equilibrium of (7.2) with the property that some solutions starting arbitrarily close to the equilibrium move away from it is said to be *unstable*. These definitions are completely analogous to those given in

Section 7.5 for first-order differential equations. We speak of local asymptotic stability to distinguish from global asymptotic stability, which is the property that *all* solutions, not merely those with initial value sufficiently close to the equilibrium, approach the equilibrium. If we speak of asymptotic stability of an equilibrium we will mean local asymptotic stability unless we specify that the asymptotic stability is global.

The fundamental property of the linearization which we will use to study stability of equilibria is the following result, which we state without proof. The proof may be found in any text which covers the qualitative study of nonlinear differential equations. Here we suppose  $F$  and  $G$  to be twice differentiable, that is, smooth enough for the linearization to give a correct picture.

**LINEARIZATION THEOREM:** If  $(y_\infty, z_\infty)$  is an equilibrium of the system

$$y' = F(y, z), \quad z' = G(y, z)$$

and if every solution of the linearization at this equilibrium approaches zero as  $t \rightarrow \infty$ , then the equilibrium  $(y_\infty, z_\infty)$  is (locally) asymptotically stable. If the linearization has unbounded solutions, then the equilibrium  $(y_\infty, z_\infty)$  is unstable.

For a first-order differential equation  $y' = g(y)$  at an equilibrium  $y_\infty$ , the linearization is the first-order linear differential equation  $u' = g'(y_\infty)u$ . We may solve this differential equation by separation of variables and see that all solutions approach zero if  $g'(y_\infty) < 0$ , and there are unbounded solutions if  $g'(y_\infty) > 0$ . We have seen in Section 7.5, without recourse to the linearization, that the equilibrium is locally asymptotically stable if  $g'(y_\infty) < 0$  and unstable if  $g'(y_\infty) > 0$ . The linearization theorem is valid for systems of any dimension and is the approach needed for the study of stability of equilibria for systems of dimension higher than 1.

Note that there is a case where the theorem above does not draw any conclusions, namely the case where the linearization about the equilibrium is neither asymptotically stable nor unstable. In this case, the equilibrium of the original (nonlinear) system may be asymptotically stable, unstable, or neither.

**Example 2.** For each equilibrium of the system

$$y' = z, \quad z' = -2(y^2 - 1)z - y$$

determine whether the equilibrium is asymptotically stable or unstable.

**Solution:** The equilibria are the solutions of  $z = 0$ ,  $-2(y^2 - 1)z - y = 0$ , and thus the only equilibrium is  $(0, 0)$ . Since  $\frac{\partial}{\partial y}[z] = 0$ ,  $\frac{\partial}{\partial z}[z] = 1$ , and

$$\frac{\partial}{\partial y}[-2(y^2 - 1)z - y] = -4yz - 1, \quad \frac{\partial}{\partial z}[-2(y^2 - 1)z - y] = -2(y^2 - 1),$$

the linearization at  $(0, 0)$  is  $u' = 0u + 1v = v$ ,  $v' = -u + 2v$ . We can actually solve this system of equations outright by using a clever trick to reduce it to a

single equation. We subtract the first equation from the second to give  $(v-u)' = (v-u)$ . This is a first-order differential equation for  $(v-u)$ , whose solution is  $v-u = c_1 e^t$ . This partial result is already enough to tell us that the equilibrium is unstable, as since the difference between  $u$  and  $v$  grows exponentially, at least one of them must therefore grow exponentially as well. As there are unbounded solutions, we conclude that the equilibrium  $(0,0)$  is unstable.  $\square$

### 7.3 Solution of Linear Systems with Constant Coefficients

We have seen in the preceding section that the stability of an equilibrium of a system of differential equations is determined by the behavior of solutions of the system's linearization at the equilibrium. This linearization is a linear system with constant coefficients (recall that a linear system has right-hand sides linear in  $y$  and  $z$ ). Thus, in order to be able to decide whether an equilibrium is asymptotically stable, we need to be able to solve linear systems with constant coefficients. We were able to do this in the examples of the preceding section because the linearizations took a simple form with one of the equations of the system containing only a single variable. In this section we shall develop a more general technique.

The problem we wish to solve is a general two-dimensional linear system with constant coefficients,

$$\begin{aligned}y' &= ay + bz, \\z' &= cy + dz,\end{aligned}\tag{7.6}$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  are constants. We look for solutions of the form

$$y = Y e^{\lambda t}, \quad z = Z e^{\lambda t},\tag{7.7}$$

where  $\lambda$ ,  $Y$ , and  $Z$  are constants to be determined, with  $Y$  and  $Z$  not both zero. When we substitute the form (7.7) into the system (7.6), using  $y' = \lambda Y e^{\lambda t}$ ,  $z' = \lambda Z e^{\lambda t}$ , we obtain two conditions

$$\begin{aligned}\lambda Y e^{\lambda t} &= a Y e^{\lambda t} + b Z e^{\lambda t}, \\ \lambda Z e^{\lambda t} &= c Y e^{\lambda t} + d Z e^{\lambda t},\end{aligned}$$

which must be satisfied for all  $t$ . Because  $e^{\lambda t} \neq 0$  for all  $t$ , we may divide these equations by  $e^{\lambda t}$  to obtain a system of two equations which do not depend on  $t$ , namely

$$\begin{aligned}\lambda Y &= a Y + b Z, \\ \lambda Z &= c Y + d Z\end{aligned}$$

or

$$\begin{aligned}(a - \lambda)Y + bZ &= 0, \\ cY + (d - \lambda)Z &= 0.\end{aligned}\tag{7.8}$$

The pair of equations (7.8) is a system of two homogeneous linear algebraic equations for the unknowns  $Y$  and  $Z$ . For certain values of the parameter  $\lambda$  this system will have a solution other than the obvious solution  $Y = 0, Z = 0$ . In order that the system (7.8) have a non-trivial solution for  $Y$  and  $Z$ , it is necessary that the determinant of the coefficient matrix, which is  $(a - \lambda)(d - \lambda) - bc$ , be equal to zero (this result comes from linear algebra). This gives a quadratic equation, called the *characteristic equation*, of the system (7.6) for  $\lambda$ . We may rewrite the characteristic equation as

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0. \quad (7.9)$$

We will assume that  $ad - bc \neq 0$ , which is equivalent to the assumption that  $\lambda = 0$  is not a root of (7.9). Our reason for this assumption is the following: If  $\lambda = 0$  is a root (or, equivalently,  $ad - bc = 0$ ) the equilibrium conditions will reduce to a single equation since each equation is a constant multiple of the other and there is effectively only one equilibrium equation. Consequently, there will be a line of non-isolated equilibria. We do not wish to explore this problem in part because the treatment of non-isolated equilibria is complicated and in part because it is not possible to apply the linearization theorem of the previous section when the linearization of a system about an equilibrium is of this form.

If  $ad - bc \neq 0$ , the characteristic equation has two roots  $\lambda_1$  and  $\lambda_2$ , which may be real and distinct, real and equal, or complex conjugates. If  $\lambda_1$  and  $\lambda_2$  are the roots of (7.9), then there is a solution  $(Y_1, Z_1)$  of (7.8) corresponding to the root  $\lambda_1$ , and a solution  $(Y_2, Z_2)$  of (7.8) corresponding to the root  $\lambda_2$ . These, in turn, give us two solutions

$$\begin{aligned} y &= Y_1 e^{\lambda_1 t}, & z &= Z_1 e^{\lambda_1 t} \\ y &= Y_2 e^{\lambda_2 t}, & z &= Z_2 e^{\lambda_2 t} \end{aligned}$$

of the system (7.6). We note that if  $\lambda$  is a root of (7.9), then equations (7.8) reduce to a single equation. Thus we may make an arbitrary choice for one of  $Y, Z$  and the other is then determined by (7.8).

Because the system (7.6) is linear, it is easy to see that every constant multiple of a solution of (7.8) is also a solution and also that the sum of two solutions of (7.8) is also a solution. It is possible to show that if we have two different solutions of the system (7.6), then every solution of (7.6) is a constant multiple of the first solution plus a constant multiple of the second solution. By “different” we mean that neither solution is a constant multiple of the other. Here, if the roots  $\lambda_1$  and  $\lambda_2$  of the characteristic equation (7.9) are distinct, we do have two different solutions of (7.6), and then every solution of system (7.6) has the form

$$\begin{aligned} y &= K_1 Y_1 e^{\lambda_1 t} + K_2 Y_2 e^{\lambda_2 t}, \\ z &= K_1 Z_1 e^{\lambda_1 t} + K_2 Z_2 e^{\lambda_2 t} \end{aligned} \quad (7.10)$$

for some constants  $K_1$  and  $K_2$ . The form (7.10) with two arbitrary constants  $K_1$  and  $K_2$  is called the *general solution* of the system (7.6). If initial values

$y(0)$  and  $z(0)$  are specified, these two initial values may be used to determine values for the constants  $K_1$  and  $K_2$ , and thus to obtain a *particular solution* in the family (7.10).

**Example 1.** Find the general solution of the system

$$y' = -y - 2z, \quad z' = y - 4z$$

and also the solution such that  $y(0) = 3$ ,  $z(0) = 1$ .

**Solution:** Here  $a = -1$ ,  $b = -2$ ,  $c = 1$ ,  $d = -4$ , so that  $a + d = -5$ ,  $ad - bc = 6$ , and the characteristic equation is  $\lambda^2 + 5\lambda + 6 = 0$ , with roots  $\lambda_1 = -2$ ,  $\lambda_2 = -3$ . With  $\lambda = -2$ , both equations of the algebraic system (7.8) are  $Y - 2Z = 0$ , and we may take  $Y = 2$ ,  $Z = 1$ . The resulting solution of (7.6) is  $y = 2e^{-2t}$ ,  $z = e^{-2t}$ . With  $\lambda = -3$ , both equations of the algebraic system (7.8) are  $Y - Z = 0$ , and we may take  $Y = 1$ ,  $Z = 1$ . The resulting solution of (7.6) is  $y = e^{-3t}$ ,  $z = e^{-3t}$ . Thus the general solution of the system is

$$y = 2K_1e^{-2t} + K_2e^{-3t}, \quad z = K_1e^{-2t} + K_2e^{-3t}.$$

To satisfy the initial conditions, we substitute  $t = 0$ ,  $y = 3$ ,  $z = 1$  into this form, obtaining a pair of equations  $2K_1 + K_2 = 3$ ,  $K_1 + K_2 = 1$ . We may subtract the second of these from the first to give  $K_1 = 2$ , and then  $K_2 = -1$ . This gives, as the solution of the initial value problem,

$$y = 4e^{-2t} - e^{-3t}, \quad z = 2e^{-2t} - e^{-3t}. \quad \square$$

In the language of linear algebra,  $\lambda$  is an *eigenvalue* and  $(Y, Z)$  is a corresponding *eigenvector* of the  $2 \times 2$  matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Note that  $a + d$  is the *trace*, denoted by  $trA$ , and  $ad - bc$  is the *determinant*, denoted by  $detA$ , of the matrix  $A$ . The sum of the eigenvalues of  $A$  is the trace, and the product of the eigenvalues is the determinant. This is seen easily by writing (7.9) as

$$\lambda^2 - trA\lambda + detA = 0$$

We will consider vectors to be column vectors, that is, matrices with two rows and one column. Suppose the matrix  $A$  has two distinct real eigenvalues  $\lambda_1, \lambda_2$  with corresponding eigenvectors

$$\begin{bmatrix} Y_1 \\ Z_1 \end{bmatrix} \qquad \qquad \qquad \begin{bmatrix} Y_2 \\ Z_2 \end{bmatrix}$$

respectively (this approach can also be used if  $\lambda_1 = \lambda_2$  provided there are two independent corresponding eigenvectors). We define the matrix

$$P = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix}$$

and then if we make the change of variable

$$\begin{bmatrix} y \\ z \end{bmatrix} = P \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

the system (7.6) is transformed to the system

$$\begin{aligned} u' &= \lambda_1 u \\ v' &= \lambda_2 v \end{aligned} \tag{7.11}$$

Since the system (7.11) is uncoupled, it may easily be solved explicitly to give

$$u = K_1 e^{\lambda_1 t}, \quad v = K_2 e^{\lambda_2 t}$$

We may then transform back to the original variables  $y, z$  to give the solution (7.10). The linear algebra approach gives the same result that was obtained by our initial approach of trying to find exponential solutions. Since it is constructive, it does not make use of the result which we stated without proof that if we have two different solutions of the system (7.6), then every solution of (7.6) is a constant multiple of the first solution plus a constant multiple of the second solution.

If the characteristic equation (7.9) has a double root, the method we have used gives only one solution of the system (7.6), and we need to find a second solution in order to form the general solution. In order to see what form the second solution must have we consider the special case  $c = 0$  of (7.6). Then (7.6) becomes

$$\begin{aligned} y' &= ay + bz, \\ z' &= dz. \end{aligned} \tag{7.12}$$

Then the eigenvalues of the corresponding matrix  $A$  are  $a$  and  $d$ ; if  $a = d$  the characteristic equation has a double root. However, the assumption  $c = 0$  means that we can solve the system (7.12) recursively. Every solution of the second equation in (7.12) has the form

$$z = Z e^{dt}$$

and we may substitute this into the first equation to give

$$y' = ay + bZ e^{dt}$$

This first order linear differential equation is easily solved. We multiply the equation by  $e^{-at}$  to give

$$y' e^{-at} - ay e^{-at} = bZ e^{(d-a)t} \tag{7.13}$$

If  $a \neq d$ , integration of (7.13) gives

$$y e^{-at} = \frac{bZ}{d-a} e^{(d-a)t} + Y$$

where  $Y$  is a constant of integration. From this we obtain the solution

$$\begin{aligned}y &= \frac{bZ}{d}e^{dt} + Ye^{at} \\z &= Ze^{dt}\end{aligned}$$

which is equivalent to the solution (7.10) obtained earlier; note that here  $\lambda_1 = a, \lambda_2 = d$ .

A more interesting case arises when  $a = d$ , so that the two roots of the characteristic equation are equal. Then (7.13) becomes

$$y'e^{-at} - aye^{-at} = bZ \quad (7.14)$$

and integration gives

$$ye^{-at} = bZt + Y$$

where  $Y$  is a constant of integration. From this we obtain the solution

$$\begin{aligned}y &= bZte^{at} + Ye^{at} \\z &= Ze^{at}\end{aligned}$$

This suggests that if there is a double root  $\lambda$  of the characteristic equation the needed second solution of (7.6) will contain terms  $te^{\lambda t}$  as well as  $e^{\lambda t}$ . It is possible to show (and the reader can verify) that if  $\lambda$  is a double root of (7.9), we must have  $\lambda = \frac{a+d}{2}$  and in addition to the solution  $y = Y_1e^{\lambda t}, z = Z_1e^{\lambda t}$  of (7.6) there is a second solution, of the form

$$y = (Y_2 + Y_1t)e^{\lambda t}, \quad z = (Z_2 + Z_1t)e^{\lambda t}$$

where  $Y_1, Z_1$  are as in (7.8) and  $Y_2, Z_2$  are given by

$$\begin{aligned}(a - \lambda)Y_2 + bZ_2 &= Y_1, \\cY_2 + (d - \lambda)Z_2 &= Z_1.\end{aligned}$$

Thus the general solution of the system (7.6) in the case of equal roots is

$$\begin{aligned}y &= (K_1Y_1 + K_2Y_2)e^{\lambda t} + K_2Y_1te^{\lambda t}, \\z &= (K_1Z_1 + K_2Z_2)e^{\lambda t} + K_2Z_1te^{\lambda t}.\end{aligned} \quad (7.15)$$

Note that if (7.9) has a single root and  $b = 0$ , then  $\lambda = a = d$ , and we have  $Y_1 = 0, Z_1 = cY_2$ , so that the general solution becomes

$$y = K_3e^{\lambda t}, \quad z = K_4e^{\lambda t} + cK_3te^{\lambda t},$$

where  $K_3 \equiv K_2Y_2$  and  $K_4 \equiv cK_1Y_2 + K_2Z_2$  are arbitrary constants. Likewise if (7.9) has a single root and  $c = 0$ , then  $\lambda = a = d, Z_1 = 0, Y_1 = bZ_2$ , and the general solution reduces to

$$y = K_5e^{\lambda t} + bK_6te^{\lambda t}, \quad z = K_6e^{\lambda t}$$

with arbitrary constants  $K_5 \equiv bK_1Z_2 + K_2Y_2$  and  $K_6 \equiv K_2Z_2$ . If  $b$  and  $c$  are both zero, the system becomes uncoupled

$$y' = ay, \quad z' = dz$$

and is easily solved by integration to give  $y = K_1e^{at}$ ,  $z = K_2e^{dt}$ . This is the solution in all cases, regardless of whether the characteristic equation has distinct roots or equal roots.

**Example 2.** Find the general solution of the system

$$y' = z, \quad z' = -y + 2z$$

and also the solution such that  $y(0) = 2$ ,  $z(0) = 3$ .

**Solution:** Since  $a = 0$ ,  $b = 1$ ,  $c = -1$ ,  $d = 2$ , we have  $a + d = 2$ ,  $ad - bc = 1$ . The characteristic equation is  $\lambda^2 - 2\lambda + 1 = 0$ , with a double root  $\lambda = 1$ . With  $\lambda = 1$ , both equations of the system (7.8) are  $Y - Z = 0$ , and we may take  $Y = 1$ ,  $Z = 1$  to give the solution  $y = e^t$ ,  $z = e^t$  of (7.6). Substituting these values into the equations for  $Y_2$ ,  $Z_2$ , we find that they reduce to the single equation  $-Y_2 + Z_2 = 1$ , so we may take  $Y_2 = 0$ ,  $Z_2 = 1$ . Equations (7.15) now give the general solution  $y = K_1e^t + K_2te^t$ ,  $z = (K_1 + K_2)e^t + K_2te^t$ . To find the solution with  $y(0) = 2$ ,  $z(0) = 3$ , we substitute  $t = 0$ ,  $y = 2$ ,  $z = 3$  into this form, obtaining the pair of equations  $K_1 = 2$ ,  $K_1 + K_2 = 3$ . Then  $K_2 = 1$ , and the solution satisfying the initial conditions is  $y = 2e^t + te^t$ ,  $z = 3e^t + te^t$ .  $\square$

From a linear algebra perspective we may make a change of variable of the system (7.6) just as in the case where we had two eigenvectors. We define the matrix

$$P = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix}$$

with

$$\begin{bmatrix} Y_1 \\ Z_1 \end{bmatrix}$$

an eigenvector of the matrix  $A$  as before, but with

$$\begin{bmatrix} Y_2 \\ Z_2 \end{bmatrix}$$

chosen arbitrarily, except that the determinant of  $P$  must be non-zero. Then under the change of variable

$$\begin{bmatrix} y \\ z \end{bmatrix} = P \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 \\ Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

the system (7.6) is transformed to a system of the form (7.6) with  $c = 0$  which may be solved explicitly, as we have seen.

A particularly clever choice of

$$\begin{bmatrix} Y_2 \\ Z_2 \end{bmatrix}$$

is the solution of the system of linear equations

$$\begin{aligned} aY_2 + bZ_2 &= \lambda Y_2 + Y_1 \\ cY_2 + dZ_2 &= \lambda Z_2 + Y_2 \end{aligned}$$

It is possible to show that this system always has a solution. With this choice of  $Y_2, Z_2$ , the system (7.6) is transformed to

$$\begin{aligned} u' &= \lambda_1 u + v \\ v' &= \lambda_2 v \end{aligned} \tag{7.16}$$

If we solve the system (7.16) recursively by the method used to solve the system (7.12) and the solution is then transformed back to the original variables, we obtain the solution (7.15).

Another complication arises if the characteristic equation (7.9) has complex roots. While the general solution of (7.6) is still given by (7.10), in this case the constants  $\lambda_1$  and  $\lambda_2$  are complex, and the solution is in terms of complex functions. Complex exponentials, however, can be defined with the aid of trigonometric functions ( $e^{i\theta} \equiv \cos \theta + i \sin \theta$  for real  $\theta$ ), so it is still possible to give the solution of (7.6) in terms of real exponential and trigonometric functions. In this case, if the characteristic equation (7.9) has conjugate complex roots  $\lambda = \alpha \pm i\beta$ , where  $\alpha$  and  $\beta$  are real and  $\beta > 0$ , equations (7.10) become

$$\begin{aligned} y &= (K_1 Y_1 + K_2 Y_2)e^{\alpha t} \cos \beta t + i(K_1 Y_1 - K_2 Y_2)e^{\alpha t} \sin \beta t, \\ z &= (K_1 Z_1 + K_2 Z_2)e^{\alpha t} \cos \beta t + i(K_1 Z_1 - K_2 Z_2)e^{\alpha t} \sin \beta t. \end{aligned}$$

We can eliminate the imaginary coefficients by defining  $Q_1 \equiv i(K_1 - K_2)$ ,  $Q_2 \equiv K_1 + K_2$ , and taking  $(a - \lambda_i)Y_i + bZ_i = 0$  for  $i = 1, 2$  from (7.10) to arrive at the form

$$\begin{aligned} y &= Q_1 b e^{\alpha t} \sin \beta t + Q_2 b e^{\alpha t} \cos \beta t, \\ z &= -[Q_1(a - \alpha) - Q_2 \beta]e^{\alpha t} \sin \beta t + [Q_1 \beta - Q_2(a - \alpha)]e^{\alpha t} \cos \beta t. \end{aligned}$$

**Example 3.** Find the general solution of the system

$$y' = -2z, \quad z' = y + 2z$$

and also the solution with  $y(0) = -2$ ,  $z(0) = 0$ .

**Solution:** We have  $a = 0$ ,  $b = -2$ ,  $c = 1$ ,  $d = 2$ , and the characteristic equation is  $\lambda^2 - 2\lambda + 2 = 0$ , with roots  $\lambda = 1 \pm i$ . The general solution is then

$$y = -2K_1 e^t \sin t - 2K_2 e^t \cos t, \quad z = (K_1 - K_2)e^t \sin t + (K_1 + K_2)e^t \cos t.$$

To find the solution with  $y(0) = -2$ ,  $z(0) = 0$ , we substitute  $t = 0$ ,  $y = -2$ ,  $z = 0$  into this form, obtaining  $-2K_2 = -2$ ,  $K_1 + K_2 = 0$ , whose solution is  $K_1 = -1$ ,  $K_2 = 1$ . This gives the particular solution

$$y = 2e^t \sin t - 2e^t \cos t, \quad z = -2e^t \sin t. \quad \square$$

In many applications, especially in analyzing stability of an equilibrium, the precise form of the solution of a linear system is less important to us than the qualitative behavior of solutions. It will turn out that often the crucial question is whether all solutions of a linear system approach zero as  $t \rightarrow \infty$ . The nature of the origin as an equilibrium of the linear system (7.6) depends on the roots of the characteristic equation (7.9). If both roots of (7.9) are real and negative then the solutions of (7.6) are combinations of negative exponentials. This implies that all orbits approach the origin and the origin is asymptotically stable. If the roots of (7.9) are real and of opposite sign, then there are solutions which are positive exponentials and solutions which are negative exponentials. Thus there are solutions approaching the origin and solutions moving away from the origin, and the origin is unstable. If the roots of (7.9) are complex, the orbits approach the origin if their real part are negative. We conclude that the origin is an asymptotically stable equilibrium of (7.6) if both roots of (7.9) have negative real part and unstable if at least one root of (7.9) has positive real part.

Solution of systems of linear differential equations with constant coefficients can be carried out more economically with the use of vectors and matrices. A system

$$\begin{aligned} y_1' &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ y_2' &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\vdots \\ y_n' &= a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{aligned}$$

may be written in the form

$$y' = Ay, \tag{7.17}$$

where  $A$  is an  $n \times n$  matrix and  $y$  and  $y'$  are column vectors,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

We attempt to find solutions of the form  $y = e^{\lambda t}c$  with  $c$  a constant column vector. Substituting this form into (7.17) we obtain the condition

$$\lambda e^{\lambda t}c = Ae^{\lambda t}c,$$

or

$$Ac = \lambda c.$$

Thus  $\lambda$  be an eigenvalue of the matrix  $A$  and  $c$  must be a corresponding eigenvector. If the eigenvalues of  $A$  are distinct, this procedure generates  $n$  solutions of (7.17) and it can be shown that every solution is a linear combination of these solutions. This explains the exponential solutions that we obtained. If there are multiple eigenvalues of  $A$ , it may be necessary to obtain additional solutions, and these will be exponential functions multiplied by powers of  $t$ . If some eigenvalues are complex, the corresponding complex exponential solutions may be replaced by products of exponential functions and sines or cosines.

## 7.4 Stability of Equilibria

In order to apply the linearization theorem of Section 8.2 to questions of stability of an equilibrium, we must determine conditions under which all solutions of a linear system with constant coefficients

$$\begin{aligned}y' &= ay + bz, \\z' &= cy + dz\end{aligned}\tag{7.18}$$

approach zero as  $t \rightarrow \infty$ . As we saw in Section 8.3, the nature of the solutions of (7.18) is determined by the roots of the characteristic equation

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0.\tag{7.19}$$

If the roots  $\lambda_1$  and  $\lambda_2$  of (7.19) are real, then the solutions of (7.18) are made up of terms  $e^{\lambda_1 t}$  and  $e^{\lambda_2 t}$ , or  $e^{\lambda_1 t}$  and  $te^{\lambda_1 t}$  if the roots are equal. In order that all solutions of (7.18) approach zero, we require  $\lambda_1 < 0$  and  $\lambda_2 < 0$ , so that the terms will be negative exponentials. If the roots are complex conjugates,  $\lambda = \alpha \pm i\beta$ , then in order that all solutions of (7.18) approach zero, we require  $\alpha < 0$ . Thus if the roots of the characteristic equation have negative real part, all solutions of the system (7.18) approach zero as  $t \rightarrow \infty$ . In a similar manner, we may see that if a root of the characteristic equation has positive real part, then (7.18) has unbounded solutions.

It turns out, however, that it is not necessary to solve the characteristic equation in order to determine whether all solutions of (7.18) approach zero, as there is a useful criterion in terms of the coefficients of the characteristic equation. The basic result is that the roots of a quadratic equation  $\lambda^2 + a_1\lambda + a_2 = 0$  have negative real part if and only if  $a_1 > 0$  and  $a_2 > 0$ . Applying this to the characteristic equation (7.19) and the system (7.18), we obtain the following result for linear systems with constant coefficients:

**STABILITY THEOREM FOR LINEAR SYSTEMS:** Every solution of the linear system with constant coefficients (7.18) approaches zero as  $t \rightarrow \infty$  if and only if the *trace*  $a + d$  of the coefficient matrix of the system is negative

and the *determinant*  $ad - bc$  of the system's coefficient matrix is positive. If either the trace is positive or the determinant is negative, there is at least one unbounded solution.

**Example 1.** Determine whether all solutions tend to zero or whether there are unbounded solutions for each of the following systems:

$$(i) \quad u' = -u - 2v, \quad v' = u - 4v$$

$$(ii) \quad u' = v, \quad v' = -u - 2v$$

$$(iii) \quad u' = -2v, \quad v' = u + 2v$$

**Solution:** (i) The characteristic equation is  $\lambda^2 + 5\lambda + 6 = 0$ , with roots  $\lambda = -2$ ,  $\lambda = -3$ . Thus all solutions tend to zero. Alternatively, since the trace of the coefficient matrix is  $-5 < 0$  and the determinant is  $6 > 0$ , the stability theorem gives the same conclusion. For (ii), the characteristic equation is  $\lambda^2 + 2\lambda + 1 = 0$  with a double root  $\lambda = -1$ , and thus all solutions tend to zero. For (iii), the characteristic equation is  $\lambda^2 - 2\lambda + 2 = 0$ , and since the trace is positive, there are unbounded solutions. As we indicated in the previous section, we could also have drawn this conclusion from a phase portrait.  $\square$

If we apply the stability theorem for linear systems to the linearization

$$\begin{aligned} u' &= F_y(y_\infty, z_\infty)u + F_z(y_\infty, z_\infty)v, \\ v' &= G_y(y_\infty, z_\infty)u + G_z(y_\infty, z_\infty)v \end{aligned} \quad (7.20)$$

of a system

$$y' = F(y, z), \quad z' = G(y, z) \quad (7.21)$$

at an equilibrium  $(y_\infty, z_\infty)$ , we obtain the following result.

**EQUILIBRIUM STABILITY THEOREM:** Let  $(y_\infty, z_\infty)$  be an equilibrium of a system  $y' = F(y, z)$ ,  $z' = G(y, z)$ , with  $F$  and  $G$  twice differentiable. Then if

$$F_y(y_\infty, z_\infty) + G_z(y_\infty, z_\infty) < 0 \quad (7.22)$$

and

$$F_y(y_\infty, z_\infty)G_z(y_\infty, z_\infty) - F_z(y_\infty, z_\infty)G_y(y_\infty, z_\infty) > 0, \quad (7.23)$$

the equilibrium  $(y_\infty, z_\infty)$  is locally asymptotically stable. If either

$$F_y(y_\infty, z_\infty) + G_z(y_\infty, z_\infty) > 0$$

or

$$F_y(y_\infty, z_\infty)G_z(y_\infty, z_\infty) - F_z(y_\infty, z_\infty)G_y(y_\infty, z_\infty) < 0$$

the equilibrium  $(y_\infty, z_\infty)$  is unstable.

**Example 2.** Determine whether each equilibrium of the system

$$y' = z, \quad z' = 2(y^2 - 1)z - y$$

is locally asymptotically stable or unstable.

**Solution:** The equilibria are the solutions of  $z = 0$ ,  $2(y^2 - 1)z - y = 0$ , and thus the only equilibrium is  $(0,0)$ . Here  $F(y, z) = z$ , with partial derivatives 0 and 1 respectively, and  $G(y, z) = 2(y^2 - 1)z - y$ , with partial derivatives  $4yz - 1$  and  $2(y^2 - 1)$  respectively. Therefore the community matrix at the equilibrium is

$$\begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$$

with trace  $-2$  and determinant 1, as in Example 1(ii). Thus the equilibrium  $(0,0)$  is locally asymptotically stable.  $\square$

**Example 3.** Determine whether each equilibrium of the system

$$\begin{aligned} y' &= y(1 - 2y - z) \\ z' &= z(1 - y - 2z) \end{aligned}$$

is locally asymptotically stable or unstable.

**Solution:** The equilibria are the solutions of  $y(1 - 2y - z) = 0$ ,  $z(1 - y - 2z) = 0$ . One solution is  $(0,0)$ ; a second is the solution of  $y = 0$ ,  $1 - y - 2z = 0$ , which is  $(0, \frac{1}{2})$ ; a third is the solution of  $z = 0$ ,  $1 - 2y - z = 0$ , which is  $(\frac{1}{2}, 0)$ ; and a fourth is the solution of  $1 - 2y - z = 0$ ,  $1 - y - 2z = 0$ , which is  $(\frac{1}{3}, \frac{1}{3})$ . The community matrix at an equilibrium  $(y_\infty, z_\infty)$  is

$$\begin{bmatrix} 1 - 4y_\infty - z_\infty & -y_\infty \\ -z_\infty & 1 - y_\infty - 4z_\infty \end{bmatrix}.$$

At  $(0,0)$ , this matrix has trace 1 and determinant 1, and thus the equilibrium is unstable. At  $(0, \frac{1}{2})$ , this matrix has trace  $-\frac{1}{2}$  and determinant  $-\frac{1}{2}$ , and thus the equilibrium is unstable. At  $(\frac{1}{2}, 0)$ , this matrix has trace  $-\frac{1}{2}$  and determinant  $-\frac{1}{2}$ , and thus the equilibrium is unstable. At  $(\frac{1}{3}, \frac{1}{3})$ , this matrix has trace  $-\frac{4}{3}$  and determinant  $\frac{1}{3}$ , and thus this equilibrium is locally asymptotically stable.  $\square$

The careful reader will have noticed that, like the linearization theorem of Section 8.2, the equilibrium stability theorem given above has a hole of sorts in its result, in that the theorem says nothing about the stability of equilibria for which the trace and determinant lie on the boundary of conditions (7.22) and (7.23)—in other words, for which the linearization has solutions which do not approach zero as  $t \rightarrow \infty$  but stay bounded. The reason for this “hole” is that in such cases, the linearization does not give enough information to determine stability.

If all orbits beginning near an equilibrium remain near the equilibrium for  $t \geq 0$ , but some orbits do not approach the equilibrium as  $t \rightarrow \infty$ , the equilibrium

is said to be *stable*, or sometimes *neutrally stable*. If the origin is neutrally stable for the linearization at an equilibrium, then the equilibrium may also be neutrally stable for the nonlinear system. However, it is also possible for the origin to be neutrally stable for the linearization at an equilibrium, while the equilibrium is asymptotically stable or unstable. Thus neutral stability of the origin for a linearization at an equilibrium gives no information about the stability of the equilibrium.

We have seen in Section 7.5 that a solution of an autonomous first-order differential equation is either unbounded or approaches a limit as  $t \rightarrow \infty$ . For an autonomous system of two first-order differential equations, these same two possibilities exist. In addition, however, there is the possibility of an orbit which is a closed curve, corresponding to a periodic solution. Such an orbit is called a *periodic orbit* because it is traversed repeatedly.

There is a remarkable result which says essentially that these are the only possibilities.

**POINCARÉ-BENDIXSON THEOREM:** A bounded orbit of a system of two first-order differential equations which does not approach an equilibrium as  $t \rightarrow \infty$  either is a periodic orbit or approaches a periodic orbit as  $t \rightarrow \infty$ .

It is possible to show that a periodic orbit must enclose an equilibrium point in its interior. In many examples, there is an unstable equilibrium, and orbits beginning near this equilibrium spiral out towards a periodic orbit. A periodic orbit which is approached by other (non-periodic) orbits is called a *limit cycle*. One example of a limit cycle involves the system

$$\begin{aligned}y' &= y(1 - y^2 - z^2) - z, \\z' &= z(1 - y^2 - z^2) + y,\end{aligned}$$

which has its only equilibrium at the origin. The equilibrium is unstable, and Figure 7.1 illustrates the fact that all orbits not beginning at the origin spiral counterclockwise [in or out] toward the unit circle  $y^2 + z^2 = 1$ .

In many applications the functions  $y(t)$  and  $z(t)$  are restricted by the nature of the problem to non-negative values. For example, this is the case if  $y(t)$  and  $z(t)$  are population sizes. In such a case, only the first quadrant  $y \geq 0, z \geq 0$  of the phase plane is of interest. For a system

$$y' = F(y, z), \quad z' = G(y, z)$$

which has  $F(0, z) \geq 0$  for  $z \geq 0$  and  $G(y, 0) \geq 0$  for  $y \geq 0$ , then since  $y' \geq 0$  along the positive  $z$ -axis (where  $y = 0$ ) and  $z' \geq 0$  along the positive  $y$ -axis (where  $z = 0$ ), no orbit can leave the first quadrant by crossing one of the axes. The Poincaré-Bendixson Theorem may then be applied to orbits in the first quadrant. In such a case, the first quadrant is called an *invariant set*, a region with the property that orbits must remain in the region.

If instead  $F$  and  $G$  are identically zero along the respective [half-]axes, then  $y' = 0$  for  $\{y = 0, z \geq 0\}$ , and  $z' = 0$  for  $\{y \geq 0, z = 0\}$ . In this case, orbits

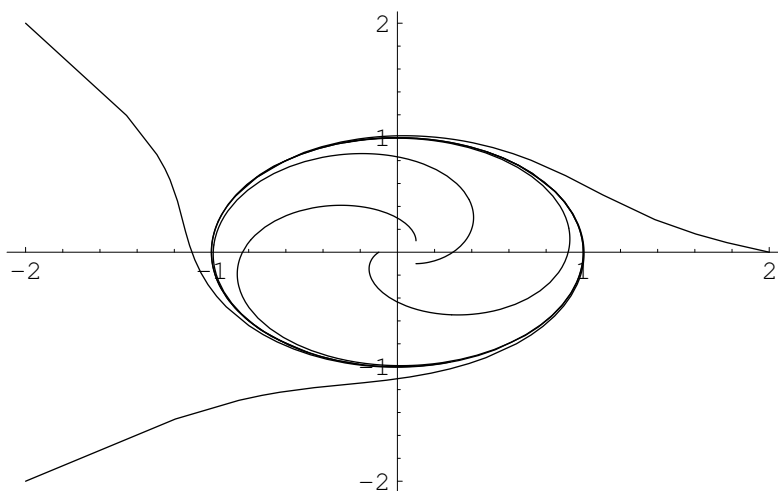


Figure 7.1: Trajectories approaching a limit cycle

which begin on an axis must remain on that axis, and orbits beginning in the interior of the first quadrant (with  $y(0) > 0$ ,  $z(0) > 0$ ) must remain in the interior of the first quadrant (i.e.,  $y(t) > 0$  and  $z(t) > 0$  for  $t \geq 0$ ). If there is no equilibrium in the first quadrant, there cannot be a periodic orbit, because a periodic orbit *must* enclose an equilibrium. Thus, if there is no equilibrium in the first quadrant every orbit must be unbounded.

**Example 4.** Show that every orbit in the region  $y > 0$ ,  $z > 0$  of the system

$$\begin{aligned} y' &= y(2 - y) - \frac{yz}{y+1}, \\ z' &= 4\frac{yz}{y+1} - z \end{aligned}$$

approaches a periodic orbit as  $t \rightarrow \infty$ .

**Solution:** We have  $y' = 0$  when  $y = 0$ , and  $z' = 0$  when  $z = 0$ , so orbits starting in the first quadrant remain in the first quadrant. Equilibria are the solutions of either  $y = 0$  or  $2 - y = \frac{z}{y+1}$ , and either  $z = 0$  or  $\frac{4y}{y+1} = 1$ . If  $y = 0$ , we must also have  $z = 0$ . If  $2 - y = \frac{z}{y+1}$ , we could have  $z = 0$ , which implies  $y = 2$ , or  $y = \frac{1}{3}$ , which implies  $z = \frac{20}{9}$ . Thus there are three equilibria, namely  $(0,0)$ ,  $(2,0)$ , and  $(\frac{1}{3}, \frac{20}{9})$ . By checking the values of the trace and determinant of the community matrix, which is

$$\begin{bmatrix} 2 - 2y_\infty - \frac{z_\infty}{(1+y_\infty)^2} & -\frac{y_\infty}{y_\infty+1} \\ \frac{4z_\infty}{(y_\infty+1)^2} & \frac{4y_\infty}{y_\infty+1} - 1 \end{bmatrix},$$

we may see that each of the three equilibria is unstable. In order to apply the Poincaré-Bendixson Theorem, we must show that every orbit starting in the first quadrant of the phase plane is bounded.

To show this, we might like to show that  $y'$  and  $z'$  are negative when  $y$  and/or  $z$  are sufficiently large, but a glance at the equations tells us this isn't necessarily so. Therefore we instead consider some positive combination of  $y$  and  $z$  whose time derivative does become negative far enough from the origin. In particular, consider the function  $V(y, z) = 4y + z$ . If an orbit is unbounded, then along this orbit the function  $V(y, z)$  must also be unbounded. The derivative of  $V(y, z)$  along an orbit is

$$\frac{d}{dt}V[y(t), z(t)] = 4y'(t) + z'(t) = 4y(2 - y) - z.$$

This is negative except in the bounded region defined by the inequality  $z < 4y(2 - y)$ . Therefore the function  $V(y, z)$  cannot become unbounded, because it is decreasing ( $dV/dt < 0$ ) whenever it becomes large ( $z > 4y(2 - y)$ , which is true, for example, whenever  $V > 9$ ). This proves that all orbits of the system are bounded. Now we may apply the Poincaré-Bendixson Theorem to see that every orbit approaches a limit cycle.  $\square$

## 7.5 Some Applications in Population Biology and Epidemiology

The formulation of models for two interacting species depends on the nature of the interaction as well as on the assumptions about the behavior of each population in the absence of the other population. In this section, we shall examine a model for species in a predator-prey relation and an epidemic model in which a population is divided into two classes.

### 7.5.1 Predator-prey systems

Let us consider the interaction of a prey species  $y$  and a predator species  $z$ . We assume that in the absence of predators the prey population would obey a logistic equation. We assume that the rate of prey consumption per predator increases with prey population size but is bounded as the prey population becomes unbounded, that is, that there is a maximum rate of consumption per predator no matter how plentiful the food supply. Beyond a certain point, the prey population no longer limits the resources of the predators. For example, we may assume that the rate of consumption of prey per predator has the form  $\frac{qy}{y+A}$  where  $q$  and  $A$  are positive constants. We assume that in the absence of prey the predator population would die out at an exponential rate. In the equation for  $z'$ , we also incorporate a term proportional to  $\frac{yz}{z+A}$ , representing the conversion of food (prey) into predator biomass. This leads us to a model of the form

$$\begin{aligned} y' &= ry \left(1 - \frac{y}{K}\right) - \frac{qyz}{y+A}, \\ z' &= sz \left(\frac{y}{y+A} - \frac{J}{J+A}\right). \end{aligned} \tag{7.24}$$

The term  $\frac{qyz}{y+A}$  in the first equation of (7.24) is called the *predator functional response*, and the term  $\frac{scy}{y+A}$  (replacing  $cyz$ ) in the second equation of (7.24) is called the *predator numerical response*; the constant  $\frac{s}{q}$  is the conversion efficiency of prey into predators.

Here we have rewritten the natural decay rate of the predator population in terms of  $J$ , which we can see from (7.24) is the minimum prey population required to sustain the predator population ( $z' \geq 0$ ). As  $J$  decreases, so does the rate at which the predator population would die out in the absence of the prey. In the following two examples, we shall see that the parameter  $J$  (or, equivalently,  $\mu$ ) is capable of changing the nature of the system's behavior.

**Example 1.** Determine the qualitative behavior of a predator-prey system modelled by the differential equations

$$\begin{aligned}y' &= y \left(1 - \frac{y}{30}\right) - \frac{yz}{y+10}, \\z' &= z \left(\frac{y}{y+10} - \frac{3}{5}\right).\end{aligned}$$

**Solution:** Equilibria are solutions of the pair of equations

$$\begin{aligned}y \left(1 - \frac{y}{30} - \frac{z}{y+10}\right) &= 0, \\z \left(\frac{y}{y+10} - \frac{3}{5}\right) &= 0.\end{aligned}\tag{7.25}$$

One equilibrium is given by  $y = 0$ ,  $z = 0$ . If  $z = 0$ , (7.25) implies

$$y \left(1 - \frac{y}{30}\right) = 0,$$

and thus another equilibrium is given by  $y = 30$ ,  $z = 0$ . An equilibrium with  $y$  and  $z$  both positive satisfies

$$1 - \frac{y}{30} - \frac{z}{y+10} = 0, \quad \frac{y}{y+10} = \frac{3}{5}.$$

The second of these equations is  $5y = 3y + 30$ , or  $y = 15$ , and substitution into the first equation gives  $1 - \frac{1}{2} - \frac{z}{25} = 0$ , or  $z = 12.5$ . Thus a third equilibrium is given by  $y = 15$ ,  $z = 12.5$ .

We now use the equilibrium stability theorem of Section 8.4 with

$$F(y, z) = y \left(1 - \frac{y}{30}\right) - \frac{yz}{z+10}, \quad G(y, z) = z \left(\frac{y}{y+10} - \frac{3}{5}\right).$$

Then the partial derivatives are

$$\begin{aligned}F_y(y, z) &= 1 - \frac{y}{15} - \frac{10z}{(y+10)^2}, & F_z(y, z) &= -\frac{y}{y+10}, \\G_y(y, z) &= \frac{10z}{(y+10)^2}, & G_z(y, z) &= \frac{y}{y+10} - \frac{3}{5}.\end{aligned}$$

At the equilibrium (0,0), the community matrix is

$$\begin{bmatrix} 1 & 0 \\ 0 & -\frac{3}{5} \end{bmatrix},$$

and thus the equilibrium is unstable. At the equilibrium (30,0), the community matrix is

$$\begin{bmatrix} -1 & -\frac{3}{20} \\ 0 & \frac{3}{20} \end{bmatrix},$$

and since its determinant is  $-\frac{3}{20} < 0$  this equilibrium is also unstable. At the equilibrium (15,12.5), the community matrix is

$$\begin{bmatrix} -\frac{1}{5} & -\frac{3}{5} \\ \frac{1}{5} & 0 \end{bmatrix},$$

and since this has trace  $-\frac{1}{5} < 0$  and determinant  $\frac{3}{25} > 0$ , this equilibrium is asymptotically stable. It is possible to show that every orbit with  $y(0) > 0$  and  $z(0) > 0$ , not just those which start close to (15,12.5), approaches this equilibrium. Thus predator and prey coexist here, with prey at half their natural carrying capacity.  $\square$

**Example 2.** Suppose now that the environment of the pond in Example 2 improves in such a way that the predator fish tend to live longer (or die off more slowly), so that the natural per capita mortality rate drops from  $\frac{3}{5}$  (in per time units) to  $\frac{1}{3}$ . Determine the qualitative behavior of this new predator-prey system, modelled by the differential equations

$$\begin{aligned} y' &= y \left( 1 - \frac{y}{30} \right) - \frac{yz}{y+10} \\ z' &= z \left( \frac{y}{y+10} - \frac{1}{3} \right). \end{aligned}$$

**Solution:** Equilibria are solutions of the pair of equations

$$\begin{aligned} y \left( 1 - \frac{y}{30} - \frac{z}{y+10} \right) &= 0 \\ z \left( \frac{y}{y+10} - \frac{1}{3} \right) &= 0. \end{aligned}$$

As in Example 1, there is an equilibrium at (0,0) and a second, predator-free ( $z = 0$ ) equilibrium with  $y = 30$ . An equilibrium with  $y$  and  $z$  both positive satisfies

$$1 - \frac{y}{30} - \frac{z}{y+10} = 0, \quad \frac{y}{y+10} = \frac{1}{3}.$$

The second of these equations is  $3y = y + 10$ , or  $y = 5$ , and substitution into the first equation gives  $1 - \frac{5}{30} - \frac{z}{15} = 0$ , or  $z = 12.5$ . Thus a third equilibrium, representing coexistence, is given by  $y = 5$ ,  $z = 12.5$ .

The calculation of the community matrix is the same as in Example 1, except that  $\frac{2}{5}$  is replaced by  $\frac{1}{3}$  in  $G(y, z)$ . Thus the community matrix at  $(0,0)$  is

$$\begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{3} \end{bmatrix},$$

and this equilibrium is unstable. The community matrix at  $(30,0)$  is

$$\begin{bmatrix} -1 & -\frac{3}{4} \\ 0 & \frac{5}{12} \end{bmatrix},$$

and since this has determinant  $-\frac{5}{12} < 0$ , this equilibrium is unstable. The community matrix at  $(5,12.5)$  is

$$\begin{bmatrix} \frac{1}{5+9} & -\frac{1}{3} \\ \frac{5+9}{9} & 0 \end{bmatrix}.$$

Since this matrix has positive trace, the equilibrium  $(5,12.5)$  is also unstable, and the system has no asymptotically stable equilibrium. In order to show that all orbits in the first quadrant are bounded, we apply the technique introduced in Example 4, Section 8.4, adding the two equations of the model to obtain

$$(y+z)' = y \left(1 - \frac{y}{30}\right) - \frac{z}{3}.$$

Thus  $y+z$  is decreasing except in the bounded region defined by  $\frac{z}{3} < y \left(1 - \frac{y}{30}\right)$ . In order for an orbit to be unbounded,  $y+z$  must be unbounded, and, as in Example 4, Section 8.4, this is impossible since  $y+z$  is decreasing whenever  $y+z$  is large. Thus all orbits in the first quadrant are bounded, and the Poincaré-Bendixson Theorem may be applied to show that there must be a limit cycle (with the equilibrium  $(5,12.5)$  in its interior) to which every orbit tends. Thus the two species co-exist, but their population sizes fluctuate periodically. Some orbits are shown in Figure 7.2.  $\square$

Examples 1 and 2 show that for a model of the form (7.24) there may be periodic orbits, or every orbit may approach an equilibrium. Which behavior occurs depends on the values of the parameters in the model rather than on the form of the model.

## 7.5.2 An epidemiological model

In Section 7.4 we considered a model for the spread of an infectious disease in which a population was divided into susceptibles and infectives; the underlying assumptions were that there was a rate of contracting the infection which depended on the number of susceptibles and the number of infectives, that there was a rate of recovery depending on the number of infectives, and that on recovery infectives returned to the susceptible class. In other words, it was

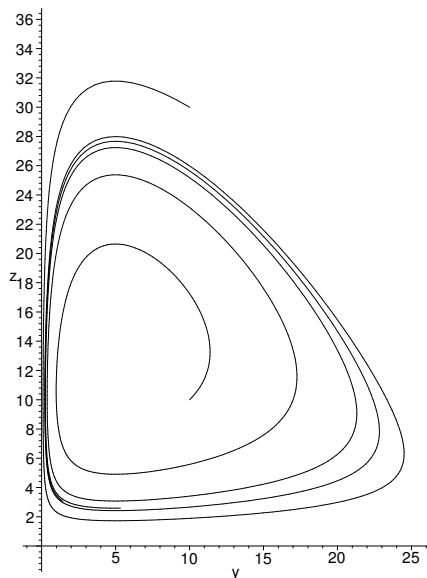


Figure 7.2: Some orbits of the system in Example 2

assumed that there was no immunity against re-infection after recovery from the infection.

In this example, we shall consider some models for the spread of infectious diseases which include a third class, of removed members. Many diseases, especially diseases caused by viral agents, including smallpox, measles, and rubella (German measles), provide immunity against re-infection.

We let  $S(t)$  denote the number of susceptibles,  $I(t)$  the number of infectives, and  $R(t)$  the number of removed members. We assume that the population has constant total size  $K$ , so that  $S(t) + I(t) + R(t) = K$ . We will derive differential equations expressing the rate of change of the size of each of the three classes, but in each case one of the equations may be eliminated since we can use the above relation to find  $S$ ,  $I$  or  $R$  in terms of the other two. Thus we will obtain a system of two differential equations to describe the spread of diseases for which there is a removed class.

A model was proposed by W. O. Kermack and A. G. McKendrick [W. O. Kermack & A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. London* 115 (1927), 700–721] to explain the rapid rise and fall of cases frequently observed in epidemics, including the Great Plague of 1665–66 in England, cholera in London in 1865, and plague in Bombay in 1906. This model is

$$\begin{aligned} S' &= -\beta SI, \\ I' &= \beta SI - \gamma I, \\ R' &= \gamma I. \end{aligned} \tag{7.26}$$

The only difference from the model of Section 7.4.1 (Equation (6.25)) is that the term  $\gamma I$  now represents a rate of transition from the class  $I$  to the class  $R$ , instead of a rate of return to the class  $S$ . The rate of recoveries in unit time is  $\gamma I$ , and the rate of transmission of infection from infectives to susceptibles is  $\beta SI$ . Note that this model is only appropriate if the duration of an outbreak is short enough that demographics (natural births and deaths) can be ignored.

We consider the model as a system of two equations, viewing  $R$  as determined by  $S$  and  $I$ ,  $R = K - S - I$ , since the first two equations do not involve  $R$ :

$$\begin{aligned} S' &= -\beta SI, \\ I' &= \beta SI - \gamma I \end{aligned} \tag{7.27}$$

The equilibria of the system (7.27) are the solutions of the pair of equations  $\beta SI = 0$ ,  $\beta SI - \gamma I = 0$ . The first of these implies that either  $S = 0$  or  $I = 0$ . If  $S = 0$ , the second equation is satisfied only if  $I = 0$ , while if  $I = 0$  the second equation is satisfied for every  $S$ . Thus there is a line of equilibria  $(S_\infty, 0)$  with  $S_\infty$  arbitrary,  $0 \leq S_\infty \leq K$ . If we compute the linearization of the system (7.27) at an equilibrium  $(S_\infty, 0)$ , we obtain

$$\begin{aligned} u' &= -\beta S_\infty v, \\ v' &= (\beta S_\infty - \gamma) v, \end{aligned}$$

and the linearization theorem of Section 8.4 cannot be applied.

In order to obtain an understanding of the qualitative behavior of solutions of the system (7.27), we observe from (7.27) that  $S' < 0$  whenever  $S > 0$ ,  $I > 0$ . This means that the function  $S(t)$  decreases for all  $t$ . In addition,  $I' < 0$  whenever  $I > 0$ ,  $\beta S < \gamma$ , while  $I' < 0$  if  $I < 0$ ,  $\beta S > \gamma$ . Thus if  $S(0) < \frac{\gamma}{\beta}$ ,  $S(t)$  remains less than  $\frac{\gamma}{\beta}$  for all  $t$ , and  $I(t)$  decreases to zero as  $t$  increases. However, if  $S(0) > \frac{\gamma}{\beta}$ , then  $I(t)$  increases so long as  $\beta S > \gamma$ , and thus  $I(t)$  increases initially before decreasing to zero. We think of introducing a small number of infectives into a susceptible population so that  $I(0) = \epsilon > 0$ ,  $S(0) = K - \epsilon$ . Then if  $\beta K < \gamma$ ,  $I(t)$  decreases monotonically to zero and the infection dies out. On the other hand, if  $\beta K > \gamma$ , an epidemic occurs, as  $S(0) > \frac{\gamma}{\beta}$  (for  $\epsilon$  small), so  $I(t)$  increases to a maximum and then decreases to zero. This is another threshold theorem of Kermack and McKendrick with the threshold quantity  $\beta K / \gamma$ . This threshold quantity distinguishes between two possible behaviors just like the threshold quantity in Section 8.4, but the possible behaviors are not the same as in Section 8.4.

One might suppose that the reason for the eventual disappearance of the infection in the epidemic case is that all susceptibles become infected, but observations of epidemics indicate that this is not the case. The model (7.27) agrees with observation in that it implies that the limiting value  $S(\infty) = \lim_{t \rightarrow \infty} S(t)$  of every solution of the system (7.27) obeys  $S(\infty) > 0$ . We may see this by

calculating

$$\begin{aligned}\frac{d}{dt} \left[ S(t) + I(t) - \frac{\gamma}{\beta} \ln S(t) \right] &= S'(t) + I'(t) - \frac{\gamma}{\beta} \frac{S'(t)}{S(t)} \\ &= -\beta SI + [\beta SI - \gamma I] - \frac{\gamma}{\beta} (-\beta I) = 0\end{aligned}$$

(motivated by observing that  $S' + I' = -\gamma I$ , and finding a function of  $S$  and  $I$  whose time derivative is  $\gamma I$ ). Thus  $S(t) + I(t) - \frac{\gamma}{\beta} \ln S(t)$  is a constant. Since  $I(\infty) = 0$  and  $I(0) \approx 0$ , this gives

$$S(\infty) - \frac{\gamma}{\beta} \ln S(\infty) = S(0) - \frac{\gamma}{\beta} \ln S(0).$$

It follows that

$$S(0) - S(\infty) = \frac{\gamma}{\beta} [\ln S(0) - \ln S(\infty)] = \frac{\gamma}{\beta} \ln \frac{S(0)}{S(\infty)}$$

and therefore

$$\frac{\beta}{\gamma} = \frac{\ln \left[ \frac{S(0)}{S(\infty)} \right]}{S(0) - S(\infty)}. \quad (7.28)$$

The quantity  $\beta S(0)/\gamma$  is known as the *contact number*. Not only does (7.28) imply  $S(\infty) > 0$ , but it also gives a means of estimating the contact rate  $\beta$ , which generally can not be measured directly. By making a serological survey (testing for immune responses in the blood) in the population before and after an epidemic, one may estimate  $S(0)$  and  $S(\infty)$ , and then (7.28) gives  $\beta/\gamma$ . If the mean infective period  $1/\gamma$  is known as well, then  $\beta$  can be calculated. The contact rate  $\beta$  depends on the disease as well as on other factors such as the rate of mixing in the population.

For example, the village of Eyam in England maintained isolation from other villages during the Great Plague of 1665–66, and its population decreased from 350 to 83 during the course of the epidemic. There is reason to believe that there were actually two separate epidemics, the first of which reduced the susceptible population to 254. By substituting  $S(0) = 254$ ,  $S(\infty) = 83$  into (7.28), we obtain

$$\frac{\beta}{\gamma} = \frac{\ln \frac{254}{83}}{254 - 83} = 6.54 \times 10^{-3}.$$

The infective period was 11 days, or 0.3667 months. Using a month as the unit of time we obtain the estimate  $\beta = 0.0178$ . This data, with 7 initial infectives, gives the phase portrait of Figure 7.3, traversed from right to left as time progressed and the number of susceptibles decreased. Note that in this case infected individuals were removed to the  $R$  class through death for the most part, rather than recovery with immunity. Our simple model (7.27) still describes this process, however different the interpretation may be, as the  $R$  class of the model simply includes individuals no longer involved in the spread of the disease.

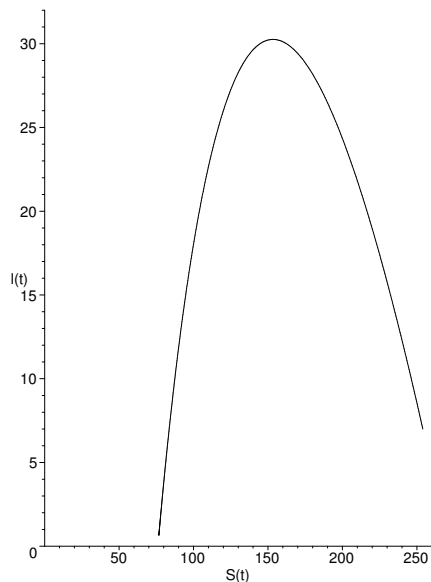


Figure 7.3: A phase portrait for model (7.27)

The criterion  $\beta K/\gamma > 1$  for the establishment of a disease can also be expressed as the requirement that the susceptible population density exceeds a certain critical value  $\frac{\gamma}{\beta}$ . For fox rabies in Europe, observations indicate a critical population density of approximately 1 fox/km<sup>2</sup>; rabies dies out in regions which are more sparsely populated. This data together with the average life expectancy of 5 days for a rabid fox gives the estimate  $\beta \approx 72$  km<sup>2</sup>/fox year.

In order to avoid an epidemic, it is necessary to reduce below 1 the quantity  $\beta K/\gamma$ , which is called the *basic reproduction number* and often denoted by  $R_0$ . This may sometimes be achieved by immunization, which has the effect of transferring members from the susceptible class to the removed class and thus reducing  $S(0)$ . If we immunize a fraction  $p$  of the susceptible population, we would replace  $K$  by  $(1-p)K$ , and this would give a basic reproductive number  $\beta K(1-p)/\gamma$ . In order to make this basic reproductive number less than 1, we require  $\beta K(1-p)/\gamma < 1$ . This is equivalent to  $1-p < \frac{\gamma}{\beta K}$ , or  $p > 1 - \frac{\gamma}{\beta K}$ .

A population is said to have *herd immunity* if a sufficiently large fraction has been immunized to reduce the basic reproductive number below 1 and thus assure that the disease will not spread if an infective is introduced into the population. The only infectious disease for which this has actually been achieved worldwide is smallpox. For measles, epidemiological data in the U.S. indicates a basic reproductive number ranging from 5.4 to 6.3 in rural areas, requiring vaccination of 81.5% to 84.1% of the population to achieve herd immunity. In urban areas, the basic reproductive number ranges from 8.3 to 13.0, requiring vaccination of 88.0% to 92.3% of the population. As measles vaccination is only about 95% effective (for vaccination at age 15 months) and not all people are

willing to allow vaccination, it is impossible in practice to achieve herd immunity. For smallpox, the basic reproductive number is about 5, requiring 80% vaccination to achieve herd immunity. This is feasible because the consequences of smallpox are dire enough to encourage immunization.

It may be important to know the maximum number of infectives at any given time, for example to be able to arrange enough facilities for isolation and treatment. From the model (7.27), we know that the maximum of  $I$  occurs when  $S = \frac{\gamma}{\beta}$ , and also that the quantity  $S + I - \frac{\gamma}{\beta} \ln S$  is constant. Thus if  $t^*$  is the time when  $S = \frac{\gamma}{\beta}$ , we have

$$S(t^*) + I(t^*) - \frac{\gamma}{\beta} \ln S(t^*) = S(0) + I(0) - \frac{\gamma}{\beta} \ln S(0)$$

and

$$\frac{\gamma}{\beta} + I(t^*) - \frac{\gamma}{\beta} \ln \frac{\gamma}{\beta} = S(0) - \frac{\gamma}{\beta} \ln S(0).$$

From this we conclude that  $I(t^*)$ , the maximum number of infectives, is given by

$$I(t^*) = S(0) - \frac{\gamma}{\beta} \ln S(0) + \frac{\gamma}{\beta} \ln \frac{\gamma}{\beta} - \frac{\gamma}{\beta} = S(0) - \frac{\gamma}{\beta} - \frac{\gamma}{\beta} \ln[\beta S(0)/\gamma]. \quad (7.29)$$

For the Great Plague in Eyam, this gives a maximum infective population of 30.4, confirmed by the phase portrait of Figure 7.3.

## 7.6 Some exercises

In each of Exercises 1–3, describe the orbits of the given system.

1.  $y' = yz^2, z' = zy^2$
2.  $y' = e^{-z}, z' = e^y$
3.  $y' = ye^z, z' = yze^{-y}$

In each of Exercises 4–8, find the linearization of the given system at each equilibrium.

4.  $y' = y - z, z' = y + z - 2$
5.  $y' = z, z' = y + z - 1$
6.  $y' = z + 1, z' = y^2 + z$
7.  $y' = z^2 - 8y, z' = y - 2$
8.  $y' = e^{-z}, z' = e^y$

In each of Exercises 9–12, for each equilibrium of the given system determine whether the equilibrium is asymptotically stable or unstable.

9.  $y' = -2y, z' = -z$   
 10.  $y' = y, z' = -z$   
 11.  $y' = -y, z' = y^2 - z$   
 12.  $y' = y + z, z' = z - 1$

In each of Exercises 13–20, find the general solution of the given system by analytic solution, use a computer algebra system to examine the behavior of solutions, and classify the origin as a node, saddle point, centre, or spiral point.

13.  $y' = y + 5z, z' = y - 3z$   
 14.  $y' = y - z, z' = z$   
 15.  $y' = 4y, z' = 2y + 4z$   
 16.  $y' = y - z, z' = 4y - 3z$   
 17.  $y' = y + 2z, z' = -3y + 6z$   
 18.  $y' = 3y + 5z, z' = -5y + 3z$   
 19.  $y' = z, z' = -y$   
 20.  $y' = y + z, z' = z$
21. Obtain the general solution of the system  $y' = ay, z' = cy + az$  by solving  $y' = ay$ , substituting the result into  $z' = cy + az$  and solving.
22. \* Consider the system

$$\begin{aligned}y' &= y - z, \\z' &= 3y - 3z,\end{aligned}$$

for which the condition  $ad - bc \neq 0$  is violated.

- (a) Show that every point of the line  $z = y$  is an equilibrium.  
 (b) Use the fact that  $z' = 3y'$  to deduce that  $z = 3y + c_1$  for some constant  $c_1$ .  
 (c) Use the result of part (b) to eliminate  $z$  from the system and obtain a first-order linear differential equation for  $y$ .  
 (d) Solve for  $y$  and obtain the solution  $y = -\frac{c_1}{2} + c_2e^{-2t}$ ,  $z = -\frac{c_1}{2} + 3c_2e^{-2t}$ .  
 (e) Show that as  $t \rightarrow \infty$  every orbit approaches the point  $(-\frac{c_1}{2}, -\frac{c_1}{2})$  on the line of equilibria, and that the slope of the line joining this point to any point on the orbit is the constant 3.

- (f) Use the information obtained to sketch the phase portrait of the system.

In Exercises 23–26, for each equilibrium of the given system determine whether the equilibrium is asymptotically stable or unstable.

23.  $y' = y - z, z' = y + z - 2$

24.  $y' = z, z' = y + z - 1$

25.  $y' = z + 1, z' = y^2 + z$

26.  $y' = e^{-z}, z' = e^y$



Part IV

**FURTHER TOPICS IN  
CALCULUS**



## Chapter 8

# Double Integrals

### 8.1 Double integrals over a rectangle

In elementary calculus, the definition of the definite integral is motivated by the idea of the area below a curve. Now we wish to introduce the concept of the double integral, motivated by the idea of the volume below a surface. We wish to describe the volume below a surface given by a function  $z = F(x, y)$  of two variables above a region  $R$  of the  $xy$ -plane. In this section we consider the special case in which the region  $R$  is the rectangle defined by the inequalities  $a \leq x \leq b$ ,  $c \leq y \leq d$ , where  $a, b, c, d$  are given constants with  $a < b$  and  $c < d$ . In the next section we shall extend the concept to more general plane regions  $D$ .

First we partition the rectangle  $R$  into subrectangles by partitioning the intervals  $a \leq x \leq b$  and  $c \leq y \leq d$ . We partition the interval  $a \leq x \leq b$  into  $m$  subintervals by defining

$$x_0 = a < x_1 < x_2 < \cdots < x_m = b.$$

The  $i$ -th interval is  $x_{i-1} \leq x \leq x_i$  [ $i = 1, 2, \dots, m$ ]. Similarly, we partition the interval  $c \leq y \leq d$  into  $n$  subintervals by defining

$$y_0 = c < y_1 < y_2 < \cdots < y_n = d.$$

The  $j$ -th interval is  $y_{j-1} \leq y \leq y_j$  [ $j = 1, 2, \dots, n$ ]. We now have  $mn$  rectangles, described by

$$r_{ij} = \{(x, y) \mid x_{i-1} \leq x \leq x_i, y_{j-1} \leq y \leq y_j\}.$$

We let

$$\Delta x_i = x_i - x_{i-1} \quad (i = 1, 2, \dots, m), \quad \Delta y_j = y_j - y_{j-1} \quad (j = 1, 2, \dots, n)$$

and then the area of the rectangle  $r_{ij}$  is

$$\Delta A_{ij} = \Delta x_i \Delta y_j \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n).$$

If  $F(x, y)$  is a non-negative function defined on the rectangle  $R$ , according to our intuitive idea of volume the portion of the volume under the surface  $z = F(x, y)$  over the rectangle  $r_{ij}$  should be approximately

$$F(u_i, v_j)\Delta A_j = F(u_i, v_i)\Delta x_i\Delta y_j,$$

where  $u_i$  and  $v_i$  are any values such that

$$x_{i-1} \leq u_i \leq x_i, \quad y_{j-1} \leq v_j \leq y_j$$

that is, where  $(u_i, v_i)$  is a point in the subrectangle  $r_{ij}$ . This suggests that the volume under the surface  $z = F(x, y)$  over the rectangle  $R$  should be approximated by the double sum of these expressions over both  $i$  and  $j$ , and we define the volume to be the limit of this sum as the lengths of all subintervals approach zero, written

$$\Phi = \lim_{\Delta x_i \rightarrow 0, \Delta y_j \rightarrow 0} \sum_{i,j} F(u_i, v_i)\Delta x_i\Delta y_j \quad (8.1)$$

*provided this limit exists.* We shall not go into detail about the nature of this limit, which is a rather more complicated concept than the basic idea of the limit of a function (which is itself a rather subtle idea). It is similar in nature to the kind of limit involved in the definition of the definite integral.

The expression (8.1), while motivated by the idea of volume, is defined as a limit of sum without using any geometric properties of volumes. We now abstract from this idea and define the double integral of the function  $F(x, y)$  over the rectangle  $R$ , written  $\int \int_R F(x, y)dA$ , by

$$\int \int_R F(x, y)dA = \lim_{\Delta x_i \rightarrow 0, \Delta y_j \rightarrow 0} \sum_{i,j} F(u_i, v_j)\Delta x_i\Delta y_j \quad (8.2)$$

provided this limit exists. Also, as we are no longer constrained by our intuitive idea of volume, we discard the assumption that  $F(x, y) \geq 0$  over  $R$ . The limit must be defined in such a way that it is independent both of the way in which the lengths of the subintervals  $x_{i-1} \leq x \leq x_i$  and  $y_{j-1} \leq y \leq y_j$  tend to zero and of the choice of the values  $u_i$  and  $v_j$ .

The definition (8.2) can be useful only if it can be shown that there is a reasonable class of functions for which the double integral exists (i.e., for which the limit (8.2) exists) and if we can develop techniques for calculating double integrals. Obviously, a proof of the existence of the double integral requires an understanding of the nature of the limit in the definition (8.2). For this reason, we shall state a result on existence of the double integral with no attempt to explain the ideas involved in the proof.

**THEOREM 1.** The double integral of a function  $F(x, y)$  over a rectangle  $R$ , defined by (8.2), exists if the function  $F(x, y)$  is bounded on the rectangle  $R$  and is continuous at all points of  $R$ .

This result gives a substantial class of integrable functions and would be sufficient for our purposes if we were planning to confine our attention to double

integrals over rectangles. However, in order to define double integrals over more general plane regions in the next section we need an even more general result for double integrals over a rectangle. It will be necessary to consider the double integral over a rectangle of a function which has points of discontinuity so long as there are “not too many” points of discontinuity. By “not too many” points of discontinuity we mean that the set of all points at which the function  $F(x, y)$  fails to be continuous is a set of points in the plane having “zero area”. By a set of points having “zero area” we mean a set of points which can be enclosed in a collection of rectangles whose area can be made arbitrarily small. A smooth curve is such a set, but it is possible to describe weird curves which do not have “zero area”. The curves that we shall encounter all have “zero area”, and the functions that we shall consider are all continuous except possibly on a set of points having “zero area”. We shall not attempt to go into more detail concerning these concepts, but shall merely state the generalization of Theorem 1 that we require for the next section.

**THEOREM 2.** The double integral of a function  $F(x, y)$  over a rectangle  $R$  exists if the function  $F(x, y)$  is bounded on the rectangle  $R$  and is continuous at all points of  $R$  except possibly a set of points having “zero area”.

The double integral has the following properties, analogous to the corresponding properties for single integrals

$$\begin{aligned} \int \int_R [F(x, y) + G(x, y)] dA &= \int \int_R F(x, y) dA + \int \int_R G(x, y) dA \\ \int \int_R kF(x, y) dA &= k \int \int_R F(x, y) dA \\ \int \int_R F(x, y) dA &\geq 0 \quad \text{if } F(x, y) \geq 0 \text{ on } R \end{aligned} \quad (8.3)$$

Here,  $F(x, y)$  and  $G(x, y)$  are any integrable functions and  $k$  is any constant. These facts may appear to be “obvious” but their proofs depend on the nature of the limit involved in the definition of the double integral and thus are not as easy as one might think. A useful consequence of (8.3) is that if  $F(x, y)$  and  $G(x, y)$  are integrable functions such that  $F(x, y) \geq G(x, y)$  at all points of  $R$ , then

$$\int \int_R F(x, y) dA \geq \int \int_R G(x, y) dA.$$

The definition (8.2) is too cumbersome to serve as a practical means for the calculation of a double integral. It is possible to reduce the calculation of a double integral to two successive calculations of single (definite) integrals. This enables us to use techniques developed previously for the calculation of single integrals in the calculation of double integrals.

We may calculate a double integral from the definition ((8.2) by summing on  $i$  for each fixed  $j$  and then summing the results on  $j$ . This amounts to dividing the rectangle  $R$  into horizontal strips (indexed by  $j$ ) and treating each

horizontal strip separately. The sum

$$\sum_{i=1}^m F(u_i, v_j) \Delta x_i \quad (8.4)$$

is a sum of the form used in the definition of the definite integral, and its limit as  $\Delta x_i \rightarrow 0$  is

$$\int_a^b F(x, v_j) dx. \quad (8.5)$$

The calculation of ((8.2) requires the multiplication of each expression ((8.4) by  $\Delta y_j$ , summation on  $j$ , and then a limiting process. If we replace ((8.4) by ((8.5) in this procedure, we obtain

$$\iint_R F(x, y) dA = \lim_{\Delta y_j \rightarrow 0} \sum_{j=1}^n \left[ \int_a^b F(x, v_j) dx \right] \Delta y_j.$$

The expression

$$\sum_{j=1}^n \left[ \int_a^b F(x, v_j) dx \right] \Delta y_j$$

is itself a sum of the form used in the definition of the definite integral of the function  $\int_a^b F(x, y) dx$  of  $y$ , and its limit as  $\Delta y_j \rightarrow 0$  is

$$\int_c^d \left[ \int_a^b F(x, y) dx \right] dy.$$

This suggests the validity of the equation

$$\iint_R F(x, y) dA = \int_c^d \left[ \int_a^b F(x, y) dx \right] dy \quad (8.6)$$

and it is possible to prove that the relation (8.6) is true provided the function  $F(x, y)$  is bounded on the rectangle  $R$  and is continuous at all points of  $R$  except possibly a set of points having “zero area”. The expression

$$\int_c^d \left[ \int_a^b F(x, y) dx \right] dy$$

is called an *iterated integral*; it is calculated by first finding the definite integral of  $F(x, y)$  with respect to  $x$ , giving a function of  $y$ , and then finding the definite integral of this function of  $y$  with respect to  $y$ .

We could use a similar approach but summing first on  $j$  for each fixed  $i$  and then summing the results on  $i$ , or dividing the rectangle  $R$  into vertical strips. This would give an iterated integral

$$\int_a^b \left[ \int_c^d F(x, y) dy \right] dx \quad (8.7)$$

and the double integral is also equal to this iterated integral.

**THEOREM 3.** If  $F(x, y)$  is bounded on the rectangle  $R$  defined by  $a \leq x \leq b$ ,  $c \leq y \leq d$  and is continuous at all points of  $R$  except possibly at a set of points having “zero area”, then

$$\begin{aligned} \int \int_R F(x, y) da &= \int_c^d \left[ \int_a^b F(x, y) dx \right] dy \\ &= \int_a^b \left[ \int_c^d F(x, y) dy \right] dx \end{aligned}$$

It is important to remember that the limits of integration  $a$  and  $b$  belong with the integration with respect to  $x$  and the limits of integration  $c$  and  $d$  belong with the integration with respect to  $y$ . The iterated integrals in (??) are sometimes written without brackets in the forms

$$\int_c^d \int_a^b F(x, y) dx dy, \quad \int_a^b \int_c^d F(x, y) dy dx$$

and the order of integration is the same as the order of the “differentials”  $dx, dy$ . In order to evaluate a double integral, one may evaluate either of the iterated integrals, and it is possible that one is much more difficult than the other.

**EXAMPLE 1.** Evaluate  $\int_R \int (x^2 + xy) dA$ , where  $R$  is the rectangle  $0 \leq x \leq 2$ ,  $1 \leq y \leq 3$ .

Solution. We write the double integral as an iterated integral

$$\int_1^3 \left[ \int_0^2 (x^2 + xy) dx \right] dy$$

and evaluate, starting with the inner integral

$$\int_0^2 (x^2 + xy) dx = \left[ \frac{x^3}{3} + \frac{x^2 y}{2} \right]_{x=0}^{x=2} = \frac{8}{3} + 2y.$$

Then

$$\begin{aligned} \int \int_R (x^2 + xy) dA &= \int_1^3 \left[ \frac{8}{3} + 2y \right] dy \\ &= \left[ \frac{8}{3}y + y^2 \right]_1^3 = (8 + 9) - \left( \frac{8}{3} + 1 \right) = 13\frac{1}{3}. \end{aligned}$$

We could equally well have integrated first with respect to  $y$  and then with

respect to  $x$ ,

$$\begin{aligned}
 \int_R \int (x^2 + xy) dA &= \int_0^2 \left[ \int_1^3 (x^2 + xy) dy \right] dx \\
 &= \int_0^2 \left[ x^2 y + \frac{xy^2}{2} \right]_{y=1}^{y=3} dx \\
 &= \int_0^2 \left[ \left( 3x^2 + \frac{9}{2}x \right) - \left( x^2 + \frac{1}{2}x \right) \right] dx \\
 &= \int_0^2 (2x^2 + 4x) dx \\
 &= \left[ \frac{2}{3}x^3 + 2x^2 \right]_0^2 = \frac{16}{3} + 8 = 13\frac{1}{3}
 \end{aligned}$$

Observe that while the final result is the same, the intermediate steps in the two approaches are not identical.

## 8.2 Double integrals over more general regions

If we attempt to define a double integral over a plane region  $D$  which is not a rectangle in the same way as we defined a double integral over a rectangle, we encounter a difficulty when we try to divide the region into rectangles. Wherever the boundary of the region is not horizontal or vertical, there are rectangle which are partly inside the region and partly outside the region

We will get around this difficulty by embedding the region  $D$  in a larger rectangle  $R$  and forming the double integral over the rectangle  $R$ . This will require extending the definition of the function to be integrated from  $D$  to the larger region  $R$  in such a way that the double integral does not depend on the choice of the rectangle  $R$ , and we do this in the obvious way – by defining the function to be zero outside the original region  $D$ . Thus formally we define the double integral

$$\int_D \int F(x, y) dA$$

in the following way:

1. (i) Let  $R$  be a rectangle which contains the region  $D$ .
2. (ii) Define the function  $\hat{F}(x, y)$  on  $R$  by

$$\hat{F}(x, y) = \begin{cases} F(x, y) & \text{if } (x, y) \in D, \\ 0 & \text{if } (x, y) \notin D. \end{cases}$$

3. (iii) Define

$$\int_D \int F(x, y) dA = \int_R \int \hat{F}(x, y) dA.$$

For this definition to be useful we need to know that the integral  $\int_R \int F(\hat{x}, y) dA$  exists and that its value does not depend on the choice of the rectangle  $R$ , which it is possible to prove provided the shape of the region  $D$  is not too complicated and the function  $F(x, y)$  is reasonably well-behaved. We also need to know how to calculate the integral  $\int_R \int \hat{F}(x, y) dA$  as an iterated integral.

In Theorem 2 of the preceding section we stated the result that the double integral  $\int_R \int \hat{F}(x, y) dA$  exists if the function  $\hat{F}(x, y)$  is bounded on the rectangle  $R$  and is continuous at all points of  $R$  except possibly a set of points having “zero area”. If  $F(x, y)$  is bounded on the region  $D$ , then  $\hat{F}(x, y)$  is bounded on the rectangle  $R$ , because of the way in which the extension  $\hat{F}(x, y)$  is defined. Also, the points of discontinuity of  $\hat{F}(x, y)$  in  $R$  are the points of discontinuity of  $F(x, y)$  in  $D$  together with the boundary points of  $D$ . If the set of boundary points of  $D$  has “zero area”, and if the set of points of discontinuity of  $F(x, y)$  in  $D$  has “zero area”, then the set of points of discontinuity of  $\hat{F}(x, y)$  in  $R$  also has “zero area”. Thus we obtain the following existence result from Theorem 2 of the preceding section.

**THEOREM 1.** If the function  $F(x, y)$  is bounded on the region  $D$ , if the function  $F(x, y)$  is continuous at all points of  $D$  except possibly a set of points having “zero area”, and if the boundary of  $D$  is a set of points having “zero area”, then the double integral

$$\iint_R F(x, y) dA$$

exists.

In order to see how to write a double integral as an iterated integral, let us consider a region  $D$  which is bounded above and below by smooth curves  $y = g(x)$  and  $y = f(x)$  which intersect at  $x = a$  and  $x = b$  (Figure 8.1).

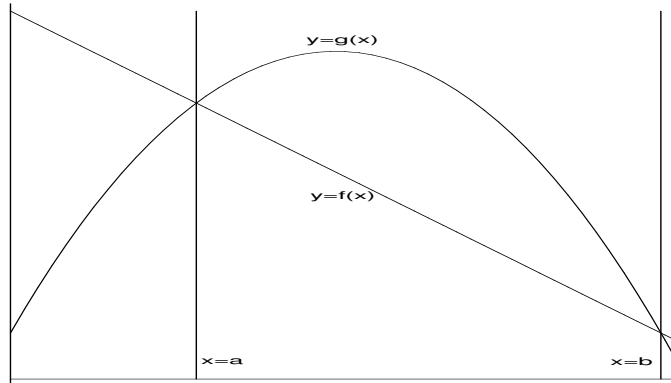


Figure 8.1: A region bounded by two curves

We choose  $c$  below the minimum of  $y = f(x)$  on  $a \leq x \leq b$  and  $d$  above the maximum of  $y = g(x)$  on  $a \leq x \leq b$  to give a rectangle  $R$  described by the

inequalities  $a \leq x \leq b$ ,  $c \leq y \leq d$  which contains  $D$ . According to the definition (1),

$$\begin{aligned} \iint_D F(x, y) dA &= \iint_R \hat{F}(x, y) dA \\ &= \int_a^b \left[ \int_c^d \hat{F}(x, y) dy \right] dx \end{aligned}$$

with  $\hat{F}(x, y)$  defined by

$$\hat{F}(x, y) = \begin{cases} 0, & c < y < f(x) \\ F(x, y), & h_1(x) \leq y \leq g(x) \\ 0, & h_2(x) < y < d. \end{cases}$$

for every  $x$ ,  $a \leq x \leq b$ . Then (8.3) gives

$$\begin{aligned} \int_c^d \hat{F}(x, y) dy &= \int_c^{f(x)} \hat{F}(x, y) dy + \int_{f(x)}^{g(x)} \hat{F}(x, y) dy \\ &\quad + \int_{g(x)}^d \hat{F}(x, y) dy \\ &= \int_c^{f(x)} 0 dy + \int_{h_1(x)}^{g(x)} F(x, y) dy + \int_{g(x)}^d 0 dy \\ &= \int_{f(x)}^{g(x)} F(x, y) dy. \end{aligned}$$

Combining (8.2) and (8.4), we see that

$$\iint_D F(x, y) dA = \int_a^b \left[ \int_{f(x)}^{g(x)} F(x, y) dy \right] dx.$$

**EXAMPLE 1.** Evaluate the integral  $\int_D \int xy dA$ , where  $D$  is the semicircle given by  $0 \leq y \leq \sqrt{1-x^2}$ ,  $-1 \leq x \leq 1$ .

Solution. According to (8.5), the desired double integral is equal to the iterated integral

$$\int_{-1}^1 \left[ \int_0^{\sqrt{1-x^2}} xy dy \right] dx.$$

To evaluate the inner integral (with respect to  $y$ ), we treat  $x$  as a constant, so that

$$\int_0^{\sqrt{1-x^2}} xy dy = \left. \frac{1}{2} xy^2 \right|_{y=0}^{y=\sqrt{1-x^2}} = \frac{1}{2} x(1-x^2),$$

noting that the limits of integration depend on  $x$ . Then

$$\begin{aligned}\iint_D xy dA &= \int_{-1}^1 \frac{1}{2}x(1-x^2)dx \\ &= \frac{1}{2} \int_{-1}^1 (x-x^3)dx \\ &= \frac{1}{2} \left[ \frac{x^2}{2} - \frac{x^4}{4} \right]_{-1}^1 = \frac{1}{2} \left[ \left( \frac{1}{2} - \frac{1}{4} \right) - \left( \frac{1}{2} - \frac{1}{4} \right) \right] = 0.\end{aligned}$$

If the region  $D$  is bounded by curves  $x = p(y)$  and  $x = q(y)$  which intersect at  $y = c$  and  $y = d$ , then a double integral of a function  $f(x, y)$  over  $D$  can be written as an iterated integral

$$\iint_D F(x, y) dA = \int_c^d \left[ \int_{p(y)}^{q(y)} F(x, y) dx \right] dy.$$

This leads to the possibility of two different iterated integrals each of which is equal to the desired double integral. One of these integrals may be easier to evaluate, and it may be necessary to try both to see which one is manageable.

It is essential to note that in an iterated integral of the form (8.5), with the inner (first) integration with respect to  $y$ , the limits of integration in the inner integral may depend on  $x$  but the limits of integration in the outer (second) integration must be constants. Similarly, in an iterated integral of the form (8.6) with the inner integration in the inner integral may depend on  $y$  (but not on  $x$ ) and the limits of integration in the outer integral must be constants.

**EXAMPLE 2.** Evaluate the iterated integral

$$\int_0^1 \left[ \int_y^1 e^{-x^2} dx \right] dy.$$

Solution. As it is not possible to find an indefinite integral of  $e^{-x^2}$ , we convert this iterated integral to a different iterated integral. In order to do this, we must interpret the iterated integral as a double integral and identify the region  $D$  of integration. In this case, we may see that the region is given by  $y \leq x \leq 1$ ,  $0 \leq y \leq 1$ , and thus is bounded by the curves  $x = y$  and  $x = 1$  for  $0 \leq y \leq 1$ . Another description of this region is as the region bounded by the curves  $y = x$  and  $y = 1$  for  $0 \leq x \leq 1$ . Thus we may write the iterated integral as

$$\begin{aligned}\int_0^1 \left[ \int_0^x e^{-x^2} dy \right] dx &= \int_0^1 \left[ ye^{-x^2} \right]_{y=0}^{y=x} dx \\ &= \int_0^1 xe^{-x^2} dx.\end{aligned}$$

To evaluate this integral, we must make the substitution  $u = x^2$ , so that  $x = 0$  corresponds to  $u = 0$ ,  $x = 1$  corresponds to  $u = 1$ , and  $\frac{du}{dx} = 2x$ . The integral

becomes

$$\frac{1}{2} \int_0^1 e^{-u} du = -\frac{1}{2} e^{-u} \Big|_0^1 = \frac{1}{2}(1 - e^{-1}).$$

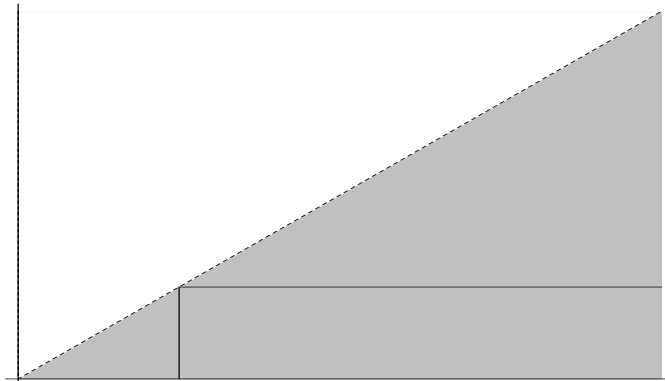


Figure 8.2: A triangular region of integration

A type of iterated integral that arises frequently in applications is the integral of a function over an infinite triangular region.

**EXAMPLE 3.** Show that

$$\int_0^{\infty} \left[ \int_0^x F(x, y) dy \right] dx = \int_0^{\infty} \left[ \int_y^{\infty} F(x, y) dx \right] dy.$$

**Solution.** The given integrals are iterated integrals corresponding to a double integral over the infinite triangular region bounded by the lines  $y = 0$ ,  $y = x$  (Figure 8.2). The iterated integral on the left side describes integration with respect to  $y$  (vertically) from 0 to  $x$  followed by integration with respect to  $x$  (horizontally) over all positive  $x$ . The iterated integral on the right side describes integration with respect to  $x$  (horizontally) from  $y$  to  $\infty$  followed by integration with respect to  $y$  (vertically) over all positive  $y$ . Since both these iterated integrals describe integration over the same region, they are equal.

### 8.3 Some exercises

In each of Exercises 1–4, evaluate the given iterated integral.

1.  $\int_0^2 \left[ \int_0^1 (1 - y^2) dy \right] dx$
2.  $\int_0^1 \left[ \int_0^2 (1 - y^2) dx \right] dy$
3.  $\int_1^2 \left[ \int_1^4 (x^2 + y^2 + 1) dx \right] dy$

$$4. \int_0^1 \left[ \int_0^1 (1 - xy) dy \right] dx.$$

In each of Exercises 5–8, evaluate the given double integral over the given rectangle  $R$ .

$$5. \int \int_R (1 - y^2) dA, \quad 0 \leq x \leq 2, 0 \leq y \leq 1$$

$$6. \int \int_R x^2 y^2 dA, \quad -1 \leq x \leq 1, -2 \leq y \leq 0$$

$$7. \int \int_R y e^{xy} dA, \quad 0 \leq x \leq 1, 0 \leq y \leq 2$$

$$8. \int \int_R x e^{-xy} dA, \quad -1 \leq x \leq 1, 0 \leq y \leq 1$$

9. Evaluate the double integral  $\int_R \int f(x, y) dA$ , where  $R$  is the rectangle  $0 \leq x \leq 1, 0 \leq y \leq 1$  and  $f(x, y)$  is defined by

$$f(x, y) = \begin{cases} e^{-x^2} & \text{if } x \geq y \\ 0 & \text{if } x < y \end{cases}$$

In each of Exercises 10–13, evaluate the given iterated integral (transforming to another iterated integral if necessary)

$$10. \int_0^2 \left[ \int_1^{\sqrt{y}} xy dx \right] dy$$

$$11. \int_0^1 \left[ \int_0^{1/y} ye^{xy} dx \right] dy$$

$$12. \int_0^2 \left[ \int_{x^2}^{2x} (2x + y) dy \right] dx$$

$$13. \int_0^1 \left[ \int_0^{\sqrt{1-x^2}} \sqrt{1-y^2} dy \right] dx$$

In each of Exercises 14–17, evaluate the given double integral over the region  $D$  described by the given inequalities.

$$14. \int \int_D dA, \quad \sqrt{y} \leq x \leq 2, 0 \leq y \leq 4$$

$$15. \int \int_D dA, \quad 1 \leq y \leq e^x, 0 \leq x \leq 2$$

$$16. \int \int_D (x + y) dA, \quad x^2 \leq y \leq x, 0 \leq x \leq 1$$

$$17. \int \int_D (x^2 + y^2) dA, \quad 0 \leq y \leq x^3, 0 \leq x \leq 2.$$

18. Show that if  $f(x)$  is a function  $y$  having a continuous second derivative and such that  $f(x_0) = f'(x_0) = 0$ , so that

$$\begin{aligned}f'(t) &= \int_{x_0}^t f''(u) du \\f(x) &= \int_{x_0}^x f'(t) dt = \int_{x_0}^x \left[ \int_{x_0}^t f''(u) du \right] dt,\end{aligned}$$

then

$$f(x) = \int_{x_0}^x (x-u)f''(u) du.$$

[Hint: Interchange the order of integration.]

19. By writing

$$\ln u = \int_1^u \frac{dt}{t}$$

and interchanging the order of integration, evaluate

$$\int_1^x \ln u du.$$

## Chapter 9

# Expansions of Functions in Power Series

In Section 2.5 we described the linear approximation of a function  $f(x)$  at a point  $x_0$ , and we made use of the linear approximation in various ways. The underlying idea was that near  $x_0$  the function  $f(x)$  could be approximated by the linear function

$$f(x_0) + (x - x_0)f'(x_0),$$

and that some properties of the function could be inferred from properties of this simpler linear function.

In this chapter we extend the idea of linear approximation to approximation by polynomials. Such approximations may be expected to be closer (having smaller error) than linear approximations. In addition, we will obtain explicit estimates for the error in an approximation. Later in the chapter we shall study the properties of infinite series, culminating in the representation of a function as the sum of an infinite series of powers of  $(x - x_0)$ , or as the limit of a polynomial approximation as the degree of the polynomial increases.

### 9.1 The Mean Value Theorem

The mean value theorem says that under suitable conditions, the secant line joining two points  $(a, f(a))$  and  $(b, f(b))$  on the graph of a function  $y = f(x)$  is parallel to the tangent line to the curve  $y = f(x)$  at some point  $c$  between  $a$  and  $b$ . Recall that the idea of the linear approximation was to approximate the graph of the curve  $y = f(x)$  by the tangent line at  $x = a$ . In the mean value theorem we have equality rather than approximation but we pay a price because we do not know the point  $c$ .

The mean value theorem is one of the most important results in calculus because it is needed to prove several results which may be considered obvious. However, proofs in mathematics must be pinned down; the word “obvious”

in mathematics often means “I’m sure it must be true but I don’t know how to prove it”, and sometimes “obvious” statements turn out to be false. For example, it was thought for many years that a function which is continuous everywhere must be differentiable except possibly at a finite number of points. This is not true; there are examples of functions which are continuous at every point but do not have a derivative at any point. Intuition in mathematics can lead one astray, and complete proofs are essential. This does not mean that every fact in every mathematics course must be established rigorously, but it does mean that there should be an awareness of the need for proofs. It also means that statements of theorems should include the hypotheses under which the theorem is valid.

**THEOREM 1** (Mean Value Theorem). If the function  $f(x)$  is continuous on the closed interval  $a \leq x \leq b$  and if it has a derivative  $f'(x)$  on the open interval  $a < x < b$ , then there exists at least one number  $c$  with  $a < c < b$  such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \quad (9.1)$$

An equivalent formulation of the relation (9.1) is

$$f(b) = f(a) + (b - a)f'(c). \quad (9.2)$$

If we replace  $a$  by  $x_0$  and  $b$  by  $x$  in (9.2), we obtain

$$f(x) = f(x_0) + (x - x_0)f'(c),$$

which we should compare with the linear approximation idea, that  $f(x)$  is approximated by (not equal to)

$$f(x_0) + (x - x_0)f'(x_0).$$

The point of the mean value theorem is that there is a number  $c$ , not that one should find it.

**EXAMPLE 1:** Let us apply the mean value theorem to the function  $f(x) = e^{-x}$  which is continuous and differentiable for all  $x$ ,  $-\infty < x < \infty$ , and has derivative  $f'(x) = -e^{-x}$ . Thus, taking  $x_0 = 0$ , for every  $x$  we have, according to the mean value theorem,

$$e^{-x} - 1 = -e^{-c}x$$

for some  $c$  between 0 and  $x$ . If  $x > 0$ , so that  $c > 0$ ,  $0 < e^{-c} < 1$ , and

$$0 < 1 - e^{-x} < x.$$

The mean value theorem is useful for proving many “obvious” results in calculus. One example of an “obvious” theorem whose proof requires the use

of the mean value theorem is the fact that if the derivative of a function is positive on an interval then the function is increasing on the interval. This result is needed in curve sketching, to translate information about the sign of the derivative into information about the nature of the graph. There is a corresponding fact that if the derivative of a function is negative on an interval then the function is decreasing on that interval.

In studying the indefinite integral we make heavy use of the fact that the only functions whose derivative is identically zero are constant functions, another result whose proof requires the mean value theorem. This is the justification for the use of a constant of integration; if we have one function with a prescribed derivative, then every function with the same derivative may be obtained by adding an arbitrary constant.

Another way of looking at the mean value theorem is to consider the form (9.2) as consisting of two parts, namely an explicit estimate  $f(a)$  for  $f(b)$  and an error term  $(b-a)f'(c)$  which is undetermined because  $c$  is not known exactly. If we think of  $a$  as a fixed base point and  $b$  as a variable, for example by replacing  $a$  by  $x_0$  and  $b$  by  $x$ , we may write (9.2) in the form

$$f(x) = f(x_0) + (x - x_0)f'(c) \quad (9.3)$$

with  $c$  between  $x_0$  and  $x$ . Recall that the linear approximation to  $f(x)$  [Section 2.5] is

$$f(x_0) + (x - x_0)f'(x_0), \quad (9.4)$$

but this is an approximation to  $f(x)$  without an error estimate. We think of (9.3) as a less refined approximation  $f(x_0)$  to  $f(x)$  but with the error estimate  $(x - x_0)f'(c)$ . In the next two sections, we shall extend the mean value theorem to give the linear approximation along with an error estimate, as well as higher order approximations.

A partial extension, less precise than the result to be obtained in the next section, may be obtained by applying the mean value theorem a second time in (9.3) to  $f'(c)$ , obtaining

$$f'(c) - f'(x_0) = (c - x_0)f''(d) \quad (9.5)$$

for some  $d$  between  $x_0$  and  $c$ . Substitution of (9.5) into (9.3) gives

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)[f'(x_0) + (c - x_0)f''(d)] \\ &= f(x_0) + (x - x_0)f'(x_0) + (x - x_0)(c - x_0)f''(d). \end{aligned} \quad (9.6)$$

The relation (9.6) gives  $f(x)$  as the sum of the linear approximation (9.4) and the error term

$$(x - x_0)(c - x_0)f''(d), \quad (9.7)$$

Taylor's theorem, to be described in the next section will give  $f(x)$  as the sum of the linear approximation (9.4) and an error term with a different form. The error term (9.7) is at most  $M|x - x_0|^2$ , if  $|f''(d)| \leq M$  for  $x_0 \leq d \leq x$ , since  $|c - x_0| < |x - x_0|$  and this estimate is sufficient for many applications.

## 9.2 Taylor Polynomials

The linear approximation to a function  $f(x)$  at a point  $x_0$ , introduced in Section 2.5, may be described as the linear function which has the same value and the same derivative as  $f(x)$  at  $x_0$ . If we attempt to find a linear function, or polynomial of degree 1,

$$P_1(x) = c_0 + c_1x \quad (9.8)$$

such that

$$P_1(x_0) = f(x_0), \quad P_1'(x_0) = f'(x_0),$$

we have two equations to determine the coefficients  $a_0$  and  $a_1$ . The condition  $P_1(x_0) = f(x_0)$  gives

$$c_0 + c_1x_0 = f(x_0), \quad (9.9)$$

and since  $P_1'(x) = c_1$  the condition  $P_1'(x_0) = f'(x_0)$  gives

$$c_1 = f'(x_0). \quad (9.10)$$

Substitution of (9.10) into (9.9) gives

$$c_0 = f(x_0) - c_1x_0 = f(x_0) - x_0f'(x_0)$$

and then we obtain

$$\begin{aligned} P_1(x) = c_0 + c_1x &= f(x_0) - x_0f'(x_0) + f'(x_0)x \\ &= f(x_0) + (x - x_0)f'(x_0), \end{aligned}$$

which is the linear approximation. The algebra would have been simpler if we had written the linear function  $P_1(x)$  in the form

$$P_1(x) = a_0 + a_1(x - x_0). \quad (9.11)$$

Then the condition  $P_1(x_0) = f(x_0)$  would give

$$a_0 = f(x_0) \quad (9.12)$$

and since  $P_1'(x) = a_1$ , the condition  $P_1'(x_0) = f'(x_0)$  would give

$$a_1 = f'(x_0).$$

In this section, we consider polynomial approximations to a function  $f(x)$  obtained by specifying conditions at a base point  $x_0$ . For example, let us find the quadratic (of degree 2) polynomial  $P_2(x)$  which has the same value, derivative, and second derivative as  $f(x)$  at  $x_0$ . If we write

$$P_2(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2. \quad (9.13)$$

The condition  $P_2(x_0) = f(x_0)$  gives

$$a_0 = f(x_0). \quad (9.14)$$

Since  $P_2'(x) = a_1 + 2a_2(x - x_0)$ , the condition  $P_2'(x_0) = f'(x_0)$  gives

$$a_1 = f'(x_0). \quad (9.15)$$

Since  $P_2''(x) = 2a_2$ , the condition  $P_2''(x_0) = f''(x_0)$  gives  $2a_2 = f''(x_0)$  or

$$a_2 = \frac{1}{2}f''(x_0). \quad (9.16)$$

Substitution of (9.14), (9.15), (9.16) into (9.13) gives

$$P_2(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2f''(x_0). \quad (9.17)$$

Because the conditions (9.14), (9.15) are the same as (9.11) (9.12) respectively, comparison of (9.13) with (9.8) shows that the constant and linear terms in  $P_2(x)$  are the same as constant and linear terms in  $P_1(x)$ . Thus the quadratic approximation  $P_2(x)$  is the linear approximation  $P_1(x)$  plus an additional second-degree term.

We may carry this process further. The polynomial  $P_3(x)$  of degree 3 such that

$$P_3(x_0) = f(x_0), P_3'(x_0) = f'(x_0), P_3''(x_0) = f''(x_0), P_3'''(x_0) = f'''(x_0)$$

is given by

$$P_3(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2f''(x_0) + \frac{1}{6}(x - x_0)^3f'''(x_0).$$

To see this, we assume the form

$$P_3(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 \quad (9.18)$$

and solve for the coefficients  $a_0, a_1, a_2, a_3$ . The calculation of  $a_0, a_1, a_2$  is exactly the same as in (9.14), (9.15), (9.16). To find  $a_3$ , we differentiate (9.18) three times, obtaining

$$\begin{aligned} P_3'(x) &= a_1 + 2a_2(x - x_0) + 3a_3(x - x_0)^2 \\ P_3''(x) &= 2a_2 + 3 \cdot 2a_3(x - x_0) \\ P_3'''(x) &= 3 \cdot 2a_3, \end{aligned}$$

and then use the condition  $P_3'''(x_0) = f'''(x_0)$  to obtain

$$a_3 = \frac{1}{3 \cdot 2}f'''(x_0) = \frac{1}{6}f'''(x_0).$$

For purposes of extension, it is convenient to write this in the form

$$a_3 = \frac{1}{3!}f'''(x_0).$$

[Recall that if  $k$  is a positive integer,  $k!$  means the product of the integers from 1 to  $k$ , so that  $1! = 1$ ,  $2! = 1 \cdot 2 = 2$ ,  $3! = 1 \cdot 2 \cdot 3 = 6$ ,  $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$ , etc.; it is conventional to define  $0! = 1$ .]

We now define the *Taylor polynomial of degree  $n$  for  $f(x)$  at the base point  $x_0$*  as the polynomial  $P_n(x)$  of degree  $n$  such that

$$P_n(x_0) = f(x_0), P'_n(x_0) = f'(x_0), \dots, P_n^{(n)}(x_0) = f^{(n)}(x_0). \quad (9.19)$$

As a polynomial of degree  $n$  has  $(n+1)$  coefficients the  $(n+1)$  conditions (9.19) determine the polynomial  $P_n(x)$  completely. If we follow the same process for determining coefficients as we have used in the cases  $n = 1, 2, 3$ , we obtain

$$\begin{aligned} P_n(x) &= f(x_0) + (x - x_0) \frac{f'(x_0)}{1!} + (x - x_0)^2 \frac{f''(x_0)}{2!} + \dots \\ &+ (x - x_0)^n \frac{f^{(n)}(x_0)}{n!}. \end{aligned} \quad (9.20)$$

**EXAMPLE 1.** Calculate the Taylor polynomials of degrees 1, 2, 3, and 4 for the function  $f(x) = e^x$  at the base point 0.

Solution. Since the derivative of every order of  $e^x$  is  $e^x$ , with value 1 at  $x = 0$ , the formula (9.20) gives

$$\begin{aligned} P_1(x) &= 1 + \frac{x}{1!}, & P_2(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!} \\ P_3(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!}, & P_4(x) &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!}. \end{aligned}$$

In fact, we may easily see that

$$P_n(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} = \sum_{k=0}^n \frac{x^k}{k!}.$$

**EXAMPLE 2.** Use the approximations obtained in Example 1 to estimate the value of  $e$ .

Solution. Since  $e$  is the value of the function  $e^x$  for  $x = 1$ , we calculate as approximations  $P_1(1) = 1 + \frac{1}{1!} = 2$ ,  $P_2(1) = 1 + \frac{1}{1!} + \frac{1}{2!} = 2\frac{1}{2}$ ,  $P_3(1) = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} = 2\frac{2}{3}$ ,  $P_4(1) = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} = 2\frac{17}{24} = 2.7083$ .

As we know a value for  $e$ , namely  $2.71828\dots$ , we might conjecture that the approximation  $P_n(x)$  to  $f(x)$  becomes more accurate in general as  $n$  is increased, and that we may approximate the value of a function as accurately as we wish by making the degree of the approximating Taylor polynomial large enough. To validate (or disprove) such a conjecture we would need an estimate of the error in the approximation (the difference between the function and the Taylor polynomial) and how it depends on the degree  $n$  of the Taylor polynomial. An approach to the general error estimate question will be developed in the next section, but it may be possible to obtain an explicit formula for the error for some functions.

**EXAMPLE 3.** Find the Taylor polynomial  $P_n(x)$  of the function  $f(x) = \frac{1}{1-x}$  at the base point 0.

Solution. We have

$$\begin{aligned} f(x) &= \frac{1}{1-x}, & f(0) &= 1 \\ f'(x) &= \frac{1}{(1-x)^2}, & f'(0) &= 1 \\ f''(x) &= \frac{2}{(1-x)^3}, & f''(0) &= 2 \\ f'''(x) &= \frac{3 \cdot 2}{(1-x)^4}, & f'''(0) &= 3 \end{aligned}$$

and we may see that in general,

$$f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}}, \quad f^{(k)}(0) = k!.$$

Thus the Taylor polynomial of degree  $n$  is

$$\begin{aligned} P_n(x) &= 1 + x + \frac{2}{2!}x^2 + \frac{3!}{3!}x^3 + \cdots + \frac{n!}{n!}x^n \\ &= 1 + x + x^2 + \cdots + x^n. \end{aligned} \tag{9.21}$$

It is possible to obtain a different but equivalent expression for  $P_n(x)$  in Example 3 above by a device used for summing a geometric series. We multiply the equation (9.21) by  $x$ , obtaining

$$xP_n(x) = x + x^2 + \cdots + x^n + x^{n+1} \tag{9.22}$$

and subtract (9.22) from (9.21) obtaining

$$(1-x)P_n(x) = 1 - x^{n+1}.$$

Thus

$$P_n(x) = \frac{1}{1-x} - \frac{x^{n+1}}{1-x}. \tag{9.23}$$

The relation (9.23) says that the difference between the function  $f(x) = 1/1-x$  and the Taylor polynomial  $P_n(x)$  is  $\frac{x^{n+1}}{1-x}$ . From this explicit formula for the error we see that if  $|x| < 1$ , so that  $x^{n+1} \rightarrow 0$  as  $n \rightarrow \infty$ , the error decreases to zero as  $n \rightarrow \infty$ . Thus on the interval  $-1 < x < 1$ , the Taylor polynomial  $P_n(x)$  approximates the function  $f(x) = \frac{1}{1-x}$  with an error which approaches zero as  $n \rightarrow \infty$ .

The main uses of Taylor polynomial approximations are not to estimate the value of a function at a given point. Currently available technology, such as inexpensive electronic pocket calculators, provides easier methods. The importance of Taylor polynomials is in approximating functions over an interval, and this will require error estimates which are valid over an interval.

**EXAMPLE 4.** Find the Taylor polynomial  $P_n(x)$  of the function  $f(x) = \ln(1+x)$  at the base point 0.

Solution. We have

$$\begin{aligned} f(x) &= \ln(1+x), & f(0) &= \ln 1 = 0 \\ f'(x) &= \frac{1}{1+x}, & f'(0) &= 1 \\ f''(x) &= \frac{-1}{(1+x)^2}, & f''(0) &= -1 \\ f'''(x) &= \frac{(-1)(-2)}{(1+x)^2} = \frac{2!}{(1+x)^2}, & f'''(0) &= 2!. \end{aligned}$$

In general,

$$f^{(k)}(x) = (-1)^{k-1} \frac{(k-1)!}{(1+x)^{k-1}}, \quad f^{(k)}(0) = (-1)^{k-1} (k-1)!.$$

Thus the Taylor polynomial is

$$\begin{aligned} P_n(x) &= \frac{x}{1!} - \frac{x^2}{2!} + \frac{2!x^3}{3!} - \cdots + (-1)^{n-1} \frac{(n-1)!}{n!} x^n \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1} \frac{x^n}{n}. \end{aligned}$$

**EXAMPLE 5.** Find the Taylor polynomial  $P_n(x)$  of the function  $f(x) = e^{-x}$  at the base point 0.

Solution. We have

$$\begin{aligned} f(x) &= e^{-x}, & f(0) &= 1 \\ f'(x) &= -e^{-x}, & f'(0) &= -1 \\ f''(x) &= e^{-x}, & f''(0) &= 1 \\ f'''(x) &= -e^{-x}, & f'''(0) &= -1 \\ f^{(4)}(x) &= e^{-x}, & f^{(4)}(0) &= 1 \text{ etc.} \end{aligned}$$

Thus  $P_n(x)$  is an alternating sum

$$1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots.$$

**EXAMPLE 6.** Find the Taylor polynomial  $P_n(x)$  of the function  $f(x) = (1+x)^p$  at the base point 0.

Solution. We have

$$\begin{aligned} f(x) &= (1+x)^p, & f(0) &= 1 \\ f'(x) &= p(1+x)^{p-1}, & f'(0) &= p \\ f''(x) &= p(p-1)(1+x)^{p-2}, & f''(0) &= p(p-1) \\ f'''(x) &= p(p-1)(p-2)(1+x)^{p-3}, & f'''(0) &= p(p-1)(p-2). \end{aligned}$$

Thus

$$\begin{aligned} P_0(x) &= 1, & P_1(x) &= 1 + px, & P_2(x) &= 1 + px + \frac{p(p-1)}{2!}x^2 \\ P_3(x) &= 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3, \text{ etc.} \end{aligned}$$

Note that if  $p$  is a positive integer,  $f^{(k)}(0) = 0$  for  $k \geq p + 1$  because

$$f^{(k)}(0) = p(p-1)(p-2)\cdots(p-k+1).$$

Thus  $P_n(x)$  is a polynomial of fixed degree  $p$  for every  $n \geq p$ . However, if  $p$  is not a positive integer,  $P_n(x)$  will always have degree  $n$  no matter how large  $n$  is. Although we speak of  $P_n(x)$  as a polynomial of degree  $n$ , it is possible for  $P_n(x)$  to be a polynomial of degree less than  $p$ .

We conclude with an example of a function which we can not evaluate directly, for which the Taylor polynomials give a useful means of approximating values of the function.

**EXAMPLE 7.** Find the Taylor polynomials of degree 0, 1, 2, and 3 of the function  $f(x) = \int_0^x e^{-t^2} dt$  at the base point 0, and use them to estimate  $f(0.2)$ . Solution. We have

$$\begin{aligned} f(x) &= \int_0^x e^{-t^2} dt, & f(0) &= 0 \\ f'(x) &= e^{-x^2}, & f'(0) &= 1 \\ f''(x) &= -2xe^{-x^2}, & f''(0) &= 0 \\ f'''(x) &= -2xe^{-x^2} - (2x)(-2x)e^{-x^2}, & f'''(0) &= -2. \end{aligned}$$

Thus

$$\begin{aligned} P_0(x) &= 0, & P_0(0.2) &= 0 \\ P_1(x) &= x, & P_1(0.2) &= 0.2 \\ P_2(x) &= x, & P_2(0.2) &= 0.2 \\ P_3(x) &= x - \frac{2}{3!}x^3 = x - \frac{x^3}{3}, & P_3(0.2) &= 0.2 - 0.00267 = 0.19733. \end{aligned}$$

The values  $P_0(0.2)$ ,  $P_1(0.2)$ ,  $P_2(0.2)$ ,  $P_3(0.2)$  are the respective approximations to  $f(0.2) = \int_0^{0.2} e^{-t^2} dt$ .

### 9.3 Taylor's Theorem

We have calculated Taylor polynomials for a given function, thinking of them as approximations to the function. The questions which we explore in this section is how good an approximation to  $f(x)$  is the Taylor polynomial  $P_n(x)$ . The Taylor

polynomial  $P_n(x)$  at base point  $x_0$  is designed to approximate the function  $f(x)$  as well as possible at the base point  $x_0$ , being defined by the conditions that  $P_n(x)$  and its derivatives up to order  $(n-1)$  should be respectively equal to  $f(x)$  and its derivatives up to order  $(n-1)$  at the point  $x_0$ . We shall obtain an estimate for the difference between  $f(x)$  and  $P_n(x)$  on an interval containing  $x_0$ , and this will measure the accuracy of the approximation on that interval.

Let us define

$$R_n(x) = f(x) - P_n(x).$$

Then  $R_n(x)$  is the error made in approximating  $f(x)$  by  $P_n(x)$ , and is the quantity which we wish to estimate. We may interpret the mean value theorem as saying

$$R_0(x) = (x - x_0)f'(c)$$

for some point  $c$  between  $x_0$  and  $x$ , and we may interpret the relation (7) of Section 4.1 as saying

$$R_1(x) = (x - x_0)(c - x_0)f''(d)$$

for points  $c$  and  $d$  with  $c$  between  $x_0$  and  $x$  and  $d$  between  $x_0$  and  $c$ . Taylor's theorem expresses  $R_n(x)$  in terms of the values of  $f^{(n+1)}(t)$  as  $t$  ranges over the interval from  $x_0$  to  $x$ .

**Taylor's Theorem.** Suppose that  $f(t), f'(t), \dots, f^{(n+1)}(t)$  are continuous on  $x_0 \leq t \leq x$ , where  $n$  is a given positive integer. Let

$$\begin{aligned} P_n(x) &= f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2!} + \dots + (x - x_0)^n \frac{f^{(n)}(x_0)}{n!} \\ R_n(x) &= f(x) - P_n(x). \end{aligned}$$

Then

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}. \quad (9.24)$$

for some point  $c$  between  $x_0$  and  $x$ .

We do not give the proof of Taylor's theorem, which may be found in many calculus texts. The form (9.24) for  $R_n(x)$  is known as the *Lagrange form of the remainder*. Just as in the mean value theorem (which is the case  $n = 0$  of Taylor's theorem), the actual value of  $c$  is not important. The bounds given by (9.26) below provide a more useful version of Taylor's theorem as an estimate of  $R_n(x)$ .

**Corollary to Taylor's Theorem.** If

$$|f^{(n+1)}(t)| \leq M \quad (9.25)$$

for all  $t$  between  $x_0$  and  $x$ , then the remainder  $R_n(x)$  satisfies the estimate

$$|R_n(x)| \leq \frac{M}{(n+1)!} |x - x_0|^{n+1}. \quad (9.26)$$

The proof of this corollary is an immediate consequence of (9.25) since  $|f^{(n+1)}(c)| \leq M$ .

**EXAMPLE 1.** Estimate the error in approximating  $e^{-x}$  by  $1 - x$  on the interval  $0 \leq x \leq 1$ .

Solution. We take  $f(x) = e^{-x}$  and use the base point 0. Then we may consider  $1 - x$  as the Taylor polynomial of degree 1. Since  $f''(x) = e^{-x}$ , and  $e^{-x} \leq 1$  for  $0 \leq x \leq 1$ , the estimate (9.26) with  $n = 1$  gives

$$|R_1(x)| \leq \frac{1}{2!}|x|^2 \leq \frac{1}{2} = 0.5.$$

**EXAMPLE 2.** Estimate the error in approximating  $e^{-x}$  by  $1 - x + \frac{x^2}{2}$  on the interval  $0 \leq x \leq 1$ .

Solution. As in Example 1 we take  $f(x) = e^{-x}$  and use the base point 0. Then  $1 - x + \frac{x^2}{2}$  is the Taylor polynomial of degree 2. Since  $|f^{(3)}(x)| \leq 1$  for  $0 \leq x \leq 1$ , the estimate (9.26) with  $n = 2$  gives

$$|R_2(x)| \leq \frac{1}{3!}|x|^3 \leq \frac{1}{6} = 0.167.$$

Thus for  $0 \leq x \leq 1$  we have

$$|e^{-x} - 1 + x - \frac{x^2}{2}| \leq 0.167$$

In fact,

$$e^{-x} - 1 + x - \frac{x^2}{2} = 0.132.$$

**EXAMPLE 3.** Obtain a bound for the error in the estimate 0.19733 for  $\int_0^{0.2} e^{-t^2} dt$  [Example 7, Section 5.2].

Solution. The estimate 0.19733 was obtained from the Taylor polynomial of degree 3. Thus we must estimate  $R_3(0.2)$ . In Example 7, Section 5.2 with  $f(x) = \int_0^x e^{-t^2} dt$ , we found

$$f'''(x) = -2e^{-x^2} + 4x^2e^{-x^2}.$$

Therefore,

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

For  $0 \leq x \leq 0.2$  we have

$$|f^{(4)}(x)| = e^{-x^2}|12x - 8x^3| \leq |12x - 8x^3|.$$

It is not difficult to verify that the function  $12x - 8x^3$  is monotone decreasing for  $0 \leq x \leq 0.2$ , from 0 at  $x = 0$  to -2.335 at  $x = 0.2$ . Thus  $|f^{(4)}(x)| \leq 2.336$  for  $0 \leq x \leq 0.2$ . Now the estimate (9.26) with  $n = 3$  gives

$$|R_3(0.2)| \leq \frac{2.336}{4!}(0.2)^4 = 0.0001557.$$

This shows that

$$\left| \int_0^{0.2} e^{-t^2} dt - 0.19733 \right| \leq 0.0001557,$$

or that  $\int_0^{0.2} e^{-t^2}$  is between  $0.19733 - 0.00016 = 0.19717$ , and  $0.19733 + 0.00016 = 0.19749$ .

## 9.4 The Taylor Series of a Function

In Example 3, Section 5.2 we calculated the Taylor polynomial  $P_n(x)$  of degree  $n$  for the function

$$f(x) = \frac{1}{1-x}$$

as

$$P_n(x) = 1 + x + x^2 + \cdots + x^n.$$

We also calculated the difference between  $f(x)$  and  $P_n(x)$ , obtaining

$$1 + x + x^2 + \cdots + x^n = \frac{1}{1-x} - \frac{x^{n+1}}{1-x},$$

so that

$$f(x) = P_n(x) + \frac{x^{n+1}}{1-x}.$$

It is possible to prove that  $x^{n+1} \rightarrow 0$  as  $n \rightarrow \infty$  for each  $x$  with  $-1 < x < 1$ , or  $|x| < 1$ , and from this we see that the difference between the function  $f(x)$  and the Taylor polynomial  $P_n(x)$  tends to zero as  $n \rightarrow \infty$  for every  $x$  in the interval  $-1 < x < 1$ . This means that, at least formally, we can write

$$f(x) = \lim_{n \rightarrow \infty} P_n(x)$$

or

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{k=0}^{\infty} x^k$$

on the interval  $-1 < x < 1$ .

In Example 1, Section 5.2, we calculated the Taylor polynomial  $P_n(x)$  of degree  $n$  of the function

$$f(x) = e^x$$

as

$$P_n(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} = \sum_{k=0}^n \frac{x^k}{k!}.$$

In Section 5.3, we established Taylor's theorem, estimating the difference between a function  $f(x)$  and its Taylor polynomial  $P_n(x)$ . According to this theorem, if we write

$$f(x) = P_n(x) + R_n(x),$$

then the "remainder"  $R_n(x)$  is given by

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1}$$

for some  $c$  between 0 and  $x$ . For the function  $f(x) = e^x$ , the derivative of every order is  $e^x$  and the remainder takes the form

$$R_n(x) = \frac{e^c}{(n+1)!} x^{n+1}.$$

On any interval  $-A \leq x \leq A$ , since  $|c| < |x|$  we have

$$|R_n(x)| \leq \frac{e^{|x|}}{(n+1)!} |x|^{n+1} \leq \frac{e^A A^{n+1}}{(n+1)!}.$$

It is possible to show that  $A^{n+1}/(n+1)! \rightarrow 0$  as  $n \rightarrow \infty$  for every  $A > 0$ , and this implies that the remainder  $R_n(x)$ , and therefore the difference between the function  $f(x)$  and the Taylor polynomial  $P_n(x)$ , tends to zero as  $n \rightarrow \infty$  for every  $x$  in an interval  $-A \leq x \leq A$  with  $A$  arbitrary. Thus, at least formally, we can write

$$f(x) = \lim_{n \rightarrow \infty} P_n(x)$$

or

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

on any interval  $-A \leq x \leq A$ , and therefore on the interval  $-\infty < x < \infty$ .

In general, we have seen in Section 5.3 that if the function  $f(x)$  has derivatives of all orders, then it has a Taylor polynomial

$$\begin{aligned} P_n(x) &= f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \\ &= \sum_{k=0}^n \frac{f^{(k)}(0)}{k!}x^k \end{aligned} \quad (9.27)$$

for every non-negative integer  $n$ . We now define the *Taylor series* of  $f(x)$  (with base point 0) as  $\lim_{n \rightarrow \infty} P_n(x)$ , represented symbolically as

$$\lim_{n \rightarrow \infty} P_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \quad (9.28)$$

In particular, we have shown that the Taylor series of the function  $1/(1-x)$  is  $\sum_{k=0}^{\infty} x^k$  and that the Taylor series of the function  $e^x$  is  $\sum_{k=0}^{\infty} x^k/k!$ . Because

the remainder  $R_n(x)$  in Taylor's theorem for each of these functions approaches zero on some interval, we have also shown that the Taylor series for each function represents the function.

Similarly, we may define the Taylor series of  $f(x)$  with base point  $x_0$  as

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \quad (9.29)$$

When we write  $\lim_{n \rightarrow \infty} P_n(x)$ , we mean the function of  $x$  whose value for any given  $x$  is the limit of the sequence of numbers  $P_n(x)$ . For those values of  $x$  for which this sequence has a limit, a function  $P(x)$  is defined by

$$P(x) = \lim_{n \rightarrow \infty} P_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

We wish to study the relation between this function  $P(x)$  and the originally given function  $f(x)$ . In particular, we wish to investigate the extent to which functions can be combined algebraically, differentiated, and integrated by performing these operations on the Taylor series of the functions.

We may use the same approach for many other functions. If  $f(x)$  is a function whose derivatives of all orders exist on some interval we can use (9.28) to construct the Taylor series of  $f(x)$ , obtaining a power series  $\sum_{k=0}^{\infty} a_k x^k$  whose coefficients are given by

$$a_k = \frac{f^{(k)}(0)}{k!} \quad (k = 0, 1, 2, \dots).$$

In order for the Taylor series of a function  $f(x)$  to converge to the function  $f(x)$ , it is necessary that

$$\lim_{n \rightarrow \infty} R_n(x) = 0. \quad (9.30)$$

According to Taylor's theorem,

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1}$$

for some  $c$  between 0 and  $x$ . If there is a number  $M$ , independent of  $n$ , such that

$$|f^{(n+1)}(t)| \leq M \quad (9.31)$$

for  $t$  in some interval, then

$$|R_n(x)| \leq M \frac{|x|^{n+1}}{(n+1)!}$$

and because  $\lim_{n \rightarrow \infty} \frac{|x|^{n+1}}{(n+1)!} = 0$ , this shows that  $\lim_{n \rightarrow \infty} R_n(x) = 0$  on that interval and thus that the Taylor series of  $f(x)$  converges to  $f(x)$  on the interval. Even if

(9.31) is not satisfied, it is possible that (9.30) may be satisfied on some interval. A function for which (9.30) holds on some interval containing 0 in its interior is said to be *analytic* at 0; an analytic function is one which can be expanded in a Taylor series and which is the sum of its Taylor series on some interval containing 0 in its interior.

All the functions which we shall encounter are analytic, and we will normally not check the details of proving that (9.30) holds. However, the reader should not assume that every function possessing derivatives of all orders is analytic. The standard counter-example is

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

It is possible to calculate the derivatives of this function and to show that  $f^{(k)}(0) = 0$ , ( $k = 0, 1, 2, \dots$ ). Thus all terms of the Taylor series of  $f(x)$  have coefficient 0, and  $f(x)$  is a non-zero function whose Taylor series is the zero series. Obviously the Taylor series converges for all  $x$ ; equally obviously, the Taylor series does not converge to  $f(x)$  except at  $x = 0$ .

Some of the Taylor series expansions we can derive from calculations made earlier are

$$\begin{aligned} \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots [Example4, Section5.2] \\ (1+x)^p &= 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \dots [Example6, Section5.2] \\ e^x &= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots [Example1, Section5.2] \end{aligned}$$

All these functions are analytic at 0, with the expansion valid if  $-\infty < x < \infty$  for  $e^x$ , and if  $-1 < x < 1$  for  $\ln(1+x)$ . For  $(1+x)^p$  the expansion is valid if  $-1 < x < 1$  unless  $p$  is a positive integer, in which case the series terminates and becomes a polynomial, valid for  $-\infty < x < \infty$ . In general, the Taylor series (with base point 0) of an analytic function converges to the function on an interval  $-R < x < R$  for some number  $R$  called the *radius of convergence* of the series. It is possible to have  $R = \infty$ , as for the function  $e^x$ , or to have a finite value of  $R$ , as for the functions  $1/(1-x)$ ,  $\ln(1+x)$ , and  $(1+x)^p$  (if  $p$  is not a positive integer).

Our principal reason for studying the Taylor series expansions of functions is that it is possible to perform operations on functions by performing the same operations on their Taylor series, treating Taylor series as if they were polynomials and operating term by term. This is possible both for algebraic operations and the calculus operations of differentiation and integration. It is possible to establish the following results.

**THEOREM 1.** Let  $f(x)$  be a function having a Taylor series expansion  $\sum_{k=0}^{\infty} a_k x^k$  which converges to  $f(x)$  and let  $g(x)$  be a function having a Taylor series expansion  $\sum_{k=0}^{\infty} b_k x^k$  which converges to  $g(x)$ , both expansions being

valid on some interval  $-R < x < R$ , so that

$$f(x) = \sum_{k=0}^{\infty} a_k x^k, \quad g(x) = \sum_{k=0}^{\infty} b_k x^k, \quad (|x| < R).$$

Then

$$\begin{aligned} f(x) + g(x) &= \sum_{k=0}^{\infty} (a_k + b_k) x^k \\ f(x)g(x) &= \left( \sum_{k=0}^{\infty} a_k x^k \right) \left( \sum_{k=0}^{\infty} b_k x^k \right) \\ &= (a_0 + a_1 x + a_2 x^2 + \cdots)(b_0 b_1 x + b_2 x^2 x \cdots) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) x + (a_0 b_2 + a_1 b_1 + a_2 b_0) x^2 + \cdots \\ &= \sum_{e=0}^{\infty} c_e x^e, \end{aligned}$$

where  $c_k = \sum_{\ell=0}^k a_{\ell} b_{k-\ell}$ .

$$\begin{aligned} f\{g(x)\} &= a_0 + a_1 g(x) + a_2 g(x)^2 + \cdots \\ &= a_0 + a_1 \{b_0 + b_1 x + b_2 x^2 \cdots\} + a_2 \{b_0 + b_1 x + b_2 x^2 + \cdots\}^2 + \cdots \end{aligned}$$

$$\begin{aligned} f'(x) &= \sum_{k=0}^{\infty} k a_k x^{k-1} \\ \int_0^x f(x) dt &= \int_0^x \left[ \sum_{k=0}^{\infty} a_k t^k \right] dt = \sum_{k=0}^{\infty} a_k \int_0^x t^k dt \\ &= \sum_{k=0}^{\infty} \frac{a_k}{k+1} x^{k+1} \end{aligned}$$

with all expansions being valid for  $-R < x < R$ .

**EXAMPLE 1.** Find the Taylor series expansion of  $x e^x$ .

Solution. While we could obtain the desired expansion by calculating successive derivatives at 0 of the function  $x e^x$ , it is easier to begin with the expansion

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad (-\infty < x < \infty)$$

and multiply by  $x$ ; multiplication by  $x$  has the effect of raising each exponent in the series by 1. Thus

$$x e^x = \sum_{k=0}^{\infty} \frac{x^{k+1}}{k!} = x + \frac{x^2}{1!} + \frac{x^3}{2!} + \cdots$$

and this representation is valid so all  $x$ .

**EXAMPLE 2.** Find the Taylor series expansion of the function  $\int_0^x e^{-t^2} dt$ .

Solution. We begin by writing

$$e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots.$$

Then we replace  $t$  by  $-t^2$  to give

$$e^{-t^2} = 1 - \frac{t^2}{1!} + \frac{t^4}{2!} - \frac{t^6}{3!} + \cdots.$$

Finally, we integrate term by term to give

$$\int_0^x e^{-t^2} dt = x - \frac{x^3}{3 \cdot 1!} + \frac{x^5}{5 \cdot 2!} - \frac{x^7}{7 \cdot 3!} + \cdots$$

with this expansion being valid for all  $x$ . [Compare Example 7, Section 5.2.]

We have dealt exclusively with expansions at the base point 0, using the form (9.28) with powers of  $x$ . We can also expand functions in Taylor series at a base point  $x_0$ , using the form (9.29) with powers of  $(x - x_0)$ . A function having a Taylor series expansion at the base point  $x_0$  which converges to the function on some interval with  $x_0$  in its interior is said to be *analytic at  $x_0$* .

**EXAMPLE 3.** Find the Taylor series expansion of the function  $f(x) = 1/(1-x)$  at the base point  $-1$  and determine its interval of validity.

Solution We have seen in Example 4, Section 5.2 that  $f'(x) = \frac{1}{(1-x)^2}$ ,  $f''(x) = \frac{2}{(1-x)^3}$ , and in general that  $f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}}$ . Substituting  $x = -1$ , we obtain  $f(-1) = \frac{1}{2}$ ,  $f'(-1) = \frac{1}{2^2}$ , and in general  $f^{(k)}(-1) = \frac{k!}{2^{k+1}}$ . Thus the Taylor series expansion at the base point  $-1$  is

$$\begin{aligned} & f(-1) + \frac{f'(-1)}{1!}(x+1) + \frac{f''(-1)}{2!}(x+1)^2 + \cdots + \frac{f^{(k)}(-1)}{k!}(x+1)^k + \cdots \\ &= \frac{1}{2} + \frac{1}{2^2}(x+1) + \frac{1}{2^3}(x+1)^2 + \cdots + \frac{1}{2^{k+1}}(x+1)^k + \cdots \\ &= \sum_{k=0}^{\infty} \frac{1}{2} \cdot \left(\frac{x+1}{2}\right)^k. \end{aligned}$$

This is a geometric series with first term  $1/2$  and ratio  $(x+1)/2$ . Thus it converges to the sum

$$\frac{\frac{1}{2}}{1 - \frac{x+1}{2}} = \frac{1}{2 - (x+1)}$$

provided  $\frac{|x+1|}{2} < 1$ , or  $-1 < \frac{x+1}{2} < 1$ , or  $-2 < x+1 < 2$ , or  $-3 < x < 1$ .

## 9.5 Convergence of Power Series

In Example 1, Section 5.2, we showed that the Taylor polynomial  $P_n(x)$  for the function  $e^x$  with base point 0 is

$$1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} = \sum_{k=0}^n \frac{x^k}{k!}.$$

Thus the Taylor series of the function  $e^x$  is

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}. \quad (9.32)$$

Viewing (9.32) purely as an infinite series, paying no attention to its relation to the function  $e^x$ , we may ask for what values of  $x$  it converges. We will see that the series (9.32) converges absolutely for all  $x$ ,  $-\infty < x < \infty$ . This means that the series defines a function  $f(x)$  of  $x$  for  $-\infty < x < \infty$ , namely the function whose value at  $x$  is the sum of the series of  $x$ . What is the relation of this function  $f(x)$  to  $e^x$ ?

As we have seen in Section 5.1, since

$$e^x = \sum_{k=0}^n \frac{x^k}{k!} + R_n(x)$$

and the remainder  $R_n(x) \rightarrow 0$  as  $n \rightarrow \infty$  on every interval  $-A \leq x \leq A$  with  $A > 0$ , the function  $f(x)$  represented by the power series (9.32) must be  $e^x$ . Thus the Taylor series of  $e^x$  converges to  $e^x$  on every closed bounded interval  $-A \leq x \leq A$ .

We may use the same approach for many other functions. If  $f(x)$  is a function whose derivatives of all orders exist on some interval, we may construct the Taylor series of  $f(x)$ , obtaining a power series

$$\sum_{k=0}^{\infty} a_k x^k,$$

with  $a_k = \frac{f^{(k)}(0)}{k!}$  ( $k = 0, 1, 2, \dots$ ). It is possible to prove that if we define

$$R = \frac{1}{\lim_{k \rightarrow \infty} \left| \frac{a_{k+1}}{a_k} \right|} = \lim_{k \rightarrow \infty} \left| \frac{a_k}{a_{k+1}} \right|, \quad (9.33)$$

(provided this limit exists), then the power series converges if  $|x| < R$ , that is, for  $x$  in the interval  $-R < x < R$ . Actually, a stronger statement is true, namely that the series converges *absolutely*, that is, the series of absolute values

$$\sum_{k=0}^{\infty} |a_k| |x|^k$$

converges. Even if the limit (9.33) does not exist, it is possible to show that every power series  $\sum_{k=0}^{\infty} a_k x^k$  has a *radius of convergence*  $R$  such that the series converges absolutely if  $|x| < R$  and (if  $R$  is finite) diverges if  $|x| > R$ . For  $|x| = R$ , that is for  $x = \pm R$ , the series may converge or diverge. The interval  $-R < x < R$  on which the series converges is known as the *interval of convergence*. Since the Taylor series (9.32) for the function  $e^x$  converges for all  $x$ , the radius of convergence  $R$  is infinite and the interval of convergence is  $-\infty < x < \infty$ .

**EXAMPLE 1.** Find the interval of convergence of the Taylor series  $1 + x + x^2 + \cdots$  of the function  $1/(1-x)$ .

Solution. We know that this geometric series converges for  $|x| < 1$  and diverges for  $|x| \geq 1$ . Thus the interval of convergence is  $-1 < x < 1$ . Observe that the function  $1/(1-x)$  is defined on a larger interval  $-\infty < x < 1$  than the interval of convergence of its Taylor series. However, the Taylor series represents the function only on the interval  $-1 < x < 1$ . In Section 5.2 we obtained the explicit expression  $x^{n+1}/(1-x)$  for the remainder  $R_n(x)$  in the Taylor approximation of this function. Because  $\lim_{n \rightarrow \infty} |x|^{n+1} = 0$  if  $|x| < 1$ ,  $\lim_{n \rightarrow \infty} R_n(x) = 0$  on the interval of convergence of the Taylor series and thus

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \quad (-1 < x < 1). \quad (9.34)$$

**EXAMPLE 2.** Find the interval of convergence of the power series  $\sum_{k=0}^{\infty} k! x^k$

Solution. Calculation of  $R$  using (9.33) gives  $R = 0$ , and thus the series diverges if  $|x| > 0$ . For  $x = 0$ , the series converges because all terms of the series are zero, but there is no interval on which the series converges.

**EXAMPLE 3.** Find the interval of convergence of the Taylor expansion of the function  $f(x) = 1/(1-x)$  at the base point  $-1$ .

Solution. We have seen in Example 4, Section 5.2 that the Taylor series expansion at the base point  $-1$  is

$$\sum_{k=0}^{\infty} \frac{1}{2} \cdot \left( \frac{x+1}{2} \right)^k,$$

and also that this is a geometric series with first term  $\frac{1}{2}$  and ratio  $\frac{x+1}{2}$ . Thus it converges to the sum

$$\frac{\frac{1}{2}}{1 - \frac{x+1}{2}} = \frac{1}{2 - (x+1)}$$

provided  $\frac{|x+1|}{2} < 1$ , or  $-1 < \frac{x+1}{2} < 1$ , or  $-2 < x+1 < 2$ , or  $-3 < x < 1$ . Alternately, we may use (9.33) to calculate  $R = 1$ .

## 9.6 Some exercises

1. Is it possible to find a differentiable function  $f(x)$  with  $-1 \leq f'(x) \leq 1$  for all  $x$ ,  $f(0) = 0$ , and  $f(1) = 2$ ?
2. Show that the equation  $x^3 + ax + b = 0$  with  $a > 0$  and  $b$  arbitrary can not have more than one real root.
3. Consider the function  $f(x) = \frac{x+1}{x-1}$  for which  $f(0) = -1$ ,  $f(2) = 3$  and  $f'(x) < 0$  for all  $x$ . Is it possible to find  $c$ ,  $0 < c < 2$  such that  $f(2) - f(0) = 2f'(c)$ ? If not, why does this not contradict the mean value theorem?
4. (Generalized mean value theorem). Let  $f(x)$  and  $g(x)$  be differentiable functions for  $a \leq x \leq b$ . By applying the mean value theorem to the function  $F(x) = \{f(b) - f(a)\}g(x) - \{g(b) - g(a)\}f(x)$  show that

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}$$

for some  $c$ ,  $a < c < b$ .

5. (Mean value theorem for integrals). By applying the mean value theorem to the function  $F(x) = \int_a^x f(t)dt$  show that  $\int_a^b f(t)dt = (b - a)f(c)$  for some  $c$ ,  $a < c < b$ .

In each of Exercises 5–7 find the Taylor polynomial of the given function  $f(x)$  of the given degree  $n$  at the base point 0.

6.  $f(x) = e^{-x}$ ,  $n = 3$
7.  $f(x) = (1 + x)^{1/2}$ ,  $n = 3$
8.  $f(x) = xe^x$ ,  $n = 3$

In each of Exercises 8–11 find the Taylor polynomial of the given function  $f(x)$  of the given degree  $n$  at the given base point  $x_0$ .

9.  $f(x) = \ln x$ ,  $n = 3$ ,  $x_0 = 1$
10.  $f(x) = e^x$ ,  $n = 3$ ,  $x_0 = 1$
11.  $f(x) = x^{1/2}$ ,  $n = 3$ ,  $x_0 = 1$
12.  $f(x) = e^{-x}$ ,  $n = 3$ ,  $x_0 = \frac{1}{e}$ .
13. (a) Find the Taylor polynomial  $P_n(x)$  of the function  $f(x) = \frac{1}{1+x}$  at the base point 0.  
(b) Obtain an expression for  $f(x) - P_n(x)$ .

14. Find the Taylor polynomial  $P_4(x)$  of the function  $f(x) = e^{x^2}$  at the base point 0
- directly from the definition
  - by finding the Taylor polynomial of the function  $e^x$  and then replacing  $x$  by  $x^2$ .
15. Estimate the error in approximating the given function  $f(x) = xe^x$  by the Taylor polynomial  $P_3(x)$  with base point 0 on the interval  $0 \leq x \leq 0.1$
16. How good is the approximation  $1 + x + \frac{x^2}{2}$  for  $e^x$  on the interval  $|x| < 0.1$ ?

In each of Exercises 16-19, find the Taylor series expansion of the given function  $f(x)$  and the interval on which this expansion is valid.

17.  $f(x) = e^{x^2}$
18.  $f(x) = \frac{1}{1+x^2}$
19.  $f(x) = (1+x)^p$
20.  $f(x) = \ln(1-x)$
21. Define the function  $F(x)$  to be the sum of the power series

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{k+2}}{k+1}$$

- Find the power series expansion of the function  $G(x) = F(x)/x$
- Find the power series expansion of the function  $G'(x)$  and find explicitly (in closed form) the sum of this series.
- Integrate the result of part (c) to find  $G(x)$  (noting that  $G(0) = 0$ ).
- Find  $F(x) = xG(x)$  in closed form

In each of Exercises 21-26, determine the values of  $x$  for which the given series converges absolutely.

22.  $\sum_{k=0}^{\infty} kx^k$
23.  $\sum_{k=0}^{\infty} (-1)^k x^k$
24.  $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k$
25.  $1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \dots + \frac{p(p-1)\dots(p-k+1)}{k!}x^k + \dots$
26.  $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^k$