A Mathematics Survival Guide: Complex Analysis & Differential Equations

Dimitri D. Vvedensky
The Blackett Laboratory,
Imperial College, London SW7 2AZ

First Term (2010-2011)

Contents

1 Complex Numbers **1**
1.1 Imaginary and Complex Numbers 2
1.2 Algebra of Complex Numbers 4
1.2.1 Binary Composition Operations 4
1.2.2 Complex Conjugation 7
1.3 The Complex Plane 8
1.4 Euler's Formula 9

2 Functions of Complex Variables **13**
2.1 Powers of Complex Numbers 14
2.2 Roots of Complex Numbers 16
2.3 Complex Power Series 19
2.4 The Complex Exponential 22
2.4.1 The Cauchy Product 22
2.4.2 Products of Complex Exponentials 23
2.5 Complex Trigonometric Functions 24
2.6 The Complex Logarithm 27
2.7 Complex Powers 29

3 First-Order Differential Equations **35**
3.1 Notation and Nomenclature 36
3.2 Radioactive Decay 37
3.2.1 Method 1: Trial solution 37
3.2.2 Method 2: Separation of Variables 39
3.3 Spread of Epidemics 40
3.4 Nonlinearity and Chaos* 43

4 Second-Order Ordinary Differential Equations **47**
4.1 The Characteristic Equation 48

4.1.1 Case I: Real Distinct Roots 49
4.1.2 Case II: Degenerate Roots 50
4.1.3 Case III: Complex Conjugate Roots 52
4.2 The Harmonic Oscillator 53
4.2.1 Case I: $\gamma^2 - 4\omega_0^2 > 0$ 54
4.2.2 Case II: $\gamma^2 - 4\omega_0^2 = 0$ 55
4.2.3 Case III: $\gamma^2 - 4\omega_0^2 < 0$ 56
4.3 Inhomogeneous Equations 59
4.3.1 Method of Solution 59
4.3.2 Resonance in a Driven Harmonic Oscillator 60
4.4 Summary 62

Chapter 1

Complex Numbers

Complex analysis is one of the most awe-inspiring areas of mathematics. Beginning with the notion of an imaginary unit, there follows an abundance of useful and unexpected results, methods, and concepts. The story begins in the mid 1550s, when the Italian mathematician Girolamo Cardano posed a problem that could not be solved with real numbers, namely, the existence of two numbers whose sum is 10 and whose product is 40. If we call these numbers x and y, then we are looking for a solution of

$$x + y = 10 \quad \text{and} \quad xy = 40\,.$$

By solving these simultaneous equations for, say, y in terms of x, we obtain a quadratic equation in x:

$$x^2 - 10x + 40 = 0\,.$$

The solutions to this equation yield

$$x = 5 \pm \sqrt{-15}\,, \qquad y = 5 \mp \sqrt{-15}\,,$$

from which we see directly that the sum of these numbers is 10 and their product is 40. Cardano did not pursue this, concluding that this result was 'as subtle as it is useless.'

Complex numbers did not arise from this example, but in connection with the solution to cubic equations. Cardano presented formulae for the solutions of certain cubic (and quartic) equations, within which there are square roots of numbers that could be negative. Cardano had serious misgivings about expressions such as $2 + \sqrt{-2}$ and, in fact, referred to thinking about them as 'mental torture.' Rafael Bombelli introduced the symbol i for $\sqrt{-1}$ in 1572 and René Descartes

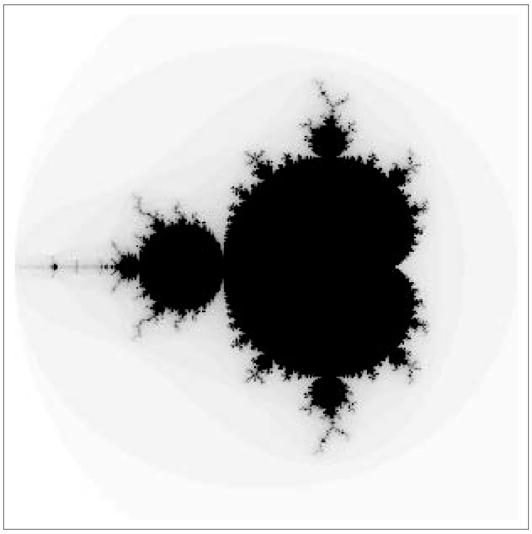


Figure 1.1: The Mandelbrot set, which is a set of complex numbers whose boundary is a fractal.

in 1637 called numbers such as $a + \sqrt{-b}$, where a is any real number and b is a *positive* number, *imaginary numbers*. The term *complex number*, which is the modern term for such numbers, seemed to have originated with Carl Friedrich Gauss in 1831, who also popularized the idea of endowing imaginary quantities with a 'real' existence as points in a plane.

The usage of complex numbers has developed tremendously in the intervening years, and now forms a natural part of coordinate systems, vectors, matrices, and quantum mechanics. As we discover more about advanced physics, complex numbers continue to become ever more significant. Recent examples include fractals, such as the Mandelbrot set (Fig. 1.1), and string theory, which purports to have the potential to be a 'theory of everything.'

1.1 Imaginary and Complex Numbers

The solution of the quadratic equation

$$ax^2 + bx + c = 0\,, \tag{1.1}$$

in which a, b, and c are real constants, is the familiar formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\,. \tag{1.2}$$

The key quantity here is the discriminant: $b^2 - 4ac$. If $b^2 - 4ac > 0$, then there are two real solutions of Eq. (1.1) while, if $b^2 - 4ac = 0$, there is only a single real solution. What happens if $b^2 - 4ac < 0$? There is clearly no *real* solution of Eq. (1.1) in this case. This is the situation faced by Cardano in his formula for the solution of cubic and quartic equations. If we introduce the symbol

$$i \equiv \sqrt{-1}\,,$$

such that $i^2 = -1$, then we have at least a symbolic solution to this impasse. The symbol i is called the **imaginary unit**.

EXAMPLE 1.1. Consider the quadratic equation

$$x^2 + 4 = 0\,. \tag{1.3}$$

There are no real solutions to this equation, but by writing this equation as

$$x^2 = -4\,,$$

we have that

$$x = \sqrt{-4} = \sqrt{-1 \times 4} = \sqrt{-1} \times \sqrt{4} = \pm 2\,i\,. \tag{1.4}$$

Thus, there are two solutions to Eq. (1.3): $x = -2\,i$ and $x = 2\,i$. ▮

Expressions of the form $a\,i$, in which a is any real number, are called **imaginary numbers**.

EXAMPLE 1.2. Consider the quadratic equation

$$2x^2 - 2x + 1 = 0\,.$$

According to the quadratic formula (1.2), the solutions to this equation are

$$x = \frac{2 \pm \sqrt{4-8}}{4} = \frac{2 \pm \sqrt{-4}}{4} = \frac{1}{2} \pm \sqrt{-\frac{1}{4}}\,.$$

By proceeding as in (1.4), we find that the two solutions of this equation are

$$x = \frac{1}{2} - \frac{i}{2}\,, \qquad x = \frac{1}{2} + \frac{i}{2}\,.$$

▮

Expressions of the form $a + b\,i$, in which a and b are real numbers, are called **complex numbers**. The number a is called the **real part** of the complex number and the number b, the coefficient of i, is called the **imaginary part**. When a complex number is a variable, the conventional notation is $z = x + i\,y$, where x, the real part of z is denoted as $\mathrm{Re}(z)$ and y, the imaginary part of z, as $\mathrm{Im}(z)$:

$$\boxed{z = x + i\,y\,; \quad \mathrm{Re}(z) = x\,, \quad \mathrm{Im}(z) = y} \tag{1.5}$$

1.2 Algebra of Complex Numbers

The algebra of complex numbers is similar to that for real numbers, with the proviso that the imaginary unit i has the property that $i^2 = -1$. The rules below show how to add, subtract, multiply, and divide complex numbers to obtain a result that is of the form $x + i\,y$. In deriving these rules, we will need two properties of complex numbers. Two complex numbers $a + i\,b$ and $c + i\,d$, in which a, b, c, and d are real numbers, are equal if and only if their real and imaginary parts are separately equal:

$$a + i\,b = c + i\,d \quad \text{if and only if} \quad a = c \quad \text{and} \quad b = d\,.$$

As a special case of this statement, we have that a complex number $a + i\,b$ is equal to zero if only and only if the real and imaginary parts are each equal to zero:

$$a + i\,b = 0 \quad \text{if and only if} \quad a = 0 \quad \text{and} \quad b = 0\,. \tag{1.6}$$

1.2.1 Binary Composition Operations

Consider the addition of two complex numbers $a + i\,b$ and $c + i\,d$, in which a, b, c, and d are real numbers:

$$(a + i\,b) + (c + i\,d) = (a + c) + i\,(b + d)\,, \tag{1.7}$$

that is, the real and imaginary parts are added separately. The rule for subtraction is similarly applied:

$$(a + i\,b) - (c + i\,d) = (a - c) + i\,(b - d)\,.$$

The multiplication of two complex numbers proceeds with the usual rule for distributive property of multiplication over addition:

$$\begin{aligned}(a+i\,b)(c+i\,d) &= a(c+i\,d)+i\,b(c+i\,d)\\ &= ac+i\,ad+i\,bc+i^2bd\\ &= (ac-bd)+i\,(ad+bc)\,. \end{aligned} \tag{1.8}$$

The division of complex numbers proceeds in two steps. Consider the quotient

$$\frac{a+i\,b}{c+i\,d}\,.$$

We first multiply the numerator and denominator by the quantity $c-i\,d$,

$$\left(\frac{a+i\,b}{c+i\,d}\right)\left(\frac{c-i\,d}{c-i\,d}\right)=\frac{(a+i\,b)(c-i\,d)}{(c+i\,d)(c-i\,d)}\,.$$

and then carry out the multiplication in the numerator and denominator, using Eq. (1.8):

$$\begin{aligned}\frac{a+i\,b}{c+i\,d} &= \frac{(ac+bd)+i\,(bc-ad)}{c^2+d^2}\\ &= \frac{ac+bd}{c^2+d^2}+i\,\frac{bc-ad}{c^2+d^2}\,,\end{aligned} \tag{1.9}$$

where we must mandate that $c+i\,d\neq 0$ for this expression to be meaningful which, according to Eq. (1.6), means that $c\neq 0$ and $d\neq 0$. Notice that, in obtaining this quotient, we have used the fact that $i\times(-i)=1$, which is a particular case of Eq. (1.8) with $a=c=0$, $b=1$ and $d=-1$. Note also the following special cases of Eq. (1.9). If $a=1$ and $b=0$, we obtain the reciprocal of a complex number as

$$\frac{1}{c+i\,d}=\frac{c}{c^2+d^2}-i\,\frac{d}{c^2+d^2}\,. \tag{1.10}$$

If in this equation we first set $c=0$, then setting $d=1$ and $d=-1$ in turn, produces

$$\frac{1}{i}=-i \quad \text{and} \quad \frac{1}{-i}=i\,,$$

respectively.

To summarize, the algebraic rules for combining complex numbers are

For real numbers a, b, c, and d

$$(a + i\,b) \pm (c + i\,d) = (a \pm c) + i\,(b \pm d)\,,$$

$$(a + i\,b)(c + i\,d) = (ac - bd) + i\,(ad + bc)\,,$$

$$\frac{a + i\,b}{c + i\,d} = \frac{ac + bd}{c^2 + d^2} + i\,\frac{bc - ad}{c^2 + d^2}\,,$$

where, in the last operation, $c \neq 0$ and $d \neq 0$.

EXAMPLE 1.3. The sum of the complex numbers $1 + i$ and $2 - 3\,i$ is carried out according to Eq. (1.7):

$$(1 + i) + (2 - 3\,i) = 3 - 2\,i\,.$$

The square of the complex number $1 + i$ is calculated according to the product rule in Eq. (1.8):

$$\begin{aligned}(1 + i)^2 &= (1 + i)(1 + i) = 1 + 2\,i + i^2 \\ &= 1 + 2\,i - 1 = 2\,i\,.\end{aligned}$$

Finally, the quotient of the complex numbers $1 - i$ and $1 + i$ is calculated according to Eq. (1.9):

$$\begin{aligned}\frac{1 - i}{1 + i} &= \frac{(1 - i)(1 - i)}{(1 + i)(1 - i)} \\ &= \frac{1 - 2\,i + (-i)^2}{1 + i - i + i(-i)} = -\frac{2\,i}{2} = -i\,.\end{aligned}$$

▮

1.2.2 Complex Conjugation

Given a complex number $z = x + i\,y$, the complex conjugate of z, denoted by z^*, is the complex number

$$\boxed{z^* = x - i\,y\,.}$$

The addition rule for complex numbers in Eq. (1.8) can be used to obtain the real part of a complex number as

$$z + z^* = (x + i\,y) + (x - i\,y) = 2x\,,$$

so

$$\boxed{\mathrm{Re}(z) = \frac{z + z^*}{2}\,.}$$

Similarly, the imaginary part is obtained as

$$z - z^* = (x + i\,y) - (x - i\,y) = 2\,i\,y\,,$$

so

$$\boxed{\mathrm{Im}(z) = \frac{z - z^*}{2\,i}\,.}$$

Another useful property of the conjugate is

$$z\,z^* = (x + i\,y)(x - i\,y) = x^2 + y^2\,.$$

In particular, the reciprocal of z can be written as

$$\frac{1}{z} = \frac{1}{z}\left(\frac{z^*}{z^*}\right) = \frac{z^*}{zz^*} = \frac{x}{x^2 + y^2} - i\,\frac{y}{x^2 + y^2}\,,$$

which agrees with Eq. (1.10).

1.3 The Complex Plane

The conventional notation $z = x + i\,y$ of a generic complex number suggests a graphical representation of complex numbers based on Cartesian coordinates wherein z is associated with the ordered pair (x, y). The real axis corresponds to the abscissa in this coordinate system and the imaginary part to the ordinate, as depicted in Fig. 1.2(a). Thus, every point corresponds to a complex number. When the x-y plane is used to represent complex numbers in this way, it is referred to as the **complex plane** or as an **Argand diagram**.

An alternative representation of complex numbers that, for many purposes, is more convenient than rectangular coordinates, is one based on *polar coordinates*. The basic construction is shown in Fig. 1.2(b). A straight line runs from the origin to the point (x, y). This line is characterized by two quantities: a length r that measures its length, and an angle θ that specifies the angle between this line and the x-axis. By convention, *positive* angles are taken in the *counterclockwise* direction from the x-axis. The relationship between the rectangular and the polar representations of complex numbers can be obtained from basic trigonometry. The rectangular coordinates corresponding to the point (r, θ) are

$$x = r\cos\theta\,, \qquad y = r\sin\theta\,,$$

and the polar coordinates of a point (x, y) are

$$r = \sqrt{x^2 + y^2}\,, \qquad \theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{1.11}$$

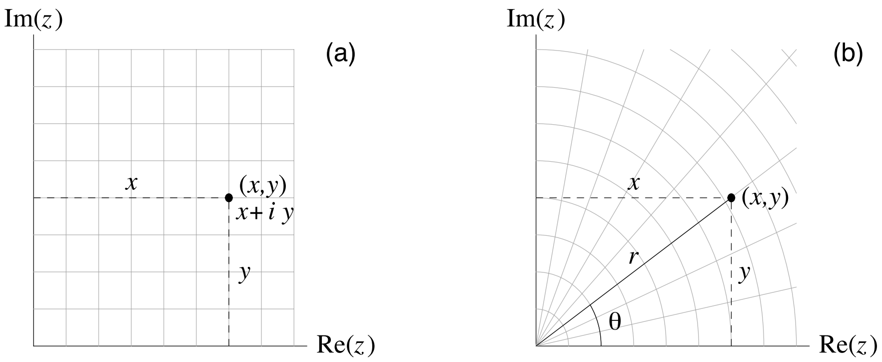


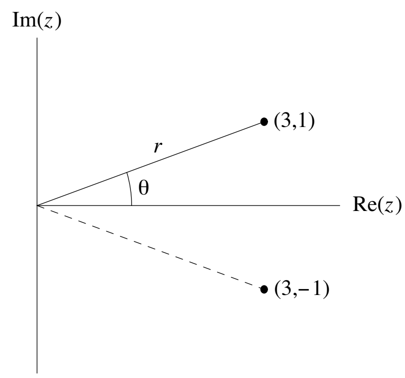
Figure 1.2: The representation of a complex number $z = x + i\,y$ in (a) rectangular and (b) polar coordinates.

Hence, we can write the polar form of a complex number $z = x + i\,y$ as

$$\boxed{z = x + i\,y = r(\cos\theta + i\,\sin\theta)\,.} \tag{1.12}$$

The quantity r is called the **modulus** or **magnitude** of z, and θ is called the **argument** or **phase**. In the next section we will find a more compact way of writing the right-hand side of this expression that will have many far-reaching consequences.

EXAMPLE 1.4. Consider the complex number $z = 3 + i$, whose position in the complex plane is shown at right. The polar representation of this number is specified by the modulus r and argument θ, which are obtained from the relations in (1.11) as



$$r = \sqrt{3^2 + 1^2} = \sqrt{10}\,,$$

$$\theta = \tan^{-1}\left(\tfrac{1}{3}\right).$$

Also shown in the figure at right is the complex conjugate $z^* = 3 - i$ of z. The radius is the same as that for z, but the argument is now $-\theta$. This is immediately apparent from Eq. (1.12):

$$z^* = x - i\,y = r(\cos\theta - i\,\sin\theta) = r\big[\cos(-\theta) + i\,\sin(-\theta)\big]\,.$$

Notice that the complex conjugate of a complex number is obtained by reflection across the real axis. ▮

1.4 Euler's Formula

The polar representation of a complex number in Eq. (1.12) contains the factor $\cos\theta + i\,\sin\theta$. We will show in this section that this factor has special properties. Suppose that we differentiate with respect to θ:

$$\frac{d}{d\theta}(\cos\theta + i\,\sin\theta) = -\sin\theta + i\,\cos\theta\,.$$

By writing the negative sign in front of the sine term as i^2, we obtain

$$\frac{d}{d\theta}(\cos\theta + i\,\sin\theta) = i^2\sin\theta + i\,\cos\theta = i(\cos\theta + i\,\sin\theta)\,,$$

so the effect of the differentiation is simply to multiply the original expression by a factor of i. Each successive derivative of this expression yields another multiplicative factor of i. Thus, the n derivative is

$$\frac{d^n}{dx^n}(\cos\theta + i\,\sin\theta) = i^n(\cos\theta + i\,\sin\theta)\,.$$

We now have the required mathematical input to perform a Taylor series expansion of the function $\cos\theta + i\,\sin\theta$ about $\theta = 0$. Recall that, for a function $f(x)$, the Taylor series about $x = 0$ (which is called a Maclaurin series) has the following form:

$$\begin{aligned} f(x) &= f(0) + f^{(1)}(0)x + \frac{1}{2!}f^{(2)}(0)x^2 + \frac{1}{3!}f^{(3)}(0)x^3 + \cdots \\ &= \sum_{n=0}^{\infty} f^{(n)}(0)\frac{x^n}{n!}\,, \end{aligned}$$

in which the 'zeroth derivative' is the function itself and $0! = 1$. The notation $f^{(n)}(0)$ means that we take the n derivative of f and then set $x = 0$. For the function at hand, we have that $\cos 0 = 1$ and $\sin 0 = 0$. Hence, for $n = 1, 2, \ldots,$

$$\left[\frac{d^n}{dx^n}(\cos\theta + i\,\sin\theta)\right]\bigg|_{\theta=0} = i^n\,,$$

from which we obtain the following Taylor series:

$$\begin{aligned} \cos\theta + i\,\sin\theta &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \cdots \\ &= \sum_{n=0}^{\infty}\frac{(i\theta)^n}{n!}\,. \end{aligned} \tag{1.13}$$

To appreciate the significance of this result, consider the Taylor series for e^x. Using the fact that

$$\frac{d(e^x)}{dx} = e^x\,,$$

and $e^0 = 1$, we obtain

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} . \tag{1.14}$$

The comparison of Eqs. (1.13) and (1.14) suggests the identification

$$\boxed{\cos\theta + i\,\sin\theta = e^{i\,\theta}\,.} \tag{1.15}$$

This is known as **Euler's formula**. The revolutionary nature of this formula becomes evident when we evaluate both sides for $\theta = \pi$. With $\cos\pi = -1$ and $\sin\pi = 0$, we obtain $e^{i\pi} = -1$. The exponential function e^x for real x is *never* negative. But, by allowing for an imaginary variable in the argument of the exponential, negative values arise quite naturally. Indeed, the familiar rules for the manipulation of the exponential function, which can derived from the Taylor series (1.14),[1] combined with imaginary arguments, leads to some new and useful results. Equations (1.12) and the polar representation in (1.15) imply an especially compact representation of $z = x + i\,y$:

$$\boxed{z = r\,e^{i\theta}\,.} \tag{1.16}$$

EXAMPLE 1.5. Consider the multiplication of two complex numbers with unit modulus ($r = 1$) and with arguments θ and ϕ:

$$\cos\theta + i\,\sin\theta = e^{i\,\theta}$$
$$\cos\phi + i\,\sin\phi = e^{i\,\phi}\,.$$

The product of these equations is

$$(\cos\theta + i\,\sin\theta)(\cos\phi + i\,\sin\phi) = e^{i\,\theta}e^{i\,\phi}\,.$$

Expanding the left-hand side according to Eq. (1.8) yields

$$\cos\theta\cos\phi - \sin\theta\sin\phi + i(\sin\theta\cos\phi + \cos\theta\sin\phi)\,.$$

[1] We have, for the moment, side-stepped the question of convergence of the series (1.13). This will addressed in the next chapter

By using the standard rule for the product of exponential functions, i.e. $e^a e^b = e^{a+b}$, together with Eq. (1.15), we obtain

$$e^{i\theta} e^{i\phi} = e^{i(\theta+\phi)} = \cos(\theta+\phi) + i\,\sin(\theta+\phi)\,.$$

Equating the real and imaginary parts in the last two equations produces

$$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi\,,$$

$$\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi\,,$$

which are the trigonometric identities for the sine and cosine of the sum of two angles! Other identities can be derived by following analogous steps. Quite apart from providing an additional level of confidence in our usage of Eq. (1.15), this procedure is far easier to apply than the conventional method of deriving such identities. ▮

Finally, we note that the representation (1.16) is especially convenient for taking products and quotients of complex numbers. For any two complex numbers $z = r\,e^{i\theta}$ and $z' = r' e^{i\theta'}$, we have

$$z\,z' = r r' e^{i(\theta+\theta')}\,,$$

$$\frac{z}{z'} = \frac{r}{r'}\,e^{i(\theta-\theta')}\,.$$

Chapter 2

Functions of Complex Variables

The binary operations of addition, subtraction, multiplication, and division of complex numbers are the basis for the assembly of composite expressions of complex quantities, such as polynomials and power series. This opens the way to extending functions of real variables to functions of variables that extend over the complex plane. Functions f of independent variables x and y that depend only on the combination $z = x + iy$ are called **functions of a complex variable** and are denoted by $f(z)$. We will be concerned in this course with what are called the elementary functions: powers and roots, trigonometric functions and their inverses, exponential and logarithmic functions, as well combinations of such functions. These functions can be defined by writing, for example, $\sin z$, e^z, and $\log z$, so that they become complex-valued quantities, having real and imaginary parts. Complex-valued functions can exhibit some quite unexpected behavior compared to their real counterparts, though most of the standard functions are real when their arguments are real. An obvious exception is the square root function, which becomes imaginary for negative arguments.

We will begin this chapter be examining the powers and roots of complex numbers. Since we can multiply z by itself and by any other complex number, we can form any polynomial in z and, by extension, any power series. This will enable us to define the power series of functions such as the exponential and trigonometric functions by their Taylor series expansions. By adapting the discussion of the convergence of real series to complex-valued series, the Taylor series of elementary functions of complex variables will be shown to retain the convergence properties of their real-valued counterparts. Since the properties of these functions can be derived from their Taylor series, they retain their familiar (and useful) properties for real arguments, while exhibiting a richer analytic structure with complex arguments.

2.1 Powers of Complex Numbers

As we noted in Sec. 1.4, products of complex numbers are most easily carried out in the polar form $z = r\,e^{i\theta}$. The n-fold product of z is

$$z^n = \left(r\,e^{i\theta}\right)^n = r^n e^{in\theta}\,.$$

Note that, for the particular case that $r = 1$, we have that

$$\begin{aligned}\left(e^{i\theta}\right)^n &= (\cos\theta + i\,\sin\theta)^n \\ &= e^{in\theta} = \cos n\theta + i\,\sin n\theta\,,\end{aligned}$$

which leads to **De Moivre's theorem**:

$$\boxed{(\cos\theta + i\,\sin\theta)^n = \cos n\theta + i\,\sin n\theta\,.} \tag{2.1}$$

EXAMPLE 2.1. Consider De Moivre's theorem for $n = 2$:

$$\begin{aligned}(\cos\theta + i\,\sin\theta)^2 &= \cos^2\theta - \sin^2\theta + 2i\,\sin\theta\cos\theta \\ &= \cos 2\theta + i\,\sin 2\theta\,.\end{aligned}$$

By equating real and imaginary parts in this equation, we obtain

$$\begin{aligned}\cos 2\theta &= \cos^2\theta - \sin^2\theta\,, \\ \sin 2\theta &= 2\sin\theta\cos\theta\,,\end{aligned}$$

which are standard double-angle trigonometric identities. Higher multiple-angle identities are derived with an analogous procedure. ∎

Powers of complex numbers are used in polynomials and powers series. We consider first an example of successive powers of complex numbers.

EXAMPLE 2.2. Consider the powers of $z = 1 + i$. The successive powers z^n are straightforward to calculate:

$$z^2 = 2\,i, \quad z^3 = -2 + 2\,i, \quad z^4 = -4\,, \quad z^5 = -4 - 4\,i, \quad z^6 = -8\,i.$$

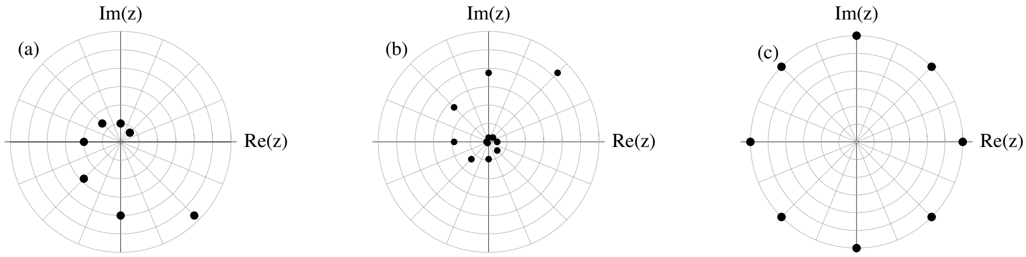


Figure 2.1: The powers z^n of (a) $z = 1 + i$, (b) $z = \frac{1}{2}(1 + i)$, and (c) $z = (1 + i)/\sqrt{2}$. In each case, successive points are rotated by $\frac{1}{4}\pi$ in the counterclockwise direction.

These points are plotted in the complex plane in Fig. 2.1(a). The expanding spiral of the sequence of these powers is evident. The reason for this type of structure can be seen from the polar representation

$$1 + i = \sqrt{2}\, e^{\frac{1}{4}i\pi} ,$$

in terms of which the nth power is

$$(1 + i)^n = (\sqrt{2})^n\, e^{\frac{1}{4}in\pi} = 2^{n/2}\, e^{\frac{1}{4}in\pi} .$$

Thus, each successive power results in a rotation by $\frac{1}{4}\pi$ in the counterclockwise direction and, since $\sqrt{2} > 1$, a *larger* radius, producing the expanding spiral in Fig. 2.1(a).

Alternatively, if we now consider the sequence of powers of $\frac{1}{2}(1 + i)$, the corresponding polar representation yields

$$\left[\tfrac{1}{2}(1 + i)\right]^n = \frac{1}{\sqrt{2}}\, e^{\frac{1}{4}in\pi} = 2^{-n/2}\, e^{\frac{1}{4}in\pi} .$$

Each successive power still results in a rotation by $\frac{1}{4}\pi$ in the counterclockwise direction but, because the modulus $\frac{1}{2}\sqrt{2} < 1$, the corresponding points produce a spiral that converges toward the origin, as shown in Fig. 2.1(b).

The marginal case is obtained for

$$z = \frac{1 + i}{\sqrt{2}} ,$$

which has modulus unity. The polar representations of the successive powers of this complex number are

$$\left[\frac{1}{\sqrt{2}}(1 + i)\right]^n = e^{\frac{1}{4}in\pi} ,$$

which produces a cycle of eight points that are again separated by $\frac{1}{4}\pi$ [Fig. 2.1(c)]. ∎

From these examples, the general expression for the powers of a complex number $x + iy = r\,e^{i\theta}$ is

$$(x + iy)^n = r^n\,e^{in\theta}\,.$$

We can use De Moivre's theorem (2.1) to express these products in rectangular form as

$$\boxed{(x + iy)^n = r^n\,e^{in\theta} = r^n\big(\cos n\theta + i\,\sin n\theta\big)\,.}$$

By invoking the usual rules for exponents, this expression is valid for *all* integers.

2.2 Roots of Complex Numbers

We now consider the inverse process of taking powers of complex numbers, taking their roots. The general problem is to find a solution of the equation $z^n = a + ib$, where $z = x + iy$ and a and b are fixed real numbers. We will focus on the case of integer n in this section, and leave the more general case for a later section. We again proceed by working through an example.

EXAMPLE 2.3. Consider the solution of $z^4 = 16$, i.e. the fourth root of 16. Two solutions, $z = 2$ and $z = -2$, can be obtained without reference to complex numbers. But, to obtain all the roots of this equation, we must use complex numbers. We begin by writing the equation in terms of the polar representation of complex numbers: $z = r\,e^{i\theta}$:

$$(r\,e^{i\theta})^4 = r^4\,e^{4i\theta} = 16\,. \tag{2.2}$$

The equation for the modulus, $r^4 = 16$, yields $r = 2$, since $r \geq 0$ always. To obtain the solutions for the argument, we first write

$$16 = 16\,e^{2n\pi i}\,, \tag{2.3}$$

for $n = 0, 1, 2, \ldots$, which clearly leaves the value of the modulus unaffected since $e^{2n\pi i} = 1$ for any integer n. Thus, our solution for r is therefore also unaffected. However, the solutions for the arguments of our roots now read

$$4\theta = 0,\ 2\pi,\ 4\pi,\ 6\pi,\ 8\pi, \ldots\,. \tag{2.4}$$

The reason for considering the additional angles in (2.3) becomes apparent when we divide both sides of this equation by 4 to obtain the solutions

$$\theta = 0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2},\ 2\pi, \ldots . \tag{2.5}$$

Since $\theta = 0$ and $\theta = 2\pi$ correspond to the same angle, and since each successive rotation by 2π on the right-hand side of (2.4) yields an additive factor of $\frac{1}{4}\pi$ in (2.5), we find that there are four *distinct* values of θ. These yield four solutions of (2.2):

$$\begin{aligned} z_1 &= 2\,e^{i0} = 2\,, \\ z_2 &= 2\,e^{\frac{1}{2}i\pi} = 2\,i\,, \\ z_3 &= 2\,e^{i\pi} = -\,2\,, \\ z_4 &= 2\,e^{\frac{3}{2}i\pi} = -\,2\,i\,. \end{aligned}$$

Thus, in addition to two real roots, we have also found two imaginary roots, which are complex conjugates (cf. Problem 8 of Classwork 1). ▮

EXAMPLE 2.4. Consider the solution of $z^3 = 1 + i$, i.e. the cube roots of $1 + i$. The most expedient method of solution is again based on the polar representation of complex numbers:

$$z = r\,e^{i\theta}\,, \qquad 1 + i = \sqrt{2}\,e^{\frac{1}{4}i\pi}\,. \tag{2.6}$$

The equation to be solved is

$$\begin{aligned} (r\,e^{i\theta})^3 &= r^3\,e^{3i\theta} \\ &= \sqrt{2}\,e^{\frac{1}{4}i\pi} = \sqrt{2}\,e^{\frac{1}{4}i\pi + 2n\pi i}\,, \end{aligned} \tag{2.7}$$

where we have again added multiples of 2π to the argument of $1+i$. The equation for the modulus, $r^3 = \sqrt{2}$, yields $2^{1/6}$. For the argument, we have

$$3\theta = \frac{\pi}{4}, \frac{\pi}{4} + 2\pi,\ \frac{\pi}{4} + 4\pi,\ \frac{\pi}{4} + 6\pi, \ldots .$$

Dividing both sides of this equation by 3, we obtain the solutions

$$\begin{aligned} \theta &= \frac{\pi}{12}, \frac{\pi}{12} + \frac{2\pi}{3}, \frac{\pi}{12} + \frac{4\pi}{3}, \frac{\pi}{12} + \frac{6\pi}{3}, \ldots \\ &= \frac{\pi}{12}, \frac{9\pi}{12}, \frac{17\pi}{12}, \frac{25\pi}{12}, \ldots . \end{aligned}$$

Since

$$\frac{25\pi}{12} = \frac{\pi}{12} + 2\pi,$$

there are three distinct values of θ:

$$\theta = \frac{\pi}{12}, \frac{9\pi}{12}, \frac{17\pi}{12} = 15^\circ, \ 135^\circ, \ 255^\circ \, .$$

The three solutions of Eq. (2.7) are therefore given by

$$z_1 = 2^{1/6} \, e^{\frac{1}{12} i\pi}$$
$$z_2 = 2^{1/6} e^{\frac{9}{12} i\pi}$$
$$z_3 = 2^{1/6} e^{\frac{17}{12} i\pi} \, .$$

▮

There are several points about the calculations in Examples 2.2 and 2.3 that are general characteristics of roots of complex numbers. The nth root of a complex number has n distinct solutions, whose points in the complex plane are separated by an angle of $2\pi/n$. When connected by straight lines, these points form the vertices of a regular n-sided polygon. Figure 2.2 shows the roots calculated in Examples 2.3 and 2.4. In Example 2.3, the quartic roots of 16 lie on a circle of radius 2, separated by $\frac{1}{2}\pi = 90^\circ$, forming a square, while the cube roots of $1 + i$ in Example 2.3 lie on a circle of radius $2^{1/6}$ and are separated by $\frac{2}{3}\pi = 120^\circ$, forming an equilateral triangle.

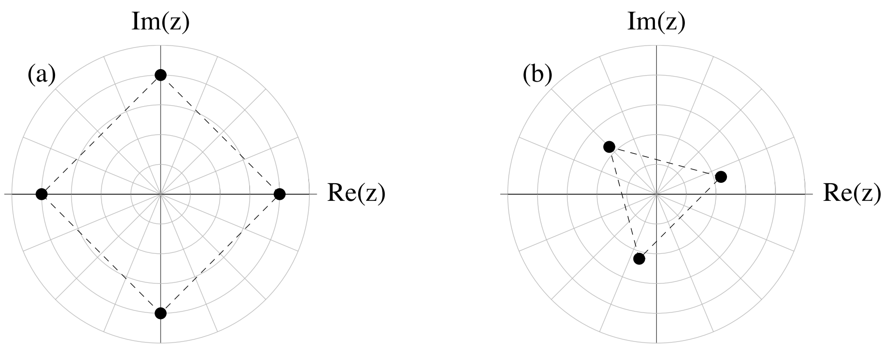


Figure 2.2: The solutions of (a) $z = 16^{1/4}$ and (b) $z = (1 + i)^{1/3}$ plotted in the complex plane.

We can now deduce the general expression for the roots of a complex number. The solutions of $z^n = a + ib$, where a and b are fixed real numbers, are obtained by first writing this equation in polar form: $r^n\, e^{in\theta} = \rho\, e^{i\phi}$, where ρ and ϕ are the modulus and argument, respectively, of $a + ib$. The nth roots of this number each have modulus $\rho^{1/n}$ and arguments given by the n values

$$\theta = \frac{\phi}{n}, \ \frac{\phi + 2\pi}{n}, \ \frac{\phi + 4\pi}{n}, \ldots, \frac{\phi + 2\pi(n-1)}{n} .$$

By invoking De Moivre's theorem (2.1) we can write this result as follows. Given a complex number $a + i\,b$ whose polar form is $\rho\, e^{i\phi}$, the solution of $z^n = \rho\, e^{i\phi}$ is given by the n complex numbers z_k, for $k = 0, \ldots, n-1$:

$$\begin{aligned} z_k &= \rho^{1/n} \exp\left[i\left(\frac{\phi}{n} + \frac{2k\pi}{n}\right)\right] \\ &= \rho^{1/n}\left[\cos\left(\frac{\phi}{n} + \frac{2k\pi}{n}\right) + i\,\sin\left(\frac{\phi}{n} + \frac{2k\pi}{n}\right)\right]. \end{aligned}$$

We can now understand how the plot of the nth roots of a complex number $\rho\, e^{i\phi}$ appear on the complex plane. There are n equally spaced points on a circle of radius $\rho^{1/n}$, with adjacent points separated by an angle of $2\pi/n$. The line that connects the points is a regular n-sided polygon, so the roots are the vertices of this polygon The polygon is tilted by ϕ/n, so that if $\phi = 0$, i.e. of the number whose root is taken is real, at least one vertex (root) lies on the real axis. Thus, even without doing any calculations, the general features of the nth roots of a complex number can be easily identified.

2.3 Complex Power Series

A polynomial of order n in the real variable x has the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n ,$$

in which the coefficients a_k are real numbers. A polynomial is a continuous function of x and is finite for any finite value of x. Only for $x \to \pm\infty$ does the polynomial become infinite. The derivatives of all orders exist and are continuous, although all but the first n are identically zero.

Maclaurin showed that the notion of a polynomial could be generalized to a function represented by an infinite series

$$a_0 + a_1 x + a_2 x^2 + \cdots , \tag{2.8}$$

where

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{1}{n!}\frac{d^n f}{dx^n}\bigg|_{x=0} .$$

If the derivatives are bounded, then the series converges to $f(x)$ everywhere. Remarkably, the values of the function everywhere are determined by the values of the function and its derivatives at the origin. By the 'convergence' of an infinite series, we mean that the sum of enough terms is as close to a fixed value S as required, where S is the sum of the series. In other words, the sum of the series approaches the value S as a limit. Taylor showed that the expansion could be taken about any point $x = a$, not just the origin, thus generalizing the Maclaurin series. We have based our discussion on real variables x, but polynomials and infinite series can be constructed where x is replaced by the complex variable $z = x + iy$.

There are several tests that determine if an infinite series converges. One of the simplest to apply is the **ratio test**. For an infinite series of complex numbers $\sum_{n=0}^{\infty} A_n$, consider the ratio ρ, defined by

$$\rho = \lim_{n\to\infty} \left| \frac{A_{n+1}}{A_n} \right| .$$

The complex series converges if $\rho < 1$ and diverges if $\rho > 1$. If $\rho = 1$, the test is inconclusive and another test for convergence must be used.

EXAMPLE 2.5. Consider the complex geometric series

$$\sum_{n=0}^{\infty} z^n = 1 + z + z^2 + \cdots . \tag{2.9}$$

In the notation used above, $A_n = z^n$. Then, the quantity ρ in the ratio test is

$$\rho = \lim_{n\to\infty} \left| \frac{z^{n+1}}{z^n} \right| = |z| .$$

Thus, the geometric series (2.9) converges if $|z| < 1$. Since $|z| = 1$ can be written in polar form as $\sqrt{x^2 + y^2} = 1$, which is equivalent to $x^2 + y^2 = 1$, the geometric series converges for complex numbers z within a circle of unit radius centered at

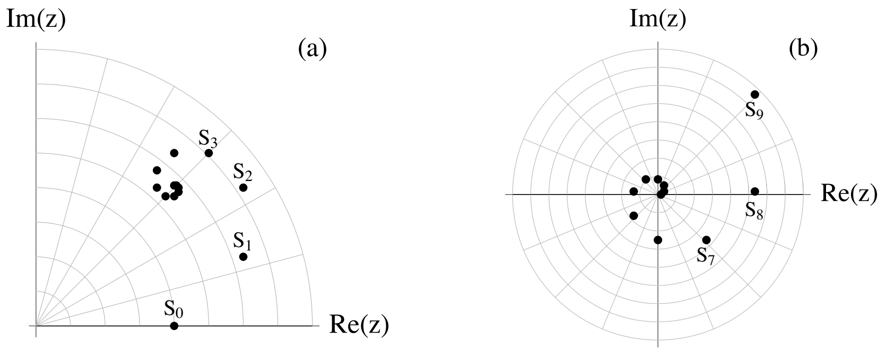


Figure 2.3: The partial sums of the geometric series (2.9) for (a) $z = \frac{1}{2}(1+i)$ and (b) $z = 1+i$. In (a) the partial sums are seen to converge toward $1+i$, which is the sum of the series, while in (b) the partial sums form an expanding spiral because $1+i$ lies outside the radius of convergence of this series.

the origin. For this reason, we refer to the **radius of convergence** of a series (even when the series resides on the real line).

Figure 2.3 shows the sequence of partial sums

$$S_N = \sum_{n=0}^{N} z^n$$

for $z = \frac{1}{2}(1+i)$, which lies within the radius of convergence of the geometric series (2.9), and for $z = 1+i$, which lies outside of this radius. The convergence in the first case is seen by the spiral that converges toward $1+i$, which is the sum of the series. In the second case, however, the partial sums lie on a diverging spiral, which is indicative of the divergence of the series. ∎

For our purposes, the most important infinite series is the Maclaurin series for e^z, which is defined by

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}\,. \tag{2.10}$$

This is a generalization of the function e^x for real x and a generalization of the function $e^{i\theta}$ that we introduced in Sec. 1.4. In the notation of the ratio test, we have $A_n = z^n/n!$, so

$$\rho = \lim_{n\to\infty} \left| \frac{z^{n+1} n!}{z^n (n+1)!} \right| = \lim_{n\to\infty} \frac{|z|}{n+1} = 0\,.$$

Thus, for any complex number with a finite modulus, we obtain $\rho < 1$, so the series (2.10) has an infinite radius of convergence, and we can now confidently make the identification

$$\boxed{e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}}\,, \tag{2.11}$$

for any complex number $z = x + iy$. This representation is valid over the entire complex plane and subsumes the special cases just mentioned, namely, e^x and $e^{i\theta}$. As an immediate consequence, we have

$$e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos y + i \sin y)\,.$$

We will explore the complex exponential in more detail in the next section.

2.4 The Complex Exponential

We have seen in the preceding section that the complex function e^z converges everywhere in the complex plane. In this section, we will show that this function has the all the properties of the real function e^x and, moreover, that these follow from the series representation (2.11).

2.4.1 The Cauchy Product

We first derive the product of power series. Suppose that we have two series,

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots\,.$$

$$\sum_{n=0}^{\infty} b_n = b_0 + b_1 + b_2 + \cdots\,.$$

Their product is formed as follows:

$$\begin{aligned}\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) &= \big(a_0 + a_1 + a_2 + \cdots\big)\big(b_0 + b_1 + b_2 + \cdots\big) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots\,.\end{aligned}$$

Note that this product has been written as an ascending series of linear combinations of the $a_i b_j$ such that $i + j = n$. Hence, we can write this product as a sum as

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right). \tag{2.12}$$

This is known as the **Cauchy product** of two series.

2.4.2 Products of Complex Exponentials

Consider the familiar rule: $e^{z_1+z_2} = e^{z_1}\, e^{z_2}$ for complex numbers z_1 and z_2. That this follows from the power series (2.11) can be seen by applying the Cauchy product:

$$e^{z_1}\, e^{z_2} = \left(\sum_{n=0}^{\infty} \frac{z_1^n}{n!}\right)\left(\sum_{n=0}^{\infty} \frac{z_2^n}{n!}\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{z_1^k}{k!}\frac{z_2^{n-k}}{(n-k)!}\,. \tag{2.13}$$

The right-hand side has been obtained from (2.12) by making the identifications

$$a_n = \frac{z_1^n}{n!}\,, \qquad b_n = \frac{z_2^n}{n!}\,.$$

The interior sum on the right-hand side of (2.13) can be simplified by multiplying and dividing by $n!$:

$$\sum_{k=0}^{n} \frac{z_1^k}{k!}\frac{z_2^{n-k}}{(n-k)!} = \frac{n!}{n!}\sum_{k=0}^{n} \frac{z_1^k}{k!}\frac{z_2^{n-k}}{(n-k)!} = \frac{1}{n!}\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} z_1^k z_2^{n-k}\,.$$

By recalling the binomial theorem,

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^k b^{n-k}\,,$$

we see that

$$\frac{1}{n!}\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} z_1^k z_2^{n-k} = \frac{(z_1+z_2)^n}{n!}\,,$$

from which we conclude that

$$e^{z_1}\, e^{z_2} = \sum_{n=0}^{\infty} \frac{(z_1+z_2)^n}{n!} = e^{z_1+z_2}\,. \tag{2.14}$$

As special cases, we have

$$e^z = e^{x+iy} = e^x\, e^{iy} = e^x(\cos y + i \sin y)\,,$$

and

$$\frac{1}{e^z} = \frac{1}{e^x(\cos y + i\sin y)} = \frac{e^{-x}}{\cos y + i \sin y}\left(\frac{\cos y - i\sin y}{\cos y - i \sin y}\right)$$
$$= e^{-x}(\cos y - i\sin y) = e^{-x}\, e^{-iy} = e^{-(x+iy)} = e^{-z}\,.$$

Thus, combining this result with that in (2.14), we find that, for any integer n, $(e^z)^n = e^{nz}$.

The main properties of the complex exponential function are summarized below:

$$e^z = e^x(\cos y + i\sin y)\,,$$

$$|e^z| = e^x > 0\,,$$

$$e^{z_1+z_2} = e^{z_1}\, e^{z_2}\,,$$

$$e^{-z} = \frac{1}{e^z}\,,$$

$$(e^z)^n = e^{nz}\,, \quad \text{for any integer } n\,.$$

2.5 Complex Trigonometric Functions

The properties of the complex exponential function can be used to define trigonometric functions with complex arguments. We begin with Euler's formula (1.15):

$$e^{i\theta} = \cos\theta + i\sin\theta\,, \tag{2.15}$$

$$e^{-i\theta} = \cos\theta - i\sin\theta\,. \tag{2.16}$$

Taking the sum of these equations causes the imaginary part to vanish, leaving

$$e^{i\theta} + e^{-i\theta} = 2\cos\theta\,,$$

or,

$$\boxed{\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}\,.}$$

Similarly, subtracting (2.16) from (2.15) yields

$$e^{i\theta} - e^{-i\theta} = 2i\,\sin\theta\,,$$

or,

$$\boxed{\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}\,.}$$

With the trigonometric functions expressed in terms of the exponential function, we can now examine the properties of these functions with complex arguments. Consider first the complex cosine function:

$$\begin{aligned}\cos z = \cos(x+iy) &= \tfrac{1}{2}\big[e^{i(x+iy)} + e^{-i(x+iy)}\big]\\ &= \tfrac{1}{2}\big(e^{ix}\,e^{-y} + e^{-ix}\,e^{y}\big)\\ &= \tfrac{1}{2}\big[(\cos x + i\,\sin x)\,e^{-y} + (\cos x - i\,\sin x)\,e^{y}\big]\\ &= \cos x\left(\frac{e^{y}+e^{-y}}{2}\right) - i\,\sin x\left(\frac{e^{y}-e^{-y}}{2}\right)\\ &= \cos x\cosh y - i\,\sin x\sinh y\,, \qquad (2.17)\end{aligned}$$

where we have used (2.28) to identify the hyperbolic functions $\cosh y$ and $\sinh y$. Since the trigonometric and hyperbolic functions are real-valued functions, we have that

$$\mathrm{Re}(\cos z) = \cos x\cosh y\,, \qquad \mathrm{Im}(\cos z) = -\sin x\sinh y\,.$$

Note, in particular, the special case where $x = 0$:

$$\boxed{\cos iy = \cosh y\,,}$$

so the hyperbolic cosine corresponds to the cosine of an imaginary angle. We could also have deduced this from the exponential representations of these functions. A similar calculation to that leading to (2.17) for the complex sine function yields

$$\sin(x + iy) = \sin x \cosh y + i \cos x \sinh y\,, \tag{2.18}$$

and

$$\boxed{\sin iy = i \sinh y\,.}$$

Equations (2.17) and (2.18) can be used to determine the moduli of $\cos z$ and $\sin z$. For $\cos z$, we have

$$\begin{aligned} |\cos z|^2 &= \cos^2 x \cosh^2 y + \sin^2 x \sinh^2 y \\ &= \cos^2 x (1 + \sinh^2 y) + \sin^2 x \sinh^2 y \\ &= \cos^2 x + \sinh^2 y\,(\cos^2 x + \sin^2 x) \\ &= \cos^2 x + \sinh^2 y\,, \end{aligned}$$

which implies that

$$|\cos z| = \sqrt{\cos^2 x + \sinh^2 y}\,.$$

A similar calculation produces

$$|\sin z| = \sqrt{\sin^2 x + \sinh^2 y}\,.$$

These relations show that $\cos z$ and $\sin z$ are unbounded. For example,

$$|\cos z| = \sqrt{\cos^2 x + \sinh^2 y} \ge \sqrt{\sinh^2 y} = |\sinh y|\,.$$

Referring to (2.28), we see that, since

$$\sinh y = \frac{e^y - e^{-y}}{2} \,,$$

then, as $y \to \infty$, $\sinh y \to \infty$ and, as $y \to -\infty$, $\sinh y \to -\infty$. Hence, $|\cos z| \to \infty$ as $|\mathrm{Im}(z)| \to \infty$.

All of the properties of the complex trigonometric functions follow from the basic relations in (2.17) and (2.18):

$$\cos z = \cos x \cosh y - i \sin x \sinh y \,,$$

$$\sin z = \sin x \cosh y + i \cos x \sinh y \,.$$

The other standard trigonometric functions can be obtained from these relations. For example, the complex tangent $\tan z$ and the complex secant $\sec z$ are defined as

$$\tan z = \frac{\sin z}{\cos z} \,, \qquad \sec z = \frac{1}{\cos z} \,, \ (\cos z \neq 0) \,.$$

2.6 The Complex Logarithm

For real variables, the logarithm is defined as the inverse of the exponential function. Accordingly, if $y = e^x$, then $x = \ln y$, where 'ln' signifies the natural logarithm, i.e. the logarithm to base e. The same relationship exists between the complex exponential and the complex logarithm. For complex numbers z and w,

$$z = e^w \quad \longrightarrow \quad w = \ln z \,. \tag{2.19}$$

We see immediately from this definition that, for complex numbers z_1, z_2, w_1, and w_2, where $z_1 = e^{w_1}$ and $z_2 = e^{w_2}$,

$$z_1 \, z_2 = e^{w_1} \, e^{w_2} = e^{w_1 + w_2} \,,$$

which, according to (2.19), implies that

$$\ln z_1 \, z_2 = w_1 + w_2 = \ln z_1 + \ln z_2 \,,$$

so the familiar rule for the logarithm of a product is recovered. The evaluation of the complex logarithm is most naturally carried out in terms of the polar representation $z = r\,e^{i\theta}$:

$$
\begin{aligned}
w = \ln z &= \ln(r\,e^{i\theta}) \\
&= \ln r + \ln e^{i\theta} \\
&= \ln r + i\theta\,.
\end{aligned}
\tag{2.20}
$$

The first term on the right-hand side of this equation is the usual logarithm of the real positive number r. The second term has an inherent ambiguity, which can be seen from the polar representation. Since a rotation of θ by any integer multiple of 2π leaves the polar representation unaltered, i.e.

$$z = r\,e^{i\theta} = r\,e^{i(\theta+2n\pi)}\,,$$

for any integer n, the logarithm is multi-valued, since for each value of n

$$w = \ln r + i(\theta + 2n\pi)\,,$$

corresponds to the same point. To define a *unique* logarithm associated with a complex number, we must restrict the range of θ to an interval of length 2π. There are two such intervals that are commonly used: (i) $0 \le \theta < 2\pi$, and (ii) $-\pi < \theta \le \pi$. We will use (i) here. Thus,

$$\boxed{\ln z = \ln r + i\theta\,, \quad 0 \le \theta < 2\pi\,.} \tag{2.21}$$

EXAMPLE 2.6. Unlike the logarithm for real arguments, the complex logarithm can also be defined for *negative* real numbers. Suppose we have a negative number x, which we represent as $-|x|$. Since the negative real numbers correspond to $\theta = \pi$ in the complex plane,

$$\ln(-|x|) = \ln|x| - i\pi.$$

Consider the logarithm of $z = 1 + i$. In polar form, $z = \sqrt{2}\,e^{\frac{1}{4}i\pi}$, so

$$\ln(1+i) = \ln\sqrt{2} + \tfrac{1}{4}\,i\pi.$$

∎

2.7 Complex Powers

In analogy with real powers, we can define a complex power of a non-zero complex number by utilizing the inverse relationship between the exponential and the logarithm:

$$z^a = e^{\ln z^a} = e^{a \ln z}\,, \tag{2.22}$$

in which we interpret the logarithm as in the preceding section in terms of its principal value.

EXAMPLE 2.7. Consider the complex number $(1+i)^2$. We will evaluate this quantity in two ways: by direct expansion and by applying (2.22). We first calculate

$$(1+i)^2 = (1+i)(1+i) = 1 + 2i - 1 = 2i.$$

Alternatively, with $1 + i = \sqrt{2}\, e^{\frac{1}{4}i\pi}$, we have

$$(1+i)^2 = e^{2\ln(1+i)} = e^{2(\ln\sqrt{2}+\frac{1}{4}i\pi)} = e^{\ln 2}\, e^{\frac{1}{2}i\pi} = 2i,$$

which agrees with the result of the direct expansion.

Consider now the evaluation of i^i. With $i = e^{\frac{1}{2}i\pi}$, we obtain

$$i^i = e^{i \ln i} = e^{i(\frac{1}{2}i\pi)} = e^{-\frac{1}{2}\pi}\,,$$

which is a real number! Finally, we calculate $(1+i)^{1+i}$:

$$\begin{aligned}
(1+i)^{1+i} &= e^{(1+i)\ln(1+i)} \\
&= e^{(1+i)(\ln\sqrt{2}+\frac{1}{4}i\pi)} \\
&= e^{(1+i)\ln\sqrt{2}+\frac{1}{4}(1+i)i\pi} \\
&= e^{\ln\sqrt{2}-\frac{1}{4}\pi}\, e^{i(\ln\sqrt{2}+\frac{1}{4}\pi)} \\
&= \sqrt{2}\, e^{-\frac{1}{4}\pi}\big[\cos\big(\ln\sqrt{2}+\tfrac{1}{4}\pi\big) + \sin\big(\ln\sqrt{2}+\tfrac{1}{4}\pi\big)\big]\,.
\end{aligned}$$

Appendix 1: Hyperbolic Functions*

Trigonometric functions are defined by the diagram shown in the left panel of Fig. 2.4 as the coordinates of a point that lies on the unit circle centered at the origin. The coordinates of this point are given by the cosine and sine functions in terms of an angle θ with respect to the x-axis in the counterclockwise direction: $(\cos\theta, \sin\theta)$. The standard properties of trigonometric functions follow from this construction, e.g. $\cos^2\theta + \sin^2\theta = 1$. This also explains why trigonometric functions are sometimes referred to as 'circular' functions.

Hyperbolic functions are based on the construction in the right panel in Fig. 2.4. The coordinates of any point on the positive branch of the hyperbola $x^2 - y^2 = 1$ (i.e. the branch for which $x > 0$) are defined in terms of the hyperbolic sine and cosine and the hyperbolic angle t as $(\cosh t, \sinh t)$. Explicit formulas for these hyperbolic functions can be obtained by showing that the hyperbolic angle is equal to twice shaded area in Fig. 2.4. To calculate this area, we first note that the equation of the line from the origin to $(\cosh t, \sinh t)$ on the hyperbola is

$$y = \frac{\sinh t}{\cosh t}\, x\,,$$

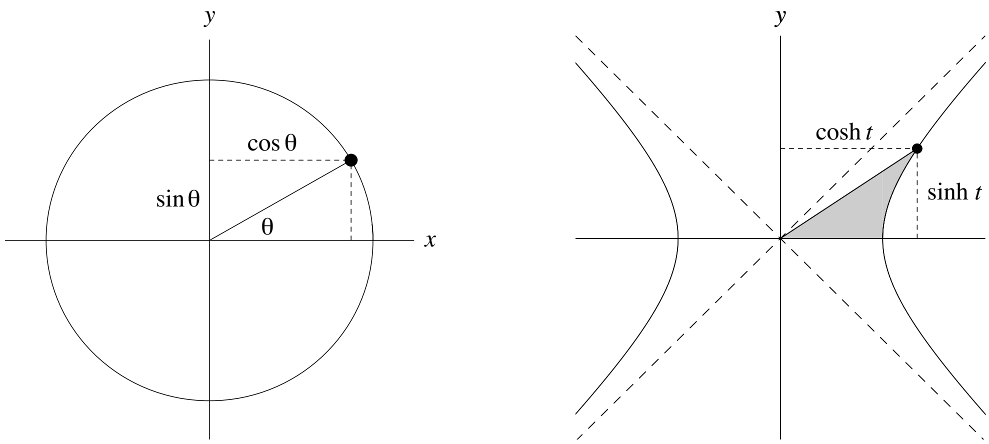


Figure 2.4: The constructions used to define trigonometric and hyperbolic functions. The left panel shows the circle $x^2 + y^2 = 1$ on which the coordinates of any point are given by $(\cos\theta, \sin\theta)$, where θ is the orientation of the point with respect to the x-axis. The right panel shows the analogous construction for hyperbolic functions. Any point on the positive branch of the hyperbola $x^2 - y^2 = 1$ can be represented by a point t, which corresponds to twice the shaded area in the diagram. The coordinates of this point are $(\cosh t, \sinh t)$.

where the range of x is $0 \le x \le \cosh t$. Therefore, the hyperbolic angle t is determined by calculating the area that is shown shaded in Fig. 2.4, which is twice the area associated with this angle. The shaded area in Fig. 2.4 is given by

$$t = 2\left(\int_0^{\cosh t} \frac{\sinh t}{\cosh t}\, x\, dx - \int_1^{\cosh t} \sqrt{x^2-1}\, dx\right). \tag{2.23}$$

The first integral on the right-hand side of this equation is straightforward to evaluate, and we obtain

$$\begin{aligned}\int_0^{\cosh t} \frac{\sinh t}{\cosh t}\, x\, dx &= \frac{\sinh t}{\cosh t}\int_0^{\cosh t} x\, dx \\ &= \frac{\sinh t}{\cosh t}\left(\frac{x^2}{2}\bigg|_0^{\cosh t}\right) = \tfrac{1}{2}\sinh t\cosh t\,. \end{aligned} \tag{2.24}$$

The second integral on the right-hand side of (2.23) is carried out by invoking the indefinite integral,

$$\int \sqrt{x^2-1}\, dx = \tfrac{1}{2}x\sqrt{x^2-1} - \tfrac{1}{2}\ln\big(x + \sqrt{x^2-1}\,\big),$$

which is derived in Appendix 2. We thereby obtain

$$\begin{aligned}\int_1^{\cosh t} \sqrt{x^2-1}\, dx &= \Big[\tfrac{1}{2}x\sqrt{x^2-1} - \tfrac{1}{2}\ln\big(x + \sqrt{x^2-1}\,\big)\Big]\bigg|_1^{\cosh t} \\ &= \tfrac{1}{2}\cosh t\sqrt{\cosh^2 t - 1} - \tfrac{1}{2}\ln\big(\cosh t + \sqrt{\cosh^2 t - 1}\,\big) \\ &= \tfrac{1}{2}\sinh t\cosh t - \tfrac{1}{2}\ln(\cosh t + \sinh t) \end{aligned} \tag{2.25}$$

Substitution of (2.24) and (2.25) into (2.23) yields

$$t = \ln(\cosh t + \sinh t)\,. \tag{2.26}$$

A similar calculation for the hyperbolic angle $-t$, which has the coordinates $(\cosh t, -\sinh t)$, yields

$$-t = \ln(\cosh t - \sinh t)\,. \tag{2.27}$$

Exponentiating (2.26) and (2.27) and solving for $\cosh t$ and $\sinh t$ yields

$$\cosh t = \frac{e^t + e^{-t}}{2}, \qquad \sinh t = \frac{e^t - e^{-t}}{2}. \tag{2.28}$$

Hyperbolic functions have some formal similarities with trigonometric functions, but also appreciable differences. For example, because the hyperbolic angle t lies on the hyperbola $x^2 - y^2 = 1$, we have that

$$\cosh^2 t - \sinh^2 t = 1 .$$

Their derivatives are calculated as

$$\frac{d\cosh t}{dt} = \frac{e^t - e^{-t}}{2} = \sinh t ,$$

$$\frac{d\sinh t}{dt} = \frac{e^t + e^{-t}}{2} = \cosh t .$$

Functions analogous to their trigonometric counterparts can also be defined for hyperbolic functions, e.g.

$$\tanh t = \frac{\sinh t}{\cosh t} .$$

Hyperbolic functions occur in several applications:

1. Advanced treatments of special relativity.
2. Solutions to several fundamental differential equations in mathematical physics.
3. The evaluation of integrals by methods analogous to trigonometric substitution.

Appendix 2: Integral for Hyperbolic Functions*

The calculation of the hyperbolic angle in (2.23) requires the evaluation of the following integral:

$$\int \sqrt{x^2 - 1}\, dx\,. \tag{2.29}$$

We proceed by changing the integration variable from x to s according to

$$x = \tfrac{1}{2}(e^s + e^{-s})\,. \tag{2.30}$$

The integrand becomes

$$\begin{aligned}
\sqrt{x^2 - 1} &= \left\{\left[\tfrac{1}{2}(e^s + e^{-s})\right]^2 - 1\right\}^{1/2} \\
&= \left(\tfrac{1}{4}e^{2s} + \tfrac{1}{2} + \tfrac{1}{4}e^{-2s} - 1\right)^{1/2} \\
&= \left(\tfrac{1}{4}e^{2s} - \tfrac{1}{2} + \tfrac{1}{4}e^{-2s}\right)^{1/2} \\
&= \left\{\left[\tfrac{1}{2}(e^s - e^{-s})\right]^2\right\}^{1/2} \\
&= \tfrac{1}{2}(e^s - e^{-s})\,.
\end{aligned}$$

The integration element is transformed to

$$dx = \tfrac{1}{2}(e^s - e^{-s})\, ds\,.$$

The integral in (2.29) is thereby written in terms of s as

$$\begin{aligned}
\frac{1}{4}\int (e^s - e^{-s})^2\, ds &= \frac{1}{4}\int (e^{2s} - 2 + e^{-2s})\, ds \\
&= \frac{1}{4}\left(\frac{e^{2s}}{2} - 2s - \frac{e^{-2s}}{2}\right) \\
&= \tfrac{1}{8}\, e^{2s} - \tfrac{1}{2}\, s - \tfrac{1}{8}\, e^{-2s}\,,
\end{aligned}$$

where we have omitted the arbitrary additive constant. We must now express this in terms of the original variable. By writing the transformation (2.30) as

$$e^s + e^{-s} = 2x\,,$$

and multiplying by e^s, we obtain a quadratic equation in e^s:

$$e^{2s} - 2x\,e^s + 1 = 0\,.$$

The solution for e^s is given by the quadratic formula:

$$e^s = \tfrac{1}{2}\big(2x \pm \sqrt{4x^2 - 4}\,\big) = x \pm \sqrt{x^2 - 1}\,.$$

Since e^{-s} is a solution of the *same* quadratic equation, and since $e^s > e^{-s}$, we must have that

$$e^{\pm s} = x \pm \sqrt{x^2 - 1}\,.$$

Hence, we find that

$$\begin{aligned}
&\tfrac{1}{8}\,e^{2s} - \tfrac{1}{2}\,s - \tfrac{1}{8}\,e^{-2s} \\
&= \tfrac{1}{8}\big(x + \sqrt{x^2 - 1}\,\big)^2 - \tfrac{1}{2}\ln\big(x + \sqrt{x^2 - 1}\,\big) - \tfrac{1}{8}\big(x - \sqrt{x^2 - 1}\,\big)^2 \\
&= \tfrac{1}{2}x\sqrt{x^2 - 1} - \tfrac{1}{2}\ln\big(x + \sqrt{x^2 - 1}\,\big)\,.
\end{aligned}$$

Thus, the integral in (2.29) is given by

$$\boxed{\int \sqrt{x^2 - 1}\,dx = \tfrac{1}{2}x\sqrt{x^2 - 1} - \tfrac{1}{2}\ln\big(x + \sqrt{x^2 - 1}\,\big)\,.}$$

Chapter 3

First-Order Differential Equations

Many physical phenomena are described in terms of a function whose value at a given point depends on its values at neighboring points. Thus, the equation determining this function contains derivatives of the function, such as a first derivative to indicate the slope or a velocity, a second derivative to indicate the curvature or an acceleration, and so on. Such an equation, which establishes a relation between the function and its derivatives, is called a **differential equation**.

There are two main types of differential equation. A differential equation for a function of a single independent variable contains only *ordinary* derivatives of that function and is called an **ordinary differential equation**. A differential equation for a function of two or more independent variables contains *partial* derivatives of the function and therefore is called a **partial differential equation**. In this course, we will be concerned with ordinary differential equations.

Ordinary differential equations were introduced to describe the motion of *discrete* particles under the action of known applied forces. The groundwork for such applications was provided by Newton's work on mechanics, particularly the second law of motion, and the development of calculus by Newton and Leibniz. Such differential equations are expressed with time as the independent variable and the coordinates of the particles as the dependent variables. The study of phenomena associated with *continuous* media, such as the motion of fluids and the transmission of sound and other disturbances established the need for partial differential equations. In such cases, the independent variables are the position and time coordinates of points within the medium and the dependent variables are the quantities associated with the medium, such as the velocity of a fluid and its density.

The fundamental equations at the heart of almost all areas of science and engineering are expressed as differential equations. Among the best known of these are Newton's second law of motion in mechanics, Maxwell's equations in electromagnetism, Schrödinger's equation and Dirac's equations in quantum mechanics, the Navier–Stokes equation in fluid mechanics and aerodynamics, Einstein's equations in general relativity, the Fokker–Planck equation in nonequilibrium statistical mechanics, the Hodgkin–Huxley equation in cellular biology, and the Black–Scholes equation in quantitative finance. The widespread use of differential equations is evident in many aspects of modern life, including weather prediction, transportation, communication, and macroeconomic forecasting, to name just a few. In all of these cases, the differential equations embody the characteristics of specific natural or social phenomena, often manifesting unexpected complexity, which are most clearly revealed by examining their solutions in particular cases.

3.1 Notation and Nomenclature

An ordinary differential equation for a function y of a single independent variable x is a functional relationship between x, y and the derivatives of y. The **order** of a differential equation is the order of the highest derivative appearing in the differential equation. For example, the most general form of a first-order ordinary differential equation is

$$F(x, y, y') = 0\,,$$

which, for the equations we will study, is written as

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

The general form of an nth-order ordinary differential equation is given by the expression

$$F\Big[x, y, y', \ldots, y^{(n)}\Big] = 0\,. \tag{3.1}$$

If the function F in these equations is a polynomial in the highest-order derivative of y appearing in its argument list, then the **degree** of the differential equation is the power to which this highest derivative is raised, i.e. the degree of that polynomial. An equation is said to be **linear** if F is of first degree in y and in each of the derivatives appearing as arguments of F. Thus, the general form of a linear nth-order ordinary differential equation is

$$a_n\frac{d^n y}{dx^n} + \cdots + a_1\frac{dy}{dx} + a_0 y = f(x)\,, \tag{3.2}$$

where $f(x)$ and the coefficients $a_1, \ldots, a_n$ are known quantities. If $f = 0$ this differential equation is said to be **homogeneous**; otherwise, it is **inhomogeneous**. In the next chapter, we will examine the solutions of second-order equations, both with and without the homogeneous term:

$$a_2 \frac{d^2 y}{dx^2} + a_1 \frac{dy}{dx} + a_0 y = f(x) .$$

3.2 Radioactive Decay

We begin our discussion with first-order equations. Our first example is based on the phenomenon of radioactive decay. We denote by $Q(t)$ the amount of material present at time t. This material decays at a rate r proportional to the amount of material present. The differential equation that describes this process is

$$\frac{dQ}{dt} = -r\,Q \,, \tag{3.3}$$

where the minus sign indicates that the amount of material decreases with time. We will solve this equation by using two standard methods.

3.2.1 Method 1: Trial solution

We attempt to solve this equation with a solution of the form $Q(t) = e^{mt}$, where m is a constant that is to be determined. The method of trial solutions with exponential functions is based on the following property:

$$\frac{d^n e^{mx}}{dx^n} = m^n \, e^{mx} \,,$$

for any positive integer n. Substituting our trial solution into Eq. (3.3), we find

$$\frac{dQ}{dt} = m e^{mt} = -r e^{mt} \,,$$

or,

$$(m + r) e^{mt} = 0 \,.$$

We cannot set $e^{mt} = 0$ because that would yield the trivial solution. However, we can satisfy this equation and obtain a solution if we set $m = -r$ (since the exponential is nonzero for finite x and finite m). The most general solution we can write for Eq. (3.3) is, therefore,

$$Q(t) = A\,e^{-rt}\,,$$

where A is *any* constant. We can determine A by appealing to the physical situation described by our differential equation. If we set $t = 0$, then $Q(0)$ corresponds to the amount of material initially present, which we denote by Q_0. Accordingly, $Q(0) = A = Q_0$. Thus, the solution of Eq. (3.3) for the amount of material at time t is

$$\boxed{Q(t) = Q_0 e^{-rt}\,.} \tag{3.4}$$

This shows that a unique solution is obtained not just by solving the differential equation, but by also imposing **initial conditions** that are appropriate for circumstances of the physical problem at hand. The solution (3.4) is plotted in Fig. 3.1. The characteristic exponential decay is clearly evident. With increasing r, the rate of decay is considerably faster because this factor appears in the argument of an exponential function.

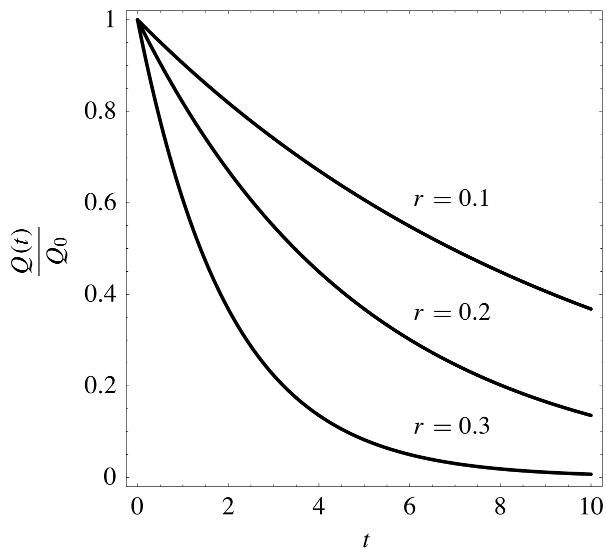


Figure 3.1: The solution in Eq. (3.4) plotted as $Q(t)/Q_0$ against t for three values of the rate constant r. With increasing r, the amount of material at time t decreases substantially.

3.2.2 Method 2: Separation of Variables

The separation of variables is straightforward to set up and carry out, but some aspects of this implementation require a justification. Suppose that we have Q at some time t. We can estimate Q at some late time $t + \Delta t$ by performing a Taylor series to first order in Δt:

$$Q(t + \Delta t) = Q(t) + \frac{dQ}{dt}\Delta t\,.$$

We can now use the equation (3.3) governing radioactive decay to substitute for dQ/dt whereupon, after a simple rearrangement, we obtain

$$\frac{Q(t + \Delta t) - Q(t)}{Q(t)} = -r\,\Delta t\,. \tag{3.5}$$

This equation is valid at any time t and becomes more accurate as Δt becomes smaller. Suppose that we require the solution Q to (3.3) from an initial time 0 to some later time t. We divide this time interval into N equal segments $\Delta t_N = t/N$, so that the nth time increment $t_n = n\Delta t_N$, with $t_0 = 0$ and $t_N = t$. For each n (3.5) can be written as

$$\frac{Q(t_{n+1}) - Q(t_n)}{Q(t_n)} = -r\,\Delta t\,.$$

We now sum this equation over n,

$$\sum_{n=0}^{N-1} \frac{Q(t_{n+1}) - Q(t_n)}{Q(t_n)} = -r \sum_{n=0}^{N-1} \Delta t\,,$$

where we have taken the sum only to $N - 1$ because Q is evaluated at t_{n+1} which, at this upper limit, corresponds to $t_N = t$. Both sides of this equation are discrete approximations to integrals (called Riemann sums). By decreasing Δt toward zero (i.e. making N larger) these sums provide correspondingly better approximations to the integrals and, in the limit that $N \to \infty$, we have

$$\int_{Q(0)}^{Q(t)} \frac{dQ'}{Q'} = -r \int_0^t dt'\,. \tag{3.6}$$

The primes on the integration variables have been introduced to avoid confusion with the limits of integration. Because the dependent variable (Q) appears only on the left-hand side of the equation, and the independent variable (t) appears only on the right-hand side, these variables are said to have been *separated* and the resulting equation can be integrated directly. Thus, with $Q(0) = Q_0$, we can integrate (3.6) to obtain

$$\ln Q' \bigg|_{Q_0}^{Q(\tau)} = \ln\left[\frac{Q(t)}{Q_0}\right] = -rt\,,$$

or, after exponentiating and solving for $Q(t)$,

$$Q(t) = Q_0 e^{-rt}\,,$$

which is the same as the solution in (3.4).

Although our development of the separation of variables method looks somewhat cumbersome, there is a short-cut that considerably simplifies the procedure. We begin by rearranging the equation (3.3) for radioactive decay as

$$\frac{dQ}{Q} = -r\,dt\,. \tag{3.7}$$

Equation (3.5) is the discrete analogue of this equation, which has been obtained by interpreting the derivative dQ/dt as a fraction, rather than as an operation on a function. This can be justified only as an intermediate step toward integrating this equation from $t = 0$, where $Q = Q(0) = Q_0$, to some later time t, where $Q = Q(t)$:

$$\int_{Q(0)}^{Q(t)} \frac{dQ'}{Q'} = -r\int_0^t dt'\,.$$

which is (3.6). The virtue of this somewhat loose interpretation of the mathematical formulation of the separation of variables method is that it is easy to apply and one can usually identify differential equations that are separable directly by inspection.

In summary, the advantage of the trial solution method is that it can be applied to higher-order equations, as we will show in the next chapter for second-order equations, but only to *linear* equations. The separation of variables method can be applied to certain types of nonlinear equations, as we will shown in the next section, but only to *first-order* equations.

3.3 Spread of Epidemics

A timely example of the use of first-order differential equations is to the spread of epidemics, first used by Daniel Bernoulli in 1760 to model the spread of smallpox.

We will construct a simple model of an epidemic and then solve the resulting differential equation. This is essentially the way that models in epidemiology as constructed: assumptions and known characteristics are used to build a model, which is then solved and various scenarios are tested (e.g. inoculation or isolation) to develop strategies on how to respond to an epidemic.

Consider model for the spread of a disease in which a population that is divided into two groups: a fraction x that has no disease, but is susceptible to the disease, and a fraction y that has the disease and can infect others. We suppose that everyone belongs to only one of these groups, so $x + y = 1$. We now make three assumptions about how the disease is spread:

1. The disease spreads only by direct contact between infected and uninfected individuals. Direct can be taken to mean 'close proximity,' as in crowds, where colds and influenza viruses can easily spread.

2. The fraction of infected individuals increases at a rate α proportional to such contacts.

3. Both groups move freely among one another, so the number of direct contacts is xy. This is another way of saying that the x and y populations are uncorrelated.

The differential equation that embodies these assumptions is

$$\frac{dy}{dt} = \alpha xy = \alpha(1 - y)y\,, \tag{3.8}$$

where α is a constant that specifies the 'efficiency' of the spreading at the point of contact, i.e. the likelihood of disease transmission once direct contact has occurred, and we have used the fact that $x + y = 1$ to eliminate x in favor of y. In accordance with our experience in finding the solution for radioactive decay, we must supplement this equation by specifying the fraction of infected individuals initially: $y(0) = y_0$.

Equation (3.8) is a first-order *nonlinear* differential equation. Thus, we cannot use the trial solution method as formulated above. However, this equation can be arranged as

$$\frac{dy}{y(1 - y)} = \alpha\, dt\,,$$

so we can use the separation of variables method. Integrating both sides of the equation with respect to the indicated variables from $t = 0$, where $y = y(0) = y_0$

to a later time t, where $y = y(t)$, we obtain

$$\int_{y_0}^{y(t)} \frac{dy'}{y'(1-y')} = \alpha \int_0^t dt' = \alpha\, t\,.$$

The left-hand side of this equation can be integrated by the method of partial fractions. We first write

$$\frac{1}{y(1-y)} = \frac{A}{y} + \frac{B}{1-y}\,,$$

which implies that

$$A(1-y) + By = 1\,.$$

Choosing $y = 0$ yields $A = 1$, and choosing $y = 1$ yields $B = 1$, so

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

with which we obtain

$$\begin{aligned}
\alpha\, t &= \int_{y_0}^{y(t)} \frac{dy'}{y'} + \int_{y_0}^{y(t)} \frac{dy'}{1-y'} \\
&= \ln y' \Big|_{y_0}^{y(t)} - \ln(1-y')\Big|_{y_0}^{y(t)} \\
&= \ln\left[\frac{y(t)}{1-y(t)}\frac{1-y_0}{y_0}\right].
\end{aligned}$$

Solving for $y(t)$ yields,

$$\boxed{y(t) = \frac{y_0\, e^{\alpha t}}{1 - y_0(1-e^{\alpha t})}\,.} \tag{3.9}$$

As $t \to \infty$, $y(t) \to 1$, provided that $y_0 \neq 0$ (Fig. 3.2). In other words, all of the population eventually becomes infected unless there is no infection initially. As long as $y_0 \neq 0$, no matter how small, the entire population becomes infected. Accordingly, the point $y = 1$ is said to be **stable** and the point $y = 0$ is said to be **unstable**.

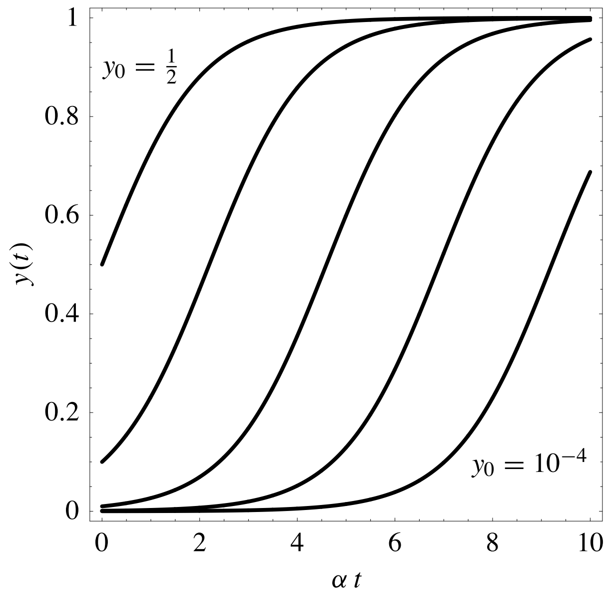


Figure 3.2: The solution in Eq. (3.9) shown as a function of αt for values of y_0 in the range $10^{-4} \leq y_0 \leq \frac{1}{2}$. As y_0 decreases toward zero, the solution remains near $y = 0$ for longer times, while as y_0 increases toward unity, the solution approaches $y = 1$ for shorter times.

3.4 Nonlinearity and Chaos*

Most of the methods for solving ordinary and partial differential equations have been developed for *linear* equations. This is because of the superposition principle, which mandates that, given any two solutions y_1 and y_2 of a linear differential equation, any linear combination $a\,y_1 + b\,y_2$, in which a and b are constants, is also a solution of the same equation. For nonlinear equations, however, linear superposition cannot be applied to generate new solutions, so general approaches to finding solutions are far less abundant than for linear equations. A change of variables can sometimes be found that transforms a nonlinear equation into a linear equation, or some other *ad hoc* technique may yield a solution for a particular equation, but finding the solutions of most nonlinear equations generally requires new techniques or a resort to numerical integration. Indeed, developing numerical methods for finding solutions to nonlinear equations is an active research area.

Linear differential equations usually only provide approximations to physical situations. We saw in Sec. 3.3 how nonlinear equations arise quite naturally even in the description of a physical situation. In that case, we were still able to obtain an analytic solution because the nonlinear equation was separable. Any attempt to bypass the nonlinearity of this equation would lose the essential character of the solutions, i.e. epidemiology is an inherently an nonlinear phenomenon.

While nonlinear equations are generally more difficult to solve than linear equations, their solutions often yield new phenomena that have no linear analogue. One striking example of this is based on the following set of three coupled first-order nonlinear ordinary differential equations:

$$\frac{dx}{dt} = \sigma(y - x)\,, \qquad \frac{dy}{dt} = rx - y - xz\,, \qquad \frac{dz}{dt} = xy - bz\,, \tag{3.10}$$

in which σ, r, and b are constants. These are called the **Lorenz equations**, after Edward Lorenz, a meteorologist who was working on the problem of weather prediction. His equations are a much reduced form of the equations used in fluid systems, but he had at his disposal (1963) only a primitive (by today's standards) computer. Lorenz discovered several interesting and unexpected properties of the solutions to his equations.

Figure 3.3 shows solution for x, y, and z for the Lorenz equations with $\sigma = 3$, $r = 26.5$, and $b = 1$. Most apparent from this diagram is that the solutions show quite irregular behavior, with occasional periods of regularity followed by sudden jumps to another type of regular behavior, but with no apparent pattern. Because of this behavior, and the extreme sensitivity of the solutions to the initial conditions, Lorenz had discovered the phenomenon of chaos. The trajectories of the solutions in Fig. 3.3, which is plotted in the x-y plane in Fig. 3.4, are even more revealing. The double spiral is typical of the solutions of the Lorenz equations. At the time of Lorenz's work, there were only two kinds of order previously known: a steady state, in which the variables do not change, and periodic behavior, in which the system goes into a loop, repeating itself indefinitely. Lorenz's equations were definitely ordered in that they always followed a spiral. They never settled down to a single point, but since they never repeated the same thing, they could not be called periodic either. This is now known as the **Lorenz attractor**.

An interesting footnote to this discovery is that Lorenz published a paper describing his findings.[1] He included the unpredictability of the weather, and discussed the types of equations that caused this of behavior. But he published his

[1] E. N. Lorenz, 'Deterministic nonperiodic flow,' *J. Atmos. Sci.* **20**, 130–141 (1963).

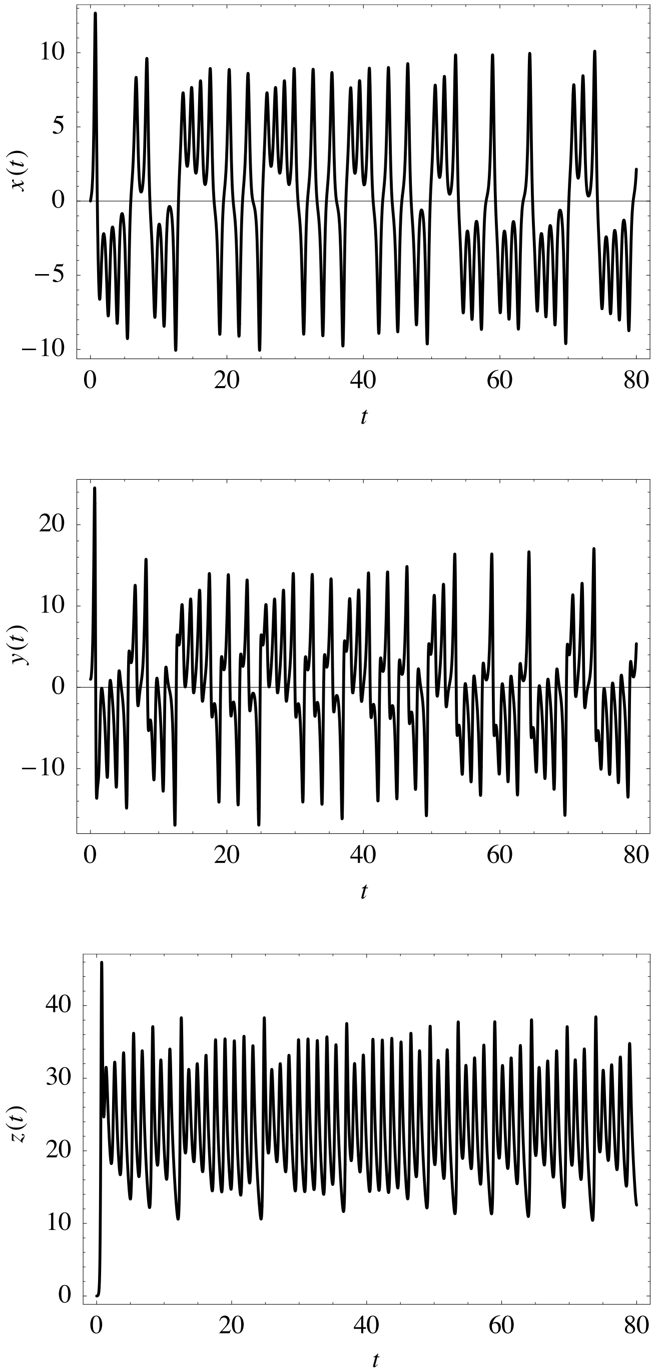


Figure 3.3: The solutions x, y, and z for the Lorenz equations (3.10) with $\sigma = 3$, $r = 26.5$, and $b = 1$.

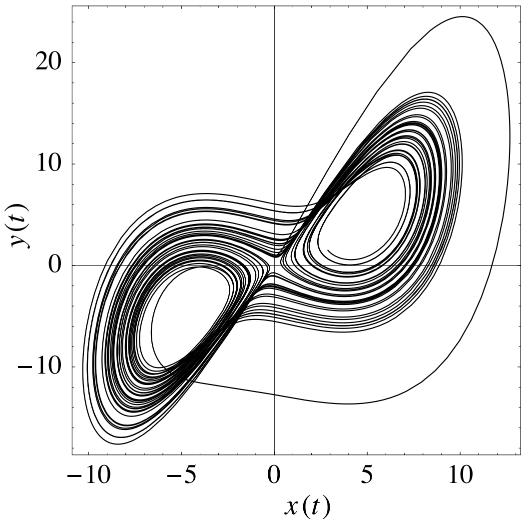


Figure 3.4: The solutions of the Lorenz equations in Fig. 3.3 plotted in the x-y plane/

work in a meteorological journal and, as a result, Lorenz's revolutionary discoveries were not acknowledged until years later, by which time they were rediscovered by others.

Chapter 4

Second-Order Ordinary Differential Equations

Among the simplest higher-order ordinary differential equations are linear homogeneous equations with constant coefficients. The second-order versions of these equations occur in many applications in science and engineering. Among the most prevalent of these involve Newton's second law of motion in mechanics and the flow of charge in electrical circuits. Such equations also occur as special cases of certain partial differential equations, for example, in quantum mechanics and problems of heat conduction. The solutions of second-order differential equations with constant coefficients will be shown to have the important feature of being expressible in terms of exponential functions, and their method of solution has evident extensions to higher-order equations with constant coefficients.

In this chapter, we illustrate the solution of equations with constant coefficients by focussing on second-order equations. The general form of a second-order linear homogeneous ordinary differential equation with constant coefficients is

$$a\,\frac{d^2y}{dx^2}+b\,\frac{dy}{dx}+c\,y=0\,, \tag{4.1}$$

where a, b and c are known real constants. The most important property of this equation is *linearity*. This means that if we have two solutions y_1 and y_2, then any linear combination of y_1 and y_2 is also a solution of this equation, i.e.

$$y(x)=Ay_1(x)+By_2(x)$$

is a solution for *any* choice of constants A and B. This is a direct consequence of the linearity of derivatives, which allows us to write

$$\begin{aligned}
&a\frac{d^2y}{dx^2}+b\frac{dy}{dx}+cy\\
&=a\frac{d^2}{dx^2}(Ay_1+By_2)+b\frac{d}{dx}(Ay_1+By_2)+c(Ay_1+By_2)\\
&=A\left(a\frac{d^2y_1}{dx^2}+b\frac{dy_1}{dx}+cy_1\right)+B\left(a\frac{d^2y_2}{dx^2}+b\frac{dy_2}{dx}+cy_2\right)\\
&=0\,. \qquad (4.2)
\end{aligned}$$

The last equality follows from the fact that y_1 and y_2 are solutions of (4.1), so the coefficients of A and B are equal to zero. The same steps can be used to verify this statement for the more general case of a linear differential equation with variable coefficients. This is the **superposition principle** for linear differential equations. It lies at the heart of both the theory of these equations and the methodologies that have been developed for solving them.

4.1 The Characteristic Equation

The solution of (4.1) will be obtained by the method of trial solutions. The recursive property of derivatives of the exponential function,

$$\frac{d}{dx}(e^{mx})=m\,e^{mx}\,,\qquad \frac{d^2}{dx^2}(e^{mx})=m^2e^{mx}\,, \qquad (4.3)$$

suggests that the trial solution method used for solving first-order equations in Sec. 3.2.1 can be applied to higher-order equations with constant coefficients. Suppose we try this for the differential equation (4.1). We substitute our trial solution e^{mx} into this equation and choose m by requiring the resulting expression to equal zero, i.e. that this function solves the equation. Substituting the derivatives in (4.3) into (4.1) yields

$$a\,\frac{d^2}{dx^2}(e^{mx})+b\,\frac{d}{dx}(e^{mx})+c\,(e^{mx})=(am^2+bm+c)\,e^{mx}\,. \qquad (4.4)$$

For the function e^{mx} to be a solution of (4.1), the coefficient of e^{mx} on the right-hand side of this equation must vanish (since the exponential is nonzero for finite

x). Thus, m must be chosen to be a root of the quadratic equation

$$am^2 + bm + c = 0\,. \tag{4.5}$$

This is the **characteristic equation** of the differential equation (4.1) and the left-hand side of this equation is called the **characteristic polynomial**. The roots of the characteristic equation, which are given by the quadratic formula,

$$m = -\frac{b}{2a} \pm \frac{1}{2a}\sqrt{b^2 - 4ac}\,,$$

yield solutions of (4.1). This is the power of this trial solution method – the solution of a differential equation has been reduced to finding the roots of quadratic equation.

By their appearance in the discriminant in this equation, the coefficients a, b and c are seen to be the central quantities for determining the number and type of roots of the characteristic polynomial and, through these roots, the behavior of the exponential solutions. in direct analogy to the discussion in Sec. 1.1, there are three cases to consider.

4.1.1 Case I: Real Distinct Roots

If $b^2 - 4ac > 0$, there are two distinct real roots of the characteristic equation, which we denote by m_1 and m_2. There result two distinct solutions of (4.1):

$$y_1(x) = e^{m_1 x}\,, \qquad y_2(x) = e^{m_2 x}\,. \tag{4.6}$$

According to the procedure in (4.2), we can use these solutions to form a more general solution of (4.1) by forming the linear combination

$$y(x) = Ae^{m_1 x} + Be^{m_2 x}\,,$$

where A and B are arbitrary constants. This is called the **general solution** of the differential equation. The constants A and B are determined by specifying initial conditions. Because there are two arbitrary constants in the solution, two initial conditions are needed to obtain a unique solution. These are taken as the function y and its derivative at the evaluated at the origin,

$$y(0) = y_0\,, \qquad \left.\frac{dy}{dx}\right|_{x=0} \equiv y'(0) = y_0'\,, \tag{4.7}$$

although they can be specified at any point. Depending on the signs of m_1 and m_2, the solutions exhibit either exponential growth or exponential decay as functions of x.

4.1.2 Case II: Degenerate Roots

If $b^2 - 4ac = 0$, there is only a single real root, $m_1 = -b/(2a)$, of the characteristic equation. Thus, this method produces only one solution of (4.1):

$$y_1(x) = e^{m_1 x}\,. \tag{4.8}$$

We seem to have arrived at an impasse. The case of two real roots of the characteristic equation provided two distinct solutions with which we can obtain a unique solution from the general solution for a particular set of two initial conditions. A similar situation will arise in the case where the characteristic equation yields complex conjugate roots. However, if the discriminant vanishes, we appear to have only a single solution, which cannot be reconciled with two initial conditions. Thus, the method of trial solutions has failed to provide two solutions. But we can extend this method by making a few simple observations that will yield a second solution in a form that enables us to deal with the case of a vanishing discriminant in an analogous manner to the other two cases.

We begin by returning to (4.4), the left-hand side of which is

$$a\frac{d^2(e^{mx})}{dx^2} + b\frac{d(e^{mx})}{dx} + c(e^{mx})\,. \tag{4.9}$$

This equation equals zero only if we set $m = m_1$, which shows that $e^{m_1 x}$ is a solution:

$$\left.\left[a\frac{d^2(e^{mx})}{dx^2} + b\frac{d(e^{mx})}{dx} + c(e^{mx})\right]\right|_{m=m_1}$$
$$= a\frac{d^2(e^{m_1 x})}{dx^2} + b\frac{d(e^{m_1 x})}{dx} + c(e^{m_1 x}) = 0$$

Since m is a continuous variable, we can differentiate (4.12) with respect to m *before* setting m equal to m_1.

$$\left\{\frac{d}{dm}\left[a\frac{d^2(e^{mx})}{dx^2}+b\frac{d(e^{mx})}{dx}+c(e^{mx})\right]\right\}\bigg|_{m=m_1}$$

The order of the derivatives with respect to x and to m is immaterial, so we can take the m derivatives before the x derivatives, in which case we obtain,

$$a\,\frac{d^2}{dx^2}\left[\frac{d(e^{mx})}{dm}\right]+b\,\frac{d}{dx}\left[\frac{d(e^{mx})}{dm}\right]+c\left[\frac{d(e^{mx})}{dm}\right]$$

$$=a\,\frac{d^2}{dx^2}(x\,e^{mx})+b\,\frac{d}{dx}(x\,e^{mx})+c\,(x\,e^{mx})\,.$$

The remaining derivatives are straightforward to calculate:

$$\frac{d}{dx}(x\,e^{mx}) = e^{mx}+mx\,e^{mx}\,,$$

$$\frac{d^2}{dx^2}(x\,e^{mx}) = 2m\,e^{mx}+m^2x\,e^{mx}\,,$$

whereupon we obtain

$$a\,\frac{d^2}{dx^2}\left[\frac{d(e^{mx})}{dm}\right]+b\,\frac{d}{dx}\left[\frac{d(e^{mx})}{dm}\right]+c\left[\frac{d(e^{mx})}{dm}\right]$$

$$=(2am+b)\,e^{mx}+(am^2+bm+c)x\,e^{mx}\,. \qquad (4.10)$$

By setting $m=m_1$ and using the fact that $m_1=-b/(2a)$, we find

$$2am_1+b=-b+b=0\,,$$

$$am_1^2+bm_1+c=\frac{b^2}{4a}-\frac{b^2}{2a}+c=-\frac{b^2-4ac}{2}=0\,.$$

so the coefficient of each term on the right-hand side of (4.10) vanishes, leaving

$$a\,\frac{d^2}{dx^2}(x\,e^{m_1x})+b\,\frac{d}{dx}(x\,e^{m_1x})+c\,(x\,e^{m_1x})=0\,,$$

which shows that our second solution is, in this case,

$$y_2(x)=x\,e^{m_1x}\,.$$

The general solution is

$$\boxed{y(x) = (A + Bx)e^{m_1 x} \, .}$$

Similar to those in (4.6), the solutions y_1 and y_2 exhibit either exponential growth or decay, depending on the sign of m_1.

4.1.3 Case III: Complex Conjugate Roots

If $b^2 - 4ac < 0$, there are two complex roots, m_1 and m_2, which are complex conjugates: $m_2 = m_1^*$ (Problem 8, Classwork 1). The two solutions of (4.1) are thus given by

$$y_1(x) = e^{m_1 x}, \qquad y_2(x) = e^{m_1^* x} \, , \tag{4.11}$$

so the general solution is

$$\boxed{y(x) = A\, e^{m_1 x} + B\, e^{m_2 x} \, .}$$

Since m_1 and m_2 are complex numbers, y_1 and y_2 are complex-valued functions. However, we can express the solutions to (4.1) solely in terms of real functions by utilizing Euler's formula (1.15). With m_1 and m_2 expressed in terms of their real and imaginary parts as

$$m_1 = \alpha + i\beta, \qquad m_2 = \alpha - i\beta \, ,$$

where α and β are real, we first write the solutions in (4.11) as

$$\begin{aligned} y_1(x) &= e^{(\alpha + i\beta)x} = e^{\alpha x}(\cos\beta x + i \sin\beta x) \, , \\ y_2(x) &= e^{(\alpha - i\beta)x} = e^{\alpha x}(\cos\beta x - i \sin\beta x) \, . \end{aligned}$$

Thus, linear combinations of y_1 and y_2 can be written as

$$\begin{aligned} y(x) &= Ay_1(x) + By_2(x) \\ &= A\Big[e^{\alpha x}(\cos\beta x + i \sin\beta)\Big] + B\Big[e^{\alpha x}(\cos\beta x - i \sin\beta)\Big] \\ &= (A + B)\, e^{\alpha x} \cos\beta x + i(A - B)\, e^{\alpha x} \sin\beta x \, . \end{aligned}$$

Since A and B are arbitrary quantities, then $A + B$ and $i(A - B)$ are as well, so we can write the general solution in an alternative form based on the real solutions

$$\tilde{y}_1(x) = e^{\alpha x}\cos\beta x \qquad \tilde{y}_2(x) = e^{\alpha x}\sin\beta x\,, \tag{4.12}$$

as

$$\boxed{y(x) = C\,e^{\alpha x}\cos\beta x + D\,e^{\alpha x}\sin\beta x\,,}$$

in which C and D are arbitrary constants whose values are obtained from the initial conditions (4.7). These solutions show that the imaginary parts of m_1 and m_2 produce oscillatory behavior and their real parts, if nonzero, modulate this with either exponential growth or decay. The choice of whether to use the real solutions in (4.11) or their complex counterparts in (4.12) is largely a matter of taste and convenience. In the next section we will show how of the three types of solution to the characteristic equation arise in a physical setting.

4.2 The Harmonic Oscillator

Consider the harmonic oscillator in Fig. 4.1, which consists of a mass m_0 attached to a spring with stiffness k and damping γ. Once displaced from equilibrium, there are two forces acting on the mass: the gravitational force $m_0 g$ acting downward, and the forces $-kx$ and $-r\dot{x}$ from the spring, which always act in *opposition* to the motion (which is the reason for the minus signs). Newton's second law of motion for the position x of the oscillator is thus given by

$$m_0\frac{d^2x}{dt^2} = -kx - r\frac{dx}{dt} - m_0 g\,,$$

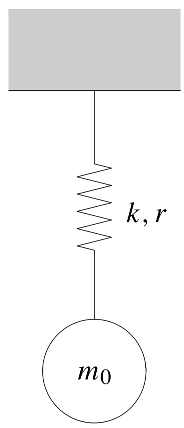


Figure 4.1: A mass m_0 attached to a harmonic spring with stiffness k and damping r.

which we rearrange as

$$\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x + g = 0\,,$$

in which

$$\gamma = \frac{r}{m_0}, \qquad \omega_0^2 = \frac{k}{m_0},$$

and ω_0 is the natural frequency of the oscillator. The constant factor g in this equation, which originates from the force $m_0 g$ due to gravity in Newton's second law, can be eliminated by shifting the position of the oscillator by $-m_0 g/k$, which is the equilibrium position of the oscillator. In other words, this is the solution obtained with $\dot{x} = 0$ and $\ddot{x} = 0$, where the overdots indicate derivatives with respect to time. We will not consider the constant term further, so the solution we will obtain will represent the *displacement* from this position. The equation to be solved is

$$\boxed{\frac{dx^2}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x = 0\,.} \tag{4.13}$$

To obtain a specific solution for the position of the oscillator, we must supplement this equation with two initial conditions. We take

$$x(0) = x_0\,, \qquad \left.\frac{dx}{dt}\right|_{t=0} = 0\,, \tag{4.14}$$

which correspond to an initial position of x_0, but with no initial velocity. Equation (4.13) has the form of Eq. (4.1), with

$$a = 1\,, \qquad b = \gamma\,, \qquad c = \omega_0^2\,,$$

so the solutions are determined by the solving the characteristic equation (4.5), to obtain the roots

$$m = \frac{1}{2}\Big(-\gamma \pm \sqrt{\gamma^2 - 4\omega_0^2}\Big)\,.$$

The three cases discussed above lead to the types of solution discussed in the next three sections.

4.2.1 Case I: $\gamma^2 - 4\omega_0^2 > 0$

In this case, we have $\gamma > 2\omega_0$, so the damping dominates the oscillations of the mass. Hence, this is referred to as the **overdamped** case. We obtain two real roots

m_1 and m_2 given by

$$m_1 = \frac{1}{2}\left(-\gamma - \sqrt{\gamma^2 - 4\omega_0^2}\right),$$
$$m_2 = \frac{1}{2}\left(-\gamma + \sqrt{\gamma^2 - 4\omega_0^2}\right),$$

so $m_1 < 0$ and $m_2 < 0$, and the general solution is

$$x(t) = Ae^{m_1 t} + Be^{m_2 t}.$$

The initial conditions in Eq. (4.14),

$$x(0) = A + B = x_0, \qquad x'(0) = m_1 A + m_2 B = 0,$$

yield

$$A = \frac{-m_2 x_0}{m_1 - m_2}, \qquad B = \frac{m_1 x_0}{m_1 - m_2}.$$

The solution for the position of the oscillator is therefore obtained as

$$\boxed{x(t) = \frac{x_0}{m_1 - m_2}\left(m_1 e^{m_2 t} - m_2 e^{m_1 t}\right).}$$

4.2.2 Case II: $\gamma^2 - 4\omega_0^2 = 0$

In this case, we have that $\gamma = 2\omega_0$, so the damping and oscillations are balanced. This is called the **critically damped** case. We obtain a single real root m_1,

$$m_1 = -\tfrac{1}{2}\gamma,$$

and the general solution is

$$x(t) = (A + Bt)e^{-\frac{1}{2}\gamma t}.$$

The initial conditions in Eq. (4.14),

$$x(0) = A = x_0, \qquad x'(0) = B - \tfrac{1}{2}\gamma = 0,$$

yield

$$A = x_0 \,, \qquad B = \tfrac{1}{2} x_0 \gamma \,.$$

The solution for the position of the oscillator is therefore obtained as

$$\boxed{x(t) = x_0 \left(1 + \tfrac{1}{2}\gamma t\right) e^{-\frac{1}{2}\gamma t} \,.}$$

4.2.3 Case III: $\gamma^2 - 4\omega_0^2 < 0$

Here, we have that $\gamma < 2\omega_0$. The oscillations dominate the damping, so this is called the **underdamped** case. We obtain two roots m_1 and m_2 that are complex conjugates, given by

$$m_1 = \frac{1}{2}\left(-\gamma - i\sqrt{4\omega_0^2 - \gamma^2}\right) ,$$

$$m_2 = \frac{1}{2}\left(-\gamma + i\sqrt{4\omega_0^2 - \gamma^2}\right) ,$$

and the general solution is

$$x(t) = Ae^{m_1 t} + Be^{m_2 t} \,.$$

The initial conditions in Eq. (4.14),

$$x(0) = A + B = x_0 \,, \qquad x'(0) = m_1 A + m_2 B = 0 \,,$$

yield

$$A = \frac{-m_2 x_0}{m_1 - m_2} \,, \qquad B = \frac{m_1 x_0}{m_1 - m_2} \,.$$

The solution for the position of the oscillator is therefore obtained as

$$\boxed{x(t) = \frac{x_0}{m_1 - m_2}\left(m_1 e^{m_2 t} - m_2 e^{m_1 t}\right) .} \qquad (4.15)$$

We can write this solution in a more physically transparent form by using the polar form of m_1 and m_2. Using the notation

$$m_1 = -\tfrac{1}{2}\gamma - i\tfrac{1}{2}\Gamma\,, \qquad m_2 = -\tfrac{1}{2}\gamma + i\tfrac{1}{2}\Gamma\,,$$

the magnitude of both roots is calculated as

$$(\tfrac{1}{4}\gamma^2 + \tfrac{1}{4}\Gamma^2)^{1/2} = (\omega_0^2)^{1/2} = \omega_0\,.$$

Thus,

$$m_1 = \omega_0\, e^{i\phi}\,, \qquad m_2 = \omega_0\, e^{-i\phi}\,,$$

where

$$\cos\phi = -\frac{\gamma}{2\omega_0}\,, \qquad \sin\phi = -\frac{\Gamma}{2\omega_0}\,.$$

The solution in (4.15) can now be written as

$$\begin{aligned} x(t) &= -\frac{x_0}{2i\Gamma}\big[2\omega_0\, e^{i\phi}\, e^{-\frac{1}{2}(\gamma - i\Gamma)t} - 2\omega_0\, e^{-i\phi}\, e^{-\frac{1}{2}(\gamma + i\Gamma)t}\big] \\ &= -\frac{x_0\omega_0}{i\Gamma}\, e^{-\frac{1}{2}\gamma t}\big[e^{i(\phi + \frac{1}{2}\Gamma t)} - e^{-i(\phi + \frac{1}{2}\Gamma t)}\big] \\ &= -\frac{2x_0\omega_0}{\Gamma}\, e^{-\frac{1}{2}\gamma t}\sin\big(\phi + \tfrac{1}{2}\Gamma t\big)\,. \end{aligned}$$

This demonstrates that the solution obtained is, indeed, real, and that the quantity ϕ enters as a phase angle into the argument of the periodic part of the solution.

The three types of solutions are shown in Fig. 4.2 for the displacement $x(t)$ with the initial conditions in (4.14) with $x(0) = 1$ and $x'(0) = 0$. All three solutions therefore begin at $x = 1$ with zero initial slope. The overdamped case decays to zero monotonically, as does the critically damped solution, although the decay of the latter solution is seen to be much faster. The underdamped solution shows several periods of oscillation that have a decaying envelope. For all three cases, the equilibrium position is reached as $t \to \infty$ because of the presence of the damping.

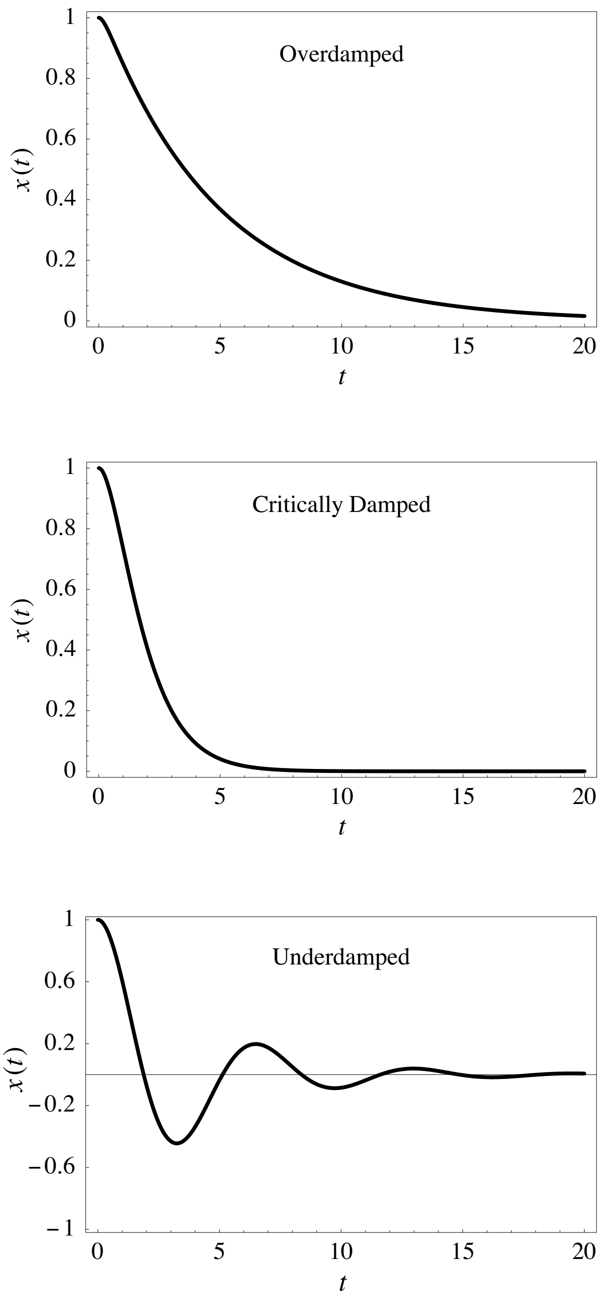


Figure 4.2: The three types of solution for a damped harmonic oscillator, showing he displacement $x(t)$ for the overdamped, critically damped, and underdamped cases.

4.3 Inhomogeneous Equations

4.3.1 Method of Solution

One of the most striking manifestations of driven systems is the phenomenon of resonance. This motivates the discussion of equations of the form

$$a\,\frac{d^2y}{dx^2}+b\,\frac{dy}{dx}+c\,y=f(x)\,,$$

which are called **inhomogeneous** because the function $f(t)$ on the right-hand side of this equation is specified independently of the solution. Such equations are solved by first supposing that there are two independent solutions $y^{(1)}(x)$ and $y^{(2)}(x)$ of this equation:

$$a\,\frac{d^2y^{(1)}}{dx^2}+b\,\frac{dy^{(1)}}{dx}+c\,y^{(1)}=f(x)\,, \tag{4.16}$$

$$a\,\frac{d^2y^{(2)}}{dx^2}+b\,\frac{dy^{(2)}}{dx}+c\,y^{(2)}=f(x)\,. \tag{4.17}$$

If we subtract one equation from the other, say Eq. (4.16) from (4.17), we obtain

$$a\frac{d^2[y^{(2)}-y^{(1)}]}{dt^2}+b\frac{d[y^{(2)}-y^{(1)}]}{dt}+[y^{(2)}-y^{(1)}]=0\,,$$

i.e. the difference $y^{(2)}-y^{(1)}$ is a solution of the *homogeneous* equation! If we denote the general solution of the homogeneous equation by $Ay_1(x)+By_2(x)$, we conclude from the foregoing that

$$y^{(2)}(x)=Ay_1(x)+By_2(x)+y^{(1)}(x)\,.$$

This suggests the following method of solution. Find a solution $y_p(x)$ of the inhomogeneous equation, called a **particular solution**, by any means. The general solution $y(x)$ of the inhomogeneous equation is then given by

$$\boxed{y(x)=Ay_1(x)+By_2(x)+y_p(x)\,,}$$

in which y_1 and y_2 are solutions of the corresponding homogeneous equation.

4.3.2 Resonance in a Driven Harmonic Oscillator

To illustrate the solution of inhomogeneous equations, we consider an undamped harmonic oscillator driven by an external sinusoidal force:

$$m_0\frac{d^2x}{dt^2}+kx=F_0\cos\omega t\,,$$

where m_0 is the mass, k is the spring constant, x is the position of the mass, t is the time, F_0 is the amplitude of the driving force with frequency ω. Upon dividing through by m_0, we can write this equation as

$$\frac{d^2x}{dt^2}+\omega_0^2x=\frac{F_0}{m_0}\cos(\omega t)\,, \tag{4.18}$$

where $\omega_0=(k/m_0)^{1/2}$ is the natural frequency of the oscillator. From the discussion in the preceding equation, we know that the solution of the corresponding *homogeneous* equation (i.e. the equation obtained by setting $F_0=0$), is

$$x(t)=A\cos\omega_0t+B\sin\omega_0t\,,$$

where A and B are arbitrary constants obtained by specifying two initial conditions (the initial position and velocity of the mass). The most general solution of the inhomogeneous equation is the sum of the general solution of the homogeneous and a particular solution $x_p(t)$ of the inhomogeneous equation:

$$x(t)=A\cos\omega_0t+B\sin\omega_0t+x_p(t)\,.$$

To determine $x_p(t)$ for this equation, we attempt a solution of the form

$$x_p(t)=C\cos\omega t\,.$$

The required derivatives are

$$\frac{dx_p}{dt}=-C\omega\sin\omega t,\qquad \frac{d^2x_p}{dt^2}=-C\omega^2\cos\omega t\,.$$

Substitution of these expressions into Eq. (4.18),

$$-C\omega^2\cos\omega t+C\omega_0^2\cos\omega t=\frac{F_0}{m_0}\cos\omega t\,,$$

cancelling the common factor of $\cos(\omega t)$, and solving for C, yields

$$x_p(t)=\frac{F_0}{m_0(\omega_0^2-\omega^2)}\cos\omega t\,.$$

Note that, as $\omega \to \omega_0$, the solution becomes unbounded. This is called **resonance**. In the presence of damping, the solutions remain finite, but still become large when the resonance condition is fulfilled. The damping of oscillations close to resonance is an important engineering problem, as evidenced by the famous collapse of the Tacoma Narrows Bridge and, more recently, by the re-design of the Millennium Bridge to incorporate damping.

The general solution to the inhomogeneous equation (4.18) is therefore given by

$$\boxed{x(t) = A\cos\omega_0 t + B\sin\omega_0 t + \frac{F_0}{m_0(\omega_0^2 - \omega^2)}\cos\omega t\,.} \tag{4.19}$$

To solve the initial-value problem, we consider the initial condition corresponding to the mass being initially at rest:

$$x(0) = 0, \qquad \left.\frac{dx}{dt}\right|_{t=0} = 0\,.$$

Substituting these conditions into the general solution yields

$$x(0) = A + \frac{F_0}{m_0(\omega_0^2 - \omega^2)} = 0\,,$$

so,

$$A = -\frac{F_0}{m_0(\omega_0^2 - \omega^2)}\,,$$

and

$$\left.\frac{dx}{dt}\right|_{t=0} = \omega_0 B = 0\,,$$

which yields $B = 0$. Thus, the solution to the initial-value problem is

$$x(t) = \frac{F_0}{m_0(\omega_0^2 - \omega^2)}\left(\cos\omega t - \cos\omega_0 t\right)\,.$$

This expression can be written in a physically more transparent form by using the trigonometric identity,

$$\cos(A - B) - \cos(A + B) = 2\sin A\sin B\,.$$

By setting $A - B = \omega t$ and $A + B = \omega_0 t$ and solving for A and B, we obtain

$$x(t) = \frac{2F_0}{m_0(\omega_0^2 - \omega^2)} \sin\left[\tfrac{1}{2}(\omega_0 + \omega)t\right] \sin\left[\tfrac{1}{2}(\omega_0 - \omega)t\right] , \tag{4.20}$$

which represents the solution as a frequency-dependent amplitude and two sinusoidal factors.

Figure 4.3 shows a plot of the quantity $X(t) = m_0 x(t)/2F_0$ for $\omega_0 = 1$ and $\omega = 0.9$. The solution is shows oscillatory behavior, as expected, but the most striking feature of this plot is the phenomenon of 'beats,' resulting from the superposition of a high-frequency oscillation, $\sin[\frac{1}{2}(\omega_0 + \omega)t]$, and a lower frequency envelope, $\sin[\frac{1}{2}(\omega_0 - \omega)t]$.

4.4 Summary

We can both summarize and generalize the main results of this chapter as follows. The solution of *any* nth-order ordinary differential equation depends, in general, on n arbitrary constants $c_1, c_2, \ldots, c_n$:

$$y = \varphi(x; c_1, c_2, \ldots, c_n) \tag{4.21}$$

Thus, to obtain a unique solution for a particular problem, it is necessary to supplement the differential equation with auxiliary conditions. A common choice is

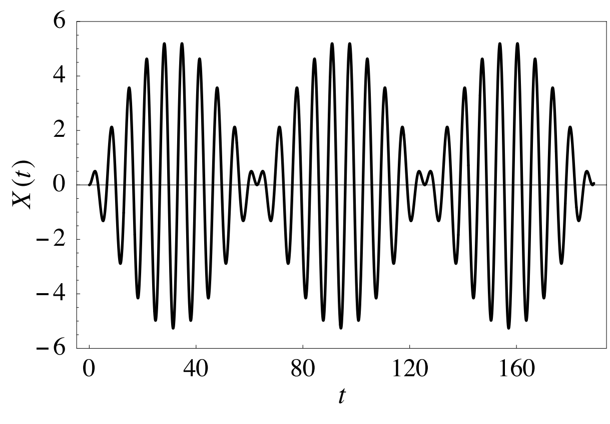


Figure 4.3: The quantity $X(t) = m_0 x(t)/2F_0$, where $x(t)$ is the solution in (4.20), for a undamped harmonic oscillator with a natural frequency $\omega_0 = 1$ driven by a sinusoidal force with a frequency $\omega = 0.9$.

for these constants to be determined from the initial values of the solution y and its first $n-1$ derivatives at some initial point x_0:

$$y(x_0) = A_0, \quad y'(x_0) = A_1, \quad \dots \quad y^{(n)}(x_0) = A_n$$

The expression in (4.21) is a general solution if it possible to satisfy these initial conditions for arbitrary values of the y_i with an appropriate choice of the c_j. This usually requires the solution of a system of algebraic equations.

For homogeneous linear nth-order equations,

$$a_n \frac{d^n y}{dx^n} + \cdots + a_1 \frac{dy}{dx} + a_0 y = 0$$

the general solution can be formed from any n linearly independent solutions $y_1, y_2, \dots, y_n$ of this equation:

$$y = c_1 y_1 + c_2 y_2 + \cdots + c_n y_n$$

The determination of the c_j from the initial conditions now reduces to the solution of a system of n linear algebraic equations.