Mathematical methods for physics and engineering

A comprehensive guide
Second edition

K. F. Riley, M. P. Hobson and S. J. Bence

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1998
Reprinted 1998 (with minor corrections), 2000 (twice), 2001
Second edition published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeset in TEX Monotype Times [EPC]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
Riley, K. F. (Kenneth Franklin), 1936–
Mathematical methods for physics and engineering / Ken Riley,
Mike Hobson, and Stephen Bence.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 81372 7 (HB) – ISBN 0 521 89067 5 (PB)
1. Mathematical analysis. I. Hobson, M. P. (Michael Paul), 1967– .
II. Bence, S. J. (Stephen John), 1972– . III. Title.
QA300.R495 2002
515′.1–dc21 2002018922 CIP

ISBN 0 521 81372 7 hardback
ISBN 0 521 89067 5 paperback

Contents

1

Preliminary algebra

This opening chapter reviews the basic algebra of which a working knowledge is presumed in the rest of the book. Many students will be familiar with much, if not all, of it, but recent changes in what is studied during secondary education mean that it cannot be taken for granted that they will already have a mastery of all the topics presented here. The reader may assess which areas need further study or revision by attempting the exercises at the end of the chapter. The main areas covered are polynomial equations and the related topic of partial fractions, curve sketching, coordinate geometry, trigonometric identities and the notions of proof by induction or contradiction.

1.1 Simple functions and equations

It is normal practice when starting the mathematical investigation of a physical problem to assign an algebraic symbol to the quantity whose value is sought, either numerically or as an explicit algebraic expression. For the sake of definiteness, in this chapter we will use x to denote this quantity most of the time. Subsequent steps in the analysis involve applying a combination of known laws, consistency conditions and (possibly) given constraints to derive one or more equations satisfied by x. These equations may take many forms, ranging from a simple polynomial equation to, say, a partial differential equation with several boundary conditions. Some of the more complicated possibilities are treated in the later chapters of this book, but for the present we will be concerned with techniques for the solution of relatively straightforward algebraic equations.

1.1.1 Polynomials and polynomial equations

Firstly we consider the simplest type of equation, a *polynomial equation*, in which a *polynomial* expression in x, denoted by $f(x)$, is set equal to zero and thereby forms an equation which is satisfied by particular values of x, called the *roots* of the equation:

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0. \tag{1.1}$$

Here n is an integer > 0, called the *degree* of both the polynomial and the equation, and the known coefficients $a_0, a_1, \ldots, a_n$ are real quantities with $a_n \neq 0$.

Equations such as (1.1) arise frequently in physical problems, the coefficients a_i being determined by the physical properties of the system under study. What is needed is to find some or all of the roots of (1.1), i.e. the x-values, α_k, that satisfy $f(\alpha_k) = 0$; here k is an index that, as we shall see later, can take up to n different values, i.e. $k = 1, 2, \ldots, n$. The roots of the polynomial equation can equally well be described as the zeroes of the polynomial. When they are *real*, they correspond to the points at which a graph of $f(x)$ crosses the x-axis. Roots that are complex (see chapter 3) do not have such a graphical interpretation.

For polynomial equations containing powers of x greater than x^4 general methods do not exist for obtaining explicit expressions for the roots α_k. Even for $n = 3$ and $n = 4$ the prescriptions for obtaining the roots are sufficiently complicated that it is usually preferable to obtain exact or approximate values by other methods. Only for $n = 1$ and $n = 2$ can closed-form solutions be given. These results will be well known to the reader, but they are given here for the sake of completeness. For $n = 1$, (1.1) reduces to the *linear* equation

$$a_1 x + a_0 = 0; \tag{1.2}$$

the solution (root) is $\alpha_1 = -a_0/a_1$. For $n = 2$, (1.1) reduces to the *quadratic* equation

$$a_2 x^2 + a_1 x + a_0 = 0; \tag{1.3}$$

the two roots α_1 and α_2 are given by

$$\alpha_{1,2} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2 a_0}}{2a_2}. \tag{1.4}$$

When discussing specifically quadratic equations, as opposed to more general polynomial equations, it is usual to write the equation in one of the two notations

$$ax^2 + bx + c = 0, \qquad ax^2 + 2bx + c = 0, \tag{1.5}$$

with respective explicit pairs of solutions

$$\alpha_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \qquad \alpha_{1,2} = \frac{-b \pm \sqrt{b^2 - ac}}{a}. \tag{1.6}$$

Of course, these two notations are entirely equivalent and the only important point is to associate each form of answer with the corresponding form of equation; most people keep to one form, to avoid any possible confusion.

If the value of the quantity appearing under the square root sign is positive then both roots are real; if it is negative then the roots form a complex conjugate pair, i.e. they are of the form $p \pm iq$ with p and q real (see chapter 3); if it has zero value then the two roots are equal and special considerations usually arise.

Thus linear and quadratic equations can be dealt with in a cut-and-dried way. We now turn to methods for obtaining partial information about the roots of higher-degree polynomial equations. In some circumstances the knowledge that an equation has a root lying in a certain range, or that it has no real roots at all, is all that is actually required. For example, in the design of electronic circuits it is necessary to know whether the current in a proposed circuit will break into spontaneous oscillation. To test this, it is sufficient to establish whether a certain polynomial equation, whose coefficients are determined by the physical parameters of the circuit, has a root with a positive real part (see chapter 3); complete determination of all the roots is not needed for this purpose. If the complete set of roots of a polynomial equation is required, it can usually be obtained to any desired accuracy by numerical methods such as those described in chapter 28.

There is no explicit step-by-step approach to finding the roots of a general polynomial equation such as (1.1). In most cases analytic methods yield only information *about* the roots, rather than their exact values. To explain the relevant techniques we will consider a particular example, 'thinking aloud' on paper and expanding on special points about methods and lines of reasoning. In more routine situations such comment would be absent and the whole process briefer and more tightly focussed.

Example: the cubic case

Let us investigate the roots of the equation

$$g(x) = 4x^3 + 3x^2 - 6x - 1 = 0 \tag{1.7}$$

or, in an alternative phrasing, investigate the zeroes of $g(x)$. We note first of all that this is a *cubic* equation. It can be seen that for x large and positive $g(x)$ will be large and positive and, equally, that for x large and negative $g(x)$ will be large and negative. Therefore, intuitively (or, more formally, by continuity) $g(x)$ must cross the x-axis at least once and so $g(x) = 0$ must have at least one real root. Furthermore, it can be shown that if $f(x)$ is an nth-degree polynomial then the graph of $f(x)$ must cross the x-axis an even or odd number of times as x varies between $-\infty$ and $+\infty$, according to whether n itself is even or odd. Thus a polynomial of odd degree always has at least one real root, but one of even degree may have no real root. A small complication, discussed later in this section, occurs when repeated roots arise.

Having established that $g(x) = 0$ has at least one real root, we may ask how many real roots it *could* have. To answer this we need one of the fundamental theorems of algebra, mentioned above:

An nth-degree polynomial equation has exactly n roots.

It should be noted that this does not imply that there are n *real* roots (only that there are not more than n); some of the roots may be of the form $p + iq$.

To make the above theorem plausible and to see what is meant by repeated roots, let us suppose that the nth-degree polynomial equation $f(x) = 0$, (1.1), has r roots $\alpha_1, \alpha_2, \ldots, \alpha_r$, considered distinct for the moment. That is, we suppose that $f(\alpha_k) = 0$ for $k = 1, 2, \ldots, r$, so that $f(x)$ vanishes only when x is equal to one of the r values α_k. But the same can be said for the function

$$F(x) = A(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_r), \tag{1.8}$$

in which A is a non-zero constant; $F(x)$ can clearly be multiplied out to form a polynomial expression.

We now call upon a second fundamental result in algebra: that if two polynomial functions $f(x)$ and $F(x)$ have equal values for *all* values of x, then their coefficients are equal on a term-by-term basis. In other words, we can equate the coefficients of each and every power of x in the two expressions (1.8) and (1.1); in particular we can equate the coefficients of the highest power of x. From this we have $Ax^r \equiv a_n x^n$ and thus that $r = n$ and $A = a_n$. As r is both equal to n and to the number of roots of $f(x) = 0$, we conclude that the nth-degree polynomial $f(x) = 0$ has n roots. (Although this line of reasoning may make the theorem plausible, it does not constitute a proof since we have not shown that it is permissible to write $f(x)$ in the form of equation (1.8).)

We next note that the condition $f(\alpha_k) = 0$ for $k = 1, 2, \ldots, r$, could also be met if (1.8) were replaced by

$$F(x) = A(x - \alpha_1)^{m_1}(x - \alpha_2)^{m_2}\cdots(x - \alpha_r)^{m_r}, \tag{1.9}$$

with $A = a_n$. In (1.9) the m_k are integers ≥ 1 and are known as the multiplicities of the roots, m_k being the multiplicity of α_k. Expanding the right-hand side (RHS) leads to a polynomial of degree $m_1 + m_2 + \cdots + m_r$. This sum must be equal to n. Thus, if any of the m_k is greater than unity then the number of *distinct* roots, r, is less than n; the total number of roots remains at n, but one or more of the α_k counts more than once. For example, the equation

$$F(x) = A(x - \alpha_1)^2(x - \alpha_2)^3(x - \alpha_3)(x - \alpha_4) = 0$$

has exactly seven roots, α_1 being a double root and α_2 a triple root, whilst α_3 and α_4 are unrepeated (*simple*) roots.

We can now say that our particular equation (1.7) has either one or three real roots but in the latter case it may be that not all the roots are distinct. To decide

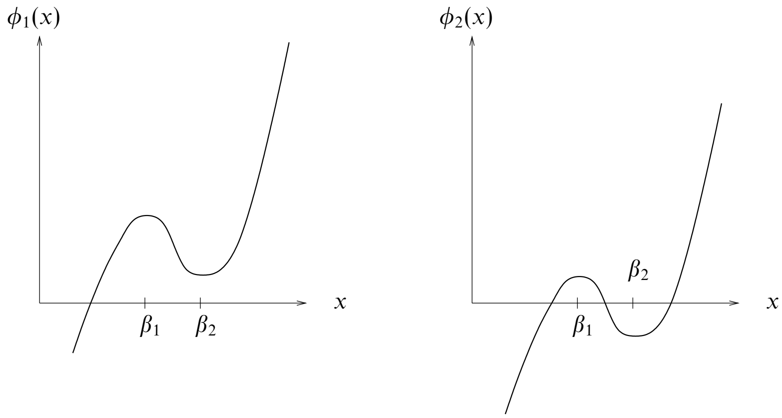

Figure 1.1 Two curves $\phi_1(x)$ and $\phi_2(x)$, both with zero derivatives at the same values of x, but with different numbers of real solutions to $\phi_i(x) = 0$.

how many real roots the equation has, we need to anticipate two ideas from the next chapter. The first of these is the notion of the derivative of a function, and the second is a result known as Rolle's theorem.

The *derivative* $f'(x)$ of a function $f(x)$ measures the slope of the tangent to the graph of $f(x)$ at that value of x (see figure 2.1 in the next chapter). For the moment, the reader with no prior knowledge of calculus is asked to accept that the derivative of ax^n is nax^{n-1}, so that the derivative $g'(x)$ of the curve $g(x) = 4x^3 + 3x^2 - 6x - 1$ is given by $g'(x) = 12x^2 + 6x - 6$. Similar expressions for the derivatives of other polynomials are used later in this chapter.

Rolle's theorem states that if $f(x)$ has equal values at two different values of x then at some point between these two x-values its derivative is equal to zero; i.e. the tangent to its graph is parallel to the x-axis at that point (see figure 2.2).

Having briefly mentioned the derivative of a function and Rolle's theorem, we now use them to establish whether $g(x)$ has one or three real zeroes. If $g(x) = 0$ does have three real roots α_k, i.e. $g(\alpha_k) = 0$ for $k = 1, 2, 3$, then it follows from Rolle's theorem that between any consecutive pair of them (say α_1 and α_2) there must be some real value of x at which $g'(x) = 0$. Similarly, there must be a further zero of $g'(x)$ lying between α_2 and α_3. Thus a *necessary* condition for three real roots of $g(x) = 0$ is that $g'(x) = 0$ itself has two real roots.

However, this condition on the number of roots of $g'(x) = 0$, whilst necessary, is not *sufficient* to guarantee three real roots of $g(x) = 0$. This can be seen by inspecting the cubic curves in figure 1.1. For each of the two functions $\phi_1(x)$ and $\phi_2(x)$, the derivative is equal to zero at both $x = \beta_1$ and $x = \beta_2$. Clearly, though, $\phi_2(x) = 0$ has three real roots whilst $\phi_1(x) = 0$ has only one. It is easy to see that the crucial difference is that $\phi_1(\beta_1)$ and $\phi_1(\beta_2)$ have the same sign, whilst $\phi_2(\beta_1)$ and $\phi_2(\beta_2)$ have opposite signs.

It will be apparent that for some equations, $\phi(x) = 0$ say, $\phi'(x)$ equals zero at a value of x for which $\phi(x)$ is also zero. Then the graph of $\phi(x)$ just touches the x-axis. When this happens the value of x so found is, in fact, a double real root of the polynomial equation (corresponding to one of the m_k in (1.9) having the value 2) and must be counted twice when determining the number of real roots.

Finally, then, we are in a position to decide the number of real roots of the equation

$$g(x) = 4x^3 + 3x^2 - 6x - 1 = 0.$$

The equation $g'(x) = 0$, with $g'(x) = 12x^2 + 6x - 6$, is a quadratic equation with explicit solutions[§]

$$\beta_{1,2} = \frac{-3 \pm \sqrt{9 + 72}}{12},$$

so that $\beta_1 = -1$ and $\beta_2 = \frac{1}{2}$. The corresponding values of $g(x)$ are $g(\beta_1) = 4$ and $g(\beta_2) = -\frac{11}{4}$, which are of opposite sign. This indicates that $4x^3 + 3x^2 - 6x - 1 = 0$ has three real roots, one lying in the range $-1 < x < \frac{1}{2}$ and the others one on each side of that range.

The techniques we have developed above have been used to tackle a cubic equation, but they can be applied to polynomial equations $f(x) = 0$ of degree greater than 3. However, much of the analysis centres around the equation $f'(x) = 0$ and this, itself, being then a polynomial equation of degree 3 or more either has no closed-form general solution or one that is complicated to evaluate. Thus the amount of information that can be obtained about the roots of $f(x) = 0$ is correspondingly reduced.

A more general case

To illustrate what can (and cannot) be done in the more general case we now investigate as far as possible the real roots of

$$f(x) = x^7 + 5x^6 + x^4 - x^3 + x^2 - 2 = 0.$$

The following points can be made.

(i) This is a seventh-degree polynomial equation; therefore the number of real roots is 1, 3, 5 or 7.

(ii) $f(0)$ is negative whilst $f(\infty) = +\infty$, so there must be at least one positive root.

[§] The two roots β_1, β_2 are written as $\beta_{1,2}$. By convention β_1 refers to the upper symbol in $\pm$, β_2 to the lower symbol.

(iii) The equation $f'(x) = 0$ can be written as $x(7x^5 + 30x^4 + 4x^2 - 3x + 2) = 0$ and thus $x = 0$ is a root. The derivative of $f'(x)$, denoted by $f''(x)$, equals $42x^5 + 150x^4 + 12x^2 - 6x + 2$. That $f'(x)$ is zero whilst $f''(x)$ is positive at $x = 0$ indicates (subsection 2.1.8) that $f(x)$ has a minimum there. This, together with the facts that $f(0)$ is negative and $f(\infty) = \infty$, implies that the total number of real roots to the right of $x = 0$ must be odd. Since the total number of real roots must be odd, the number to the left must be even (0, 2, 4 or 6).

This is about all that can be deduced by *simple* analytic methods in this case, although some further progress can be made in the ways indicated in exercise 1.3.

There are, in fact, more sophisticated tests that examine the relative signs of successive terms in an equation such as (1.1), and in quantities derived from them, to place limits on the numbers and positions of roots. But they are not prerequisites for the remainder of this book and will not be pursued further here.

We conclude this section with a worked example which demonstrates that the practical application of the ideas developed so far can be both short and decisive.

►*For what values of k, if any, does*

$$f(x) = x^3 - 3x^2 + 6x + k = 0$$

have three real roots?

Firstly we study the equation $f'(x) = 0$, i.e. $3x^2 - 6x + 6 = 0$. This is a quadratic equation but, using (1.6), because $6^2 < 4 \times 3 \times 6$, it can have no real roots. Therefore, it follows immediately that $f(x)$ has no maximum or minimum; consequently $f(x) = 0$ cannot have more than one real root, whatever the value of k. ◄

1.1.2 Factorising polynomials

In the previous subsection we saw how a polynomial with r given distinct zeroes α_k could be constructed as the product of factors containing those zeroes:

$$\begin{aligned} f(x) &= a_n(x - \alpha_1)^{m_1}(x - \alpha_2)^{m_2} \cdots (x - \alpha_r)^{m_r} \\ &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \end{aligned} \tag{1.10}$$

with $m_1 + m_2 + \cdots + m_r = n$, the degree of the polynomial. It will cause no loss of generality in what follows to suppose that all the zeroes are simple, i.e. all $m_k = 1$ and $r = n$, and this we will do.

Sometimes it is desirable to be able to reverse this process, in particular when one exact zero has been found by some method and the remaining zeroes are to be investigated. Suppose that we have located one zero, α; it is then possible to write (1.10) as

$$f(x) = (x - \alpha)f_1(x), \tag{1.11}$$

where $f_1(x)$ is a polynomial of degree $n-1$. How can we find $f_1(x)$? The procedure is much more complicated to describe in a general form than to carry out for an equation with given numerical coefficients a_i. If such manipulations are too complicated to be carried out mentally, they could be laid out along the lines of an algebraic 'long division' sum. However, a more compact form of calculation is as follows. Write $f_1(x)$ as

$$f_1(x) = b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + b_{n-3}x^{n-3} + \cdots + b_1 x + b_0.$$

Substitution of this form into (1.11) and subsequent comparison of the coefficients of x^p for $p = n,\ n-1,\ \ldots,\ 1,\ 0$ with those in the second line of (1.10) generates the series of equations

$$\begin{aligned} b_{n-1} &= a_n, \\ b_{n-2} - \alpha b_{n-1} &= a_{n-1}, \\ b_{n-3} - \alpha b_{n-2} &= a_{n-2}, \\ &\ \vdots \\ b_0 - \alpha b_1 &= a_1, \\ -\alpha b_0 &= a_0. \end{aligned}$$

These can be solved successively for the b_j, starting either from the top or from the bottom of the series. In either case the final equation used serves as a check; if it is not satisfied, at least one mistake has been made in the computation – or α is not a zero of $f(x) = 0$. We now illustrate this procedure with a worked example.

►*Determine by inspection the simple roots of the equation*

$$f(x) = 3x^4 - x^3 - 10x^2 - 2x + 4 = 0$$

and hence, by factorisation, find the rest of its roots.

From the pattern of coefficients it can be seen that $x = -1$ is a solution to the equation. We therefore write

$$f(x) = (x+1)(b_3x^3 + b_2x^2 + b_1x + b_0),$$

where

$$\begin{aligned} b_3 &= 3, \\ b_2 + b_3 &= -1, \\ b_1 + b_2 &= -10, \\ b_0 + b_1 &= -2, \\ b_0 &= 4. \end{aligned}$$

These equations give $b_3 = 3, b_2 = -4, b_1 = -6, b_0 = 4$ (check) and so

$$f(x) = (x+1)f_1(x) = (x+1)(3x^3 - 4x^2 - 6x + 4).$$

We now note that $f_1(x) = 0$ if x is set equal to 2. Thus $x - 2$ is a factor of $f_1(x)$, which therefore can be written as

$$f_1(x) = (x-2)f_2(x) = (x-2)(c_2x^2 + c_1x + c_0)$$

with

$$\begin{aligned} c_2 &= 3, \\ c_1 - 2c_2 &= -4, \\ c_0 - 2c_1 &= -6, \\ -2c_0 &= 4. \end{aligned}$$

These equations determine $f_2(x)$ as $3x^2 + 2x - 2$. Since $f_2(x) = 0$ is a quadratic equation, its solutions can be written explicitly as

$$x = \frac{-1 \pm \sqrt{1+6}}{3}.$$

Thus the four roots of $f(x) = 0$ are $-1, 2, \frac{1}{3}(-1+\sqrt{7})$ and $\frac{1}{3}(-1-\sqrt{7})$. ◄

1.1.3 Properties of roots

From the fact that a polynomial equation can be written in any of the alternative forms

$$\begin{aligned} f(x) &= a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0, \\ f(x) &= a_n(x-\alpha_1)^{m_1}(x-\alpha_2)^{m_2}\cdots(x-\alpha_r)^{m_r} = 0, \\ f(x) &= a_n(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n) = 0, \end{aligned}$$

it follows that it must be possible to express the coefficients a_i in terms of the roots α_k. To take the most obvious example, comparison of the constant terms (formally the coefficient of x^0) in the first and third expressions shows that

$$a_n(-\alpha_1)(-\alpha_2)\cdots(-\alpha_n) = a_0,$$

or, using the product notation,

$$\prod_{k=1}^{n} \alpha_k = (-1)^n \frac{a_0}{a_n}. \tag{1.12}$$

Only slightly less obvious is a result obtained by comparing the coefficients of x^{n-1} in the same two expressions of the polynomial:

$$\sum_{k=1}^{n} \alpha_k = -\frac{a_{n-1}}{a_n}. \tag{1.13}$$

Comparing the coefficients of other powers of x yields further results, though they are of less general use than the two just given. One such, which the reader may wish to derive, is

$$\sum_{j=1}^{n}\sum_{k>j}^{n} \alpha_j\alpha_k = \frac{a_{n-2}}{a_n}. \tag{1.14}$$

In the case of a quadratic equation these root properties are used sufficiently often that they are worth stating explicitly, as follows. If the roots of the quadratic equation $ax^2 + bx + c = 0$ are α_1 and α_2 then

$$\alpha_1 + \alpha_2 = -\frac{b}{a},$$
$$\alpha_1\alpha_2 = \frac{c}{a}.$$

If the alternative standard form for the quadratic is used, b is replaced by $2b$ in both the equation and the first of these results.

►*Find a cubic equation whose roots are* $-4, 3$ *and* 5.

From results (1.12) – (1.14) we can compute that, arbitrarily setting $a_3 = 1$,

$$-a_2 = \sum_{k=1}^{3} \alpha_k = 4, \qquad a_1 = \sum_{j=1}^{3}\sum_{k>j}^{3} \alpha_j\alpha_k = -17, \qquad a_0 = (-1)^3 \prod_{k=1}^{3} \alpha_k = 60.$$

Thus a possible cubic equation is $x^3 + (-4)x^2 + (-17)x + (60) = 0$. Of course, any multiple of $x^3 - 4x^2 - 17x + 60 = 0$ will do just as well. ◄

1.2 Trigonometric identities

So many of the applications of mathematics to physics and engineering are concerned with periodic, and in particular sinusoidal, behaviour that a sure and ready handling of the corresponding mathematical functions is an essential skill. Even situations with no obvious periodicity are often expressed in terms of periodic functions for the purposes of analysis. Later in this book whole chapters are devoted to developing the techniques involved, but as a necessary prerequisite we here establish (or remind the reader of) some standard identities with which he or she should be fully familiar, so that the manipulation of expressions containing sinusoids becomes automatic and reliable. So as to emphasise the angular nature of the argument of a sinusoid we will denote it in this section by θ rather than x.

1.2.1 Single-angle identities

We give without proof the basic identity satisfied by the sinusoidal functions $\sin\theta$ and $\cos\theta$, namely

$$\cos^2\theta + \sin^2\theta = 1. \tag{1.15}$$

If $\sin\theta$ and $\cos\theta$ have been defined geometrically in terms of the coordinates of a point on a circle, a reference to the name of Pythagoras will suffice to establish this result. If they have been defined by means of series (with θ expressed in radians) then the reader should refer to Euler's equation (3.23) on page 96, and note that $e^{i\theta}$ has unit modulus if θ is real.

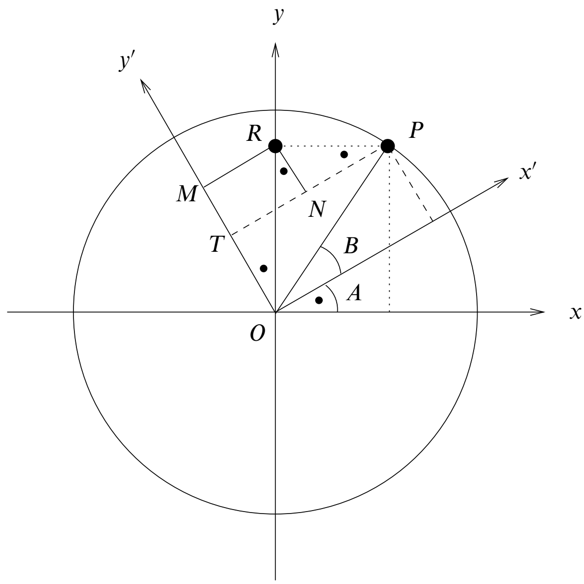

Figure 1.2 Illustration of the compound-angle identities. Refer to the main text for details.

Other standard single-angle formulae derived from (1.15) by dividing through by various powers of $\sin\theta$ and $\cos\theta$ are

$$1 + \tan^2\theta = \sec^2\theta, \tag{1.16}$$

$$\cot^2\theta + 1 = \mathrm{cosec}\,^2\theta. \tag{1.17}$$

1.2.2 Compound-angle identities

The basis for building expressions for the sinusoidal functions of compound angles are those for the sum and difference of just two angles, since all other cases can be built up from these, in principle. Later we will see that a study of complex numbers can provide a more efficient approach in some cases.

To prove the basic formulae for the sine and cosine of a compound angle $A+B$ in terms of the sines and cosines of A and B, we consider the construction shown in figure 1.2. It shows two sets of axes, Oxy and $Ox'y'$, with a common origin but rotated with respect to each other through an angle A. The point P lies on the unit circle centred on the common origin O and has coordinates $\cos(A+B), \sin(A+B)$ with respect to the axes Oxy and coordinates $\cos B, \sin B$ with respect to the axes $Ox'y'$.

Parallels to the axes Oxy (dotted lines) and $Ox'y'$ (broken lines) have been drawn through P. Further parallels (MR and RN) to the $Ox'y'$ axes have been drawn through R, the point $(0, \sin(A + B))$ in the Oxy system. That all the angles marked with the symbol • are equal to A follows from the simple geometry of right-angled triangles and crossing lines.

We now determine the coordinates of P in terms of lengths in the figure, expressing those lengths in terms of both sets of coordinates:

(i) $\cos B = x' = TN + NP = MR + NP$
$$= OR \sin A + RP \cos A = \sin(A + B) \sin A + \cos(A + B) \cos A;$$
(ii) $\sin B = y' = OM - TM = OM - NR$
$$= OR \cos A - RP \sin A = \sin(A + B) \cos A - \cos(A + B) \sin A.$$

Now, if equation (i) is multiplied by $\sin A$ and added to equation (ii) multiplied by $\cos A$, the result is

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

Similarly, if equation (ii) is multiplied by $\sin A$ and subtracted from equation (i) multiplied by $\cos A$, the result is

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

Corresponding graphically based results can be derived for the sines and cosines of the difference of two angles; however, they are more easily obtained by setting B to $-B$ in the previous results and remembering that $\sin B$ becomes $-\sin B$ whilst $\cos B$ is unchanged. The four results may be summarised by

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B \tag{1.18}$$
$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B. \tag{1.19}$$

Standard results can be deduced from these by setting one of the two angles equal to π or to $\pi/2$:

$$\sin(\pi - \theta) = \sin \theta, \qquad \cos(\pi - \theta) = -\cos \theta, \tag{1.20}$$
$$\sin\left(\tfrac{1}{2}\pi - \theta\right) = \cos \theta, \qquad \cos\left(\tfrac{1}{2}\pi - \theta\right) = \sin \theta, \tag{1.21}$$

From these basic results many more can be derived. An immediate deduction, obtained by taking the ratio of the two equations (1.18) and (1.19) and then dividing both the numerator and denominator of this ratio by $\cos A \cos B$, is

$$\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A \tan B}. \tag{1.22}$$

One application of this result is a test for whether two lines on a graph are orthogonal (perpendicular); more generally, it determines the angle between them. The standard notation for a straight-line graph is $y = mx + c$, in which m is the slope of the graph and c is its intercept on the y-axis. It should be noted that the slope m is also the tangent of the angle the line makes with the x-axis.

Consequently the angle θ_{12} between two such straight-line graphs is equal to the difference in the angles they individually make with the x-axis, and the tangent of that angle is given by (1.22):

$$\tan\theta_{12} = \frac{\tan\theta_1 - \tan\theta_2}{1 + \tan\theta_1 \tan\theta_2} = \frac{m_1 - m_2}{1 + m_1 m_2}. \tag{1.23}$$

For the lines to be orthogonal we must have $\theta_{12} = \pi/2$, i.e. the final fraction on the RHS of the above equation must equal ∞, and so

$$m_1 m_2 = -1. \tag{1.24}$$

A kind of inversion of equations (1.18) and (1.19) enables the sum or difference of two sines or cosines to be expressed as the product of two sinusoids; the procedure is typified by the following. Adding together the expressions given by (1.18) for $\sin(A+B)$ and $\sin(A-B)$ yields

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

If we now write $A + B = C$ and $A - B = D$, this becomes

$$\sin C + \sin D = 2\sin\left(\frac{C+D}{2}\right)\cos\left(\frac{C-D}{2}\right). \tag{1.25}$$

In a similar way each of the following equations can be derived:

$$\sin C - \sin D = 2\cos\left(\frac{C+D}{2}\right)\sin\left(\frac{C-D}{2}\right), \tag{1.26}$$

$$\cos C + \cos D = 2\cos\left(\frac{C+D}{2}\right)\cos\left(\frac{C-D}{2}\right), \tag{1.27}$$

$$\cos C - \cos D = -2\sin\left(\frac{C+D}{2}\right)\sin\left(\frac{C-D}{2}\right). \tag{1.28}$$

The minus sign on the right of the last of these equations should be noted; it may help to avoid overlooking this 'oddity' to recall that if $C > D$ then $\cos C < \cos D$.

1.2.3 Double- and half-angle identities

Double-angle and half-angle identities are needed so often in practical calculations that they should be committed to memory by any physical scientist. They can be obtained by setting B equal to A in results (1.18) and (1.19). When this is done, and use made of equation (1.15), the following results are obtained:

$$\sin 2\theta = 2\sin\theta\cos\theta, \tag{1.29}$$

$$\begin{aligned}\cos 2\theta &= \cos^2\theta - \sin^2\theta \\ &= 2\cos^2\theta - 1 \\ &= 1 - 2\sin^2\theta,\end{aligned} \tag{1.30}$$

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta}. \tag{1.31}$$

A further set of identities enables sinusoidal functions of θ to be expressed as polynomial functions of a variable $t = \tan(\theta/2)$. They are not used in their primary role until the next chapter, but we give a derivation of them here for reference.

If $t = \tan(\theta/2)$, then it follows from (1.16) that $1+t^2 = \sec^2(\theta/2)$ and $\cos(\theta/2) = (1+t^2)^{-1/2}$, whilst $\sin(\theta/2) = t(1+t^2)^{-1/2}$. Now, using (1.29) and (1.30), we may write:

$$\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} = \frac{2t}{1+t^2}, \tag{1.32}$$

$$\cos\theta = \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} = \frac{1-t^2}{1+t^2}, \tag{1.33}$$

$$\tan\theta = \frac{2t}{1-t^2}. \tag{1.34}$$

It can be further shown that the derivative of θ with respect to t takes the algebraic form $2/(1+t^2)$. This completes a package of results that enables expressions involving sinusoids, particularly when they appear as integrands, to be cast in more convenient algebraic forms. The proof of the derivative property and examples of use of the above results are given in subsection (2.2.7).

We conclude this section with a worked example which is of such a commonly occurring form that it might be considered a standard procedure.

►*Solve for θ the equation*

$$a\sin\theta + b\cos\theta = k,$$

where a, b and k are given real quantities.

To solve this equation we make use of result (1.18) by setting $a = K\cos\phi$ and $b = K\sin\phi$ for suitable values of K and ϕ. We then have

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

with

$$K^2 = a^2 + b^2 \qquad \text{and} \qquad \phi = \tan^{-1}\frac{b}{a}.$$

Whether ϕ lies in $0 \le \phi \le \pi$ or in $-\pi < \phi < 0$ has to be determined by the individual signs of a and b. The solution is thus

$$\theta = \sin^{-1}\left(\frac{k}{K}\right) - \phi,$$

with K and ϕ as given above. Notice that there is no real solution to the original equation if $|k| > |K| = (a^2 + b^2)^{1/2}$. ◀

1.3 Coordinate geometry

We have already mentioned the standard form for a straight-line graph, namely

$$y = mx + c, \tag{1.35}$$

representing a linear relationship between the independent variable x and the dependent variable y. The slope m is equal to the tangent of the angle the line makes with the x-axis whilst c is the intercept on the y-axis.

An alternative form for the equation of a straight line is

$$ax + by + k = 0, \tag{1.36}$$

to which (1.35) is clearly connected by

$$m = -\frac{a}{b} \qquad \text{and} \qquad c = -\frac{k}{b}.$$

This form treats x and y on a more symmetrical basis, the intercepts on the two axes being $-k/a$ and $-k/b$ respectively.

A power relationship between two variables, i.e. one of the form $y = Ax^n$, can also be cast into straight-line form by taking the logarithms of both sides. Whilst it is normal in mathematical work to use natural logarithms (to base e, written $\ln x$), for practical investigations logarithms to base 10 are often employed. In either case the form is the same, but it needs to be remembered which has been used when recovering the value of A from fitted data. In the mathematical (base e) form, the power relationship becomes

$$\ln y = n \ln x + \ln A. \tag{1.37}$$

Now the slope gives the power n, whilst the intercept on the $\ln y$ axis is $\ln A$, which yields A, either by exponentiation or by taking antilogarithms.

The other standard coordinate forms of two-dimensional curves that students should know and recognise are those concerned with the *conic sections* – so called because they can all be obtained by taking suitable sections across a (double) cone. Because the conic sections can take many different orientations and scalings their general form is complex,

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0, \tag{1.38}$$

but each can be represented by one of four generic forms, an ellipse, a parabola, a hyperbola or, the degenerate form, a pair of straight lines. If they are reduced to their standard representations, in which axes of symmetry are made to coincide with the coordinate axes, the first three take the forms

$$\frac{(x-\alpha)^2}{a^2} + \frac{(y-\beta)^2}{b^2} = 1 \qquad \text{(ellipse)}, \tag{1.39}$$

$$(y-\beta)^2 = 4a(x-\alpha) \qquad \text{(parabola)}, \tag{1.40}$$

$$\frac{(x-\alpha)^2}{a^2} - \frac{(y-\beta)^2}{b^2} = 1 \qquad \text{(hyperbola)}. \tag{1.41}$$

Here, (α, β) gives the position of the 'centre' of the curve, usually taken as the origin $(0, 0)$ when this does not conflict with any imposed conditions. The parabola equation given is that for a curve symmetric about a line parallel to the x-axis. For one symmetrical about a parallel to the y-axis the equation would read $(x-\alpha)^2 = 4a(y-\beta)$.

Of course, the circle is the special case of an ellipse in which $b = a$ and the equation takes the form

$$(x-\alpha)^2 + (y-\beta)^2 = a^2. \tag{1.42}$$

The distinguishing characteristic of this equation is that when it is expressed in the form (1.38) the coefficients of x^2 and y^2 are equal and that of xy is zero; this property is not changed by any reorientation or scaling and so acts to identify a general conic as a circle.

Definitions of the conic sections in terms of geometrical properties are also available; for example, a parabola can be defined as the locus of a point that is always at the same distance from a given straight line (the *directrix*) as it is from a given point (the *focus*). When these properties are expressed in Cartesian coordinates the above equations are obtained. For a circle, the defining property is that all points on the curve are a distance a from (α, β); (1.42) expresses this requirement very directly. In the following worked example we derive the equation for a parabola.

►*Find the equation of a parabola that has the line $x = -a$ as its directrix and the point $(a, 0)$ as its focus.*

Figure 1.3 shows the situation in Cartesian coordinates. Expressing the defining requirement that PN and PF are equal in length gives

$$(x+a) = [(x-a)^2 + y^2]^{1/2} \quad \Rightarrow \quad (x+a)^2 = (x-a)^2 + y^2$$

which, on expansion of the squared terms, immediately gives $y^2 = 4ax$. This is (1.40) with α and β both set equal to zero. ◄

Although the algebra is more complicated, the same method can be used to derive the equations for the ellipse and the hyperbola. In these cases the distance from the fixed point is a definite fraction, e, known as the *eccentricity*, of the distance from the fixed line. For an ellipse $0 < e < 1$, for a circle $e = 0$, and for a hyperbola $e > 1$. The parabola corresponds to the case $e = 1$.

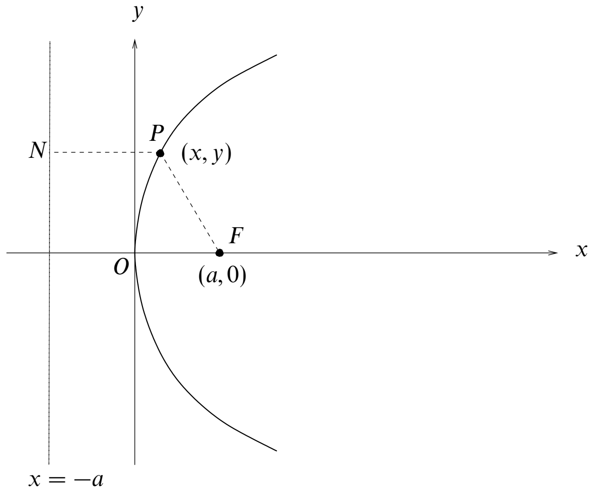

Figure 1.3 Construction of a parabola using the point $(a, 0)$ as the focus and the line $x = -a$ as the directrix.

The values of a and b (with $a \geq b$) in equation (1.39) for an ellipse are related to e through

$$e^2 = \frac{a^2 - b^2}{a^2}$$

and give the lengths of the semi-axes of the ellipse. If the ellipse is centred on the origin, i.e. $\alpha = \beta = 0$, then the focus is $(-ae, 0)$ and the directrix is the line $x = -a/e$.

For each conic section curve, although we have two variables, x and y, they are not independent, since if one is given then the other can be determined. However, determining y when x is given, say, involves solving a quadratic equation on each occasion, and so it is convenient to have *parametric* representations of the curves. A parametric representation allows each point on a curve to be associated with a unique value of a *single* parameter t. The simplest parametric representations for the conic sections are as given below, though that for the hyperbola uses hyperbolic functions, not formally introduced until chapter 3. That they do give valid parameterizations can be verified by substituting them into the standard forms (1.39)–(1.41); in each case the standard form is reduced to an algebraic or trigonometric identity.

$$\begin{aligned}
x &= \alpha + a\cos\phi, & y &= \beta + b\sin\phi & &\text{(ellipse)},\\
x &= \alpha + at^2, & y &= \beta + 2at & &\text{(parabola)},\\
x &= \alpha + a\cosh\phi, & y &= \beta + b\sinh\phi & &\text{(hyperbola)}.
\end{aligned}$$

As a final example illustrating several topics from this section we now prove the well-known result that the angle subtended by a diameter at any point on a circle is a right angle.

►*Taking the diameter to be the line joining $Q = (-a, 0)$ and $R = (a, 0)$ and the point P to be any point on the circle $x^2 + y^2 = a^2$, prove that angle QPR is a right angle.*

If P is the point (x, y), the slope of the line QP is

$$m_1 = \frac{y - 0}{x - (-a)} = \frac{y}{x + a}.$$

That of RP is

$$m_2 = \frac{y - 0}{x - (a)} = \frac{y}{x - a}.$$

Thus

$$m_1 m_2 = \frac{y^2}{x^2 - a^2}.$$

But, since P is on the circle, $y^2 = a^2 - x^2$ and consequently $m_1 m_2 = -1$. From result (1.24) this implies that QP and RP are orthogonal and that QPR is therefore a right angle. Note that this is true for *any* point P on the circle. ◄

1.4 Partial fractions

In subsequent chapters, and in particular when we come to study integration in chapter 2, we will need to express a function $f(x)$ that is the ratio of two polynomials in a more manageable form. To remove some potential complexity from our discussion we will assume that all the coefficients in the polynomials are real, although this is not an essential simplification.

The behaviour of $f(x)$ is crucially determined by the location of the zeroes of its denominator, i.e. if $f(x)$ is written as $f(x) = g(x)/h(x)$ where both $g(x)$ and $h(x)$ are polynomials,[§] then $f(x)$ changes extremely rapidly when x is close to those values α_i that are the roots of $h(x) = 0$. To make such behaviour explicit, we write $f(x)$ as a sum of terms such as $A/(x - \alpha)^n$, in which A is a constant, α is one of the α_i that satisfy $h(\alpha_i) = 0$ and n is a positive integer. Writing a function in this way is known as expressing it in *partial fractions*.

Suppose, for the sake of definiteness, that we wish to express the function

$$f(x) = \frac{4x + 2}{x^2 + 3x + 2}$$

[§] It is assumed that the ratio has been reduced so that $g(x)$ and $h(x)$ do not contain any common factors, i.e. there is no value of x that makes both vanish at the same time. We may also assume without any loss of generality that the coefficient of the highest power of x in $h(x)$ has been made equal to unity, if necessary, by dividing both numerator and denominator by the coefficient of this highest power.

in partial fractions, i.e. to write it as

$$f(x) = \frac{g(x)}{h(x)} = \frac{4x+2}{x^2+3x+2} = \frac{A_1}{(x-\alpha_1)^{n_1}} + \frac{A_2}{(x-\alpha_2)^{n_2}} + \cdots . \tag{1.43}$$

The first question that arises is that of how many terms there should be on the right-hand side (RHS). Although some complications occur when $h(x)$ has repeated roots (these are considered below) it is clear that $f(x)$ only becomes infinite at the *two* values of x, α_1 and α_2, that make $h(x) = 0$. Consequently the RHS can only become infinite at the same two values of x and therefore contains only two partial fractions – these are the ones shown explicitly. This argument can be trivially extended (again temporarily ignoring the possibility of repeated roots of $h(x)$) to show that if $h(x)$ is a polynomial of degree n then there should be n terms on the RHS, each containing a different root α_i of the equation $h(\alpha_i) = 0$.

A second general question concerns the appropriate values of the n_i. This is answered by putting the RHS over a common denominator, which will clearly have to be the product $(x-\alpha_1)^{n_1}(x-\alpha_2)^{n_2}\cdots$. Comparison of the highest power of x in this new RHS with the same power in $h(x)$ shows that $n_1 + n_2 + \cdots = n$. This result holds whether or not $h(x) = 0$ has repeated roots and, although we do not give a rigorous proof, strongly suggests the following correct conclusions.

- The number of terms on the RHS is equal to the number of *distinct* roots of $h(x) = 0$, each term having a different root α_i in its denominator $(x-\alpha_i)^{n_i}$.
- If α_i is a multiple root of $h(x) = 0$ then the value to be assigned to n_i in (1.43) is that of m_i when $h(x)$ is written in the product form (1.9). Further, as discussed on p. 23, A_i has to be replaced by a polynomial of degree $m_i - 1$. This is also formally true for non-repeated roots, since then both m_i and n_i are equal to unity.

Returning to our specific example we note that the denominator $h(x)$ has zeroes at $x = \alpha_1 = -1$ and $x = \alpha_2 = -2$; these x-values are the simple (non-repeated) roots of $h(x) = 0$. Thus the partial fraction expansion will be of the form

$$\frac{4x+2}{x^2+3x+2} = \frac{A_1}{x+1} + \frac{A_2}{x+2}. \tag{1.44}$$

We now list several methods available for determining the coefficients A_1 and A_2. We also remind the reader that, as with all the explicit examples and techniques described, these methods are to be considered as models for the handling of any ratio of polynomials, with or without characteristics that make it a special case.

(i) The RHS can be put over a common denominator, in this case $(x+1)(x+2)$, and then the coefficients of the various powers of x can be equated in the numerators on both sides of the equation. This leads to

$$4x + 2 = A_1(x + 2) + A_2(x + 1),$$
$$4 = A_1 + A_2 \qquad 2 = 2A_1 + A_2.$$

Solving the simultaneous equations for A_1 and A_2 gives $A_1 = -2$ and $A_2 = 6$.

(ii) A second method is to substitute two (or more generally n) different values of x into each side of (1.44) and so obtain two (or n) simultaneous equations for the two (or n) constants A_i. To justify this practical way of proceeding it is necessary, strictly speaking, to appeal to method (i) above, which establishes that there are unique values for A_1 and A_2 valid for all values of x. It is normally very convenient to take zero as one of the values of x, but of course any set will do. Suppose in the present case that we use the values $x = 0$ and $x = 1$ and substitute in (1.44). The resulting equations are

$$\frac{2}{2} = \frac{A_1}{1} + \frac{A_2}{2},$$
$$\frac{6}{6} = \frac{A_1}{2} + \frac{A_2}{3},$$

which on solution give $A_1 = -2$ and $A_2 = 6$, as before. The reader can easily verify that any other pair of values for x (except for a pair that includes α_1 or α_2) gives the same values for A_1 and A_2.

(iii) The very reason why method (ii) fails if x is chosen as one of the roots α_i of $h(x) = 0$ can be made the basis for determining the values of the A_i corresponding to non-multiple roots without having to solve simultaneous equations. The method is conceptually more difficult than the other methods presented here, and needs results from the theory of complex variables (chapter 20) to justify it. However, we give a practical 'cookbook' recipe for determining the coefficients.

(a) To determine the coefficient A_k, imagine the denominator $h(x)$ written as the product $(x - \alpha_1)(x - \alpha_2)\cdots(x - \alpha_n)$, with any m-fold repeated root giving rise to m factors in parentheses.

(b) Now set x equal to α_k and evaluate the expression obtained after omitting the factor that reads $\alpha_k - \alpha_k$.

(c) Divide the value so obtained into $g(\alpha_k)$; the result is the required coefficient A_k.

For our specific example we find that in step (a) that $h(x) = (x + 1)(x + 2)$ and that in evaluating A_1 step (b) yields $-1 + 2$, i.e. 1. Since $g(-1) = 4(-1) + 2 = -2$, step (c) gives A_1 as $(-2)/(1)$, i.e in agreement with our other evaluations. In a similar way A_2 is evaluated as $(-6)/(-1) = 6$.