UNIVERSAL
LIBRARY
OU_152752
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 517/K99M Accession No. G10167

Author Kynch, G. J.

Title Mathematics for the chemist. 1965

This book should be returned on or before the date last marked below.

MATHEMATICS FOR THE CHEMIST

MATHEMATICS FOR THE CHEMIST

MATHEMATICAL ANALYSIS FOR CHEMISTS PHYSICISTS AND CHEMICAL ENGINEERS

G. J. KYNCH

Professor of Applied Mathematics
University College of Wales, Aberystwyth

LONDON
BUTTERWORTHS SCIENTIFIC PUBLICATIONS
1955

BUTTERWORTHS PUBLICATIONS LTD.
88 KINGSWAY, LONDON, W.C.2

AFRICA: BUTTERWORTH & CO. (AFRICA) LTD.
DURBAN: 33/35 Beach Grove

AUSTRALIA: BUTTERWORTH & CO. (AUSTRALIA) LTD.
SYDNEY: 8 O'Connell Street
MELBOURNE: 430 Bourke Street
BRISBANE: 240 Queen Street

CANADA: BUTTERWORTH & CO. (CANADA) LTD.
TORONTO: 1367 Danforth Avenue

NEW ZEALAND: BUTTERWORTH & CO. (AUSTRALIA) LTD.
WELLINGTON: 49/51 Ballance Street
AUCKLAND: 35 High Street

U.S.A. Edition published by
ACADEMIC PRESS INC. PUBLISHERS
125 EAST 23RD STREET
NEW YORK 10, NEW YORK

First Impression, 1955

Second Impression, 1956

Set in Monotype Baskerville type
Made and printed in Great Britain by
The Pitman Press, Bath

PREFACE

MANY chemists have expressed to me the need for a book on mathematics for chemists, a view supported by the popularity of the book by Mellor with those fortunate enough to possess a copy, although it was published about forty years ago. Moreover, the use of mathematics in chemistry is steadily increasing, so that students learn far more mathematics nowadays, and older chemists are often in need of a link between their chemistry and the new types of mathematics in vogue.

This book has three purposes: one, to meet the needs of students taking advanced or Honours Courses in Chemistry at universities and technical colleges; two, to provide a text where the emphasis is on the applications of mathematics, using the language of chemists wherever possible; and three to provide a reference book for practising chemists and biochemists who themselves wish to apply mathematics and understand new theories.

To students I would say: here you will find nearly all the syllabus for the Honours Course in Chemistry at the University of London, also some other items which will be useful later. There are no long lists of examples to work, but there are many examples relating the mathematics to the chemistry which you are learning. The mathematical principles are explained briefly. I have often been told that chemists do not want proofs, but do want to know when a method can be used and when it cannot. I have compromised, so that bald statements about conditions are amplified to make them intelligible where this is desirable.

To practising chemists I would say: the two questions you have most frequently put to me have been: How did he do this? and, How would you do this? So you will find here the formulae which seem so elementary to the student fresh from school, and many of the short cuts loved by the mathematically inclined. Mathematics, like many other things, is forgotten if it is not practised, and the elementary parts seem most difficult (as I found myself recently when trying to do some very, very elementary chemistry). As far as possible every step in the worked examples has been included.

I regret that numerical integration and other numerical techniques have not been discussed more fully. In my opinion chemists often know far too little about these matters, but the detailed working of examples, which is so necessary to a proper explanation, would have made the book too long. To make up for

this the numerical background of the work included has been emphasized as much as possible. Chapters on the theory of probability and errors have been omitted for the same reason.

A book for chemists cannot be written without help from chemists, and this book owes much to the chemistry students who have suffered from my lectures but gave me a better understanding of their difficulties, and the many colleagues in chemistry departments who have refused to accept the simplest mathematical answer to their problems but demanded an account intelligible to them.

More especially I wish to thank Drs MANSELL DAVIES, BIRCUMSHAW and ORVILLE THOMAS who have read all or part of the manuscript. Their criticisms and suggestions have led to many improvements in the exposition and choice of examples.

My thanks are also due to the publishers for their help and advice in the preparation of the diagrams and manuscript.

G. J. K.

University College of Wales
Aberystwyth, Wales

February 1954

CONTENTS

		PAGE
	PREFACE	v
1.	INTRODUCTION	1
2.	APPROXIMATE SOLUTION OF EQUATIONS	14
3.	SEQUENCES AND LIMITS	25
4.	DIFFERENTIATION	35
5.	APPLICATIONS OF DIFFERENTIATION	51
6.	RATIONAL FUNCTIONS	63
7.	CIRCULAR FUNCTIONS	79
8.	EXPONENTIAL, LOGARITHMIC AND HYPERBOLIC FUNCTIONS	90
9.	PROPERTIES OF CURVES	105
10.	INTEGRATION	124
11.	METHODS OF INTEGRATION	144
12.	INTEGRALS	162
13.	SERIES	178
14.	COMPLEX NUMBERS	210
15.	DIFFERENTIAL EQUATIONS	223
16.	FUNCTIONS OF SEVERAL VARIABLES	254
17.	SURFACE AND VOLUME DISTRIBUTIONS	278
18.	DETERMINANTS	301
19.	VECTORS	318
20.	PARTIAL DIFFERENTIAL EQUATIONS	337
	INDEX	351

CHAPTER 1

INTRODUCTION

Functions—Dependent and independent variables—Notation—Constants—Errors—Significant figures—Errors in variables—Arithmetical approximations—Approximate formulae—Graphs—Use of log *and* log-log *graphs—Terminology for graphs—Polar diagrams—Equations of curves—Curve sketching—Odd and even functions*

1.01. FUNCTIONS. If one quantity A is determined by another quantity B, then A is said to be a function of B, or symbolically, that $A = f(B)$*. It is sufficient to give a table of values of A and B, but tables can be condensed by the use of formulae. When the relation between A and B is given by a formula then the form of the function $f(B)$ is known.

If B has a discrete set of values it is called a discrete variable, and if it has a continuous range of values, it is a continuous variable.

Example 1. Table I is a list of boiling points of normal fatty hydrocarbons, C_nH_{2n+2} for various values of n. Here A is the boiling point (b. pt) and B the value of n. Each hydrocarbon is obtained in a pure form so that the points only correspond to integral values of n.

Table I

n	5	6	8	10	11	12
B. pt (°K)	309	342	399	447	469	489

The relation between the b. pt and n, within the accuracy of *Table I*, is

$$\text{B. pt} = 745{\cdot}4 \log (n + 4{\cdot}4) - 416{\cdot}3.$$

This can be used to predict values of the b. pt for other values of n either between those given (interpolation) or outside the range of values in *Table I* (extrapolation), assuming that the formula is generally true. Care must be taken. Sometimes the curves for even and odd values of n are different.

Formulae of the type given by this example are very useful in

* $f(B)$ is read 'function of B'. The use of parentheses in a functional relation of this sort is not to be confused with their use in multiplication, e.g. $a(b + c)$. Experience and context help to distinguish the two uses.

organic chemistry to summarize vast tables. In physical chemistry it is more common for the variable B to be continuous. Since experimental data are of necessity only obtained for a limited choice of values of B, except in very few cases, it is necessary to assume regularity to deduce a formula.

Example 2. Table II gives the volume v of a given mass of gas at constant temperature, for various pressures. The formula

$$(p + a/v^2)(v - b) = R(\theta + 272{\cdot}5),$$

due to van der Waals, fits this and other data with $a = 0{\cdot}00874$, $b = 0{\cdot}0023$, $R = 0{\cdot}00369$ and $\theta = 32$. The constants a and b depend on the gas, R only on its mass, and θ is the temperature in °C. In this formula v is measured as a fraction of its value at standard temperature and pressure (s.t.p.).

Table II

v	0·1	0·05	0·025	0·020	0·010	0·0075	0·005
p	10·6	23·4	34·6	41·9	58·7	62·0	67·2

1.02. Dependent and Independent Variables. The independent variable is the quantity B adjusted experimentally. The quantity A then measured is the dependent variable. From the mathematical point of view they are interchangeable: if A is a function of B, then B is a function of A. In formulae also their roles can be interchanged.

Example 1. If $y = ax + b$ gives y as a function of x, then $x = (y - b)/a$ gives x as a function of y.

1.03. Notation. The equation $y = f(x)$ always means that there is a formula for calculating y from x. Thus $f(x) = x^2 + 3x + 4$ means that four is added to three times the value of x and then to the square of x. This same procedure is carried out whatever symbol takes the place of x inside the parentheses.

Example 1. If $f(x) = x^2 + 3x + 4$, then $f(a) = a^2 + 3a + 4$,

or $f(2) = 2^2 + 3.2 + 4 = 14$,

or $f(0) = 0 + 0 + 4 = 4$,

or $f(x + 1) = (x + 1)^2 + 3(x + 1) + 4 = x^2 + 5x + 8$,

or $f(x^2) = x^4 + 3x^2 + 4$,

or $f(xy) = x^2y^2 + 3xy + 4$.

Example 2. If $f(x) = \sin x$, then $f(90°) = \sin\ (90°) = 1$.

Thus in one sense the form of the function is more important than the symbol occurring in it. It is common to speak of quadratic functions, exponential functions, and so on.

The notations $f(x)$, $g(x)$, $F(x)$ are all used to denote 'some function of x.'

1.04. CONSTANTS. A variable can be given a number of values, but a constant has only one value in a given problem, i.e. a numerical value. This remains true even when symbols are used to denote constants. Sooner or later the correct value has to be inserted for the symbol.

Experimentally, when A is found to be a function of B, the apparatus is controlled so that other quantities are kept constant in value. If these quantities are C, D . . . then it is clearly implied that A is also a function of these and that circumstances may arise sooner or later where C, D . . . are variables (see Chapter 16). Thus a symbol may denote a quantity which is constant in one problem and variable in another. The mathematician, who is not concerned with what the symbols represent, can avoid this difficulty by always choosing symbols such as x, y, u, v . . . at the end of the alphabet (any alphabet!) as variables, and a, b, c . . . at the beginning as constants. This is not possible in applied mathematics, where it is customary to use a special symbol for each quantity. The result is that a symbol such as c is naturally chosen for concentration, even though it is in universal use to denote a constant. A golden rule with a formula, then, is first to decide which symbols denote constants and which symbols denote variables.

Finally, an arbitrary constant means a number or constant the value of which is not yet known, but must be determined before a problem is completely and uniquely solved. Arbitrary constants are usually denoted by capitals A, B, C . . . but reasons for not doing this are given in section 10.11.

Example 1. The equation $pv = RT$, for a perfect gas, is a relation between the pressure p and volume v, when R and T are constants.

Example 2. In $y = ax + b$, a and b are unspecified, but known, constants, i.e. it may be that $y = 2x + 3$, or $y = x + 1$, or even $y = 2x$ (when $b = 0$).

If, however, we write $y = ax + B$, then we imply that a is known but B has still to be found.

1.05. ERRORS. In the calculation of y using a formula $y = f(x)$ and a derived value of x, errors can appear in the result on account of:

(*i*) mistakes in arithmetic,
(*ii*) the use of an approximate value of x,
(*iii*) approximations due to the method of calculation.

The important thing is that the error should be within the desired limits.

Mistakes in arithmetic are often difficult to discover even when checks can be made using another method. They are avoided only by systematic and careful work. This cannot be overemphasized, since the time spent on a fresh calculation is just that time wasted. It does not mean that checks are not valuable to correct the occasional mistake. Two types of checks are advisable:

(*i*) a rough answer to prevent 'coarse errors,' and
(*ii*) checks at important stages of the work.

The absolute error is defined as the correction to be added to the approximate answer y_0 to give the correct answer y, or

$$E = y - y_0.$$

The relative error is $(y - y_0)/y_0$.

The percentage error is $100 \,.\, (y - y_0)/y_0$.

Note that y_0 and not y is used in the denominator.

'To an accuracy of 0·05' means that the error E lies between $-$ 0·05 and $+$ 0·05, i.e. $y = y_0 \pm 0{\cdot}05$.

'To an accuracy of 5 per cent,' or $y = y_0(1 \pm 5/100)$, means that the percentage error lies between $-$ 5 and $+$ 5.

'An accuracy of one part in 75,' i.e. $y = y_0(1 \pm 1/75)$ means that the relative error lies between $-$ 1/75 and $+$ 1/75.

Unless otherwise stated 'error' always means absolute error.

1.06. SIGNIFICANT FIGURES. Tables of values such as sines or logarithms give only approximate values. Their accuracy is usually assessed by the number of significant figures. A table giving values to, say, six significant figures gives the six-figure number nearest to the correct answer, not counting any initial zeros.

Example 1. 0·5738147 is 0·5738 to four significant figures and 0·574 to three significant figures.

Example 2. 0·008726 is 0·00873 to three significant figures.

Thus a table giving $\sin 26° = 0{\cdot}4384$ does not tell us more than that

$$0{\cdot}43835 \leqslant \sin 26° \leqslant 0{\cdot}43845.$$

The error (or correction) in using a table correct to four significant figures is $\pm$ 5 in the fifth significant position.

In reducing the number of significant figures there is sometimes an ambiguity when the last figure is a 5 which is resolved as in the next example.

Example 3. 0·00785 becomes 0·0079 on reduction to two significant figures: 115 becomes 120.

Another point to be noticed is the distinction between the number of decimal places and the number of significant figures. If two numbers are given to the same number of decimal places their absolute errors are about the same, whereas two numbers to the same number of significant figures have about the same relative error.

1.07. ERRORS IN VARIABLES. An error in a variable or function of a variable carries through a calculation and combines with other errors. The rules for combining errors in algebraic processes are simple:

(*i*) in addition or subtraction $(a \pm b)$, the possible absolute errors are added,

(*ii*) in multiplication (ab) and division (a/b) the possible relative errors are added, when small*.

These results follow from the table below, where a has an approximate value a_0, an absolute error as large as $\pm\ \varepsilon_a$ or relative error $r = (\varepsilon_a/a_0)$; and b has an approximate value b_0, an absolute error as large as $\pm\ \varepsilon_b$ and relative error $s = \varepsilon_b/b_0$.

$a + b$	can	$(a_0 - \varepsilon_a) + (b_0 - \varepsilon_b)$	or	$(a_0 + \varepsilon_a) + (b_0 + \varepsilon_b)$
$a - b$	be	$(a_0 - \varepsilon_a) - (b_0 + \varepsilon_b)$	as	$(a_0 + \varepsilon_a) - (b_0 - \varepsilon_b)$
ab	as	$a_0(1 - r)b_0(1 - s)$	large	$a_0(1 + r)b_0(1 + s)$
$\frac{a}{b}$	small as	$\frac{a_0}{b_0} \cdot \frac{1 - r}{1 + s}$	as	$\frac{a_0}{b_0} \cdot \frac{1 + r}{1 - s}$

Hence in the sum or difference the absolute error can be as large as $\pm\ (\varepsilon_a + \varepsilon_b)$.

In the product ab the relative error lies between

$$-(r + s) + rs \quad \text{and} \quad +(r + s) + rs$$

or $\pm\ (r + s)$ when r and s are small.

* The true errors do not combine in this way.

In the quotient the relative error lies between

$$-\frac{r+s}{1+s} \quad \text{and} \quad \frac{r+s}{1-s}$$

or $\pm (r+s)$ when s is small.

Example 1. If $y = 64 \sin \theta$, where $\theta = 29°$, the relative error in y is that in sin 29°. To four significant figures sin 29° = 0·4848, with an error of 0·00005 and a relative error of 0·00005/0·48 or 1 in 10^4.

Example 2. If $y = 64 \sin \theta$, the relative error in using tables with four-figure accuracy is 0·00005/sin θ, i.e. greater than 1 in 10^4 if θ is less than 30°, and less than this if θ is greater than 30°.

When several algebraic operations are involved, the absolute errors can increase very rapidly. The relative errors do not increase as rapidly unless subtractions are involved.

Example 3. Consider $y = ax^2 + bx + c$ where $a = 1$, $b = 2$, $c = 1$ and $x = 3$ with possible small percentage errors of α, β, γ and P respectively.

The assumed values of ax^2, bx and c are 9, 6, 1 and $y = 16$.

Using the rules given above, the percentage errors in these terms are $(2P + \alpha)$, $(P + \alpha\beta)$ and γ respectively. The absolute errors ($\times$ 100) are $9(2P + \alpha)$, $6(P + \beta)$, γ. Hence the possible absolute error in y is

$$(24P + 9\alpha + 6\beta + \gamma)/100.$$

If $\alpha = \beta = \gamma = P = 2$ per cent this is 0·8, so that $y = 16 \pm 0{\cdot}8$ with a relative error of 5 per cent, the bulk of which comes from the biggest term.

Example 4. However, when $y = ax^2 - bx + c$, using the same values of the constants and of x, the relative error is considerably larger. The absolute error is still 0·8 but, since $y = 9 - 6 + 1 = 4$, the relative error is now 20 per cent.

More general methods of calculating errors appear in Sections 4.09, 5.07 and 16.05.

1.08. ARITHMETICAL APPROXIMATIONS. If a and b are known only to an accuracy of 1 per cent and $y = ab$, then y is known only to within 2 per cent. Hence an arithmetical process of multiplying a by b need not retain all the figures, and the process of contracted multiplication is used. The degree of contraction should not be such as to affect the error in the answer appreciably, and this often means that it is necessary to work to four significant figures. By proper study, the errors implicit in a problem can be used to simplify the calculations.

1.09. APPROXIMATE FORMULAE. An alternative approach to the problem of calculation is to replace an exact formula by an approximate formula giving the same answer to the desired accuracy. Finding such formulae is an important part of applied mathematics, but it is relatively advanced and is considered in later chapters.

Example 1. The equation $y = \sin x$ can be replaced by

$$y = x - x^3/6 + x^5/120$$

with an accuracy of 0·5 per cent when $x < \frac{1}{2}\pi$ radians (90°). This is very useful when the book of sine tables is lost.

1.10. GRAPHS. These not only give a concise picture of the relation between two variables but they can be used to read values of the variables, when many are required to no very great accuracy. Thus, when $y = x/(1 + x^2)$ is needed for a large number of values of x, it is easy to calculate it for simple values of x, plot the results on a graph and read from the graph the values of y for other values of x. A well drawn curve on a sheet of graph paper 10″ by 8″ should give an answer accurate to almost 0·1 per cent. When greater accuracy is required the graphical method is not possible. Moreover, to achieve this accuracy certain simple rules must be applied. These are:

(*i*) The scales should be chosen so that the curve covers the paper available.

(*ii*) The scales should be convenient so that mistakes are not made in plotting or reading values.

(*iii*) Accurate plotting is essential.

(*iv*) Plotting known values with a cross of two fine lines involves some error, which determines the accuracy of the curve drawn through the crosses. Hence the calculated values which are plotted need be no more accurate than they can be plotted, nor can the readings from the graph be more accurate than this*.

(*v*) A graph sometimes shows up simple mistakes in calculations; then it appears to contain kinks when it 'ought to be' smooth.

1.11. USE OF SINGLE AND DOUBLE LOGARITHMIC SCALES. When one or both variables cover a wide range of values it is often advisable to plot the logarithms of their values, e.g. $\log y$ against x, or $\log y$ against $\log x$. The resulting graphs have different shapes from the elementary graph of y against x, but even this is usually an advantage,

* We are not discussing graphs of experimental points subject to experimental errors, but graphs summarizing calculations.

the shape often being simpler. In the examples given below it is shown when a straight line is obtained in these modified graphs, but the reader should always be certain that the lines are straight before using the graph to determine an approximation to the function. For certain ranges of values special graph paper can be obtained.

It should be realized that equal absolute errors in $\log y$ in different parts of the graph correspond to equal relative errors in y, so that it is an advantage to use this form of graph when it is desired to maintain a given relative accuracy.

Example 1. If $y = a \times 10^{mx}$, then $\log_{10} y = \log_{10} a + mx$. A straight line of slope m is obtained when $\log_{10} y$ is plotted against x.

Example 2. If $y = ax^n$, then $\log_{10} y = \log_{10} a + n \log_{10} x$. A straight line is obtained when $\log_{10} y$ is plotted against $\log_{10} x$, the slope being n.

Example 3. If the relative error in measuring y is r then the true value is $y = y_0(1 \pm r)$, where y_0 is the approximate value. Hence

$$\log_{10} y = \log_{10} y_0 + \log_{10} (1 \pm r).$$

There is an absolute error in $\log_{10} y$ determined by r. It is shown later that $\log_{10} (1 \pm r) = \pm 0{\cdot}42r$ (approximately) when r is small.

1.12. TERMINOLOGY FOR GRAPHS. At this point a summary is given of some terms commonly used to describe graphs and functions. It

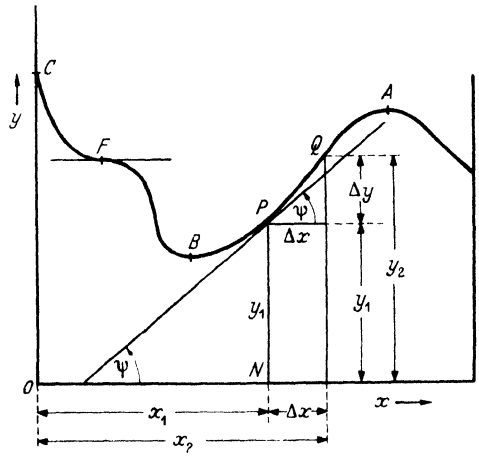

Figure 1.01

is assumed that y is plotted vertically against x horizontally as in *Figure 1.01.*

Abscissa, e.g. *ON*, measures a value of x.

Ordinate, e.g. *NP*, measures a value of y.

As x changes from x_1 to x_2, y changes from y_1 to y_2. The suffix 1 denotes the initial value, and suffix 2 the final value.

$\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$ (final less initial value) are the changes in x and y. The symbol Δx is read 'delta x' and not delta times x, just as $f(x)$ means 'function of x' and not f times x.

$|x|$ or $|\Delta x|$, read as 'mod x' or 'mod Δx,' means the numerical value of x or Δx as the case may be, ignoring its sign, so that a 'mod' or modulus is always positive.

The *slope* of a tangent or chord ($m = \tan \psi$) is determined by the angle ψ that it makes with the positive x axis. Both m and ψ are positive when y increases with x along the chord or tangent, and negative when y decreases as x increases.

The variable y has a greatest value (C) and a least value (B) in any limited range of values of x. If x is unlimited this may not be so.

A *maximum* (A) of y occurs where it exceeds the value at neighbouring points on the curve.

A *minimum* value (B) occurs where y is less than at neighbouring points.

On some curves y has several maxima and minima. If the curve is continuous they must occur alternately. The greatest and least values may not be maxima or minima, respectively.

At a maximum or minimum $\psi = 0$ and $m = 0$ for curves with a continuously turning tangent.

$y = |x|$ has a minimum at $x = 0$, but m is not defined.

The equation $m = 0$ may be true at other points on the graph such as F.

A graph has positive *curvature* where it is concave upwards, and negative curvature where it is convex upwards. For positive curvature m increases with x, and for negative curvature m decreases as x increases.

A point of inflection occurs where the curvature changes sign.

1.13. POLAR DIAGRAMS. A relation $r = f(\theta)$, where θ is an angle specifying direction, is conveniently represented on a polar diagram.

On this diagram the graph is such that a line OP drawn at any angle θ with a given line OA, intersects the curve at a point P, and the length of OP represents the value of r calculated from the formula $r = f(\theta)$. A variation of r with θ now appears as a variation of the distance from the origin of the point P as the angle θ changes (cf. *Figure 9.05*, p. 110).

In this diagram (r, θ) are called polar coordinates, and suitable graph paper can be obtained on which to draw it.

1.14. EQUATIONS OF CURVES. Given any plane curve, it is possible to choose an origin and Cartesian axes, and, by using the properties of the curve, to deduce the equation of the curve, i.e. the relation between the coordinates x and y of points on the curve.

It is also possible to use polar coordinates to describe the same curve, or indeed one of a number of other ways, each having its special advantages for certain purposes.

1.15. CURVE SKETCHING. A theoretical approach to a chemical or physical problem leads to a formula, e.g. $y = f(x)$. This formula may, and generally does, contain arbitrary constants. A comparison with experiment is made in at least two stages, (*i*) to find if qualitative agreement is possible, and (*ii*) to adjust the arbitrary constants to give a good fit.

If this comparison is made using graphs, then in the first stage, a graph of the theoretical result is needed which shows certain critical features, but need not be correct numerically. Such graphs are called sketches.

The crucial test of a theory often lies in extreme conditions. Hence the important features of a graph are:

(*i*) behaviour near the origin (when x and y are small);
(*ii*) changes of sign of the dependent variable;
(*iii*) behaviour for large values of x;
(*iv*) behaviour for large values of y;

By behaviour we mean the answer to such questions as: 'Is the variable almost constant, corresponding perhaps to saturation, or does it increase or decrease?' If it decreases numerically, does it decrease as $(1/x)$ or $(1/x^2)$ or e^{-x}, as x increases?

(*v*) the greatest and least values, or maxima or minima if there are any;
(*vi*) more rarely, the curvature and change of sign of curvature are significant.

Example 1. It is of great importance to know whether the vapour pressure of liquid helium varies as the first or third power of the absolute temperature at very low temperatures.

Example 2. The force between two atoms or particles is quite different in form for charged and uncharged pairs when they are widely separated.

Example 3. An exponential law of decay is an inevitable result of some theories.

Example 4. When a substance is being made by one process and lost by another, then its amount is steady (in many problems a maximum) when the two rates are equal. Then the conditions for a maximum can be compared with experimental conditions.

In sketching curves, therefore, the particulars (i)–(v) are first obtained, and any others of importance, and the results inserted roughly on a graph. A smooth curve is roughly sketched to fit these features, without worrying too much about numerical values, and this is usually a good enough approximation to the true graph for the purposes required.

1.16. EVEN AND ODD FUNCTIONS. A function $f(x)$ is an even function of x if its value is unchanged when the value of x changes sign.

Even: $$f(-x) = f(x).$$

It is an odd function of x if it only changes sign when the value of x changes sign.

Odd: $$f(-x) = -f(x).$$

The product or ratio of two even or two odd functions is even: the product or ratio of an odd and an even function is odd.

Example 1. Even functions: x^2, $1/(x^2+1)$, e^{-x^2}, $\cos x$, $x \sin x$.

Example 2. Odd functions: x, $\sin x$, $x \cos x$, $x/(x^2+1)$.

Any function $f(x)$ can be expressed as the sum of two functions, one of them even and the other odd. For it is the sum of

$$f_e(x) = \tfrac{1}{2}\{f(x) + f(-x)\}, \text{ all } x$$

and $$f_o(x) = \tfrac{1}{2}\{f(x) - f(-x)\}, \text{ all } x.$$

$f_e(x)$ is an even function of x since

$$f_e(-x) = \tfrac{1}{2}\{f(-x) + f(x)\} = f_e(x).$$

$f_o(x)$ is an odd function of x since

$$f_o(-x) = \tfrac{1}{2}\{f(-x) - f(x)\} = -f_o(x).$$

Example 3. $f(x) = \sqrt{\dfrac{1+x}{1-x}}$. Then $f(-x) = \sqrt{\dfrac{1-x}{1+x}}$.

$$f_e(x) = \frac{1}{2}\left\{\sqrt{\frac{1+x}{1-x}} + \sqrt{\frac{1-x}{1+x}}\right\} = \frac{1}{2}\left\{\frac{(1+x)+(1-x)}{\sqrt{(1-x^2)}}\right\} = \frac{1}{\sqrt{(1-x^2)}},$$

$$f_o(x) = \frac{1}{2}\left\{\sqrt{\frac{1+x}{1-x}} - \sqrt{\frac{1-x}{1+x}}\right\} = \frac{1}{2}\left\{\frac{(1+x)-(1-x)}{\sqrt{(1-x^2)}}\right\} = \frac{x}{\sqrt{(1-x^2)}}.$$

EXAMPLES I

1. Fit the data of *Table I* as accurately as possible with a linear relation b. pt $= an + b$, and compare its accuracy with the formula given in the text by plotting the relative errors against n for the two approximations.

2. The heat of crystallization (Q) of saturated monocarboxyl acids is a function of the number (n) of carbon atoms:

n	:	6	7	8	9	10	11	12,
Q	:	3·60	3·58	5·11	4·85	6·69	5·99	8·75.

Extend this table if you can, and show that (approximately)

$$n \text{ even}: \quad Q = 1{\cdot}030n - 3{\cdot}61,$$
$$n \text{ odd}: \quad Q = 0{\cdot}965n - 4{\cdot}49.$$

Does the error justify the use of three figures in these approximate formulae?

Find the melting point (m. pt) (T°A) of the substances, and show that they oscillate in the same way as Q, but the ratio (Q/T) (entropy change) is a 'continuous function' of n.

3. A bar of iron 100 cm long is heated through 100C° and its increase in length is measured. The coefficient of linear expansion is

$$\alpha = 10{\cdot}6 \times 10^{-6}/\text{C}°.$$

Find the percentage error in α if an error of 1 mm is made in the increase in length, and the absolute accuracy of this measurement if α is to be correct to 1 per cent.

4. To find the equivalent of sodium, x parts of sodium chloride are precipitated with 1 part of silver and weighed as silver chloride. Show that the relative error in the equivalent is a factor K greater than that in x, where

$$K = \frac{(\text{equiv. of Na}) + (\text{equiv. of Cl})}{(\text{equiv. of Na})} \doteqdot 2{\cdot}5.$$

5. If x g of a metallic carbonate reacts with hydrochloric acid and produces y cm³ of carbon dioxide at s.t.p., the equivalent of the metal is

$$E = 11{\cdot}2x/y - 30.$$

Find the absolute and relative errors in the calculated values of E if there is a relative error of 1 per cent in y, when (*i*) $E = 20$ and (*ii*) $E = 44$.

6. What is meant by the assertions (*i*) the volume of a gas at constant pressure is a linear function of temperature, and (*ii*) the resistance R of a platinum wire is a quadratic function of the absolute temperature T?

7. Sketch $y = x/(1 + x^2)$ for positive values of x. Describe how it differs from $y = 2x^2/(1 + x^3)$ (*i*) for very small values of x, and (*ii*) for large values of x.

8. A square bar of specific gravity s, floating lengthwise in water with two sides vertical, is in stable equilibrium if $(1/6s) > 1 - s$. By sketching the two curves $y = (1 - s)$ and $y = (1/6s)$ illustrate the range of values of s

satisfying this condition. What is the effect of changing the liquid from water to a liquid with specific gravity less than unity?

9. The negative energy of an ion of polarizability α_a, in the field due to a molecule of dipole moment μ_b directed at an angle θ with the line joining their centres, is $E = \frac{1}{2}\mu_b{}^2\alpha_a(1 + \cos^2\theta)/r^6$, where r is the distance between them. Draw a polar diagram of E as a function of θ.

10. Draw polar diagrams of the Legendre polynomials $r = P_n(\cos\theta)$ and their squares $r = P_n{}^2$, when (*i*) $P_0 = 1$, (*ii*) $P_1 = \cos\theta$, (*iii*) $P_2 = \frac{1}{2}(3\cos^2\theta - 1)$, and (*iv*) $P_3 = \frac{1}{2}(5\cos^3\theta - 3\cos\theta)$.

REFERENCES

The author has found helpful treatments of the subject matter of this book in the following:

CLEMENTS, G. R. and WILSON, L. T. *Manual of Mathematics and Mechanics.* McGraw-Hill Book Co., New York, 1937.

DANIELS, F. *Mathematical Preparation for Physical Chemistry.* McGraw-Hill Book Co., New York, 1935.

GEARY, A., LOWRY, H. V. and HAYDEN, H. A. *Advanced Mathematics for Technical Students.* Volume I. Longmans, Green and Co., London, 1945.

LOWRY, H. V. and HAYDEN, H. A. *Advanced Mathematics for Technical Students.* Volume II. Longmans, Green and Co., London, 1950.

GRANVILLE, W. A., SMITH, P. F. and LONGLEY, W. R. *Introduction to the Differential and Integral Calculus.* G. Ginn and Co., London and New York, 1941.

LAMB, H. *Infinitesimal Calculus.* Cambridge University Press, London, 1951.

MELLOR, J. W. *Higher Mathematics for Students of Physics and Chemistry.* Longmans, Green and Co., London, 1912.

PARTINGTON, J. R. *An Advanced Treatise on Physical Chemistry.* Longmans, Green and Co., London, 1951.

PIPES, L. A. *Applied Mathematics for Engineers and Physicists.* McGraw-Hill Book Co., New York, 1946.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering.* McGraw-Hill Book Co., New York, 1939.

CHAPTER 2

APPROXIMATE SOLUTION OF EQUATIONS

Introduction—Use of graphs—Chord method—Newton's tangent method—Chord and tangent method—Iterative methods—Chord and tangent iterative method—Newton's tangent iterative method—General remarks—Comparison of methods—Decimal method—Square roots—General iterative method—Small roots and large roots

2.01. INTRODUCTION. The problem of solving an equation is a very common one and some methods of doing so are given in this chapter. Just two illustrations must suffice of the ways in which solutions are obtained. First, from an equation $y = f(x)$ it may be necessary to find a value of x where y is zero, or a maximum, or satisfies some other condition. Secondly, it may be necessary to solve an equation to obtain the values of the dependent variable.

To solve an equation of the general form $f(x) = 0$, it is usual to sketch $f(x)$ and calculate it for simple values of x near a likely root of the equation $f(x) = 0$. It may happen that one of these values is a root, but the primary purpose is to use these values to calculate the root approximately, using the rule: if $f(x)$ is continuous and has values of opposite sign for two values of x, then $f(x) = 0$ for at least one intermediate value of x.

2.02. USE OF GRAPHS. A graph of $y = f(x)$ cuts the x axis at values of x where $f(x) = 0$.

Example 1. The following table gives some values of $y = \cos x - x$ (x in radians) obtained from elementary tables:

x	0·5	0·6	0·7	0·8	0·9	
y	0·3776	0·2250	0·0648	− 0·10303	− 0·2784	± 0·00015

A graph drawn using this table gives $x = 0{\cdot}739$ as the root of $\cos x = x$. (The reader should check or improve this, by drawing the graph. The author used centimetre paper.)

The roots of an equation which is written in the form $f_1(x) = f_2(x)$ are the values of x at the points of intersection of the graphs $y = f_1(x)$ and $y = f_2(x)$.

Example 2. The equation $x^2 - 10x + 1 = 0$ can be written $10 - x = (1/x)$ and its roots are the values of x at the intersection of the line $x + y = 10$ and the curve $y = (1/x)$. Both of these are easily drawn (draw them!) and the roots are approximately $x = 0{\cdot}1$ and $x = 9{\cdot}9$.

In each example the roots are found by the intersection of two curves, or a line and a curve. The greater the angle between the curves on the graphs, the more accurately the roots are located. If necessary the scale can be adjusted to achieve this.

This method is especially useful with equations involving an adjustable parameter.

Example 3. A rod of length $2l$ resting with one end on a plane of inclination α and the other passing through a smooth ring, makes an angle ϕ with the horizontal given by

$$l \cos \phi \sin^2 (\phi + \alpha) = a,$$

where a is the distance of the ring from the plane.

The roots ϕ of this equation can be estimated by drawing $y = (l/a) \cos \phi$ and $y = \operatorname{cosec}^2 (\phi + \alpha)$ against ϕ, for various values of the ratio $k = (l/a)$ (*Figure 2.01*). It is at once clear that these

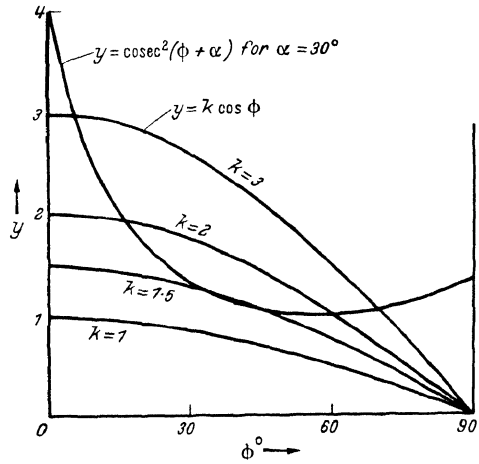

Figure 2.01

curves do not intersect if k is too small, that they touch if k reaches a certain value, and intersect at two points if k is not much larger than this. For even bigger values of k there is only one root. It is assumed that ϕ lies between 0° and 90°.

Example 4. Consider the positive roots of the equation $\tan x = mx$ (*Figure 2.02*) for various values of m, by plotting the curves $y = \tan x$ and $y = mx$.

When $m = 0$, the roots are $x = 0, \pi, 2\pi, 3\pi. \ . \ .$

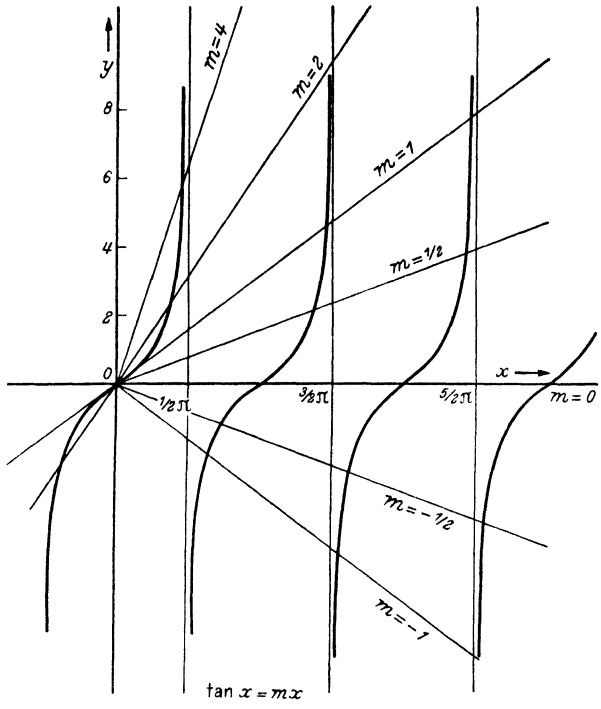

Figure 2.02

Now consider the lowest root $x = \pi$ as m decreases and becomes negative. This root decreases steadily until, when m is very large and negative, it has descended to little more than $x = \frac{1}{2}\pi$.

Now switch m to a large positive value. This root is now a little less then $\frac{1}{2}\pi$ and still decreases with m until it is zero when $m = 1$, since the slope of $y = \tan x$ is unity at $x = 0$. The role of lowest non-zero root is now taken over by the next root which lies between π and $\frac{3}{2}\pi$, and as m decreases from 1 to 0 this root decreases to π.

For the same cycle of values of m, every root follows the same cycle. It should be noticed, however, that the root which is 2π (or any larger root) when $m = 0$ very quickly descends to $\frac{3}{2}\pi$ as m becomes negative, remains approximately $\frac{3}{2}\pi$ when m switches to

positive values, and only decreases still further to π when m is quite small, i.e. for all except very small values of m, the larger roots are approximately $(n + \frac{1}{2})\pi$.

2.03. CHORD METHOD. If A and B are two points on a graph $y = f(x)$ (*Figure 2.03*) on opposite sides of the x axis, the chord AB intersects the axis at a point M, not far from the point P where the curve cuts it. Hence the value $x = x_C$ at M is an approximation to a root of $f(x) = 0$.

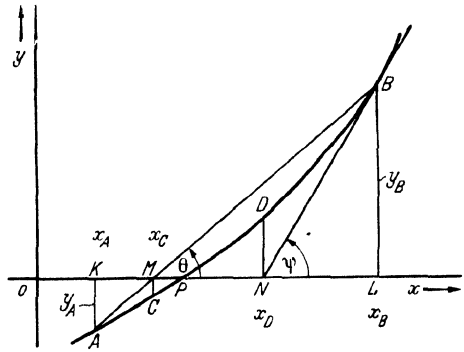

Figure 2.03

An algebraic formula for x_C can be used to calculate it without reference to the graph. From the diagram,

$$\cot\theta = \frac{KM}{AK} = \frac{x_C - x_A}{(-y_A)},$$

and

$$\cot\theta = \frac{KL}{AK + LB} = \frac{x_B - x_A}{y_B - y_A}.$$

Equating these two

$$x_{approx.} = x_C = x_A - (x_B - x_A) \times \frac{y_A}{y_B - y_A}.$$

Example 1. Using the table in *Example 1* of Section 2.02 with $x_A = 0{\cdot}7$ and $x_B = 0{\cdot}8$,

$$x_C = 0{\cdot}7 + (0{\cdot}1) \times \frac{0{\cdot}0648}{0{\cdot}1678} = 0{\cdot}7386 \pm (0{\cdot}0001\ ?).$$

The graph ought to give a better answer than this since it allows for curvature, but it is not easy to read accurately.

The error $\pm$ 0·0001 refers to the value of x_C: the error between x_C and the root is not known.

2.04. NEWTON'S TANGENT METHOD. The tangent at any point B (*Figure 2.03*, p. 17) cuts the axis at a point N near the root P provided B is not too far away. Thus $x = x_D$ is the tangent approximation to the root.

If $m_B = \tan \psi$ is the slope of the tangent at B, then $ON = OL - NL$, or

$$x_{approx.} = x_D = x_B - \frac{y_B}{m_B}.$$

2.05. CHORD AND TANGENT METHOD. This method depends on the fact that if A and B lie on opposite sides of P and not too far away, then M and N lie on opposite sides of P. If x_D is greater than the root and x_C is less, presumably the mean of the two is a better approximation than either. Hence the suggestion that the root is

$$\begin{aligned} x_{approx.} &= \tfrac{1}{2}(x_C + x_D) \\ &= \frac{1}{2}\left\{x_A - \frac{x_B - x_A}{y_B - y_A} \times y_A + x_B - \frac{y_B}{m_B}\right\}. \end{aligned}$$

2.06. ITERATIVE METHODS. The following methods are designed to give successive approximations of greater and greater accuracy. An iterative method uses each approximation to obtain a better one.

2.07. CHORD AND TANGENT ITERATIVE METHOD. Graphically the chord and tangent method starts with two points A and B on either side of the root at P, and leads to points M and N on the x axis on either side of P (*Figure 2.03*).

Now there are two points C and D on the curve, corresponding to M and N, and they can be used instead of A and B for a second application of the chord-and-tangent method, leading to two more x coordinates of points E and F. Algebraically, the steps are:

(*i*) calculate y_A and y_B using x_A and x_B;

(*ii*) calculate $x_C = x_A - y_A(x_B - x_A)/(y_B - y_A)$ and
$x_D = x_B - (y_B/m_B)$;

(*iii*) calculate y_C and y_D from x_C and x_D;

(*iv*) calculate $x_E = x_C - y_C(x_D - x_C)/(y_D - y_C)$ and
$x_F = x_D - (y_D/m_D)$.

The last calculation (*iv*) is similar to (*ii*).

This process can be repeated any number of times, the error being reduced each time. It need be continued only until the two values of x obtained are equal, to the required accuracy, since the root lies between them.

2.08. NEWTON'S TANGENT ITERATIVE METHOD. This is similar to the previous method but uses tangents only. The procedure is to calculate:

(*i*) y_B from x_B and find m_B, the slope of the tangent;

(*ii*) $x_D = (x - y/m)_B$;

(*iii*) y_D from x_D and find m_D;

(*iv*) $x_F = (x - y/m)_D$, etc.

This is quicker than the chord-and-tangent method but suffers from the defect that the error at any stage is unknown. It is customary to proceed until two successive answers are equal to the required accuracy.

2.09. GENERAL REMARKS. When tangents are used in these approximations, their slopes m are either measured graphically or calculated (Section 5.03). Small errors do not matter for a simple reason. The main purpose of the method is to produce successive pairs of values of x between which the root lies. Their exact position does not matter. Hence small errors in the calculation do not matter. For example, in Newton's tangent method, iterated, it is possible to use the same slope m_B at every stage without attempting to calculate m_D, etc. This is an extreme example, however, and borders on laziness!

This argument must not be carried too far. Unnecessary errors may mean further iterations.

2.10. COMPARISON OF METHODS. We apply these methods to a simple problem, e.g. the calculation of $\sqrt{2}$, i.e. the root of $y = x^2 - 2 = 0$.

The reader should draw a graph of $y = x^2 - 2$ between $x = 0$ and $x = 3$ and convince himself (*i*) that the root lies between $x = 1{\cdot}4$ and $x = 1{\cdot}5$, (*ii*) that the slope of the tangent at $x = 1{\cdot}5$ is very nearly $m = 3$ and does not vary very much between $x = 1{\cdot}4$ and $x = 1{\cdot}5$.

(*i*) Chord method:

Let	$x_A = 1{\cdot}4,$	$x_B = 1{\cdot}5.$
Then	$y_A = -0{\cdot}04,$	$y_B = 0{\cdot}25.$

Hence

$$x_{approx.} = x_C = x_A - y_A \frac{x_B - x_A}{y_B - y_A} = 1{\cdot}4 + 0{\cdot}04 \times \frac{0{\cdot}1}{0{\cdot}29}$$
$$= 1{\cdot}4 + 0{\cdot}0138 = 1{\cdot}414.$$

(*ii*) Tangent method:

$$x_B = 1{\cdot}5, \quad y_B = 0{\cdot}25, \quad m_B = 3,$$
$$x_D = x_B - y_B/m_B = 1{\cdot}5 - 0{\cdot}25/3 = 1{\cdot}417 \text{ approx.}$$

neglecting extra decimal places.

(*iii*) Chord and tangent method:

$$x_{approx.} = \tfrac{1}{2}(x_C + x_D) = 1{\cdot}416 \text{ approx.}$$

Let us now use iterative methods to calculate the root to four decimal places. Notice that we use the slope $m = 3$ for all tangents to simplify the division, even though it is slightly inaccurate. The graph is not accurate enough to improve this value appreciably.

(*iv*) Chord and tangent iterated. We start with the calculated values:

$$x_C = 1{\cdot}414, \quad x_D = 1{\cdot}417.$$

Then $y_C = (1{\cdot}414)^2 - 2 = -0{\cdot}00060.$

$y_D = (1{\cdot}417)^2 - 2 = +0{\cdot}00789.$

Then $$x_E = x_C - y_C \times \frac{x_D - x_C}{y_D - y_C}$$
$$= 1{\cdot}414 + 0{\cdot}00060 \times \frac{0{\cdot}00300}{0{\cdot}00849}$$
$$= 1{\cdot}414 + 0{\cdot}00021 = 1{\cdot}41421.$$
$$x_F = x_D - y_D/m = 1{\cdot}417 - 0{\cdot}00789/3$$
$$= 1{\cdot}417 - 0{\cdot}00263 = 1{\cdot}41437.$$
$$y_E = (1{\cdot}41421)^2 - 2 = -0{\cdot}00001$$
$$y_F = (1{\cdot}41437)^2 - 2 = +0{\cdot}00044.$$

Comparing these two it is clear that the root is very close to x_E so that we can say that

$$\sqrt{2} = 1{\cdot}4142 \text{ to four decimal places.}$$

(*v*) Tangent method iterated. As above we should calculate:

$$x_D = (x - y/m)_B = 1{\cdot}417.$$
$$y_D = 0{\cdot}00789.$$

$$x_F = (x - y/m)_D = 1{\cdot}417 - 0{\cdot}00263 = 1{\cdot}41437.$$
$$y_F = 0{\cdot}00044.$$
$$x_H = (x - y/m)_F = 1{\cdot}41437 - 0{\cdot}00015 = 1{\cdot}41422 \text{ etc.}$$
$$\text{giving } \sqrt{2} = 1{\cdot}4142.$$

2.11. DECIMAL METHOD. This iterative method is practicable with polynomials only and advances step by step in a definite way, increasing the accuracy by a known amount each time. These qualities are shared with any good numerical method.

We simply calculate successive decimal places in the answer, and illustrate the method with the well known equation $x^2 - 2 = 0$.

(*i*) The root lies between 1 and 2.

(*ii*) Let $x = 1 + y_1$. The equation for y_1 is $(1 + y_1)^2 - 2 = 0$
i.e. $$y_1^2 + 2y_1 - 1 = 0$$
which has a root between 0 and 1. (Why?)

(*iii*) Let $z_1 = 10y_1$, so that $(z_1/10)^2 + 2(z_1/10) - 1 = 0$ or $z_1^2 + 20z_1 - 100 = 0$. This equation must have a root between 0 and 10. By trial it is found that it lies between 4 and 5 since the left-hand side (LHS) $= -4$ when $z_1 = 4$ and 25 when $z_1 = 5$.

(*iv*) Put $z_1 = 4 + y_2$. The equation for y_2 is
$$(y_2 + 4)^2 + 20(4 + y_2) - 100 = y_2^2 + 28y_2 - 4 = 0.$$

(*v*) Put $z_2 = 10y_2$, i.e. $z_2^2 + 280z_2 - 400 = 0$.
This equation has a root between 0 and 10, which is nearer 0. Actually it is between 1 and 2.

(*vi*) Put $z_2 = 1 + y_3$, $z_3 = 10y_3$. It is found that z_3 lies between 4 and 5.

This may seem complicated: but the nearest integers to z_1, z_2, z_3 are the figures in successive decimal places in the root. Since these are 1, 4, 1, 4 in order we find
$$\sqrt{2} = 1{\cdot}414 \ldots$$

2.12. SQUARE ROOTS. The usual method of calculating square roots on a calculating machine is little different from the following iterative method. Let x_0 be an approximation to $\sqrt{M}$. If $x_0 < \sqrt{M}$, then $M/x_0 > \sqrt{M}$, and a better approximation is
$$x_1 = \tfrac{1}{2}(x_0 + M/x_0).$$
Other approximations are calculated by the same rule
$$x_2 = \tfrac{1}{2}(x_1 + M/x_1) \text{ etc.}$$

This method is a good one. For if $x = \sqrt{M}$, then

$$\begin{aligned} x - x_1 &= x - \tfrac{1}{2}(x_0 + M/x_0) \\ &= (2xx_0 - x_0^2 - x^2)/2x_0 \quad \text{since } M = x^2 \\ &= -(x - x_0)^2/2x_0. \end{aligned}$$

i.e.
$$\frac{x_1 - x}{x - x_0} = \frac{x - x_0}{2x_0}.$$

Now if $|x - x_0| < 1$ the RHS is less than $1/2\sqrt{M}$, so that the errors go down by a factor not larger than $1/2\sqrt{M}$ at each stage. Actually the gain in accuracy is bigger than this. If x_0 is correct to p decimal places, then $|x - x_0| < 1/10^p$ and

$$|x_1 - x| < 1/10^{2p} \quad \text{if } M > 1,$$

i.e. x_1 is correct to $2p$ decimal places.

2.13. General Iterative Method. The last section is an example of a general method applicable to an equation of the form

$$x = f(x).$$

Successive approximations to a particular root are calculated according to the rule

$$x_{n+1} = f(x_n),$$

starting from some value x_0, e.g. $x_1 = f(x_0)$, $x_2 = f(x_1)$, $x_3 = f(x_2)$. . . .

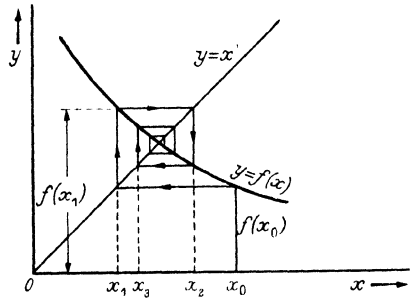

Figure 2.04

A graph of the line $y = x$ and $y = f(x)$ near the root not only gives an initial approximate value x_0 but shows whether the method will yield a result. It is essential that the curve $y = f(x)$ should not rise more steeply than $y = x$, nor fall more steeply than $y = -x$, near the root.

The reader should try to understand the meaning of *Figure 2.04.*

2.14. SMALL ROOTS AND LARGE ROOTS. For small roots an equation is separated into large terms and small terms.

Thus when $x \ll 1$, it follows that $x^2 \ll x$, $x^3 \ll x$, so that some terms can be neglected, if their coefficients are not too large.

Example 1. Find the smaller root of $x^2 - 10x + 1 = 0$.

When x is small, $x^2 \ll x$ and we write approximately

$$-10x + 1 = 0 \quad \text{i.e. } x = 0{\cdot}1.$$

To improve on this we introduce x^2 as a correction:

$$+10x = 1 + x^2$$

i.e.
$$x = 0{\cdot}1(1 + x^2)$$
$$= 0{\cdot}1(1 + (0{\cdot}1)^2) = 0{\cdot}101.$$

and again $x = 0{\cdot}1(1 + x^2) = 0{\cdot}1(1 + (0{\cdot}101)^2) = 0{\cdot}10102$ etc. . . .

This is an iteration

$$x_{n+1} = 0{\cdot}1(1 + x_n^2).$$

Large roots are dealt with in the same way after changing from x to $y = (1/x)$. If x is large, y must be small.

EXAMPLES II

1. Solve the equations $y = 1 - x$ and $y = 1/6x$ by the iteration formula $x_{n+1} = 1 - 1/6x_n$ using $x_0 = 1$ as the initial approximation. Compare the results of four iterations with the graphical method (Chapter 1, Example *8*) and the exact solution. Can the other root be obtained by the same iteration method?

2. A unit sphere is divided into three equal parts by two planes distant x from the centre, if $3x^2 - 9x + 2 = 0$. By a graphical approximation and iteration ($x_{n+1} = \frac{2}{9} + \frac{1}{3}x_n^2$) show that $x = 0{\cdot}2261$ to four decimal places.

3. A donkey is tethered to a peg at the edge of a circular field of unit radius. The length of tether x is such that the donkey can graze on one half of the field. Show that $x = 1{\cdot}159$.

4. The angle x between two chords AB and BC dividing a circle into three equal parts satisfies the equation $y = x + \sin x - \frac{1}{3}\pi = 0$. Draw a graph of y accurately between 25° and 40° (N.B. work in radians), and use Newton's method (if necessary twice) to obtain $x = 0{\cdot}536$ radian, correct to three decimal places.

5. Use the decimal method to solve $x^3 - 2 = 0$ correct to three decimal places.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

HARTREE, D. R. *Numerical Analysis.* Clarendon Press, Oxford, 1952.

JENNINGS, J. *Mathematical Solution of Engineering Problems.* E. and F. N. Spon, Ltd, London, 1951.

LEVY, H. and BAGGOTT, E. *Numerical Studies in Differential Equations.* Watts and Co., London, 1934.

MILNE, W. E. *Numerical Calculus.* Princeton University Press, Princeton, N.J., 1948.

CHAPTER 3

SEQUENCES AND LIMITS

Sequences—Infinity—Limits and convergence—Sequences of ratios—Orders of magnitude—Convergence tests—Washing precipitates—Comments—Functions of a continuous variable—Limits of functions. Continuity—Indeterminate values—Functions of sequences—Three important limits

3.01. SEQUENCES. Successive approximations s_1, s_2. . . . can be written in the form of a table:

n:	1	2	3	4	5	. . .
S:	s_1	s_2	s_3	s_4	s_5	. . .

where they appear as the values of a function S of a discrete variable. A graph can be drawn to represent this function, or a formula obtained of the form $S = f(n)$.

The numbers s_1, s_2, s_3 . . . form a sequence and the expression $f(n)$ is called the nth term of the sequence. Its values for $n = 1, 2, 3$ . . . give successive terms.

In this chapter we discuss the terms in the sequence S as n gets larger and larger, a process which is called 'allowing n to tend to infinity' or more briefly 'as $n \to \infty$.' The following examples illustrate how this problem can arise.

Example 1. Given a set of numbers s_1, s_2 . . . obtained by an iterative method of solving an equation, how can it be shown that they are successive approximations to the correct answer, and what is this answer?

Example 2. A series of melting points M is known for the hydrocarbons C_nH_{2n}. What happens to M for very heavy hydrocarbons (n large)?

Example 3. The energy of a vibrating molecule has a sequence of values E_0, E_1, E_2, . . . arranged in order of increasing magnitude. E_0 is the lowest energy of the zero-point vibration and E_n is the energy of the nth vibrational level ($n = 1, 2, 3$ . . .). The value of $E_n - E_0$ as $n \to \infty$ is the dissociation energy of the molecule. What is this value?

3.02. INFINITY. A very brief discussion of this concept is necessary. Here we use infinity (∞) only in the sense of the previous section.

A quantity 'tends to ∞' or 'becomes infinite' when it assumes a succession of values getting larger and larger, so that ultimately it exceeds any number, however large that number may be. See also the discussion of divergence below.

3.03. LIMITS AND CONVERGENCE. The behaviour of a sequence S as $n \to \infty$ cannot be determined from a finite number of values, since these are not necessarily typical of what happens when n is really large. It is therefore absolutely necessary to have a formula $S = f(n)$ from which every term can be calculated, or a rule which explains in some other way how the terms can be derived.

There are three possibilities, for large n:

(*i*) The terms oscillate.
(*ii*) The values s_n get closer and closer to some definite value s.
(*iii*) They get larger and larger (either positive or negative).

Usually the problem is to distinguish between (*ii*) and (*iii*).

When (*ii*) occurs, the sequence is said to be convergent, and the number s is called the 'limiting value' or 'limit'. We write

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} f(n) = s,$$

which is read, 'the limit, as n tends to infinity, of s_n'. Normally the limit s is not any of the numbers s_n.

The following is a formal definition of convergence:
A sequence S converges to a limit s if only a finite number of the differences $s - s_n$ exceeds in magnitude any chosen number, however small.

When (*iii*) occurs the sequence is said to diverge:
A sequence S diverges to $(+)$ infinity if only a finite number of terms is less than any chosen positive number, however large. The definition of a sequence which diverges to minus infinity is similar. We write

$$\lim_{n\to\infty} f(n) = \infty.$$

There remain sequences which neither converge nor diverge in the defined sense of these terms. These are usually termed divergent by the mathematician, but the term oscillatory is more descriptive.

Example 1. When $s_n = (-)^n/n$ the sequence is $-1, \frac{1}{2}, -\frac{1}{3}, \frac{1}{4}, -\frac{1}{5}, \ldots$ and converges to the limit 0.

Example 2. The sequence $s_n = (-r)^n$ is $-r, r^2, -r^3 \ldots$, which converges to zero when r is less than one, and oscillates when r is equal to or greater than one.

Example 3. Let $s_n = A(1 - e^{-n^2/a^2})$, where A, e and a are constants. As n gets larger, e^{-n^2/a^2} gets smaller, so that the sequence has the limiting value A. Hence

$$\lim_{n\to\infty} A(1 - e^{-n^2/a^2}) = A.$$

Example 4. The sequences $s_n = 2n - 3$; $-1, +1, 3, 5, 7 \ldots$

$s_n = \log n$; $0, \log 2, \log 3, \ldots$

both diverge. The latter is said to diverge slowly (see below).

Example 5. Any polynomial $s_n = a + bn + cn^2 + \ldots$ diverges.

Example 6. The sequence $s_n = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2}$, i.e. $s_n - s_{n-1} = \frac{1}{n^2}$, converges to the limit $\pi^2/8$. (See Chapter 13.)

Example 7. The sequence $s_n = (1 + 1/n)^n$ is $2, \frac{9}{4}, \frac{64}{27}, \frac{625}{256}, \ldots$ and tends to the important number $e = 2{\cdot}71828 \ldots$, the so-called natural number. This answer is not obvious, as $(1 + 1/n)$ decreases as the power n increases, and it is not clear which gains control (cf. Section 8.01).

Many limits are obvious, but the last two examples show why a definition of convergence is necessary. Such a definition can give a test of convergence, and sometimes gives the value of the limit.

3.04. SEQUENCES OF RATIOS. The behaviour of a sequence of ratios of the form $s_n = (a_n/b_n)$ is usually clear when that of the separate sequences a_n and b_n is known.

Example 1. $s_n = r^n/n$ $(r < 1)$. Here $a_n = r^n \to 0$ and $b_n = n \to \infty$ as $n \to \infty$. Hence s_n converges to zero.

In general, if $\lim a_n = a$, and $\lim b_n = b$, then $\lim (a_n/b_n) = (a/b)$ as $n \to \infty$, unless $b = 0$, when the sequence diverges, or both $a = 0$ and $b = 0$, when the ratio is indeterminate.

Doubts only arise when a_n and b_n either both converge to the limit zero, or both diverge. The following example shows one way of dealing with this situation:

$$\frac{a_n}{b_n} = \frac{n+1}{2n+1} \to \frac{\infty}{\infty} \quad (?) \qquad \frac{a_n}{b_n} = \frac{1 + 1/n}{2 + 1/n} \to \frac{1}{2}.$$

$$\frac{a_n}{b_n} = \frac{1/n + 1/n^2}{2/n + 1/n^2} \to \frac{0}{0} \quad (?)$$

These three sequences being identical, they have the same limit $\frac{1}{2}$.

The moral is that a sequence and its limit are not changed by any algebraic alteration of the form of the function $f(n)$ describing the terms.

Ratios such as $0/0$ and ∞/∞ should in fact never be used. They have no meaning and to reach such an impasse implies erroneous reasoning.

Example 2. $\operatorname{Lim}_{n\to\infty} \dfrac{n^3+4}{2n^3+n} = \lim_{n\to\infty} \dfrac{1+4/n^3}{2+1/n^2} = \dfrac{1}{2}.$

Example 3. $\operatorname{Lim}_{n\to\infty} \dfrac{a_0+a_1n+\ldots+a_pn^p}{b_0+b_1n+\ldots+b_qn^q} = \lim_{n\to\infty} \dfrac{\dfrac{a_0}{n^p}+\ldots+a_p}{\dfrac{b_0}{n^q}+\ldots+b_q}\, n^{p-q}.$

As $n \to \infty$, the numerator tends to a_p, and the denominator to b_q. Hence the sequence diverges if $p > q$, converges to a_p/b_q when $p = q$, and to zero if $p < q$.

Example 4. $\operatorname{Lim}_{n\to\infty} n/\log_{10} n = \infty$ in spite of the large denominator. (This cannot be proved without using the ideas of later sections.) Let $n = 10^m$ so that the limit is that of $10^m/m$, which is obviously divergent.

3.05. Orders of Magnitude. If the sequence a_n/b_n has a finite limit, the sequence a_n is said to be of the same order of magnitude as b_n, when n is large.

If the sequence a_n/b_n tends to zero, then a_n is said to be of lower order of magnitude than b_n when n is large; or when both diverge, then a_n tends to infinity more slowly than b_n.

This characteristic enables us to compare sequences with the same limit zero or infinity.

Example 1. If $a_n = 1/n^2$, $b_n = 1/(n+1)$, then

$$s_n = a_n/b_n = (n+1)/n^2 \to 0$$

as $n \to \infty$. Hence the sequence $(1/n^2)$ is of a lower order of magnitude than $1/(n+1)$.

Example 2. If $a_n = (n+1)/n^2$ and $b_n = (n^2+1)/(2n^3+1)$, then

$$s_n = \frac{a_n}{b_n} = \frac{(n+1)(2n^3+1)}{n^2(n^2+1)}$$

tends to the limit 2 so that the two sequences are of the same order of magnitude, that of $(1/n)$.

Example 3. If $a_n = 1000/n$ and $b_n = 1/(n+1)$ then

$$s_n = a_n/b_n = 1000(n+1)/n \to 1000,$$

so that a_n is said to be of the same order of magnitude as b_n, when n is large, even though it is 1000 times as large numerically.

3.06. CONVERGENCE TESTS. Many of these tests are simple enough, though they should be proved to be consistent with the definition of convergence. The following tests are amongst the more useful:

(*i*) A sequence converges if every term lies between the two preceding terms, and the difference between successive terms tends to zero [*Figure 3.01* (a)].

(*ii*) A monotonic increasing sequence (each term bigger than the one before), where every term is less than some number M, converges to a limit $s \leqslant M$ [*Figure 3.01* (b)].

(*iii*) A monotonic decreasing sequence, every term greater than a number M, converges to a limit $s \geqslant M$.

(*iv*) If two sequences S and T are such that $s_n - t_n \to 0$ as $n \to \infty$, and S is convergent, then T converges to the same limit.

(*v*) If the ratio of successive terms $\left| s_{n+1}/s_n \right| \leq \alpha$, where $\alpha < 1$, after a finite number of terms, then the series converges to zero.

(*vi*) Comparison test: If S and T are two monotonic increasing sequences such that $t_n < as_n$ for every n, after a finite number of terms, where a is some constant, then T converges if S converges.

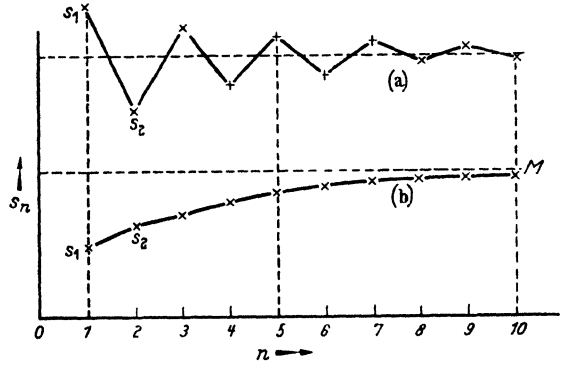

Figure 3.01

More advanced tests are to be found in advanced books. The reader should notice how they compare with the definition of convergence, and also note the severely arithmetical nature of the definition.

3.07. WASHING PRECIPITATES. The solvent left in a precipitate after draining may contain an impurity in concentration c_0. If the volume of solvent left is a the amount of impurity is $(c_0 a)$.

Washing the precipitate reduces the concentration. By adding a volume x of pure solvent, shaking and draining, the concentration of impurity is reduced to

$$c_1 = \frac{ac_0}{a + x_1}$$

both in the solvent removed and in the drained precipitate. A further washing using a volume x_2 of pure solvent reduces the concentration to

$$c_2 = c_0 \frac{a}{a + x_1} \cdot \frac{a}{a + x_2},$$

and after n washings,

$$c_n = c_0 \frac{a}{a + x_1} \cdot \frac{a}{a + x_2} \cdot \ldots \frac{a}{a + x_n}.$$

A total volume $v = x_1 + x_2 + \ldots x_n$ of pure solvent has been used.

If $x_1 = x_2 = \ldots = x_n = x$, then

$$c_n = c_0 \left(\frac{a}{a + x}\right)^n.$$

The concentrations form a sequence tending to zero with increasing n.

Let us now consider the concentration after a volume v of pure solvent has been used in n washes. Then $x = (v/n)$ and

$$c_n = c_0 \left(\frac{a}{a + v/n}\right)^n.$$

This is another sequence which decreases as n increases, i.e. a large number of small washes removes more impurity than a small number of large washes. When n is large

$$\frac{c_n}{c_0} = \left(1 + \frac{v}{na}\right)^{-n} = \left\{\left(1 + \frac{v}{na}\right)^{na/v}\right\}^{-v/a} \to e^{-v/a}, \text{ as } \frac{na}{v} \to \infty,$$

using *Example 7* of Section 3.03 (and 8.01).

According to the definition of a limit c_n/c_0 does not differ much from $e^{-v/a}$ when n is large, so that the value can be used to measure the effect of a large number of washes. Thus, to reduce the concentration by a factor of 1000, $e^{v/a}$ must be greater than 1000, or v/a must be greater than 7.

3.08. COMMENTS. The reader has probably noticed that we have not yet used the convergence tests, nor have we evaluated any limits of a difficult nature. This is deliberate. They appear later in this chapter. The most important things to remember about sequences are (*i*) the simpler tests, so that they can be rapidly applied when there is doubt of the convergence, and (*ii*) that the later terms of a sequence are very close indeed to the limit when it is convergent, so that either can be used as an approximation to the other.

3.09. FUNCTIONS OF A CONTINUOUS VARIABLE. A function $y = f(x)$ of a continuous variable x can converge, diverge, or oscillate as x tends to infinity, just as can a function of a discrete variable.

Example 1. $y = \sin x/x$: $\lim_{x\to\infty} (\sin x/x) = 0$.

Example 2. $y = x^2/(x+1)$: $\lim_{x\to\infty} (x^2/(x+1)) = \infty$.

Example 3. $y = \sin x$: oscillates as $x \to \infty$.

It is not possible to 'put $x = \infty$', although it is convenient to talk loosely as though it is. Hence to say 'when $x \to \infty$' always means 'when x is given any sequence of values tending to infinity'.

The statement $\lim_{x\to\infty} f(x) = a$ means that the same limit is obtained for *any* sequence of values of x tending to infinity and not merely the sequence 1, 2, 3, . . . This is the real distinction between limits of functions of a continuous variable and functions of a discrete variable. It can be made clearer by the following.

Example 4. For $x = 1, 2, 3, \ldots$ the function $f(x) = \sin \pi x$ has the values 0, 0, 0, . . . converging to 0 ! But when $x = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ it has the values $+1, -1, +1, -1, \ldots$ which oscillate as x tends to infinity. In fact by choosing a suitable sequence of values of x, any limit between -1 and $+1$ can be obtained.

3.10. LIMITS OF FUNCTIONS. CONTINUITY. When x has a sequence of values $a_1, a_2, \ldots$ converging to a limit a, then $y = f(x)$ has a sequence of values

$$y_1 = f(a_1), \quad y_2 = f(a_2), \quad y_3 = f(a_3), \ \ldots$$

Normally these converge to a limit, which is the value of $y = f(x)$ when $x = a$.

$$\lim_{x\to a} f(x) = f(a).$$

If this is true for all sequences of values of x converging to a, then the function $f(x)$ is continuous at $x = a$.

It is not always true. For some discontinuous functions one limit is obtained for x values less than a, and another limit for values of x greater than a. These two limits are

$$\lim_{x\to a^-} f(x) = f(a-0) \quad \text{and} \quad \lim_{x\to a^+} f(x) = f(a+0).$$

Other difficulties arise when $f(x)$ is a fraction, or is not determined for some reason at $x = a$.

Example 1. $\operatorname{Lim}_{x\to 1} \dfrac{x+1}{x^2+x+1} = \dfrac{1+1}{1+1+1} = \dfrac{2}{3}.$

Example 2. $\operatorname{Lim}_{x\to\infty} x^n = \infty \ (n>0), \quad = 1 \ (n=0), \quad = 0 \ (n<0).$

3.11. INDETERMINATE VALUES. In physical and chemical problems, when a function is not determined for some reason at $x = a$, it is usually best to use either

$$\lim_{x\to a} f(x) \quad \text{or} \quad \tfrac{1}{2}\{\lim_{x\to a^-} f(x) + \lim_{x\to a^+} f(x)\},$$

the second reducing to the first if the two limits in it are equal.

The commonest functions of this type reduce to the form $0/0$ if x is put equal to some particular value. These are dealt with as in Section 3.04.

Example 1. $f(x) = (x^2-4)/(x-2)$ is not defined at $x = 2$, but

$$\lim_{x\to 2} \frac{x^2-4}{x-2} = \lim_{x\to 2} (x+2) = 4.$$

Example 2. $\operatorname{Lim}_{x\to 0} \dfrac{x+x^3}{x+x^2} = \lim_{x\to 0} \dfrac{1+x^2}{1+x} = 1.$

Example 3.

$$\operatorname{Lim}_{x\to 0} \frac{\sqrt{(a+x)}-\sqrt{a}}{x}$$

$$= \lim_{x\to 0} \frac{\{\sqrt{(a+x)}-\sqrt{a}\}\{\sqrt{(a+x)}+\sqrt{a}\}}{x\{\sqrt{(a+x)}+\sqrt{a}\}}$$

$$= \lim_{x\to 0} \frac{(a+x)-a}{x\{\sqrt{(a+x)}+\sqrt{a}\}}$$

$$= \lim_{x\to 0} \frac{1}{\sqrt{(a+x)}+\sqrt{a}} = \frac{1}{2\sqrt{a}}.$$

3.12. FUNCTIONS OF SEQUENCES. When it is necessary to combine two sequences, as in Section 3.04, or by any other algebraic method, the limit of the combination is obtained by combining the limits of the

sequences. The same, of course, is true for a function of the numbers of a sequence, provided the function is continuous at the limit, e.g.

(*i*) $\lim_{n\to\infty}\{a_n + b_n\} = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$.

(*ii*) $\lim_{n\to\infty}\{\log a_n\} = \log\{\lim_{n\to\infty} a_n\}$.

3.13. THREE IMPORTANT LIMITS.

(*i*) $\lim_{a-b\to 0} \dfrac{a^n - b^n}{a - b} = na^{n-1}$.

For by long division

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \ldots + b^{n-1}).$$

This limit can be expressed, when $b = x$ and $a = x + h$, as

$$\lim_{h\to 0} \frac{(x + h)^n - x^n}{h} = nx^{n-1}.$$

(*ii*) $\lim_{\theta\to 0} \dfrac{\theta}{\sin\theta} = 1$, when θ is in radians.

Since $\sin\theta$ is defined geometrically, the proof is geometrical. From *Figure 7.01* (p. 79) it is clear that

$$\triangle OBN < \text{sector } AOB < \triangle OBC,$$

i.e. $\quad \frac{1}{2}a\cos\theta \,.\, a\sin\theta < (\theta/2\pi) \,.\, \pi a^2 < \frac{1}{2}a \,.\, a\tan\theta,$

or $\quad \cos\theta < \theta/\sin\theta < \sec\theta.$

As θ tends to zero, $\cos\theta \to 1$, $\sec\theta \to 1$, hence $\theta/\sin\theta \to 1$.

(*iii*) $\lim_{h\to 0} (a^h - 1)/h = \log_e a$.

Let h have the sequence of values $h_n = \log_a(1 + 1/n)$, i.e. $a^{h_n} - 1 = 1/n$.

Then $\lim_{n\to\infty} h_n = 0$.

Hence
$$\begin{aligned}\lim_{h\to 0} \frac{h}{a^h - 1} &= \lim_{n\to\infty} \frac{\log_a(1 + 1/n)}{(1/n)},\\ &= \lim_{n\to\infty}\{\log_a(1 + 1/n)^n\},\\ &= \log_a\{\lim_{n\to\infty}(1 + 1/n)^n\},\\ &= \log_a e.\end{aligned}$$

The required limit follows on taking reciprocals, using

$$\log_a e = 1/\log_e a.$$

EXAMPLES III

1. According to the Morse approximation, the vibrational energy of a diatomic molecule can have one of a finite sequence of values,

$$E_n = (n + \tfrac{1}{2})\hbar\nu - (n + \tfrac{1}{2})^2(\hbar\nu)^2/(4D),$$

where $\hbar$ is Planck's constant, ν the frequency of small vibrations, D the dissociation energy and n an integer.

This formula is to be used only as long as E increases with n. Find the largest value n_0 of n in the sequence, and show that $E_{n_0} = D$.

2. The value of a sequence of determinants D_n, $(n = 0, 1, 2, 3, \ldots)$ are obtained from a recurrence relation $D_n = 5D_{n-1} - 6D_{n-2}$. Find the sequence of values up to $n = 5$, when $D_0 = 1$ and $D_1 = 2$.

Prove, by direct substitution in the equation, that the values could be of the form $D_n = Ak^n$, provided that $k = 2$ or $k = 3$.

3. A sequence of numbers s_n satisfies the recurrence relation,

$$as_n + bs_{n-1} + cs_{n-2} = 0 \quad (n = 2, 3, 4, \ldots).$$

Show that the possible values of the sequences must have a formula of the type $s_n = Ap^n + Bq^n$, where p and q are the roots of the equation $ax^2 + bx + c = 0$ (assume $b^2 > 4ac$). Express A and B in terms of s_0 and s_1.

REFERENCE

The author has found useful information on the subject matter of this chapter in the following book:

BROMWICH, T. J. *An Introduction to the Theory of Infinite Series.* Macmillan and Co. Ltd, London, 1926.

CHAPTER 4

DIFFERENTIATION

Changes and rates of change—Graphical representation—Instantaneous rates of change—Differentiation—Derivative or derived function—Basic derivatives—Derivatives from first principles—Changes in related functions—Laws of small changes—Rules of differentiation—Relations between rates of change—Examples of use of rules—Derivatives with respect to functions—Function of a function rule—Chain rule—Differentials—Equality of terms to first order—Higher derivatives—Functions with simple higher derivatives—Leibnitz rule—Higher derivatives of functions—Notation—Logarithmic differentiation—Implicit differentiation—Interchange of variables

4.01. CHANGES AND RATES OF CHANGE. The differential calculus is the result of a particular attitude to three familiar ideas, namely, change, the possibility of relationships between changes, and rate of change. The essential concept is that it is much easier to work with relations between changes, when these relations are considered in the limit where the changes tend to zero. To correspond with this, a distinction is made between an average rate of change, which is the ratio of two changes, and an instantaneous rate of change, which is the limit of this ratio. To make this clear let us consider some examples. In this chapter, we assume that only two quantities are involved, one a function of the other.

An average rate, or rate of change, is the ratio of two changes which occur during the transition from an initial to a final state, and it depends on both the initial and final state. For example, an average rate of burning is measured by the ratio of the fuel consumed over a definite period to the time taken. Similarly, average velocity is the rate of change of position with time, which is measured by the ratio (distance gone)/(time taken), both these quantities being changes. Its dependence on the initial and final circumstances appears in the fact that there are many average velocities. The average velocity of a car on a journey need not be the same as its average velocity between two telegraph poles, or the indication of a speedometer, which gives an average speed over a distance of a few feet.

Let us now consider two examples where time is not involved. The specific heat of a substance is measured as the rate of change of heat content with temperature, per unit mass. Normally the change

in temperature is quite small, and for some purposes it must be small, e.g. to find the change of the specific heat of a substance with temperature, the change of temperature used for the determination of the specific heat should be smaller than that used when comparing specific heats.

This dependence of average rates on both the initial and final circumstances is eliminated experimentally by making the changes very small, and theoretically by the use of the differential calculus, where the changes considered tend to zero. It may happen that, experimentally, extremely small changes have little meaning; for example, the rate of burning of a gas has no meaning unless the time involved is sufficiently large for the reaction of a fairly large number of the molecules taking part. This does not alter the value of the idea of infinitesimal changes, just as the abstract geometrical point or line is a valuable idea even though it differs from a real point or line. As long as its limitations are realized, the ideal viewpoint has the tremendous advantage of simplicity in use.

4.02. GRAPHICAL REPRESENTATION. Graphically, rates of change appear as the slopes of lines. Assume that a graph is drawn of the

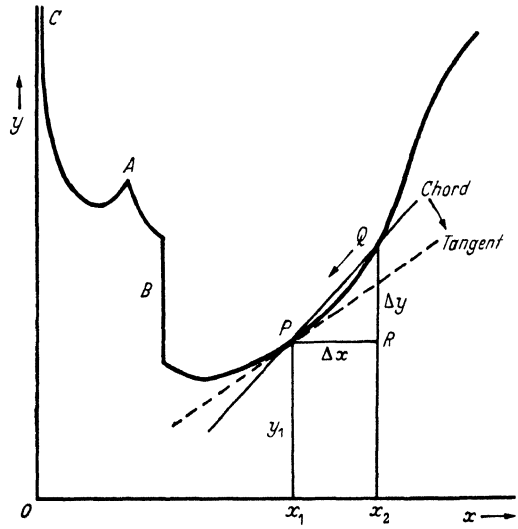

Figure 4.01

relation between y and x. A change of x from the value x_1 to the value x_2 corresponds to a change in y from the value y_1 to the value y_2, or from the point P to the point Q on the graph (*Figure 4.01*).

The average rate of change of y with x is measured by

$$\Delta y/\Delta x = RQ/PR,$$

or the slope of the chord PQ.

As the change Δx gets smaller, either by moving P on, or Q back, or both, the chord shifts its position, and in the limit when $\Delta x \to 0$ it becomes a tangent.

The instantaneous rate of change of y with x is therefore the slope of the tangent.

Difficulties can be expected at certain points, e.g. in the region of the point A or the point B. These can be removed by admitting the possibility of drawing two tangents at some points. A more serious difficulty arises at C, if y tends to infinity, since P and Q must be on the same side, but the definition is maintained by the introduction of a vertical tangent with infinite slope. Finally, a smooth curve has been assumed, whereas the actual curve may vary so rapidly that no tangent can be drawn at certain points.

4.03. INSTANTANEOUS RATES OF CHANGE. The definitions of the previous section can be expressed analytically for precise calculations.

The change in any quantity is defined as the final value less the initial value. Consequently its sign, when it is not zero, indicates whether the final value is greater than or less than the initial value.

Denoting the changes in x and y which correspond to one another by Δx and Δy, the average rate of change of y with x is

$$\Delta y/\Delta x,$$

and the instantaneous rate of change is denoted by

$$\frac{dy}{dx} = \operatorname*{limit}_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

if the limit exists. Whereas the average value depends on both the initial and final values of x, the limit only depends on their common value as $\Delta x \to 0$.

The reason for the notation dy/dx is that by definition $\Delta y/\Delta x$ differs from its limit by an arbitrarily small amount if Δx is sufficiently small. From one point of view $\Delta y/\Delta x$ is a very close approximation to the limit dy/dx, and from another point of view dy/dx is a very close approximation to the actual ratio $\Delta y/\Delta x$ of the very small changes. Both points of view are valuable.

If $$\Delta y/\Delta x = (dy/dx)(1 + r),$$

then $$\Delta y = (dy/dx)\Delta x(1 + r).$$

Thus r is the relative error in using the limit instead of the actual ratio; it is the relative error, *not* the absolute error, which tends to zero with Δx.

4.04. DIFFERENTIATION. In the calculation of the limit it is customary to use the symbol x as the initial value, so that $x + \Delta x$ is the final value of the variable (this double meaning of x as initial value and variable sometimes causes confusion). The corresponding values are y and $y + \Delta y$.

Thus, if $y = f(x)$, then $y + \Delta y = f(x + \Delta x)$. In taking the limit the initial value is kept fixed.

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x),$$

where $f'(x)$ is the limiting value.

As explained in Section 4.02, functions which are not smooth can give different tangents, and dy/dx can have different values, for positive values of Δx and negative values of Δx. If this happens the two limits are denoted by $f'(x + 0)$ and $f'(x - 0)$.

In the calculation of $f'(x)$ it is not necessary to use the symbol Δx for the change in x. If any other symbol simplifies the expressions obtained, it is better to use it. The value of the limit does not depend on Δx but only on the initial value of x. In a few of the following sections we find it convenient to use

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$

4.05. DERIVATIVE OR DERIVED FUNCTION. The limit $f'(x)$ is a function of the initial value x, which is called the derivative of $f(x)$, i.e. the function derived from it. Its value for a particular value of x, say $x = a$, is written

$$\left(\frac{dy}{dx}\right)_a = f'(a).$$

4.06. BASIC DERIVATIVES. We can prove, using the three limits of Section 3.13, the three basic derivatives:

(*i*) $y = x^n$: $dy/dx = nx^{n-1}$. (Cf. Sections 4.12 and 4.24.)

(*ii*) $y = \sin x$: $dy/dx = \cos x$. (Section 7.08.)

(*iii*) $y = e^x$: $dy/dx = e^x$. (Section 8.06.)

Most derivatives can be calculated from these using the rules given below.

4.07. DERIVATIVES FROM FIRST PRINCIPLES. This is the calculation of derivatives without using the rules given below or the basic derivatives given above. It is a test of manipulation to find limits. As several examples are given in the course of this text only one is reproduced here.

Example 1. $y = x^{1/2} = \sqrt{x}$. Some juggling is required to remove x from the denominator before taking the limit ($\Delta x = h$).

$$\frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{h} = \frac{(x+h)^{1/2} - x^{1/2}}{h} \to \frac{1}{2x^{1/2}},$$

as explained in *Example 4*, Section 3.11.

4.08. CHANGES IN RELATED FUNCTIONS. If one quantity u changes to $u + \Delta u$ and another v to $v + \Delta v$, the product $y = uv$ changes to

$$y + \Delta y = uv + \Delta(uv) = (u + \Delta u)(v + \Delta v),$$

i.e.

$$\Delta y = \Delta(uv) = (u + \Delta u)(v + \Delta v) - uv = u\Delta v + v\Delta u + \Delta u \Delta v.$$

This formula gives the change in a product due to changes in its factors. Other simple relations derived in the same way are given in *Table III.* They are the same as those given in Section 1.08, when discussing errors, provided Δu and Δv are regarded not as changes but as errors, i.e. the differences between the true and approximate values (see also Section 5.08).

4.09. LAWS OF SMALL CHANGES. The two expressions,

$$\Delta(uv) = u\Delta v + v\Delta u + \Delta u \Delta v,$$

and

$$\frac{1}{v + \Delta v} = \frac{1}{v} - \frac{\Delta v}{v^2} + \frac{(\Delta v)^2}{v^3} - \frac{(\Delta v)^3}{v^4} + \ldots \quad (\Delta v < v),$$

can be divided into terms of various orders in Δu, Δv.

Zero order: not containing Δu or Δv.

First order: containing either one factor Δu, or Δv.

Second order: containing $(\Delta u)^2$, $(\Delta u \Delta v)$ or $(\Delta v)^2$ as a factor.

And so on for higher orders, if they occur.

The ratio of two terms of the same order, such as $u\Delta v$ and $v\Delta u$, remains finite as Δu and Δv get small, whereas the ratio of a higher order to a lower order term tends to zero.

Table III

Item	y	Δy	$(\Delta y)_1$	$(\Delta y)_2$
I	$u + v$	$\Delta u + \Delta v$	$\Delta u + \Delta v$	0
II	$u - v$	$\Delta u - \Delta v$	$\Delta u - \Delta v$	0
III	cu	$c\Delta u$	$c\Delta u$	0
IV	uv	$u\Delta v + v\Delta u + \Delta u\Delta v$	$u\Delta v + v\Delta u$	$\Delta u\Delta v$
V	$\frac{1}{v}$	$-\frac{\Delta v}{v(v + \Delta v)}$	$-\frac{\Delta v}{v^2}$	$\frac{(\Delta v)^2}{v^3}$
VI	$\frac{u}{v}$	$\frac{v\Delta u - u\Delta v}{v(v + \Delta v)}$	$\frac{u\Delta v - u\Delta v}{v^2}$	$\frac{1}{v^3}\{u(\Delta v)^2 - v\Delta u\Delta v\}$

Thus, when Δu and Δv are sufficiently small, the higher order terms are numerically negligible compared with the lower order terms. Hence, if Δy is divided up in the same way, as given in *Table III*,

$$\Delta y = (\Delta y)_1 + (\Delta y)_2 + (\Delta y)_3 + \cdots$$

then the first term $(\Delta y)_1$ is a first approximation to Δy for small changes, $(\Delta y)_2$ a second order correction, and so on.

Example 1. To first order the change in an area $A = ab$ of a rectangle with sides a and b is $\Delta A = a\Delta b + b\Delta a$.

Example 2. To first order the change in the area $A = 4\pi r^2$ of a sphere of radius r, due to a small change Δr in the radius, is

$$(\Delta A)_1 = 4\pi \cdot 2r\Delta r = 8\pi r\Delta r.$$

The average rate of change of $y = f(x)$ due to a change Δx in x can be separated in the same way,

$$\frac{\Delta y}{\Delta x} = \frac{(\Delta y)_1}{\Delta x} + \frac{(\Delta y)_2}{\Delta x} + \cdots,$$

where $(\Delta y)_1$ is first order in Δx, $(\Delta y)_2$ is second order, etc. Hence $(\Delta y)_1/\Delta x$ is zero order, $(\Delta y)_2/\Delta x$ is first order. The calculation of these terms is given in Section 13.10, where the whole matter is discussed thoroughly. For the present it is sufficient to notice that the instantaneous rate of change, obtained by allowing $\Delta x \to 0$, reduces to

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{(\Delta y)_1}{\Delta x},$$

i.e. the first order change divided by Δx.

4.10. RULES OF DIFFERENTIATION. More complicated functions can be obtained from simple ones by addition, multiplication and division, and their derivatives can be calculated using the rules of *Table IV* for just these combinations.

Table IV

Item	y	dy/dx
I	$u + v$	$\frac{du}{dx} + \frac{dv}{dx}$
II	cu	$c\frac{du}{dx}$
III	uv	$u\frac{dv}{dx} + v\frac{du}{dx}$
IV	$\frac{1}{v}$	$-\frac{1}{v^2}\frac{dv}{dx}$
V	$\frac{u}{v}$	$\frac{1}{v^2}\left(v\frac{du}{dx} - u\frac{dv}{dx}\right)$
VI	u^2	$2u\frac{du}{dx}$
VII	$uvw \ldots$	$\frac{du}{dx}vw + u\frac{dv}{dx}w \ldots + uv\frac{dw}{dx}\ldots$

As a typical example, let $y = uv$, where u and v are functions of x. Then using the first order change of y,

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{(\Delta y)_1}{\Delta x} = \lim_{\Delta x \to 0} \frac{u\Delta v + v\Delta u}{\Delta x} = u\frac{dv}{dx} + v\frac{du}{dx}.$$

4.11. RELATIONS BETWEEN RATES OF CHANGE. The rules just given can be summarized by saying that the rate of change of y with x is the sum of the rates due to the separate components of y, whether they be factors or terms.

Example 1. The mass of a body is the product of its density ρ and its volume v. Hence $m = \rho v$, and the rate of change with time

$$\frac{dm}{dt} = \rho\frac{dv}{dt} + v\frac{d\rho}{dt}.$$

Example 2. The amount M of a substance increases at a rate a due to a process A, at a rate b due to a process B, and decreases at a rate

c due to a process C. Then due to A alone $dM/dt = +a$, due to B alone $dM/dt = b$, and due to C alone $dM/dt = -c$ (the minus because M decreases). Due to all three

$$dM/dt = a + b - c.$$

4.12. Examples of Use of Rules.

Example 1. $y = a + bx + cx^2 + dx^3 + \ldots + kx^n$,

$$dy/dx = b + 2cx + 3dx^2 + \ldots + nkx^{n-1}.$$

Example 2. $y = x \sin x$ ($u = x$, $v = \sin x$, so that $du/dx = 1$ and $dv/dx = \cos x$). Then $dy/dx = 1 \,.\, \sin x + x \,.\, \cos x$.

Example 3. $y = e^x \,.\, \sin x$, where $u = e^x$, $v = \sin x$ and $du/dx = e^x$, $dv/dx = \cos x$. Then $dy/dx = e^x \,.\, \sin x + e^x \,.\, \cos x = e^x(\sin x + \cos x)$.

Example 4. $y = \dfrac{1}{\sin x}$ $(v = \sin x)$;

$$\frac{dy}{dx} = -\frac{1}{\sin^2 x}\cos x = -\operatorname{cosec} x \,.\, \cot x.$$

Example 5. $y = x^{-n} = 1/x^n$ $(v = x^n)$;

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

Example 6. $y = \sin x/x^2$, so that $(u = \sin x, v = x^2)$,

$$du/dx = \cos x,\ dv/dx = 2x,$$

and

$$\frac{dy}{dx} = \frac{v\,du/dx - u\,dv/dx}{v^2}$$

$$= \frac{x^2\cos x - 2x\sin x}{x^4} = \frac{x\cos x - 2\sin x}{x^3}.$$

Example 7. $y = \sin^2 x = u^2$, where $u = \sin x$,

$$dy/dx = 2u\,du/dx = 2\sin x\cos x.$$

4.13. Derivatives with Respect to Functions. A relation giving a derivative, such as $y = \sin u$, $(dy/du) = \cos u$, is true whether u is an independent variable or depends upon another. It gives the rate of change of y with u, whatever the cause of the latter. When u is a

given function of x, the derivative dy/du is taken with respect to that function. For example:

$$\begin{aligned} u = x: &\qquad y = \sin x, &&\qquad dy/dx = \cos x. \\ u = x^2: &\qquad y = \sin (x^2), &&\qquad dy/d(x^2) = \cos (x^2). \\ u = \log x: &\qquad y = \sin (\log x), &&\qquad dy/d(\log x) = \cos (\log x). \end{aligned}$$

Expressed verbally, the derivative of a sine is always a cosine.

Similarly using the exponential function $y = e^{-u}$, $dy/du = -e^{-u}$,

$$\begin{aligned} u = x: &\qquad y = e^{-x}, &&\qquad dy/dx = -e^{-x}. \\ u = x^2: &\qquad y = e^{-x^2}, &&\qquad dy/d(x^2) = -e^{-x^2}. \\ u = \alpha/T: &\qquad y = e^{-(\alpha/T)}, &&\qquad dy/d(\alpha/T) = -e^{-\alpha/T}. \end{aligned}$$

or, with a simple power,

$$v = (\sin x)^3, \qquad \frac{dy}{d(\sin x)} = 3(\sin x)^2.$$

4.14. FUNCTION OF A FUNCTION (FF) RULE. This is a very useful rule. Let y be a function of u, where u is a function of x, as in the previous paragraph. By the elimination of u, y can be expressed in terms of x alone. The FF rule gives a relation between the derivatives,

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

To prove this, note that a value of x determines both u and y. If x changes by Δx, let the changes in the other variables be Δu and Δy. Provided the derivatives exist, so that the changes Δx, Δu and Δy at least tend to zero simultaneously, the relation between the derivatives is obtained by considering the identity

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x}$$

in the limit when $\Delta x \to 0$.

Example 1. If $y = (1 + x^2)^{1/2} = u^{1/2}$, where $u = 1 + x^2$, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = \frac{1}{2} u^{-1/2} \frac{du}{dx} = \frac{1}{2} (1 + x^2)^{-1/2} \cdot 2x = \frac{x}{(1 + x^2)^{1/2}}.$$

The working can be abbreviated by not introducing the symbol u, as in the previous paragraph.

Example 2. $y = \sin(x^2)$: $\dfrac{dy}{dx} = \dfrac{dy}{d(x^2)} \cdot \dfrac{d(x^2)}{dx} = \cos(x^2) \cdot 2x.$

Example 3. $y = e^{\sin x}$: $\dfrac{dy}{dx} = \dfrac{dy}{d(\sin x)} \cdot \dfrac{d(\sin x)}{dx} = e^{\sin x} \cdot \cos x.$

4.15. CHAIN RULE. The FF rule is easily extended to a series of successive derivatives, when necessary. Thus if y is a function of u, which is a function of v, which is a function of x, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \quad \text{and} \quad \frac{du}{dx} = \frac{du}{dv} \cdot \frac{dv}{dx},$$

i.e.

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

Example 1. $y = \sin^2 3x$, then

$$\frac{dy}{dx} = \frac{dy}{d(\sin 3x)} \cdot \frac{d(\sin 3x)}{d(3x)} \cdot \frac{d(3x)}{dx},$$

$$= 2 \sin 3x \cdot \cos 3x \cdot 3 = 6 \sin 3x \cos 3x.$$

Example 2. $y = (1 + e^{-x})^{5/2}$, i.e. a five-halves power,

$$\frac{dy}{dx} = \frac{dy}{d(1 + e^{-x})} \cdot \frac{d(1 + e^{-x})}{d(e^{-x})} \cdot \frac{d(e^{-x})}{d(-x)} \cdot \frac{d(-x)}{dx}$$

$$= \tfrac{5}{2}(1 + e^{-x})^{3/2} \cdot 1 \cdot e^{-x} \cdot (-1)$$

$$= -\tfrac{5}{2} e^{-x}(1 + e^{-x})^{3/2}.$$

After some practice this sort of calculation is simple.

4.16. DIFFERENTIALS. The differential of a variable x is written dx and represents the change in the variable to first order in small quantities.

The differential of a function $f(x)$ is defined as

$$df(x) = f'(x)dx,$$

where $f'(x)$ is the derivative of the function.

Example 1. If $f(x) = x^3$, then $df(x) = f'(x)dx = 3x^2dx$.

Example 2. $d(\sin x) = \cos x dx$; $d(\log x) = (1/x)dx$; $d(x \sin x) = (\sin x + x \cos x)dx$.

According to the section on small changes, with $y = f(x)$, the differential $f'(x)dx$ is the change in y to first order, when x changes

by dx. Hence differentials of functions represent small changes in the functions, to first order of small quantities.

The notation dy/dx, for the derivative, views it as the ratio of two differentials. For if $y = f(x)$, their differentials are equal, i.e.

$$dy = f'(x)dx \quad \text{or} \quad dy/dx = f'(x).$$

Table IV, giving changes to first order, can also be written in terms of differentials, e.g. if

$$y = uv, \text{ then } dy = udv + vdu.$$

4.17 EQUALITY OF TERMS TO FIRST ORDER. Two quantities which differ by a small quantity of the second order are equal to first order in small quantities.

Thus for a small change in x, the changes Δy and dy are equal to first order, although they differ by a small amount.

Similarly, when $y = f(u)$ and $u = g(x)$, a change Δx in x results in changes Δu and Δy in the other variables. The relations between the differentials are as follows:

$$(i)\ dy = \left(\frac{dy}{dx}\right) dx, \quad (ii)\ du = \left(\frac{du}{dx}\right) dx, \quad (iii)\ dy = \left(\frac{dy}{du}\right) du.$$

Let us now assume that $dx = \Delta x$, the actual change in x. The calculations of dy and du from (i) and (ii) are then made. The same value of the differential dy is obtained from (iii) by using this value of du.

However, a differential of y can be calculated using the actual change Δu of the variable u rather than the approximation du. This gives a different value of dy, but the two are equal to first order.

Thus we see that, whilst the differentials are consistent amongst themselves, it is permissible at any point to replace them by the correct changes, without changing the results to first order.

4.18. HIGHER DERIVATIVES. The derivative of $y = f(x)$ is a function of x which may be differentiable. The result,

$$\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2} = f''(x),$$

is called the second derivative of y, or of the function $f(x)$. By further differentiation higher derivatives are obtained.

Except for polynomials and a few other functions, higher derivatives are cumbersome expressions. Consequently is it usually better

not to evaluate each as a function of x, but to find recurrence relations between the derivatives. Only when a derivative is required for all values of x is it worth while collecting the whole expression and simplifying it.

4.19. Functions with Simple Higher Derivatives. A few functions have derivatives easily expressed to any order. Proofs by induction (Section 6.18) are simplest, as in the two worked examples. The reader should prove the others.

Example 1. $y = 2x^2 + 2$, $dy/dx = 4x$, $d^2y/dx^2 = 4$, $d^3y/dx^3 = 0$, etc.

Example 2. $y = \dfrac{1}{x+2}$, $\dfrac{dy}{dx} = -\dfrac{1}{(x+2)^2}$, $\dfrac{d^2y}{dx^2} = +\dfrac{2}{(x+2)^3}$.

Let $\dfrac{d^ny}{dx^n} = (-)^n \dfrac{n!}{(x+2)^{n+1}}$, then $\dfrac{d^{n+1}y}{dx^{n+1}} = (-)^n(-)\dfrac{(n+1)n!}{(x+2)^{n+1}}$

$$= (-)^{n+1}\frac{(n+1)!}{(x+2)^{n+2}}.$$

Hence this is the general form of the derivative.

Example 3. $y = e^{ax}\sin(bx + c)$.

Let $$\frac{d^ny}{dx^n} = (a^2+b^2)^{\frac{1}{2}n}e^{ax}\sin(bx + c + n\varepsilon),$$

where $$\cos\varepsilon = \frac{a}{(a^2+b^2)^{1/2}}, \sin\varepsilon = \frac{b}{(a^2+b^2)^{1/2}}.$$

This is certainly true for $n = 0$: assuming that it is true for any n,

$$\frac{d^{n+1}y}{dx^{n+1}} = (a^2+b^2)^{\frac{1}{2}n}\{ae^{ax}\sin(bx + c + n\varepsilon) + e^{ax}b\cos(bx + c + n\varepsilon)\},$$

$$= (a^2+b^2)^{\frac{1}{2}n}e^{ax}(a^2+b^2)^{\frac{1}{2}}\{\cos\varepsilon\sin(bx + c + n\varepsilon) + \sin\varepsilon\cos(bx + c + n\varepsilon)\},$$

$$= (a^2+b^2)^{\frac{1}{2}(n+1)}e^{ax}\sin\{bx + c + (n+1)\varepsilon\}.$$

It is therefore true for the next value of n.

List of formulae:

(i) $y = x^m$, $\dfrac{d^ny}{dx^n} = m(m-1)\ldots(m-n+1)x^{m-n}$ $(n > 0)$.

I m is an integer,

$$\frac{d^my}{dx^m} = m!, \quad \frac{d^ny}{dx^n} = 0 \quad (n > m).$$

(*ii*) $y = \dfrac{1}{ax+b}, \quad \dfrac{d^n y}{dx^n} = (-)^n \dfrac{n!\,a^n}{(ax+b)^{n+1}}.$

(*iii*) $y = \dfrac{1}{(ax+b)^\alpha}, \quad \dfrac{d^n y}{dx^n} = (-)^n \dfrac{(n+\alpha-1)!}{(\alpha-1)!} \dfrac{a^m}{(ax+b)^{\alpha+n}}.$

(*iv*) $y = e^{mx}, \quad \dfrac{d^n y}{dx^n} = m^n e^{mx}.$

(*v*) $y = \sin(bx+c), \quad \dfrac{d^n y}{dx^n} = b^n \sin(bx + c + \tfrac{1}{2}n\pi).$

(*vi*) Rational functions can be expressed as a sum of terms (*ii*) and (*iii*).

Example 3. $y = \dfrac{1}{x^2-1} = \dfrac{1}{2}\left\{\dfrac{1}{x-1} - \dfrac{1}{x+1}\right\}$. Hence, using (*ii*),

$$\frac{d^n y}{dx^n} = \frac{1}{2}\, n!(-)^n \left\{\frac{1}{(x-1)^{n+1}} - \frac{1}{(x+1)^{n+1}}\right\}.$$

4.20. LEIBNITZ RULE. This is similar to the Binomial Theorem (Section 6.12) and is proved by induction in the same way.

$$\frac{d^n}{dx^n}(uv) = u\frac{d^n v}{dx^n} + n\frac{du}{dx}\frac{d^{n-1}v}{dx^{n-1}} + \frac{n(n-1)}{1.2}\frac{d^2u}{dx^2}\frac{d^{n-2}v}{dx^{n-2}} + \ldots + \frac{d^n u}{dx^n}v.$$

It is useful when all but a few terms are zero, i.e. when u is a polynomial of low degree.

Example 1. If $y = x^2e^x$, let $u = x^2$ and $v = e^x$. Then

$$\frac{d^n}{dx^n}(x^2e^x) = x^2\frac{d^n(e^x)}{dx^n} + n(2x)\frac{d^{n-1}(e^x)}{dx^{n-1}} + \frac{n(n-1)}{1.2}2\frac{d^{n-2}(e^x)}{dx^{n-2}} + 0$$
$$= x^2e^x + 2nxe^x + n(n-1)e^x.$$

4.21. HIGHER DERIVATIVES OF FUNCTIONS. The FF rule is often employed when the form of the final function is unknown.

Example 1. $\dfrac{d}{dx}(y^2) = \dfrac{d(y^2)}{dy}\cdot\dfrac{dy}{dx} = 2y\dfrac{dy}{dx}.$

Example 2. $\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right)^2 = 2\dfrac{dy}{dx}\cdot\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right) = 2\dfrac{dy}{dx}\dfrac{d^2y}{dx^2}.$

Example 3. $\dfrac{d}{dx}\left(y^2\dfrac{dy}{dx}\right) = \dfrac{d}{dx}(y^2)\cdot\dfrac{dy}{dx} + y^2\dfrac{d}{dx}\left(\dfrac{dy}{dx}\right)$ (as a product),

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

Example 4.
$$\frac{d}{dx}(e^v) = e^v \frac{dy}{dx}.$$
$$\frac{d^2}{dx^2}(e^v) = \frac{d}{dx}\left(e^v \frac{dy}{dx}\right) = \frac{d}{dx}(e^v)\frac{dy}{dx} + e^v \frac{d^2y}{dx^2},$$
$$= e^v\left\{\left(\frac{dy}{dx}\right)^2 + \frac{d^2y}{dx^2}\right\}.$$

4.22. NOTATION. Alternative notations used for higher derivatives have as their aim the simplification of expression, e.g. y' or y_1 for dy/dx, and y'' or y_2 for (d^2y/dx^2).

Example 1. (Cf. *Examples 1* and *3* above.)
$$\frac{d}{dx}(y^2) = 2yy_1.$$
$$\frac{d}{dx}(y^2y_1) = (2yy_1)y_1 + y^2y_2 = y(2y_1^2 + yy_2).$$

Example 2. $d(u_mv_n)/dx = u_{m+1}v_n + u_mv_{n+1}$, and Leibnitz's rule is
$$(uv)_n = uv_n + nu_1v_{n-1} + \tfrac{1}{2}n(n-1)u_2v_{n-2} + \ldots + u_nv.$$
In mechanics, where suffices are employed for other purposes and derivatives are usually taken with respect to time, a dot notation is employed.
$$\frac{dy}{dt} = \dot{y}, \quad \frac{d^2y}{dt^2} = \ddot{y}.$$

4.23. LOGARITHMIC DIFFERENTIATION. When $y = uvw\ldots$ so that $\log y = \log u + \log v + \log w + \ldots$, then (Section 8.08),
$$\frac{1}{y}\frac{dy}{dx} = \frac{1}{u}\frac{du}{dx} + \frac{1}{v}\frac{dv}{dx} + \frac{1}{w}\frac{dw}{dx} + \ldots$$
This is very convenient when y has a number of factors, and numerical values of the derivatives are required.

Example 1. $y = (1+x)^3/(1+x^2)$, and
$$\log y = 3\log(1+x) - \log(1+x^2).$$
Then
$$\frac{1}{y}\frac{dy}{dx} = \frac{3}{1+x} - \frac{2x}{1+x^2} = \frac{3-2x+x^2}{(1+x)(1+x^2)}.$$
This can be differentiated again to give higher derivatives.
$$\frac{d}{dx}\left(\frac{1}{y}\frac{dy}{dx}\right) = \frac{1}{y}\frac{d^2y}{dx^2} - \frac{1}{y^2}\left(\frac{dy}{dx}\right)^2 = -\frac{3}{(1+x)^2} - \frac{2}{1+x^2} + \frac{4x^2}{(1+x^2)^2}.$$

4.24. IMPLICIT DIFFERENTIATION. Using the FF rule we can find dy/dx from equations such as $y^2 = \sin x$, without first solving for y.

Example 1. If $y^2 - 2xy + 1 = 0$, then $y = x \pm \sqrt{(1 + x^2)}$. Rather than differentiate this solution we can find the derivatives from the quadratic equation directly. It gives

$$2y\frac{dy}{dx} - 2\left(x\frac{dy}{dx} + y\right) = 0,$$

so that

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

If $y = x + \sqrt{(1 + x^2)}$, then $\dfrac{dy}{dx} = \dfrac{y}{y - x} = \dfrac{x + \sqrt{(1 + x^2)}}{\sqrt{(1 + x^2)}}$.

If $y = x - \sqrt{(1 + x^2)}$, then $\dfrac{dy}{dx} = \dfrac{y}{y - x} = \dfrac{x - \sqrt{(1 + x^2)}}{-\sqrt{(1 + x^2)}}$.

It is also possible to calculate higher derivatives from

$$(y - x)\frac{dy}{dx} - y = 0.$$

Differentiating this equation, which has no fractions,

$$\left\{(y - x)\frac{d^2y}{dx^2} + \left(\frac{dy}{dx} - 1\right)\frac{dy}{dx}\right\} - \frac{dy}{dx} = 0,$$

i.e.

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

For example when $x = 0$, it is easily found that

$$y = 1, \quad dy/dx = 1, \quad d^2y/dx^2 = 1.$$

Example 2. The derivative of $y = x^n$ has been shown to be $dy/dx = nx^{n-1}$, when n is a positive integer (Section 3.13) or a negative integer (Section 4.12). To prove that it is true when n is a rational fraction, i.e. the ratio p/q of two integers, we write $y = x^n$ in the form $y^q = x^p$ and then differentiate.

4.25. INTERCHANGE OF VARIABLES. It is quite common in theoretical work to interchange the independent and dependent variables, to regard x as a function of y rather than y as a function of x. The derivatives of x with regard to y are related to those of y with regard to x.

They can all be obtained using the FF rule, but dx/dy can

easily be obtained from the definition. Since $\Delta x/\Delta y = 1/(\Delta y/\Delta x)$ it follows, in the limit when $\Delta x \to 0$ and $\Delta y \to 0$, that

$$dx/dy = 1/(dy/dx).$$

The next relation is more complicated. Let $dy/dx = p$, so that $dx/dy = (1/p)$. Then

$$\frac{d^2x}{dy^2} = \frac{d}{dy}\left(\frac{1}{p}\right) = -\frac{1}{p^2}\frac{dp}{dy}.$$

But $p = dy/dx$ is a function of x, so that $\dfrac{dp}{dy} = \dfrac{dp}{dx}\dfrac{dx}{dy} = \dfrac{1}{p}\dfrac{dp}{dx}$.

$$\therefore \quad \frac{d^2x}{dy^2} = -\frac{1}{p^3}\frac{dp}{dx} = -\left(\frac{d^2y}{dx^2}\right)\Big/\left(\frac{dy}{dx}\right)^3.$$

For third derivatives a similar method is employed.

EXAMPLES IV

1. Water falls at a constant rate v cm^3/sec on to a horizontal circular plate, of radius r cm, and flows over the edge at a rate kx^3 cm^3/sec per unit length around the edge, where x is the depth of water at the edge. Find the rate of increase of x with time (t) as a function of x, assuming that the water on the disc is in the form of a flat disc of depth x everywhere, and that the two flows balance when $x = d$, if $r = 1$.

2. The frequency of light emitted by a hydrogen atom has one of the values

$$\nu = \frac{2\pi^2 e^4 \mu_H}{h^3}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right),$$

where the quantum numbers n_1 and n_2 are integers. What does this become, approximately, if $n_2 - n_1 = 1$ and both n_1 and n_2 are large.

3. In special relativity the energy E and momentum p of a free particle are related by the equation $E^2 = c^2(m^2c^2 + p^2)$, where m is the mass and c the velocity of light. If the velocity is $v = (dE/dp)$, prove that

$$p = mv(1 - v^2/c^2)^{-1/2}.$$

4. If $x = \sin y$, prove that $(1 - x^2)(d^2y/dx^2) = x(dy/dx)$. Differentiate this equation once, twice, and then n times using the Leibnitz theorem, and deduce from these equations the values of all the derivatives of y, when $x = 0$.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

EDWARDS, J. *An Elementary Treatise on the Differential Calculus.* Macmillan and Co. Ltd, London, 1908.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering.* Chapter 9. McGraw-Hill Book Co. Inc., New York, 1939.

CHAPTER 5

APPLICATIONS OF DIFFERENTIATION

Introduction—Geometrical interpretation—Approximations to curves. Newton's method—Increasing, decreasing and stationary functions—Curvature—Small changes—Calculation of errors—Maxima and minima—Second derivatives—Alternate maxima and minima—Greatest and least values—Other stationary values—Equilibrium conditions and stability—Rolle's theorem—Mean value theorem—L'Hôpital's rule for limiting values

5.01. INTRODUCTION. Having discussed ways of finding derivatives, we now consider their uses. Amongst the applications are those methods of solving equations already discussed in Chapter 2, where the slopes of curves were required and found graphically. More are given in this chapter. Some require lengthy discussion and are left to later chapters.

5.02. GEOMETRICAL INTERPRETATION. It has already been mentioned that the slope at any point P on the graph of the function $y = f(x)$ is given by

$$m = \tan \psi = \left(\frac{dy}{dx}\right)_P.$$

Assuming that the coordinates of P are (x_P, y_P), the equation of the tangent to the curve at P is

$$y - y_P = m(x - x_P) = \left(\frac{dx}{dy}\right)_P (x - x_P),$$

and the equation of the normal, which is perpendicular to the tangent and has a slope $-(1/m)$, is

$$m(y - y_P) + (x - x_P) = 0.$$

5.03. APPROXIMATIONS TO CURVES. NEWTON'S METHOD. Any curve which is to approximate to a given curve near a point P is, in general, first made to pass through P. Then it is made to have the same slope at the point P. The simplest of such curves is the tangent at P. The next step would seem to be to make the values of d^2y/dx^2 the same at the point P, but this cannot be done using a straight line, for which this derivative is zero. The process can be continued and is discussed in full in Chapter 13.

The tangent approximation can be very useful. It is the basis of

Newton's method of solving equations given in Section 2.04. Given an approximate root x_B of the equation $f(x) = 0$, we obtained a better approximation

$$x_D = x_B - y_B/m_B,$$

where m_B is the slope of the tangent at B. Using the relation $m = dy/dx$, this becomes

$$x_D = \{x - y/(dy/dx)\}_B,$$

which can be worked out arithmetically without using a graph. The iterative method, and the chord-and-tangent method, can be modified in the same way.

5.04. INCREASING, DECREASING AND STATIONARY FUNCTIONS. In Section 4.03 it was pointed out that $\dfrac{\Delta y}{\Delta x} = \dfrac{dy}{dx}(1 + r)$, where r tends to zero with Δx. Hence, when Δx is small and positive, Δy has the same sign as dy/dx.

(dy/dx) positive: y increases with x.
(dy/dx) negative: y decreases with x.
$(dy/dx) = 0$: y is stationary.

Example 1. The viscosity μ of a mixture of fluids A and B in concentrations c and $(1 - c)$ respectively, is

$$\frac{1}{\mu} = \frac{c}{\mu_A} + \frac{1 - c}{\mu_B},$$

where μ_A and μ_B are the viscosities of the pure fluids.

$$\therefore \qquad -\frac{1}{\mu^2}\frac{d\mu}{dc} = \frac{1}{\mu_A} - \frac{1}{\mu_B},$$

so that μ increases steadily with concentration if $\mu_A > \mu_B$, and decreases steadily if $\mu_A < \mu_B$.

Example 2. In a reversible reaction $A + B \rightleftharpoons C + D$, the concentration x of the products increases with time according to the formula

$$dx/dt = k_1(a - x)(b - x) - k_2x^2,$$

where k_1, k_2, a and b are positive, and $a < b$.

If $x = 0$, initially, then dx/dt is positive. Thus x increases and continues to increase until $dx/dt = 0$. This must occur before $x = a$. The equilibrium or stationary value is given by the equation

i.e. $$k_1(a - x)(b - x) = k_2x^2, \quad (0 < x < a).$$

Example 3. In a first order chemical reaction the amount x of the material changes according to the equation $(dx/dt) = -kx$, where k is the rate constant. That the equation gives a steady decrease is shown by the fact that (dx/dt) is negative.

5.05. CURVATURE. The graph of $y = f(x)$ is:

concave upwards when dy/dx increases algebraically with x,

i.e. $$\frac{d^2y}{dx^2} > 0;$$

convex upwards when dy/dx decreases as x increases,

i.e. $$\frac{d^2y}{dx^2} < 0.$$

Thus:

concave to the x axis when $\frac{d^2y}{dx^2}$ and y have opposite signs;

convex to the x axis when $\frac{d^2y}{dx^2}$ and y have the same sign.

Example 1. If $y = \sin x$, then $\frac{d^2y}{dx^2} = -y$. Hence its graph is concave towards the x axis.

Example 2. The graph of y, where $\frac{d^2y}{dx^2} + \{E - f(x)\}y = 0$, is concave to the x axis for those values of x for which $f(x) < E$. If this is true for all x, then y must oscillate about the x axis (cf. $y = \sin x$); for, starting from the origin, it must bend round to cut the x axis at some point (since it is concave to that axis). Past this point, y changes sign, but the graph is still concave to the x axis and cuts it again, and so on.

Example 3. In *Example 3* of the previous section, $dx/dt = -kx$, so that $(d^2x/dt^2) = -k(dx/dt) = k^2x$. Hence not only does x decrease with decreasing time, but the curve is everywhere convex to the x axis.

Points of inflection, where there is no curvature, are points where $d^2y/dx^2 = 0$.

5.06. SMALL CHANGES. Combining the results of Sections 4.16 and 4.03, we see that the differential $dy = f'(x)dx$ is an approximation to the change in $y = f(x)$ as x changes to $x + dx$. The values of dx for which it is useful depend on the accuracy required but, provided

dy/dx exists, it is correct to any accuracy for some small range of values of dx.

This can be put in a different form by a change in notation. Let the initial value be $x = a$ and the final value $x + dx = a + h$. The approximation is now

$$f(a + h) - f(a) = \Delta y = hf'(a),$$

or

$$f(a + h) = f(a) + hf'(a),$$

to first order.

If asked to calculate sin 31°, we normally react by assuming that it is not far from sin 30° $= \frac{1}{2}$, and follow this up by an inquiry about the difference between sin 31° and sin 30°. The above formula gives the answer to this question (approximately).

Let $f(x) = \sin x$ (x in radians), so that $f'(x) = \cos x$. The values of the variables are $a = \frac{1}{6}\pi$ radians (30°), and $h = (\pi/180)$ radians (1°). Thus

$$\sin 31° - \sin 30° = f(a + h) - f(a) = hf'(a)$$

$$= \frac{\pi}{180} \cdot \cos 30° = \frac{\pi}{180} \cdot \frac{\sqrt{3}}{2} = 0{\cdot}0151,$$

i.e. $\sin 31° = 0{\cdot}515$ approximately.

(There is no point in adding further decimal places.)

The difficulty about this approximation is that the error is unknown (cf. Section 5.15).

5.07. Calculation of Errors. If y is evaluated from the formula $y = f(x)$, using a value x with an error of $\pm dx$, the true value of y lies between $f(x + dx)$ and $f(x - dx)$. The error in y lies between

$$f(x + dx) - f(x) \quad \text{and} \quad f(x) - f(x - dx).$$

Using the simplest approximation for these differences, the error in y is between $\pm E$, where (apart from sign)

$$E = f'(x)dx = dy.$$

In other words, when the errors are small, the relation between them is that between the differentials. This is an extension of the result of Section 1.08.

The relative error in y is the differential of $\ln y$ (Section 8.08). It is

$$E/y = dy/y = d(\ln y).$$

Example 1. According to Poiseuille's law, the viscosity μ of a fluid can be calculated from the volume flow Q down a circular tube of radius r, using the formula

$$\mu = (\pi\rho C/8Q)r^4,$$

where C is the pressure drop per unit length and ρ is the density of the fluid. A measurement of r being liable to error, the error in μ is

$$d\mu = (\pi\rho C/8Q) \,.\, 4r^3 dr.$$

The relative error is

$$d\mu/\mu = 4(dr/r).$$

To measure μ correct to 1 per cent, it is necessary to measure the radius of the tube to an accuracy of $\frac{1}{4}$ per cent, i.e. 0·0025 mm for a capillary of 1 mm radius.

Example 2. If $y = kx(1 - x)$ is used to determine the constant k from given values of y and x, then

$$k = y/\{x(1 - x)\}, \quad \text{i.e. } \ln k = \ln y - \ln x - \ln (1 - x).$$

For small alterations in y and x, the relative error in k is

$$\begin{aligned} dk/k &= dy/y - dx/x + dx/(1 - x), \\ &= dy/y + (dx/x)\{(2x - 1)/(1 - x)\}. \end{aligned}$$

Assuming that y is known accurately, so that $dy = 0$, then an error dx in the measurement of x gives a relative error $(dx/x)\{(2x - 1)/(1 - x)\}$ which is least when $x = \frac{1}{2}$ (this approximation actually gives zero), and becomes very large when x is small or nearly equal to unity.

Notice that the relative errors of x and k have a ratio which is finite when x is small.

5.08. MAXIMA AND MINIMA. A function $y = f(x)$ has a maximum at $x = a$ if its value there is greater than for neighbouring values of x, and a minimum if it is less than at neighbouring values. For simple functions their position can be guessed, and proved by algebra.

Example 1. $y = \cos x$ has a maximum at $x = 0$, since $\cos x \leqslant 1$.

Example 2. $y = ax^2 - 2bx + c$ has a maximum at $x = b/a$ when a is negative, and a minimum there when a is positive. For

$$y = (1/a)\{a^2x^2 - 2abx + ac\} = (1/a)\{(ax - b)^2 + ac - b^2\},$$

which is greater than its value for $x = b/a$ when a is positive and less than this value when a is negative.

Example 3. The function $y = x/(1 + x^2)$ obviously has a maximum somewhere, because it is zero when $x = 0$ and when $x \to \infty$, and is positive between. This maximum value is $\frac{1}{2}$ when $x = 1$; for

$$\tfrac{1}{2} - y = \tfrac{1}{2} - x/(1 + x^2) = (1 - x)^2/2(1 + x^2),$$

which is always positive, except at $x = 1$, where it is zero.

When a function y has a derivative, maxima and minima are amongst the stationary values of y, i.e. $dy/dx = 0$. Otherwise for small values of dx, $\Delta y = (dy/dx)dx$ would change sign, in contradiction to our definition. For the converse of this, see Section 5.12.

In point of fact, if the function is continuous, it is dy/dx which changes sign at a maximum or a minimum*. It changes from plus to minus as x increases through a maximum and from minus to plus as x increases through a minimum.

Example 4. The function $y = xe^{-x}$ is zero when $x = 0$, and as $x \to \infty$. Differentiate to find the maximum.

$$(dy/dx) = e^{-x}(1 - x),$$

which is zero at $x = 1$. As there is only one stationary value it must be the maximum. The maximum is therefore $y = (1/e)$ when $x = 1$.

Example 5. For some molecular processes, the velocity of the reaction is proportional to

$$f(E) = p(E) \,.\, aE^{1/2} \,.\, e^{-E/kT}.$$

Roughly speaking, the second and third factors give the probability that the molecular system has an available energy E for the reaction, and $p(E)$ is the probability that the reaction occurs when it has that energy. Normally $p(E)E^{1/2}$ increases steadily with E from the value 0 when $E = 0$, whereas the exponential decreases steadily to zero as $E \to \infty$.

The product $f(E)$ therefore increases steadily with increasing E for small values, and in general decreases for large values. It has a maximum when $df/dE = 0$, or, since f is not zero, when

$$d(\log f)/dE = (1/f)(df/dE) = 0,$$

i.e.
$$(1/p)(dp/dE) + 1/2E - 1/kT = 0.$$

* This is true even for a function such as $y = |x|$, which is continuous, but has the two derivatives $f'(0^+)$ and $f'(0^-)$, at $x = 0$.

Example 6. If $y = x/(1 + x^2)$, then

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

The stationary values ($dy/dx = 0$) occur when $x = +1$ and $x = -1$. As x increases through $+1$, dy/dx decreases from positive to negative, i.e. this is a maximum.

As x *increases* (algebraically) through -1, dy/dx increases from negative to positive, i.e. this is a minimum.

5.09. SECOND DERIVATIVES. Since dy/dx is decreasing at a maximum, normally d^2y/dx^2 is negative. At a minimum dy/dx is increasing and d^2y/dx^2 is positive. Thus it is possible to distinguish a maximum or a minimum by the sign of the second derivative.

Sometimes it happens that this second derivative is zero even at a maximum or minimum. When this happens, the conditions given on page 55 should be used.

Example 1. Use the last example $y = x/(1 + x^2)$, where

$$dy/dx = (1 - x^2)/(1 + x^2)^2$$

and

$$\frac{d^2y}{dx^2} = \frac{-2x}{(1 + x^2)^2} - (1 - x^2)\frac{2 \cdot 2x}{(1 + x^2)^3}.$$

At $x = 1$, $dy/dx = 0$ and $d^2y/dx^2 = -\frac{1}{2}$; this, being negative, indicates a maximum.

At $x = -1$, $dy/dx = 0$ and $d^2y/dx^2 = +\frac{1}{2}$; i.e. positive, indicating a minimum.

Example 2. If $y = x^4$, then $dy/dx = 4x^3$, and $d^2y/dx^2 = 12x^2$.

At $x = 0$, all three are zero but there is clearly a minimum.

5.10. ALTERNATE MAXIMA AND MINIMA. A useful practical point is that a continuous function must have a minimum between two maxima, and vice versa, i.e. maxima and minima occur alternately. Hence, arranging the values of x where $f(x)$ is either a maximum or a minimum in order of magnitude, it is only necessary to decide whether one of them is a maximum to answer the question for them all.

5.11. GREATEST AND LEAST VALUES. The following remarks are fairly obvious, but are worth mention.

If dy/dx is positive and tends to zero as $x \to \infty$, the greatest or least value of y may be reached asymptotically as $x \to \infty$.

Example 1. $y = a(1 - e^{-kt})$, so that $dy/dt = ake^{-kt}$. For this function the derivative is positive for all values of t so that y steadily increases with t, but the derivative tends to zero as t increases, corresponding to the fact that y tends asymptotically to the value a in this limit.

When the range of values of the variable appropriate to a problem is limited, the greatest and least values may not occur at the maxima or minima, if they exist, but at the extremes of the range.

5.12. Other Stationary Values. The derivative (dy/dx) may be zero at a point where y is neither a maximum nor a minimum, but increases (or decreases) with x on both sides of the point.

At such stationary values dy/dx is zero but does not change sign, so that dy/dx is either a maximum or a minimum. Hence these values only occur where d^2y/dx^2 is zero as well as dy/dx.

Note, however, that $d^2y/dx^2 = 0$ does not exclude the possibility of a maximum or a minimum, as is shown by *Example 2*, Section 5.09, the reason being that dy/dx can change sign, even when d^2y/dx^2 is zero.

5.13. Equilibrium Conditions and Stability. Problems of this type are excellent examples of the work of the last few sections. Usually there is a system having its state described by a variable (x). For each value of x it has a potential V from which there is derived a force F acting on the system and tending to change the value of x. This force F is a function of x, the sign of which is normally chosen so that F is positive when it tends to increase the value of x and negative when it tends to decrease the value of x.

Assume that

$$F = -dV/dx.$$

Then F is positive when V decreases with increasing x and it is negative when V increases with x. Hence the system is acted on by a force tending to change x so that the potential V decreases. It tends to a state where V is as small as possible.

The equilibrium positions of the system are those where no force acts, i.e.

$$dV/dx = 0.$$

These are the states where V has its stationary values, and is a maximum, minimum or neither.

A stable position of equilibrium is one where a slight movement or change of state brings into play small forces tending to restore

the system to the equilibrium position. Stable equilibrium therefore occurs where V is a minimum.

Other equilibrium positions are unstable since, if V is a maximum, any slight change decreases V and brings into play forces tending to decrease V still more, and if the stationary value is neither a maximum nor a minimum, then a slight movement one way does the same thing*.

The only exception to these results arises when the values of the variable x are limited. A position of stable equilibrium may then occur at one of the extreme values of x where the potential has a least value, but dV/dx is not zero.

Example 1. The potential of the forces between the two atoms of a diatomic molecule in its normal state is a function of the distance x between the atoms. It is a good approximation to write

$$V = \frac{A}{x^n} - \frac{B}{x^6}, \quad (A, B > 0),$$

where $n = 12$, though it may have other values between 9 and 13 for some molecules. The equilibrium distance is that where

$$\frac{dV}{dx} = 0, \text{ i.e. } -\frac{nA}{x^{n+1}} + \frac{6B}{x^7} = 0, \quad \text{or } x^{n-6} = \frac{nA}{6B}.$$

If this distance is called $x = a$, so that $a^{n-6} = (nA/6B)$, then we can let $A = Ca^n$ so that $B = (nAa^{5-n}/6) = nCa^6/6$. Then V has the form

$$V = C\left(\frac{a^n}{x^n} - \frac{na^6}{6x^6}\right).$$

It is easily verified that V has a minimum when $x = a$.

5.14. ROLLE'S THEOREM. This simple but useful theorem states that a continuous curve $y = f(x)$ cutting the x axis at two points $x = a$ and $x = b$ has at least one maximum or minimum at some point between, or is zero for all values of x. An analytical proof is hardly necessary.

More formally, if $y = f(x)$ is zero when $x = a$ and $x = b$, and is continuous and differentiable between these values, then $dy/dx = 0$ for at least one value $x = \xi$, between a and b. (It may also be zero at $x = a$ or b.)

* Sometimes the forces are so related to the potential that $F = +(dV/dx)$. Then the positions of stable equilibrium are those where V is greatest.

5.15. MEAN VALUE THEOREM. If A and B are two points on a continuous curve $y = f(x)$ with a continuously turning tangent, then the chord AB is parallel to the tangent at some point between (draw a diagram to illustrate this).

This theorem is a disguised form of Rolle's theorem, and its proof depends on that theorem, applied to the new function,

$$f(x) - f(a) - \frac{x - a}{b - a}\{f(b) - f(a)\},$$

which gives the difference between the y values on the curve and on the chord. Differentiating this function, which vanishes at $x = a$ and $x = b$, and applying Rolle's theorem,

$$f'(\xi) - \frac{1}{b - a}\{f(b) - f(a)\} = 0,$$

i.e.
$$f(b) - f(a) = (b - a)f'(\xi),$$

where ξ lies between a and b. An alternative form, obtained by putting $b = a + h$, is

$$f(a + h) = f(a) + hf'(\xi).$$

The difficulty that ξ is not known can be turned to advantage.

Example 1. The theorem indicates the possible error in the approximate formula (where ξ is replaced by a),

$$f(a + h) = f(a) + h \,.\, f'(a),$$

given in Section 5.06. Since ξ lies between a and b, the value of $f'(\xi)$ lies between the greatest and least values of the derivative $f'(x)$ between these values of x. Hence the error in using the approximate formula is certainly no greater than h times the difference between these greatest and least values.

For example, in the calculation of sin 31° of that section the answer is not in error by more than

$$\frac{\pi}{180}\{\cos 30° - \cos 31°\}.$$

Example 2. Giving ξ an arbitrary value in the expression $f(a) + hf'(\xi)$ *may* give the right value of $f(a + h)$ but normally only gives an approximation. One such has already been considered, namely, $\xi = a$, because it is simple and convenient. It is shown in the examples that, in one sense, the best value is $\xi = a + \frac{1}{2}h$.

5.16. L'Hôpital's Rule for Limiting Values. This rule states that the limiting value of the ratio $f(x)/g(x)$ as $x \to a$, when both $f(x)$ and $g(x)$ are zero at $x = a$, is

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

From the Mean Value Theorem,

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{(x-a)(df/dx)_{\xi_1}}{(x-a)(dg/dx)_{\xi_2}},$$

where ξ_1 and ξ_2 certainly lie between a and x. Since $f(a)$ and $g(a)$ are zero,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{\substack{x\to a \\ \xi_1,\,\xi_2}} \frac{(df/dx)_{\xi_1}}{(dg/dx)_{\xi_2}} = \lim_{\xi\to a} \frac{(df/dx)_{\xi}}{(dg/dx)_{\xi}} = \lim_{x\to a} \frac{df/dx}{dg/dx},$$

the second step being allowed by Section 3.12. If the derivatives are not both zero at $x = a$, the limit is known at once. Otherwise the theorem is applied a second time.

Example 1. $\operatorname{Lim}_{x\to\infty} xe^{-x} = \lim_{x\to\infty} \dfrac{x}{e^x} = \lim_{x\to\infty} \dfrac{1}{e^x} = 0.$

Example 2. $\operatorname{Lim}_{x\to 0} \dfrac{1 - \cos x}{x^2} = \lim_{x\to 0} \dfrac{\sin x}{2x} = \lim_{x\to 0} \dfrac{\cos x}{2} = \frac{1}{2},$

applying the rule twice.

EXAMPLES V

1. If y increases as x decreases, what is the sign of the slope $\tan\psi$?

2. Sketch a curve with a minimum and a least value, the two being distinct.

3. A function y satisfies the equation $d^2y/dx^2 + f(x)y = 0$, where $f(x)$ is positive when $x < a$, and negative when $x > a$. Sketch y for positive values of x given that $y = 0$ when $x = 0$, that it is zero for one other value of x, and that $y \to 0$ as $x \to \infty$.

4. The curve $y = ax^3 + bx^2 + cx + d$ passes through the point (1, 1), touches the x axis at (2, 0) and has slope 4 at $x = 0$. Find a, b, c and d and prove that y has a maximum when $x = \frac{2}{3}$.

5. The rate of discharge Q from a circular pipe is

$$Q = KD^{5/2}\sqrt{\{(\theta - \sin\theta)^3/\theta\}},$$

where D is the diameter of the pipe, K is a constant and θ the angle subtended at the centre by the wetted perimeter. Show that, for maximum rate of discharge, θ satisfies the equation

$$\theta(3\cot\theta - 2\operatorname{cosec}\theta) = 1 \text{ or } 3\cos\theta - 2 = (\sin\theta/\theta),$$

and solve this equation approximately.

6. The velocity of capillary waves on relatively deep water is given by the formula

$$c^2 = g\lambda/2\pi + 2\pi T/\lambda\rho.$$

For what wavelength λ is this a minimum?

7. Van der Waals's equation for a gas is $(p + a/v^2)(v - b) = RT$, where a and b are constants. This is a cubic equation for v given p and T. Show that there are three equal roots v_c when $p = p_c$, $T = T_c$, where $a = 3p_cv_c^2$, $3b = v_c$ and $R = 8p_cv_c/3T_c$.

From the derivative dp/dv, show that the p/v curve at constant temperature T has two stationary values when $RT = 2a(v - b)^2/v^3$, and by plotting this function deduce that they only occur when $T < T_c = (8a/27Rb)$.

8. From the gas equation of Dieterici $p(v - b) = RTe^{-A/RTv}$ find the maxima and minima in the p/v curve, and show that they do not exist above a temperature $T_c = A/4bR$, the corresponding values at this limit being the critical values $v_c = 2b$ and $p_c = RT_c/e^2v_c$, where e is the natural number.

9. Using L'Hôpital's rule prove that, when $f(x)$ has two equal roots at $x = a$, then $f(x)/(x - a)^2$ tends to a finite limit at $x = a$.

Deduce that $y = f(x)$ touches the x axis at $x = a$.

10. Prove that, when h is a variable,

$$\lim_{h\to 0} \frac{f(a + h) - f(a) - hf'(a + \lambda h)}{h^2} = (\tfrac{1}{2} - \lambda)f''(a).$$

Deduce that the approximation $f(a + h) = f(a) + hf'(a + \lambda h)$ is correct to second order in the small quantity h when $\lambda = \frac{1}{2}$, provided $f(x)$ can be differentiated three times.

11. According to the elementary kinetic theory of gases, the root mean square molecular velocity of a gas is $v = (3p/\rho)^{1/2}$ at pressure p and density ρ. If p is accurately known, but ρ is known to 0·1 per cent, what is the relative error in v?

12. In the Birge–Sponer method the energy levels of a molecule are assumed to satisfy the equations

$$E_n = x - (x^2/a), \quad x = (n + \tfrac{1}{2})h\nu, \quad (n = 0, 1, 2 \ldots),$$

where E increases with x, the dissociation energy being the maximum value of E, namely $\frac{1}{4}a$.

If in fact $E_n = x - x^2/a + \beta x^3/a^2$, and the dissociation energy is the maximum value of E, find the value of β so that the new value does not differ by more than 10 per cent from the previous value. Also deduce the relative accuracy of measurements of E_n if the existence of the term β is to be detected for values of x equal to $\frac{1}{8}a$.

REFERENCE

The author has found a helpful treatment of the subject matter of this chapter in the following book:

LAMB, H. *Infinitesimal Calculus.* Cambridge University Press, London, 1951.

CHAPTER 6

RATIONAL FUNCTIONS

Linear functions—Quadratic functions—Polynomials—Remainder theorem—Factorizing polynomials—Identical polynomials—Graphs of simple powers—Comparison of powers—Graphs of polynomials—Approximations near roots—Simple extension—Binomial theorem—Rational functions—Graphs of rational functions—Method of partial fractions—Rules for partial fractions—Sums involving integers—Method of mathematical induction

6.01. LINEAR FUNCTIONS. y is a linear function of x if $y = ax + b$, where a and b are constants, and a is not zero.

Example 1. According to Charles's law the volume v of a gas at constant pressure varies linearly with temperature, i.e. $v = v_0(1 + \alpha t)$, where v_0 is the volume when $t = 0$ in °C, and α is the coefficient of expansion.

6.02. QUADRATIC FUNCTIONS. $y = ax^2 + bx + c$, by completing the square, can be written as a square plus a number,

$$y = (1/a)(a^2x^2 + abx + ac),$$
$$= (1/a)\{(ax + \tfrac{1}{2}b)^2 + (ac - \tfrac{1}{4}b^2)\}.$$

It is a complete square if $b^2 = 4ac$.

The quadratic equation $ax^2 + bx + c = 0$ has:

(*i*) one root $x = 0$, if $c = 0$;

(*ii*) equal roots $x = -(b/2a)$, when it is a complete square;

(*iii*) two unequal roots, when $b^2 > 4ac$, which are

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a} = \frac{-\tfrac{1}{2}b \pm \sqrt{\{(\tfrac{1}{2}b)^2 - ac\}}}{a},$$
$$= \frac{2c}{-b \pm \sqrt{(b^2 - 4ac)}}.$$

The form involving $\frac{1}{2}b$ is useful when b is even;

(*iv*) no real roots, when $b^2 < 4ac$, but two conjugate complex roots

$$x = \frac{-b \pm i\sqrt{(4ac - b^2)}}{2a},$$

where i can be put inside the root with $i^2 = -1$.

6.03. POLYNOMIALS. A polynomial in x is a sum of multiples of powers of x, the degree being the highest power. A quadratic function $ax^2 + bx + c$ is a polynomial of the second degree, a cubic function is a polynomial of the third degree, and so on. The function

$$P(x) = a_0 + a_1x + a_2x^2 + \ldots a_mx^m,$$

for any values of $a_0, a_1, a_2, \ldots a_m$ (except $a_m = 0$) is a polynomial of the mth degree.

6.04. REMAINDER THEOREM. If $P(x)$ is any polynomial in x, then $P(a)$ is the remainder when $P(x)$ is divided by $(x - a)$.

For by long division $P(x)/(x - a) = Q(x) + R/(x - a)$, where $Q(x)$ is the dividend and R the remainder. R is simply a number, independent of x. Hence $P(x) = Q(x)(x - a) + R$. Put $x = a$, then $P(a) = R$.

Corollary. When $P(a) = R = 0$, then $P(x) = Q(x)(x - a)$, so that $(x - a)$ is a factor of $P(x)$.

6.05. FACTORIZING POLYNOMIALS. A fundamental theorem of algebra states that the equation $a_0 + a_1x + \ldots a_mx^m = 0$ has exactly m roots, real or complex, and no more. If the coefficients $a_0 \ldots a_m$ are real, the complex roots occur in pairs of conjugates as in quadratic equations with complex roots. According to the Remainder Theorem it follows that every polynomial can be expressed as a product of factors. A real root $x = a$ implies a factor $(x - a)$ of the polynomial, and a pair of complex roots $b \pm ic$ corresponds to a quadratic factor $(x - b)^2 + c^2$.

Example 1.
$$x^2 - 1 = (x - 1)(x + 1),$$
$$x^3 - 1 = (x - 1)(x^2 + x + 1),$$
$$x^3 + 1 = (x + 1)(x^2 - x + 1),$$
$$x^4 - 1 = (x - 1)(x + 1)(x^2 + 1),$$
$$x^4 + 1 = (x^2 - x\sqrt{2} + 1)(x^2 + x\sqrt{2} + 1).$$

The Remainder Theorem can be used to factorize polynomials.

Example 2. $y = x^3 - 3x^2 + 2x - 6$ is zero when $x = 3$, so that $(x - 3)$ is a factor. By writing $y = (x - 3)(ax^2 + bx + c)$ and multiplying out it is easily seen that $a = 1$, $c = 2$ and finally $b = 0$. Hence $y = (x - 3)(x^2 + 2)$ and there is one linear factor.

Example 3. $y = x^3 - 3x + 2$ is zero when $x = 1$ and -2.

There must be a third factor, and examination of the terms in x^3 and the number $+ 2$, shows that this must be $(x - 1)$. Hence

$$x^3 - 3x + 2 = (x - 1)^2(x + 2).$$

Relations between the coefficients in the polynomials and its roots are found in the following manner. Let x_1 and x_2 be the roots of $ax^2 + bx + c = 0$. Then

$$\begin{aligned} ax^2 + bx + c &= a(x - x_1)(x - x_2), \\ &= ax^2 - a(x_1 + x_2)x + ax_1x_2. \end{aligned}$$

Hence (see next section),

$$x_1 + x_2 = -(b/a), \quad \text{and} \quad x_1x_2 = (c/a).$$

The product of the roots is c/a and their sum is $-(b/a)$. This is true even with complex roots.

6.06. IDENTICAL POLYNOMIALS. Two polynomials of the same degree n which are equal for $(n + 1)$ or more distinct values of x are identical, i.e. the coefficients of each power of x are the same.

This is so because their difference is a polynomial $P(x)$ of degree $m \leqslant n$. If $x_1, x_2, \ldots x_m$ are among the values of x which make the polynomials equal, then they make $P(x)$ zero. Hence

$$P(x) = k(x - x_1)(x - x_2) \ldots (x - x_m),$$

where k is the coefficient of the highest power of x in $P(x)$. But the polynomials are equal for at least one more value of x, and this value can only make $P(x) = 0$ if $k = 0$. Thus $P(x) = 0$ and the two polynomials are identical.

This result can be used to express polynomials in various forms.

Example 1. Express $x^2 + 3x + 4$ in the form $A(x - 1)(x + 3) + B(x - 1)^2 + C(x + 3)^2$.

Since these are to be equal for all values of x, they are equal when $x = 1$, i.e. $8 = 16C$, so that $C = \frac{1}{2}$. Similarly when $x = -3$, $4 = B(-4)^2$ or $B = \frac{1}{4}$.

To find A the coefficients of x^2 on both sides are equated. This gives $1 = A + B + C$, so that $A = \frac{1}{4}$. In all

$$x^2 + 3x + 4 = \tfrac{1}{4}(x - 1)(x + 3) + \tfrac{1}{4}(x - 1)^2 + \tfrac{1}{2}(x + 3)^2.$$

6.07. GRAPHS OF SIMPLE POWERS. $y = x^n$. Several of these are shown in *Figures 6.01* and *6.02*. It is very useful to be able to sketch

them quickly, especially the negative powers, and to notice such simple points as the following (it is assumed that x is positive):

(*i*) For positive powers x^n increases steadily with x, the curves for $n > 1$ being concave upwards, and those for $n < 1$ being convex upwards.

(*ii*) Only $y = x$ cuts the axis at an angle. Curves for $n > 1$ touch the x axis at the origin and those for $n < 1$ touch the y axis there. For $dy/dx = nx^{n-1}$ tends to zero at $x = 0$ when $n > 1$, but tends to infinity when $n < 1$.

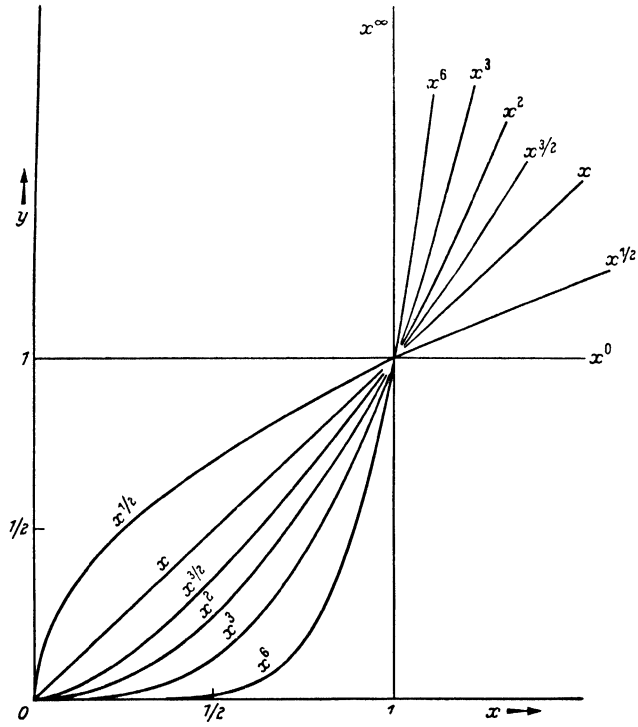

Figure 6.01

(*iii*) The curve for simple fractional powers, e.g. $y = x^{1/2}$, are obtained from those for integral powers, e.g. $y = x^2$, simply by exchanging x and y.

(*iv*) When x is large, $y = 1/x^3$ is smaller than $y = 1/x^2$ which has the smaller power in the denominator, but when x is very small $y = 1/x^3$ is the larger.

The values of these powers for negative values of x are easily deduced, except that for some powers they do not exist, i.e. for $x^{m/n}$, where m and n are integers, and n is even.

The curves,

$$y = mx^n,$$

are similar in shape to those given above except that the scale of y is changed.

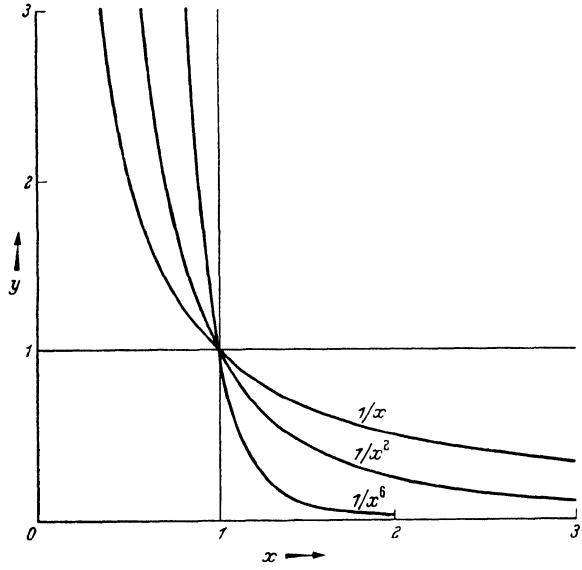

Figure 6.02

The curve,

$$y = b + m(x - a)^n,$$

is obtained by moving the curve $y = mx^n$ bodily so that the origin is now at the point $x = a$, $y = b$. Hence it is said to have the same shape near the point (a, b) as $y = x^n$ has near the origin.

6.08. COMPARISON OF POWERS. Consider two functions $y = ax^m$ and $y = bx^n$.

If $n > m$, the second is larger than the first when x is large enough, since

$$\frac{bx^n}{ax^m} = \frac{b}{a}x^{n-m} \to \infty \text{ as } x \to \infty.$$

On the other hand, the first is the larger when x is small enough, since

$$\frac{bx^n}{ax^m} = \frac{b}{a}x^{n-m} \to 0 \text{ as } x \to 0.$$

This is true whatever the numerical values of a and b.

Example 1. $y = a/x^3 - b/x^5$. When x is very large the first term is the more important, so that $y = a/x^3$ approximately.

When x is small enough the second term is the larger and $y = -b/x^5$ approximately, since $x^5 \ll x^3$, when x is small. The complete function changes sign once and is zero when $x^2 = (b/a)$.

6.09. GRAPHS OF POLYNOMIALS.

$$y = P(x) = a + bx + cx^2 + \ldots + kx^n.$$

These are easily drawn:

(*i*) find and plot the roots of $P(x) = 0$, where $y = 0$. This may involve numerical work;

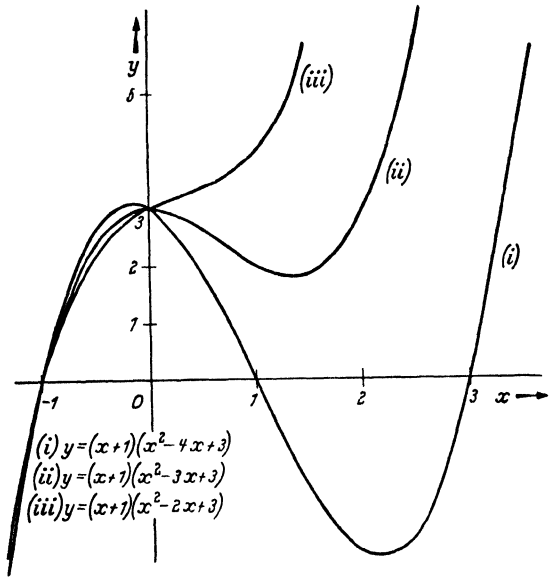

Figure 6.03

(*ii*) find the sign of y (+ or −) between these roots;

(*iii*) draw a continuous curve to cut the x axis only at the roots, with the correct sign between the roots, and going to

infinity with x. Since y is continuous it can only change sign by the curve crossing the x axis;

(*iv*) it helps to note that y changes sign across a single or triple root of $P(x) = 0$, but retains its sign across a double or quadruple root;

(*v*) if n is even, $y \to \infty$ as $x \to \pm \infty$. If n is odd, $y \to \infty$ as $x \to \infty$ and $y \to -\infty$ as $x \to -\infty$.

In sketching these curves no more bends should be introduced than are absolutely necessary.

(*vi*) A bend due to a maximum and minimum without the curve crossing the x axis is detected by examining the derivative dy/dx.

(*vii*) Similar bends without a maximum or minimum can occur with a quadratic factor in the polynomial with complex roots. To illustrate this we have drawn in *Figure 6.03* some curves of functions which differ by the variation of one parameter.

6.10. APPROXIMATIONS NEAR ROOTS. The remark (*iv*) of the previous section arises from the result: when $x = a$ is a simple root of $P(x) = 0$ then

$$y = P(x) = (x - a)Q(x),$$

where $Q(x)$ is a polynomial not zero when $x = a$. When $(x - a)$ is small, $Q(x) = Q(a) +$ small correction and

$$y = (x - a)Q(a) \text{ approximately.}$$

Hence y changes sign as x passes through the value $x = a$.

Similarly near a double root, $x = b$, say,

$$\begin{aligned} y = P(x) &= (x - b)^2 R(x), \\ &= (x - b)^2 R(b) \text{ approximately,} \end{aligned}$$

when $(x - b)$ is small, so that y does not change sign as x passes through the value b.

6.11. SIMPLE EXTENSION. The graph of $y^2 = f(x)$ can be derived from the values of $f(x)$ given by its graph. The values of y, where

$$y = \pm \{f(x)\}^{1/2},$$

are obtained by imagining what happens to $f(x)$ on taking a square root and remembering that the two signs $(\pm)$ correspond to two points on the curve, one above and one below the x axis. Where $f(x)$ is negative, y has no real values.

The shape of the curve near the roots of $f(x)$ often causes trouble, but there is no need for this if the approximations of the previous section are used, and a continuous curve is drawn (with no unnecessary bends!).

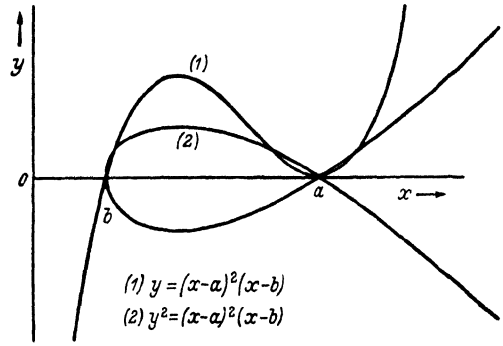

Figure 6.04

Example 1. In *Figure 6.04* we have drawn $f(x) = (x - a)^2(x - b)$, where a is greater than b, together with the graph of $y^2 = (x - a)^2(x - b)$.

Near $x = a$, according to the previous section,

$y^2 = (x - a)^2(a - b)$ or $y = \pm (x - a)\sqrt{(a - b)}$ approximately,

giving two straight lines crossing the x axis as shown.

Near $x = b$, $y^2 = (b - a)^2(x - b)$ approximately. The curve does not exist for $x < b$ because the RHS is then negative; near $x = b$ it has the same shape as $y^2 = x$ near the origin, as discussed in Section 6.08 (see *Figure 6.01*).

6.12. Binomial Theorem. This well known theorem states that, when n is a positive integer,

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1 \cdot 2} a^{n-2}b^2 + \ldots + b^n,$$

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{1 \cdot 2} x^2 + \ldots + x^n.$$

The second form is obtained from the first by writing $a = 1$ and $b = x$. The coefficient of any power x^r (or $a^{n-r}b^r$) is

$$\frac{n \cdot (n-1) \ldots (n-r+1)}{1 \cdot 2 \cdot 3 \ldots r}.$$

It has just r integers above and below.

To prove this, note that

$$(a + b)^n = (a + b)(a + b)(a + b) \ldots n \text{ factors},$$

and when multiplied out consists of a lot of terms obtained by either taking an a or b from each factor. The coefficient of $a^{n-r}b^r$ is therefore the number of ways of taking a from $(n - r)$ factors and b from the remaining r factors, which is ${}^nC_r$ in the usual notation, or the expression given above. It is also denoted by $\binom{n}{r}$ and satisfies the relation

$$\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}.$$

Example 1. The biggest term in $(1 + x)^n$ is the first term, when $x < 1/n$. Otherwise it is the first term which is greater than the next. Now the term in x^{r+1} is the term in x^r multiplied by $(n - r)/(r + 1) \cdot x$ and is the smaller when $(n - r)x < r + 1$, i.e. $r > \alpha = (nx - 1)/(x + 1)$. Hence the value of r for the greatest term is the integer between α and $\alpha + 1$. This result is used in probability theory and also in statistical mechanics, where n is very large and x small in such a way that $nx \gg 1$. Then $\alpha = nx$.

Example 2. In the expansion of $(a + b)^n$ the quantities a and b can be replaced by any expressions, e.g. $(1 + \sin x)^n$, $(1 + x^2)^n$, $\{1 - (x - x^2)\}^3$ and so on.

6.13. RATIONAL FUNCTIONS. A rational function of x is constructed by multiplication and division and addition of polynomials in x. The result simplifies to either a polynomial in x or a ratio of two polynomials.

For rational functions such as $(2 + x^2)/(1 + x)$, where the degree of the numerator is higher than that of the denominator, a process of long division gives a polynomial and a *proper* algebraic fraction, in which the degree of the numerator is less than that of the denominator, e.g.

$$(2 + x^2)/(1 + x) = (x - 1) + 3/(1 + x).$$

6.14. GRAPHS OF RATIONAL FUNCTIONS. If they are not polynomials these functions can be written as a single fraction. To sketch such a function y, remember:

(*i*) that $y = 0$ when the numerator is zero;

(*ii*) that $y \to \infty$ when the denominator is zero, or when $x \to \infty$, if the numerator has the higher degree;

(*iii*) to find y when $x \to \infty$, and its sign;
(*iv*) to find the sign of y between values of x and y derived from (*i*) and (*ii*);
(*v*) to find dy/dx where necessary, and the stationary points where $dy/dx = 0$.

The data from (*i*) to (*iv*) above are usually sufficient to sketch the curve, but (*v*) is sometimes helpful to determine difficult features. It is also occasionally useful to find the points of inflection $d^2y/dx^2 = 0$.

Example 1. $y = ax/(x^2 + b^2)$:

(*i*) $y = 0$ only when $x = 0$;
(*ii*) $y \to \infty$ nowhere;
(*iii*) as $x \to \infty$, $y \to (ax/x^2) = a/x \to 0$ and y is positive;
(*iv*) $y < 0$ when $x < 0$; $x > 0$ implies $y > 0$.

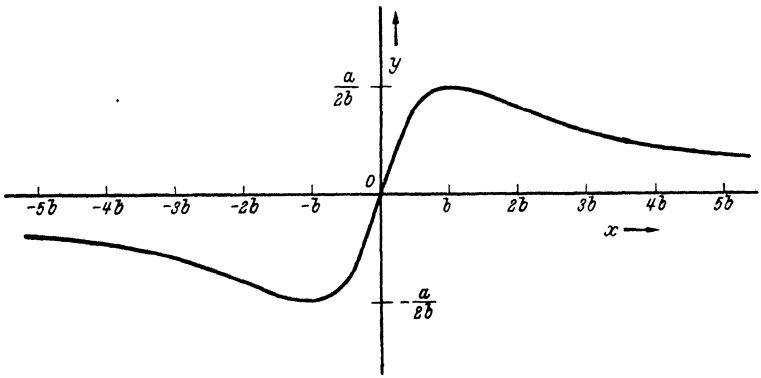

Figure 6.05

The curve based on these results is shown in *Figure 6.05.* It clearly has a maximum and a minimum, which occur at $x = b$ and $x = -b$ (cf. *Example 1*, Section 5.09).

6.15. Method of Partial Fractions. The algebraic analogue of the arithmetical equation, $2/35 = 1/5 - 1/7$, is to express a *proper* rational fraction as a sum of terms the denominators of which are factors of the original denominator.

The following examples show how this is done for such fractions. Of the various methods described, *Method I* can always be used, but the others speed up at least part of the calculation. The main

problem is to factorize the denominator, and then find the numerators of the various terms. In the examples it is always assumed that the factors are simple, to avoid a great deal of the numerical work, but this is by no means true in practice.

Method I. Find A and B if

$$\frac{4x - 11}{x^2 - 5x + 6} = \frac{4x - 11}{(x - 2)(x - 3)} = \frac{A}{x - 2} + \frac{B}{x - 3}.$$

Multiply both sides by the whole denominator $(x - 2)(x - 3)$,

$$4x - 11 = A(x - 3) + B(x - 2).$$

This equation is true for all values of x. Hence the coefficients of each power of x on the two sides are equal, i.e.

$$\begin{aligned} 4 &= A + B, \\ -11 &= -3A - 2B, \end{aligned} \qquad \text{i.e. } A = 3,\ B = 1.$$

$$\therefore \qquad \frac{4x - 11}{x^2 - 5x + 6} = \frac{3}{x - 2} + \frac{1}{x - 3}.$$

This general method can be used to prove that the numbers $A, B, \ldots$ inserted according to the rules given below are determined uniquely. The general proof is lengthy and full of symbols, so we prefer to give examples to illustrate the method.

Method II. For single linear factors

$$\frac{4x - 11}{(x - 2)(x - 3)} = \frac{A}{x - 2} + \frac{B}{x - 3}.$$

Multiply by $(x - 2)$: $\dfrac{4x - 11}{x - 3} = A + B\dfrac{x - 2}{x - 3}.$

Put $x = 2$: $\qquad \dfrac{8 - 11}{2 - 3} = A, \quad \text{i.e. } A = 3.$

Multiply by $(x - 3)$: $\dfrac{4x - 11}{x - 2} = A\dfrac{x - 3}{x - 2} + B.$

Put $x = 3$: $\qquad \dfrac{12 - 11}{3 - 2} = B, \quad \text{i.e. } B = 1.$

$$\therefore \qquad \frac{4x - 11}{(x - 2)(x - 3)} = \frac{3}{x - 2} + \frac{1}{x - 3}.$$

Method III. A combination of *I* and *II*, with a quadratic factor.

$$\frac{3(1+x)}{x^3-1} = \frac{3(1+x)}{(x-1)(x^2+x+1)} = \frac{A}{x-1} + \frac{Bx+C}{x^2+x+1}.$$

Multiply by (x^3-1):

$$3(1+x) = A(x^2+x+1) + (Bx+C)(x-1).$$

To find A, put $x=1$: $6 = A(1+1+1)$, i.e. $A=2$.
To find B, use coefficients of x^2: $0 = A+B$, i.e. $B=-A=-2$.
To find C, equate numbers: $3 = A-C$, i.e. $C=A-3=-1$.
The coefficients of x on both sides should also be equal,

$$3 = A - B + C.$$

We have not used this equation to find C, but it could be so used. As A, B and C have been determined, it can be used as a check on the arithmetic: thus $A-B+C = 2-(-2)+(-1) = 3$.

Method IV. For multiple factors. As always, factorize the denominator first.

$$\frac{5x+1}{x^3-3x+2} = \frac{5x+1}{(x-1)^2(x+2)} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{C}{x+2}.$$

Multiply by $(x-1)^2(x+2)$:

$$5x+1 = A(x-1)(x+2) + B(x+2) + C(x-1)^2.$$

Find C by putting $x=-2$: $-10+1 = C(-3)^2$, i.e. $C=-1$.
Find B by putting $x=1$: $5+1 = B(1+2)$, i.e. $B=2$.
Equate coefficients of x^2: $0 = A+C$, i.e. $A=-C=1$.
As a check, the numbers on the RHS are

$$-2A+2B+C = -2+4-1 = 1,$$

which is correct.

Method V. Occasionally useful, this method uses symmetry of the functions. According to the rules (below),

$$\frac{1}{(x^2+1)(x^2+4)} = \frac{Ax+B}{x^2+1} + \frac{Cx+D}{x^2+4}.$$

Changing x to $-x$ on both sides (this leaves x^2 unchanged):

$$\frac{1}{(x^2+1)(x^2+4)} = \frac{-Ax+B}{x^2+1} + \frac{-Cx+D}{x^2+4}.$$

These two expressions are compatible only when $A = 0$ and $C = 0$. Dropping A and C we proceed:

$$\frac{1}{(x^2+1)(x^2+4)} = \frac{B}{x^2+1} + \frac{D}{x^2+4}.$$

Multiply by (x^2+1): $\dfrac{1}{x^2+4} = B + D\dfrac{x^2+1}{x^2+4}$.

Put $x^2 = -1$ (!): $\frac{1}{3} = B$, i.e. $B = \frac{1}{3}$.

Similarly, multiplying by (x^2+4) and putting $x^2 = -4$ gives $D = -\frac{1}{3}$.

$$\therefore \quad \frac{1}{(x^2+1)(x^2+4)} = \frac{1}{3}\left\{\frac{1}{x^2+1} - \frac{1}{x^2+4}\right\}.$$

In general, however, use *Method III* with quadratic factors.

Method VI. In this method, quadratic factors are factorized using conjugate complex numbers (see Section 14.02).

$$\frac{1}{(x-1)(x^2+1)} = \frac{1}{(x-1)(x-i)(x+i)}, \ (i^2 = -1),$$
$$= \frac{A}{x-1} + \frac{B}{x-i} + \frac{C}{x+i}.$$

Multiply by $(x-1)$ and put $x = 1$: $A = \dfrac{1}{(1+i)(1-i)} = \dfrac{1}{2}$.

Multiply by $(x-i)$ and put $x = i$:

$$B = \frac{1}{(i-1)(2i)} = \frac{1}{4}(-1+i);$$

and similarly $\quad C = \frac{1}{4}(-1-i)$.

$$\therefore \quad \frac{1}{(x-1)(x^2+1)} = \frac{\frac{1}{2}}{x-1} + \frac{1}{4}\left\{\frac{-1+i}{x-i} + \frac{-1-i}{x+i}\right\}$$
$$= \frac{1}{2(x-1)} - \frac{x+1}{2(x^2+1)}.$$

6.16. RULES FOR PARTIAL FRACTIONS. In these examples the following rules are used:

(*i*) in each partial fraction the degree of the numerator is assumed to be *one* lower than that of the factor in the denominator, e.g. $\dfrac{A}{x-\beta}$, $\quad \dfrac{Ax+B}{x^2+x+1}$, $\quad \dfrac{A}{(x-1)^2}$;

(*ii*) there is one term for each linear or quadratic factor;

(*iii*) a multiple factor in the denominator has a number of terms, one for every power of that factor which can occur, e.g. a factor $(x-1)^2$ in the denominator gives the two terms $A/(x-1)+B/(x-1)^2$.

These rules ensure that the number of unknowns A, B, . . . is initially just one greater than the degree of the original denominator, which is the condition that these unknowns are determined uniquely.

6.17. Sums Involving Integers. The following results are useful:

$$1+2+3+\ldots+n = \tfrac{1}{2}n(n+1);$$
$$1^2+2^2+3^2+\ldots+n^2 = \tfrac{1}{6}n(n+1)(2n+1);$$
$$1^3+2^3+3^3+\ldots+n^3 = \tfrac{1}{4}n^2(n+1)^2;$$
$$1^4+2^4+3^4+\ldots+n^4 = \tfrac{1}{30}n(n+1)(2n+1)(3n^2+3n-1).$$

6.18. Method of Mathematical Induction. The formulae of the previous section are easily proved for $n = 1, 2, \ldots$. To prove them for large values of n, a step-by-step method can be used, known as the method of mathematical induction. Assuming a formula is true for $n = 1$, a method is devised to show that it is true for $n = 2$. Then knowing it to be true for $n = 2$, the method is used to prove it true when $n = 3$, and so on.

The practical objection, that the time taken to prove in this way that the formula is true for, say, $n = 1000$ is excessive, can be avoided quite simply. If all the steps are of the same type, they can all be carried through in one operation, if the method can be used to prove that the formula, being true for $n = m$, is also true for $n = m + 1$, whatever the value of m. If the formula is in fact true for $n = 1$, then it is true for $n = 2, 3, \ldots$ and as large as we please.

Expressed in symbols we assert: if $f(n)$ and $F(n)$ are two functions of n which are certainly equal when $n = 1$, then they are equal for all integral values of n if, assuming $f(m) = F(m)$, it can be proved from their definitions that $f(m+1) = F(m+1)$.

Example 1. Prove that

$$1^2+2^2+3^2+\ldots+n^2 = \tfrac{1}{6}n(n+1)(2n+1).$$

Let $f(n) = 1^2+2^2+\ldots+n^2$ and $F(n) = \frac{1}{6}n(n+1)(2n+1)$.

They are equal when $n = 1$. Assume that they are equal when $n = m$. Then from the definitions,

$$\begin{aligned} F(m+1) &= \tfrac{1}{6}(m+1)(m+2)(2m+3), \\ \text{and} \quad f(m+1) &= 1^2 + 2^2 + 3^2 + \ldots + m^2 + (m+1)^2, \\ &= f(m) + (m+1)^2, \\ &= F(m) + (m+1)^2 \quad \text{(cf. assumption)}, \\ &= \tfrac{1}{6}m(m+1)(2m+1) + (m+1)^2, \\ &= \tfrac{1}{6}(m+1)\{(2m^2+m) + 6(m+1)\}, \\ &= \tfrac{1}{6}(m+1)(2m^2+7m+6), \\ &= \tfrac{1}{6}(m+1)(2m+2)(2m+3) = F(m+1). \quad \text{Q.E.D.} \end{aligned}$$

Example 2. Prove that

$$\tfrac{1}{2} + \cos\phi + \cos 2\phi + \ldots + \cos n\phi = \sin(n + \tfrac{1}{2})\phi/(2\sin\tfrac{1}{2}\phi).$$

This equation is true when $n = 0$.

Let $f(n) =$ LHS and $F(n) =$ RHS and assume $f(m) = F(m)$.

Then

$$\begin{aligned} f(m+1) &= f(m) + \cos(m+1)\phi = F(m) + \cos(m+1)\phi, \\ &= \sin(m+\tfrac{1}{2})\phi/(2\sin\tfrac{1}{2}\phi) + \cos(m+1)\phi, \\ &= \{\sin(m+\tfrac{1}{2})\phi + 2\cos(m+1)\phi\sin\tfrac{1}{2}\phi\}/2\sin\tfrac{1}{2}\phi, \\ &= \{\sin(m+\tfrac{1}{2})\phi + \sin(m+\tfrac{3}{2})\phi - \sin(m+\tfrac{1}{2})\phi\}/2\sin\tfrac{1}{2}\phi, \\ &\qquad \text{(cf. Section 7.02.)} \\ &= \sin(m+\tfrac{3}{2})\phi/(2\sin\tfrac{1}{2}\phi) = F(m+1). \end{aligned}$$

N.B. It should be mentioned that this is neither the only nor the best method of proving the results in these examples.

EXAMPLES VI

1. Sketch the following functions over the ranges of values of x indicated for each: (*i*) $y = ax^{4/3}$ $(-\infty < x < \infty)$; (*ii*) $y = b(x-a)^{3/2}$ $(x > a)$; (*iii*) $y = b/(x-a)^{3/2}$ $(x > a)$; (*iv*) $y = 1 - a/(x-a)$ $(a > 0, 0 < x < \infty)$.

2. The efficiency of a screw is $E = t(1 - \mu t)/(\mu + t)$ where t is the pitch of the screw and μ the coefficient of friction. Plot E between $t = 0$ and $t = 1/\mu$ and find t for maximum efficiency.

3. The equilibrium constant of the reaction $PCl_5 \rightleftharpoons PCl_3 + Cl_2$ is $K = x^2/\{(a-x)V\}$, where x is the amount of PCl_3 and Cl_2, a the original amount of PCl_5 and V the volume. Plot V against x for $0 \leqslant x < a$.

4. Starting with 1 mole of ammonia, the reaction $2NH_3 \rightleftharpoons N_2 + 3H_2$

proceeds at pressure P until there are $\frac{1}{2}x$ moles of nitrogen and $\frac{3}{2}x$ moles of hydrogen, where

$$P^2 = 16(1 - x)^2(1 + x)^2/(27Kx^4),$$

and K is the reaction constant. Plot P against x between $x = 0$ and $x = 1$.

5. Sketch the curve known as the cissoid, $y^2(2a - x) = x^3$. Where does it cut a circle of radius a having its centre at the point $x = a, y = 0$?

6. Sketch the strophoid $y^2 = x^2(a + x)/(a - x)$, and the parabola $x^{1/2} + y^{1/2} = a^{1/2}$.

7. The equilibrium constant for a reaction $A + B \rightleftharpoons C + D$ is

$$K = x^2/(a - x)(b - x).$$

Plot K against x over a suitable range, assuming that $a < b$.

8. Sketch the three curves:

$$y = ax/(a^2 + x^2), \quad y = x^2/(a^2 + x^2), \quad y = x/(a + x).$$

9. Express $n^3 + 6n^2 + 15n + 15$ in the form

$$An(n - 1)(n - 2) + Bn(n - 1) + Cn + D,$$

using the Remainder Theorem as much as possible.

Find a general method of expressing a polynomial in n of the mth degree, in the form

$$P(n) = n!\left\{\frac{A}{(n - m)!} + \frac{B}{(n - m + 1)!} + \ldots + \frac{M}{n!}\right\}.$$

10. Express as partial fractions: (*i*) $\dfrac{4x - 2}{x^3 - x^2 - 2x}$, (*ii*) $\dfrac{x - 3}{x^3 - x^2}$, (*iii*) $\dfrac{x^2}{(1 - x)^3}$, (*iv*) $\dfrac{1}{x^2 + x^4}$, (*v*) $\dfrac{x^2 + x}{(x - 1)(x^2 + 1)}$, (*vi*) $\dfrac{x(x^2 + 3)}{(x^2 + 1)^2}$.

Chapter 7

CIRCULAR FUNCTIONS

Geometrical definitions—Identities—Graphs of functions—Finding values—Relations in a triangle—Solution of triangles—Solution of equations—Derivatives—Inverse sine function—Other inverse functions—Addition of inverse functions

7.01. Geometrical Definition. The circular functions of an acute angle θ are defined in terms of a right-angled triangle with one angle θ (*Figure 7.01*). There are other definitions not depending on geometry but using the differential equation, $d^2y/dx^2 = -y$, or integrals. Their properties can also be easily obtained using de Moivre's theorem (Chapter 14).

In general, if P is a point on a circle of radius r with its centre at the origin of Cartesian coordinates, then its coordinates are

$$x = r\cos\theta, \quad y = r\sin\theta.$$

Thus x and $\cos\theta$ are negative together, and so are y and $\sin\theta$. The angle θ is measured anticlockwise from OX. As P moves round the

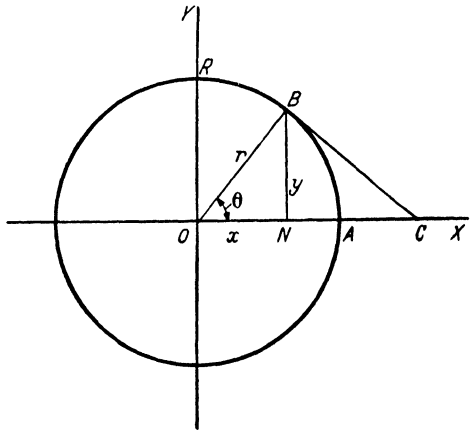

Figure 7.01

circle anticlockwise, θ increases. Angles larger than 360°, or 2π radians, are obtained when P has traversed the circle once and begins again. Negative angles are measured in the opposite sense.

The other circular, or trigonometric, functions are

$$\tan\theta = \frac{1}{\cot\theta} = \frac{\sin\theta}{\cos\theta}, \qquad \sec\theta = \frac{1}{\cos\theta}, \qquad \operatorname{cosec}\theta = \frac{1}{\sin\theta}.$$

Occasionally the following are met (e.g. in astronomy)

$$\text{versine } \theta = \text{vers } \theta = 2\sin^2\tfrac{1}{2}\theta, \quad \text{haversine } \theta = \text{hav } \theta = \sin^2\tfrac{1}{2}\theta.$$

The trigonometric (or circular) functions of the angles θ, $\theta \pm 2\pi$, $\theta \pm 2n\pi$ (n any integer) are equal because they are all represented by the same point P on the circle.

$\sin\theta$ and $\tan\theta$ are odd functions of θ, and $\cos\theta$, $\sec\theta$ are even functions. Furthermore:

$$\sin(-\theta) = -\sin\theta, \qquad \cos(-\theta) = \cos\theta,$$
$$\tan(-\theta) = -\tan\theta.$$

$$\sin(\theta + \tfrac{1}{2}\pi) = \cos\theta, \qquad \cos(\theta + \tfrac{1}{2}\pi) = -\sin\theta,$$
$$\tan(\theta + \tfrac{1}{2}\pi) = -\cot\theta.$$

$$\sin(\theta + \pi) = -\sin\theta, \qquad \cos(\theta + \pi) = -\cos\theta,$$
$$\tan(\theta + \pi) = \tan\theta.$$

$$\sin(\pi - \theta) = \sin\theta, \qquad \cos(\pi - \theta) = -\cos\theta,$$
$$\tan(\pi - \theta) = -\tan\theta.$$

7.02. IDENTITIES. The last-mentioned equations are special cases of the following identities which are true for all values of A, B and θ.

$$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B.$$
$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B. \quad (N.B.\ \text{signs.})$$
$$\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A\tan B}.$$

$$\sin A\cos B = \tfrac{1}{2}\{\sin(A + B) + \sin(A - B)\}.$$
$$\cos A\sin B = \tfrac{1}{2}\{\sin(A + B) - \sin(A - B)\}.$$
$$\cos A\cos B = \tfrac{1}{2}\{\cos(A + B) + \cos(A - B)\}.$$
$$\sin A\sin B = \tfrac{1}{2}\{\cos(A - B) - \cos(A + B)\}.$$

$$\sin A + \sin B = 2\sin\tfrac{1}{2}(A + B)\cos\tfrac{1}{2}(A - B).$$
$$\sin A - \sin B = 2\sin\tfrac{1}{2}(A - B)\cos\tfrac{1}{2}(A + B).$$
$$\cos A + \cos B = 2\cos\tfrac{1}{2}(A + B)\cos\tfrac{1}{2}(A - B).$$
$$\cos A - \cos B = -2\sin\tfrac{1}{2}(A + B)\sin\tfrac{1}{2}(A - B).$$

$$\sin 2\theta = 2 \sin \theta \cos \theta.$$

$$\sin^2 \theta + \cos^2 \theta = 1. \qquad 2 \cos^2 \theta - 1 = \cos 2\theta.$$

$$\cos^2 \theta - \sin^2 \theta = \cos 2\theta. \qquad 1 - 2 \sin^2 \theta = \cos 2\theta.$$

$$\sec^2 \theta = 1 + \tan^2 \theta. \qquad \operatorname{cosec}^2 \theta = 1 + \cot^2 \theta.$$

$$\tan 2\theta = 2 \tan \theta/(1 - \tan^2 \theta).$$

$$\sin \theta = 2 \sin \tfrac{1}{2}\theta \cos \tfrac{1}{2}\theta = (2 \tan \tfrac{1}{2}\theta)/(1 + \tan^2 \tfrac{1}{2}\theta).$$

$$\cos \theta = \cos^2 \tfrac{1}{2}\theta - \sin^2 \tfrac{1}{2}\theta = (1 - \tan^2 \tfrac{1}{2}\theta)/(1 + \tan^2 \tfrac{1}{2}\theta).$$

$$2 \cos^2 \tfrac{1}{2}\theta = 1 + \cos \theta. \qquad \cos \theta = \{\tfrac{1}{2}(1 + \cos 2\theta)\}^{1/2}.$$

$$\operatorname{vers} \theta = 2 \sin^2 \tfrac{1}{2}\theta = 1 - \cos \theta. \qquad \sin \theta = \{\tfrac{1}{2}(1 - \cos 2\theta)\}^{1/2}.$$

$$\tan^2 \tfrac{1}{2}\theta = (1 - \cos \theta)/(1 + \cos \theta).$$

$$\tan \theta = \{(1 - \cos 2\theta)/(1 + \cos 2\theta)\}^{1/2}.$$

$$4 \sin^3 \theta = 3 \sin \theta - \sin 3\theta. \qquad 4 \cos^3 \theta = 3 \cos \theta + \cos 3\theta.$$

$$4 \sin^2 \theta \cos \theta = \cos \theta - \cos 3\theta. \qquad 4 \cos^2 0 \sin \theta = \sin \theta + \sin 3\theta.$$

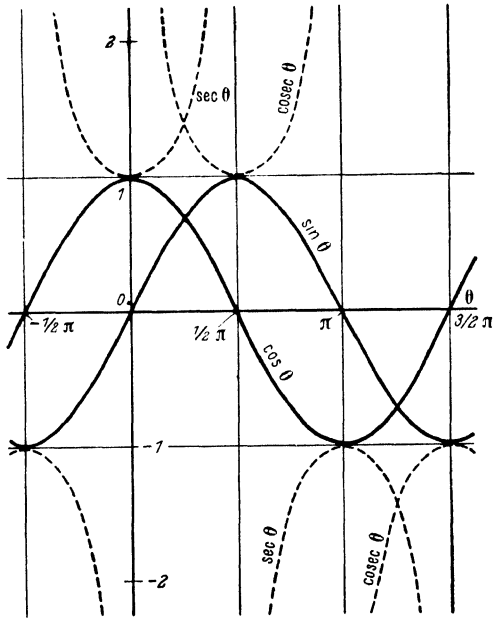

Figure 7.02

7.03. GRAPHS OF FUNCTIONS. See *Figures 7.02* and *7.03*. These functions are all periodic functions with a period of 2π, i.e. they are functions $f(\theta)$ such that

$$f(\theta + 2\pi) = f(\theta).$$

The graphs are only plotted, therefore, from $\theta = -\frac{1}{2}\pi$ to $\theta = \frac{3}{2}\pi$. In addition $\tan\theta$ is periodic over an interval π, i.e.

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

Numerically, $\sin\theta$ and $\cos\theta$ have values between ± 1, whereas $\sec\theta$ and $\operatorname{cosec}\theta$ exceed the value 1 numerically, and $\tan\theta$ and $\cot\theta$ have all values.

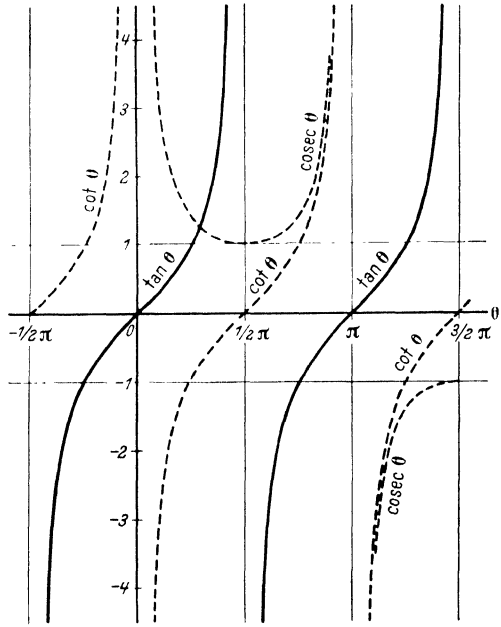

Figure 7.03

7.04. FINDING VALUES. The simplest problem of all, to find $\cos\theta$ or any other function, given the value of $\sin\theta$, is best solved by drawing a right-angled triangle as in *Figure 7.01* and inserting suitable values for the sides. This is easier than using a formula such as $\cos\theta = \pm(1 - \sin^2\theta)^{1/2}$.

7.05. RELATIONS IN A TRIANGLE. Sides a, b, c and angles A, B, C.

Sum of interior angles $A + B + C = 180°$.

Perimeter $(2s)$ $\qquad 2s = a + b + c.$

Sine law $\qquad \dfrac{\sin A}{a} = \dfrac{\sin B}{b} = \dfrac{\sin C}{c}.$

Cosine law

$$a^2 = b^2 + c^2 - 2bc \cos A \quad \text{or} \quad \cos A = (b^2 + c^2 - a^2)/2bc.$$

Area

$$\Delta = \tfrac{1}{2}bc \sin A = \tfrac{1}{2}a^2 \frac{\sin B \sin C}{\sin (B + C)} = \{s(s-a)(s-b)(s-c)\}^{1/2}.$$

Radius of inscribed circle

$$r = \Delta/s = (s - a) \tan \tfrac{1}{2}A.$$

Radius of circumscribing circle

$$R = abc/4\Delta.$$

$$\sin \tfrac{1}{2}A = \left\{\frac{(s-b)(s-c)}{bc}\right\}^{1/2}, \quad \cos \tfrac{1}{2}A = \left\{\frac{s(s-a)}{bc}\right\}^{1/2},$$

$$\text{hav } A = (s-b)(s-c)/bc, \qquad \frac{a-b}{a+b} = \frac{\tan \tfrac{1}{2}(A-B)}{\tan \tfrac{1}{2}(A+B)}.$$

7.06. SOLUTION OF TRIANGLES. Right-angled triangles are solved by inspection (*Figure 7.01*). A condition for a right-angle at A is $a^2 = b^2 + c^2$.

Other triangles are solved as follows:

I. Given one side and two angles, e.g. a, B, C. The procedure is to find $A = 180° - (B + C)$; then $b = a \sin B/\sin A$, $c = a \sin C/\sin A$.

II. Given two sides and the included angle, e.g. b, c, A.

Find B and C from $\tfrac{1}{2}(B + C) = 90° - \tfrac{1}{2}A$,

and $\tan \tfrac{1}{2}(B - C) = \tan \tfrac{1}{2}(B + C) \cdot (b - c)/(b + c)$.

Then find $a = b \sin A/\sin B$.

Alternatively, $a^2 = b^2 + c^2 - 2bc \cos A$, then $\sin B = b \sin A/a$, $C = 180° - (A + B)$.

III. Given three sides: a, b, c. (*i*) Calculate r and s (above). Then $\tan \tfrac{1}{2}A = r/(s - a)$ and similarly for B and C, or use the formula

$$\tan \tfrac{1}{2}A = \sqrt{[\{(s-b)(s-c)\}/\{s(s-a)\}]}.$$

(*ii*) Calculate the largest angle using the cosine law.

7.07. SOLUTION OF EQUATIONS. (*i*) Given both $\sin \theta = s$ and $\cos \theta = c$, where $s^2 + c^2 = 1$, there is only one solution, $\theta = \alpha$, between 0 and 2π. Other solutions are

$$\theta = \alpha + 2n\pi, \ (n \text{ any integer}).$$

(*ii*) The equation $\sin\theta = s$, where $|s| \leqslant 1$, has an infinity of solutions $\alpha + 2n\pi$, where α is any solution between 0 and 2π. In fact there are two solutions between 0 and 2π: if α is one the other is $(\pi - \alpha)$, since $\sin(\pi - \alpha) = \sin\alpha = s$. Altogether the solutions are

$$\theta = \alpha + 2n\pi \quad \text{and} \quad \theta = (2n+1)\pi - \alpha.$$

The equation $\cos\theta = c$ has solutions

$$\theta = \pm\alpha + 2n\pi,$$

where α is one solution.

The equation $\tan\theta = t$ has solutions

$$\theta = \alpha + n\pi,$$

where $\tan\alpha = t$.

(*iii*) Given any two numbers a and b, positive or negative, there is an angle β such that

$$\cos\beta = a/(a^2 + b^2)^{1/2}, \quad \sin\beta = b/(a^2 + b^2)^{1/2}.$$

(*iv*) The equation $a\cos\theta + b\sin\theta = c$ has solutions if $c^2 \leqslant (a^2 + b^2)$. Using (*iii*) it can be written

$$(a^2 + b^2)^{1/2}(\cos\theta\cos\beta + \sin\theta\sin\beta) = c,$$

i.e.
$$\cos(\theta - \beta) = c/(a^2 + b^2)^{1/2}.$$

(*v*) To solve such an equation as $\sin 4\theta = \cos 2\theta$ we write it as

$$\cos(\tfrac{1}{2}\pi - 4\theta) = \cos 2\theta,$$

i.e. either

$$\tfrac{1}{2}\pi - 4\theta = 2\theta \pm 2n\pi \quad \text{or} \quad \tfrac{1}{2}\pi - 4\theta = -2\theta \pm 2n\pi,$$

i.e.
$$\theta = \tfrac{1}{6}(\tfrac{1}{2} \pm 2n)\pi \quad \text{or} \quad \theta = \tfrac{1}{2}(\tfrac{1}{2} \pm 2n)\pi.$$

(*vi*) Quadratic equations, such as $2\cos^2\theta + \cos\theta - 1 = 0$, are solved first for $\cos\theta$ (two values) and then for θ from those roots which are not more than unity.

The equations $A\cos^2\theta - 2H\sin\theta\cos\theta + B\sin^2\theta = 0$ can either be written in terms of $t = \tan\theta$,

$$A - 2Ht + Bt^2 = 0,$$

whence t and θ can be found provided $H^2 > AB$; or it can be written as

$$\tfrac{1}{2}A(1 + \cos 2\theta) - H\sin 2\theta + \tfrac{1}{2}B(1 - \cos 2\theta) = 0,$$

i.e.
$$\tfrac{1}{2}(A - B)\cos 2\theta - H\sin 2\theta + \tfrac{1}{2}(A + B) = 0,$$

an equation for 2θ.

Most equations, however, have to be solved numerically, and this is always true of "mixed" equations such as $\frac{1}{3}x = \pi(1 - \cos x)$.

7.08. DERIVATIVES.

$$\frac{d}{dx}(\sin x) = \cos x, \qquad \frac{d}{dx}(\operatorname{cosec} x) = -\operatorname{cosec} x \cot x,$$

$$\frac{d}{dx}(\cos x) = -\sin x, \quad \frac{d}{dx}(\sec x) = \sec x \tan x,$$

$$\frac{d}{dx}(\tan x) = \sec^2 x, \qquad \frac{d}{dx}(\cot x) = -\operatorname{cosec}^2 x.$$

The first is proved from first principles using the limit (Section 3.13), $\lim_{\theta \to 0} \dfrac{\sin \theta}{\theta} = 1$ (θ in radians). Let $y = \sin x$;

$$\frac{\Delta y}{\Delta x} = \frac{\sin (x + \Delta x) - \sin x}{\Delta x} = \frac{2 \cos (x + \frac{1}{2}\Delta x) \sin \frac{1}{2}\Delta x}{\Delta x}.$$

If $\theta = \frac{1}{2}\Delta x$, then $\dfrac{dy}{dx} = \lim_{\theta \to 0} \dfrac{\cos (x + \theta) \sin \theta}{\theta} = \cos x.$

The other derivatives are proved either from first principles using the same limit, or by using the derivative of $\sin x$ and the rules of differentiation, e.g.

$$\frac{d}{dx}(\cos x) = \frac{d}{dx} \sin (x + \tfrac{1}{2}\pi) = \cos (x + \tfrac{1}{2}\pi) = -\sin x.$$

$$\frac{d}{dx}(\sec x) = \frac{d}{dx}\left(\frac{1}{\cos x}\right) = -\frac{1}{\cos^2 x}(-\sin x) = \frac{\sin x}{\cos^2 x} = \sec x \tan x.$$

It is best to learn these derivatives by heart, even though in practice, for example, it is more convenient to use $\dfrac{d}{dx}(\sec x) = \dfrac{\sin x}{\cos^2 x}$ than $\sec x \tan x$.

Example 1.

$$\frac{d}{dx}(\sin 2x) = 2 \cos 2x \text{ and } \frac{d}{dx}(2 \sin x \cos x) = 2 \cos^2 x - 2 \sin^2 x.$$

Since $\sin 2x = 2 \sin x \cos x$, then $\cos 2x = \cos^2 x - \sin^2 x$.

Example 2. $\dfrac{d}{dx}(\cos 2x) = -2 \sin 2x = -4 \sin x \cos x,$

and $\dfrac{d}{dx}(2 \cos^2 x) = -4 \sin x \cos x$, since $\cos 2x = 2 \cos^2 x - 1$.

Example 3. $\dfrac{d}{dx}(\cos^m x) = -m \cos^{m-1} x \sin x.$

Example 4. $\dfrac{d}{dx}(\tan^n x) = n \tan^{n-1} x \sec^2 x,$

$= n \tan^{n-1} x(1 + \tan^2 x),$

$= n(\tan^{n-1} x + \tan^{n+1} x).$

7.09. INVERSE SINE FUNCTION. A knowledge of the inverse trigonometric functions is necessary for integration. The inverse tangent is

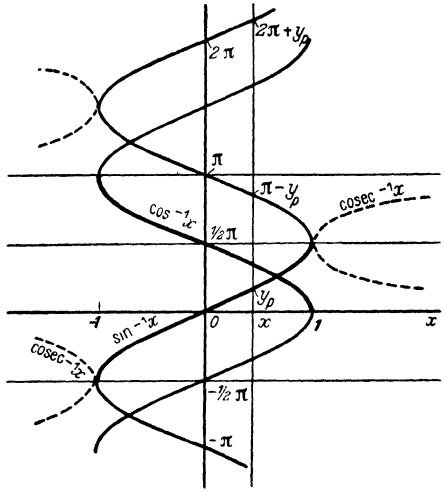

Figure 7.04

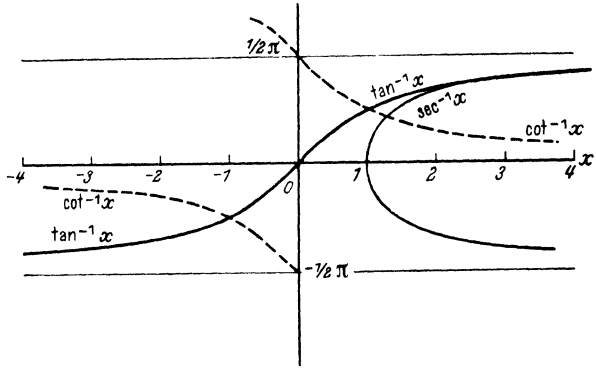

Figure 7.05

the most important but in this section the inverse sine is discussed as a better illustration.

The solution of the equation $\sin y = x$ is written

$$y = \sin^{-1} x.$$

It is not to be confused with $\dfrac{1}{\sin x} = (\sin x)^{-1}$. Some scientists avoid this confusion by the use of the notation $y = \text{arc sin } x$.

The function $y = \sin^{-1} x$ is called the inverse of $y = \sin x$ because the roles of x and y are interchanged. The graph of the function (shown in *Figure 7.04*) is most easily obtained this way. It is multi-valued.

There is one angle y between $-\frac{1}{2}\pi$ and $+\frac{1}{2}\pi$ such that $\sin y = x$ for every value of x between -1 and $+1$. This is the Principal Value y_P, and the graph of Principal Values increases steadily as x increases from -1 to $+1$.

There are other values, solutions of $\sin y = x = \sin y_P$, given by

$$\left.\begin{array}{ll} \text{either } (i): & y = y_P + 2n\pi \\ \text{or } (ii): & y = \pi - y_P + 2n\pi \end{array}\right\} n = 1, 2, 3 \ldots$$

Hence in any problem it is always necessary to say which values are to be used, and if this is not done it is assumed that Principal Values are to be used. This is to ensure that y is a continuous function of x.

From the graph it can be seen that tangents to the curve at values (i) are parallel and slope one way, whereas tangents at values (ii) slope the other way. These two values appear when the derivative is calculated.

The derivative of $\sin y = x$ is

$$\cos y \cdot \frac{dy}{dx} = 1, \quad \text{i.e. } \frac{dy}{dx} = \frac{1}{\cos y}.$$

For Principal Values, $\cos y_P$ is positive and

$$\cos y_P = \{1 - \sin^2 y_P\}^{1/2} = (1 - x^2)^{1/2}; \quad \frac{dy_P}{dx} = \frac{1}{(1 - x^2)^{1/2}}.$$

This is also true for all values (i). For the values (ii)

$$\cos y = -\cos y_P = -(1 - x^2)^{1/2}; \quad \frac{dy}{dx} = -\frac{1}{(1 - x^2)^{1/2}}.$$

A constant use of Principal Values (P.V.) removes this ambiguity.

7.10. Other Inverse Functions. The graphs of these functions are given in *Figure 7.05* and their properties are tabulated in *Table V*. They are derived as in the last section.

Table V

y	*Principal Value* y_P	*Principal derivative* dy_P/dx	*Other values*	*General derivative* dy/dx
$\sin^{-1} x$	$-90°$ *to* $+90°$	$\frac{1}{(1-x^2)^{1/2}}$	$(-)^n y_P + n\pi$	$\frac{1}{\cos y}$
$\cos^{-1} x$	$0°$ *to* $180°$	$-\frac{1}{(1-x^2)^{1/2}}$	$\pm y_P + 2n\pi$	$-\frac{1}{\sin y}$
$\tan^{-1} x$	$-90°$ *to* $90°$	$\frac{1}{1+x^2}$	$y_P + n\pi$	*same*
$\operatorname{cosec}^{-1} x$	$-90°$ *to* $90°$	$-\frac{1}{x(x^2-1)^{1/2}}$	$(-)^n y_P + n\pi$	$\frac{1}{x^2 \cos y}$
$\sec^{-1} x$	0 *to* $180°$	$\frac{1}{x(x^2-1)^{1/2}}$	$\pm y_P + 2n\pi$	$-\frac{1}{x^2 \sin y}$
$\cot^{-1} x$	$-90°$ *to* $90°$	$-\frac{1}{1+x^2}$	$y_P + n\pi$	*same*

The Principal Value of $\tan^{-1} x$ rises gradually from $-\frac{1}{2}\pi$ to $+\frac{1}{2}\pi$ as x increases from $-\infty$ to $+\infty$. The graph is S-shaped. At the origin the Principal Value is zero and the slope is

$$\left(\frac{dy}{dx}\right)_0 = \left(\frac{1}{1+x^2}\right)_0 = 1.$$

By shifting the origin and changing the scale, a curve of similar shape can be obtained which rises from 0 to any value b as x increases from $-\infty$ to $+\infty$, and has the value $\frac{1}{2}b$ at $x = a$. This is

$$y = \tfrac{1}{2}b\{1 + (2/\pi)\tan^{-1}(x - a)\}.$$

The slope at $x = a$ is (b/π). To obtain a similar curve with a given slope m at $x = a$, we write

$$y = \tfrac{1}{2}b\{1 + (2/\pi)\tan^{-1}(\pi m/b)(x - a)\}.$$

7.11. Addition of Inverse Functions. The sum $(\tan^{-1}\alpha + \tan^{-1}\beta)$ is the sum of two angles the value of which can be expressed as a single inverse function. Thus

$$\tan^{-1}\alpha + \tan^{-1}\beta = \tan^{-1}\{(\alpha + \beta)/(1 - \alpha\beta)\}.$$

To prove this, let $\tan^{-1}\alpha = A$, $\tan^{-1}\beta = B$. Then $\tan A = \alpha$,

$\tan B = \beta$. The simplest function of the angle $(A + B)$ that can be found is

$$\tan (A + B) = \frac{\tan A + \tan B}{1 - \tan A \tan B} = \frac{\alpha + \beta}{1 - \alpha\beta},$$

whence the result follows. In the same way

$$\cos^{-1} x = \tfrac{1}{2}\pi + \sin^{-1} x, \quad \text{(P.V.)}$$
$$\tan^{-1} x + \cot^{-1} x = \tfrac{1}{2}\pi. \quad \text{(P.V.)}$$

EXAMPLES VII

1. Solve completely the equations $\cos\theta = \cos 2\theta$ and $\cos\theta = \sin 2\theta$.

2. Prove the formula $\sin 3\theta = 3\sin\theta - 4\sin^3\theta$. By the substitution $x = 2A\sin\theta$, show that the cubic equation $x^3 - 3A^2x + B = 0$ can be written $B = 2A^3 \sin 3\theta$. Explain how to solve the equation when $B \leqslant 2A^3$ and solve the equation $x^3 - 3x + 1 = 0$ in this way.

Adjust the procedure to show that a cubic equation,

$$ax^3 + 3bx^2 + 3cx + d = 0,$$

known to have three roots, can be solved with $x = b/a + 2A\sin\theta$, and

$$A^2 = \frac{ac - b^2}{a^2}, \quad B = \frac{a^2d - 3abc + 2b^3}{a^3}.$$

3. Show by two methods that the maximum value of $a\cos\theta + b\sin\theta$ is $(a^2 + b^2)^{1/2}$.

4. A sphere is dropped into a full conical glass of depth a and semi-vertical angle α. Find the radius of the sphere for maximum overflow.

5. Sketch as accurately as possible between $\theta = 0$ and $\theta = \pi$: (*i*) $y = \sin\theta$, (*ii*) $y = \sin\theta + \frac{1}{3}\sin 3\theta$, (*iii*) $y = \sin\theta + \frac{1}{3}\sin 3\theta + \frac{1}{5}\sin 5\theta$.

Compare each with the line $y = 2\theta$.

6. Sketch the graphs of $y = \dfrac{\sin x}{x}$ and $y = \dfrac{\sin x}{x^2} - \dfrac{\cos x}{x}$.

7. A curve to be fitted, using the formula

$$y = \alpha + \beta \tan^{-1}\{(x - a)/b\},$$

has (*i*) $y \to 6$ as $x \to \infty$, (*ii*) $y \to 0$ as $x \to -\infty$, (*iii*) a point of inflection at $x = 2$ where (*iv*) the slope is $\frac{3}{2}$. Sketch the curve and find y when $x = 0$.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

DURELL, C. V. and ROBSON, A. *Advanced Trigonometry.* G. Bell and Sons Ltd, London, 1938.

GRANVILLE, W. A., SMITH, P. F. and LONGLEY, W. R. *Introduction to the Differential and Integral Calculus.* G. Ginn and Co., London and New York, 1941.

JAHNKE, E. and EMDE, F. *Tables of Advanced Functions.* Teubner, Leipzig, 1948.

MILNE-THOMSON, L. and COMRIE, L. J. *Standard Four-Figure Mathematical Tables.* Macmillan and Co. Ltd, London, 1936.

CHAPTER 8

EXPONENTIAL, LOGARITHMIC AND HYPERBOLIC FUNCTIONS

Natural number e—Exponential functions—Exponential increase—Exponential decay—Combination of exponentials—Derivative of $y = e^x$—Logarithmic function—Derivative of $y = \ln x$—Hyperbolic functions—Derivatives—Hyperbolic and circular functions—Properties of hyperbolic functions—Inverse hyperbolic functions—Derivatives of inverse functions

8.01. NATURAL NUMBER. $e = 2{\cdot}71828.$

$$\log_{10} e = 0{\cdot}434294, \quad \log_e 10 = 2{\cdot}302585.$$

In the following chapters $\log_e a$ is written $\ln a$, $\log_{10} a$ is written $\log a$, unless there is ambiguity.

The number e is defined as

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.$$

To prove that the sequence, $S_n = \left(1 + \frac{1}{n}\right)^n$, is convergent, we use the Binomial Theorem,

$$S_n = 1 + n\left(\frac{1}{n}\right) + \frac{n(n-1)}{1.2}\left(\frac{1}{n}\right)^2 + \frac{n(n-1)(n-2)}{1.2.3}\left(\frac{1}{n}\right)^3 + \ldots + \left(\frac{1}{n}\right)^n,$$

$$= 1 + 1 + \frac{1}{1.2}\left(1 - \frac{1}{n}\right) + \frac{1}{1.2.3}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \ldots + \left(\frac{1}{n}\right)^n.$$

Thus S_n is expressed as the sum of $(n + 1)$ terms. Consider any one term, say the third or fourth. If n is increased by one, each of its factors $\left(1 - \frac{1}{n}\right)$, $\left(1 - \frac{2}{n}\right)$, increases in value, and therefore the term increases in value. In fact, every term in S_{n+1} is greater than or equal to the corresponding term of S_n, and S_{n+1} has an extra term at the end. Hence $S_{n+1} > S_n$, i.e. the series is monotonic increasing.

Also by comparing terms

$$S_n < 1 + 1 + \frac{1}{1.2} + \frac{1}{1.2.2} + \frac{1}{1.2.2.2} + \cdots + \frac{1}{2^{n-1}}.$$
$$< 3.$$

Hence from the convergence test in Section 3.06 (*ii*) the sequence converges.

8.02. EXPONENTIAL FUNCTIONS. These are of the form $y = a^x$ or $y = a^{-x}$, where a is a given number. Graphs are drawn in *Figure 8.01* for $a = 1, 2, e, 10$.

As long as $a > 1$, a^x increases with x, and a^{-x} decreases to zero. Thus a^x is exponentially increasing and a^{-x} is exponentially decreasing.

Example 1. $y = a^x$ can be written $y = e^{mx}$ where $m = \ln a$. The number m is positive when $a > 1$ and negative when $a < 1$.

Example 2. Graphs of $y = b^x, \left(b = \frac{1}{2}, \frac{1}{e}, \frac{1}{10}\right)$, are derived from those of *Figure 8.01* by reversing the sign of x,

$$a^x = \left(\frac{1}{a}\right)^{-x}.$$

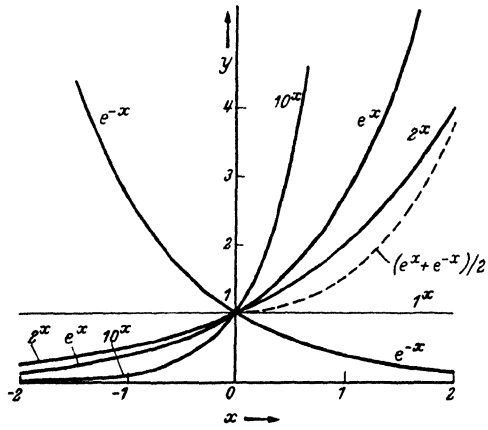

Figure 8.01

8.03. EXPONENTIAL INCREASE. Let us assume that a repetitive process, such as a chain reaction, increases the amount of a substance by a fraction $1/n$ at each operation.

If the original amount is a_0, the amount after one operation is $a_1 = a_0\left(1 + \frac{1}{n}\right)$, and after the second operation $a_2 = a_0\left(1 + \frac{1}{n}\right)^2$. After N operations it is

$$a_N = a_0\left(1 + \frac{1}{n}\right)^N.$$

When the fractional increase in any operation is small, so that n is large, this can be changed to an exponential form. By definition, the natural number e differs little from $\left(1 + \frac{1}{n}\right)^n$ when n is large.

$$\therefore \qquad a_N = a_0\left\{\left(1 + \frac{1}{n}\right)^n\right\}^{N/n} = a_0 e^x,$$

where $x = N/n$. Thus the amount increases exponentially with the number of operations, in this approximation.

The variation with time is obtained by introducing the time τ of each operation. The gain per unit time in one operation is

$$\frac{\Delta a_N}{\tau} = \frac{a_{N+1} - a_N}{\tau} = \frac{(a_N/n)}{\tau} = \frac{a_N}{n\tau},$$

and the time for N operations is $t = N\tau = (n\tau)x$, so that

$$a_N = a_0 e^{kt}, \text{ approximately,}$$

where $k = 1/n\tau$.

An exponential increase is also obtained in any continuous process where the rate of increase is proportional to the amount present. This corresponds to the limit $\tau \to 0$ and $n \to \infty$ in the last two equations, k remaining finite, so that they become

$$\frac{da}{dt} = \lim_{\tau \to 0} \frac{\Delta a}{\tau} = ka \quad \text{and} \quad a = a_0 e^{kt}.$$

Such a process occurs in any chain reaction or autocatalytic reaction, where the active agent is reproduced by the action of the products rather faster than it is used up.

8.04. EXPONENTIAL DECAY. By the same method it is possible to show that a distillation or a washing process leads to an exponential decrease when the amount of a substance is reduced by a given fraction at every operation. This has been considered already in Section 3.07.

The corresponding continuous process is one where the rate of loss is proportional to the amount present. For example, in a

unimolecular decomposition (a first-order chemical reaction) the concentration c at time t changes to $(c + \Delta c)$ in a small time Δt, where

$$\Delta c = -kc\Delta t, \qquad \text{i.e. } dc/dt = -kc.$$

The minus sign indicates a decrease with time. The variation of c with time is

$$c = c_0 e^{-kt},$$

where $c = c_0$ at time $t = 0$.

Another example of this type of process is radioactive decay, where the number of atoms decaying per unit time in a given sample is proportional to the number present at that instant.

8.05. COMBINATION OF EXPONENTIALS. In a complicated decay process, it often happens that the quantity of importance is a sum of several decreasing exponential functions,

$$y = a_1 e^{-m_1 t} + a_2 e^{-m_2 t} + a_3 e^{-m_3 t} + \ldots$$

It is assumed that the terms are arranged in order of increasing exponents, i.e. $m_1 < m_2 < m_3 \ldots$.

When t is small the important term is clearly that with the largest value of a, since any exponential $e^{-mt} \fallingdotseq 1$ when t is small.

When t is large enough, however, the first term, with the smallest m, is most important, i.e.

$$y = a_1 e^{-m_1 t},$$

approximately, so that $\ln y = \ln a_1 - m_1 t$ and the graph of $\ln y$ against t becomes a straight line.

The reason for this lies in the result $e^{-x} \to 0$ as $x \to \infty$, for this implies that

$$e^{-(m_2 - m_1)t} \to 0 \text{ as } t \to \infty, \text{ if } m_2 > m_1.$$

Thus all the terms become small compared with the first if t is large enough.

8.06. DERIVATIVE OF $y = e^x$. It has been shown in Section 8.03 that

$$y = e^x, \quad dy/dx = e^x = y,$$

but a formal proof is now given.

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{e^{x+h} - e^x}{h} = e^x \lim_{h \to 0} \frac{e^h - 1}{h} = e^x.$$

The limit has already been found in Section 3.13 (*ii*).

Geometrically this value for the derivative implies that the subtangent ($y \div dy/dx$) equals unity at all points of the curve $y = e^x$.

Other exponential functions are differentiated using the FF rule, e.g.

$$y = ae^{mx}, \qquad dy/dx = mae^{mx};$$
$$y = 10^x = e^{x \ln 10}, \quad dy/dx = \ln 10 \,.\, e^{x \ln 10} = 10^x \,.\, \ln 10.$$

8.07. LOGARITHMIC FUNCTION $y = \ln x$. This means $x = e^y$ and is the inverse of the exponential function. It is defined for positive values of x only, since $e^y > 0$ for all values of y.

Logarithmic functions occur in various forms:

$$y = a \ln x, \qquad (x > 0),$$
$$y = a + k \ln (x - c) = k \ln (x - c)/b, \quad (x > c),$$
$$y = a + k \ln (c - x) = k \ln (c - x)/b, \quad (x < c),$$

where $a = k \ln (1/b)$, i.e. $b = e^{-a/k}$.

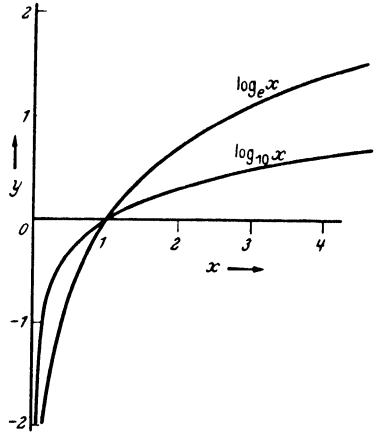

Figure 8.02

The graph (*Figure 8.02*) illustrates the following properties of $\ln x$:

(*i*) it increases steadily with x;
(*ii*) $\ln x \to -\infty$ as $x \to 0$;
(*iii*) $\ln x = 0$ when $x = 1$;
(*iv*) $\ln x \to +\infty$ as $x \to \infty$.

The increase with x is very slow when x is large, e.g. increasing x from 1 to 2·7 and increasing x from 1000 to 2700 increase $\ln x$ by the same amount. The slow rate of increase can also be made clear by comparing $\ln x$ with a small power of x, say x^n. By L'Hôpital's theorem (for the derivative see the next section),

$$\lim_{x\to\infty} \frac{\ln x}{x^n} = \lim_{x\to\infty} \frac{1/x}{nx^{n-1}} = \lim_{x\to\infty} \frac{1}{nx^n} = 0,$$

however small n may be, as long as it is positive. In other words $\ln x$ increases more slowly than any positive power, if x is sufficiently large.

8.08. Derivative of $y = \ln x$. Since $x = e^y$, then

$$\frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{e^y} = \frac{1}{x}.$$

Thus $x^{-1} = 1/x$ is the derivative not of a power of x, but of $\ln x$.

Using the FF rule we derive

$$\frac{d}{dx}(\ln u) = \frac{1}{u}\frac{du}{dx},$$

which is extremely useful and is the basis of logarithmic differentiation (Section 4.23).

To differentiate logarithms with another base we simply change the base, e.g.

$$y = \log_{10} x = \log_e x/\log_e 10,$$

$$\therefore \qquad \frac{dy}{dx} = \frac{1}{x \log_e 10} = \frac{\log_{10} e}{x}.$$

8.09. Hyperbolic Functions. In Section 1.16, it was shown that any function $f(x)$ can be separated into $f_e(x)$, an even part, and $f_o(x)$, an odd part.

The hyperbolic cosine (cosh x) is the even part of e^x and the hyperbolic sine (sinh x) is the odd part.

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}), \qquad \sinh x = \tfrac{1}{2}(e^x - e^{-x}),$$

$$\cosh(-x) = \cosh x, \qquad \sinh(-x) = -\sinh x$$

Other hyperbolic functions are derived from these as for the circular functions:

$$\tanh x = 1/\coth x = \sinh x/\cosh x,$$
$$\operatorname{sech} x = 1/\cosh x, \ \operatorname{cosech} x = 1/\sinh x.$$

These curves are drawn in *Figures 8.03* and *8.04*.

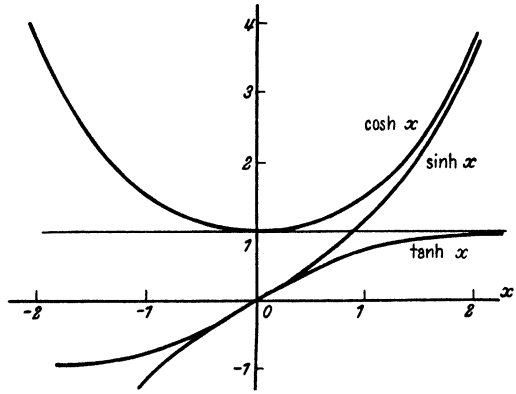

Figure 8.03

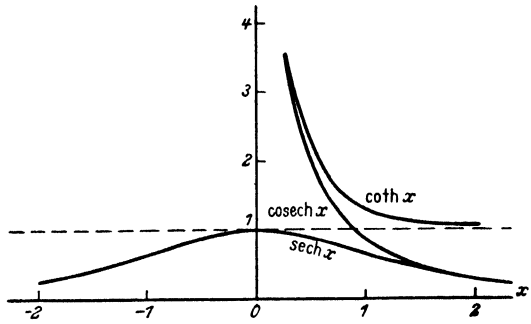

Figure 8.04

8.10. DERIVATIVES. From the definitions in terms of exponentials:

$$\frac{d}{dx}(\sinh x) = \frac{d}{dx}\tfrac{1}{2}(e^x - e^{-x}) = \tfrac{1}{2}(e^x + e^{-x}) = \cosh x,$$

$$\frac{d}{dx}(\cosh x) = \frac{d}{dx}\tfrac{1}{2}(e^x + e^{-x}) = \tfrac{1}{2}(e^x - e^{-x}) = \sinh x.$$

Using the product and quotient formulae, the derivatives of the other hyperbolic functions are obtained, e.g. from Section 8.12,

$$\frac{d}{dx}(\tanh x) = \frac{d}{dx}\left(\frac{\sinh x}{\cosh x}\right),$$
$$= \frac{\cosh x(\cosh x) - \sinh x(\sinh x)}{\cosh^2 x} = \frac{1}{\cosh^2 x},$$
$$= \operatorname{sech}^2 x.$$

Similarly:

$$\frac{d}{dx}(\operatorname{cosech} x) = -\frac{\cosh x}{\sinh^2 x} = -\coth x \operatorname{cosech} x,$$

$$\frac{d}{dx}(\operatorname{sech} x) = -\frac{\sinh x}{\cosh^2 x} = -\tanh x \operatorname{sech} x,$$

$$\frac{d}{dx}(\coth x) = -\operatorname{cosech}^2 x.$$

It is usually best to express these in terms of sinh x and cosh x.

8.11. Hyperbolic and Circular Functions. The similarity between these two groups of functions, suggested by their names and properties, can be described geometrically using the rectangular hyperbola and a circle. An analytical method is also given in Chapter 14, where the circular functions are expressed as exponential functions of complex numbers.

A third method, similar to the second, shows that they play identical roles in similar problems. For this we need the derivatives.

Consider: (*i*) $y = a\cos mx + b\sin mx,$
$$dy/dx = -am\sin mx + bm\cos mx,$$
$$d^2y/dx^2 = -am^2\cos mx - bm^2\sin mx,$$
$$= -m^2y.$$

(*ii*) $y = a\cosh mx + b\sinh mx,$
$$dy/dx = ma\sinh mx + mb\cosh mx,$$
$$d^2y/dx^2 = m^2a\cosh mx + m^2b\cosh mx,$$
$$= m^2y.$$

For both functions it is clear that

$$y = a,\ dy/dx = mb, \text{ when } x = 0$$

Comparing (*i*) and (*ii*) we see that a solution of

$$d^2y/dx^2 = \alpha y,$$

such that $y = 1$ and $dy/dx = 0$ ($a = 1$, $b = 0$), when $x = 0$, is:

$$y = \cos mx, \text{ when } \alpha = -m^2 \text{ is negative;}$$
$$y = \cosh mx, \text{ when } \alpha = +m^2 \text{ is positive;}$$

and a solution such that $y = 0$, $dy/dx = 1$ ($a = 0$, $b = 1$), when $x = 0$, is:

$$y = (1/m) \sin mx, \text{ when } \alpha = -m^2 \text{ is negative;}$$
$$y = (1/m) \sinh mx, \text{ when } \alpha = +m^2 \text{ is positive.}$$

8.12. PROPERTIES OF HYPERBOLIC FUNCTIONS. The hyperbolic functions satisfy identities similar to those of the trigonometric or circular functions. A simple rule for obtaining the former from the latter is: place an 'h' after the symbol and note that a product containing one sine remains unchanged, but a term containing two (and sometimes three) sines changes its sign. The reason is given in Section 14.14.

From the definitions, $\cosh x = \frac{1}{2}(e^x + e^{-x})$ and $\sinh x = \frac{1}{2}(e^x - e^{-x})$, all the following results are proved.

When $x = 0$, $e^x = e^{-x} = 1$, so that

$$\cosh (0) = 1, \quad \sinh (0) = 0.$$

When x is large and positive, e^x is large and e^{-x} small;

$$\sinh x \to \tfrac{1}{2}e^x, \quad \cosh x \to \tfrac{1}{2}e^x.$$

Also it follows that $|\sinh x| < \cosh x$ and $|\tanh x| < 1$. Adding and subtracting:

$$\cosh x + \sinh x = e^x,$$
$$\cosh x - \sinh x = e^{-x},$$

and multiplying these two

$$\cosh^2 x - \sinh^2 x = 1, \quad (\text{cf. } \cos^2 x + \sin^2 x = 1).$$

From this equation there follows:

$$\tanh^2 x + \operatorname{sech}^2 x = 1,$$
$$\coth^2 x - \operatorname{cosech}^2 x = 1.$$

As $|\tanh x| < 1$ there is always an angle θ such that:

$$\tanh x = \sin \theta, \quad \operatorname{sech} x = \cos \theta, \quad |\theta| < 90°;$$
$$\sinh x = \tan \theta, \quad \cosh x = \sec \theta.$$

The relations given below are proved directly as in the example

$$\sinh(A+B) = \sinh A \cosh B + \cosh A \sinh B.$$

Proof: $\text{RHS} = \frac{1}{2}(e^A - e^{-A})\frac{1}{2}(e^B + e^{-B}) + \frac{1}{2}(e^A + e^{-A})\frac{1}{2}(e^B - e^{-B})$

$= \frac{1}{4}\{(e^{A+B} - e^{-A+B} + e^{A-B} - e^{-A-B})$

$+ (e^{A+B} + e^{-A+B} - e^{A-B} - e^{-A-B})\}$

$= \frac{1}{2}(e^{A+B} - e^{-(A+B)}),$

$= \sinh(A+B).$

$$\sinh(A \pm B) = \sinh A \cosh B \pm \cosh A \sinh B.$$

$$\cosh(A \pm B) = \cosh A \cosh B \pm \sinh A \sinh B$$

(note signs).

Hence:
$$2\sinh A \cosh B = \sinh(A+B) + \sinh(A-B).$$
$$2\cosh A \cosh B = \cosh(A+B) + \cosh(A-B).$$
$$2\sinh A \sinh B = \cosh(A+B) - \cosh(A-B).$$
$$\sinh A + \sinh B = 2\sinh \tfrac{1}{2}(A+B)\cosh \tfrac{1}{2}(A-B).$$
$$\sinh A - \sinh B = 2\cosh \tfrac{1}{2}(A+B)\sinh \tfrac{1}{2}(A-B).$$
$$\cosh A + \cosh B = 2\cosh \tfrac{1}{2}(A+B)\cosh \tfrac{1}{2}(A-B).$$
$$\cosh A - \cosh B = 2\sinh \tfrac{1}{2}(A+B)\sinh \tfrac{1}{2}(A-B).$$
$$\tanh(A+B) = (\tanh A + \tanh B)/(1 + \tanh A \tanh B).$$

By writing $A = B = x$, we deduce that:

$$\cosh^2 x + \sinh^2 x = \cosh 2x \quad \text{and} \quad 2\sinh x \cosh x = \sinh 2x,$$
$$\cosh x = \cosh^2 \tfrac{1}{2}x + \sinh^2 \tfrac{1}{2}x \quad \text{and} \quad \sinh x = 2\sinh \tfrac{1}{2}x \cosh \tfrac{1}{2}x.$$

Using $\cosh^2 \frac{1}{2}x - \sinh^2 \frac{1}{2}x = 1$, the last formula but one gives:

$$\cosh^2 \tfrac{1}{2}x = \tfrac{1}{2}(\cosh x + 1),$$
$$\sinh^2 \tfrac{1}{2}x = \tfrac{1}{2}(\cosh x - 1).$$

Also $\sinh x = 2t/(1 - t^2)$, $\cosh x = (1 + t^2)/(1 - t^2)$,
where $t = \tanh \frac{1}{2}x$..

8.13. INVERSE HYPERBOLIC FUNCTIONS. Each hyperbolic function has an inverse, which can be expressed in logarithmic form:

(i) $y = \sinh^{-1} x = \ln\{\sqrt{(x^2+1)} + x\} = -\ln\{\sqrt{(x^2+1)} - x\}$,

Thus $\sinh y = x$,

$\therefore \cosh y = \sqrt{(1 + \sinh^2 y)} = \sqrt{(1 + x^2)}$, (since $\cosh y > 0$).

$$\therefore \qquad e^{y} = \sinh y + \cosh y = x + \sqrt{(x^2 + 1)},$$

$$e^{-y} = \cosh y - \sinh y = \sqrt{(x^2 + 1)} - x.$$

$$\therefore \quad y = \ln \{\sqrt{(x^2 + 1)} + x\} = -\ln \{\sqrt{(x^2 + 1)} - x\}.$$

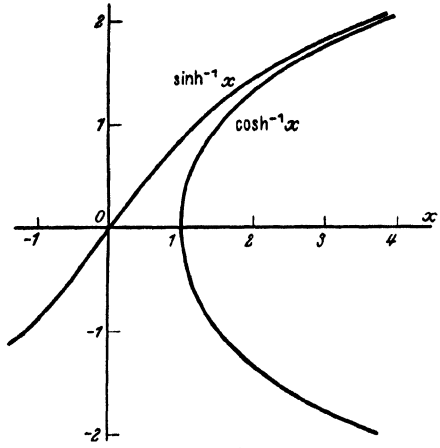

Figure 8.05

(*ii*) $y = \cosh^{-1} x = \begin{cases} \ln \{x + \sqrt{(x^2 - 1)}\}, & (\text{when } \sinh y > 0), \\ \ln \{x - \sqrt{(x^2 - 1)}\}, & (\text{when } \sinh y < 0). \end{cases}$

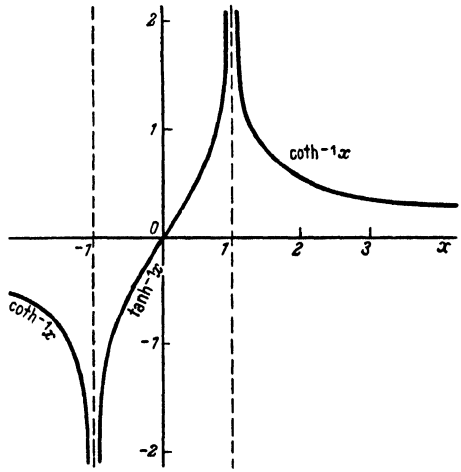

Figure 8.06

This function is double-valued because $\cosh y = \cosh(-y)$. The Principal Value is the positive value.

This is proved as follows:

$$\cosh y = x,$$
$$\sinh y = \sqrt{(\cosh^2 y - 1)} = \sqrt{(x^2 - 1)}, \quad (\text{when } y > 0).$$
$$\therefore \quad e^y = \cosh y + \sinh y = x + \sqrt{(x^2 - 1)}.$$

The other value is obtained by changing to $\sinh y = -\sqrt{(x^2 - 1)}$. The two values of $\cosh^{-1} x$ are equal and opposite since

$$\{x + \sqrt{(x^2 - 1)}\}\{x - \sqrt{(x^2 - 1)}\} = x^2 - (x^2 - 1) = 1,$$

so that the sum of their logarithms is $\ln (1) = 0$.

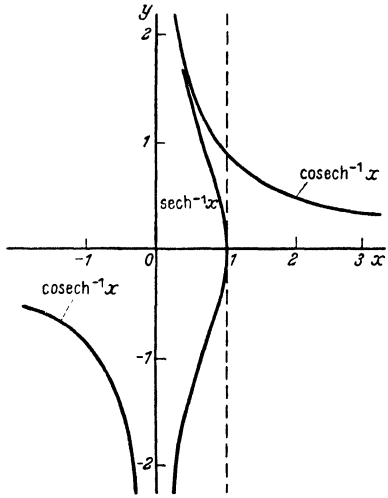

Figure 8.07

(*iii*) $y = \tanh^{-1} x = \frac{1}{2} \ln (1 + x)/(1 - x) = \ln \sqrt{\{(1 + x)/(1 - x)\}}$.

Since $\tanh y = x$,

$$\operatorname{sech} y = \sqrt{(1 - \tanh^2 y)} = \sqrt{(1 - x^2)}.$$
$$\therefore \quad \sinh y = x/\sqrt{(1 - x^2)} \quad \text{and} \quad \cosh y = 1/\sqrt{(1 - x^2)}.$$
$$\therefore \quad e^y = (x + 1)/\sqrt{(1 - x^2)} = \sqrt{\{(x + 1)/(x - 1)\}}.$$

The graphs of these functions are given in *Figures 8.05, 8.06* and *8.07*, together with those of

$$\operatorname{cosech}^{-1} x, \quad \operatorname{sech}^{-1} x, \quad \coth^{-1} x.$$

These results are summarized in *Table VI.*

Table VI

y	*Logarithmic form*	*Values of* x	*Derivative*
$\sinh^{-1} x$	$\ln\{x + (1 + x^2)^{1/2}\}$	*all*	$1/(1 + x^2)^{1/2}$
$\cosh^{-1} x$	$\ln\{x + (x^2 - 1)^{1/2}\}$	$\lvert x\rvert > 1$	$1/(x^2 - 1)^{1/2}$ *(P.V.)*
$\tanh^{-1} x$	$\frac{1}{2}\ln(1 + x)/(1 - x)$	$\lvert x\rvert < 1$	$1/(1 - x^2)$
$\operatorname{cosech}^{-1} x$	$\ln\{1 + (1 + x^2)^{1/2}\} - \ln x$	*all*	$-1/\{x(1 + x^2)^{1/2}\}$
$\operatorname{sech}^{-1} x$	$\ln\{1 + (1 - x^2)^{1/2}\} - \ln x$	$\lvert x\rvert < 1$	$-1/\{x(1 - x^2)^{1/2}\}$
$\coth^{-1} x$	$\frac{1}{2}\ln(x + 1)/(x - 1)$	$\lvert x\rvert > 1$	$1/(1 - x^2)$

8.14. DERIVATIVES OF INVERSE FUNCTIONS. These can easily be obtained using Section 4.24, e.g.

$$y = \sinh^{-1} x \quad \text{means } x = \sinh y, \quad \text{or} \quad \cosh y(dy/dx) = 1.$$

But $\cosh y = \sqrt{(1 + x^2)}$,

$$\therefore \qquad dy/dx = 1/\sqrt{(1 + x^2)}.$$

The same result is obtained in a longer way by differentiating the logarithmic form.

The relation between $\tan^{-1} x$ and $\tanh^{-1} x$ is very close. Their graphs are not very similar, if the Principal Value of $\tan^{-1} x$ is used, but their derivatives are similar:

$$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1 + x^2}, \quad \frac{d}{dx}(\tanh^{-1} x) = \frac{1}{1 - x^2} \quad (|x| < 1),$$

$$\frac{d}{dx}(\coth^{-1} x) = \frac{1}{1 - x^2} \quad (|x| > 1).$$

EXAMPLES VIII

1. From the definition of a logarithm prove *(i)* if $y = e^{\log x}$, then $y = x$, *(ii)* if $\log y_0 - \log y = kct$, then $y = y_0 e^{-kct}$, *(iii)* if $a(x + \log y) = b \log x + m$, then $x^b = y^a e^{ax - m}$.

2. A radioactive substance decays exponentially. The time for the amount to decrease by a factor $\frac{1}{2}$, i.e. the half-life, is τ. If y_0 is the amount present at $t = 0$, show that the amount at time t is

$$y = y_0(\tfrac{1}{2})^{t/\tau} = y_0 e^{-(t \log 2)/\tau}.$$

3. If $f = ae^{-b/T}$, where a and b are constants, find $d(\log f)/d(\log T)$.

4. The Magnus formula for vapour pressure p is $p = ab^{\theta/(\theta+r)}$. Find $dp/d\theta$.

5. Find the ratio of successive stationary values of $x = Ae^{-\lambda t}\sin \omega t$, and sketch this function by multiplying the curves $Ae^{-\lambda t}$ and $\sin \omega t$.

6. According to Wien's law the energy between wavelengths λ and $(\lambda + d\lambda)$ emitted per unit time by a 'black body' at temperature T is $E_\lambda d\lambda$ where $E_\lambda = (k/\lambda^5)e^{-(a/\lambda T)}$.

Plot E_λ against λ, and prove that it tends to zero as $\lambda \to \infty$ and as $\lambda \to 0$ (hint: put $\lambda = 1/x$), and has a maximum when $\lambda = (a/5T)$.

7. According to the one-vibration model of Einstein, the energy of a crystal at temperature T is $E = 3R(\hbar\nu/k)/(e^{\hbar\nu/kT} - 1)$.

Plot E/R against T assuming that $\hbar/k = 4{\cdot}8 \times 10^{-11}$ and $\nu = 10^{13}$. Find the specific heat $c = dE/dT$ and plot this against T.

8. Show that the values of the sequence S_n defined by the recurrence relation,

$$S_n = 2\alpha S_{n-1} - \beta^2 S_{n-2},$$

are of the form

$$S_n = \beta^n \{Ae^{n\theta} + Be^{-n\theta}\},$$

where $\alpha = \beta \cosh \theta$, if $\alpha > \beta$. Find S_n when $S_0 = 1$, $S_1 = 2\alpha$. (cf. Qn *8*, Chapter 14.)

9. Prove that the derivative of $\sinh^{-1}(\sqrt{x})$ is one-half that of

$$\log \{x + \tfrac{1}{2} + \sqrt{(x^2 + x)}\},$$

and explain why, by expressing $\sinh^{-1}(\sqrt{x})$ in logarithmic form.

10. Repeat Qn *7*, Chapter 7, when the curve is fitted by the formula

$$y = \alpha + \beta \tanh^{-1}(x - a)/b.$$

11. Sketch Langevin's function $y = \coth x - 1/x$ for positive values of x.

12. The speed of water, flowing from a tap which is suddenly turned on, varies initially according to the law $a(dv/dt) = (v_0^2 - v^2)$, where a is a constant. Show that $v = v_0 \tanh (v_0 t/a)$ is a solution such that $v = 0$ when $t = 0$.

13. Sketch graphs to show that each of the following equations has an infinite number of roots, the larger roots having the approximate values stated (n is an integer):

$$(i)\ x = \tan x, [\tfrac{1}{2}(2n + 1)\pi]: \quad (ii)\ \tanh x = \tan x, [(n + \tfrac{1}{2})\pi],$$
$$(iii)\ \cosh x = \sec x, [(n + \tfrac{1}{2})\pi].$$

The author has found useful information on exponential and hyperbolic functions in the following books:

PARTINGTON, J. R. *An Advanced Treatise on Physical Chemistry.* Longmans, Green and Co., London, 1951.

PIPES, L. A. *Applied Mathematics for Engineers and Physicists.* McGraw-Hill Book Co. Inc., New York, 1946.

JAHNKE, E. and EMDE, F. *Tables of Advanced Functions.* Teubner, Leipzig, 1948.

MILNE-THOMSON, L. M. and COMRIE, L. J. *Standard Four-Figure Mathematical Tables.* Macmillan and Co. Ltd, London, 1936.

MELLOR, J. W. *Higher Mathematics for Students of Physics and Chemistry.* Chapter VI. Longmans, Green and Co., London, 1912.

CHAPTER 9

PROPERTIES OF CURVES

Introduction—Curves in Cartesian coordinates—Change of origin—Rotation of axes—Change of origin and rotation of axes—Tangent and normal—Distance along curves—Curvature—Polar coordinates—Short list of curves—Triangular coordinates—Curvilinear and generalized coordinates—Two points—The straight line—Two straight lines—Conic sections—Circle—Parabola—Ellipse—Hyperbola

9.01. INTRODUCTION. This chapter is a collection of the properties of curves, discussed first in terms of Cartesian coordinates, and then in terms of polar coordinates. At the end of the chapter the properties of conic sections are collected together, since these arise in mechanical problems and especially in the 'classical' pictures used in atomic physics. Between the two there is a section containing lists of curves of various types.

9.02. CURVES IN CARTESIAN COORDINATES. The equation of a curve in Cartesian coordinates has the form $y = f(x)$, or $\phi(x, y) = 0$, where $\phi(x, y)$ stands for a function of the two coordinates.

The equation can also be given in parametric form, i.e. both x and y may be given as functions of a parameter, e.g. t or θ, which has a value at each point of the curve. The relation between x and y is obtained by eliminating t (or θ) between these two expressions. This method not only has the advantage that t often has a definite meaning for the curve, but it leaves the symbols x and y free for general purposes.

Example 1. Parametric equations of the circle $x^2 + y^2 = a^2$ are $x = a \cos \theta, y = a \sin \theta$.

The form of the equation, or equations, naturally depends on the position of the curve relative to the axes and the origin. If the position of these is optional, it may be possible to choose them relative to the curve so that its equation is especially simple for the discussion of its properties. This is the standard form of the equation in Cartesian coordinates.

9.03. CHANGE OF ORIGIN. Let a new origin O' be chosen with coordinates (h, k) relative to the old axes, and new axes $O'X'$, $O'Y'$ be drawn parallel to the old axes (*Figure 9.01*). The coordinates

of any point P relative to the old axes and to the new axes are related by the equations

$$x = x' + h, \quad y = y' + k.$$

Hence the equation of a curve $y = f(x)$ becomes $y' + k = f(x' + h)$ in the new coordinates. The parametric equations $x = f(t)$, $y = g(t)$ become

$$x' = x - h = f(t) - h,$$
$$y' = y - k = g(t) - k.$$

9.04. Rotation of Axes. New axes can be introduced at an angle θ, anticlockwise, with the old, without changing the origin (*Figure 9.02*). The old coordinates of any point P in terms of the new are

$$x = x'' \cos\theta - y'' \sin\theta,$$
$$y = x'' \sin\theta + y'' \cos\theta.$$

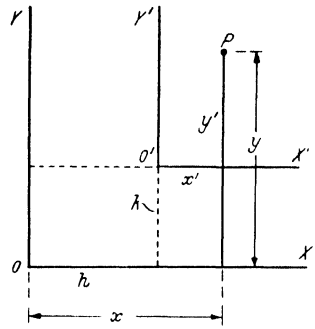

Figure 9.01

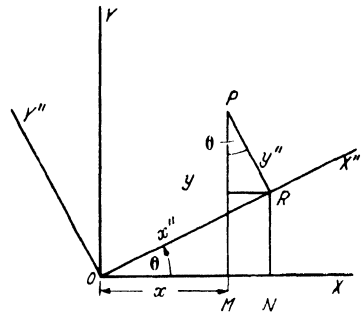

Figure 9.02

To change parametric equations it is better to find x'' and y'' in terms of x and y. This can be done either by solving the above equations or more simply by arguing that a return from the axis OX'' to the axis OX means a rotation in the opposite sense, and is obtained by changing θ to $-\theta$ in the above equations.

$$x'' = x \cos\theta + y \sin\theta, \quad y'' = -x \sin\theta + y \cos\theta.$$

Example 1. If $\theta = 45°$, then

$$x = (1/\sqrt{2})(x'' - y''), \quad \text{and} \quad y = (1/\sqrt{2})(x'' + y'').$$

The equation $xy = k^2$ becomes

$$\tfrac{1}{2}(x'' - y'')(x'' + y'') = k^2, \quad \text{or} \quad x''^2 - y''^2 = 2k^2.$$

Example 2. The equation $x^2 + y^2 = a^2$ becomes $x''^2 + y''^2 = a^2$ for all values of θ. Thus the curve (a circle) has the same shape relative to all axes.

9.05. CHANGE OF ORIGIN AND ROTATION OF AXES. If $O'X'$ and $O'Y'$ are new axes obtained by shifting the origin to (h, k) and rotating the axes anticlockwise through an angle θ, then

$$x = h + x' \cos\theta - y' \sin\theta,$$
$$y = k + x' \sin\theta + y' \cos\theta.$$

9.06. TANGENT AND NORMAL. At any point P on $y = f(x)$ the slope of the tangent is $(dy/dx) = f'(x)$ and that of the normal is $-1/(dy/dx) = -1/\{f'(x)\}$. Their equations are therefore (cf. Section 5.02):

$$\text{Tangent:} \qquad y - y_P = (dy/dx)_P(x - x_P),$$
$$\text{Normal:} \quad (dy/dx)_P(y - y_P) = -(x - x_P).$$

Given the parametric form $x = f(t), y = g(t)$,

$$\left(\frac{dy}{dx}\right)_P = \left(\frac{\dot{y}}{\dot{x}}\right)_P, \text{ where } \dot{y} = \frac{dy}{dt} \text{ and } \dot{x} = \frac{dx}{dt},$$

and the equations become:

$$\text{Tangent:} \quad (y - y_P)\dot{x}_P = (x - x_P)\dot{y}_P,$$
$$\text{Normal:} \quad (y - y_P)\dot{y}_P = -(x - x_P)\dot{x}_P.$$

Example 1. For the ellipse $x^2/a^2 + y^2/b^2 = 1$, by implicit differentiation, $\frac{2x}{a^2} + \frac{2y}{b^2}\frac{dy}{dx} = 0$ or $\frac{dy}{dx} = -\frac{b^2x}{a^2y}$. Hence the equation of the tangent is

$$y - y_P = -\frac{b^2x_P}{a^2y_P}(x - x_P),$$

or

$$\frac{xx_P}{a^2} + \frac{yy_P}{b^2} = \frac{x_P^2}{a^2} + \frac{y_P^2}{b^2} = 1.$$

The equation of the normal is

$$-\frac{b^2x_P}{a^2y_P}(y - y_P) = -(x - x_P),$$

or

$$a^2y_Px - b^2x_Py = (a^2 - b^2)x_Py_P.$$

Example 2. The parametric equations of the ellipse of *Example 1* above are $x = a\cos\theta$ and $y = b\sin\theta$. Hence $\dot{x} = dx/d\theta = -a\sin\theta$, $\dot{y} = (dy/d\theta) = b\cos\theta$ and the equations given above become

$$ay\sin\theta + bx\cos\theta = ab \text{ (Tangent)},$$
$$by\cos\theta - ax\sin\theta = (b^2 - a^2)\sin\theta\cos\theta \text{ (Normal)}.$$

As checks on these equations, we note that they represent lines through the point $x = a\cos\theta, y = b\sin\theta$, the coordinates of which satisfy them. Moreover, when $x = 0$, i.e. $\cos\theta = 0$, the tangent is parallel to the y axis, as it should be.

The equation of the normal can be derived from that of the tangent, $ay\sin\theta + bx\cos\theta = ab$, by the argument that any line perpendicular to the tangent has an equation $by\cos\theta - ax\sin\theta = C$ (Section 9.15). C is then fixed by the condition that this line must pass through P if it is the normal.

9.07. Distance along Curves. A good parameter to use (when it can be calculated) is the distance along the arc of a curve from any chosen point A on the curve. If s is the distance to a point P and $(s + ds)$ the distance to Q, then ds is the arc length from P to Q.

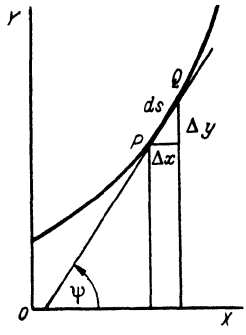

Figure 9.03

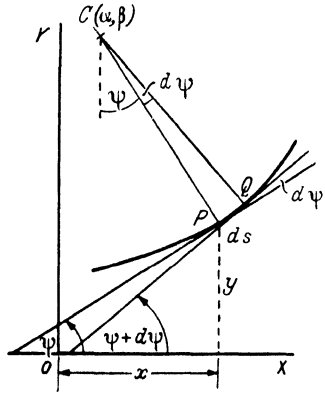

Figure 9.04

If the curve has a continuously turning tangent and Q is near to P, then (*Figure 9.03*)

$$ds = \text{chord } PQ, \text{ to first order} = \{(\Delta x)^2 + (\Delta y)^2\}^{1/2}.$$

To the same accuracy, we can also use

$$(ds)^2 = (dx)^2 + (dy)^2.$$

If a tangent drawn from P in the sense s increasing makes an angle ψ with the x axis, then

$$dx = ds \cos \psi, \quad dy = ds \sin \psi.$$

Example 1. For the circle $x = a \cos \theta, y = a \sin \theta$, the arc distance is $s = a\theta$, if measured from the point $\theta = 0$. The slope of the tangent requires that $\psi = s/a + \frac{1}{2}\pi$.

Example 2. The curve $x = c \sinh^{-1}(s/c), y = \sqrt{(s^2 + c^2)}$ is the coshine curve (catenary) $y = c \cosh (x/c)$. For

$$y = \sqrt{(s^2 + c^2)} = c \cosh (x/c).$$

Moreover $\dfrac{dx}{ds} = \dfrac{c}{\sqrt{(c^2 + s^2)}} = \dfrac{c}{y}, \quad \dfrac{dy}{ds} = \dfrac{s}{\sqrt{(s^2 + c^2)}} = \dfrac{s}{y},$

so that $$(dx)^2 + (dy)^2 = (ds/y)^2(c^2 + s^2) = (ds)^2.$$

9.08. CURVATURE. This has already been discussed qualitatively (Section 1.12). A quantitative measure of curvature is the rate of change of direction of the tangent with distance along the curve.

$$\text{Curvature} = d\psi/ds.$$

Now $d\psi$ is not only the angle between the tangents at P and Q (*Figure 9.04*) but also the angle between the normals. If these normals intersect at a point C distant R from the curve (*N.B.* $CP = CQ$ to first order) then $ds = PQ = Rd\psi$, i.e.

$$R = ds/d\psi = 1/\text{Curvature}.$$

R is called the radius of curvature and C the centre of curvature.

For a curve in Cartesian coordinates, $(dy/dx) = \tan \psi$,

$$\frac{d^2y}{dx^2} = \sec^2 \psi \frac{d\psi}{dx} = \sec^2 \psi \cdot \frac{d\psi}{ds} \cdot \frac{ds}{dx} = \sec^3 \psi \cdot \frac{d\psi}{ds}.$$

Hence the curvature is

$$\frac{1}{R} = \frac{d\psi}{ds} = \cos^3 \psi \frac{d^2y}{dx^2} = \frac{d^2y/dx^2}{\{1 + (dy/dx)^2\}^{3/2}}.$$

Using parametric equations,

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{\dot{y}}{\dot{x}}, \quad \frac{d^2y}{dx^2} = \frac{d}{dt}\left(\frac{\dot{y}}{\dot{x}}\right)\frac{dt}{dx} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\dot{x}^3}.$$

$$\text{Curvature} = \frac{1}{R} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\pm \{\dot{x}^2 + \dot{y}^2)^{3/2}}, \quad (\pm \text{ as } \dot{x} \text{ is } \pm).$$

In practice, dy/dx is often small, and d^2y/dx^2 is a sufficient approximation to the curvature.

The coordinates (α, β) of the centre of curvature C for the point $P(x, y)$ on the curve are

$$\alpha = x - R\sin\psi = x - \frac{dy}{dx}\frac{1+(dy/dx)^2}{d^2y/dx^2} = x - \dot{y}\frac{\dot{x}^2+\dot{y}^2}{\dot{y}\dot{x}-\ddot{x}\dot{y}},$$

$$\beta = y + R\cos\psi = y + \frac{1+(dy/dx)^2}{d^2y/dx^2} = y + \dot{x}\frac{\dot{x}^2+\dot{y}^2}{\dot{y}\dot{x}-\ddot{x}\dot{y}}.$$

The second and third of these formulae are valid, whatever the sign of the radius of curvature R.

The evolute of a curve is the locus of the centres of curvature $C(\alpha, \beta)$ for all points on the curve.

The curve itself is an involute of this locus.

9.09. Polar Coordinates. Where the curve has a natural centre, it is often more convenient to use polar coordinates. The centre is chosen as the origin or pole O, and a line OA through the pole is the fixed line from which angles are measured (*Figure 9.05*). The polar coordinates of any point P are

$$r = OP, \quad \text{and} \quad \theta = \angle POA.$$
$$(r \text{ from } 0 \text{ to } \infty) \qquad (\theta \text{ from } 0 \text{ to } 2\pi \text{ or } -\pi \text{ to } +\pi)$$

The circles r = constant, and straight lines, θ = constant, through the origin, form the background of curves on which polar coordinates are plotted.

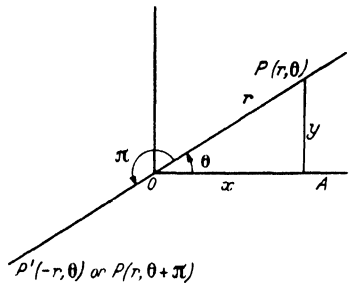

Figure 9.05

These coordinates are often compared with Cartesian coordinates with the same origin and with OA as the x axis.

Then $\quad x = r\cos\theta, \quad y = r\sin\theta.$

A curve is described by a relation $r = f(\theta)$ between the polar

coordinates of points on the curve. Curves having simpler equations, using polar rather than Cartesian coordinates, are closed curves, such as ellipses or ovals, flower or petal-shaped curves and spirals.

A point moving along a spiral (e.g. $r = a/\theta$) goes more than once round the origin. Instead of making r a multi-valued function of θ, it is better to allow θ to have values which are greater than 2π, or negative. This means, however, that the same point P is obtained for (r, θ), $(r, \theta \pm 2\pi)$, etc.

For a similar reason it is a simplification sometimes to allow negative values of r, when they are naturally obtained from the equation, though normally they are excluded on the grounds that there are no points in any direction θ with negative values of r. Negative values of r are measured in the opposite sense to positive values. Thus the point P' in *Figure 9.05* has coordinates, either $(-r, \theta)$ or $(r, \theta + \pi)$.

Geometrical properties not involving angles are not easily discussed in polar coordinates, because the equation of a straight line is not simple. A line parallel to the line $\theta = \alpha$ at a distance d from the origin has the equation

$$r = d/\sin(\theta - \alpha).$$

With certain changes, curve sketching follows the same pattern as in Cartesian coordinates. In the next two examples we discuss the two possibilities, $r \to 0$ and $r \to \infty$, as $\theta \to \alpha$.

Example 1. $r = a \cos 2\theta$. As θ increases from 0 to $\frac{1}{4}\pi$, r decreases from a to zero. Note that this part of the curve does not cross the line $\theta = \alpha$ at any point, but is tangential to it as it comes into the origin. This is a trivial but an easy pitfall.

Example 2. In the reciprocal spiral $r = a/(\theta - \alpha)$, $r \to \infty$ as θ decreases to α. Hence the curve is asymptotic to the line $\theta = \alpha$ or another line parallel to it. This asymptote can be found by comparing the equation of the curve with the equation of a line in this direction. If the asymptote is a distance d from the origin its equation is $r = d/\sin(\theta - \alpha)$. From the equation of the curve,

$$d = \lim_{\theta \to \alpha} r \sin(\theta - \alpha) = \lim_{\theta \to \alpha} a \frac{\sin(\theta - \alpha)}{\theta - \alpha} = a.$$

Now let us consider small displacements along a polar curve. A displacement (*Figure 9.06*) from $P(r, \theta)$ to $Q(r + dr, \theta + d\theta)$ can be made by first increasing θ and then changing r, e.g.

a change $d\theta$ means a transverse displacement (arc PN) $= rd\theta$;

a change dr means a radial displacement $NQ = dr$.

Replacing the arc by a straight line gives a triangle PNQ with a right-angle at N, in the limit $d\theta \to 0$.

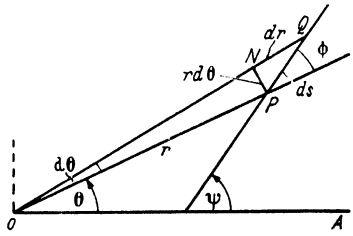

Figure 9.06

Hence ϕ, the angle between the radius vector produced and the tangent drawn in the sense θ increasing, is such that

$$\cot\phi \to \frac{NQ}{PN} = \frac{dr}{rd\theta} = \frac{dq}{d\theta}, \text{ where } q = \log r.$$

The angle between the tangent and the fixed line OA is

$$\psi = \phi + \theta.$$

The stationary values of r are those where $(dr/d\theta) = 0$ and $\phi = \frac{1}{2}\pi$. They usually correspond to points on the curve which are a maximum or minimum distance from the origin. In mechanics they are known as apse points.

An expression for the arc length ds follows from the relation $PQ^2 = NP^2 + NQ^2$. It is

$$ds = \{dr^2 + r^2 d\theta^2\}^{1/2},$$

or

$$ds/r = \{dq^2 + d\theta^2\}^{1/2},$$

$$dr = ds\cos\phi, \quad d\theta = (ds/r)\sin\phi.$$

To calculate curvature we use the similarity between the equations for ϕ and (ds/r) and those in Cartesian coordinates, and proceed as in Section 9.08, using $(dq/d\theta) = \cot\phi$,

$$\frac{d^2q}{d\theta^2} = -\operatorname{cosec}^2\phi\frac{d\phi}{d\theta} = -\operatorname{cosec}^3\phi\left(r\frac{d\phi}{ds}\right) = -r\frac{d\phi}{ds}\left\{1 + \left(\frac{dq}{d\theta}\right)^2\right\}^{3/2}.$$

Hence the curvature is

$$\frac{d\psi}{ds} = \frac{d\theta}{ds} + \frac{d\phi}{ds} = \frac{1}{r}\frac{1 + (dq/d\theta)^2 - (d^2q/d\theta^2)}{\{1 + (dq/d\theta)^2\}^{3/2}}.$$

An expression using parametric equations is obtained in the examples at the end of this chapter.

Example 3. Cardioid $r = a(1 + \cos\theta)$: then

$$\cot\phi = \frac{1}{r}\frac{dr}{d\theta} = \frac{-a\sin\theta}{a(1+\cos\theta)}$$
$$= -\tan\tfrac{1}{2}\phi = \cot\tfrac{1}{2}(\theta - \pi) = \cot\tfrac{1}{2}(\theta + \pi).$$

Of the two possibilities, $\phi = \frac{1}{2}(\theta + \pi)$ and $\phi = \frac{1}{2}(\theta - \pi)$, the first is seen to be correct for the upper half of the curve, where θ lies between 0 and π, and either the first or the second is correct for the lower half, according as θ is negative or is greater than π.

Example 4. For the equiangular spiral $r = ae^{b\theta}$, $(dr/d\theta) = br$. Hence $\cot\phi = b$, i.e. the tangent makes a constant angle with the radius vector. Hence the name. The curvature is

$$\frac{d\psi}{ds} = \frac{d\phi}{ds} + \frac{d\theta}{ds} = \frac{d\theta}{ds} = \frac{1}{r}\frac{1}{(1+b^2)^{1/2}},$$

and is inversely proportional to r.

Example 5. For the reciprocal spiral $r = a/(\theta - \alpha)$,

$$\frac{dr}{d\theta} = -\frac{a}{(\theta-\alpha)^2} = -\frac{r^2}{a}.$$

Hence $\tan\phi = r(d\theta/dr) = -a/r$ and $\phi \to 0$ as $\theta \to \alpha$ and $r \to \infty$. This also shows that there is an asymptote in the direction $\theta = \alpha$. (Cf. *Example 2.*)

9.10. SHORT LIST OF CURVES. It would be satisfactory to include a selection of diagrams of well known curves at this point. This would mean repeating a large number which have been given in various parts of this text. Such curves are listed by name in the index, but as an aid to the reader who is not familiar with the names and also to the reader who is looking for curves with certain characteristics, these and other curves are collected below under various headings*.

(*i*) Curves passing through zero at $x = 0$ and $x = \infty$, with one maximum, include:

$$cx/(x^2 + b^2), \quad cx^m/(x^n + a^n) \quad (n > m), \quad xe^{-mx}.$$

(*ii*) Decay curves tending steadily to zero include:

$a/(x + b)^n$, $c/(x^2 + b^2)$ (witch of Agnesi), ae^{-mx}, a sech mx, $a/(1 + e^{m(x-a)})$, tractrix $y = a$ sech t, $x = a(t - \tanh t)$.

---

* See also REFERENCES, p. 123.

(*iii*) S-shaped curves, zero at the origin, are:

$a(1 - e^{-mx})$, $a \tanh mx = a(1 - e^{-2mx})/(1 + e^{2mx})$, $a \tan^{-1} mx$. $a/(1 + e^{2mx})$ decreases from a, when $x \to -\infty$, to zero as $x \to \infty$.

(*iv*) Potential curves with a minimum, tending to zero as $x \to \infty$:

$$\frac{a}{x^m} - \frac{b}{x^n} \ (m > n), \text{ or } a\left\{\left(\frac{x_e}{x}\right)^m - \frac{m}{n}\left(\frac{x_e}{x}\right)^n\right\},$$

(usually used in kinetic theory).

$$D\{1 - e^{-a(x-x_e)}\}^2 - D, \quad \text{(Morse function)}.$$

$$e^{-mx}(1 - mx)/mx.$$

(*v*) Potential curves without a minimum:

$$\pm 1/x \text{ (Coulomb)}, \quad \pm e^{-r/a}/r \text{ (meson)}.$$

(*vi*) Ovals:

$$r = a + b \cos\theta, \quad r = a(1 + \cos\theta),$$

$$r(1 + e \cos\theta) = l \text{ (eccentricity } e < 1), \quad \text{(ellipse)}.$$

(*vii*) Spirals:

$r = a\theta$ (Archimedes, or parallel), $r = ae^{b\theta}$ (equiangular); with asymptotes, $r = a/\sin n\theta$, $r = a/\theta$ (reciprocal), $r = a/\sinh\theta$.

(*viii*) Roses:

$r^2 = a^2 \cos 2\theta$ (lemniscate),
$r = a \cos n\theta$ ($n = 2, 3, \ldots$), $r = a \sin n\theta$ ($n = 2, 3, \ldots$).

(*ix*) Special curves:

Cissoid: $y^2(2a - x) = x^3$; $x = 2a \sin^2 t$, $y = 2a \sin^2 t \cdot \tan t$.

Folium of Descartes: $x^3 + y^3 - 3axy = 0$;

$$x = 3at/(1 + t^3),$$
$$y = 3at^2/(1 + t^3).$$

Strophoid: $y^2(a - x) = x^2(a + x)$;

$$x = a \cos 2t,$$
$$y = a \cos 2t \cos t.$$

9.11. Triangular Coordinates. Draw a triangle with sides a, b and c. Let P be a point inside the triangle, distant x, y and z from the sides taken in order (*Figure 9.07*).

Let Δ be the area of the triangle. The areas of the smaller triangles CPB, CPA, APB are $\frac{1}{2}ax$, $\frac{1}{2}by$ and $\frac{1}{2}cz$ respectively. Hence

$$ax + by + cz = 2\Delta.$$

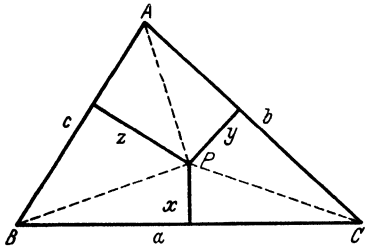

Figure 9.07

As P moves about inside the triangle, the triplet (x, y, z) takes on all sets of positive values consistent with this equation, e.g. the maximum value of x is the perpendicular from A to BC. The triangle can therefore be used to represent three quantities satisfying a condition of this sort and is so used to represent the amounts of the components involved in phase changes.

Example 1. Consider a gas mixture with three components enclosed in a volume V. Let x, y and z be the volumes of the components. Then

$$x + y + z = V.$$

The maximum value of x or y or z is V.

If an equilateral triangle $(a = b = c)$ is drawn, its area is $\Delta = \frac{1}{4}a^2\sqrt{3}$. If x, y and z are the distances from the sides of a point P inside the triangle, then

$$x + y + z = 2\Delta/a = \tfrac{1}{2}a\sqrt{3}.$$

If the scale is chosen so that the length a represents a volume $2V/\sqrt{3}$, then any possible mixture is represented by a point P, whose distances from the sides give the volumes of the components.

Example 2. If three gases have densities d_1, d_2 and d_3 and their masses are m_1, m_2 and m_3, then

$$m_1/d_1 + m_2/d_2 + m_3/d_3 = V.$$

Draw a triangle with sides $a = k/d_1$, $b = k/d_2$, $c = k/d_3$, where k is a scale factor. For this triangle

$$\frac{x}{d_1} + \frac{y}{d_2} + \frac{z}{d_3} = \frac{\Delta}{k}.$$

The values (x, y, z) represent the masses of a mixture, if the scale is properly chosen. The maximum value of x is $\Delta d_1/k$, and that of m_1 is $d_1 V$. Hence the scale for measuring m_1 in terms of the distance x is such that

$$m_1 = d_1 V x/(\Delta d_1/k) = (kV/\Delta)x.$$

This is the same for all three components.

Example 3. In the previous example the scales required for the lengths of the sides and the lengths of the perpendiculars are not the same. Since they represent different quantities this does not matter. If it is necessary that the same scales should be used the following procedure is adopted.

In the equation $ax + by + cz = \Delta$, the area is

$$\Delta = \{s(s-a)(s-b)(s-c)\}^{1/2}$$

in units which are the square of the units of length of side. Any equation $\alpha x + \beta y + \gamma z = \delta$ can be put in this form by introducing a factor K such that $a = K\alpha$, $b = K\beta$, $c = K\gamma$, where K is so chosen that $\Delta = K\delta$.

If $\theta = \frac{1}{2}(\alpha + \beta + \gamma)$ and $s = K\theta$, then

$$K^2\delta^2 = \Delta^2 = s(s-a)(s-b)(s-c) = K^4\theta(\theta-\alpha)(\theta-\beta)(\theta-\gamma)$$

or
$$K^2 = \delta^2/\{\theta(\theta-\alpha)(\theta-\beta)(\theta-\gamma)\}.$$

If three quantities x, y and z, satisfying the equation $ax + by + cz = \Delta$, are also subject to a further restriction imposing another relation between the variables, the point P is in general restricted to a curve lying within the triangle. It is easy to see that a linear equation $Ax + By + Cz = D$ restricts P so that it moves on a straight line. To find the position of this line it is only necessary to find the points where it cuts the sides, i.e. either $x = 0$, $y = 0$ or $z = 0$.

9.12. Curvilinear and Generalized Coordinates. In Cartesian coordinates one set of lines, $x =$ constant, are drawn and another set of lines, $y =$ constant, at right angles to them. Any point P lies at the intersection of two lines, one from one set and one from the other, and its coordinates are the x and y values attached to these two lines.

Polar coordinates can be described in the same way. A set of concentric circles $x^2 + y^2 = r^2$ can be drawn for values of r from 0 to ∞ and a set of lines $y = x \tan\theta$ from the origin for values of θ

from 0 to 2π. Every point P lies at the intersection of one circle and one line, and the polar coordinates of P are the corresponding values of r and θ.

Other sets of curvilinear coordinates are used for special purposes. In every case there are two sets of curves, e.g. $f(x, y) = \lambda$ and $g(x, y) = \mu$, for various values of λ and μ, such that only one curve from each set passes through each point P (except possibly at special points such as the origin in polar coordinates). The appropriate values of λ and μ are the curvilinear coordinates of P. Usually the two sets of curves are such that every member of one set intersects every member of the other set at right angles. For example, let

$$x = c \cosh \lambda \cos \mu, \quad y = c \sinh \lambda \sin \mu.$$

The curves $\lambda =$ constant are the ellipses with a common focus

$$x^2/(c^2 \cosh^2 \lambda) + y^2/(c^2 \sinh^2 \lambda) = 1,$$

and the curves $\mu =$ constant are the hyperbolae

$$x^2/c^2 \cos^2 \mu - y^2/c^2 \sin^2 \mu = 1.$$

Every ellipse intersects every hyperbola at right angles.

Two other ways of specifying plane curves are also mentioned briefly:

(*i*) The intrinsic equation of a curve is $s = f(\psi)$, where s is the arc length along the curve and ψ gives the direction of the tangent (cf. Section 9.07).

(*ii*) The pedal equation is $r = f(p)$, where r is the distance from the origin of any point P on the curve, and p is the length of the perpendicular from the origin to the tangent at P. This form has considerable merit for some curves and is useful in discussing collision problems in molecular physics.

9.13. Two Points. The projections of AB on the axes are (see *Figure 9.08*, p. 118)

$$MN = x_2 - x_1, \quad \text{and} \quad KL = y_2 - y_1,$$

$$\text{and } AB = r = \{(x_2 - x_1)^2 + (y_2 - y_1)^2\}^{1/2}.$$

The direction cosines of AB are:

$$\left.\begin{aligned} \cos \alpha &= (x_2 - x_1)/r \\ \cos \beta &= (y_2 - y_1)/r \end{aligned}\right\} (0 < \alpha, \beta < \pi).$$

A point Q dividing AB into segments in the ratio $AQ : QB = m_1 : m_2$ has coordinates,

$$x = (m_2x_1 + m_1x_2)/(m_1 + m_2), \quad y = (m_2y_1 + m_1y_2)/(m_1 + m_2).$$

The slope of AB is $\tan\theta = (y_2 - y_1)/(x_2 - x_1)$ (*N.B.* θ and α are not always the same).

The area of OAB is $\frac{1}{2}(x_1y_2 - x_2y_1)$, apart from sign.

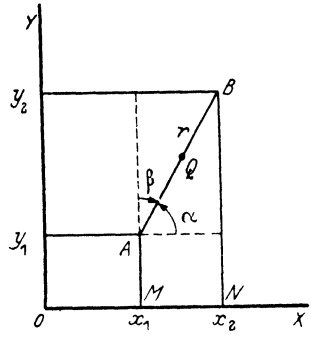

Figure 9.08

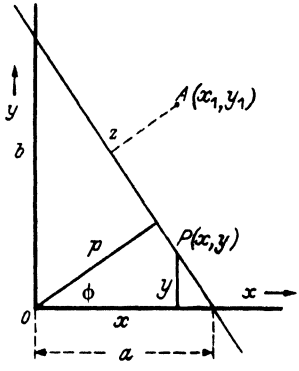

Figure 9.09

9.14. The Straight Line. Let a and b be the intercepts on the axes (*Figure 9.09*), m the slope and p the perpendicular from the origin which is inclined at an angle ϕ to the x axis.

A line $x = a$ is parallel to OY, and a line $y = b$ is parallel to OX.

General equation: $Ax + By + C = 0$.

Slope form: $y = mx + b$, $(B \neq 0)$.

Intercept form: $x/a + y/b = 1$.

Perpendicular form: $x\cos\phi + y\sin\phi = p$.

In terms of A, B and C,

$$m = -A/B, \quad a = -C/A, \quad b = -C/B,$$

$$\frac{\cos\phi}{A} = \frac{\sin\phi}{B} = \frac{p}{-C} = \frac{1}{\sqrt{(A^2 + B^2)}}.$$

The distance of $A(x_1, y_1)$ from the line (see *Figure 9.09*) is

$$z = x_1\cos\phi + y_1\sin\phi - p = (Ax_1 + By_1 + C)/(A^2 + B^2)^{1/2}.$$

9.15. Two Straight Lines. Two lines with slopes m and m' are parallel if $m = m'$, and perpendicular if $mm' = -1$: otherwise an angle β between them is given by

$$\tan\beta = (m - m')/(1 + mm').$$

The two lines $Ax + By + C = 0$ and $A'x + B'y + C' = 0$ are:

(*i*) identical if $A : B : C = A' : B' : C'$;
(*ii*) parallel if $A : B = A' : B'$, but $C \neq C'$;
(*iii*) perpendicular if $AA' = -BB'$,

e.g. $Ax + By + C = 0$ and $Bx - Ay + C' = 0$ are perpendicular.

The line $(Ax + By + C) + \lambda(A'x + B'y + C') = 0$ passes through the point of intersection of the two lines, for all values of λ (or is parallel to them, if they are parallel).

9.16. CONIC SECTIONS. A conic section is the curve of intersection of a plane and a cone. It is also defined as the locus of points P the distance of which from a fixed point, the focus F, is a constant multiple of its distance from a fixed line, the directrix (*Figures 9.10, 9.11* and *9.12*, pp. 121 and 122), i.e.

$$PF = e \,.\, PM.$$

The latus rectum is the chord $LL' = 2l$ parallel to the directrix. The axis is a line XFD through F perpendicular to the directrix. The curves are symmetrical about the axis.

The polar equation, where $FP = r$ and $XFP = \theta$, is

$$r = l/(1 - e\cos\theta) = l/(1 + e\cos\phi).$$

In Cartesian coordinates the equation is of the second degree

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0.$$

The different curves are:

(*i*) Circle: $e = 0$; $a = b$, $h = 0$.
(*ii*) Ellipse: $0 < e < 1$; $ab > h^2$.
(*iii*) Parabola: $e = 1$; $ab = h^2$.
(*iv*) Hyperbola: $e > 1$; $ab < h^2$.

In Cartesian form the tangent to the curve at P is

$$axx_P + h(xy_P + yx_P) + byy_P + g(x + x_P) + f(y + y_P) + c = 0.$$

Chords of slope m are bisected by the diameter

$$(ax + hy + g) + m(hx + by + f) = 0.$$

For a parabola ($h^2 = ab$) the diameters for all m are parallel: otherwise they all pass through the centre, the point of intersection of

$$ax + hy + g = 0 \quad \text{and} \quad hx + by + f = 0.$$

The major and minor axes are those diameters which are perpendilar to the chords they bisect: hence the corresponding slopes m satisfy the equation (cf. Section 9.15)

$$\{-(a + mh)/(h + mb)\}m = -1,$$

i.e.
$$m^2 + m(a - b)/h - 1 = 0.$$

The term in xy is eliminated from the general equation by rotating the axes anticlockwise through an angle θ, where

$$\tan 2\theta = 2h/(a - b).$$

If the resulting equation is

$$a'x^2 + b'y^2 + 2g'x + 2f'y + c' = 0,$$

the terms in x and y are removed by changing the origin to the centre $x = -g'/a', y = -f'/b'$, unless a' or b' is zero (parabola).

9.17. CIRCLE. If the centre is at the origin,

$$x = a\cos\theta, \quad y = a\sin\theta, \quad x^2 + y^2 = a^2.$$

a is the radius, θ is the angle between the radius and the x axis.

If the centre lies at (h, k): $(x - h)^2 + (y - k)^2 = a^2$.

The general second order equation represents a circle if there is no term in xy and if the coefficients of x^2 and y^2 are equal.

The tangent is $x_P x + y_P y = a^2$, or $x\cos\theta + y\sin\theta = a$.

The normal is $y/x = y_P/x_P$, or $y/x = \tan\theta$.

9.18. PARABOLA. With the axis of symmetry as the x axis, and the tangent at the vertex as the y axis, the equation is

$$y^2 = 4ax.$$

The distance of the focus F from the vertex A is a.

Parametric form: $x = at^2$, $y = 2at$.

Latus rectum: $LL' = 4a$.

Directrix: $AD = a$.

Tangent at $P(t)$ is $y_P y = 2a(x + x_P)$, or $ty = x + at^2$.

Normal is $2ay + xy_P = y_P(2a + x_P)$, or $y + tx = at^3 + 2at$.

Tangent bisects the angle FPN ($OT = ON$).

If the axis of the parabola is vertical, i.e. axes interchanged, the equation is $x^2 = 4ay$, and x and y are interchanged in all the equations given above.

The equation $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ represents a parabola if $h^2 = ab$, i.e. if the quadratic terms form a perfect square.

Polar equation: if $FP = r$, $XFP = \theta$,

$$r = l/(1 - \cos\theta) = \tfrac{1}{2}l \operatorname{cosec}^2 \tfrac{1}{2}\theta.$$

If $FP = r$, $AFP = \phi$,

$$r = l/(1 + \cos\phi) = \tfrac{1}{2}l \sec^2 \tfrac{1}{2}\phi.$$

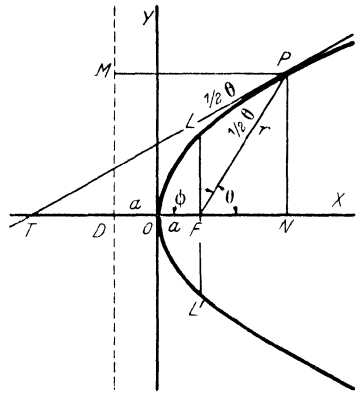

Figure 9.10

9.19. ELLIPSE. This is a closed curve with a second focus F'. The major axis $AA' = 2a$ passes through F and F'. The minor axis BB bisects AA' and FF' at right angles.

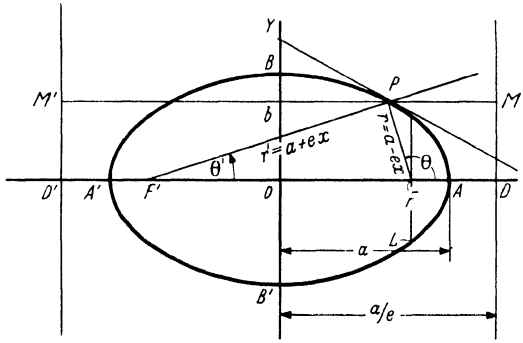

Figure 9.11

Eccentricity ($e < 1$): $b^2 = a^2(1 - e^2)$.
Directrix through D: $OD = OD' = a/e$.
Focus at F: $OF = OF' = ae$.
Focal distances: $FP = a - ex$; $F'P = a + ex$.

Their sum is constant: $FP = F'P = 2a$.
Semi-latus rectum: $l = b^2/a = a(1 - e^2)$.

$$AF = a(1 - e) = l/(1 + e).$$
$$A'F = AF' = a(1 + e) = l/(1 - e).$$
$$AF \,.\, AF' = b^2.$$

Area of ellipse: $\pi ab = \pi a^2(1 - e^2)^{1/2} = \pi b^2/(1 - e^2)^{1/2}$
$= \pi a^{3/2} l^{1/2}$.

Parametric equations: $x = a \cos \theta = a(1 - t^2)/(1 + t^2)$,
$y = b \sin \theta = 2bt/(1 + t^2)$.

Tangent: $xx_P/a^2 + yy_P/b^2 = 1$.
Normal: $a^2x/x_P - b^2y/y_P = a^2 - b^2$.

The tangent bisects the angle between FP and $F'P$ produced.
The director circle has the equation $x^2 + y^2 = a^2$.
The mid points of chords $y = mx + c$ of slope m lie on the diameter $y = m'x$, where $mm' = -(b^2/a^2)$, provided $c^2 < a^2m^2 + b^2$. If $c^2 = a^2m^2 + b^2$, the line touches the ellipse at $x = -a^2m/c$, $y = b^2/c$.
Polar equations: referred to focus F: $r = l/(1 - e \cos \theta)$;
referred to focus F': $r' = l/(1 + e \cos \theta')$.

9.20. HYPERBOLA. This has two branches (*Figure 9.12*). Each branch has an interior and an exterior focus (e.g. F and F'). The

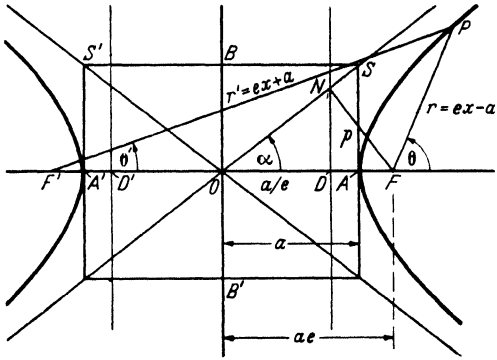

Figure 9.12

major axis $AA' = 2a$ lies on FF' and the minor axis $BB' = 2b$ bisects AA' at right angles. Using the major and minor axes as coordinate axes the equation is

$$x^2/a^2 - y^2/b^2 = 1.$$

Eccentricity ($e > 1$): $b^2 = a^2(e^2 - 1)$.
Asymptotes OS and OS':

$$OS = OS' = ae.$$

Angle 2α between asymptotes: $\cos \alpha = OA/OS = 1/e$,

$$b = ae \sin \alpha.$$

Focus at F: $OF = OS = ae$.
Perpendicular from focus to asymptote:

$$p = ON = OF \sin \alpha = b, \quad ON = OA = a.$$

Focal distances: $FP = ex - a$, $F'P = ex + a$.
Their difference is constant: $F'P - FP = 2a$.
Directrix through D: $OD = a/e$, passes through $\mathcal{N}$.
Semi-latus rectum: $FL = l = b^2/a = a(e^2 - 1)$.
Parameters: (*i*) $x = a \cosh \phi$, $y = b \sinh \phi$;

(*ii*) $x = a \sec \beta$, $y = b \tan \beta$;

(*iii*) $x = 2a(t^2 + 1)/(t^2 - 1)$, $y = 2bt/(t^2 - 1)$.

Tangent: $xx_P/a^2 - yy_P/b^2 = 1$.
Normal: $a^2x/x_P + b^2y/y_P = 1$.
The tangent at P bisects the angle FPF'.

The mid points of parallel chords $y = mx + c$ of slope m ($> \tan \alpha$) lie on the diameter $y = m'x$, where $mm' = b^2/a^2$, provided $c^2 > a^2m^2 - b^2$. If $c^2 = a^2m^2 - b^2$ the line touches the hyperbola at $x^2 = a^2m/c$, $y = -b^2/c$.

Polar equations: with pole F and $\theta = \angle XFP$, $r = l/(1 - e \cos \theta)$;

with pole F' and $\theta' = \angle XF'P$, $r' = l/(1 + e \cos \theta')$.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

FROST, P. *An Elementary Treatise on Curve Tracing.* Macmillan and Co. Ltd, London, 1872.

GRANVILLE, W. A., SMITH, P. F. and LONGLEY, W. R. *Elements of the Differential and Integral Calculus.* G. Ginn and Co., London and New York, 1941.

LAMB, H. *Infinitesimal Calculus.* Cambridge University Press Ltd, London, 1951.

CHAPTER 10

INTEGRATION

Introduction—Numerical integration—Approximations to an area—Mean values—Limits of sums—Integrals—Evaluation of definite integrals—Indefinite and particular integrals—Applications—Motion of a point—Forms of the constant of integration—Interpretation of arbitrary constants—Charge and current—Summation over parts—Areas—Volumes—Lengths of arc—Areas of curved surfaces—Work done by pressure—Mass of a rod of variable density—Mass of suspension—Time for a reaction—Heat given to substance—Numerical evaluation of definite integrals—Euler–Maclaurin summation formula

10.01. INTRODUCTION. Integration is a method of summing or adding a sequence of small changes or small parts. Rather surprisingly, the following three examples are mathematically almost identical:

(*i*) Find the kinetic energy T of a particle moving on a line and acted on by a force $F = f(x)$ when its position is x. It starts from the origin with kinetic energy T_0.

(*ii*) Find the area between the graph of $y = f(x)$, the x axis and the ordinates at $x = a$ and $x = b$.

(*iii*) A great many temperature recordings are made at equal intervals over a period of several days. When plotted as points on a graph (temperature T versus time t) these points lie on a curve $T = f(t)$. Find the average temperature.

Under each of these headings there is a large group of similar problems and the fact that they are all treated the same way explains the importance of integration. The similarity is clear when we try to solve the problems numerically, and this we now do.

10.02. NUMERICAL INTEGRATION. In the first problem the relation between force and kinetic energy is $dT/dx = F = f(x)$. Geometrically, the problem of finding T is equivalent to constructing a graph of T against x through the point A ($x = 0$, $T = T_0$) given the slope of the curve for every value of x (*Figure 10.01*).

An approximate graph can be drawn as follows. Calculate $f(x)$ for a series of equidistant values of x, e.g.

$$x = 0{\cdot}05 \quad 0{\cdot}15 \quad 0{\cdot}25 \quad 0{\cdot}35 \quad \ldots$$

$$f(x) = f_1 \quad f_2 \quad f_3 \quad f_4 \quad \ldots$$

Draw a line AB of slope f_1 from $x = 0$ to $x = 0{\cdot}1$ (*Figure 10.01*): then a line BC of slope f_2 from $x = 0{\cdot}1$ to $x = 0{\cdot}2$: then a line CD with slope f_3 from $x = 0{\cdot}2$ to $x = 0{\cdot}3$, and so on.

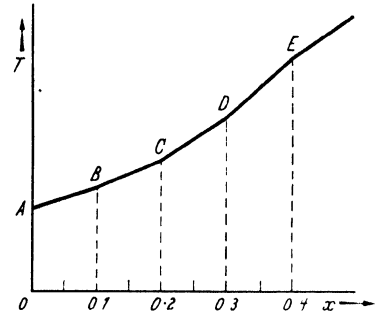

Figure 10.01

The result is a continuous curve, but one with the correct slope only at the points $x = 0{\cdot}05, 0{\cdot}15, 0{\cdot}25$, etc. Since the slope is correct at these points the curve is not badly in error, especially at the start, but there are errors, and these almost certainly increase with x. To reduce them we can increase the number of points where the slope is correct by reducing the intervals in x, so that the lines AB, BC, CD . . . are shorter. If there is any meaning in taking the limit as the intervals tend to zero, presumably the result is a curve with a slope $f(x)$ everywhere.

It should be noted that the starting value T_0 does not affect the shape of the curve in this problem. Altering its value merely shifts the curve up or down the T axis.

In effect we have summed the small changes in T using the approximate formula

$$T(a + h) - T(a) = \Delta T = h\,.f(a + \tfrac{1}{2}h),$$

where every $h = 0{\cdot}1$. Writing dx instead of h, the change in T from $x = 0$ to $x = 0{\cdot}1$ is $f_1 dx_1$ where $dx_1 = 0{\cdot}1$; the change from $x = 0{\cdot}1$ to $x = 0{\cdot}2$ is $f_2 dx_2$, where $dx_2 = 0{\cdot}1$, and so on. The total change in T at the end of n steps, when $x = (0{\cdot}1)n$ is then

$$T - T_0 = f_1 dx_1 + f_2 dx_2 + \ldots + f_m dx_m \text{ (approximately)},$$

$$= \sum_{m=1}^{n} f_m dx_m,$$

where Σ means 'sum'. The rule for calculating each 'f' is the 'halfway' rule, $f_m = f(x)$ at the mid point of each interval dx.

There is the expectation that, for a fixed value of x, the sum Σ ends to a limit which is the true value of T as the intervals dx are decreased and the number n correspondingly increased.

It is not really necessary to draw the graph. Given the rule for calculating the numbers f_m, the sums can be found numerically.

10.03. APPROXIMATIONS TO AN AREA. If the graph $y = f(x)$ is drawn on squared paper, the area between the curve, the x axis and ordinates at $x = a$ and $x = b$ can be found by counting squares. Three methods of counting are in use: (*i*) number of squares wholly inside the area, (*ii*) number wholly or partly inside, (*iii*) number wholly inside plus an estimate of the area of those cut by the curve. The accuracy of the method is related to the differences between these counts, and increases as the size of the squares decreases. The answer for smaller and smaller squares gets close to what we intuitively feel to be the area, and we can meet the difficulty of defining an area with a curved boundary by defining the area as the limit of the area of the counted squares as their size tends to zero.

To express this in the form of a numerical calculation we do not quite follow this procedure (which is left until Chapter 16) but use a simplified version. Divide the interval from a to b on the x axis into n parts each of length $dx = (b - a)/n$. They are numbered dx_1, dx_2, . . . dx_n. The ordinates at the ends of all the parts divide the area into n strips each of width dx (*Figure 10.02*).

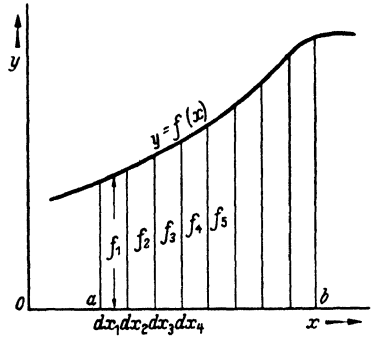

Figure 10.02

In an approximate calculation of the area each strip is replaced by a rectangle on the same base, with a height f suitably chosen for each one. Thus the first strip is replaced by a rectangle of area

$f_1 dx_1$, the second by a rectangle of area $f_2 dx_2$ and so on, the sum being

$$S = f_1 dx_1 + f_2 dx_2 + \ldots + f_n dx_n = \sum_{m=1}^{n} f_m dx_m.$$

This is an approximation to the area. A decrease in the size of the squares of the graph paper now corresponds to a decrease in the width dx of the strips, and an increase in the number n.

It remains to give the rule for choosing the values f_m for each strip. Let $x_m = a + dx_1 + dx_2 + \ldots + dx_{m-1}$. Then f_m is a mean height for the mth strip, and is related to the values of $f(x)$ in the particular interval from x_m to $x_m + dx_m$.

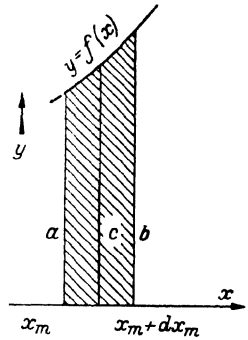

Figure 10.03

Let a and b be the parallel sides of the strip and c the height midway between them. Three simple rules are:

Half-way rule: $f = c$.

Trapezium rule: $f = \frac{1}{2}(a + b)$.

Simpson's rule: $f = \frac{1}{6}(a + 4c + b)$.

In terms of $f(x)$ we see that

$$a = f(x_m), \quad b = f(x_m + dx_m), \quad c = f(x_m + \tfrac{1}{2}dx_m).$$

If we use the half-way rule the expression obtained for the area is identical with that of the previous section, for equal intervals dx and the same number of strips n, apart from a change in the initial value of x. If one sum, obtained in the previous section, tends to a limit as $n \to \infty$, then so does the other, obtained in this section, and it is reasonable to define the area required as this limiting value as $n \to \infty$.

10.04. MEAN VALUES. Finally we turn to the third problem. The temperature readings are at intervals τ of time. Let us measure time

t so that the first reading is half-way through the first interval τ, the second half-way through the second interval τ, and so on. If there are n readings, then the time at the end of the last interval is $t = n\tau$.

Now the recorded temperatures can be obtained from the formula $T = f(t)$. Since each time is half-way through an interval they are clearly identical in form with the values $f_1, f_2, \ldots f_n$, used in the previous examples.

The average recorded temperature is now

$$\bar{f} = (1/n)(f_1 + f_2 + \ldots + f_n) = \{\tau/(b-a)\}\{f_1 + f_2 + \ldots + f_n\}$$

$$= \{1/(b-a)\}\{f_1 dt_1 + f_2 dt_2 + \ldots + f_n dt_n\} = \{1/(b-a)\} \sum_{m=1}^{n} f_m dt_m,$$

if we use $dt_1, dt_2, dt_3, \ldots dt_n$, all equal to τ, to represent the time intervals in order.

The difference between this sum and those of the previous sections is that now $\tau = dt$ is a given size. It is possible to take the readings more frequently and in the limit $\tau \to 0$, the recording is continuous and the correspondence is complete. If τ is small, however, the analogy is useful because, from the definition of a limit, this limit, if it can be calculated, is an approximation to the value of the sum, which is closer the smaller the value of τ.

10.05. Limits of Sums. In the three previous sections a sum has been obtained

$$S = f_1 dx_1 + f_2 dx_2 + \ldots + f_n dx_n = \sum_{x=a}^{b} f' dx,$$

where $$dx_1 + dx_2 + \ldots + dx_n = \sum_{x=a}^{b} dx = b - a.$$

The notation for the sums Σ is different from that of the previous sections. The suffix is omitted for simplicity, and f is written f'. The interval x_m to $x_m + dx_m$ is now described as any interval from x to $(x + dx)$.

Four questions about such sums are:

(*i*) whether a limiting value exists as the number of intervals $n \to \infty$,

(*ii*) if the limit depends on the rule for choosing f', in view of the variety of the choice suggested in the section on areas,

iii) does the limit exist and have the same value when the intervals dx are not equal, and

(*iv*) most important of all, how do we obtain the limiting value when it exists?

The answers to the first three are left as exercises at the end of the chapter, except for one point in (*ii*). For any interval x to $(x + dx)$ the choice of f' must clearly be one of the values of $f(x)$ in that interval, because it represents the mean height of a strip (see *Figure 10.03*). Thus $f' = f(x')$, where $x \leqslant x' \leqslant x + dx$. It therefore lies between f_M and f_L, the greatest and least values of $f(x)$ in the interval. Consequently S lies between $S_M = \Sigma f_M dx$ and $S_L = \Sigma f_L dx$. We prove in the examples that these two have the same limit, and that S has this limiting value, provided:

(*i*) every $dx \to 0$ in the limit,
(*ii*) the total variation of $f(x)$ up and down, between $x = a$ and $x = b$, is finite.

10.06. INTEGRALS. This limit is called the definite integral of $f(x)$ from $x = a$ to $x = b$. It is written $\int_a^b f(x)dx$, i.e. as every $dx \to 0$,

$$\sum_{x=a}^{b} f'dx \to \int_a^b f(x)dx.$$

The numbers a and b are the limits of integration. The upper limit is b and the lower limit is a.

Its evaluation has been suggested in Section 10.02. In that example the integral is the change in the kinetic energy T from $x = a$ to $x = b$, when $dT/dx = f(x)$.

If any function $F(x)$ exists such that $dF/dx = f(x)$, then the value of the integral

$$\int_a^b f(x)dx = \Big[F(x)\Big]_a^b = F(b) - F(a).$$

Any function $F(x)$ of this sort is called an integral of $f(x)$,

$$F(x) = \int f(x)dx.$$

Thus all three types of problem are reduced to a single question: to find an integral of the function $f(x)$ and its values at $x = a$ and $x = b$.

10.07. EVALUATION OF DEFINITE INTEGRALS. This is quite simple when $f(x)$ is recognized as the derivative of another function.

Example 1. $$\int_1^2 4x^3dx = \Big[x^4\Big]_1^2 = 2^4 - 1^4 = 15.$$

Common integrals of this sort are called standard forms. They all correspond to standard derivatives, although the form may be a little different and adapted for integration.

There are rules of integration, derived from the rules of differentiation, making it possible to change vast numbers of integrals into combinations of simpler ones. These rules, methods and results are given in the next chapters.

There remain vast numbers of functions, even simple expressions such as e^{-x^2}, $\sin x/x$, which are not derivatives of simple functions. This is often worrying to students who have been led to believe at school that all integrals are functions they can recognize, but there is no reason for this. Such integrals are evaluated numerically. If any occurs frequently enough it is given a name and tabulated, e.g. $y = \operatorname{erf} x$ is a function having the derivative $dy/dx = e^{-x^2}$. Some definite integrals are evaluated by tricks peculiar to integration. It is these more difficult integrals, which establish integration as a subject quite distinct from differentiation.

What is really surprising is the extent to which we can get along without complications, relying on simple functions.

10.08. INDEFINITE AND PARTICULAR INTEGRALS. An integral $F(x) = \int f(x)dx$ has been defined as a function such that $dF/dx = f(x)$. The integral is not unique, but two integrals only differ by a constant, since their derivatives are the same. If $F(x)$ is one integral, then $F(x) + C$ is another, for any value of C; the number C is called the constant of integration and is an arbitrary constant. This means that its value can be adjusted to suit other requirements.

The integral $\int f(x)dx = F(x) + C$ is called the indefinite integral of $f(x)$, and a particular integral is one that can be derived from it by giving C a definite value, e.g. $F(x)$, $F(x) + 2$.

The particular integral which is zero when $x = a$ is

$$F(x) - F(a) = \int_a^x f(x)dx,$$

obtained by choosing $C = -F(a)$.

The particular integral with the value β when $x = a$ is

$$F(x) + \beta - F(a) = \beta + \int_a^x f(x)dx,$$

obtained by choosing $C = \beta - F(a)$.

To omit C is equivalent to writing $C = 0$, i.e. choosing a particular integral which may not suit a problem. Without definite evidence that $C = 0$, as in one of the examples below, never assume that it is so.

Example 1. Both $(x + a)^2$ and $(x^2 + 2ax)$ are integrals of $2(x + a)$. Their difference is $(x + a)^2 - (x^2 + 2ax) = a^2$, which is a constant.

Example 2. Since $d(x^4)/dx = 4x^3$, then $F(x) = x^4$ is an integral of $4x^3$. The indefinite integral is $x^4 + C$. The integral which is 1 when $x = 2$ is $x^4 + 1 - (2)^4 = x^4 - 15$.

Example 3. Since $d(\sin x)/dx = \cos x$, the indefinite integral of $\cos x$ is $(\sin x + C)$. The particular integral which is unity when $x = \frac{1}{2}\pi$ is

$$\sin x + 1 - \sin(\tfrac{1}{2}\pi) = \sin x,$$

so that $C = 0$.

10.09. APPLICATIONS. The applications listed below fall into three groups, as they resemble the three examples of Sections 10.02, 10.03 and 10.04:

(*i*) solution of equations $dy/dx = f(x)$ (Sections 10.10–10.13),
(*ii*) problems involving sums of parts (Sections 10.14–10.24),
(*iii*) problems involving averages and sums (Section 10.25).

10.10. MOTION OF A POINT. Let us find the position at any time later, of a point which starts from the origin $x = 0$ with a velocity u, and has an acceleration $f \cos pt$ at time t, where f and p are given constants.

If x is the position coordinate, the velocity $v = (dx/dt)$, then the acceleration is (dv/dt), in the sense x increasing. Thus

$$dv/dt = f \cos pt,$$

and $x = 0$, $v = u$ at $t = 0$.

Integrating the acceleration equation we obtain the indefinite integral

$$v = \int f \cos pt \, dt = (f/p) \sin pt + A,$$

where A is the constant of integration. The condition $v = u$ when $t = 0$ (and $\sin pt = 0$) requires that $A = u$. The particular integral needed is therefore

$$v = dx/dt = (f/p) \sin pt + u.$$

A second integration gives

$$x = -(f/p^2) \cos pt + ut + B,$$

where B is to be chosen so that $x = 0$ when $t = 0$, i.e. so that

$$0 = -(f/p^2) + B, \quad \text{or} \quad B = f/p^2.$$

Hence $x = (f/p^2)(1 - \cos pt) + ut,$

is the expression required. To obtain this we need enough information to determine two constants of integration.

10.11. Forms of the Constant of Integration. This constant can always be written in many equivalent ways. Thus the equation $(dx/dt) = (f/p) \sin pt + u$ can be integrated to:

(*i*) $x = -(f/p^2) \cos pt + ut + B;$

(*ii*) $x = -(f/p^2) \cos pt + u(t - \beta);$

(*iii*) $x = (f/p^2)(1 - \cos pt) + ut + b.$

The integrals obtained from all three are exactly the same. For example, putting $B = a$ in the first gives the same integral as putting $\beta = -a/u$ in the second, or $b = a - (f/p^2)$ in the third.

The form of the constant of integration is therefore a matter of convenience.

10.12. Interpretation of Arbitrary Constants. Integration of the equation $dv/dt = f \cos pt$ gives

$$v = (f/p) \sin pt + A.$$

The constant A has a definite meaning or interpretation; $v = A$ when $t = 0$, so that A, however determined, is always the initial velocity.

Now it is not necessary to use symbols such as A, B, etc. to denote constants of integration. It is more sensible to express them in a form suggested by their meaning. Instead of A in the above equation it is better to use v_0 so that it is written

$$v = (f/p) \sin pt + v_0,$$

since v_0 suggests an initial velocity (v when $t = 0$).

Similarly, on integrating this equation to find x, the results

$$x = -(f/p^2) \cos pt + v_0 t + B,$$
$$x = (f/p^2)(1 - \cos pt) + v_0 t + x_0,$$

are equivalent if $B = (f/p^2 + x_0)$. We prefer the second, or something like it, because x_0 has a definite meaning, being the value of x when $t = 0$ (cf. end of Section 10.08).

These considerations are not important mathematically, but they are to the scientist who wishes to bring symbols to life.

10.13. CHARGE AND CURRENT. The current from a discharging condenser is found to decay exponentially with time according to the law $i = i_0 e^{-\lambda t}$. If it is completely discharged at the end of the experiment, what is the charge on it at any time t, and what is the initial charge?

Let $q = f(t)$ be the charge on the condenser at time t. At time $(t + dt)$ it has changed to $(q + dq)$. In this time the charge which has left the condenser to flow round the circuit is $q - (q + dq) = -dq$. The current = (charge)/(time), so that

$$i = -dq/dt.$$

(The minus sign arises because q decreases with time.)

$$\therefore \qquad dq/dt = -i = -i_0 e^{-\lambda t}.$$

Integrating,

$$q = -\int i_0 e^{-\lambda t} dt = (i_0/\lambda) e^{-\lambda t} + A.$$

As the condenser is finally discharged, $q \to 0$ as $t \to \infty$. But $e^{-\lambda t} \to 0$ as $t \to \infty$; hence we must choose $A = 0$. The current at time t is therefore

$$q = (i_0/\lambda) e^{-\lambda t},$$

and the initial charge, obtained by putting $t = 0$, is $q = (i_0/\lambda)$.

10.14. SUMMATION OVER PARTS. As might be expected from the example on areas in Section 10.03 the technique in these problems is quite different from that employed in the previous group. The technique used is explained now to prevent repetition later, and to enable the reader to refer back to it. First re-read Section 10.03.

Integration can be applied to the calculation of a quantity such as an area if:

(*i*) it can be divided into parts labelled by a coordinate x, which lies between a small interval x and $(x + dx)$ for each part; and the whole is the sum of the parts, as x increases from a smaller value a to a larger value b;

(*ii*) the value for each part can be written $f . dx$ or $f(x)dx$ approximately, in the sense that the relative error tends to zero with dx, where f is some value of $f(x)$ between x and $(x + dx)$.

The total value is then given by the limit

$$\sum_{x=a}^{b} f(x)dx \to \int_a^b f(x)dx.$$

In most examples we do not evaluate the integrals, but construct them up to the point of evaluation.

10.15. AREAS. The following questions arise in the calculation of areas.

Example 1. Find the area between the x axis and $y = x/(1 + x^2)$ from the origin up to the ordinate where y is a maximum. The curve is shown in *Figure 6.05.*

Since the derivative $dy/dx = (1 - x^2)/(1 + x^2)^2$ is zero when $x = \pm 1$, the maximum is at $x = 1$. Dividing the area into strips by ordinates at intervals dx, as in the last diagram, the area of each strip is $y\,dx$ to first order, and the total area is

$$\sum_{x=0}^{1} y\,dx = \int_0^1 \frac{x\,dx}{1 + x^2} = \left[\tfrac{1}{2} \ln (1 + x^2)\right]_0^1 = \tfrac{1}{2} \ln 2.$$

Example 2. To find the area between the two curves, $y = x$ and $y = x^2$, when x is positive, we notice that these two curves intersect at $x = 0$ and $x = 1$, so that the area required lies between these two values of x (*Figure 10.04*). When it has been divided into strips by ordinates at intervals dx, the strip between x and $(x + dx)$ extends upwards from the curve $y = x^2$ to the curve $y = x$, so that its length is $(x - x^2)$. To first order, its area is $dA = (x - x^2)dx$. The total area is

$$\sum_{x=0}^{1} (x - x^2)dx = \int_0^1 (x - x^2)dx = \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{1}{6}.$$

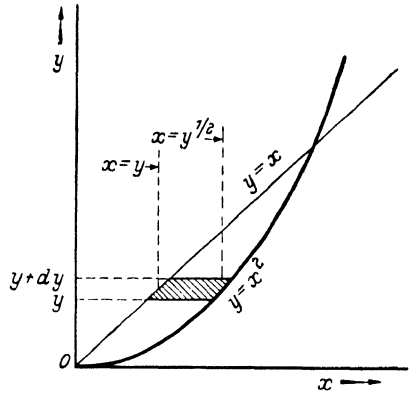

Figure 10.04

Example 3. The area calculated in the previous example can be divided into strips by lines parallel to the x axis at intervals dy, as shown in the *Figure 10.04.* The strip between y and $(y + dy)$

extends horizontally from the line $x = y$ to the curve $x = y^{1/2}$, so that its length is $(y^{1/2} - y)$ (*N.B.* this is positive as y is less than one). The total area is

$$\sum_{y=0}^{1} (y^{1/2} - y)dy = \int_0^1 (y^{1/2} - y)dy = \tfrac{2}{3} - \tfrac{1}{2} = \tfrac{1}{6}.$$

With a little ingenuity areas can often be expressed in simple terms. Sometimes definite integrals can be regarded as areas and evaluated this way.

Example 4. The integral $\int_0^x \sqrt{(r^2 - x^2)}dx$ is the area between the x axis, the circle $x^2 + y^2 = r^2$ and an ordinate, in fact the area $ONBR$ in *Figure 7.01.* This area is the sum of a sector ORB and a triangle OBN the areas of which can be written down at once. Hence

$$\int_0^x (r^2 - x^2)^{1/2}dx = \tfrac{1}{2}r^2\theta + \tfrac{1}{2}xy = \tfrac{1}{2}r^2 \sin^{-1}(x/r) + \tfrac{1}{2}x(r^2 - x^2)^{1/2}.$$

In order that all integrals can be regarded as areas it is necessary to introduce the convention that areas below the x axis are negative. For example, the area between the curve $y = \sin x$ and the x axis, from $x = 0$ to $x = \pi$, is

$$\int_0^\pi y\,dx = \Big[-\cos x\Big]_0^\pi = 2.$$

The area between the curve and the axis between $x = \pi$ and $x = 2\pi$ is

$$\int_\pi^{2\pi} y\,dx = \Big[-\cos x\Big]_\pi^{2\pi} = -2,$$

which is negative because $y = \sin x$ is negative. Hence the integral,

$$\int_0^{2\pi} y\,dx = \Big[-\cos x\Big]_0^{2\pi} = 0,$$

does not give the sum of these areas in the usual sense but their difference, i.e. the area above the x axis equals the area below.

10.16. VOLUMES. Volumes can be calculated by an integration. Consider the sphere obtained by rotating the circle $x^2 + y^2 = a^2$ about the x axis. By taking cross sections normal to the x axis at intervals dx along it, this sphere is divided into slices. (The reader should draw this for himself, cf. *Figure 10.05*, p. 138.) The section

at a distance x from the centre is a circle of radius y and area $A = \pi y^2 = \pi(a^2 - x^2)$.

The slice between x and $(x + dx)$ is not a cylinder, but its volume only differs slightly (first order) from that of a cylinder of uniform section A, and thickness dx. To first order, therefore, its volume is $A\,dx = \pi(a^2 - x^2)dx$.

The total volume is defined as

$$V = \lim_{*} \Sigma A\,dx = \int_{-a}^{+a} \pi(a^2 - x^2)dx = 4\pi a^3/3.$$

The limits $x = \pm a$ define the ends of the sphere.

The same method can be applied to any volume having cross sections A normal to a particular direction which can be given by a formula, as in the case of the sphere. Instead of giving further examples of this, however, we calculate the volume of the sphere by a different method.

By drawing concentric spheres of increasing radius r inside the given sphere it can be divided into spherical shells. The shell between two spheres of internal radius r and external radius $(r + dr)$ is approximately a curved sheet of thickness dr and area $4\pi r^2$ (the area of a sphere). To first order, therefore, its volume is $4\pi r^2 dr$. Summing up over such shells making up the whole sphere, from $r = 0$ to $r = a$, we obtain a total volume

$$\sum_{r=0}^{a} 4\pi r^2 dr = \int_0^a 4\pi r^2 dr = 4\pi a^3/3.$$

10.17. Lengths of Arc. The length of a curved line cannot be found algebraically in the same way as the length of a straight line. It is defined as follows. Divide the given curved line, which is assumed to be the segment AB of a given curve, into short lengths PQ (*Figure 9.03*, Section 9.07). The length of the arc PQ is not known, but the length of the chord $PQ = \Delta s$ can be calculated as explained in Section 9.07 and the difference is slight, for short lengths, if the slope of the curve changes continuously. The sum of all the chords such as PQ from A to B is, say, $\sum_A^B \Delta s$.

The length of the curved line is now defined as the limit of this sum as the number of divisions increases and the length of every part tends to zero,

$$\lim_{\Delta s \to 0} \left(\sum_A^B \Delta s\right) = \int_A^B ds.$$

(*i*) Cartesian coordinates. Let the curve be $x = f(t)$, $y = g(t)$, where t is a parameter increasing steadily from A to B. Then to first order (Section 9.07)

$$\Delta s = ds = \{(dx)^2 + (dy)^2\}^{1/2},$$
$$= \{(dx/dt)^2 + (dy/dt)^2\}^{1/2}dt.$$
$$\text{Arc } AB = \sum_A^B ds = \int_{t_A}^{t_B} \{(dx/dt)^2 + (dy/dt)^2\}^{1/2}dt.$$

If the equation of the curve is $y = f(x)$, put $t = x$, $dx/dt = 1$.

(*ii*) Polar coordinates. If the curve is $r = f(\theta)$ in polar coordinates, then $ds^2 = dr^2 + r^2d\theta^2$ (Section 9.09), and

$$\text{Arc } AB = \sum_A^B ds = \int_{\theta_A}^{\theta_B} \{(dr/d\theta)^2 + r^2\}^{1/2}d\theta.$$

Owing to the square root, arc lengths are usually impossible to evaluate, except numerically.

Various properties of arcs can be expressed in terms of integrals. Thus the mean distance of a curved line from, say, the x axis, is obtained by first dividing it into short lengths PQ, multiplying the length of each by its distance from the x axis ($y\,ds$ to first order) adding all the results and dividing by the length of the curve, i.e.

$$y_g = \frac{\sum_A^B y\,ds}{\text{arc } AB} = \frac{1}{\text{arc}} \int_{t_A}^{t_B} y \left\{ \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right\}^{1/2} dt.$$

Similarly,

$$x_g = \frac{\sum_A^B x\,ds}{\text{arc } AB} = \frac{1}{\text{arc}} \int_{t_A}^{t_B} x \left\{ \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right\}^{1/2} dt.$$

The point with coordinates (x_g, y_g) is the mean centre of the arc. If the equation is given in polar coordinates both x (or y) and ds have to be expressed in terms of polar coordinates but the method is the same.

10.18. AREAS OF CURVED SURFACES. Consider the sphere of *Figure 10.05*, divided into slices. The curved surface of the slice between x and $(x + dx)$ is a strip of length $2\pi y$ (the perimeter of the section) and width $PQ = ds$ (approximately). Its area is therefore $2\pi y\,ds$, to first order. The total curved surface is

$$\Sigma 2\pi y\,ds \to \int 2\pi y\,ds,$$

with suitable limits of integration.

The surface of the sphere can also be divided into zones corresponding to values of the angle θ, where $x = a \cos \theta$, $y = a \sin \theta$. The part between values θ and $(\theta + d\theta)$ has a perimeter $2\pi a \sin \theta$, and a width $PQ = a\, d\theta$, resulting in an area

$$dS = 2a^2 \sin \theta \, d\theta.$$

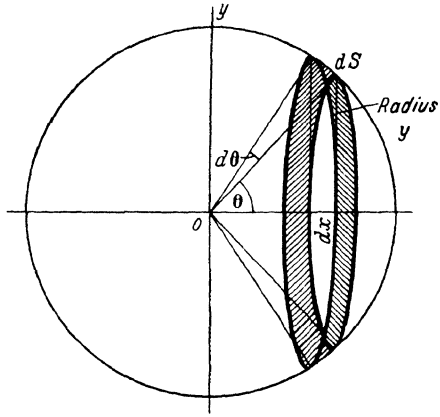

Figure 10.05

The curved surface of the whole sphere is obtained by summing from $\theta = 0$ to $\theta = \pi$, thus,

$$\text{area of sphere} = \int_0^{\pi} 2\pi a \sin \theta \,.\, a\, d\theta = 2\pi a^2 \Big[-\cos \theta \Big]_0^{\pi} = 4\pi a^2.$$

The zone between $\theta = 0$ and $\theta = \alpha$, i.e. from $x = a \cos \alpha$ to $x = a$, has an area $2\pi a^2(1 - \cos \alpha)$.

This integral is one where a serious error can be made. Although the volume of a slice is that of a circular cylinder of thickness dx and radius y, to first order, its surface area does not equal that of the cylinder, which is $2\pi y\, dx$. In fact the ratio of these two approximations is ds/dx which does not tend to unity as $dx \to 0$, but to sec ψ, where ψ is the inclination of the tangent. The approximation $2\pi y\, ds$ is correct to first order, the approximation $2\pi y\, dx$ is not. The reader should convince himself about this.

10.19. Work Done by Pressure. The pressure of a certain mass of gas expanding adiabatically is $p = kv^{\gamma}$ when its volume is v, where k and γ are constants. Let A be the area of a part of the boundary which moves outwards a small distance h during the expansion. The

force on it is pA and the work done on this force is pAh, to first order, any variation in pressure being a higher order effect if h is small. Now Ah = volume covered by this part of the boundary, so that, adding up over the whole boundary, the total work done against the pressure is p times (total volume covered by boundary),

$$= p \text{ (increase in volume of gas)},$$
$$= p\,dv, \text{ say.}$$

Thus the total work done when the gas expands from a volume v_1 to a volume v_2 is

$$\sum_{v=v_1}^{v_2} p\,dv = \int_{v_1}^{v_2} p\,dv = \int_{v_1}^{v_2} kv^\gamma dv = \frac{k}{\gamma+1}\left[v^{\gamma+1}\right]_{v_1}^{v_2}$$
$$= \frac{k}{\gamma+1}(v_2^{\gamma+1} - v_1^{\gamma+1}) = \frac{1}{\gamma+1}(p_2v_2 - p_1v_1),$$

where p_1 and p_2 are the initial and final pressures, respectively.

10.20. Mass of a Rod of Variable Density. Let $m(x)$ give the mass per unit length of the rod at a distance x from one end. The total length is L.

If the rod were divided into short lengths, that between x and $(x + dx)$ has a mass $m(x)dx$ to first order. The total mass is therefore

$$\sum_{x=0}^{L} m\,dx = \int_0^L m(x)dx.$$

10.21. Mass of Suspension. The concentration c, or mass per unit volume of solution, of suspended particles, in a suspension in a cylinder of height h and cross section A, may depend on the depth x below the surface.

Imagine the suspension from $x = 0$ to $x = h$ divided into a number of layers of thickness dx. The layer between x and $(x + dx)$ has a volume $A\,dx$, containing a mass $dm = cA\,dx$ of suspended particles, to first order. The total mass is therefore

$$\Sigma dm = \sum_{x=0}^{h} cA\,dx = \int_0^h cA\,dx.$$

10.22. Time for a Reaction. In a chemical reaction the rate of decrease of one component is given by the formula

$$dx/dt = -kx^n,$$

where x is the concentration of the component, k is a constant and n is the order of the reaction. The time T for half-reaction is the time for the concentration to be reduced to one half of its original value x_0.

The time for the concentration to change from x to $(x + dx)$ is $dt = -dx/kx^n$. The time for the required change is, therefore,

$$T = \sum_{x=x_0}^{\frac{1}{2}x_0} (-dx/kx^n) = -\int_{x_0}^{\frac{1}{2}x_0} \frac{dx}{kx^n} = \left[\frac{1}{k(n-1)x^{n-1}}\right]_{x_0}^{\frac{1}{2}x_0},$$

i.e. $$T = \frac{1}{k(n-1)}\left\{\left(\frac{2}{x_0}\right)^{n-1} - \left(\frac{1}{x_0}\right)^{n-1}\right\},$$

or $$kT = \frac{2^{n-1} - 1}{(n-1){x_0}^{n-1}}.$$

Experimentally the value of n is found by plotting $\ln T$ against $\ln x_0$.

10.23. Heat given to Substance. The specific heat c_p of a substance is defined as $c_p = \lim_{\Delta\theta \to 0} (\Delta Q/\Delta\theta)$, where ΔQ is the heat required to raise the temperature $\Delta\theta$, at constant pressure.

The heat required to raise the temperature from θ_1 to θ_2 can be calculated in terms of c_p. Dividing the temperature rise into a number of small steps $d\theta$, the heat required for each step is $dQ = mc_p d\theta$, to first order, where m is the mass. The total heat required is

$$Q = \sum_{\theta=\theta_1}^{\theta_2} mc_p d\theta = m\int_{\theta_1}^{\theta_2} c_p d\theta.$$

10.24. Numerical Evaluation of Definite Integrals. We shall not dwell on this important subject. The simplest method is to replace the integral by a sum, using the approximate formulae given in Section 10.03. The accuracy depends on the number of intervals chosen, and Simpson's rule is often used because it is usually sufficiently accurate for most purposes. Fuller discussion of these points is to be found in books on numerical analysis.

10.25. Euler–Maclaurin Summation Formula. Using the trapezium rule to evaluate a definite integral (cf. last section and Section 10.03),

$$\int_{x_0}^{x_0+nh} f(x)dx = h\{\tfrac{1}{2}f_0 + f_1 + f_2 + \ldots + f_{n-1} + \tfrac{1}{2}f_n\} - E,$$

where f_r is the value of $f(x)$ when $x = x_0 + rh$, and E is the error.

This formula also gives the value of a sum in terms of an integral, and is used very frequently for this purpose. It is the answer to the problem, discussed in Section 10.04, of how to find the mean value of a large number of readings $f_0, f_1, f_2, \ldots$ which can be expressed by a formula $f(x)$. For this purpose it is changed to

$$f_0 + f_1 + f_2 + \ldots + f_n = \tfrac{1}{2}(f_0 + f_n) + \frac{E}{h} + \frac{1}{h}\int_{x_0}^{x_0+nh} f(x)dx.$$

The Euler–Maclaurin formula gives an expression for E,

$$E = B_2 \frac{h^2}{2!}(f_n' - f_0') + B_4 \frac{h^4}{4!}(f_n''' - f_0''') + B_6 \frac{h^6}{6!}(f_n^{\mathrm{v}} - f_0^{\mathrm{v}}) + \ldots$$

where $f', f''', \ldots$ are the first, third, $\ldots$ derivatives of $f(x)$, and B_{2r} are the Bernoulli numbers,

$$B_2 = \tfrac{1}{6},\ B_4 = -\tfrac{1}{30},\ B_6 = \tfrac{1}{42},\ B_8 = -\tfrac{1}{30},\ B_{10} = \tfrac{5}{66}, \ldots$$

This shows that the error is small when $f(x)$ and its derivatives are small or zero at both ends of the range of integration.

Example 1. The sum $1 + e^{-a} + e^{-2a} + \ldots e^{-na}$ is obtained when $x_0 = 0$, $h = a$ and $f(x) = e^{-x}$. The Euler–Maclaurin theorem gives the value

$$\frac{1}{2}(1 + e^{-na}) + \frac{E}{a} + \frac{1}{a}\int_0^{na} e^{-x}dx = \frac{1}{2}(1 + e^{-na}) + \frac{1}{a}(1 - e^{-na}) + \frac{E}{a}$$

where

$$E = -\frac{a^2}{12}(e^{-na} - 1) + \frac{a^4}{30 \cdot 24}(e^{-na} - 1) - \frac{a^6}{42 \cdot 720}(e^{-na} - 1) \ldots$$

In the limit $n \to \infty$, the geometric series

$$1 + e^{-a} + e^{-2a} + \ldots = \frac{1}{2} + \frac{1}{a} + \frac{E}{a},$$

$$E = \frac{a^2}{12} - \frac{a^4}{30 \cdot 24} + \frac{a^6}{42 \cdot 720} - \ldots$$

In this example, the answer is known. It is in fact

$$1/(1 - e^{-a}),$$

and the expression given is an expansion of this in powers of a. This example is not a good one, but is useful because the answer is known and can be used to determine the numbers $B_2, B_4, \ldots$

EXAMPLES X

1. Prove, by two distinct methods, that the volume of a right circular cone of height h and base radius a is $\frac{1}{3}\pi a^2 h$.

2. The density of population of a town is $nr/(a^2 + r^2)$ per square mile at a distance r miles from the centre of the town, where n and a are constants. Show that the population of a town with perimeter radius ka miles is

$$\int_0^{ka} \frac{2\pi n r^2 dr}{a^2 + r^2} = 2\pi n a(k - \tan^{-1} k).$$

3. Derive the formula $\int 2\pi r \sin\theta\{(dr/d\theta)^2 + r^2\}^{1/2} d\theta$ for a surface of revolution about the line $\theta = 0$, when the generating curve is given in polar coordinates.

Deduce the area of (*i*) the sphere $r = a$, and (*ii*) the sphere $r = 2a \cos\theta$.

4. The mean moment of a collection of magnets, each of moment μ at temperature T in a magnetic field of strength H is, according to Langevin,

$$\bar{\mu} = \mu \int_0^{\pi} e^{\mu H \cos\theta/kT} \cos\theta \sin\theta \, d\theta \Big/ \int_0^{\pi} e^{\mu H \cos\theta/kT} \sin\theta \, d\theta.$$

Find $\bar{\mu}$.

5. A perfect gas expands isothermally from a volume V_1 to a volume V_2. Prove that the work done against the pressure is $RT \ln (V_2/V_1)$.

6. In the notation of Section 10.05, show that $f_M - f_L$ for each interval dx is less than or equal to the variation of $f(x)$, up and down, in the interval.

Prove that $S_M - S_L = \Sigma(f_M - f_L)dx \leqslant dx_m \Sigma|f_M - f_L| \leqslant C dx_m$, where dx_m is the largest of the n intervals dx, and C is the total variation of $f(x)$, up and down, between $x = a$ and $x = b$. Deduce that $S_M - S_L \to 0$ as $dx_m \to 0$ if C is finite.

7. Prove that S_M decreases and S_L increases, when the number of intervals n is increased by dividing one interval into two smaller intervals. Deduce that S_M, S_L, S tend to limits as $n \to \infty$ by this method and, using the previous example, deduce that these limits are equal.

8. Show that the limit of S_M and S_L is independent of the choice of intervals dx by comparing (*i*) values S_M for a division $dx_1, dx_2, \ldots dx_n$ with n intervals ($n = n_1, n_2, n_3, \ldots \to \infty$), (*ii*) values S_M' for a division $dx_1', dx_2', \ldots dx_m'$ into m intervals ($m = m_1, m_2, m_3, \ldots \to \infty$), (*iii*) a superposition of (*i*) and (*ii*) giving S_m'' for a division into n'' intervals ($n'' = n + m$, in general $= n_1 + m_1, n_1' + m_1', \ldots \to \infty$).

Using the previous example prove that $S_L \leqslant S_L'' \leqslant S_M'' \leqslant S_M$, $S_L' < S_L'' < S_M'' < S_M'$, and deduce that in the limit $n \to \infty$, $m \to \infty$, $S = S' = S''$.

9. Use the Euler–Maclaurin theorem to sum $\ln n! = \sum_{m=1}^{n} \ln m$ and to derive the Stirling approximation

$$n! = cn^{n+\frac{1}{2}} e^{-n}(1 + 1/12n),$$

where c is a constant (*N.B.* $c^2 = 2\pi$).

10. The quantum partition function for a rigid rotator is

$$\mathcal{Z} = \sum_{j=0}^{\infty} (2j + 1)e^{-j(j+1)\alpha}$$

where $\alpha = (h^2/8\pi^2 IkT)$. Using the Euler–Maclaurin formula up to B_6, with $f(x) = e^{-\alpha x(x+1)}$ and $h = 1$ (the integration is direct), prove that, when α is small,

$$\mathcal{Z} = \frac{1}{\alpha}\left(1 + \frac{\alpha}{3} + \frac{\alpha^2}{15} + \frac{4\alpha^3}{315} + \cdots\right).$$

The author has found helpful treatments of the subject matter of this chapter in the following books:

GRANVILLE, W. A., SMITH, P. F. and LONGLEY, W. R. *Elements of the Differential and Integral Calculus.* G. Ginn and Co., London and New York, 1941.

HARTREE, D. R. *Numerical Analysis.* Clarendon Press, Oxford, 1952.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering.* Chapter 1. McGraw-Hill Book Co. Inc., New York, 1939.

CHAPTER 11

METHODS OF INTEGRATION

Introduction—Algebraic changes—Logarithmic integrals—Change of variable—Extensions of standard formulae—General remarks—Integration by parts—Substitution to simplify integrals—Different answers—Definite integrals—Positive and negative integrals—Change of variable—Even and odd functions—Periodic functions—Mean value theorems for integrals—Infinite limits and infinite integrands—Derivatives with respect to limits—Mean values of functions—Reduction formulae—Gamma function—Integrals as functions—Integrals in the kinetic theory of gases

11.01. INTRODUCTION. The first part of this chapter includes methods of obtaining integrals of various functions, or expressing the integrals in a convenient form. The second part discusses various problems relating to definite integrals, and their properties.

Some of the standard formulae SF 1–27 (pp. 162–164) are used in the worked examples. In the main, however, only the simplest are used. Remember that the value of an integral can always be checked by differentiation.

11.02. ALGEBRAIC CHANGES. The integrand, or function to be integrated, is changed in form by the use of algebraic or other identities.

$$\int\{(1-x)/x\}dx = \int(1/x - 1)dx = \ln x - x.$$

$$\int \tan^2 x\, dx = \int(\sec^2 x - 1)dx = \tan x - x.$$

$$\int 2 \sin x \cos 2x\, dx = \int(\sin 3x - \sin x)dx = \cos x - \tfrac{1}{3}\cos 3x.$$

11.03. LOGARITHMIC INTEGRALS. A fraction having as its numerator the derivative of the denominator can be integrated by the formula,

$$\int \frac{du/dx}{u}\, dx = \ln (u),$$

which results from $d(\ln u)/dx = (du/dx)/u$.

$$\int \frac{x}{1+x^2}\, dx = \frac{1}{2}\int \frac{d(1+x^2)}{1+x^2} = \frac{1}{2}\ln (1+x^2).$$

$$\int \frac{1}{1+e^{-x}}\, dx = \int \frac{e^x}{1+e^x}\, dx = \ln (1+e^x).$$

$$\int \cot x\, dx = \int \frac{\cos x}{\sin x}\, dx = \ln (\sin x).$$

$$\int \tan x\, dx = \int \frac{\tan x \sec x}{\sec x}\, dx = \ln (\sec x).$$

11.04. CHANGE OF VARIABLE. If u is a function of x, or vice versa, and

$$\int f(x)dx = \int g(u)du,$$

we say that the variable of integration has been changed from x to u.

Let F denote either integral: then $dF/dx = f(x)$ and $dF/du = g(u)$, or, in terms of differentials,

$$dF = f(x)dx = g(u)du.$$

In a change of variable, using the chosen relation between x and u, the product $f(x)dx$ is changed to the form $g(u)du$.

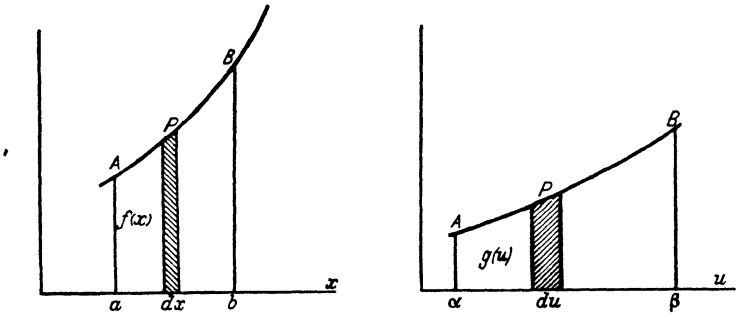

Figure 11.01

Graphically (*Figure 11.01*) this means that, by a change from one graph with x as abscissa to another with u as abscissa, any strip between x and $(x + dx)$ on one graph determines a strip between corresponding values u and $(u + du)$ on the other. Thus the width of the strip is changed from dx to du. To obtain a strip of the same area the height must be changed from $f(x)$ to $g(u)$ where

$$f(x)dx = g(u)du.$$

Given this change in heights, an area in one diagram equals an area in the other.

The limits of integration must also be changed. The first strip at $x = a$ in one starts at the corresponding value $u = \alpha$ in the

other, and the last strip $x = b$ ends at the corresponding value $u = \beta$ in the other.

$$\int_a^b f(x)dx = \int_\alpha^\beta g(u)du.$$

11.05. EXTENSIONS OF STANDARD FORMULAE. Using a simple change of variable many integrals can be expressed as standard forms (cf. Section 12.03).

$$\int u^n du = \frac{u^{n+1}}{n+1} \quad (n + 1 \neq 0), \text{ (SF 1)}.$$

$u = \cos x$; $\int \sin^3 x \cos x\, dx = \int \sin^3 x\, d(\sin x) = \frac{1}{4}\sin^4 x$.

$u = x^2 + 3$; $\int (x^2+3)^2 x\, dx = \int \frac{1}{2}(x^2+3)^2 d(x^2+3)$
$= \frac{1}{6}(x^2+3)^3$.

$$\int e^u du = e^u, \text{ (SF 3)}.$$

$u = \sin x$; $\int e^{\sin x} \cos x\, dx = \int e^{\sin x}\, d(\sin x) = e^{\sin x}$.

$u = -x^2$; $\int x e^{-x^2} dx = -\frac{1}{2}\int e^{-x^2} d(-x^2) = -\frac{1}{2}e^{-x^2}$.

$$\int \frac{du}{\sqrt{(1-u^2)}} = \sin^{-1} u, \text{ (SF 24)}.$$

$$u = x^2; \quad \int \frac{x\, dx}{\sqrt{(1-x^4)}} = \frac{1}{2}\int \frac{d(x^2)}{\sqrt{(1-x^4)}} = \frac{1}{2}\sin^{-1}(x^2).$$

$$u = e\sin\theta; \quad \int \frac{\cos\theta\, d\theta}{\sqrt{(1-e^2\sin^2\theta)}} = \frac{1}{e}\int \frac{d(e\sin\theta)}{\sqrt{(1-e^2\sin^2\theta)}},$$

$$= \frac{1}{e}\sin^{-1}(e\sin\theta).$$

$$u = e^{-\frac{1}{2}x}; \quad \int \frac{dx}{(e^x-1)^{1/2}} = \int \frac{e^{-\frac{1}{2}x}\, dx}{(1-e^{-x})^{1/2}} = -2\sin^{-1}(e^{-\frac{1}{2}x}).$$

The substitution $u = ax + b$, $du = a\, dx$, is very common.

$$\int f(ax+b)dx = \frac{1}{a}\int f(u)du.$$

$$\int (ax+b)^n dx = \frac{1}{a(n+1)}(ax+b)^{n+1}, \quad (n \neq -1).$$

$$\int \frac{dx}{x^2+4x+20} = \int \frac{d(x+2)}{(x+2)^2+16} = \frac{1}{4}\tan^{-1}\frac{x+2}{4}, \quad \text{(SF 20)}.$$

11.06. GENERAL REMARKS. The methods given so far, and certain 'forcing' methods, which ensure the evaluation of certain classes of integral (some are given in the next chapter), are the only methods which guarantee that an integral can be expressed in terms of known functions. It is true, of course, that their scope is extended by a greater knowledge of differentiation or of special functions.

For the rest we proceed by trial and error, trying to change the integral into another which we can integrate. A certain amount of judgment enables us to assess the chance of success, but there remains some doubt as to the result in many cases. To a large extent such judgment is obtained by experience, but this is the point where the art of the mathematician comes into play.

11.07. INTEGRATION BY PARTS. One integral can be related to another, when part of its integrand can be integrated, by the formula

$$\int \left(v \frac{du}{dx}\right) dx = uv - \int \left(u \frac{dv}{dx}\right) dx,$$

or

$$\int v\, du = uv - \int u\, dv.$$

This formula is obtained by integrating $d(uv) = v\, du + u\, dv$.

The same formula for definite integrals is illustrated in *Figure 11.02*.

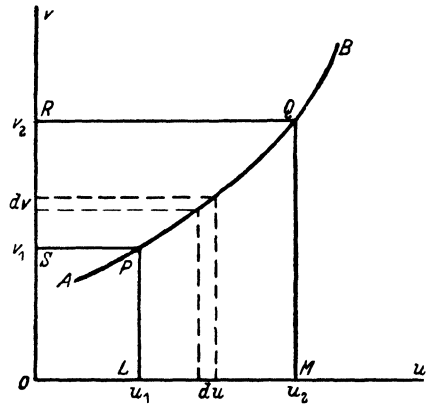

Figure 11.02

$$\text{Area } PLMQ + \text{Area } PSRQ = u_2v_2 - u_1v_1,$$

i.e.

$$\int_{u_1}^{u_2} v\, du + \int_{v_1}^{v_2} u\, dv = \Big[uv\Big]_1^2.$$

If the curve AB relating u and v is defined by parametric equations $u = f(x)$, $v = g(x)$ with x as parameter, this equation is

$$\int_{x_1}^{x_2} v \frac{du}{dx}\,dx = \Big[uv\Big]_{x_1}^{x_2} - \int_{x_1}^{x_2} u \frac{dv}{dx}\,dx.$$

This formula is normally used only when the derivative of v is simpler than v itself.

$$\begin{aligned}\int xe^x\,dx = \int x\,d(e^x) &= xe^x - \int e^x\,d(x),\\ &= xe^x - e^x.\end{aligned}$$

$$\begin{aligned}\int x\cos x\,dx = \int x\,d(\sin x) &= x\sin x - \int \sin x\,dx,\\ &= x\sin x + \cos x.\end{aligned}$$

If there is a choice of factors v and du, one usually gives a worse integral, the other a better. In doubtful cases, work out both and see which is the better.

In the example above

$$\begin{aligned}\int x\cos x\,dx = \int \cos x\,d(\tfrac{1}{2}x^2) &= \tfrac{1}{2}x^2\cos x - \int \tfrac{1}{2}x^2\,d(\cos x),\\ &= \tfrac{1}{2}x^2\cos x + \tfrac{1}{2}\int x^2\sin x\,dx.\end{aligned}$$

This is quite correct, but the last integral is no better than the first.

In the next two examples, integration by parts is used to remove an awkward function by differentiation.

$$\begin{aligned}I = \int x\ln x\,dx = \int \ln x\,d(\tfrac{1}{2}x^2) &= \tfrac{1}{2}x^2\ln x - \int (\tfrac{1}{2}x^2)d(\log x),\\ &= \tfrac{1}{2}x^2\ln x - \int \tfrac{1}{2}x^2(1/x)\,dx,\\ &= \tfrac{1}{2}x^2\ln x - \tfrac{1}{4}x^2.\end{aligned}$$

$$\begin{aligned}I = \int \tan^{-1}x\,dx &= x\tan^{-1}x - \int x\,d(\tan^{-1}x), \quad [N.B.\ dx = d(x).]\\ &= x\tan^{-1}x - \int \frac{x}{1+x^2}\,dx,\\ &= x\tan^{-1}x - \tfrac{1}{2}\ln(1+x^2).\end{aligned}$$

Sometimes repeated integration by parts is needed to complete a process started by the first application. From two of the above examples

$$\begin{aligned}\int x^2\sin x\,dx &= -x^2\cos x + \int 2x\cos x\,dx,\\ &= -x^2\cos x + 2x\sin x + 2\cos x.\end{aligned}$$

The large part of an integral can sometimes be obtained by integration by parts.

$$\int e^{ax} f(x)\,dx = \frac{1}{a}\,e^{ax} f(x) - \frac{1}{a}\int e^{ax}\frac{df}{dx}\,dx,$$

$$= \frac{1}{a}\,e^{ax} f(x) - \frac{1}{a^2}\,e^{ax}\frac{df}{dx} + \frac{1}{a^2}\int e^{ax}\frac{d^2f}{dx^2}\,dx.$$

If the derivatives of $f(x)$ are not much larger in value than $f(x)$, the first term is the biggest for large values of a.

The following result is used in Chapter 13 ($f'(x) = \dfrac{df}{dx}$, etc.).

$$f(b) - f(a) = \int_a^b f'(x)\,dx = -\int_a^b f'(x)\,d(b-x),$$

$$= \Big[-f'(x)(b-x)\Big]_a^b + \int_a^b f''(x)(b-x)\,dx,$$

$$= f'(a)(b-a) - \left[f''(x)\,\frac{(b-x)^2}{2}\right]_a^b + \frac{1}{2!}\int_a^b f'''(x)(b-x)^2\,dx,$$

$$= (b-a)f'(a) + \frac{1}{2!}(b-a)^2 f''(a) + \frac{1}{2!}\int_a^b f'''(x)(b-x)^2\,dx.$$

Integrating by parts n times,

$$f(b) = f(a) + (b-a)f'(a) + \frac{1}{2!}(b-a)^2 f''(a) + \frac{1}{3!}(b-a)^3 f'''(a) + \ldots + \frac{1}{n!} f^{(n)}(a)(b-a)^n + \frac{1}{n!}\int_a^b f^{(n+1)}(x)(b-x)^n\,dx.$$

11.08. SUBSTITUTION TO SIMPLIFY INTEGRALS. A change of variable can eliminate an awkward feature, such as a surd or logarithm.

Example 1. $I = \int \ln x\,dx$. Put $x = e^v$, then $y = \ln x$, $dx = e^v dy$.

$$I = \int y(e^v dy) = ye^v - \int e^v dy = e^v(y-1) = x(\ln x - 1).$$

Example 2. $I = \dfrac{dx}{x\sqrt{(x-4)}}$. Put $t = (x-4)^{1/2}$ to remove the surd. Then $x = t^2 + 4$, $dx = 2t\,dt$.

$$I = \int \frac{2t\,dt}{(t^2+4)t} = \int \frac{2dt}{t^2+4} = \tan^{-1}\tfrac{1}{2}t = \tan^{-1}\sqrt{\frac{x-4}{4}}.$$

Example 3. $I = \int \frac{dx}{a^2 + x^2}$. Put $x = a \tan \theta$, whence $a^2 + x^2 = a^2 + a^2\tan^2 \theta = a^2 \sec^2 \theta$, and $dx = a \sec^2 \theta \, d\theta$.

$$\therefore \quad I = \int \frac{a \sec^2 \theta \, d\theta}{a^2 \sec^2 \theta} = \int \frac{1}{a} d\theta = \frac{\theta}{a} = \frac{1}{a} \tan^{-1} \frac{x}{a}.$$

Example 4. $I = \int \frac{d\theta}{\sin \theta}$ (SF 11, and Section 12.04). Put $t = \tan \frac{1}{2}\theta$. Then $\sin \theta = 2t/(1 + t^2)$ (Section 7.02) and $d\theta = d(2 \tan^{-1} t) = \frac{2dt}{1 + t^2}$.

$$\therefore \quad I = \int \frac{1 + t^2}{2t} \left(\frac{2dt}{1 + t^2} \right) = \int \frac{dt}{t} = \ln t = \ln (\tan \tfrac{1}{2}\theta).$$

Example 5. $I = \int \frac{x}{(1 + x)^4} dx$. Put $u = 1 + x$. $I = \int \frac{u - 1}{u^4} du$, etc.

An example follows where two surds are removed.

Example 6. $I = \int \frac{dx}{(1 + x)^{1/2} + (1 + x)^{1/3}}$.

Put $1 + x = u^6$, $dx = 6u^5 du$.

$$I = \int \frac{6u^5 du}{u^3 + u^2} = \int \frac{6u^3}{1 + u} du = 6 \int \left\{ u^2 - u + 1 - \frac{1}{1 + u} \right\} du,$$

$$= 2u^3 - 3u^2 + 6u - 6 \ln (1 + u),$$

$$= 2(1 + x)^{1/2} - 3(1 + x)^{1/3} + 6(1 + x)^{1/6} - 6 \ln \{1 + (1 + x)^{1/6}\}.$$

Example 7. $\int \frac{dx}{\{(a - x)(x - b)\}^{1/2}}$.

Put $x = a \sin^2 \theta + b \cos^2 \theta$, $(b < x < a)$.

Then $(a - x) = a - a \sin^2 \theta - b \cos^2 \theta = (a - b) \cos^2 \theta$,

$(x - b) = a \sin^2 \theta + b \cos^2 \theta - b = (a - b) \sin^2 \theta$.

$$I = \int \frac{(a - b) 2 \sin \theta \cos \theta \, d\theta}{(a - b) \sin \theta \cos \theta} = \int 2d\theta = 2\theta = 2 \tan^{-1} \left(\frac{x - b}{a - x} \right)^{1/2}.$$

11.09. Different Answers. If different methods give different answers, this may be because the constants of integration are different. (The value of this constant is easily found by putting $x = 0$ or some other value.) It may also be because of an identity.

Example 1. $\int \frac{dx}{3x} = \frac{1}{3}\int \frac{dx}{x} = \frac{1}{3}\ln x$ and $\int \frac{dx}{3x} = \frac{1}{3}\int \frac{d(3x)}{3x} = \frac{1}{3}\ln(3x)$ differ by a constant, namely $\frac{1}{3}\ln 3$.

Example 2. $$\int \frac{dx}{(1+x^2)^{1/2}} = \ln\{x + \sqrt{(x^2+1)}\},$$
$$= -\ln\{\sqrt{(x^2+1)} - x\} = \sinh^{-1} x.$$

Cf. Section 8.13.

A more subtle difference arises when the integrals are different because they apply for different values of x. This occurs when the integrand has a discontinuity.

Example 3. $$\int \frac{dx}{x-a} = \ln(x-a) \quad \text{when} \quad x > a,$$
$$\ln(a-x) \quad \text{when} \quad x < a.$$

Example 4. $$\int \frac{dx}{a^2 - x^2} = \frac{1}{a}\tanh^{-1}\frac{x}{a} \quad \text{when} \quad x < a,$$
$$= \frac{1}{a}\coth^{-1}\frac{x}{a} \quad \text{when} \quad x > a.$$

11.10. DEFINITE INTEGRALS. The evaluation of a definite integral by an analytical method involves the insertion of limits as well as the discovery of the indefinite integral. There are problems connected with the limits and also with the whole process, which are discussed in the following sections.

First some elementary results, which can be proved either regarding the definite integral as the limit of a sum, or using the equation

$$\int_a^b f(x)dx = \Big[F(x)\Big]_a^b = F(b) - F(a).$$

(*i*) Division of range of integration:

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx.$$

This is used when $f(x)$ is expressed by different formulae in the two parts of the range.

(*ii*) Reversal of limits: $\int_a^b f(x)dx = -\int_b^a f(x)dx.$

This is usually used to make the upper limit the larger.

11.11. Positive and Negative Integrals. If $dF/dx = f(x)$, then $F(x)$ increases with x, if $f(x)$ is positive, and decreases when $f(x)$ is negative. Consequently the definite integral,

$$\int_a^b f(x)dx = F(b) - F(a), \ (a < b),$$

is positive if $f(x)$ is positive when $a < x < b$,

negative if $f(x)$ is negative when $a < x < b$.

This result is useful to check integrals when a negative sign may be forgotten.

$$\int_0^\pi \sin x \, dx = \Big[-\cos x \Big]_0^\pi = -\cos \pi + \cos 0 = 2 \text{ must be positive.}$$

11.12. Change of Variable. The change of the limits when the variable of integration is altered has been explained in Section 11.07.

One point, which deserves mention, only arises with definite integrals. The value of a definite integral does not depend on the symbol used for the variable of integration, but only on the limits and the form of the integrand. Thus:

$$\int_0^\infty e^{-x}dx = \int_0^\infty e^{-y}dy = \int_0^\infty e^{-t}dt. \quad (\text{Put } x = t = y.)$$

$$\int_0^x e^{-x}dx = \int_0^x e^{-y}dy = \int_0^x e^{-t}dt = 1 - e^{-x}.$$

Such simple changes of variable simplify the appearance of integrals.

Put $x = \left(\frac{m}{2kT}\right)^{1/2} v$: $\qquad \int_0^\infty e^{-mv^2/2kT}dv = \left(\frac{2kT}{m}\right)^{1/2} \int_0^\infty e^{-x^2}dx.$

11.13. Even and Odd Functions. If $f(x)$ is an even function of x,

$$\int_{-a}^{a} f(x)dx = 2\int_0^a f(x)dx,$$

and if $f(x)$ is an odd function of x (Section 1.16),

$$\int_{-a}^{+a} f(x)dx = 0.$$

These results are obvious from the graphs (*Figure 11.03*). For an even function the areas on either side of the y axis are equal and can

be added, since the graph is symmetrical about the y axis. If the function is odd, they are equal and can be cancelled.

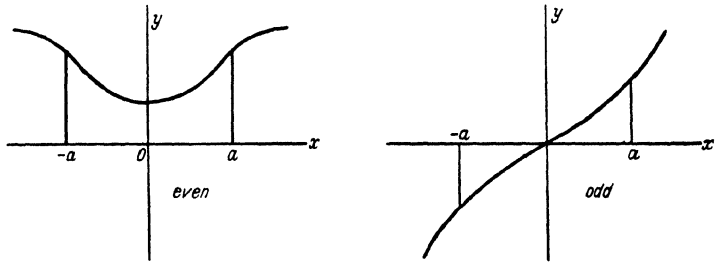

Figure 11.03

Analytically, for the first, when $f(x) = f(-x)$, by a change of variable to $x' = -x$,

$$\int_{-a}^{0} f(x)dx = \int_{-a}^{0} f(-x)dx = -\int_{a}^{0} f(x')d(x') = \int_{0}^{a} f(x')dx' = \int_{0}^{a} f(x)dx.$$

11.14. PERIODIC FUNCTIONS. If $f(x)$ is periodic, with a period L in x, then

$$f(x + L) = f(x), \quad \text{(all } x\text{)}.$$

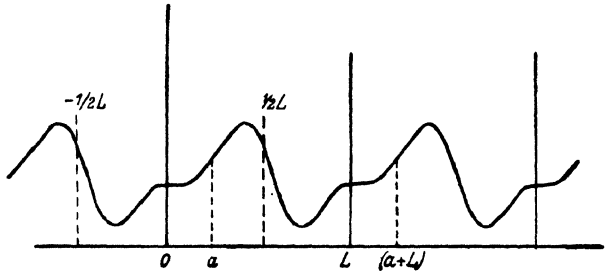

Figure 11.04

It follows (*Figure 11.04*) that

$$\int_{0}^{a} f(x)dx = \int_{L}^{L+a} f(x)dx, \text{ etc.}$$

Moreover

$$\int_{0}^{L} f(x)dx = \int_{a}^{L+a} f(x)dx = \int_{L+a}^{2L+a} f(x)dx = \ldots = \int_{-\frac{1}{2}L}^{\frac{1}{2}L} f(x)dx,$$

and

$$\int_{0}^{nL} f(x)dx = n\int_{0}^{L} f(x)dx, \quad (n \text{ an integer}).$$

Example 1. $\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos\theta\, d\theta = 2\int_{0}^{\frac{1}{2}\pi} \cos\theta\, d\theta,$

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \sin^n\theta\, d\theta = 2\int_{0}^{\frac{1}{2}\pi} \sin^n\theta\, d\theta \text{ if } n \text{ is even, but } 0 \text{ if } n \text{ is odd.}$$

11.15. MEAN VALUE THEOREMS FOR INTEGRALS. To obtain approximate values of integrals and avoid numerical work, the Mean Value Theorems can be used. If $f(x)$ is continuous,

$$\int_a^b f(x)dx = (b-a)f(\xi), \quad (a < \xi < b).$$

Another form is probably more useful. When $g(x) > 0$,

$$\int_a^b f(x)g(x)dx = f(\xi)\int_a^b g(x)dx, \quad (a < \xi < b).$$

To prove the first, let M and L be the greatest and least values of $f(x)$ between $x = a$ and $x = b$. Then

$$\int_a^b (M - f(x))dx > 0 \quad \text{and} \quad \int_a^b (f(x) - L)dx > 0.$$

Hence the integral lies between $M(b-a)$ and $L(b-a)$, and is therefore $f(\xi)(b-a)$, where ξ is between a and b.

Example 1. $\int_0^{\frac{1}{2}\pi} e^{-x/20}\sin x\, dx = e^{-\xi/20}\int_0^{\frac{1}{2}\pi} \sin x\, dx = e^{-\xi/20},$ where $0 < \xi < \frac{1}{2}\pi$. Since $e^{-\xi/20}$ lies between $e^0 = 1$ and $e^{-\pi/40} = 0{\cdot}93$, the value of the integral is 0·96, within 4 per cent.

11.16. INFINITE LIMITS AND INFINITE INTEGRANDS. Infinite limits do not normally give trouble, but one should always remember the definition when there is difficulty.

$$\int_a^\infty f(x)dx = \lim_{b\to\infty} F(b) - F(a).$$

Example 1. $\int_0^\infty e^{-E/kT}\, dE = \left[-kTe^{-E/kT}\right]_0^\infty = -(-kT) = kT.$

If $F(b)$ diverges as $b \to \infty$ the integral has no meaning, but when it oscillates a mathematical physicist should be consulted. For example $\int_0^b \sin x\, dx = (1 - \cos b)$ oscillates as $b \to \infty$: but when this occurs

in physical problems, there are often reasons to believe that it means

$$\lim_{m\to 0}\int_0^\infty e^{-mx}\sin x\,dx = \lim_{m\to 0}\frac{1}{1+m^2} = 1.$$

The factor e^{-mx} is called a convergence factor.

Similarly, integrals across a point where the integrand diverges should not normally occur in elementary work, but the integrand sometimes diverges at one of the limits. Let the divergence occur at $x = c$. Then we define

$$\int_a^c f(x)dx = \lim_{\varepsilon\to 0}\int_a^{c-\varepsilon} f(x)dx,$$

and the integral has a meaning when the limit exists. E.g.,

$$\int_0^1 \frac{dx}{(1-x^2)^{1/2}} = \Big[\sin^{-1}x\Big]_0^1 = \frac{\pi}{2}.$$

11.17. Derivatives with Respect to Limits. The definite integral, $\int_a^b f(x)dx = F(b) - F(a)$, is a function of its limits. From this equation, or the definition of the integral as a sum,

$$\frac{d}{db}\int_a^b f(x)dx = \frac{d}{db}F(b) = f(b), \quad (a \text{ constant}),$$

$$\frac{d}{da}\int_a^b f(x)dx = -\frac{d}{da}F(a) = -f(a), \quad (b \text{ constant}).$$

Thus $$\frac{d}{dx}\int_0^x t^3dt = x^3, \quad \text{and} \quad \frac{d}{dx}\int_x^\infty e^{-t^2}dt = -e^{-x^2}.$$

11.18. Mean Values of Functions. The mean value of a function $y = f(x)$ between $x = a$ and $x = b$ is defined as

$$\bar{y} = \frac{1}{b-a}\int_a^b f(x)dx.$$

The mean square value of the function is

$$\overline{y^2} = \frac{1}{b-a}\int_a^b \{f(x)\}^2dx.$$

Example 1. The mean value of $\sin x$ between 0 and π is

$$\frac{1}{\pi}\int_0^\pi \sin x\,dx = \frac{1}{\pi}\Big[-\cos x\Big]_0^\pi = \frac{2}{\pi}.$$

Example 2. The mean square values of $\sin x$ and $\cos x$ between 0 and π are equal.

$$\frac{1}{\pi}\int_0^\pi \sin^2 x\,dx = \frac{1}{2\pi}\left[x - \sin x\cos x\right]_0^\pi = \frac{1}{2}.$$

$$\frac{1}{\pi}\int_0^\pi (\sin^2 x + \cos^2 x)dx = \frac{1}{\pi}\int_0^\pi dx = 1.$$

The root mean square value is $\{\overline{y^2}\}^{1/2}$, which is not usually the same as $\bar{y}$.

11.19. REDUCTION FORMULAE. A group of integrals, such as

$$I_n = \int x^n e^{-x}dx \ (n = 0, 1, 2, 3, \ldots),$$

is likely to have such values that there are relations between the members of the group.

A reduction formula expresses an integral with a given value of n in terms of others with lower values of n. By using a reduction formula for different values of n it is possible to build up the value of any member of the series. For the example above

$$\int x^n e^{-x}dx = -x^n e^{-x} + n\int x^{n-1}e^{-x}dx,$$

i.e.
$$I_n = -x^n e^{-x} + nI_{n-1}.$$

Now $I_0 = \int e^{-x}\,dx = -e^{-x}$. Hence using the formula,

with $n = 1$, $I_1 = -xe^{-x} + I_0 = -(x + 1)e^{-x}$,

$n = 2$, $I_2 = -x^2e^{-x} + 2I_1 = -(x^2 + 2x + 2)e^{-x}$,

$n = 3$, $I_3 = -x^3e^{-x} + 3I_2 = -(x^3 + 3x^2 + 6x + 6)e^{-x}$.

The notation I_n is used for an integral containing a parameter n.

Examples appear in I.34, 39, 48, 49 of the next chapter.

Integration by parts is a common method of obtaining reduction formulae, and altering the integrand is another.

$$\begin{aligned} I_n = \int \tan^n\theta\,d\theta &= \int \tan^{n-2}\theta\tan^2\theta\,d\theta, \qquad (n \geqslant 2) \\ &= \int \tan^{n-2}\theta\,(\sec^2\theta - 1)d\theta, \\ &= \frac{1}{n-1}\tan^{n-1}\theta - I_{n-2}. \end{aligned}$$

Integration by parts sometimes leads back to the original integral with a different coefficient.

$$I_n = \int \sec^n \theta \, d\theta = \int \sec^{n-2} \theta \, (1 + \tan^2 \theta) d\theta,$$
$$= \int \sec^{n-2} \theta \, d\theta + \int (\sec^{n-2} \theta \tan \theta) \tan \theta \, d\theta,$$
$$= I_{n-2} + \left(\frac{1}{n-2} \sec^{n-2} \theta\right) \tan \theta - \int \frac{\sec^{n-2} \theta}{n-2} \sec^2 \theta \, d\theta,$$
$$= I_{n-2} + \frac{1}{n-2} \sec^{n-2} \theta \tan \theta - \frac{1}{n-2} I_n.$$

$$\therefore (n-2) I_n = (n-2) I_{n-2} + \sec^{n-2} \theta \tan \theta - I_n,$$

i e. $$I_n = \sec^{n-2} \theta \tan \theta / (n-1) + \{(n-2)/(n-1)\} I_{n-2}. \quad (n \geqslant 2.)$$

A more elegant method is to think of an expression, the derivative of which is a combination of the integrands of I_n. This is similar to integration by parts, but it is more flexible.

Example 1. $I_n = \int \cos^n \theta \, d\theta$: $\cos^n \theta$ appears in the derivative of $\cos^{n-1} \theta \sin \theta$.

$$d(\cos^{n-1} \theta \sin \theta)/d\theta = \cos^{n-1} \theta \cos \theta - (n-1) \cos^{n-2} \theta \sin^2 \theta,$$
$$= \cos^n \theta - (n-1)(\cos^{n-2} \theta - \cos^n \theta),$$
$$= n \cos^n \theta - (n-1) \cos^{n-2} \theta.$$

$$\therefore \quad n \int \cos^n \theta \, d\theta - (n-1) \int \cos^{n-2} \theta \, d\theta = \cos^{n-1} \theta \sin \theta,$$

i.e. $$I_n = (1/n) \cos^{n-1} \theta \sin \theta + \{(n-1)/n\} I_{n-2}.$$

Recurrence relations between definite integrals are often simpler than those between indefinite integrals, e.g.

$$\int_0^{\frac{1}{2}\pi} \cos^n \theta \, d\theta = \left[\frac{1}{n} \cos^{n-1} \theta \sin \theta\right]_0^{\pi/2} + \frac{n-1}{n} \int_0^{\frac{1}{2}\pi} \cos^{n-2} \theta \, d\theta,$$
$$= \frac{n-1}{n} \int_0^{\frac{1}{2}\pi} \cos^{n-2} \theta \, d\theta.$$

A list of recurrence relations appears in the next chapter.

11.20. GAMMA FUNCTION. The gamma function $\Gamma(n)$ is defined as

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx \quad (n > 0),$$

when n is positive, and by the recurrence relation

$$n\Gamma(n) = \Gamma(n+1)$$

when n is negative, unless it is a negative integer or zero, when $\Gamma(n)$

becomes infinite. We can prove that this relation is also true when n is positive.

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x}dx = \Big[-x^n e^{-x}\Big]_0^\infty + \int_0^\infty nx^{n-1}e^{-x}dx,$$

$$= n\int_0^\infty x^{n-1}e^{-x}dx = n\Gamma(n), \quad (n > 0);$$

$$\Gamma(1) = \int_0^\infty e^{-x}dx = \Big[-e^{-x}\Big]_0^\infty = 1.$$

When n is a positive integer,

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = n(n-1)\ \ldots\ 2.1\Gamma(1)$$

$$= n!$$

The gamma function is a generalization of the factorials. Even when n is not an integer, the same method is used to express $\Gamma(n)$ in terms of another $\Gamma(n)$ with $0 < n < 1$.

The value of $\Gamma(\frac{1}{2})$ can be found by changing the variable of integration to y, where $x = y^2$.

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty x^{-1/2}e^{-x}dx = \int_0^\infty \frac{1}{y}\,e^{-y^2}(2y\,dy) = 2\int_0^\infty e^{-y^2}dy = \sqrt{\pi},$$

(cf. Section 17.10).

It has been proved also that

$$\Gamma(n)\,.\,\Gamma(1-n) = \pi/\sin n\pi, \quad (0 < n < 1).$$

11.21. INTEGRALS AS FUNCTIONS. Definite integrals are functions of parameters in the integrand other than the variable of integration.

The dependence of the integral on the parameter can often be found by a change in the variable of integration, without evaluating the integral completely.

Example 1. Put $ax = y$: then

$$I(a) = \int_0^\infty x^n e^{-ax}dx = a^{-(n+1)}\int_0^\infty y^n e^{-y}dy,$$

i.e. the integral is inversely proportional to a^{n+1}.

The integral, as a function of a parameter α, has a derivative.

$$I(\alpha) = \int_a^b f(x, \alpha)dx, \qquad I(\alpha + d\alpha) = \int_a^b f(x, \alpha + d\alpha)dx.$$

$$\therefore \quad \frac{dI}{d\alpha} = \lim_{d\alpha \to 0}\int_a^b \frac{f(x, \alpha + d\alpha) - f(x, \alpha)}{d\alpha}\,dx.$$

If $f(x, \alpha)$ is a continuous function of both variables (cf. Section 16.06)

$$\frac{dI}{d\alpha} = \int_a^b \left(\frac{df(x, \alpha)}{d\alpha}\right) dx.$$

This is illustrated in *Figure 11.05*, together with the effect of changing the limits. This formula suggests a method of evaluating integrals.

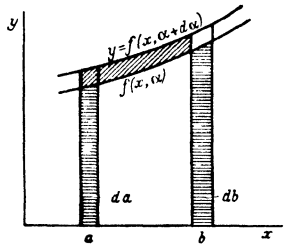

Figure 11.05

Example 2. $I = \int_0^\infty e^{-\alpha x} dx = (1/\alpha)$ is a function of α. The derivative with respect to α is

$$-\frac{1}{\alpha^2} = \int_0^\infty \frac{d}{d\alpha}(e^{-\alpha x}) dx = \int_0^\infty (-xe^{-\alpha x}) dx.$$

$$\therefore \qquad \int_0^\infty xe^{-\alpha x} dx = \frac{1}{\alpha^2}.$$

Proceeding in this way, after n differentiations, we deduce that

$$\int_0^\infty x^n e^{-\alpha x} dx = (n!)/\alpha^{n+1}.$$

Example 3. Evaluate $\int_0^\infty \frac{dx}{(x^2+a^2)^2}$. Consider the integral,

$$\int_0^\infty \frac{dx}{x^2+a^2} = \left[\frac{1}{a}\tan^{-1}\frac{x}{a}\right]_0^\infty = \frac{\pi}{2a}.$$

Differentiate: $$\int_0^\infty -\frac{2a}{(x^2+a^2)^2} dx = -\frac{\pi}{2a^2}.$$

$$\therefore \qquad \int_0^\infty \frac{dx}{(x^2+a^2)^2} = \frac{\pi}{4a^3}.$$

11.22. INTEGRALS IN THE KINETIC THEORY OF GASES. The following integrals constantly appear with $\alpha = (m/2kT)$.

$$A_n = \int_0^\infty v^{2n} e^{-\alpha v^2} dv, \quad B_n = \int_0^\infty v^{2n+1} e^{-\alpha v^2} dv,$$

and their values a_n and b_n when $\alpha = 1$.

Both A_n and B_n are functions of α; their dependence on α can be obtained by changing the variable of integration to $x = \alpha^{1/2} v$, so that $x^2 = \alpha v^2$, e.g.

$$A_n = \int_0^\infty v^{2n} e^{-\alpha v^2} dv = \int_0^\infty \frac{x^n}{\alpha^n} e^{-x^2} \frac{dx}{\alpha^{1/2}} = \frac{a_n}{\alpha^{n+\frac{1}{2}}}; \quad B_n = b_n/\alpha^{n+1},$$

since $a_n = \int_0^\infty x^{2n} e^{-x^2} dx, \quad b_n = \int_0^\infty x^{2n+1} e^{-x^2} dx$ (in terms of x).

When $n = 0$ the values are:

$$a_0 = \int_0^\infty e^{-x^2} dx = \tfrac{1}{2}\sqrt{\pi}. \qquad \therefore A_0 = \frac{a_0}{\alpha^{1/2}} = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}.$$

$$b_0 = \int_0^\infty x e^{-x^2} dx = \tfrac{1}{2}. \qquad B_0 = \frac{b_0}{\alpha} = \frac{1}{2\alpha}.$$

The other values are obtained by differentiating A and B with respect to α (cf. Section 16.01).

$$\frac{dA_n}{d\alpha} = \frac{d}{d\alpha} \int_0^\infty v^{2n} e^{-\alpha v^2} dv = -\int_0^\infty v^{2n+2} e^{-\alpha v^2} dv = -A_{n+1}.$$

Similarly, $$\frac{dB_n}{d\alpha} = -B_{n+1}.$$

In this way we obtain:

$$A_n(\alpha) = \int v^{2n} e^{-\alpha v^2} dv = \frac{1.3.5 \ldots (2n-1)}{2^{n+1}} \sqrt{\frac{\pi}{\alpha^{2n+1}}},$$

$$a_n = \frac{1.3 \ldots (2n-1)\sqrt{\pi}}{2^{n+1}}.$$

$$B_n(\alpha) = \int v^{2n+1} e^{-\alpha v^2} dv = \frac{n!}{2\alpha^{n+1}}, \qquad b_n = \tfrac{1}{2}(n!).$$

EXAMPLES XI

1. Integrating by parts, show that

$$\int \sec^3 \theta \, d\theta = \tan\theta \sec\theta - \int (\sec^2\theta - 1) \sec\theta \, d\theta.$$

Deduce the value of $\int \sec^3\theta \, d\theta$ in terms of $\int \sec\theta \, d\theta$.

2. Using Wien's law (Example *6*, Chapter 8) show that the total energy is

$$\int_0^\infty E_\lambda d\lambda = 6kT^4/a^4.$$

Use $x = a/\lambda T$ as a variable.

3. In the virial expansion $pv = \text{const.}\left(1 - \frac{b}{v} + \frac{c}{v^2} \cdots\right)$, for the pressure of a gas in terms of the volume, the coefficient b for a gas at low pressure is

$$b = \int_0^\infty 2\pi r^2(1 - e^{-V/kT})dr,$$

where $V = f(r)$ is the potential energy of the interaction of two molecules. Using $x = (\lambda a^3/r^3)$ as a variable, evaluate this when $V = \infty$ $(r < a)$, $V = (\lambda^2 a^6 kT/r^6)$, $(r > a)$. (Use the results of Section 11.22.)

4. The Sommerfeld theory of the atom asserts that the angular momentum p_θ has values $k\hbar$ where k is an integer, and the radial momentum, p, is such that

$$\frac{1}{2\pi}\int_0^{2\pi} p\frac{dr}{d\theta}\,d\theta = (n - k)\hbar.$$

Given $p = \frac{p_\theta}{r^2}\frac{dr}{d\theta}$, evaluate this integral for the ellipse $r(1 + e\cos\theta) = l$ and deduce that $k = n(1 - e^2)^{1/2}$ (use SF 83).

5. The elastic scattering of electrons by neutral heavy atoms has an approximate cross section,

$$\sigma = \int_0^\pi 2\pi f^2 \sin\theta\, d\theta,$$

where $f(\theta) = (2mZe^2/\hbar^2 K)\int_0^\infty \sin Kr \,.\, e^{-r/a}dr$, and $K = 2k\sin\frac{1}{2}\theta$.

Prove that $$\sigma = \frac{16\pi m^2 Z^2 e^4 a^4}{\hbar^4(1 + 4k^2a^2)}.$$

6. Optically active solute rotates the plane of polarization, of a beam of light passing through a solution, through an angle $\theta = \alpha c_0 l$, where $c = c_0$ is the concentration and l the length of the path of the beam. If c is not uniform, show that the same formula holds if c_0 is the average value of c with respect to distance over the whole path.

CHAPTER 12

INTEGRALS

Introduction—Standard forms—Rational functions—Trigonometric integrals—Exponential functions—Logarithmic functions—Integrals involving linear surds—Quadratic surds (i)—Quadratic surds (ii)—Definite integrals—Reduction formulae

12.01. INTRODUCTION. This chapter is primarily a list of indefinite and definite integrals, but it contains also formal and other methods of integrating:

(*i*) rational functions of x,
(*ii*) trigonometric functions not involving surds,
(*iii*) hyperbolic functions not involving surds,
(*iv*) functions involving x and surds of the forms $\sqrt{(bx + c)}$ and $\sqrt{(ax^2 + bx + c)}$.

For reference the integrals are numbered, the standard forms having letters SF and the remainder the letter I. For reference, symbols are used instead of numbers in the integrands.

The reader who needs practice in integration should work them using numbers instead of symbols. If he encounters difficulties, he should differentiate the answer and work backwards.

This list is by no means complete and it is advisable for the reader to prepare his own list to meet his special needs.

12.02. STANDARD FORMS. Integrals, which are not the reverse of formulae for derivatives already given, are proved in the sections indicated. The notation is used to remind the reader of the change of variable $\int f(u)\,du = \int f(u)\,(du/dx)\,dx$.

1. $\int u^n du = \dfrac{u^{n+1}}{n+1}, \quad (n + 1 \neq 0).$

2. $\int \dfrac{du}{u} = \log u, \qquad (u > 0).$
$\qquad = \ln(-u), \ (u < 0).$

3. $\int e^u du = e^u.$

4. $\int \cos u\, du = \sin u.$

5. $\int \sin u\, du = -\cos u.$

6. $\int \sec^2 u\, du = \tan u.$

7. $\int \operatorname{cosec}^2 u\, du = -\cot u.$

8. $\int \tan u\, du = \ln(\sec u),$ (Section 11.03).

9. $\int \cot u\, du = \ln(\sin u),$ (Section 11.03).

10. $\int \sec u\, du = \ln(\sec u + \tan u),$
$= \ln\{\tan(\frac{1}{2}u + \frac{1}{4}\pi)\},$
$= \ln \dfrac{1 + \sin u}{\cos u}.$

11. $\int \operatorname{cosec} u\, du = \ln(\operatorname{cosec} u - \cot u),$
$= \ln\{\tan \tfrac{1}{2}u\},$
$= \ln \dfrac{1 - \cos u}{\sin u},$ (Section 11.08).

12. $\int \cosh u\, du = \sinh u.$

13. $\int \sinh u\, du = \cosh u.$

14. $\int \operatorname{sech}^2 u\, du = \tanh u.$

15. $\int \operatorname{cosech}^2 u\, du = -\coth u.$

16. $\int \tanh u\, du = \ln(\cosh u),$ (cf. Section 11.03).

16. $\int \coth u\, du = \ln(\sinh u),$ (cf. Section 11.03).

18. $\int \operatorname{sech} u\, du = \sin^{-1}(\tanh u) = 2\tan^{-1}(\tanh \tfrac{1}{2}u),$
or $2\tan^{-1}(e^u).$

19. $\int \operatorname{cosech} u\, du = \ln(\tanh \tfrac{1}{2}u).$

20. $\displaystyle\int \frac{du}{a^2 + u^2} = \frac{1}{2a}\tan^{-1}\left(\frac{u}{a}\right).$

21. $\displaystyle\int \frac{du}{a^2 - u^2}\ (|u| < a) = \frac{1}{2a}\ln\frac{a+u}{a-u} = \frac{1}{a}\tanh^{-1}\frac{u}{a}.$

22. $\displaystyle\int \frac{du}{u^2 - a^2}\ (|u| > a) = \frac{1}{2a}\ln\frac{u+a}{u-a} = -\frac{1}{a}\coth^{-1}\frac{u}{a}.$

23. $\displaystyle\int \frac{\alpha u + \beta}{au^2 + bu + c}\, du = \int \frac{Cv + D}{v^2 + \lambda}\, dv,$
where $v = au + \frac{1}{2}b$, $\lambda = ac - \frac{1}{4}b^2.$

24. $\displaystyle\int \frac{u\, du}{u^2 + a^2} = \frac{1}{2}\ln\left(1 + \frac{u^2}{a^2}\right).$

25. $\int \frac{du}{\sqrt{(a^2 + u^2)}} = \sinh^{-1}\frac{u}{a}$, or $\ln\{u + (u^2 + a^2)^{1/2}\}$.

26. $\int \frac{du}{\sqrt{(a^2 - u^2)}} = \sin^{-1}\frac{u}{a}$, or $-\cos^{-1}\frac{u}{a}$.

27. $\int \frac{du}{\sqrt{(u^2 - a^2)}} = \cosh^{-1}\frac{u}{a}$, or $\ln\{u + (u^2 - a^2)^{1/2}\}$.

12.03. RATIONAL FUNCTION OF x. By the method of partial fractions (Section 6.18), a rational function can be expressed as a polynomial plus a sum of fractions, i.e. SF 20–24, or

$$\int \frac{A}{(ax+b)^n}\,dx, \quad \int \frac{Bx + C}{(x^2 + 2bx + c)^n}\,dx = \int \frac{B(x+b) + (C - bB)}{(x^2 + 2bx + c)^n}\,dx \qquad (c > b^2,\ n > 0).$$

The following types of integral occur:

28. $\int \frac{dx}{ax - b} = \frac{1}{a}\ln(ax - b)$, $(ax > b)$; or $\frac{1}{a}\ln(b - ax)$, $(ax < b)$.

29. $\int \frac{dx}{(ax - b)^n} = -\frac{1}{(n-1)a}\frac{1}{(ax-b)^{n-1}}$, $(n > 1)$.

30. $\int \frac{dx}{a^2 + b^2x^2} = \frac{1}{ab}\tan^{-1}\frac{bx}{a}$, (SF 20).

31. $\int \frac{x\,dx}{a^2 + x^2} = \frac{1}{2}\ln(a^2 + x^2)$.

32. $\int \frac{dx}{(a^2 + x^2)^2} = \frac{1}{2a^2}\left\{\frac{x}{x^2 + a^2} + \frac{1}{a}\tan^{-1}\frac{x}{a}\right\}$.

33. $\int \frac{x\,dx}{(a^2 + x^2)^2} = -\frac{1}{2(a^2 + x^2)}$.

34. $\int \frac{(n-1)\,dx}{(a^2 + x^2)^n} = \frac{1}{2a^2}\left\{\frac{x}{(x^2 + a^2)^{n-1}} + \int \frac{(2n-3)\,dx}{(a^2 + x^2)^{n-1}}\right\}$, $(n > 1)$.

35. $\int \frac{x\,dx}{(a^2 + x^2)^n} = -\frac{1}{2(n-1)}\cdot\frac{1}{(a^2 + x^2)^{n-1}}$.

36. $\int \frac{x + b}{x^2 + 2bx + c}\,dx = \frac{1}{2}\ln(x^2 + 2bx + c)$.

37. $$\int \frac{x+b}{(x^2+2bx+c)^n}\,dx = \frac{-1}{2(n-1)}\,\frac{1}{(x^2+2bx+c)^{n-1}}, \quad (n \neq 1).$$

38. $$\int \frac{dx}{x^2+2bx+c} = \int \frac{d(x+b)}{(x+b)^2+(c+b^2)},$$
$$= \frac{1}{(c-b^2)^{1/2}} \tan^{-1} \frac{x-b}{(c-b^2)^{1/2}}.$$

39. $$\int \frac{dx}{(x^2+2bx+c)^n} = \frac{x+b}{(n-1)(c-b^2)(x^2+2bx+c)^{n-1}},$$
$$+ \frac{2n-3}{(n-1)(c-b^2)} \int \frac{dx}{(x^2+2bx+c)^{n-1}}, \quad (n > 1).$$

The following integrals are examples of the method.

40. $$\int \frac{x\,dx}{ax-b} = \frac{x}{a} + \frac{b}{a^2} \ln \left| 1 - \frac{ax}{b} \right|.$$

41. $$\int \frac{dx}{x\,(ax+b)} = \frac{1}{b} \ln \frac{x}{ax+b}.$$

42. $$\int \frac{x^2 dx}{ax-b} = \frac{x^2}{2a} + \frac{bx}{a^2} + \frac{b^2}{a^3} \ln \left| 1 - \frac{ax}{b} \right|.$$

43. $$\int \frac{dx}{x(ax+b)^2} = \frac{1}{b(ax+b)} - \frac{1}{b^2} \ln \frac{ax+b}{x}.$$

44. $$\int \frac{dx}{(ax+b)(a'x+b')} = \frac{1}{ab'-a'b} \ln \frac{ax+b}{a'x+b'}, \quad \left(\frac{a}{a'} \neq \frac{b}{b'}\right).$$

45. $$\int \frac{dx}{(a^2-x^2)^2} = \frac{x}{2a^2(a^2-x^2)} + \frac{1}{4a^3} \ln \left| \frac{a+x}{a-x} \right|.$$

46. $$\int \frac{dx}{x(ax^2+c)} = \frac{1}{2c} \ln \frac{x^2}{ax^2+c}.$$

47. $$\int \frac{dx}{x^2(ax^2+c)} = -\frac{1}{cx} - \frac{a}{c} \int \frac{dx}{ax^2+c}.$$

48. $$(n-1) \int \frac{dx}{x^n(ax^2+c)} = -\frac{1}{cx^{n-1}} - \frac{a}{c} \int \frac{(n-1)\,dx}{(ax^2+c)x^{n-2}}, \quad (n > 1).$$

49. $$(n-1) \int \frac{dx}{(ax^2+c)^n} = \frac{x}{2c(ax^2+c)^{n-1}} + \frac{2n-3}{2c} \int \frac{dx}{(ax^2+c)^{n-1}},$$
$(n > 1)$. (cf. I.41.)

50. $$\int \frac{dx}{x^3 - a^3} = \frac{1}{3a^2}\int\left\{\frac{1}{x-a} - \frac{x+2a}{x^2+ax+a^2}\right\}dx,$$

and using (I.36) and (I.38),

$$= \frac{1}{6a^2}\left\{2\ln(x-a) - \ln(x^2+ax+a^2) - 3a\cdot\frac{2}{a\sqrt{3}}\tan^{-1}\frac{x+\frac{1}{2}a}{\frac{1}{2}a\sqrt{3}}\right.,$$

$$= \frac{1}{6a^2}\left\{\ln\frac{(x-a)^3}{x^3-a^3} - 2\sqrt{3}\tan^{-1}\frac{2x+a}{a\sqrt{3}}\right\}.$$

51. $$\int \frac{x\,dx}{x^3 - a^3} = \frac{1}{6a}\left\{\ln\frac{(x-a)^3}{x^3-a^3} + 2\sqrt{3}\tan^{-1}\frac{2x+a}{a\sqrt{3}}\right\}.$$

52. $$\int \frac{x^2\,dx}{x^3 - a^3} = \tfrac{1}{3}\ln(x^3 - a^3), \quad (x < a).$$

53. $$\int \frac{dx}{x^3 + a^3}$$ etc.: change (a) to $(-a)$ in (I.52)–(I.54).

12.04. Trigonometric Integrals. The use of identities and integration by parts is very useful.

Difficult rational functions of $\sin\theta$ and $\cos\theta$ (cf. SF 10 and 11) are reduced to rational functions of $t = \tan\frac{1}{2}\theta$, and then integrated. The quantities occurring are (Section 7.02):

$$\sin\theta = \frac{2t}{1+t^2}, \quad \cos\theta = \frac{1-t^2}{1+t^2}, \quad d\theta = \frac{2}{1+t^2}\,dt.$$

There are first the standard formulae SF 4–11.

54. $$\int \sin^2\theta\,d\theta = \int\tfrac{1}{2}(1 - \cos 2\theta)d\theta = \tfrac{1}{2}\theta - \tfrac{1}{4}\sin 2\theta,$$
$$= \tfrac{1}{2}(\theta - \sin\theta\cos\theta).$$

55. $$\int \cos^2\theta\,d\theta = \tfrac{1}{2}(\theta + \sin\theta\cos\theta).$$

56. $$\int \sin^2 p\theta\,d\theta = \frac{1}{2p}(p\theta - \sin p\theta\cos p\theta).$$

57. $$\int \tan^2\theta\,d\theta = \tan\theta - \theta,$$ (Section 11.02).

58. $$\int \cot^2\theta\,d\theta = -\cot\theta - \theta.$$

59. $$\int \sin^3\theta\,d\theta = -\tfrac{1}{12}\cos 3\theta - \tfrac{3}{4}\cos\theta.$$

60. $$\int \cos^3\theta\,d\theta = \tfrac{1}{12}\sin 3\theta + \tfrac{3}{4}\sin\theta.$$

61. $$\int \tan^3\theta\,d\theta = \int(\sec^2\theta - 1)\tan\theta\,d\theta = \tfrac{1}{2}\tan^2\theta + \ln(\cos\theta).$$

62. $\int \sec^3\theta\, d\theta = \frac{1}{2}\sec\theta\tan\theta + \frac{1}{2}\ln(\sec\theta + \tan\theta)$.

(Example *1*, Chapter 11, and SF 10.)

63. $\displaystyle\int \frac{1+\cos\theta}{\sin\theta}\,d\theta = \int \frac{\cos\frac{1}{2}\theta}{\sin\frac{1}{2}\theta}\,d\theta = 2\ln(\sin\tfrac{1}{2}\theta)$, or $\ln(1-\cos\theta)$.

64. $\displaystyle\int \frac{\sin\theta}{1-\cos\theta}\,d\theta = \ln(1-\cos\theta)$.

65. $\displaystyle\int \sin^n\theta\cos\theta\,d\theta = \frac{\sin^{n+1}\theta}{n+1}$, $(n+1\neq 0)$.

66. $\displaystyle\int \cos^n\theta\sin\theta\,d\theta = -\frac{\cos^{n+1}\theta}{n+1}$, $(n+1\neq 0)$.

(There are many reduction formulae, used mainly as definite integrals, see Section 11.19 and below.)

67. $\displaystyle\int \sin a\theta\sin b\theta\,d\theta = \frac{1}{2}\left\{\frac{\sin(b-a)\theta}{b-a} - \frac{\sin(b+a)\theta}{b+a}\right\}$,

$a\neq b$. (cf. I.58).

68. $\displaystyle\int \cos a\theta\cos b\theta\,d\theta = \frac{1}{2}\left\{\frac{\sin(b-a)\theta}{b-a} + \frac{\sin(b+a)\theta}{b+a}\right\}$, $a\neq b$.

69. $\displaystyle\int \sin a\theta\cos b\theta\,d\theta = -\frac{1}{2}\left\{\frac{\cos(a+b)\theta}{a+b} + \frac{\cos(a-b)\theta}{a-b}\right\}$, $a\neq b$.

70. $\displaystyle\int \frac{d\theta}{a^2\cos^2\theta + b^2\sin^2\theta} = \int \frac{\sec^2\theta\,d\theta}{a^2+b^2\tan^2\theta} = \frac{1}{ab}\tan^{-1}\left(\frac{b}{a}\tan\theta\right)$.

71. $\displaystyle\int \frac{d\theta}{a^2\cos^2\theta - b^2\sin^2\theta} = \frac{1}{ab}\tanh^{-1}\left(\frac{b}{a}\tan\theta\right)$, $\tan\theta < \frac{a}{b}$:

$\displaystyle = \frac{1}{ab}\coth^{-1}\left(\frac{b}{a}\tan\theta\right)$, $\tan\theta > \frac{a}{b}$.

72. $\int \sin^{-1}x\,dx = x\sin^{-1}x + (1-x^2)^{1/2}$, (integration by parts).

73. $\int \cos^{-1}x\,dx = x\cos^{-1}x - (1-x^2)^{1/2}$.

74. $\int \tan^{-1}x\,dx = x\tan^{-1}x - \frac{1}{2}\ln(1+x^2)$.

75. $\int \cot^{-1}x\,dx = x\cot^{-1}x + \frac{1}{2}\ln(1+x^2)$.

76. $\int \operatorname{cosec}^{-1}x\,dx = x\operatorname{cosec}^{-1}x + \ln\{x + (x^2-1)^{1/2}\}$

$= x\operatorname{cosec}^{-1}x + \cosh^{-1}x$.

77. $\int \sec^{-1}x\,dx = x\sec^{-1}x - \ln\{x + (x^2-1)^{1/2}\}$

$= x\sec^{-1}x - \cosh^{-1}x$.

For (I.78)–(I.83) see beginning of this section.

78. $\int \frac{d\theta}{\sin\theta} = \ln(\tan\tfrac{1}{2}\theta) = -\ln(\operatorname{cosec}\theta + \cot\theta)$, (SF 11).

79. $\int \frac{d\theta}{\cos\theta} = \ln\{\tan(\tfrac{1}{2}\theta + \tfrac{1}{4}\pi)\} = \ln(\sec\theta + \tan\theta)$, (SF 10).

80. $\int \frac{d\theta}{a\sin\theta + b\cos\theta} = \int \frac{d\theta}{(a^2+b^2)^{1/2}\sin(\theta+\alpha)}$

(where $a = (a^2+b^2)^{1/2}\cos\alpha$, $b = (a^2+b^2)^{1/2}\sin\alpha$)

$$= \frac{1}{(a^2+b^2)^{1/2}} \ln\tan(\tfrac{1}{2}\theta + \tfrac{1}{2}\alpha).$$

81. $\int \frac{d\theta}{a\sin\theta + b\cos\theta} = \frac{1}{(a^2+b^2)^{1/2}} \ln \frac{\{(a^2+b^2)^{1/2} - a\} + b\tan\frac{1}{2}\theta}{\{a + (a^2+b^2)^{1/2}\} - b\tan\frac{1}{2}\theta}.$

82. $\int \frac{d\theta}{a + b\sin\theta} = \frac{2}{(a^2-b^2)^{1/2}} \tan^{-1}\left\{\frac{(a^2-b^2)^{1/2}}{a+b}\tan(\tfrac{1}{2}\theta - \tfrac{1}{4}\pi)\right\}$ $a^2 > b^2$.

$$= \frac{2}{(b^2-a^2)^{1/2}} \tanh^{-1}\left\{\frac{(b^2-a^2)^{1/2}}{a+b}\tan(\tfrac{1}{2}\theta - \tfrac{1}{4}\pi)\right\}$$ $a^2 < b^2$.

or $\coth^{-1}$ if $\{\ \ldots\ \} > 1$.

83. $\int \frac{d\theta}{a + b\cos\theta}$: replace $\frac{1}{2}\theta - \frac{1}{4}\pi$ in (I.82) by $\frac{1}{2}\theta$.

12.05. Exponential Functions.

84. $\int a^x\,dx = a^x/\ln a$.

85. $\int xe^{mx}\,dx = (e^{mx}/m^2)(mx - 1)$.

86. $\int x^n e^{mx}\,dx = \frac{x^n e^{mx}}{m} - \frac{n}{m}\int x^{n-1}e^{mx}\,dx,$

$$= \frac{x^n e^{mx}}{m}\left\{1 - \frac{n}{mx} + \frac{n(n-1)}{m^2x^2} - \ldots \pm \frac{n!}{(mx)^n}\right\}.$$

87. $\int e^{ax}\sin bx\,dx = e^{ax}\frac{a\sin bx - b\cos bx}{a^2+b^2},$

$$= e^{ax}\frac{\sin(bx-\varepsilon)}{(a^2+b^2)^{1/2}}, \quad \left(\tan\varepsilon = \frac{b}{a}\right).$$

88. $\int e^{ax} \cos bx\, dx = e^{ax} \dfrac{a \cos bx + b \sin bx}{a^2 + b^2}$

$$= \frac{1}{(a^2 + b^2)^{1/2}} e^{ax} \cos (bx - \varepsilon).$$

89. $\int e^{ax} \sin^n bx\, dx = \dfrac{e^{ax} \sin^{n-1} bx}{a^2 + n^2 b^2} (a \sin bx - nb \cos bx)$

$$+ \frac{n(n-1)b^2}{a^2 + n^2 b^2} \int e^{ax} \sin^{n-2} bx\, dx.$$

90. $\int e^{ax} \cos^n bx\, dx = \dfrac{e^{ax} \cos^{n-1} bx}{a^2 + n^2 b^2} (a \cos bx + nb \sin bx)$

$$+ \frac{n(n-1)b^2}{a^2 + n^2 b^2} \int e^{ax} \cos^{n-2} bx\, dx.$$

91. $\int e^{ax} \sinh ax\, dx = \dfrac{1}{4a} (e^{2ax} - 2ax)$, $\int e^{ax} \cosh ax\, dx = \dfrac{1}{4a} (e^{2ax} + 2ax)$.

92. $\int e^{ax} \sinh bx\, dx = \dfrac{e^{ax}}{a^2 - b^2} (a \sinh bx - b \cosh bx)$, $|a| \neq |b|$.

(*N.B* It is better to express sinh bx in terms of exponentials.)

12.06. LOGARITHMIC FUNCTIONS.

93. $\int \ln (ax)\, dx = x \ln (ax) - x$.

94. $\int (\ln ax)^n\, dx = x(\ln ax)^n - n \int (\ln ax)^{n-1}\, dx$.

95. $\displaystyle\int \frac{\ln (ax)}{x}\, dx = \tfrac{1}{2}(\ln ax)^2$.

96. $\int x^n \ln ax\, dx = \dfrac{x^{n+1}}{n+1} (\ln ax) - \dfrac{x^{n+1}}{(n+1)^2}$, $(n \neq -1)$.

97. $\displaystyle\int \frac{dx}{x \ln x} = \ln (\ln x)$.

12.07. HYPERBOLIC FUNCTIONS. These are integrated in the same way as trigonometric functions. A change into exponentials is often convenient, and answers are often more easily used in that form.

Difficult rational functions of sinh θ and cosh θ are reduced to rational functions of $t = \tanh \frac{1}{2}\theta$, and then integrated. The quantities needed are

$$\sinh \theta = \frac{2t}{1 - t^2}, \quad \cosh \theta = \frac{1 + t^2}{1 - t^2}, \quad d\theta = \frac{2}{1 - t^2}\, dt.$$

There are first the standard formulae SF 12–19.

98. $\int \operatorname{sech}\theta\, d\theta = \int \frac{2}{e^{\theta}+e^{-\theta}}\, d\theta = \int \frac{2d(e^{\theta})}{e^{2\theta}+1} = 2\tan^{-1}(e^{\theta})$, (SF 18).

$$\int \operatorname{sech}\theta\, d\theta = \int \frac{1-t^2}{1+t^2}\cdot\frac{2dt}{1-t^2} = 2\tan^{-1}(t) = 2\tan^{-1}(\tanh\tfrac{1}{2}\theta).$$

By putting $\theta = 0$ we find that these integrals differ by $\frac{1}{2}\pi$.

99. $\int \sinh^2\theta\, d\theta = \int \tfrac{1}{2}(\cosh 2\theta - 1)d\theta = \tfrac{1}{2}(\sinh\theta\cosh\theta - \theta)$.

100. $\int \cosh^2\theta\, d\theta = \tfrac{1}{2}(\sinh\theta\cosh\theta + \theta)$.

101. $\int \tanh^2\theta\, d\theta = \theta - \tanh\theta$.

102. $\int \coth^2\theta\, d\theta = \theta - \coth\theta$.

103. $\int \operatorname{sech}^2\theta\, d\theta = \tanh\theta$.

104. $\int \operatorname{cosech}^2\theta\, d\theta = -\coth\theta$.

105. $\int \sinh^n\theta\cosh\theta\, d\theta = \frac{\sinh^{n+1}\theta}{n+1}$.

106. $\int \cosh^n\theta\sinh\theta\, d\theta = \frac{\cosh^{n+1}\theta}{n+1}$.

107. $\int \sinh a\theta\sinh b\theta\, d\theta = \frac{\sinh(a+b)\theta}{2(a+b)} - \frac{\sinh(a-b)\theta}{2(a-b)}$, $(a \neq b)$.

108. $\int \sinh a\theta\cosh b\theta\, d\theta = \frac{\cosh(a+b)\theta}{2(a+b)} + \frac{\cosh(a-b)\theta}{2(a-b)}$.

109. $\int \cosh a\theta\cosh b\theta\, d\theta = \frac{\sinh(a+b)\theta}{2(a+b)} + \frac{\sinh(a-b)\theta}{2(a-b)}$.

110. $\int \sinh^{-1}x\, dx = \int \ln\{x + (1+x^2)^{1/2}\}dx,$
$= x\sinh^{-1}x - (1+x^2)^{1/2}.$

111. $\int \cosh^{-1}x\, dx = \int \ln\{x + (x^2-1)^{1/2}\}dx,$
$= x\cosh^{-1}x - (x^2-1)^{1/2}, \quad (|x| > 1).$

112. $\int \tanh^{-1}x\, dx = \int \ln\sqrt{\{(1+x)/(1-x)\}}dx,$
$= x\tanh^{-1}x + \tfrac{1}{2}\ln(1-x^2), \quad (|x| < 1).$

113. $\int \coth^{-1}x\, dx = \int \ln\sqrt{\{(x+1)/(x-1)\}}dx,$
$= x\coth^{-1}x + \tfrac{1}{2}\ln(x^2-1), \quad (|x| > 1).$

114. $\displaystyle\int \frac{d\theta}{a + b\sinh\theta} = \int \frac{2du}{bu^2 + 2au - b}$, when $u = e^{\theta}$.

12.08. INTEGRALS INVOLVING LINEAR SURDS. $(ax + b)^{1/2}$.

115. $\displaystyle\int \frac{dx}{(ax + b)^{1/2}} = \frac{2}{a}(ax + b)^{1/2}$.

116. $\displaystyle\int \frac{x\,dx}{(ax + b)^{n+\frac{1}{2}}} = \frac{1}{a}\int \frac{dx}{(ax + b)^{n-\frac{1}{2}}} - \frac{b}{a}\int \frac{dx}{(ax + b)^{n+\frac{1}{2}}}$.

117. $\displaystyle\int \frac{dx}{x(ax + b)^{n+\frac{1}{2}}} = \frac{1}{(n - \frac{1}{2})b(ax + b)^{n-\frac{1}{2}}} + \frac{1}{b}\int \frac{dx}{x(ax + b)^{n-\frac{1}{2}}}$.

118. $\displaystyle\int \frac{dx}{x(ax + b)^{1/2}} = -\frac{2}{b^{1/2}}\tanh^{-1}\left(\frac{ax + b}{b}\right)^{1/2}$, $b > ax + b > 0$,

and $\displaystyle -\frac{2}{b^{1/2}}\coth^{-1}\left(\frac{ax + b}{b}\right)^{1/2}$, $ax + b > b > 0$.

119. $\displaystyle\int \frac{dx}{x^p(ax + b)^{1/2}} = -\frac{(ax + b)^{1/2}}{(p - 1)bx^{p-1}} - \frac{2p - 3}{2p - 2}\cdot\frac{a}{b}\int \frac{dx}{x^{p-1}(ax + b)^{1/2}}$.

The general method of evaluating integrals of rational functions of x and $(ax + b)^{1/2}$ is to use the substitution

$$t = (ax + b)^{1/2}, \quad x = \frac{1}{a}(t^2 - b), \quad a\,dx = 2t\,dt.$$

12.09. QUADRATIC SURDS (*i*), involving $\sqrt{(Ax^2 + B)}$.

$$\int \frac{x}{(a^2 + x^2)^{1/2}}\,dx = (a^2 + x^2)^{1/2}, \quad \int \frac{x\,dx}{(Ax^2 + B)^{1/2}} = \frac{1}{A}(Ax^2 + B)^{1/2}.$$

The standard formulae SF 25–27 and most of the examples below are obtained using substitutions of the following form:

$\sqrt{(a^2 - x^2)}$, $|x| \leqslant a$: put $x = a\sin\theta$, $(a^2 - x^2)^{1/2} = a\cos\theta$, $dx = a\cos\theta\,d\theta$,

or $x = a\tanh\theta$, or $x = a\,\mathrm{sech}\,\theta$.

$\sqrt{(a^2 + x^2)}$, all x: put $x = a \sinh \theta$, $(a^2 + x^2)^{1/2} = a \cosh \theta$, $dx = a \cosh \theta \, d\theta$,
or $x = a \tan \theta$, or $x = a \operatorname{cosech} \theta$.

$\sqrt{(x^2 - a^2)}$, $|x| > a$: put $x = a \cosh \theta$, $(x^2 - a^2)^{1/2} = a \sinh \theta$, $dx = a \sinh \theta \, d\theta$,
or $x = a \sec \theta$, or $x = a \coth \theta$.

The first substitution in each case is the most useful, but the others are needed occasionally.

120. $\int (a^2 - x^2)^{1/2} dx = \frac{1}{2}a^2 \sin^{-1}(x/a) + \frac{1}{2}x(a^2 - x^2)^{1/2}$.

121. $\int (a^2 + x^2)^{1/2} dx = \frac{1}{2}a^2 \sinh^{-1}(x/a) + \frac{1}{2}x(x^2 + a^2)^{1/2}$,
or $\frac{1}{2}a^2 \ln \{x + (a^2 + x^2)^{1/2}\} + \frac{1}{2}x(a^2 + x^2)^{1/2}$.

122. $\int (x^2 - a^2)^{1/2} dx = \frac{1}{2}x(x^2 - a^2)^{1/2} - \frac{1}{2}a^2 \cosh^{-1}(x/a)$,
or $\frac{1}{2}x(x^2 - a^2)^{1/2} - \frac{1}{2}a^2 \ln \{x + (x^2 - a^2)^{1/2}\}$.

123. $\displaystyle\int \frac{dx}{x(a^2 - x^2)^{1/2}} = -\frac{1}{a} \operatorname{sech}^{-1}\left(\frac{x}{a}\right) = -\frac{1}{a} \cosh^{-1}\frac{a}{x}$,
$(x = a \operatorname{sech} \theta)$.

124. $\displaystyle\int \frac{dx}{x(a^2 + x^2)^{1/2}} = -\frac{1}{a} \operatorname{cosech}^{-1}\left(\frac{x}{a}\right) = -\frac{1}{a} \sinh^{-1}\left(\frac{a}{x}\right)$.

125. $\displaystyle\int \frac{dx}{x(x^2 - a^2)^{1/2}} = \frac{1}{a} \sec^{-1}\left(\frac{x}{a}\right) = \frac{1}{a} \cos^{-1}\left(\frac{a}{x}\right)$.

Functions involving x and one of these surds can be integrated using the same substitutions.

126. $\displaystyle\int \frac{x \, dx}{(a^2 - x^2)^{3/2}} = \frac{1}{(a^2 - x^2)^{1/2}}$.

127. $\displaystyle\int \frac{dx}{(a^2 - x^2)^{3/2}} = \frac{x}{a^2(a^2 - x^2)^{1/2}}$.

128. $\displaystyle\int \frac{x \, dx}{(a^2 + x^2)^{3/2}} = -\frac{1}{(a^2 + x^2)^{1/2}}$.

129. $\displaystyle\int \frac{dx}{(x^2 + a^2)^{3/2}} = \frac{x}{a^2(a^2 + x^2)^{1/2}}$.

130. $\displaystyle\int \frac{x \, dx}{(x^2 - a^2)^{3/2}} = -\frac{1}{(x^2 - a^2)^{1/2}}$.

131. $$\int \frac{dx}{(x^2 - a^2)^{3/2}} = -\frac{x}{a^2(x^2 - a^2)^{1/2}}.$$

132. $$\int \frac{x^2}{(a^2 - x^2)^{1/2}}\,dx = \int\left\{\frac{a^2}{(a^2 - x^2)^{1/2}} - (a^2 - x^2)^{1/2}\right\}dx,$$
$$= \frac{1}{2}a^2 \sin^{-1}\left(\frac{x}{a}\right) - \frac{1}{2}x(a^2 - x^2)^{1/2}.$$

133. $$\int \frac{1}{x + (a^2 + x^2)^{1/2}}\,dx = \int \frac{x - (a^2 + x^2)^{1/2}}{[x + (a^2 + x^2)^{1/2}][x - (a^2 + x^2)^{1/2}]}\,dx,$$
$$= \int \frac{x - (a^2 + x^2)^{1/2}}{x^2 - (a^2 + x^2)}\,dx,$$
$$= 1/a^2 \int\{(a^2 + x^2)^{1/2} - x\}dx,$$
$$= \frac{1}{2}\ln\frac{x + (x^2 + a^2)^{1/2}}{a}$$
$$+ \frac{1}{2a^2}x(a^2 + x^2)^{1/2} - \frac{x^2}{2a^2}.$$

The last example illustrates the fact that surds can always be removed from the denominator.

12.10. QUADRATIC SURDS (*ii*), involving $\sqrt{(Ax^2 + Bx + c)}$. Simple integrals involving $\sqrt{(Ax^2 + Bx + c)}$ can be reduced to standard forms by the substitution $u = x + (B/2A)$ (or a multiple of this to avoid fractions). This expresses the quadratic as a sum of squares of a type discussed in the previous section.

134. $$\int \frac{dx}{(9 - 8x - x^2)^{1/2}} = \int \frac{d(x + 4)}{\{25 - (x + 4)^2\}^{1/2}} = \int \frac{du}{(25 - u^2)^{1/2}},$$
$$\left(u = x + 4 = \sin^{-1}\frac{u}{5} = \sin^{-1}\frac{x + 4}{5}\right).$$

135. $$\int \frac{dx}{(7 - 4x - 3x^2)^{1/2}} = \int \frac{\sqrt{3}\,dx}{(21 - 12x - 9x^2)^{1/2}},$$
$$= \int \frac{\sqrt{3} \cdot \frac{1}{3}d(3x + 2)}{\{25 - (3x + 2)^2\}^{1/2}},$$
$$= \frac{1}{\sqrt{3}}\sin^{-1}\frac{3x + 2}{5}.$$

136. $$\int \frac{dx}{(x^2 - 6x + 10)^{1/2}} = \int \frac{d(x - 3)}{\{(x - 3)^2 + 1\}^{1/2}} = \sinh^{-1}(x - 3).$$

General formulae are:

$$137.\quad \int \frac{dx}{(ax^2+bx+c)^{1/2}} = \frac{1}{\sqrt{a}} \ln\left\{\frac{b}{2a} + x + \left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)^{1/2}\right\}, \quad (a>0).$$

$$= \frac{1}{\sqrt{a}} \sinh^{-1} \frac{2ax+b}{(4ac-b^2)^{1/2}}, \quad \text{where } 4ac > b^2,$$

$$\text{and} = \frac{1}{\sqrt{a}} \cosh^{-1} \frac{2ax+b}{(b^2-4ac)^{1/2}}, \quad \text{where } 4ac < b^2.$$

$$138.\quad \int \frac{dx}{(c+bx-ax^2)^{1/2}} = \frac{1}{\sqrt{a}} \sin^{-1} \frac{2ax-b}{(b^2-4ac)^{1/2}}, \quad (a>0,\ b^2>4ac).$$

$$139.\quad \int \frac{ax+\frac{1}{2}b}{(ax^2+bx+c)^{1/2}}\,dx = (ax^2+bx+c)^{1/2}.$$

$$\text{i.e.} \int \frac{Ax+B}{(ax^2+bx+c)^{1/2}}\,dx = \int \frac{(A/a)(ax+\frac{1}{2}b) + (B - Ab/2a)}{(ax^2+bx+c)^{1/2}}\,dx,$$

$$= \frac{A}{a}(ax^2+bx+c)^{1/2} + \left(B - \frac{Ab}{2a}\right)\int \frac{dx}{(ax^2+bx+c)^{1/2}}.$$

More complicated integrals can also be solved using this substitution to remove the term in x, and the earlier substitutions then remove the surd.

Example 1. $\displaystyle\int \frac{x+1}{(x^2+x+1)^{1/2}}\,dx.$

Since $x^2 + x + 1 = (x+\frac{1}{2})^2 + \frac{3}{4}$,

let $x + \frac{1}{2} = (\sqrt{3}/2)\sinh\theta$, (double substitution).

$$(x^2+x+1)^{1/2} = (\tfrac{3}{4}\sinh^2\theta + \tfrac{3}{4})^{1/2} = (\sqrt{3}/2)\cosh\theta,$$

$$dx = (\sqrt{3}/2)\cosh\theta\,d\theta.$$

$$I = \int \frac{\{(\sqrt{3}/2)\sinh\theta - \frac{1}{2}\} + 1}{(\sqrt{3}/2)\cosh\theta}\left(\frac{\sqrt{3}}{2}\cosh\theta\,d\theta\right),$$

$$= \int \left(\frac{\sqrt{3}}{2}\sinh\theta - \frac{1}{2}\right)d\theta,$$

$$= (\sqrt{3}/2)\cosh\theta - \tfrac{1}{2}\theta,$$

$$= (x^2+x+1)^{1/2} - \tfrac{1}{2}\sinh^{-1}(2x+1)/\sqrt{3}.$$

By means of algebraic tricks, such as those given, it is always possible to reduce integrals to simple forms, although the algebra may be heavy.

140. $\displaystyle\int \frac{dx}{\sqrt{\{x(ax+b)\}}} = \frac{1}{\sqrt{a}} \cosh^{-1} \frac{2ax+b}{b}, \quad a > 0.$

141. $\displaystyle\int \frac{dx}{\sqrt{\{x(b-ax)\}}} = \frac{1}{\sqrt{a}} \sin^{-1} \frac{2ax-b}{b}, \quad a > 0.$

142. $\displaystyle\int \frac{dx}{(x-\alpha)\sqrt{(ax^2+bx+c)}}$

$$= -\frac{1}{\lambda} \ln \left[\frac{\sqrt{(ax^2+bx+c)} + \lambda}{x-\alpha} + \frac{b+2a\alpha}{2\lambda} \right],$$

when $a\alpha^2 + b\alpha + c = \lambda^2 > 0$:

$$= \frac{1}{\lambda'} \sin^{-1} \left\{ \frac{(b+2a\alpha)(x-\alpha) - 2\lambda'^2}{(x-\alpha)(b^2-4ac)^{1/2}} \right\},$$

when $a\alpha^2 + b\alpha + c = -\lambda'^2 < 0$ and $b^2 > 4ac$.

12.11. DEFINITE INTEGRALS. See also: Reduction Formulae, Section 12.12.

143. $\displaystyle\int_0^\infty \frac{dx}{a^2+x^2} = \frac{\pi}{2a}.$

144. $\displaystyle\int_0^\infty \frac{dx}{(a^2+x^2)(b^2+x^2)} = \frac{\pi}{2ab(a+b)}.$

145. $\displaystyle\int_0^\infty x^n e^{-ax}\, dx = \frac{\Gamma(n+1)}{a^{n+1}} = \frac{n!}{a^{n+1}}, \quad (n \text{ integral}).$

146. $\displaystyle\int_0^\infty e^{-\alpha x^2}\, dx = \frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2}.$

147. $\displaystyle\int_0^\infty x^{2n} e^{-\alpha x^2}\, dx = \frac{1 \cdot 3 \cdot 5 \ldots (2n-1)}{2^{n+1}} \cdot \frac{\sqrt{\pi}}{\alpha^{n+\frac{1}{2}}}.$

148. $\displaystyle\int_0^\infty x e^{-\alpha x^2}\, dx = 1/(2\alpha).$

149. $\displaystyle\int_0^\infty x^{2n+1} e^{-\alpha x^2}\, dx = \frac{n!}{2\alpha^{n+1}}.$

150. $\displaystyle\int_0^\infty x^p e^{-x^2/a^2}\,dx = \tfrac{1}{2}a^{p+1}\Gamma\left(\frac{p+1}{2}\right), \quad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}.$

151. $\displaystyle\int_0^\infty \frac{\sin ax}{x}\,dx = \tfrac{1}{2}\pi, \quad (a > 0).$

152. $\displaystyle\int_0^\infty \frac{\sin^2 ax}{x^2}\,dx = \tfrac{1}{2}a\pi, \quad (a > 0).$

153. $\displaystyle\int_0^\infty \frac{\sin x}{x^{1/2}}\,dx = 2\int_0^\infty \sin(x^2)dx = \left(\frac{\pi}{2}\right)^{1/2}.$

154. $\displaystyle\frac{1}{2\pi}\int_0^{2\pi} \sin m\theta \sin n\theta\, d\theta = \tfrac{1}{2}\ (m = n),\ 0\ (m \neq n)$

155. $\displaystyle\frac{1}{2\pi}\int_0^{2\pi} \cos m\theta \cos n\theta\, d\theta = \tfrac{1}{2}\ (m = n),\ 0\ (m \neq n)$

m and n integers > 0.

156. $\displaystyle\frac{1}{2\pi}\int_0^{2\pi} \sin m\theta \cos n\theta\, d\theta = 0.$

157. $\displaystyle\int_0^{\frac{1}{2}\pi} \sin^n\theta\, d\theta = \int_0^{\frac{1}{2}\pi} \cos^n\theta\, d\theta = \frac{(n-1)(n-3)\ldots 1}{n(n-2)\ldots 2}\cdot\frac{\pi}{2},$ (n even),

$\displaystyle= \frac{(n-1)(n-3)\ldots 2}{n(n-2)\ldots 3}\cdot 1,$ (n odd).

158. $\displaystyle\int_0^{\pi} \sin^n\theta\, d\theta = 2\int_0^{\frac{1}{2}\pi} \sin^n\theta\, d\theta, \quad \int_0^{\pi} \cos^n\theta\, d\theta = 2\int_0^{\frac{1}{2}\pi} \cos^n\theta\, d\theta,$ (n even),

$= 0,$ (n odd).

12.12. Reduction Formulae.

159. $I_n = \int x^n e^{-\alpha x}\,dx$: $\displaystyle I_n = -\frac{x^n}{\alpha}e^{-\alpha x} + \frac{n}{\alpha}I_{n-1}$: $\displaystyle I_0 = -\frac{1}{\alpha}e^{-\alpha x}.$

160. $\displaystyle I_n = \int_0^\infty x^n e^{-\alpha x}\,dx$: $I_n = (n/\alpha)I_{n-1} = n!/\alpha^{n+1}$,

$I_0 = 1/\alpha.$

161. $I_n = \int \tan^n\theta\, d\theta$: $I_n + I_{n-2} = \{1/(n-1)\}\tan^{n-1}\theta$:

$I_1 = \ln(\sec\theta)$, $I_0 = \theta.$

162. $I_n = \int \cos^n \theta \, d\theta$: $nI_n - (n-1)I_{n-2} = \cos^{n-1}\theta \sin\theta$.

163. $I_n = \int_0^{\frac{1}{2}\pi} \cos^n \theta \, d\theta$: $I_n = \{(n-1)/n\}I_{n-2}$: $I_1 = 1$, $I_0 = \frac{1}{2}\pi$.

164. $I_n = \int \sin^n \theta \, d\theta$: $nI_n - (n-1)I_{n-2} = -\sin^{n-1}\theta \cos\theta$.

165. $I_n = \int_0^{\frac{1}{2}\pi} \sin^n \theta \, d\theta$: $I_n = \{(n-1)/n\}I_{n-2}$: $I_1 = 1$, $I_0 = \frac{1}{2}\pi$.

REFERENCES

The author has found useful information in the following books:

CLEMENTS, G. R. and WILSON, L. T. *Manual of Mathematics and Mechanics.* McGraw-Hill Book Co., New York, 1937.

DWIGHT, H. B. *Tables of Integrals and other Mathematical Data.* Macmillan and Co., New York, 1937.

GROBNER, W. and HÖFREITER, N. *Integraltafel.* Springer Verlag, Vienna, 1950.

PEIRCE, B. O. *A Short Table of Integrals.* Ginn and Co., 2nd Edn, London, 1910.

CHAPTER 13

SERIES

Series—Series of positive terms—Series of positive and negative terms—Absolute convergence—Approximations and errors—Estimate of length of calculation—Partially known series—Series of functions—Power series. Radius of convergence—Taylor's theorem and Maclaurin's theorem—Examples of Maclaurin's series—Evaluation of power series. Errors—Linear and quadratic approximations—Approximations near $x = a$—Calculation of Maclaurin's series—Legendre polynomials—Divergent functions—Asymptotic expansions—Fourier series—Sum of a Fourier series—Errors of Fourier series—Alteration of range—Arbitrary functions and Fourier series—Numerical series obtained from Fourier series—Expansions in series of orthogonal functions—Fourier sine series—Fourier cosine series—Legendre polynomials—Proof of Fourier's theorem

13.01. SERIES. A series is a sum of numbers, or terms,

$$a_1 + a_2 + a_3 + \ldots a_m,$$

with a rule for calculating each term a_n $(n = 1, 2, \ldots m)$. Summing the series means finding its numerical value, or a short formula for that value. A parameter appearing in the terms also appears in the sum. Typical examples are:

Arithmetic series: $1 + 2 + 3 \ldots + m = \frac{1}{2}m(m + 1)$.

Geometric series: $1 + r + r^2 + r^3 + \ldots r^m = \dfrac{1 - r^{m+1}}{1 - r}$.

Binomial series: $(1 + x)^m = 1 + mx + \dfrac{m(m - 1)}{1.2}x^2 + \ldots x^m$.

Series with large numbers of terms appear in most molecular statistical problems, e.g.

$$1 + 2 + 3 + \ldots + \mathcal{N} = \tfrac{1}{2}\mathcal{N}(\mathcal{N} + 1),$$

$$1/1 + 1/2^2 + 1/3^2 + \ldots + 1/\mathcal{N}^2 = \pi^2/8 \text{ (approximately ?)},$$

where $\mathcal{N}$ is very large, as large as 10^9 or more. These two sums are treated in very different ways. When $\mathcal{N}$ is very large, the first sum is also large, and approximately $\frac{1}{2}\mathcal{N}^2$, with a relative error of $1/\mathcal{N}$, which is very small, whereas the later terms in the second would not appear to affect the answer very much, so that it makes little absolute error to replace $\mathcal{N}$ by any other very large number, or

even to imagine that it is infinite. Sums of the first type are distinguished by the order of largeness of the approximate values attributed to them, i.e. the sum given is of order N^2.

Finally, series are obtained which are infinite, e.g. in the molecular theory of solids, the sum,

$$e^{-E_1/kT} + 3e^{-E_2/kT} + 5e^{-E_3/kT} + \ldots,$$

appears in a calculation of the specific heat, where the energies E_n increase steadily with n and appear to be defined for all positive values $n = 1, 2, 3 \ldots$.

Such series are also obtained by a method of successive approximations which leads to successive corrections, e.g. starting with $s_1 = a_1$ a correction a_2 is calculated to give the next approximation $s_2 = a_1 + a_2$, then another correction a_3 to give $s_3 = a_1 + a_2 + a_3$, and so on. Although such a calculation is always stopped at a finite stage, in principle it is believed that it could be carried on indefinitely to give an infinite series $a_1 + a_2 + a_3 + \ldots$ which would be the correct answer.

Example 1. In decimal approximations, each decimal added is a correction;

$$2{\cdot}71828 = 2 + 0{\cdot}7 + 0{\cdot}01 + 0{\cdot}008 + 0{\cdot}0002 + \ldots$$

In this chapter we shall be concerned mainly with infinite series, although much of our explanation applies to finite series, especially that part connected with approximations. By analogy with the last example, given an infinite series

$$a_1 + a_2 + a_3 + \ldots$$

we construct the partial sums of the series,

$$s_n = a_1 + a_2 + a_3 + \ldots a_n \text{ for } n = +1, 2, \ldots$$

If the sequence s_n tends to a limit s, then the series is said to be convergent and to have the sum s. If the sequence s_n diverges, then the series is said to be divergent. For divergent series, we try to get the order of magnitude of the partial sums when n is very large.

Although 'divergent' series appear in theoretical work, usually only a partial sum is required and can be dealt with as just described. Much more frequently series are obtained which are believed to be convergent. When convergent, they can be used to obtain approximations: the fact that the sequence s_n has a limit s means that $(s - s_n)$ becomes small when n is large, i.e. that the partial sum is an approximation to the correct answer. In view of later work in this

chapter we now consider rather carefully what series are convergent, and the errors involved in using partial sums as approximations. To do this we distinguish different types, but there is one simple condition affecting them all, namely that a series cannot be convergent unless $a_n \to 0$ as $n \to \infty$.

13.02. SERIES OF POSITIVE TERMS. When all the terms of a series Σa_n (an abbreviation for $a_1 + a_2 + \ldots$) are positive the partial sums s_n increase steadily with n. To prove convergence (Section 3.06) it is only necessary to find an upper bound. There are three simple tests for convergence in terms of the numbers a_n which the reader can easily verify. They are:

(*i*) *Comparison Test.* Every term a_n of Σa_n is less than the corresponding term b_n of another sequence Σb_n which is known to be convergent.

(*ii*) *Ratio Test.* The geometric series $1 + r + r^2 + r^3 + \ldots$ is convergent when $r < 1$, so that, by comparison, any series is convergent when $(a_{n+1}/a_n) < r < 1$.

(*iii*) *Limit Ratio Test.* This takes advantage of the fact that convergence depends on the later terms of a series. It asserts that a series is convergent if $\lim_{n\to\infty} (a_{n+1}/a_n) = \alpha < 1$.

The ratio tests do not express any opinion on series where a_{n+1}/a_n tends to 1. For these more sensitive tests are required.

The alteration or rearrangement of a finite number of terms does not affect the value of the later partial sums by more than a finite amount, and therefore does not affect the convergence. Hence the tests given above can be applied, even if they are not satisfied for the earlier terms, and even when a *finite* number of terms are negative.

It can be proved that an infinite rearrangement of a series of positive terms does not affect the sum. This is very important because it means that such series can be *added*, subtracted and *multiplied* by rearranging the terms in any way, without affecting the answer. The rearrangement required is usually that of moving all the large terms to the beginning. For example, a product of two series might be calculated as follows

$$(a_1 + a_2 + a_3 \ldots)(b_1 + b_2 + b_3 \ldots) = a_1b_1 + (a_1b_2 + a_2b_1) + (a_3b_1 + a_2b_2 + a_1b_3) + \ldots$$

e.g. $$(1 + r + r^2 + r^3 \ldots)(1 + r + r^2 + r^3 \ldots) = 1 + 2r + 3r^2 + 4r^3 + \ldots$$

13.03. SERIES OF POSITIVE AND NEGATIVE TERMS. It is necessary to distinguish series with a finite number of negative terms from those with an infinite number of negative terms.

As explained in the last section, the convergence of a series with a finite number of negative terms is not essentially different from that of the series obtained by making all the terms positive. The change merely alters the sum by a finite amount. This is not true when the series has an infinity of both positive and negative terms.

Amongst such series there is an important class which is always convergent. These series have successive terms with alternate signs and decreasing in magnitude. Examples are given below.

Other series, with runs of positive and negative terms, are often very difficult to handle and it is not easy to prove their convergence. However, when the positive terms by themselves form a sequence $\Sigma\alpha_n$ which is convergent, and the negative terms by themselves form a sequence $-\Sigma\beta_n$ which is convergent, then the actual series is merely a rearranged form of $\Sigma\alpha_n - \Sigma\beta_n$ and, according to the previous section dealing with the combination of series of positive terms, this rearranged series is not only convergent, but its sum is the difference of the sums of the two series.

To test convergence it is not necessary to test the two series $\Sigma\alpha$ and $\Sigma\beta$ separately. It is simpler to start with the original series $a_1 + a_2 + a_3 \ldots$ and make all the signs positive. This gives the new series $|a_1| + |a_2| + |a_3| + \ldots$, which is simply a rearrangement of $\Sigma\alpha + \Sigma\beta$. If this series is convergent then $\Sigma\alpha$ and $\Sigma\beta$ must be separately convergent, as they are all series of positive terms.

Since series of this type can always be expressed as the difference of two convergent series of positive terms, they can be added and subtracted and multiplied with rearrangement of the terms without altering the answer.

13.04. ABSOLUTE CONVERGENCE. A series $a_1 + a_2 + a_3 + \ldots$ is said to be absolutely convergent when the series of positive terms $|a_1| + |a_2| + |a_3| + \ldots$ is convergent. Thus a series which is absolutely convergent is either a convergent series of positive terms, or a series of positive and negative terms of the type considered at the end of the last section. To emphasize the importance of this concept it is worth while considering those series of positive and negative terms which are not absolutely convergent.

Any rearrangement of such a series is liable to give a series with a different sum. In fact, it can be proved that it is always possible to find an infinite rearrangement such that the sum of the rearranged series has any value previously assigned.

This is not odd. The really odd thing is that a series of positive terms or an absolutely convergent series can be rearranged without radically changing its sum, and to show that there are difficulties we consider quite a simple example.

Example 1. The series $1 + 1/2 + 1/3 + 1/4 + \ldots + 1/n \ldots$ is not convergent. To prove this it is only necessary to note that the partial sums increase steadily with the number of terms, and that the partial sum of 2^n terms is

$$\begin{aligned} s_{2^n} &= 1 + 1/2 + 1/3 + 1/4 + 1/5 + \ldots + 1/2^n, \\ &> 1 + 1/2 + (1/4 + 1/4) + (1/8 + 1/8 + 1/8 + 1/8) + \\ &\qquad \ldots + (\ldots + 1/2^n), \\ &> 1 + 1/2 + 1/2 + 1/2 + \ldots + 1/2, \end{aligned}$$

i.e. $s_{2^n} > 1 + \frac{1}{2}n$.

Hence the partial sums can be made as large as we please by choosing n sufficiently large.

It follows that the series

$$1 - 1/2 + 1/3 - 1/4 + 1/5 - 1/6 + \ldots = \lambda$$

is not absolutely convergent. It is *convergent*, however, because it consists of terms of alternate signs decreasing in magnitude.

Now consider the following series obtained by rearranging the terms of the last

$$1 + 1/3 - 1/2 + 1/5 + 1/7 - 1/4 + 1/9 + 1/11 - 1/6 + \ldots = \tfrac{3}{2}\lambda.$$

As indicated, the sum of this rearranged series is 50 per cent larger than that from which it was obtained. There is no trick in this. To obtain this result, after the subtraction of a term with an even denominator, enough positive terms are added, taken in order, to make the partial sums just greater than $3\lambda/2$; then another negative term is inserted. Since the series of positive terms by itself diverges this can always be done.

13.05. Approximations and Errors. The partial sums s_n of a series,

$$s_n = a_1 + a_2 + a_3 + \ldots + a_n,$$

are approximations to the sum of the complete series, whether finite or infinite. If the calculation of the sum only proceeds as far as n terms, the error involved is

$$E_n = s - s_n = a_{n+1} + a_{n+2} + a_{n+3} + \ldots,$$

and the problem is to estimate this error, remembering that the sum s is not known.

When a formula for all the terms a_n is given, then a convergence test often provides an upper limit for E_n, as in the following example.

Example 1. A series for the number e is

$$1 + 1 + 1/2 + 1/3! + 1/4! + \ldots,$$

where $a_{n+1} = (1/n!)$. A calculation of e using only seven terms of this series means an error

$$E = 1/7! + 1/8! + 1/9! + \ldots = (1/7!)\,(1 + 1/8 + 1/8.9 + \ldots),$$
$$< (1/7!)\,(1 + 1/8 + 1/8^2 + \ldots), \quad \text{(ratio test)}$$

i.e. $$E < 8/(7 \,.\, 7!) = 0{\cdot}0003 \text{ approximately.}$$

By the comparison test, the fact that E has an upper bound shows that the series is convergent.

When the numbers a_n are found by a method of successive approximations, the calculation of E_n is a little more difficult. The terms a_n after those calculated are not known precisely. Only the method of calculating them is known.

Example 2. The solution of the equation $x = \cos x$ by the recurrence relation $s_{n+1} = \cos s_n$ gives successive corrections

$$a_1 + a_2 + a_3 + \ldots,$$

i.e. $$a_{n+1} = s_{n+1} - s_n = \cos s_n - \cos s_{n-1},$$
$$= -\sin s'(s_n - s_{n-1}) = -a_n \sin s',$$

where s' lies between s_n and s_{n-1}, by the Mean Value Theorem (Section 5.15). Thus successive terms of the series alternate in sign and decrease in magnitude. If we explicitly introduce the signs of the terms, the error after m terms is

$$E_n = a_{n+1} - a_{n+2} + a_{n+3} - \ldots$$
$$= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \ldots$$
$$= a_{n+1} - a_{n+2} + (a_{n+3} - a_{n+4}) + \ldots$$

which is certainly less than a_{n+1} (and therefore a_n) and greater than $a_{n+1} - a_{n+2}$, because the terms decrease in size. As these limits tend to zero as $n \to 0$ this proves that the series converges, as well as evaluating the error.

13.06. ESTIMATE OF LENGTH OF CALCULATION. When the error E_n is known even approximately as a function of n, it is possible before

commencing the calculation to estimate how many terms are needed to give the required accuracy, and therefore the amount of work involved. This is an important factor in deciding between two possible methods of solving a problem, and may provide an incentive for finding a quicker method.

Example 1. Using the series for e in *Example 1* of the previous section let us suppose that e is required to 6 decimal places. Then, as before,

$$E_n = \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \frac{1}{(n+3)!} + \ldots$$

$$< \frac{1}{(n+1)!}\left\{1 + \frac{1}{n+2} + \frac{1}{(n+2)^2} + \ldots\right\},$$

i.e. $$E_n < \frac{1}{(n+1)!}\,\frac{1}{1 - 1/(n+2)} \quad \text{or} \quad \frac{n+2}{(n+1)(n+1)!}.$$

For this accuracy, it is usually sufficient that E_n and therefore the upper bound should be less than 5×10^{-7}. Hence

$$(n+2) < 5 \times 10^{-7}(n+1)\,.\,(n+1)!$$

and after a little arithmetic we find that $n \geqslant 9$, i.e. nine terms are sufficient.

13.07. Partially Known Series. After examining these exact methods for calculating errors, it is advisable to review two techniques which are often employed, although there is little to be said for them unless they are backed up by 'theoretical intuition', or additional information.

When a few terms $a_1 + a_2 + \ldots + a_n$ of a series are known or calculated, it is still possible to use the methods just described if the remaining terms can be estimated, using some care when the signs of the terms change. If nothing useful can be discovered about the terms we proceed by experience and faith.

For example, if the terms already known are decreasing satisfactorily, it is assumed that this decrease continues at the same rate, i.e. that the remaining terms decrease no less rapidly than a geometric series. Thus if $a_1 = 6$, $a_2 = 11$, $a_3 = 3{\cdot}2$, $a_4 = 1{\cdot}25$, $a_5 = 0{\cdot}23$, then it might be guessed that the remaining terms decrease by a factor $\frac{1}{2}$ from one to the next. This guess leads to an estimated error of

$$a_6 + a_7 + \ldots = a_5(\tfrac{1}{2} + \tfrac{1}{4} + \ldots) = \tfrac{1}{2}a_5 = 0{\cdot}1.$$

Another 'practical' definition of convergence is that a series is

convergent if the Nth term is less than the desired error. The number N depends on the person using the series, how much time he has, and whether any other methods are available for doing the calculation. For an accuracy of 1 : 1000, N usually is of the order 4 or 5, and rarely exceeds 10. In effect, the terms are going down rapidly and the test is similar to that given in the previous paragraphs.

It is often difficult to criticize these methods, as they may be based on other knowledge or experience, but they should be avoided when possible, and any additional knowledge should be used quantitatively to allow an estimate to be made which is reliable. They are unsound mathematically and the results should not be relied upon.

13.08. SERIES OF FUNCTIONS. When the terms of a series depend on a variable x the result is a series of functions $\Sigma a_n(x)$, which may converge for certain values of x.

There are two important groups of series of this type. The first is the Power Series of the form

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots$$

or $$a_0 + a_1(x - a) + a_2(x - a)^2 + a_3(x - a)^3 + \ldots$$

where the approximations obtained by using partial sums improve near $x = 0$ or $x = a$ respectively, because of the properties of powers. This is called a point expansion. The second is an expansion such as a Fourier series

$$a_1 \sin x + a_2 \sin 2x + a_3 \sin 3x + \ldots$$

where the error involved in using a partial sum is an average error.

One general difficulty with such series is that their sum $S(x)$ is not necessarily a continuous function of x, even when each term is continuous and the series is convergent. (This applies only to the sum of the series and not to the partial sums $S_n(x)$ of n terms, which are necessarily continuous when each term is continuous.) We do not wish to discuss this problem, except to mention that criteria can be found so that the sum $S(x)$ itself is continuous, and that this is satisfied by power series, provided that x lies within the range of values for which the power series is convergent. This means that power series have sums which are continuous and differentiable any number of times, term by term, as explained below.

13.09. POWER SERIES. RADIUS OF CONVERGENCE. Only power series in x are discussed below. A power series in $(x - a)$ or any other quantity is treated in exactly the same way.

A power series of integral powers of x is

$$S(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots$$

(fractional or negative powers can also be used). This series is absolutely convergent, by the Limit Ratio Test, when

$$\lim_{n\to\infty}\left|\frac{a_{n+1}x^{n+1}}{a_nx^n}\right| = |x| \lim_{n\to\infty}\frac{a_{n+1}}{a_n} = \frac{|x|}{R} < 1.$$

The number $R = \lim_{n\to\infty}(a_n/a_{n+1})$ is called the radius of convergence.

The series is absolutely convergent when $|x| < R$, and divergent when $|x| > R$. This test is not sensitive enough to consider the convergence when $x = R$. We shall not discuss the case $x = R$ any more as power series are rarely used for values of x nearly equal to the radius of convergence.

For some series R is finite, for others infinite, which means that the series converge for every value of x, and for other series R may be zero, so that the series is useless.

When $|x| < R$, the sum $S(x)$ of the series is a continuous function of x, as explained in the previous section, and has derivatives which can be obtained as series by differentiating term by term the series for $S(x)$.

$$dS/dx = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \ldots + na_nx^{n-1} + \ldots$$

$$d^2S/dx^2 = 2a_2 + 6a_3x + 12a_4x^2 + \ldots + n(n-1)a_nx^{n-2} + \ldots$$

$$d^3S/dx^3 = 6a_3 + 24a_4x + \ldots + n(n-1)(n-2)a_nx^{n-3} + \ldots$$

and so on. It can be shown easily that all these series are also absolutely convergent when $|x| < R$.

Similarly the indefinite integral is

$$\int S(x)dx = C + a_0x + \tfrac{1}{2}a_1x^2 + \tfrac{1}{3}a_2x^3 + \ldots + \frac{1}{n}a_nx^{n+1} + \ldots$$

These results lead to the following relations between the coefficients a_n of the series and the derivatives of the sum $S(x)$. For when $x = 0$, we find that

$$a_0 = S(0),\ a_1 = \left(\frac{dS}{dx}\right)_0,\ a_2 = \frac{1}{2!}\left(\frac{d^2S}{dx^2}\right)_0,\ a_n = \frac{1}{n!}\left(\frac{d^nS}{dx^n}\right)_0.$$

13.10. TAYLOR'S THEOREM AND MACLAURIN'S THEOREM. It has already been proved in Section 11.07 that

$$f(b) = f(a) + (b-a)\left(\frac{df}{dx}\right)_a + \frac{(b-a)^2}{2!}\left(\frac{d^2f}{dx^2}\right)_a + \cdots + \frac{(b-a)^n}{n!}\left(\frac{d^nf}{dx^n}\right)_a + E_n,$$

where $E_n = \int_a^b \frac{(b-t)^n}{n!}\left(\frac{d^{n+1}f}{dt^{n+1}}\right) dt$ (in the definite integral the variable of integration has been changed for convenience). This is Taylor's theorem.

By writing x instead of b, another form of the theorem is obtained

$$f(x) = f(a) + (x-a)\left(\frac{df}{dx}\right)_a + \frac{(x-a)^2}{2!}\left(\frac{d^2f}{dx^2}\right)_a + \cdots + \frac{(x-a)^n}{n!}\left(\frac{d^nf}{dx^n}\right)_a + E_n(x).$$

A third form is obtained by writing $x - a = h$.

$$f(a+h) = f(a) + h\left(\frac{df}{dx}\right)_a + \frac{h^2}{2!}\left(\frac{d^2f}{dx^2}\right)_a + \cdots + \frac{h^n}{n!}\left(\frac{d^nf}{dx^n}\right)_a + E_n(h).$$

Finally, when $a = 0$, Maclaurin's theorem is obtained

$$f(x) = f(0) + x\left(\frac{df}{dx}\right)_0 + \frac{x^2}{2!}\left(\frac{d^2f}{dx^2}\right)_0 + \cdots + \frac{x^n}{n!}\left(\frac{d^nf}{dx^n}\right)_0 + E_n(x).$$

These results are of considerable importance, for they assert that any function that can be differentiated $(n+1)$ times can be expressed as a power series (in $b-a$, $x-a$ or h), up to terms of the nth degree, with an error E_n, and an explicit expression is given for this error. This error cannot be calculated exactly, but it can often be estimated when the nth derivative is known approximately, using the Mean Value Theorem for integrals. For an approximation with a given accuracy, the procedure is obviously to choose the least value of n which makes E_n as small as required. Of the various possibilities in the behaviour of E_n as n varies, only two are mentioned.

For some functions E_n first decreases as n increases, and then starts to increase and goes on increasing indefinitely. In this case the best approximation is obtained with the value of n that makes E_n least, and it must be hoped that this is within the bounds allowed.

On the other hand, for some values of x (or b), E_n may tend to zero as $n \to \infty$. When this happens, all the right hand sides

become infinite power series, the first in $(b - a)$, the second in $(x - a)$, the third in h and the last in x. The function has then been expressed as a power series, within the radius of convergence of the series, e.g. Maclaurin's series

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$$

where $$a_n = \frac{1}{n!}\left(\frac{d^n f}{dx^n}\right)_0, \quad |x| < R, \quad E_n \to 0.$$

This can only be done when the function has an infinity of derivatives, all finite at $x = 0$.

The theorem is necessary for this reason. The series on the right has a sum $S(x)$, when $|x| < R$, and its derivatives at $x = 0$ are (previous section)

$$\left(\frac{d^n S}{dx^n}\right)_0 = n!\, a_n = \left(\frac{d^n f}{dx^n}\right)_0.$$

It does not, however, follow that $S(x)$ is identical with $f(x)$. This is the problem mentioned earlier, that successive approximations may converge to an answer but it may be the wrong answer. Maclaurin's theorem (or Taylor's theorem for the other series) gives an assurance that there is no ambiguity when $E_n \to 0$; two functions with the same derivatives at $x = 0$ are identical.

To make the implications of this result quite plain, consider the Maclaurin's series for e^x, i.e. $1 + x + (x^2/2!) + (x^3/3!) + \ldots$ with an infinite radius of convergence R. This means that the value of e^x even for $x = 10^9$ is determined by its value and those of all its derivatives at the origin $x = 0$, a result that is not obvious.

13.11. Examples of Maclaurin's Series.

Binomial Series:

$$(1 + x)^m = 1 + mx + \frac{m(m-1)}{1.2} x^2 + \ldots \binom{n}{m} x^n + \ldots$$

where $$\binom{n}{m} = \frac{n(n-1)\ldots(n-m+1)}{1.2\ldots m}. \quad |x| < 1, \text{ all } n.$$

This series terminates at x^n when n is a positive integer, and becomes the Binomial Theorem (valid for all x). Amongst the common special series derived from it are:

$$\frac{1}{(1-x)^n} = 1 + nx + \frac{n(n+1)}{1.2} x^2 + \frac{n(n+1)(n+2)}{1.2.3} x^3 + \ldots \qquad |x| < 1.$$

$$\sqrt{(1+x)} = 1 + \tfrac{1}{2} x - \tfrac{1}{8} x^2 + \tfrac{1}{16} x^3 - \tfrac{5}{128} x^4 + \ldots \qquad |x| < 1.$$

Trigonometric Series:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots \qquad \text{any } |x|.$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots \qquad \text{any } x.$$

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \ldots \qquad |x| < \frac{\pi}{4}.$$

$$\tan^{-1} x = x - \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 - \ldots \text{(P.V.)}. \qquad |x| \leqslant 1.$$

Exponential Series:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots$$

$$e^{-x} = 1 - x + \frac{1}{2!}x^2 - \frac{x^3}{3!} + \ldots \qquad \text{all } x.$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots$$

Logarithmic Series:

$$\log_e (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \qquad |x| < 1.$$

$$\log_e \frac{1}{(1 - x)} = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots \qquad |x| < 1.$$

$$\tanh^{-1} x = \tfrac{1}{2} \log_e \frac{1 + x}{1 - x} = x + \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 + \ldots \quad |x| < 1.$$

13.12. EVALUATION OF POWER SERIES: ERRORS. Power series are used to evaluate functions.

Example 1. The series for e^x gives

$$e^{1/2} = 1 + \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{3!}\left(\frac{1}{2}\right)^3 + \frac{1}{4!}\left(\frac{1}{2}\right)^4 + \frac{1}{5!}\left(\frac{1}{2}\right)^5 + \ldots$$

$$= 1{\cdot}6487 \text{ to six terms.}$$

Example 2. Using $\ln(1+x) = x - \frac{1}{2}x^2 + \ldots$,

$$\ln(3/2) = \frac{1}{2} - \frac{1}{2.4} + \frac{1}{3.8} - \frac{1}{4.16} + \frac{1}{5.32} - \frac{1}{6.64} + \ldots$$
$$= 0{\cdot}40 \ (4\ ?) \text{ to six terms.}$$

It is better and quicker to use $\ln\{(1+x)/(1-x)\}$ with $x = 1/5$.

$$\ln\frac{1+x}{1-x} = 2\left(x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \ldots\right),$$
$$= 2\left(\frac{1}{5} + \frac{1}{3.5^3} + \frac{1}{5.5^5} + \ldots\right),$$
$$= 0{\cdot}40546 \text{ to three terms.}$$

Example 3. $\pi/4 = (\tan^{-1}x)_{x=1} = 1 - 1/3 + 1/5 - 1/7 + \ldots$

The convergence of a power series ensures that the partial sums are approximations to the sum of the series, the accuracy increasing with the number of terms evaluated.

The error in using the partial sum $s_n = a_0 + a_1x + \ldots + a_nx^n$, is

$$E_n(x) = a_{n+1}x^{n+1} + a_{n+2}x^{n+2} + \ldots$$

It is clear that:

(*i*) for given x, $E_n(x)$ finally decreases with increasing n;
(*ii*) for given n, $E_n(x)$ decreases as x decreases.

The series is said to converge rapidly when only a few terms are needed for a given accuracy (few here usually means less than 10). When many terms are needed it is said to converge slowly.

(*iii*) for given accuracy, more terms are needed for large values of x than for small.

A discussion of errors is exactly the same as in Section 13.05, except that x occurs in the expression for the error. The remarks in Section 13.07 also apply, but only with the same force when x is nearly equal to the radius of convergence R. The wise chemist does not use a power series in this region, but looks for an alternative expression.

At this point it is convenient to mention that we are discussing theoretical formulae. Fitting experimental data by means of a power series is a powerful tool, e.g. at low temperatures, vapour pressure might be expressed in the form

$$p = aT + bT^2 + cT^3 + \ldots$$

The numbers a, b, c are now chosen empirically to fit the data and their values often depend on the number of terms included. Our considerations do not apply to these *polynomials*, until the experimenter, as mathematician, begins to extrapolate his results, when he is probably quite aware of the deficiency of the method.

Example 4. How many terms of the series for $\sinh x$ are needed to give an accuracy of 10^{-4} when $x < 0{\cdot}1$? If n terms are used (Section 13.11) the error is

$$E_n = \frac{x^{2n+1}}{(2n+1)!} + \frac{x^{2n+3}}{(2n+3)!} + \ldots < \frac{x^{2n+1}}{(2n+1)!}(1 + x^2 + \ldots),$$

i.e.
$$E_n < \frac{x^{2n+1}}{(2n+1)!(1-x^2)} \quad (x < 1).$$

This is greatest when $x = 0{\cdot}1$: to obtain $E_n < 0{\cdot}5 \times 10^{-4}$ only requires $n = 2$, i.e. $x + (x^3/6)$ gives the desired accuracy.

Example 5. The approximation $\sinh x = x + (x^3/6) + (x^5/120)$ is accurate to 10^{-4} provided x is such that $E_3 < 0{\cdot}00005$. From the previous example, this requires

$$x^7 < \frac{1}{2 \times 10^5} 7!(1 - x^2), \quad \text{or} \quad x^7 < \frac{7!}{2 \times 10^5} \sim 4 \times 10^{-4}.$$

i.e. $x < 0{\cdot}3$.

Example 6. If $\sin x$ is only required to 1 per cent accuracy, the following approximations are adequate:

(*i*) $\sin x = x$, $\quad x < 0{\cdot}3$ radian (18°)

(*ii*) $\sin x = x - \frac{1}{6}x^3$, $\quad x < 1$ radian (60°)

(*iii*) $\sin x = x - \frac{1}{6}x^3 + \frac{1}{120}x^5$, $x < \frac{1}{2}\pi$ radians (90°).

Since the terms alternate in sign, the error in this series for $\sin x$ is always less than the last term omitted, provided x is not very large. Thus $E_3 < x^7/7!$ in (*iii*), is less than 0·05 when $x^7 < 7!/200$, i.e. 25·2.

13.13. Linear and Quadratic Approximations. If

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots$$

and not all of a_0, a_1, a_2 are zero, the simplest approximations are often useful, on the assumption that x is sufficiently small:

$$\text{linear: } f(x) = a_0 + a_1x;$$
$$\text{quadratic: } f(x) = a_0 + a_1x + a_2x^2.$$

Example 1. $\sin x = x$, $\ln(1 + x) = x$,

$\cos x = 1 - \frac{1}{2}x^2$, $\sinh x = x$,

$\tan x = x$, $\cosh x = 1 + \frac{1}{2}x^2$,

$(1 + x)^n = 1 + nx$, or $1 + nx + \frac{1}{2}n(n - 1)x^2$.

13.14. APPROXIMATIONS NEAR $x = a$. Taylor's series with $x = a + h$ gives a series of approximations to a function near $x = a$, when h is small.

$$\text{Linear: } f(x) = f(a + h) = f(a) + hf'(a);$$

$$\text{Quadratic: } f(x) = f(a + h) = f(a) + hf'(a) + \tfrac{1}{2}h^2f''(a),$$

and so on. The greater the number of terms included, the greater the value of h can be for a given accuracy.

These equations form the best starting point for a discussion of maxima and minima. At a stationary value $x = a$, $f'(a) = 0$ and the quadratic approximation is

$$f(x) = f(a + h) = f(a) + \tfrac{1}{2}h^2f''(a),$$

for very small h. This shows that there is a minimum when $f''(a) > 0$ and a maximum when $f''(a) < 0$.

If $f''(a) = 0$ the next approximation must be used. If $f^{(m)}(a)$ is the first non-vanishing derivative at the stationary point, then

$$f(a + h) = f(a) + \frac{h^m}{m!} f^{(m)}(a).$$

This gives a maximum or minimum only when m is even; but not when m is odd.

13.15. CALCULATION OF MACLAURIN'S SERIES.

$$f(x) = f(0) + x\left(\frac{df}{dx}\right)_0 + \frac{1}{2!}x^2\left(\frac{d^2f}{dx^2}\right)_0 + \ldots + \frac{x^n}{n!}\left(\frac{d^nf}{dx^n}\right)_0 + \ldots$$

$$= f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \ldots + \frac{1}{n!}f^{(n)}(0)x^n + \ldots$$

The calculation of many derivatives by direct differentiation is very tedious. There are ways of reducing the labour.

Recurrence relations between the derivatives, using Leibnitz' theorem if necessary

Example 1. To find $f(x) = \tan^{-1} x$ in powers of x up to terms in x^4, use

$$\frac{df}{dx} = f'(x) = \frac{1}{1+x^2}, \quad \text{or} \quad (1+x^2)f'(x) = 1.$$

whence $f(0) = 0, f'(0) = 1$.

Differentiating this equation, $(1+x^2)f''(x) + 2xf'(x) = 0$: i.e. when $x = 0, f''(0) = 0$.

Again, $(1+x^2)f'''(x) + 4xf''(x) + 2f'(x) = 0.$

Put $x = 0$: $f'''(0) + 2f'(0) = 0, \quad \therefore f'''(0) = -2.$

Again, $(1+x^2)f''''(x) + 6xf'''(x) + 6f''(x) = 0.$

Put $x = 0, f''''(x) + 0 + 0 = 0, f''''(0) = 0.$

$$\therefore \tan^{-1} x = 0 + 1 \,.\, x + 0 \,.\, x^2 + \frac{1}{3!}(-2)x^3 + 0 \,.\, x^4 + \ldots$$

$$= x - \tfrac{1}{3}x^3 + 0x^4, \text{ to terms in } x^4.$$

In general, however, it is better to use Maclaurin's theorem as a proof that a function can be expressed as a power series

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots$$

and to use the properties of $f(x)$ and the series to find the coefficients directly. The following examples offer some methods of doing this.

Odd or even functions

The power series of odd or even functions are also odd or even, respectively, and therefore contain only odd or even powers of x.

$$f_e(x) = f_e(-x) = a_0 + a_2 x^2 + a_4 x^4 + \ldots$$
$$f_o(x) = -f_o(-x) = a_1 x + a_3 x^3 + a_5 x^5 + \ldots$$

Use of derivatives

Example 2. Sin x is an odd function of x. Let

$$\sin x = a_1 x + a_3 x^3 + a_5 x^5 + \ldots$$

Differentiate: $\cos x = a_1 + 3a_3 x^2 + 5a_5 x^4 + \ldots$

Again: $-\sin x = 3.2a_3 x + 5.4a_5 x^3 + \ldots$

Since there is only one series for $\sin x$:

$$3.2a_3 = -a_1\text{: } 5.4a_5 = -a_3\text{: } 7.6a_7 = -a_5\text{: etc.}$$

Now a_1 can be found from the series for $\cos x$: when $x = 0$, $\cos(0) = 1$; therefore

$$a_1 = 1,\ a_3 = -1/3!,\ a_5 = 1/5!,\ a_7 = -1/7!, \ldots$$

We now have series for $\sin x$ and $\cos x$

$$\sin x = x - x^3/3! + x^5/5! - x^7/7! + \ldots$$
$$\cos x = 1 - x^2/2! + x^4/4! - x^6/6! + \ldots$$

The radius of convergence is $R = \lim_{n\to\infty} (a_{2n-1}/a_{2n+1})$

$$= \lim_{n\to\infty} (2n-1)(2n+1) = \infty.$$

N.B. In calculating R we use successive terms in the series.

Derivatives or integrals of known series (especially binomial)

Example 3. Since

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - \ldots \qquad (|x| < 1).$$

$$\frac{1}{(1+x)^2} = -\frac{d}{dx}\left(\frac{1}{1+x}\right) = 1 - 2x + 3x^2 - 4x^3 + \ldots \qquad (|x| < 1).$$

$$\ln(1+x) = \int_0^x \frac{dx}{1+x} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \qquad (|x| < 1).$$

$$\tan^{-1} x = \int_0^x \frac{dx}{1+x^2} = \int_0^x (1 - x^2 + x^4 - \ldots)dx,$$
$$= x - \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 - \tfrac{1}{7}x^7 + \ldots \qquad (|x| < 1).$$

$$\sin^{-1} x = \int_0^x (1 - x^2)^{-1/2}\, dx$$
$$= \int_0^x (1 + \tfrac{1}{2}x^2 + \tfrac{3}{8}x^4 + \tfrac{1\,5}{3\,2}x^6 + \ldots)dx,$$
$$= x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{15x^7}{224} + \ldots$$

Substitutions in known series (cf. Section 13.11 for series)

$$e^{x^2} = 1 + (x^2) + \frac{1}{2!}(x^2)^2 + \frac{1}{3!}(x^2)^3 + \ldots$$

$$\sin(\sqrt{x}) = x^{1/2} - \frac{1}{3!}x^{3/2} + \frac{1}{5!}x^{5/2} - \frac{1}{7!}x^{7/2} + \ldots$$

$$\ln(1 + x + x^2) = (x + x^2) - \tfrac{1}{2}(x + x^2)^2 + \tfrac{1}{3}(x + x^2)^3 - \ldots \qquad (|x + x^2| < 1),$$

$$= x + \tfrac{1}{2}x^2 - \tfrac{2}{3}x^3 + \ldots \qquad (|x| < 1).$$

N.B. Change of radius of convergence after rearrangement.

Algebraic operations on known series

$$\ln(1 + x + x^2) = \ln(1 - x^3) - \ln(1 - x)$$
$$= -(x^3 + \tfrac{1}{2}x^6 + \ldots) + (x + \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 + \ldots)$$
$$= x + \tfrac{1}{2}x^2 - \tfrac{2}{3}x^3 + \tfrac{1}{4}x^4 + \tfrac{1}{5}x^5 - \tfrac{1}{3}x^6 + \ldots$$

$$e^{-x}\sin x = (1 + x + \tfrac{1}{2}x^2 + \tfrac{1}{6}x^3 + \ldots)(x - \tfrac{1}{6}x^3 + \ldots)$$
$$= x + x^2 + \tfrac{1}{3}x^3 + 0x^4 + \ldots$$

Example 4. Let $\tan x = a_1x + a_3x^3 + \ldots$.

Since $\sin x = \tan x \,(\cos x)$, then

$$(x - \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 - \ldots)$$
$$= (a_1x + a_3x^3 + a_5x^5 + \ldots)(1 - \tfrac{1}{2}x^2 + \tfrac{1}{24}x^4 - \ldots).$$

Multiply out and compare coefficients. Then $a = 1$, $a_3 = \frac{1}{2}a_1 - \frac{1}{6} = \frac{1}{3}$, etc.

$$\tan x = x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \ldots$$

13.16. LEGENDRE POLYNOMIALS. The potential at P (*Figure 13.01*) due to unit electric charge at A is

$$V = \frac{1}{\rho} = \frac{1}{(a^2 - 2aR\cos\theta + R^2)^{1/2}} = \frac{1}{a(1 - 2x\cos\theta + x^2)^{1/2}}, \qquad (x = (R/a)).$$

When $x < 1$ this can be expanded in powers of x

$$aV = \{1 - (2x\cos\theta - x^2)\}^{-1/2},$$
$$= 1 + \tfrac{1}{2}(2x\cos\theta - x^2) + \tfrac{3}{8}(2x\cos\theta - x^2)^2 + \tfrac{5}{16}(2x\cos\theta - x^2)^3 + \ldots$$
$$= 1 + x\cos\theta + \tfrac{1}{2}x^2(3\cos^2\theta - 1) + \tfrac{1}{2}x^3(5\cos^3\theta - 3) + \ldots$$

The coefficient of x^n is called the Legendre polynomial $P_n(\cos\theta)$ in $\cos\theta$.

$$P_0(\cos\theta) = 1,$$
$$P_1(\cos\theta) = \cos\theta,$$
$$P_2(\cos\theta) = \tfrac{1}{2}(3\cos^2\theta - 1),$$
$$P_3(\cos\theta) = \tfrac{1}{2}(5\cos^3\theta - 3).$$

$$\therefore\ V = (a^2 - 2aR\cos\theta + R^2)^{-1/2} = \sum_{n=0}^{\infty} R^n P_n(\cos\theta)/a^{n+1},\quad (R < a).$$

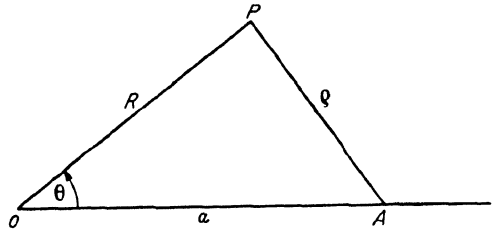

Figure 13.01

When $R > a$, we can expand in powers of (a/R). Interchanging R and a everywhere,

$$V = (a^2 - 2aR\cos\theta + R^2)^{-1/2} = \sum_{n=0}^{\infty} a^n P_n(\cos\theta)/R^{n+1},\quad (R > a).$$

13.17. Divergent Functions. Functions such as $\log x$ and e^{-x}/x have no Maclaurin series, since they diverge at the origin, and so do their derivatives: $x \ln x$ has divergent derivatives.

This difficulty can often (but not always!) be met by isolating the terms which cause the trouble, e.g.

$$\ln x(1-x) = \ln x + \ln(1-x) = \ln x - (x + \tfrac{1}{2}x^2 + \ldots),$$

$$\frac{e^{-x}}{x} = \frac{1}{x}(1 - x + \tfrac{1}{2}x^2 - \ldots) = \frac{1}{x} - 1 + \frac{x}{2} - \frac{x^2}{6} + \ldots$$

The part isolated is the 'divergent' part of the function and indicates the nature of the divergence, e.g. $\ln x(1-x)$ has a logarithmic divergence near $x = 0$ and $x = 1$, whereas e^{-x}/x diverges as $(1/x)$ near the origin. Note that

$$\frac{\sinh x}{x} = \frac{1}{x}\left(x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots\right) = 1 + \frac{x^2}{6} + \frac{x^4}{120} + \ldots$$

has finite derivatives in the limit $x \to 0$.

13.18. Asymptotic Expansion. This type of expansion cannot be discussed at all thoroughly here. It is a series having partial sums which give successive approximations for large values of x. Sometimes it is a power series in $(1/x)$, but usually other functions are involved. If it is a power series in $y = (1/x)$ the usual methods are employed, using y as variable, e.g.

$$\frac{1}{1+x} = \frac{1}{1+(1/y)} = \frac{y}{1+y} = y(1 - y + y^2 - \ldots), \quad (y < 1).$$
$$= \frac{1}{x} - \frac{1}{x^2} + \frac{1}{x^3} - \frac{1}{x^4} + \ldots, \quad (x > 1).$$

By integration by parts we find when x is large

$$\int_x^\infty e^{-t^2} dt = \int_x^\infty (-2te^{-t^2}) \left(-\frac{1}{2t}\right) dt = \left[-\frac{1}{2t} e^{-t^2}\right]_x^\infty - \int_x^\infty \frac{1}{2t^2} e^{-t^2} dt,$$
$$= \frac{1}{2x} e^{-x^2} + \left[\frac{1}{4t^3} e^{-t^2}\right]_x^\infty + \int_x^\infty \frac{3}{4t^5} e^{-t^2} dt$$
$$= e^{-x^2} \left(\frac{1}{2x} - \frac{1}{4x^3} + \frac{3}{8x^5} \ldots\right).$$

13.19. Fourier Series. A Fourier series is a finite or infinite series of the form,

$$S(\theta) = a_0 + a_1 \cos\theta + a_2 \cos 2\theta + a_3 \cos 3\theta + \ldots a_n \cos n\theta + \ldots$$
$$+ b_1 \sin\theta + b_2 \sin 2\theta + b_3 \sin 3\theta + \ldots b_n \sin n\theta + \ldots$$

or

$$S(\theta) = c_0 + c_1 \sin(\theta - \alpha_1) + c_2 \sin(2\theta - \alpha_2) + \ldots$$

These two are completely equivalent.

Assuming that the series converges, or is finite, the sum $S(\theta)$ has various properties. It is a *periodic* function, with a period 2π in θ, because

$$\sin n(\theta + 2\pi) = \sin n\theta, \quad \cos n(\theta + 2\pi) = \cos n\theta,$$

i.e.

$$S(\theta + 2\pi) = S(\theta).$$

It is an odd function of θ when every coefficient a_n is zero, and an even function of θ when every coefficient b_n is zero.

These coefficients can be expressed in terms of the sum by integration, because of some remarkable properties of the circular functions embodied in the integrals I.154–56. Briefly, these state that the mean value of the products of any two of the sequence of functions used in the Fourier series between $\theta = 0$ and 2π is zero, and

the mean value of the square of each is $\frac{1}{2}$ (except $1 = \cos 0\theta$, which has a mean value 1).

$$\frac{1}{2\pi}\int_0^{2\pi} \cos^2 m\theta\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} \sin^2 m\theta\, d\theta = \tfrac{1}{2}, \qquad \frac{1}{2\pi}\int_0^{2\pi} d\theta = 1,$$

(m, n, integers).

$$\frac{1}{2\pi}\int_0^{2\pi} \cos m\theta \cos n\theta\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} \sin m\theta \sin n\theta\, d\theta = 0$$

$(m \neq n)$ $\qquad$ $(m \neq n)$

$$\frac{1}{2\pi}\int_0^{2\pi} \sin m\theta \cos n\theta\, d\theta = 0.$$

Consequently, the average between $\theta = 0$ and 2π of every term in a Fourier series multiplied by $\sin m\theta$ is zero, except the term $b_m \sin m\theta$, the average of which multiplied by this function is $\frac{1}{2}b_m$. Assuming that the average of the whole series multiplied by $\sin m\theta$ is the sum of the averages of the individual terms, we find that

$$\tfrac{1}{2}b_m = \frac{1}{2\pi}\int_0^{2\pi} S(\theta) \sin m\theta\, d\theta.$$

Similarly by averaging every term multiplied by $\cos m\theta$ (including $m = 0$)

$$\tfrac{1}{2}a_m = \frac{1}{2\pi}\int_0^{2\pi} S(\theta) \cos m\theta\, d\theta, \qquad a_0 = \frac{1}{2\pi}\int_0^{2\pi} S(\theta) d\theta.$$

By giving m the values 1, 2, 3, . . . we obtain every coefficient, in terms of the sum $S(\theta)$ multiplied by that sine or cosine which multiplies the coefficient in the series. This rule is even simpler than Maclaurin's theorem for power series.

Because $S(\theta)$ is periodic, only its values over a range of 2π are required. Also the range of integration need not be 0 to 2π, as given, but $-\pi$ to $+\pi$ or any convenient range of 2π (cf. Section 11.14).

13.20. Sum of a Fourier Series. If the above Fourier series is written down with coefficients a_n and b_n calculated by the above formulae, with any periodic function $f(\theta)$ in the integrands instead of $S(\theta)$, it can be proved that the series does in fact have $f(\theta)$ as its sum for an extraordinary variety of functions. Owing to the fact that the coefficients are obtained by integration the function can be discontinuous. Suppose the function $f(\theta)$ has a discontinuity at $\theta = \alpha$, changing from $f(\alpha - 0)$ for values slightly less than α to $f(\alpha + 0)$ for values slightly greater than α.

As θ increases to α, the sum of the Fourier series equals $f(\theta)$ all the

time and therefore tends to the limit $f(\alpha - 0)$. As θ decreases to α, the sum of the series tends to the limit $f(\alpha + 0)$. But when $\theta = \alpha$, the Fourier series has a sum which is a compromise,

$$S(\alpha) = \tfrac{1}{2}\{f(\alpha - 0) + f(\alpha + 0)\}.$$

Thus the Fourier series has a discontinuous sum. This distinguishes Fourier series from power series, which are continuous and differentiable within the radius of convergence.

Example 1. $f(\theta) = \sin^2 \theta$ can be expressed as a finite series without the use of the general method.

$$\sin^2 \theta = \tfrac{1}{2} - \tfrac{1}{2} \cos 2\theta.$$

Example 2. The periodic function which is $f(\theta) = \theta$ between $\theta = -\pi$ and $+\pi$ can be expressed in the form

$$\theta = 2\{\sin \theta - \tfrac{1}{2} \sin 2\theta + \tfrac{1}{3} \sin 3\theta - \tfrac{1}{4} \sin 4\theta + \ldots\}.$$

By elementary integrations,

$$\tfrac{1}{2}b_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} \theta \sin m\theta \, d\theta = \frac{1}{\pi}\int_{0}^{\pi} \theta \sin m\theta \, d\theta = \frac{(-1)^{m+1}}{m}.$$

There are no terms in $\cos n\theta$ because the function is an odd function of θ.

13.21. ERRORS OF FOURIER SERIES. Let the Fourier series be $S(\theta) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta)$. If only the partial sum $S_m(\theta)$ up to terms in $(m\theta)$ is used as an approximation, there is an error

$$E_m(\theta) = S(\theta) - S_m(\theta) = S(\theta) - a_0 - \sum_{n=1}^{m} (a_n \cos n\theta + b_n \sin n\theta).$$

Whereas for a power series the effect of taking more terms is to improve the accuracy near a particular point, in a Fourier series it has the effect of improving the average accuracy over one period. In fact, the mean square error over one period, from $\theta = 0$ to $\theta = 2\pi$ is

$$\overline{E_m^2} = \frac{1}{2\pi}\int_0^{2\pi} [S - S_m]^2 d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} [S^2(\theta) - 2S(\theta)\{a_0 + \sum_1^m (a_n \cos n\theta + b_n \sin n\theta)\}$$

$$+ \{a_0 + \sum_1^m a_n \cos n\theta + b_n \sin n\theta\}^2] d\theta.$$

Except the first, all these integrals are mean values which have already been calculated, and

$$\overline{E_m^2} = \frac{1}{2\pi}\int_0^{2\pi} S^2(\theta)d\theta - 2a_0^2 - \sum_1^m (a_n^2 + b_n^2) + a_0^2 + \tfrac{1}{2}\sum_1^m (a_n^2 + b_n^2)$$

$$= \overline{S^2(\theta)} - \{a_0^2 + \tfrac{1}{2}\sum_{m=1}^n a_n^2 + b_n^2\}.$$

Thus the inclusion of each term, $a_n \cos n\theta$ reduces the mean square error by $\frac{1}{2}a_n^2$, and the inclusion of each term, $b_n \sin n\theta$ reduces it by $\frac{1}{2}b_n^2$.

We now prove that the coefficients are so chosen that the mean square error is least for a series expansion of this sort. Let us assume that the coefficients of the terms up to $m\theta$ have been chosen in this way, and an attempt is made to reduce the error still further by adding an extra term in $\cos (m + 1)\theta$, making in all the series

$$S'_m(\theta) = a_0 + \sum_{n=1}^m (a_n \cos n\theta + b_n \sin n\theta) + A \cos (m + 1)\theta.$$

The new mean square error with this extra term is

$$\overline{E'^2} = \overline{E_m^2} - \frac{1}{2\pi}\int_0^{2\pi} 2A \cos (m + 1)\theta\{S - S_m(\theta)\}d\theta + \frac{1}{2\pi}\int_0^{2\pi} A^2 \cos^2 (m + 1)\theta \, d\theta$$

$$= \overline{E_m^2} - Aa_{m+1} + \tfrac{1}{2}A^2,$$

which is a minimum when $A = a_{m+1}$.

In this connection it should be noticed that the working shows that this value of A is obtained, whatever terms may have been included in the sum. That is to say, even if some of the terms before $\cos (m + 1)\theta$ have been omitted, or later terms have been included, the value of A which reduces the mean square error by the greatest amount is always the same.

If the series converges to the function $f(\theta)$, then the error tends to zero as m tends to infinity, so that

$$\frac{1}{2\pi}\int_0^{2\pi} S^2(\theta)d\theta = a_0^2 + \tfrac{1}{2}\sum_0^\infty (a_n^2 + b_n^2).$$

13.22. ALTERATION OF RANGE. By the substitution $\theta = 2\pi x/L$, the Fourier series $S(\theta)$ becomes a function with a period L in x:

$$f(x) = a_0 + a_1 \cos\frac{2\pi x}{L} + a_2 \cos\frac{4\pi x}{L} + \ldots + a_n \cos\frac{2\pi n x}{L} + \ldots$$

$$+ b_1 \sin\frac{2\pi x}{L} + b_2 \sin\frac{4\pi x}{L} + \ldots + b_n \sin\frac{2\pi n x}{L} + \ldots$$

The formulae for the coefficients in terms of mean values are the same as before, since mean values are clearly not affected by the change in scale involved in going from θ to x.

$$a_0 = \frac{1}{L}\int_0^L f(x)dx, \quad \tfrac{1}{2}a_n = \frac{1}{L}\int_0^L f(x)\cos\frac{2\pi n x}{L}\,dx,$$

$$\tfrac{1}{2}b_n = \frac{1}{L}\int_0^L f(x)\sin\frac{2\pi n x}{L}\,dx.$$

13.23. ARBITRARY FUNCTIONS AND FOURIER SERIES. The fact that a Fourier series happens to be a periodic function should not blind us to the fact that quite general functions can be expressed in terms of a Fourier series whose sum equals the function over a limited range.

For example, a function $f(\theta)$ which is only required in the limited range 0 to 2π can be expressed as a Fourier series by substituting $f(\theta)$ for $S(\theta)$ in the formulae for the coefficients given in Section 13.19. Then:

(*i*) the sum of the Fourier series obtained is $f(\theta)$ between $\theta = 0$ and 2π;

(*ii*) the sum of the Fourier series is a periodic function with period 2π in θ, and is therefore not equal to $f(\theta)$ outside this range unless $f(\theta)$ is also periodic with the same period.

By changing the beginning of the range, and by alteration of the range, the same can be done for $f(x)$ over any range.

Example 1. Between $-\pi$ and $+\pi$, $f(\theta) = \cos\frac{1}{2}\theta$ is

$$\cos\tfrac{1}{2}\theta = \frac{2}{\pi}\left\{1 + \frac{2}{1.3}\cos\theta - \frac{2}{3.5}\cos 2\theta + \frac{2}{5.7}\cos 3\theta - \ldots\right\},$$

$$(-\pi \leqslant \theta \leqslant \pi),$$

the coefficients being

$$b_n = 0, \; a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos\tfrac{1}{2}\theta\,d\theta, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi}\cos\tfrac{1}{2}\theta\cos n\theta\,d\theta.$$

Between 0 and 2π the same function is

$$\cos \tfrac{1}{2}\theta = \frac{4}{\pi}\left\{\frac{2}{1.3}\sin\theta + \frac{4}{3.5}\sin 2\theta + \frac{6}{5.7}\sin 3\theta + \ldots\right\}, \qquad (0 < \theta < 2\pi),$$

the coefficients being

$$a_0 = a_n = 0,\ \tfrac{1}{2}b_n = \frac{1}{2\pi}\int_0^{2\pi} \cos\tfrac{1}{2}\theta \sin n\theta\, d\theta = \frac{4n}{\pi(2n-1)(2n+1)}.$$

(*N.B.* At $\theta = 0$ the two are not equal, cf. Section 13.20. Draw a diagram.)

13.24. NUMERICAL SERIES OBTAINED FROM FOURIER SERIES. By giving the variable special values, the sums of some useful series are obtained. For example we find, putting $\theta = \frac{1}{2}\pi$ in *Example 2* of Section 13.20, that

$$\frac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \ldots$$

Rather surprisingly, the same result can be obtained by putting $\theta = 0$ in the completely different series in the last section,

$$1 = \frac{2}{\pi}\left\{1 + \frac{2}{1.3} - \frac{2}{3.5} + \ldots\right\} = \frac{2}{\pi}\{1 + (1 - \tfrac{1}{3}) - (\tfrac{1}{3} - \tfrac{1}{5}) + \ldots\},$$

$$= \frac{4}{\pi}\{1 - \tfrac{1}{3} + \tfrac{1}{5} - \ldots\}.$$

Similarly, from the series $(-\pi \leqslant \theta \leqslant \pi)$,

$$\theta^2 = \tfrac{1}{3}\pi^2 + 4\sum_{m=1}^{\infty}(-)^m \frac{\cos m\theta}{m^2},$$

$$= \tfrac{1}{3}\pi^2 + 4\left\{-\frac{\cos\theta}{1^2} + \frac{\cos 2\theta}{2^2} - \ldots\right\},$$

we find the following results:

$$\theta = 0: \qquad 1 - \frac{1}{2^2} + \frac{1}{3^2} - \ldots = \frac{\pi^2}{12}.$$

$$\theta = \pi: \qquad 1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = \frac{\pi^2}{6}.$$

Adding these two,

$$1 + \frac{1}{3^2} + \frac{1}{5^2} + \ldots = \frac{\pi^2}{8}.$$

Finally, using the result at the end of Section 13.21,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\theta^4\,d\theta = \frac{\pi^4}{5} = a_0^2 + \tfrac{1}{2}\Sigma a_n^2 = \frac{\pi^4}{9} + 8\left(\frac{1}{1^4} + \frac{1}{2^4} + \ldots\right),$$

i.e.
$$1 + \frac{1}{2^4} + \frac{1}{3^4} + \ldots = \frac{\pi^4}{90}.$$

13.25. EXPANSIONS IN SERIES OF ORTHOGONAL FUNCTIONS. The expansion of an arbitrary function in terms of a Fourier series over a limited range of values is not the only expansion of this form. A problem which often arises in the solution of partial differential equations is that of expressing a function $F(x)$ between, say $x = a$ and $x = b$ in a series of the form

$$F(x) = a_1 f_1(x) + a_2 f_2(x) + \ldots + a_n f_n(x) + \ldots$$

where $f_1(x)$, $f_2(x)$, $\ldots$ $f_n(x)$, $\ldots$ form an ordered sequence of orthogonal functions.

By orthogonal we mean that, for any two different members of the set,

$$\int_a^b f_m(x) f_n(x)\,dx = 0, \quad m \neq n.$$

We shall also require

$$\int_a^b \{f_m(x)\}^2 dx = \mathcal{N}_m, \quad (m = 1, 2, 3, \ldots).$$

The orthogonality condition is satisfied by the functions 1, $\cos\theta$, $\sin\theta$, $\ldots$ used in a Fourier series if $a = 0$, $b = 2\pi$. These functions therefore form an orthogonal set, and a brief reference to earlier sections shows that it is this very property which makes it easy to obtain the coefficients in a Fourier series.

This suggests that the same method should be employed to find the coefficients in the new type of series. To find any coefficient a_n, multiply both sides by the function $f_n(x)$ associated with it in the series and integrate from $x = a$ to $x = b$.

$$\int_a^b F(x) f_n(x)\,dx = \int_a^b \{a_1 f_1(x) + a_2 f_2(x) + \ldots\} f_n(x)\,dx,$$

$$= a_n \int_a^b f_n(x)^2 dx, \text{ other terms being zero.}$$

This leads to a series of the form $\Sigma a_n f_n(x)$, where

$$a_n = \frac{1}{\mathcal{N}_n}\int_a^b F(x)\,f_n(x)\,dx, \qquad \mathcal{N}_n = \int_a^b \{f_n(x)\}^2 dx.$$

$\mathcal{N}_n$ is often called a normalizing factor, and the functions $f_n(x)$ are said to be normalized to unity if $\mathcal{N}_n = 1$.

The series with these coefficients may or may not converge to the function $F(x)$ when x lies between a and b. In either event, it can be shown that the partial sum,

$$F_n(x) = a_1 f_1(x) + \ldots + a_n f_n(x),$$

is an approximation to $F(x)$ with an error $F(x) - F_n(x) = E_n(x)$ whose mean square value between $x = a$ and $x = b$ is

$$\begin{aligned}(b-a)\overline{E_n^2} &= \int_a^b \{F(x) - F_n(x)\}^2 dx,\\ &= \int_a^b \{F(x)\}^2 dx - (a_1^2\mathcal{N}_1 + a_2^2\mathcal{N}_2 + \ldots + a_n^2\mathcal{N}_n).\end{aligned}$$

Thus the error diminishes with every term included.

If the series converges to the function required,

$$\lim_{n\to\infty}\overline{E_n^2} = 0, \quad \text{i.e.} \quad \sum_1^\infty a_n^2\mathcal{N}_n = \int_a^b F(x)^2 dx.$$

When this is true the set of functions $f_n(x)$ is said to form a complete set. (If it is not true there is at least one function $F(x)$ which is orthogonal to all the members of the set.)

13.26. Fourier Sine Series. The sine functions

$$f(x) = \sin\frac{n\pi x}{a}, \quad (n = 1, 2, 3, \ldots),$$

form a complete orthogonal set over the range $0 < x < a$. They are orthogonal, since

$$\int_0^a \sin\frac{m\pi x}{a}\sin\frac{n\pi x}{a}\,dx = \begin{cases} 0, & (m \neq n).\\ \frac{1}{2}a, & (m = n).\end{cases}$$

Thus a function $F(x)$ can be expanded in the form

$$F(x) = a_1\sin\frac{\pi x}{a} + a_2\sin\frac{2\pi x}{a} + \ldots$$

where $\frac{1}{2}a\,.\,a_n = \int_0^a F(x)\sin\frac{n\pi x}{a}\,dx.$

Example 1.

$$F(x) = 1 = \frac{4}{\pi}\left\{\sin\frac{\pi x}{a} + \frac{1}{3}\sin\frac{3\pi x}{a} + \ldots\right\}, \quad (0 < x < a).$$

13.27. FOURIER COSINE SERIES. $f_n = \cos(n\pi x/a)$ $(n = 0, 1, 2, \ldots)$.

$$F(x) = a_0 + \sum_1^\infty a_n \cos(n\pi x/a),$$

$$a_0 = \frac{1}{a}\int_0^a F(x)dx, \quad a_n = \frac{2}{a}\int_0^a F(x)\cos\frac{n\pi x}{a}\,dx.$$

These cosines are such that their derivatives vanish both at $x = 0$ and at $x = a$.

It may seem odd that the same function can be expressed in the same range as either a Fourier series of sines, as in the previous section, or as a series of cosines. This is because the two series only equal the function in the given range of values of x. Outside this range the two series are not always equal, e.g. between $x = -a$ and $x = 0$, and they are only equal to the function $F(x)$ outside the range if it has some special property.

Example 1.

$$\sin\theta = \frac{2}{\pi} - \frac{4}{\pi}\left\{\frac{\cos 2\theta}{1.3} + \frac{\cos 4\theta}{3.5} + \frac{\cos 6\theta}{5.7} + \ldots\right\}, \quad (0 < \theta < \pi).$$

13.28. LEGENDRE POLYNOMIALS. The polynomials defined in Section 13.16 form an orthogonal set.

$$P_0 = 1,\ P_1 = x,\ P_2 = \tfrac{1}{2}(3x^2 - 1),\ P_3 = \tfrac{1}{2}(5x^3 - 3x),\ \ldots$$

$$f(x) = \sum_0^\infty a_n P_n(x),\ (-1 \leqslant x \leqslant 1):\ a_n = \frac{2n+1}{2}\int_{-1}^1 f(x)P_n(x)dx.$$

$$\int_{-1}^1 P_m(x)P_n(x)dx = \begin{cases} 0, & (m \neq n). \\ 2/(2n+1), & (m = n). \end{cases}$$

They are often used with distributions in direction, where $x = \cos\theta$.

$$f(\theta) = \Sigma a_n P_n(\cos\theta), \quad (0 \leqslant \theta \leqslant \pi):$$

$$a_n = \frac{2n+1}{2}\int_0^\pi f(\theta)P_n(\cos\theta)\sin\theta\,d\theta.$$

13.29. PROOF OF FOURIER'S THEOREM. The following proof is simple and interesting. We prove that the partial sums $F_m(\theta)$ tend to $f(\theta)$ as $m \to \infty$.

$$F_m(\theta) = b_0 + \sum_{n=1}^{m} (b_n \cos n\theta + a_n \sin n\theta),$$

where $b_0 = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)d\theta, \qquad \frac{1}{2}b_n = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)\cos n\theta\, d\theta,$

and $\frac{1}{2}a_n = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)\sin n\theta\, d\theta.$

We write $\theta = \alpha$ in $F_m(\alpha)$ and substitute these expressions for the coefficients

$$F_m(\alpha) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)\,[1 + 2\sum_{n=1}^{m}(\cos n\theta \cos n\alpha + \sin n\theta \sin n\alpha)]d\theta,$$

$$= \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)\,[1 + \sum_{n=1}^{m} 2\cos n(\theta - \alpha)]d\theta.$$

This can be simplified using the identity

$$1 + 2\cos\beta + 2\cos 2\beta + \ldots + 2\cos n\beta = \sin(n + 1)\beta/\sin\beta,$$

with $\beta = \theta - \alpha$. Then

$$F_m(\alpha) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(\theta)\,\frac{\sin(m+1)(\theta-\alpha)}{\sin(\theta-\alpha)}\,d\theta$$

$$= \frac{1}{2\pi}\int_{-(m+1)(\pi+\alpha)}^{(m+1)(\pi-\alpha)} f\left(\alpha + \frac{\phi}{m+1}\right)\frac{\sin\phi}{\sin\left(\frac{\phi}{m+1}\right)}\,\frac{d\phi}{m+1},$$

where $\phi = (m + 1)(\theta - \alpha)$.

The integrand is now divided into the factors which change relatively slowly over the range of integration, and factors which oscillate rapidly. In case $f(\theta)$ is discontinuous at $\theta = \alpha$, it is advisable to divide the range of integration into two parts for ϕ positive and ϕ negative.

For ϕ positive, the integral is simplified using the Mean Value Theorem for integrals (Section 11.15). We assume that α is neither π nor $-\pi$.

This integral is

$$\frac{1}{2\pi}\int_0^{(m+1)(\pi-\alpha)}\left[f\left(\alpha+\frac{\phi}{m+1}\right)\frac{\dfrac{\phi}{m+1}}{\sin\left(\dfrac{\phi}{m+1}\right)}\right]\frac{\sin\phi}{\phi}\,d\phi$$

$$=\frac{1}{2\pi}\left[f\left(\alpha+\frac{\phi}{m+1}\right)\frac{\dfrac{\phi}{m+1}}{\sin\left(\dfrac{\phi}{m+1}\right)}\right]_{\phi=\phi_1}\cdot\int_0^{(m+1)(\pi-\alpha)}\frac{\sin\phi}{\phi}\,d\phi,$$

where ϕ_1 lies between 0 and $(m+1)(\pi-\alpha)$. This is a very large range, but as $(\sin\phi/\phi)$ oscillates $\frac{1}{2}(m+1)$ times in the range of integration and decreases steadily as ϕ increases, it would seem that the value of ϕ_1 should be finite and fairly small whatever the value of m. Hence, in the limit $m\to\infty$,

$$\int_0^{(m+1)(\pi-\alpha)}\frac{\sin\phi}{\phi}\,d\phi\to\int_0^{\infty}\frac{\sin\phi}{\phi}\,d\phi=\pi,$$

$$f\left(\alpha+\frac{\phi_1}{m+1}\right)\to f(\alpha+0),\quad\frac{\dfrac{\phi_1}{m+1}}{\sin\left(\dfrac{\phi_1}{m+1}\right)}\to 1.$$

The whole expression tends to

$$\frac{1}{2\pi}f(\alpha+0)\pi=\tfrac{1}{2}f(\alpha+0).$$

The integral for negative values of ϕ is treated in the same way and in the limit $m\to\infty$ becomes $\frac{1}{2}f(\alpha-0)$. Hence, in all,

$$\lim_{m\to\infty}F_m(\alpha)=\tfrac{1}{2}\{f(\alpha+0)+f(\alpha-0)\}.$$

This is the value predicted earlier, which becomes $f(\alpha)$ when the function is continuous.

EXAMPLES XIII

1. In Einstein's approximate model of a crystal the partition function f is the series

$$f=e^{-\frac{1}{2}x}+e^{-\frac{3}{2}x}+e^{-\frac{5}{2}x}+\ldots,$$

where $x=h\nu/kT$. Sum the series and calculate the internal energy

$$E=NkT^2\frac{\partial}{\partial T}(\log f),$$

where N is the number of atoms.

2. Given the heat constant

$$\Delta H = \Delta H_0 + \alpha T + \beta T^2 + \gamma T^3 + \ldots,$$

calculate ΔF from the Gibbs–Helmholtz equation,

$$\Delta F - \Delta H = T\frac{\partial}{\partial T}(\Delta F),$$

as a function of temperature at constant pressure.

3. If the pressure change $p_2 - p_1$ is small, show that

$$\log\frac{p_2}{p_1} = \frac{p_2 - p_1}{p_1} - \frac{1}{2}\left(\frac{p_2 - p_1}{p_1}\right)^2 + \ldots$$

$$= \frac{p_2 - p_1}{p_2 + p_1} + \frac{1}{3}\left(\frac{p_2 - p_1}{p_2 + p_1}\right)^3 + \ldots$$

4. Expand the Langevin function in powers of y,

$$L(y) = \coth y - \frac{1}{y} = \frac{1}{3}y - \frac{1}{45}y^3 + \ldots$$

and verify that $L(y) \to 1$ as $y \to \infty$.

5. Find a power series for $y = \dfrac{\sin^{-1} x}{(1 - x^2)^{1/2}}$, where $y = 0$ when $x = 0$, proving and using the equation $(1 - x^2)(dy/dx) = 1 + xy$.

6. Show that the power series $y = a_0 + a_2x^2 + a_4x^4 + \ldots$ satisfies the equation $xy'' + y' + xy = 0$ if

$$a_2 = -\frac{1}{2^2}a_0, \quad a_4 = \frac{1}{2^2 4^2}a_0, \quad a_6 = -\frac{1}{2^2 4^2 6^2}a_0, \ldots$$

7. Expand $(e^x - 1)^{-1}$ in powers of e^{-x} (*N.B.* not e^x) where $x > 0$. Hence find the value of the Debye specific heat function

$$D = \frac{3}{a^3}\left[\frac{\pi^4}{15} - \int_a^\infty \frac{x^3}{e^x - 1}\,dx\right]$$

to the first terms in e^{-a}, where $a = (\theta/T) \gg 1$, θ being the Debye temperature, and T the absolute temperature.

8. From the Fourier cosine series for θ^4 between 0 and π, prove that

$$\sum_1^\infty \frac{1}{n^4} = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \ldots = \frac{\pi^4}{90}.$$

Use this to evaluate $\displaystyle\int_0^\infty \frac{x^3}{e^x - 1}\,dx$.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

PIPES, L. A. *Applied Mathematics for Engineers and Physicists*, Chapter 3. McGraw-Hill Book Co., New York, 1946.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering*, Chapter 5. McGraw-Hill Book Co., New York, 1939.

FRANKLIN, L. *Fourier Methods*. McGraw-Hill Book Co., New York, 1949.

CHURCHILL, R. V. *Fourier Series and Boundary Value Problems*. McGraw-Hill Book Co., New York, 1941.

RELTON, F. E. *Applied Bessel Functions*. Blackie and Sons, Ltd, London and Glasgow, 1948.

CHAPTER 14

COMPLEX NUMBERS

Introduction—Algebra of complex numbers—Conjugate complex numbers. Modulus—Division—Argand diagram—de Moivre's theorems—Complex circular functions—Complex numbers as vectors—Modulus and argument—Power series—Complex exponential function—Logarithmic function—Hyperbolic functions—Circular functions—Powers and roots of z

14.01. INTRODUCTION. It is easy to obtain expressions involving $\sqrt{-1} = (-1)^{1/2}$ by ordinary algebraic processes. For example, $x^2 + 2bx + c = (x + b)^2 + (c - b^2)$ is not zero for any positive or negative value of x when $c > b^2$, but the formal solution of the quadratic equation $x^2 + 2bx + c = 0$ is

$$x = -b \pm (b^2 - c)^{1/2} = -b \pm (c - b^2)^{1/2} \sqrt{-1}.$$

It is a triumph of mathematicians to find not only satisfaction in the formal discussion of quantities of this sort but practical advantages in their use, and even a theoretical necessity in the case of Schrödinger's wave mechanics.

Consider one formal advantage in allowing such quantities. An equation of the nth degree in z,

$$a + bz + cz^2 + \ldots + kz^n = 0,$$

now possesses exactly n roots, so that the LHS can always be written as a product of n linear factors, by the Remainder Theorem (Section 6.04). This means in turn that quadratic denominators need not appear in the method of partial fractions (Section 6.15). It also means that a number has n nth roots.

Example 1. The square roots of -1 satisfy $z^2 = -1$, i.e. $z = \pm (-1)^{1/2}$. The cube roots of $+1$ satisfy $z^3 = 1$, or

$$z^3 - 1 = (z - 1)(z^2 + z + 1) = 0,$$

i.e.
$$z = -1 \quad \text{or} \quad \tfrac{1}{2}\{-1 \pm (-3)^{1/2}\}.$$

The fourth roots of -1 are obtained by solving the equation

$$z^4 + 1 = (z^2 - z\sqrt{2} + 1)(z^2 + z\sqrt{2} + 1) = 0.$$

As a second example, consider the two equations obtained in

Section 8.11, when the hyperbolic and circular functions were compared:

$$d^2z/dt^2 = m^2z: \qquad d^2z/dt^2 = -m^2z.$$

If $z = \cosh mt$ or e^{mt} is a solution of the first, then $z = \cosh (mt\sqrt{-1})$ or $e^{mt\sqrt{-1}}$ ought to be a solution of the second.

The purpose of this chapter is to elucidate and give the general principles in the use of these quantities, beginning with their treatment as numbers.

14.02. ALGEBRA OF COMPLEX NUMBERS. Fundamentally a complex number z is a pair of numbers, one called its real part (a) and the other its imaginary part (b). Combinations of two complex numbers $z_1(a_1, b_1)$ and $z_2(a_2, b_2)$ are defined as follows:

			Real Part	*Imaginary Part*
Equality	$z_1 = z_2$	implies	$a_1 = a_2$	$b_1 = b_2$
Addition	$z_1 + z_2$	has	$a_1 + a_2$	$b_1 + b_2$
Subtraction	$z_1 - z_2$	has	$a_1 - a_2$	$b_1 - b_2$
Product	z_1z_2	has	$(a_1a_2 - b_1b_2)$	$(a_1b_2 + a_2b_1)$

Note especially that the equality of complex numbers gives two ordinary equations. These complicated rules are more conveniently expressed using a change in notation. If the complex numbers are written in the form $z_1 = a_1 + ib_1$ and $z_2 = a_2 + ib_2$ and combined according to the ordinary rules of algebra, then all the above rules are satisfied provided that:

(i) i^2 is replaced by -1 whenever it occurs, and
(ii) the real part of the answer is the part which does not contain i and the imaginary part of the answer is the coefficient of i.

Equality $\quad a_1 + ib_1 = a_2 + ib_2 \quad$ implies $\quad a_1 = a_2,\ b_1 = b_2.$

Addition $\quad (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2)$, cf. above.

Subtraction $\quad (a_1 + ib_1) - (a_2 + ib_2) = (a_1 - a_2) + i(b_1 - b_2).$

Product $\quad (a_1 + ib_1)(a_2 + ib_2) = a_1a_2 + ib_1a_2 + a_1ib_2 + i^2b_1b_2,$

$$= (a_1a_2 - b_1b_2) + i(b_1a_2 + a_1b_2).$$

This list corresponds exactly to the earlier list. The symbol i distinguishes the real and imaginary parts and i^2 is replaced by -1 to fit the law of multiplication.

An ordinary, or real, number can be regarded as a complex number with its imaginary part zero $(a + 0i)$, because such

complex numbers combine in the same way as real ones (put $b_1 = b_2 = 0$ in the list above!).

Complex numbers $0 + bi = ib$ having their real part zero are called pure imaginary numbers.

Any complex number $a + ib$ is the sum of a real number $a + 0i$ and an imaginary number $0 + ib$.

14.03. CONJUGATE COMPLEX NUMBERS. MODULUS. The two complex numbers $z = a + ib$ and $z^\star = a - ib$ are said to be conjugate, i.e. the conjugate complex of z is obtained by changing the sign of its imaginary part, or changing the sign of i, which comes to the same thing.

The sum and product of conjugate complex numbers are both real:

$$z + z^\star = (a + ib) + (a - ib) = 2a,$$
$$zz^\star = (a + ib)(a - ib) = a^2 - i^2b^2 = a^2 + b^2.$$

The modulus of z is defined as $|z| = (a^2 + b^2)^{1/2} = (zz^\star)^{1/2}$. It is positive.

Example 1. $|3 + 4i| = (3^2 + 4^2)^{1/2} = 5.$

$|\cos\theta + i\sin\theta| = (\cos^2\theta + \sin^2\theta)^{1/2} = 1.$

$|z^n|$ is $(z^n z^{\star n})^{1/2} = (zz^\star)^{\frac{1}{2}n} = |z|^n.$

14.04. DIVISION. The quotient (z_1/z_2) is defined from first principles as a complex number z such that $zz_2 = z_1$.

Hence, as in ordinary algebra, the quotient can be multiplied above and below by the same complex number z_3 without changing its value (for $zz_2 = z_1$ implies $zz_2z_3 = z_1z_3$, i.e. $z = z_1z_3/z_2z_3$).

This rule makes it possible to find the real and imaginary parts of a fraction without any trouble. Division by a real number is easy,

$$(a_1 + ib_1)/a_2 = a_1/a_2 + i \,.\, b_1/a_2,$$

and the denominator can always be made real by multiplication above and below by its conjugate complex.

$$\frac{c + id}{a + ib} = \frac{(c + id)(a - ib)}{(a + ib)(a - ib)} = \frac{(ca + bd) + i(ad - bc)}{a^2 + b^2}.$$

Example 1. $\dfrac{i}{1 + i} = \dfrac{i(1 - i)}{1 - i^2} = \dfrac{1 + i}{2}.$

Example 2. $\dfrac{1}{\cos\theta + i\sin\theta} = \dfrac{\cos\theta - i\sin\theta}{\cos^2\theta - i^2\sin^2\theta} = \cos\theta - i\sin\theta.$

Example 3. The absorption of light by a dielectric can be described by a complex dielectric constant $\hat{n} = n(1 - ik)$. The real part n is the usual refractive index, and k is the absorption index. On the simplest theory of the dispersion of light of frequency $(\omega/2\pi)$,

$$\hat{n} = 1 + \alpha/(\omega_0^2 - \omega^2 + i\gamma\omega),$$

where α, ω_0 and γ are characteristics of the medium. Then

$$n(1 - ik) = \hat{n} = 1 + \alpha \frac{(\omega_0^2 - \omega^2) - i\gamma\omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2},$$

i.e. $$n = 1 + \alpha \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2}, \quad kn = \frac{\gamma\omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2}.$$

14.05. ARGAND DIAGRAM. On the Argand diagram a complex number $z = a + ib$ is represented by a point P with coordinates $x = a, y = b$. It is also represented by the directed line OP, or any parallel and equal line.

In *Figure 14.01*, two complex numbers z_1 and z_2 are represented by points P_1 and P_2 and their sum by Q. It is clear that OP_1QP_2 is a parallelogram.

The modulus of a complex number represented by P on the Argand diagram is the length of the line OP from the origin. From

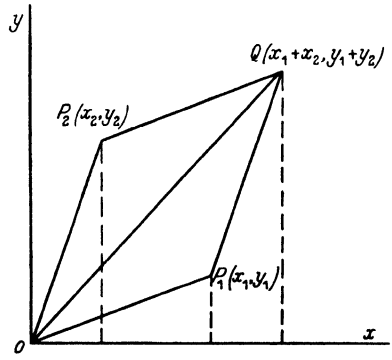

Figure 14.01

the *Figure 14.01* it follows that the modulus of the sum of two complex numbers is less than the sum of their moduli

$$OQ < OP_1 + OP_2,$$

and greater than their difference.

Example 1.

$$|a_0 + a_1 z + a_2 z^2 + \ldots a_n z^n| \leqslant |a_0| + |a_1 z| + |a_2 z^2| + \ldots + |a_n z^n|.$$

14.06. DE MOIVRE'S THEOREMS. By direct multiplication it is easy to verify that

(*1*) $(\cos\alpha + i\sin\alpha)(\cos\beta \pm i\sin\beta) = \cos(\alpha \pm \beta) + i\sin(\alpha \pm \beta)$.

By repeated application of this rule, with $\alpha = \beta$,

(*2*) $(\cos\alpha + i\sin\alpha)^p = \cos p\alpha + i\sin p\alpha, \quad (p = 1, 2, 3, \ldots)$.

Using the result $(\cos\beta + i\sin\beta)^{-1} = \cos\beta - i\sin\beta$, from the previous section,

(*3*) $\dfrac{\cos\alpha + i\sin\alpha}{\cos\beta + i\sin\beta} = \cos(\alpha - \beta) + i\sin(\alpha - \beta)$, from (*1*).

(*4*) $\dfrac{1}{(\cos\beta + i\sin\beta)^p} = \cos p\beta - i\sin p\beta$, from (*2*).

These are known as de Moivre's theorems.

14.07. COMPLEX CIRCULAR FUNCTIONS. The complex circular function $e^{i\theta}$ is defined by a general method as the function whose Maclaurin series is that of e^x when $x = i\theta$. We then find, using $i^2 = -1$, $i^3 = -i$, $i^4 = +1$, etc., that

$$e^{i\theta} = 1 + (i\theta) + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \ldots$$

$$= 1 + i\theta - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + i\frac{\theta^5}{5!} - \ldots$$

$$= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots\right)$$

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

The conjugate complex is

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

so that $$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}, \quad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

These results have widespread implications, since exponentials are so easy to handle. The reader should verify that de Moivre's

theorems now become trivial, e.g. (*1*) becomes $e^{i\alpha} \cdot e^{i\beta} = e^{i(\alpha+\beta)}$. The properties of circular functions given in Section 7.02 can all be proved using these formulae in the same way as properties of the hyperbolic functions were proved (Section 8.12). Some formulae are more easily obtained by this new method.

Example 1. Express $\cos^4\theta$ in terms of cosines of multiples of θ. By the Binomial Theorem,

$$(\cos\theta)^4 = \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^4 = \tfrac{1}{16}\,(e^{4i\theta} + 4e^{3i\theta} \cdot e^{-i\theta} + 6e^{2i\theta} \cdot e^{-2i\theta} + 4e^{i\theta} \cdot e^{-3i\theta} + e^{-4i\theta}),$$

$$= \tfrac{1}{16}\,\{(e^{4i\theta} + e^{-4i\theta}) + 4(e^{2i\theta} + e^{-2i\theta}) + 6\},$$

$$= \tfrac{1}{8}(\cos 4\theta + 4\cos 2\theta + 3).$$

Note the 3, rather than 6, from the last term.

Example 2. Sum the trigonometric series,

$$1 + \cos\alpha + \cos 2\alpha + \ldots + \cos n\alpha.$$

(*N.B.* There is a quicker alternative solution to this problem.)

This is the real part of $1 + e^{i\alpha} + e^{2i\alpha} + e^{3i\alpha} + \ldots + e^{in\alpha}$

$$= \frac{e^{i(n+1)\alpha} - 1}{e^{i\alpha} - 1} = \frac{e^{i(n+\frac{1}{2})\alpha} - e^{-\frac{1}{2}i\alpha}}{e^{\frac{1}{2}i\alpha} - e^{-\frac{1}{2}i\alpha}} = \frac{e^{i(n+\frac{1}{2})\alpha} - e^{-\frac{1}{2}i\alpha}}{2i\sin\frac{1}{2}\alpha}.$$

The real part of this is $\dfrac{\sin(n+\frac{1}{2})\alpha}{2\sin\frac{1}{2}\alpha} - \frac{1}{2}$.

N.B. the useful result $e^{i\alpha} - 1 = 2ie^{\frac{1}{2}i\alpha}\sin\frac{1}{2}\alpha$.

These formulae can also be used in the calculation of derivatives and integrals, using the relation

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

i.e.

$$\frac{d}{d\theta}(e^{i\theta}) = ie^{i\theta}.$$

14.08. Complex Numbers as Vectors. In an Argand diagram (*Figure 14.02*) the directed line or vector $\overrightarrow{OP}$ represents the complex number

$$z = a + ib = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Multiplication of z by $(\cos\alpha + i\sin\alpha) = e^{i\alpha}$ corresponds on the diagram to a rotation of the vector OP anticlockwise through an angle α to the position OQ. For the complex number becomes

$$ze^{i\alpha} = re^{i(\theta+\alpha)},$$

which is represented by OQ.

Similarly the division of z by $e^{i\alpha}$ rotates the vector clockwise through an angle α.

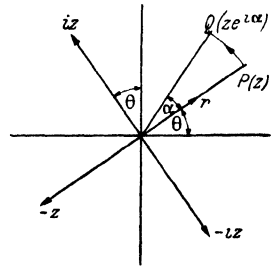

Figure 14.02

In the diagram three special cases are shown: (*i*) multiplication by i ($\alpha = 90°$, $e^{i\alpha} = i$) means an anticlockwise rotation through 90°; (*ii*) division by i (or multiplication by $-i$) implies a clockwise rotation through 90°; (*iii*) multiplication by $-1 = i^2$ ($\alpha = 180°$) rotates the vector through 180°.

Oscillating functions, such as $c\cos\omega t$, can also be represented by vectors. The function $c\cos\omega t$ is the real part of the complex number $ce^{i\omega t}$, which is represented by a vector OP, as above, with $r = c$ and $\theta = \omega t$. Its variation with time means that the vector rotates with angular velocity ω about the origin.

The original quantity, being the real part of the complex number, is represented by the projection of OP on the x axis.

Example 1. In a dielectric a static electric field E leads to an electric induction D, the ratio $D/E = \kappa$ being the dielectric constant.

In an oscillating electric field $E = E_0\cos\omega t$, the induction often has the form $D = D_0\cos(\omega t - \phi)$ where ϕ is a constant. The dielectric properties clearly involve both D_0/E_0 and ϕ. They are conveniently combined if we regard E and D as the real parts of

$$\hat{E} = E_0e^{i\omega t} \text{ and } \hat{D} = D_0e^{i(\omega t-\phi)}.$$

Then $$\hat{\kappa} = (\hat{D}/\hat{E}) = (D_0/E_0)e^{-i\phi}.$$

(A circumflex accent is used to denote this transition from real to complex quantities.)

14.09. MODULUS AND ARGUMENT. As in the previous section, any complex number can be represented in the form (*Figure 14.02*)

$$z = a + ib = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

This is most easily done on the Argand diagram, where (r, θ) are polar coordinates of the point (a, b).

The modulus of z is $|z| = r$, and the argument of z is $\theta = \arg z$. They are chosen so that $a = r\cos\theta$, $b = r\sin\theta$, i.e.

$$r = (a^2 + b^2)^{1/2};\ \cos\theta = (a/r),\ \sin\theta = (b/r).$$

The modulus is unique, but not the argument θ, which can be increased or decreased by a multiple of 2π without changing the complex number. In particular,

$$e^{2\pi ni} = \cos(2\pi n) + i\sin(2\pi n) = 1, \quad (n = 0, \pm 1, \pm 2, \ldots).$$

Multiplication and division are very easy when the complex numbers are known by their moduli and arguments.

$$z_1z_2 = (r_1e^{i\theta_1})(r_2e^{i\theta_2}) = (r_1r_2)e^{i(\theta_1+\theta_2)}: \quad \frac{z_1}{z_2} = \frac{r_1e^{i\theta_1}}{r_2e^{i\theta_2}} = \frac{r_1}{r_2}e^{i(\theta_1-\theta_2)}.$$

Summarizing the results:

(*1*) Equal complex numbers have equal moduli, and their arguments are equal or differ only by a multiple of 2π.

(*2*) In multiplication, moduli multiply and arguments add.

(*3*) In division, moduli divide and arguments subtract.

(*4*) Multiplication by $\cos\alpha + i\sin\alpha$ increases the argument by α.

(*5*) The modulus of a sum is not greater than and is usually less than the sum of the moduli (Section 14.04),

$$|z_1 + z_2| \leqslant |z_1| + |z_2|.$$

14.10. POWER SERIES. To obtain a function of a complex number z which is an analogue of a function $f(x)$ of a variable x, it is usual to express $f(x)$ as a power series $a_0 + a_1x + a_2x^2 + \ldots$ and then replace x by z.

$$f(z) = a_0 + a_1z + a_2z^2 + a_3z^3 + \ldots + a_nz^n + \ldots$$

These power series have real coefficients a_n: there are other power series in z with complex coefficients.

The partial sums

$$s_n = a_0 + a_1z + a_2z^2 + \ldots + a_nz^n, \quad (n = 0, 1, 2, \ldots),$$

can be separated into their real and imaginary parts, $s_n = u_n + iv_n$,

and the series is said to converge to the value $s = u + iv$ if the real and imaginary parts converge to u and v respectively. These two conditions can be reduced to a single condition which is equivalent:

$$(u - u_n)^2 + (v - v_n)^2 \to 0, \text{ or } |s - s_n| \to 0 \text{ as } n \to \infty.$$

In the same way, the condition for the absolute convergence of s, that u and v are absolutely convergent, reduces to the single condition that

$$\begin{aligned} &|a_0| + |a_1 z| + |a_2 z^2| + \ldots + |a_n z^n| + \ldots \\ &= |a_0| + |a_1| r + |a_2| r^2 + \ldots + |a_n| r^n + \ldots \end{aligned}$$

is convergent, where $|z| = r$. Now this is a real series, which converges (Section 13.09) when

$$|z| = r < R, \quad \text{where} \quad R = \lim_{n \to \infty} \left| \frac{a_n}{a_{n+1}} \right|.$$

The results previously given for the real series $\Sigma a_n x^n$ now appear as special results for $\Sigma a_n z^n$ when a and z are real.

The radius of convergence R has a definite meaning on the Argand diagram. Points which represent values of z for which the series is absolutely convergent lie inside a circle of radius R about the origin as centre, since $|z| < R$. The series usually converges at most points on the circumference of this circle, but the series, and the function it represents, diverges at some point or points.

14.11. Complex Exponential Function. This function is

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \ldots + \frac{z^n}{n!} + \ldots$$

It is absolutely convergent for all values of z. (Why?)

When $z = x$ is real, it is the ordinary exponential function and when $z = iy$ is imaginary it is the complex circular function (Section 14.07). In general, when $z = x + iy$,

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

Thus complex exponentials correspond to real oscillating functions having amplitudes which increase or decrease exponentially.

Example 1. When ω is real, the function $\hat{E} = E_0 e^{i\omega t}$ represents an oscillating electric field in complex form. When $\omega = p + iq$ is complex, with a *positive* imaginary part, it represents an oscillating electric field with a decaying amplitude

$$\hat{E} = E_0 e^{i(p+iq)t} = E_0 e^{-qt} e^{ipt}.$$

14.12. LOGARITHMIC FUNCTION. $w = \ln z$. This can be defined as a solution of the equation $e^w = z$, or, indirectly, by the series

$$\ln (1 + z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \ldots$$

The first is more direct: let $w = u + iv$ and $z = re^{i\alpha}$. Then

$$e^w = e^{u+iv} = re^{i\alpha}.$$

i.e. $\quad u = \ln r, \quad v = \alpha + 2\pi n, \quad (n = 0, \pm 1, \pm 2 \ldots).$

Thus $\ln z$ is not single-valued. The real part is unique, but not the imaginary part. To make v unique it is customary to restrict its values to some range of 2π.

14.13. HYPERBOLIC FUNCTIONS. These functions are defined by

$$\cosh z = \tfrac{1}{2}(e^z + e^{-z}), \quad \sinh z = \tfrac{1}{2}(e^z - e^{-z}),$$

and satisfy all the formulae given in Section 8.12.

The hyperbolic functions with imaginary argument, $z = iy$, are related to the circular functions:

$$\cosh iy = \tfrac{1}{2}(e^{iy} + e^{-iy}) = \cos y,$$
$$\sinh iy = \tfrac{1}{2}(e^{iy} - e^{-iy}) = i \sin y.$$

Thus $\cosh iy$ is real, in spite of appearances, and $\sinh iy$ is imaginary. The real and imaginary parts of $\cosh z$ and $\sinh z$ can now be found.

$$\begin{aligned}\cosh z &= \cosh (x + iy) = \cosh x \cosh iy + \sinh x \sinh iy \\ &= \cosh x \cos y + i \sinh x \sin y.\end{aligned}$$

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

14.14. CIRCULAR FUNCTIONS. By analogy with $\sin \theta$ and $\cos \theta$

$$\sin z = \frac{1}{2i}(e^{iz} - e^{-iz}), \quad \cos z = \tfrac{1}{2}(e^{iz} + e^{-iz}).$$

These definitions ensure that they reduce to the usual functions when z is real. All the usual trigonometric formulae hold for both real and complex values of z.

The circular functions with imaginary argument $z = iy$ are related to the hyperbolic functions.

$$\sin (iy) = \frac{1}{2i}(e^{-y} - e^{y}) = \tfrac{1}{2}i(e^{y} - e^{-y}) = i \sinh y,$$

$$\cos (iy) = \tfrac{1}{2}(e^{-y} + e^{y}) = \cosh y.$$

Thus sin (iy) is imaginary and cos (iy) is real. It might be added that these relations also hold when y is not real, but replaced by a complex number!

The real and imaginary parts are:

$$\begin{aligned}\sin(x+iy) &= \sin x \cos iy + \cos x \sin iy \\ &= \sin x \cosh y + i \cos x \sinh y, \\ \cos(x+iy) &= \cos x \cosh y - i \sin x \sinh y.\end{aligned}$$

14.15. Powers and Roots of z. According to Section 14.09, the positive and negative powers of z have the properties,

$$|z^n| = |z|^n, \quad \arg(z^n) = n \arg z.$$

The argument can be increased or reduced by multiples of 2π to bring it to a convenient size.

It follows from these relations that other complex numbers have the same nth power, if (i) they have the same modulus as z and (ii) their arguments differ from arg z by a multiple of $(2\pi/n)$.

Including z itself, there are just n different complex numbers with the same nth power. For starting with z and increasing its argument by $2\pi/n$, we obtain a second number with the desired property, and repeating this process twice, three times and more, we finally at the nth stage arrive at a number having its argument 2π greater than z. This number is identical with z, and going any further gives no new results. Furthermore, by decreasing the argument of z no new numbers are obtained with the same nth power.

To find the nth roots of a complex number $a + ib$, therefore, we first express it in the form $r_0(\cos\alpha + i \sin\alpha)$. One root is

$$z_0 = r_0^{1/n}\left(\cos\frac{\alpha}{n} + i\sin\frac{\alpha}{n}\right)$$

as may easily be verified. This is the Principal Root. The other roots, making n altogether, are obtained by increasing the argument of z_0 by multiples of $2\pi/n$.

$$\begin{aligned}z_r &= z_0\left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^r, \\ &= r_0^{1/n}\left\{\cos\frac{\alpha+2\pi r}{n} + i\sin\frac{\alpha+2\pi r}{n}\right\},\end{aligned} \qquad r = 1, 2, \ldots (n-1).$$

Example 1. The n complex nth roots of unity $(z = 1)$ are obtained by writing $r_0 = 1$ and $\alpha = 0$,

$$\left(\cos\frac{2\pi r}{n} + i\sin\frac{2\pi r}{n}\right), \quad r = 0, 1, 2, \ldots (n-1).$$

Example 2. On an Argand diagram the roots are equally spaced on the circumference of a circle (see *Example 2*, Section 14.04).

Thus, any complex number has exactly n complex roots. A real number must also have just n complex nth roots, but usually one of these is real, sometimes two (even roots of a positive number) and sometimes none (even roots of a negative number).

Approximate values of principal roots of a complex number with a very small modulus, $|z| \ll 1$, can be found using the Binomial Theorem,

$$(1 + z)^m = 1 + mz + \tfrac{1}{2}m(m - 1)z^2 \ldots$$

with $m = (1/n)$.

EXAMPLES XIV

1. Show that the locus of $z = ae^{i\theta}$ in the Argand diagram, where θ is a real parameter, is a circle of radius a.

Find the real and imaginary parts of $w = \frac{1}{2}\left(z + \frac{1}{z}\right)$ as functions of θ, and show that the locus of w in its Argand diagram is an ellipse (assume $a \geqslant 1$). What property of the ellipse remains fixed as a changes? What happens when $a = 1$?

2. Prove that the series $1 + r\cos\theta + r^2\cos 2\theta + \ldots$,

and $$r\sin\theta + r^2\sin 2\theta + \ldots,$$

are convergent when $r < 1$ and find their sums.

3. Show that the Fourier coefficients a_n and b_n in the expansion

$$f(x) = \sum_0^\infty (a_n \cos nx + b_n \sin nx), \quad (-\pi < x \leqslant \pi),$$

are the real and imaginary parts of

$$\frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)e^{inx}dx, \quad (n \geqslant 0).$$

4. If $z = e^{i\theta}$, prove that

$$\cos^n\theta = \frac{1}{2^n}\left(z + \frac{1}{z}\right)^n = \frac{1}{2^{n-1}}\{\cos n\theta + {}^nC_2\cos(n-2)\theta + \ldots\}.$$

and deduce the last term when n is odd or even.

5. Prove that $\cos n\theta = c^n - {}^nC_2\, c^{n-2}s^2 + {}^nC_4\, c^{n-4}s^4 - \ldots,$

and $\sin n\theta = nc^{n-1}s - {}^nC_3 c^{n-3}s^3 + \ldots,$

where $c = \cos\theta$, $s = \sin\theta$, using $e^{i\theta} = c + is$.

6. Find the modulus and argument of $w = a\,dz/d\theta + bz$, when $z = Ae^{(\alpha+i\beta)\theta}$ when A, α, β, a and b are real.

Obtain the real and imaginary parts of z and w.

7. Solve Qn *8* of Chapter 8 when $\alpha = \beta \cos\theta < \beta$.

8. Express $1/(x^2 + a^2)$ in partial fractions using complex roots of $(x^2 + a^2) = 0$, and deduce that

$$\int \frac{dx}{x^2 + a^2} = \text{imaginary part of } \frac{1}{2ia} \int \frac{dx}{x - ia}.$$

Evaluate the integral by this means.

9. If $z = x + iy$ is a function of t such that

$$a\frac{d^2z}{dt^2} + b\frac{dz}{dt} + cz = 0,$$

where a, b and c are *real* functions of t, show that x and y separately satisfy this equation.

What equations do x and y satisfy if b is imaginary?

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

DURELL, C. V. and ROBSON, A. *Advanced Trigonometry.* G. Bell and Sons, Ltd, London, 1938.

GREEN, S. L. *The Theory and Use of the Complex Variable.* Sir Isaac Pitman and Sons, Ltd, London, 1948.

PIPES. L. A. *Applied Mathematics for Engineers and Physicists.* McGraw-Hill Book Co, Inc, New York, 1946.

CHAPTER 15

DIFFERENTIAL EQUATIONS

Introduction—Terms used for differential equations—First-order equations—Absorption processes—Chemical reactions—First-order reaction—Second-order reaction—Reversible reactions—Some general remarks—Linear equations—Special linear equation—Further remarks—Successive reactions—Simultaneous equations—Steady heat flow—Diffusion equation—Cooling sphere—Second-order differential equations—Simple harmonic equation—Diatomic molecule—Exponential solutions—Damped harmonic motion—Heavily damped motion—Forced oscillations—Chain theory of reactions—Reducible equations—Bessel's equation—Solution by power series—Equations reducible to Bessel's equation

15.01. INTRODUCTION. In this chapter the main emphasis is on the various applications of differential equations, although the solution of all the common types is given. An attempt is also made to discuss the various types of difficulty which can occur in obtaining the solutions and how they are overcome.

15.02. TERMS USED FOR DIFFERENTIAL EQUATIONS. An *ordinary* differential equation contains two variables, e.g. x and y, and one is chosen as the independent variable, the other as the dependent variable.

Order: the index (n) of the highest derivative ($d^n y/dx^n$).

Degree: the highest power of the highest derivative in the equation.

Solution: any function $y = f(x)$ which satisfies the equation.

General solution: the general solution of an equation of order n is a solution containing n arbitrary constants.

Particular solution: a solution obtained from the general solution by assigning special values to some or all of the arbitrary constants.

Singular solution: any solution which is not a particular solution, i.e. obtained from the general solution. These rarely occur, but when they do, are usually of considerable importance and physical significance.

Solution of problem: this is usually a particular solution, obtained by using the conditions imposed by the problem to determine the arbitrary constants in the general solution.

Guesswork: this plays an important role in obtaining the general solutions of many equations, but not in finding the solution of many

problems, because the particular solution obtained by guesswork may not satisfy the conditions of the problem.

15.03. First-order Equations. Here is a list of the types of first-order equation the solutions of which can be expressed in terms of integrals (which then have to be evaluated). Examples of them appear in the sections below.

A first-order equation contains only x, y, and dy/dx, and can be written $(dy/dx) = \phi(x, y)$, where ϕ is some function of x and y. Equations of degree higher than one have to be solved for dy/dx to express them in this form. It can also be expressed as

$$dy = \phi(x, y)dx.$$

(*i*) *Simple.* An equation $dy = f(x)dx$, which can be integrated at once.

(*ii*) *Exact.* The equation can be written $P\,dx + Q\,dy = 0$, where P and Q are functions of x and y such that $\partial P/\partial y = \partial Q/\partial x$ (cf. Section 16.16).

(*iii*) *Separable.* The equation can be written $f(x)dx = g(y)dy$, and is solved by integrating both sides.

(*iv*) *Homogeneous.* In this context this term means that $\phi(x, y)$ can be reduced to a function of v alone, by writing $y = vx$. This type rarely occurs in practice and when it does, it is solved as a separable equation with v and x as variables. It is worth mention as the first equation which is more easily solved by changing the dependent variable from y to another variable, in this case v.

(*v*) *Linear in y.* Equation can be written $dy/dx + Py = Q$, where P and Q are independent of y (see Section 15.10).

Equations linear in x can be written as $dx/dy + Px = Q$, where P and Q are functions of y.

(*vi*) *Bernoulli's equation.* $(dy/dx) + Py = Qy^n$. This can be expressed and solved as a linear equation in z where $z = y^{1-n}$. Its terms contain either dy/dx or y or *one* other power of y.

If an equation, e.g. $dy/dx = y/x$, can be put under a number of these headings the easiest method of solution is generally the method which appears earliest in the list.

15.04. Absorption Processes. A typical absorption process is that of light passing through a solution. In its simplest form, the intensity of light decreases away from the source owing to interactions with particles or molecules, each interaction causing a definite loss. The probability of each interaction is proportional to the intensity of the light, and the number of interactions is proportional to the number

of absorbers, i.e. concentration times volume. Hence the decrease in intensity I of a parallel beam, per unit area, in a distance dx along the beam is

$$\begin{aligned} -dI &= \text{constant (intensity) (conc.) (volume)} \\ &= k \,.\, I \,.\, c \,.\, dx, \end{aligned}$$

where k is a constant depending on the medium and c is the concentration. This leads to a differential equation,

$$dI/dx = -kcI.$$

When c is a constant, this is an equation for exponential decay with the solution

$$I = I_0 e^{-kcx},$$

where I_0 is the intensity at $x = 0$.

In general, as a separable equation, it is put in the form

$$dI/I = -kc\,dx.$$

Integrate, choosing the constant of integration such that $I = I_0$ when $x = 0$ (cf. Section 10.08);

$$\ln I = \ln I_0 - k\int_0^x c\,dx,$$

i.e.

$$I = I_0 e^{-k\int_0^x c\,dx}.$$

When c is constant, this reduces to the answer given earlier.

15.05. CHEMICAL REACTIONS. The velocity of a single chemical reaction is measured by the rate of change with time of the concentration x of one of the reacting substances. This concentration determines the extent to which the reaction has occurred and the concentrations of all the other reactants, provided the volume is known or is determined by x. From the Law of Mass Action or simple kinetic theory it is usually inferred that the reaction rate is proportional to products of the amounts of the reacting molecules. Then the differential equation for x has the form of a separable equation,

$$dx/dt = f(x).$$

In the following example it is assumed for simplicity that there are no complicating factors, such as a variation with temperature, which affects the constants of proportionality, or mixtures of gases at constant pressure.

15.06. First-order Reaction. $A \rightarrow B + C$. (The order of a reaction is a chemical concept and not to be confused with the order of the differential equation.) In a spontaneous decomposition of this sort, the reaction rate is proportional to the concentration of A and is independent of the substances formed. If this concentration is initially a, decreasing to x at time t, then its rate of decrease is

$$dx/dt = -kx,$$

with the solution

$$x = ae^{-kt}, \quad \text{or} \quad kt = \ln (a/x).$$

If x is measured in mole/cm^3 or similar units, then

$$a - x = a(1 - e^{-kt})$$

gives the amount of B and C formed in the same units.

Since

$$\ln x = \ln a - kt,$$

experimental data would be used by plotting $\ln x$ against t. Alternatively the values of $(1/t) \ln (a/x)$ could be calculated at various times. The values of k deduced in this way should be the same.

15.07. Second-order Reactions. $A + B \rightarrow C + D$. The reaction rate is now proportional to the product of the concentrations of A and B. Let these be a and b initially and let x be the concentration change at time t, all in moles/unit volume. Then $(a - x)$ and $(b - x)$ are the concentrations at time t, respectively, and

$$dx/dt = k(a - x)(b - x).$$

If $a \neq b$ we separate the variables as follows,

$$k\,dt = \frac{dx}{(a - x)(b - x)} = \frac{dx}{b - a}\left\{\frac{1}{a - x} - \frac{1}{b - x}\right\}.$$

Integrate term by term, choosing the constant of integration so that $t = 0$ when $x = 0$;

$$kt = \frac{1}{b - a}\{-\ln (a - x) + \ln (b - x) + \ln a - \ln b\}$$

$$= \frac{1}{b - a} \ln \frac{a(b - x)}{b(a - x)} = \frac{1}{b - a} \ln \frac{1 - \frac{x}{b}}{1 - \frac{x}{a}}.$$

As a check on the calculation, note that $x/a > x/b$ when $x > 0$ and $a < b$, so that t is positive and tends to infinity as $x \rightarrow a$.

If B is in excess initially ($b \gg a$), then its concentration hardly changes during the reaction: the change from b to the final value $(b - a)$ is small. If this is so, the differential equation is hardly distinguishable from the equation

$$dx/dt = kb(a - x), \quad (b \gg a),$$

and the same can be verified for the solution. As this equation is that of a first-order reaction with a velocity constant kb, the solution is very similar to that of the previous section.

When $b = a$, the equation is

$$dx/dt = k(a - x)^2, \qquad \text{i.e. } k\,dt = dx/(a - x)^2.$$

Integrating, so that $t = 0$ when $x = 0$,

$$kt = 1/(a - x) - 1/a = x/\{a(a - x)\}.$$

15.08. REVERSIBLE REACTIONS. $2A \leftrightharpoons B + C$. The velocity consists of two terms (*i*) $k_1(a - x)^2$ due to the decomposition of A, with

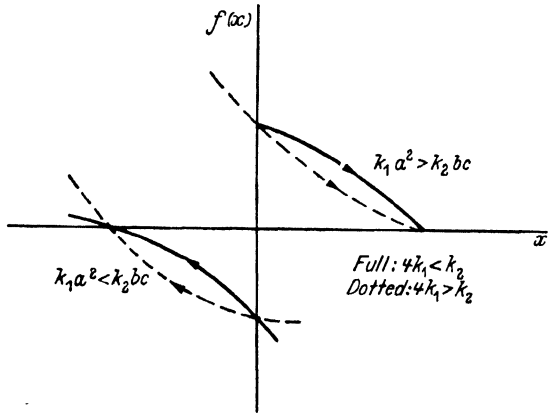

Figure 15.01

concentration a initially and x at time t, and (*ii*) $-k_2(b + \frac{1}{2}x)(c + \frac{1}{2}x)$ due to the recombination of B and C, the initial concentrations of which are b and c respectively (in moles/unit volume),

i.e. $$dx/dt = k_1(a - x)^2 - k_2(b + \tfrac{1}{2}x)(c + \tfrac{1}{2}x) = f(x).$$

If the reaction proceeds so that x increases, then $k_1a^2 > k_2bc$ and x increases until $dx/dt = 0$, i.e. until

$$k_1(a - x)^2 = k_2(b + \tfrac{1}{2}x)(c + \tfrac{1}{2}x),$$

which must occur for $x < a$. In *Figure 15.01* the graph of $f(x)$ is

plotted against x, when $f(x)$ is either positive or negative initially. If $4k_1 > k_2$, we find that the curves are concave upwards. If $4k_1 < k_2$ the curves are convex upwards (full lines in figure). If $4k_1 = k_2$ they are straight lines, and the equation is

$$dx/dt = k_1(a^2 - 4bc) - 2k_1(a + b + c)x,$$

which resembles that of a first-order reaction.

15.09. Some General Remarks. A reaction, such that $dx/dt = f(x)$, has an initial velocity $f(0) = \alpha$ and proceeds until $x = \beta$, the first positive root of $f(x) = 0$ (if $\alpha > 0$). Except in the initial stages the increase in x depends on the shape of the graph of $f(x)$ between $x = 0$ and $x = \beta$. If it is approximately a straight line, x increases as in Section 15.06, if approximately a quadratic as in Section 15.07, and so on. Since $f(x)$ is sensitive to the values of the initial concentrations $a, b, \ldots$ and can behave in various ways depending on their relative magnitudes, it is desirable to use this fact, when investigating a reaction.

One way of doing this is to make all the initial concentrations equal: for example, for an irreversible reaction between n different molecules of equal initial concentration,

$$dx/dt = k(a - x)^n,$$

an equation which was considered in Section 10.22.

15.10. Linear Equations. An equation is linear in the dependent variable when the variable and its derivatives are never multiplied in the same term and do not occur above the first power. Hence a first-order equation for y is at the worst of the form

$$\alpha\, dy/dx + \beta y = \gamma,$$

where α, β, γ are constants, or functions of x. (The terms containing dy/dx are collected together, those containing y, and finally those containing neither.)

(*i*) The equation is separable when β and γ are constants.

Example 1. $x\, dy/dx + y = 1$, i.e. $dy/(1 - y) = dx/x$ or $x(1 - y) = C$.

(*ii*) The equation is separable when $\gamma = 0$.

Example 2. $e^x\, dy/dx + y \cos x = 0$, or $dy/y = -e^{-x} \cos x \,.\, dx$.

$$\therefore \qquad \log y = \tfrac{1}{2}e^{-x}(\cos x - \sin x) + C.$$

(*iii*) It can be integrated at once when $\beta = d\alpha/dx$; the equation then has the form

$$\alpha \frac{dy}{dx} + \frac{d\alpha}{dx} y = \frac{d}{dx}(\alpha y) = \gamma, \qquad \text{i.e. } \alpha y = \int \gamma \, dx + C.$$

Example 3. $\cos x(dy/dx) - \sin x \,.\, y = x$, is $d(y \cos x)/dx = x$.
Hence $y \,.\, \cos x = \frac{1}{2}x^2 + C$.

(*iv*) In general, first divide by α so that the equation is

$$dy/dx + Py = Q,$$

where P and Q are functions of x. Then multiply by $e^{\int P dx}$,

$$e^{\int P\,dx} \,.\, \frac{dy}{dx} + Pe^{\int P\,dx} y = Qe^{\int P\,dx}.$$

The LHS is the derivative of a product,

$$\frac{d}{dx}\{ye^{\int P\,dx}\} = Qe^{\int P\,dx},$$

i.e.
$$ye^{\int P\,dx} = C + \int Qe^{\int P\,dx}\,dx.$$

15.11. SPECIAL LINEAR EQUATION. $dy/dx + ay = Q$, where a is a constant. With $P = a$, the solution is

$$y \,.\, e^{ax} = C + \int Qe^{ax}\,dx,$$

i.e.
$$y = e^{-ax}\{C + \int Qe^{ax}\,dx\}.$$

Several examples are given below.

15.12. FURTHER REMARKS. This section is separated from the others because the following observations apply to all linear equations of any order.

The solution of a linear equation shows that it contains:

(*i*) one term for each term in the function Q, and
(*ii*) a term which appears even when $Q = 0$.

This can be proved directly. Let

$$dy/dx + Py = Q_1 + Q_2. \qquad \ldots\ldots(1)$$

Let y_1 be any particular solution of

$$dy_1/dx + Py_1 = Q_1, \qquad \ldots\ldots(2)$$

and y_2 be any particular solution of

$$dy_2/dx + Py_2 = Q_2, \qquad \ldots\ldots(3)$$

and finally y_0 be the general solution (with an arbitrary constant) of

$$dy_0/dx + Py_0 = 0. \quad \ldots.(4)$$

Then, by adding equations (2), (3) and (4) it is clear that

$$y = y_0 + y_1 + y_2$$

is the general solution of equation (1)—general because it contains one arbitrary constant.

It is therefore possible to solve equation (1) without using the general method, by solving the homogeneous equation (4) with $Q = 0$, and then adding any solutions of equations (2) and (3). The latter can often be guessed, for simple functions Q, by an elementary knowledge of differentiation. If they cannot be guessed, the general method can always be used.

15.13. Successive Reactions. A simple solution can be obtained for a series of first-order irreversible reactions,

$$\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C} \rightarrow \ldots,$$

such as occurs in a radioactive decay series.

The concentration of A is calculated as in Section 15.06,

$$F(t) = a(1 - e^{-\lambda t}),$$

where λ is the velocity constant, and a its initial concentration.

The concentration of B at time t is $\{b + F(t) - x\}$ where (*i*) b is the initial amount, (*ii*) $F(t)$ is the amount gained by the first reaction, and (*iii*) x is transformed by the second reaction. The velocity of the second reaction is

$$dx/dt = k(b + F(t) - x),$$

i.e.
$$dx/dt + kx = k(a + b) - kae^{-\lambda t}.$$

To solve this equation, first use the method of the last section. We add

(*i*) the general solution of $(dx/dt) + kx = 0$, i.e. $x = Ae^{-kt}$;

(*ii*) a particular solution of $dx/dt + kx = k(a + b)$, i.e. $x = a + b$;

(*iii*) a particular solution of $dx/dt + kx = -kae^{-\lambda t}$.

One might expect the last equation to have a solution of the form $x = Be^{-\lambda t}$, unless $k = \lambda$ (why not?), provided B is suitably chosen, so we substitute this in the equation and find

$$dx/dt + kx = B(k - \lambda)e^{-\lambda t} = -kae^{-\lambda t}.$$

This is true if $B = ka/(\lambda - k)$, $(\lambda \neq k)$.

In all, therefore, the general solution ($k \neq \lambda$) is

$$x = Ce^{-kt} + (a + b) + kae^{-\lambda t}/(\lambda - k).$$

To make $x = 0$ when $t = 0$ requires $C = -(a + b) - ka/(\lambda - k)$.

$$x = (a + b)(1 - e^{-kt}) - ka(e^{-kt} - e^{-\lambda t})/(\lambda - k), \quad (k \neq \lambda).$$

As one might expect, since it arises from a negative term $-kae^{-\lambda t}$ in the differential equation, the last term is negative whether k is greater or less than λ.

This guesswork method breaks down when $k = \lambda$, but the general method gives

$$x = e^{-kt}[C + \int\{k(a + b) - kae^{-\lambda t}\}e^{kt}\,dt].$$

When k and λ are not equal this gives the earlier result. When $k = \lambda$, then $e^{-\lambda t}e^{kt} = 1$ and

$$\begin{aligned} x &= e^{-kt}\{C + (a + b)e^{kt} - kat\}, \\ &= (a + b)(1 - e^{-kt}) - kate^{-kt}, \quad (x = 0, t = 0). \end{aligned}$$

The reason why the trial method failed is now obvious: we looked for the wrong type of particular solution.

If there are further reactions, their products are calculated in the same way.

When the first reaction is second-order, the expression for $F(t)$ is so complicated that the next step needs numerical evaluation. If the second reaction is second-order it has to be integrated numerically. When the reactions are reversible the equations are usually extremely complicated except when they are first-order both ways, as sometimes happens when radiation is involved.

15.14. SIMULTANEOUS EQUATIONS. In a cyclic reaction it is possible to obtain several equations which cannot be solved successively, but simultaneously. This also occurs for two reversible reactions such as

$$\underset{x}{\mathrm{A}} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \underset{y}{\mathrm{B}} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \underset{z}{\mathrm{C}},$$

where the concentrations satisfy such equations as

$$\begin{aligned} dx/dt &= k_2y - k_1x, \\ dy/dt &= k_4z - k_3y, \\ x + y + z &= \text{constant}. \end{aligned}$$

The procedure is to eliminate all the variables except one and solve the equation for that one. The other variables can then be determined.

15.15. STEADY HEAT FLOW. If the two sides of a flat plate of thickness d are maintained at temperatures θ_1 and θ_2 (less than θ_1), there is a steady flow of heat from the hot to the cold plate of amount $Q = \kappa(\theta_1 - \theta_2)/d$ per unit area, where κ is the thermal conductivity. Fourier's law asserts that this can be generalized, so that temperature variations cause a heat flow in any direction which is κ times the rate of decrease of temperature in that direction, per unit area normal to it. When the temperatures are not changing with time the heat flow is steady, and there is the further condition that there must be no accumulation of heat with time at any point.

Consider a uniform circular tube of inner radius a and outer radius b, containing fluid at a constant temperature θ_1, placed in surroundings at a temperature θ_0. The temperature θ in the body of the tube only varies with the distance r from the axis of the tube and heat flows either radially from, or towards, this axis Per unit length, the heat flow Q outwards across a cylindrical surface of radius r about the axis is

$$\begin{aligned} Q &= \kappa(\text{area})(\text{radial temperature decrease}), \\ &= -\kappa \cdot 2\pi r \cdot (d\theta/dr). \end{aligned}$$

As there is no accumulation of heat, this is the same for all radii and equals the rate of loss of heat from the fluid inside the tube. Separating the variables

$$-2\pi\kappa\, d\theta = Q\, dr/r.$$

Integrating, so that $\theta = \theta_1$ when $r = a$,

$$2\pi\kappa(\theta_1 - \theta) = Q \ln (r/a).$$

If heat is lost by convection to the surroundings, then $\theta = \theta_0$ at $r = b$, i.e. Q is given by the equation

$$2\pi\kappa(\theta_1 - \theta_0) = Q \ln (b/a).$$

If heat is lost by radiation, Newton's law of cooling states that Q is proportional to the temperature difference between the outer surface and the surroundings. Hence, if $\theta = \theta_2$ at $r = b$,

$$Q = h(\theta_2 - \theta_0),$$

where h is a constant. From the earlier equation,

$$2\pi\kappa(\theta_1 - \theta_2) = Q \ln (b/a).$$

Eliminating θ_2, the heat loss is now given by

$$2\pi\kappa\{\theta_1 - \theta_0 - Q/h\} = Q \ln (b/a).$$

15.16. DIFFUSION EQUATION. Diffusion is a process which tends to remove concentration differences. According to Fick's law a variation of concentration c in the x direction causes a flow of amount $-D(dc/dx)$ per unit area in the x direction. The sign gives the direction of flow as in the heat problem.

This law is identical with that of heat flow. Consequently problems where the concentration does not vary with time are also the same if the diffusion constant D is substituted for the heat conductivity κ.

15.17. COOLING SPHERE. By Newton's law of cooling, a sphere with a surface temperature θ in surroundings at temperature θ_0 loses in time dt, by radiation, an amount of heat

$$dQ = h(\theta - \theta_0)dt,$$

where h is a constant. If the sphere has a mass m and specific heat c and cools so slowly that the temperature is uniform throughout, this causes the temperature to change from θ to $(\theta + d\theta)$, where

$$dQ = -mc\, d\theta.$$

Equating these two expressions for dQ,

$$d\theta/dt = -k(\theta - \theta_0), \text{ where } k = h/mc.$$

In isothermal surroundings, θ_0 = constant, and the sphere ultimately attains that temperature. To emphasize this we write the equation

$$d(\theta - \theta_0)/dt = -k(\theta - \theta_0), \quad \text{i.e. } \theta - \theta_0 = Ae^{-kt},$$

as in Section 15.06. The constant A is the initial temperature difference. If the sphere has a temperature θ_1 initially, $A = \theta_1 - \theta_0$ and

$$\theta - \theta_0 = (\theta_1 - \theta_0)e^{-kt}.$$

The temperature difference decreases exponentially to zero with time.

If the temperature of the surroundings varies with time, then, from Section 15.11 with $a = k$ and $Q = k\theta_0$, the solution is

$$\theta = e^{-kt}\{\theta_1 + \int_0^t k\theta_0 e^{kt}\, dt\},$$

where θ_0 is a function of t.

In certain cases, however, we can proceed in a simpler manner. Assume that $\theta_0 = a + b \cos pt$, and therefore oscillates. Then

$$d\theta/dt + k\theta = k\theta_0 = ka + kb \cos pt.$$

The general solution is (Section 15.12)

$$\theta = Ae^{-kt} + a + f(t).$$

The origin of the first two terms should now be obvious. The last term $f(t)$ is any solution of

$$d\theta/dt + k\theta = kb \cos pt.$$

An oscillating solution seems likely: let

$$\theta = f(t) = C \cos (pt - \alpha).$$

(Both C and α are necessary: one could try $A \cos pt + B \sin pt$.) Then

$$\begin{aligned} d\theta/dt + k\theta &= C\{-p \sin (pt - \alpha) + k \cos (pt - \alpha)\}, \\ &= C(p^2 + k^2)^{1/2} \cos (pt - \alpha + \varepsilon), \end{aligned}$$

where $(k^2 + p^2)^{1/2} \cos \varepsilon = k$ and $(k^2 + p^2)^{1/2} \sin \varepsilon = p$ {Section 7.07 (*iv*)}.

This expression equals $kb \cos pt$, and the trial is a success, if

$$C(p^2 + k^2)^{1/2} = kb, \quad \text{i.e. } C = b \cos \varepsilon \quad \text{and} \quad \alpha = \varepsilon.$$

The solution is, altogether,

$$\theta = Ae^{-kt} + a + b \cos \varepsilon \cos (pt - \varepsilon).$$

Choosing A so that $\theta = \theta_1$ when $t = 0$,

$$\theta = (\theta_1 - a - b \cos^2 \varepsilon)e^{-kt} + a + b \cos \varepsilon \cos (pt - \varepsilon).$$

Thus θ does oscillate with θ_0 but it lags behind by a time (ε/p) and its amplitude of oscillation is $b \cos \varepsilon$ rather than b. This amplitude is approximately the same when $p \ll k$ and ε is a small angle, but it is much smaller when $p \gg k$ and $\cos \varepsilon$ is small.

15.18. Second-order Differential Equations. By successive differentiation of

$$y = e^{-kx}(A + Bx)$$

expressions for the derivatives are obtained containing A and B. Any relation between these, which is independent of A and B, is the

expression of a common property of all the functions y that can be obtained by giving special values to these two arbitrary constants. To obtain such a relation at least two derivatives are required, e.g.

$$dy/dx = -ke^{-kx}(A + Bx) + Be^{-kx} = e^{-kx}\{(B - Ak) - kBx\};$$
$$d^2y/dx^2 = -ke^{-kx}\{-kA - kBx + B\} + e^{-kx}(-kB)$$
$$= e^{-kx}(k^2A - 2kB + k^2Bx).$$

Eliminating A and B between the three equations,

$$d^2y/dx^2 + 2k\,dy/dx + k^2y = 0.$$

From examples of this type it is deduced that a function with two arbitrary constants cannot satisfy a differential equation of less than the second order, and conversely, that an equation of the second order has a general solution with two arbitrary constants.

Consider the linear second-order equation (Sections 15.10 and 15.12)

$$a\,d^2y/dx^2 + b\,dy/dx + cy = F(x),$$

where a, b, c are at most functions of the independent variable x.

As explained previously, the general solution is the sum of a Particular Integral (PI, any solution of the equation) and a Complementary Function (CF) which is the general solution when $F = 0$,

$$a\,d^2y/dx^2 + b\,dy/dx + cy = 0.$$

The CF is divisible into parts. If y_1 and y_2 are any two solutions of this equation which are independent (one is not a multiple of the other), then it is easily proved that

$$y = Ay_1 + By_2$$

is a solution. This is true only because the equation is linear and its RHS is *zero*. As this expression contains two arbitrary constants it is the general solution of its equation and the desired CF.

The solution of a linear second-order differential equation is therefore resolved into the discovery of three functions, or only two when $F = 0$.

15.19. Simple Harmonic Equation. A mass m is suspended by a spring obeying Hooke's law, i.e. if its length is greater than the natural length there is a tension $= k$(extension), where k is the force constant, or stiffness. In equilibrium the weight of the mass just balances the tension, and the extension is d, where $mg = kd$ (in absolute units).

If the mass is moving and is a distance y below the equilibrium position, the extra tension ky gives it an acceleration upwards, i.e.

$$d^2y/dt^2 = -ky/m = -\omega^2 y, \text{ where } \omega^2 = k/m.$$

Any equation of this form is a simple harmonic equation, and the variation of y with t is called a simple harmonic oscillation even when y and t have different meanings.

It is satisfied by $y = e^{i\omega t} = \cos \omega t + i \sin \omega t$, so that $\cos \omega t$ and $\sin \omega t$ are each a solution. According to the last section and Examples XIV, No. 9, the general solution is

$$y = A \cos \omega t + B \sin \omega t.$$

If $A = C \cos \phi$, $B = C \sin \phi$, the solution becomes

$$y = C \cos (\omega t - \phi),$$

where C and ϕ are new arbitrary constants. This function is periodic: the periodic time $2\pi/\omega$ is the increase in t, to increase the phase angle $(\omega t - \phi)$ by 2π. The frequency is $\nu = \omega/2\pi$, C is the amplitude and ϕ the phase. If the mass starts from rest $\phi = 0$, whereas if it starts from the origin $\phi = \pm 90°$.

15.20. DIATOMIC MOLECULE. In *Figure 15.02* there is plotted the potential energy $V(x)$ of a diatomic molecule as a function of the

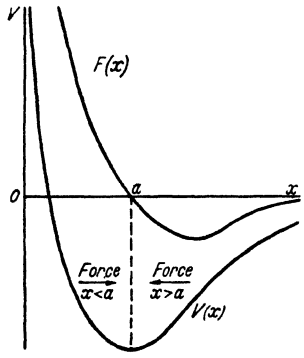

Figure 15.02

internuclear distance x. The corresponding force $F = -(dV/dx)$ is also plotted.

If the atoms have masses m_1 and m_2 and move along the line of centres, without rotation, the equations of motion are

$$m_1 d^2x_1/dt^2 = F(x), \quad m_2 d^2x_2/dt^2 = -F(x),$$

where x_1 and x_2 are the coordinates (*Figure 15.02*). Since $x = x_2 - x_1$, we find

$$\mu\, d^2x/dt^2 = F(x),$$

where $\mu = m_1m_2/(m_1 + m_2)$ is the reduced mass. $F(x)$ is positive for a repulsive force and negative for an attractive force.

The equilibrium distance $x = a$ is given by $d^2x/dt^2 = 0$, i.e. $F(a) = 0$; and when x does not differ much from a, by Taylor's theorem,

$$F(x) = F(a) + (x - a)(dF/dx)_{x=a} = (x - a)(dF/dx)_a.$$

If the atoms form a stable molecule, this force is attractive when $x > a$ and repulsive when $x < a$. Hence $(dF/dx)_a = F'(a)$ is negative.

Let us consider the motion when $x - a$ is small. Let $y = x - a$ and $(dF/dx)_a = -k$. Then

$$\mu\, d^2x/dt^2 = (x - a)(dF/dx)_a \text{ becomes } \mu\, d^2y/dt^2 = -ky,$$

which represents simple harmonic vibrations, $\ddot{y} = -\omega^2 y$, with $\omega^2 = k/\mu$ and a frequency

$$\nu = (1/2\pi)\sqrt{(k/\mu)}.$$

When the molecule is rotating with an angular momentum H, there are modifications to the value of the equilibrium distance a and the frequency. The equation of motion along the line of centres is now

$$\mu\, d^2x/dt^2 = G(x), \text{ where } G(x) = F(x) + H^2/\mu x^3,$$

so that $F(a) + H^2/\mu a^3 = 0$ and $k = -F'(a) + (3H^2/\mu a^4)$.

In general the small oscillations of any mechanical system, without friction, about a position of relative equilibrium which is stable are always simple harmonic (this is not generally true for large amplitudes).

15.21. EXPONENTIAL SOLUTIONS. The result of the last section depends completely on the fact that the equilibrium is stable so that dF/dx was negative at the equilibrium position. If it had been positive the equation for y would have been

$$d^2y/dt^2 = \omega^2 y,$$

which has exponential solutions $e^{-\omega t}$ and $e^{\omega t}$, and therefore a general solution

$$y = Ae^{\omega t} + Be^{-\omega t}.$$

This does not oscillate with time. If the system starts from the equilibrium position, A cannot be zero (why?) and y increases rapidly with t until the approximations are no longer valid.

15.22. Damped Harmonic Motion. Frictional forces introduce an extra term into the oscillation equation proportional to the velocity (at low speeds) and tending to prevent the motion. The result is an equation of the form

$$d^2y/dt^2 + 2\lambda\, dy/dt + \omega_0^2 y = 0.$$

This equation has been solved when $\lambda = 0$ (Section 15.19), $\lambda = \omega_0$ (Section 15.18), and can easily be solved when $\omega_0 = 0$. From these three solutions and the fact that λ and ω are constants, the mathematical approach is to find solutions of the form $y = e^{mt}$. After substitution, we find that m must satisfy the equation,

$$m^2 + 2\lambda m + \omega = 0, \text{ i.e. } m = -\lambda \pm \sqrt{(\lambda^2 - \omega_0^2)}.$$

A more experimental approach is to argue that a small value of λ is not likely to change appreciably the oscillatory character of the motion, but is likely to reduce the amplitude. We try

$$y = z \,.\, e^{-\mu t}$$

choosing μ so that z represents an oscillation when λ is small. Then

$$\omega_0^2 y + 2\lambda \frac{dy}{dt} + \frac{d^2y}{dt^2} = \omega_0^2 z e^{-\mu t} + 2\lambda \left\{ \frac{dz}{dt} e^{-\mu t} - \mu z e^{-\mu t} \right\} + \left\{ \frac{d^2z}{dt^2} e^{-\mu t} - 2\mu \frac{dz}{dt} e^{-\mu t} + \mu^2 z e^{-\mu t} \right\} = 0$$

i.e. $$e^{-\mu t} \left\{ \frac{d^2z}{dt^2} + 2(\lambda - \mu) \frac{dz}{dt} + (\omega_0^2 - 2\mu\lambda + \mu^2) z \right\} = 0.$$

Since $e^{-\mu t}$ is not zero, the required equation for z is obtained if

$$\mu = \lambda, \text{ i.e. } \omega_0^2 - 2\lambda\mu + \lambda^2 = \omega^2 - \lambda^2.$$

The equation and the solution become

$$d^2z/dt^2 = -\omega^2 z, \qquad z = C \cos(\omega t - \phi),$$

where $$\omega^2 = \omega_0^2 - \lambda^2 > 0.$$

Hence $$y = Ce^{-\lambda t} \cos(\omega t - \phi).$$

The effect of the damping factor λ, when $\lambda < \omega_0$, is to introduce a decaying amplitude $Ce^{-\lambda t}$ and reduce the frequency $(\omega_0/2\pi)$ to $\omega/2\pi = (\omega_0^2 - \lambda^2)^{1/2}/2\pi$.

Using *Example 3*, Section 4.19, and $\lambda^2 + \omega^2 = \omega_0^2$,

$$dy/dt = -e^{-\lambda t}C\omega_0 \sin(\omega t - \phi + \varepsilon), \text{ where } \tan\varepsilon = \lambda/\omega,$$

so that stationary values of y (maxima and minima) occur when

$$\omega t - \phi + \varepsilon = n\pi,$$

i.e. at intervals of π/ω. At the stationary values,

$$\begin{aligned} t_n &= (1/\omega)(n\pi + \phi - \varepsilon), \\ y_n &= e^{-\lambda t_n}C\cos(\omega t_n - \phi) = e^{-\lambda t_n}C\cos(n\pi - \varepsilon), \\ &= (-)^n e^{-\lambda t_n}C\cos\varepsilon. \end{aligned}$$

The logarithmic decrement is the logarithm of the ratio of two successive stationary values,

$$-\frac{y_{n+1}}{y_n} = e^{-\lambda(t_{n+1}-t_n)} = e^{-\pi\lambda/\omega},$$

$$\log \text{dec} = \log|y_n/y_{n+1}| = \pi\lambda/\omega.$$

The decay time is the time in which the amplitude $Ce^{-\lambda t}$ decreases by a factor $(1/e)$, namely,

$$\tau = 1/\lambda.$$

15.23. HEAVILY DAMPED MOTION. When $\lambda = \omega_0$, the equation for z is $d^2z/dt^2 = 0$, giving

$$y = e^{-\lambda t}z = e^{-\omega_0 t}(At + B), \quad (\lambda = \omega_0),$$

as given in Section 15.18. This is critical damping.

When $\lambda > \omega_0$, the equation for z is

$$d^2z/dt^2 = (\lambda^2 - \omega_0^2)z = q^2z, \quad q = (\lambda^2 - \omega_0^2)^{1/2};$$

whence $z = Ae^{-qt} + Be^{qt}$ and $y = Ae^{-(\lambda+q)t} + Be^{-(\lambda-q)t}$.

Since $(\lambda - q)(\lambda + q) = \omega_0^2$, then $\lambda - q < \omega_0$, so that the second exponential decreases more slowly than the exponential $e^{-\omega_0 t}$ for critical damping. In other words, the value of λ which gives the least value of y after a very long time is that for critical damping.

15.24. FORCED OSCILLATIONS. Mechanical systems are often set into oscillatory motion by outside forces, as when an electron bound in an atom is acted upon by the electric field of a light wave. When the free system performs damped (or undamped) oscillations the resulting motion is called a forced oscillation.

Consider a system with an external force $mF(t)$ so that its equation of motion is

$$d^2y/dt^2 + 2\lambda\, dy/dt + \omega_0^2 y = F(t).$$

Let us assume that this external force is periodic,

$$F = b \sin pt.$$

The general solution, which gives all possible motions, is a sum of the complementary function and a particular integral. The former is now known, so we concentrate on the latter. We postulate a solution

$$y = C \sin (pt - \alpha).$$

$$\therefore \quad \omega_0^2 y + 2\lambda \frac{dy}{dt} + \frac{d^2y}{dt^2}$$

$$= C\{(w_0^2 - p^2) \sin (pt - \alpha) + 2\lambda p \cos (pt - \alpha)\},$$

$$= C\{(\omega_0^2 - p^2)^2 + 4\lambda^2 p^2\} \sin (pt - \alpha + \varepsilon),$$

where $$\{(\omega_0^2 - p^2)^2 + 4\lambda^2 p^2\}^{1/2} \begin{matrix}\cos \varepsilon \\ \sin \varepsilon\end{matrix} = \begin{matrix}(\omega_0^2 - p^2) \\ 2\lambda p\end{matrix}.$$

If this equals $F = b \sin pt$, then $\alpha = \varepsilon$ and

$$y = C \sin (pt - \varepsilon); \; C = b/\{(\omega_0^2 - p^2)^2 + 4\lambda^2 p^2\}^{1/2}.$$

In *Figure 15.03* we have plotted values of ε and $(C\omega_0^2/b)$ against p/ω_0 for two values of λ. The phase ε increases steadily with p from 0 to π, but the increase is greatest near $p = \omega_0$ when λ is small.

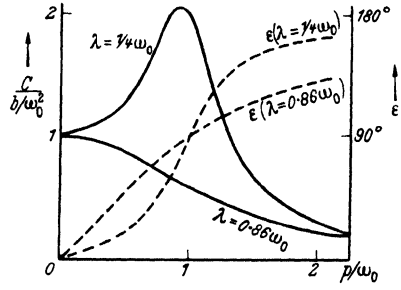

Figure 15.03

The C/p curve has a maximum when λ is not too large and is called a resonance curve. The position of the maximum gives the free oscillation period approximately, and the height and width of

the peak in the curve are a measure of λ. In fact the amplitude C has the value (b/ω_0^2) when $p = 0$ and tends to zero as $p \to \infty$. From its derivative with respect to p we find that it has a maximum value

$$C_{max} = b/2\lambda\omega, \quad \text{when } p^2 = \omega^0 - 2\lambda^2, \quad (\omega^2 = \omega_0^2 - \lambda^2).$$

The maximum only occurs if this value of p^2 is positive, i.e.

$$\lambda^2 < \tfrac{1}{2}\omega_0^2.$$

15.25. CHAIN THEORY OF REACTIONS. A chemical reaction occurs because there are several 'active' molecules with sufficient kinetic energy to react. In each molecular reaction this kinetic energy first becomes internal energy and is then reconverted into the kinetic energy of the products.

Ordinary kinetic theory, as used to obtain the equations of Section 11.05, assumes that this is shared out, or dissipated, but is not lost, and by ordinary thermal collision processes results in some 'passive' molecules being made active. If n is the number of active molecules at any instant and τ the mean lifetime before each reacts, the velocity of the reaction, or number of reactions per unit time, is n/τ. The variation of n with time is described by the equation

$$dn/dt = n_0 - n/\tau,$$

where, per unit time, n_0 is the number of active molecules created thermally and n/τ is the loss by reaction. If n_0 does not depend on the time, the solution is

$$n = \tau n_0 + (\mathcal{N} - \tau n_0)e^{-t/\tau},$$

where $n = \mathcal{N}$ at $t = 0$. Hence $n \to \tau n_0$, and the rate of the reaction is $(n/\tau) = n_0$.

In fact n_0 is not constant, but decreases as the reactants are used up, and may increase in an exothermal reaction when the mean kinetic energy of the products is greater than the average, so that the temperature rises. When τ is quite small, however, $n = \tau n_0$ approximately and this is used in the Section 15.05 ff.

The essential feature of this theory is that there is no spatial relation between one molecular reaction and another for which it is ultimately responsible.

In some reactions this type of theory is useless. Thus, in nuclear reactions, a large energy is needed to cause a reaction, and the reaction only goes at a reasonable rate without outside help if some of the products have even larger energy and cause another reaction before they dissipate their energy by collisions.

The chain theory of chemical reactions discusses fast reactions between these two extremes. In addition to the active molecules created thermally, it is assumed that there are chains of molecular reactions where the products of one reaction are the creators of active molecules, by handing their relatively large kinetic energy directly to passive molecules. These react, and their products continue the chain with a succession of reactions, spatially related.

At any instant, the end of a chain is an active molecule which may continue the chain (or not) and there are active molecules, thermally produced, which may start new chains. Thus the rate of change of the number of active molecules n—or chains and potential chains—is

$$dn/dt = n_0 - \alpha n + \kappa\beta n,$$

where $(\alpha + \beta)$ is the total probability (per unit time) that the active molecules react, αn is the number of molecules which have their energy dissipated thermally, βn the number which hand their energy to passive molecules which can react, and κ the average number of active molecules produced in each process of that type. The velocity remains $n/\tau = n(\alpha + \beta)$.

When α, β and κ are constants, the solution of this equation is

$$n = \frac{n_0}{\alpha - \kappa\beta} + \left(\mathcal{N} - \frac{n_0}{\alpha - \kappa\beta}\right) e^{-(\alpha - \kappa\beta)t},$$

where $n = \mathcal{N}$ at $t = 0$. If the thermal process is dominant $\alpha > \kappa\beta$, n tends to the value $n_0/(\alpha - \kappa\beta)$, and the velocity tends to

$$n_0(\alpha + \beta)/(\alpha - \kappa\beta),$$

which may be considerably greater than the 'thermal' rate n_0, obtained when $\beta = 0$.

When $\kappa\beta > \alpha$, however, n increases exponentially with time. The solution can better be written with $\gamma = \kappa\beta - \alpha$, i.e.

$$n = (\mathcal{N} + n_0/\gamma)e^{\gamma t} - n_0/\gamma.$$

The equation has often to be modified, e.g. so that α and $\kappa\beta$ vary with n. We shall not discuss these in detail as they are all first-order separable equations from the mathematical point of view. For example, if $\alpha = \alpha_0 n$ and $\beta = 0$, the equation becomes

$$dn/dt = n_0 - \alpha_0 n^2,$$

so that n tends to the value $n_1 = (n_0/\alpha_0)^{1/2}$. The solution is

$$\alpha_0 t = \int dn/(n_1^2 - n^2) = A + \tanh^{-1}(n/n_1), \quad (I.21).$$

$$\therefore\ n/n_1 = \tanh(\alpha_0 t + A)$$
$$= (\tanh \alpha_0 t + \tanh A)/(1 + \tanh \alpha_0 t \tanh A).$$

If $n = \mathcal{N}$ when $t = 0$, then $\tanh A = (\mathcal{N}/n_1)$, whence

$$n/n_1 = (n_1 \tanh \alpha_0 t + \mathcal{N})/(n_1 + \mathcal{N} \tanh \alpha_0 t).$$

This equation has been derived assuming $n < n_1$, and therefore $\mathcal{N} < n_1$; but the final result is obviously true whatever the value of $\mathcal{N}$.

For a reaction between two plane walls, a term must be added to allow for those ends of chains which reach the walls and cannot continue by the direct process. This is in the nature of diffusion and the equation becomes

$$dn/dt = D\, d^2n/dx^2 + n_0 - (\alpha - \kappa\beta)n,$$

where $n = 0$ at the walls ($x = \pm \frac{1}{2}d$, say). This can be solved in the steady state $dn/dt = 0$ for n as a function of x, when α and $\kappa\beta$ are constants. Proceeding as in Sections 15.20–23, we find that the solutions such that $n = 0$ at the walls are

$$\alpha - \kappa\beta = Dp^2 > 0, \qquad n = \frac{n_0}{Dp^2}\left\{1 - \frac{\cosh \omega x}{\cosh \frac{1}{2}\omega d}\right\};$$

$$\alpha - \kappa\beta = -D\omega^2 < 0, \quad n = \frac{n_0}{D\omega^2}\left\{\frac{\cos \omega x}{\cos \frac{1}{2}\omega d} - 1\right\}.$$

15.26. REDUCIBLE EQUATIONS. The equation $d^2x/dt^2 = F(x)$ of Section 15.20 often occurs in mechanics and can be reduced to two first-order separable equations. This is achieved by introducing

$$v = dx/dt, \quad dv/dt = F(x),$$

i.e.
$$dt = dx/v = dv/F(x).$$

The second equality yields the equation $v\, dv = F(x)dx$, which can be integrated at once to give

$$\tfrac{1}{2}v^2 = \int F(x)dx + \text{constant}.$$

This is of the nature of an energy equation. The first equality $dt = dx/v$ can now be integrated to give t in terms of x.

15.27. BESSEL'S EQUATION. Although Bessel's equation,

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{n^2}{x^2}\right)y = 0,$$

cannot be solved in terms of elementary functions, in general, some

properties of the solution can be obtained without solving the equation.

The equation for $u = yx^{1/2}$ contains no term in du/dx (it is in *normal* form). Thus,

$$\frac{du}{dx} = \frac{dy}{dx}x^{1/2} + y \cdot \tfrac{1}{2}x^{-1/2},$$

$$\frac{d^2u}{dx^2} = \left(\frac{d^2y}{dx^2} \cdot x^{1/2} + \frac{dy}{dx} \cdot \tfrac{1}{2}x^{-1/2}\right) + \left(\frac{dy}{dx} \cdot \tfrac{1}{2}x^{-1/2} - y \cdot \tfrac{1}{4}x^{-3/2}\right),$$

$$= x^{1/2}\left\{\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \frac{y}{4x^2}\right\},$$

$$= x^{1/2}\left\{-\left(1 - \frac{n^2}{x^2}\right)y - \frac{y}{4x^2}\right\},$$

$$= \left(\frac{n^2 - \frac{1}{4}}{x^2} - 1\right)u.$$

If $n = \frac{1}{2}$, then $d^2u/dx^2 = -u$, i.e. $u = A \sin x + B \cos x$,

or $$y = (A \sin x + B \cos x)/\sqrt{x}.$$

This solution oscillates with a period 2π in x, and an amplitude decreasing slowly as x increases.

When n has some other value, similar deductions can be made. Using the method of Section 5.05, (d^2u/dx^2) has the same sign as u when $x^2 < (n^2 - \frac{1}{4})$, and the graph is convex to the x axis, starting from the origin. When $x^2 > (n^2 - \frac{1}{4})$, d^2u/dx^2 is opposite in sign to u, the graph is concave towards the x axis and oscillates about it. When x^2 is very much greater than $(n^2 - \frac{1}{4})$ the equation is approximately

$$d^2u/dx^2 = -u,$$

which is the same as for $n = \frac{1}{2}$ and has the same solution.

15.28. Solution by Power Series. A solution of the Bessel equation,

$$x^2\, d^2y/dx^2 + x\, dy/dx + (x^2 - n^2)y = 0,$$

can be obtained in the form of a power series, which we write as

$$y = a_c x^c + a_{c+1}x^{c+1} + a_{c+2}x^{c+2} + \ldots = \sum_{m=c}^{\infty} a_m x^m.$$

The first term $y = a_c x^c$ is the approximation to y near the origin, neglecting higher powers of x. The value of c can therefore be

obtained by the condition that x^c is a solution of the equation, when x is small:

$$\text{i.e. } x^2\,d^2y/dx^2 + x\,dy/dx + (x^2 - n^2)y$$
$$= c(c-1)a_c x^c + ca_c x^c + (x^2 - n^2)a_c x^c,$$
$$= (c^2 - n^2)a_c x^c, \text{ neglecting the term in } x^{c+2}.$$

This is zero, and $y = x^c$ satisfies the equation, when $c = \pm n$. The negative power is unexpected, but the two values give two distinct solutions, unless n is an integer.

The complete power series $y = \sum_{m=c}^{\infty} a_m x^m$ is a solution when

$$x^2\,d^2y/dx^2 + x\,dy/dx + (x^2 - n^2)y$$
$$\equiv \sum_{m=c}^{\infty} [m(m-1)a_m x^m + ma_m x^m + (x^2 - n^2)a_m x^m],$$
$$= \sum_{m=c}^{\infty} [(m^2 - n^2)a_m x^m + a_m x^{m+2}] = 0.$$

If this is zero for every value of x, the coefficient of every power of x must vanish. Examination of the complete coefficients of the higher powers x^m shows that they vanish, if we choose the coefficients so that

$$(m^2 - n^2)a_m + a_{m-2} = 0, \quad m = (c+2), (c+3), \ldots$$

The only terms left are those in x^c and x^{c+1}, namely

$$(c^2 - n^2)a_c x^c + \{(c+1)^2 - n^2\}a_{c+1}x^{c+1}.$$

The first vanishes because we choose $c = \pm n$; but then the second only vanishes when

$$a_{c+1} = 0,$$

except in the special case $c = -\frac{1}{2}$, $n = +\frac{1}{2}$, which we have solved already. This gives two series, for $c = +n$ and $c = -n$.

The recurrence relation $a_m(m^2 - n^2) + a_{m-2} = 0$ relates alternate coefficients in the series as multiples of one another. They form two groups, one determined by the first coefficient a_c and the other by the second coefficient a_{c+1}. As $a_{c+1} = 0$ all the second group is zero, and only the other group of powers remains.

$$a_m = -\frac{a_{m-2}}{(m-n)(m+n)}, \quad m = c+2, c+4, c+6, \ldots$$

It is now necessary to use the values of c.

(i) $c = +n$: let $a_c = a_n = A$.

Then $a_{n+2} = -\frac{a_n}{2(2n+2)} = -\frac{A}{2^2(n+1)}$;

$$a_{n+4} = -\frac{a_{n+2}}{4(2n+4)} = \frac{A}{2^4 \cdot 2(n+1)(n+2)}.$$

$$a_{n+6} = -\frac{a_{n+4}}{6(2n+6)} = -\frac{A}{2^6 \cdot 3!(n+1)(n+2)(n+3)}.$$

$$\therefore \quad y = Ax^n \left\{1 - \frac{x^2}{2^2(n+1)} + \frac{x^4}{2^4 \cdot 2!(n+1)(n+2)} - \frac{x^6}{2^6 \cdot 3!(n+1)(n+2)(n+3)} + \ldots\right\}.$$

The particular solution with $A = 1/2^n\Gamma(n+1)$ is the Bessel function of the first kind.

$$J_n(x) = \frac{x^n}{2^n\Gamma(n+1)}\left\{1 - \frac{x^2}{2^2(n+1)} + \ldots\right\}$$

$$= \frac{x^n}{2^n\Gamma(n+1)} - \frac{x^{n+2}}{2^{n+2}\Gamma(n+2)} + \frac{x^{n+4}}{2^{n+2}\Gamma(3)\Gamma(n+3)} - \ldots$$

$$J_0(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2 4^2} - \frac{x^6}{2^2 4^2 6^2} + \ldots$$

The series are convergent for all values of x.

(*ii*) $c = -n$: let $a_c = a_{-n} = B$. Then,

$$a_{-n+2} = -\frac{a_{-n}}{(-2n+2) \cdot 2} = -\frac{B}{2^2(1-n)},$$

$$a_{-n+4} = -\frac{a_{-n+2}}{(-2n+4) \cdot 4} = \frac{B}{2^4 \cdot 2!(1-n)(2-n)}, \text{ etc.}$$

$$y = Bx^{-n}\left\{1 - \frac{x^2}{2^2(1-n)} + \frac{x^4}{2^4 \cdot 2!(1-n)(2-n)} + \ldots\right\}.$$

It is obtained from the previous solution by substituting $-n$ for $+n$. In particular, when $B = \frac{1}{2^{-n}\Gamma(1-n)}$, the solution is

$$J_{-n}(x) = \frac{x^{-n}}{2^{-n}\Gamma(1-n)} - \frac{x^{-2n+2}}{2^{-n+2}\Gamma(2-n)} + \ldots$$

Given these two solutions, the general solution of Bessel's equation is

$$y = aJ_n(x) + bJ_{-n}(x).$$

Instead of $J_{-n}(x)$, a second solution is often used which is

$$Y_n(x) = \frac{1}{\sin n\pi}\{\cos n\pi\, J_n(x) - J_{-n}(x)\}.$$

When n is an integer, the second power series $J_{-n}(x)$ is useless because the coefficients in the series $\Sigma a_m x^m$ become infinite when $m = n$. A second solution can then be obtained from Y,

$$Y_n(x) = \lim_{r\to n} Y_r(x), \text{ when } n \text{ is an integer.}$$

This is important because Bessel functions of integral order n are very common.

The solution $J_n(x)$ is finite or zero at the origin and the other solution Y_n or J_{-n} diverges there. In problems where the solution is finite at $r = 0$ only the first solution $J_n(x)$ is required.

Other solutions are the Hankel functions:

$$H_n^{(1)}(x) = J_n(x) + iY_n(x),$$
$$H_n^{(2)}(x) = J_n(x) - iY_n(x),$$

which reduce to the complex exponentials, $e^{\pm ix}/x^{1/2}$, when $n = \frac{1}{2}$.

15.29. EQUATIONS REDUCIBLE TO BESSEL'S EQUATION. The equation,

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(k^2 - \frac{n^2}{x^2}\right)y = 0,$$

becomes Bessel's equation by changing the independent variable to $x' = kx$. The solution is

$$y = aJ_n(kx) + bY_n(kx).$$

The following equation appears in spherical problems,

$$\frac{d^2y}{dx^2} + \frac{2}{x}\frac{dy}{dx} + \left(k^2 - \frac{n(n+1)}{x^2}\right)y = 0, \quad (n = 0, 1, 2, 3, \ldots).$$

By the substitution $u = yx^{1/2}$ it becomes

$$\frac{d^2u}{dx^2} + \frac{1}{x}\frac{du}{dx} + \left(k^2 - \frac{(n+\frac{1}{2})^2}{x^2}\right)u = 0,$$

with the solution

$$y = \frac{u}{x^{1/2}} = a\frac{J_{n+\frac{1}{2}}(kx)}{x^{1/2}} + b\frac{Y_{n+\frac{1}{2}}(kx)}{x^{1/2}}.$$

For large values of x,

$$\frac{1}{\sqrt{x}} J_{n+\frac{1}{2}}(kx) = \left(\frac{2}{\pi k}\right)^{1/2} \frac{\cos\,(kx - \frac{1}{2}(n+1)\pi)}{x},$$

$$\frac{1}{\sqrt{x}} Y_{n+\frac{1}{2}}(kx) = \left(\frac{2}{\pi k}\right)^{1/2} \frac{\sin\,(kx - \frac{1}{2}(n+1)\pi)}{x}.$$

These solutions can be expressed in terms of sines and cosines, as in Section 13.18.

Another equation, which appears at regular intervals, is obtained by replacing k^2 by $-k^2$, i.e. k by ik,

$$\frac{d^2y}{dx^2} + \frac{2}{x}\frac{dy}{dx} - \left(k^2 + \frac{n(n+1)}{x^2}\right) y = 0.$$

The solution is

$$y = \frac{1}{\sqrt{x}} \{aJ_{n+\frac{1}{2}}(ikx) + bY_{n+\frac{1}{2}}(ikx)\},$$

which can be expressed in terms of hyperbolic sines and cosines.

EXAMPLES XV

1. Assuming that the vapour has a much lower density than the liquid and behaves as a perfect gas, the Clausius–Clapeyron equation for the vapour pressure of a liquid as a function of temperature reduces to $dp/dT = qp/RT^2$, where q is the latent heat of vaporization, which is assumed to be constant. Find p as a function of T.

2. If the problem of Section 15.15 is modified by the liberation of heat between the walls of the tube, due to an electric current, at a steady rate q per unit volume, show that the heat flow per unit length at radius r is $Q + \pi(r^2 - a^2)q$, where Q is the loss from the inside. Deduce the value of Q when the inside and outside temperatures are θ_1 and θ_2.

3. In a 'Maxwell' solid, the relation between stress S and strain σ is $dS/dt = k\,d\sigma/dt - S/\tau$, where k and τ are constants. If σ is a given function of the time, find S as a function of time. Obtain solutions when (*i*) $\sigma =$ constant, (*ii*) σ increases linearly with time.

4. In a simple theory of dispersion of light by atoms, the following equation is used for the motion of an electron oscillating due to the electric field of the radiation,

$$m(d^2x/dt^2) = -kx + eE\cos\omega t.$$

Find the amplitude of the forced oscillation of frequency $\omega/2\pi$.

5. Find the values of ω so that the solutions of $d^2y/d\theta^2 = -\omega^2 y$ have the same values at $\theta = 2\pi$ as at $\theta = 0$.

6. A function y is required such that

$$x < a\colon\ d^2y/dx^2 = -\omega^2 y,$$

$$x > a\colon\ d^2y/dx^2 = k^2 y,$$

and $y = 0$ at $x = 0$ and as $x \to \infty$. At $x = a$ both y and dy/dx are continuous. Show that this is only possible if $\omega a \cot \omega a = -ka$.

7. Reduce the Debye equation, $d^2\psi/dx^2 = K \sinh (Ze\psi/kT)$, to a first-order equation for $p = d\psi/dx$ as a function of ψ and integrate it to find p.

Solve the equation completely when $Ze\psi \ll kT$, if $\psi \to 0$ as $x \to \infty$, and $\psi = \alpha$ when $x = a$, by assuming that $\sinh (Ze\psi/kT) = Ze\psi/kT$, approximately.

8. Using the method of Section 15.19, find the frequency of small oscillations of a diatomic molecule about the position of equilibrium ($r = a$), using the expression for the potential energy due to Morse,

$$V(r) = D\{1 - e^{-\alpha(r-a)}\}^2.$$

9. A rather more difficult problem than the previous one is to calculate the period of small vibrations, when the diatomic molecule is rotating about its mass centre with angular momentum H. The effective potential energy is now

$$V(r) = D(1 - e^{-\alpha(r-a)})^2 + (H^2/2\mu r^2),$$

where μ is the reduced mass. Find an equation for the new position of relative equilibrium ($dV/dr = 0$), and the new expression for the period. By an expansion in powers of H^2, which is assumed to be small, obtain the corrections to the formula of Example *8*.

10. Find a solution in series of the equation,

$$(1 - x^2)y'' - 2xy' + l(l + 1)y = 0,$$

which is a *finite* series with a highest power x^l in x, when l is a positive integer. Compare the results for $l = 0, 1, 2, 3$ with the Legendre polynomials of Section 13.28.

11. Solve the problem of damped simple harmonic motion, given in Section 15.22, by looking for a solution $x = e^{mt}$, where m is a complex number, and taking real and imaginary parts of the result.

12. If $y = e^{mx}$ is a solution of $y'' - 2my' + m^2x = 0$, prove that $y = d(e^{mx})/dm = xe^{mx}$ is also a solution.

The following series of examples contains solutions of the simpler first-order differential equations arising from chemical reactions: x denotes the amount of reactant lost and $x = 0$ at $t = 0$. In reversible reactions it can be negative. a, b, c denote initial concentrations, and it is assumed that $a \leqslant b \leqslant c$.

13. $\left.\begin{matrix} \text{A} \to \text{B} \\ \text{A} \to \text{B} + \text{C} \ldots \end{matrix}\right\} \dfrac{dx}{dt} = k(a - x),\ \dfrac{x}{a} = 1 - e^{-kt},\ k = \dfrac{1}{t}\log\dfrac{a}{a - x}.$

14. $A + B \rightarrow C + D$ . . .: $\dfrac{dx}{dt} = k(a - x)(b - x)$,

$$x = ab \cdot \frac{e^{-kat} - e^{-kbt}}{be^{-kat} - ae^{-kbt}} = a\,\frac{1 - e^{-k(b-a)t}}{1 - (a/b)e^{-k(b-a)t}}.$$

$$(b - a)kt = \ln \{a(b - x)\}/\{b(a - x)\}.$$

When $b \gg a$ (cf. Example *13*), $x = a(1 - e^{-kbt})$, $bkt = \log a/(a - x)$ (approximately).

15. $A + B \rightarrow$ . . . $(a = b)$
$2A \rightarrow$ . . .
$\Big\}\ \dfrac{dx}{dt} = k(a - x)^2,\ \dfrac{x}{a} = 1 - \dfrac{1}{1 + akt},\ kt = \dfrac{x}{a(a - x)}.$

($2a$ = amount of A)

16. $A + B + C \rightarrow$ . . .: $dx/dt = k(a - x)(b - x)(c - x)$.

$$kt = \frac{1}{(b - a)(c - a)} \ln \frac{a}{a - x} - \frac{1}{(b - a)(c - b)} \log \frac{b}{b - x} + \frac{1}{(c - a)(c - b)} \log \frac{c}{c - x}.$$

$c \gg b > a$: $ckt = \dfrac{1}{b - a} \log \dfrac{a(b - x)}{b(a - x)}$, (cf. Example *14*).

$c > b \gg a$: $bckt = \log \dfrac{a}{a - x}$, (cf. Example *13*).

17. $A + B + C \rightarrow D + . . .$: $A + 2B \rightarrow D + . . .$
$(a < b = c)$ $(a < b)$

$$dx/dt = k(a - x)(b - x)^2.$$

$$kt = \frac{1}{(b - a)^2} \left\{ \ln \frac{a(b - x)}{b(a - x)} - \frac{x(b - a)}{b(b - x)} \right\},$$

reducing, when $a \ll b$, to $kt = \dfrac{1}{b^2} \ln \dfrac{a}{a - x}$, (cf. Example *13*).

18. $A + B + C \rightarrow D + . . .$: or $2A + C \rightarrow D + . . .$
$a = b < c$ $a < c$

$$dx/dt = k(a - x)^2(c - x).$$

$$kt = \frac{1}{(c - a)^2} \left\{ \log \frac{a(c - x)}{c(a - x)} + \frac{x(c - a)}{a(a - x)} \right\},$$

reducing, when $a \ll c$, to $kact = x/(a - x)$, (cf. Example *14*).

19. $A + B + C \to D + \ldots$: $A + 2B \to D + \ldots$: $3A \to D + \ldots$
$(a = b = c)$ $(a = \frac{1}{2}b)$ $(a + b + c = 3a)$

$$dx/dt = k(a - x)^3.$$

$$kt = \frac{1}{2a^2}\left\{\left(\frac{a}{a - x}\right)^2 - 1\right\} = \frac{x(a - \frac{1}{2}x)}{a^2(a - x)^2}.$$

$$\frac{x}{a} = 1 - \frac{1}{(1 + 2a^2kt)^{1/2}}.$$

20. $nA \to D + \ldots$ or similar reactions with n molecules:

$$dx/dt = k(a - x)^n: \quad (n - 1)a^{n-1}kt = \{a/(a - x)\}^{n-1} - 1.$$

21. Reversible reactions, $A \rightleftharpoons B$: $dx/dt = k_1(a - x) - k_2(a + x)$.

$$\text{As } t \to \infty,\ x \to \lambda = \frac{k_1a - k_2b}{k_1 + k_2} = a - \frac{k_2(a + b)}{k_1 + k_2} = \frac{k_1(a + b)}{k_1 + k_2} - b.$$

$$x = \lambda(1 - e^{-\kappa t}), \text{ where } \kappa = k_1 + k_2.$$

λ is positive when $a/b > k_2/k_1$ and negative when $a/b < k_2/k_1$.

22. $A \rightleftharpoons B + C$: $dx/dt = k_1(a - x) - k_2(b + x)(c + x)$.

$$x = \frac{\lambda}{\alpha + \beta \coth k_2\beta t} = \frac{\beta^2 - \alpha^2}{\alpha + \beta \coth k_2\beta t}; \quad 2\alpha = b + c + k_1/k_2,$$
$$\beta = +(\alpha^2 + \lambda)^{1/2}.$$

$$2k_2\beta t = \log\frac{1 - (\alpha - \beta)x/\lambda}{1 - (\alpha + \beta)x/\lambda} = \log\frac{1 - x/(\alpha + \beta)}{1 - x/(\alpha - \beta)}; \quad \lambda = (k_1/k_2)a - bc.$$

apply even when λ is negative.

23. $A + B \rightleftharpoons C + D$: $dx/dt = k_1(a - x)(b - x) - k_2(c - x)(d - x)$, $(k_1 \neq k_2)$.

$$x = \frac{2\lambda\alpha}{\alpha + \beta \coth \beta(k_1 - k_2)t}$$

$$= \frac{\alpha^2 - \beta^2}{\alpha + \beta \coth \beta(k_1 - k_2)t}; \quad 2\alpha = \frac{k_1(a + b) + k_2(c + d)}{k_1 - k_2},$$
$$\beta = +(\alpha^2 - 2\lambda\alpha)^{1/2},$$

or
$$2\beta(k_1 - k_2)t = \log\frac{1 - (\alpha - \beta)x/2\lambda\alpha}{1 - (\alpha + \beta)x/2\lambda\alpha},$$

$$= \log\frac{1 - x/(\alpha + \beta)}{1 - x/(\alpha - \beta)}; \quad 2\alpha\lambda = \frac{k_1ab - k_2cd}{k_1 - k_2},$$

apply even when α and λ are negative: β is always positive.

When $k_1 = k_2$, $x = \lambda(1 - e^{-kt})$,

where λ is defined above and $k = k_1(a + b) + k_2(c + d)$.

24. Successive reactions: $\underset{a-x}{A} \xrightarrow{k_1} \underset{x-y}{B} \xrightarrow{k_2} \underset{y}{C}$. This reaction proceeds according to the equations $dx/dt = k_1(a - x)$, $dy/dt = k_2(x - y)$, where a is the initial concentration of A, x the amount transformed to B, and y the amount of B transformed to C, only A being present initially.

Prove that

$$x/a = (1 - e^{-k_1 t}),\ y/a = (1 - e^{-k_2 t}) + \{k_2/(k_2 - k_1)\}\{e^{-k_2 t} - e^{-k_1 t}\}.$$

25. In the hydrolysis of succinyldichloride by cholinesterase the following set of equations is used:

$$z = 2x_0 - 2x - y\colon\ k_1 e_1 = xE,\ k_2 e_2 = yE,\ k_3 e_3 = zE,$$
$$E = e_0 - e_1 - e_2 - e_3\colon\ dx/dt = -ke_1,\ dy/dt = ke_1 - k'e_2.$$

k_1, k_2, k_3, k, k' and e_0, x_0 are constants. Given $x = x_0$, $y = z = 0$ at $t = 0$ the equations determine x, y, z, e_1, e_2, e_3 as functions of the time.

Find dy/dx as a function of x and y by eliminating the remaining variables and solve the equations completely.

$$(\alpha - 1)y = x - x^\alpha/x_0^{\alpha-1}, \quad \alpha = k_1 k'/k_2 k,$$

$$\frac{k}{k_1}t = \frac{1}{e_0}\left(1 + \frac{2x_0}{e_3}\right)\ln\frac{x_0}{x} + \left\{\frac{1}{e_1} - \frac{2}{e_2} + \frac{1}{(\alpha-1)}\left(\frac{1}{e_2} - \frac{1}{e_3}\right)\right\}(x_0 - x)$$
$$- \frac{x_0}{\alpha(\alpha-1)}\left(\frac{1}{e_2} - \frac{1}{e_3}\right)\left(1 - \frac{x^\alpha}{x_0^\alpha}\right).$$

26. In a reaction with velocity $dx/dt = kf(x)$, $k = \dfrac{1}{t}\displaystyle\int_0^x \frac{dx}{f(x)} = \frac{\phi(x)}{t}$.

If k is calculated from one simultaneous measurement of x and t, find the error, due to an error Δx in the measurement of x and an error Δt in the measurement of t. $\{\Delta k = \Delta x/tf(x) - \Delta t\phi(x)/t^2\}$.

Find the relative error in k in Example *20*, when $f(x) = (a - x)^n$, assuming $\Delta t = 0$, and Δx the same for all times, and prove that it is least when $x/a = 1 - (1/n)^{1/(n-1)}$.

27. In the successive reversible reactions $\underset{(a-x)}{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \underset{(x-y)}{B} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \underset{y}{C}$ show that

$$dx/dt = k_1 a - (k_1 + k_2)x + k_2 y, \quad dy/dt = k_3 x - (k_3 + k_4)y.$$

Find the equilibrium values of x and y, and by eliminating y show that x satisfies the equation

$$d^2x/dt^2 + (k_1 + k_2 + k_3 + k_4)dx/dt + \{(k_1 + k_2)(k_3 + k_4) - k_2 k_3\}x$$
$$= k_1(k_3 + k_4)a.$$

What form has the solution of this equation?

28. The following equations were used to trace the metabolism of benzoic acid, and to determine the amount (g_T) excreted.

$$-da/dt = k_1 a,\ dm/dt = dg/dt + k_2,\ dg/dt = k_3 m',\ m + m' = B + A - a.$$

Solve these equations assuming $m' > 0$ and $k_2 < k_1 a$, given $a = A$ and $m = 0$ when $t = 0$. Find an equation for the time T when $m' = 0$, and show that at this time $g_T = A(1 - e^{k_1 T}) + B - k_2 T$.

The author has found helpful treatments of the subject matter of this chapter in the following books:

PIAGGIO, H. T. H. *Elementary Treatise on Differential Equations.* G. Bell and Sons, London, 1949.

PIPES, L. A. *Applied Mathematics for Engineers and Physicists.* Chapter IV. McGraw-Hill Book Co., New York, 1946.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering.* Chapters II and III. McGraw-Hill Book Co., New York, 1939.

JAEGER, J. C. *An Introduction to Applied Mathematics.* Clarendon Press, Oxford, 1950.

CONDON, E. U. and MORSE, P. *Quantum Mechanics.* McGraw-Hill Book Co., New York, 1932.

MOELWYN-HUGHES, E. A. *Physical Chemistry: an Introduction.* Clarendon Press, Oxford, 1940.

RELTON, F. E. *Applied Differential Equations.* Blackie and Sons, London and Glasgow, 1948.

SEMENOFF, P. *Chemical Kinetics and Chain Reactions.* Clarendon Press, Oxford, 1935.

CHAPTER 16

FUNCTIONS OF SEVERAL VARIABLES

Introduction—Solid representation—Contour representation—Higher derivatives—Small changes. Differentials—Calculation of derivatives—Total derivatives—Function of function rules—Thermodynamic functions—Relations between derivatives—Order of differentiation—Maxwell relations—Integrals—Heat changes—Work done by a force—Condition of integrability—Potential energy—Functions of state—Integrating factors—Maclaurin's and Taylor's series—Stationary values. Undetermined multipliers—Maxima and minima—Repeated integrals

16.01. INTRODUCTION. The number of degrees of freedom of a system is the number of variables needed to specify it completely, and it depends on the conditions applied to the system. For example, a particle moving in space has three degrees of freedom, because three coordinates are needed to specify its position; when restricted to move in a plane it only has two degrees of freedom, because only two coordinates are then needed. If it moves along a line it only has one degree of freedom.

Similarly the beginnings of thermodynamics lie in the observation that the volume and temperature of a substance can be varied independently, that these two describe the state of the substance, and other properties of the state such as the pressure, internal energy, etc., can be expressed in terms of them. In this system there are two degrees of freedom. Later on, when investigations are made of other characteristics, such as surface or magnetic properties, it is discovered that there are other independent variables, but for many purposes the conditions imposed do not allow these to vary.

In earlier chapters we have only discussed functions of one variable, describing systems with one degree of freedom. The problem in this chapter is to extend these results to functions of several variables, describing systems with more than one degree of freedom. It is found that the new features which arise when more than one variable is changed can be expressed in terms of earlier results: a new notation is used to make this easier.

A function of two variables, p and v, is written $\phi(p, v)$ with a comma between the symbols p and v. For a system with more than two degrees of freedom, extra symbols are inserted in the parentheses.

The derivative, or rate of change, of ϕ with respect to p, when v is fixed, is now written $(\partial\phi/\partial p)_v$ and called a *partial* derivative. This

new notation shows that ϕ is a function of two variables, that p is varied and v is fixed. There is another partial derivative $(\partial\phi/\partial v)_p$, the rate of change of ϕ with v when p is fixed. (The suffices are often omitted when they are obvious or cumbersome to insert.) A function has as many partial derivatives as there are independent variables.

Example 1. $U = pv$: then $(\partial U/\partial p)_v = v$ and $(\partial U/\partial v)_p = p$.

If $V = \tan^{-1}(x/y)$: then $\partial V/\partial x = y/(x^2 + y^2)$ and

$$\frac{\partial V}{\partial y} = \frac{1}{1 + x^2/y^2}\left(-\frac{x}{y^2}\right) = -\frac{x}{x^2 + y^2}.$$

Integral notation is the same as before: $\int \phi(p, v)dv$ means the integral with respect to v, keeping p fixed, unless the contrary is indicated.

16.02. SOLID REPRESENTATION. A function of position in a plane, $V = \phi(x, y)$, can be represented by a solid model on the plane as base, with an upper surface the height of which above the plane at any point equals the value of V at that point.

Similarly, any function $U = \phi(p, v)$ can be illustrated by a solid model on a plane base, using p, v as coordinates in the base rather than x, y.

Example 1. In *Figure 16.01* the function $V = \sin(x - ct)$ is shown in perspective with $c = 1$, using coordinates (x, t). Sections at $t = 0, \frac{1}{2}\pi, \pi$ are the curves $V = \sin x, -\cos x, -\sin x$ respectively.

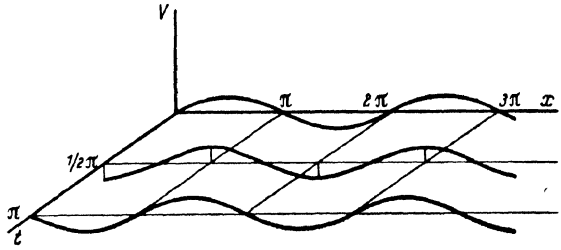

Figure 16.01

16.03. CONTOUR REPRESENTATION. It is more convenient to illustrate $V = \phi(x, y)$ or $U = \phi(p, v)$ by drawing contours, which are level sections $V = \text{const.}$ of the solid model, projected on to the plane base, as in an ordinary map. The contours are drawn for convenient values of V and labelled accordingly.

In *Figure 16.02* contours are drawn for $V = x^2 + y^2$, which are circles, and $V = x^2 - y^2$, which are the rectangular hyperbolae $x^2 - y^2 = a^2$ when V is positive and the hyperbolae $y^2 - x^2 = a^2$ when V is negative. (The contour $V = 0$ consists of two lines $x = \pm y$.) Notice that the solid model of the first would have a minimum at the origin, whereas the second would have a saddle

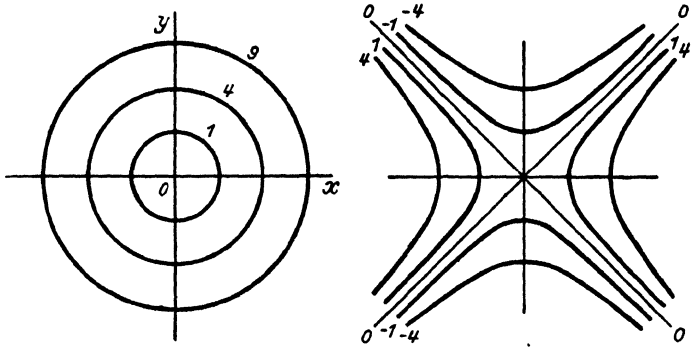

Figure 16.02

shape where V increases both ways along the x axis and decreases both ways along the y axis.

The contours of $V = e^{x^2+y^2}$ are the same circles as $V = x^2 + y^2$, but labelled differently.

16.04. Higher Derivatives. With a function of two or more variables, a new feature appears in 'mixed' higher derivatives. Thus, given $U = \phi(p, v)$ we calculate $\partial U/\partial p$ and $\partial U/\partial v$. Generally, each is a function of p and v and has two derivatives, making four in all,

$$\frac{\partial}{\partial p}\left(\frac{\partial U}{\partial p}\right), \quad \frac{\partial}{\partial v}\left(\frac{\partial U}{\partial p}\right), \quad \frac{\partial}{\partial p}\left(\frac{\partial U}{\partial v}\right), \quad \frac{\partial}{\partial v}\left(\frac{\partial U}{\partial v}\right)$$

i.e.

$$\frac{\partial^2 U}{\partial p^2}, \quad \frac{\partial^2 U}{\partial v\,\partial p}, \quad \frac{\partial^2 U}{\partial p\,\partial v}, \quad \frac{\partial^2 U}{\partial v^2}.$$

It is unreasonable to insert suffices in higher derivatives.

The number of higher derivatives increases very rapidly with their order, but there is one simplification discussed in Section 16.06.

16.05. Small Changes. Differentials. The change in a function due to changes in several variables can be expressed as a sum of

changes due to each variable, the contribution of each being independent of the rest (Distributive Law). Consider (*Figure 16.03*) the change in $F = \phi(v, T)$ from $A(v, T)$ to $C(v + \delta v, T + \delta T)$. Either

$$\Delta F = F_C - F_A = (F_B - F_A) + (F_C - F_B),$$

or

$$\Delta F = F_C - F_A = (F_C - F_D) + (F_D - F_A).$$

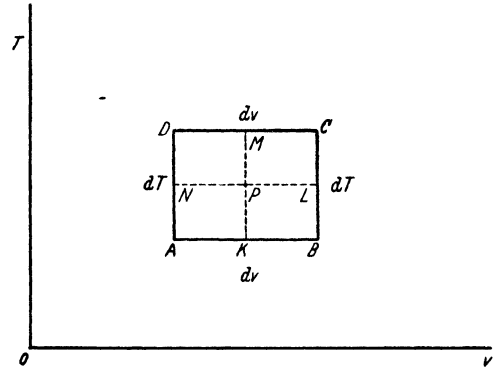

Figure 16.03

In each equation, the first term is due to a change of v only, the second due to a change in T only. Consider the first equation: to first order,

$$F_B - F_A = (\partial F/\partial v)_A \delta v,$$

$$F_C - F_B = (\partial F/\partial T)_B \delta T = (\partial F/\partial T)_A \delta T.$$

In the second expression, the difference between $(\partial F/\partial T)_B$ and $(\partial F/\partial T)_A$ is only of first order and is neglected when multiplied by δT. In all

$$\Delta F = (\partial F/\partial v)_A \delta v + (\partial F/\partial T)_A \delta T.$$

The same result is obtained from the other expression for ΔF. Thus we have proved the initial statement.

The total differential of $F = \phi(v, T)$ is defined as

$$dF = (\partial F/\partial v)_T dv + (\partial F/\partial T)_v dT,$$

with one term for each independent variable. If one or more terms are omitted it is called a partial differential, e.g. $dF_v = (\partial F/\partial v)_T dv$.

When dv and dT are the changes to first order in v and T, respectively, then dF is the change in F to first order.

This definition is an extension of the differential of a function of one variable and does not conflict with the rules given there.

Example 1. $V = x^3 + 3xy + y^3$.

(*i*) $dV = (\partial V/\partial x)_y dx + (\partial V/\partial y)_x dy,$
$= 3(x^2 + y)dx + 3(x + y^2)dy.$

(*ii*) $dV = d(x^3) + 3d(xy) + d(y^3),$
$= 3x^2 dx + 3(x\,dy + y\,dx) + 3y^2\,dy,$
$= 3(x^2 + y)dx + 3(x + y^2)dy.$

Example 2. (For the reader.) Repeat the remarks of this section for $V = \phi(x, y, z)$.

16.06. CALCULATION OF DERIVATIVES. The basic distribution law for differentials given in Section 16.05 has to be adapted to different problems. The two elementary principles in the calculation of derivatives are:

(*i*) a derivative $\partial U/\partial s$ can be obtained only under conditions which allow s to be varied and U to vary with it. If these conditions, whatever form they take, are denoted by a suffix C, then

$$\left(\frac{\partial U}{\partial s}\right)_C = \lim_{\Delta s \to 0} \left(\frac{\Delta U}{\Delta s}\right)_C = \frac{dU_C}{ds_C},$$

where dU_C and ds_C are the differentials of U and s, since the second-order corrections vanish in the limit.

(*ii*) If $U, V, W, \ldots$ are all functions of s alone under the same conditions C, then a relation between their differentials $dU, dV, dW, \ldots$, on division by ds, becomes a relation between their derivatives $(\partial U/\partial s)_C, (\partial V/\partial s)_C, \ldots$

Example 1. By the first law of thermodynamics $\delta Q = dE + p dv$. Hence, under any conditions C so that all quantities are related to temperature, the specific heat is

$$c = \left(\frac{dQ}{dT}\right)_C = \left(\frac{\partial E}{\partial T}\right)_C + p\left(\frac{\partial v}{\partial T}\right)_C.$$

At constant pressure $c_p = \left(\frac{\partial E}{\partial T}\right)_p + p\left(\frac{\partial v}{\partial T}\right)_p.$

At constant volume ($dv = 0$), $\delta Q = dE$ and

$$c_v = \left(\frac{\partial E}{\partial T}\right)_v.$$

16.07. TOTAL DERIVATIVES. Let $V = \phi(x, y, z, \ldots)$, where $x, y, z, \ldots$ describe the state of the system. Assume that it is so controlled that $x, y, z, \ldots$ are all functions of a single parameter t. Then

$$dV = \frac{\partial V}{\partial x}\,dx + \frac{\partial V}{\partial y}\,dy + \frac{\partial V}{\partial z}\,dz + \ldots$$

Divide by dt and take the limit $dt \to 0$;

$$\frac{dV}{dt} = \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt} + \ldots$$

This is the total derivative of V with respect to t. Each variable contributes one term to the total rate of change, and the total rate of change is the sum of them all. It should be noticed that each contribution is independent of the others, e.g. the term $(\partial V/\partial x)(dx/dt)$ due to the change of x appears whether or not the other variables $y, z, \ldots$ change with t. This is a generalization of Section 4.14.

An important special case is that where t is one of the independent variables, e.g. if y and z are functions of x,

$$\frac{dV}{dx} = \frac{\partial V}{\partial x} + \frac{\partial V}{\partial y}\frac{dy}{dx} + \frac{\partial V}{\partial z}\frac{dz}{dx}.$$

$(\partial V/\partial x)$ is the partial rate of change of V with x; (dV/dx) is the total rate of change when y and z also vary with x.

Example 1. $p = \phi(v, T)$. If v and T change with time, then

$$\frac{dp}{dt} = \left(\frac{\partial p}{\partial v}\right)_T \frac{dv}{dt} + \left(\frac{\partial p}{\partial T}\right)_v \frac{dT}{dt}.$$

Example 2. $p = \phi(v, T)$. When a gas expands with increasing temperature, the rate of increase of pressure with temperature is

$$\frac{dp}{dT} = \left(\frac{\partial p}{\partial T}\right)_v + \left(\frac{\partial p}{\partial v}\right)_T \frac{dv}{dT}.$$

Example 3. Elements of fluid moving in the x direction have velocities $u = \phi(x, t)$ depending on their position x and the time t when they arrive there. The acceleration, or rate of change of velocity of one element, is in part due to its change of position, by $dx = u\,dt$ in time dt.

$$\frac{du}{dt} = \left(\frac{\partial u}{\partial t}\right)_x + \left(\frac{\partial u}{\partial x}\right)_t \frac{dx}{dt}: \quad u = \frac{dx}{dt}.$$

16.08. FUNCTION OF FUNCTION RULES. Some further examples are given of these: the similarities should be noted rather than the differences. The reader should try to interpret *Figure 16.04* p 261 using these sections.

(*i*) A system is partially controlled so that everything depends on two parameters s and t.

Let $U = \phi(x, y, z, \ldots)$, as in the previous section, and keep s constant. Then everything depends on t.

$$\left(\frac{\partial U}{\partial t}\right)_s = \frac{\partial U}{\partial x}\left(\frac{\partial x}{\partial t}\right)_s + \frac{\partial U}{\partial y}\left(\frac{\partial y}{\partial t}\right)_s + \frac{\partial U}{\partial z}\left(\frac{\partial z}{\partial t}\right)_s + \ldots$$

Similarly,

$$\left(\frac{\partial U}{\partial s}\right)_t = \frac{\partial U}{\partial x}\left(\frac{\partial x}{\partial s}\right)_t + \frac{\partial U}{\partial y}\left(\frac{\partial y}{\partial s}\right)_t + \frac{\partial U}{\partial z}\left(\frac{\partial z}{\partial s}\right)_t + \ldots$$

When $U = \phi(x, y)$ is a function of only two variables, these formulae are used for a change of variable from (x, y) to (s, t).

Example 1. Let (x, y) be Cartesian coordinates and (r, θ) polar coordinates, so that $x = r\cos\theta$ and $y = r\sin\theta$. If $U = \phi(x, y)$ is expressed in terms of r and θ, then

$$\left(\frac{\partial U}{\partial r}\right)_\theta = \frac{\partial U}{\partial x}\left(\frac{\partial x}{\partial r}\right)_\theta + \frac{\partial U}{\partial y}\left(\frac{\partial y}{\partial r}\right)_\theta = \cos\theta\,\frac{\partial U}{\partial x} + \sin\theta\,\frac{\partial U}{\partial y},$$

$$\left(\frac{\partial U}{\partial \theta}\right)_r = \frac{\partial U}{\partial x}\left(\frac{\partial x}{\partial \theta}\right)_r + \frac{\partial U}{\partial y}\left(\frac{\partial y}{\partial \theta}\right)_r = -r\sin\theta\,\frac{\partial U}{\partial x} + r\cos\theta\,\frac{\partial U}{\partial y}.$$

(*ii*) A property U of a system is a function of a second property S: to find $\partial U/\partial x$, etc., given S as a function of x, y, z.

Under all conditions $dU = \dfrac{dU}{dS}\,dS$. Hence for variations due to x alone ($y, z, \ldots$ fixed),

$$(\partial U/\partial x)_{y,z,\ldots} = (dU/dS)(\partial S/\partial x)_{y,z,\ldots}.$$

and similarly

$$\partial U/\partial y = (dU/dS)\partial S/\partial y, \quad \partial U/\partial z = (dU/dS)\partial S/\partial z, \text{ etc.}$$

Example 2. The internal energy E of a substance is a function of temperature alone, and $T = \phi(p, v)$. Then

$$\left(\frac{\partial E}{\partial v}\right)_p = \frac{dE}{dT}\left(\frac{\partial T}{\partial v}\right)_p; \quad \left(\frac{\partial E}{\partial p}\right)_v = \frac{dE}{dT}\left(\frac{\partial T}{\partial p}\right)_v.$$

A more general example is the following.

Example 3. Owing to external forces the pressure in an isothermal gas varies with coordinates (x, y, z). Find $(\partial E/\partial x)_{y,z}$ in terms of the known derivatives, given the internal energy E and temperature T as functions of pressure and volume.

Using an extension of (i),

$$\frac{\partial E}{\partial x} = \left(\frac{\partial E}{\partial p}\right)_v \frac{\partial p}{\partial x} + \left(\frac{\partial E}{\partial v}\right)_p \frac{\partial v}{\partial x}.$$

To eliminate $\partial v/\partial x$ we use the isothermal nature of the gas.

$$\frac{\partial T}{\partial x} = \left(\frac{\partial T}{\partial p}\right)_v \frac{\partial p}{\partial x} + \left(\frac{\partial T}{\partial v}\right)_p \frac{\partial v}{\partial x} = 0.$$

$$\therefore \quad \frac{\partial E}{\partial x} = \left\{\left(\frac{\partial E}{\partial p}\right)_v - \left(\frac{\partial E}{\partial v}\right)_p \frac{(\partial T/\partial p)_v}{(\partial T/\partial v)_p}\right\} \frac{\partial p}{\partial x}.$$

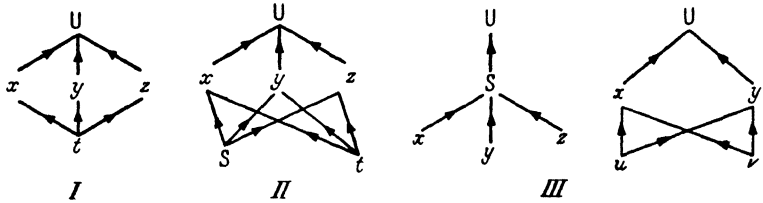

Figure 16.04

16.09. THERMODYNAMIC FUNCTIONS. The position here is complicated by the numbers of functions used. Consider the simplest system with two degrees of freedom. In addition to pressure p, volume v and absolute temperature T, there are the entropy S and the internal energy E such that

$$dE = TdS - pdv. \qquad \ldots.(1)$$

There are three other energy functions:

$$F = E - TS: \quad dF = dE - TdS - SdT = -SdT - pdv, \qquad \ldots.(2)$$

$$H = E + pv: \quad dH = dE + pdv + vdp = TdS + vdp, \qquad \ldots.(3)$$

$$G = H - TS: \quad dG = dH - TdS - SdT = -SdT + vdp. \qquad \ldots.(4)$$

Any two of these can be used to describe the state, and there is the general problem of expressing derivatives in terms of those with simple meanings, which 'everyone' can recognize. This usually means derivatives with respect to v, T, and p.

An elementary method which follows the general procedure described above is the following. From the meaning of differentials we can derive from the differential equations above (e.g. from equation (1)) such results as

$$(\partial E/\partial S)_v = T, \quad (\partial E/\partial v)_S = -p.$$

Let us now assume that we require a derivative such as

$$X = (\partial H/\partial G)_S,$$

using p and T as independent variables. This is equivalent to the two equations

$$XdG = dH, \quad dS = 0.$$

We now express both in terms of dp and dT and equate the two ratios dp/dT obtained. Using equations (3) and (4) they become

$$X(-SdT + vdp) = TdS + vdp = vdp \ (dS = 0),$$

and

$$dS = (\partial S/\partial T)_p dT + (\partial S/\partial p)_T dp = 0.$$

From these two equations,

$$\frac{dp}{dT} = -\frac{(\partial S/\partial T)_p}{(\partial S/\partial p)_T} = \frac{-XS}{v - Xv}.$$

The solution of this equation for X gives

$$X = \left(\frac{\partial H}{\partial G}\right)_S = \frac{v(\partial S/\partial T)_p}{v(\partial S/\partial T)_p + S(\partial S/\partial p)_T}.$$

These formulae are usually expressed in terms of known quantities using such relations as those for the specific heats or compressibility,

$$c_p = (dQ/dT)_p = T(\partial S/\partial T)_p,$$
$$c_v = (dQ/dT)_v = T(\partial S/\partial T)_v.$$

See also Sections 16.10 and 16.12.

16.10. Relations between Derivatives. The equations

$$p = \phi(v, T) \text{ and } dp = (\partial p/\partial v)_T dv + (\partial p/\partial T)_v dT,$$

relate the three variables p, v and T and their differentials. They can be used to obtain other derivatives.

At constant temperature ($dT = 0$), we obtain the rate of change of pressure with volume, or the rate of change of volume with pressure,

$$\left(\frac{\partial p}{\partial v}\right)_T \text{ and } \left(\frac{\partial v}{\partial p}\right)_T = \left(\frac{dv}{dp}\right)_{dT=0} = \frac{1}{(\partial p/\partial v)_T}.$$

Similarly,

$$\left(\frac{\partial T}{\partial p}\right)_v = \left(\frac{dT}{dp}\right)_{dv=0} = \frac{1}{(\partial p/\partial T)_v}.$$

Finally, when $dp = 0$, we obtain the rate of change of volume with temperature at constant pressure

$$\left(\frac{\partial v}{\partial T}\right)_p = \left(\frac{dv}{dT}\right)_{dp=0} = -\frac{(\partial p/\partial T)_v}{(\partial p/\partial v)_T} = -\left(\frac{\partial p}{\partial T}\right)_v \left(\frac{\partial v}{\partial p}\right)_T.$$

The minus sign is due to the different conditions for the various derivatives.

Since $\lambda = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_p$ is the coefficient of thermal expansion and $k = -\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T$ is the isothermal compressibility,

$$\left(\frac{\partial p}{\partial T}\right)_v = -\left(\frac{\partial v}{\partial T}\right)_p \Big/ \left(\frac{\partial v}{\partial p}\right)_T = \lambda/k.$$

16.11. ORDER OF DIFFERENTIATION. In Section 16.05 the change in $F = \phi(v, T)$ was calculated to first order. The same result was

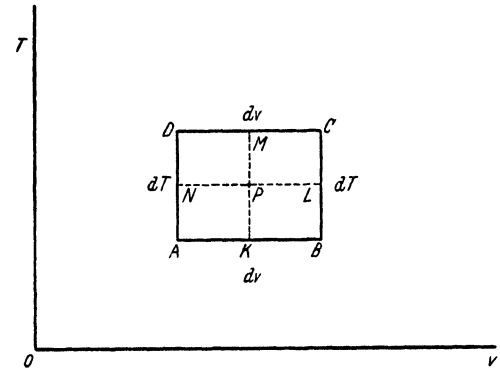

Figure 16.05

obtained for the two paths *ABC* and *ADC* (*Figure 16.05*). Of course the same result must be obtained to any order, since

$$\Delta F = \phi(v + \delta v, T + \delta T) - \phi(v, T)$$

is unique. This has remarkable consequences! To second order (Section 5.15)

$$\Delta F = (F_B - F_A) + (F_C - F_B) = \alpha_K \delta v + \beta_L \delta T;$$
$$\Delta F = (F_C - F_D) + (F_D - F_A) = \alpha_M \delta v + \beta_N \delta T,$$

where $\alpha = (\partial F/\partial v)_T, \quad \beta = (\partial F/\partial T)_v.$

The difference of the two, to second order, is

$$(\alpha_M - \alpha_K)\delta v - (\beta_L - \beta_N)\delta T = (\partial \alpha/\partial T)_G \delta T \delta v - (\partial \beta/\partial v)_G \delta v \delta T.$$

The two values of ΔF are equal only if this is zero, i.e.

$$(\partial \alpha/\partial T)_v = (\partial \beta/\partial v)_T,$$

or

$$\frac{\partial}{\partial T}\left(\frac{\partial F}{\partial v}\right) = \frac{\partial}{\partial v}\left(\frac{\partial F}{\partial T}\right).$$

This important result means that in the calculation of higher derivatives the order of differentiation does not affect the final result.

The differences ΔF for various paths can be equated to third or higher order but no essentially new result is obtained. For example, third-order terms show that

$$\frac{\partial^3 F}{\partial v^2 \partial T} = \frac{\partial^3 F}{\partial v \partial v \partial T} = \frac{\partial^3 F}{\partial v \partial T \partial v} = \frac{\partial^3 F}{\partial T \partial v \partial v},$$

which merely corresponds once again to the assertion that the order of differentiation does not matter.

Summing up, if $dF = \alpha\, dv + \beta\, dT$ is the differential of $F = \phi(v, T)$,

then $$(\partial \alpha/\partial T)_v = (\partial \beta/\partial v)_T.$$

When an expression $(\alpha\, dv + \beta\, dT)$ satisfies this condition it is said to be an exact differential.

16.12. Maxwell Relations. The thermodynamic relations given in Section 16.09 combined with the last result give the following Maxwell relations:

$$dF = -SdT - pdv: \quad (\partial S/\partial v)_T = (\partial p/\partial T)_v,$$
$$dE = TdS - pdv: \quad (\partial T/\partial v)_S = -(\partial p/\partial S)_v,$$
$$dH = TdS + vdp: \quad (\partial T/\partial p)_S = (\partial v/\partial S)_p,$$
$$dG = -SdT + vdp: \quad -(\partial S/\partial p)_T = (\partial v/\partial T)_p.$$

Three Variables. If $V = \phi(x, y, z)$ and

$$dV = \alpha\, dx + \beta\, dy + \gamma\, dz,$$

then $$\alpha = \partial V/\partial x, \quad \beta = \partial V/\partial y, \quad \gamma = \partial V/\partial z,$$

and $$\frac{\partial \beta}{\partial z} = \frac{\partial \gamma}{\partial y}, \quad \frac{\partial \gamma}{\partial x} = \frac{\partial \alpha}{\partial z}, \quad \frac{\partial \alpha}{\partial y} = \frac{\partial \beta}{\partial x}.$$

16.13. INTEGRALS. Any function of two or more variables can be integrated with respect to one, keeping the others fixed, by the methods of integration described in Chapter 10. In fact, most of the integrals evaluated were of that form.

Example 1. $$\int_0^a xy\,dx = \left[\tfrac{1}{2}x^2y\right]_0^a = \tfrac{1}{2}a^2y.$$

$$\int_0^b xy\,dy = \left[\tfrac{1}{2}xy^2\right]_0^b = \tfrac{1}{2}b^2x.$$

Constants of integration are less well defined. An integral $F = \int \phi(x, y)dx$ (y constant) is any function with the partial derivative $\partial F/\partial x = \phi(x, y)$. Integrals can differ by any function independent of x, i.e. not only constants, but any function of y.

Example 2. $\int xy\,dx = \frac{1}{2}x^2y + g(y)$, where $g(y)$ is an arbitrary function of y.

A new type of integral is

$$F = \int(\alpha\,dx + \beta\,dy),$$

such that $dF = \alpha\,dx + \beta\,dy$. This only exists where $\partial\alpha/\partial y = \partial\beta/\partial x$, as we now show.

16.14. HEAT CHANGES. To raise the temperature of a substance $d\theta$ at constant volume, the heat required is $\delta Q = c\,d\theta$, where c is the specific heat at constant volume. To raise the temperature from θ_1 to θ_2 at constant volume, the heat required is

$$Q = \int_{\theta_1}^{\theta_2} c\,d\theta, \quad (v \text{ constant}).$$

Similarly to increase the volume from v_1 to v_2, at constant temperature, the heat required is

$$Q = \sum_{v_1}^{v_2} l\,dv = \int_{v_1}^{v_2} l\,dv,$$

where l is the latent heat for volume changes at constant temperature.

Now suppose the temperature is raised from θ_1 to θ_2 and the volume also changed from v_1 to v_2, i.e. the substance changes from

the state $A(\theta_1, v_1)$ to the state $B(\theta_2, v_2)$ (*Figure 16.06*). This can be done in various ways. We could increase the temperature first and then the volume (ACB), or the volume first and then the temperature (ADB).

In practice it is more likely that they both change together, e.g. at constant pressure. Unless c and l are constants, it is necessary to specify the relation between volume and temperature during the change. This relation corresponds to a definite path from A to B in

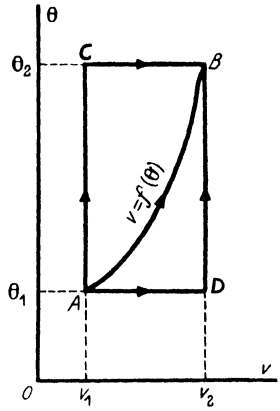

Figure 16.06

Figure 16.06. For each small step in the total change we can calculate dv, c and l in terms of θ and $d\theta$, calculate $\delta Q = c\,d\theta + l\,dv$ and by summing obtain the total amount of heat,

$$Q = \int_A^B (c\,d\theta + l\,dv).$$

The result depends in general on the way the change is carried out, i.e. on the 'path' from A to B.

16.15. Work done by a Force. Let α and β be the x- and y-components of a force acting on a particle at the point (x, y) in a plane. If the particle moves to the point $(x + dx, y + dy)$, the work done on the body by the force is defined (*Figure 16.07*, p 268) as

$$\delta W = \alpha\,dx + \beta\,dy.$$

If the body moves along a chosen path from A to B, the total work done is

$$W = \Sigma \delta W = \int_A^B (\alpha\,dx + \beta\,dy).$$

To evaluate this when the force varies, we need to know how (α, β) depend on the position of the body (x, y) and the equation of the path, as well as the position of A and B.

16.16. CONDITION OF INTEGRABILITY. The meaning of 'line integral',

$$\int_A^B (\alpha\, dx + \beta\, dy),$$

should be clear from the two examples just given, as well as the method of evaluation.

In certain circumstances the integration is very easy. In the simpler problem with one variable, $\int_a^b f(x)dx$ was evaluated in the form $F(b) - F(a)$ by finding a function $F(x)$ such that $dF = f(x)dx$. By exactly the same argument we can write

$$\int_A^B (\alpha\, dx + \beta\, dy) = F(B) - F(A) = F(x_2, y_2) - F(x_1, y_1),$$

for all positions of $B(x_2, y_2)$ and $A(x_1, y_1)$, provided it is possible to find a function $F(x, y)$ such that

$$dF = \alpha\, dx + \beta\, dy,$$

for all dx and dy, i.e. $\partial F/\partial x = \alpha$ and $\partial F/\partial y = \beta$.

In Section 16.11, it was shown that this cannot be done unless

$$\partial\alpha/\partial y = \partial\beta/\partial x.$$

This is called the condition of integrability, since the function $F(x, y)$ can always be found when it is satisfied. One method of proving this is given in *Example 10* (p 276). When this condition is satisfied $(\alpha\, dx + \beta\, dy)$ is called an exact differential.

That some condition must be satisfied is clear, since we are in effect integrating simultaneously with respect to two variables. There are severe limitations on the integral. If the integral is $F(B) - F(A)$ it depends only on the position of A and B and not on the path from one to the other. This fact explains why the condition of integrability is the same as the condition which was obtained in Section 16.11.

Consider *Figure 16.07*. It was shown in Section 16.11 that the value of $(\alpha\, dx + \beta\, dy)$ was exactly the same for the two paths EFG and EHG between neighbouring points E and G when

$$\partial\alpha/\partial y = \partial\beta/\partial x.$$

Hence,

$$W = \int_A^B (\alpha\, dx + \beta\, dy)$$

is the same for the two paths *AEFGB* and *AEHGB*. Now a succession of small distortions of the path such as that from *EFG* to *EHG* would change the path *AEB* into any other path *ACB*, and the integral does

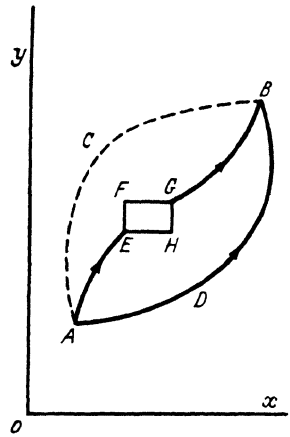

Figure 16.07

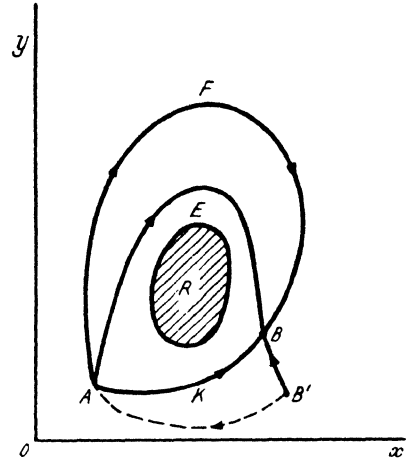

Figure 16.08

not change in value as long as the condition of integrability is satisfied.

We can go further. If this condition is not satisfied, it was shown in Section 16.11 that the change in W is a change in path from *EFG* to *EHG* represented by

$$\left(\frac{\partial \alpha}{\partial y} - \frac{\partial \beta}{\partial x}\right) dx\, dy,$$

which is proportional to the area $dx\, dy$ covered by the distortion. For a large deformation in the path the total change in W is

$$\Delta W = \Sigma \left(\frac{\partial \alpha}{\partial y} - \frac{\partial \beta}{\partial x}\right) dx\, dy,$$

which is a multiple of the total area between the old and the new paths. This proves that the change in W is in fact appreciable, in general, unless the condition of integrability is satisfied everywhere between paths. It also proves that W is the same for any two paths, even when this condition is not satisfied everywhere, provided the regions R where this occurs do not lie between the paths.

Example 1. Consider $(3x + y)dy + (3y + 2x)dx$ where

$$\alpha = 3y + 2x, \quad \beta = 3x + y.$$

Then $\partial\alpha/\partial y = 3$ and $\partial\beta/\partial x = 3$, so that this is an exact differential. In fact

$$d\phi = (3y + 2x)dx + (3x + y)dy = 2x\,dx + 3(x\,dy + y\,dx) + y\,dy,$$

i.e.
$$d\phi = d(x^2) + 3d(xy) + d(\tfrac{1}{2}y^2),$$
$$\phi = x^2 + 3xy + \tfrac{1}{2}y^2 + \text{const.}$$

Example 2. Consider $y^2dx + xy\,dy$, where $\alpha = y^2$ and $\beta = xy$. Then $\partial\alpha/\partial y = 2y$ and $\partial\beta/\partial x = y$, so that this is not an exact differential. In this case it is easily verified that the integral,

$$\int(y^2dx + xy\,dy),$$

depends on the path between two points.

16.17. POTENTIAL ENERGY. When α and β are the components of a force (Section 16.15), $W = \int_A^B (\alpha\,dx + \beta\,dy)$ is the work done. If $\partial\alpha/\partial y = \partial\beta/\partial x$, then there is a function $V = \phi(x, y)$, known as the potential energy, such that

$$\alpha = -\partial V/\partial x, \quad \beta = -\partial V/\partial y, \quad \text{and } \alpha\,dx + \beta\,dy = -dV;$$

the work done is

$$W = \int_A^B (-dV) = V_A - V_B.$$

16.18. FUNCTIONS OF STATE. In thermodynamics it is necessary to distinguish between such quantities as pressure p and internal energy E, which are functions of state, and those which are not.

If the state is defined by volume and temperature, then there is an internal energy function $E = \phi(v, \theta)$ such that the change in internal energy is measured by the change in this function. For a small change,

$$dE = p'\,dv + q'\,d\theta$$

is an exact differential.

There is no such function to determine the heat given to a body since this depends on the change of state, and the manner in which it is accomplished;

$$\delta Q = c\,d\theta + l\,dv,$$

is not an exact differential.

16.19. INTEGRATING FACTORS. Even though $\alpha\,dx + \beta\,dy$ is not an exact differential, it can be proved that it is always possible (though not necessarily easy) to find a function $\mu(x, y)$ such that

$$(\mu\alpha)dx + (\mu\beta)dy$$

is an exact differential. This function is called an integrating factor.

In thermodynamics,

$$\delta Q = l\,dv + c\,d\theta$$

is not an exact differential. The second law of thermodynamics (Kelvin's principle) shows that there is an integrating factor which is a function of the temperature alone. Denoting this by $(1/T)$, where T is a function of θ, we derive

$$dS = \delta Q/T = (l/T)dv + (c/T)d\theta.$$

This is an exact differential: S is the entropy.

For three variables, $\alpha\,dx + \beta\,dy + \gamma\,dz$ is an exact differential if

$$\frac{\partial\alpha}{\partial y} = \frac{\partial\beta}{\partial x}, \quad \frac{\partial\beta}{\partial z} = \frac{\partial\gamma}{\partial y}, \quad \frac{\partial\gamma}{\partial x} = \frac{\partial\alpha}{\partial z}.$$

Otherwise there is an integrating factor only when

$$\gamma\left(\frac{\partial\alpha}{\partial y} - \frac{\partial\beta}{\partial x}\right) + \alpha\left(\frac{\partial\beta}{\partial z} - \frac{\partial\gamma}{\partial y}\right) + \beta\left(\frac{\partial\gamma}{\partial x} - \frac{\partial\alpha}{\partial z}\right) = 0.$$

In thermodynamics, because of the second law, $(1/T)$ is always an integrating factor, however many variables there may be.

16.20. MACLAURIN'S AND TAYLOR'S SERIES. Power series are essentially the same as those considered in Chapter 13 but everything is now in terms of two variables, each raised to a power. Problems of convergence are much more involved and are not discussed here.

Using the suffix A to indicate that the derivatives are calculated at $x = a, y = b$, Taylor's series for $\phi(x, y)$ becomes

$$\phi(a + h, b + k) = \phi(a, b)$$
$$+ \left(h\frac{\partial\phi}{\partial x} + k\frac{\partial\phi}{\partial y}\right)_A + \frac{1}{2!}\left(h^2\frac{\partial^2\phi}{\partial x^2} + 2hk\frac{\partial^2\phi}{\partial x\,\partial y} + k^2\frac{\partial^2\phi}{\partial y^2}\right)_A$$
$$+ \frac{1}{3!}\left(h^3\frac{\partial^3\phi}{\partial x^3} + 3h^2k\frac{\partial^3\phi}{\partial x^2\,\partial y} + 3hk^2\frac{\partial^3\phi}{\partial x\,\partial y^2} + k^3\frac{\partial^3\phi}{\partial y^3}\right)_A + \cdots$$

Maclaurin's series is

$$\phi(x,y) = \phi(0,0) + \left(x\frac{\partial\phi}{\partial x} + y\frac{\partial\phi}{\partial y}\right)_0 + \frac{1}{2!}\left(x^2\frac{\partial^2\phi}{\partial x^2} + 2xy\frac{\partial^2\phi}{\partial x\,\partial y} + y^2\frac{\partial^2\phi}{\partial y^2}\right)_0 + \ldots$$

Proof: given a, b, h, k, there is a function $f(t) = \phi(a + th, b + tk)$ which becomes $f(0) = \phi(a, b)$ when $t = 0$, and $f(1) = \phi(a+h, b+k)$ when $t = 1$. By Maclaurin's theorem,

$$f(t) = f(0) + t\left(\frac{df}{dt}\right)_0 + \frac{1}{2!}t^2\left(\frac{d^2f}{dt^2}\right)_0 + \ldots + \frac{1}{n!}t^n\left(\frac{d^nf}{dt^n}\right)_0 + \ldots$$

and if this is convergent when $t = 1$,

$$f(1) = f(0) + \left(\frac{df}{dt}\right)_0 + \frac{1}{2!}\left(\frac{d^2f}{dt^2}\right)_0 + \ldots$$

Now $f(t) = \phi(x, y)$, where $x = a + th$, $y = b + tk$. Hence,

$$\frac{df}{dt} = \frac{\partial\phi}{\partial x}\frac{dx}{dt} + \frac{\partial\phi}{\partial y}\frac{dy}{dt} = h\frac{\partial\phi}{\partial x} + k\frac{\partial\phi}{\partial y};$$

$$\frac{d^2f}{dt^2} = \frac{d}{dt}\left(\frac{df}{dt}\right) = h\frac{\partial}{\partial x}\left(\frac{df}{dt}\right) + k\frac{\partial}{\partial y}\left(\frac{df}{dt}\right),$$

$$= h^2\frac{\partial^2\phi}{\partial x^2} + 2hk\frac{\partial^2\phi}{\partial x\,\partial y} + k^2\frac{\partial^2\phi}{\partial y^2}, \text{ etc.}$$

Putting $t = 0$ on both sides leads to the result stated.

16.21. STATIONARY VALUES. UNDETERMINED MULTIPLIERS. A stationary value of a function ϕ of a number of variables occurs for a set of values of the variables such that $d\phi = 0$ for small changes in any or all of the variables. For example, the stationary values of $\phi(x, y, z)$ are determined by the condition,

$$d\phi = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial z}dz = 0, \quad (\text{all } dx, dy, dz),$$

i.e. $$\frac{\partial\phi}{\partial x} = 0, \quad \frac{\partial\phi}{\partial y} = 0, \quad \frac{\partial\phi}{\partial z} = 0.$$

These three equations give all possible sets of values of (x, y, z).

Stationary values are often required for a function of several variables which are *not* independent, because of a condition imposed

upon them. The obvious method is to use the condition to eliminate one variable.

Example 1. Find the stationary values of $\phi = xyz$, when

$$x + y + 2z = a.$$

Eliminating x, then $\phi = yz(a - y - 2z)$, which is stationary when

$$\partial\phi/\partial y = az - 2yz - 2z^2 = 0 \text{ and } \partial\phi/\partial z = ay - y^2 - 4yz = 0.$$

Hence $\left.\begin{matrix}\text{either} & z = 0 \\ \text{or} & y + z = \frac{1}{2}a\end{matrix}\right\}$ and $\begin{cases}\text{either} & y = 0, \\ \text{or} & y + 4z = a,\end{cases}$

i.e. (*i*) $y = 0, z = 0, x = a$, (*ii*) $y = 0, z = \frac{1}{2}a, x = 0$,
(*iii*) $z = 0, y = a, x = 0$, (*iv*) $y = \frac{1}{3}a, z = \frac{1}{6}a, x = \frac{1}{3}a$.

The elegant method of undetermined multipliers due to Lagrange carries out the elimination after differentiation. Thus stationary values of $\phi(x, y, z)$, subject to a condition $f(x, y, z) = c$, occur when

$$d\phi = \frac{\partial\phi}{\partial x}\,dx + \frac{\partial\phi}{\partial y}\,dy + \frac{\partial\phi}{\partial z}\,dz = 0,$$

where $f(x, y, z) = c$, and $df = \dfrac{\partial f}{\partial x}\,dx + \dfrac{\partial f}{\partial y}\,dy + \dfrac{\partial f}{\partial z}\,dz = 0.$

If we wish to eliminate x, we also eliminate dx using the last equation to express it in terms of dy and dz. To do this first combine $d\phi$ and df using a multiplier λ,

$$d\phi - \lambda\,df = \left(\frac{\partial\phi}{\partial x} - \lambda\frac{\partial f}{\partial x}\right)dx + \left(\frac{\partial\phi}{\partial y} - \lambda\frac{\partial f}{\partial y}\right)dy + \left(\frac{\partial\phi}{\partial z} - \lambda\frac{\partial f}{\partial z}\right)dz = 0.$$

Now choose λ so that

$$\frac{\partial\phi}{\partial x} - \lambda\frac{\partial f}{\partial x} = 0.$$

The rest of the expression must now (after eliminating x) vanish for all values of dy and dz, i.e.

$$\frac{\partial\phi}{\partial y} - \lambda\frac{\partial f}{\partial y} = 0, \quad \frac{\partial\phi}{\partial z} - \lambda\frac{\partial f}{\partial z} = 0.$$

Since the form of these equations is the same whether we actually eliminate x or not, we have a beautifully symmetric result:

stationary values of $\phi(x, y, z)$, subject to $f(x, y, z) = c$, occur for those values of (x, y, z) which make

$$f(x, y, z) = c; \quad \frac{\partial \phi}{\partial x} - \lambda \frac{\partial f}{\partial x} = 0, \quad \frac{\partial \phi}{\partial y} - \lambda \frac{\partial f}{\partial y} = 0, \quad \frac{\partial \phi}{\partial z} - \lambda \frac{\partial f}{\partial z} = 0.$$

Example 2. The stationary values of $\phi = xyz$ subject to

$$x + y + 2z = a,$$

occur when

$$x + y + 2z = a; \quad yz - \lambda = 0, \quad xz - \lambda = 0, \quad xy - 2\lambda = 0,$$

i.e.

$$\lambda = yz = xz = \tfrac{1}{2}xy.$$

$\therefore$ either $x = 0, y = 0, (z = \frac{1}{2}a);$
or $x = 0, z = 0, (y = a);$
or $y = 0, z = 0, (x = a);$
$\Big\}$ $(\lambda = 0).$
or $x = y = 2z = \frac{1}{3}a, \quad (\lambda = a^2/18).$

When there are two conditions, there are two functions f and two multipliers.

16.22. MAXIMA AND MINIMA. When a function $\phi(x, y)$ is known to have a maximum, this can be found by making it a maximum with regard to one variable, keeping the others fixed; having found this maximum, then make ϕ a maximum for variations of the remaining variable.

A more satisfactory examination of maxima and minima is aided by Taylor's theorem. The maxima and minima are amongst the stationary points $x = a, y = b$ where

$$\partial\phi/\partial x = 0, \quad \partial\phi/\partial y = 0.$$

Near such a point, the Taylor series asserts that

$$\begin{aligned} \Delta\phi &= \phi(a + h, b + k) - \phi(a, b), \\ &= \tfrac{1}{2}\left\{h^2 \frac{\partial^2\phi}{\partial x^2} + 2hk \frac{\partial^2\phi}{\partial x\, \partial y} + k^2 \frac{\partial^2\phi}{\partial y^2}\right\} + \ldots \\ &= \tfrac{1}{2}(Ah^2 + 2Hhk + Bk^2) \text{ to second order.} \end{aligned}$$

The first derivatives vanish, and extra symbols have been introduced:

$$A = \frac{\partial^2\phi}{\partial x^2}, \; H = \frac{\partial^2\phi}{\partial x\, \partial y}, \; B = \frac{\partial^2\phi}{\partial y^2}; \text{ at } x = a, y = b.$$

Unless A, B, H are all zero, the higher-order terms are negligible for small h and k. Now

$$Ah^2 + 2Hhk + Bk^2 = (Ah + Bk)^2/A + \{(AB - H^2)/A\}k^2.$$

Hence there is at $x = a, y = b$:

(*i*) Maximum: $\Delta\phi > 0$, all h, k when $A > 0$, $B > 0$, $AB > H^2$;

(*ii*) Minimum: $\Delta\phi < 0$, all h, k when $A < 0$, $B < 0$, $AB > H^2$;

(*iii*) Saddle point: $AB < H^2$.

Near a maximum or minimum the contours are ellipses, as indicated by *Figure 16.02*. The saddle point is a new idea. It arises because $\Delta\phi$ is positive for some values of (h, k) and negative for others. From the stationary point, ϕ increases in some directions and decreases in others. The two lines along which $\Delta\phi = 0$ are given by the equation,

$$2\Delta\phi = Ah^2 + 2Hhk + Bk^2 = 0, \quad (AB < H^2);$$
$$h/k = \{-H \pm \sqrt{(H^2 - AB)}\}/A.$$

These are the 'level' lines.

The neighbouring contours, where $\Delta\phi$ is small, are asymptotic to the level lines and are approximately hyperbolae (*Figure 16.02*). The major and minor axes of these curves bisect the angles between the level lines, and ϕ decreases most rapidly along one and increases most rapidly along the other.

16.23. Repeated Integrals. A repeated integral is the result of integrating a function more than once, not necessarily with respect to the same variable. Consider especially the definite integral,

$$\int_a^b \phi(x, y)\,dx, \quad (y \text{ constant}),$$

where a and b are independent of x. It is in general a function of y, either because ϕ is a function of y or because the limits a and b can be functions of y. Its integral with respect to y, between limits c and d, is a definite integral which is written

$$\int_c^d \int_a^b \phi(x, y)\,dx\,dy, \quad \text{or} \quad \int_c^d dy \int_a^b \phi(x, y)\,dx.$$

If the first is used, and it is preferable for some purposes, the innermost limits and differential refer to the first integration (a, b and dx) and the next set (c, d and dy) to the second. Since a and b need not

be constants it is important to attribute them to the right variable (see the next chapter).

Example 1.

$$\int_0^1 2xy\,dx = \Big[x^2y\Big]_0^1 = y: \quad \int_0^2\int_0^1 2xy\,dx\,dy = \int_0^2 \Big[x^2y\Big]_0^1 dy = \int_0^2 y\,dy = 2.$$

Example 2.

$$\int_0^y 2xy\,dx = \Big[x^2y\Big]_0^y = y^3: \quad \int_0^2\int_0^y 2xy\,dx\,dy = \int_0^2 \Big[x^2y\Big]_0^y dy = \int_0^2 y^3\,dy = 4.$$

EXAMPLES XVI

1. Show that $V = (Ar^n + Br^{-n})\sin(n\theta - \varepsilon)$ satisfies the differential equation,

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} = 0,$$

where A, B, ε and n are arbitrary constants. What limitations are imposed on V if (*i*) it is to be single-valued and (*ii*) finite at the origin?

2. Show that $\phi = 1/r$, where $r^2 = x^2 + y^2 + z^2$, satisfies the equation

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0.$$

Show that the functions $\frac{\partial}{\partial x}\left(\frac{1}{r}\right), \frac{\partial}{\partial y}\left(\frac{1}{r}\right), \frac{\partial}{\partial z}\left(\frac{1}{r}\right)$ and higher derivatives also satisfy this equation.

Calculate these quantities as far as the third derivatives, and show that there are four relations between the nineteen new solutions obtained in this way.

3. Find z, given $\partial^2 z/(\partial x\,\partial y) = \sin x \sin y$, where $\partial z/\partial y = -2\sin y$ when $x = 0$, and $z = 0$ when $y = \pi(n + \frac{1}{2})$, where n is an integer.

4. A uniform beam of length l is clamped at one end and supported at a distance y from it. It carries a load W at a distance x from the clamped end. The deflection at the free end is

$$f = (W/4EI)\{x^2(y - x)(l - y)\}/y.$$

Find its maximum value as x varies and y is kept constant and its maximum value for any value of y, together with the values of x and y for this maximum.

5. A factory has a rectangular floor plan and a flat roof, enclosing a volume of 864,000 ft^3. The front wall costs 3 times, and roof 1·5 times as much per unit area as the other walls. Find the length of frontage and the height for the least cost.

6. A box is tied once each way by a string of length l, excluding knots. Find the maximum volume of the box, using the method of undetermined multipliers.

7. For any substance,

$$\delta Q = TdS = l_v dv + c_v dT = l_p dp + c_p dT.$$

Prove that:

(*i*) $c_p = c_v + l_v(\partial v/\partial T)_p$;

(*ii*) $l_v = T(\partial S/\partial v)_T$, $l_p = T(\partial S/\partial p)_T$;

(*iii*) $c_v = -l_v(\partial v/\partial T)_S$, $c_p = -l_p(\partial p/\partial T)_S$;

(*iv*) $\dfrac{c_p}{c_v} = \dfrac{(\partial p/\partial v)_S}{(\partial p/\partial v)_T}$, using Section 16.10.

8. Prove that, at constant temperature,

$$E_B - E_A = \int_{v_A}^{v_B} \{T(\partial p/\partial T)_v - p\}dv.$$

9. Debye's theory of the specific heat of a solid yields a formula for the thermal energy,

$$E = E_0 + 9N\frac{\hbar}{\nu^3}\int_0^{\nu} \frac{x^3 dx}{(e^{\hbar x/kT} - 1)},$$

where E_0 and ν are constants.

Find the specific heat $c_v = (\partial E/\partial T)$ and prove that

$$E = E_0 + \tfrac{3}{4}RT\left\{\frac{c_v}{3R} + 3\frac{\theta}{T}(e^{\theta/T} - 1)\right\},$$

where $R = Nk$ and $\theta = (\hbar\nu/k)$.

10. Given a differential $\alpha\,dx + \beta\,dy$ and a function $F(x, y)$, such that $\partial F/\partial x = \alpha$, show that either $\partial F/\partial y = \beta$ or $(\beta - \partial F/\partial y)$ is a function $f(y)$ of y alone, if $\partial\alpha/\partial y = \partial\beta/\partial x$. Deduce that $d\phi = \alpha\,dx + \beta\,dy$, if

$$\phi = F(x, y) + \int f(y)dy + C,$$

where C is a constant.

Use this method to find ϕ when $d\phi = (x^2 + y)dx + (y^2 + x)dy$.

11. Prove that $(\partial E/\partial v)_T = T^2(\partial(p/T)/\partial T)_v$ and find a general form for E when $pv = F(T)$, where $F(T)$ is an arbitrary function of the absolute temperature T.

12. Legendre polynomials. Find $G = \partial F/\partial x$ and $H = \partial F/\partial h$, where $F = (1 - 2xh + h^2)^{-1/2}$. Prove that $(x - h)G = hH$ and

$$(1 - xh)G = hF + h^2H.$$

Writing $F = \Sigma h^n P_n(x)$ in these equations and equating powers of h, prove that

$$x\frac{dP_n}{dx} - \frac{dP_{n-1}}{dx} = nP_n, \quad \frac{dP_n}{dx} - x\frac{dP_{n-1}}{dx} = nP_{n-1}.$$

REFERENCE

The author has found a helpful discussion of the subject matter of this chapter in the following book:

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering*. Chapter IV. McGraw-Hill Book Co., New York, 1939.

CHAPTER 17

SURFACE AND VOLUME DISTRIBUTIONS

Introduction—Fluid flow—Electric potential due to a surface charge—Hydrostatic pressure. Properties of areas—Variable pressure—Double integrals—Volume of a solid—Polar coordinates—Separable functions—Integrals over the whole plane—Mass distributions—Volume integrals—Cartesian coordinates—Axial coordinates—Spherical polar coordinates—Velocity space—Solid angles. Distribution in direction—Phase space

17.01. INTRODUCTION. This chapter is concerned with properties of distributions over areas and volumes. In the definitions of total values of, for example, a distribution over an area, it is necessary to divide the area into very small areas dS each centred on a point, and obtain a sum $\Sigma\phi\, dS$ where ϕ is determined by the distribution. In the evaluation of these sums by integration (cf. Chapter 10), it is not necessary that dS should be centred on a point, but only so small that ϕ does not vary over the element of area by more than a first-order small quantity.

17.02. FLUID FLOW. Water flowing slowly along a circular tube of radius a has a velocity,

$$v = v_0(a^2 - r^2)/a^2,$$

at a distance r from the axis. The flow per unit time across a small part dS of a normal section (*Figure 17.01*) is $v\, dS$ and the total flow is

$$Q = \Sigma v\, dS.$$

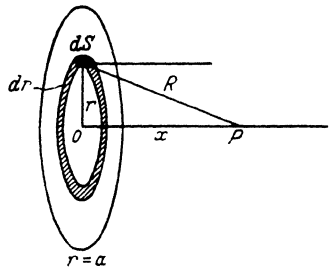

Figure 17.01

As v only depends on r, we can choose as a suitable element of area the ring between two circles of radius r and $(r + dr)$,

$$dS = 2\pi r\, dr.$$

The total flow is obtained by summing over rings from $r = 0$ to $r = a$,

$$Q = \sum_{r=0}^{a} v\, dS = \int_0^a v_0 \frac{a^2 - r^2}{a^2} 2\pi r\, dr = \tfrac{1}{2} v_0 \,.\, \pi a^2.$$

The mean velocity of flow is therefore $\frac{1}{2}v_0$.

17.03. Electric Potential due to a Surface Charge. At a distance R from a point charge q, the electric potential is q/R. Consider the potential due to a charge q distributed uniformly over a circular disc of radius a, i.e. a charge density $\sigma = q/\pi a^2$, per unit area.

At points a long way from the disc, the charge can still be treated as a point charge. At points P nearer the disc some allowance must be made for the variation of R from one point on the disc to another.

Divide the disc into small areas dS (*Figure 17.01*). The charge $\sigma\, dS$ can be regarded as a point charge, with a potential $(\sigma\, dS/R)$ at P. The total potential at P is

$$V = \Sigma(\sigma\, dS/R) = \sigma\Sigma(dS/R).$$

Assume that P is on the axis of the disc, $OP = x$. In the summation an area dS can be chosen so that R is constant over it. Now R is constant for points on the disc on a circle with O as centre. Hence we divide the disc into concentric rings. For the ring between circles of radii r and $(r + dr)$,

$$dS = 2\pi r\, dr, \qquad R = (x^2 + r^2)^{1/2}.$$

$$V = \sigma \int_0^a \frac{2\pi r\, dr}{(x^2 + r^2)^{1/2}} = \frac{2q}{a^2} \Big[(x^2 + r^2)^{1/2} \Big]_{r=0}^{a} = \frac{2q}{a^2} \{(x^2 + a^2)^{1/2} - x\}$$

$$= \frac{2q}{(x^2 + a^2)^{1/2} + x}.$$

Hence $V \to q/x$, as $x \to \infty$, as required at long distances.

17.04. Hydrostatic Pressure. Properties of Areas. The force on one side of a plane vertical area immersed in a liquid is distributed over the area. When the pressure variation is appreciable, the resultant force is obtained by adding the forces on small elements dS. Since the area is plane these forces are parallel. The force on each element is $p\, dS$ normal to the area, so that the total force is

$$F = \Sigma p\, dS.$$

According to the laws of statics, the resultant force acts at a point C, the centre of pressure, so that its moment about any axis is the total

moment of the distributed forces. Choosing horizontal and vertical axes Ox and Oy as in *Figure 17.02*,

$$Fx_c = \Sigma\, xp\, dS, \quad Fy_c = \Sigma\, yp\, dS.$$

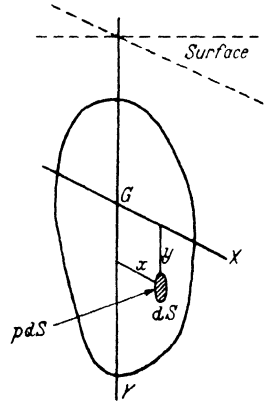

Figure 17.02

Let us first assume that the pressure is in fact uniform. In practice this is very often true as an approximation. The total thrust is

$$F = pS,$$

where S is the area; the coordinates of G, the centre of pressure, are $x_c = x_g, y_c = y_g$ [*Figure 17.03* (a)],

$$Sx_g = \Sigma x\, dS,\ Sy_g = \Sigma y\, dS.$$

The pressure p does not appear in these expressions.

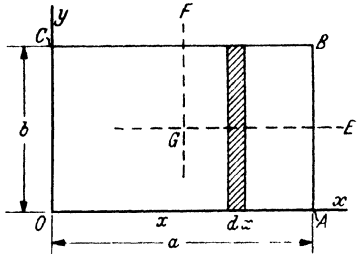

Figure 17.03 (a)

This point G is the mean centre of the area, and only depends on its shape. Its coordinates x_g and y_g are the average values of x and y over the area. The two sums $\Sigma x\, dS$ and $\Sigma y\, dS$ are called the first

moments of the area about the y- and x-axes, respectively, and it can easily be deduced that their values are known at once when G is located.

That the position of G is unique is a consequence of the laws of statics, but it can be proved directly. The moment of the resultant thrust on the area about a line LL is $\Sigma\, zp\, dS$, where z is the distance of the area dS from the line. Now in *Figure 17.03* (b)

$$z = x \sin\alpha + y \cos\alpha + d.$$

Hence

$$\begin{aligned}\Sigma pz\, dS &= p \sin\alpha\, \Sigma x\, dS + p \cos\alpha\, \Sigma y\, dS + pd\, \Sigma\, dS,\\ &= pS(x_g \sin\alpha + y_g \cos\alpha + d),\\ &= Fz_g,\end{aligned}$$

where z_g is the distance of G from the line. Thus a force F acting at G has the correct moment about any line.

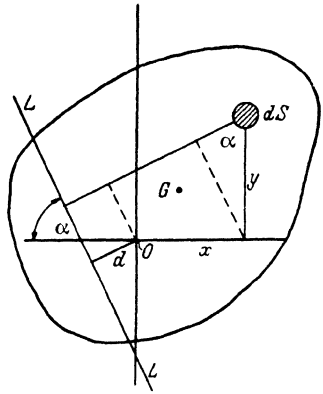

Figure 17.03 (b)

The position of G is often easily determined without long calculations. If the area is symmetric about a line LL then G lies on that line, since it is obvious that the moment of the forces due to a uniform pressure must be zero about it. (The same result can be proved more formally using the sum and reversing the direction in which z is measured positive.) Thus the position of G is partially determined when the area is symmetric about one line, and its position is completely determined when it is symmetric about two distinct lines, as in the case of a circle or rectangle or any regular polygon.

Example 1. Prove that the mean centre of a triangle lies on a median, by dividing the triangle into strips parallel to one side and

considering the forces on each strip. Deduce the position of the mean centre.

A useful point to remember is that, for any area, $\Sigma x\, dS = 0$ if the y axis passes through G, and that $\Sigma y\, dS = 0$ if the x axis passes through G.

17.05. VARIABLE PRESSURE. In many problems the liquid pressure varies with depth below the surface and this variation must be allowed for in the calculation. Since it does only depend on depth, however, the calculation of the total force F and the value of y_c for the centre of pressure can be made, dividing the area into strips parallel to the surface. The calculation of x_c is more difficult. However, it is zero for any area with a vertical axis of symmetry, which is chosen as the y axis, by the same argument as was used to fix the position of the mean centre in the last paragraph.

For a liquid of uniform density ρ, the pressure increases linearly with depth. Let us assume that the mean centre has been found and that axes Gx and Gy are drawn through it. Then the pressure at a depth y below G is $p = p_g + \rho g y$. The total force on the area is

$$F = \Sigma(p_g + \rho g y)dS = p_g S + \rho g y_g = p_g S$$

since $y_g = 0$, i.e. the total force is the area times the pressure at the mean centre of the area.

The coordinates of the centre of pressure are:

$$Fx_c = \Sigma x(p_g + \rho g y)dS = \rho g H,$$

where $H = \Sigma\, xy\, dS$, and

$$Fy_c = \Sigma y(p_g + \rho g y)dS = \rho g B,$$

where $B = \Sigma y^2 dS$.

The second sum $\Sigma y^2 dS$ and the similar sum $\Sigma\, x^2 dS$ are known as the second moments (or moments of inertia) of the area about the axes. When divided by the area they give the mean square values of y and x over the area. The other sum H is called the mixed second moment, and is zero, as noted above, for areas with a vertical axis of symmetry.

17.06. DOUBLE INTEGRALS. In the foregoing sections we encountered, but did not evaluate, integrals of the form

$$\Sigma\, \phi\, dS,$$

where dS was a small area near a point P and ϕ was a function of the coordinates (x, y) of P. Such sums lead to double integrals in general.

To carry out the summation it is first necessary to divide the area in an orderly fashion.

Using Cartesian coordinates this is most conveniently done by means of a network of lines normal to the x axis at intervals dx and lines normal to the y axis at intervals dy. The whole plane is divided into small rectangles and the element of area is one of them.

$$dS = dx\,dy, \qquad \text{(Cartesian coordinates)}.$$

Any area S (*Figure 17.04*) contains a large number of them. The sum $\Sigma\phi\ dS$ over these rectangles differs from the sum $\Sigma\phi\ dS$ over the whole area by the omission of small areas near the boundary. In

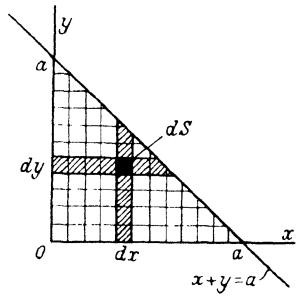

Figure 17.04

the limit, all dx, $dy \to 0$, the sum of the areas omitted normally tends to zero and with it the contribution to the sum. Hence we write

$$\lim_{dx,\,dy\to 0} \Sigma\phi\ dS = \int\int_S \phi(x, y)\,dx\,dy.$$

This is a double integral, evaluated by repeated integration, as explained below. Once this stage is reached the method of evaluation depends not so much on the function $\phi(x, y)$ as on the shape of the area.

Example 1. Calculate the mixed second moment,

$$H = \Sigma\, xy\ dS, \quad (\phi = xy),$$

for the area bounded by the x axis, the y axis and the line $x + y = a$ [*Figure 17.04*. The sum over the rectangle must be carried out in an orderly fashion. Let us first sum elements in rows parallel to

the x axis. The row between y and $(y + dy)$ (marked in the diagram) extends from $x = 0$ to $x = a - y$, where it meets the line $x + y = a$.

$$\sum_{row} \phi \, dS = \sum_{x=0}^{a-y} \phi \, dx \, dy \to \int_0^{a-y} \phi \, dx \, dy, \quad (y \text{ and } dy \text{ constant}).$$

This holds for every row from $y = 0$ up to the corner $y = a$. Summing the results for every row,

$$\sum_S \phi \, dS = \sum_{y=0}^{a} \left(\sum_{x=0}^{a-y} \phi \, dx \right) dy = \int_0^a \int_0^{a-y} \phi \, dx \, dy.$$

The inner limits refer to the first integration dx, the outer limits to the second integration dy. Since $\phi = xy$,

$$\int_0^a \int_0^{a-y} xy \, dx \, dy = \int_0^a \left[\tfrac{1}{2} x^2 y \right]_0^{a-y} dy = \int_0^a \tfrac{1}{2}(a - y)^2 y \, dy = \tfrac{1}{24} a^4.$$

If the integrals are summed first in rows parallel to the y axis, the row between x and $x + dx$ extends from $y = 0$ to $y = a - x$, giving

$$\left(\sum_{y=0}^{a-x} \phi \, dy \right) dx = \int_0^{a-x} \phi \, dy \, dx, \quad (x \text{ and } dx \text{ constant}),$$

and the sum over rows, from $x = 0$ to $x = a$,

$$\sum_S \phi \, dS = \int_0^a \int_0^{a-x} \phi \, dy \, dx, \quad (y \text{ first, then } x).$$

Although the limits have changed with the order of integration, the integral has the same value as before, whatever the function ϕ, as long as the area of integration is the same.

Example 2. If the area of integration is a rectangle enclosed by the lines $x = a, x = b, y = c, y = d$, the limits remain the same when the order of integration is changed.

$$\sum \phi \, dS = \int_c^d \int_a^b \phi(x, y) dx \, dy = \int_a^b \int_c^d \phi(x, y) dy \, dx.$$

17.07. Volume of a Solid. The double integral of $U = \phi(x, y)$ over an area equals the volume between the upper surface and the base of a solid model representing the function, and vice versa.

To define the volume more closely, let us assume it to be a vertical cylinder (*Figure 17.05*) of base area S in the base plane. The area S is that between two curves $y = u(x)$ and $y = v(x)$, extending from $x = a$ to $x = b$.

The volume is divided into slices by vertical planes normal to the x axis, at intervals dx. If $A(x)$ is the area of the section at x, the

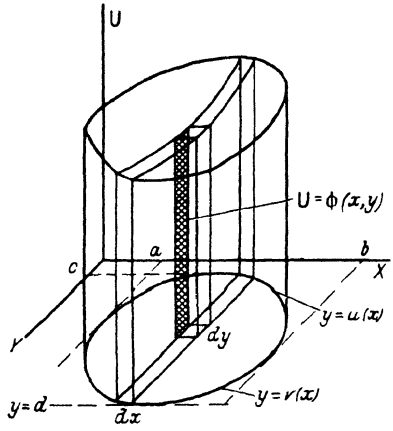

Figure 17.05

volume of the slice between x and $(x + dx)$ is $A(x)dx$, to first order, as in Chapter 10, so that the total volume is

$$\sum_{x=a}^{b} A(x)dx = \int_a^b A(x)dx.$$

Now the section $A(x)$ is an area with a line as base from $y = u(x)$ to $y = v(x)$, the height of which at every point from the base is $U(x, y)$, where x is fixed and y varies along the *base* line. Dividing this area into strips by vertical lines at intervals dy, its area by integration is

$$A(x) = \int_{u(x)}^{v(x)} \phi(x, y)dy, \quad (x \text{ constant}).$$

Substituting this formula in the earlier expression, the volume is

$$\int_a^b \left\{ \int_{u(x)}^{v(x)} \phi(x, y)dy \right\} dx = \int_a^b \int_{u(x)}^{v(x)} \phi(x, y)dy\, dx = \iint_S \phi\, dS.$$

Example 1. (The reader should draw this to make it clearer.) Find the volume between the planes $x + y + z = 1$ and the coordinate planes $x = 0, y = 0, z = 0$. This is a tetrahedron.

The plane $x + y + z = 1$ intersects the base plane $z = 0$ in a line $x + y = 1$, so that the base area S is triangular, and lies between $y = 0$ and $y = 1 - x$, from $x = 0$ to $x = 1$. At any point in this area the height of the tetrahedron is $z = 1 - x - y$.

The area of the section normal to the x axis between x and $(x + dx)$ is

$$A(x) = \int_0^{1-x} z\, dy.$$

The total volume $V = \Sigma A\, dx$ is

$$V = \int_0^1 \int_0^{1-x} (1 - x - y)dy\, dx = \int_0^1 \Big[-\tfrac{1}{2}(1 - x - y)^2 \Big]_0^{1-x} dx,$$

$$= \int_0^1 \tfrac{1}{2}(1 - x)^2 dx = \tfrac{1}{6}.$$

17.08. Polar Coordinates. It is neither necessary nor advisable always to divide an area into rectangles using Cartesian coordinates. Other divisions using curvilinear coordinates are possible.

In polar coordinates, the plane is divided by lines drawn radially from the origin (θ = constant) at intervals $d\theta$, and by circles around the origin (r = constant) at intervals dr. The former divide it into sectors and the latter into circular strips (*Figure 17.06*). The two together give small areas with two straight sides of length dr and two curved sides of lengths $r\, d\theta$ and $(r + dr)d\theta$. To first order the latter are equal and the element of area is a small rectangle with

$$dS = r\, d\theta\, dr, \qquad \text{(Polar coordinates).}$$

Then $\Sigma\phi\, dS$ becomes

$$\Sigma\, \phi\, r\, d\theta\, dr.$$

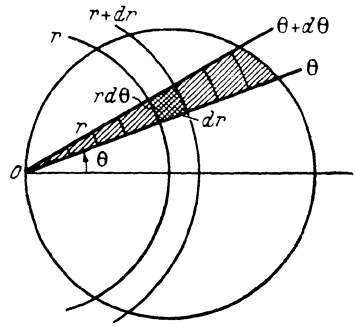

Figure 17.06

Example 1. The first moment of the circle $r = 2a \cos\theta$ about the tangent at the origin (*Figure 17.06*) is

$$\Sigma\, x\, dS = \Sigma\, r \cos\theta \,.\, r\, d\theta\, dr.$$

Keeping θ and $d\theta$ fixed and summing radially from the origin ($r = 0$) to the circle ($r = 2a\cos\theta$), gives the first moment of a sector of the area. Then adding up for different values of θ from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$ covers the whole area and gives the first moment of the circle.

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi}\int_0^{2a\cos\theta} r^2\cos\theta\, dr\, d\theta,$$

$$= \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \tfrac{1}{3}(2a\cos\theta)^3\cos\theta\, d\theta = \tfrac{8}{3}\cdot 2\int_0^{\frac{1}{2}\pi}\cos^4\theta\, d\theta,$$

$$= \pi a^3, \quad \text{(using I.157)}.$$

N.B. The double integral of $\phi(r, \theta)$ with respect to r and θ means the integral $\int\int\phi(r, \theta)d\theta\, dr$. The integral over an area $\int\int\phi(r, \theta)r\, d\theta\, dr$ is therefore the double integral of $r\phi(r, \theta)$.

17.09. Separable Functions. When a function $\phi(x, y)$ is separable in the variables, it can be written in the form

$$\phi(x, y) = f(x)\cdot g(y).$$

When a repeated integral has a function of this sort as integrand, and constant limits,

$$\int_c^d\int_a^b \phi(x, y)dx\, dy = \int_c^d \left\{\int_a^b f(x)g(y)dx\right\}_{y\text{ const.}} dy,$$

$$= \int_c^d g(y)\left\{\int_a^b f(x)dx\right\} dy = \int_c^d g(y)dy\cdot\int_a^b f(x)dx,$$

i.e. the integral can be factorized.

Similarly, when $\phi(r, \theta) = f(r)\cdot g(\theta)$, with constant limits,

$$\int_0^a\int_0^{2\pi} f(r)g(\theta)r\, d\theta\, dr = \int_0^a f(r)r\, dr\cdot\int_0^{2\pi} g(\theta)d\theta,$$

over a circle $r = a$.

17.10. Integrals over the Whole Plane. An integral over the whole plane means an integral over a large area S the perimeter of which tends to infinity in all directions. In Cartesian coordinates the area chosen for this purpose would be a rectangle, and in polar coordinates the area chosen would be a circle, but the integral only

has a meaning if the answers are the same whatever the shape of the area.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(x,y)\,dx\,dy \text{ (Cartesian)}: \quad \int_0^{\infty}\int_0^{2\pi}\phi(r,\theta)r\,d\theta\,dr \text{ (polar)}.$$

Example 1. Over the whole plane the integral of

$$\phi = e^{-r^2} = e^{-(x^2+y^2)},$$

is: (Polar coordinates)

$$\int_0^{\infty}\int_0^{2\pi} e^{-r^2} r\,d\theta\,dr = \int_0^{\infty} e^{-r^2} 2\pi r\,dr = 2\pi\left[-\tfrac{1}{2}e^{-r^2}\right]_0^{\infty} = \pi.$$

(Cartesian coordinates)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2-y^2}\,dx\,dy = \int_{-\infty}^{\infty} e^{-x^2}\,dx \,.\int_{-\infty}^{\infty} e^{-y^2}\,dy = A^2,$$

where $A = \int_{-\infty}^{\infty} e^{-x^2}\,dx$. Since the integrals are equal, $A = \sqrt{\pi}$.

17.11. MASS DISTRIBUTIONS. A mass dm of small dimensions at a distance z from a plane is said to have a first moment $z\,dm$ and a second moment $z^2\,dm$ about the plane. If it is a distance r from a line it has a moment of inertia $r^2\,dm$ about the line.

To define these quantities for a large body of density ρ, which need not be uniform, although it often is, we divide the body into small volumes. If dv is a small volume near a point P, where the density is ρ, then its mass is $dm = \rho\,dv$. By summing the moments of these small volumes the moments for the whole body are obtained, e.g. the first moment about the plane is $\Sigma\, z\rho\,dv$. The distance of the mass centre from this plane is defined as z_g where mz_g is the first moment about the plane.

Example 1. The density of a right circular cone is proportional to the distance z along the axis from the vertex, and reaches a maximum ρ_0 at the base, which is a distance h from the vertex O and has a radius a (*Figure 17.07*). Its volume is $\frac{1}{3}\pi a^2 h = V$.

The density at z is $\rho = \rho_0 z/h$: the mass of a volume element dv is ρdv. Hence the total mass is $\Sigma\rho_0 z dv/h$.

Divide the volume into slices over which ρ is constant. The slice between z and $(z + dz)$ has a radius (az/h) (why?), so that its area is $\pi a^2 z^2/h^2$ and its volume is

$$dv = \pi a^2 z^2 dz/h^2.$$

Hence

$$m = \Sigma\rho\, dv = \int_0^h \rho_0 \frac{z}{h} \cdot \frac{\pi a^2 z^2}{h^2}\, dz = \frac{\pi}{4}\rho_0 a^2 h = \tfrac{3}{4}\rho_0 V.$$

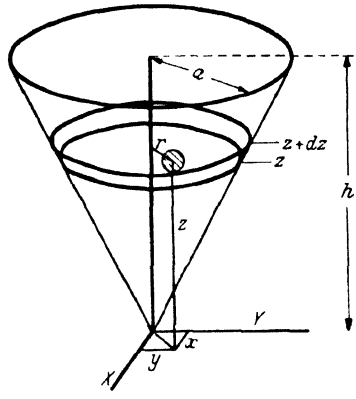

Figure 17.07

The mass centre G lies on the axis. Its distance from O is z_g, where

$$mz_g = \Sigma z\rho\, dv = \int_0^h \rho_0 \frac{z^2}{h} \cdot \frac{\pi a^2 z^2}{h^2}\, dz = \tfrac{1}{5}\rho_0 a^2 h^2 = \tfrac{4}{5}mh.$$

Thus $z_g = 4h/5$.

17.12. VOLUME INTEGRALS. If F is any quantity with a value at every point of a volume, then the volume integral of F over the volume is the sum, over all parts of the volume, of the products $F\,dv$ of the volume dv of each small part and the value of F there. The volume Sdv must be so small that F is constant over it to first order. We write

$$\sum_V F\, dv \to \int\int\int_V F\, dv.$$

This integral, divided by the volume V, is the mean value of F over the volume. Sometimes, of course, as in the previous section, the volume integral itself has a definite meaning.

The evaluation of these integrals depends on the shape of the volume and the form of the function F as a function of position. There are three types of coordinate system in fairly common use which we now describe.

17.13. CARTESIAN COORDINATES. Using Cartesian coordinates with perpendicular axes OX, OY, OZ, a volume is divided into slices by planes x = constant, at intervals dx; these are divided into rods by another set of planes, y = constant, at intervals dy, and finally into

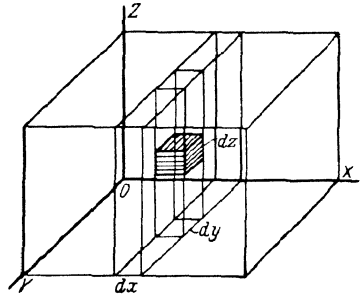

Figure 17.08

small rectangular blocks by planes at intervals dz (*Figure 17.08*). The volume element is

$$dv = dx\,dy\,dz, \qquad \text{(Cartesian coordinates)}$$

and

$$\Sigma F\,dv = \iiint_V F(x, y, z)\,dx\,dy\,dz.$$

The summation proceeds as in earlier sections over first one coordinate, then another and finally the third, inserting the appropriate limits on the way which correspond to the given volume.

Example 1. For the rectangular block with sides a, b and c (*Figure 17.08*) we can sum first over small blocks with different z but the same x and y. This gives a rod extending from $z = 0$ to $z = c$. The value for this rod is

$$\sum_{z=0}^{c} F\,dx\,dy\,dz = dx\,dy \int_0^c F(x, y, z)\,dz.$$

We now sum over rods with different values of y but the same value of x and dx to form a slice. The value for the slice is

$$\sum_{y=0}^{b} \left(\sum_{z=0}^{c} F\,dz \right) dy\,dx = dx \int_0^b \int_0^c F(x, y, z)\,dz\,dy.$$

Finally by a summation over slices, the value for the whole block is obtained

$$\int_0^a \int_0^b \int_0^c F(x, y, z)\,dz\,dy\,dx.$$

For example, if the block is of material of uniform density ρ, the moment of inertia of a volume element about the z axis is

$$dI = (x^2 + y^2)\rho\, dv.$$

The total moment of inertia is

$$\begin{aligned}\Sigma(x^2 + y^2)\rho\, dv &= \int_0^a \int_0^b \int_0^c \rho(x^2 + y^2)dz\, dy\, dx,\\ &= \rho\int_0^a x^2\, dx \int_0^b dy \int_0^c dz + \rho\int_0^a dx \int_0^b y^2\, dy \int_0^c dz,\\ &= \rho \,.\, \tfrac{1}{3}a^3 \,.\, b \,.\, c + \rho \,.\, a \,.\, \tfrac{1}{3}b^3 \,.\, c,\\ &= \tfrac{1}{3}(a^2 + b^2)\rho abc = \tfrac{1}{3}(a^2 + b^2)m.\end{aligned}$$

Note that the integral with constant limits factorizes when F does.

17.14. AXIAL COORDINATES. These are more suitable for cylinders, or volume integrals of functions with axial symmetry.

The coordinate z is measured along an axis as before (*Figure 17.09*) but polar coordinates (r, ϕ) are used in the plane normal to this axis: r is the distance from the axis, and ϕ the azimuthal angle about it.

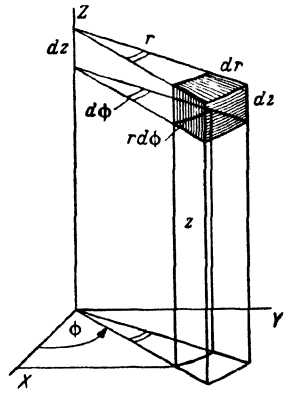

Figure 17.09

(*i*) Planes z = const. at intervals dz divide space into slices.
(*ii*) Planes ϕ = const. through the axis at intervals $d\phi$ divide space into sectors.
(*iii*) Cylinders r = const. at intervals dr divide it into thin cylindrical shells.

Combining these into pairs we can get various volume elements. Using all three gives a small block near P of volume

$$dv = r\,d\phi\,.\,dr\,.\,dz. \qquad \text{(Axial coordinates).}$$

Example 1. A mass of gas m is in a circular cylinder of radius a and length L, rotating about its axis with angular velocity ω. The pressure in the gas satisfies the equation $dp/dr = \rho\omega^2 r$ and for an isothermal gas $p = \rho RT/M$, this integrates to $\rho = \rho_0 e^{kr^2}$, where $k = (M\omega^2/2RT)$ and ρ_0 is a constant (the density at the axis). To express ρ_0 in terms of m we use the formula $m = \Sigma\rho\,dv$. As ρ is a function of r, we divide the cylinder into cylindrical shells. The shell between r and $(r + dr)$ has a volume

$$dv = 2\pi r\,dr\,.\,L.$$

Hence $$m = \sum_{r=0}^{a} \rho\,dv = \int_0^a \rho_0 e^{kr^2} 2\pi r\,dr\,.\,L = \pi\rho_0 L\left[e^{kr^2}/k\right]_0^a.$$

$$\therefore \qquad \rho_0 = mk/\pi L(e^{ka^2} - 1)$$
$$= mM\omega^2/2\pi LRT(e^{Ma^2\omega^2/2RT} - 1).$$

Example 2. The electric charge density in an infinite cylinder of radius a from $z = 0$ to $z = \infty$, is $\rho = Ae^{-kz}r^2\cos^2\phi$. The total charge is

$$q = \Sigma\rho\,dv = \Sigma\rho\,r\,d\phi\,dr\,dz.$$

The order of summation does not matter much. For given r and z the cylinder contains volume elements from $\phi = 0$ all the way round to $\phi = 2\pi$, forming a ring of rectangular section $dr\,dz$. The charge in this ring is

$$r\,dr\,dz\int_0^{2\pi} \rho\,d\phi, \quad (r, z \text{ constant}).$$

Summing over rings with the same z but different radii r gives the charge in a slice between z and $(z + dz)$,

$$dz\int_0^a \int_0^{2\pi} \rho r\,d\phi\,dr, \quad (z \text{ and } dz \text{ constant}).$$

Summing from $z = 0$ to $z = \infty$ gives the total charge,

$$\int_0^\infty \int_0^a \int_0^{2\pi} Ae^{-kz}r^2\cos^2\phi\,.\,r\,d\phi\,dr\,dz,$$
$$= A\int_0^\infty e^{-kz}\,dz\int_0^a r^3\,dr\int_0^{2\pi}\cos^2\phi\,d\phi = \pi Aa^4/4k.$$

17.15. SPHERICAL POLAR COORDINATES (R, θ, ϕ). These are used for problems with a natural centre, or involving spheres.

R is the distance of a point P from the origin O.

An axis OZ is chosen in a suitable direction when necessary and $\theta = \angle ZOP$.

The angle ϕ is the azimuthal angle about the axis OZ.

$$0 \leqslant R < \infty; \quad 0 \leqslant \theta \leqslant \pi; \quad 0 \leqslant \phi < 2\pi.$$

The angles θ, ϕ give the direction of OP (comparing with axial coordinates $z = R\cos\theta, r = R\sin\theta$). $(\frac{1}{2}\pi - \theta)$ is the latitude, and θ is sometimes called the co-latitude on a sphere.

Two spheres of radii R and $(R + dR)$ enclose a spherical shell of thickness dR and volume,

$$dv_R = 4\pi R^2\, dR.$$

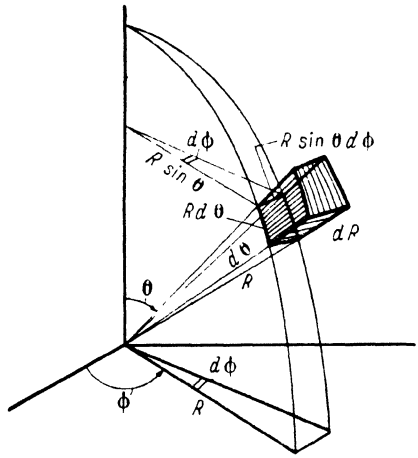

Figure 17.10

Cones of semi-vertical angle θ at intervals $d\theta$ divide this shell into zones of volume

$$dv = 2\pi R\sin\theta \,.\, Rd\theta \,.\, dR$$

$$= 2\pi R^2 \sin\theta\, d\theta\, dR.$$

Further division by planes through the axis, e.g. between ϕ and $\phi + d\phi$, divide this into small blocks of volume

$$dv = R^2 \sin\theta\, d\phi\, d\theta\, dR.$$

Example 1. An electric charge distribution of density ρ has an electric dipole moment with a component in the z direction

$$\mu_z = \Sigma\, \rho z\, dv = \Sigma\, \rho R \cos\theta\, dv,$$

relative to a given point O. If the charge extends over the whole of space and $\rho = \rho_0 e^{-kR}\cos\theta$, this is

$$\mu_z = \Sigma \rho_0 e^{-kR}\cos\theta \,.\, R\cos\theta\, dv = \iiint \rho_0 e^{-kR} R\cos^2\theta (R^2 \sin\theta\, d\phi\, d\theta\, dR).$$

In an integration over all space, all values of the coordinates are possible and

$$\begin{aligned}
\mu_z &= \int_0^\infty \int_0^\pi \int_0^{2\pi} \rho_0 e^{-kR} \,.\, R^3 \cos^2\theta \sin\theta\, d\phi\, d\theta\, dR, \\
&= \int_0^\infty \int_0^\pi 2\pi\rho_0 e^{-kR} \,.\, R^3 \cos^2\theta \sin\theta\, d\theta\, dR, \quad \left(\int_0^{2\pi} d\phi = 2\pi\right), \\
&= \int_0^\infty (4\pi\rho_0/3) e^{-kR} R^3\, dR, \quad \left(\int_0^\pi \cos^2\theta \sin\theta\, d\theta = 2/3\right), \\
&= 8\pi\rho_0/k^4.
\end{aligned}$$

Example 2. If the total charge density is $\rho = \rho_0 e^{-kR}/kR$ at a distance R from O the total charge can be evaluated using the volume element,

$$dv = 4\pi R^2\, dR,$$

between spheres of radii R and $R + dR$, since ρ is constant over the volume. Hence the charge inside a sphere of radius a is

$$q_a = \int_0^a (\rho_0 e^{-kR}/kR) 4\pi R^2\, dR = \frac{4\pi\rho_0}{k^3}\{1 - e^{-ka}(1 + ka)\}.$$

The total charge is

$$q = \lim_{a\to\infty} q_a = 4\pi\rho_0/k^3.$$

An interesting point is that the total charge in the whole of space is finite, even though ρ diverges at the origin $R = 0$. The reason is that the region in which ρ is very large is only a very minute one, as shown by the fact that $dv = 4\pi R^2\, dR \to 0$ as R decreases, for a given thickness dR.

17.16. VELOCITY SPACE. Velocity space is the name given to a three-dimensional diagram where the coordinates represent the velocity of a body. A body with velocity components (v_x, v_y, v_z) in

the x, y and z directions is represented by a point P with these three components as Cartesian coordinates. Remembering that distances now represent velocity components we can use the previous diagrams (*Figures 17.08–10*).

For example, Maxwell's kinetic theory of gases states that the number of molecules with velocity components between (v_x, v_y, v_z) and $(v_x + dv_x, v_y + dv_y, v_z + dv_z)$ is

$$dN = Ae^{-\alpha(v_x^2+v_y^2+v_z^2)}\, dv_x\, dv_y\, dv_z, \quad (\alpha = m/2kT).$$

Every one of these molecules is represented by a point in velocity space, in the small element of volume $d\tau = dv_x\, dv_y\, dv_z$, near the point P, i.e. if

$$dN = f d\tau,$$

where $f = Ae^{-\alpha(v_x^2+v_y^2+v_z^2)}$, then f is the number of points per unit volume, or the density of points in velocity space.

The total number of molecules (points) is

$$\begin{aligned} N = \Sigma f\, d\tau &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Ae^{-\alpha(v_x^2+v_y^2+v_z^2)}\, dv_x\, dv_y\, dv_z, \\ &= A\left[\int_{-\infty}^{\infty} e^{-\alpha v_x^2}\, dx\right]^3, \\ &= A(\pi/\alpha)^{3/2} = A(2\pi kT/m)^{3/2}, \quad \text{(cf. Section 17.09).} \end{aligned}$$

If we do not carry out the final integration over v_x, but only over v_y and v_z, the answer is the number of molecules with a velocity in the x direction between v_x and $v_x + dv_x$, whatever their other velocity components.

$$dN_x = Ae^{-\alpha v_x^2}\, dv_x(\pi/\alpha) = N(\alpha/\pi)^{1/2}e^{-\alpha v_x^2}\, dv_x.$$

i.e. the fraction is $dN_x/N = (\alpha/\pi)^{1/2}e^{-\alpha v_x^2}\, dv_x$.

The mean value of any function of the velocity, $\phi(v_x, v_y, v_z)$, is defined as the sum of its values for all the molecules, divided by N. For points in a volume element $d\tau$ the sum of the values is just $(\phi f\, d\tau)$ to first order, and the mean value is therefore

$$\frac{1}{N}\int\int\int \phi f\, d\tau.$$

Example 1. The mean square velocity in the x direction is

$$\frac{1}{N}\int\int\int v_x^2 f\, d\tau = \frac{1}{2\alpha} = \frac{kT}{m}.$$

It is not necessary to use Cartesian coordinates in velocity space. The line OP (*Figure 17.10*) represents the velocity in magnitude and direction. If spherical polar coordinates are used (q, θ, ϕ), then $q = (v_x^2 + v_y^2 + v_z^2)^{1/2}$ is the speed and the angles (θ, ϕ) give the direction. The density of points is unchanged by a change of coordinates, and is therefore, in spherical polar coordinates

$$f = Ae^{-q^2},$$

which is independent of direction. Using the volume element $d\tau = q^2 \sin\theta \, d\theta \, d\phi \, dq$ in polar coordinates (cf. Section 17.15) we find that

$$dN = f \, d\tau = Ae^{-\alpha q^2} \cdot q^2 \sin\theta \, d\phi \, d\theta \, dq,$$

is the number of molecules (points) with speeds between q and $(q + dq)$ travelling in directions between (θ, ϕ) and $(\theta + d\theta, \phi + d\phi)$. By integration over ϕ we obtain the number moving in directions making angles between θ and $\theta + d\theta$ with the z axis;

$$dN_\theta = 2\pi Ae^{-\alpha q^2} q^2 \sin\theta \, d\theta \, dq.$$

Finally let us find the velocity distribution function $F(q)$ such that the number of molecules with speeds between q and $(q + dq)$ is

$$dN_q = F(q)dq.$$

In velocity space dN_q is the number of points between two spheres of radii q and $(q + dq)$. Now the volume between the two spheres is $d\tau = 4\pi q^2 \, dq$ and the density of points is the same throughout the volume. Hence the number of points is

$$dN_q = Ae^{-\alpha q^2} \cdot 4\pi q^2 \, dq, \quad \text{i.e. } F(q) = 4\pi A q^2 e^{-\alpha q^2}.$$

It can easily be verified that $\int_0^\infty F(q)dq = N$.

17.17. Solid Angles. Distribution in Direction. Consider a number of molecules, or particles, emitted from a point, or leaving a very small volume. Their directions are represented by lines drawn from the point. Consider those lying within a small cone with the point as vertex (*Figure 17.11*).

The spread of the directions is measured by the area cut by the cone from a unit sphere with O as centre. This area is called the solid angle subtended at O by the cone. A small solid angle is denoted by $d\Omega$. Given enough particles, the number with directions lying within the cone is

$$dn = f \, d\Omega,$$

where f is a function giving the distribution of the particles, and its value depends on the direction of the axis of the cone. By summing over cones in all directions we obtain the total number of particles being emitted $n = \Sigma f\, d\Omega$. (The sum of the values of $d\Omega$ is 4π, the area of the unit sphere.)

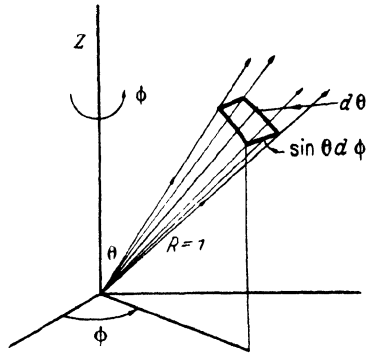

Figure 17.11

Direction is normally specified by the polar angles (θ, ϕ) measured relative to a convenient axis OZ. From the previous section the area of a unit sphere cut off by the cone limited by the directions $\theta \to (\theta + d\theta)$, $\phi \to (\phi + d\phi)$ is

$$d\Omega = \sin\theta\, d\phi\, d\theta, \quad \text{i.e. } dn = f(\theta, \phi) \sin\theta\, d\phi\, d\theta.$$

If this number is integrated over all values of ϕ, we obtain

$$dn_\theta = \int_0^{2\pi} f(\theta, \phi) d\phi \,.\, \sin\theta\, d\theta.$$

This is the number emitted in directions at angles between θ and $(\theta + d\theta)$ with the axis, irrespective of their ϕ values. It is usually written $dn_\theta = F(\theta) d\Omega_\theta$,

where $F(\theta) = \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} f(\theta, \phi) d\phi$ and $d\Omega_\theta = 2\pi \sin\theta\, d\theta.$

$F(\theta)$ is the mean value of $f(\theta, \phi)$ over all values of ϕ, and $d\Omega_\theta$ is the solid angle between θ and $(\theta + d\theta)$.

The total number emitted is

$$n = \int_0^{\pi} F(\theta)\, 2\pi \sin\theta\, d\theta.$$

17.18. PHASE SPACE. In general, as a particle moves along a line, its coordinate x and its momentum $p = mv$ both change. The motion can be illustrated by a graph of p against x, using the time as parameter along the curve.

Since $v = (dx/dt)$, the slope of the curve is

$$\frac{dp}{dx} = \frac{dp/dt}{dx/dt} = \frac{1}{v}\frac{d}{dt}(mv) = \frac{mF}{p},$$

where F is the force on the particle.

Example 1. For simple harmonic motion, where $F = -kx$, the curves are the ellipses $(p^2/2m) + \frac{1}{2}kx^2 = \text{constant} = E$.

The two-dimensional space in which the curve is drawn is called the phase space of the motion. The motion is described by a curve in phase space.

Given a large number of particles moving along the same line, or parallel lines, each has a motion described by a curve in phase space. At any instant they are all represented by points in phase space. If there are enough of them, the distribution can be described by a density function $f(p, x)$, where $dN = f(p, x)dS$ gives the number of points on an area dS. For example $f(p, x)\, dxdp$ gives the number between x and $(x + dx)$ with a momentum between p and $(p + dp)$.

Let all the particles be acted on by the same force. If we follow a group of neighbouring particles, initially occupying a region $dS = dx\, dp$ of phase space, it is found that the laws of mechanics are such that they continue to occupy the same area of phase space, when the force is conservative, although the shape of the area is changing.

There are extensions to this law and for this reason phase space plays an important part in statistical mechanics, which deals with the motion of a large number of particles.

A particle moving in space with coordinates (x, y, z) and momentum components (p_x, p_y, p_z) has a six-dimensional phase space with a volume element

$$d\tau = dx\, dy\, dz\, dp_x\, dp_y\, dp_z,$$

or

$$d\tau = R^2 \sin\theta\, d\theta\, d\phi\, dR \,.\, p^2 \sin\theta'\, d\theta'\, d\phi'\, dp,$$

using polar coordinates (R, θ, ϕ) for position and (p, θ', ϕ') for momentum.

EXAMPLES XVII

1. In classical statistical mechanics the partition function for rotational, motion of a rigid diatomic molecule (rigid rotator) is

$$f = \frac{2\pi}{\hbar^2\sigma}\sqrt{(2\pi IkT)}\int_0^{\pi}\int_{-\infty}^{+\infty} e^{-p^2/(2IkT\sin^2\theta)}\,dp\,d\theta,$$

where $\sigma = 2$ for a symmetric molecule and $\sigma = 1$ for an asymmetric molecule. Prove that $f = (8\pi^2 IkT/\hbar\sigma)$.

2. Evaluate the integral,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\alpha(x^2-xy+y^2)}\,dy\,dx.$$

Evaluate it by inserting $x = x'\cos\theta + y'\sin\theta, y = -x'\sin\theta + y'\cos\theta$ and choosing θ so that $x^2 - xy + y^2 = \alpha' x'^2 + \beta' y'^2$ where α' and β' have to be determined.

$$\left[\text{Hint: } \int_{-\infty}^{+\infty} e^{-(y-a)^2}\,dy = \int_{-\infty}^{+\infty} e^{-y^2}\,dy = \sqrt{\pi}.\right]$$

3. Prove that the moment of inertia of a uniform sphere of radius a and density ρ about the z axis ($\theta = 0$), in spherical polar coordinates with the origin at the centre of the sphere, is

$$\int_0^a\int_0^{\pi}\int_0^{2\pi}(\rho r^2\cos^2\theta)r^2\sin\theta\,d\phi\,d\theta\,dr.$$

Show that it is $\frac{2}{5}Ma^2$, where M is the mass of the sphere.

4. Using coordinates as in the previous example, prove that the mean centre of a uniform hemisphere of radius a is a distance z from the plane face, where

$$\tfrac{2}{3}\pi a^3 z = \int_0^a\int_0^{\frac{1}{2}\pi}\int_0^{2\pi} r\cos\theta\, r^2\sin\theta\,d\phi\,d\theta\,dr.$$

5. The wave-function $\psi = Are^{-(r/a)}\sin\theta\cos\phi$ is to be normalized, i.e. the volume integral of ψ^2 over the whole of space is to be unity,

$$\int_0^{\infty}\int_0^{\pi}\int_0^{2\pi}\psi^2 r^2\sin\theta\,d\phi\,d\theta\,dr = 1.$$

Find A.

6. Find the average value of $\sin^2\theta\sin^2\phi$ on the surface of a unit sphere.

7. Verify that the volume integral of $f = e^{-(x^2+y^2+z^2)}$ over the whole of space has the same value in either of the forms

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\,dx\,dy\,dz, \quad \int_{-\infty}^{+\infty}\int_0^{\infty}\int_0^{2\pi} f\,r\,d\theta\,dr\,dz,$$

$$\int_0^{\infty}\int_0^{\pi}\int_0^{2\pi} f\,r^2\sin\theta\,d\phi\,d\theta\,dr$$

corresponding to Cartesian, axial and spherical polar coordinates.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

GEARY, A., LOWRY, H. V. and HAYDEN, H. A. *Advanced Mathematics for Technical Students*. Vol. II, Chapter X. Longmans, Green and Co., London, 1950.

LAMB, H. *Infinitesimal Calculus*, Chapter IX. Cambridge University Press, London, 1951.

CHAPTER 18

DETERMINANTS

Linear simultaneous equations—Determinants—Solution in terms of determinants—Homogeneous equations—Characteristic equation—Theory of vibrations—General theory of vibrations—Energy levels of molecules—Symmetric and antisymmetric determinants—Variational derivation of a symmetric determinant—Permutations—Minors and cofactors—Expansion of determinant—Higher-order determinants—Solution of simultaneous equations—Homogeneous equations

18.01. LINEAR SIMULTANEOUS EQUATIONS. Any pair of simultaneous equations for two variables x, y is of the form

$$a_1x + b_1y = c_1,$$
$$a_2x + b_2y = c_2.$$

By eliminating one or the other variable, we obtain for x and y,

$$(a_1b_2 - a_2b_1)x = (c_1b_2 - c_2b_1),$$
$$(a_1b_2 - a_2b_1)y = (a_1c_2 - a_2c_1).$$

Thus the coefficients only appear in certain combinations. Both x and y are multiplied by the same combination. The solution can be written

$$\frac{x}{(c_1b_2 - c_2b_1)} = \frac{y}{(a_1c_2 - a_2c_1)} = \frac{1}{(a_1b_2 - a_2b_1)}.$$

The same happens when we solve the three equations,

$$a_1x + b_1y + c_1z = d_1,$$
$$a_2x + b_2y + c_2z = d_2,$$
$$a_3x + b_3y + c_3z = d_3.$$

The solution is tedious, and we merely state the answer for x.

$$\{a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1)\}x$$
$$= \{d_1(b_2c_3 - b_3c_2) - d_2(b_1c_3 - b_3c_1) + d_3(b_1c_2 - b_2c_1)\}.$$

Here again the coefficients only occur in certain forms. When we have examined these forms, a simple method of solving simultaneous equations can be given.

Example 1. Prove the last formula for x.

Example 2. From the solution for x, guess the solution for y and check your result by actually solving the equations for y.

18.02. DETERMINANTS. A determinant is a combination of numbers or symbols of a certain type.

A determinant of the second order contains four numbers.

It is written $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ and stands for $(a_1b_2 - a_2b_1)$.

Example 1.

$$\begin{vmatrix} 1 & 4 \\ 2 & 9 \end{vmatrix} = 1.9 - 2.4 = 1: \quad \begin{vmatrix} 3 & -5 \\ 1 & -2 \end{vmatrix} = 3(-2) - 1(-5) = -1.$$

Example 2. If $\begin{vmatrix} \lambda & b \\ c & \lambda \end{vmatrix} = 0$, then $\lambda^2 = bc$, or $\lambda = \pm (bc)^{1/2}$.

Example 3. If $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0$, then $a_1b_2 - a_2b_1 = 0$ and $\frac{a_1}{a_2} = \frac{b_1}{b_2}$.

A typical determinant of the *third* order is

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}$$

$$= a_1(b_2c_3 - b_3c_2) - a_1(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1).$$

Since the position of a number or symbol in the determinant determines how it appears in the answer, the correct placing of the numbers in the array is of some importance. The reason for using a square array is clear from the method of evaluation.

Notice the alternate signs $+$, $-$, $+$ in front of the a's.

Example 4.

$$\begin{vmatrix} 1 & 3 & 1 \\ 2 & 1 & 4 \\ 3 & 2 & 1 \end{vmatrix} = 1 \begin{vmatrix} 1 & 4 \\ 2 & 1 \end{vmatrix} - 2 \begin{vmatrix} 3 & 1 \\ 2 & 1 \end{vmatrix} + 3 \begin{vmatrix} 3 & 1 \\ 1 & 4 \end{vmatrix}$$

$$= -7 - 2 + 33 = 24.$$

Example 5.

$$\begin{vmatrix} 0 & b & -c \\ -b & 0 & a \\ c & -a & 0 \end{vmatrix} = 0 \begin{vmatrix} 0 & a \\ -a & 0 \end{vmatrix} - (-b) \begin{vmatrix} b & -c \\ -a & 0 \end{vmatrix} + c \begin{vmatrix} b & -c \\ 0 & a \end{vmatrix} = 0.$$

Example 6.

$$\begin{vmatrix} \lambda & b & -c \\ -b & \lambda & a \\ c & -a & \lambda \end{vmatrix} = \lambda \begin{vmatrix} \lambda & a \\ -a & \lambda \end{vmatrix} - (-b) \begin{vmatrix} b & -c \\ -a & \lambda \end{vmatrix} + c \begin{vmatrix} b & -c \\ \lambda & a \end{vmatrix}$$

$$= \lambda(\lambda^2 + a^2) + b(b\lambda - ac) + c(ab + \lambda c),$$

$$= \lambda(\lambda^2 + a^2 + b^2 + c^2).$$

3.03. SOLUTION IN TERMS OF DETERMINANTS. In terms of determinants, the solution of the two equations,

$$a_1x + b_1y = c_1,$$

$$a_2x + b_2y = c_2,$$

$Dx = A$, $Dy = B$, where

$$D = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad A = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}, \quad B = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}.$$

In the same way the solution of the equations,

$$a_\alpha x + b_\alpha y + c_\alpha z = d_\alpha, \quad (\alpha = 1, 2, 3),$$

now $Dx = A$, $Dy = B$, $Dz = C$, where A, B, C and D are third-order determinants.

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \quad A = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}, \quad B = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}.$$

can be written down by analogy with A and B.

18.04. HOMOGENEOUS EQUATIONS. When it is applied to the three homogeneous equations,

$$\begin{aligned} a_1x + b_1y + c_1z &= 0, \\ a_2x + b_2y + c_2z &= 0, \\ a_3x + b_3y + c_3z &= 0, \end{aligned}$$

the method of the last section gives the three equations,

$$Dx = 0, \quad Dy = 0, \quad Dz = 0.$$

These equations have no solutions other than $x = y = z = 0$, unless

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = 0.$$

The same condition can be considered in a different light. The equations do not determine the absolute values of x, y and z but only the ratios $x : y : z$. These ratios are found by solving only two of the three equations. They are either compatible or incompatible with the remaining equation. If they are incompatible, then $D \neq 0$ and the variables are necessarily zero. If they are compatible, then $D = 0$.

18.05. CHARACTERISTIC EQUATION. The characteristic equation of the third-order determinant D is

$$\Delta(\lambda) = \begin{vmatrix} a_1 - \lambda & b_1 & c_1 \\ a_2 & b_2 - \lambda & c_2 \\ a_3 & b_3 & c_3 - \lambda \end{vmatrix} = 0.$$

The determinant differs from D by the appearance of λ on the principal diagonal.

The roots of this equation are the characteristic roots of the determinant. They are the values of λ for which the following equations have non-zero solutions,

$$\begin{aligned} a_1x + b_1y + c_1z &= \lambda x, \\ a_2x + b_2y + c_2z &= \lambda y, \\ a_3x + b_3y + c_3z &= \lambda z. \end{aligned}$$

The characteristic roots can always be found by expanding the determinant $\Delta(\lambda)$. They can sometimes be guessed, using some of the properties of determinants, especially those which say when determinants are zero!

18.06. THEORY OF VIBRATIONS. A symmetric linear molecule *ABC* can vibrate along its length. The displacements in this direction from the equilibrium positions are x, y and z as shown in *Figure 18.01*.

The changes in the distances *AB* and *BC* due to these displacements bring into play forces F and G between the atoms. These changes are $(y - x)$ and $(z - y)$, respectively, so that the forces are

$$F = k(y - x), \quad G = k(z - y),$$

where k is a force constant, or force per unit extension, which is the same for F and G since the molecule is symmetrical and the atoms *A* and *C* are identical. It is assumed that forces are proportional to extensions, which is only true for small displacements, and also that

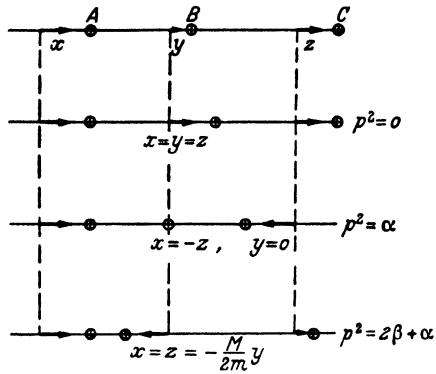

Figure 18.01

a change in the length of *AB* does not affect the force between *B* and *C*, and vice versa. Otherwise we should write, for example,

$$F = k(y - x) + h(z - y), \quad G = k(z - y) + h(y - x),$$

where h is smaller than k. The equations of motion, assuming that $h = 0$, are

$$m\ddot{x} = F = k(y - x),$$
$$M\ddot{y} = G - F = k(x + z - 2y),$$
$$m\ddot{z} = -G = k(y - z).$$

We now look for solutions of these equations for undamped vibrations, where all the atoms vibrate in phase, with the same frequency $\nu = p/2\pi$, e.g.

$$x = a\cos(pt + \varepsilon), \quad y = b\cos(pt + \varepsilon), \quad z = c\cos(pt + \varepsilon).$$
$$\ddot{x} = -p^2x, \quad \ddot{y} = -p^2y, \quad \ddot{z} = -p^2z.$$

If we substitute these expressions into the differential equations, we find that every term contains a factor $\cos(pt + \varepsilon)$. This shows that our assumption might be a correct one. Divided by this common factor the equations become

$$\begin{aligned} -mp^2a &= k(b - a), \\ -Mp^2b &= k(a + c - 2b), \\ -mp^2c &= k(b - c). \end{aligned}$$

If $(k/m) = \alpha$ and $(k/M) = \beta$, then

$$\begin{aligned} (\alpha - p^2)a \qquad - \alpha b \qquad\qquad &= 0, \\ -\beta a + (2\beta - p^2)b - \beta \quad c &= 0, \\ -\alpha b + (\alpha - p^2)c &= 0. \end{aligned}$$

These equations have a non-zero solution only when

$$\begin{vmatrix} \alpha - p^2 & -\alpha & 0 \\ -\beta & 2\beta - p^2 & -\beta \\ 0 & -\alpha & \alpha - p^2 \end{vmatrix} = 0.$$

According to Section 18.02, this is

$$(\alpha - p^2)\begin{vmatrix} 2\beta - p^2 & -\beta \\ -\alpha & \alpha - p^2 \end{vmatrix} + \beta\begin{vmatrix} -\alpha & 0 \\ -\alpha & \alpha - p^2 \end{vmatrix} + 0 = 0$$

or $$(\alpha - p^2)\{p^4 - (2\beta + \alpha)p^2 + 2\alpha\beta\} - \alpha\beta(\alpha - p^2) = 0,$$

i.e. $$(\alpha - p^2)p^2\{p^2 - (2\beta + \alpha)\} = 0.$$

The roots of this equation are

$$p^2 = 0, \quad p^2 = \alpha, \quad p^2 = (2\beta + \alpha).$$

None of these roots is negative, so each gives a possible value of p. For each root, the corresponding values of a, b, c can be found.

(*i*) $p^2 = 0$. This gives $a = b = c$ and $x = y = z$. If all atoms move in the same way, there is motion of the molecule as a

whole along the x axis, the atoms remaining at their equilibrium distances. Although it is not a vibration, it is a possible motion and therefore appears in the answer. It often appears in vibration problems, and could be avoided by choosing different coordinates to describe the motion. Its appearance does not matter as long as it is recognized for what it is.

(ii) $p^2 = \alpha$: the equations give in order,

$$-\alpha b = 0, \quad -\beta(a+c) + (2\beta - \alpha)b = 0, \quad b = 0.$$

so that $b = 0$ and $a = -c$, or

$$y = 0, \quad x = -z = a\cos(pt + \varepsilon).$$

The middle atom remains at rest and the end atoms move in and out together, the same distance in opposite directions.

(iii) $p^2 = 2\beta + \alpha$: the equations now give,

$$-2\beta a - \alpha\beta = 0, \quad -\beta(a+c) - \alpha b = 0, \quad -\alpha b - 2\beta c = 0.$$

These are consistent (if they are not, there is a mistake) and

$$\frac{a}{\alpha} = \frac{b}{-2\beta} = \frac{c}{\alpha},$$

$$\therefore \qquad \frac{x}{M} = -\frac{y}{2m} = \frac{z}{M} = \frac{a}{M}\cos(pt + \varepsilon).$$

This vibration is such that A and C move together the same way, and B moves just far enough in the opposite direction to keep the mass centre of the three at rest.

If we had assumed initially that the mass centre did not move, i.e. $mx + My + mz = 0$, the solution *(i)* would not have appeared.

18.07. General Theory of Vibrations. If the small motion of any system is described by coordinates $x_1, x_2, x_3, \ldots x_n$, of any type whatever, and the kinetic energy is

$$T = \tfrac{1}{2}(m_{11}\dot{x}_1^2 + m_{22}\dot{x}_2^2 + \ldots + 2m_{12}\dot{x}_1\dot{x}_2 + \ldots),$$

and the potential energy,

$$V = \tfrac{1}{2}(b_{11}x_1^2 + b_{22}x_2^2 + \ldots + 2b_{12}x_1x_2 + \ldots),$$

it has been proved by Lagrange that the normal vibrations, where all the coordinates are of the same form $x_i = a_i \sin(pt + \varepsilon)$ $(i = 1, 2, \ldots n)$, are obtained from the n equations

$$(b_{\alpha 1} - p^2 m_{\alpha 1})x_1 + (b_{\alpha 2} - p^2 m_{\alpha 2})x_2 + \ldots + (b_{\alpha n} - p^2 m_{\alpha n})x_n = 0 \quad (\alpha = 1, 2, 3, \ldots n).$$

These equations (see below) can be solved by means of determinants, and the condition for non-zero solutions obtained, just as above, to yield the required values of p^2.

18.08. Energy Levels of Molecules. In the chemical theory of valency, electrons are distributed to the molecules so that the energy is a minimum, in order to get the ground state, or equilibrium state of the molecule. It often happens that a number of distinct structures of this sort all have the same energy E, or practically the same energy E, provided certain interactions between the electrons are ignored. In order to find the correct ground state it is necessary to include these interactions.

Let there be three such structures, labelled 1, 2, 3. When the interactions are included, quantum mechanics asserts that there are states of energy $E + e$, involving mixtures of the three in the proportionate amounts $x : y : z$, provided the following equations are satisfied:

$$\begin{aligned}(v_{11} - en_{11})x_1 + (v_{12} - en_{12})x_2 + (v_{13} - en_{13})x_3 &= 0,\\ (v_{21} - en_{21})x_1 + (v_{22} - en_{22})x_2 + (v_{23} - en_{23})x_3 &= 0,\\ (v_{31} - en_{31})x_1 + (v_{32} - en_{32})x_2 + (v_{33} - en_{33})x_3 &= 0.\end{aligned}$$

The numbers v_{ij} are due to the newly introduced interactions, n_{ii} are normalizing factors and $n_{ij} (i \neq j)$ are overlap integrals. Often the overlap integrals are zero and the calculation made so that $n_{11} = n_{22} = n_{33} = 1$.

The equations are compatible when the determinant of the coefficients of x_1, x_2, x_3 is zero. In general, this gives three values of e. The lowest value of e (algebraically) corresponds to the ground state, and the corresponding values of x_1^2, x_2^2, x_3^2 measure the fractions of 1, 2, 3 in that state. The difference between the smallest of (v_{11}, v_{22}, v_{33}) and e is called the resonance energy.

Example 1. The equations

$$\begin{aligned}(v - e)x_1 + \beta x_2 &= 0,\\ \beta x_1 + (w - e)x_2 + \beta x_3 &= 0,\\ \beta x_2 + (v - e)x_3 &= 0,\end{aligned}$$

require

$$\begin{vmatrix} v - e & \beta & 0 \\ \beta & w - e & \beta \\ 0 & \beta & v - e \end{vmatrix} = 0.$$

This equation has three roots

$$e_0 = v, \quad e_+, e_- = \tfrac{1}{2}(v + w) \pm \tfrac{1}{2}|v - w|\{1 + 8\beta^2/(v - w)^2\}^{1/2}.$$

The order depends on the values of v, w and β but e_- is always the lowest. The resonance energy is

$$\tfrac{1}{2}|v - w|\{\sqrt{(1 + 8\beta^2/(v - w)^2)} - 1\}.$$

This depends on the difference of v and w, but not on whether one or the other is smaller.

18.09. SYMMETRIC AND ANTISYMMETRIC DETERMINANTS. The determinant,

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

is symmetric when $a_{ij} = a_{ji}$ for all i and j. In detail,

$$a_{12} = a_{21}, \; a_{13} = a_{31}, \; a_{23} = a_{32}.$$

It is *antisymmetric* when $a_{ij} = -a_{ji}$, for all i and j. In detail,

$$a_{11} = a_{22} = a_{33} = 0,$$

$$a_{12} = -a_{21}, \quad a_{13} = -a_{31}, \quad a_{23} = -a_{32}.$$

18.10. VARIATIONAL DERIVATION OF A SYMMETRIC DETERMINANT. When there are relations between the coefficients, $a_{12} = a_{21}$, $a_{13} = a_{31}$ and $a_{23} = a_{32}$ the equations,

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= \lambda x, \\ a_{21}x + a_{22}y + a_{23}z &= \lambda y, \\ a_{31}x + a_{32}y + a_{33}z &= \lambda z, \end{aligned} \qquad \ldots . (1)$$

are identical with the equations,

$$\frac{\partial}{\partial x}(I - \lambda r^2) = 0, \quad \frac{\partial}{\partial y}(I - \lambda r^2) = 0, \quad \frac{\partial}{\partial z}(I - \lambda r^2) = 0,$$

where $I = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz$,

and $r^2 = x^2 + y^2 + z^2$.

The solutions of the simultaneous equations are therefore the points where $(I - \lambda r^2)$ has stationary values. Normally there is only one point, namely the origin $x = y = z = 0$, but when λ is one of the characteristic roots of the determinant of the coefficient a_{ij} there are other stationary points.

For each root, the equations (1) can be solved for the ratios $x : y : z$. For the root $\lambda = \lambda_1$ let the result be

$$\frac{x}{l_1} = \frac{y}{m_1} = \frac{z}{n_1} = r. \qquad \ldots . (2)$$

The common ratio r is the distance from the origin, if the numbers l_1, m_1, n_1 are so chosen that $l_1^2 + m_1^2 + n_1^2 = 1$, which can always be done as only ratios are involved.

Hence, for $\lambda = \lambda_1$, we obtain $d(I - \lambda_1 r^2) = 0$, for any small change (dx, dy, dz) starting from a point on the line defined by equations (2). Some interesting results can be obtained from particular types of small change.

(*i*) $d(I - \lambda_1 r^2) = 0$ for a displacement along the line (2). Thus $I - \lambda_1 r^2$ is constant along this line: since it is zero at the origin it is zero all along the line. In other words, when λ_1 is a characteristic root and $x : y : z = l_1 : m_1 : n_1$, then $\lambda_1 = (I/r^2)$.

(*ii*) The line through the origin described by equations (2) meets the surface $I =$ constant at two points. For any displacement from one of these points $d(I - \lambda_1 r^2) = 0$. If the displacement is along the surface $I =$ constant, then $dI = 0$ as well, and therefore $d(r^2) = 0$. This means that the surfaces $I =$ constant are normal to the line (2).

(*iii*) The values of the ratio I/r^2 only depend on the ratios x/r, y/r, z/r; and its changes, as x, y and z are varied, are not altered by the limitation that $r =$ constant. But then the stationary values of I/r^2 are those of I or of $I - \lambda r^2$, which are in the directions (2).

In all, therefore, to each root λ_1, there is a direction in which I/r^2 is stationary for changes in x, y, z, and this direction is normal to the surfaces $I =$ constant.

Example 1. The directions normal to the surface $I =$ constant are the directions of the principal axes of this quadric surface.

Example 2. The problems discussed in Sections 18.06 and 18.08 can be rephrased using an expression I, so that the values of the characteristic roots appear as the stationary values of a ratio (I/r^2).

This method of calculating the values of the roots is called a variation method.

18.11. PERMUTATIONS. In this section and the next we discuss third-order determinants in such a manner that the discussion applies to high-order determinants, such as arise from the solution of more than three simultaneous equations. The necessity for such an extension has already been suggested in Sections 18.07 and 18.08.

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1.$$

Every term in D is a product of numbers or elements, one from each column, so that no two elements come from the same row. D itself is obtained by adding all possible products of this sort with a $\pm$ sign attached.

Since an 'a' can be chosen in three ways and, when this is done, there remain two rows from which to choose the 'b', there are altogether $3.2.1 = 6$ terms.

The principal term $a_1b_2c_3$ is the product of the elements on the principal diagonal. When the others are written (as above) in order of letters or columns from which the elements come, their suffices form some rearrangement or permutation of the numbers 1, 2, 3.

The principal term has a $+$ sign. The signs attached to the other terms are determined by a simple rule: if one product differs from another simply by an interchange of the suffices on two letters, it has the opposite sign. (In terms of the array, this means that the two terms contain the same elements from all the columns except two, and for those two the rows are interchanged.)

Example 1. $a_1b_3c_2$ and $a_2b_3c_1$ have opposite signs.

This rule leads to no contradictions owing to the group property of permutations, and it gives the sign uniquely.

Now consider the expression for D in the light of these remarks. Exchanging two suffices, say 1 and 2, changes every term to another term with the opposite sign. Hence if D' is the expression after the exchange, then $D' = -D$. In terms of the array of elements, this corresponds to exchanging the elements of the two rows 1 and 2 (keeping the elements in the same columns, of course).

Thus changing two rows of a determinant D leads to a determinant $D' = -D$.

Example 2.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = - \begin{vmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{vmatrix} = + \begin{vmatrix} 7 & 8 & 9 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{vmatrix}.$$

It follows at once that $D = 0$, if two rows are identical.

Example 3.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{vmatrix}.$$

Hence both are zero.

Example 4. When expanded

$$D = \begin{vmatrix} a & b & c \\ a^2 & b^2 & c^2 \\ b+c & c+a & a+b \end{vmatrix}$$

must have $(a - b)$, $(b - c)$, $(c - a)$ as factors, since it vanishes when $a = b$ or $b = c$ or $c = a$.

18.12. Minors and Cofactors. The minor of any element is the determinant left, when the row and column containing that element are deleted, and the remainder is closed up to a smaller square array without altering the relative positions of the parts. The minor of an element such as b_2 is denoted by the symbol B_2.

Example 1.

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

The minor of a_2 is $A_2 = \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix}$,

and that of b_2 is $B_2 = \begin{vmatrix} a_1 & c_1 \\ a_3 & c_3 \end{vmatrix}$.

The *cofactor* of an element is its coefficient in the expansion of the determinant. Cofactors are denoted here by Greek letters.

Example 2. The cofactor of a_2 is $\alpha_2 = b_3c_1 - b_1c_3$ and that of b_2 is $\beta_2 = a_1c_3 - a_3c_1$.

Since the terms in which an element occurs contain no two elements from the same row or column, its cofactor is a sum of all products of elements from other rows and columns, satisfying the $\pm$ rule for a determinant. Thus the cofactor and minor of the same element are almost identical.

The cofactor is a minor with a $+$ or $-$ sign attached, the sign being easily found by considering the sign of the principal term in the expansion of the minor.

Example 3. The principal term in the minor of a_1 is b_2c_3. Since the product $a_1b_2c_3$ has a positive sign in D, then b_2c_3 has a positive sign in the cofactor of a_1. Hence

$$\alpha_1 = A_1.$$

Example 4. The principal term in the minor of b_3 is a_1c_2. Since the product $a_1b_3c_2$ has a negative sign in D (it is opposite in sign to $a_1b_2c_3$) then a_1c_2 has a negative sign in the cofactor. Hence $\beta_3 = -B_3$.

The cofactor of an element is obtained from the minor by attaching a sign, which is placed in the position of the element in the following array.

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}.$$

18.13. EXPANSION OF DETERMINANT. We return to the rule given in Section 18.02. In terms of the cofactors of the elements of the first column,

$$D = a_1\alpha_1 + a_2\alpha_2 + a_3\alpha_3,$$

because every term in D contains one factor from that column, and must therefore occur in one of the cofactors, and no term occurs twice.

Any other column of elements could be used equally well.

$$D = b_1\beta_1 + b_2\beta_2 + b_3\beta_3 = c_1\gamma_1 + c_2\gamma_2 + c_3\gamma_3.$$

Finally we note that the symmetry in rows and columns means

that any property of the determinant which applies to columns also applies to rows. Hence we can write

$$\begin{aligned} D &= a_1\alpha_1 + b_1\beta_1 + c_1\gamma_1, \\ &= a_2\alpha_2 + b_2\beta_2 + c_3\gamma_3, \\ &= a_3\alpha_3 + b_3\beta_3 + c_3\gamma_3. \end{aligned}$$

The advantage of being able to choose between different ways of expanding a determinant is that it is possible to take advantage of any zeros in some of the rows or columns.

It may now be noted that if the cofactors of any row (or column) are multiplied by the elements of another row (or column) and added the result is zero.

For example

$$b_1\alpha_1 + b_2\alpha_2 + b_3\alpha_3 = \begin{vmatrix} b_1 & b_1 & c_1 \\ b_2 & b_2 & c_2 \\ b_3 & b_3 & c_3 \end{vmatrix} = 0,$$

because the sum is a determinant with two columns identical.

18.14. HIGHER-ORDER DETERMINANTS. A determinant of order n is the value of a combination of $n \times n$ elements in n rows and n columns.

$$D_n = \begin{vmatrix} a_1 & b_1 & c_1 & \cdots\cdots & k_1 \\ a_2 & b_2 & c_2 & \cdots\cdots & k_2 \\ \cdot & \cdot & \cdot & \cdots\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots\cdot & \cdot \\ a_n & b_n & c_n & \cdots\cdot & k_n \end{vmatrix} : \text{Signs of cofactors} \begin{vmatrix} + & - & + & - & \cdots\cdot \\ - & + & - & + & \cdots\cdot \\ + & - & + & - & \cdots\cdot \\ - & + & - & + & \cdots\cdot \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{vmatrix}$$

As in the previous two sections, D_n is defined as the sum of $n!$ terms each containing n factors or elements, one from each row and column. It changes sign when two rows are exchanged and is zero if two rows (or columns) are equal. Minors and cofactors are defined as before, and the signs to be attached to minors to convert them to cofactors are given above. The determinant can be expanded in terms of the elements of any row (or column) multiplied by their cofactors, e.g.

$$\begin{aligned} D_n &= a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n, \\ &= a_1A_1 - a_2A_2 + a_3A_3 - \cdots + (-)^n a_nA_n. \end{aligned}$$

18.15. SOLUTION OF SIMULTANEOUS EQUATIONS. Consider the n equations,

$$a_1x + b_1y + c_1z + \ldots + k_1w = p_1,$$
$$a_2x + b_2y + c_2z + \ldots + k_2w = p_2,$$
$$\ldots\ldots\ldots\ldots$$
$$a_nx + b_ny + c_nz + \ldots + k_nw = p_n.$$

The determinant D_n of the previous section is the determinant of the coefficients of the unknowns $x, y, \ldots . w$.

The equations can be multiplied in order by the cofactors α_1, α_2, α_3, . . . of the elements $a_1, a_2, \ldots a_n$, in this determinant and then added. The total coefficient of x in the sum is

$$a_1\alpha_1 + a_2\alpha_2 + \ldots + a_n\alpha_n = D.$$

The coefficient of y in the sum is

$$b_1\alpha_1 + b_2\alpha_2 + \ldots + b_n\alpha_n = 0,$$

because it is the expansion of a determinant with two columns the same. In exactly the same way the coefficients of $z, \ldots w$ are all zero, so that the sum reduces to

$$(a_1\alpha_1 + a_2\alpha_2 + \ldots + a_n\alpha_n)x = (p_1\alpha_1 + p_2\alpha_2 + \ldots + p_n\alpha_n)$$

or

$$Dx = A.$$

A is the same as D with the numbers $p_1, p_2, \ldots p_n$, in the first column in place of $a_1, a_2, \ldots a_n$.

The value of y is obtained by multiplying the original equations by the cofactors $\beta_1, \beta_2, \ldots \beta_n$ and adding. Then we find that

$$Dy = B,$$

where B is the determinant D with the elements $b_1, b_2, \ldots b_n$ in the second column replaced by $p_1, p_2, \ldots p_n$.

Repeating this for all the unknowns we write the final solution in the form

$$\frac{x}{A} = \frac{y}{B} = \frac{z}{C} = \ldots = \frac{w}{K} = \frac{1}{D}.$$

18.16. HOMOGENEOUS EQUATIONS. The homogeneous equations, obtained by putting $p_1 = p_2 = p_3 = \ldots = p_n = 0$ in the equations of the previous section, have a solution where not all of $x, y, z, \ldots w$ are zero only if

$$D = 0.$$

EXAMPLES XVIII

1. Solve the equations: $\lambda x = kx + \omega_3 y - \omega_2 z,$

$$\lambda y = -\omega_3 x + ky + \omega_1 z,$$

$$\lambda z = \omega_2 x - \omega_1 y + kz,$$

assuming that they have a non-zero solution. Use the abbreviation $\Omega^2 = \omega_1^2 + \omega_2^2 + \omega_3^2$. (N.B. the solutions are complex.)

2. The component, in a given direction, of the angular momentum of an atom has a value $m\hbar$, where m is a root of the equation

$$\begin{vmatrix} -m & 1 & 0 \\ 1 & -m & 1 \\ 0 & 1 & -m \end{vmatrix} = 0.$$

Find the values of m.

3. The potential energy of a particle, with coordinates (x, y), is

$$V = \tfrac{1}{2}(ax^2 + 2hxy + y^2),$$

and its kinetic energy is $T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)$. Find the periods of normal oscillations when $h^2 < ab$, and the ratio $x : y$ for each type of oscillation.

4. To a 50 cm³ sample gas mixture of carbon monoxide, methane and acetylene, a volume of 150 cm³ of oxygen is added. After combustion and cooling, the residual volume is 122·5 cm³. Absorption of carbon dioxide by caustic potash solution reduces the volume to 57·5 cm³. Find the composition of the sample, the oxygen consumed during combustion and the oxygen unused, assuming the reactions:

$$2CO + O_2 \rightarrow 2CO_2,\ CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O,$$

$$2C_2H_2 + 5O_2 \rightarrow 4CO_2 + 2H_2O.$$

5. By guessing values of x, a or b which make the determinant zero, or otherwise, show that

$$\begin{vmatrix} x & x^2 & a+b \\ a & a^2 & x+b \\ b & b^2 & a+x \end{vmatrix} = (b-a)(x-a)(x-b)(x+a+b).$$

6. Determinants D_n, with n rows and columns ($n = 1, 2, 3, \ldots$) have the form

$$D_n = \begin{vmatrix} 2\alpha & \beta & 0 & \ldots & 0 \\ \beta & 2\alpha & \beta & \ldots & 0 \\ 0 & \beta & 2\alpha & & \\ \vdots & \vdots & \beta & & \\ \vdots & \vdots & \vdots & 2\alpha & \beta \\ 0 & 0 & 0 & \beta & 2\alpha \end{vmatrix}$$

Prove that

$$D_n = 2\alpha D_{n-1} - \beta^2 D_{n-2}.$$

Using *Example 8*, Chapter 8, prove that

$$D_n = \beta^n(A \cosh n\theta + B \sinh n\theta) = \beta^n \sinh (n + 1)\theta/\sinh \theta,$$

where $\alpha = \beta \cosh \theta > \beta$ and $A = 1$, $B = \coth \theta$. Find the corresponding formula when $\alpha = \beta \cos \alpha < \beta$. (N.B. $\theta = i\alpha$.)

REFERENCES

The author has found helpful information on the subject matter of this chapter in the following books:

AITKEN, A. C. *Determinants and Matrices.* Oliver and Boyd, Edinburgh, 1939.

COWLEY, W. L. *Advanced Practical Mathematics*, Chapters 10 and 11. Sir Isaac Pitman and Sons, London, 1934.

PAULING, L. and WILSON, J. B. *Introduction to Quantum Mechanics.* McGraw-Hill Book Co., New York, 1935.

CHAPTER 19

VECTORS

Scalars and vectors—Position vector—Areas as vectors—Rotations—Velocity and flux vectors—Thermal gradients—Heat flow—Potential gradients—Concentration gradients—Vector algebra—Vector algebra using components—Definition of vectors—Vectors in mechanics—Moments. Vector products—Fields—Scalar fields—Vector fields—Activity—Conservative fields—Irrotational fields—Circulation—Flux or flow across an area—Divergence of a vector field—Solenoidal fields—Vector fields in general—Double derivatives—Generalized vectors

19.01. SCALARS AND VECTORS. A *scalar* is a quantity with magnitude only, such as temperature or any form of energy.

A *vector* is a quantity with magnitude and direction, such as velocity or the flow of energy or material.

The idea of a vector was developed by Gibbs, the vector being regarded as a single quantity, and it gives such an excellent pictorial representation that it is now in common use.

19.02. POSITION VECTOR. Given an origin O and Cartesian axes OX, OY, OZ, the coordinates of any point P are (x, y, z). The line **OP** in the sense from O to P has magnitude and direction, and is therefore a vector, the position vector of P. The magnitude of this vector is its length $r = OP$.

The projections of the line OP on the coordinate axes (*Figure 19.01*) are the coordinates of P, with suitable signs attached. These are the components of the vector along the axes. If α, β, γ are the angles XOP, YOP, ZOP, to be chosen so that they are less than 180°, then $OL = x = r \cos \alpha$ and similarly,

$$x = r \cos \alpha, \quad y = r \cos \beta, \quad z = r \cos \gamma.$$

These components, or coordinates, are positive when the angles are less than 90°, and negative when the angles are greater than 90°.

The three numbers $\cos \alpha$, $\cos \beta$ and $\cos \gamma$ are called the direction cosines of OP relative to the axes and usually denoted by the symbols (l, m, n) in that order.

The component of the vector $\mathbf{r} = \mathbf{OP}$ in any other direction OA is defined as $r \cos \theta$, where $\theta = POA$.

The magnitude of the vector is determined from its Cartesian components by the relation,

$$r^2 = x^2 + y^2 + z^2.$$

It follows that there is a relation between the direction cosines,

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1.$$

Now consider changes in the position vector. Let Q be another point with position vector $\mathbf{r}' = \mathbf{OQ}$, with Cartesian components (x', y', z'). The displacement $\mathbf{PQ}$ from P to Q, in that sense, is also a vector $\mathbf{d}$. Its Cartesian components are $(x' - x, y' - y, z' - z)$.

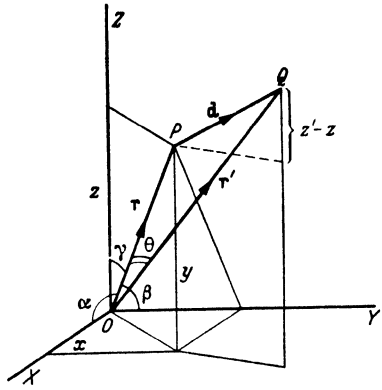

Figure 19.01

Similarly the component of $\mathbf{d}$ in *any* direction is the difference of the components of $\mathbf{r}'$ and $\mathbf{r}$. We therefore write

$$\mathbf{d} = \mathbf{r}' - \mathbf{r},$$

or

$$\mathbf{r}' = \mathbf{r} + \mathbf{d}.$$

This is the law of addition of vectors. If $\mathbf{r} = \mathbf{OP}$ and $\mathbf{d} = \mathbf{PQ}$, then $\mathbf{r} + \mathbf{d} = \mathbf{OQ}$. The points O, P and Q do not necessarily lie in a straight line, so that the magnitude of $\mathbf{r}'$ is not the sum of the magnitudes of $\mathbf{r}$ and $\mathbf{d}$. In fact, using the cosine theorem,

$$r'^2 = r^2 + d^2 + 2rd \cos \theta,$$

where θ is the angle between OP and PQ.

The displacement $\mathbf{QP}$ should obviously be $-\mathbf{d}$. That is, changing the sign of a vector gives an equal vector in the opposite direction. Similarly $\mathbf{PO} = -\mathbf{r}$ and the equation $\mathbf{d} = \mathbf{r}' - \mathbf{r}$ can be written

$$\mathbf{d} = \mathbf{r}' + (-\mathbf{r}) \quad \text{or} \quad \mathbf{PQ} = \mathbf{PO} + \mathbf{OQ}.$$

19.03. AREAS AS VECTORS. A plane area can be represented by a vector along its normal. The magnitude of the vector is the magnitude of the area, but the direction of this vector along the normal is a matter of choice. It is usually related (*Figure 19.06*) to a certain direction of travel around the boundary of the area, so that a right-handed screw moves in the direction of the vector when twisted in the direction of travel around the boundary. The component of this vector in any direction equals the area of the projection of the given area on a plane perpendicular to the direction.

19.04. ROTATIONS. A rotation through an angle $d\alpha$ about an axis OA can be represented by a vector of magnitude $d\alpha$ along the axis. Once again the sense of the vector along this axis is a matter of choice, and it is usually chosen, as for areas, in the direction of motion of a right-handed screw which undergoes the rotation.

19.05. VELOCITY AND FLUX VECTORS. The velocity of a point P is a vector $\mathbf{v}$. Since velocity is displacement per unit time, it is easy to show, using the results of Section 19.02, that its components along three Cartesian axes are the rates of change of the coordinates of P with time.

To describe the motion of a fluid, the velocity $\mathbf{v}$ needs to be known at all points in the fluid, and in general it varies from point to point. The fluid crossing a small area dS near a point P per unit time is the fluid contained in a cylinder on the base dS and of length v in the direction of the velocity (*Figure 19.02*). The volume of this cylinder is (base)(height) $= dS \,.\, v \cos\theta$.

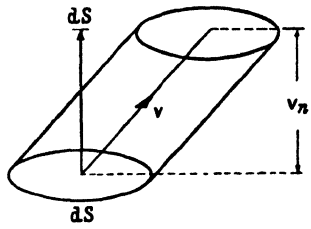

Figure 19.02

The angle θ is that between the vector $\mathbf{v}$ and the vector $d\mathbf{S}$ representing the area. The combination just obtained is called the scalar product of these vectors and is written

$$\mathbf{v} \,.\, d\mathbf{S} = v\, dS \cos\theta.$$

The flow per unit area, or volume flux, is the component of $\mathbf{v}$ normal

to the area. Hence **v** is also called the volume flux vector, to be used to calculate the flow across any area.

The mass flux vector is $\rho\mathbf{v}$ where ρ is the density. It is just ρ times as large as **v** and in the same direction.

19.06. THERMAL GRADIENTS. Variation of temperature in a body implies temperature gradients and heat flow. The temperature at a point P is a function of the position of P and $T = f(x, y, z)$, if the position is given by coordinates (x, y, z).

Consider two isothermal surfaces at temperatures T and $T + dT$. From a point P on one to a point Q on the other is a distance $PQ = ds$. If PN is the normal at P, then $PN = dn = ds \cos\theta$ (*Figure 19.03*).

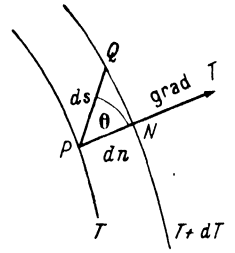

Figure 19.03

The temperature gradient at P in the direction PQ is defined as the rate of change of temperature with distance in that direction, i.e.

$$\frac{dT}{ds} = \frac{dT}{(dn/\cos\theta)} = \frac{dT}{dn}\cos\theta.$$

Now (dT/dn) is the temperature gradient, in the same sense, along the normal. Hence (dT/ds) is the component of a vector of magnitude (dT/dn) along the normal to the isothermal at P, in the direction in which the temperature increases. This vector is the

$$\text{gradient of } T = \text{grad } T.$$

If PQ is parallel to the x axis, then $ds = dx$. The partial derivative $\partial T/\partial x$ is the rate of change of temperature with distance in the direction x-increasing. Thus the three partial derivatives $\partial T/\partial x$, $\partial T/\partial y$, $\partial T/\partial z$ are the components of grad T in the directions of the coordinate axes. Just as in the example of the position vector in Section 19.02 the magnitude of the gradient vector is therefore

$$|\text{grad } T|^2 = \left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2,$$

The partial derivatives, divided by this magnitude, are the direction cosines of the vector relative to the three Cartesian axes.

19.07. HEAT FLOW. According to Fourier's law, the heat flow by conduction across an area dS normal to a direction PQ is $-\kappa(dT/ds)$ per unit area, where κ is the thermal conductivity, i.e. in all,

$$-\kappa\,(\partial T/\partial s)dS = -\kappa \operatorname{grad} T \,.\, d\mathbf{S}.$$

Thus $-\kappa \operatorname{grad} T$ is a heat flux vector.

19.08. POTENTIAL GRADIENTS. A force $\mathbf{F}$ on a body is a vector. If the force on a body at a point P is such that:

(*i*) it depends on the position of the body,
(*ii*) its components are those of $-\operatorname{grad} V$,

where V is a function of the position of P, then V is called the potential energy of the force.

19.09. CONCENTRATION GRADIENTS. Fick's law of diffusion, in its general form, says that a concentration variation from point to point gives rise to a flow of material, which can be calculated from a flux vector $-D \operatorname{grad} c$, where D is the coefficient of diffusion and c is the concentration. This means that the flow across an area dS due to diffusion is

$$-D \operatorname{grad} c \,.\, d\mathbf{S}$$

or $-D(\partial c/\partial n)dS$, where $(\partial c/\partial n)$ is the rate of change of concentration with distance along the normal to the area.

19.10. VECTOR ALGEBRA. Formal rules for combining vectors are now stated. Most of them have already been illustrated.

The vectors $\mathbf{a}$ and $\mathbf{b}$ can be represented in magnitude and direction by two lines OA and AB, respectively; the line OB then represents their sum. A line parallel and equal to AB, such as OC, would also represent the vector $\mathbf{b}$. The difference $\mathbf{b} - \mathbf{a}$ is represented by AC (*Figure 19.04*).

The component of the sum $\mathbf{a} + \mathbf{b}$ of two vectors, in any direction, is the sum of the components of $\mathbf{a}$ and $\mathbf{b}$ in that direction.

The scalar product of two vectors is

$$\mathbf{a} \,.\, \mathbf{b} = \mathbf{b} \,.\, \mathbf{a} = ab \cos\theta,$$

where θ is the angle between them, that is, the angle between any

two lines which represent them. The scalar product equals the magnitude of **a** times the component of **b** in the direction of **a**, and is zero when the two vectors are perpendicular.

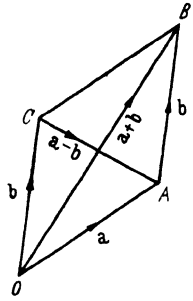

Figure 19.04

The component of $(\mathbf{b} + \mathbf{c})$ in the direction of **a** is the sum of the components of **b** and **c** in the direction of **a**. Multiply the three by the magnitude of **a**. Then we obtain the equation

$$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c},$$

i.e. in scalar products, parentheses can be removed as in ordinary algebra

19.11. VECTOR ALGEBRA USING COMPONENTS. Let $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ be three unit vectors (i.e. vectors of unit magnitude), perpendicular to one another. For example, they may point along the three axes OX, OY, OZ, when Cartesian coordinates are being used to specify position (*Figure 19.05*, p 324).

The components (a_1, a_2, a_3) of **a** along these three directions are referred to as orthogonal components of the vector **a**. Because the vectors **n** are unit vectors,

$$a_1 = \mathbf{a} \cdot \mathbf{n}_1, \quad a_2 = \mathbf{a} \cdot \mathbf{n}_2, \quad a_3 = \mathbf{a} \cdot \mathbf{n}_3.$$

Moreover, **a** is the sum of three vectors,

$$\mathbf{a} = a_1\mathbf{n}_1 + a_2\mathbf{n}_2 + a_3\mathbf{n}_3.$$

Similarly, if **b** is a second vector with orthogonal components (b_1, b_2, b_3) in the same directions,

$$\mathbf{b} = b_1\mathbf{n}_1 + b_2\mathbf{n}_2 + b_3\mathbf{n}_3.$$

Adding and multiplying these two vectors **a** and **b**,

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1)\mathbf{n}_1 + (a_2 + b_2)\mathbf{n}_2 + (a_3 + b_3)\mathbf{n}_3.$$
$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot (b_1\mathbf{n}_1 + b_2\mathbf{n}_2 + b_3\mathbf{n}_3),$$
$$= a_1b_1 + a_2b_2 + a_3b_3.$$

In this way the scalar product can be obtained directly from the orthogonal components.

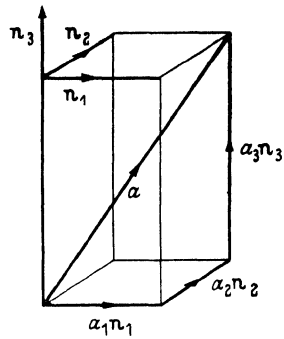

Figure 19.05

Referred to the three unit vectors, the direction cosines of **a** are:

$$l = a_1/a, \quad m = a_2/a, \quad n = a_3/a.$$

Divide both sides of the expression for the scalar product by the magnitude a, and introduce these direction cosines;

$$(\mathbf{b} \cdot \mathbf{a})/a = b \cos \theta = b_1l + b_2m + b_3n.$$

This is the component of any vector **b** in the direction (l, m, n).

19.12. Definition of Vectors. A vector has a value associated with every direction, which has been called its component in that direction. Not every quantity with a value associated with every direction is a vector, but only those having components obtained according to the law given above (component $= a \cos \theta$). Thus if **a** has three components (a_1, a_2, a_3) in three orthogonal directions, then its component in any other direction with cosines (l, m, n) must be

$$a_n = la_1 + ma_2 + na_3.$$

Geometrically, if a line OA represents a vector a, and OP its component a_n in some direction making an angle θ with OA, then the

relation $a_n = a \cos \theta$ implies that P lies on a sphere with OA as diameter.

If these relations are not satisfied, the quantity is not a vector. A body has a moment of inertia about every axis through its mass centre, and these are not related in the same way as the components of a vector (e.g. they are never negative).

19.13. VECTORS IN MECHANICS. In mechanics the position vector of a moving point $\mathbf{r}$ is a function of the time. Its velocity $\mathbf{v} = d\mathbf{r}/dt$ is a vector, and its acceleration $\mathbf{f} = d\mathbf{v}/dt$. Newton's law of motion takes the form,

$$\text{Force} = \mathbf{F} = m\mathbf{f},$$

where m is the mass. Forces are vectors.

The work done by a force $\mathbf{F}$ acting on a body which moves a distance $\mathbf{d}$ is defined as

$$F \cos \theta \, . \, d = \mathbf{F} \, . \, \mathbf{d}.$$

It is the product of the displacement and the component of the force in the direction of the displacement.

19.14. MOMENTS. VECTOR PRODUCTS. Consider a force applied at a point P of a large body, free to rotate about a point O. This force tends to turn the body about an axis OA through O normal to the plane of the vectors $\mathbf{OP}$ and $\mathbf{F}$. Let the direction OA be such that the sense of rotation is clockwise looking along that line. (Draw this.)

The moment of the force about OA is defined as the product $Fp = Fr \sin \theta$, where θ is the angle between $\mathbf{r} = \mathbf{OP}$ and $\mathbf{F}$. This moment is represented by a vector along the line OA of magnitude $Fr \sin \theta$, namely

$$\mathbf{r} \wedge \mathbf{F}.$$

It is the vector product of $\mathbf{r}$ and $\mathbf{F}$, in that order.

Vector products are defined for any two vectors. In *Figure 19.06* the vector products $\mathbf{a} \wedge \mathbf{b}$ and $\mathbf{b} \wedge \mathbf{a}$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ are represented. They are constructed by the same rules as $\mathbf{r} \wedge \mathbf{F}$. It is very obvious that they are in opposite directions,

$$\mathbf{a} \wedge \mathbf{b} = -\mathbf{b} \wedge \mathbf{a}.$$

Thus the order of the vectors in a vector product is very important.

This diagram illustrates another example of a vector product. The area of the parallelogram $OABC$ is a vector $\mathbf{S} = \mathbf{a} \wedge \mathbf{b}$, if the normal is drawn one way, or the vector $\mathbf{S} = \mathbf{b} \wedge \mathbf{a}$ if it is drawn the

other way. Both these have the right magnitude ($ab \sin\theta$) and they are directed along the normal.

The vector product $\mathbf{a} \wedge \mathbf{b}$ is zero when $\sin\theta = 0$, i.e. $\mathbf{a}$ and $\mathbf{b}$ are parallel or opposite (e.g. $\mathbf{a} \wedge \mathbf{a} = 0$).

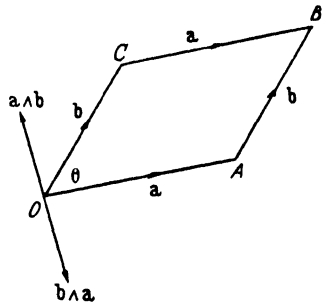

Figure 19.06

Draw three unit orthogonal vectors $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$ such that $\mathbf{n}_2 \wedge \mathbf{n}_3$ is in the direction of $\mathbf{n}_1$. Then it is found that $\mathbf{n}_3 \wedge \mathbf{n}_1$ is in the direction of $\mathbf{n}_2$ and that $\mathbf{n}_1 \wedge \mathbf{n}_2$ is in the direction of $\mathbf{n}_3$. The three vectors ($\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$), in that order, form a right-handed set. Their vector products can now be easily written down. The magnitude of $\mathbf{n}_2 \wedge \mathbf{n}_3$ is $|n_2||n_3|\sin\theta = \sin 90° = 1$, and the same for the other two. Hence

$$\mathbf{n}_2 \wedge \mathbf{n}_3 = \mathbf{n}_1, \qquad \mathbf{n}_3 \wedge \mathbf{n}_1 = \mathbf{n}_2, \qquad \mathbf{n}_1 \wedge \mathbf{n}_2 = \mathbf{n}_3.$$

$$\mathbf{n}_3 \wedge \mathbf{n}_2 = -\mathbf{n}_1, \quad \mathbf{n}_1 \wedge \mathbf{n}_3 = -\mathbf{n}_2, \quad \mathbf{n}_2 \wedge \mathbf{n}_1 = -\mathbf{n}_3.$$

Also

$$\mathbf{n}_1 \wedge \mathbf{n}_1 = \mathbf{n}_2 \wedge \mathbf{n}_2 = \mathbf{n}_3 \wedge \mathbf{n}_3 = 0.$$

In terms of the orthogonal components

$$\begin{aligned}\mathbf{a} \wedge \mathbf{b} &= (a_1\mathbf{n}_1 + a_2\mathbf{n}_2 + a_3\mathbf{n}_3) \wedge (b_1\mathbf{n}_1 + b_2\mathbf{n}_2 + b_3\mathbf{n}_3) \\ &= \mathbf{n}_1(a_2b_3 - a_3b_2) + \mathbf{n}_2(a_3b_1 - a_1b_3) + \mathbf{n}_3(a_1b_2 - a_2b_1).\end{aligned}$$

$$\therefore \qquad \mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} \mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

It can be proved that it is permissible to remove the parentheses in the first line.

19.15. FIELDS. The assertion that a quantity, either scalar or vector, has a value at every point in a certain region is equivalent to saying that the scalar, or the components of the vector, are functions of position. Even more explicitly, their value or values at any point P are functions of the coordinates of P. Viewed as a whole, the values at all points form a collection, or field, of values.

In the next few sections we examine some properties of fields, either by forming the derivatives of the functions describing them, which give local properties of the field, or by calculating integrals which give properties of the field as a whole.

Examples of scalar fields are temperature, mass density, electric charge density, and potentials. Examples of vector fields include the velocity of a moving fluid, and gravitational force, the values of which at every point are vectors. The electromagnetic field consists of the values of both the electric and magnetic field at each point.

19.16. SCALAR FIELDS. A scalar field ϕ can be expressed as

$$\phi = f(x, y, z),$$

where (x, y, z) are the Cartesian coordinates of the points, or it can be expressed in terms of other coordinates. The properties of ϕ in which we are interested are those which do not depend on the particular coordinates used. An abbreviated notation $\phi = f(\mathbf{r})$ can be used, where $\mathbf{r}$ is the position vector of the point.

When a scalar field is integrated over a volume, it is being treated in the same way as a density, and the result is a bulk property. For example, the density ρ of a body is a scalar function of position and is used to calculate properties of the body as a whole, such as the mass or its moments of inertia, by means of volume integrals (see Section 17.12).

Scalar fields can also be regarded as potentials. Our chief interest then lies in their variation from point to point, which gives rise to forces producing motion of one sort or another in the direction of decreasing (or sometimes increasing) potential. The gravitational potential gives rise to the mechanical force of gravity, and the electric potential to electrical forces. Variations in temperature give rise to thermal forces and a flow of heat to regions of lower temperature. (In a reaction at constant volume, the thermodynamic potential tends to decrease to a minimum.)

The variation of a scalar ϕ is described by its gradient vector grad ϕ. As described in Section 19.09, the scalar product $\mathbf{n} \cdot \text{grad } \phi$ gives the rate of change of ϕ with distance in the direction $\mathbf{n}$.

Some quantities behave as potentials in one context and as densities in another. Thus on the one hand, by integration, the function describing the local concentration of solute gives the total amount of solute, and on the other hand, by differentiation, it gives rise to diffusion.

19.17. VECTOR FIELDS. In earlier sections it has been explained that the volume flow of a liquid is described by a velocity vector $\mathbf{v}$, that heat conduction is described by a gradient $-\kappa \operatorname{grad} \phi$, and so on. All these are examples of vector fields of flow.

The other main type of vector field is the field of force, where the vector at every point represents a force at that point.

Vector fields of either type can be divided into:

(*i*) conservative and irrotational fields;
(*ii*) solenoidal or source-free fields;
(*iii*) those which can be expressed in two parts, one solenoidal and one irrotational.

These terms are explained in the following sections.

19.18. ACTIVITY. If $\mathbf{F}$ is a mechanical force acting on a particle, the work done by the force when the particle is displaced $d\mathbf{r}$ is the scalar product $\mathbf{F} \cdot d\mathbf{r}$ (see Section 19.13). The total work done by the force, as the particle moves along some path from a point A to a point B, is the sum of small amounts for all the small displacements making up the whole path.

$$W = \sum_{A}^{B} \mathbf{F} \cdot d\mathbf{r} = \int_{A}^{B} \mathbf{F} \cdot d\mathbf{r}.$$

An identical calculation can be made using any vector field $\mathbf{F}$ although the result may not have such a well defined meaning. We call the scalar product $\mathbf{F} \cdot d\mathbf{r}$, and the integral, the *activity* of the field.

Example 1. If $\mathbf{F} = \mathbf{E}$ is the electrostatic field, then the activity $\int \mathbf{E} \cdot d\mathbf{r}$ is the electromotive force from A to B.

In evaluating such an integral using Cartesian coordinates (x, y, z),

$$\mathbf{F} \cdot d\mathbf{r} = \alpha\, dx + \beta\, dy + \gamma\, dz,$$

where α, β, γ are the Cartesian components of $\mathbf{F}$ in the x, y and z directions. The integral is therefore the line integral discussed in Section 16.16 and is evaluated in the same way. The value usually depends on the path from A to B.

For the reverse motion from B to A along the same path, every $d\mathbf{r}$ changes sign, so that the integral, or the activity, also changes sign.

19.19. CONSERVATIVE FIELDS. In mechanics, a force field is conservative if the work done by the force on a particle which moves from A to B is independent of the path.

The same definition is applied to all fields, that they are conservative when the activity,

$$\int_A^B \mathbf{F} \cdot d\mathbf{r},$$

is independent of the path from A to B.

This can be expressed in another form. Consider two paths ACB and ADB, and let $W(ACB)$ and $W(ADB)$ be the activities along them. Then the difference is zero.

$$W(ACB) - W(ADB) = W(ACB) + W(BDA) = W(ACBDA) = 0.$$

Thus the activity (or work) along the closed path $ACBDA$ is zero. This is usually written

$$\oint \mathbf{F} \cdot d\mathbf{r} = 0.$$

19.20. IRROTATIONAL FIELDS. In Section 16.16 we considered the limitations on the field, when it is conservative only near a point. The conclusion, using Cartesian coordinates, was that this is only true if $\mathbf{F} \cdot d\mathbf{r}$ is an exact differential; in full

$$\alpha\, dx + \beta\, dy + \gamma\, dz = -\, d\phi,$$

i.e. $$\alpha = -\,\partial\phi/\partial x, \quad \beta = -\,\partial\phi/\partial y, \quad \gamma = -\,\partial\phi/\partial z.$$

It is customary to insert a minus sign, so that the force is in the direction of decreasing potential ϕ, but it is not necessary to do so.

In vector form, which is preferable because it is true in any system of coordinates,

$$\mathbf{F} = -\,\text{grad}\,\phi.$$

For a path which crosses a region where this is true,

$$\int_A^B \mathbf{F} \cdot d\mathbf{r} = \int_A^B -\, d\phi = \phi_A - \phi_B.$$

In such a region the force field is said to be irrotational. A conservative field is one which is irrotational everywhere.

Example 1. The gravitational force is a conservative field, and so is the electrostatic force field.

Example 2. The forces between atoms or molecules are electrical in origin and have a potential in the static approximation where radiation terms are neglected.

Example 3. The magnetic field H due to a loop of wire carrying an electric current is irrotational in the space outside the wire, but it is not in the space occupied by the wire.

The condition that a field is irrotational, i.e. that $\mathbf{F} \cdot d\mathbf{r}$ is an exact differential, was given in Section 16.19. It is that

$$\frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} = 0, \quad \frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x} = 0, \quad \frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y} = 0.$$

It can be shown that the three quantities on the left-hand sides of these equations are the components of a vector, in the x, y and z directions respectively. This vector is known as the curl or rotation of the field $\mathbf{F}$ and is written rot $\mathbf{F}$, or more usually curl $\mathbf{F}$. Its Cartesian components are*,

$$(\text{curl } \mathbf{F})_x = \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}; \quad (\text{curl } \mathbf{F})_y = \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x};$$

$$(\text{curl } \mathbf{F})_z = \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}.$$

19.21. CIRCULATION. To explain the name rotation or curl, given to the vector introduced in the previous section, we shall consider a field of flow, e.g. the vector $\mathbf{v}$ describing the velocity of a moving fluid. The circulation around a closed curve is defined as,

$$\Gamma = \oint \mathbf{v} \cdot d\mathbf{r}.$$

The name activity is hardly applicable, because $\mathbf{v} \cdot d\mathbf{r} = v_t \, ds$, where v_t is the component of the velocity along the tangent to the curve, but it gives what is vaguely described as the flow along and around the closed curve.

Circulation is associated with rotation. Consider a liquid rotating in a bucket as a rigid body, with angular velocity ω about a vertical

* Because of the form of these components, the notation $\nabla \wedge \mathbf{F}$ is widely used, where ∇ has Cartesian components $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$. It can, however, be misleading when Cartesian components are not used.

axis. The velocity at a distance r from the axis is $v = \omega r$, along the tangent to a circle of radius r about the axis.

For this circle the circulation is

$$\Gamma = \Sigma v_t \, ds = \Sigma \omega r \, . \, ds = \omega r(\Sigma \, ds),$$

since the velocity is constant along the circle. Finally $\Sigma \, ds = 2\pi r$, the circumference of the circle, so that

$$\Gamma = \omega r \, . \, 2\pi r = 2\omega(\text{area of circle}).$$

The same result, circulation $= 2\omega$(area), can be proved for any horizontal closed curve, for this type of motion.

A similar result can be obtained for more general motions than this simple one, provided that the area is small. Around a small area dS, near a point P, having its normal in the direction n, the circulation is

$$d\Gamma = 2\omega_n \, dS,$$

where ω_n can be calculated and is independent of the size and shape of the area, provided only that it is small. From the vector properties of the areas, it can be proved that ω_n varies with the direction of the normal to the area as if it were the component of a vector. This vector is called the vorticity vector $\boldsymbol{\omega}$ for the flow.

Comparing this sort of calculation with that made in Sections 16.11 and 16.16, in a very similar problem, it can be shown that

$$\boldsymbol{\omega} = \tfrac{1}{2} \operatorname{curl} \mathbf{v}.$$

Thus the component of the curl of a vector field in a given direction is calculated from the circulation around a small area normal to the given direction, the shape of the area depending on the position coordinates being used.

19.22. Flux, or Flow across an Area. In dealing with fields of flow, such as the velocity of a moving fluid, a natural quantity to calculate is the flow across an area.

In Section 19.15 it was shown that the volume flow of a fluid across an element of area $d\mathbf{S}$ was given by the scalar product $(\mathbf{v} \, . \, d\mathbf{S})$: the flow across a large area S is therefore obtained by dividing it into small parts, and summing the flows across these parts.

$$\text{Flow} = \Sigma v_n \, dS = \int\int_S v_n \, dS = \int\int_S \mathbf{v} \, . \, d\mathbf{S}.$$

Since v_n is positive for flow one way across the area and negative for flow the other way across the area, it follows that the integral is

positive when there is flow one way and negative when there is flow the other way across the area, and that it gives, in general, the net flow across the area.

In the same way the flux of any vector field **F** across an area is defined as

$$\Sigma F_n \, dS = \int\int_S F_n \, dS = \iint \mathbf{F} \cdot d\mathbf{S}.$$

19.23. Divergence of a Vector Field (div **F**). When the surface S is closed, i.e. it is the bounding surface of a volume V, the surface integral $\iint F_n \, dS$ represents the net flux out of the volume, when the normals are drawn outwards from the enclosed volume, or the net flux in when the normals are drawn inwards.

The divergence of the vector field at any point P is defined as the net flux, per unit volume, from a small volume near the point P: that is, given a small volume V around P with a surface S,

$$(\text{div } \mathbf{F})_P = \lim_{S, V \to 0} \left\{ \frac{1}{V} \iint F_n \, dS \right\}.$$

This is independent of the shape of the volume.

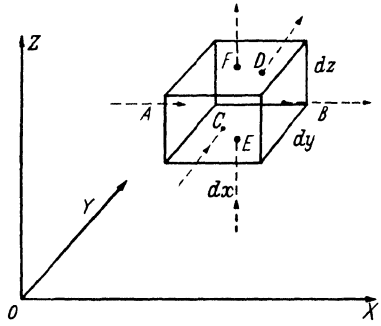

Figure 19.07

We can calculate the scalar div **F**, in Cartesian coordinates, using a small rectangular block with sides dx, dy, dz parallel to the axes (*Figure 19.07*). The flux out of the face B less the flux in across the face A is

$$(F_x)_B \, dy \, dz - (F_x)_A \, dy \, dz = (\partial F_x/\partial x)_P \, dx \, dy \, dz.$$

The flux out across D less the flux in across C is

$$(F_y)_D \, dz \, dx - (F_y)_C \, dz \, dx = (\partial F_y/\partial y)_P \, dy \, dz \, dx.$$

The flux out across F less the flux in across E is

$$(F_z)_F\,dx\,dy - (F_z)_E\,dx\,dy = (\partial F_z/\partial z)_P\,dz\,dx\,dy.$$

Adding these three and dividing by the volume $dx\,dy\,dz$,

$$\operatorname{div}\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

19.24. SOLENOIDAL FIELDS. A vector field is solenoidal when $\operatorname{div}\mathbf{F} = 0$, or the flux across all closed surfaces is zero.

The simplest example of a solenoidal field is the flow of an incompressible fluid. Since the amount of fluid in a given region is fixed, the flux into the region across one part of the boundary is balanced by the flux out across the rest.

Another field which is solenoidal is the magnetic induction **B** and a third is the electric induction **D** when there are no electric charges.

The properties of solenoidal fields can be illustrated in another way. In a vector field **F**, lines can be drawn in space which are in the direction of **F** at every point through which they pass. For a flow field these lines are called streamlines, and for a field of force they are called lines of force.

A tube of force is a tube formed from the lines of force passing through a closed curve. There is no flux across the curved surface of such a tube because **F** is everywhere along the surface and its normal component is zero. The only flux in or out of the tube between two sections A and B is that across these two sections. For a solenoidal field, therefore, the flux across A must be equal to that across B. This is true wherever the section B is cut: hence the flux across any section of a tube has a unique value which is called the strength of the tube.

This means that the tube cannot end in space at a place where **F** is finite. Its cross section at the end would tend to zero and with it the flux across the section there. Thus these tubes, for a solenoidal field, either go to infinity, or form closed rings, or end at a region where **F** is infinite. These regions are the sources of the field.

19.25. VECTOR FIELDS IN GENERAL. The types of vector field discussed so far have been irrotational or solenoidal. The first implies that there is a scalar potential ϕ;

$$\operatorname{curl}\mathbf{F} = 0, \text{ then } \mathbf{F} = -\operatorname{grad}\phi.$$

The second can be proved to imply that there is a vector potential **A**;

$$\operatorname{div}\mathbf{F} = 0, \text{ then } \mathbf{F} = \operatorname{curl}\mathbf{A}.$$

Many fields are neither, but it is generally true that any vector field can be expressed as the sum of two parts, one irrotational and the other solenoidal.

This division into two parts is not unique, because the two conditions are not mutually exclusive. It is quite possible for a vector field to be both irrotational and solenoidal.

19.26. DOUBLE DERIVATIVES. Since grad ϕ is a vector field, we form

$$\nabla^2\phi = \text{div (grad } \phi) \text{ and curl (grad } \phi) = 0.$$

The last equation has already been used. The first, in Cartesian form, can be obtained as follows:

$$\mathbf{F} = \text{grad } \phi \text{ and has components } \frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z};$$

$$\text{div } \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z};$$

Hence
$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}.$$

This combination of double derivatives is called the Laplacian of ϕ.

Since curl $\mathbf{F}$ is a vector field, we can form

$$\text{div (curl } \mathbf{F}) = 0,$$

and
$$\text{curl (curl } \mathbf{F}) = \text{grad (div } \mathbf{F}) - \nabla^2\mathbf{F},$$

where $\nabla^2\mathbf{F}$ has e.g. Cartesian components ∇^2F_x, ∇^2F_y, ∇^2F_z.

All these results are easily verified, using Cartesian components.

19.27. GENERALIZED VECTORS. Any two variables u, v can be plotted in a u/v diagram as the Cartesian coordinates of a point P, of which the position vector $\mathbf{OP}$ is a vector having Cartesian components u and v. Any changes in u and v can now be discussed in terms of the changes in this position vector.

Similarly, any three functions can be treated as the components of a three-dimensional vector.

Example 1. Consider the simultaneous equations

$$(dc_1/dt) = \beta c_3 - \gamma c_2,$$
$$(dc_2/dt) = \gamma c_1 - \alpha c_3,$$
$$(dc_3/dt) = \alpha c_2 - \beta c_1.$$

If (c_1, c_2, c_3) are the components of a vector $\mathbf{c}$ and (α, β, γ) are the components of a vector $\boldsymbol{\omega}$, then these equations are the Cartesian components of the vector equation

$$d\mathbf{c}/dt = \boldsymbol{\omega} \wedge \mathbf{c}, \quad \text{i.e. } d\mathbf{c} = \boldsymbol{\omega} \wedge \mathbf{c}\, dt.$$

The second equation shows that any small change in $\mathbf{c}$ is perpendicular both to $\boldsymbol{\omega}$ and to $\mathbf{c}$. Thus in a diagram where a point P represents the vector $\mathbf{c}$, the point moves with time on a circle of unknown radius (c) with O as centre, in a plane perpendicular to $\boldsymbol{\omega}$, moving clockwise about the direction of $\boldsymbol{\omega}$.

EXAMPLES XIX

1. In Cartesian coordinates, the equation of a surface is $z = f(x, y)$. From the differential equation, $dz = (\partial f/\partial x)dx + (\partial f/\partial y)dy$, deduce that the direction of the normal at any point P on the surface is that of a vector with components $(\partial f/\partial x, \partial f/\partial y, -1)$.

What form would this result have if the equation of the surface were $\phi(x, y, z) = 0$.

2. Find grad V, when $V = (x^2 + y^2 + z^2)^{-1/2}$ and

$$V = \{(x - a)^2 + y^2 + z^2\}^{-1/2}.$$

Illustrate the results.

3. Atoms lie at the points of a rectangular lattice with coordinates $x = la$, $y = mb$, $z = nc$ where l, m, n are integers. Three atoms A, B, C lie at the points $(l_1, 0, 0)$, $(0, m_2, 0)$, $(0, 0, n_3)$ respectively. Using symbols $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ for unit vectors along Ox, Oy, Oz, write down vectors **AB**, **BC**, **CA** representing the lines joining them, and a vector normal to the plane ABC (e.g. $\mathbf{AB} \wedge \mathbf{AC}$).

4. The sides OA, OB and OC of a unit cube are three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. Obtain vectors representing the diagonals of the faces, and the four body diagonals joining opposite corners. What is the angle between any two of the body diagonals? (Use the scalar product.)

5. The position vector $\mathbf{r}$ of a moving point varies with time t. Draw a diagram to illustrate its change $\Delta\mathbf{r}$, in magnitude and direction, in a short time Δt.

Prove that $\dfrac{d\mathbf{r}}{dt} = \lim\limits_{\Delta t \to 0} \dfrac{\Delta \mathbf{r}}{\Delta t}$ is a vector in the direction of motion of the moving point, with a magnitude equal to its speed.

6. An electron of charge e moving with velocity $\mathbf{v}$ in a magnetic field $\mathbf{H}$, is acted on by a force $(e/c)\mathbf{v} \wedge \mathbf{H}$. Show that no work is done on the charge as it moves and its speed is constant.

7. Show that the work done by a force $\mathbf{F}$ in a small displacement equals F times the component of the displacement in the direction of the force.

A central force is a force at a point P directed along the line OP from the origin. Show that the work done in a small displacement is $F\,dr$, where dr

is the increase in distance from O. Deduce that the work done by a central force of magnitude $f(r)$ in a displacement from A to B is $\int_a^b f(r)\,dr$ where $a = OA$, $b = OB$, and that the force has a potential.

8. The magnetic field due to an electric current i in a long straight wire is derived from a vector potential $\mathbf{A}$ by the formula $\mathbf{H} = \text{curl } \mathbf{A}$, where

$$A_x = A_y = 0,\ A_z = (i/c) \log (x^2 + y^2).$$

Find the Cartesian components of $\mathbf{H}$ and show that $\mathbf{H} = -\text{grad } \phi$ where $\phi = \tan^{-1}(x/y)$.

REFERENCES

There are many books on vectors and vector methods. Most are more suitable for physicists. The author suggests the following three:

HAGUE, B. *Vector Methods.* 4th Edn. Methuen Monographs, London, 1950.
GANS, R. *Vector Analysis.* Blackie and Sons, London, 1932.
WEATHERBURN, C. E. *Elementary Vector Analysis.* G. Bell and Sons, London, 1920.

CHAPTER 20

PARTIAL DIFFERENTIAL EQUATIONS

Introduction—Diffusion equation—Diffusion between plates—Heat equation—Cooling sphere—Temperature oscillations along a bar—Separation of variables—Two-dimensional heat flow using polar coordinates—Propagation or wave equations—Conclusion

20.01. INTRODUCTION. The solution of partial differential equations by simple methods is not easy, because the fitting of the solution to the initial conditions is a difficult operation. The general method is to obtain as many solutions of the equation as possible, of a type likely to be required, and then to combine them to give the actual solution. This is possible because most theories give linear equations. It can be made more formal as the method of separating the variables, or in other ways.

20.02. DIFFUSION EQUATION. In Section 15.16 a brief mention was made of diffusion problems which were time-independent. It is now possible to discuss time variations.

During the flow of particles or molecules in solution, the total mass is conserved. The flow into a given volume across part of the surface and the flow out across the rest either balance, or the difference equals the change of mass inside and causes a change in the concentration.

If the concentration is c the amount of material in a small volume dv is $c\,dv$. If the flux vector $\mathbf{F}$ describes the flow of particles, the net flow from this small volume per unit time is $(\operatorname{div}\mathbf{F})dv$. Equating this to the rate of decrease of the amount inside leads to an equation

$$(\operatorname{div}\mathbf{F})dv = -\partial(c\,dv)/\partial t = -(\partial c/\partial t)dv,$$

i.e.

$$\operatorname{div}\mathbf{F} = -\partial c/\partial t.$$

When the flow is governed by diffusion, it satisfies Fick's law, as stated in Section 15.16. The general form of this law is

$$\mathbf{F} = -D\operatorname{grad} c,$$

where D is the diffusion coefficient. These two equations are combined to eliminate $\mathbf{F}$.

$$\operatorname{div}(D\operatorname{grad} c) = \partial c/\partial t.$$

This is the Diffusion Equation.

When D is a constant it simplifies to

$$D \operatorname{div} (\operatorname{grad} c) = D\nabla^2 c = \partial c/\partial t.$$

The form of $\nabla^2 c$ depends on the coordinates used to describe the flow and the form of the flow. When the flow is in the x direction due to a concentration variation in that direction, then

$$F_x = -D(\partial c/\partial x).$$

Also div $\mathbf{F} = (\partial F_x/\partial x)$, since $\mathbf{F}$ has only one component, so that the diffusion equation becomes

$$D(\partial^2 c/\partial x^2) = \partial c/\partial t.$$

20.03. Diffusion between Plates. Consider a solution between two plates at $x = 0$ and $x = d$, the concentration of particles depending on x. Then

$$D(\partial^2 c/\partial x^2) = \partial c/\partial t.$$

For steady flow $\partial c/\partial t = 0$, and this equation becomes $d^2c/dx^2 = 0$,

i.e. $$c = Ax + B.$$

The numbers A and B depend on conditions at the boundaries.

When the initial concentration is not of this form, the flow takes time to adjust itself. For example, when the solute cannot pass through the plates,

$$D(\partial c/\partial x) = 0 \text{ at } x = 0 \text{ and } x = d, \quad (\text{all } t),$$

and the steady state is $c =$ constant. If the initial concentration is $c = f(x)$, so that the mean concentration initially is

$$c_0 = \frac{1}{d}\int_0^d f(x)dx,$$

then presumably c_0 is the steady value to which the concentration tends as time goes on. We therefore write

$$c = c_0 + c',$$

where $c' \to 0$ as $t \to \infty$, and also satisfies the diffusion equation. This equation involves $\partial c/\partial t$, so we assume that

$$c' = e^{-\lambda t}\phi(x).$$

To determine λ and $\phi(x)$ we equate

$$D\frac{\partial^2 c'}{\partial x^2} = De^{-\lambda t}\frac{d^2\phi}{dx^2} \quad \text{and} \quad \frac{\partial c'}{\partial t} = -\lambda e^{-\lambda t}\phi,$$

$$\therefore \qquad d^2\phi/dx^2 = -p^2\phi, \quad \text{where } \lambda = Dp^2.$$

This equation has the general solution $\phi = A \cos px + B \sin px$, whence

$$c' = e^{-\lambda t}(A \cos px + B \sin px).$$

The conditions that $\partial c/\partial x = 0$ at $x = 0$ and at $x = d$ become

$$0 = e^{-\lambda t}(-Ap \sin pd + Bp \cos pd),$$
$$0 = e^{-\lambda t}(0 + Bp),$$

so that $B = 0$, and either $A = 0$ or $\sin pd = 0$. Rejecting $A = 0$ which leads to no solution at all, we are left with the condition

$$\sin pd = 0 \quad \text{or} \quad p = n\pi/d, \quad \lambda = n^2\pi^2 D/d^2, \quad (n = 1, 2, 3, \ldots).$$

Thus λ is determined, but not uniquely, and

$$c' = A_n e^{-n^2\pi^2 Dt/d^2} \cos (n\pi x/d), \quad (n = 1, 2, 3, \ldots).$$

We have obtained a number of solutions but unfortunately none of them agrees with the remaining condition, that

$$c' = c - c_0 = f(x) - c_0, \text{ at } t = 0,$$

unless $f(x)$ has a special form.

However, it is not necessary to look for further solutions. As the diffusion equation is linear, not only do these solutions separately satisfy it but their sum also satisfies it, i.e.

$$c' = c - c_0 = \sum_n A_n e^{-n^2\pi^2 Dt/d^2} \cos (n\pi x/d).$$

This gives, when $t = 0$, between $x = 0$ and $x = d$

$$c - c_0 = f(x) - c_0 = \sum_{n=1}^{\infty} A_n \cos (n\pi x/d),$$

and by Fourier's theorem it is possible to find numbers A_n for this to be true. Their values are (Section 13.27)

$$A_n = \frac{1}{d}\int_0^d f(x) \cos \frac{n\pi x}{d}\, dx.$$

(N.B. The term $n = 0$ in the Fourier series is the number c_0.)

The whole problem is now solved. The concentration has been expressed as c_0 together with terms tending to zero as time increases. The only criticism which can be levelled at the solution is that the series may converge slowly when t is small, and the numerical evaluation may be tedious. This is a valid criticism, but an alternative method to be used for small values of t takes us beyond the limits of this volume.

20.04. HEAT EQUATION. According to Fourier's law, heat flow by conduction is measured by a flux vector $\mathbf{F} = -\kappa \operatorname{grad} \theta$ where κ is the thermal conductivity and θ is the temperature.

Now $-\operatorname{div} \mathbf{F}\, dv$ is the rate of flow of heat *into* a small volume dv, and if ρ is the density and c the specific heat the corresponding rate of increase of temperature is

$$\frac{\partial \theta}{\partial t} = \frac{1}{\rho c\, dv}(-\operatorname{div} \mathbf{F}\, dv) = -\frac{1}{\rho c} \operatorname{div} \mathbf{F},$$
$$= (1/\rho c) \operatorname{div} (\kappa \operatorname{grad} \theta).$$

If the material is homogeneous, κ is independent of position and

$$\frac{\partial \theta}{\partial t} = D \operatorname{div} (\operatorname{grad} \theta) = D \nabla^2 \theta, \quad \text{where } D = \kappa/\rho c.$$

This equation is identical in form with the diffusion equation and has identical solutions when D is constant in both cases. Hence problems of heat flow and diffusion correspond even to their time variations.

20.05. COOLING SPHERE. This problem was solved in Section 15.17 assuming that κ is so large and the rate of heat loss so small that the temperature in the sphere is uniform. It is now solved exactly.

It is assumed that the initial temperature of the sphere, if it varies at all, only varies with distance r from the centre. If the sphere is isotropic, there is no reason why the temperature should not remain a function of r, so that the flow of heat is then radially from the centre. The flux is

$$F = -\kappa(\partial\theta/\partial r).$$

Consider two spheres, one of radius r and the other of radius $(r + dr)$, both inside the sphere. The heat flow from the region between them across the larger sphere, less the heat flow into this region across the inner sphere, is

$$(4\pi r^2 F)_{r+dr} - (4\pi r^2 F)_r = 4\pi \frac{\partial}{\partial r}(r^2 F) dr.$$

The volume of the region is $4\pi r^2\, dr$, so that for this type of flow

$$\operatorname{div} \mathbf{F} = \frac{1}{r^2} \frac{\partial}{\partial r}(r^2 F).$$

The diffusion equation is

$$D \nabla^2 \theta = \frac{D}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial \theta}{\partial r}\right) = \frac{\partial \theta}{\partial t}.$$

Because of the Leibnitz theorem,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r^2\frac{\partial\theta}{\partial r}\right) = r\frac{\partial^2\theta}{\partial r^2} + 2\frac{\partial\theta}{\partial r} = \frac{\partial^2}{\partial r^2}(r\theta),$$

and the equation can be written

$$D\frac{\partial^2}{\partial r^2}(r\theta) = r\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial t}(r\theta).$$

This is similar in form to the equation $D(\partial^2\theta/\partial x^2) = \partial\theta/\partial t$ for linear flow.

Assuming that θ_0 is constant, and $\theta' = \theta - \theta_0$ we find, as before, (p 339) that there are solutions of the form

$$r\theta' = e^{-\lambda t}(A\cos pr + B\sin pr), \quad \lambda = Dp^2.$$

The boundary condition is that the heat flowing to the outer surface at $r = a$ is lost by radiation, i.e.

$$-4\pi a^2\kappa\left(\frac{\partial\theta}{\partial r}\right)_{r=a} = h(\theta - \theta_0) = h\theta'.$$

A second condition satisfied by the solutions is not immediately obvious. Near the centre of the sphere, where r is very small, the value of θ' is

$$\theta' = e^{-\lambda t}(A + Bpr)/r,$$

which remains finite at $r = 0$ only when $A = 0$. Thus the condition that the temperature is *finite* at the centre is an important restriction on the solutions.

The final condition is that the temperature is correct initially, say,

$$\theta' = f(r) \text{ initially, at } t = 0.$$

To satisfy the second condition the original solution reduces to $(A = 0)$

$$r\theta' = e^{-\lambda t}(B\sin pr),$$

and the first condition requires that

$$-4\pi a^2\kappa e^{-\lambda t}B\frac{\partial}{\partial r}\left(\frac{\sin pr}{r}\right)_{r=a} = he^{-\lambda t}B\left(\frac{\sin pr}{r}\right)_{r=a},$$

or

$$\tan pa = \frac{1}{1 - (h/4\pi a\kappa)}pa = mpa.$$

The solutions of this equation are obtained by plotting the curves

$y = \tan x$ and $y = mx$ (Section 2.02). Corresponding to each positive root there is a possible value of p and of $\lambda = Dp^2 = (\kappa p^2/\rho c)$.

In general the third condition cannot be satisfied by θ' for just one value of p, but only by a sum of the solutions with all values of p,

$$\theta = \sum_n B_n e^{-\lambda_n t} \sin (p_n r/r).$$

It then takes the form

$$\sum_{n=1}^{\infty} B_n \sin p_n r = rf(r).$$

To find a particular coefficient B_m, it is necessary to proceed as in Section 13.10 on Fourier series, that is, to multiply by $\sin p_m r$ and integrate from $r = 0$ to $r = a$. This gives

$$\int_0^a rf(r) \sin p_m r \, dr = \int_0^a (\sum_n B_n \sin p_n r) \sin p_m r \, dr.$$

We must prove that every term B_n in the series, except B_m, gives zero. In fact the integral multiplied by $(p_n^2 - p_m^2)$, which is not zero, is

$$\begin{aligned}(p_n^2 - p_m^2)B_n \int_0^a &\sin p_n r \sin p_m r \, dr \\ &= B_n \Big[p_m \sin p_n r \cos p_m r - p_n \sin p_m r \cos p_n r \Big]_0^a \\ &= (1/a)B_n \sin p_m a \sin p_n a \{p_m a \cot p_m a - p_n a \cot p_n a\}.\end{aligned}$$

The expression in the bracket is zero, because the condition which determines the values of p ensures that $(pa \cot pa)$ has the same value for every root.

$$\therefore \quad \int_0^a rf(r) \sin p_m r \, dr = B_m \int_0^a \sin^2 p_m r \, dr,$$

i.e.
$$B_m = \int_0^a rf(r) \sin p_m r \, dr \Big/ \int_0^a \sin^2 p_m r \, dr.$$

The formula for B_m gives all the coefficients $m = 1, 2, \ldots$ It is no accident that only one term remains after integration as in the Fourier series. It always happens when the solutions satisfy a condition on the boundary which is linear in θ, and independent of the time.

Let us now examine the roots of the equation

$$\tan pa = m \, . \, pa, \quad \text{where } m = 1/(1 - h/4\pi a\kappa).$$

This has already been done in Section 2.02. When $\kappa \ll (h/4\pi a)$ m is

negative but small, and the roots are approximately $pa = n\pi$, or $\lambda = (n^2\pi^2\kappa/\rho ca^2)$. The mass of the sphere is $M = \frac{4}{3}\pi a^3\rho$, so that

$$\lambda = n^2\pi^2\kappa \,.\, 4\pi a/3Mc = \tfrac{1}{3}n^2\pi^2\, 4\pi a\kappa/Mc.$$

As κ increases, m remains negative but gets larger, and the roots diminish in magnitude. When $\kappa = (h/4\pi a)$, m becomes infinite and the nth root has decreased to $pa = (n - \frac{1}{2})\pi$. The decrease in the roots continues as κ increases, even although m is now positive.

Finally, if κ is much larger than $(h/4\pi a)$, i.e. heat conduction within the sphere is much faster than the heat loss from the surface, the conditions are those considered in Section 15.17. The function m is now only a little greater than unity, and the smallest root p is a small number. Using the expansion, $\tan pa = pa + \frac{1}{3}(pa)^3$, when pa is small, the equation for this root is

$$pa + \tfrac{1}{3}(pa)^3 = \frac{1}{1 - h/4\pi a\kappa}\, pa = pa\left(1 + \frac{h}{4\pi a\kappa}\right), \text{ approximately,}$$

$$\text{i.e.} \qquad (pa)^2 = \frac{3h}{4\pi a\kappa} \ll 1 \quad \text{and} \quad \lambda = \frac{\kappa p^2}{\rho c} = \frac{3h}{4\pi a^3\rho c} = \frac{h}{Mc}.$$

The other roots give much larger values of p and λ, the nth root still being approximately $pa = (n - \frac{1}{2})\pi$.

We can now justify the method of that earlier section. As t increases, the terms with large values of p and λ decrease most rapidly, so that for large values of t

$$\theta = \Sigma B_n e^{-\lambda_n t}\frac{\sin p_n r}{r} \to B_1 e^{-\lambda_1 t}\frac{\sin p_1 r}{r},$$

where λ_1 is the smallest value. Using the smallest value and remembering that $p_1 r \leqslant p_1 a \ll 1$, then

$$\theta \fallingdotseq B_1 p_1 e^{-(ht/Mc)},$$

which is the form of solution used previously.

To complete the demonstration the value of B_1 has to be obtained. Since $p_1 a \ll 1$, for the lowest root,

$$B_1 p_1 = \frac{\int_0^a rf(r)\sin p_1 r\, dr}{\int_0^a \sin^2 p_1 r\, dr}\, p_1 = \frac{\int_0^a r^2 p_1 f(r)dr}{\int_0^a p_1^2 r^2\, dr}\, p_1,$$

$$= \frac{\int_0^a r^2 f(r)dr}{\frac{1}{3}a^3} = \frac{1}{\frac{4}{3}\pi a^3}\int_0^a 4\pi r^2 f(r)dr,$$

which is the average temperature of the sphere initially. In fact, it was assumed in Section 15.17 that the temperature of the sphere was uniform, so that this new result is rather more general, although it is an obvious generalization.

20.06. TEMPERATURE OSCILLATIONS ALONG A BAR. An insulated rod, extending from $x = 0$ to $x = L$, is at a uniform temperature T_0. If the temperature at $x = 0$ is now made to oscillate,

$$T = T_0 + \alpha \sin \omega t = T_0 + \text{i.p. of } \alpha e^{i\omega t},$$

we expect the temperature ultimately to oscillate at every point of the bar. This problem is easily solved when $L \to \infty$.

In view of the periodicity ω a trial solution $T = e^{i\omega t + mx}$ is assumed. This satisfies the heat equation $D(\partial^2 T/\partial x^2) = \partial T/\partial t$ if

$$Dm^2 = i\omega, \quad \text{i.e. } m = \pm (1 + i)p, \text{ where } p^2 = \frac{\omega}{2D} = \frac{\rho c \omega}{2\kappa}.$$

Combining these two solutions with (complex) constants,

$$T - T_0 = \text{i.p. of } Ae^{i\omega t + (1+i)px} + Be^{i\omega t - (1+i)px}.$$

Now T should remain finite even as $x \to L \to \infty$, so that A must be zero. Taking the imaginary part of the term in $B = B_1 + iB_2$,

$$T - T_0 = e^{-px}\{B_1 \sin (\omega t - px) + B_2 \cos (\omega t - px)\},$$

which satisfies the condition at $x = 0$ when $B_1 = \alpha$, $B_2 = 0$.

Thus the temperature oscillation at the point x has an amplitude αe^{-px}.

20.07. SEPARATION OF VARIABLES. This is a generalization of the method used so far, which makes it easier to obtain solutions. The essential points are that, amongst all the solutions which are functions of several variables, (*i*) it seeks only those which are products of factors, each a function of one variable, (*ii*) it obtains differential equations for these functions, and (*iii*) from all the solutions obtained this way those are chosen which satisfy certain general conditions in the problem.

For example, the heat equation $\partial^2 T/\partial x^2 = (1/D)\partial T/\partial t$ has solutions of the required form $T = f(x) \cdot g(t)$ if

$$\frac{d^2 f}{dx^2} \cdot g(t) = \frac{1}{D} f(x) \cdot \frac{dg}{dt}.$$

To obtain equations for $f(x)$ and $g(t)$ separately, we use the fact that they are functions of different variables, which can be varied separately. Division by the product $f(x) \cdot g(t)$ puts the equation into the revised form

$$\frac{1}{f(x)}\frac{d^2f}{dx^2} = \frac{1}{Dg(t)}\frac{dg}{dt}.$$

Now, as x varies, the RHS does not change, so that $f(x)$ must be chosen so that the LHS does not change, i.e.

$$\frac{1}{f}\frac{d^2f}{dx^2} = C \quad \text{or} \quad \frac{d^2f}{dx^2} = Cf,$$

where C is a constant. But this also requires that

$$\frac{1}{Dg(t)}\frac{dg}{dt} = C \quad \text{or} \quad \frac{dg}{dt} = CDg.$$

Thus we have separate equations for f and g. Their solutions are

$$f = Ae^{x\sqrt{C}} + Be^{-x\sqrt{C}}, \quad g = Ke^{CDt}$$

$$\text{i.e. } \theta = e^{CDt}(A'e^{x\sqrt{C}} + B'e^{-x\sqrt{C}}),$$

the numbers AK and BK being replaced by A' and B'.

This expression holds for any value of C, positive, negative, zero or complex, and sums of such expressions for various values of C are also solutions. In Section 20.03 we used solutions with $C = 0$ and C negative, and in Section 20.06 a solution where C was complex.

20.08. TWO-DIMENSIONAL HEAT FLOW USING POLAR COORDINATES. Let there be heat flow around an axis and radially from it; the temperature is assumed not to vary along the axis, so that there is no flow in that direction. Taking a section of unit length along the axis and using the element of volume shown in *Figure 20.01*, the reader should be able to prove that

$$\text{div } S = \frac{1}{r\,d\theta\,dr}\left\{\frac{\partial}{\partial r}(rS_r\,d\theta)dr + \frac{\partial}{\partial\theta}(S_\theta\,dr)d\theta\right\},$$

$$= \frac{1}{r}\frac{\partial}{\partial r}(rS_r) + \frac{1}{r}\frac{\partial S_\theta}{\partial\theta},$$

where $S_r = -\kappa\dfrac{\partial T}{\partial r}, \quad S_\theta = -\kappa\dfrac{\partial T}{r\partial\theta}.$

The radial and transverse flux of heat are proportional to the rate of decrease of temperature with distance. Hence

$$\kappa \nabla^2 T = -\operatorname{div} S = \kappa \left\{ \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial T}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 T}{\partial \theta^2} \right\} = \frac{1}{\rho c} \frac{\partial T}{\partial t}.$$

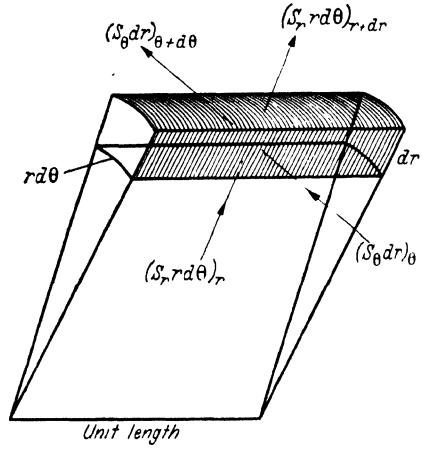

Figure 20.01

We solve this equation by the separation of variables, assuming a solution of the form $T = f(r) \,.\, g(\theta) \,.\, h(t)$. We substitute this into the equation and find that

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{df}{dr} \right) gh + \frac{1}{r^2} f \,.\, \frac{d^2 y}{d\theta^2} \,.\, h = \frac{1}{D} fg \,.\, \frac{dh}{dt}, \qquad D = \frac{\rho c}{\kappa}.$$

Dividing by $T = (fgh)$,

$$\frac{1}{rf} \frac{d}{dr} \left(r \frac{df}{dr} \right) + \frac{1}{r^2 g} \frac{d^2 g}{d\theta^2} = \frac{1}{Dh} \frac{dh}{dt}.$$

Variation of t leaves the LHS unchanged, so that

$$\frac{1}{Dh} \frac{dh}{dt} = -C; \quad h = e^{-CDt},$$

where C is any constant. Therefore

$$\frac{1}{rf} \frac{d}{dr} \left(r \frac{df}{dr} \right) + \frac{1}{r^2 g} \frac{d^2 g}{d\theta^2} = -C.$$

Multiply by r^2 to separate the variables, and rearrange,

$$\frac{r}{f}\frac{d}{dr}\left(r\frac{df}{dr}\right) + Cr^2 = -\frac{1}{g}\frac{d^2g}{d\theta^2}.$$

The LHS does not vary with θ; hence

$$\frac{1}{g}\frac{d^2g}{d\theta^2} = -\alpha, \quad \text{i.e.} \quad \frac{d^2g}{d\theta^2} = -\alpha g,$$

where α is any constant. Finally

$$\frac{r}{f}\frac{d}{dr}\left(r\frac{df}{dr}\right) + Cr^2 = \alpha,$$

i.e.

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{dr}{df} + \left(C - \frac{\alpha}{r^2}\right)f = 0,$$

which is Bessel's equation of order $\alpha^{1/2}$. The whole solution is

$$T = f(r)g(\theta)e^{-CDt},$$

where

$$\frac{d^2g}{d\theta^2} = -\alpha g; \quad \frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} + \left(C - \frac{\alpha}{r^2}\right)f = 0.$$

For the steady state $C = 0$, and for most other problems it is positive, but it can be negative or complex. The allowed values of α also depend very much on the particular problem. For example, assume that the flow is across a circular tube around the axis. Then θ can go from 0 to 2π and, much more important, the solutions must be periodic with a period 2π in θ. Now the equation for $g(\theta)$ only gives periodic solutions when α is positive, or zero. (If α is negative it has exponential solutions which are not periodic.) If $\alpha = p^2$ the solutions are of the form,

$$g = A\cos(p\theta - \beta).$$

These have the correct period of 2π only when p is an integer n, $n = 0, 1, 2, \ldots$ ($n = 0$ gives the solution independent of θ).

For this problem therefore

$$T = Ae^{-CDt}\cos(n\theta - \beta)f(r),$$

where

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} + \left(C - \frac{n^2}{r^2}\right)f = 0.$$

When $C = 0$ this equation for $f(r)$ has the solution

$$f = Ar^n + B/r^n,$$

and when C is positive it is Bessel's equation of order n with the solution

$$f = AJ_n(r\sqrt{C}) + BY_n(r\sqrt{C}).$$

If the temperature is finite at the origin, the coefficients B must be zero.

20.09. PROPAGATION OR WAVE EQUATIONS. In this section flows are considered which do not satisfy the Diffusion Law.

Consider particles in a fluid, the concentration at time $t = 0$ being $c = f(x)$. Let us assume that they all move in the x direction with the same velocity u. Then at any later time t all the particles have moved the same distance ut, so that the concentration of particles at the point x is that of particles initially at the point $(x - ut)$. Hence, at any time t,

$$c = f(x - ut).$$

This function describes the propagation of concentration values with velocity u in the x direction.

Let us now express this in terms of the conservation equation $\partial S/\partial x = -\partial c/\partial t$ given in Section 20.02. The flux of particles in the x direction is $S = cu$ per unit area, so that

$$\partial c/\partial t + u(\partial c/\partial x) = 0,$$

which is the differential equation having the solution $c = f(x - ut)$.

Similarly the expression $c = g(x + ut)$ is the solution of the equation

$$\partial c/\partial t - u(\partial c/\partial x) = 0.$$

It is easily verified that the combined expression,

$$c = f(x - ut) + g(x + ut),$$

satisfies the more complicated equation,

$$\partial^2 c/\partial t^2 = u^2(\partial^2 c/\partial x^2).$$

This is known as the Wave Equation.

The wave equation is obtained when

$$\frac{\partial S}{\partial x} = -\frac{\partial c}{\partial t}, \quad \frac{\partial S}{\partial t} = -u^2\frac{\partial c}{\partial x}.$$

The first of these is the Conservation Law. The second is an alternative to diffusion when $(\partial c/\partial x)$ is in the nature of a mechanical force and is therefore equated to the rate of change of flow with time.

This implies, for example, that uniform concentration means steady flow (rather than no flow, as in diffusion). It arises in the small compression waves in a gas, which can be superposed on the general motion of the gas. The function c is then the density of the gas, and $u^2(\partial c/\partial x)$ is a pressure gradient.

20.10. CONCLUSION. As in other chapters of this book, there are equations which could be discussed more fully, and some which have not been discussed at all. The wave equation has whole books devoted to it, and Laplace's equation $\nabla^2\phi = 0$ has not been discussed at all. I hope that this chapter will enable the discerning reader to understand more comprehensive books on the subject.

EXAMPLES XX

1. Owing to diffusion, solute flows steadily from one plate ($x = 0$) where the concentration is c_1 to another ($x = d$) where the concentration is c_2. Find the concentration between the plates. The concentration at $x = d$ is suddenly reduced to zero and remains zero. Find the concentration between the plates as a function of x and t, and the flow across the plate $x = d$ as a function of time.

2. A bar is heated to a temperature θ_0 and then insulated, except at the ends where the temperature is reduced to zero. The length is L and the cross section is uniform. Assuming heat conduction across the ends, calculate the temperature everywhere as a function of time.

3. State the diffusion problem which corresponds to the heat problem of Section 20.05.

4. In a spherical cell of radius a one substance of concentration c is converted into another at a rate kc per unit volume, and is replaced by diffusion across the boundary $r = a$. Show that the equations governing the concentration are

$$\partial c/\partial t = -\operatorname{div} S - kc; \quad S = -D \operatorname{grad} c.$$

Assuming radial flow, $\{\nabla^2 = \partial^2/\partial r^2 + (2/r)(\partial/\partial r)\}$, find the concentration everywhere in the steady state, assuming that the concentration c_0 on the boundary $r = a$ is constant. Determine the variation with time if the boundary becomes impermeable to the substance.

5. Heat flow along a bar which is incompletely insulated along its length is given by

$$\frac{\partial\theta}{\partial t} = D\frac{\partial^2\theta}{\partial x^2} - h(\theta - \theta_0),$$

Terms due to the radial variation in temperature can be neglected when h is small.

Find solutions of this equation for a bar of length l, insulated at one end and maintained at a given temperature at the other.

6. Show that the diffusion equation $\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial c}{\partial x}\right)$, where D is a function of c, has a solution of the form $c = f(u)$ where $u = xt^{-1/2}$, if

$$\frac{d}{du}\left(D\frac{dc}{du}\right) = -\tfrac{1}{2}u\frac{dc}{du}.$$

Prove that one solution of this equation, when D is constant, is

$$c = \tfrac{1}{2}a\left\{1 - \frac{2}{\sqrt{\pi}}\int_0^{u/2\sqrt{D}} e^{-v^2}\,dv\right\},$$

and verify that this solution, in the limit $t \to 0$, gives $c = 0$ where x is positive, and $c = a$ where x is negative.

What diffusion problem does this illustrate?

7. Show that Laplace's equation in polar coordinates,

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} = 0,$$

has solutions of the form $V = (Ar^n + Br^{-n})\sin(n\theta + \alpha)$.

8. Solve the wave equation $\partial^2\phi/\partial t^2 = u^2\,\partial^2\phi/\partial x^2$ by the method of separating the variables, and find solutions so that either (*i*) $\phi = 0$ at $x = 0$ and $x = a$, or (*ii*) $\phi = 1$ when $x = 0$ and $\phi \to 0$ as $x \to \infty$.

REFERENCES

The author has found helpful treatments of the subject matter of this chapter in the following books:

PIAGGIO, H. T. H. *Elementary Treatise on Differential Equations.* G. Bell and Sons, London, 1949.

PIPES, L. A. *Applied Mathematics for Engineers and Physicists*, Chapters XV to XVIII. McGraw-Hill Book Co., New York, 1946.

SHERWOOD, T. K. and REID, C. E. *Applied Mathematics in Chemical Engineering*, Chapter VI. McGraw-Hill Book Co., New York, 1939.

CARSLAW, H. S. *Introduction to the Mathematical Theory of the Conduction of Heat.* Macmillan and Co., London, 1921.

JAEGER, J. C. *An Introduction to Applied Mathematics.* Clarendon Press, Oxford, 1951.

BARRER, R. M. *Diffusion in and through Solids.* Cambridge University Press, London, 1941.

COULSON, C. A. *Waves.* Oliver and Boyd, Edinburgh, 1941.

MARGENAU, H. and MURPHY, G. M. *Mathematics of Physics and Chemistry*, Chapter 13. Van Nostrand, New York, 1943.

INDEX

A page number in italics gives the main position of the item to which it refers.

Absolute convergence, 181
Absorption of light, 224
Activity in a field, 328
Adiabatic expansion, 138
Approximations, 5, 182
— linear, 191
— quadratic, 191
— to functions, 191
Arbitrary constants, 3, 132
Arc lengths, 108, 136
— mean centre, 137
Areas, approximate values, 126
— as vectors, 320, 325
— by integration, 134
— mean centre, 280
— moments of, 281–2
— of curved surfaces, 137
— properties of, 279
Argand diagram, 213
Argument, of complex number, 217
Asymptotic expansions, 197
Atomic theory, 161
Averages (*see* Mean values), 124
Axial coordinates, 291

Bernoulli, 224
— numbers, 141
Bessel functions, $J_n(x)$, $Y_n(x)$, 246–7
Bessel's equations, 243
Binomial theorem, 70, 188
Black-body radiation, 103

Calculations, estimate of length, 183
Cartesian coordinates, in plane, change of origin, 105
— — rotation of axes, 106
— in space, 290
Centre of curvature, 110
Centre of pressure, 280, 282
Chain reactions, 91, 241
Chain rule, 44
Change of variable, 152
— differentiation, 43
— in integration, 145
— multiple integrals, 286
Changes, *35*
— in related functions, 39
—, small, 40, 44
Characteristic equation, *304*
Characteristic roots, 304
— variation method, 309
Chemical reactions (*see* Reactions), 225
Circle, coordinate geometry, 119–20
Circular functions, *79*
— (complex), 214, 219
— derivatives, 85
— expansions in terms of, 204–5
— graphs, 81–2
— identities, 80
— integrals, 163, 166–8
— inverse functions, 86, 88
— — addition of, 88
— solution of equations, 83
Circulation, 330
Cofactors, 312
Compatible simultaneous equations, 304
Complementary function, 235
Complex numbers, 210
— conjugate, 212
— modulus and argument, 217
— powers, 220
— roots, 220
Configuration space, *290*
Conic sections, Cartesian, 119
— polar equations, 119
— (*see also* Ellipse, Hyperbola, Parabola, Circle)
Conjugate complex numbers, 212
Conservation laws, 337, 348
Conservative forces and fields, 329
Constants, arbitrary, 3
— of integration, 130, 132
Continuity, 31
Continuous variable, 1
Contour maps, 255
Convergence, *26*
— absolute, 181
— of series, 179
— radius of, 186
— tests, 29
Cooling, Newton's law of, 232
Coordinate geometry, *117*
— of lines, 118
— of points, 117
Coordinate, arc lengths (*see* Arc)
— (*see* Cartesian, Polar), generalized, 116
— polar, 9
Corrections, *18*, 40, 179
— first order, 40
— second order, 40
— simultaneous, 256

Critical damping, 239
Crystals (*see* Molecular properties), 207
Cubic equations, 89
Curl, 330
Curvature, 9, 53, *109*
Curved surfaces, areas of, 137
Curves, in polar coordinates, 110–12
— properties of, 105 ff
— sketching of, 9, 10
Cylindrical polar coordinates, 291

DAMPING factor, 238
Debye, 208, 276
— equation, 249
Decay curves, 92, 102, 113
Decreasing functions (*see* Functions)
Definite integrals (*see* Integrals), 129 ff
Degrees of freedom, 254
De Moivre's theorem, 214
Density, variable, 292
Dependent variables, 2
Derivatives, *38*
— applications, 51
— circular functions, 85
— functional, 42
— higher, 45
— implicit, 49
— interchange of variables, 49
— interpretation of, 51
— logarithmic, 48
— notation, 38, 48
— of integrals, 155, 159
— relations between partial derivatives, 262
— several variables, 258
Derived function (*see* Derivative), *38*
Determinants, *301*
— antisymmetric, 309
— cofactors, 312
— evaluation, 34
— general theory of, 311
— minors, 312
— symmetric, 309
Dieterici, 62
Differential equations (*see* Partial), 222
— change of dependent variable, 224, 238, 244
— first-order, 224 ff
— linear, 224, 228
— power series solution of, 244
— reducible, 243
— terminology, 223
Differentials, *44*
— exact, 270
— integrable relations, 268
— non-integrable relations, 268
— several variables, 256
— total, 257
Differentiation, *37*
— implicit, 49
— rules of, 41, 45
— several variables, *254*
Diffusion, 322
— equation, 350
— — steady, 233
Direction cosines, 318
Direction, distribution in, 296
Discrete variable, 1
Dispersion of light, 213, 216, 248
Displacement vector, 319
Divergence, 26
— of field, div **F**, 332
— of functions, 196
— of series, 179
— types, 196
Donkeys, 23
Double integrals, *282*

e, natural number, 27, 90
Electric charge, 279
— distribution, 292, 294
— potential due to, 196
Electricity, currents, 133
Electron scattering, 161
Ellipse, coordinate geometry, 121
Energy (*see* Potential, Work)
— levels (*see* Molecular vibrations), 161
— relativistic, 50
Equations, cubic, 89
— graphical solution, *14*
— of conic sections, 119–22
— solution of, 14, 51
Equilibrium, 58
— constants (*see* Reaction rates)
Errors, *4*, 8, 54
— combinations of, 5
— Gaussian integrals, 160
— in series, 182
— logarithmic, 54
— mean square error, 199
Euler–Maclaurin summation formula, 140
Even functions, 11
— and odd functions, 152
— series, 193
Evolute, 110
Exact differentials, 264, 270
Exponential decay, *92*, 133, 224
Exponential functions, 30, *90*
— (complex), 218
— graphs, 91
— integrals, 168
— series, 189
Exponential increase, 91
Exponentials, derivatives, 93
— sums of, 93

FACTORIAL representation of polynomials, 78
Factorials, 142, 157
Factorization, 64
Fields, scalar, 327
— vector, *328*
First moment, of area, 281
— of mass, 288
First order (*see* Orders of magnitude *and* Changes, small)
Fluid, flow, 278
— motion, acceleration, 259
— — rotation, 330
Flux vectors, 320–1, 331
Formulae, derivatives, 38, 39
— errors, 5, 39
— higher derivatives, 46
— rules of differentiation, 41
Fourier series, 185, 197, 221
Fourier's heat law, 232
— theorem, proof, 206
Functions, approximations to, 191
— complex, 214, *218*
— continuous, 31
— decreasing, 52
— definition, 1
— even, 11
— even and odd, series, 193
— increasing, 52
— limits of, 31
— notation, 2
— odd, 11
— of sequences, 32
— of several variables, *254*
— — contour and solid models, 255
— — integration, *265*
— — Maclaurin's and Taylor's series, 270
— — maxima and minima, 273
— — power series, 271
— — stationary values, 271
— periodic (*see* Periodic), 153

GAMMA function, $\Gamma(x)$, 157
Gas equations, 62
Gradient vector, grad T, 321
Graph plotting, 6
Graphs (*see also* Curves), 6, 9, *105*
— hyperbolic functions, 96
— inverse circular functions, 86
— inverse hyperbolic functions, 99–100
— logarithmic, 6
— of polynomials, *65*, 67
— of powers, 66
— of rational functions, 71
— two variables, 255
Graphical solutions, *14*
Greatest and least values (*see* Maxima), 57

HALF-lifetime, 102
Hankel functions, $H^{(1)}(x)$, $H^{(2)}(x)$, 247
Harmonic oscillator (*see* Oscillations), 235
Heat equation, 340
Heat flow, 232, 233, 322
Higher derivatives, 45
— of functions, 47
— simple forms, 46
Homogeneous simultaneous equations, 304
Hydrostatics, 279
Hyperbola, coordinate geometry, 122
Hyperbolic functions, *95*, 109
— derivatives, 96
— graphs, 96
— identities and properties, 98–9
— integrals, 169
— inverse functions, derivatives, table, 102
— — graphs and logarithmic forms, 99–101
— relation to circular functions, 97
— series, 189

IMPLICIT differentiation, 49
Increasing functions (*see* Functions), 52
Indefinite integrals, 130
Independent variables, 2
Indeterminate values, 27, 32
Index, 351
Induction, mathematical, 76
Infinite integrals, multiple, 287
Infinite series (*see* Series)
Infinity, 25
— in integration, 154
Inflection, 53
Integers, sums of powers, 76, 202, 208
Integrability, 267
Integrals (*see also* Double, Multiple, Surface, Triple, Volume),
— as functions, 158
— definite, 129
— derivatives of, 155, 159
— double, *282*
— indefinite, 130
— Mean Value Theorem, 154
— particular, 130. 235
— properties of definite integrals, 151
— standard forms, 130
— surface, *278*
— triple, *289*
Integral tables, 162 ff
— definite integrals, 175
— exponential, 162, 168
— hyperbolic, 163, 169
— linear surds, $\sqrt{(ax + b)}$, 171

Integral tables, logarithmic, 169
— quadratic surds, $\sqrt{(ax^2 + bx + c)}$, 172–4
— rational functions, 163–4
— reduction formula, 176
— standard forms, 162
— trigonometric, 163, 166
Integrating factors, 229, 270
Integration (*see also* Arc lengths, Areas, Mass, Volumes), 124
— numerical, 124, 127, 140
— techniques, algebraic changes, 144
— — change of variables, 145, 152
— — integration by parts, 147
— — integration by substitutions, 147
— — logarithmic, 144
— — periodic functions, 154
— — reduction formulae, 156
— — standard formulae, 146
Intermolecular forces (*see* Molecular vibrations), 59
Intrinsic equation of curve, 117
Inverse functions, addition, 88
— circular (trigonometric), 86
— hyperbolic, 99–102
— —, table of, 102
Involute, 110
Irrotational field, rot $\mathbf{F} = 0$, 329
Isothermal expansion, 142
Iterative method, *18*

KINETIC theory, 295
— of gases, integrals, 160

LAGRANGE multipliers, 271
Langevin function, 103, 208
Laplace's equation, $\nabla^2\phi = 0$, 350
— solutions, 276
Laplacian operator, ∇^2, 334
Latent heats, 248
Legendre polynomials, $P_n(x)$, 195, 205, 249, 276
Leibnitz rule, 47
Lengths of arc, 108
L'Hôpital's rule, 61
Limits, *26*
— evaluation of, 61
— of functions, 31
— of integration, 129
— of summations, 128
Linear differential equations, 224, 228 ff
Linear surds, integrals, 171
Lines, relations for straight line, 118
— relations for two lines, 118
Logarithmic differentiation, 48
Logarithmic function, *94*
— (complex), 219
— derivative, 95
Logarithmic function, graphs, 94
— integrals, 169

MACLAURIN's series, 187
— errors, 187
— methods, 193 ff
— several variables, 270
Magnetic properties, 142
Mass Action Law, 225
Mass calculations, 139
Mass centre, 288
Mass distributions, *288*, 292
Mathematical induction, 76
Maxima and minima, *55*, 192
— several variables, 273
Maximum, 9
Maxwell relations, 264
Maxwell solid, 248
Mean centre, areas, 280, 286
— of arc, 137
Mean values, 127, 140
— of functions, 155
Mean Value Theorem, derivatives, 60
— for integrals, 154
Mechanics, 325
— linear motion, 131
Melting points, 25
Minima (*see also* Maxima), *55*
Minimum, 9
Minors, 312
Mixed second moment, 283
Modulus, 9, 212
Molecular distribution, 296
Molecular properties, 142
Molecular velocities, 295
Molecular vibrations, 25, 34, 62, 103, 207, 236, 249, *308*
Moment of force, 325
Moment of inertia, 282, 291, 299
Monotonic sequences, 29
Morse, 34, 247
Multiple integrals, *278*
— repeated integrals, 274

NATURAL number (*see* '*e*'), 27, 90
Newton's law of cooling, 232
Newton's method, 19, 51
Newton's tangent method, 18
Normal, equations of, 51
Normal vibrations, 305
Normals, *107*
Normalization, 299
Numerical integration, 124, 127, 140

ODD functions, 11
— series, 193
Optical absorption, 213
Optical rotation, 161
Order of differentiation, 263

Orders of magnitude, *28*, 40, 45, 67
— exponentials, 93
Orthogonal functions, 203
Orthogonality, 203
Oscillations, 305
— about stable equilibrium, 237
— critically damped, 239
— damped, 238
— forced, 239
— simple harmonic, 234
Oscillatory sequences, 26

PARABOLA, coordinate geometry, 119, 120
Parameters, 105
Parametric equations, 107–8
Partial derivatives, 254
— change of variables, *258*
Partial differential equations, 337
— polar coordinates, 337, 345
Partial fractions, *72*
Partial sums, 182
Particular Integral, 235
Particular solutions, 223
Partition function (*see* Statistical mechanics), 207
Pedal equation of curve, 117
Periodic functions, 153, 197 ff
Phase, 234
— changes, 115
— in resonance, 240
— space, 298
Plasticity, 248
Points of inflection (*see* Inflection), 53
Poisseuille, 55
Polar coordinates, 9, *110*, 345
— curvature, 112
— double integrals, 286
— slopes of curves, 112
Polynomials, *64*
Position vector, 318
Potential energy, 236, 269
Potential functions, 114
Potentials, 58, 327
— electric, 279
Power series, 185 ff
— complex numbers, 217
— derivatives and integrals, 186
— methods, 193
— two variables, 271
Precipitates, 30
Pressure, 138
Principal Values of inverse functions, *87*

QUADRATIC equations, 63
Quadratic surds, integrals, 172–4
Quantum mechanics, 53, 142, 299, 308
Quantum theory, 161

RADIOACTIVE decay series, 230 (*see also* Sums of exponentials, 93)
Radius of convergence 185
Radius of curvature, 110
Rates of change, *34*
— relations between, 41
Rates of reaction, 225 ff
Ratio test, 180
Ratios, sequences of, 27
Rational functions, *71*
— integrals, 163–4
Reaction, half-lifetime, 139
Reaction rates, 52–56, 77
— differential equations, 225 ff
— solution of equations, 249–53
Reactions, chain theory, 241
— first-order, 225
— reversible, 227
— second-order, 226
— simultaneous, 231
— successive, 230
Recurrence relation, 34, 103, 183
Reduction formulae, 156
— table of, 176
Remainder Theorem, 63
Repeated integrals (*see* Multiple integrals)
Resonance, 240
— curve, 240
— energy, 308
Rolle's theorem, 59
Roots of equations (*see* Solution of equations), 14, 51
— complex, 220
Rotations, as vectors, 320
— rot **F**, 330

S-SHAPED curves, 114
Scalar fields, 327
Scalar (quantity), 318
Scalar product, 320, 323, 324
Second moments of area, 282
Semi-convergent series, 187
Separable functions, 287, 344
Separation of variables, 344
Sequences, *25*
— functions of, 32
— monotonic, 29
— of ratios, 27
Series, 178
— addition and multiplication, 180
— convergence tests, 180
— numerical, 202
— powers of integers, 76
Sigmoid curves, 113
Simple harmonic motion (*see* Oscillations), 235
— period, 236, 237
— phase, 236

Simpson's rule, 127
Simultaneous equations, 315
— homogeneous, 315
Simultaneous linear equations, *301*
— homogeneous, 304
Simultaneous reactions (*see* Reactions)
Slope, 9
Small changes (*see* Changes), 53
— simultaneous, 256
Solenoidal, 333
Solid angles, 296
Solid models, 255
Solution of equations, 51, 301
— trigonometric, 83
Space coordinates, *290*
Special curves, 78
— list, 112
— spirals and ovals, 111
Specific heats, 140, 208, 276
Sphere, 135, 138, 337
Spherical polar coordinates, 293
Stability, 58
Standard forms for integration, 162
Standard formulae, extensions to, 146
Stationary values, 52
Statistical mechanics, 71, 142–3, 161, 202, 207, 295, 298
Stirling's formula for $n!$, 142
Substitutions in integrals, 149, 168–74
Successive approximations, 25, 179
— (*see* Iterative methods), 18
Summations, 133, 140
— (*see* Integration), 125
Surds, in integrals, *171* ff
Surface integrals, *278*

TABLES, 1
— of integrals, 162 ff
Tangents, 9, *107*
— equations of, 51
Taylor's series, 187 ff
— several variables, 270
Taylor's theorem, 149
Tetrahedron, volume of, 285
Thermal capacity, 140, 265
Thermodynamic relations, *260–4*, 276
Thermodynamics, 103, 161
Thermodynamics, functions of state, 269
Total derivative, 259
Total differential, 257
Trapezium rule, 127
Triangles, solution of, 83
— trigonometric relations, 83
Triangular coordinates, 114
Trigonometric functions (*see* Circular functions), *79*
— integrals, 163, 166–8
Trigonometric identities, 80
Trigonometry, relations for triangle, 83
Triple integrals, *289*
Tube of force, 333

UNDETERMINED multipliers, 271

VAN DER WAALS' equation, 2, 62
Vapour pressure, 103
Variable, continuous, 1
— dependent, 2
— discrete, 1
— independent, 2
Variation method, 309
Vector algebra, 319, 322
Vector components, 318, 323 ff
Vector gradient, 321
Vector product, 325
Vectors, 215, *318*
— generalized, 334
— scalar product, 323
Velocity, distributions, 295
— mean square, 295
— space, 294
— vectors, 320
Vibrations, 305
Virial coefficients, 161
Viscosity, 52, 55
Volume integrals, 284
Volumes, 135
Vorticity, 330

WASHINGS, 30
Wave equation, 348
Waves, capillary, 62
Work, done by force, 266
— evaluation of, 138