A COURSE

OF

PURE MATHEMATICS

BY

G. H. HARDY, M.A., F.R.S.

FELLOW OF NEW COLLEGE

SAVILIAN PROFESSOR OF GEOMETRY IN THE UNIVERSITY OF OXFORD

LATE FELLOW OF TRINITY COLLEGE, CAMBRIDGE

THIRD EDITION

Cambridge

at the University Press

1921

First Edition 1908
Second Edition 1914
Third Edition 1921

PREFACE TO THE THIRD EDITION

No extensive changes have been made in this edition. The most important are in §§ 80–82, which I have rewritten in accordance with suggestions made by Mr S. Pollard.

The earlier editions contained no satisfactory account of the genesis of the circular functions. I have made some attempt to meet this objection in § 158 and Appendix III. Appendix IV is also an addition.

It is curious to note how the character of the criticisms I have had to meet has changed. I was too meticulous and pedantic for my pupils of fifteen years ago: I am altogether too popular for the Trinity scholar of to-day. I need hardly say that I find such criticisms very gratifying, as the best evidence that the book has to some extent fulfilled the purpose with which it was written.

G. H. H.

August 1921

EXTRACT FROM THE PREFACE TO THE SECOND EDITION

The principal changes made in this edition are as follows. I have inserted in Chapter I a sketch of Dedekind's theory of real numbers, and a proof of Weierstrass's theorem concerning points of condensation; in Chapter IV an account of 'limits of indetermination' and the 'general principle of convergence'; in Chapter V a proof of the 'Heine-Borel Theorem', Heine's theorem concerning uniform continuity, and the fundamental theorem concerning implicit functions; in Chapter VI some additional matter concerning the integration of algebraical functions; and in Chapter VII a section on differentials. I have also rewritten in a more general form the sections which deal with the definition of the definite integral. In order to find space for these insertions I have deleted a good deal of the analytical geometry and formal trigonometry contained in Chapters II and III of the first edition. These changes have naturally involved a large number of minor alterations.

G. H. H.

October 1914

EXTRACT FROM THE PREFACE TO THE FIRST EDITION

This book has been designed primarily for the use of first year students at the Universities whose abilities reach or approach something like what is usually described as 'scholarship standard'. I hope that it may be useful to other classes of readers, but it is this class whose wants I have considered first. It is in any case a book for mathematicians: I have nowhere made any attempt to meet the needs of students of engineering or indeed any class of students whose interests are not primarily mathematical.

I regard the book as being really elementary. There are plenty of hard examples (mainly at the ends of the chapters): to these I have added, wherever space permitted, an outline of the solution. But I have done my best to avoid the inclusion of anything that involves really difficult ideas. For instance, I make no use of the 'principle of convergence': uniform convergence, double series, infinite products, are never alluded to: and I prove no general theorems whatever concerning the inversion of limit-operations—I never even define $\dfrac{\partial^2 f}{\partial x\, \partial y}$ and $\dfrac{\partial^2 f}{\partial y\, \partial x}$. In the last two chapters I have occasion once or twice to integrate a power-series, but I have confined myself to the very simplest cases and given a special discussion in each instance. Anyone who has read this book will be in a position to read with profit Dr Bromwich's *Infinite Series*, where a full and adequate discussion of all these points will be found.

September 1908

CONTENTS

CHAPTER I

REAL VARIABLES

CHAPTER II

FUNCTIONS OF REAL VARIABLES

CHAPTER III

COMPLEX NUMBERS

CHAPTER IV

LIMITS OF FUNCTIONS OF A POSITIVE INTEGRAL VARIABLE

CHAPTER V

LIMITS OF FUNCTIONS OF A CONTINUOUS VARIABLE. CONTINUOUS AND DISCONTINUOUS FUNCTIONS

CHAPTER VI

DERIVATIVES AND INTEGRALS

CHAPTER VII

ADDITIONAL THEOREMS IN THE DIFFERENTIAL AND INTEGRAL CALCULUS

CHAPTER VIII

THE CONVERGENCE OF INFINITE SERIES AND INFINITE INTEGRALS

CHAPTER IX

THE LOGARITHMIC AND EXPONENTIAL FUNCTIONS OF A REAL VARIABLE

CHAPTER X

THE GENERAL THEORY OF THE LOGARITHMIC, EXPONENTIAL, AND CIRCULAR FUNCTIONS

CHAPTER I

REAL VARIABLES

1. Rational numbers. A fraction $r = p/q$, where p and q are positive or negative integers, is called a *rational number*. We can suppose (i) that p and q have no common factor, as if they have a common factor we can divide each of them by it, and (ii) that q is positive, since

$$p/(-q) = (-p)/q, \quad (-p)/(-q) = p/q.$$

To the rational numbers thus defined we may add the 'rational number 0' obtained by taking $p = 0$.

We assume that the reader is familiar with the ordinary arithmetical rules for the manipulation of rational numbers. The examples which follow demand no knowledge beyond this.

Examples I. 1. If r and s are rational numbers, then $r + s$, $r - s$, rs, and r/s are rational numbers, unless in the last case $s = 0$ (when r/s is of course meaningless).

2. If λ, m, and n are positive rational numbers, and $m > n$, then $\lambda(m^2 - n^2)$, $2\lambda mn$, and $\lambda(m^2 + n^2)$ are positive rational numbers. Hence show how to determine any number of right-angled triangles the lengths of all of whose sides are rational.

3. Any terminated decimal represents a rational number whose denominator contains no factors other than 2 or 5. Conversely, any such rational number can be expressed, and in one way only, as a terminated decimal.

[The general theory of decimals will be considered in Ch. IV.]

4. The positive rational numbers may be arranged in the form of a simple series as follows:

$$\tfrac{1}{1},\quad \tfrac{2}{1},\quad \tfrac{1}{2},\quad \tfrac{3}{1},\quad \tfrac{2}{2},\quad \tfrac{1}{3},\quad \tfrac{4}{1},\quad \tfrac{3}{2},\quad \tfrac{2}{3},\quad \tfrac{1}{4},\ \dots.$$

Show that p/q is the $[\frac{1}{2}(p + q - 1)(p + q - 2) + q]$th term of the series.

[In this series every rational number is repeated indefinitely. Thus 1 occurs as $\frac{1}{1}$, $\frac{2}{2}$, $\frac{3}{3}$, ... We can of course avoid this by omitting every number which has already occurred in a simpler form, but then the problem of determining the precise position of p/q becomes more complicated.]

2. The representation of rational numbers by points on a line. It is convenient, in many branches of mathematical analysis, to make a good deal of use of geometrical illustrations.

The use of geometrical illustrations in this way does not, of course, imply that analysis has any sort of dependence upon geometry: they are illustrations and nothing more, and are employed merely for the sake of clearness of exposition. This being so, it is not necessary that we should attempt any logical analysis of the ordinary notions of elementary geometry; we may be content to suppose, however far it may be from the truth, that we know what they mean.

Assuming, then, that we know what is meant by a *straight line*, a *segment* of a line, and the *length* of a segment, let us take a straight line Λ, produced indefinitely in both directions, and a segment A_0A_1 of any length. We call A_0 the *origin*, or *the point* 0, and A_1 *the point* 1, and we regard these points as representing the numbers 0 and 1.

In order to obtain a point which shall represent a positive rational number $r = p/q$, we choose the point A_r such that

$$A_0A_r/A_0A_1 = r,$$

A_0A_r being a stretch of the line extending in the same direction along the line as A_0A_1, a direction which we shall suppose to be from left to right when, as in Fig. 1, the line is drawn horizontally across the paper. In order to obtain a point to represent a negative rational number $r = -s$,

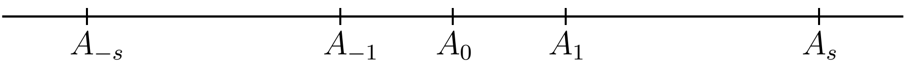

Fig. 1.

it is natural to regard length as a magnitude capable of sign, positive if the length is measured in one direction (that of A_0A_1), and negative if measured in the other, so that $AB = -BA$; and to take as the point representing r the point A_{-s} such that

$$A_0A_{-s} = -A_{-s}A_0 = -A_0A_s.$$

We thus obtain a point A_r on the line corresponding to every rational value of r, positive or negative, and such that

$$A_0 A_r = r \cdot A_0 A_1;$$

and if, as is natural, we take $A_0 A_1$ as our unit of length, and write $A_0 A_1 = 1$, then we have

$$A_0 A_r = r.$$

We shall call the points A_r the *rational points* of the line.

3. Irrational numbers. If the reader will mark off on the line all the points corresponding to the rational numbers whose denominators are 1, 2, 3, ... in succession, he will readily convince himself that he can cover the line with rational points as closely as he likes. We can state this more precisely as follows: *if we take any segment BC on Λ, we can find as many rational points as we please on BC.*

Suppose, for example, that BC falls within the segment $A_1 A_2$. It is evident that if we choose a positive integer k so that

$$k \cdot BC > 1,^* \tag{1}$$

and divide $A_1 A_2$ into k equal parts, then at least one of the points of division (say P) must fall inside BC, without coinciding with either B or C. For if this were not so, BC would be entirely included in one of the k parts into which $A_1 A_2$ has been divided, which contradicts the supposition (1). But P obviously corresponds to a rational number whose denominator is k. Thus at least one rational point P lies between B and C. But then we can find another such point Q between B and P, another between B and Q, and so on indefinitely; *i.e.*, as we asserted above, we can find as many as we please. We may express this by saying that BC includes *infinitely many* rational points.

---

*The assumption that this is possible is equivalent to the assumption of what is known as the Axiom of Archimedes.

The meaning of such phrases as '*infinitely many*' or '*an infinity of*', in such sentences as 'BC includes infinitely many rational points' or 'there are an infinity of rational points on BC' or 'there are an infinity of positive integers', will be considered more closely in Ch. IV. The assertion 'there are an infinity of positive integers' means 'given any positive integer n, however large, we can find more than n positive integers'. This is plainly true whatever n may be, *e.g.* for $n = 100{,}000$ or $100{,}000{,}000$. The assertion means exactly the same as 'we can find *as many positive integers as we please*'.

The reader will easily convince himself of the truth of the following assertion, which is substantially equivalent to what was proved in the second paragraph of this section: given any rational number r, and any positive integer n, we can find another rational number lying on either side of r and differing from r by less than $1/n$. It is merely to express this differently to say that we can find a rational number lying on either side of r and differing from r *by as little as we please*. Again, given any two rational numbers r and s, we can interpolate between them a chain of rational numbers in which any two consecutive terms differ by as little as we please, that is to say by less than $1/n$, where n is any positive integer assigned beforehand.

From these considerations the reader might be tempted to infer that an adequate view of the nature of the line could be obtained by imagining it to be formed simply by the rational points which lie on it. And it is certainly the case that if we imagine the line to be made up solely of the rational points, and all other points (if there are any such) to be eliminated, the figure which remained would possess most of the properties which common sense attributes to the straight line, and would, to put the matter roughly, look and behave very much like a line.

A little further consideration, however, shows that this view would involve us in serious difficulties.

Let us look at the matter for a moment with the eye of common sense, and consider some of the properties which we may reasonably expect a straight line to possess if it is to satisfy the idea which we have formed of it in elementary geometry.

The straight line must be composed of points, and any segment of it by all the points which lie between its end points. With any such segment must be associated a certain entity called its *length*, which must be a *quantity* capable of *numerical measurement* in terms of any standard or unit length, and these lengths must be capable of combination with one another, according to the ordinary rules of algebra, by means of addition or multiplication. Again, it must be possible to construct a line whose length is the sum or product of any two given lengths. If the length PQ, along a given line, is a, and the length QR, along the same straight line, is b, the length PR must be $a+b$. Moreover, if the lengths OP, OQ, along one straight line, are 1 and a, and the length OR along another straight line is b, and if we determine the length OS by Euclid's construction (Euc. VI. 12) for a fourth proportional to the lines OP, OQ, OR, this length must be ab, the algebraical fourth proportional to 1, a, b. And it is hardly necessary to remark that the sums and products thus defined must obey the ordinary 'laws of algebra'; viz.

$$a+b=b+a, \quad a+(b+c)=(a+b)+c,$$
$$ab=ba, \quad a(bc)=(ab)c, \quad a(b+c)=ab+ac.$$

The lengths of our lines must also obey a number of obvious laws concerning inequalities as well as equalities: thus if A, B, C are three points lying along Λ from left to right, we must have $AB<AC$, and so on. Moreover it must be possible, on our fundamental line Λ, to find a point P such that A_0P is equal to any segment whatever taken along Λ or along any other straight line. All these properties of a line, and more, are involved in the presuppositions of our elementary geometry.

Now it is very easy to see that the idea of a straight line as composed of a series of points, each corresponding to a rational number, cannot possibly satisfy all these requirements. There are various elementary geometrical constructions, for example, which purport to construct a length x such that $x^2=2$. For instance, we may construct an isosceles right-angled triangle ABC such that $AB=AC=1$. Then if $BC=x$, $x^2=2$. Or we may determine the length x by means of Euclid's construction (Euc. VI. 13) for a mean proportional to 1 and 2, as indicated in the figure. Our requirements therefore involve the existence of a length measured by a number x,

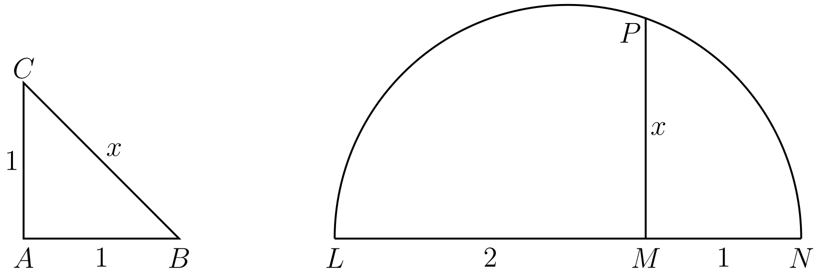

Fig. 2.

and a point P on Λ such that

$$A_0P = x, \quad x^2 = 2.$$

But it is easy to see that *there is no rational number such that its square is* 2. In fact we may go further and say that there is no rational number whose square is m/n, where m/n is any positive fraction in its lowest terms, unless m and n are both perfect squares.

For suppose, if possible, that

$$p^2/q^2 = m/n,$$

p having no factor in common with q, and m no factor in common with n. Then $np^2 = mq^2$. Every factor of q^2 must divide np^2, and as p and q have no common factor, every factor of q^2 must divide n. Hence $n = \lambda q^2$, where λ is an integer. But this involves $m = \lambda p^2$: and as m and n have no common factor, λ must be unity. Thus $m = p^2$, $n = q^2$, as was to be proved. In particular it follows, by taking $n = 1$, that an integer cannot be the square of a rational number, unless that rational number is itself integral.

It appears then that our requirements involve the existence of a number x and a point P, not one of the rational points already constructed, such that $A_0P = x$, $x^2 = 2$; and (as the reader will remember from elementary algebra) we write $x = \sqrt{2}$.

The following alternative proof that no rational number can have its square equal to 2 is interesting.

Suppose, if possible, that p/q is a positive fraction, in its lowest terms, such that $(p/q)^2 = 2$ or $p^2 = 2q^2$. It is easy to see that this involves $(2q - p)^2 = 2(p - q)^2$; and so $(2q - p)/(p - q)$ is another fraction having the same property. But clearly $q < p < 2q$, and so $p - q < q$. Hence there is another fraction equal to p/q and having a smaller denominator, which contradicts the assumption that p/q is in its lowest terms.

Examples II. 1. Show that no rational number can have its cube equal to 2.

2. Prove generally that a rational fraction p/q in its lowest terms cannot be the cube of a rational number unless p and q are both perfect cubes.

3. A more general proposition, which is due to Gauss and includes those which precede as particular cases, is the following: *an algebraical equation*

$$x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_n = 0,$$

with integral coefficients, cannot have a rational but non-integral root.

[For suppose that the equation has a root a/b, where a and b are integers without a common factor, and b is positive. Writing a/b for x, and multiplying by b^{n-1}, we obtain

$$-\frac{a^n}{b} = p_1 a^{n-1} + p_2 a^{n-2} b + \cdots + p_n b^{n-1},$$

a fraction in its lowest terms equal to an integer, which is absurd. Thus $b = 1$, and the root is a. It is evident that a must be a divisor of p_n.]

4. Show that if $p_n = 1$ and neither of

$$1 + p_1 + p_2 + p_3 + \dots, \quad 1 - p_1 + p_2 - p_3 + \dots$$

is zero, then the equation cannot have a rational root.

5. Find the rational roots (if any) of

$$x^4 - 4x^3 - 8x^2 + 13x + 10 = 0.$$

[The roots can only be integral, and so ± 1, ± 2, ± 5, ± 10 are the only possibilities: whether these are roots can be determined by trial. It is clear that we can in this way determine the rational roots of any such equation.]

4. Irrational numbers (*continued*). The result of our geometrical representation of the rational numbers is therefore to suggest the desirability of enlarging our conception of 'number' by the introduction of further numbers of a new kind.

The same conclusion might have been reached without the use of geometrical language. One of the central problems of algebra is that of the solution of equations, such as

$$x^2 = 1, \quad x^2 = 2.$$

The first equation has the two rational roots 1 and -1. But, if our conception of number is to be limited to the rational numbers, we can only say that the second equation has no roots; and the same is the case with such equations as $x^3 = 2$, $x^4 = 7$. These facts are plainly sufficient to make some generalisation of our idea of number desirable, if it should prove to be possible.

Let us consider more closely the equation $x^2 = 2$.

We have already seen that there is no rational number x which satisfies this equation. The square of any rational number is either less than or greater than 2. We can therefore divide the rational numbers into two classes, one containing the numbers whose squares are less than 2, and the other those whose squares are greater than 2. We shall confine our attention to the *positive* rational numbers, and we shall call these two classes *the class* L, or *the lower class*, or *the left-hand class*, and *the class* R, or *the upper class*, or *the right-hand class*. It is obvious that every member of R is greater than all the members of L. Moreover it is easy to convince ourselves that we can find a member of the class L whose square, though less than 2, differs from 2 by as little as we please, and a member of R whose square, though greater than 2, also differs from 2 by as little as we please. In fact, if we carry out the ordinary arithmetical process for the extraction of the square root of 2, we obtain a series of rational numbers, viz.

$$1, \quad 1.4, \quad 1.41, \quad 1.414, \quad 1.4142, \ \dots$$

whose squares

$$1, \quad 1.96, \quad 1.9881, \quad 1.999\,396, \quad 1.999\,961\,64, \ \dots$$

are all less than 2, but approach nearer and nearer to it; and by taking a sufficient number of the figures given by the process we can obtain as close an approximation as we want. And if we increase the last figure, in each of the approximations given above, by unity, we obtain a series of rational numbers

$$2, \quad 1.5, \quad 1.42, \quad 1.415, \quad 1.4143, \; \dots$$

whose squares

$$4, \quad 2.25, \quad 2.0164, \quad 2.002\,225, \quad 2.000\,244\,49, \; \dots$$

are all greater than 2 but approximate to 2 as closely as we please.

The reasoning which precedes, although it will probably convince the reader, is hardly of the precise character required by modern mathematics. We can supply a formal proof as follows. In the first place, we can find a member of L and a member of R, differing by as little as we please. For we saw in § 3 that, given any two rational numbers a and b, we can construct a chain of rational numbers, of which a and b are the first and last, and in which any two consecutive numbers differ by as little as we please. Let us then take a member x of L and a member y of R, and interpolate between them a chain of rational numbers of which x is the first and y the last, and in which any two consecutive numbers differ by less than δ, δ being any positive rational number as small as we please, such as .01 or .0001 or .000 001. In this chain there must be a last which belongs to L and a first which belongs to R, and these two numbers differ by less than δ.

We can now prove that *an x can be found in L and a y in R such that $2-x^2$ and y^2-2 are as small as we please*, say less than δ. Substituting $\frac{1}{4}\delta$ for δ in the argument which precedes, we see that we can choose x and y so that $y-x < \frac{1}{4}\delta$; and we may plainly suppose that both x and y are less than 2. Thus

$$y+x < 4, \quad y^2-x^2 = (y-x)(y+x) < 4(y-x) < \delta;$$

and since $x^2 < 2$ and $y^2 > 2$ it follows *a fortiori* that $2-x^2$ and y^2-2 are each less than δ.

It follows also that *there can be no largest member of L or smallest member of R*. For if x is any member of L, then $x^2 < 2$. Suppose that $x^2 = 2-\delta$. Then we can find a member x_1 of L such that x_1^2 differs from 2 by less than δ, and so $x_1^2 > x^2$ or $x_1 > x$. Thus there are larger members of L than x; and as x is *any* member of L, it follows that no member of L can be larger than all the rest. Hence L has no largest member, and similarly R has no smallest.

5. Irrational numbers (*continued*). We have thus divided the positive rational numbers into two classes, L and R, such that (i) every member of R is greater than every member of L, (ii) we can find a member of L and a member of R whose difference is as small as we please, (iii) L has no greatest and R no least member. Our common-sense notion of the attributes of a straight line, the requirements of our elementary geometry and our elementary algebra, alike demand *the existence of a number x greater than all the members of L and less than all the members of R, and of a corresponding point P on Λ such that P divides the points which correspond to members of L from those which correspond to members of R.*

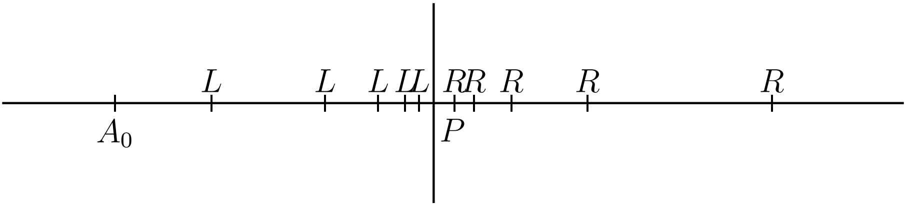

Fig. 3.

Let us suppose for a moment that there is such a number x, and that it may be operated upon in accordance with the laws of algebra, so that, for example, x^2 has a definite meaning. Then x^2 cannot be either less than or greater than 2. For suppose, for example, that x^2 is less than 2. Then it follows from what precedes that we can find a positive rational number ξ such that ξ^2 lies between x^2 and 2. That is to say, we can find a member of L greater than x; and this contradicts the supposition that x divides the members of L from those of R. Thus x^2 cannot be less than 2, and similarly it cannot be greater than 2. We are therefore driven to the conclusion that $x^2 = 2$, and that x is the number which in algebra we denote by $\sqrt{2}$. And of course this number $\sqrt{2}$ is not rational, for no rational number has its square equal to 2. It is the simplest example of what is called an **irrational** number.

But the preceding argument may be applied to equations other than $x^2 = 2$, almost word for word; for example to $x^2 = N$, where N is any integer which is not a perfect square, or to

$$x^3 = 3, \quad x^3 = 7, \quad x^4 = 23,$$

or, as we shall see later on, to $x^3 = 3x + 8$. We are thus led to believe in the existence of irrational numbers x and points P on Λ such that x satisfies equations such as these, even when these lengths cannot (as $\sqrt{2}$ can) be constructed by means of elementary geometrical methods.

The reader will no doubt remember that in treatises on elementary algebra the root of such an equation as $x^q = n$ is denoted by $\sqrt[q]{n}$ or $n^{1/q}$, and that a meaning is attached to such symbols as

$$n^{p/q}, \quad n^{-p/q}$$

by means of the equations

$$n^{p/q} = (n^{1/q})^p, \quad n^{p/q} n^{-p/q} = 1.$$

And he will remember how, in virtue of these definitions, the 'laws of indices' such as

$$n^r \times n^s = n^{r+s}, \quad (n^r)^s = n^{rs}$$

are extended so as to cover the case in which r and s are any rational numbers whatever.

The reader may now follow one or other of two alternative courses. He may, if he pleases, be content to assume that 'irrational numbers' such as $\sqrt{2}$, $\sqrt[3]{3}$, … exist and are amenable to the algebraical laws with which he is familiar.* If he does this he will be able to avoid the more abstract discussions of the next few sections, and may pass on at once to §§ 13 *et seq.*

If, on the other hand, he is not disposed to adopt so *naive* an attitude, he will be well advised to pay careful attention to the sections which follow, in which these questions receive fuller consideration.†

*This is the point of view which was adopted in the first edition of this book.

†In these sections I have borrowed freely from Appendix I of Bromwich's *Infinite Series*.

Examples III. 1. Find the difference between 2 and the squares of the decimals given in § 4 as approximations to $\sqrt{2}$.

2. Find the differences between 2 and the squares of

$$\frac{1}{1},\quad \frac{3}{2},\quad \frac{7}{5},\quad \frac{17}{12},\quad \frac{41}{29},\quad \frac{99}{70}.$$

3. Show that if m/n is a good approximation to $\sqrt{2}$, then $(m+2n)/(m+n)$ is a better one, and that the errors in the two cases are in opposite directions. Apply this result to continue the series of approximations in the last example.

4. If x and y are approximations to $\sqrt{2}$, by defect and by excess respectively, and $2-x^2<\delta$, $y^2-2<\delta$, then $y-x<\delta$.

5. The equation $x^2=4$ is satisfied by $x=2$. Examine how far the argument of the preceding sections applies to this equation (writing 4 for 2 throughout). [If we define the classes L, R as before, they do not include *all* rational numbers. The rational number 2 is an exception, since 2^2 is neither less than or greater than 4.]

6. Irrational numbers (*continued*). In § 4 we discussed a special mode of division of the positive rational numbers x into two classes, such that $x^2<2$ for the members of one class and $x^2>2$ for those of the others. Such a mode of division is called a **section** of the numbers in question. It is plain that we could equally well construct a section in which the numbers of the two classes were characterised by the inequalities $x^3<2$ and $x^3>2$, or $x^4<7$ and $x^4>7$. Let us now attempt to state the principles of the construction of such 'sections' of the positive rational numbers in quite general terms.

Suppose that P and Q stand for two properties which are mutually exclusive and one of which must be possessed by every positive rational number. Further, suppose that every such number which possesses P is less than any such number which possesses Q. Thus P might be the property '$x^2<2$' and Q the property '$x^2>2$.' Then we call the numbers which possess P the lower or left-hand class L and those which possess Q the upper or right-hand class R. In general both classes will exist; but it may happen in special cases that one is non-existent and that every number belongs to the other. This would obviously happen, for example, if P

(or Q) were the property of being rational, or of being positive. For the present, however, we shall confine ourselves to cases in which both classes do exist; and then it follows, as in § 4, that we can find a member of L and a member of R whose difference is as small as we please.

In the particular case which we considered in § 4, L had no greatest member and R no least. This question of the existence of greatest or least members of the classes is of the utmost importance. We observe first that it is impossible in any case that L should have a greatest member *and* R a least. For if l were the greatest member of L, and r the least of R, so that $l < r$, then $\frac{1}{2}(l + r)$ would be a positive rational number lying between l and r, and so could belong neither to L nor to R; and this contradicts our assumption that every such number belongs to one class or to the other. This being so, there are but three possibilities, which are mutually exclusive. Either (i) L has a greatest member l, or (ii) R has a least member r, or (iii) L has no greatest member and R no least.

The section of § 4 gives an example of the last possibility. An example of the first is obtained by taking P to be '$x^2 \leqq 1$' and Q to be '$x^2 > 1$'; here $l = 1$. If P is '$x^2 < 1$' and Q is '$x^2 \geqq 1$', we have an example of the second possibility, with $r = 1$. It should be observed that we do not obtain a section at all by taking P to be '$x^2 < 1$' and Q to be '$x^2 > 1$'; for the special number 1 escapes classification (cf. Ex. III. 5).

7. Irrational numbers (*continued*). In the first two cases we say that the section *corresponds* to a positive rational number a, which is l in the one case and r in the other. Conversely, it is clear that to any such number a corresponds a section which we shall denote by α.* For we might take P and Q to be the properties expressed by

$$x \leqq a, \quad x > a$$

respectively, or by $x < a$ and $x \geqq a$. In the first case a would be the greatest member of L, and in the second case the least member of R.

*It will be convenient to denote a section, corresponding to a rational number denoted by an English letter, by the corresponding Greek letter.

There are in fact just two sections corresponding to any positive rational number. In order to avoid ambiguity we select one of them; let us select that in which the number itself belongs to the *upper* class. In other words, let us agree that we will consider only sections in which the lower class L has no greatest number.

There being this correspondence between the positive rational numbers and the sections defined by means of them, it would be perfectly legitimate, for mathematical purposes, to replace the numbers by the sections, and to regard the symbols which occur in our formulae as standing for the sections instead of for the numbers. Thus, for example, $\alpha > \alpha'$ would mean the same as $a > a'$, if α and α' are the sections which correspond to a and a'.

But when we have in this way substituted sections of rational numbers for the rational numbers themselves, we are almost forced to a generalisation of our number system. For there are sections (such as that of § 4) which do *not* correspond to any rational number. The aggregate of sections is a larger aggregate than that of the positive rational numbers; it includes sections corresponding to all these numbers, and more besides. It is this fact which we make the basis of our generalisation of the idea of number. We accordingly frame the following definitions, which will however be modified in the next section, and must therefore be regarded as temporary and provisional.

A section of the positive rational numbers, in which both classes exist and the lower class has no greatest member, is called a **positive real number**.

A positive real number which does not correspond to a positive rational number is called a positive **irrational** *number.*

8. Real numbers. We have confined ourselves so far to certain sections of the positive rational numbers, which we have agreed provisionally to call 'positive real numbers.' Before we frame our final definitions, we must alter our point of view a little. We shall consider sections, or divisions into two classes, not merely of the positive rational numbers, but of all rational numbers, including zero. We may then repeat all that we have said about sections of the positive rational numbers in §§ 6, 7, merely omitting the word positive occasionally.

Definitions. *A section of the rational numbers, in which both classes exist and the lower class has no greatest member, is called a* **real number**, *or simply a* **number**.

A real number which does not correspond to a rational number is called an **irrational** *number.*

If the real number does correspond to a rational number, we shall use the term 'rational' as applying to the real number also.

The term 'rational number' will, as a result of our definitions, be ambiguous; it may mean the rational number of § 1, or the corresponding real number. If we say that $\frac{1}{2} > \frac{1}{3}$, we may be asserting either of two different propositions, one a proposition of elementary arithmetic, the other a proposition concerning sections of the rational numbers. Ambiguities of this kind are common in mathematics, and are perfectly harmless, since the relations between different propositions are exactly the same whichever interpretation is attached to the propositions themselves. From $\frac{1}{2} > \frac{1}{3}$ and $\frac{1}{3} > \frac{1}{4}$ we can infer $\frac{1}{2} > \frac{1}{4}$; the inference is in no way affected by any doubt as to whether $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$ are arithmetical fractions or real numbers. Sometimes, of course, the context in which (*e.g.*) '$\frac{1}{2}$' occurs is sufficient to fix its interpretation. When we say (see § 9) that $\frac{1}{2} < \sqrt{\frac{1}{3}}$, we *must* mean by '$\frac{1}{2}$' the real number $\frac{1}{2}$.

The reader should observe, moreover, that no particular logical importance is to be attached to the precise form of definition of a 'real number' that we have adopted. We defined a 'real number' as being a section, *i.e.* a pair of classes. We might equally well have defined it as being the lower, or the upper, class; indeed it would be easy to define an infinity of classes of entities each of which would possess the properties of the class of real numbers. What is essential in mathematics is that its symbols should be capable of *some* interpretation; generally they are capable of *many*, and then, so far as mathematics is concerned, it does not matter which we adopt. Mr Bertrand Russell has said that 'mathematics is the science in which we do not know what we are talking about, and do not care whether what we say about it is true', a remark which is expressed in the form of a paradox but which in reality embodies a number of important truths. It would take too long to analyse the meaning of Mr Russell's epigram in detail, but one at any rate of its implications is this, that the symbols of mathematics are capable of varying interpretations, and that we are in general at liberty to adopt whichever we prefer.

There are now three cases to distinguish. It may happen that all negative rational numbers belong to the lower class and zero and all positive rational numbers to the upper. We describe this section as the **real number zero**. Or again it may happen that the lower class includes some positive numbers. Such a section we describe as a **positive real number**. Finally it may happen that some negative numbers belong to the upper class. Such a section we describe as a **negative real number**.*

The difference between our present definition of a positive real number a and that of § 7 amounts to the addition to the lower class of zero and all the negative rational numbers. An example of a negative real number is given by taking the property P of § 6 to be $x + 1 < 0$ and Q to be $x + 1 \geqq 0$. This section plainly corresponds to the negative rational number -1. If we took P to be $x^3 < -2$ and Q to be $x^3 > -2$, we should obtain a negative real number which is not rational.

9. Relations of magnitude between real numbers. It is plain that, now that we have extended our conception of number, we are bound to make corresponding extensions of our conceptions of equality, inequality, addition, multiplication, and so on. We have to show that these ideas can be applied to the new numbers, and that, when this extension of them is made, all the ordinary laws of algebra retain their validity, so that we can operate with real numbers in general in exactly the same way as with the rational numbers of § 1. To do all this systematically would occupy a

*There are also sections in which every number belongs to the lower or to the upper class. The reader may be tempted to ask why we do not regard these sections also as defining numbers, which we might call the *real numbers positive and negative infinity.*

There is no logical objection to such a procedure, but it proves to be inconvenient in practice. The most natural definitions of addition and multiplication do not work in a satisfactory way. Moreover, for a beginner, the chief difficulty in the elements of analysis is that of learning to attach precise senses to phrases containing the word ‘infinity’; and experience seems to show that he is likely to be confused by any addition to their number.

considerable space, and we shall be content to indicate summarily how a more systematic discussion would proceed.

We denote a real number by a Greek letter such as α, β, γ, ...; the rational numbers of its lower and upper classes by the corresponding English letters a, A; b, B; c, C; .... The classes themselves we denote by (a), (A), ...

If α and β are two real numbers, there are three possibilities:

(i) every a is a b and every A a B; in this case (a) is identical with (b) and (A) with (B);

(ii) every a is a b, but not all A's are B's; in this case (a) is a proper part of (b),* and (B) a proper part of (A);

(iii) every A is a B, but not all a's are b's.

These three cases may be indicated graphically as in Fig. 4.

In case (i) we write $\alpha = \beta$, in case (ii) $\alpha < \beta$, and in case (iii) $\alpha > \beta$. It is clear that, when α and β are both rational, these definitions agree

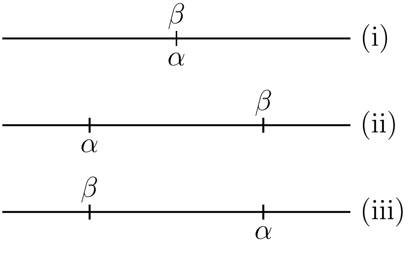

Fig. 4.

with the ideas of equality and inequality between rational numbers which we began by taking for granted; and that any positive number is greater than any negative number.

It will be convenient to define at this stage the negative $-\alpha$ of a positive number α. If (a), (A) are the classes which constitute α, we can define another section of the rational numbers by putting all numbers $-A$ in the lower class and all numbers $-a$ in the upper. The real number thus defined, which is clearly negative, we denote by $-\alpha$. Similarly we can define $-\alpha$

*_I.e._ is included in but not identical with (b).

when α is negative or zero; if α is negative, $-\alpha$ is positive. It is plain also that $-(-\alpha) = \alpha$. Of the two numbers α and $-\alpha$ one is always positive (unless $\alpha = 0$). The one which is positive we denote by $|\alpha|$ and call the *modulus* of α.

Examples IV. 1. Prove that $0 = -0$.

2. Prove that $\beta = \alpha$, $\beta < \alpha$, or $\beta > \alpha$ according as $\alpha = \beta$, $\alpha > \beta$, or $\alpha < \beta$.

3. If $\alpha = \beta$ and $\beta = \gamma$, then $\alpha = \gamma$.

4. If $\alpha \leqq \beta$, $\beta < \gamma$, or $\alpha < \beta$, $\beta \leqq \gamma$, then $\alpha < \gamma$.

5. Prove that $-\beta = -\alpha$, $-\beta < -\alpha$, or $-\beta > -\alpha$, according as $\alpha = \beta$, $\alpha < \beta$, or $\alpha > \beta$.

6. Prove that $\alpha > 0$ if α is positive, and $\alpha < 0$ if α is negative.

7. Prove that $\alpha \leqq |\alpha|$.

8. Prove that $1 < \sqrt{2} < \sqrt{3} < 2$.

9. Prove that, if α and β are two different real numbers, we can always find an infinity of rational numbers lying between α and β.

[All these results are immediate consequences of our definitions.]

10. Algebraical operations with real numbers. We now proceed to define the meaning of the elementary algebraical operations such as addition, as applied to real numbers in general.

(i) *Addition.* In order to define the sum of two numbers α and β, we consider the following two classes: (i) the class (c) formed by all sums $c = a + b$, (ii) the class (C) formed by all sums $C = A + B$. Plainly $c < C$ in all cases.

Again, there cannot be more than one rational number which does not belong either to (c) or to (C). For suppose there were two, say r and s, and let s be the greater. Then both r and s must be greater than every c and less than every C; and so $C - c$ cannot be less than $s - r$. But

$$C - c = (A - a) + (B - b);$$

and we can choose a, b, A, B so that both $A - a$ and $B - b$ are as small as we like; and this plainly contradicts our hypothesis.

If every rational number belongs to (c) or to (C), the classes (c), (C) form a section of the rational numbers, that is to say, a number γ. If there is one which does not, we add it to (C). We have now a section or real number γ, which must clearly be rational, since it corresponds to the least member of (C). *In any case we call γ the sum of α and β, and write*

$$\gamma = \alpha + \beta.$$

If both α and β are rational, they are the least members of the upper classes (A) and (B). In this case it is clear that $\alpha + \beta$ is the least member of (C), so that our definition agrees with our previous ideas of addition.

(ii) *Subtraction.* We define $\alpha - \beta$ by the equation

$$\alpha - \beta = \alpha + (-\beta).$$

The idea of subtraction accordingly presents no fresh difficulties.

Examples V. 1. Prove that $\alpha + (-\alpha) = 0$.

2. Prove that $\alpha + 0 = 0 + \alpha = \alpha$.

3. Prove that $\alpha + \beta = \beta + \alpha$. [This follows at once from the fact that the classes $(a + b)$ and $(b + a)$, or $(A + B)$ and $(B + A)$, are the same, since, *e.g.*, $a + b = b + a$ when a and b are rational.]

4. Prove that $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$.

5. Prove that $\alpha - \alpha = 0$.

6. Prove that $\alpha - \beta = -(\beta - \alpha)$.

7. From the definition of subtraction, and Exs. 4, 1, and 2 above, it follows that

$$(\alpha - \beta) + \beta = \{\alpha + (-\beta)\} + \beta = \alpha + \{(-\beta) + \beta\} = \alpha + 0 = \alpha.$$

We might therefore define the difference $\alpha - \beta = \gamma$ by the equation $\gamma + \beta = \alpha$.

8. Prove that $\alpha - (\beta - \gamma) = \alpha - \beta + \gamma$.

9. Give a definition of subtraction which does not depend upon a previous definition of addition. [To define $\gamma = \alpha - \beta$, form the classes (c), (C) for which

$c = a - B$, $C = A - b$. It is easy to show that this definition is equivalent to that which we adopted in the text.]

10. Prove that

$$\big||\alpha| - |\beta|\big| \leqq |\alpha \pm \beta| \leqq |\alpha| + |\beta|.$$

11. Algebraical operations with real numbers (*continued*). (iii) *Multiplication.* When we come to multiplication, it is most convenient to confine ourselves to *positive* numbers (among which we may include 0) in the first instance, and to go back for a moment to the sections of positive rational numbers only which we considered in §§ 4–7. We may then follow practically the same road as in the case of addition, taking (c) to be (ab) and (C) to be (AB). The argument is the same, except when we are proving that all rational numbers with at most one exception must belong to (c) or (C). This depends, as in the case of addition, on showing that we can choose a, A, b, and B so that $C - c$ is as small as we please. Here we use the identity

$$C - c = AB - ab = (A - a)B + a(B - b).$$

Finally we include negative numbers within the scope of our definition by agreeing that, if α and β are positive, then

$$(-\alpha)\beta = -\alpha\beta, \quad \alpha(-\beta) = -\alpha\beta, \quad (-\alpha)(-\beta) = \alpha\beta.$$

(iv) *Division.* In order to define division, we begin by defining the reciprocal $1/\alpha$ of a number α (other than zero). Confining ourselves in the first instance to positive numbers and sections of positive rational numbers, we define the reciprocal of a positive number α by means of the lower class $(1/A)$ and the upper class $(1/a)$. We then define the reciprocal of a negative number $-\alpha$ by the equation $1/(-\alpha) = -(1/\alpha)$. Finally we define α/β by the equation

$$\alpha/\beta = \alpha \times (1/\beta).$$

We are then in a position to apply to all real numbers, rational or irrational, the whole of the ideas and methods of elementary algebra. Naturally we do not propose to carry out this task in detail. It will be more profitable and more interesting to turn our attention to some special, but particularly important, classes of irrational numbers.

Examples VI. Prove the theorems expressed by the following formulae:

1. $\alpha \times 0 = 0 \times \alpha = 0$.
2. $\alpha \times 1 = 1 \times \alpha = \alpha$.
3. $\alpha \times (1/\alpha) = 1$.
4. $\alpha\beta = \beta\alpha$.
5. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.
6. $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
7. $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$.
8. $|\alpha\beta| = |\alpha|\,|\beta|$.

12. The number $\sqrt{2}$. Let us now return for a moment to the particular irrational number which we discussed in §§ 4–5. We there constructed a section by means of the inequalities $x^2 < 2$, $x^2 > 2$. This was a section of the positive rational numbers only; but we replace it (as was explained in § 8) by a section of all the rational numbers. We denote the section or number thus defined by the symbol $\sqrt{2}$.

The classes by means of which the product of $\sqrt{2}$ by itself is defined are (i) (aa'), where a and a' are positive rational numbers whose squares are less than 2, (ii) (AA'), where A and A' are positive rational numbers whose squares are greater than 2. These classes exhaust all positive rational numbers save one, which can only be 2 itself. Thus

$$(\sqrt{2})^2 = \sqrt{2}\sqrt{2} = 2.$$

Again

$$(-\sqrt{2})^2 = (-\sqrt{2})(-\sqrt{2}) = \sqrt{2}\sqrt{2} = (\sqrt{2})^2 = 2.$$

Thus *the equation $x^2 = 2$ has the two roots $\sqrt{2}$ and $-\sqrt{2}$*. Similarly we could discuss the equations $x^2 = 3$, $x^3 = 7, \dots$ and the corresponding irrational numbers $\sqrt{3}$, $-\sqrt{3}$, $\sqrt[3]{7}, \dots$.

13. Quadratic surds. A number of the form $\pm\sqrt{a}$, where a is a positive rational number which is not the square of another rational number, is called a *pure quadratic surd*. A number of the form $a \pm \sqrt{b}$, where a is rational, and $\sqrt{b}$ is a pure quadratic surd, is sometimes called a mixed quadratic surd.

The two numbers $a \pm \sqrt{b}$ are the roots of the quadratic equation

$$x^2 - 2ax + a^2 - b = 0.$$

Conversely, the equation $x^2 + 2px + q = 0$, where p and q are rational, and $p^2 - q > 0$, has as its roots the two quadratic surds $-p \pm \sqrt{p^2 - q}$.

The only kind of irrational numbers whose existence was suggested by the geometrical considerations of § 3 are these quadratic surds, pure and mixed, and the more complicated irrationals which may be expressed in a form involving the repeated extraction of square roots, such as

$$\sqrt{2} + \sqrt{2 + \sqrt{2}} + \sqrt{2 + \sqrt{2 + \sqrt{2}}}.$$

It is easy to construct geometrically a line whose length is equal to any number of this form, as the reader will easily see for himself. That irrational numbers of these kinds *only* can be constructed by Euclidean methods (*i.e.* by geometrical constructions with ruler and compasses) is a point the proof of which must be deferred for the present.* This property of quadratic surds makes them especially interesting.

Examples VII. 1. Give geometrical constructions for

$$\sqrt{2}, \quad \sqrt{2 + \sqrt{2}}, \quad \sqrt{2 + \sqrt{2 + \sqrt{2}}}.$$

2. The quadratic equation $ax^2 + 2bx + c = 0$ has two real roots† if $b^2 - ac > 0$.

---

*See Ch. II, Misc. Exs. 22.

†*I.e.* there are two values of x for which $ax^2 + 2bx + c = 0$. If $b^2 - ac < 0$ there are no such values of x. The reader will remember that in books on elementary algebra the equation is said to have two 'complex' roots. The meaning to be attached to this statement will be explained in Ch. III.

When $b^2 = ac$ the equation has only one root. For the sake of uniformity it is generally said in this case to have 'two equal' roots, but this is a mere convention.

Suppose a, b, c rational. Nothing is lost by taking all three to be integers, for we can multiply the equation by the least common multiple of their denominators.

The reader will remember that the roots are $\{-b \pm \sqrt{b^2 - ac}\}/a$. It is easy to construct these lengths geometrically, first constructing $\sqrt{b^2 - ac}$. A much more elegant, though less straightforward, construction is the following.

Draw a circle of unit radius, a diameter PQ, and the tangents at the ends of the diameters.

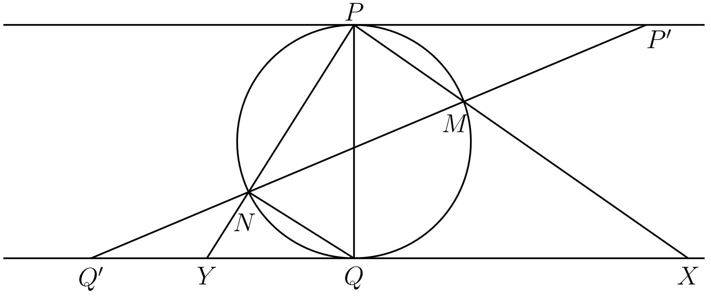

Fig. 5.

*Take $PP' = -2a/b$ and $QQ' = -c/2b$, having regard to sign.** *Join $P'Q'$, cutting the circle in M and N. Draw PM and PN, cutting QQ' in X and Y. Then QX and QY are the roots of the equation with their proper signs.*†

The proof is simple and we leave it as an exercise to the reader. Another, perhaps even simpler, construction is the following. *Take a line AB of unit length. Draw $BC = -2b/a$ perpendicular to AB, and $CD = c/a$ perpendicular to BC and in the same direction as BA. On AD as diameter describe a circle cutting BC in X and Y. Then BX and BY are the roots.*

3. If ac is positive PP' and QQ' will be drawn in the same direction. Verify that $P'Q'$ will not meet the circle if $b^2 < ac$, while if $b^2 = ac$ it will be a tangent. Verify also that if $b^2 = ac$ the circle in the second construction will touch BC.

4. Prove that

$$\sqrt{pq} = \sqrt{p} \times \sqrt{q}, \quad \sqrt{p^2 q} = p\sqrt{q}.$$

---

*The figure is drawn to suit the case in which b and c have the same and a the opposite sign. The reader should draw figures for other cases.

†I have taken this construction from Klein's *Leçons sur certaines questions de géométrie élémentaire* (French translation by J. Griess, Paris, 1896).

14. Some theorems concerning quadratic surds. Two pure quadratic surds are said to be *similar* if they can be expressed as rational multiples of the same surd, and otherwise to be *dissimilar*. Thus

$$\sqrt{8} = 2\sqrt{2}, \quad \sqrt{\tfrac{25}{2}} = \tfrac{5}{2}\sqrt{2},$$

and so $\sqrt{8}$, $\sqrt{\frac{25}{2}}$ are similar surds. On the other hand, if M and N are integers which have no common factor, and neither of which is a perfect square, $\sqrt{M}$ and $\sqrt{N}$ are dissimilar surds. For suppose, if possible,

$$\sqrt{M} = \frac{p}{q}\sqrt{\frac{t}{u}}, \quad \sqrt{N} = \frac{r}{s}\sqrt{\frac{t}{u}},$$

where all the letters denote integers.

Then $\sqrt{MN}$ is evidently rational, and therefore (Ex. II. 3) integral. Thus $MN = P^2$, where P is an integer. Let a, b, c, ... be the prime factors of P, so that

$$MN = a^{2\alpha} b^{2\beta} c^{2\gamma} \ \ldots,$$

where α, β, γ, ... are positive integers. Then MN is divisible by $a^{2\alpha}$, and therefore either (1) M is divisible by $a^{2\alpha}$, or (2) N is divisible by $a^{2\alpha}$, or (3) M and N are both divisible by a. The last case may be ruled out, since M and N have no common factor. This argument may be applied to each of the factors $a^{2\alpha}$, $b^{2\beta}$, $c^{2\gamma}$, ..., so that M must be divisible by some of these factors and N by the remainder. Thus

$$M = P_1^2, \quad N = P_2^2,$$

where P_1^2 denotes the product of some of the factors $a^{2\alpha}$, $b^{2\beta}$, $c^{2\gamma}$, ... and P_2^2 the product of the rest. Hence M and N are both perfect squares, which is contrary to our hypothesis.

THEOREM. *If A, B, C, D are rational and*

$$A + \sqrt{B} = C + \sqrt{D},$$

then either (i) $A = C$, $B = D$ *or* (ii) *B and D are both squares of rational numbers.*

For $B - D$ is rational, and so is

$$\sqrt{B} - \sqrt{D} = C - A.$$

If B is not equal to D (in which case it is obvious that A is also equal to C), it follows that

$$\sqrt{B} + \sqrt{D} = (B - D)/(\sqrt{B} - \sqrt{D})$$

is also rational. Hence $\sqrt{B}$ and $\sqrt{D}$ are rational.

Corollary. *If* $A + \sqrt{B} = C + \sqrt{D}$, *then* $A - \sqrt{B} = C - \sqrt{D}$ (*unless* $\sqrt{B}$ *and* $\sqrt{D}$ *are both rational*).

Examples VIII. 1. Prove *ab initio* that $\sqrt{2}$ and $\sqrt{3}$ are not similar surds.

2. Prove that $\sqrt{a}$ and $\sqrt{1/a}$, where a is rational, are similar surds (unless both are rational).

3. If a and b are rational, then $\sqrt{a}+\sqrt{b}$ cannot be rational unless $\sqrt{a}$ and $\sqrt{b}$ are rational. The same is true of $\sqrt{a} - \sqrt{b}$, unless $a = b$.

4. If

$$\sqrt{A} + \sqrt{B} = \sqrt{C} + \sqrt{D},$$

then either (a) $A = C$ and $B = D$, or (b) $A = D$ and $B = C$, or (c) $\sqrt{A}$, $\sqrt{B}$, $\sqrt{C}$, $\sqrt{D}$ are all rational or all similar surds. [Square the given equation and apply the theorem above.]

5. Neither $(a + \sqrt{b})^3$ nor $(a - \sqrt{b})^3$ can be rational unless $\sqrt{b}$ is rational.

6. Prove that if $x = p + \sqrt{q}$, where p and q are rational, then x^m, where m is any integer, can be expressed in the form $P + Q\sqrt{q}$, where P and Q are rational. For example,

$$(p + \sqrt{q})^2 = p^2 + q + 2p\sqrt{q}, \quad (p + \sqrt{q})^3 = p^3 + 3pq + (3p^2 + q)\sqrt{q}.$$

Deduce that any polynomial in x with rational coefficients (*i.e.* any expression of the form

$$a_0x^n + a_1x^{n-1} + \cdots + a_n,$$

where a_0, ..., a_n are rational numbers) can be expressed in the form $P+Q\sqrt{q}$.

7. If $a+\sqrt{b}$, where b is not a perfect square, is the root of an algebraical equation with rational coefficients, then $a-\sqrt{b}$ is another root of the same equation.

8. Express $1/(p+\sqrt{q})$ in the form prescribed in Ex. 6. [Multiply numerator and denominator by $p-\sqrt{q}$.]

9. Deduce from Exs. 6 and 8 that any expression of the form $G(x)/H(x)$, where $G(x)$ and $H(x)$ are polynomials in x with rational coefficients, can be expressed in the form $P+Q\sqrt{q}$, where P and Q are rational.

10. If p, q, and p^2-q are positive, we can express $\sqrt{p+\sqrt{q}}$ in the form $\sqrt{x}+\sqrt{y}$, where

$$x=\tfrac{1}{2}\{p+\sqrt{p^2-q}\},\quad y=\tfrac{1}{2}\{p-\sqrt{p^2-q}\}.$$

11. Determine the conditions that it may be possible to express $\sqrt{p+\sqrt{q}}$, where p and q are rational, in the form $\sqrt{x}+\sqrt{y}$, where x and y are rational.

12. If a^2-b is positive, the necessary and sufficient conditions that

$$\sqrt{a+\sqrt{b}}+\sqrt{a-\sqrt{b}}$$

should be rational are that a^2-b and $\frac{1}{2}\{a+\sqrt{a^2-b}\}$ should both be squares of rational numbers.

15. The continuum. The aggregate of all real numbers, rational and irrational, is called the **arithmetical continuum**.

It is convenient to suppose that the straight line Λ of § 2 is composed of points corresponding to all the numbers of the arithmetical continuum, and of no others.* The points of the line, the aggregate of which may be said to constitute the **linear continuum**, then supply us with a convenient image of the arithmetical continuum.

We have considered in some detail the chief properties of a few classes of real numbers, such, for example, as rational numbers or quadratic surds.

*This supposition is merely a hypothesis adopted (i) because it suffices for the purposes of our geometry and (ii) because it provides us with convenient geometrical illustrations of analytical processes. As we use geometrical language only for purposes of illustration, it is not part of our business to study the foundations of geometry.

We add a few further examples to show how very special these particular classes of numbers are, and how, to put it roughly, they comprise only a minute fraction of the infinite variety of numbers which constitute the continuum.

(i) Let us consider a more complicated surd expression such as

$$z = \sqrt[3]{4+\sqrt{15}} + \sqrt[3]{4-\sqrt{15}}.$$

Our argument for supposing that the expression for z has a meaning might be as follows. We first show, as in § 12, that there is a number $y = \sqrt{15}$ such that $y^2 = 15$, and we can then, as in § 10, define the numbers $4+\sqrt{15}$, $4-\sqrt{15}$. Now consider the equation in z_1,

$$z_1^3 = 4+\sqrt{15}.$$

The right-hand side of this equation is not rational: but exactly the same reasoning which leads us to suppose that there is a real number x such that $x^3 = 2$ (or any other rational number) also leads us to the conclusion that there is a number z_1 such that $z_1^3 = 4+\sqrt{15}$. We thus define $z_1 = \sqrt[3]{4+\sqrt{15}}$, and similarly we can define $z_2 = \sqrt[3]{4-\sqrt{15}}$; and then, as in § 10, we define $z = z_1 + z_2$.

Now it is easy to verify that

$$z^3 = 3z + 8.$$

And we might have given a direct proof of the existence of a unique number z such that $z^3 = 3z+8$. It is easy to see that there cannot be two such numbers. For if $z_1^3 = 3z_1 + 8$ and $z_2^3 = 3z_2 + 8$, we find on subtracting and dividing by $z_1 - z_2$ that $z_1^2 + z_1 z_2 + z_2^2 = 3$. But if z_1 and z_2 are positive $z_1^3 > 8$, $z_2^3 > 8$ and therefore $z_1 > 2$, $z_2 > 2$, $z_1^2 + z_1 z_2 + z_2^2 > 12$, and so the equation just found is impossible. And it is easy to see that neither z_1 nor z_2 can be negative. For if z_1 is negative and equal to $-\zeta$, ζ is positive and $\zeta^3 - 3\zeta + 8 = 0$, or $3 - \zeta^2 = 8/\zeta$. Hence $3 - \zeta^2 > 0$, and so $\zeta < 2$. But then $8/\zeta > 4$, and so $8/\zeta$ cannot be equal to $3-\zeta^2$, which is less than 3.

Hence there is at most one z such that $z^3 = 3z + 8$. And it cannot be rational. For any rational root of this equation must be integral and a factor of 8 (Ex. II. 3), and it is easy to verify that no one of 1, 2, 4, 8 is a root.

Thus $z^3 = 3z + 8$ has at most one root and that root, if it exists, is positive and not rational. We can now divide the positive rational numbers x into two classes L, R according as $x^3 < 3x + 8$ or $x^3 > 3x + 8$. It is easy to see that if $x^3 > 3x + 8$ and y is any number greater than x, then also $y^3 > 3y + 8$. For suppose if possible $y^3 \leqq 3y+8$. Then since $x^3 > 3x+8$ we obtain on subtracting $y^3 - x^3 < 3(y - x)$, or $y^2 + xy + x^2 < 3$, which is impossible; for y is positive and $x > 2$ (since $x^3 > 8$). Similarly we can show that if $x^3 < 3x + 8$ and $y < x$ then also $y^3 < 3y + 8$.

Finally, it is evident that the classes L and R both exist; and they form a section of the positive rational numbers or positive real number z which satisfies the equation $z^3 = 3z + 8$. The reader who knows how to solve cubic equations by Cardan's method will be able to obtain the explicit expression of z directly from the equation.

(ii) The direct argument applied above to the equation $x^3 = 3x + 8$ could be applied (though the application would be a little more difficult) to the equation

$$x^5 = x + 16,$$

and would lead us to the conclusion that a unique positive real number exists which satisfies this equation. In this case, however, it is not possible to obtain a simple explicit expression for x composed of any combination of surds. It can in fact be proved (though the proof is difficult) that it is *generally* impossible to find such an expression for the root of an equation of higher degree than 4. Thus, besides irrational numbers which can be expressed as pure or mixed quadratic or other surds, or combinations of such surds, there are others which are roots of algebraical equations but cannot be so expressed. It is only in very special cases that such expressions can be found.

(iii) But even when we have added to our list of irrational numbers roots of equations (such as $x^5 = x+16$) which cannot be explicitly expressed as surds, we have not exhausted the different kinds of irrational numbers contained in the continuum. Let us draw a circle whose diameter is equal to A_0A_1, *i.e.* to unity. It is natural to suppose* that the circumference of such a circle has a length capable of numerical measurement. This length

*A proof will be found in Ch. VII.

is usually denoted by π. And it has been shown* (though the proof is unfortunately long and difficult) that this number π is not the root of any algebraical equation with integral coefficients, such, for example, as

$$\pi^2 = n, \quad \pi^3 = n, \quad \pi^5 = \pi + n,$$

where n is an integer. In this way it is possible to define a number which is not rational nor yet belongs to any of the classes of irrational numbers which we have so far considered. And this number π is no isolated or exceptional case. Any number of other examples can be constructed. In fact it is only special classes of irrational numbers which are roots of equations of this kind, just as it is only a still smaller class which can be expressed by means of surds.

16. The continuous real variable. The 'real numbers' may be regarded from two points of view. We may think of them *as an aggregate*, the 'arithmetical continuum' defined in the preceding section, or *individually*. And when we think of them individually, we may think either of a particular *specified* number (such as 1, $-\frac{1}{2}$, $\sqrt{2}$, or π) or we may think of *any* number, *an unspecified* number, *the number* x. This last is our point of view when we make such assertions as 'x is a number', 'x is the measure of a length', 'x may be rational or irrational'. The x which occurs in propositions such as these is called *the continuous real variable*: and the individual numbers are called the *values* of the variable.

A 'variable', however, need not necessarily be continuous. Instead of considering the aggregate of *all* real numbers, we might consider some partial aggregate contained in the former aggregate, such as the aggregate of rational numbers, or the aggregate of positive integers. Let us take the last case. Then in statements about *any* positive integer, or *an unspecified* positive integer, such as 'n is either odd or even', n is called the variable, a *positive integral variable*, and the individual positive integers are its values.

Naturally 'x' and 'n' are only examples of variables, the variable whose 'field of variation' is formed by all the real numbers, and that whose field is

*See Hobson's *Trigonometry* (3rd edition), pp. 305 *et seq.*, or the same writer's *Squaring the Circle* (Cambridge, 1913).

formed by the positive integers. These are the most important examples, but we have often to consider other cases. In the theory of decimals, for instance, we may denote by x any figure in the expression of any number as a decimal. Then x is a variable, but a variable which has only ten different values, viz. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. The reader should think of other examples of variables with different fields of variation. He will find interesting examples in ordinary life: policeman x, the driver of cab x, the year x, the xth day of the week. The values of these variables are naturally not numbers.

17. Sections of the real numbers. In §§ 4–7 we considered 'sections' of the rational numbers, *i.e.* modes of division of the rational numbers (or of the positive rational numbers only) into two classes L and R possessing the following characteristic properties:

(i) that every number of the type considered belongs to one and only one of the two classes;

(ii) that both classes exist;

(iii) that any member of L is less than any member of R.

It is plainly possible to apply the same idea to the aggregate of all real numbers, and the process is, as the reader will find in later chapters, of very great importance.

Let us then suppose* that P and Q are two properties which are mutually exclusive, and one of which is possessed by every real number. Further let us suppose that any number which possesses P is less than any which possesses Q. We call the numbers which possess P the *lower* or *left-hand class* L, and those which possess Q the *upper* or *right-hand class* R.

Thus P might be $x \leqq \sqrt{2}$ and Q be $x > \sqrt{2}$. It is important to observe that a pair of properties which suffice to define a section of the rational numbers

*The discussion which follows is in many ways similar to that of § 6. We have not attempted to avoid a certain amount of repetition. The idea of a 'section,' first brought into prominence in Dedekind's famous pamphlet *Stetigkeit und irrationale Zahlen*, is one which can, and indeed must, be grasped by every reader of this book, even if he be one of those who prefer to omit the discussion of the notion of an irrational number contained in §§ 6–12.

may not suffice to define one of the real numbers. This is so, for example, with the pair '$x < \sqrt{2}$' and '$x > \sqrt{2}$' or (if we confine ourselves to positive numbers) with '$x^2 < 2$' and '$x^2 > 2$'. Every rational number possesses one or other of the properties, but not every real number, since in either case $\sqrt{2}$ escapes classification.

There are now two possibilities.* Either L has a greatest member l, or R has a least member r. *Both* of these events cannot occur. For if L had a greatest member l, and R a least member r, the number $\frac{1}{2}(l + r)$ would be greater than all members of L and less than all members of R, and so could not belong to either class. On the other hand *one* event must occur.†

For let L_1 and R_1 denote the classes formed from L and R by taking only the rational members of L and R. Then the classes L_1 and R_1 form a section of the rational numbers. There are now two cases to distinguish.

It may happen that L_1 has a greatest member α. In this case α must be also the greatest member of L. For if not, we could find a greater, say β. There are rational numbers lying between α and β, and these, being less than β, belong to L, and therefore to L_1; and this is plainly a contradiction. Hence α is the greatest member of L.

On the other hand it may happen that L_1 has no greatest member. In this case the section of the rational numbers formed by L_1 and R_1 is a real number α. This number α must belong to L or to R. If it belongs to L we can show, precisely as before, that it is the greatest member of L, and similarly, if it belongs to R, it is the least member of R.

Thus in any case either L has a greatest member or R a least. Any section of the real numbers therefore 'corresponds' to a real number in the sense in which a section of the rational numbers sometimes, but not always, corresponds to a rational number. This conclusion is of very great importance; for it shows that the consideration of sections of all the real numbers does not lead to any further generalisation of our idea of number. Starting from the rational numbers, we found that the idea of a section of the rational numbers led us to a new conception of a number, that of a real number, more general than that of a rational number; and it might have

*There were three in § 6.

†This was not the case in § 6.

been expected that the idea of a section of the real numbers would have led us to a conception more general still. The discussion which precedes shows that this is not the case, and that the aggregate of real numbers, or the continuum, has a kind of completeness which the aggregate of the rational numbers lacked, a completeness which is expressed in technical language by saying that the continuum is closed.

The result which we have just proved may be stated as follows:

Dedekind's Theorem. *If the real numbers are divided into two classes L and R in such a way that*

(i) *every number belongs to one or other of the two classes,*

(ii) *each class contains at least one number,*

(iii) *any member of L is less than any member of R,*

then there is a number α, which has the property that all the numbers less than it belong to L and all the numbers greater than it to R. The number α itself may belong to either class.

In applications we have often to consider sections not of *all* numbers but of all those contained in an *interval* $[\beta, \gamma]$, that is to say of all numbers x such that $\beta \leqq x \leqq \gamma$. A 'section' of such numbers is of course a division of them into two classes possessing the properties (i), (ii), and (iii). Such a section may be converted into a section of *all* numbers by adding to L all numbers less than β and to R all numbers greater than γ. It is clear that the conclusion stated in Dedekind's Theorem still holds if we substitute 'the real numbers of the interval $[\beta, \gamma]$' for 'the real numbers', and that the number α in this case satisfies the inequalities $\beta \leqq \alpha \leqq \gamma$.

18. Points of accumulation. A system of real numbers, or of the points on a straight line corresponding to them, defined in any way whatever, is called an **aggregate** or **set** of numbers or points. The set might consist, for example, of all the positive integers, or of all the rational points.

It is most convenient here to use the language of geometry.* Suppose

*The reader will hardly require to be reminded that this course is adopted solely for reasons of linguistic convenience.

then that we are given a set of points, which we will denote by S. Take any point ξ, which may or may not belong to S. Then there are two possibilities. Either (i) it is possible to choose a positive number δ so that the interval $[\xi-\delta, \xi+\delta]$ does not contain any point of S, other than ξ itself,* or (ii) this is not possible.

Suppose, for example, that S consists of the points corresponding to all the positive integers. If ξ is itself a positive integer, we can take δ to be any number less than 1, and (i) will be true; or, if ξ is halfway between two positive integers, we can take δ to be any number less than $\frac{1}{2}$. On the other hand, if S consists of all the rational points, then, whatever the value of ξ, (ii) is true; for any interval whatever contains an infinity of rational points.

Let us suppose that (ii) is true. Then any interval $[\xi-\delta, \xi+\delta]$, however small its length, contains at least one point ξ_1 which belongs to S and does not coincide with ξ; and this whether ξ itself be a member of S or not. In this case we shall say that ξ is a **point of accumulation** of S. It is easy to see that the interval $[\xi-\delta, \xi+\delta]$ must contain, not merely one, but infinitely many points of S. For, when we have determined ξ_1, we can take an interval $[\xi-\delta_1, \xi+\delta_1]$ surrounding ξ but not reaching as far as ξ_1. But this interval also must contain a point, say ξ_2, which is a member of S and does not coincide with ξ. Obviously we may repeat this argument, with ξ_2 in the place of ξ_1; and so on indefinitely. In this way we can determine as many points

$$\xi_1, \quad \xi_2, \quad \xi_3, \quad \dots$$

as we please, all belonging to S, and all lying inside the interval $[\xi-\delta, \xi+\delta]$.

A point of accumulation of S may or may not be itself a point of S. The examples which follow illustrate the various possibilities.

Examples IX. 1. If S consists of the points corresponding to the positive integers, or all the integers, there are no points of accumulation.

2. If S consists of all the rational points, every point of the line is a point of accumulation.

*This clause is of course unnecessary if ξ does not itself belong to S.

3. If S consists of the points $1, \frac{1}{2}, \frac{1}{3}, \dots$, there is one point of accumulation, viz. the origin.

4. If S consists of all the positive rational points, the points of accumulation are the origin and all positive points of the line.

19. Weierstrass's Theorem. The general theory of sets of points is of the utmost interest and importance in the higher branches of analysis; but it is for the most part too difficult to be included in a book such as this. There is however one fundamental theorem which is easily deduced from Dedekind's Theorem and which we shall require later.

THEOREM. *If a set S contains infinitely many points, and is entirely situated in an interval $[\alpha, \beta]$, then at least one point of the interval is a point of accumulation of S.*

We divide the points of the line Λ into two classes in the following manner. The point P belongs to L if there are an infinity of points of S to the right of P, and to R in the contrary case. Then it is evident that conditions (i) and (iii) of Dedekind's Theorem are satisfied; and since α belongs to L and β to R, condition (ii) is satisfied also.

Hence there is a point ξ such that, however small be δ, $\xi - \delta$ belongs to L and $\xi + \delta$ to R, so that the interval $[\xi - \delta, \xi + \delta]$ contains an infinity of points of S. Hence ξ is a point of accumulation of S.

This point may of course coincide with α or β, as for instance when $\alpha = 0$, $\beta = 1$, and S consists of the points $1, \frac{1}{2}, \frac{1}{3}, \dots$. In this case 0 is the sole point of accumulation.

MISCELLANEOUS EXAMPLES ON CHAPTER I.

1. What are the conditions that $ax + by + cz = 0$, (1) for all values of x, y, z; (2) for all values of x, y, z subject to $\alpha x + \beta y + \gamma z = 0$; (3) for all values of x, y, z subject to both $\alpha x + \beta y + \gamma z = 0$ and $Ax + By + Cz = 0$?

2. Any positive rational number can be expressed in one and only one way in the form

$$a_1 + \frac{a_2}{1 \cdot 2} + \frac{a_3}{1 \cdot 2 \cdot 3} + \dots + \frac{a_k}{1 \cdot 2 \cdot 3 \dots k},$$

where a_1, a_2, ..., a_k are integers, and

$$0 \leqq a_1, \quad 0 \leqq a_2 < 2, \quad 0 \leqq a_3 < 3, \ \ldots \quad 0 < a_k < k.$$

3. Any positive rational number can be expressed in one and one way only as a simple continued fraction

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}},$$

where a_1, $a_2, \ldots$ are positive integers, of which the first only may be zero.

[Accounts of the theory of such continued fractions will be found in text-books of algebra. For further information as to modes of representation of rational and irrational numbers, see Hobson, *Theory of Functions of a Real Variable*, pp. 45–49.]

4. Find the rational roots (if any) of $9x^3 - 6x^2 + 15x - 10 = 0$.

5. A line AB is divided at C *in aurea sectione* (Euc. II. 11)—*i.e.* so that $AB \cdot AC = BC^2$. Show that the ratio AC/AB is irrational.

[A direct geometrical proof will be found in Bromwich's *Infinite Series*, § 143, p. 363.]

6. A is irrational. In what circumstances can $\dfrac{aA+b}{cA+d}$, where a, b, c, d are rational, be rational?

7. **Some elementary inequalities.** In what follows a_1, $a_2, \ldots$ denote positive numbers (including zero) and p, $q, \ldots$ positive integers. Since $a_1^p - a_2^p$ and $a_1^q - a_2^q$ have the same sign, we have $(a_1^p - a_2^p)(a_1^q - a_2^q) \geqq 0$, or

$$a_1^{p+q} + a_2^{p+q} \geqq a_1^p a_2^q + a_1^q a_2^p, \tag{1}$$

an inequality which may also be written in the form

$$\frac{a_1^{p+q} + a_2^{p+q}}{2} \geqq \left(\frac{a_1^p + a_2^p}{2}\right)\left(\frac{a_1^q + a_2^q}{2}\right). \tag{2}$$

By repeated application of this formula we obtain

$$\frac{a_1^{p+q+r+\ldots} + a_2^{p+q+r+\ldots}}{2} \geqq \left(\frac{a_1^p + a_2^p}{2}\right)\left(\frac{a_1^q + a_2^q}{2}\right)\left(\frac{a_1^r + a_2^r}{2}\right)\ldots, \tag{3}$$

and in particular

$$\frac{a_1^p + a_2^p}{2} \geqq \left(\frac{a_1 + a_2}{2}\right)^p. \tag{4}$$

When $p = q = 1$ in (1), or $p = 2$ in (4), the inequalities are merely different forms of the inequality $a_1^2 + a_2^2 \geqq 2a_1a_2$, which expresses the fact that the arithmetic mean of two positive numbers is not less than their geometric mean.

8. **Generalisations for n numbers.** If we write down the $\frac{1}{2}n(n-1)$ inequalities of the type (1) which can be formed with n numbers $a_1, a_2, \ldots, a_n$, and add the results, we obtain the inequality

$$n\sum a^{p+q} \geqq \sum a^p \sum a^q, \tag{5}$$

or

$$\left(\sum a^{p+q}\right)/n \geqq \left\{\left(\sum a^p\right)/n\right\}\left\{\left(\sum a^q\right)/n\right\}. \tag{6}$$

Hence we can deduce an obvious extension of (3) which the reader may formulate for himself, and in particular the inequality

$$\left(\sum a^p\right)/n \geqq \left\{\left(\sum a\right)/n\right\}^p. \tag{7}$$

9. **The general form of the theorem concerning the arithmetic and geometric means.** An inequality of a slightly different character is that which asserts that the arithmetic mean of $a_1, a_2, \ldots, a_n$ is not less than their geometric mean. Suppose that a_r and a_s are the greatest and least of the a's (if there are several greatest or least a's we may choose any of them indifferently), and let G be their geometric mean. We may suppose $G > 0$, as the truth of the proposition is obvious when $G = 0$. If now we replace a_r and a_s by

$$a_r' = G, \quad a_s' = a_r a_s / G,$$

we do not alter the value of the geometric mean; and, since

$$a_r' + a_s' - a_r - a_s = (a_r - G)(a_s - G)/G \leqq 0,$$

we certainly do not increase the arithmetic mean.

It is clear that we may repeat this argument until we have replaced each of $a_1, a_2, \ldots, a_n$ by G; at most n repetitions will be necessary. As the final value of the arithmetic mean is G, the initial value cannot have been less.

10. **Schwarz's inequality.** Suppose that a_1, a_2, $\dots$, a_n and b_1, b_2, $\dots$, b_n are any two sets of numbers positive or negative. It is easy to verify the identity

$$\left(\sum a_r b_r\right)^2 = \sum a_r^2 \sum a_s^2 - \sum (a_r b_s - a_s b_r)^2,$$

where r and s assume the values 1, 2, $\dots$, n. It follows that

$$\left(\sum a_r b_r\right)^2 \leqq \sum a_r^2 \sum b_r^2,$$

an inequality usually known as Schwarz's (though due originally to Cauchy).

11. If a_1, a_2, $\dots$, a_n are all positive, and $s_n = a_1 + a_2 + \dots + a_n$, then

$$(1 + a_1)(1 + a_2) \dots (1 + a_n) \leqq 1 + s_n + \frac{s_n^2}{2!} + \dots + \frac{s_n^n}{n!}.$$

(*Math. Trip.* 1909.)

12. If a_1, a_2, $\dots$, a_n and b_1, b_2, $\dots$, b_n are two sets of positive numbers, arranged in descending order of magnitude, then

$$(a_1 + a_2 + \dots + a_n)(b_1 + b_2 + \dots + b_n) \leqq n(a_1 b_1 + a_2 b_2 + \dots + a_n b_n).$$

13. If a, b, c, $\dots$ k and A, B, C, $\dots$ K are two sets of numbers, and all of the first set are positive, then

$$\frac{aA + bB + \dots + kK}{a + b + \dots + k}$$

lies between the algebraically least and greatest of A, B, $\dots$, K.

14. If $\sqrt{p}$, $\sqrt{q}$ are dissimilar surds, and $a + b\sqrt{p} + c\sqrt{q} + d\sqrt{pq} = 0$, where a, b, c, d are rational, then $a = 0$, $b = 0$, $c = 0$, $d = 0$.

[Express $\sqrt{p}$ in the form $M + N\sqrt{q}$, where M and N are rational, and apply the theorem of § 14.]

15. Show that if $a\sqrt{2} + b\sqrt{3} + c\sqrt{5} = 0$, where a, b, c are rational numbers, then $a = 0$, $b = 0$, $c = 0$.

16. Any polynomial in $\sqrt{p}$ and $\sqrt{q}$, with rational coefficients (*i.e.* any sum of a finite number of terms of the form $A(\sqrt{p})^m(\sqrt{q})^n$, where m and n are integers, and A rational), can be expressed in the form

$$a + b\sqrt{p} + c\sqrt{q} + d\sqrt{pq},$$

where a, b, c, d are rational.

17. Express $\dfrac{a+b\sqrt{p}+c\sqrt{q}}{d+e\sqrt{p}+f\sqrt{q}}$, where a, b, etc. are rational, in the form

$$A+B\sqrt{p}+C\sqrt{q}+D\sqrt{pq},$$

where A, B, C, D are rational.

[Evidently

$$\begin{aligned}\frac{a+b\sqrt{p}+c\sqrt{q}}{d+e\sqrt{p}+f\sqrt{q}} &= \frac{(a+b\sqrt{p}+c\sqrt{q})(d+e\sqrt{p}-f\sqrt{q})}{(d+e\sqrt{p})^2-f^2q}\\ &= \frac{\alpha+\beta\sqrt{p}+\gamma\sqrt{q}+\delta\sqrt{pq}}{\epsilon+\zeta\sqrt{p}},\end{aligned}$$

where α, β, etc. are rational numbers which can easily be found. The required reduction may now be easily completed by multiplication of numerator and denominator by $\epsilon-\zeta\sqrt{p}$. For example, prove that

$$\frac{1}{1+\sqrt{2}+\sqrt{3}} = \frac{1}{2}+\frac{1}{4}\sqrt{2}-\frac{1}{4}\sqrt{6}.]$$

18. If a, b, x, y are rational numbers such that

$$(ay-bx)^2+4(a-x)(b-y)=0,$$

then either (i) $x=a$, $y=b$ or (ii) $1-ab$ and $1-xy$ are squares of rational numbers. (*Math. Trip.* 1903.)

19. If all the values of x and y given by

$$ax^2+2hxy+by^2=1, \quad a'x^2+2h'xy+b'y^2=1$$

(where a, h, b, a', h', b' are rational) are rational, then

$$(h-h')^2-(a-a')(b-b'), \quad (ab'-a'b)^2+4(ah'-a'h)(bh'-b'h)$$

are both squares of rational numbers. (*Math. Trip.* 1899.)

20. Show that $\sqrt{2}$ and $\sqrt{3}$ are cubic functions of $\sqrt{2}+\sqrt{3}$, with rational coefficients, and that $\sqrt{2}-\sqrt{6}+3$ is the ratio of two linear functions of $\sqrt{2}+\sqrt{3}$. (*Math. Trip.* 1905.)

21. The expression

$$\sqrt{a + 2m\sqrt{a - m^2}} + \sqrt{a - 2m\sqrt{a - m^2}}$$

is equal to $2m$ if $2m^2 > a > m^2$, and to $2\sqrt{a - m^2}$ if $a > 2m^2$.

22. Show that any polynomial in $\sqrt[3]{2}$, with rational coefficients, can be expressed in the form

$$a + b\sqrt[3]{2} + c\sqrt[3]{4},$$

where a, b, c are rational.

More generally, if p is any rational number, any polynomial in $\sqrt[m]{p}$ with rational coefficients can be expressed in the form

$$a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{m-1}\alpha^{m-1},$$

where a_0, $a_1, \ldots$ are rational and $\alpha = \sqrt[m]{p}$. For any such polynomial is of the form

$$b_0 + b_1\alpha + b_2\alpha^2 + \cdots + b_k\alpha^k,$$

where the b's are rational. If $k \leqq m - 1$, this is already of the form required. If $k > m-1$, let α^r be any power of α higher than the $(m-1)$th. Then $r = \lambda m + s$, where λ is an integer and $0 \leqq s \leqq m - 1$; and $\alpha^r = \alpha^{\lambda m + s} = p^\lambda \alpha^s$. Hence we can get rid of all powers of α higher than the $(m - 1)$th.

23. Express $(\sqrt[3]{2} - 1)^5$ and $(\sqrt[3]{2} - 1)/(\sqrt[3]{2} + 1)$ in the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, where a, b, c are rational. [Multiply numerator and denominator of the second expression by $\sqrt[3]{4} - \sqrt[3]{2} + 1$.]

24. If

$$a + b\sqrt[3]{2} + c\sqrt[3]{4} = 0,$$

where a, b, c are rational, then $a = 0$, $b = 0$, $c = 0$.

[Let $y = \sqrt[3]{2}$. Then $y^3 = 2$ and

$$cy^2 + by + a = 0.$$

Hence $2cy^2 + 2by + ay^3 = 0$ or

$$ay^2 + 2cy + 2b = 0.$$

Multiplying these two quadratic equations by a and c and subtracting, we obtain $(ab - 2c^2)y + a^2 - 2bc = 0$, or $y = -(a^2 - 2bc)/(ab - 2c^2)$, a rational number, which is impossible. The only alternative is that $ab - 2c^2 = 0$, $a^2 - 2bc = 0$.

Hence $ab = 2c^2$, $a^4 = 4b^2c^2$. If neither a nor b is zero, we can divide the second equation by the first, which gives $a^3 = 2b^3$: and this is impossible, since $\sqrt[3]{2}$ cannot be equal to the rational number a/b. Hence $ab = 0$, $c = 0$, and it follows from the original equation that a, b, and c are all zero.

As a corollary, if $a + b\sqrt[3]{2} + c\sqrt[3]{4} = d + e\sqrt[3]{2} + f\sqrt[3]{4}$, then $a = d$, $b = e$, $c = f$.

It may be proved, more generally, that if

$$a_0 + a_1 p^{1/m} + \cdots + a_{m-1} p^{(m-1)/m} = 0,$$

p not being a perfect mth power, then $a_0 = a_1 = \cdots = a_{m-1} = 0$; but the proof is less simple.]

25. If $A + \sqrt[3]{B} = C + \sqrt[3]{D}$, then either $A = C$, $B = D$, or B and D are both cubes of rational numbers.

26. If $\sqrt[3]{A} + \sqrt[3]{B} + \sqrt[3]{C} = 0$, then either one of A, B, C is zero, and the other two equal and opposite, or $\sqrt[3]{A}$, $\sqrt[3]{B}$, $\sqrt[3]{C}$ are rational multiples of the same surd $\sqrt[3]{X}$.

27. Find rational numbers α, β such that

$$\sqrt[3]{7 + 5\sqrt{2}} = \alpha + \beta\sqrt{2}.$$

28. If $(a - b^3)b > 0$, then

$$\sqrt[3]{a + \frac{9b^3 + a}{3b}\sqrt{\frac{a - b^3}{3b}}} + \sqrt[3]{a - \frac{9b^3 + a}{3b}\sqrt{\frac{a - b^3}{3b}}}$$

is rational. [Each of the numbers under a cube root is of the form

$$\left\{\alpha + \beta\sqrt{\frac{a - b^3}{3b}}\right\}^3$$

where α and β are rational.]

29. If $\alpha = \sqrt[n]{p}$, any polynomial in α is the root of an equation of degree n, with rational coefficients.

[We can express the polynomial (x say) in the form

$$x = l_1 + m_1\alpha + \cdots + r_1\alpha^{(n-1)},$$

where l_1, $m_1, \dots$ are rational, as in Ex. 22.

Similarly

$$x^2 = l_2 + m_2 a + \dots + r_2 a^{(n-1)},$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$x^n = l_n + m_n a + \dots + r_n a^{(n-1)}.$$

Hence

$$L_1 x + L_2 x^2 + \dots + L_n x^n = \Delta,$$

where Δ is the determinant

$$\begin{vmatrix} l_1 & m_1 & \dots & r_1 \\ l_2 & m_2 & \dots & r_2 \\ \dots & \dots & \dots & \dots \\ l_n & m_n & \dots & r_n \end{vmatrix}$$

and L_1, $L_2, \dots$ the minors of l_1, $l_2, \dots$.]

30. Apply this process to $x = p + \sqrt{q}$, and deduce the theorem of § 14.

31. Show that $y = a + bp^{1/3} + cp^{2/3}$ satisfies the equation

$$y^3 - 3ay^2 + 3y(a^2 - bcp) - a^3 - b^3 p - c^3 p^2 + 3abcp = 0.$$

32. **Algebraical numbers.** We have seen that some irrational numbers (such as $\sqrt{2}$) are roots of equations of the type

$$a_0 x^n + a_1 x^{n-1} + \dots + a_n = 0,$$

where a_0, a_1, $\dots$, a_n are integers. Such irrational numbers are called *algebraical* numbers: all other irrational numbers, such as π (§ 15), are called *transcendental* numbers. Show that if x is an algebraical number, then so are kx, where k is any rational number, and $x^{m/n}$, where m and n are any integers.

33. If x and y are algebraical numbers, then so are $x + y$, $x - y$, xy and x/y. [We have equations

$$a_0 x^m + a_1 x^{m-1} + \dots + a_m = 0,$$
$$b_0 y^n + b_1 y^{n-1} + \dots + b_n = 0,$$

where the a's and b's are integers. Write $x+y=z$, $y=z-x$ in the second, and eliminate x. We thus get an equation of similar form

$$c_0 z^p + c_1 z^{p-1} + \cdots + c_p = 0,$$

satisfied by z. Similarly for the other cases.]

34. If

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0,$$

where $a_0,\ a_1,\ \dots,\ a_n$ are any algebraical numbers, then x is an algebraical number. [We have $n+1$ equations of the type

$$a_{0,r} a_r^{m_r} + a_{1,r} a_r^{m_r-1} + \cdots + a_{m_r,r} = 0 \quad (r = 0,\ 1,\ \dots,\ n),$$

in which the coefficients $a_{0,r},\ a_{1,r}, \dots$ are integers. Eliminate $a_0,\ a_1,\ \dots,\ a_n$ between these and the original equation for x.]

35. Apply this process to the equation $x^2 - 2x\sqrt{2} + \sqrt{3} = 0$.

[The result is $x^8 - 16x^6 + 58x^4 - 48x^2 + 9 = 0$.]

36. Find equations, with rational coefficients, satisfied by

$$1+\sqrt{2}+\sqrt{3}, \quad \frac{\sqrt{3}+\sqrt{2}}{\sqrt{3}-\sqrt{2}}, \quad \sqrt{\sqrt{3}+\sqrt{2}} + \sqrt{\sqrt{3}-\sqrt{2}}, \quad \sqrt[3]{2}+\sqrt[3]{3}.$$

37. If $x^3 = x+1$, then $x^{3n} = a_n x + b_n + c_n/x$, where

$$a_{n+1} = a_n + b_n, \quad b_{n+1} = a_n + b_n + c_n, \quad c_{n+1} = a_n + c_n.$$

38. If $x^6 + x^5 - 2x^4 - x^3 + x^2 + 1 = 0$ and $y = x^4 - x^2 + x - 1$, then y satisfies a quadratic equation with rational coefficients. (*Math. Trip.* 1903.)

[It will be found that $y^2 + y + 1 = 0$.]

CHAPTER II

FUNCTIONS OF REAL VARIABLES

20. The idea of a function. Suppose that x and y are two continuous real variables, which we may suppose to be represented geometrically by distances $A_0P = x$, $B_0Q = y$ measured from fixed points A_0, B_0 along two straight lines Λ, M. And let us suppose that the positions of the points P and Q are not independent, but connected by a relation which we can imagine to be expressed as a relation between x and y: so that, when P and x are known, Q and y are also known. We might, for example, suppose that $y = x$, or $y = 2x$, or $\frac{1}{2}x$, or $x^2 + 1$. In all of these cases the value of x determines that of y. Or again, we might suppose that the relation between x and y is given, not by means of an explicit formula for y in terms of x, but by means of a geometrical construction which enables us to determine Q when P is known.

In these circumstances y is said to be a *function* of x. This notion of functional dependence of one variable upon another is perhaps the most important in the whole range of higher mathematics. In order to enable the reader to be certain that he understands it clearly, we shall, in this chapter, illustrate it by means of a large number of examples.

But before we proceed to do this, we must point out that the simple examples of functions mentioned above possess three characteristics which are by no means involved in the general idea of a function, viz.:

(1) y is determined *for every value of* x;

(2) to each value of x for which y is given corresponds *one and only one value of* y;

(3) the relation between x and y is expressed by means of *an analytical formula*, from which the value of y corresponding to a given value of x can be calculated by direct substitution of the latter.

It is indeed the case that these particular characteristics are possessed by many of the most important functions. But the consideration of the following examples will make it clear that they are by no means essential to a function. All that is essential is that there should be some relation between x and y such that to some values of x at any rate correspond values of y.

Examples X. 1. Let $y = x$ or $2x$ or $\frac{1}{2}x$ or $x^2 + 1$. Nothing further need be said at present about cases such as these.

2. Let $y = 0$ whatever be the value of x. Then y is a function of x, for we can give x any value, and the corresponding value of y (viz. 0) is known. In this case the functional relation makes the same value of y correspond to all values of x. The same would be true were y equal to 1 or $-\frac{1}{2}$ or $\sqrt{2}$ instead of 0. Such a function of x is called *a constant.*

3. Let $y^2 = x$. Then if x is positive this equation defines *two* values of y corresponding to each value of x, viz. $\pm\sqrt{x}$. If $x = 0$, $y = 0$. Hence to the particular value 0 of x corresponds *one* and only one value of y. But if x is negative there is *no* value of y which satisfies the equation. That is to say, the function y is not defined for negative values of x. This function therefore possesses the characteristic (3), but neither (1) nor (2).

4. Consider a volume of gas maintained at a constant temperature and contained in a cylinder closed by a sliding piston.*

Let A be the area of the cross section of the piston and W its weight. The gas, held in a state of compression by the piston, exerts a certain pressure p_0 per unit of area on the piston, which balances the weight W, so that

$$W = Ap_0.$$

Let v_0 be the volume of the gas when the system is thus in equilibrium. If additional weight is placed upon the piston the latter is forced downwards. The volume (v) of the gas diminishes; the pressure (p) which it exerts upon unit area of the piston increases. Boyle's experimental law asserts that the product of p and v is very nearly constant, a correspondence which, if exact, would be represented by an equation of the type

$$pv = a, \tag{i}$$

where a is a number which can be determined approximately by experiment.

Boyle's law, however, only gives a reasonable approximation to the facts provided the gas is not compressed too much. When v is decreased and p increased

*I borrow this instructive example from Prof. H. S. Carslaw's *Introduction to the Calculus.*

beyond a certain point, the relation between them is no longer expressed with tolerable exactness by the equation (i). It is known that a much better approximation to the true relation can then be found by means of what is known as 'van der Waals' law', expressed by the equation

$$\left(p+\frac{\alpha}{v^2}\right)(v-\beta)=\gamma, \tag{ii}$$

where α, β, γ are numbers which can also be determined approximately by experiment.

Of course the two equations, even taken together, do not give anything like a complete account of the relation between p and v. This relation is no doubt in reality much more complicated, and its form changes, as v varies, from a form nearly equivalent to (i) to a form nearly equivalent to (ii). But, from a mathematical point of view, there is nothing to prevent us from contemplating an ideal state of things in which, for all values of v not less than a certain value V, (i) would be exactly true, and (ii) exactly true for all values of v less than V. And then we might regard the two equations as together defining p as a function of v. It is an example of a function which for some values of v is defined by one formula and for other values of v is defined by another.

This function possesses the characteristic (2); to any value of v only one value of p corresponds: but it does not possess (1). For p is not defined as a function of v for negative values of v; a 'negative volume' means nothing, and so negative values of v do not present themselves for consideration at all.

5. Suppose that a perfectly elastic ball is dropped (without rotation) from a height $\frac{1}{2}g\tau^2$ on to a fixed horizontal plane, and rebounds continually.

The ordinary formulae of elementary dynamics, with which the reader is probably familiar, show that $h=\frac{1}{2}gt^2$ if $0 \leqq t \leqq \tau$, $h=\frac{1}{2}g(2\tau-t)^2$ if $\tau \leqq t \leqq 3\tau$, and generally

$$h=\tfrac{1}{2}g(2n\tau-t)^2$$

if $(2n-1)\tau \leqq t \leqq (2n+1)\tau$, h being the depth of the ball, at time t, below its original position. Obviously h is a function of t which is only defined for positive values of t.

6. Suppose that y is defined as being *the largest prime factor of* x. This is an instance of a definition which only applies to a particular class of values of x, viz. *integral* values. 'The largest prime factor of $\frac{11}{3}$ or of $\sqrt{2}$ or of π' means nothing, and so our defining relation fails to define for such values of x as these.

Thus this function does not possess the characteristic (1). It does possess (2), but not (3), as there is no simple formula which expresses y in terms of x.

7. Let y be defined as *the denominator of x when x is expressed in its lowest terms*. This is an example of a function which is defined if and only if x is *rational*. Thus $y = 7$ if $x = -11/7$: but y is not defined for $x = \sqrt{2}$, 'the denominator of $\sqrt{2}$' being a meaningless form of words.

8. Let y be defined as *the height in inches of policeman Cx, in the Metropolitan Police, at* 5.30 P.M. *on* 8 *Aug.* 1907. Then y is defined for a certain number of integral values of x, viz. 1, 2, ..., N, where N is the total number of policemen in division C at that particular moment of time.

21. The graphical representation of functions. Suppose that the variable y is a function of the variable x. It will generally be open to us also to regard x as a function of y, in virtue of the functional relation between x and y. But for the present we shall look at this relation from the first point of view. We shall then call x the *independent variable* and y the *dependent variable*; and, when the particular form of the functional relation is not specified, we shall express it by writing

$$y = f(x)$$

(or $F(x)$, $\phi(x)$, $\psi(x)$, ..., as the case may be).

The nature of particular functions may, in very many cases, be illustrated and made easily intelligible as follows. Draw two lines OX, OY at right angles to one another and produced indefinitely in both directions. We can represent values of x and y by distances measured from O along the lines OX, OY respectively, regard being paid, of course, to sign, and the positive directions of measurement being those indicated by arrows in Fig. 6.

Let a be any value of x for which y is defined and has (let us suppose) the single value b. Take $OA = a$, $OB = b$, and complete the rectangle $OAPB$. Imagine the point P marked on the diagram. This marking of the point P may be regarded as showing that the value of y for $x = a$ is b.

If to the value a of x correspond several values of y (say b, b', b''), we have, instead of the single point P, a number of points P, P', P''.

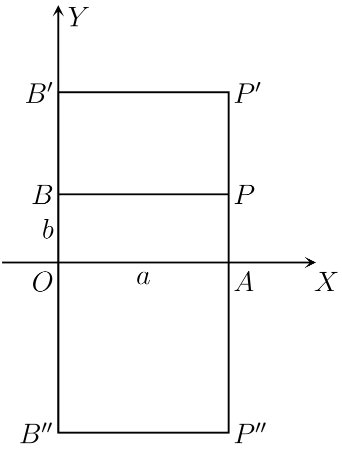

Fig. 6.

We shall call P the *point* (a, b); a and b the *coordinates of P referred to the axes OX, OY*; a the *abscissa*, b the *ordinate* of P; OX and OY the *axis of x* and the *axis of y*, or together the *axes of coordinates*, and O the *origin of coordinates*, or simply the *origin*.

Let us now suppose that for all values a of x for which y is defined, the value b (or values b, b', b'', ...) of y, and the corresponding point P (or points P, P', P'', ...), have been determined. We call the aggregate of all these points the **graph** of the function y.

To take a very simple example, suppose that y is defined as a function of x by the equation

$$Ax + By + C = 0, \tag{1}$$

where A, B, C are any fixed numbers.* Then y is a function of x which possesses all the characteristics (1), (2), (3) of § 20. It is easy to show that *the graph of y is a straight line.* The reader is in all probability familiar with one or other of the various proofs of this proposition which are given in text-books of Analytical Geometry.

We shall sometimes use another mode of expression. We shall say that

*If $B = 0$, y does not occur in the equation. We must then regard y as a function of x defined for one value only of x, viz. $x = -C/A$, and then having *all* values.

when x and y vary in such a way that equation (1) is always true, *the locus of the point* (x, y) *is a straight line*, and we shall call (1) *the equation of the locus*, and say that the equation *represents* the locus. This use of the terms 'locus', 'equation of the locus' is quite general, and may be applied whenever the relation between x and y is capable of being represented by an analytical formula.

The equation $Ax+By+C = 0$ is *the general equation of the first degree*, for $Ax + By + C$ is the most general polynomial in x and y which does not involve any terms of degree higher than the first in x and y. Hence *the general equation of the first degree represents a straight line.* It is equally easy to prove the converse proposition that *the equation of any straight line is of the first degree.*

We may mention a few further examples of interesting geometrical loci defined by equations. An equation of the form

$$(x - \alpha)^2 + (y - \beta)^2 = \rho^2,$$

or

$$x^2 + y^2 + 2Gx + 2Fy + C = 0,$$

where $G^2 + F^2 - C > 0$, represents a circle. The equation

$$Ax^2 + 2Hxy + By^2 + 2Gx + 2Fy + C = 0$$

(*the general equation of the second degree*) represents, assuming that the coefficients satisfy certain inequalities, a conic section, *i.e.* an ellipse, parabola, or hyperbola. For further discussion of these loci we must refer to books on Analytical Geometry.

22. Polar coordinates. In what precedes we have determined the position of P by the lengths of its coordinates $OM = x$, $MP = y$. If $OP = r$ and $MOP = \theta$, θ being an angle between 0 and 2π (measured in the positive direction), it is evident that

$$x = r\cos\theta, \qquad y = r\sin\theta,$$
$$r = \sqrt{x^2 + y^2}, \quad \cos\theta : \sin\theta : 1 :: x : y : r,$$

and that the position of P is equally well determined by a knowledge of r and θ. We call r and θ the *polar coordinates* of P. The former, it should be observed, is essentially positive.*

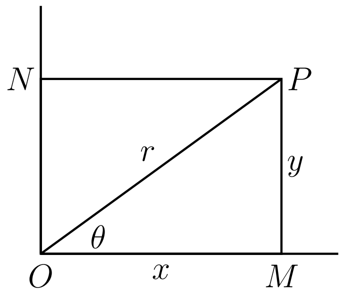

Fig. 7.

If P moves on a locus there will be some relation between r and θ, say $r = f(\theta)$ or $\theta = F(r)$. This we call the *polar equation* of the locus. The polar equation may be deduced from the (x, y) equation (or *vice versa*) by means of the formulae above.

Thus the polar equation of a straight line is of the form

$$r\cos(\theta - \alpha) = p,$$

where p and α are constants. The equation $r = 2a\cos\theta$ represents a circle passing through the origin; and the general equation of a circle is of the form

$$r^2 + c^2 - 2rc\cos(\theta - \alpha) = A^2,$$

where A, c, and α are constants.

*Polar coordinates are sometimes defined so that r may be positive or negative. In this case two pairs of coordinates—*e.g.* $(1, 0)$ and $(-1, \pi)$—correspond to the same point. The distinction between the two systems may be illustrated by means of the equation $l/r = 1 - e\cos\theta$, where $l > 0$, $e > 1$. According to our definitions r must be positive and therefore $\cos\theta < 1/e$: the equation represents one branch only of a hyperbola, the other having the equation $-l/r = 1 - e\cos\theta$. With the system of coordinates which admits negative values of r, the equation represents the whole hyperbola.

23. Further examples of functions and their graphical representation. The examples which follow will give the reader a better notion of the infinite variety of possible types of functions.

A. **Polynomials.** A *polynomial* in x is a function of the form

$$a_0x^m + a_1x^{m-1} + \cdots + a_m,$$

where a_0, a_1, ..., a_m are constants. The simplest polynomials are the simple powers $y = x$, x^2, x^3, ..., x^m, ... The graph of the function x^m is of two distinct types, according as m is even or odd.

First let $m = 2$. Then three points on the graph are $(0, 0)$, $(1, 1)$, $(-1, 1)$. Any number of additional points on the graph may be found by assigning other special values to x: thus the values

$$x = \tfrac{1}{2},\quad 2,\quad 3,\quad -\tfrac{1}{2},\quad -2,\quad 3$$

give

$$y = \tfrac{1}{4},\quad 4,\quad 9,\quad \tfrac{1}{4},\quad 4,\quad 9.$$

If the reader will plot off a fair number of points on the graph, he will be led to conjecture that the form of the graph is something like that shown in Fig. 8. If he draws a curve through the special points which he has proved to lie on the graph and then tests its accuracy by giving x new values, and calculating the corresponding values of y, he will find that they lie as near to the curve as it is reasonable to expect, when the inevitable inaccuracies of drawing are considered. The curve is of course a parabola.

There is, however, one fundamental question which we cannot answer adequately at present. The reader has no doubt some notion as to what is meant by a *continuous* curve, a curve without breaks or jumps; such a curve, in fact, as is roughly represented in Fig. 8. The question is whether the graph of the function $y = x^2$ is in fact such a curve. This cannot be *proved* by merely constructing any number of isolated points on the curve, although the more such points we construct the more probable it will appear.

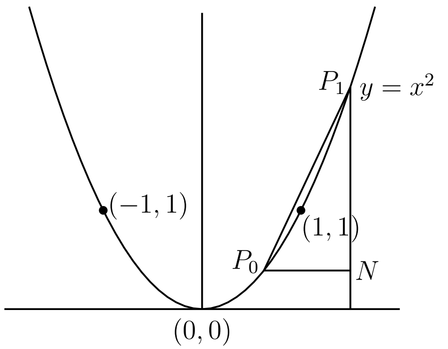

Fig. 8.

This question cannot be discussed properly until Ch. V. In that chapter we shall consider in detail what our common sense idea of continuity really means, and how we can prove that such graphs as the one now considered, and others which we shall consider later on in this chapter, are really continuous curves. For the present the reader may be content to draw his curves as common sense dictates.

It is easy to see that the curve $y = x^2$ is everywhere convex to the axis of x. Let P_0, P_1 (Fig. 8) be the points (x_0, x_0^2), (x_1, x_1^2). Then the coordinates of a point on the chord P_0P_1 are $x = \lambda x_0 + \mu x_1$, $y = \lambda x_0^2 + \mu x_1^2$, where λ and μ are positive numbers whose sum is 1. And

$$y - x^2 = (\lambda + \mu)(\lambda x_0^2 + \mu x_1^2) - (\lambda x_0 + \mu x_1)^2 = \lambda\mu(x_1 - x_0)^2 \geqq 0,$$

so that the chord lies entirely above the curve.

The curve $y = x^4$ is similar to $y = x^2$ in general appearance, but flatter near O, and steeper beyond the points A, A' (Fig. 9), and $y = x^m$, where m is even and greater than 4, is still more so. As m gets larger and larger the flatness and steepness grow more and more pronounced, until the curve is practically indistinguishable from the thick line in the figure.

The reader should next consider the curves given by $y = x^m$, when m is odd. The fundamental difference between the two cases is that whereas when m is even $(-x)^m = x^m$, so that the curve is symmetrical about OY, when m is odd $(-x)^m = -x^m$, so that y is negative when x is negative.

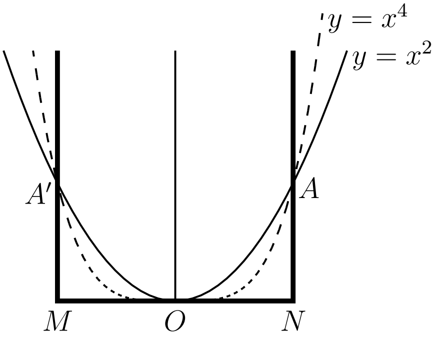

Fig. 9.

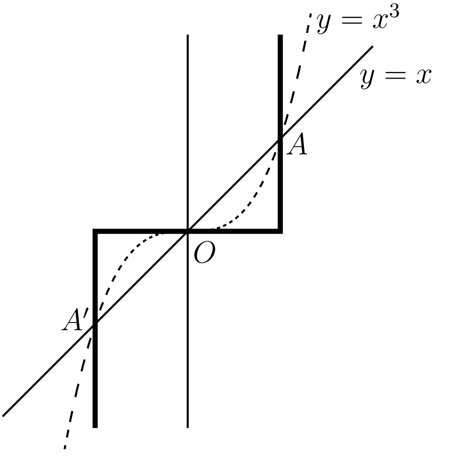

Fig. 10.

Fig. 10 shows the curves $y = x$, $y = x^3$, and the form to which $y = x^m$ approximates for larger odd values of m.

It is now easy to see how (theoretically at any rate) the graph of any polynomial may be constructed. In the first place, from the graph of $y = x^m$ we can at once derive that of Cx^m, where C is a constant, by multiplying the ordinate of every point of the curve by C. And if we know the graphs of $f(x)$ and $F(x)$, we can find that of $f(x) + F(x)$ by taking the ordinate of every point to be the sum of the ordinates of the corresponding points on the two original curves.

The drawing of graphs of polynomials is however so much facilitated by the use of more advanced methods, which will be explained later on, that we shall not pursue the subject further here.

Examples XI. 1. Trace the curves $y = 7x^4$, $y = 3x^5$, $y = x^{10}$.

[The reader should draw the curves carefully, and all three should be drawn in one figure.* He will then realise how rapidly the higher powers of x increase,

*It will be found convenient to take the scale of measurement along the axis of y a good deal smaller than that along the axis of x, in order to prevent the figure becoming of an awkward size.

as x gets larger and larger, and will see that, in such a polynomial as

$$x^{10} + 3x^5 + 7x^4$$

(or even $x^{10} + 30x^5 + 700x^4$), it is the *first* term which is of really preponderant importance when x is fairly large. Thus even when $x = 4$, $x^{10} > 1{,}000{,}000$, while $30x^5 < 35{,}000$ and $700x^4 < 180{,}000$; while if $x = 10$ the preponderance of the first term is still more marked.]

2. Compare the relative magnitudes of x^{12}, $1{,}000{,}000x^6$, $1{,}000{,}000{,}000{,}000x$ when $x = 1$, 10, 100, etc.

[The reader should make up a number of examples of this type for himself. This idea of the *relative rate of growth* of different functions of x is one with which we shall often be concerned in the following chapters.]

3. Draw the graph of $ax^2 + 2bx + c$.

[Here $y - \{(ac - b^2)/a\} = a\{x + (b/a)\}^2$. If we take new axes parallel to the old and passing through the point $x = -b/a$, $y = (ac - b^2)/a$, the new equation is $y' = ax'^2$. The curve is a parabola.]

4. Trace the curves $y = x^3 - 3x + 1$, $y = x^2(x - 1)$, $y = x(x - 1)^2$.

24. B. **Rational Functions.** The class of functions which ranks next to that of polynomials in simplicity and importance is that of *rational functions*. A rational function is the quotient of one polynomial by another: thus if $P(x)$, $Q(x)$ are polynomials, we may denote the general rational function by

$$R(x) = \frac{P(x)}{Q(x)}.$$

In the particular case when $Q(x)$ reduces to unity or any other constant (*i.e.* does not involve x), $R(x)$ reduces to a polynomial: thus the class of rational functions includes that of polynomials as a sub-class. The following points concerning the definition should be noticed.

(1) We usually suppose that $P(x)$ and $Q(x)$ have no common factor $x + a$ or $x^p + ax^{p-1} + bx^{p-2} + \cdots + k$, all such factors being removed by division.

(2) It should however be observed that this removal of common factors *does as a rule change the function*. Consider for example the function x/x, which is a rational function. On removing the common factor x we obtain $1/1 = 1$. But the

original function is not *always* equal to 1: it is equal to 1 only so long as $x \neq 0$. If $x = 0$ it takes the form $0/0$, which is meaningless. Thus the function x/x is equal to 1 if $x \neq 0$ and is undefined when $x = 0$. It therefore differs from the function 1, which is *always* equal to 1.

(3) Such a function as

$$\left(\frac{1}{x+1} + \frac{1}{x-1}\right) \Big/ \left(\frac{1}{x} + \frac{1}{x-2}\right)$$

may be reduced, by the ordinary rules of algebra, to the form

$$\frac{x^2(x-2)}{(x-1)^2(x+1)},$$

which is a rational function of the standard form. But here again it must be noticed that the reduction is not *always* legitimate. In order to calculate the value of a function for a given value of x we must substitute the value for x in the function *in the form in which it is given*. In the case of this function the values $x = -1$, 1, 0, 2 all lead to a meaningless expression, and so the function is not defined for these values. The same is true of the reduced form, so far as the values -1 and 1 are concerned. But $x = 0$ and $x = 2$ give the value 0. Thus once more the two functions are not the same.

(4) But, as appears from the particular example considered under (3), there will generally be a certain number of values of x for which the function is not defined even when it has been reduced to a rational function of the standard form. These are the values of x (if any) for which the denominator vanishes. Thus $(x^2 - 7)/(x^2 - 3x + 2)$ is not defined when $x = 1$ or 2.

(5) Generally we agree, in dealing with expressions such as those considered in (2) and (3), to disregard the exceptional values of x for which such processes of simplification as were used there are illegitimate, and to reduce our function to the standard form of rational function. The reader will easily verify that (on this understanding) the sum, product, or quotient of two rational functions may themselves be reduced to rational functions of the standard type. And generally *a rational function of a rational function is itself a rational function*: *i.e.* if in $z = P(y)/Q(y)$, where P and Q are polynomials, we substitute $y = P_1(x)/Q_1(x)$, we obtain on simplification an equation of the form $z = P_2(x)/Q_2(x)$.

(6) It is in no way presupposed in the definition of a rational function that the constants which occur as coefficients should be rational *numbers*. The word

rational has reference solely to the way in which the variable x appears in the function. Thus

$$\frac{x^2 + x + \sqrt{3}}{x\sqrt[3]{2} - \pi}$$

is a rational function.

The use of the word rational arises as follows. The rational function $P(x)/Q(x)$ may be generated from x by a finite number of operations upon x, including only multiplication of x by itself or a constant, addition of terms thus obtained and division of one function, obtained by such multiplications and additions, by another. In so far as the variable x is concerned, this procedure is very much like that by which all rational numbers can be obtained from unity, a procedure exemplified in the equation

$$\frac{5}{3} = \frac{1+1+1+1+1}{1+1+1}.$$

Again, *any* function which can be deduced from x by the elementary operations mentioned above using at each stage of the process functions which have already been obtained from x in the same way, can be reduced to the standard type of rational function. The most general kind of function which can be obtained in this way is sufficiently illustrated by the example

$$\left(\frac{x}{x^2+1} + \frac{2x+7}{x^2 + \dfrac{11x - 3\sqrt{2}}{9x+1}}\right) \Bigg/ \left(17 + \frac{2}{x^3}\right),$$

which can obviously be reduced to the standard type of rational function.

25. The drawing of graphs of rational functions, even more than that of polynomials, is immensely facilitated by the use of methods depending upon the differential calculus. We shall therefore content ourselves at present with a very few examples.

Examples XII. 1. Draw the graphs of $y = 1/x$, $y = 1/x^2$, $y = 1/x^3$, ...

[The figures show the graphs of the first two curves. It should be observed that since $1/0$, $1/0^2$, ... are meaningless expressions, these functions are not defined for $x = 0$.]

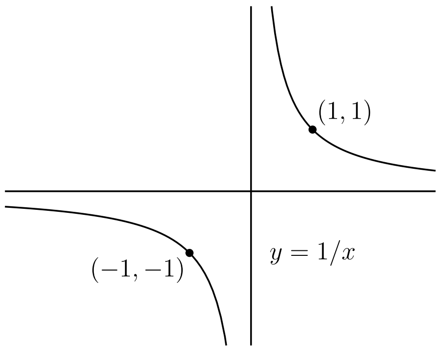

Fig. 11.

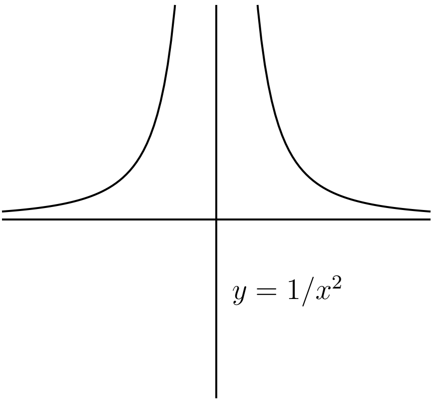

Fig. 12.

2. Trace $y = x + (1/x)$, $x - (1/x)$, $x^2 + (1/x^2)$, $x^2 - (1/x^2)$ and $ax + (b/x)$ taking various values, positive and negative, for a and b.

3. Trace

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

4. Trace $y = 1/(x-a)(x-b)$, $1/(x-a)(x-b)(x-c)$, where $a < b < c$.

5. Sketch the general form assumed by the curves $y = 1/x^m$ as m becomes larger and larger, considering separately the cases in which m is odd or even.

26. C. **Explicit Algebraical Functions.** The next important class of functions is that of *explicit algebraical functions*. These are functions which can be generated from x by a finite number of operations such as those used in generating rational functions, together with a finite number of operations of root extraction. Thus

$$\frac{\sqrt{1+x} - \sqrt[3]{1-x}}{\sqrt{1+x} + \sqrt[3]{1-x}}, \quad \sqrt{x} + \sqrt{x + \sqrt{x}}, \quad \left(\frac{x^2 + x + \sqrt{3}}{x\sqrt[3]{2} - \pi}\right)^{\frac{2}{3}}$$

are explicit algebraical functions, and so is $x^{m/n}$ (*i.e.* $\sqrt[n]{x^m}$), where m and n are any integers.

It should be noticed that there is an ambiguity of notation involved in such an equation as $y = \sqrt{x}$. We have, up to the present, regarded (*e.g.*) $\sqrt{2}$ as denoting the *positive* square root of 2, and it would be natural to denote by $\sqrt{x}$, where x is any positive number, the positive square root of x, in which case $y = \sqrt{x}$ would be a one-valued function of x. It is however often more convenient to regard $\sqrt{x}$ as standing for the two-valued function whose two values are the positive and negative square roots of x.

The reader will observe that, when this course is adopted, the function $\sqrt{x}$ differs fundamentally from rational functions in two respects. In the first place a rational function is always defined for all values of x with a certain number of isolated exceptions. But $\sqrt{x}$ is undefined for a *whole range* of values of x (*i.e.* all negative values). Secondly the function, when x has a value for which it is defined, has generally two values of opposite signs.

The function $\sqrt[3]{x}$, on the other hand, is one-valued and defined for all values of x.

Examples XIII. 1. $\sqrt{(x-a)(b-x)}$, where $a < b$, is defined only for $a \leqq x \leqq b$. If $a < x < b$ it has two values: if $x = a$ or b only one, viz. 0.

2. Consider similarly

$$\sqrt{(x-a)(x-b)(x-c)} \quad (a < b < c),$$
$$\sqrt{x(x^2-a^2)}, \quad \sqrt[3]{(x-a)^2(b-x)} \quad (a < b),$$
$$\frac{\sqrt{1+x}-\sqrt{1-x}}{\sqrt{1+x}+\sqrt{1-x}}, \quad \sqrt{x+\sqrt{x}}.$$

3. Trace the curves $y^2 = x$, $y^3 = x$, $y^2 = x^3$.

4. Draw the graphs of the functions

$$y = \sqrt{a^2-x^2}, \quad y = b\sqrt{1-(x^2/a^2)}.$$

27. D. **Implicit Algebraical Functions.** It is easy to verify that if

$$y = \frac{\sqrt{1+x}-\sqrt[3]{1-x}}{\sqrt{1+x}+\sqrt[3]{1-x}},$$

then

$$\left(\frac{1+y}{1-y}\right)^6 = \frac{(1+x)^3}{(1-x)^2};$$

or if

$$y = \sqrt{x} + \sqrt{x+\sqrt{x}},$$

then

$$y^4 - (4y^2 + 4y + 1)x = 0.$$

Each of these equations may be expressed in the form

$$y^m + R_1 y^{m-1} + \cdots + R_m = 0, \tag{1}$$

where R_1, R_2, ..., R_m are rational functions of x: and the reader will easily verify that, if y is any one of the functions considered in the last set of examples, y satisfies an equation of this form. It is naturally suggested that the same is true of any explicit algebraic function. And this is in fact true, and indeed not difficult to prove, though we shall not delay to write out a formal proof here. An example should make clear to the reader the lines on which such a proof would proceed. Let

$$y = \frac{x + \sqrt{x} + \sqrt{x+\sqrt{x}} + \sqrt[3]{1+x}}{x - \sqrt{x} + \sqrt{x+\sqrt{x}} - \sqrt[3]{1+x}}.$$

Then we have the equations

$$y = \frac{x+u+v+w}{x-u+v-w},$$
$$u^2 = x, \quad v^2 = x+u, \quad w^3 = 1+x,$$

and we have only to eliminate u, v, w between these equations in order to obtain an equation of the form desired.

We are therefore led to give the following definition: *a function $y = f(x)$ will be said to be an algebraical function of x if it is the root of an equation such as* (1), *i.e. the root of an equation of the m^{th} degree in y, whose coefficients are rational functions of x.* There is plainly no loss of generality in supposing the first coefficient to be unity.

This class of functions includes all the explicit algebraical functions considered in § 26. But it also includes other functions which cannot be expressed as explicit algebraical functions. For it is known that in general such an equation as (1) cannot be solved explicitly for y in terms of x, when m is greater than 4, though such a solution is always possible if $m = 1$, 2, 3, or 4 and in special cases for higher values of m.

The definition of an algebraical function should be compared with that of an algebraical number given in the last chapter (Misc. Exs. 32).

Examples XIV. 1. If $m = 1$, y is a rational function.

2. If $m = 2$, the equation is $y^2 + R_1 y + R_2 = 0$, so that

$$y = \tfrac{1}{2}\{-R_1 \pm \sqrt{R_1^2 - 4R_2}\}.$$

This function is defined for all values of x for which $R_1^2 \geqq 4R_2$. It has two values if $R_1^2 > 4R_2$ and one if $R_1^2 = 4R_2$.

If $m = 3$ or 4, we can use the methods explained in treatises on Algebra for the solution of cubic and biquadratic equations. But as a rule the process is complicated and the results inconvenient in form, and we can generally study the properties of the function better by means of the original equation.

3. Consider the functions defined by the equations

$$y^2 - 2y - x^2 = 0, \quad y^2 - 2y + x^2 = 0, \quad y^4 - 2y^2 + x^2 = 0,$$

in each case obtaining y as an explicit function of x, and stating for what values of x it is defined.

4. Find algebraical equations, with coefficients rational in x, satisfied by each of the functions

$$\sqrt{x} + \sqrt{1/x}, \quad \sqrt[3]{x} + \sqrt[3]{1/x}, \quad \sqrt{x + \sqrt{x}}, \quad \sqrt{x + \sqrt{x + \sqrt{x}}}.$$

5. Consider the equation $y^4 = x^2$.

[Here $y^2 = \pm x$. If x is positive, $y = \sqrt{x}$: if negative, $y = \sqrt{-x}$. Thus the function has two values for all values of x save $x = 0$.]

6. An algebraical function of an algebraical function of x is itself an algebraical function of x.

[For we have

$$y^m + R_1(z)y^{m-1} + \dots + R_m(z) = 0,$$

where

$$z^n + S_1(x)z^{n-1} + \dots + S_n(x) = 0.$$

Eliminating z we find an equation of the form

$$y^p + T_1(x)y^{p-1} + \dots + T_p(x) = 0.$$

Here all the capital letters denote rational functions.]

7. An example should perhaps be given of an algebraical function which cannot be expressed in an explicit algebraical form. Such an example is the function y defined by the equation

$$y^5 - y - x = 0.$$

But the proof that we cannot find an explicit algebraical expression for y in terms of x is difficult, and cannot be attempted here.

28. Transcendental functions. All functions of x which are not rational or even algebraical are called *transcendental* functions. This class of functions, being defined in so purely negative a manner, naturally includes an infinite variety of whole kinds of functions of varying degrees of simplicity and importance. Among these we can at present distinguish two kinds which are particularly interesting.

E. **The direct and inverse trigonometrical or circular functions.** These are the sine and cosine functions of elementary trigonometry, and their inverses, and the functions derived from them. We may assume provisionally that the reader is familiar with their most important properties.*

*The definitions of the circular functions given in elementary trigonometry presuppose that any sector of a circle has associated with it a definite number called its *area*. How this assumption is justified will appear in Ch. VII.

Examples XV. 1. Draw the graphs of $\cos x$, $\sin x$, and $a\cos x + b\sin x$.

[Since $a\cos x + b\sin x = \beta\cos(x-\alpha)$, where $\beta = \sqrt{a^2+b^2}$, and α is an angle whose cosine and sine are $a/\sqrt{a^2+b^2}$ and $b/\sqrt{a^2+b^2}$, the graphs of these three functions are similar in character.]

2. Draw the graphs of $\cos^2 x$, $\sin^2 x$, $a\cos^2 x + b\sin^2 x$.

3. Suppose the graphs of $f(x)$ and $F(x)$ drawn. Then the graph of

$$f(x)\cos^2 x + F(x)\sin^2 x$$

is a wavy curve which oscillates between the curves $y = f(x)$, $y = F(x)$. Draw the graph when $f(x) = x$, $F(x) = x^2$.

4. Show that the graph of $\cos px + \cos qx$ lies between those of $2\cos\frac{1}{2}(p-q)x$ and $-2\cos\frac{1}{2}(p+q)x$, touching each in turn. Sketch the graph when $(p-q)/(p+q)$ is small. (*Math. Trip.* 1908.)

5. Draw the graphs of $x + \sin x$, $(1/x) + \sin x$, $x\sin x$, $(\sin x)/x$.

6. Draw the graph of $\sin(1/x)$.

[If $y = \sin(1/x)$, then $y = 0$ when $x = 1/m\pi$, where m is any integer. Similarly $y = 1$ when $x = 1/(2m+\frac{1}{2})\pi$ and $y = -1$ when $x = 1/(2m-\frac{1}{2})\pi$. The curve is entirely comprised between the lines $y = -1$ and $y = 1$ (Fig. 13). It oscillates up and down, the rapidity of the oscillations becoming greater and greater as x approaches 0. For $x = 0$ the function is undefined. When x is large y is small.* The negative half of the curve is similar in character to the positive half.]

7. Draw the graph of $x\sin(1/x)$.

[This curve is comprised between the lines $y = -x$ and $y = x$ just as the last curve is comprised between the lines $y = -1$ and $y = 1$ (Fig. 14).]

8. Draw the graphs of $x^2\sin(1/x)$, $(1/x)\sin(1/x)$, $\sin^2(1/x)$, $\{x\sin(1/x)\}^2$, $a\cos^2(1/x) + b\sin^2(1/x)$, $\sin x + \sin(1/x)$, $\sin x\sin(1/x)$.

9. Draw the graphs of $\cos x^2$, $\sin x^2$, $a\cos x^2 + b\sin x^2$.

10. Draw the graphs of $\arccos x$ and $\arcsin x$.

[If $y = \arccos x$, $x = \cos y$. This enables us to draw the graph of x, considered as a function of y, and the same curve shows y as a function of x. It is clear that y is only defined for $-1 \leqq x \leqq 1$, and is infinitely many-valued for these values of x. As the reader no doubt remembers, there is, when $-1 < x < 1$, a

*See Chs. IV and V for explanations as to the precise meaning of this phrase.

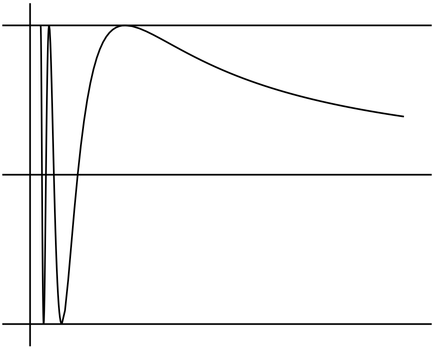

Fig. 13.

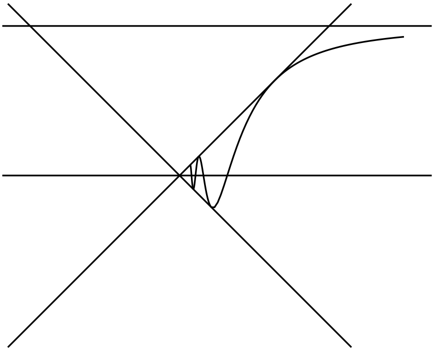

Fig. 14.

value of y between 0 and π, say α, and the other values of y are given by the formula $2n\pi \pm \alpha$, where n is any integer, positive or negative.]

11. Draw the graphs of

$$\tan x, \quad \cot x, \quad \sec x, \quad \operatorname{cosec} x, \quad \tan^2 x, \quad \cot^2 x, \quad \sec^2 x, \quad \operatorname{cosec}^2 x.$$

12. Draw the graphs of $\arctan x$, $\operatorname{arc\,cot} x$, $\operatorname{arc\,sec} x$, $\operatorname{arc\,cosec} x$. Give formulae (as in Ex. 10) expressing all the values of each of these functions in terms of any particular value.

13. Draw the graphs of $\tan(1/x)$, $\cot(1/x)$, $\sec(1/x)$, $\operatorname{cosec}(1/x)$.

14. Show that $\cos x$ and $\sin x$ are not rational functions of x.

[A function is said to be *periodic*, with period a, if $f(x) = f(x + a)$ for all values of x for which $f(x)$ is defined. Thus $\cos x$ and $\sin x$ have the period 2π. It is easy to see that no periodic function can be a rational function, unless it is a constant. For suppose that

$$f(x) = P(x)/Q(x),$$

where P and Q are polynomials, and that $f(x) = f(x+a)$, each of these equations holding for all values of x. Let $f(0) = k$. Then the equation $P(x) - kQ(x) = 0$ is satisfied by an infinite number of values of x, viz. $x = 0$, a, $2a$, etc., and therefore for all values of x. Thus $f(x) = k$ for all values of x, *i.e.* $f(x)$ is a constant.]

15. Show, more generally, that no function with a period can be an algebraical function of x.

[Let the equation which defines the algebraical function be

$$y^m + R_1 y^{m-1} + \cdots + R_m = 0 \tag{1}$$

where $R_1, \ldots$ are rational functions of x. This may be put in the form

$$P_0 y^m + P_1 y^{m-1} + \cdots + P_m = 0,$$

where $P_0, P_1, \ldots$ are polynomials in x. Arguing as above, we see that

$$P_0 k^m + P_1 k^{m-1} + \cdots + P_m = 0$$

for all values of x. Hence $y = k$ satisfies the equation (1) for all values of x, and one set of values of our algebraical function reduces to a constant.

Now divide (1) by $y - k$ and repeat the argument. Our final conclusion is that our algebraical function has, for any value of x, the same set of values $k, k', \ldots$; *i.e.* it is composed of a certain number of constants.]

16. The inverse sine and inverse cosine are not rational or algebraical functions. [This follows from the fact that, for any value of x between -1 and $+1$, $\arcsin x$ and $\arccos x$ have infinitely many values.]

29. F. **Other classes of transcendental functions.** Next in importance to the trigonometrical functions come the exponential and logarithmic functions, which will be discussed in Chs. IX and X. But these functions are beyond our range at present. And most of the other classes of transcendental functions whose properties have been studied, such as the elliptic functions, Bessel's and Legendre's functions, Gamma-functions, and so forth, lie altogether beyond the scope of this book. There are however some elementary types of functions which, though of much less importance theoretically than the rational, algebraical, or trigonometrical functions, are particularly instructive as illustrations of the possible varieties of the functional relation.

Examples XVI. 1. Let $y = [x]$, where $[x]$ denotes the greatest integer not greater than x. The graph is shown in Fig. 15a. The left-hand end points of the thick lines, but not the right-hand ones, belong to the graph.

2. $y = x - [x]$. (Fig. 15b.)

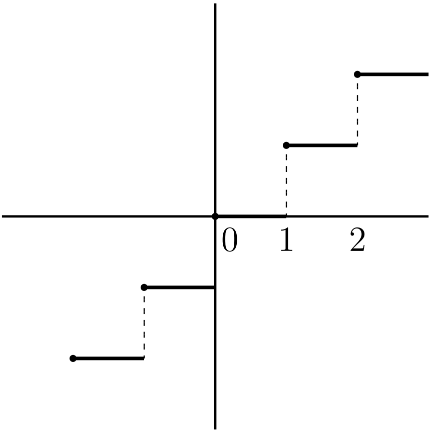

Fig. 15a.

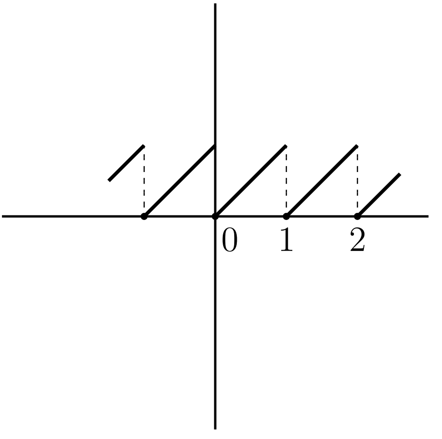

Fig. 15b.

3. $y = \sqrt{x - [x]}$. (Fig. 15c.)
4. $y = [x] + \sqrt{x - [x]}$. (Fig. 15d.)
5. $y = (x - [x])^2$, $[x] + (x - [x])^2$.
6. $y = [\sqrt{x}]$, $[x^2]$, $\sqrt{x} - [\sqrt{x}]$, $x^2 - [x^2]$, $[1 - x^2]$.

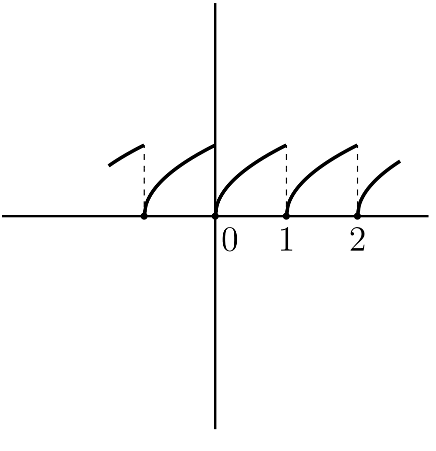

Fig. 15c.

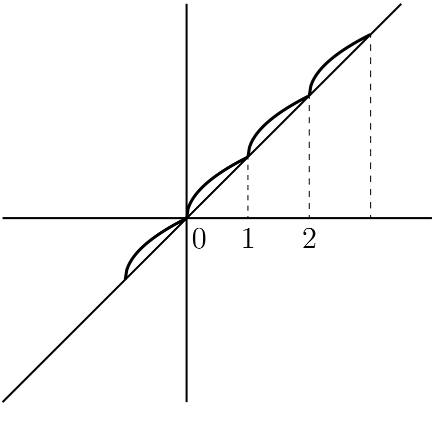

Fig. 15d.

7. Let y be defined as *the largest prime factor of* x (cf. Exs. X. 6). Then

y is defined only for integral values of x. If

$$x = 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ \dots,$$

then

$$y = 1,\ 2,\ 3,\ 2,\ 5,\ 3,\ 7,\ 2,\ 3,\quad 5,\ 11,\quad 3,\ 13,\ \dots.$$

The graph consists of a number of isolated points.

8. Let y be *the denominator of* x (Exs. x. 7). In this case y is defined only for rational values of x. We can mark off as many points on the graph as we please, but the result is not in any ordinary sense of the word a curve, and there are no points corresponding to any irrational values of x.

Draw the straight line joining the points $(N-1, N)$, (N, N), where N is a positive integer. Show that the number of points of the locus which lie on this line is equal to the number of positive integers less than and prime to N.

9. Let $y = 0$ when x is an integer, $y = x$ when x is not an integer. The graph is derived from the straight line $y = x$ by taking out the points

$$\dots\ (-1,-1),\quad (0,0),\quad (1,1),\quad (2,2),\ \dots$$

and adding the points $(-1, 0)$, $(0, 0)$, $(1, 0)$, ... on the axis of x.

The reader may possibly regard this as an unreasonable function. *Why*, he may ask, if y is equal to x for all values of x save integral values, should it not be equal to x for integral values too? The answer is simply, *why should it*? The function y does in point of fact answer to the definition of a function: there is a relation between x and y such that when x is known y is known. We are perfectly at liberty to take this relation to be what we please, however arbitrary and apparently futile. This function y is, of course, a quite different function from that one which is *always* equal to x, whatever value, integral or otherwise, x may have.

10. Let $y = 1$ when x is rational, but $y = 0$ when x is irrational. The graph consists of two series of points arranged upon the lines $y = 1$ and $y = 0$. To the eye it is not distinguishable from two continuous straight lines, but in reality an infinite number of points are missing from each line.

11. Let $y = x$ when x is irrational and $y = \sqrt{(1 + p^2)/(1 + q^2)}$ when x is a rational fraction p/q.

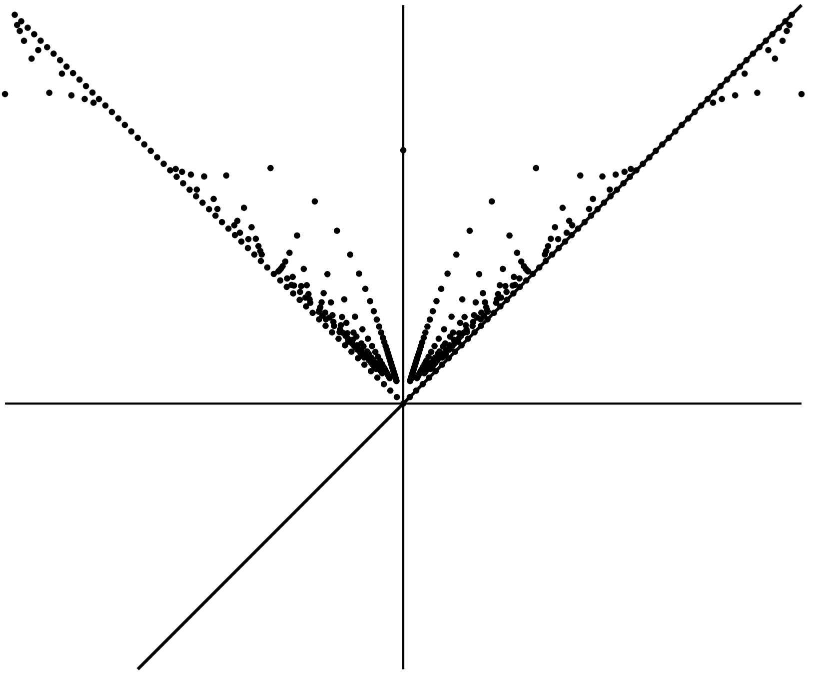

Fig. 16.

The irrational values of x contribute to the graph a curve in reality discontinuous, but apparently not to be distinguished from the straight line $y = x$.

Now consider the rational values of x. First let x be positive. Then $\sqrt{(1+p^2)/(1+q^2)}$ cannot be equal to p/q unless $p = q$, *i.e.* $x = 1$. Thus all the points which correspond to rational values of x lie off the line, except the one point $(1, 1)$. Again, if $p < q$, $\sqrt{(1+p^2)/(1+q^2)} > p/q$; if $p > q$, $\sqrt{(1+p^2)/(1+q^2)} < p/q$. Thus the points lie above the line $y = x$ if $0 < x < 1$, below if $x > 1$. If p and q are large, $\sqrt{(1+p^2)/(1+q^2)}$ is nearly equal to p/q. Near any value of x we can find any number of rational fractions with large numerators and denominators. Hence the graph contains a large number of points which crowd round the line $y = x$. Its general appearance (for positive values of x) is that of a line surrounded by a swarm of isolated points which gets denser and denser as the points approach the line.

The part of the graph which corresponds to negative values of x consists of the rest of the discontinuous line together with the reflections of all these isolated points in the axis of y. Thus to the left of the axis of y the swarm of points is not round $y = x$ but round $y = -x$, which is not itself part of the graph. See Fig. 16.

30. Graphical solution of equations containing a single unknown number. Many equations can be expressed in the form

$$f(x) = \phi(x), \tag{1}$$

where $f(x)$ and $\phi(x)$ are functions whose graphs are easy to draw. And if the curves

$$y = f(x), \quad y = \phi(x)$$

intersect in a point P whose abscissa is ξ, then ξ is a root of the equation (1).

Examples XVII. 1. **The quadratic equation $ax^2 + 2bx + c = 0$.** This may be solved graphically in a variety of ways. For instance we may draw the graphs of

$$y = ax + 2b, \quad y = -c/x,$$

whose intersections, if any, give the roots. Or we may take

$$y = x^2, \quad y = -(2bx + c)/a.$$

But the most elementary method is probably to draw the circle

$$a(x^2 + y^2) + 2bx + c = 0,$$

whose centre is $(-b/a, 0)$ and radius $\{\sqrt{b^2 - ac}\}/a$. The abscissae of its intersections with the axis of x are the roots of the equation.

2. Solve by any of these methods

$$x^2 + 2x - 3 = 0, \quad x^2 - 7x + 4 = 0, \quad 3x^2 + 2x - 2 = 0.$$

3. **The equation $x^m + ax + b = 0$.** This may be solved by constructing the curves $y = x^m$, $y = -ax - b$. Verify the following table for the number of

roots of

$$x^m + ax + b = 0:$$

$$(a)\, m\ \textit{even} \begin{cases} b \text{ positive, } \textit{two or none}, \\ b \text{ negative, } \textit{two}; \end{cases}$$

$$(b)\, m\ \textit{odd} \begin{cases} a \text{ positive, } \textit{one}, \\ a \text{ negative, } \textit{three or one}. \end{cases}$$

Construct numerical examples to illustrate all possible cases.

4. Show that the equation $\tan x = ax + b$ has always an infinite number of roots.

5. Determine the number of roots of

$$\sin x = x, \quad \sin x = \tfrac{1}{3}x, \quad \sin x = \tfrac{1}{8}x, \quad \sin x = \tfrac{1}{120}x.$$

6. Show that if a is small and positive (*e.g.* $a = .01$), the equation

$$x - a = \tfrac{1}{2}\pi \sin^2 x$$

has three roots. Consider also the case in which a is small and negative. Explain how the number of roots varies as a varies.

31. Functions of two variables and their graphical representation. In § 20 we considered two variables connected by a relation. We may similarly consider *three* variables (x, y, and z) connected by a relation such that when the values of x and y are both given, the value or values of z are known. In this case we call z a *function of the two variables* x and y; x and y the *independent* variables, z the *dependent* variable; and we express this dependence of z upon x and y by writing

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

The remarks of § 20 may all be applied, *mutatis mutandis*, to this more complicated case.

The method of representing such functions of two variables graphically is exactly the same in principle as in the case of functions of a single variable. We must take three axes, OX, OY, OZ in space of three dimensions, each axis being perpendicular to the other two. The point (a, b, c) is the point whose distances from the planes YOZ, ZOX, XOY, measured parallel to OX, OY, OZ, are a, b, and c. Regard must of course be paid to sign, lengths measured in the directions OX, OY, OZ being regarded as positive. The definitions of *coordinates*, *axes*, *origin* are the same as before.

Now let

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

As x and y vary, the point (x, y, z) will move in space. The aggregate of all the positions it assumes is called the *locus* of the point (x, y, z) or the *graph* of the function $z = f(x, y)$. When the relation between x, y, and z which defines z can be expressed in an analytical formula, this formula is called the *equation* of the locus. It is easy to show, for example, that the equation

$$Ax + By + Cz + D = 0$$

(*the general equation of the first degree*) represents a *plane*, and that the equation of any plane is of this form. The equation

$$(x - \alpha)^2 + (y - \beta)^2 + (z - \gamma)^2 = \rho^2,$$

or

$$x^2 + y^2 + z^2 + 2Fx + 2Gy + 2Hz + C = 0,$$

where $F^2 + G^2 + H^2 - C > 0$, represents a *sphere*; and so on. For proofs of these propositions we must again refer to text-books of Analytical Geometry.

32. Curves in a plane. We have hitherto used the notation

$$y = f(x) \tag{1}$$

to express functional dependence of y upon x. It is evident that this notation is most appropriate in the case in which y is expressed explicitly in terms of x by means of a formula, as when for example

$$y = x^2, \quad \sin x, \quad a\cos^2 x + b\sin^2 x.$$

We have however very often to deal with functional relations which it is impossible or inconvenient to express in this form. If, for example, $y^5-y-x=0$ or $x^5+y^5-ay=0$, it is known to be impossible to express y explicitly as an algebraical function of x. If

$$x^2+y^2+2Gx+2Fy+C=0,$$

y can indeed be so expressed, viz. by the formula

$$y=-F+\sqrt{F^2-x^2-2Gx-C};$$

but the functional dependence of y upon x is better and more simply expressed by the original equation.

It will be observed that in these two cases the functional relation is fully expressed *by equating a function of the two variables x and y to zero*, *i.e.* by means of an equation

$$f(x,y)=0. \tag{2}$$

We shall adopt this equation as the standard method of expressing the functional relation. It includes the equation (1) as a special case, since $y-f(x)$ is a special form of a function of x and y. We can then speak of the locus of the point (x,y) subject to $f(x,y)=0$, the graph of the function y defined by $f(x,y)=0$, the curve or locus $f(x,y)=0$, and the equation of this curve or locus.

There is another method of representing curves which is often useful. Suppose that x and y are both functions of a third variable t, which is to be regarded as essentially auxiliary and devoid of any particular geometrical significance. We may write

$$x=f(t), \quad y=F(t). \tag{3}$$

If a particular value is assigned to t, the corresponding values of x and of y are known. Each pair of such values defines a point (x,y). If we construct all the points which correspond in this way to different values of t, we obtain *the graph of the locus defined by the equations* (3). Suppose for example

$$x = a\cos t, \quad y = a\sin t.$$

Let t vary from 0 to 2π. Then it is easy to see that the point (x, y) describes the circle whose centre is the origin and whose radius is a. If t varies beyond these limits, (x, y) describes the circle over and over again. We can in this case at once obtain a direct relation between x and y by squaring and adding: we find that $x^2 + y^2 = a^2$, t being now eliminated.

Examples XVIII. 1. The points of intersection of the two curves whose equations are $f(x, y) = 0$, $\phi(x, y) = 0$, where f and ϕ are polynomials, can be determined if these equations can be solved as a pair of simultaneous equations in x and y. The solution generally consists of a finite number of pairs of values of x and y. The two equations therefore generally represent a finite number of isolated points.

2. Trace the curves $(x + y)^2 = 1$, $xy = 1$, $x^2 - y^2 = 1$.

3. The curve $f(x, y) + \lambda\phi(x, y) = 0$ represents a curve passing through the points of intersection of $f = 0$ and $\phi = 0$.

4. What loci are represented by

$$(\alpha)\ x = at + b, \quad y = ct + d, \qquad (\beta)\ x/a = 2t/(1 + t^2), \quad y/a = (1 - t^2)/(1 + t^2),$$

when t varies through all real values?

33. Loci in space. In space of three dimensions there are two fundamentally different kinds of loci, of which the simplest examples are the plane and the straight line.

A particle which moves along a straight line has only *one degree of freedom.* Its direction of motion is fixed; its position can be completely fixed by one measurement of position, *e.g.* by its distance from a fixed point on the line. If we take the line as our fundamental line Λ of Chap. I, the position of any of its points is determined by a single coordinate x. A particle which moves in a plane, on the other hand, has *two* degrees of freedom; its position can only be fixed by the determination of two coordinates.

A locus represented by a single equation

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

plainly belongs to the second of these two classes of loci, and is called a *surface*. It may or may not (in the obvious simple cases it will) satisfy our common-sense notion of what a surface should be.

The considerations of § 31 may evidently be generalised so as to give definitions of a function $f(x, y, z)$ of *three* variables (or of functions of any number of variables). And as in § 32 we agreed to adopt $f(x, y) = 0$ as the standard form of the equation of a plane curve, so now we shall agree to adopt

$$f(x, y, z) = 0$$

as the standard form of equation of a surface.

The locus represented by *two* equations of the form $z = f(x, y)$ or $f(x, y, z) = 0$ belongs to the first class of loci, and is called a *curve*. Thus a *straight line* may be represented by two equations of the type $Ax + By + Cz + D = 0$. A *circle* in space may be regarded as the intersection of a sphere and a plane; it may therefore be represented by two equations of the forms

$$(x - \alpha)^2 + (y - \beta)^2 + (z - \gamma)^2 = \rho^2, \quad Ax + By + Cz + D = 0.$$

Examples XIX. 1. What is represented by *three* equations of the type $f(x, y, z) = 0$?

2. Three linear equations in general represent a single point. What are the exceptional cases?

3. What are the equations of a plane curve $f(x, y) = 0$ in the plane XOY, when regarded as a curve in space? [$f(x, y) = 0$, $z = 0$.]

4. **Cylinders.** What is the meaning of a single equation $f(x, y) = 0$, considered as a locus in space of three dimensions?

[All points on the surface satisfy $f(x, y) = 0$, whatever be the value of z. The curve $f(x, y) = 0$, $z = 0$ is the curve in which the locus cuts the plane XOY. The locus is the surface formed by drawing lines parallel to OZ through all points of this curve. Such a surface is called a *cylinder*.]

5. **Graphical representation of a surface on a plane. Contour Maps.** It might seem to be impossible to represent a surface adequately by a drawing on a plane; and so indeed it is: but a very fair notion of the nature of the surface may often be obtained as follows. Let the equation of the surface be $z = f(x, y)$.

If we give z a particular value a, we have an equation $f(x, y) = a$, which we may regard as determining a plane curve on the paper. We trace this curve and mark it (a). Actually the curve (a) is the projection on the plane XOY of the section of the surface by the plane $z = a$. We do this for all values of a (practically, of course, for a selection of values of a). We obtain some such figure as is shown in Fig. 17. It will at once suggest a contoured Ordnance Survey map: and in fact this is the principle on which such maps are constructed. The contour line 1000 is the projection, on the plane of the sea level, of the section of the surface of the land by the plane parallel to the plane of the sea level and 1000 ft. above it.*

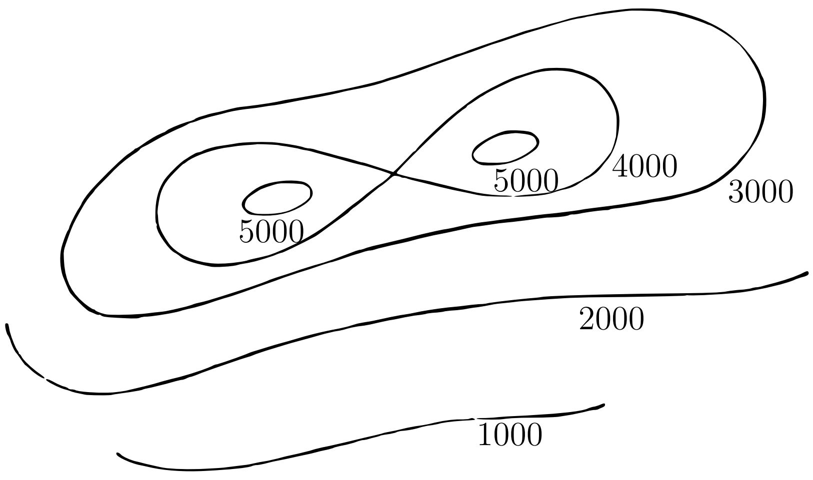

Fig. 17.

6. Draw a series of contour lines to illustrate the form of the surface $2z = 3xy$.

7. **Right circular cones.** Take the origin of coordinates at the vertex of

*We assume that the effects of the earth's curvature may be neglected.

the cone and the axis of z along the axis of the cone; and let α be the semi-vertical angle of the cone. The equation of the cone (which must be regarded as extending both ways from its vertex) is $x^2 + y^2 - z^2 \tan^2 \alpha = 0$.

8. **Surfaces of revolution in general.** The cone of Ex. 7 cuts ZOX in two lines whose equations may be combined in the equation $x^2 = z^2 \tan^2 \alpha$. That is to say, the equation of the surface generated by the revolution of the curve $y = 0$, $x^2 = z^2 \tan^2 \alpha$ round the axis of z is derived from the second of these equations by changing x^2 into $x^2 + y^2$. Show generally that the equation of the surface generated by the revolution of the curve $y = 0$, $x = f(z)$, round the axis of z, is

$$\sqrt{x^2 + y^2} = f(z).$$

9. **Cones in general.** A surface formed by straight lines passing through a fixed point is called a *cone*: the point is called the *vertex*. A particular case is given by the right circular cone of Ex. 7. Show that the equation of a cone whose vertex is O is of the form $f(z/x, z/y) = 0$, and that any equation of this form represents a cone. [If (x, y, z) lies on the cone, so must $(\lambda x, \lambda y, \lambda z)$, for any value of λ.]

10. **Ruled surfaces.** Cylinders and cones are special cases of *surfaces composed of straight lines*. Such surfaces are called *ruled surfaces*.

The two equations

$$x = az + b, \quad y = cz + d, \tag{1}$$

represent the intersection of two planes, *i.e.* a straight line. Now suppose that a, b, c, d instead of being fixed are *functions of an auxiliary variable* t. For any particular value of t the equations (1) give a line. As t varies, this line moves and generates a surface, whose equation may be found by eliminating t between the two equations (1). For instance, in Ex. 7 the equations of the line which generates the cone are

$$x = z \tan \alpha \cos t, \quad y = z \tan \alpha \sin t,$$

where t is the angle between the plane XOZ and a plane through the line and the axis of z.

Another simple example of a ruled surface may be constructed as follows. Take two sections of a right circular cylinder perpendicular to the axis and at a distance l apart (Fig. 18a). We can imagine the surface of the cylinder to be made up of a number of thin parallel rigid rods of length l, such as PQ, the ends of the rods being fastened to two circular rods of radius a.

Now let us take a third circular rod of the same radius and place it round the surface of the cylinder at a distance h from one of the first two rods (see Fig. 18a, where $Pq = h$). Unfasten the end Q of the rod PQ and turn PQ about P until Q can be fastened to the third circular rod in the position Q'. The angle $qOQ' = \alpha$ in the figure is evidently given by

$$l^2 - h^2 = qQ'^2 = \left(2a \sin \tfrac{1}{2}\alpha\right)^2.$$

Let all the other rods of which the cylinder was composed be treated in the same way. We obtain a ruled surface whose form is indicated in Fig. 18b. It is entirely built up of straight lines; but the surface is curved everywhere, and is in general shape not unlike certain forms of table-napkin rings (Fig. 18c).

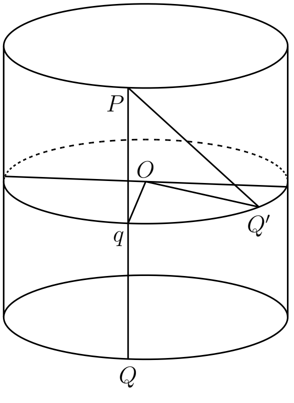

Fig. 18a.

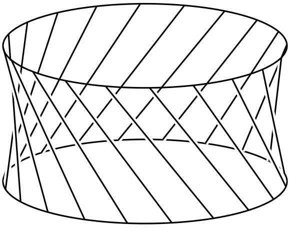

Fig. 18b.

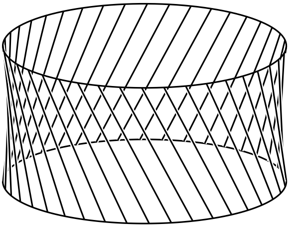

Fig. 18c.

MISCELLANEOUS EXAMPLES ON CHAPTER II.

1. Show that if $y = f(x) = (ax + b)/(cx - a)$ then $x = f(y)$.

2. If $f(x) = f(-x)$ for all values of x, $f(x)$ is called an *even* function. If $f(x) = -f(-x)$, it is called an *odd* function. Show that any function of x, defined for all values of x, is the sum of an even and an odd function of x.

[Use the identity $f(x) = \frac{1}{2}\{f(x) + f(-x)\} + \frac{1}{2}\{f(x) - f(-x)\}$.]

3. Draw the graphs of the functions

$$3\sin x + 4\cos x, \quad \sin\left(\frac{\pi}{\sqrt{2}}\sin x\right).$$

(*Math. Trip.* 1896.)

4. Draw the graphs of the functions

$$\sin x(a\cos^2 x + b\sin^2 x), \quad \frac{\sin x}{x}(a\cos^2 x + b\sin^2 x), \quad \left(\frac{\sin x}{x}\right)^2.$$

5. Draw the graphs of the functions $x[1/x]$, $[x]/x$.

6. Draw the graphs of the functions

$$\text{(i)} \quad \operatorname{arc}\cos(2x^2 - 1) - 2\operatorname{arc}\cos x,$$
$$\text{(ii)} \quad \operatorname{arc}\tan\frac{a+x}{1-ax} - \operatorname{arc}\tan a - \operatorname{arc}\tan x,$$

where the symbols $\operatorname{arc}\cos a$, $\operatorname{arc}\tan a$ denote, for any value of a, the least positive (or zero) angle, whose cosine or tangent is a.

7. Verify the following method of constructing the graph of $f\{\phi(x)\}$ by means of the line $y = x$ and the graphs of $f(x)$ and $\phi(x)$: take $OA = x$ along OX, draw AB parallel to OY to meet $y = \phi(x)$ in B, BC parallel to OX to meet $y = x$ in C, CD parallel to OY to meet $y = f(x)$ in D, and DP parallel to OX to meet AB in P; then P is a point on the graph required.

8. Show that the roots of $x^3 + px + q = 0$ are the abscissae of the points of intersection (other than the origin) of the parabola $y = x^2$ and the circle

$$x^2 + y^2 + (p-1)y + qx = 0.$$

9. The roots of $x^4 + nx^3 + px^2 + qx + r = 0$ are the abscissae of the points of intersection of the parabola $x^2 = y - \frac{1}{2}nx$ and the circle

$$x^2 + y^2 + (\tfrac{1}{8}n^2 - \tfrac{1}{2}pn + \tfrac{1}{2}n + q)x + (p - 1 - \tfrac{1}{4}n^2)y + r = 0.$$

10. Discuss the graphical solution of the equation

$$x^m + ax^2 + bx + c = 0$$

by means of the curves $y = x^m$, $y = -ax^2 - bx - c$. Draw up a table of the various possible numbers of roots.

11. Solve the equation $\sec\theta + \operatorname{cosec}\theta = 2\sqrt{2}$; and show that the equation $\sec\theta + \operatorname{cosec}\theta = c$ has two roots between 0 and 2π if $c^2 < 8$ and four if $c^2 > 8$.

12. Show that the equation

$$2x = (2n+1)\pi(1-\cos x),$$

where n is a positive integer, has $2n+3$ roots and no more, indicating their localities roughly. (*Math. Trip.* 1896.)

13. Show that the equation $\frac{2}{3}x\sin x = 1$ has four roots between $-\pi$ and π.

14. Discuss the number and values of the roots of the equations

(1) $\cot x + x - \frac{3}{2}\pi = 0$,

(2) $x^2 + \sin^2 x = 1$,

(3) $\tan x = 2x/(1+x^2)$,

(4) $\sin x - x + \frac{1}{6}x^3 = 0$,

(5) $(1-\cos x)\tan\alpha - x + \sin x = 0$.

15. The polynomial of the second degree which assumes, when $x = a$, b, c the values α, β, γ is

$$\alpha\frac{(x-b)(x-c)}{(a-b)(a-c)} + \beta\frac{(x-c)(x-a)}{(b-c)(b-a)} + \gamma\frac{(x-a)(x-b)}{(c-a)(c-b)}.$$

Give a similar formula for the polynomial of the $(n-1)$th degree which assumes, when $x = a_1, a_2, \dots a_n$, the values $\alpha_1, \alpha_2, \dots \alpha_n$.

16. Find a polynomial in x of the second degree which for the values 0, 1, 2 of x takes the values $1/c$, $1/(c+1)$, $1/(c+2)$; and show that when $x = c+2$ its value is $1/(c+1)$. (*Math. Trip.* 1911.)

17. Show that if x is a rational function of y, and y is a rational function of x, then $Axy + Bx + Cy + D = 0$.

18. If y is an algebraical function of x, then x is an algebraical function of y.

19. Verify that the equation

$$\cos\tfrac{1}{2}\pi x = 1 - \frac{x^2}{x + (x-1)\sqrt{\dfrac{2-x}{3}}}$$

is approximately true for all values of x between 0 and 1. [Take $x = 0$, $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, $\frac{5}{6}$, 1, and use tables. For which of these values is the formula exact?]

20. What is the form of the graph of the functions

$$z = [x] + [y], \quad z = x + y - [x] - [y]?$$

21. What is the form of the graph of the functions $z = \sin x + \sin y$, $z = \sin x \sin y$, $z = \sin xy$, $z = \sin(x^2 + y^2)$?

22. **Geometrical constructions for irrational numbers.** In Chapter I we indicated one or two simple geometrical constructions for a length equal to $\sqrt{2}$, starting from a given unit length. We also showed how to construct the roots of any quadratic equation $ax^2 + 2bx + c = 0$, it being supposed that we can construct lines whose lengths are equal to any of the ratios of the coefficients a, b, c, as is certainly the case if a, b, c are rational. All these constructions were what may be called Euclidean constructions; they depended on the ruler and compasses only.

It is fairly obvious that we can construct by these methods the length measured by any irrational number which is defined by any combination of square roots, however complicated. Thus

$$\sqrt[4]{\sqrt{\frac{17+3\sqrt{11}}{17-3\sqrt{11}}} - \sqrt{\frac{17-3\sqrt{11}}{17+3\sqrt{11}}}}$$

is a case in point. This expression contains a fourth root, but this is of course the square root of a square root. We should begin by constructing $\sqrt{11}$, *e.g.* as the mean between 1 and 11: then $17+3\sqrt{11}$ and $17-3\sqrt{11}$, and so on. Or these two mixed surds might be constructed directly as the roots of $x^2-34x+190 = 0$.

Conversely, *only* irrationals of this kind can be constructed by Euclidean methods. Starting from a unit length we can construct any *rational* length. And hence we can construct the line $Ax + By + C = 0$, provided that the ratios of A, B, C are rational, and the circle

$$(x - \alpha)^2 + (y - \beta)^2 = \rho^2$$

(or $x^2 + y^2 + 2gx + 2fy + c = 0$), provided that α, β, ρ are rational, a condition which implies that g, f, c are rational.

Now in any Euclidean construction each new point introduced into the figure is determined as the intersection of two lines or circles, or a line and a circle. But if the coefficients are rational, such a pair of equations as

$$Ax + By + C = 0, \quad x^2 + y^2 + 2gx + 2fy + c = 0$$

give, on solution, values of x and y of the form $m + n\sqrt{p}$, where m, n, p are rational: for if we substitute for x in terms of y in the second equation we obtain a quadratic in y with rational coefficients. Hence the coordinates of all points obtained by means of lines and circles with rational coefficients are expressible by rational numbers and quadratic surds. And so the same is true of the distance $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ between any two points so obtained.

With the irrational distances thus constructed we may proceed to construct a number of lines and circles whose coefficients may now themselves involve quadratic surds. It is evident, however, that all the lengths which we can construct by the use of such lines and circles are still expressible by square roots only, though our surd expressions may now be of a more complicated form. And this remains true however often our constructions are repeated. Hence *Euclidean methods will construct any surd expression involving square roots only, and no others.*

One of the famous problems of antiquity was that of the duplication of the cube, that is to say of the construction by Euclidean methods of a length measured by $\sqrt[3]{2}$. It can be shown that $\sqrt[3]{2}$ cannot be expressed by means of any finite combination of rational numbers and square roots, and so that the problem is an impossible one. See Hobson, *Squaring the Circle*, pp. 47 *et seq.*; the first stage of the proof, viz. the proof that $\sqrt[3]{2}$ cannot be a root of a quadratic equation $ax^2 + 2bx + c = 0$ with rational coefficients, was given in Ch. I (Misc. Exs. 24).

23. **Approximate quadrature of the circle.** Let O be the centre of a circle of radius R. On the tangent at A take $AP = \frac{11}{5}R$ and $AQ = \frac{13}{5}R$, in the same direction. On AO take $AN = OP$ and draw NM parallel to OQ and cutting AP in M. Show that

$$AM/R = \tfrac{13}{25}\sqrt{146},$$

and that to take AM as being equal to the circumference of the circle would lead to a value of π correct to five places of decimals. If R is the earth's radius, the error in supposing AM to be its circumference is less than 11 yards.

24. Show that the only lengths which can be constructed with the ruler only, starting from a given unit length, are rational lengths.

25. **Constructions for $\sqrt[3]{2}$.** O is the vertex and S the focus of the parabola $y^2 = 4x$, and P is one of its points of intersection with the parabola $x^2 = 2y$. Show that OP meets the latus rectum of the first parabola in a point Q such that $SQ = \sqrt[3]{2}$.

26. Take a circle of unit diameter, a diameter OA and the tangent at A. Draw a chord OBC cutting the circle at B and the tangent at C. On this line take $OM = BC$. Taking O as origin and OA as axis of x, show that the locus of M is the curve

$$(x^2 + y^2)x - y^2 = 0$$

(the *Cissoid of Diocles*). Sketch the curve. Take along the axis of y a length $OD = 2$. Let AD cut the curve in P and OP cut the tangent to the circle at A in Q. Show that $AQ = \sqrt[3]{2}$.

CHAPTER III

COMPLEX NUMBERS

34. Displacements along a line and in a plane. The 'real number' x, with which we have been concerned in the two preceding chapters, may be regarded from many different points of view. It may be regarded as a pure number, destitute of geometrical significance, or a geometrical significance may be attached to it in at least three different ways. It may be regarded as *the measure of a length*, viz. the length A_0P along the line Λ of Chap. I. It may be regarded as *the mark of a point*, viz. the point P whose distance from A_0 is x. Or it may be regarded as *the measure of a displacement* or *change of position* on the line Λ. It is on this last point of view that we shall now concentrate our attention.

Imagine a small particle placed at P on the line Λ and then displaced to Q. We shall call the displacement or change of position which is needed to transfer the particle from P to Q *the displacement* $\overline{PQ}$. To specify a displacement completely three things are needed, its *magnitude*, its *sense* forwards or backwards along the line, and what may be called its *point of application*, *i.e.* the original position P of the particle. But, when we are thinking merely of the change of position produced by the displacement, it is natural to disregard the point of application and to consider all displacements as equivalent whose lengths and senses are the same. Then the displacement is completely specified by the length $PQ = x$, the sense of the displacement being fixed by the sign of x. We may therefore, without ambiguity, speak of *the displacement* $[x]$,* and we may write $\overline{PQ} = [x]$.

We use the square bracket to distinguish the displacement $[x]$ from the length or number x.† If the coordinate of P is a, that of Q will be $a + x$;

---

*It is hardly necessary to caution the reader against confusing this use of the symbol $[x]$ and that of Chap. II (Exs. XVI. and Misc. Exs.).

†Strictly speaking we ought, by some similar difference of notation, to distinguish the actual length x from the number x which measures it. The reader will perhaps be inclined to consider such distinctions futile and pedantic. But increasing experience of mathematics will reveal to him the great importance of distinguishing clearly between things which, however intimately connected, are not the same. If cricket were a math-

the displacement $[x]$ therefore transfers a particle from the point a to the point $a + x$.

We come now to consider *displacements in a plane*. We may define the displacement $\overline{PQ}$ as before. But now more data are required in order to specify it completely. We require to know: (i) the *magnitude* of the displacement, *i.e.* the length of the straight line PQ; (ii) the *direction* of the displacement, which is determined by the angle which PQ makes with some fixed line in the plane; (iii) the *sense* of the displacement; and (iv) its *point of application*. Of these requirements we may disregard the fourth, if we consider two displacements as equivalent if they are the same

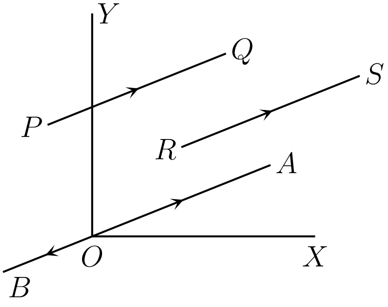

Fig. 19.

in magnitude, direction, and sense. In other words, if PQ and RS are equal and parallel, and the sense of motion from P to Q is the same as that of motion from R to S, we regard the displacements $\overline{PQ}$ and $\overline{RS}$ as equivalent, and write

$$\overline{PQ} = \overline{RS}.$$

Now let us take any pair of coordinate axes in the plane (such as OX, OY in Fig. 19). Draw a line OA equal and parallel to PQ, the sense of motion from O to A being the same as that from P to Q. Then $\overline{PQ}$ and $\overline{OA}$ are equivalent displacements. Let x and y be the coordinates

ematical science, it would be very important to distinguish between the *motion* of the batsman between the wickets, the *run* which he scores, and the *mark* which is put down in the score-book.

of A. Then it is evident that $\overline{OA}$ is completely specified if x and y are given. We call $\overline{OA}$ *the displacement* $[x, y]$ and write

$$\overline{OA} = \overline{PQ} = \overline{RS} = [x, y].$$

35. Equivalence of displacements. Multiplication of displacements by numbers. If ξ and η are the coordinates of P, and ξ' and η' those of Q, it is evident that

$$x = \xi' - \xi, \quad y = \eta' - \eta.$$

The displacement from (ξ, η) to (ξ', η') is therefore

$$[\xi' - \xi, \eta' - \eta].$$

It is clear that two displacements $[x, y]$, $[x', y']$ are equivalent if, and only if, $x = x'$, $y = y'$. Thus $[x, y] = [x', y']$ if and only if

$$x = x', \quad y = y'. \tag{1}$$

The reverse displacement $\overline{QP}$ would be $[\xi - \xi', \eta - \eta']$, and it is natural to agree that

$$[\xi - \xi', \eta - \eta'] = -[\xi' - \xi, \eta' - \eta],$$
$$\overline{QP} = -\overline{PQ},$$

these equations being really definitions of the meaning of the symbols $-[\xi' - \xi, \eta' - \eta]$, $-\overline{PQ}$. Having thus agreed that

$$-[x, y] = [-x, -y],$$

it is natural to agree further that

$$\alpha[x, y] = [\alpha x, \alpha y], \tag{2}$$

where α is any real number, positive or negative. Thus (Fig. 19) if $OB = -\frac{1}{2}OA$ then

$$\overline{OB} = -\tfrac{1}{2}\overline{OA} = -\tfrac{1}{2}[x, y] = [-\tfrac{1}{2}x, -\tfrac{1}{2}y].$$

The equations (1) and (2) define the first two important ideas connected with displacements, viz. *equivalence* of displacements, and *multiplication of displacements by numbers.*

36. Addition of displacements. We have not yet given any definition which enables us to attach any meaning to the expressions

$$\overline{PQ} + \overline{P'Q'}, \quad [x, y] + [x', y'].$$

Common sense at once suggests that we should define the sum of two displacements as the displacement which is the result of the successive application of the two given displacements. In other words, it suggests that if QQ_1 be drawn equal and parallel to $P'Q'$, so that the result of successive displacements $\overline{PQ}$, $\overline{P'Q'}$ on a particle at P is to transfer it first to Q and then to Q_1 then we should define the sum of $\overline{PQ}$ and $\overline{P'Q'}$ as being $\overline{PQ_1}$. If then we draw OA equal and parallel to PQ, and OB equal and parallel to $P'Q'$, and complete the parallelogram $OACB$, we have

$$\overline{PQ} + \overline{P'Q'} = \overline{PQ_1} = \overline{OA} + \overline{OB} = \overline{OC}.$$

Let us consider the consequences of adopting this definition. If the coordinates of B are x', y', then those of the middle point of AB are $\frac{1}{2}(x + x')$, $\frac{1}{2}(y + y')$, and those of C are $x + x'$, $y + y'$. Hence

$$[x, y] + [x', y'] = [x + x', y + y'], \tag{3}$$

which may be regarded as the symbolic definition of addition of displacements. We observe that

$$\begin{aligned}[x', y'] + [x, y] &= [x' + x, y' + y] \\ &= [x + x', y + y'] = [x, y] + [x', y']\end{aligned}$$

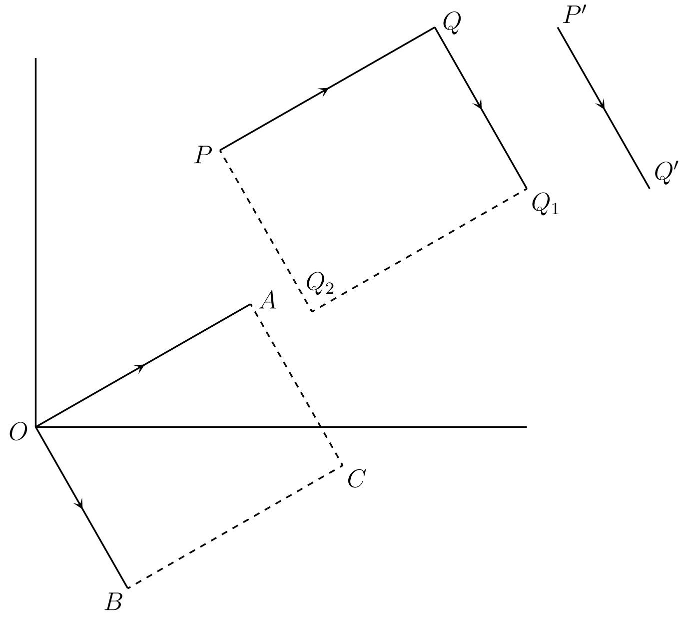

Fig. 20.

In other words, *addition of displacements obeys the commutative law* expressed in ordinary algebra by the equation $a+b = b+a$. This law expresses the obvious geometrical fact that if we move from P first through a distance PQ_2 equal and parallel to $P'Q'$, and then through a distance equal and parallel to PQ, we shall arrive at the same point Q_1 as before.

In particular

$$[x, y] = [x, 0] + [0, y]. \tag{4}$$

Here $[x, 0]$ denotes a displacement through a distance x in a direction parallel to OX. It is in fact what we previously denoted by $[x]$, when we were considering only displacements along a line. We call $[x, 0]$ and $[0, y]$ the *components* of $[x, y]$, and $[x, y]$ their *resultant*.

When we have once defined addition of two displacements, there is no further difficulty in the way of defining addition of any number. Thus, by definition,

$$[x, y] + [x', y'] + [x'', y''] = ([x, y] + [x', y']) + [x'', y'']$$
$$= [x + x', y + y'] + [x'', y''] = [x + x' + x'', y + y' + y''].$$

We define *subtraction* of displacements by the equation

$$[x, y] - [x', y'] = [x, y] + (-[x', y']), \tag{5}$$

which is the same thing as $[x, y] + [-x', -y']$ or as $[x - x', y - y']$. In particular

$$[x, y] - [x, y] = [0, 0].$$

The displacement $[0, 0]$ leaves the particle where it was; it is the *zero displacement*, and we agree to write $[0, 0] = 0$.

Examples XX. 1. Prove that

(i) $\alpha[\beta x, \beta y] = \beta[\alpha x, \alpha y] = [\alpha\beta x, \alpha\beta y]$,

(ii) $([x, y] + [x', y']) + [x'', y''] = [x, y] + ([x', y'] + [x'', y''])$,

(iii) $[x, y] + [x', y'] = [x', y'] + [x, y]$,

(iv) $(\alpha + \beta)[x, y] = \alpha[x, y] + \beta[x, y]$,

(v) $\alpha\{[x, y] + [x', y']\} = \alpha[x, y] + \alpha[x', y']$.

[We have already proved (iii). The remaining equations follow with equal ease from the definitions. The reader should in each case consider the geometrical significance of the equation, as we did above in the case of (iii).]

2. If M is the middle point of PQ, then $\overline{OM} = \frac{1}{2}(\overline{OP} + \overline{OQ})$. More generally, if M divides PQ in the ratio $\mu : \lambda$, then

$$\overline{OM} = \frac{\lambda}{\lambda + \mu}\,\overline{OP} + \frac{\mu}{\lambda + \mu}\,\overline{OQ}.$$

3. If G is the centre of mass of equal particles at P_1, P_2, ..., P_n, then

$$\overline{OG} = (\overline{OP_1} + \overline{OP_2} + \cdots + \overline{OP_n})/n.$$

4. If P, Q, R are collinear points in the plane, then it is possible to find real numbers α, β, γ, not all zero, and such that

$$\alpha \cdot \overline{OP} + \beta \cdot \overline{OQ} + \gamma \cdot \overline{OR} = 0;$$

and conversely. [This is really only another way of stating Ex. 2.]

5. If $\overline{AB}$ and $\overline{AC}$ are two displacements not in the same straight line, and

$$\alpha \cdot \overline{AB} + \beta \cdot \overline{AC} = \gamma \cdot \overline{AB} + \delta \cdot \overline{AC},$$

then $\alpha = \gamma$ and $\beta = \delta$.

[Take $\overline{AB_1} = \alpha \cdot \overline{AB}$, $\overline{AC_1} = \beta \cdot \overline{AC}$. Complete the parallelogram $AB_1P_1C_1$. Then $\overline{AP_1} = \alpha \cdot \overline{AB} + \beta \cdot \overline{AC}$. It is evident that $\overline{AP_1}$ can only be expressed in this form in one way, whence the theorem follows.]

6. $ABCD$ is a parallelogram. Through Q, a point inside the parallelogram, RQS and TQU are drawn parallel to the sides. Show that RU, TS intersect on AC.

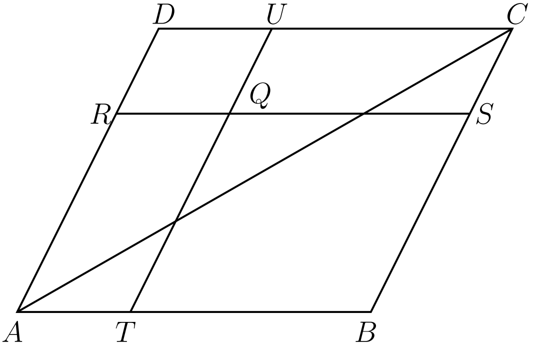

Fig. 21.

[Let the ratios $AT : AB$, $AR : AD$ be denoted by α, β. Then

$$\overline{AT} = \alpha \cdot \overline{AB}, \quad \overline{AR} = \beta \cdot \overline{AD},$$
$$\overline{AU} = \alpha \cdot \overline{AB} + \overline{AD}, \quad \overline{AS} = \overline{AB} + \beta \cdot \overline{AD}.$$

Let RU meet AC in P. Then, since R, U, P are collinear,

$$\overline{AP} = \frac{\lambda}{\lambda + \mu}\,\overline{AR} + \frac{\mu}{\lambda + \mu}\,\overline{AU},$$

where μ/λ is the ratio in which P divides RU. That is to say

$$\overline{AP} = \frac{\alpha\mu}{\lambda+\mu}\,\overline{AB} + \frac{\beta\lambda+\mu}{\lambda+\mu}\,\overline{AD}.$$

But since P lies on AC, $\overline{AP}$ is a numerical multiple of $\overline{AC}$; say

$$\overline{AP} = k\cdot\overline{AC} = k\cdot\overline{AB} + k\cdot\overline{AD}.$$

Hence (Ex. 5) $\alpha\mu = \beta\lambda + \mu = (\lambda+\mu)k$, from which we deduce

$$k = \frac{\alpha\beta}{\alpha+\beta-1}.$$

The symmetry of this result shows that a similar argument would also give

$$\overline{AP'} = \frac{\alpha\beta}{\alpha+\beta-1}\,\overline{AC},$$

if P' is the point where TS meets AC. Hence P and P' are the same point.]

7. $ABCD$ is a parallelogram, and M the middle point of AB. Show that DM trisects and is trisected by AC.*

37. Multiplication of displacements. So far we have made no attempt to attach any meaning whatever to the notion of the *product* of two displacements. The only kind of multiplication which we have considered is that in which a displacement is multiplied by a number. The expression

$$[x, y] \times [x', y']$$

so far means nothing, and we are at liberty to define it to mean anything we like. It is, however, fairly clear that if any definition of such a product is to be of any use, the product of two displacements must itself be a displacement.

We might, for example, define it as being equal to

$$[x + x', y + y'];$$

*The two preceding examples are taken from Willard Gibbs' *Vector Analysis*.

in other words, we might agree that the product of two displacements was to be always equal to their sum. But there would be two serious objections to such a definition. In the first place our definition would be futile. We should only be introducing a new method of expressing something which we can perfectly well express without it. In the second place our definition would be inconvenient and misleading for the following reasons. If α is a real number, we have already defined $\alpha[x, y]$ as $[\alpha x, \alpha y]$. Now, as we saw in § 34, the real number α may itself from one point of view be regarded as a displacement, viz. the displacement $[\alpha]$ along the axis OX, or, in our later notation, the displacement $[\alpha, 0]$. It is therefore, if not absolutely necessary, at any rate most desirable, that our definition should be such that

$$[\alpha, 0][x, y] = [\alpha x, \alpha y],$$

and the suggested definition does not give this result.

A more reasonable definition might appear to be

$$[x, y][x', y'] = [xx', yy'].$$

But this would give

$$[\alpha, 0][x, y] = [\alpha x, 0];$$

and so this definition also would be open to the second objection.

In fact, it is by no means obvious what is the best meaning to attach to the product $[x, y][x', y']$. All that is clear is (1) that, if our definition is to be of any use, this product must itself be a displacement whose coordinates depend on x and y, or in other words that we must have

$$[x, y][x', y'] = [X, Y],$$

where X and Y are functions of x, y, x', and y'; (2) that the definition must be such as to agree with the equation

$$[x, 0][x', y'] = [xx', xy'];$$

and (3) that the definition must obey the ordinary commutative, distributive, and associative laws of multiplication, so that

$$
\begin{aligned}
[x, y][x', y'] &= [x', y'][x, y], \\
([x, y] + [x', y'])[x'', y''] &= [x, y][x'', y''] + [x', y'][x'', y''], \\
[x, y]([x', y'] + [x'', y'']) &= [x, y][x', y'] + [x, y][x'', y''],
\end{aligned}
$$

and

$$[x, y]([x', y'][x'', y'']) = ([x, y][x', y'])[x'', y''].$$

38. The right definition to take is suggested as follows. We know that, if OAB, OCD are two similar triangles, the angles corresponding in the order in which they are written, then

$$OB/OA = OD/OC,$$

or $OB \cdot OC = OA \cdot OD$. This suggests that we should try to define multiplication and division of displacements in such a way that

$$\overline{OB}/\overline{OA} = \overline{OD}/\overline{OC}, \quad \overline{OB} \cdot \overline{OC} = \overline{OA} \cdot \overline{OD}.$$

Now let

$$\overline{OB} = [x, y], \quad \overline{OC} = [x', y'], \quad \overline{OD} = [X, Y],$$

and suppose that A is the point $(1, 0)$, so that $\overline{OA} = [1, 0]$. Then

$$\overline{OA} \cdot \overline{OD} = [1, 0][X, Y] = [X, Y],$$

and so

$$[x, y][x', y'] = [X, Y].$$

The product $\overline{OB} \cdot \overline{OC}$ is therefore to be defined as $\overline{OD}$, D being obtained by constructing on OC a triangle similar to OAB. In order to free this

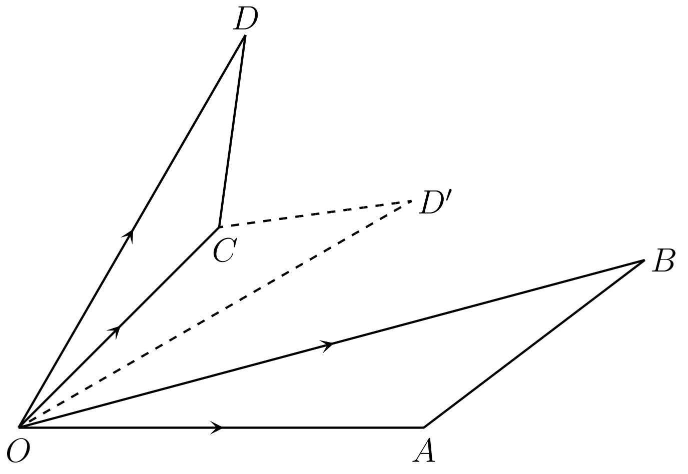

Fig. 22.

definition from ambiguity, it should be observed that on OC we can describe *two* such triangles, OCD and OCD'. We choose that for which the angle COD is equal to AOB in sign as well as in magnitude. We say that the two triangles are then *similar in the same sense*.

If the polar coordinates of B and C are (ρ, θ) and (σ, ϕ), so that

$$x = \rho\cos\theta, \quad y = \rho\sin\theta, \quad x' = \sigma\cos\phi, \quad y' = \sigma\sin\phi,$$

then the polar coordinates of D are evidently $\rho\sigma$ and $\theta + \phi$. Hence

$$X = \rho\sigma\cos(\theta + \phi) = xx' - yy',$$
$$Y = \rho\sigma\sin(\theta + \phi) = xy' + yx'.$$

The required definition is therefore

$$[x, y][x', y'] = [xx' - yy', xy' + yx']. \tag{6}$$

We observe (1) that if $y = 0$, then $X = xx'$, $Y = xy'$, as we desired; (2) that the right-hand side is not altered if we interchange x and x', and y and y', so that

$$[x, y][x', y'] = [x', y'][x, y];$$

and (3) that

$$\begin{aligned}\{[x,y]+[x',y']\}[x'',y''] &= [x+x',y+y'][x'',y'']\\ &= [(x+x')x''-(y+y')y'',(x+x')y''+(y+y')x'']\\ &= [xx''-yy'',xy''+yx'']+[x'x''-y'y'',x'y''+y'x'']\\ &= [x,y][x'',y'']+[x',y'][x'',y''].\end{aligned}$$

Similarly we can verify that all the equations at the end of § 37 are satisfied. Thus the definition (6) fulfils all the requirements which we made of it in § 37.

Example. Show directly from the geometrical definition given above that multiplication of displacements obeys the commutative and distributive laws. [Take the commutative law for example. The product $\overline{OB}\cdot\overline{OC}$ is $\overline{OD}$ (Fig. 22), COD being similar to AOB. To construct the product $\overline{OC}\cdot\overline{OB}$ we should have to construct on OB a triangle BOD_1 similar to AOC; and so what we want to prove is that D and D_1 coincide, or that BOD is similar to AOC. This is an easy piece of elementary geometry.]

39. Complex numbers. Just as to a displacement $[x]$ along OX correspond a point (x) and a real number x, so to a displacement $[x, y]$ in the plane correspond a point (x, y) and a *pair of real numbers x, y.*

We shall find it convenient to denote this pair of real numbers x, y by the symbol

$$x + yi.$$

The reason for the choice of this notation will appear later. For the present the reader must regard $x + yi$ as *simply another way of writing* $[x, y]$. The expression $x + yi$ is called a *complex number.*

We proceed next to define *equivalence*, *addition*, and *multiplication* of complex numbers. To every complex number corresponds a displacement. Two complex numbers are equivalent if the corresponding displacements are equivalent. The sum or product of two complex numbers is the complex number which corresponds to the sum or product of the two corresponding displacements. Thus

$$x + yi = x' + y'i, \tag{1}$$

if and only if $x = x'$, $y = y'$;

$$(x + yi) + (x' + y'i) = (x + x') + (y + y')i; \tag{2}$$

$$(x + yi)(x' + y'i) = xx' - yy' + (xy' + yx')i. \tag{3}$$

In particular we have, as special cases of (2) and (3),

$$x + yi = (x + 0i) + (0 + yi),$$
$$(x + 0i)(x' + y'i) = xx' + xy'i;$$

and these equations suggest that there will be no danger of confusion if, when dealing with complex numbers, we write x for $x+0i$ and yi for $0+yi$, as we shall henceforth.

Positive integral powers and polynomials of complex numbers are then defined as in ordinary algebra. Thus, by putting $x = x'$, $y = y'$ in (3), we obtain

$$(x + yi)^2 = (x + yi)(x + yi) = x^2 - y^2 + 2xyi.$$

The reader will easily verify for himself that addition and multiplication of complex numbers obey the laws of algebra expressed by the equations

$$(x + yi) + (x' + y'i) = (x' + y'i) + (x + yi),$$
$$\{(x + yi) + (x' + y'i)\} + (x'' + y''i) = (x + yi) + \{(x' + y'i) + (x'' + y''i)\},$$
$$(x + yi)(x' + y'i) = (x' + y'i)(x + yi),$$
$$(x + yi)\{(x' + y'i) + (x'' + y''i)\} = (x + yi)(x' + y'i) + (x + yi)(x'' + y''i),$$
$$\{(x + yi) + (x' + y'i)\}(x'' + y''i) = (x + yi)(x'' + y''i) + (x' + y'i)(x'' + y''i),$$
$$(x + yi)\{(x' + y'i)(x'' + y''i)\} = \{(x + yi)(x' + y'i)\}(x'' + y''i),$$

the proofs of these equations being practically the same as those of the corresponding equations for the corresponding displacements.

Subtraction and division of complex numbers are defined as in ordinary algebra. Thus we may define $(x+yi)-(x'+y'i)$ as

$$(x+yi)+\{-(x'+y'i)\}=x+yi+(-x'-y'i)=(x-x')+(y-y')i;$$

or again, as the number $\xi+\eta i$ such that

$$(x'+y'i)+(\xi+\eta i)=x+yi,$$

which leads to the same result. And $(x+yi)/(x'+y'i)$ is defined as being the complex number $\xi+\eta i$ such that

$$(x'+y'i)(\xi+\eta i)=x+yi,$$

or

$$x'\xi-y'\eta+(x'\eta+y'\xi)i=x+yi,$$

or

$$x'\xi-y'\eta=x,\quad x'\eta+y'\xi=y. \tag{4}$$

Solving these equations for ξ and η, we obtain

$$\xi=\frac{xx'+yy'}{x'^2+y'^2},\quad \eta=\frac{yx'-xy'}{x'^2+y'^2}.$$

This solution fails if x' and y' are both zero, *i.e.* if $x'+y'i=0$. Thus subtraction is always possible; division is always possible unless the divisor is zero.

Examples. (1) From a geometrical point of view, the problem of the division of the displacement $\overline{OB}$ by $\overline{OC}$ is that of finding D so that the triangles COB, AOD are similar, and this is evidently possible (and the solution unique) unless C coincides with 0, or $\overline{OC}=0$.

(2) The numbers $x+yi$, $x-yi$ are said to be *conjugate*. Verify that

$$(x+yi)(x-yi)=x^2+y^2,$$

so that the product of two conjugate numbers is real, and that

$$\frac{x+yi}{x'+y'i}=\frac{(x+yi)(x'-y'i)}{(x'+y'i)(x'-y'i)}=\frac{xx'+yy'+(x'y-xy')i}{x'^2+y'^2}.$$

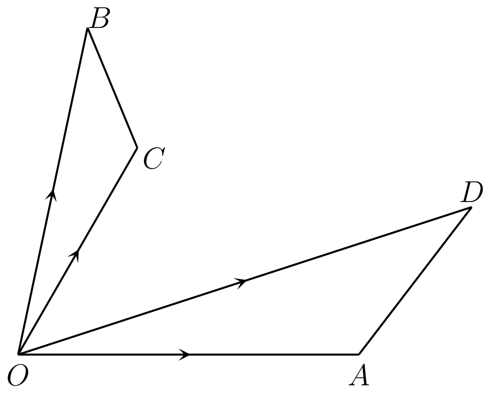

Fig. 23.

40. One most important property of real numbers is that known as *the factor theorem*, which asserts that *the product of two numbers cannot be zero unless one of the two is itself zero*. To prove that this is also true of complex numbers we put $x = 0$, $y = 0$ in the equations (4) of the preceding section. Then

$$x'\xi - y'\eta = 0, \quad x'\eta + y'\xi = 0.$$

These equations give $\xi = 0$, $\eta = 0$, *i.e.*

$$\xi + \eta i = 0,$$

unless $x' = 0$ and $y' = 0$, or $x' + y'i = 0$. Thus $x + yi$ cannot vanish unless either $x' + y'i$ or $\xi + \eta i$ vanishes.

41. The equation $i^2 = -1$. We agreed to simplify our notation by writing x instead of $x + 0i$ and yi instead of $0 + yi$. The particular complex number $1i$ we shall denote simply by i. It is the number which corresponds to a unit displacement along OY. Also

$$i^2 = ii = (0 + 1i)(0 + 1i) = (0 \cdot 0 - 1 \cdot 1) + (0 \cdot 1 + 1 \cdot 0)i = -1.$$

Similarly $(-i)^2 = -1$. Thus the complex numbers i and $-i$ satisfy the equation $x^2 = -1$.

The reader will now easily satisfy himself that the upshot of the rules for addition and multiplication of complex numbers is this, that *we operate with complex numbers in exactly the same way as with real numbers, treating the symbol i as itself a number, but replacing the product $ii = i^2$ by -1 whenever it occurs.* Thus, for example,

$$\begin{aligned}(x + yi)(x' + y'i) &= xx' + xy'i + yx'i + yy'i^2 \\ &= (xx' - yy') + (xy' + yx')i.\end{aligned}$$

42. The geometrical interpretation of multiplication by i. Since

$$(x + yi)i = -y + xi,$$

it follows that if $x + yi$ corresponds to $\overline{OP}$, and OQ is drawn equal to OP and so that POQ is a positive right angle, then $(x + yi)i$ corresponds to $\overline{OQ}$. In other words, *multiplication of a complex number by i turns the corresponding displacement through a right angle.*

We might have developed the whole theory of complex numbers from this point of view. Starting with the ideas of x as representing a displacement along OX, and of i as a symbol of operation equivalent to turning x through a right angle, we should have been led to regard yi as a displacement of magnitude y along OY. It would then have been natural to define $x + yi$ as in §§ 37 and 40, and $(x + yi)i$ would have represented the displacement obtained by turning $x + yi$ through a right angle, *i.e.* $-y + xi$. Finally, we should naturally have defined $(x+yi)x'$ as $xx'+yx'i$, $(x+yi)y'i$ as $-yy' + xy'i$, and $(x + yi)(x' + y'i)$ as the sum of these displacements, *i.e.* as

$$xx' - yy' + (xy' + yx')i.$$

43. The equations $z^2 + 1 = 0$, $az^2 + 2bz + c = 0$. There is no real number z such that $z^2 + 1 = 0$; this is expressed by saying that the equation has *no real roots*. But, as we have just seen, the two complex numbers i and $-i$ satisfy this equation. We express this by saying that the equation has *the two complex roots* i and $-i$. Since i satisfies $z^2 = -1$, it is sometimes written in the form $\sqrt{-1}$.

Complex numbers are sometimes called *imaginary*.* The expression is by no means a happily chosen one, but it is firmly established and has to be accepted. It cannot, however, be too strongly impressed upon the reader that an 'imaginary number' is no more 'imaginary', in any ordinary sense of the word, than a 'real' number; and that it is not a number at all, in the sense in which the 'real' numbers are numbers, but, as should be clear from the preceding discussion, *a pair of numbers* (x, y), united symbolically, for purposes of technical convenience, in the form $x + yi$. Such a pair of numbers is no less 'real' than any ordinary number such as $\frac{1}{2}$, or than the paper on which this is printed, or than the Solar System. Thus

$$i = 0 + 1i$$

stands for the pair of numbers $(0, 1)$, and may be represented geometrically by a point or by the displacement $[0, 1]$. And when we say that i is a root of the equation $z^2 + 1 = 0$, what we mean is simply that we have defined a method of combining such pairs of numbers (or displacements) which we call 'multiplication', and which, when we so combine $(0, 1)$ with itself, gives the result $(-1, 0)$.

Now let us consider the more general equation

$$az^2 + 2bz + c = 0,$$

where a, b, c are real numbers. If $b^2 > ac$, the ordinary method of solution gives two real roots

$$\{-b \pm \sqrt{b^2 - ac}\}/a.$$

If $b^2 < ac$, the equation has no real roots. It may be written in the form

$$\{z + (b/a)\}^2 = -(ac - b^2)/a^2,$$

an equation which is evidently satisfied if we substitute for $z + (b/a)$ either of the complex numbers $\pm i\sqrt{ac - b^2}/a$.† We express this by saying that the equation has *the two complex roots*

$$\{-b \pm i\sqrt{ac - b^2}\}/a.$$

*The phrase 'real number' was introduced as an antithesis to 'imaginary number'.

†We shall sometimes write $x + iy$ instead of $x + yi$ for convenience in printing.

If we agree as a matter of convention to say that when $b^2 = ac$ (in which case the equation is satisfied by *one* value of x only, viz. $-b/a$), the equation has *two equal roots*, we can say that *a quadratic equation with real coefficients has two roots in all cases, either two distinct real roots, or two equal real roots, or two distinct complex roots.*

The question is naturally suggested whether a quadratic equation may not, when complex roots are once admitted, have more than two roots. It is easy to see that this is not possible. Its impossibility may in fact be proved by precisely the same chain of reasoning as is used in elementary algebra to prove that an equation of the nth degree cannot have more than n real roots. Let us denote the complex number $x + yi$ by the single letter z, a convention which we may express by writing $z = x + yi$. Let $f(z)$ denote any polynomial in z, with real or complex coefficients. Then we prove in succession:

(1) that the remainder, when $f(z)$ is divided by $z - a$, a being any real or complex number, is $f(a)$;

(2) that if a is a root of the equation $f(z) = 0$, then $f(z)$ is divisible by $z - a$;

(3) that if $f(z)$ is of the nth degree, and $f(z) = 0$ has the n roots a_1, a_2, ..., a_n, then

$$f(z) = A(z - a_1)(z - a_2) \dots (z - a_n),$$

where A is a constant, real or complex, in fact the coefficient of z^n in $f(z)$. From the last result, and the theorem of § 40, it follows that $f(z)$ cannot have more than n roots.

We conclude that a quadratic equation with real coefficients has exactly two roots. We shall see later on that a similar theorem is true for an equation of any degree and with either real or complex coefficients: *an equation of the nth degree has exactly n roots.* The only point in the proof which presents any difficulty is the first, viz. the proof that any equation must have *at least one* root. This we must postpone for the present.* We may, however, at once call attention to one very interesting result of this

*See Appendix I.

theorem. In the theory of number we start from the positive integers and from the ideas of addition and multiplication and the converse operations of subtraction and division. We find that these operations are not always possible unless we admit new kinds of numbers. We can only attach a meaning to $3-7$ if we admit *negative* numbers, or to $\frac{3}{7}$ if we admit *rational fractions*. When we extend our list of arithmetical operations so as to include root extraction and the solution of equations, we find that some of them, such as that of the extraction of the square root of a number which (like 2) is not a perfect square, are not possible unless we widen our conception of a number, and admit the *irrational* numbers of Chap. I.

Others, such as the extraction of the square root of -1, are not possible unless we go still further, and admit the *complex* numbers of this chapter. And it would not be unnatural to suppose that, when we come to consider equations of higher degree, some might prove to be insoluble even by the aid of complex numbers, and that thus we might be led to the considerations of higher and higher types of, so to say, *hyper-complex* numbers. The fact that the roots of any algebraical equation whatever are ordinary complex numbers shows that this is not the case. The application of any of the ordinary algebraical operations to complex numbers will yield only complex numbers. In technical language 'the field of the complex numbers is closed for algebraical operations'.

Before we pass on to other matters, let us add that all theorems of elementary algebra which are proved merely by the application of the rules of addition and multiplication are true *whether the numbers which occur in them are real or complex*, since the rules referred to apply to complex as well as real numbers. For example, we know that, if α and β are the roots of

$$az^2 + 2bz + c = 0,$$

then

$$\alpha + \beta = -(2b/a), \quad \alpha\beta = (c/a).$$

Similarly, if α, β, γ are the roots of

$$az^3 + 3bz^2 + 3cz + d = 0,$$

then

$$\alpha + \beta + \gamma = -(3b/a), \quad \beta\gamma + \gamma\alpha + \alpha\beta = (3c/a), \quad \alpha\beta\gamma = -(d/a).$$

All such theorems as these are true whether a, b, ..., α, β, ... are real or complex.

44. Argand's diagram. Let P (Fig. 24) be the point (x, y), r the length OP, and θ the angle XOP, so that

$$x = r\cos\theta, \quad y = r\sin\theta, \quad r = \sqrt{x^2 + y^2}, \quad \cos\theta : \sin\theta : 1 :: x : y : r.$$

We denote the complex number $x + yi$ by z, as in § 43, and we call z the *complex variable*. We call P *the point* z, or the point corresponding to z;

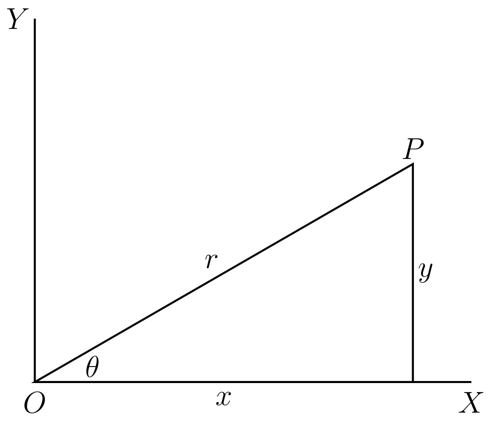

Fig. 24.

z the *argument* of P, x the *real part*, y the *imaginary part*, r the *modulus*, and θ the *amplitude* of z; and we write

$$x = \mathbf{R}(z), \quad y = \mathbf{I}(z), \quad r = |z|, \quad \theta = \operatorname{am} z.$$

When $y = 0$ we say that *z is real*, when $x = 0$ that *z is purely imaginary*. Two numbers $x + yi$, $x - yi$ which differ only in the signs of their imaginary parts, we call *conjugate*. It will be observed that the sum $2x$ of two conjugate numbers and their product $x^2 + y^2$ are both real, that they have the same modulus $\sqrt{x^2 + y^2}$ and that their product is equal to the square of the modulus of either. The roots of a quadratic with real coefficients, for example, are conjugate, when not real.

It must be observed that θ or $\operatorname{am} z$ is a many-valued function of x and y, having an infinity of values, which are angles differing by multiples of 2π.* A line originally lying along OX will, if turned through any of these angles, come to lie along OP. We shall describe that one of these angles which lies between $-\pi$ and π as the *principal value* of the amplitude of z. This definition is unambiguous except when one of the values is π, in which case $-\pi$ is also a value. In this case we must make some special provision as to which value is to be regarded as the principal value. In general, when we speak of the amplitude of z we shall, unless the contrary is stated, mean the principal value of the amplitude.

Fig. 24 is usually known as Argand's diagram.

45. De Moivre's Theorem. The following statements follow immediately from the definitions of addition and multiplication.

(1) The real (or imaginary) part of the sum of two complex numbers is equal to the sum of their real (or imaginary) parts.

(2) The modulus of the product of two complex numbers is equal to the product of their moduli.

(3) The amplitude of the product of two complex numbers is either equal to the sum of their amplitudes, or differs from it by 2π.

It should be observed that it is not always true that the principal value of $\operatorname{am}(zz')$ is the sum of the principal values of $\operatorname{am} z$ and $\operatorname{am} z'$. For example, if $z = z' = -1 + i$, then the principal values of the amplitudes of z and z' are each $\frac{3}{4}\pi$. But $zz' = -2i$, and the principal value of $\operatorname{am}(zz')$ is $-\frac{1}{2}\pi$ and not $\frac{3}{2}\pi$.

*It is evident that $|z|$ is identical with the polar coordinate r of P, and that the other polar coordinate θ is one value of $\operatorname{am} z$. This value is not necessarily the *principal* value, as defined below, for the polar coordinate of § 22 lies between 0 and 2π, and the principal value between $-\pi$ and π.

The two last theorems may be expressed in the equation

$$r(\cos\theta + i\sin\theta) \times \rho(\cos\phi + i\sin\phi) = r\rho\{\cos(\theta+\phi) + i\sin(\theta+\phi)\},$$

which may be proved at once by multiplying out and using the ordinary trigonometrical formulae for $\cos(\theta+\phi)$ and $\sin(\theta+\phi)$. More generally

$$\begin{aligned} &r_1(\cos\theta_1 + i\sin\theta_1) \times r_2(\cos\theta_2 + i\sin\theta_2) \times \dots \times r_n(\cos\theta_n + i\sin\theta_n) \\ &\quad = r_1 r_2 \dots r_n\{\cos(\theta_1+\theta_2+\dots+\theta_n) + i\sin(\theta_1+\theta_2+\dots+\theta_n)\}. \end{aligned}$$

A particularly interesting case is that in which

$$r_1 = r_2 = \dots = r_n = 1, \quad \theta_1 = \theta_2 = \dots = \theta_n = \theta.$$

We then obtain the equation

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta,$$

where n is any positive integer: a result known as *De Moivre's Theorem.**

Again, if

$$z = r(\cos\theta + i\sin\theta)$$

then

$$1/z = (\cos\theta - i\sin\theta)/r.$$

Thus the modulus of the reciprocal of z is the reciprocal of the modulus of z, and the amplitude of the reciprocal is the negative of the amplitude of z. We can now state the theorems for quotients which correspond to (2) and (3).

(4) The modulus of the quotient of two complex numbers is equal to the quotient of their moduli.

(5) The amplitude of the quotient of two complex numbers either is equal to the difference of their amplitudes, or differs from it by 2π.

---

*It will sometimes be convenient, for the sake of brevity, to denote $\cos\theta + i\sin\theta$ by $\operatorname{Cis}\theta$: in this notation, suggested by Profs. Harkness and Morley, De Moivre's theorem is expressed by the equation $(\operatorname{Cis}\theta)^n = \operatorname{Cis} n\theta$.

Again

$$(\cos\theta + i\sin\theta)^{-n} = (\cos\theta - i\sin\theta)^n$$
$$= \{\cos(-\theta) + i\sin(-\theta)\}^n$$
$$= \cos(-n\theta) + i\sin(-n\theta).$$

Hence *De Moivre's Theorem holds for all integral values of* n*, positive or negative.*

To the theorems (1)–(5) we may add the following theorem, which is also of very great importance.

(6) The modulus of the sum of any number of complex numbers is not greater than the sum of their moduli.

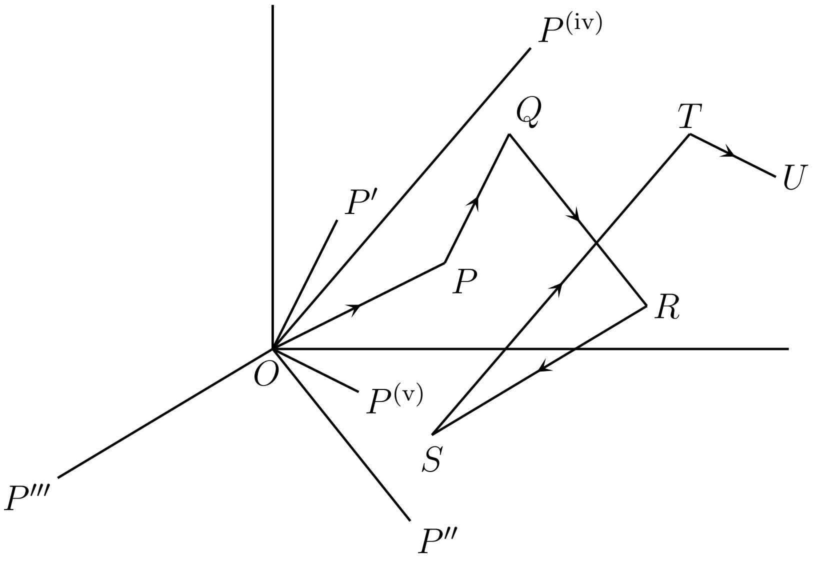

Fig. 25.

Let $\overline{OP}$, $\overline{OP'}$, ... be the displacements corresponding to the various complex numbers. Draw PQ equal and parallel to OP', QR equal and parallel to OP'', and so on. Finally we reach a point U, such that

$$\overline{OU} = \overline{OP} + \overline{OP'} + \overline{OP''} + \dots.$$

The length OU is the modulus of the sum of the complex numbers, whereas the sum of their moduli is the total length of the broken line $OPQR\dots U$, which is not less than OU.

A purely arithmetical proof of this theorem is outlined in Exs. XXI. 1.

46. We add some theorems concerning rational functions of complex numbers. A *rational function* of the complex variable z is defined exactly as is a rational function of a real variable x, viz. as the quotient of two polynomials in z.

THEOREM 1. *Any rational function $R(z)$ can be reduced to the form $X+Yi$, where X and Y are rational functions of x and y with real coefficients.*

In the first place it is evident that any polynomial $P(x+yi)$ can be reduced, in virtue of the definitions of addition and multiplication, to the form $A+Bi$, where A and B are polynomials in x and y with real coefficients. Similarly $Q(x+yi)$ can be reduced to the form $C+Di$. Hence

$$R(x+yi) = P(x+yi)/Q(x+yi)$$

can be expressed in the form

$$\begin{aligned}(A+Bi)/(C+Di) &= (A+Bi)(C-Di)/(C+Di)(C-Di)\\ &= \frac{AC+BD}{C^2+D^2} + \frac{BC-AD}{C^2+D^2}i,\end{aligned}$$

which proves the theorem.

THEOREM 2. *If $R(x+yi) = X+Yi$, R denoting a rational function as before, but with* **real** *coefficients, then $R(x-yi) = X-Yi$.*

In the first place this is easily verified for a power $(x+yi)^n$ by actual expansion. It follows by addition that the theorem is true for any polynomial with real coefficients. Hence, in the notation used above,

$$R(x-yi) = \frac{A-Bi}{C-Di} = \frac{AC+BD}{C^2+D^2} - \frac{BC-AD}{C^2+D^2}i,$$

the reduction being the same as before except that the sign of i is changed throughout. It is evident that results similar to those of Theorems 1 and 2 hold for functions of any number of complex variables.

THEOREM 3. *The roots of an equation*

$$a_0 z^n + a_1 z^{n-1} + \dots + a_n = 0,$$

whose coefficients are real, may, in so far as they are not themselves real, be arranged in conjugate pairs.

For it follows from Theorem 2 that if $x + yi$ is a root then so is $x - yi$. A particular case of this theorem is the result (§ 43) that the roots of a quadratic equation with real coefficients are either real or conjugate.

This theorem is sometimes stated as follows: *in an equation with real coefficients complex roots occur in conjugate pairs*. It should be compared with the result of Exs. VIII. 7, which may be stated as follows: *in an equation with rational coefficients irrational roots occur in conjugate pairs.**

Examples XXI. 1. Prove theorem (6) of § 45 directly from the definitions and without the aid of geometrical considerations.

[First, to prove that $|z + z'| \leqq |z| + |z'|$ is to prove that

$$(x + x')^2 + (y + y')^2 \leqq \{\sqrt{x^2 + y^2} + \sqrt{x'^2 + y'^2}\}^2.$$

The theorem is then easily extended to the general case.]

2. The one and only case in which

$$|z| + |z'| + \dots = |z + z' + \dots|,$$

is that in which the numbers z, z', ... have all the same amplitude. Prove this both geometrically and analytically.

3. The modulus of the sum of any number of complex numbers is not less than the sum of their real (or imaginary) parts.

4. If the sum and product of two complex numbers are both real, then the two numbers must either be real or conjugate.

*The numbers $a + \sqrt{b}$, $a - \sqrt{b}$, where a, b are rational, are sometimes said to be 'conjugate'.

5. If

$$a + b\sqrt{2} + (c + d\sqrt{2})i = A + B\sqrt{2} + (C + D\sqrt{2})i,$$

where a, b, c, d, A, B, C, D are real rational numbers, then

$$a = A, \quad b = B, \quad c = C, \quad d = D.$$

6. Express the following numbers in the form $A + Bi$, where A and B are real numbers:

$$(1+i)^2, \quad \left(\frac{1+i}{1-i}\right)^2, \quad \left(\frac{1-i}{1+i}\right)^2, \quad \frac{\lambda+\mu i}{\lambda-\mu i}, \quad \left(\frac{\lambda+\mu i}{\lambda-\mu i}\right)^2 - \left(\frac{\lambda-\mu i}{\lambda+\mu i}\right)^2,$$

where λ and μ are real numbers.

7. Express the following functions of $z = x + yi$ in the form $X + Yi$, where X and Y are real functions of x and y: z^2, z^3, z^n, $1/z$, $z+(1/z)$, $(\alpha+\beta z)/(\gamma+\delta z)$, where α, β, γ, δ are real numbers.

8. Find the moduli of the numbers and functions in the two preceding examples.

9. The two lines joining the points $z = a$, $z = b$ and $z = c$, $z = d$ will be perpendicular if

$$\operatorname{am}\left(\frac{a-b}{c-d}\right) = \pm\tfrac{1}{2}\pi,$$

i.e. if $(a-b)/(c-d)$ is purely imaginary. What is the condition that the lines should be parallel?

10. The three angular points of a triangle are given by $z = \alpha$, $z = \beta$, $z = \gamma$, where α, β, γ are complex numbers. Establish the following propositions:

(i) *the centre of gravity is given by* $z = \frac{1}{3}(\alpha + \beta + \gamma)$;

(ii) *the circum-centre is given by* $|z - \alpha| = |z - \beta| = |z - \gamma|$;

(iii) *the three perpendiculars from the angular points on the opposite sides meet in a point given by*

$$\mathbf{R}\left(\frac{z-\alpha}{\beta-\gamma}\right) = \mathbf{R}\left(\frac{z-\beta}{\gamma-\alpha}\right) = \mathbf{R}\left(\frac{z-\gamma}{\alpha-\beta}\right) = 0;$$

(iv) *there is a point P inside the triangle such that*

$$CBP = ACP = BAP = \omega,$$

and

$$\cot\omega = \cot A + \cot B + \cot C.$$

[To prove (iii) we observe that if A, B, C are the vertices, and P any point z, then the condition that AP should be perpendicular to BC is (Ex. 9) that $(z-\alpha)/(\beta-\gamma)$ should be purely imaginary, or that

$$\mathbf{R}(z-\alpha)\,\mathbf{R}(\beta-\gamma) + \mathbf{I}(z-\alpha)\,\mathbf{I}(\beta-\gamma) = 0.$$

This equation, and the two similar equations obtained by permuting α, β, γ cyclically, are satisfied by the same value of z, as appears from the fact that the sum of the three left-hand sides is zero.

To prove (iv), take BC parallel to the positive direction of the axis of x. Then*

$$\gamma-\beta = a, \quad \alpha-\gamma = -b\operatorname{Cis}(-C), \quad \beta-\alpha = -c\operatorname{Cis}B.$$

We have to determine z and ω from the equations

$$\frac{(z-\alpha)(\beta_0-\alpha_0)}{(z_0-\alpha_0)(\beta-\alpha)} = \frac{(z-\beta)(\gamma_0-\beta_0)}{(z_0-\beta_0)(\gamma-\beta)} = \frac{(z-\gamma)(\alpha_0-\gamma_0)}{(z_0-\gamma_0)(\alpha-\gamma)} = \operatorname{Cis}2\omega,$$

where z_0, α_0, β_0, γ_0 denote the conjugates of z, α, β, γ.

Adding the numerators and denominators of the three equal fractions, and using the equation

$$i\cot\omega = (1+\operatorname{Cis}2\omega)/(1-\operatorname{Cis}2\omega),$$

we find that

$$i\cot\omega = \frac{(\beta-\gamma)(\beta_0-\gamma_0)+(\gamma-\alpha)(\gamma_0-\alpha_0)+(\alpha-\beta)(\alpha_0-\beta_0)}{\beta\gamma_0-\beta_0\gamma+\gamma\alpha_0-\gamma_0\alpha+\alpha\beta_0-\alpha_0\beta}.$$

From this it is easily deduced that the value of $\cot\omega$ is $(a^2+b^2+c^2)/4\Delta$, where Δ is the area of the triangle; and this is equivalent to the result given.

To determine z, we multiply the numerators and denominators of the equal fractions by $(\gamma_0-\beta_0)/(\beta-\alpha)$, $(\alpha_0-\gamma_0)/(\gamma-\beta)$, $(\beta_0-\alpha_0)/(\alpha-\gamma)$, and add to form a new fraction. It will be found that

$$z = \frac{a\alpha\operatorname{Cis}A + b\beta\operatorname{Cis}B + c\gamma\operatorname{Cis}C}{a\operatorname{Cis}A + b\operatorname{Cis}B + c\operatorname{Cis}C}.]$$

*We suppose that as we go round the triangle in the direction ABC we leave it on our left.

11. The two triangles whose vertices are the points a, b, c and x, y, z respectively will be similar if

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ x & y & z \end{vmatrix} = 0$$

[The condition required is that $\overline{AB}/\overline{AC} = \overline{XY}/\overline{XZ}$ (large letters denoting the points whose arguments are the corresponding small letters), or $(b-a)/(c-a) = (y-x)/(z-x)$, which is the same as the condition given.]

12. Deduce from the last example that if the points x, y, z are collinear then we can find *real* numbers α, β, γ such that $\alpha+\beta+\gamma = 0$ and $\alpha x + \beta y + \gamma z = 0$, and conversely (cf. Exs. XX. 4). [Use the fact that in this case the triangle formed by x, y, z is similar to a certain line-triangle on the axis OX, and apply the result of the last example.]

13. **The general linear equation with complex coefficients.** The equation $\alpha z + \beta = 0$ has the one solution $z = -(\beta/\alpha)$, unless $\alpha = 0$. If we put

$$\alpha = a + Ai, \quad \beta = b + Bi, \quad z = x + yi,$$

and equate real and imaginary parts, we obtain two equations to determine the two real numbers x and y. The equation will have a real root if $y = 0$, which gives $ax + b = 0$, $Ax + B = 0$, and the condition that these equations should be consistent is $aB - bA = 0$.

14. **The general quadratic equation with complex coefficients.** This equation is

$$(a + Ai)z^2 + 2(b + Bi)z + (c + Ci) = 0.$$

Unless a and A are both zero we can divide through by $a + iA$. Hence we may consider

$$z^2 + 2(b + Bi)z + (c + Ci) = 0 \tag{1}$$

as the standard form of our equation. Putting $z = x + yi$ and equating real and imaginary parts, we obtain a pair of simultaneous equations for x and y, viz.

$$x^2 - y^2 + 2(bx - By) + c = 0, \quad 2xy + 2(by + Bx) + C = 0.$$

If we put

$$x + b = \xi, \quad y + B = \eta, \quad b^2 - B^2 - c = h, \quad 2bB - C = k,$$

these equations become

$$\xi^2 - \eta^2 = h, \quad 2\xi\eta = k.$$

Squaring and adding we obtain

$$\xi^2 + \eta^2 = \sqrt{h^2 + k^2}, \quad \xi = \pm\sqrt{\tfrac{1}{2}\{\sqrt{h^2+k^2}+h\}}, \quad \eta = \pm\sqrt{\tfrac{1}{2}\{\sqrt{h^2+k^2}-h\}}.$$

We must choose the signs so that $\xi\eta$ has the sign of k: *i.e.* if k is positive we must take like signs, if k is negative unlike signs.

Conditions for equal roots. The two roots can only be equal if both the square roots above vanish, *i.e.* if $h = 0$, $k = 0$, or if $c = b^2 - B^2$, $C = 2bB$. These conditions are equivalent to the single condition $c + Ci = (b + Bi)^2$, which obviously expresses the fact that the left-hand side of (1) is a perfect square.

Condition for a real root. If $x^2 + 2(b+Bi)x + (c+Ci) = 0$, where x is real, then $x^2 + 2bx + c = 0$, $2Bx + C = 0$. Eliminating x we find that the required condition is

$$C^2 - 4bBC + 4cB^2 = 0.$$

Condition for a purely imaginary root. This is easily found to be

$$C^2 - 4bBC - 4b^2c = 0.$$

Conditions for a pair of conjugate complex roots. Since the sum and the product of two conjugate complex numbers are both real, $b + Bi$ and $c + Ci$ must both be real, *i.e.* $B = 0$, $C = 0$. Thus the equation (1) can have a pair of conjugate complex roots only if its coefficients are real. The reader should verify this conclusion by means of the explicit expressions of the roots. Moreover, if $b^2 \geqq c$, the roots will be real even in this case. Hence for a pair of conjugate roots we must have $B = 0$, $C = 0$, $b^2 < c$.

15. **The Cubic equation.** Consider the cubic equation

$$z^3 + 3Hz + G = 0,$$

where G and H are complex numbers, it being given that the equation has (*a*) a real root, (*b*) a purely imaginary root, (*c*) a pair of conjugate roots. If $H = \lambda + \mu i$, $G = \rho + \sigma i$, we arrive at the following conclusions.

(a) *Conditions for a real root.* If μ is not zero, then the real root is $-\sigma/3\mu$, and $\sigma^3 + 27\lambda\mu^2\sigma - 27\mu^3\rho = 0$. On the other hand, if $\mu = 0$ then we must also have $\sigma = 0$, so that the coefficients of the equation are real. In this case there may be three real roots.

(b) *Conditions for a purely imaginary root.* If μ is not zero then the purely imaginary root is $(\rho/3\mu)i$, and $\rho^3 - 27\lambda\mu^2\rho - 27\mu^3\sigma = 0$. If $\mu = 0$ then also $\rho = 0$, and the root is yi, where y is given by the equation $y^3 - 3\lambda y - \sigma = 0$, which has real coefficients. In this case there may be three purely imaginary roots.

(c) *Conditions for a pair of conjugate complex roots.* Let these be $x+yi$ and $x - yi$. Then since the sum of the three roots is zero the third root must be $-2x$. From the relations between the coefficients and the roots of an equation we deduce

$$y^2 - 3x^2 = 3H, \quad 2x(x^2 + y^2) = G.$$

Hence G and H must both be real.

In each case we can either find a root (in which case the equation can be reduced to a quadratic by dividing by a known factor) or we can reduce the solution of the equation to the solution of a cubic equation with real coefficients.

16. The cubic equation $x^3+a_1x^2+a_2x+a_3 = 0$, where $a_1 = A_1+A_1'i, \dots$, has a pair of conjugate complex roots. Prove that the remaining root is $-A_1'a_3/A_3'$, unless $A_3' = 0$. Examine the case in which $A_3' = 0$.

17. Prove that if $z^3 + 3Hz + G = 0$ has two complex roots then the equation

$$8\alpha^3 + 6\alpha H - G = 0$$

has one real root which is the real part α of the complex roots of the original equation; and show that α has the same sign as G.

18. An equation of any order with complex coefficients will in general have no real roots nor pairs of conjugate complex roots. How many conditions must be satisfied by the coefficients in order that the equation should have (a) a real root, (b) a pair of conjugate roots?

19. **Coaxal circles.** In Fig. 26, let a, b, z be the arguments of A, B, P. Then

$$\operatorname{am}\frac{z-b}{z-a} = APB,$$

if the principal value of the amplitude is chosen. If the two circles shown in the figure are equal, and z', z_1, z_1' are the arguments of P', P_1, P_1', and $APB = \theta$, it is easy to see that

$$\operatorname{am}\frac{z'-b}{z'-a} = \pi - \theta, \quad \operatorname{am}\frac{z_1-b}{z_1-a} = -\theta,$$

and

$$\operatorname{am}\frac{z_1'-b}{z_1'-a} = -\pi + \theta.$$

The locus defined by the equation

$$\operatorname{am}\frac{z-b}{z-a} = \theta,$$

where θ is constant, is the arc APB. By writing $\pi - \theta$, $-\theta$, $-\pi + \theta$ for θ, we obtain the other three arcs shown.

The system of equations obtained by supposing that θ is a parameter, varying from $-\pi$ to $+\pi$, represents *the system of circles which can be drawn through the points A, B*. It should however be observed that each circle has to be divided into two parts to which correspond different values of θ.

20. Now let us consider the equation

$$\left|\frac{z-b}{z-a}\right| = \lambda, \tag{1}$$

where λ is a constant.

Let K be the point in which the tangent to the circle ABP at P meets AB. Then the triangles KPA, KBP are similar, and so

$$AP/PB = PK/BK = KA/KP = \lambda.$$

Hence $KA/KB = \lambda^2$, and therefore K is a fixed point for all positions of P which satisfy the equation (1). Also $KP^2 = KA \cdot KB$, and so is constant. Hence *the locus of P is a circle whose centre is K*.

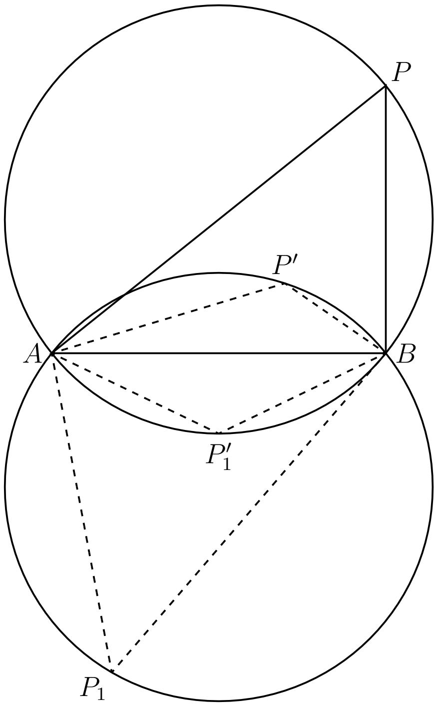

Fig. 26.

The system of equations obtained by varying λ represents a system of circles, and every circle of this system cuts at right angles every circle of the system of Ex. 19.

The system of Ex. 19 is called *a system of coaxal circles of the common point kind*. The system of Ex. 20 is called *a system of coaxal circles of the limiting point kind*, A and B being the *limiting points* of the system. If λ is very large or very small then the circle is a very small circle containing A or B in its interior.

21. **Bilinear Transformations.** Consider the equation

$$z = Z + a, \tag{1}$$

where $z = x + yi$ and $Z = X + Yi$ are two complex variables which we may suppose to be represented in two planes xoy, XOY. To every value of z corresponds one of Z, and conversely. If $a = \alpha + \beta i$ then

$$x = X + \alpha, \quad y = Y + \beta,$$

and to the point (x, y) corresponds the point (X, Y). If (x, y) describes a curve of any kind in its plane, (X, Y) describes a curve in its plane. Thus to any figure in one plane corresponds a figure in the other. A passage of this kind from a figure in the plane xoy to a figure in the plane XOY by means of a relation such as (1) between z and Z is called a *transformation*. In this particular case the relation between corresponding figures is very easily defined. The (X, Y) figure is the same in size, shape, and orientation as the (x, y) figure, but is shifted a distance α to the left, and a distance β downwards. Such a transformation is called a *translation*.

Now consider the equation

$$z = \rho Z, \tag{2}$$

where ρ is real. This gives $x = \rho X$, $y = \rho Y$. The two figures are similar and similarly situated about their respective origins, but the scale of the (x, y) figure is ρ times that of the (X, Y) figure. Such a transformation is called a *magnification*.

Finally consider the equation

$$z = (\cos\phi + i\sin\phi)Z. \tag{3}$$

It is clear that $|z| = |Z|$ and that one value of $\operatorname{am} z$ is $\operatorname{am} Z + \phi$, and that the two figures differ only in that the (x, y) figure is the (X, Y) figure turned about the origin through an angle ϕ in the positive direction. Such a transformation is called a *rotation*.

The general linear transformation

$$z = aZ + b \tag{4}$$

is a combination of the three transformations (1), (2), (3). For, if $|a| = \rho$ and $\operatorname{am} a = \phi$, we can replace (4) by the three equations

$$z = z' + b, \quad z' = \rho Z', \quad Z' = (\cos\phi + i\sin\phi)Z.$$

Thus *the general linear transformation is equivalent to the combination of a translation, a magnification, and a rotation.*

Next let us consider the transformation

$$z = 1/Z. \tag{5}$$

If $|Z| = R$ and $\operatorname{am} Z = \Theta$, then $|z| = 1/R$ and $\operatorname{am} z = -\Theta$, and to pass from the (x, y) figure to the (X, Y) figure we invert the former with respect to o, with unit radius of inversion, and then construct the image of the new figure in the axis ox (*i.e.* the symmetrical figure on the other side of ox).

Finally consider the transformation

$$z = \frac{aZ + b}{cZ + d}. \tag{6}$$

This is equivalent to the combination of the transformations

$$z = (a/c) + (bc - ad)(z'/c), \quad z' = 1/Z', \quad Z' = cZ + d,$$

i.e. to a certain combination of transformations of the types already considered.

The transformation (6) is called the *general bilinear transformation*. Solving for Z we obtain

$$Z = \frac{dz - b}{cz - a}.$$

The general bilinear transformation is the most general type of transformation for which one and only one value of z corresponds to each value of Z, and conversely.

22. *The general bilinear transformation transforms circles into circles.* This may be proved in a variety of ways. We may assume the well-known theorem in pure geometry, that inversion transforms circles into circles (which may of course in particular cases be straight lines). Or we may use the results of Exs. 19 and 20. If, *e.g.*, the (x, y) circle is

$$|(z - \sigma)/(z - \rho)| = \lambda,$$

and we substitute for z in terms of Z, we obtain

$$|(Z - \sigma')/(Z - \rho')| = \lambda',$$

where

$$\sigma' = -\frac{b - \sigma d}{a - \sigma c}, \quad \rho' = -\frac{b - \rho d}{a - \rho c}, \quad \lambda' = \left|\frac{a - \rho c}{a - \sigma c}\right| \lambda.$$

23. Consider the transformations $z = 1/Z$, $z = (1+Z)/(1-Z)$, and draw the (X, Y) curves which correspond to (1) circles whose centre is the origin, (2) straight lines through the origin.

24. The condition that the transformation $z = (aZ + b)/(cZ + d)$ should make the circle $x^2 + y^2 = 1$ correspond to a straight line in the (X, Y) plane is $|a| = |c|$.

25. **Cross ratios.** The cross ratio $(z_1, z_2; z_3, z_4)$ is defined to be

$$\frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

If the four points z_1, z_2, z_3, z_4 are on the same line, this definition agrees with that adopted in elementary geometry. There are 24 cross ratios which can be formed from z_1, z_2, z_3, z_4 by permuting the suffixes. These consist of six groups of four equal cross ratios. If one ratio is λ, then the six distinct cross ratios are λ, $1-\lambda$, $1/\lambda$, $1/(1-\lambda)$, $(\lambda-1)/\lambda$, $\lambda/(\lambda-1)$. The four points are said to be *harmonic* or *harmonically related* if any one of these is equal to -1. In this case the six ratios are -1, 2, -1, $\frac{1}{2}$, 2, $\frac{1}{2}$.

If any cross ratio is real then all are real and the four points lie on a circle. For in this case

$$\operatorname{am} \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}$$

must have one of the three values $-\pi$, 0, π, so that $\operatorname{am}\{(z_1 - z_3)/(z_1 - z_4)\}$ and $\operatorname{am}\{(z_2 - z_3)/(z_2 - z_4)\}$ must either be equal or differ by π (cf. Ex. 19).

If $(z_1, z_2; z_3, z_4) = -1$, we have the two equations

$$\operatorname{am} \frac{z_1 - z_3}{z_1 - z_4} = \pm\pi + \operatorname{am} \frac{z_2 - z_3}{z_2 - z_4}, \qquad \left|\frac{z_1 - z_3}{z_1 - z_4}\right| = \left|\frac{z_2 - z_3}{z_2 - z_4}\right|.$$

The four points A_1, A_2, A_3, A_4 lie on a circle, A_1 and A_2 being separated by A_3 and A_4. Also $A_1A_3/A_1A_4 = A_2A_3/A_2A_4$. Let O be the middle point of A_3A_4. The equation

$$\frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} = -1$$

may be put in the form

$$(z_1 + z_2)(z_3 + z_4) = 2(z_1 z_2 + z_3 z_4),$$

or, what is the same thing,

$$\{z_1 - \tfrac{1}{2}(z_3 + z_4)\}\{z_2 - \tfrac{1}{2}(z_3 + z_4)\} = \{\tfrac{1}{2}(z_3 - z_4)\}^2.$$

But this is equivalent to $\overline{OA_1} \cdot \overline{OA_2} = \overline{OA_3}^2 = \overline{OA_4}^2$. Hence OA_1 and OA_2 make equal angles with A_3A_4, and $OA_1 \cdot OA_2 = OA_3^2 = OA_4^2$. It will be observed that the relation between the pairs A_1, A_2 and A_3, A_4 is symmetrical. Hence, if O' is the middle point of A_1A_2, $O'A_3$ and $O'A_4$ are equally inclined to A_1A_2, and $O'A_3 \cdot O'A_4 = O'A_1^2 = O'A_2^2$.

26. If the points A_1, A_2 are given by $az^2 + 2bz + c = 0$, and the points A_3, A_4 by $a'z^2 + 2b'z + c' = 0$, and O is the middle point of A_3A_4, and $ac' + a'c - 2bb' = 0$, then OA_1, OA_2 are equally inclined to A_3A_4 and $OA_1 \cdot OA_2 = OA_3^2 = OA_4^2$.

(*Math. Trip.* 1901.)

27. AB, CD are two intersecting lines in Argand's diagram, and P, Q their middle points. Prove that, if AB bisects the angle CPD and $PA^2 = PB^2 = PC \cdot PD$, then CD bisects the angle AQB and $QC^2 = QD^2 = QA \cdot QB$.

(*Math. Trip.* 1909.)

28. **The condition that four points should lie on a circle.** A sufficient condition is that one (and therefore all) of the cross ratios should be real (Ex. 25); this condition is also necessary. Another form of the condition is that it should be possible to choose real numbers α, β, γ such that

$$\begin{vmatrix} 1 & 1 & 1 \\ \alpha & \beta & \gamma \\ z_1z_4 + z_2z_3 & z_2z_4 + z_3z_1 & z_3z_4 + z_1z_2 \end{vmatrix} = 0.$$

[To prove this we observe that the transformation $Z = 1/(z - z_4)$ is equivalent to an inversion with respect to the point z_4, coupled with a certain reflexion (Ex. 21). If z_1, z_2, z_3 lie on a circle through z_4, the corresponding points $Z_1 = 1/(z_1 - z_4)$, $Z_2 = 1/(z_2 - z_4)$, $Z_3 = 1/(z_3 - z_4)$ lie on a straight line. Hence (Ex. 12) we can find real numbers α', β', γ' such that $\alpha' + \beta' + \gamma' = 0$ and $\alpha'/(z_1 - z_4) + \beta'/(z_2 - z_4) + \gamma'/(z_3 - z_4) = 0$, and it is easy to prove that this is equivalent to the given condition.]

29. Prove the following analogue of De Moivre's Theorem for real numbers: if ϕ_1, ϕ_2, ϕ_3, ... is a series of positive acute angles such that

$$\tan\phi_{m+1} = \tan\phi_m \sec\phi_1 + \sec\phi_m \tan\phi_1,$$

then

$$\tan\phi_{m+n} = \tan\phi_m \sec\phi_n + \sec\phi_m \tan\phi_n,$$
$$\sec\phi_{m+n} = \sec\phi_m \sec\phi_n + \tan\phi_m \tan\phi_n,$$

and

$$\tan\phi_m + \sec\phi_m = (\tan\phi_1 + \sec\phi_1)^m.$$

[Use the method of mathematical induction.]

30. **The transformation $z = Z^m$.** In this case $r = R^m$, and θ and $m\Theta$ differ by a multiple of 2π. If Z describes a circle round the origin then z describes a circle round the origin m times.

The whole (x, y) plane corresponds to any one of m sectors in the (X, Y) plane, each of angle $2\pi/m$. To each point in the (x, y) plane correspond m points in the (X, Y) plane.

31. **Complex functions of a real variable.** If $f(t)$, $\phi(t)$ are two real functions of a real variable t defined for a certain range of values of t, we call

$$z = f(t) + i\phi(t) \tag{1}$$

a complex function of t. We can represent it graphically by drawing the curve

$$x = f(t), \quad y = \phi(t);$$

the equation of the curve may be obtained by eliminating t between these equations. If z is a polynomial in t, or rational function of t, with complex coefficients, we can express it in the form (1) and so determine the curve represented by the function.

(i) Let

$$z = a + (b - a)t,$$

where a and b are complex numbers. If $a = \alpha + \alpha' i$, $b = \beta + \beta' i$, then

$$x = \alpha + (\beta - \alpha)t, \quad y = \alpha' + (\beta' - \alpha')t.$$

The curve is the straight line joining the points $z = a$ and $z = b$. The segment between the points corresponds to the range of values of t from 0 to 1. Find the values of t which correspond to the two produced segments of the line.

(ii) If

$$z = c + \rho\left(\frac{1+ti}{1-ti}\right),$$

where ρ is positive, then the curve is the circle of centre c and radius ρ. As t varies through all real values z describes the circle once.

(iii) In general the equation $z = (a+bt)/(c+dt)$ represents a circle. This can be proved by calculating x and y and eliminating: but this process is rather cumbrous. A simpler method is obtained by using the result of Ex. 22. Let $z = (a + bZ)/(c + dZ)$, $Z = t$. As t varies Z describes a straight line, viz. the axis of X. Hence z describes a circle.

(iv) The equation

$$z = a + 2bt + ct^2$$

represents a parabola generally, a straight line if b/c is real.

(v) The equation $z = (a + 2bt + ct^2)/(\alpha + 2\beta t + \gamma t^2)$, where α, β, γ are real, represents a conic section.

[Eliminate t from

$$x = (A + 2Bt + Ct^2)/(\alpha + 2\beta t + \gamma t^2), \quad y = (A' + 2B't + C't^2)/(\alpha + 2\beta t + \gamma t^2),$$

where $A + A'i = a$, $B + B'i = b$, $C + C'i = c$.]

47. Roots of complex numbers. We have not, up to the present, attributed any meaning to symbols such as $\sqrt[n]{a}$, $a^{m/n}$, when a is a complex number, and m and n integers. It is, however, natural to adopt the definitions which are given in elementary algebra for real values of a. Thus we define $\sqrt[n]{a}$ or $a^{1/n}$, where n is a positive integer, as a number z which satisfies the equation $z^n = a$; and $a^{m/n}$, where m is an integer, as $(a^{1/n})^m$. These definitions do not prejudge the question as to whether there are or are not more than one (or any) roots of the equation.

48. Solution of the equation $z^n = a$. Let

$$a = \rho(\cos\phi + i\sin\phi),$$

where ρ is positive and ϕ is an angle such that $-\pi < \phi \leqq \pi$. If we put $z = r(\cos\theta + i\sin\theta)$, the equation takes the form

$$r^n(\cos n\theta + i\sin n\theta) = \rho(\cos\phi + i\sin\phi);$$

so that

$$r^n = \rho, \quad \cos n\theta = \cos\phi, \quad \sin n\theta = \sin\phi. \tag{1}$$

The only possible value of r is $\sqrt[n]{\rho}$, the ordinary arithmetical nth root of ρ; and in order that the last two equations should be satisfied it is necessary and sufficient that $n\theta = \phi + 2k\pi$, where k is an integer, or

$$\theta = (\phi + 2k\pi)/n.$$

If $k = pn + q$, where p and q are integers, and $0 \leqq q < n$, the value of θ is $2p\pi + (\phi + 2q\pi)/n$, and in this the value of p is a matter of indifference. Hence *the equation*

$$z^n = a = \rho(\cos\phi + i\sin\phi)$$

has n roots and n only, given by $z = r(\cos\theta + i\sin\theta)$, where

$$r = \sqrt[n]{\rho}, \quad \theta = (\phi + 2q\pi)/n, \quad (q = 0,\ 1,\ 2,\ \dots, n-1).$$

That these n roots are in reality all distinct is easily seen by plotting them on Argand's diagram. The particular root

$$\sqrt[n]{\rho}\{\cos(\phi/n) + i\sin(\phi/n)\}$$

is called the *principal value* of $\sqrt[n]{a}$.

The case in which $a = 1$, $\rho = 1$, $\phi = 0$ is of particular interest. The n roots of the equation $x^n = 1$ are

$$\cos(2q\pi/n) + i\sin(2q\pi/n), \quad (q = 0,\ 1,\ \dots, n-1).$$

These numbers are called the nth roots of unity; the principal value is unity itself. If we write ω_n for $\cos(2\pi/n) + i\sin(2\pi/n)$, we see that the nth roots of unity are

$$1, \quad \omega_n, \quad \omega_n^2, \ \dots, \quad \omega_n^{n-1}.$$

Examples XXII. 1. The two square roots of 1 are 1, -1; the three cube roots are 1, $\frac{1}{2}(-1+i\sqrt{3})$, $\frac{1}{2}(-1-i\sqrt{3})$; the four fourth roots are 1, i, -1, $-i$; and the five fifth roots are

$$\begin{gathered} 1,\quad \tfrac{1}{4}\left[\ \sqrt{5}-1+i\sqrt{10+2\sqrt{5}}\right],\quad \tfrac{1}{4}\left[-\sqrt{5}-1+i\sqrt{10-2\sqrt{5}}\right], \\ \tfrac{1}{4}\left[-\sqrt{5}-1-i\sqrt{10-2\sqrt{5}}\right],\quad \tfrac{1}{4}\left[\ \sqrt{5}-1-i\sqrt{10+2\sqrt{5}}\right]. \end{gathered}$$

2. Prove that

$$1+\omega_n+\omega_n^2+\dots+\omega_n^{n-1}=0.$$

3. Prove that

$$(x+y\omega_3+z\omega_3^2)(x+y\omega_3^2+z\omega_3)=x^2+y^2+z^2-yz-zx-xy.$$

4. The nth roots of a are the products of the nth roots of unity by the principal value of $\sqrt[n]{a}$.

5. It follows from Exs. XXI. 14 that the roots of

$$z^2=\alpha+\beta i$$

are

$$\pm\sqrt{\tfrac{1}{2}\{\sqrt{\alpha^2+\beta^2}+\alpha\}}\pm i\sqrt{\tfrac{1}{2}\{\sqrt{\alpha^2+\beta^2}-\alpha\}},$$

like or unlike signs being chosen according as β is positive or negative. Show that this result agrees with the result of § 48.

6. Show that $(x^{2m}-a^{2m})/(x^2-a^2)$ is equal to

$$\left(x^2-2ax\cos\frac{\pi}{m}+a^2\right)\left(x^2-2ax\cos\frac{2\pi}{m}+a^2\right)\dots\left(x^2-2ax\cos\frac{(m-1)\pi}{m}+a^2\right).$$

[The factors of $x^{2m}-a^{2m}$ are

$$(x-a),\quad (x-a\omega_{2m}),\quad (x-a\omega_{2m}^2),\ \dots\quad (x-a\omega_{2m}^{2m-1}).$$

The factor $x-a\omega_{2m}^m$ is $x+a$. The factors $(x-a\omega_{2m}^s)$, $(x-a\omega_{2m}^{2m-s})$ taken together give a factor $x^2-2ax\cos(s\pi/m)+a^2$.]

7. Resolve $x^{2m+1}-a^{2m+1}$, $x^{2m}+a^{2m}$, and $x^{2m+1}+a^{2m+1}$ into factors in a similar way.

8. Show that $x^{2n} - 2x^n a^n \cos\theta + a^{2n}$ is equal to

$$\left(x^2 - 2xa\cos\frac{\theta}{n} + a^2\right)\left(x^2 - 2xa\cos\frac{\theta+2\pi}{n} + a^2\right)\dots$$
$$\dots\left(x^2 - 2xa\cos\frac{\theta+2(n-1)\pi}{n} + a^2\right).$$

[Use the formula

$$x^{2n} - 2x^n a^n \cos\theta + a^{2n} = \{x^n - a^n(\cos\theta + i\sin\theta)\}\{x^n - a^n(\cos\theta - i\sin\theta)\},$$

and split up each of the last two expressions into n factors.]

9. Find all the roots of the equation $x^6 - 2x^3 + 2 = 0$. (*Math. Trip.* 1910.)

10. The problem of finding the accurate value of ω_n in a numerical form involving square roots only, as in the formula $\omega_3 = \frac{1}{2}(-1+i\sqrt{3})$, is the algebraical equivalent of the geometrical problem of inscribing a regular polygon of n sides in a circle of unit radius by Euclidean methods, *i.e.* by ruler and compasses. For this construction will be possible if and only if we can construct lengths measured by $\cos(2\pi/n)$ and $\sin(2\pi/n)$; and this is possible (Ch. II, Misc. Exs. 22) if and only if these numbers are expressible in a form involving square roots only.

Euclid gives constructions for $n = 3$, 4, 5, 6, 8, 10, 12, and 15. It is evident that the construction is possible for any value of n which can be found from these by multiplication by any power of 2. There are other special values of n for which such constructions are possible, the most interesting being $n = 17$.

49. The general form of De Moivre's Theorem. It follows from the results of the last section that if q is a positive integer then one of the values of $(\cos\theta + i\sin\theta)^{1/q}$ is

$$\cos(\theta/q) + i\sin(\theta/q).$$

Raising each of these expressions to the power p (where p is any integer positive or negative), we obtain the theorem that one of the values of $(\cos\theta + i\sin\theta)^{p/q}$ is $\cos(p\theta/q) + i\sin(p\theta/q)$, or that *if α is any rational number then one of the values of* $(\cos\theta + i\sin\theta)^\alpha$ *is*

$$\cos\alpha\theta + i\sin\alpha\theta.$$

This is a generalised form of De Moivre's Theorem (§ 45).

MISCELLANEOUS EXAMPLES ON CHAPTER III.

1. The condition that a triangle (xyz) should be equilateral is that

$$x^2 + y^2 + z^2 - yz - zx - xy = 0.$$

[Let XYZ be the triangle. The displacement $\overline{ZX}$ is $\overline{YZ}$ turned through an angle $\frac{2}{3}\pi$ in the positive or negative direction. Since $\operatorname{Cis}\frac{2}{3}\pi = \omega_3$, $\operatorname{Cis}(-\frac{2}{3}\pi) = 1/\omega_3 = \omega_3^2$, we have $x - z = (z - y)\omega_3$ or $x - z = (z - y)\omega_3^2$. Hence $x + y\omega_3 + z\omega_3^2 = 0$ or $x + y\omega_3^2 + z\omega_3 = 0$. The result follows from Exs. XXII. 3.]

2. If XYZ, $X'Y'Z'$ are two triangles, and

$$\overline{YZ} \cdot \overline{Y'Z'} = \overline{ZX} \cdot \overline{Z'X'} = \overline{XY} \cdot \overline{X'Y'},$$

then both triangles are equilateral. [From the equations

$$(y - z)(y' - z') = (z - x)(z' - x') = (x - y)(x' - y') = \kappa^2,$$

say, we deduce $\sum 1/(y' - z') = 0$, or $\sum x'^2 - \sum y'z' = 0$. Now apply the result of the last example.]

3. Similar triangles BCX, CAY, ABZ are described on the sides of a triangle ABC. Show that the centres of gravity of ABC, XYZ are coincident.

[We have $(x-c)/(b-c) = (y-a)/(c-a) = (z-b)/(a-b) = \lambda$, say. Express $\frac{1}{3}(x + y + z)$ in terms of a, b, c.]

4. If X, Y, Z are points on the sides of the triangle ABC, such that

$$BX/XC = CY/YA = AZ/ZB = r,$$

and if ABC, XYZ are similar, then either $r = 1$ or both triangles are equilateral.

5. If A, B, C, D are four points in a plane, then

$$AD \cdot BC \leqq BD \cdot CA + CD \cdot AB.$$

[Let z_1, z_2, z_3, z_4 be the complex numbers corresponding to A, B, C, D. Then we have identically

$$(x_1 - x_4)(x_2 - x_3) + (x_2 - x_4)(x_3 - x_1) + (x_3 - x_4)(x_1 - x_2) = 0.$$

Hence

$$\begin{aligned}|(x_1 - x_4)(x_2 - x_3)| &= |(x_2 - x_4)(x_3 - x_1) + (x_3 - x_4)(x_1 - x_2)| \\ &\leqq |(x_2 - x_4)(x_3 - x_1)| + |(x_3 - x_4)(x_1 - x_2)|.]\end{aligned}$$

6. Deduce Ptolemy's Theorem concerning cyclic quadrilaterals from the fact that the cross ratios of four concyclic points are real. [Use the same identity as in the last example.]

7. If $z^2 + z'^2 = 1$, then the points z, z' are ends of conjugate diameters of an ellipse whose foci are the points $1, -1$. [If CP, CD are conjugate semi-diameters of an ellipse and S, H its foci, then CD is parallel to the external bisector of the angle SPH, and $SP \cdot HP = CD^2$.]

8. Prove that $|a + b|^2 + |a - b|^2 = 2\{|a|^2 + |b|^2\}$. [This is the analytical equivalent of the geometrical theorem that, if M is the middle point of PQ, then $OP^2 + OQ^2 = 2OM^2 + 2MP^2$.]

9. Deduce from Ex. 8 that

$$|a + \sqrt{a^2 - b^2}| + |a - \sqrt{a^2 - b^2}| = |a + b| + |a - b|.$$

[If $a + \sqrt{a^2 - b^2} = z_1$, $a - \sqrt{a^2 - b^2} = z_2$, we have

$$|z_1|^2 + |z_2|^2 = \tfrac{1}{2}|z_1 + z_2|^2 + \tfrac{1}{2}|z_1 - z_2|^2 = 2|a|^2 + 2|a^2 - b^2|,$$

and so

$$(|z_1| + |z_2|)^2 = 2\{|a|^2 + |a^2 - b^2| + |b|^2\} = |a + b|^2 + |a - b|^2 + 2|a^2 - b^2|.$$

Another way of stating the result is: if z_1 and z_2 are the roots of $\alpha z^2 + 2\beta z + \gamma = 0$, then

$$|z_1| + |z_2| = (1/|\alpha|)\{(| - \beta + \sqrt{\alpha\gamma}|) + (| - \beta - \sqrt{\alpha\gamma}|)\}.]$$

10. Show that the necessary and sufficient conditions that both the roots of the equation $z^2 + az + b = 0$ should be of unit modulus are

$$|a| \leqq 2, \quad |b| = 1, \quad \operatorname{am} b = 2 \operatorname{am} a.$$

[The amplitudes have not necessarily their principal values.]

11. If $x^4+4a_1x^3+6a_2x^2+4a_3x+a_4=0$ is an equation with real coefficients and has two real and two complex roots, concyclic in the Argand diagram, then

$$a_3^2+a_1^2a_4+a_2^3-a_2a_4-2a_1a_2a_3=0.$$

12. The four roots of $a_0x^4+4a_1x^3+6a_2x^2+4a_3x+a_4=0$ will be harmonically related if

$$a_0a_3^2+a_1^2a_4+a_2^3-a_0a_2a_4-2a_1a_2a_3=0.$$

[Express $Z_{23,14}Z_{31,24}Z_{12,34}$, where $Z_{23,14}=(z_1-z_2)(z_3-z_4)+(z_1-z_3)(z_2-z_4)$ and z_1, z_2, z_3, z_4 are the roots of the equation, in terms of the coefficients.]

13. **Imaginary points and straight lines.** Let $ax+by+c=0$ be an equation with complex coefficients (which of course may be real in special cases).

If we give x any particular real or complex value, we can find the corresponding value of y. The aggregate of pairs of real or complex values of x and y which satisfy the equation is called an *imaginary straight line*; the pairs of values are called *imaginary points*, and are said *to lie on the line*. The values of x and y are called the *coordinates* of the point (x, y). When x and y are real, the point is called a *real point*: when a, b, c are all real (or can be made all real by division by a common factor), the line is called a *real line*. The points $x=\alpha+\beta i$, $y=\gamma+\delta i$ and $x=\alpha-\beta i$, $y=\gamma-\delta i$ are said to be *conjugate*; and so are the lines

$$(A+A'i)x+(B+B'i)y+C+C'i=0, \quad (A-A'i)x+(B-B'i)y+C-C'i=0.$$

Verify the following assertions:—every real line contains infinitely many pairs of conjugate imaginary points; an imaginary line in general contains one and only one real point; an imaginary line cannot contain a pair of conjugate imaginary points:—and find the conditions (a) that the line joining two given imaginary points should be real, and (b) that the point of intersection of two imaginary lines should be real.

14. Prove the identities

$$(x+y+z)(x+y\omega_3+z\omega_3^2)(x+y\omega_3^2+z\omega_3)=x^3+y^3+z^3-3xyz,$$
$$(x+y+z)(x+y\omega_5+z\omega_5^4)(x+y\omega_5^2+z\omega_5^3)(x+y\omega_5^3+z\omega_5^2)(x+y\omega_5^4+z\omega_5)$$
$$=x^5+y^5+z^5-5x^3yz+5xy^2z^2.$$

15. Solve the equations

$$x^3 - 3ax + (a^3 + 1) = 0, \quad x^5 - 5ax^3 + 5a^2x + (a^5 + 1) = 0.$$

16. If $f(x) = a_0 + a_1x + \cdots + a_kx^k$, then

$$\{f(x) + f(\omega x) + \cdots + f(\omega^{n-1}x)\}/n = a_0 + a_nx^n + a_{2n}x^{2n} + \cdots + a_{\lambda n}x^{\lambda n},$$

ω being any root of $x^n = 1$ (except $x = 1$), and λn the greatest multiple of n contained in k. Find a similar formula for $a_\mu + a_{\mu+n}x^n + a_{\mu+2n}x^{2n} + \ldots$.

17. If

$$(1 + x)^n = p_0 + p_1x + p_2x^2 + \ldots,$$

n being a positive integer, then

$$p_0 - p_2 + p_4 - \cdots = 2^{\frac{1}{2}n} \cos \tfrac{1}{4}n\pi, \quad p_1 - p_3 + p_5 - \cdots = 2^{\frac{1}{2}n} \sin \tfrac{1}{4}n\pi.$$

18. Sum the series

$$\frac{x}{2!(n-2)!} + \frac{x^2}{5!(n-5)!} + \frac{x^3}{8!(n-8)!} + \cdots + \frac{x^{n/3}}{(n-1)!},$$

n being a multiple of 3. (*Math. Trip.* 1899.)

19. If t is a complex number such that $|t| = 1$, then the point $x = (at + b)/(t - c)$ describes a circle as t varies, unless $|c| = 1$, when it describes a straight line.

20. If t varies as in the last example then the point $x = \frac{1}{2}\{at + (b/t)\}$ in general describes an ellipse whose foci are given by $x^2 = ab$, and whose axes are $|a| + |b|$ and $|a| - |b|$. But if $|a| = |b|$ then x describes the finite straight line joining the points $-\sqrt{ab}$, $\sqrt{ab}$.

21. Prove that if t is real and $z = t^2 - 1 + \sqrt{t^4 - t^2}$, then, when $t^2 < 1$, z is represented by a point which lies on the circle $x^2 + y^2 + x = 0$. Assuming that, when $t^2 > 1$, $\sqrt{t^4 - t^2}$ denotes the positive square root of $t^4 - t^2$, discuss the motion of the point which represents z, as t diminishes from a large positive value to a large negative value. (*Math. Trip.* 1912.)

22. The coefficients of the transformation $z = (aZ + b)/(cZ + d)$ are subject to the condition $ad - bc = 1$. Show that, if $c \neq 0$, there are two *fixed points* α, β, *i.e.* points unaltered by the transformation, except when $(a + d)^2 = 4$, when

there is only one fixed point α; and that in these two cases the transformation may be expressed in the forms

$$\frac{z-\alpha}{z-\beta} = K\frac{Z-\alpha}{Z-\beta}, \quad \frac{1}{z-\alpha} = \frac{1}{Z-\alpha} + K.$$

Show further that, if $c = 0$, there will be one fixed point α unless $a = d$, and that in these two cases the transformation may be expressed in the forms

$$z - \alpha = K(Z - \alpha), \quad z = Z + K.$$

Finally, if a, b, c, d are further restricted to positive integral values (including zero), show that the only transformations with less than two fixed points are of the forms $(1/z) = (1/Z) + K$, $z = Z + K$. (*Math. Trip.* 1911.)

23. Prove that the relation $z = (1 + Zi)/(Z + i)$ transforms the part of the axis of x between the points $z = 1$ and $z = -1$ into a semicircle passing through the points $Z = 1$ and $Z = -1$. Find all the figures that can be obtained from the originally selected part of the axis of x by successive applications of the transformation. (*Math. Trip.* 1912.)

24. If $z = 2Z + Z^2$ then the circle $|Z| = 1$ corresponds to a cardioid in the plane of z.

25. Discuss the transformation $z = \frac{1}{2}\{Z+(1/Z)\}$, showing in particular that to the circles $X^2 + Y^2 = \alpha^2$ correspond the confocal ellipses

$$\frac{x^2}{\left\{\dfrac{1}{2}\left(\alpha + \dfrac{1}{\alpha}\right)\right\}^2} + \frac{y^2}{\left\{\dfrac{1}{2}\left(\alpha - \dfrac{1}{\alpha}\right)\right\}^2} = 1.$$

26. If $(z + 1)^2 = 4/Z$ then the unit circle in the z-plane corresponds to the parabola $R\cos^2 \frac{1}{2}\Theta = 1$ in the Z-plane, and the inside of the circle to the outside of the parabola.

27. Show that, by means of the transformation $z = \{(Z - ci)/(Z + ci)\}^2$, the upper half of the z-plane may be made to correspond to the interior of a certain semicircle in the Z-plane.

28. If $z = Z^2 - 1$, then as z describes the circle $|z| = \kappa$, the two corresponding positions of Z each describe the Cassinian oval $\rho_1\rho_2 = \kappa$, where ρ_1, ρ_2 are the distances of Z from the points -1, 1. Trace the ovals for different values of κ.

29. Consider the relation $az^2 + 2hzZ + bZ^2 + 2gz + 2fZ + c = 0$. Show that there are two values of Z for which the corresponding values of z are equal, and *vice versa*. We call these the *branch points* in the Z and z-planes respectively. Show that, if z describes an ellipse whose foci are the branch points, then so does Z.

[We can, without loss of generality, take the given relation in the form

$$z^2 + 2zZ \cos\omega + Z^2 = 1:$$

the reader should satisfy himself that this is the case. The branch points in either plane are $\operatorname{cosec}\omega$ and $-\operatorname{cosec}\omega$. An ellipse of the form specified is given by

$$|z + \operatorname{cosec}\omega| + |z - \operatorname{cosec}\omega| = C,$$

where C is a constant. This is equivalent (Ex. 9) to

$$|z + \sqrt{z^2 - \operatorname{cosec}^2\omega}| + |z - \sqrt{z^2 - \operatorname{cosec}^2\omega}| = C.$$

Express this in terms of Z.]

30. If $z = aZ^m + bZ^n$, where m, n are positive integers and a, b real, then as Z describes the unit circle, z describes a hypo- or epi-cycloid.

31. Show that the transformation

$$z = \frac{(a + di)Z_0 + b}{cZ_0 - (a - di)},$$

where a, b, c, d are real and $a^2 + d^2 + bc > 0$, and Z_0 denotes the conjugate of Z, is equivalent to an inversion with respect to the circle

$$c(x^2 + y^2) - 2ax - 2dy - b = 0.$$

What is the geometrical interpretation of the transformation when

$$a^2 + d^2 + bc < 0?$$

32. The transformation

$$\frac{1 - z}{1 + z} = \left(\frac{1 - Z}{1 + Z}\right)^c,$$

where c is rational and $0 < c < 1$, transforms the circle $|z| = 1$ into the boundary of a circular lune of angle π/c.

CHAPTER IV

LIMITS OF FUNCTIONS OF A POSITIVE INTEGRAL VARIABLE

50. Functions of a positive integral variable. In Chapter II we discussed the notion of a function of a real variable x, and illustrated the discussion by a large number of examples of such functions. And the reader will remember that there was one important particular with regard to which the functions which we took as illustrations differed very widely. Some were defined for *all* values of x, some for *rational* values only, some for *integral* values only, and so on.

Consider, for example, the following functions: (i) x, (ii) $\sqrt{x}$, (iii) the denominator of x, (iv) the square root of the product of the numerator and the denominator of x, (v) the largest prime factor of x, (vi) the product of $\sqrt{x}$ and the largest prime factor of x, (vii) the xth prime number, (viii) the height measured in inches of convict x in Dartmoor prison.

Then the aggregates of values of x for which these functions are defined or, as we may say, the *fields of definition* of the functions, consist of (i) *all* values of x, (ii) *all positive* values of x, (iii) *all rational* values of x, (iv) *all positive rational* values of x, (v) *all integral* values of x, (vi), (vii) *all positive integral* values of x, (viii) a certain number of positive integral values of x, viz., 1, 2, …, N, where N is the total number of convicts at Dartmoor at a given moment of time.*

Now let us consider a function, such as (vii) above, which is defined for all positive integral values of x and no others. This function may be regarded from two slightly different points of view. We may consider it, as has so far been our custom, as a function of the real variable x defined for some only of the values of x, viz. positive integral values, and say that for all other values of x the definition fails. Or we may leave values of x

*In the last case N depends on the time, and convict x, where x has a definite value, is a different individual at different moments of time. Thus if we take different moments of time into consideration we have a simple example of a function $y = F(x, t)$ of two variables, defined for a certain range of values of t, viz. from the time of the establishment of Dartmoor prison to the time of its abandonment, and for a certain number of positive integral values of x, this number varying with t.

other than positive integral values entirely out of account, and regard our function as a function of the *positive integral variable* n, whose values are the positive integers

$$1,\ 2,\ 3,\ 4,\ \dots.$$

In this case we may write

$$y = \phi(n)$$

and regard y now as a function of n defined for all values of n.

It is obvious that any function of x defined for all values of x gives rise to a function of n defined for all values of n. Thus from the function $y = x^2$ we deduce the function $y = n^2$ by merely omitting from consideration all values of x other than positive integers, and the corresponding values of y. On the other hand from any function of n we can deduce any number of functions of x by merely assigning values to y, corresponding to values of x other than positive integral values, in any way we please.

51. Interpolation. The problem of determining a function of x which shall assume, for all positive integral values of x, values agreeing with those of a given function of n, is of extreme importance in higher mathematics. It is called the *problem of functional interpolation.*

Were the problem however merely that of finding *some* function of x to fulfil the condition stated, it would of course present no difficulty whatever. We could, as explained above, simply fill in the missing values as we pleased: we might indeed simply regard the given values of the function of n as *all* the values of the function of x and say that the definition of the latter function failed for all other values of x. But such purely theoretical solutions are obviously not what is usually wanted. What is usually wanted is some *formula* involving x (of as simple a kind as possible) which assumes the given values for $x = 1,\ 2,\ \dots$.

In some cases, especially when the function of n is itself defined by a formula, there is an obvious solution. If for example $y = \phi(n)$, where $\phi(n)$ is a function of n, such as n^2 or $\cos n\pi$, which would have a meaning even were n not a positive integer, we naturally take our function of x to be $y = \phi(x)$. But even in this very simple case it is easy to write down other almost equally obvious solutions of the problem. For example

$$y = \phi(x) + \sin x\pi$$

assumes the value $\phi(n)$ for $x = n$, since $\sin n\pi = 0$.

In other cases $\phi(n)$ may be defined by a formula, such as $(-1)^n$, which ceases to define for some values of x (as here in the case of fractional values of x with even denominators, or irrational values). But it may be possible to transform the formula in such a way that it does define for all values of x. In this case, for example,

$$(-1)^n = \cos n\pi,$$

if n is an integer, and the problem of interpolation is solved by the function $\cos x\pi$.

In other cases $\phi(x)$ may be defined for some values of x other than positive integers, but not for all. Thus from $y = n^n$ we are led to $y = x^x$. This expression has a meaning for some only of the remaining values of x. If for simplicity we confine ourselves to positive values of x, then x^x has a meaning for all rational values of x, in virtue of the definitions of fractional powers adopted in elementary algebra. But when x is *irrational* x^x has (so far as we are in a position to say at the present moment) no meaning at all. Thus in this case the problem of interpolation at once leads us to consider the question of extending our definitions in such a way that x^x shall have a meaning even when x is irrational. We shall see later on how the desired extension may be effected.

Again, consider the case in which

$$y = 1 \cdot 2 \dots n = n!.$$

In this case there is no obvious formula in x which reduces to $n!$ for $x = n$, as $x!$ means nothing for values of x other than the positive integers. This is a case in which attempts to solve the problem of interpolation have led to important advances in mathematics. For mathematicians have succeeded in discovering a function (the Gamma-function) which possesses the desired property and many other interesting and important properties besides.

52. Finite and infinite classes. Before we proceed further it is necessary to make a few remarks about certain ideas of an abstract and logical nature which are of constant occurrence in Pure Mathematics.

In the first place, the reader is probably familiar with the notion of **a class**. It is unnecessary to discuss here any logical difficulties which may be involved in the notion of a 'class': roughly speaking we may say that a class is the aggregate or collection of all the entities or objects which possess a certain property, simple or complex. Thus we have the class of British subjects, or members of Parliament, or positive integers, or real numbers.

Moreover, the reader has probably an idea of what is meant by a **finite** or **infinite** class. Thus the class of *British subjects* is a finite class: the aggregate of all British subjects, past, present, and future, has a finite number n, though of course we cannot tell at present the actual value of n. The class of *present British subjects*, on the other hand, has a number n which could be ascertained by counting, were the methods of the census effective enough.

On the other hand the class of positive integers is not finite but infinite. This may be expressed more precisely as follows. If n is any positive integer, such as 1000, 1,000,000 or any number we like to think of, then there are more than n positive integers. Thus, if the number we think of is 1,000,000, there are obviously at least 1,000,001 positive integers. Similarly the class of rational numbers, or of real numbers, is infinite. It is convenient to express this by saying that there are **an infinite number** of positive integers, or rational numbers, or real numbers. But the reader must be careful always to remember that by saying this we mean *simply* that the class in question has not a finite number of members such as 1000 or 1,000,000.

53. Properties possessed by a function of n for large values of n. We may now return to the 'functions of n' which we were discussing in §§ 50–51. They have many points of difference from the functions of x which we discussed in Chap. II. But there is one fundamental characteristic which the two classes of functions have in common: *the values of the variable for which they are defined form an infinite class.* It is this fact which forms the basis of all the considerations which follow and which, as we shall see in the next chapter, apply, *mutatis mutandis*, to functions of x as well.

Suppose that $\phi(n)$ is any function of n, and that P is any property which $\phi(n)$ may or may not have, such as that of being a positive integer or of being greater than 1. Consider, for each of the values $n = 1, 2, 3, \dots$, whether $\phi(n)$ has the property P or not. Then there are three possibilities:—

(a) $\phi(n)$ may have the property P for *all* values of n, or for all values of n except a finite number N of such values:

(b) $\phi(n)$ may have the property for *no* values of n, or only for a finite number N of such values:

(c) neither (a) nor (b) may be true.

If (b) is true, the values of n for which $\phi(n)$ has the property form a finite class. If (a) is true, the values of n for which $\phi(n)$ has not the property form a finite class. In the third case neither class is finite. Let us consider some particular cases.

(1) Let $\phi(n) = n$, and let P be the property of being a positive integer. Then $\phi(n)$ has the property P for all values of n.

If on the other hand P denotes the property of being a positive integer greater than or equal to 1000, then $\phi(n)$ has the property for all values of n except a finite number of values of n, viz. $1, 2, 3, \dots, 999$. In either of these cases (a) is true.

(2) If $\phi(n) = n$, and P is the property of being less than 1000, then (b) is true.

(3) If $\phi(n) = n$, and P is the property of being odd, then (c) is true. For $\phi(n)$ is odd if n is odd and even if n is even, and both the odd and the even values of n form an infinite class.

Example. Consider, in each of the following cases, whether (a), (b), or (c) is true:

(i) $\phi(n) = n$, P being the property of being a perfect square,

(ii) $\phi(n) = p_n$, where p_n denotes the nth prime number, P being the property of being odd,

(iii) $\phi(n) = p_n$, P being the property of being even,

(iv) $\phi(n) = p_n$, P being the property $\phi(n) > n$,

(v) $\phi(n) = 1 - (-1)^n(1/n)$, P being the property $\phi(n) < 1$,

(vi) $\phi(n) = 1 - (-1)^n(1/n)$, P being the property $\phi(n) < 2$,

(vii) $\phi(n) = 1000\{1 + (-1)^n\}/n$, P being the property $\phi(n) < 1$,

(viii) $\phi(n) = 1/n$, P being the property $\phi(n) < .001$,

(ix) $\phi(n) = (-1)^n/n$, P being the property $|\phi(n)| < .001$,

(x) $\phi(n) = 10{,}000/n$, or $(-1)^n 10{,}000/n$, P being either of the properties $\phi(n) < .001$ or $|\phi(n)| < .001$,

(xi) $\phi(n) = (n-1)/(n+1)$, P being the property $1 - \phi(n) < .0001$.

54. Let us now suppose that $\phi(n)$ and P are such that the assertion (a) is true, *i.e.* that $\phi(n)$ has the property P, if not for all values of n, at any rate for all values of n except a finite number N of such values. We may denote these exceptional values by

$$n_1,\ n_2,\ \dots,\ n_N.$$

There is of course no reason why these N values should be the *first* N values $1, 2, \dots, N$, though, as the preceding examples show, this is frequently the case in practice. But whether this is so or not we know that $\phi(n)$ has the property P if $n > n_N$. Thus the nth prime is odd if $n > 2$, $n = 2$ being the only exception to the statement; and $1/n < .001$ if $n > 1000$, the first 1000 values of n being the exceptions; and

$$1000\{1 + (-1)^n\}/n < 1$$

if $n > 2000$, the exceptional values being $2, 4, 6, \dots, 2000$. That is to say, in each of these cases the property is possessed *for all values of n from a definite value onwards.*

We shall frequently express this by saying that $\phi(n)$ has the property for **large**, or *very large*, or *all sufficiently large* values of n. Thus when we say that $\phi(n)$ *has the property P* (which will as a rule be a property expressed by some relation of inequality) *for large values of n*, what we mean is that we can determine some definite number, n_0 say, such that $\phi(n)$ has the property for all values of n greater than or equal to n_0. This number n_0, in the examples considered above, may be taken to be any number greater than n_N, the greatest of the exceptional numbers: it is most natural to take it to be $n_N + 1$.

Thus we may say that 'all large primes are odd', or that '$1/n$ is less than .001 for large values of n'. And the reader must make himself familiar with the use of the word *large* in statements of this kind. *Large* is in fact a word which, standing by itself, has no more absolute meaning in mathematics than in the language of common life. It is a truism that in common life a number which is large in one connection is small in another; 6 goals is a large score in a football match, but 6 runs is not a large score in a cricket match; and 400 runs is a large score, but £400 is not a large income: and so of course in mathematics *large* generally means *large enough*, and what is large enough for one purpose may not be large enough for another.

We know now what is meant by the assertion '$\phi(n)$ has the property P for large values of n'. It is with assertions of this kind that we shall be concerned throughout this chapter.

55. The phrase 'n tends to infinity'. There is a somewhat different way of looking at the matter which it is natural to adopt. Suppose that n assumes successively the values 1, 2, 3, ... The word 'successively' naturally suggests succession in time, and we may suppose n, if we like, to assume these values at successive moments of time (*e.g.* at the beginnings of successive seconds). Then as the seconds pass n gets larger and larger and there is no limit to the extent of its increase. However large a number we may think of (*e.g.* 2,147,483,647), a time will come when n has become larger than this number.

It is convenient to have a short phrase to express this unending growth of n, and we shall say that n **tends to infinity**, or $n \to \infty$, this last symbol being usually employed as an abbreviation for 'infinity'. The phrase 'tends to' like the word 'successively' naturally suggests the idea of change in time, and it is convenient to think of the variation of n as accomplished in time in the manner described above. This however is a mere matter of convenience. The variable n is a purely logical entity which has in itself nothing to do with time.

The reader cannot too strongly impress upon himself that when we say that n 'tends to ∞' we mean simply that n is supposed to assume a series of values which increase continually and without limit. **There is no**

number 'infinity': such an equation as

$$n = \infty$$

is as it stands *absolutely meaningless*: n cannot be equal to ∞, because 'equal to ∞' means nothing. So far in fact the symbol ∞ means nothing at all except in the one phrase 'tends to ∞', the meaning of which we have explained above. Later on we shall learn how to attach a meaning to other phrases involving the symbol ∞, but the reader will always have to bear in mind

(1) that ∞ *by itself* means nothing, although *phrases containing it* sometimes mean something,

(2) that in every case in which a phrase containing the symbol ∞ means something it will do so simply because we have previously attached a meaning to this particular phrase by means of a special definition.

Now it is clear that if $\phi(n)$ has the property P for large values of n, and if n 'tends to ∞', in the sense which we have just explained, then n will ultimately assume values large enough to ensure that $\phi(n)$ has the property P. And so another way of putting the question 'what properties has $\phi(n)$ for sufficiently large values of n?' is 'how does $\phi(n)$ behave as n tends to ∞?'

56. The behaviour of a function of n as n tends to infinity. We shall now proceed, in the light of the remarks made in the preceding sections, to consider the meaning of some kinds of statements which are perpetually occurring in higher mathematics. Let us consider, for example, the two following statements: (a) $1/n$ *is small for large values of* n, (b) $1 - (1/n)$ *is nearly equal to* 1 *for large values of* n. Obvious as they may seem, there is a good deal in them which will repay the reader's attention. Let us take (a) first, as being slightly the simpler.

We have already considered the statement '$1/n$ *is less than* $.01$ *for large values of* n'. This, we saw, means that the inequality $1/n < .01$ is true for all values of n greater than some definite value, in fact greater than 100. Similarly it is true that '$1/n$ *is less than* $.0001$ *for large values of* n': in fact $1/n < .0001$ if $n > 10{,}000$. And instead of .01 or .0001 we might take .000 001 or .000 000 01, or indeed any positive number we like.

It is obviously convenient to have some way of expressing the fact that *any* such statement as '$1/n$ *is less than* .01 *for large values of* n' is true, when we substitute for .01 any smaller number, such as .0001 or .000 001 or any other number we care to choose. And clearly we can do this by saying that '*however small* ϵ *may be* (provided of course it is positive), *then* $1/n < \epsilon$ *for sufficiently large values of* n'. That this is true is obvious. For $1/n < \epsilon$ if $n > 1/\epsilon$, so that our 'sufficiently large' values of n need only all be greater than $1/\epsilon$. The assertion is however a complex one, in that it really stands for the whole class of assertions which we obtain by giving to ϵ special values such as .01. And of course the smaller ϵ is, and the larger $1/\epsilon$, the larger must be the least of the 'sufficiently large' values of n: values which are sufficiently large when ϵ has one value are inadequate when it has a smaller.

The last statement italicised is what is really meant by the statement (a), that $1/n$ is small when n is large. Similarly (b) really means "*if* $\phi(n) = 1 - (1/n)$, *then the statement* '$1 - \phi(n) < \epsilon$ *for sufficiently large values of* n' *is true whatever positive value* (*such as* .01 *or* .0001) *we attribute to* ϵ". That the statement (b) is true is obvious from the fact that $1 - \phi(n) = 1/n$.

There is another way in which it is common to state the facts expressed by the assertions (a) and (b). This is suggested at once by § 55. Instead of saying '$1/n$ is small for large values of n' we say '$1/n$ tends to 0 as n tends to ∞'. Similarly we say that '$1 - (1/n)$ tends to 1 as n tends to ∞': and these statements are to be regarded as precisely equivalent to (a) and (b). Thus the statements

'$1/n$ is small when n is large',

'$1/n$ tends to 0 as n tends to ∞',

are equivalent to one another and to the more formal statement

'if ϵ is any positive number, however small, then $1/n < \epsilon$ for sufficiently large values of n', or to the still more formal statement

> 'if ϵ is any positive number, however small, then we can find a number n_0 such that $1/n < \epsilon$ for all values of n greater than or equal to n_0'.

The number n_0 which occurs in the last statement is of course a function of ϵ. We shall sometimes emphasize this fact by writing n_0 in the form $n_0(\epsilon)$.

The reader should imagine himself confronted by an opponent who questions the truth of the statement. He would name a series of numbers growing smaller and smaller. He might begin with .001. The reader would reply that $1/n < .001$ as soon as $n > 1000$. The opponent would be bound to admit this, but would try again with some smaller number, such as .000 000 1. The reader would reply that $1/n < .000\,000\,1$ as soon as $n > 10{,}000{,}000$: and so on. In this simple case it is evident that the reader would always have the better of the argument.

We shall now introduce yet another way of expressing this property of the function $1/n$. We shall say that '*the* **limit** *of* $1/n$ *as* n *tends to* ∞ *is* 0', a statement which we may express symbolically in the form

$$\lim_{n\to\infty} \frac{1}{n} = 0,$$

or simply $\lim(1/n) = 0$. We shall also sometimes write '$1/n \to 0$ as $n \to \infty$', which may be read '$1/n$ tends to 0 as n tends to ∞'; or simply '$1/n \to 0$'. In the same way we shall write

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right) = 1, \quad \lim \left(1 - \frac{1}{n}\right) = 1,$$

or $1 - (1/n) \to 1$.

57. Now let us consider a different example: let $\phi(n) = n^2$. Then 'n^2 *is large when* n *is large*'. This statement is equivalent to the more formal statements

> 'if Δ is any positive number, however large, then $n^2 > \Delta$ for sufficiently large values of n',

> 'we can find a number $n_0(\Delta)$ such that $n^2 > \Delta$ for all values of n greater than or equal to $n_0(\Delta)$'.

And it is natural in this case to say that 'n^2 tends to ∞ as n tends to ∞', or 'n^2 tends to ∞ with n', and to write

$$n^2 \to \infty.$$

Finally consider the function $\phi(n) = -n^2$. In this case $\phi(n)$ is large, but negative, when n is large, and we naturally say that '$-n^2$ tends to $-\infty$ as n tends to ∞' and write

$$-n^2 \to -\infty.$$

And the use of the symbol $-\infty$ in this sense suggests that it will sometimes be convenient to write $n^2 \to +\infty$ for $n^2 \to \infty$ and generally to use $+\infty$ instead of ∞, in order to secure greater uniformity of notation.

But we must once more repeat that in all these statements the symbols ∞, $+\infty$, $-\infty$ mean nothing whatever by themselves, and only acquire a meaning when they occur in certain special connections in virtue of the explanations which we have just given.

58. Definition of a limit. After the discussion which precedes the reader should be in a position to appreciate the general notion of a *limit.* Roughly we may say that $\phi(n)$ *tends to a limit* l *as* n *tends to* ∞ *if* $\phi(n)$ *is nearly equal to* l *when* n *is large.* But although the meaning of this statement should be clear enough after the preceding explanations, it is not, as it stands, precise enough to serve as a strict mathematical definition. It is, in fact, equivalent to a whole class of statements of the type '*for sufficiently large values of* n, $\phi(n)$ *differs from* l *by less than* ϵ'. This statement has to be true for $\epsilon = .01$ or $.0001$ or *any* positive number; and for any such value of ϵ it has to be true for *any* value of n after a certain definite value $n_0(\epsilon)$, though the smaller ϵ is the larger, as a rule, will be this value $n_0(\epsilon)$.

We accordingly frame the following formal definition:

Definition I. *The function $\phi(n)$ is said to tend to the limit l as n tends to ∞, if, however small be the positive number ϵ, $\phi(n)$ differs from l by less than ϵ for sufficiently large values of n; that is to say if, however small be the positive number ϵ, we can determine a number $n_0(\epsilon)$ corresponding to ϵ, such that $\phi(n)$ differs from l by less than ϵ for all values of n greater than or equal to $n_0(\epsilon)$.*

It is usual to denote the difference between $\phi(n)$ and l, taken positively, by $|\phi(n)-l|$. It is equal to $\phi(n)-l$ or to $l-\phi(n)$, whichever is positive, and agrees with the definition of the *modulus* of $\phi(n)-l$, as given in Chap. III, though at present we are only considering real values, positive or negative.

With this notation the definition may be stated more shortly as follows: '*if, given any positive number, ϵ, however small, we can find $n_0(\epsilon)$ so that $|\phi(n)-l| < \epsilon$ when $n \geqq n_0(\epsilon)$, then we say that $\phi(n)$ tends to the limit l as n tends to ∞, and write*

$$\lim_{n\to\infty} \phi(n) = l\text{'}.$$

Sometimes we may omit the '$n \to \infty$'; and sometimes it is convenient, for brevity, to write $\phi(n) \to l$.

The reader will find it instructive to work out, in a few simple cases, the explicit expression of n_0 as a function of ϵ. Thus if $\phi(n) = 1/n$ then $l = 0$, and the condition reduces to $1/n < \epsilon$ for $n \geqq n_0$, which is satisfied if $n_0 = 1 + [1/\epsilon]$.* There is one and only one case in which *the same* n_0 will do for *all* values of ϵ. If, from a certain value N of n onwards, $\phi(n)$ is constant, say equal to C, then it is evident that $\phi(n) - C = 0$ for $n \geqq N$, so that the inequality $|\phi(n) - C| < \epsilon$ is satisfied for $n \geqq N$ and all positive values of ϵ. And if $|\phi(n) - l| < \epsilon$ for $n \geqq N$ and all positive values of ϵ, then it is evident that $\phi(n) = l$ when $n \geqq N$, so that $\phi(n)$ is constant for all such values of n.

*Here and henceforward we shall use $[x]$ in the sense of Chap. II, *i.e.* as the greatest integer not greater than x.

59. The definition of a limit may be illustrated geometrically as follows. The graph of $\phi(n)$ consists of a number of points corresponding to the values $n = 1, 2, 3, \dots$.

Draw the line $y = l$, and the parallel lines $y = l - \epsilon$, $y = l + \epsilon$ at distance ϵ from it. Then

$$\lim_{n\to\infty} \phi(n) = l,$$

if, when once these lines have been drawn, no matter how close they may

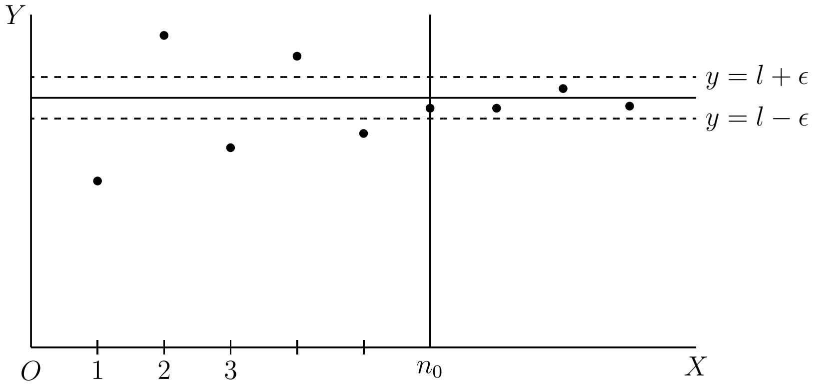

Fig. 27.

be together, we can always draw a line $x = n_0$, as in the figure, in such a way that the point of the graph on this line, and all points to the right of it, lie between them. We shall find this geometrical way of looking at our definition particularly useful when we come to deal with functions defined for all values of a real variable and not merely for positive integral values.

60. So much for functions of n which tend to a limit as n tends to ∞. We must now frame corresponding definitions for functions which, like the functions n^2 or $-n^2$, tend to positive or negative infinity. The reader should by now find no difficulty in appreciating the point of

Definition II. *The function $\phi(n)$ is said to tend to $+\infty$ (positive infinity) with n, if, when any number Δ, however large, is assigned, we* *can determine* $n_0(\Delta)$ *so that* $\phi(n) > \Delta$ *when* $n \geqq n_0(\Delta)$*; that is to say if, however large* Δ *may be,* $\phi(n) > \Delta$ *for sufficiently large values of* n*.*

Another, less precise, form of statement is '*if we can make* $\phi(n)$ *as large as we please by sufficiently increasing* n'. This is open to the objection that it obscures a fundamental point, viz. that $\phi(n)$ must be greater than Δ for *all* values of n such that $n \geqq n_0(\Delta)$, and not merely for *some* such values. But there is no harm in using this form of expression if we are clear what it means.

When $\phi(n)$ tends to $+\infty$ we write

$$\phi(n) \to +\infty.$$

We may leave it to the reader to frame the corresponding definition for functions which tend to negative infinity.

61. Some points concerning the definitions. The reader should be careful to observe the following points.

(1) We may obviously alter the values of $\phi(n)$ for any finite number of values of n, in any way we please, without in the least affecting the behaviour of $\phi(n)$ as n tends to ∞. For example $1/n$ tends to 0 as n tends to ∞. We may deduce any number of new functions from $1/n$ by altering a finite number of its values. For instance we may consider the function $\phi(n)$ which is equal to 3 for $n = 1, 2, 7, 11, 101, 107, 109, 237$ and equal to $1/n$ for all other values of n. For this function, just as for the original function $1/n$, $\lim \phi(n) = 0$. Similarly, for the function $\phi(n)$ which is equal to 3 if $n = 1, 2, 7, 11, 101, 107, 109, 237$, and to n^2 otherwise, it is true that $\phi(n) \to +\infty$.

(2) On the other hand we cannot as a rule alter an *infinite* number of the values of $\phi(n)$ without affecting fundamentally its behaviour as n tends to ∞. If for example we altered the function $1/n$ by changing its value to 1 whenever n is a multiple of 100, it would no longer be true that $\lim \phi(n) = 0$. So long as a finite number of values only were affected we could always choose the number n_0 of the definition so as to be greater than the greatest of the values of n for which $\phi(n)$ was altered. In the examples above, for instance, we could always take $n_0 > 237$, and indeed

we should be compelled to do so as soon as our imaginary opponent of § 56 had assigned a value of ϵ as small as 3 (in the first example) or a value of Δ as great as 3 (in the second). But now *however* large n_0 may be there will be greater values of n for which $\phi(n)$ has been altered.

(3) In applying the test of Definition I it is of course absolutely essential that we should have $|\phi(n) - l| < \epsilon$ not merely when $n = n_0$ but when $n \geqq n_0$, *i.e. for n_0 and for all larger values of n.* It is obvious, for example, that, if $\phi(n)$ is the function last considered, then given ϵ we can choose n_0 so that $|\phi(n)| < \epsilon$ when $n = n_0$: we have only to choose a sufficiently large value of n which is not a multiple of 100. But, when n_0 is thus chosen, it is not true that $|\phi(n)| < \epsilon$ when $n \geqq n_0$: all the multiples of 100 which are greater than n_0 are exceptions to this statement.

(4) If $\phi(n)$ is always greater than l, we can replace $|\phi(n) - l|$ by $\phi(n) - l$. Thus the test whether $1/n$ tends to the limit 0 as n tends to ∞ is simply whether $1/n < \epsilon$ when $n \geqq n_0$. If however $\phi(n) = (-1)^n/n$, then l is again 0, but $\phi(n) - l$ is sometimes positive and sometimes negative. In such a case we must state the condition in the form $|\phi(n) - l| < \epsilon$, for example, in this particular case, in the form $|\phi(n)| < \epsilon$.

(5) *The limit l may itself be one of the actual values of $\phi(n)$.* Thus if $\phi(n) = 0$ for all values of n, it is obvious that $\lim \phi(n) = 0$. Again, if we had, in (2) and (3) above, altered the value of the function, when n is a multiple of 100, to 0 instead of to 1, we should have obtained a function $\phi(n)$ which is equal to 0 when n is a multiple of 100 and to $1/n$ otherwise. The limit of this function as n tends to ∞ is still obviously zero. This limit is itself the value of the function for an infinite number of values of n, viz. all multiples of 100.

On the other hand *the limit itself need not (and in general will not) be the value of the function for any value of n.* This is sufficiently obvious in the case of $\phi(n) = 1/n$. The limit is zero; but the function is never equal to zero for any value of n.

The reader cannot impress these facts too strongly on his mind. **A limit is not a value of the function**: it is something quite distinct from these values, though it is defined by its relations to them and may possibly be equal to some of them. For the functions

$$\phi(n) = 0,\ 1,$$

the limit is equal to *all* the values of $\phi(n)$: for

$$\phi(n) = 1/n, \quad (-1)^n/n, \quad 1 + (1/n), \quad 1 + \{(-1)^n/n\}$$

it is not equal to *any* value of $\phi(n)$: for

$$\phi(n) = (\sin \tfrac{1}{2}n\pi)/n, \quad 1 + \{(\sin \tfrac{1}{2}n\pi)/n\}$$

(whose limits as n tends to ∞ are easily seen to be 0 and 1, since $\sin \frac{1}{2}n\pi$ is never numerically greater than 1) the limit is equal to the value which $\phi(n)$ assumes for all even values of n, but the values assumed for odd values of n are all different from the limit and from one another.

(6) A function may be always numerically very large when n is very large without tending either to $+\infty$ or to $-\infty$. A sufficient illustration of this is given by $\phi(n) = (-1)^n n$. A function can only tend to $+\infty$ or to $-\infty$ if, after a certain value of n, it maintains a constant sign.

Examples XXIII. Consider the behaviour of the following functions of n as n tends to ∞:

1. $\phi(n) = n^k$, where k is a positive or negative integer or rational fraction. If k is positive, then n^k tends to $+\infty$ with n. If k is negative, then $\lim n^k = 0$. If $k = 0$, then $n^k = 1$ for all values of n. Hence $\lim n^k = 1$.

The reader will find it instructive, even in so simple a case as this, to write down a formal proof that the conditions of our definitions are satisfied. Take for instance the case of $k > 0$. Let Δ be any assigned number, however large. We wish to choose n_0 so that $n^k > \Delta$ when $n \geqq n_0$. We have in fact only to take for n_0 any number greater than $\sqrt[k]{\Delta}$. If *e.g.* $k = 4$, then $n^4 > 10{,}000$ when $n \geqq 11$, $n^4 > 100{,}000{,}000$ when $n \geqq 101$, and so on.

2. $\phi(n) = p_n$, where p_n is the nth prime number. If there were only a finite number of primes then $\phi(n)$ would be defined only for a finite number of values of n. There are however, as was first shown by Euclid, infinitely many primes. Euclid's proof is as follows. If there are only a finite number of primes, let them

be 1, 2, 3, 5, 7, 11, ... N. Consider the number $1+(1\cdot 2\cdot 3\cdot 5\cdot 7\cdot 11\ldots N)$. This number is evidently not divisible by any of 2, 3, 5, ... N, since the remainder when it is divided by any of these numbers is 1. It is therefore not divisible by any prime save 1, and is therefore itself prime, which is contrary to our hypothesis.

It is moreover obvious that $\phi(n) > n$ for all values of n (save $n = 1$, 2, 3). Hence $\phi(n) \to +\infty$.

3. Let $\phi(n)$ be the number of primes less than n. Here again $\phi(n) \to +\infty$.

4. $\phi(n) = [\alpha n]$, where α is any positive number. Here

$$\phi(n) = 0 \quad (0 \leqq n < 1/\alpha), \qquad \phi(n) = 1 \quad (1/\alpha \leqq n < 2/\alpha),$$

and so on; and $\phi(n) \to +\infty$.

5. If $\phi(n) = 1{,}000{,}000/n$, then $\lim \phi(n) = 0$: and if $\psi(n) = n/1{,}000{,}000$, then $\psi(n) \to +\infty$. These conclusions are in no way affected by the fact that at first $\phi(n)$ is much larger than $\psi(n)$, being in fact larger until $n = 1{,}000{,}000$.

6. $\phi(n) = 1/\{n-(-1)^n\}$, $n-(-1)^n$, $n\{1-(-1)^n\}$. The first function tends to 0, the second to $+\infty$, the third does not tend either to a limit or to $+\infty$.

7. $\phi(n) = (\sin n\theta\pi)/n$, where θ is any real number. Here $|\phi(n)| < 1/n$, since $|\sin n\theta\pi| \leqq 1$, and $\lim \phi(n) = 0$.

8. $\phi(n) = (\sin n\theta\pi)/\sqrt{n}$, $(a\cos^2 n\theta + b\sin^2 n\theta)/n$, where a and b are any real numbers.

9. $\phi(n) = \sin n\theta\pi$. If θ is integral then $\phi(n) = 0$ for all values of n, and therefore $\lim \phi(n) = 0$.

Next let θ be rational, *e.g.* $\theta = p/q$, where p and q are positive integers. Let $n = aq + b$ where a is the quotient and b the remainder when n is divided by q. Then $\sin(np\pi/q) = (-1)^{ap}\sin(bp\pi/q)$. Suppose, for example, p even; then, as n increases from 0 to $q-1$, $\phi(n)$ takes the values

$$0, \quad \sin(p\pi/q), \quad \sin(2p\pi/q), \ \ldots \quad \sin\{(q-1)p\pi/q\}.$$

When n increases from q to $2q-1$ these values are repeated; and so also as n goes from $2q$ to $3q-1$, $3q$ to $4q-1$, and so on. Thus the values of $\phi(n)$ form *a perpetual cyclic repetition of a finite series of different values*. It is evident that when this is the case $\phi(n)$ cannot tend to a limit, nor to $+\infty$, nor to $-\infty$, as n tends to infinity.

The case in which θ is irrational is a little more difficult. It is discussed in the next set of examples.

62. Oscillating Functions. DEFINITION. *When $\phi(n)$ does not tend to a limit, nor to $+\infty$, nor to $-\infty$, as n tends to ∞, we say that $\phi(n)$* **oscillates** *as n tends to ∞.*

A function $\phi(n)$ certainly oscillates if its values form, as in the case considered in the last example above, a continual repetition of a cycle of values. But of course it may oscillate without possessing this peculiarity. Oscillation is defined in a purely negative manner: a function oscillates when it does not do certain other things.

The simplest example of an oscillatory function is given by

$$\phi(n) = (-1)^n,$$

which is equal to $+1$ when n is even and to -1 when n is odd. In this case the values recur cyclically. But consider

$$\phi(n) = (-1)^n + (1/n),$$

the values of which are

$$-1 + 1, \quad 1 + (1/2), \quad -1 + (1/3), \quad 1 + (1/4), \quad -1 + (1/5), \ \ldots.$$

When n is large every value is nearly equal to $+1$ or -1, and obviously $\phi(n)$ does not tend to a limit or to $+\infty$ or to $-\infty$, and therefore it oscillates: but the values do not recur. It is to be observed that in this case every value of $\phi(n)$ is numerically less than or equal to $3/2$. Similarly

$$\phi(n) = (-1)^n 100 + (1000/n)$$

oscillates. When n is large, every value is nearly equal to 100 or to -100. The numerically greatest value is 900 (for $n = 1$). But now consider $\phi(n) = (-1)^n n$, the values of which are -1, 2, -3, 4, -5, ... This function oscillates, for it does not tend to a limit, nor to $+\infty$, nor to $-\infty$. And in this case we cannot assign any limit beyond which the numerical value of the terms does not rise. The distinction between these two examples suggests a further definition.

DEFINITION. *If $\phi(n)$ oscillates as n tends to ∞, then $\phi(n)$ will be said to* **oscillate finitely** *or* **infinitely** *according as it is or is not possible to* *assign a number K such that all the values of $\phi(n)$ are numerically less than K, i.e. $|\phi(n)| < K$ for all values of n.*

These definitions, as well as those of §§ 58 and 60, are further illustrated in the following examples.

Examples XXIV. Consider the behaviour as n tends to ∞ of the following functions:

1. $(-1)^n$, $5 + 3(-1)^n$, $(1{,}000{,}000/n) + (-1)^n$, $1{,}000{,}000(-1)^n + (1/n)$.
2. $(-1)^n n$, $1{,}000{,}000 + (-1)^n n$.
3. $1{,}000{,}000 - n$, $(-1)^n(1{,}000{,}000 - n)$.
4. $n\{1 + (-1)^n\}$. In this case the values of $\phi(n)$ are

$$0,\quad 4,\quad 0,\quad 8,\quad 0,\quad 12,\quad 0,\quad 16,\ \dots.$$

The odd terms are all zero and the even terms tend to $+\infty$: $\phi(n)$ oscillates infinitely.

5. $n^2 + (-1)^n 2n$. The second term oscillates infinitely, but the first is very much larger than the second when n is large. In fact $\phi(n) \geqq n^2 - 2n$ and $n^2 - 2n = (n-1)^2 - 1$ is greater than any assigned value Δ if $n > 1 + \sqrt{\Delta + 1}$. Thus $\phi(n) \to +\infty$. It should be observed that in this case $\phi(2k+1)$ is always less than $\phi(2k)$, so that the function progresses to infinity by a continual series of steps forwards and backwards. It does not however 'oscillate' according to our definition of the term.

6. $n^2\{1 + (-1)^n\}$, $(-1)^n n^2 + n$, $n^3 + (-1)^n n^2$.

7. $\sin n\theta\pi$. We have already seen (Exs. XXIII. 9) that $\phi(n)$ oscillates finitely when θ is rational, unless θ is an integer, when $\phi(n) = 0$, $\phi(n) \to 0$.

The case in which θ is irrational is a little more difficult. But it is not difficult to see that $\phi(n)$ still oscillates finitely. We can without loss of generality suppose $0 < \theta < 1$. In the first place $|\phi(n)| < 1$. Hence $\phi(n)$ must oscillate finitely or tend to a limit. We shall consider whether the second alternative is really possible. Let us suppose that

$$\lim \sin n\theta\pi = l.$$

Then, however small ϵ may be, we can choose n_0 so that $\sin n\theta\pi$ lies between $l - \epsilon$ and $l + \epsilon$ for all values of n greater than or equal to n_0. Hence $\sin(n+1)\theta\pi - \sin n\theta\pi$ is numerically less than 2ϵ for all such values of n, and so $|\sin\frac{1}{2}\theta\pi\cos(n+\frac{1}{2})\theta\pi| < \epsilon$.

Hence

$$\cos(n + \tfrac{1}{2})\theta\pi = \cos n\theta\pi \cos \tfrac{1}{2}\theta\pi - \sin n\theta\pi \sin \tfrac{1}{2}\theta\pi$$

must be numerically less than $\epsilon/|\sin \frac{1}{2}\theta\pi|$. Similarly

$$\cos(n - \tfrac{1}{2})\theta\pi = \cos n\theta\pi \cos \tfrac{1}{2}\theta\pi + \sin n\theta\pi \sin \tfrac{1}{2}\theta\pi$$

must be numerically less than $\epsilon/|\sin \frac{1}{2}\theta\pi|$; and so each of $\cos n\theta\pi \cos \frac{1}{2}\theta\pi$, $\sin n\theta\pi \sin \frac{1}{2}\theta\pi$ must be numerically less than $\epsilon/|\sin \frac{1}{2}\theta\pi|$. That is to say, $\cos n\theta\pi \cos \frac{1}{2}\theta\pi$ is very small if n is large, and this can only be the case if $\cos n\theta\pi$ is very small. Similarly $\sin n\theta\pi$ must be very small, so that l must be zero. But it is impossible that $\cos n\theta\pi$ and $\sin n\theta\pi$ can *both* be very small, as the sum of their squares is unity. Thus the hypothesis that $\sin n\theta\pi$ tends to a limit l is impossible, and therefore $\sin n\theta\pi$ oscillates as n tends to ∞.

The reader should consider with particular care the argument '$\cos n\theta\pi \cos \frac{1}{2}\theta\pi$ is very small, and this can only be the case if $\cos n\theta\pi$ is very small'. Why, he may ask, should it not be the other factor $\cos \frac{1}{2}\theta\pi$ which is 'very small'? The answer is to be found, of course, in the meaning of the phrase 'very small' as used in this connection. When we say '$\phi(n)$ is very small' for large values of n, we mean that we can choose n_0 so that $\phi(n)$ is numerically smaller than *any* assigned number, if $n \geqq n_0$. Such an assertion is palpably absurd when made of a *fixed* number such as $\cos \frac{1}{2}\theta\pi$, which is not zero.

Prove similarly that $\cos n\theta\pi$ oscillates finitely, unless θ is an even integer.

8. $\sin n\theta\pi + (1/n)$, $\sin n\theta\pi + 1$, $\sin n\theta\pi + n$, $(-1)^n \sin n\theta\pi$.

9. $a\cos n\theta\pi + b\sin n\theta\pi$, $\sin^2 n\theta\pi$, $a\cos^2 n\theta\pi + b\sin^2 n\theta\pi$.

10. $a + bn + (-1)^n(c + dn) + e\cos n\theta\pi + f\sin n\theta\pi$.

11. $n\sin n\theta\pi$. If θ is integral, then $\phi(n) = 0$, $\phi(n) \to 0$. If θ is rational but not integral, or irrational, then $\phi(n)$ oscillates infinitely.

12. $n(a\cos^2 n\theta\pi + b\sin^2 n\theta\pi)$. In this case $\phi(n)$ tends to $+\infty$ if a and b are both positive, but to $-\infty$ if both are negative. Consider the special cases in which $a = 0$, $b > 0$, or $a > 0$, $b = 0$, or $a = 0$, $b = 0$. If a and b have opposite signs $\phi(n)$ generally oscillates infinitely. Consider any exceptional cases.

13. $\sin(n^2\theta\pi)$. If θ is integral, then $\phi(n) \to 0$. Otherwise $\phi(n)$ oscillates finitely, as may be shown by arguments similar to though more complex than those used in Exs. XXIII. 9 and XXIV. 7.*

*See Bromwich's *Infinite Series*, p. 485.

14. $\sin(n!\,\theta\pi)$. If θ has a rational value p/q, then $n!\,\theta$ is certainly integral for all values of n greater than or equal to q. Hence $\phi(n) \to 0$. The case in which θ is irrational cannot be dealt with without the aid of considerations of a much more difficult character.

15. $\cos(n!\,\theta\pi)$, $a\cos^2(n!\,\theta\pi) + b\sin^2(n!\,\theta\pi)$, where θ is rational.

16. $an - [bn]$, $(-1)^n(an - [bn])$.

17. $[\sqrt{n}]$, $(-1)^n[\sqrt{n}]$, $\sqrt{n} - [\sqrt{n}]$.

18. *The smallest prime factor of n.* When n is a prime, $\phi(n) = n$. When n is even, $\phi(n) = 2$. Thus $\phi(n)$ oscillates infinitely.

19. *The largest prime factor of n.*

20. *The number of days in the year n* A.D.

Examples XXV. 1. If $\phi(n) \to +\infty$ and $\psi(n) \geqq \phi(n)$ for all values of n, then $\psi(n) \to +\infty$.

2. If $\phi(n) \to 0$, and $|\psi(n)| \leqq |\phi(n)|$ for all values of n, then $\psi(n) \to 0$.

3. If $\lim |\phi(n)| = 0$, then $\lim \phi(n) = 0$.

4. If $\phi(n)$ tends to a limit or oscillates finitely, and $|\psi(n)| \leqq |\phi(n)|$ when $n \geqq n_0$, then $\psi(n)$ tends to a limit or oscillates finitely.

5. If $\phi(n)$ tends to $+\infty$, or to $-\infty$, or oscillates infinitely, and

$$|\psi(n)| \geqq |\phi(n)|$$

when $n \geqq n_0$, then $\psi(n)$ tends to $+\infty$ or to $-\infty$ or oscillates infinitely.

6. 'If $\phi(n)$ oscillates and, however great be n_0, we can find values of n greater than n_0 for which $\psi(n) > \phi(n)$, and values of n greater than n_0 for which $\psi(n) < \phi(n)$, then $\psi(n)$ oscillates'. Is this true? If not give an example to the contrary.

7. If $\phi(n) \to l$ as $n \to \infty$, then also $\phi(n+p) \to l$, p being any fixed integer. [This follows at once from the definition. Similarly we see that if $\phi(n)$ tends to $+\infty$ or $-\infty$ or oscillates so also does $\phi(n+p)$.]

8. The same conclusions hold (except in the case of oscillation) if p varies with n but is always numerically less than a fixed positive integer N; or if p varies with n in any way, so long as it is always positive.

9. Determine the least value of n_0 for which it is true that

$$(a)\ n^2 + 2n > 999{,}999 \quad (n \geqq n_0), \qquad (b)\ n^2 + 2n > 1{,}000{,}000 \quad (n \geqq n_0).$$

10. Determine the least value of n_0 for which it is true that

$$(a)\ n+(-1)^n > 1000 \quad (n \geqq n_0), \qquad (b)\ n+(-1)^n > 1{,}000{,}000 \quad (n \geqq n_0).$$

11. Determine the least value of n_0 for which it is true that

$$(a)\ n^2+2n > \Delta \quad (n \geqq n_0), \qquad (b)\ n+(-1)^n > \Delta \quad (n \geqq n_0),$$

Δ being any positive number.

[(*a*) $n_0 = [\sqrt{\Delta+1}]$: (*b*) $n_0 = 1+[\Delta]$ or $2+[\Delta]$, according as $[\Delta]$ is odd or even, *i.e.* $n_0 = 1+[\Delta]+\frac{1}{2}\{1+(-1)^{[\Delta]}\}$.]

12. Determine the least value of n_0 such that

$$(a)\ n/(n^2+1) < .0001, \qquad (b)\ (1/n)+\{(-1)^n/n^2\} < .000\,01,$$

when $n \geqq n_0$. [Let us take the latter case. In the first place

$$(1/n)+\{(-1)^n/n^2\} \leqq (n+1)/n^2,$$

and it is easy to see that the least value of n_0, such that $(n+1)/n^2 < .000\,001$ when $n \geqq n_0$, is 1,000,002. But the inequality given is satisfied by $n = 1{,}000{,}001$, and this is the value of n_0 required.]

63. Some general theorems with regard to limits. A. **The behaviour of the sum of two functions whose behaviour is known.**

THEOREM I. *If $\phi(n)$ and $\psi(n)$ tend to limits a, b, then $\phi(n)+\psi(n)$ tends to the limit $a+b$.*

This is almost obvious.* The argument which the reader will at once form in his mind is roughly this: 'when n is large, $\phi(n)$ is nearly equal to a

*There is a certain ambiguity in this phrase which the reader will do well to notice. When one says 'such and such a theorem is almost obvious' one may mean one or other of two things. One may mean 'it is difficult to doubt the truth of the theorem', 'the theorem is such as common-sense instinctively accepts', as it accepts, for example, the truth of the propositions '$2+2=4$' or 'the base-angles of an isosceles triangle are equal'. That a theorem is 'obvious' in this sense does not prove that it is true, since the most confident of the intuitive judgments of common sense are often found to be mistaken; and even if the theorem is true, the fact that it is also 'obvious' is no reason for not proving it, if a proof can be found. The object of mathematics is to prove that and $\psi(n)$ to b, and therefore their sum is nearly equal to $a+b$'. It is well to state the argument quite formally, however.

Let ϵ be any assigned positive number (*e.g.* .001, .000 000 1, ...). We require to show that a number n_0 can be found such that

$$|\phi(n)+\psi(n)-a-b| < \epsilon, \tag{1}$$

when $n \geqq n_0$. Now by a proposition proved in Chap. III (more generally indeed than we need here) the modulus of the sum of two numbers is less than or equal to the sum of their moduli. Thus

$$|\phi(n)+\psi(n)-a-b| \leqq |\phi(n)-a| + |\psi(n)-b|.$$

It follows that the desired condition will certainly be satisfied if n_0 can be so chosen that

$$|\phi(n)-a| + |\psi(n)-b| < \epsilon, \tag{2}$$

when $n \geqq n_0$. But this is certainly the case. For since $\lim \phi(n) = a$ we can, by the definition of a limit, find n_1 so that $|\phi(n)-a| < \epsilon'$ when $n \geqq n_1$, and this however small ϵ' may be. Nothing prevents our taking $\epsilon' = \frac{1}{2}\epsilon$, so that $|\phi(n)-a| < \frac{1}{2}\epsilon$ when $n \geqq n_1$. Similarly we can find n_2 so that $|\psi(n)-b| < \frac{1}{2}\epsilon$ when $n \geqq n_2$. Now take n_0 to be *the greater of the two numbers* n_1, n_2. Then $|\phi(n)-a| < \frac{1}{2}\epsilon$ and $|\psi(n)-b| < \frac{1}{2}\epsilon$ when $n \geqq n_0$, and therefore (2) is satisfied and the theorem is proved.

The argument may be concisely stated thus: since $\lim \phi(n) = a$ and $\lim \psi(n) = b$, we can choose n_1, n_2 so that

$$|\phi(n)-a| < \tfrac{1}{2}\epsilon \quad (n \geqq n_1), \qquad |\psi(n)-b| < \tfrac{1}{2}\epsilon \quad (n \geqq n_2);$$

---

certain premises imply certain conclusions; and the fact that the conclusions may be as 'obvious' as the premises never detracts from the necessity, and often not even from the interest of the proof.

But sometimes (as for example here) we mean by 'this is almost obvious' something quite different from this. We mean 'a moment's reflection should not only convince the reader of the truth of what is stated, but should also suggest to him the general lines of a rigorous proof'. And often, when a statement is 'obvious' in this sense, one may well omit the proof, not because the proof is in any sense unnecessary, but because it is a waste of time and space to state in detail what the reader can easily supply for himself.

and then, if n is not less than either n_1 or n_2,

$$|\phi(n) + \psi(n) - a - b| \leqq |\phi(n) - a| + |\psi(n) - b| < \epsilon;$$

and therefore

$$\lim\{\phi(n) + \psi(n)\} = a + b.$$

64. Results subsidiary to Theorem I. The reader should have no difficulty in verifying the following subsidiary results.

1. *If $\phi(n)$ tends to a limit, but $\psi(n)$ tends to $+\infty$ or to $-\infty$ or oscillates finitely or infinitely, then $\phi(n) + \psi(n)$ behaves like $\psi(n)$.*

2. *If $\phi(n) \to +\infty$, and $\psi(n) \to +\infty$ or oscillates finitely, then $\phi(n) + \psi(n) \to +\infty$.*

In this statement we may obviously change $+\infty$ into $-\infty$ throughout.

3. *If $\phi(n) \to +\infty$ and $\psi(n) \to -\infty$, then $\phi(n)+\psi(n)$ may tend either to a limit or to $+\infty$ or to $-\infty$ or may oscillate either finitely or infinitely.*

These five possibilities are illustrated in order by (i) $\phi(n) = n$, $\psi(n) = -n$, (ii) $\phi(n) = n^2$, $\psi(n) = -n$, (iii) $\phi(n) = n$, $\psi(n) = -n^2$, (iv) $\phi(n) = n + (-1)^n$, $\psi(n) = -n$, (v) $\phi(n) = n^2 + (-1)^n n$, $\psi(n) = -n^2$. The reader should construct additional examples of each case.

4. *If $\phi(n) \to +\infty$ and $\psi(n)$ oscillates infinitely, then $\phi(n)+\psi(n)$ may tend to $+\infty$ or oscillate infinitely, but cannot tend to a limit, or to $-\infty$, or oscillate finitely.*

For $\psi(n) = \{\phi(n) + \psi(n)\} - \phi(n)$; and, if $\phi(n) + \psi(n)$ behaved in any of the three last ways, it would follow, from the previous results, that $\psi(n) \to -\infty$, which is not the case. As examples of the two cases which are possible, consider (i) $\phi(n) = n^2$, $\psi(n) = (-1)^n n$, (ii) $\phi(n) = n$, $\psi(n) = (-1)^n n^2$. Here again the signs of $+\infty$ and $-\infty$ may be permuted throughout.

5. *If $\phi(n)$ and $\psi(n)$ both oscillate finitely, then $\phi(n)+\psi(n)$ must tend to a limit or oscillate finitely.*

As examples take

$$\text{(i)}\quad \phi(n) = (-1)^n,\quad \psi(n) = (-1)^{n+1},\qquad \text{(ii)}\quad \phi(n) = \psi(n) = (-1)^n.$$

6. *If $\phi(n)$ oscillates finitely, and $\psi(n)$ infinitely, then $\phi(n) + \psi(n)$ oscillates infinitely.*

For $\phi(n)$ is in absolute value always less than a certain constant, say K. On the other hand $\psi(n)$, since it oscillates infinitely, must assume values numerically greater than any assignable number (*e.g.* $10K$, $100K$, ...). Hence $\phi(n) + \psi(n)$ must assume values numerically greater than any assignable number (*e.g.* $9K$, $99K$, ...). Hence $\phi(n) + \psi(n)$ must either tend to $+\infty$ or $-\infty$ or oscillate infinitely. But if it tended to $+\infty$ then

$$\psi(n) = \{\phi(n) + \psi(n)\} - \phi(n)$$

would also tend to $+\infty$, in virtue of the preceding results. Thus $\phi(n) + \psi(n)$ cannot tend to $+\infty$, nor, for similar reasons, to $-\infty$: hence it oscillates infinitely.

7. *If both $\phi(n)$ and $\psi(n)$ oscillate infinitely, then $\phi(n)+\psi(n)$ may tend to a limit, or to $+\infty$, or to $-\infty$, or oscillate either finitely or infinitely.*

Suppose, for instance, that $\phi(n) = (-1)^n n$, while $\psi(n)$ is in turn each of the functions $(-1)^{n+1}n$, $\{1 + (-1)^{n+1}\}n$, $-\{1 + (-1)^n\}n$, $(-1)^{n+1}(n + 1)$, $(-1)^n n$. We thus obtain examples of all five possibilities.

The results 1–7 cover all the cases which are really distinct. Before passing on to consider the product of two functions, we may point out that the result of Theorem I may be immediately extended to the sum of three or more functions which tend to limits as $n \to \infty$.

65. B. **The behaviour of the product of two functions whose behaviour is known.** We can now prove a similar set of theorems concerning the product of two functions. The principal result is the following.

THEOREM II. *If* $\lim \phi(n) = a$ *and* $\lim \psi(n) = b$, *then*

$$\lim \phi(n)\psi(n) = ab.$$

Let

$$\phi(n) = a + \phi_1(n), \quad \psi(n) = b + \psi_1(n),$$

so that $\lim \phi_1(n) = 0$ and $\lim \psi_1(n) = 0$. Then

$$\phi(n)\psi(n) = ab + a\psi_1(n) + b\phi_1(n) + \phi_1(n)\psi_1(n).$$

Hence the numerical value of the difference $\phi(n)\psi(n) - ab$ is certainly not greater than the sum of the numerical values of $a\psi_1(n)$, $b\phi_1(n)$, $\phi_1(n)\psi_1(n)$. From this it follows that

$$\lim\{\phi(n)\psi(n) - ab\} = 0,$$

which proves the theorem.

The following is a strictly formal proof. We have

$$|\phi(n)\psi(n) - ab| \leqq |a\psi_1(n)| + |b\phi_1(n)| + |\phi_1(n)||\psi_1(n)|.$$

Assuming that neither a nor b is zero, we may choose n_0 so that

$$|\phi_1(n)| < \tfrac{1}{3}\epsilon/|b|, \quad |\psi_1(n)| < \tfrac{1}{3}\epsilon/|a|,$$

when $n \geqq n_0$. Then

$$|\phi(n)\psi(n) - ab| < \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon + \{\tfrac{1}{9}\epsilon^2/(|a||b|)\},$$

which is certainly less than ϵ if $\epsilon < \frac{1}{3}|a||b|$. That is to say we can choose n_0 so that $|\phi(n)\psi(n) - ab| < \epsilon$ when $n \geqq n_0$, and so the theorem follows. The reader should supply a proof for the case in which at least one of a and b is zero.

We need hardly point out that this theorem, like Theorem I, may be immediately extended to the product of any number of functions of n. There is also a series of subsidiary theorems concerning products analogous to those stated in § 64 for sums. We must distinguish now *six* different ways in which $\phi(n)$ may behave as n tends to ∞. It may (1) tend to a limit *other than zero*, (2) tend to zero, (3a) tend to $+\infty$, (3b) tend to $-\infty$, (4) oscillate finitely, (5) oscillate infinitely. It is not necessary, as a rule, to take account separately of (3a) and (3b), as the results for one case may be deduced from those for the other by a change of sign.

To state these subsidiary theorems at length would occupy more space than we can afford. We select the two which follow as examples, leaving the verification of them to the reader. He will find it an instructive exercise to formulate some of the remaining theorems himself.

(i) *If $\phi(n) \to +\infty$ and $\psi(n)$ oscillates finitely, then $\phi(n)\psi(n)$ must tend to $+\infty$ or to $-\infty$ or oscillate infinitely.*

Examples of these three possibilities may be obtained by taking $\phi(n)$ to be n and $\psi(n)$ to be one of the three functions $2 + (-1)^n$, $-2 - (-1)^n$, $(-1)^n$.

(ii) *If $\phi(n)$ and $\psi(n)$ oscillate finitely, then $\phi(n)\psi(n)$ must tend to a limit* (*which may be zero*) *or oscillate finitely.*

For examples, take (a) $\phi(n) = \psi(n) = (-1)^n$, ($b$) $\phi(n) = 1 + (-1)^n$, $\psi(n) = 1 - (-1)^n$, and (c) $\phi(n) = \cos\frac{1}{3}n\pi$, $\psi(n) = \sin\frac{1}{3}n\pi$.

A particular case of Theorem II which is important is that in which $\psi(n)$ is constant. The theorem then asserts simply that $\lim k\phi(n) = ka$ if $\lim \phi(n) = a$. To this we may join the subsidiary theorem that if $\phi(n) \to +\infty$ then $k\phi(n) \to +\infty$ or $k\phi(n) \to -\infty$, according as k is positive or negative, unless $k = 0$, when of course $k\phi(n) = 0$ for all values of n and $\lim k\phi(n) = 0$. And if $\phi(n)$ oscillates finitely or infinitely, then so does $k\phi(n)$, unless $k = 0$.

66. C. **The behaviour of the difference or quotient of two functions whose behaviour is known.** There is, of course, a similar set of theorems for the difference of two given functions, which are obvious corollaries from what precedes. In order to deal with the quotient

$$\frac{\phi(n)}{\psi(n)},$$

we begin with the following theorem.

THEOREM III. *If* $\lim \phi(n) = a$, *and a is not zero, then*

$$\lim \frac{1}{\phi(n)} = \frac{1}{a}.$$

Let

$$\phi(n) = a + \phi_1(n),$$

so that $\lim \phi_1(n) = 0$. Then

$$\left|\frac{1}{\phi(n)} - \frac{1}{a}\right| = \frac{|\phi_1(n)|}{|a||a + \phi_1(n)|},$$

and it is plain, since $\lim \phi_1(n) = 0$, that we can choose n_0 so that this is smaller than any assigned number ϵ when $n \geqq n_0$.

From Theorems II and III we can at once deduce the principal theorem for quotients, viz.

THEOREM IV. *If* $\lim \phi(n) = a$ *and* $\lim \psi(n) = b$, *and* b *is not zero, then*

$$\lim \frac{\phi(n)}{\psi(n)} = \frac{a}{b}.$$

The reader will again find it instructive to formulate, prove, and illustrate by examples some of the 'subsidiary theorems' corresponding to Theorems III and IV.

67. THEOREM V. *If* $R\{\phi(n), \psi(n), \chi(n), \dots\}$ *is any rational function of* $\phi(n)$, $\psi(n)$, $\chi(n)$, …, *i.e. any function of the form*

$$P\{\phi(n), \psi(n), \chi(n), \dots\}/Q\{\phi(n), \psi(n), \chi(n), \dots\},$$

where P *and* Q *denote polynomials in* $\phi(n)$, $\psi(n)$, $\chi(n)$, …: *and if*

$$\lim \phi(n) = a, \quad \lim \psi(n) = b, \quad \lim \chi(n) = c, \ \dots,$$

and

$$Q(a, b, c, \dots) \neq 0;$$

then

$$\lim R\{\phi(n), \psi(n), \chi(n), \dots\} = R(a, b, c, \dots).$$

For P is a sum of a finite number of terms of the type

$$A\{\phi(n)\}^p\{\psi(n)\}^q \dots,$$

where A is a constant and p, q, … positive integers. This term, by Theorem II (or rather by its obvious extension to the product of any number of functions) tends to the limit $Aa^pb^q \dots$, and so P tends to the limit $P(a, b, c, \dots)$, by the similar extension of Theorem I. Similarly Q tends to $Q(a, b, c, \dots)$; and the result then follows from Theorem IV.

68. The preceding general theorem may be applied to the following very important particular problem: *what is the behaviour of the most general rational function of n, viz.*

$$S(n) = \frac{a_0 n^p + a_1 n^{p-1} + \dots + a_p}{b_0 n^q + b_1 n^{q-1} + \dots + b_q},$$

as n tends to ∞*?**

In order to apply the theorem we transform $S(n)$ by writing it in the form

$$n^{p-q} \left\{ \left(a_0 + \frac{a_1}{n} + \dots + \frac{a_p}{n^p} \right) \Big/ \left(b_0 + \frac{b_1}{n} + \dots + \frac{b_q}{n^q} \right) \right\}.$$

The function in curly brackets is of the form $R\{\phi(n)\}$, where $\phi(n) = 1/n$, and therefore tends, as n tends to ∞, to the limit $R(0) = a_0/b_0$. Now $n^{p-q} \to 0$ if $p < q$; $n^{p-q} = 1$ and $n^{p-q} \to 1$ if $p = q$; and $n^{p-q} \to +\infty$ if $p > q$. Hence, by Theorem II,

$$\begin{gathered} \lim S(n) = 0 \quad (p < q), \\ \lim S(n) = a_0/b_0 \quad (p = q), \\ S(n) \to +\infty \quad (p > q,\ a_0/b_0 \textit{ positive}), \\ S(n) \to -\infty \quad (p > q,\ a_0/b_0 \textit{ negative}). \end{gathered}$$

Examples XXVI. 1. What is the behaviour of the functions

$$\left(\frac{n-1}{n+1}\right)^2, \quad (-1)^n \left(\frac{n-1}{n+1}\right)^2, \quad \frac{n^2+1}{n}, \quad (-1)^n \frac{n^2+1}{n},$$

as $n \to \infty$?

2. Which (if any) of the functions

$$\begin{gathered} 1/(\cos^2 \tfrac{1}{2}n\pi + n \sin^2 \tfrac{1}{2}n\pi), \quad 1/\{n(\cos^2 \tfrac{1}{2}n\pi + n \sin^2 \tfrac{1}{2}n\pi)\}, \\ (n \cos^2 \tfrac{1}{2}n\pi + \sin^2 \tfrac{1}{2}n\pi)/\{n(\cos^2 \tfrac{1}{2}n\pi + n \sin^2 \tfrac{1}{2}n\pi)\} \end{gathered}$$

tend to a limit as $n \to \infty$?

*We naturally suppose that neither a_0 nor b_0 is zero.

3. Denoting by $S(n)$ the general rational function of n considered above, show that in all cases

$$\lim \frac{S(n+1)}{S(n)} = 1, \quad \lim \frac{S\{n+(1/n)\}}{S(n)} = 1.$$

69. Functions of n which increase steadily with n. A special but particularly important class of functions of n is formed by those whose variation as n tends to ∞ is always in the same direction, that is to say those which always increase (or always decrease) as n increases. Since $-\phi(n)$ always increases if $\phi(n)$ always decreases, it is not necessary to consider the two kinds of functions separately; for theorems proved for one kind can at once be extended to the other.

DEFINITION. *The function $\phi(n)$ will be said to increase steadily with n if $\phi(n+1) \geqq \phi(n)$ for all values of n.*

It is to be observed that we do not exclude the case in which $\phi(n)$ has the *same* value for several values of n; all we exclude is possible *decrease.* Thus the function

$$\phi(n) = 2n + (-1)^n,$$

whose values for $n = 0, 1, 2, 3, 4, \dots$ are

$$1, 1, 5, 5, 9, 9, \dots$$

is said to increase steadily with n. Our definition would indeed include even functions which remain constant from some value of n onwards; thus $\phi(n) = 1$ steadily increases according to our definition. However, as these functions are extremely special ones, and as there can be no doubt as to their behaviour as n tends to ∞, this apparent incongruity in the definition is not a serious defect.

There is one exceedingly important theorem concerning functions of this class.

THEOREM. *If $\phi(n)$ steadily increases with n, then either* (i) $\phi(n)$ *tends to a limit as n tends to ∞, or* (ii) $\phi(n) \to +\infty$.

That is to say, while there are in general *five* alternatives as to the behaviour of a function, there are *two* only for this special kind of function.

This theorem is a simple corollary of Dedekind's Theorem (§ 17). We divide the real numbers ξ into two classes L and R, putting ξ in L or R according as $\phi(n) \geqq \xi$ for some value of n (and so of course for all greater values), or $\phi(n) < \xi$ for all values of n.

The class L certainly exists; the class R may or may not. If it does not, then, given any number Δ, however large, $\phi(n) > \Delta$ for all sufficiently large values of n, and so

$$\phi(n) \to +\infty.$$

If on the other hand R exists, the classes L and R form a section of the real numbers in the sense of § 17. Let a be the number corresponding to the section, and let ϵ be any positive number. Then $\phi(n) < a + \epsilon$ for all values of n, and so, since ϵ is arbitrary, $\phi(n) \leqq a$. On the other hand $\phi(n) > a - \epsilon$ for some value of n, and so for all sufficiently large values. Thus

$$a - \epsilon < \phi(n) \leqq a$$

for all sufficiently large values of n; *i.e.*

$$\phi(n) \to a.$$

It should be observed that in general $\phi(n) < a$ for all values of n; for if $\phi(n)$ is equal to a for any value of n it must be equal to a for all greater values of n. Thus $\phi(n)$ can never be equal to a except in the case in which the values of $\phi(n)$ are ultimately all the same. If this is so, a is the largest member of L; otherwise L has no largest member.

Cor 1. *If $\phi(n)$ increases steadily with n, then it will tend to a limit or to $+\infty$ according as it is or is not possible to find a number K such that $\phi(n) < K$ for all values of n.*

We shall find this corollary exceedingly useful later on.

Cor 2. *If $\phi(n)$ increases steadily with n, and $\phi(n) < K$ for all values of n, then $\phi(n)$ tends to a limit and this limit is less than or equal to K.*

It should be noticed that the limit may be equal to K: if *e.g.* $\phi(n) = 3 - (1/n)$, then every value of $\phi(n)$ is less than 3, but the limit is equal to 3.

Cor 3. *If $\phi(n)$ increases steadily with n, and tends to a limit, then*

$$\phi(n) \leqq \lim \phi(n)$$

for all values of n.

The reader should write out for himself the corresponding theorems and corollaries for the case in which $\phi(n)$ *decreases* as n increases.

70. The great importance of these theorems lies in the fact that they give us (what we have so far been without) a means of deciding, in a great many cases, whether a given function of n does or does not tend to a limit as $n \to \infty$, *without requiring us to be able to guess or otherwise infer beforehand what the limit is.* If we know what the limit, if there is one, must be, we can use the test

$$|\phi(n) - l| < \epsilon \quad (n \geqq n_0):$$

as for example in the case of $\phi(n) = 1/n$, where it is obvious that the limit can only be zero. But suppose we have to determine whether

$$\phi(n) = \left(1 + \frac{1}{n}\right)^n$$

tends to a limit. In this case it is not obvious what the limit, if there is one, will be: and it is evident that the test above, which involves l, cannot be used, at any rate directly, to decide whether l exists or not.

Of course the test can sometimes be used indirectly, to prove by means of a *reductio ad absurdum* that l *cannot* exist. If *e.g.* $\phi(n) = (-1)^n$, it is clear that l would have to be equal to 1 and also equal to -1, which is obviously impossible.

71. Alternative proof of Weierstrass's Theorem of § 19. The results of § 69 enable us to give an alternative proof of the important theorem proved in § 19.

If we divide PQ into two equal parts, one at least of them must contain infinitely many points of S. We select the one which does, or, if both do, we

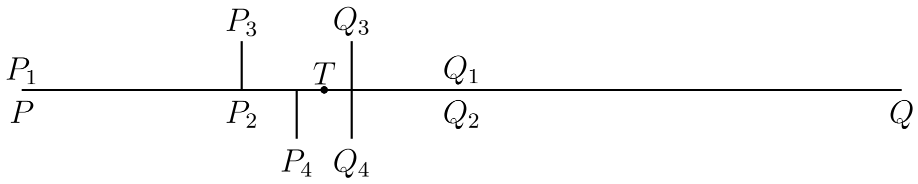

Fig. 28.

select the left-hand half; and we denote the selected half by P_1Q_1 (Fig. 28). If P_1Q_1 is the left-hand half, P_1 is the same point as P.

Similarly, if we divide P_1Q_1 into two halves, one at least of them must contain infinitely many points of S. We select the half P_2Q_2 which does so, or, if both do so, we select the left-hand half. Proceeding in this way we can define a sequence of intervals

$$PQ, \quad P_1Q_1, \quad P_2Q_2, \quad P_3Q_3, \ \dots,$$

each of which is a half of its predecessor, and each of which contains infinitely many points of S.

The points P, P_1, P_2, ... progress steadily from left to right, and so P_n tends to a limiting position T. Similarly Q_n tends to a limiting position T'. But TT' is plainly less than P_nQ_n, whatever the value of n; and P_nQ_n, being equal to $PQ/2^n$, tends to zero. Hence T' coincides with T, and P_n and Q_n both tend to T.

Then T is a point of accumulation of S. For suppose that ξ is its coordinate, and consider any interval of the type $[\xi-\epsilon, \xi+\epsilon]$. If n is sufficiently large, P_nQ_n will lie entirely inside this interval.* Hence $[\xi-\epsilon, \xi+\epsilon]$ contains infinitely many points of S.

72. The limit of x^n as n tends to ∞. Let us apply the results of § 69 to the particularly important case in which $\phi(n) = x^n$. If $x = 1$ then $\phi(n) = 1$, $\lim \phi(n) = 1$, and if $x = 0$ then $\phi(n) = 0$, $\lim \phi(n) = 0$, so that these special cases need not detain us.

First, suppose x positive. Then, since $\phi(n+1) = x\phi(n)$, $\phi(n)$ increases with n if $x > 1$, decreases as n increases if $x < 1$.

*This will certainly be the case as soon as $PQ/2^n < \epsilon$.

If $x > 1$, then x^n must tend either to a limit (which must obviously be greater than 1) or to $+\infty$. Suppose it tends to a limit l. Then $\lim \phi(n+1) = \lim \phi(n) = l$, by Exs. XXV. 7; but

$$\lim \phi(n+1) = \lim x\phi(n) = x \lim \phi(n) = xl,$$

and therefore $l = xl$: and as x and l are both greater than 1, this is impossible. Hence

$$x^n \to +\infty \quad (x > 1).$$

Example. The reader may give an alternative proof, showing by the binomial theorem that $x^n > 1 + n\delta$ if δ is positive and $x = 1 + \delta$, and so that

$$x^n \to +\infty.$$

On the other hand x^n is a decreasing function if $x < 1$, and must therefore tend to a limit or to $-\infty$. Since x^n is positive the second alternative may be ignored. Thus $\lim x^n = l$, say, and as above $l = xl$, so that l must be zero. Hence

$$\lim x^n = 0 \quad (0 < x < 1).$$

Example. Prove as in the preceding example that $(1/x)^n$ tends to $+\infty$ if $0 < x < 1$, and deduce that x^n tends to 0.

We have finally to consider the case in which x is negative. If $-1 < x < 0$ and $x = -y$, so that $0 < y < 1$, then it follows from what precedes that $\lim y^n = 0$ and therefore $\lim x^n = 0$. If $x = -1$ it is obvious that x^n oscillates, taking the values -1, 1 alternatively. Finally if $x < -1$, and $x = -y$, so that $y > 1$, then y^n tends to $+\infty$, and therefore x^n takes values, both positive and negative, numerically greater than any assigned number. Hence x^n oscillates infinitely. To sum up:

$$\begin{array}{ll}
\phi(n) = x^n \to +\infty & (x > 1), \\
\lim \phi(n) = 1 & (x = 1), \\
\lim \phi(n) = 0 & (-1 < x < 1), \\
\phi(n) \textit{ oscillates finitely} & (x = -1), \\
\phi(n) \textit{ oscillates infinitely} & (x < -1).
\end{array}$$

Examples XXVII.* 1. If $\phi(n)$ is positive and $\phi(n+1) > K\phi(n)$, where $K > 1$, for all values of n, then $\phi(n) \to +\infty$.

[For

$$\phi(n) > K\phi(n-1) > K^2\phi(n-2) \cdots > K^{n-1}\phi(1),$$

from which the conclusion follows at once, as $K^n \to \infty$.]

2. The same result is true if the conditions above stated are satisfied only when $n \geqq n_0$.

3. If $\phi(n)$ is positive and $\phi(n+1) < K\phi(n)$, where $0 < K < 1$, then $\lim \phi(n) = 0$. This result also is true if the conditions are satisfied only when $n \geqq n_0$.

4. If $|\phi(n+1)| < K|\phi(n)|$ when $n \geqq n_0$, and $0 < K < 1$, then $\lim \phi(n) = 0$.

5. If $\phi(n)$ is positive and $\lim\{\phi(n+1)\}/\{\phi(n)\} = l > 1$, then $\phi(n) \to +\infty$.

[For we can determine n_0 so that $\{\phi(n+1)\}/\{\phi(n)\} > K > 1$ when $n \geqq n_0$: we may, *e.g.*, take K halfway between 1 and l. Now apply Ex. 1.]

6. If $\lim\{\phi(n+1)\}/\{\phi(n)\} = l$, where l is numerically less than unity, then $\lim \phi(n) = 0$. [This follows from Ex. 4 as Ex. 5 follows from Ex. 1.]

7. Determine the behaviour, as $n \to \infty$, of $\phi(n) = n^r x^n$, where r is any positive integer.

[If $x = 0$ then $\phi(n) = 0$ for all values of n, and $\phi(n) \to 0$. In all other cases

$$\frac{\phi(n+1)}{\phi(n)} = \left(\frac{n+1}{n}\right)^r x \to x.$$

First suppose x positive. Then $\phi(n) \to +\infty$ if $x > 1$ (Ex. 5) and $\phi(n) \to 0$ if $x < 1$ (Ex. 6). If $x = 1$, then $\phi(n) = n^r \to +\infty$. Next suppose x negative. Then

*These examples are particularly important and several of them will be made use of later in the text. They should therefore be studied very carefully.

$|\phi(n)| = n^r|x|^n$ tends to $+\infty$ if $|x| \geqq 1$ and to 0 if $|x| < 1$. Hence $\phi(n)$ oscillates infinitely if $x \leqq -1$ and $\phi(n) \to 0$ if $-1 < x < 0$.]

8. Discuss $n^{-r}x^n$ in the same way. [The results are the same, except that $\phi(n) \to 0$ when $x = 1$ or -1.]

9. Draw up a table to show how $n^k x^n$ behaves as $n \to \infty$, for all real values of x, and all positive and negative integral values of k.

[The reader will observe that the value of k is immaterial except in the special cases when $x = 1$ or -1. Since $\lim\{(n+1)/n\}^k = 1$, whether k be positive or negative, the limit of the ratio $\phi(n+1)/\phi(n)$ depends only on x, and the behaviour of $\phi(n)$ is in general dominated by the factor x^n. The factor n^k only asserts itself when x is numerically equal to 1.]

10. Prove that if x is positive then $\sqrt[n]{x} \to 1$ as $n \to \infty$. [Suppose, *e.g.*, $x > 1$. Then x, $\sqrt{x}$, $\sqrt[3]{x}$, … is a decreasing sequence, and $\sqrt[n]{x} > 1$ for all values of n. Thus $\sqrt[n]{x} \to l$, where $l \geqq 1$. But if $l > 1$ we can find values of n, as large as we please, for which $\sqrt[n]{x} > l$ or $x > l^n$; and, since $l^n \to +\infty$ as $n \to \infty$, this is impossible.]

11. $\sqrt[n]{n} \to 1$. [For $\sqrt[n+1]{n+1} < \sqrt[n]{n}$ if $(n+1)^n < n^{n+1}$ or $\{1 + (1/n)\}^n < n$, which is certainly satisfied if $n \geqq 3$ (see § 73 for a proof). Thus $\sqrt[n]{n}$ decreases as n increases from 3 onwards, and, as it is always greater than unity, it tends to a limit which is greater than or equal to unity. But if $\sqrt[n]{n} \to l$, where $l > 1$, then $n > l^n$, which is certainly untrue for sufficiently large values of n, since $l^n/n \to +\infty$ with n (Exs. 7, 8).]

12. $\sqrt[n]{n!} \to +\infty$. [However large Δ may be, $n! > \Delta^n$ if n is large enough. For if $u_n = \Delta^n/n!$ then $u_{n+1}/u_n = \Delta/(n+1)$, which tends to zero as $n \to \infty$, so that u_n does the same (Ex. 6).]

13. Show that if $-1 < x < 1$ then

$$u_n = \frac{m(m-1)\dots(m-n+1)}{n!}x^n = \binom{m}{n}x^n$$

tends to zero as $n \to \infty$.

[If m is a positive integer, $u_n = 0$ for $n > m$. Otherwise

$$\frac{u_{n+1}}{u_n} = \frac{m-n}{n+1}x \to -x,$$

unless $x = 0$.]

73. The limit of $\left(1+\dfrac{1}{n}\right)^n$. A more difficult problem which can be solved by the help of § 69 arises when $\phi(n) = \{1 + 1/n\}^n$.

It follows from the binomial theorem* that

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= 1 + n\cdot\frac{1}{n} + \frac{n(n-1)}{1\cdot 2}\,\frac{1}{n^2} + \dots + \frac{n(n-1)\dots(n-n+1)}{1\cdot 2\dots n}\,\frac{1}{n^n}\\ &= 1 + 1 + \frac{1}{1\cdot 2}\left(1-\frac{1}{n}\right) + \frac{1}{1\cdot 2\cdot 3}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) + \dots\\ &\qquad + \frac{1}{1\cdot 2\dots n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{n-1}{n}\right).\end{aligned}$$

The $(p + 1)$th term in this expression, viz.

$$\frac{1}{1\cdot 2\dots p}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{p-1}{n}\right),$$

is positive and an increasing function of n, and the number of terms also increases with n. Hence $\left(1+\dfrac{1}{n}\right)^n$ increases with n, and so tends to a limit or to $+\infty$, as $n \to \infty$.

But

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &< 1 + 1 + \frac{1}{1\cdot 2} + \frac{1}{1\cdot 2\cdot 3} + \dots + \frac{1}{1\cdot 2\cdot 3\dots n}\\ &< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^{n-1}} < 3.\end{aligned}$$

Thus $\left(1+\dfrac{1}{n}\right)^n$ cannot tend to $+\infty$, and so

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e,$$

where e is a number such that $2 < e \leqq 3$.

---

*The binomial theorem for a positive integral exponent, which is what is used here, is a theorem of elementary algebra. The other cases of the theorem belong to the theory of infinite series, and will be considered later.

74. Some algebraical lemmas. It will be convenient to prove at this stage a number of elementary inequalities which will be useful to us later on.

(i) It is evident that if $\alpha > 1$ and r is a positive integer then

$$r\alpha^r > \alpha^{r-1} + \alpha^{r-2} + \cdots + 1.$$

Multiplying both sides of this inequality by $\alpha - 1$, we obtain

$$r\alpha^r(\alpha - 1) > \alpha^r - 1;$$

and adding $r(\alpha^r - 1)$ to each side, and dividing by $r(r+1)$, we obtain

$$\frac{\alpha^{r+1} - 1}{r+1} > \frac{\alpha^r - 1}{r} \quad (\alpha > 1). \tag{1}$$

Similarly we can prove that

$$\frac{1 - \beta^{r+1}}{r+1} < \frac{1 - \beta^r}{r} \quad (0 < \beta < 1). \tag{2}$$

It follows that if r and s are positive integers, and $r > s$, then

$$\frac{\alpha^r - 1}{r} > \frac{a^s - 1}{s}, \quad \frac{1 - \beta^r}{r} < \frac{1 - \beta^s}{s}. \tag{3}$$

Here $0 < \beta < 1 < \alpha$. In particular, when $s = 1$, we have

$$\alpha^r - 1 > r(\alpha - 1), \quad 1 - \beta^r < r(1 - \beta). \tag{4}$$

(ii) The inequalities (3) and (4) have been proved on the supposition that r and s are positive integers. But it is easy to see that they hold under the more general hypothesis that r and s are any positive rational numbers. Let us consider, for example, the first of the inequalities (3). Let $r = a/b$, $s = c/d$, where a, b, c, d are positive integers; so that $ad > bc$. If we put $\alpha = \gamma^{bd}$, the inequality takes the form

$$(\gamma^{ad} - 1)/ad > (\gamma^{bc} - 1)/bc;$$

and this we have proved already. The same argument applies to the remaining inequalities; and it can evidently be proved in a similar manner that

$$\alpha^s - 1 < s(\alpha - 1), \quad 1 - \beta^s > s(1 - \beta), \tag{5}$$

if s is a positive rational number less than 1.

(iii) In what follows it is to be understood *that all the letters denote positive numbers, that r and s are rational, and that α and r are greater than* 1, *β and s less than* 1. Writing $1/\beta$ for α, and $1/\alpha$ for β, in (4), we obtain

$$\alpha^r - 1 < r\alpha^{r-1}(\alpha - 1), \quad 1 - \beta^r > r\beta^{r-1}(1 - \beta). \tag{6}$$

Similarly, from (5), we deduce

$$\alpha^s - 1 > s\alpha^{s-1}(\alpha - 1), \quad 1 - \beta^s < s\beta^{s-1}(1 - \beta). \tag{7}$$

Combining (4) and (6), we see that

$$r\alpha^{r-1}(\alpha - 1) > \alpha^r - 1 > r(\alpha - 1). \tag{8}$$

Writing x/y for α, we obtain

$$rx^{r-1}(x - y) > x^r - y^r > ry^{r-1}(x - y) \tag{9}$$

if $x > y > 0$. And the same argument, applied to (5) and (7), leads to

$$sx^{s-1}(x - y) < x^s - y^s < sy^{s-1}(x - y). \tag{10}$$

Examples XXVIII. 1. Verify (9) for $r = 2$, 3, and (10) for $s = \frac{1}{2}$, $\frac{1}{3}$.

2. Show that (9) and (10) are also true if $y > x > 0$.

3. Show that (9) also holds for $r < 0$. [See Chrystal's *Algebra*, vol. ii, pp. 43–45.]

4. If $\phi(n) \to l$, where $l > 0$, as $n \to \infty$, then $\phi^k \to l^k$, k being any rational number.

[We may suppose that $k > 0$, in virtue of Theorem III of § 66; and that $\frac{1}{2}l < \phi < 2l$, as is certainly the case from a certain value of n onwards. If $k > 1$,

$$k\phi^{k-1}(\phi - l) > \phi^k - l^k > kl^{k-1}(\phi - l)$$

or

$$kl^{k-1}(l - \phi) > l^k - \phi^k > k\phi^{k-1}(l - \phi),$$

according as $\phi > l$ or $\phi < l$. It follows that the ratio of $|\phi^k - l^k|$ and $|\phi - l|$ lies between $k(\frac{1}{2}l)^{k-1}$ and $k(2l)^{k-1}$. The proof is similar when $0 < k < 1$. The result is still true when $l = 0$, if $k > 0$.]

5. Extend the results of Exs. XXVII. 7, 8, 9 to the case in which r or k are any rational numbers.

75. The limit of $n(\sqrt[n]{x} - 1)$. If in the first inequality (3) of § 74 we put $r = 1/(n-1)$, $s = 1/n$, we see that

$$(n-1)(\sqrt[n-1]{\alpha} - 1) > n(\sqrt[n]{\alpha} - 1)$$

when $\alpha > 1$. Thus if $\phi(n) = n(\sqrt[n]{\alpha} - 1)$ then $\phi(n)$ decreases steadily as n increases. Also $\phi(n)$ is always positive. Hence $\phi(n)$ tends to a limit l as $n \to \infty$, and $l \geqq 0$.

Again if, in the first inequality (7) of § 74, we put $s = 1/n$, we obtain

$$n(\sqrt[n]{\alpha} - 1) > \sqrt[n]{\alpha}\left(1 - \frac{1}{\alpha}\right) > 1 - \frac{1}{\alpha}.$$

Thus $l \geqq 1 - (1/\alpha) > 0$. Hence, if $\alpha > 1$, we have

$$\lim_{n\to\infty} n(\sqrt[n]{\alpha} - 1) = f(\alpha),$$

where $f(\alpha) > 0$.

Next suppose $\beta < 1$, and let $\beta = 1/\alpha$; then $n(\sqrt[n]{\beta} - 1) = -n(\sqrt{\alpha} - 1)/\sqrt[n]{\alpha}$. Now $n(\sqrt[n]{\alpha} - 1) \to f(\alpha)$, and (Exs. XXVII. 10)

$$\sqrt[n]{\alpha} \to 1.$$

Hence, if $\beta = 1/\alpha < 1$, we have

$$n(\sqrt[n]{\beta} - 1) \to -f(\alpha).$$

Finally, if $x = 1$, then $n(\sqrt[n]{x} - 1) = 0$ for all values of n.

Thus we arrive at the result: *the limit*

$$\lim n(\sqrt[n]{x} - 1)$$

defines a function of x for all positive values of x. This function $f(x)$ possesses the properties

$$f(1/x) = -f(x), \quad f(1) = 0,$$

and is positive or negative according as $x > 1$ or $x < 1$. Later on we shall be able to identify this function with the *Napierian logarithm* of x.

Example. Prove that $f(xy) = f(x) + f(y)$. [Use the equations

$$f(xy) = \lim n(\sqrt[n]{xy} - 1) = \lim\{n(\sqrt[n]{x} - 1)\sqrt[n]{y} + n(\sqrt[n]{y} - 1)\}.]$$

76. Infinite Series. Suppose that $u(n)$ is any function of n defined for all values of n. If we add up the values of $u(\nu)$ for $\nu = 1, 2, \dots n$, we obtain another function of n, viz.

$$s(n) = u(1) + u(2) + \cdots + u(n),$$

also defined for all values of n. It is generally most convenient to alter our notation slightly and write this equation in the form

$$s_n = u_1 + u_2 + \cdots + u_n,$$

or, more shortly,

$$s_n = \sum_{\nu=1}^{n} u_\nu.$$

If now we suppose that s_n tends to a limit s when n tends to ∞, we have

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} u_\nu = s.$$

This equation is usually written in one of the forms

$$\sum_{\nu=1}^{\infty} u_\nu = s, \quad u_1 + u_2 + u_3 + \cdots = s,$$

the dots denoting the indefinite continuance of the series of u's.

The meaning of the above equations, expressed roughly, is that by adding more and more of the u's together we get nearer and nearer to the limit s. More precisely, if any small positive number ϵ is chosen, we can choose $n_0(\epsilon)$ so that the sum of the first $n_0(\epsilon)$ terms, or any of greater number of terms, lies between $s - \epsilon$ and $s + \epsilon$; or in symbols

$$s - \epsilon < s_n < s + \epsilon,$$

if $n \geqq n_0(\epsilon)$. In these circumstances we shall call the series

$$u_1 + u_2 + \dots$$

a **convergent infinite series**, and we shall call s the *sum* of the series, or the *sum of all the terms* of the series.

Thus to say that the series $u_1 + u_2 + \dots$ *converges and has the sum* s, or *converges to the sum* s or simply *converges to* s, is merely another way of stating that the sum $s_n = u_1 + u_2 + \dots + u_n$ of the first n terms tends to the limit s as $n \to \infty$, and the consideration of such infinite series introduces no new ideas beyond those with which the early part of this chapter should already have made the reader familiar. In fact the sum s_n is merely a function $\phi(n)$, such as we have been considering, expressed in a particular form. Any function $\phi(n)$ may be expressed in this form, by writing

$$\phi(n) = \phi(1) + \{\phi(2) - \phi(1)\} + \dots + \{\phi(n) - \phi(n-1)\};$$

and it is sometimes convenient to say that $\phi(n)$ *converges* (instead of 'tends') to the limit l, say, as $n \to \infty$.

If $s_n \to +\infty$ or $s_n \to -\infty$, we shall say that the series $u_1 + u_2 + \dots$ is **divergent** or *diverges to* $+\infty$, or $-\infty$, as the case may be. These phrases too may be applied to any function $\phi(n)$: thus if $\phi(n) \to +\infty$ we may say that $\phi(n)$ *diverges to* $+\infty$. If s_n does not tend to a limit or to $+\infty$ or to $-\infty$, then it oscillates finitely or infinitely: in this case we say that the series $u_1 + u_2 + \dots$ oscillates finitely or infinitely.*

*The reader should be warned that the words 'divergent' and 'oscillatory' are used differently by different writers. The use of the words here agrees with that of Bromwich's *Infinite Series*. In Hobson's *Theory of Functions of a Real Variable* a series is said to oscillate only if it oscillates *finitely*, series which oscillate infinitely being classed as 'divergent'. Many foreign writers use 'divergent' as meaning merely 'not convergent'.

77. General theorems concerning infinite series. When we are dealing with infinite series we shall constantly have occasion to use the following general theorems.

(1) If $u_1+u_2+\dots$ is convergent, and has the sum s, then $a+u_1+u_2+\dots$ is convergent and has the sum $a+s$. Similarly $a+b+c+\dots+k+u_1+u_2+\dots$ is convergent and has the sum $a+b+c+\dots+k+s$.

(2) If $u_1+u_2+\dots$ is convergent and has the sum s, then $u_{m+1}+u_{m+2}+\dots$ is convergent and has the sum

$$s-u_1-u_2-\dots-u_m.$$

(3) If any series considered in (1) or (2) diverges or oscillates, then so do the others.

(4) If $u_1+u_2+\dots$ is convergent and has the sum s, then $ku_1+ku_2+\dots$ is convergent and has the sum ks.

(5) If the first series considered in (4) diverges or oscillates, then so does the second, unless $k=0$.

(6) If $u_1+u_2+\dots$ and $v_1+v_2+\dots$ are both convergent, then the series $(u_1+v_1)+(u_2+v_2)+\dots$ is convergent and its sum is the sum of the first two series.

All these theorems are almost obvious and may be proved at once from the definitions or by applying the results of §§ 63–66 to the sum $s_n=u_1+u_2+\dots+u_n$. Those which follow are of a somewhat different character.

(7) *If* $u_1+u_2+\dots$ *is convergent, then* $\lim u_n=0$.

For $u_n=s_n-s_{n-1}$, and s_n and s_{n-1} have the same limit s. Hence $\lim u_n=s-s=0$.

The reader may be tempted to think that the converse of the theorem is true and that if $\lim u_n=0$ then the series $\sum u_n$ must be convergent. That this is not the case is easily seen from an example. Let the series be

$$1+\tfrac{1}{2}+\tfrac{1}{3}+\tfrac{1}{4}+\dots$$

so that $u_n=1/n$. The sum of the first four terms is

$$1+\tfrac{1}{2}+\tfrac{1}{3}+\tfrac{1}{4}>1+\tfrac{1}{2}+\tfrac{2}{4}=1+\tfrac{1}{2}+\tfrac{1}{2}.$$

The sum of the next four terms is $\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}>\frac{4}{8}=\frac{1}{2}$; the sum of the next eight terms is greater than $\frac{8}{16}=\frac{1}{2}$, and so on. The sum of the first

$$4+4+8+16+\dots+2^n=2^{n+1}$$

terms is greater than

$$2+\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}+\dots+\tfrac{1}{2}=\tfrac{1}{2}(n+3),$$

and this increases beyond all limit with n: hence the series diverges to $+\infty$.

(8) *If $u_1+u_2+u_3+\dots$ is convergent, then so is any series formed by grouping the terms in brackets in any way to form new single terms, and the sums of the two series are the same.*

The reader will be able to supply the proof of this theorem. Here again the converse is not true. Thus $1-1+1-1+\dots$ oscillates, while

$$(1-1)+(1-1)+\dots$$

or $0+0+0+\dots$ converges to 0.

(9) *If every term u_n is positive (or zero), then the series $\sum u_n$ must either converge or diverge to $+\infty$. If it converges, its sum must be positive* (unless all the terms are zero, when of course its sum is zero).

For s_n is an increasing function of n, according to the definition of § 69, and we can apply the results of that section to s_n.

(10) *If every term u_n is positive (or zero), then the necessary and sufficient condition that the series $\sum u_n$ should be convergent is that it should be possible to find a number K such that the sum of any number of terms is less than K; and, if K can be so found, then the sum of the series is not greater than K.*

This also follows at once from § 69. It is perhaps hardly necessary to point out that the theorem is not true if the condition that every u_n is positive is not fulfilled. For example

$$1-1+1-1+\dots$$

obviously oscillates, s_n being alternately equal to 1 and to 0.

(11) *If* $u_1 + u_2 + \dots$, $v_1 + v_2 + \dots$ *are two series of positive (or zero) terms, and the second series is convergent, and if* $u_n \leqq Kv_n$, *where* K *is a constant, for all values of* n, *then the first series is also convergent, and its sum is less than or equal to* K *times that of the second.*

For if $v_1 + v_2 + \dots = t$ then $v_1 + v_2 + \dots + v_n \leqq t$ for all values of n, and so $u_1 + u_2 + \dots + u_n \leqq Kt$; which proves the theorem.

Conversely, if $\sum u_n$ *is divergent, and* $v_n \geqq Ku_n$, *then* $\sum v_n$ *is divergent.*

78. The infinite geometrical series. We shall now consider the 'geometrical' series, whose general term is $u_n = r^{n-1}$. In this case

$$s_n = 1 + r + r^2 + \dots + r^{n-1} = (1 - r^n)/(1 - r),$$

except in the special case in which $r = 1$, when

$$s_n = 1 + 1 + \dots + 1 = n.$$

In the last case $s_n \to +\infty$. In the general case s_n will tend to a limit if and only if r^n does so. Referring to the results of § 72 we see that *the series* $1 + r + r^2 + \dots$ *is convergent and has the sum* $1/(1 - r)$ *if and only if* $-1 < r < 1$.

If $r \geqq 1$, then $s_n \geqq n$, and so $s_n \to +\infty$; *i.e.* the series diverges to $+\infty$. If $r = -1$, then $s_n = 1$ or $s_n = 0$ according as n is odd or even: *i.e.* s_n oscillates finitely. If $r < -1$, then s_n oscillates infinitely. Thus, to sum up, *the series* $1 + r + r^2 + \dots$ *diverges to* $+\infty$ *if* $r \geqq 1$, *converges to* $1/(1 - r)$ *if* $-1 < r < 1$, *oscillates finitely if* $r = -1$, *and oscillates infinitely if* $r < -1$.

Examples XXIX. 1. **Recurring decimals.** The commonest example of an infinite geometric series is given by an ordinary recurring decimal. Consider, for example, the decimal $.217\overline{13}$. This stands, according to the ordinary rules of arithmetic, for

$$\frac{2}{10}+\frac{1}{10^2}+\frac{7}{10^3}+\frac{1}{10^4}+\frac{3}{10^5}+\frac{1}{10^6}+\frac{3}{10^7}+\dots = \frac{217}{1000}+\frac{13}{10^5}\bigg/\left(1-\frac{1}{10^2}\right) = \frac{2687}{12{,}375}.$$

The reader should consider where and how any of the general theorems of § 77 have been used in this reduction.

2. Show that in general

$$.a_1a_2\dots a_m\overline{\alpha_1\alpha_2\dots\alpha_n} = \frac{a_1a_2\dots a_m\alpha_1\dots\alpha_n - a_1a_2\dots a_n}{99\dots900\dots0},$$

the denominator containing n 9's and m 0's.

3. Show that a pure recurring decimal is always equal to a proper fraction whose denominator does not contain 2 or 5 as a factor.

4. A decimal with m non-recurring and n recurring decimal figures is equal to a proper fraction whose denominator is divisible by 2^m or 5^m but by no higher power of either.

5. The converses of Exs. 3, 4 are also true. Let $r = p/q$, and suppose first that q is prime to 10. If we divide all powers of 10 by q we can obtain at most q different remainders. It is therefore possible to find two numbers n_1 and n_2, where $n_1 > n_2$, such that 10^{n_1} and 10^{n_2} give the same remainder. Hence $10^{n_1} - 10^{n_2} = 10^{n_2}(10^{n_1-n_2} - 1)$ is divisible by q, and so $10^n - 1$, where $n = n_1 - n_2$, is divisible by q. Hence r may be expressed in the form $P/(10^n - 1)$, or in the form

$$\frac{P}{10^n} + \frac{P}{10^{2n}} + \dots,$$

i.e. as a pure recurring decimal with n figures. If on the other hand $q = 2^\alpha 5^\beta Q$, where Q is prime to 10, and m is the greater of α and β, then $10^m r$ has a denominator prime to 10, and is therefore expressible as the sum of an integer and a pure recurring decimal. But this is not true of $10^\mu r$, for any value of μ less than m; hence the decimal for r has exactly m non-recurring figures.

6. To the results of Exs. 2–5 we must add that of Ex. I. 3. Finally, if we observe that

$$.\overline{9} = \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \dots = 1,$$

we see that every terminating decimal can also be expressed as a mixed recurring decimal whose recurring part is composed entirely of 9's. For example, $.217 = .216\overline{9}$. Thus every proper fraction can be expressed as a recurring decimal, and conversely.

7. **Decimals in general. The expression of irrational numbers as non-recurring decimals.** Any decimal, whether recurring or not, corresponds

to a definite number between 0 and 1. For the decimal $.a_1a_2a_3a_4\dots$ stands for the series

$$\frac{a_1}{10}+\frac{a_2}{10^2}+\frac{a_3}{10^3}+\dots.$$

Since all the digits a_r are positive, the sum s_n of the first n terms of this series increases with n, and it is certainly not greater than $.\overline{9}$ or 1. Hence s_n tends to a limit between 0 and 1.

Moreover no two decimals can correspond to the same number (except in the special case noticed in Ex. 6). For suppose that $.a_1a_2a_3\dots$, $.b_1b_2b_3\dots$ are two decimals which agree as far as the figures a_{r-1}, b_{r-1}, while $a_r > b_r$. Then $a_r \geqq b_r + 1 > b_r.b_{r+1}b_{r+2}\dots$ (unless b_{r+1}, b_{r+2}, $\dots$ are all 9's), and so

$$.a_1a_2\dots a_ra_{r+1}\dots > .b_1b_2\dots b_rb_{r+1}\dots.$$

It follows that the expression of a rational fraction as a recurring decimal (Exs. 2–6) is unique. It also follows that every decimal which does not recur represents some *irrational* number between 0 and 1. Conversely, any such number can be expressed as such a decimal. For it must lie in one of the intervals

$$0,\ 1/10;\quad 1/10,\ 2/10;\ \dots;\quad 9/10,\ 1.$$

If it lies between $r/10$ and $(r+1)/10$, then the first figure is r. By subdividing this interval into 10 parts we can determine the second figure; and so on. But (Exs. 3, 4) the decimal cannot recur. Thus, for example, the decimal $1.414\dots$, obtained by the ordinary process for the extraction of $\sqrt{2}$, cannot recur.

8. The decimals $.101\,001\,000\,100\,001\,0\dots$ and $.202\,002\,000\,200\,002\,0\dots$, in which the number of zeros between two 1's or 2's increases by one at each stage, represent irrational numbers.

9. The decimal $.111\,010\,100\,010\,10\dots$, in which the nth figure is 1 if n is prime, and zero otherwise, represents an irrational number. [Since the number of primes is infinite the decimal does not terminate. Nor can it recur: for if it did we could determine m and p so that m, $m+p$, $m+2p$, $m+3p$, $\dots$ are all prime numbers; and this is absurd, since the series includes $m+mp$.]*

*All the results of Exs. XXIX may be extended, with suitable modifications, to decimals in any scale of notation. For a fuller discussion see Bromwich, *Infinite Series*, Appendix I.

Examples XXX. 1. The series $r^m + r^{m+1} + \dots$ is convergent if $-1 < r < 1$, and its sum is $1/(1-r) - 1 - r - \dots - r^{m-1}$ (§ 77, (2)).

2. The series $r^m + r^{m+1} + \dots$ is convergent if $-1 < r < 1$, and its sum is $r^m/(1-r)$ (§ 77, (4)). Verify that the results of Exs. 1 and 2 are in agreement.

3. Prove that the series $1 + 2r + 2r^2 + \dots$ is convergent, and that its sum is $(1+r)/(1-r)$, (α) by writing it in the form $-1 + 2(1 + r + r^2 + \dots)$, ($\beta$) by writing it in the form $1 + 2(r + r^2 + \dots)$, (γ) by adding the two series $1 + r + r^2 + \dots$, $r + r^2 + \dots$. In each case mention which of the theorems of § 77 are used in your proof.

4. Prove that the 'arithmetic' series

$$a + (a+b) + (a+2b) + \dots$$

is always divergent, unless both a and b are zero. Show that, if b is not zero, the series diverges to $+\infty$ or to $-\infty$ according to the sign of b, while if $b = 0$ it diverges to $+\infty$ or $-\infty$ according to the sign of a.

5. What is the sum of the series

$$(1-r) + (r-r^2) + (r^2-r^3) + \dots$$

when the series is convergent? [The series converges only if $-1 < r \leqq 1$. Its sum is 1, except when $r = 1$, when its sum is 0.]

6. Sum the series

$$r^2 + \frac{r^2}{1+r^2} + \frac{r^2}{(1+r^2)^2} + \dots.$$

[The series is always convergent. Its sum is $1 + r^2$, except when $r = 0$, when its sum is 0.]

7. If we assume that $1 + r + r^2 + \dots$ is convergent then we can prove that its sum is $1/(1-r)$ by means of § 77, (1) and (4). For if $1 + r + r^2 + \dots = s$ then

$$s = 1 + r(1 + r^2 + \dots) = 1 + rs.$$

8. Sum the series

$$r + \frac{r}{1+r} + \frac{r}{(1+r)^2} + \dots$$

when it is convergent. [The series is convergent if $-1 < 1/(1+r) < 1$, *i.e.* if $r < -2$ or if $r > 0$, and its sum is $1+r$. It is also convergent when $r = 0$, when its sum is 0.]

9. Answer the same question for the series

$$r - \frac{r}{1+r} + \frac{r}{(1+r)^2} - \dots, \qquad r + \frac{r}{1-r} + \frac{r}{(1-r)^2} + \dots,$$
$$1 - \frac{r}{1+r} + \left(\frac{r}{1+r}\right)^2 - \dots, \qquad 1 + \frac{r}{1-r} + \left(\frac{r}{1-r}\right)^2 + \dots.$$

10. Consider the convergence of the series

$$(1+r) + (r^2 + r^3) + \dots, \qquad (1 + r + r^2) + (r^3 + r^4 + r^5) + \dots,$$
$$1 - 2r + r^2 + r^3 - 2r^4 + r^5 + \dots, \qquad (1 - 2r + r^2) + (r^3 - 2r^4 + r^5) + \dots,$$

and find their sums when they are convergent.

11. If $0 \leqq a_n \leqq 1$ then the series $a_0 + a_1 r + a_2 r^2 + \dots$ is convergent for $0 \leqq r < 1$, and its sum is not greater than $1/(1-r)$.

12. If in addition the series $a_0 + a_1 + a_2 + \dots$ is convergent, then the series $a_0 + a_1 r + a_2 r^2 + \dots$ is convergent for $0 \leqq r \leqq 1$, and its sum is not greater than the lesser of $a_0 + a_1 + a_2 + \dots$ and $1/(1-r)$.

13. The series
$$1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \dots$$
is convergent. [For $1/(1 \cdot 2 \dots n) \leqq 1/2^{n-1}$.]

14. The series

$$1 + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + \dots, \qquad \frac{1}{1} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} + \dots$$

are convergent.

15. The general harmonic series

$$\frac{1}{a} + \frac{1}{a+b} + \frac{1}{a+2b} + \dots,$$

where a and b are positive, diverges to $+\infty$.

[For $u_n = 1/(a+nb) > 1/\{n(a+b)\}$. Now compare with $1 + \frac{1}{2} + \frac{1}{3} + \dots$.]

16. Show that the series

$$(u_0 - u_1) + (u_1 - u_2) + (u_2 - u_3) + \dots$$

is convergent if and only if u_n tends to a limit as $n \to \infty$.

17. If $u_1 + u_2 + u_3 + \dots$ is divergent then so is any series formed by grouping the terms in brackets in any way to form new single terms.

18. Any series, formed by taking a selection of the terms of a convergent series of positive terms, is itself convergent.

79. The representation of functions of a continuous real variable by means of limits. In the preceding sections we have frequently been concerned with limits such as

$$\lim_{n\to\infty} \phi_n(x),$$

and series such as

$$u_1(x) + u_2(x) + \dots = \lim_{n\to\infty} \{u_1(x) + u_2(x) + \dots + u_n(x)\},$$

in which the function of n whose limit we are seeking involves, besides n, another variable x. In such cases the limit is of course a function of x. Thus in § 75 we encountered the function

$$f(x) = \lim_{n\to\infty} n(\sqrt[n]{x} - 1) :$$

and the sum of the geometrical series $1 + x + x^2 + \dots$ is a function of x, viz. the function which is equal to $1/(1-x)$ if $-1 < x < 1$ and is undefined for all other values of x.

Many of the apparently 'arbitrary' or 'unnatural' functions considered in Ch. II are capable of a simple representation of this kind, as will appear from the following examples.

Examples XXXI. 1. $\phi_n(x) = x$. Here n does not appear at all in the expression of $\phi_n(x)$, and $\phi(x) = \lim \phi_n(x) = x$ for all values of x.

2. $\phi_n(x) = x/n$. Here $\phi(x) = \lim \phi_n(x) = 0$ for all values of x.

3. $\phi_n(x) = nx$. If $x > 0$, $\phi_n(x) \to +\infty$; if $x < 0$, $\phi_n(x) \to -\infty$: only when $x = 0$ has $\phi_n(x)$ a limit (viz. 0) as $n \to \infty$. Thus $\phi(x) = 0$ when $x = 0$ and is not defined for any other value of x.

4. $\phi_n(x) = 1/nx$, $nx/(nx + 1)$.

5. $\phi_n(x) = x^n$. Here $\phi(x) = 0$, $(-1 < x < 1)$; $\phi(x) = 1$, $(x = 1)$; and $\phi(x)$ is not defined for any other value of x.

6. $\phi_n(x) = x^n(1 - x)$. Here $\phi(x)$ differs from the $\phi(x)$ of Ex. 5 in that it has the value 0 when $x = 1$.

7. $\phi_n(x) = x^n/n$. Here $\phi(x)$ differs from the $\phi(x)$ of Ex. 6 in that it has the value 0 when $x = -1$ as well as when $x = 1$.

8. $\phi_n(x) = x^n/(x^n + 1)$. [$\phi(x) = 0$, $(-1 < x < 1)$; $\phi(x) = \frac{1}{2}$, $(x = 1)$; $\phi(x) = 1$, $(x < -1$ or $x > 1)$; and $\phi(x)$ is not defined when $x = -1$.]

9. $\phi_n(x) = x^n/(x^n - 1)$, $1/(x^n + 1)$, $1/(x^n - 1)$, $1/(x^n + x^{-n})$, $1/(x^n - x^{-n})$.

10. $\phi_n(x) = (x^n - 1)/(x^n + 1)$, $(nx^n - 1)/(nx^n + 1)$, $(x^n - n)/(x^n + n)$. [In the first case $\phi(x) = 1$ when $|x| > 1$, $\phi(x) = -1$ when $|x| < 1$, $\phi(x) = 0$ when $x = 1$ and $\phi(x)$ is not defined when $x = -1$. The second and third functions differ from the first in that they are defined both when $x = 1$ and when $x = -1$: the second has the value 1 and the third the value -1 for both these values of x.]

11. Construct an example in which $\phi(x) = 1$, $(|x| > 1)$; $\phi(x) = -1$, $(|x| < 1)$; and $\phi(x) = 0$, $(x = 1$ and $x = -1)$.

12. $\phi_n(x) = x\{(x^{2n} - 1)/(x^{2n} + 1)\}^2$, $n/(x^n + x^{-n} + n)$.

13. $\phi_n(x) = \{x^n f(x) + g(x)\}/(x^n + 1)$. [Here $\phi(x) = f(x)$, $(|x| > 1)$; $\phi(x) = g(x)$, $(|x| < 1)$; $\phi(x) = \frac{1}{2}\{f(x) + g(x)\}$, $(x = 1)$; and $\phi(x)$ is undefined when $x = -1$.]

14. $\phi_n(x) = (2/\pi) \arctan(nx)$. [$\phi(x) = 1$, $(x > 0)$; $\phi(x) = 0$, $(x = 0)$; $\phi(x) = -1$, $(x < 0)$. This function is important in the Theory of Numbers, and is usually denoted by *sgn* x.]

15. $\phi_n(x) = \sin nx\pi$. [$\phi(x) = 0$ when x is an integer; and $\phi(x)$ is otherwise undefined (Ex. XXIV. 7).]

16. If $\phi_n(x) = \sin(n!\, x\pi)$ then $\phi(x) = 0$ for all rational values of x (Ex. XXIV. 14). [The consideration of irrational values presents greater difficul-

ties.]

17. $\phi_n(x) = (\cos^2 x\pi)^n$. [$\phi(x) = 0$ except when x is integral, when $\phi(x) = 1$.]

18. If $N \geqq 1752$ then the number of days in the year N A.D. is

$$\lim\{365 + (\cos^2 \tfrac{1}{4}N\pi)^n - (\cos^2 \tfrac{1}{100}N\pi)^n + (\cos^2 \tfrac{1}{400}N\pi)^n\}.$$

80. The bounds of a bounded aggregate. Let S be any system or aggregate of real numbers s. If there is a number K such that $s \leqq K$ for every s of S, we say that S is *bounded above*. If there is a number k such that $s \geqq k$ for every s, we say that S is *bounded below*. If S is both bounded above and bounded below, we say simply that S is *bounded*.

Suppose first that S is bounded above (but not necessarily below). There will be an infinity of numbers which possess the property possessed by K; any number greater than K, for example, possesses it. We shall prove that *among these numbers there is a least*,* which we shall call M. This number M is not exceeded by any member of S, but every number less than M is exceeded by at least one member of S.

We divide the real numbers ξ into two classes L and R, putting ξ into L or R according as it is or is not exceeded by members of S. Then every ξ belongs to one and one only of the classes L and R. Each class exists; for any number less than any member of S belongs to L, while K belongs to R. Finally, any member of L is less than some member of S, and therefore less than any member of R. Thus the three conditions of Dedekind's Theorem (§ 17) are satisfied, and there is a number M dividing the classes.

The number M is the number whose existence we had to prove. In the first place, M cannot be exceeded by any member of S. For if there were such a member s of S, we could write $s = M + \eta$, where η is positive. The number $M + \frac{1}{2}\eta$ would then belong to L, because it is less than s, and to R, because it is greater than M; and this is impossible. On the other hand, any number less

*An infinite aggregate of numbers does not necessarily possess a least member. The set consisting of the numbers

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \dots,\ \frac{1}{n},\ \dots,$$

for example, has no least member.

than M belongs to L, and is therefore exceeded by at least one member of S. Thus M has all the properties required.

This number M we call the *upper bound* of S, and we may enunciate the following theorem. *Any aggregate S which is bounded above has an upper bound M. No member of S exceeds M; but any number less than M is exceeded by at least one member of S.*

In exactly the same way we can prove the corresponding theorem for an aggregate bounded below (but not necessarily above). *Any aggregate S which is bounded below has a lower bound m. No member of S is less than m; but there is at least one member of S which is less than any number greater than m.*

It will be observed that, when S is bounded above, $M \leqq K$, and when S is bounded below, $m \geqq k$. When S is bounded, $k \leqq m \leqq M \leqq K$.

81. The bounds of a bounded function. Suppose that $\phi(n)$ is a function of the positive integral variable n. The aggregate of all the values $\phi(n)$ defines a set S, to which we may apply all the arguments of § 80. If S is bounded above, or bounded below, or bounded, we say that $\phi(n)$ is bounded above, or bounded below, or bounded. If $\phi(n)$ is bounded above, that is to say if there is a number K such that $\phi(n) \leqq K$ for all values of n, then there is a number M such that

(i) *$\phi(n) \leqq M$ for all values of n;*

(ii) *if ϵ is any positive number then $\phi(n) > M - \epsilon$ for at least one value of n.* This number M we call the **upper bound** of $\phi(n)$. Similarly, if $\phi(n)$ is bounded below, that is to say if there is a number k such that $\phi(n) \leqq k$ for all values of n, then there is a number m such that

(i) *$\phi(n) \geqq m$ for all values of n;*

(ii) *if ϵ is any positive number then $\phi(n) < m + \epsilon$ for at least one value of n.* This number m we call the **lower bound** of $\phi(n)$.

If K exists, $M \leqq K$; if k exists, $m \geqq k$; and if both k and K exist then

$$k \leqq m \leqq M \leqq K.$$

82. The limits of indetermination of a bounded function. Suppose that $\phi(n)$ is a bounded function, and M and m its upper and lower bounds. Let us take any real number ξ, and consider now the relations of inequality which may hold between ξ and the values assumed by $\phi(n)$ for *large* values of n. There are three mutually exclusive possibilities:

(1) $\xi \geqq \phi(n)$ for all sufficiently large values of n;

(2) $\xi \leqq \phi(n)$ for all sufficiently large values of n;

(3) $\xi < \phi(n)$ for an infinity of values of n, and also $\xi > \phi(n)$ for an infinity of values of n.

In case (1) we shall say that ξ is a *superior* number, in case (2) that it is an *inferior* number, and in case (3) that it is an *intermediate* number. It is plain that no superior number can be less than m, and no inferior number greater than M.

Let us consider the aggregate of all superior numbers. It is bounded below, since none of its members are less than m, and has therefore a lower bound, which we shall denote by Λ. Similarly the aggregate of inferior numbers has an upper bound, which we denote by λ.

We call Λ and λ respectively the *upper and lower limits of indetermination of $\phi(n)$ as n tends to infinity*; and write

$$\Lambda = \overline{\lim}\, \phi(n), \quad \lambda = \underline{\lim}\, \phi(n).$$

These numbers have the following properties:

(1) $m \leqq \lambda \leqq \Lambda \leqq M$;

(2) Λ and λ are the upper and lower bounds of the aggregate of intermediate numbers, if any such exist;

(3) if ϵ is any positive number, then $\phi(n) < \Lambda + \epsilon$ for all sufficiently large values of n, and $\phi(n) > \Lambda - \epsilon$ for an infinity of values of n;

(4) similarly $\phi(n) > \lambda - \epsilon$ for all sufficiently large values of n, and $\phi(n) < \lambda + \epsilon$ for an infinity of values of n;

(5) the necessary and sufficient condition that $\phi(n)$ should tend to a limit is that $\Lambda = \lambda$, and in this case the limit is l, the common value of λ and Λ.

Of these properties, (1) is an immediate consequence of the definitions; and we can prove (2) as follows. If $\Lambda = \lambda = l$, there can be at most one intermediate number, viz. l, and there is nothing to prove. Suppose then that $\Lambda > \lambda$. Any intermediate number ξ is less than any superior and greater than any inferior number, so that $\lambda \leqq \xi \leqq \Lambda$. But if $\lambda < \xi < \Lambda$ then ξ must be intermediate, since it is plainly neither superior nor inferior. Hence there are intermediate numbers as near as we please to either λ or Λ.

To prove (3) we observe that $\Lambda + \epsilon$ is superior and $\Lambda - \epsilon$ intermediate or inferior. The result is then an immediate consequence of the definitions; and the proof of (4) is substantially the same.

Finally (5) may be proved as follows. If $\Lambda = \lambda = l$, then

$$l - \epsilon < \phi(n) < l + \epsilon$$

for every positive value of ϵ and all sufficiently large values of n, so that $\phi(n) \to l$. Conversely, if $\phi(n) \to l$, then the inequalities above written hold for all sufficiently large values of n. Hence $l - \epsilon$ is inferior and $l + \epsilon$ superior, so that

$$\lambda \geqq l - \epsilon, \quad \Lambda \leqq l + \epsilon,$$

and therefore $\Lambda - \lambda \leqq 2\epsilon$. As $\Lambda - \lambda \geqq 0$, this can only be true if $\Lambda = \lambda$.

Examples XXXII. 1. Neither Λ nor λ is affected by any alteration in any finite number of values of $\phi(n)$.

2. If $\phi(n) = a$ for all values of n, then $m = \lambda = \Lambda = M = a$.

3. If $\phi(n) = 1/n$, then $m = \lambda = \Lambda = 0$ and $M = 1$.

4. If $\phi(n) = (-1)^n$, then $m = \lambda = -1$ and $\Lambda = M = 1$.

5. If $\phi(n) = (-1)^n/n$, then $m = -1$, $\lambda = \Lambda = 0$, $M = \frac{1}{2}$.

6. If $\phi(n) = (-1)^n\{1 + (1/n)\}$, then $m = -2$, $\lambda = -1$, $\Lambda = 1$, $M = \frac{3}{2}$.

7. Let $\phi(n) = \sin n\theta\pi$, where $\theta > 0$. If θ is an integer then $m = \lambda = \Lambda = M = 0$. If θ is rational but not integral a variety of cases arise. Suppose, *e.g.*, that $\theta = p/q$, p and q being positive, odd, and prime to one another, and $q > 1$. Then $\phi(n)$ assumes the cyclical sequence of values

$$\sin(p\pi/q), \quad \sin(2p\pi/q), \ \dots, \quad \sin\{(2q-1)p\pi/q\}, \quad \sin(2qp\pi/q), \ \dots.$$

It is easily verified that the numerically greatest and least values of $\phi(n)$ are $\cos(\pi/2q)$ and $-\cos(\pi/2q)$, so that

$$m = \lambda = -\cos(\pi/2q), \quad \Lambda = M = \cos(\pi/2q).$$

The reader may discuss similarly the cases which arise when p and q are not both odd.

The case in which θ is irrational is more difficult: it may be shown that in this case $m = \lambda = -1$ and $\Lambda = M = 1$. It may also be shown that the values of $\phi(n)$ are scattered all over the interval $[-1, 1]$ in such a way that, if ξ is *any* number of the interval, then there is a sequence n_1, n_2, ... such that $\phi(n_k) \to \xi$ as $k \to \infty$.*

The results are very similar when $\phi(n)$ is the fractional part of $n\theta$.

83. The general principle of convergence for a bounded function. The results of the preceding sections enable us to formulate a very important necessary and sufficient condition that a bounded function $\phi(n)$ should tend to a limit, a condition usually referred to as *the general principle of convergence* to a limit.

THEOREM 1. *The necessary and sufficient condition that a bounded function $\phi(n)$ should tend to a limit is that, when any positive number ϵ is given, it should be possible to find a number $n_0(\epsilon)$ such that*

$$|\phi(n_2) - \phi(n_1)| < \epsilon$$

for all values of n_1 and n_2 such that $n_2 > n_1 \geqq n_0(\epsilon)$.

In the first place, the condition is *necessary*. For if $\phi(n) \to l$ then we can find n_0 so that

$$l - \tfrac{1}{2}\epsilon < \phi(n) < l + \tfrac{1}{2}\epsilon$$

when $n \geqq n_0$, and so

$$|\phi(n_2) - \phi(n_1)| < \epsilon \tag{1}$$

when $n_1 \geqq n_0$ and $n_2 \geqq n_0$.

In the second place, the condition is *sufficient*. In order to prove this we have only to show that it involves $\lambda = \Lambda$. But if $\lambda < \Lambda$ then there are, however small ϵ may be, infinitely many values of n such that $\phi(n) < \lambda + \epsilon$ and infinitely many such that $\phi(n) > \Lambda - \epsilon$; and therefore we can find values of n_1 and n_2, each greater than any assigned number n_0, and such that

$$\phi(n_2) - \phi(n_1) > \Lambda - \lambda - 2\epsilon,$$

which is greater than $\frac{1}{2}(\Lambda - \lambda)$ if ϵ is small enough. This plainly contradicts the inequality (1). Hence $\lambda = \Lambda$, and so $\phi(n)$ tends to a limit.

*A number of simple proofs of this result are given by Hardy and Littlewood, "Some Problems of Diophantine Approximation", *Acta Mathematica*, vol. xxxvii.

84. Unbounded functions. So far we have restricted ourselves to bounded functions; but the 'general principle of convergence' is the same for unbounded as for bounded functions, and the words '*a bounded function*' may be omitted from the enunciation of Theorem 1.

In the first place, if $\phi(n)$ tends to a limit l then it is certainly bounded; for all but a finite number of its values are less than $l+\epsilon$ and greater than $l-\epsilon$.

In the second place, if the condition of Theorem 1 is satisfied, we have

$$|\phi(n_2) - \phi(n_1)| < \epsilon$$

whenever $n_1 \geqq n_0$ and $n_2 \geqq n_0$. Let us choose some particular value n_1 greater than n_0. Then

$$\phi(n_1) - \epsilon < \phi(n_2) < \phi(n_1) + \epsilon$$

when $n_2 \geqq n_0$. Hence $\phi(n)$ is bounded; and so the second part of the proof of the last section applies also.

The theoretical importance of the 'general principle of convergence' can hardly be overestimated. Like the theorems of § 69, it gives us a means of deciding whether a function $\phi(n)$ tends to a limit or not, without requiring us to be able to tell beforehand what the limit, if it exists, must be; and it has not the limitations inevitable in theorems of such a special character as those of § 69. But in elementary work it is generally possible to dispense with it, and to obtain all we want from these special theorems. And it will be found that, in spite of the importance of the principle, practically no applications are made of it in the chapters which follow.* We will only remark that, if we suppose that

$$\phi(n) = s_n = u_1 + u_2 + \dots + u_n,$$

we obtain at once a necessary and sufficient condition for the convergence of an infinite series, viz:

THEOREM 2. *The necessary and sufficient condition for the convergence of the series $u_1 + u_2 + \dots$ is that, given any positive number ϵ, it should be possible to find n_0 so that*

$$|u_{n_1+1} + u_{n_1+2} + \dots + u_{n_2}| < \epsilon$$

for all values of n_1 and n_2 such that $n_2 > n_1 \geqq n_0$.

---

*A few proofs given in Ch. VIII can be simplified by the use of the principle.

85. Limits of complex functions and series of complex terms. In this chapter we have, up to the present, concerned ourselves only with real functions of n and series all of whose terms are real. There is however no difficulty in extending our ideas and definitions to the case in which the functions or the terms of the series are complex.

Suppose that $\phi(n)$ is complex and equal to

$$\rho(n) + i\sigma(n),$$

where $\rho(n)$, $\sigma(n)$ are real functions of n. Then *if $\rho(n)$ and $\sigma(n)$ converge respectively to limits r and s as $n \to \infty$, we shall say that $\phi(n)$ converges to the limit $l = r + is$, and write*

$$\lim \phi(n) = l.$$

Similarly, when u_n is complex and equal to $v_n + iw_n$, we shall say that *the series*

$$u_1 + u_2 + u_3 + \dots$$

is convergent and has the sum $l = r + is$, if the series

$$v_1 + v_2 + v_3 + \dots, \quad w_1 + w_2 + w_3 + \dots$$

are convergent and have the sums r, s respectively.

To say that $u_1 + u_2 + u_3 + \dots$ is convergent and has the sum l is of course the same as to say that the sum

$$s_n = u_1 + u_2 + \dots + u_n = (v_1 + v_2 + \dots + v_n) + i(w_1 + w_2 + \dots + w_n)$$

converges to the limit l as $n \to \infty$.

In the case of real functions and series we also gave definitions of *divergence* and *oscillation*, *finite* or *infinite*. But in the case of complex functions and series, where we have to consider the behaviour both of $\rho(n)$ and of $\sigma(n)$, there are so many possibilities that this is hardly worth while. When it is necessary to make further distinctions of this kind, we shall make them by stating the way in which the real or imaginary parts behave when taken separately.

86. The reader will find no difficulty in proving such theorems as the following, which are obvious extensions of theorems already proved for real functions and series.

(1) If $\lim \phi(n) = l$ then $\lim \phi(n+p) = l$ for any fixed value of p.

(2) If $u_1 + u_2 + \dots$ is convergent and has the sum l, then $a + b + c + \dots + k + u_1 + u_2 + \dots$ is convergent and has the sum $a + b + c + \dots + k + l$, and $u_{p+1} + u_{p+2} + \dots$ is convergent and has the sum $l - u_1 - u_2 - \dots - u_p$.

(3) If $\lim \phi(n) = l$ and $\lim \psi(n) = m$, then

$$\lim\{\phi(n) + \psi(n)\} = l + m.$$

(4) If $\lim \phi(n) = l$, then $\lim k\phi(n) = kl$.

(5) If $\lim \phi(n) = l$ and $\lim \psi(n) = m$, then $\lim \phi(n)\psi(n) = lm$.

(6) If $u_1 + u_2 + \dots$ converges to the sum l, and $v_1 + v_2 + \dots$ to the sum m, then $(u_1 + v_1) + (u_2 + v_2) + \dots$ converges to the sum $l + m$.

(7) If $u_1 + u_2 + \dots$ converges to the sum l then $ku_1 + ku_2 + \dots$ converges to the sum kl.

(8) If $u_1 + u_2 + u_3 + \dots$ is convergent then $\lim u_n = 0$.

(9) If $u_1 + u_2 + u_3 + \dots$ is convergent, then so is any series formed by grouping the terms in brackets, and the sums of the two series are the same.

As an example, let us prove theorem (5). Let

$$\phi(n) = \rho(n) + i\sigma(n), \quad \psi(n) = \rho'(n) + i\sigma'(n), \quad l = r + is, \quad m = r' + is'.$$

Then

$$\rho(n) \to r, \quad \sigma(n) \to s, \quad \rho'(n) \to r', \quad \sigma'(n) \to s'.$$

But

$$\phi(n)\psi(n) = \rho\rho' - \sigma\sigma' + i(\rho\sigma' + \rho'\sigma),$$

and

$$\rho\rho' - \sigma\sigma' \to rr' - ss', \quad \rho\sigma' + \rho'\sigma \to rs' + r's;$$

so that

$$\phi(n)\psi(n) \to rr' - ss' + i(rs' + r's),$$

i.e.

$$\phi(n)\psi(n) \to (r+is)(r'+is') = lm.$$

The following theorems are of a somewhat different character.

(10) *In order that $\phi(n) = \rho(n)+i\sigma(n)$ should converge to zero as $n \to \infty$, it is necessary and sufficient that*

$$|\phi(n)| = \sqrt{\{\rho(n)\}^2 + \{\sigma(n)\}^2}$$

should converge to zero.

If $\rho(n)$ and $\sigma(n)$ both converge to zero then it is plain that $\sqrt{\rho^2+\sigma^2}$ does so. The converse follows from the fact that the numerical value of ρ or σ cannot be greater than $\sqrt{\rho^2+\sigma^2}$.

(11) *More generally, in order that $\phi(n)$ should converge to a limit l, it is necessary and sufficient that*

$$|\phi(n) - l|$$

should converge to zero.

For $\phi(n) - l$ converges to zero, and we can apply (10).

(12) *Theorems* 1 *and* 2 *of* §§ 83–84 *are still true when $\phi(n)$ and u_n are complex.*

We have to show that the necessary and sufficient condition that $\phi(n)$ should tend to l is that

$$|\phi(n_2) - \phi(n_1)| < \epsilon \tag{1}$$

when $n_2 > n_1 \geqq n_0$.

If $\phi(n) \to l$ then $\rho(n) \to r$ and $\sigma(n) \to s$, and so we can find numbers n_0' and n_0'' depending on ϵ and such that

$$|\rho(n_2) - \rho(n_1)| < \tfrac{1}{2}\epsilon, \quad |\sigma(n_2) - \sigma(n_1)| < \tfrac{1}{2}\epsilon,$$

the first inequality holding when $n_2 > n_1 \geqq n_0'$, and the second when $n_2 > n_1 \geqq n_0''$. Hence

$$|\phi(n_2) - \phi(n_1)| \leqq |\rho(n_2) - \rho(n_1)| + |\sigma(n_2) - \sigma(n_1)| < \epsilon$$

when $n_2 > n_1 \geqq n_0$, where n_0 is the greater of n_0' and n_0''. Thus the condition (1) is *necessary*. To prove that it is *sufficient* we have only to observe that

$$|\rho(n_2) - \rho(n_1)| \leqq |\phi(n_2) - \phi(n_1)| < \epsilon$$

when $n_2 > n_1 \geqq n_0$. Thus $\rho(n)$ tends to a limit r, and in the same way it may be shown that $\sigma(n)$ tends to a limit s.

87. The limit of z^n as $n \to \infty$, z being any complex number. Let us consider the important case in which $\phi(n) = z^n$. This problem has already been discussed for real values of z in § 72.

If $z^n \to l$ then $z^{n+1} \to l$, by (1) of § 86. But, by (4) of § 86,

$$z^{n+1} = zz^n \to zl,$$

and therefore $l = zl$, which is only possible if (a) $l = 0$ or (b) $z = 1$. If $z = 1$ then $\lim z^n = 1$. Apart from this special case the limit, if it exists, can only be zero.

Now if $z = r(\cos\theta + i\sin\theta)$, where r is positive, then

$$z^n = r^n(\cos n\theta + i\sin n\theta),$$

so that $|z^n| = r^n$. Thus $|z^n|$ tends to zero if and only if $r < 1$; and it follows from (10) of § 86 that

$$\lim z^n = 0$$

if and only if $r < 1$. In no other case does z^n converge to a limit, except when $z = 1$ and $z^n \to 1$.

88. The geometric series $1 + z + z^2 + \dots$ when z is complex. Since

$$s_n = 1 + z + z^2 + \dots + z^{n-1} = (1 - z^n)/(1 - z),$$

unless $z = 1$, when the value of s_n is n, it follows that *the series* $1 + z + z^2 + \dots$ *is convergent if and only if* $r = |z| < 1$. *And its sum when convergent is* $1/(1 - z)$.

Thus if $z = r(\cos\theta + i\sin\theta) = r\,\mathrm{Cis}\,\theta$, and $r < 1$, we have

$$1 + z + z^2 + \dots = 1/(1 - r\,\mathrm{Cis}\,\theta),$$

or

$$\begin{aligned}1 + r\,\mathrm{Cis}\,\theta + r^2\,\mathrm{Cis}\,2\theta + \dots &= 1/(1 - r\,\mathrm{Cis}\,\theta)\\ &= (1 - r\cos\theta + ir\sin\theta)/(1 - 2r\cos\theta + r^2).\end{aligned}$$

Separating the real and imaginary parts, we obtain

$$1 + r\cos\theta + r^2\cos 2\theta + \dots = (1 - r\cos\theta)/(1 - 2r\cos\theta + r^2),$$
$$r\sin\theta + r^2\sin 2\theta + \dots = r\sin\theta/(1 - 2r\cos\theta + r^2),$$

provided $r < 1$. If we change θ into $\theta + \pi$, we see that these results hold also for negative values of r numerically less than 1. Thus they hold when $-1 < r < 1$.

Examples XXXIII. 1. Prove directly that $\phi(n) = r^n \cos n\theta$ converges to 0 when $r < 1$ and to 1 when $r = 1$ and θ is a multiple of 2π. Prove further that if $r = 1$ and θ is not a multiple of 2π, then $\phi(n)$ oscillates finitely; if $r > 1$ and θ is a multiple of 2π, then $\phi(n) \to +\infty$; and if $r > 1$ and θ is not a multiple of 2π, then $\phi(n)$ oscillates infinitely.

2. Establish a similar series of results for $\phi(n) = r^n \sin n\theta$.

3. Prove that

$$z^m + z^{m+1} + \dots = z^m/(1 - z),$$
$$z^m + 2z^{m+1} + 2z^{m+2} + \dots = z^m(1 + z)/(1 - z),$$

if and only if $|z| < 1$. Which of the theorems of § 86 do you use?

4. Prove that if $-1 < r < 1$ then

$$1 + 2r\cos\theta + 2r^2\cos 2\theta + \dots = (1 - r^2)/(1 - 2r\cos\theta + r^2).$$

5. The series

$$1 + \frac{z}{1+z} + \left(\frac{z}{1+z}\right)^2 + \dots$$

converges to the sum $1\Big/\left(1 - \dfrac{z}{1+z}\right) = 1 + z$ if $|z/(1+z)| < 1$. Show that this condition is equivalent to the condition that z has a real part greater than $-\frac{1}{2}$.

MISCELLANEOUS EXAMPLES ON CHAPTER IV.

1. The function $\phi(n)$ takes the values 1, 0, 0, 0, 1, 0, 0, 0, 1, ... when $n = 0, 1, 2, \dots$. Express $\phi(n)$ in terms of n by a formula which does not involve trigonometrical functions. [$\phi(n) = \frac{1}{4}\{1 + (-1)^n + i^n + (-i)^n\}$.]

2. If $\phi(n)$ steadily increases, and $\psi(n)$ steadily decreases, as n tends to ∞, and if $\psi(n) > \phi(n)$ for all values of n, then both $\phi(n)$ and $\psi(n)$ tend to limits, and $\lim \phi(n) \leqq \lim \psi(n)$. [This is an immediate corollary from § 69.]

3. Prove that, if

$$\phi(n) = \left(1 + \frac{1}{n}\right)^n, \quad \psi(n) = \left(1 - \frac{1}{n}\right)^{-n},$$

then $\phi(n+1) > \phi(n)$ and $\psi(n+1) < \psi(n)$. [The first result has already been proved in § 73.]

4. Prove also that $\psi(n) > \phi(n)$ for all values of n: and deduce (by means of the preceding examples) that both $\phi(n)$ and $\psi(n)$ tend to limits as n tends to ∞.*

5. The arithmetic mean of the products of all distinct pairs of positive integers whose sum is n is denoted by S_n. Show that $\lim(S_n/n^2) = 1/6$.

(*Math. Trip.* 1903.)

6. Prove that if $x_1 = \frac{1}{2}\{x + (A/x)\}$, $x_2 = \frac{1}{2}\{x_1 + (A/x_1)\}$, and so on, x and A being positive, then $\lim x_n = \sqrt{A}$.

[Prove first that $\dfrac{x_n - \sqrt{A}}{x_n + \sqrt{A}} = \left(\dfrac{x - \sqrt{A}}{x + \sqrt{A}}\right)^{2^n}$.]

7. If $\phi(n)$ is a positive integer for all values of n, and tends to ∞ with n, then $x^{\phi(n)}$ tends to 0 if $0 < x < 1$ and to $+\infty$ if $x > 1$. Discuss the behaviour of $x^{\phi(n)}$, as $n \to \infty$, for other values of x.

8.† If a_n increases or decreases steadily as n increases, then the same is true of $(a_1 + a_2 + \dots + a_n)/n$.

9. If $x_{n+1} = \sqrt{k + x_n}$, and k and x_1 are positive, then the sequence x_1, x_2, x_3, ... is an increasing or decreasing sequence according as x_1 is less than or

---

*A proof that $\lim\{\psi(n) - \phi(n)\} = 0$, and that therefore each function tends to the limit e, will be found in Chrystal's *Algebra*, vol. ii, p. 78. We shall however prove this in Ch. IX by a different method.

†Exs. 8–12 are taken from Bromwich's *Infinite Series*.

greater than α, the positive root of the equation $x^2 = x + k$; and in either case $x_n \to \alpha$ as $n \to \infty$.

10. If $x_{n+1} = k/(1 + x_n)$, and k and x_1 are positive, then the sequences $x_1, x_3, x_5, \dots$ and $x_2, x_4, x_6, \dots$ are one an increasing and the other a decreasing sequence, and each sequence tends to the limit α, the positive root of the equation $x^2 + x = k$.

11. The function $f(x)$ is increasing and continuous (see Ch. V) for all values of x, and a sequence $x_1, x_2, x_3, \dots$ is defined by the equation $x_{n+1} = f(x_n)$. Discuss on general graphical grounds the question as to whether x_n tends to a root of the equation $x = f(x)$. Consider in particular the case in which this equation has only one root, distinguishing the cases in which the curve $y = f(x)$ crosses the line $y = x$ from above to below and from below to above.

12. If x_1, x_2 are positive and $x_{n+1} = \frac{1}{2}(x_n + x_{n-1})$, then the sequences $x_1, x_3, x_5, \dots$ and $x_2, x_4, x_6, \dots$ are one a decreasing and the other an increasing sequence, and they have the common limit $\frac{1}{3}(x_1 + 2x_2)$.

13. Draw a graph of the function y defined by the equation

$$y = \lim_{n\to\infty} \frac{x^{2n} \sin \frac{1}{2}\pi x + x^2}{x^{2n} + 1}.$$

(*Math. Trip.* 1901.)

14. The function

$$y = \lim_{n\to\infty} \frac{1}{1 + n \sin^2 \pi x}$$

is equal to 0 except when x is an integer, and then equal to 1. The function

$$y = \lim_{n\to\infty} \frac{\psi(x) + n\phi(x) \sin^2 \pi x}{1 + n \sin^2 \pi x}$$

is equal to $\phi(x)$ unless x is an integer, and then equal to $\psi(x)$.

15. Show that the graph of the function

$$y = \lim_{n\to\infty} \frac{x^n \phi(x) + x^{-n} \psi(x)}{x^n + x^{-n}}$$

is composed of parts of the graphs of $\phi(x)$ and $\psi(x)$, together with (as a rule) two isolated points. Is y defined when (*a*) $x = 1$, (*b*) $x = -1$, (*c*) $x = 0$?

16. Prove that the function y which is equal to 0 when x is rational, and to 1 when x is irrational, may be represented in the form

$$y = \lim_{m\to\infty} sgn\,\{\sin^2(m!\,\pi x)\},$$

where

$$sgn\,x = \lim_{n\to\infty} (2/\pi)\arctan(nx),$$

as in Ex. XXXI. 14. [If x is rational then $\sin^2(m!\,\pi x)$, and therefore $sgn\,\{\sin^2(m!\,\pi x)\}$, is equal to zero from a certain value of m onwards: if x is irrational then $\sin^2(m!\,\pi x)$ is always positive, and so $sgn\,\{\sin^2(m!\,\pi x)\}$ is always equal to 1.]

Prove that y may also be represented in the form

$$1 - \lim_{m\to\infty}\bigl[\lim_{n\to\infty}\{\cos(m!\,\pi x)\}^{2n}\bigr].$$

17. Sum the series

$$\sum_1^\infty \frac{1}{\nu(\nu+1)},\quad \sum_1^\infty \frac{1}{\nu(\nu+1)\dots(\nu+k)}.$$

[Since

$$\frac{1}{\nu(\nu+1)\dots(\nu+k)} = \frac{1}{k}\left\{\frac{1}{\nu(\nu+1)\dots(\nu+k-1)} - \frac{1}{(\nu+1)(\nu+2)\dots(\nu+k)}\right\},$$

we have

$$\sum_1^n \frac{1}{\nu(\nu+1)\dots(\nu+k)} = \frac{1}{k}\left\{\frac{1}{1\cdot 2\dots k} - \frac{1}{(n+1)(n+2)\dots(n+k)}\right\}$$

and so

$$\sum_1^\infty \frac{1}{\nu(\nu+1)\dots(\nu+k)} = \frac{1}{k(k!)}.]$$

18. If $|z| < |\alpha|$, then

$$\frac{L}{z-\alpha} = -\frac{L}{\alpha}\left(1 + \frac{z}{\alpha} + \frac{z^2}{\alpha^2} + \dots\right);$$

and if $|z| > |\alpha|$, then

$$\frac{L}{z-\alpha} = \frac{L}{z}\left(1 + \frac{\alpha}{z} + \frac{\alpha^2}{z^2} + \dots\right).$$

19. **Expansion of** $(Az + B)/(az^2 + 2bz + c)$ **in powers of** z**.** Let α, β be the roots of $az^2 + 2bz + c = 0$, so that $az^2 + 2bz + c = a(z - \alpha)(z - \beta)$. We shall suppose that A, B, a, b, c are all real, and α and β unequal. It is then easy to verify that

$$\frac{Az + B}{az^2 + 2bz + c} = \frac{1}{a(\alpha - \beta)}\left(\frac{A\alpha + B}{z - \alpha} - \frac{A\beta + B}{z - \beta}\right).$$

There are two cases, according as $b^2 > ac$ or $b^2 < ac$.

(1) If $b^2 > ac$ then the roots α, β are real and distinct. If $|z|$ is less than either $|\alpha|$ or $|\beta|$ we can expand $1/(z - \alpha)$ and $1/(z - \beta)$ in ascending powers of z (Ex. 18). If $|z|$ is greater than either $|\alpha|$ or $|\beta|$ we must expand in descending powers of z; while if $|z|$ lies between $|\alpha|$ and $|\beta|$ one fraction must be expanded in ascending and one in descending powers of z. The reader should write down the actual results. If $|z|$ is equal to $|\alpha|$ or $|\beta|$ then no such expansion is possible.

(2) If $b^2 < ac$ then the roots are conjugate complex numbers (Ch. III § 43), and we can write

$$\alpha = \rho \operatorname{Cis}\phi, \quad \beta = \rho \operatorname{Cis}(-\phi),$$

where $\rho^2 = \alpha\beta = c/a$, $\rho\cos\phi = \frac{1}{2}(\alpha + \beta) = -b/a$, so that $\cos\phi = -\sqrt{b^2/ac}$, $\sin\phi = \sqrt{1 - (b^2/ac)}$.

If $|z| < \rho$ then each fraction may be expanded in ascending powers of z. The coefficient of z^n will be found to be

$$\frac{A\rho\sin n\phi + B\sin\{(n+1)\phi\}}{a\rho^{n+1}\sin\phi}.$$

If $|z| > \rho$ we obtain a similar expansion in descending powers, while if $|z| = \rho$ no such expansion is possible.

20. Show that if $|z| < 1$ then

$$1 + 2z + 3z^2 + \dots + (n+1)z^n + \dots = 1/(1-z)^2.$$

[The sum to n terms is $\dfrac{1-z^n}{(1-z)^2} - \dfrac{nz^n}{1-z}$.]

21. Expand $L/(z-\alpha)^2$ in powers of z, ascending or descending according as $|z| < |\alpha|$ or $|z| > |\alpha|$.

22. Show that if $b^2 = ac$ and $|az| < |b|$ then

$$\frac{Az+B}{az^2+2bz+c} = \sum_0^\infty p_n z^n,$$

where $p_n = \{(-a)^n/b^{n+2}\}\{(n+1)aB - nbA\}$; and find the corresponding expansion, in descending powers of z, which holds when $|az| > |b|$.

23. Verify the result of Ex. 19 in the case of the fraction $1/(1+z^2)$. [We have $1/(1+z^2) = \sum z^n \sin\{\frac{1}{2}(n+1)\pi\} = 1 - z^2 + z^4 - \dots$.]

24. Prove that if $|z| < 1$ then

$$\frac{1}{1+z+z^2} = \frac{2}{\sqrt{3}} \sum_0^\infty z^n \sin\{\tfrac{2}{3}(n+1)\pi\}.$$

25. Expand $(1+z)/(1+z^2)$, $(1+z^2)/(1+z^3)$ and $(1+z+z^2)/(1+z^4)$ in ascending powers of z. For what values of z do your results hold?

26. If $a/(a+bz+cz^2) = 1 + p_1 z + p_2 z^2 + \dots$ then

$$1 + p_1^2 z + p_2^2 z^2 + \dots = \frac{a+cz}{a-cz}\,\frac{a^2}{a^2-(b^2-2ac)z+c^2z^2}.$$

(*Math. Trip.* 1900.)

27. If $\lim\limits_{n\to\infty} s_n = l$ then

$$\lim_{n\to\infty} \frac{s_1+s_2+\dots+s_n}{n} = l.$$

[Let $s_n = l + t_n$. Then we have to prove that $(t_1+t_2+\dots+t_n)/n$ tends to zero if t_n does so.

We divide the numbers $t_1, t_2, \dots, t_n$ into two sets $t_1, t_2, \dots, t_p$ and $t_{p+1}, t_{p+2}, \dots, t_n$. Here we suppose that p is a function of n which tends to ∞ as $n \to \infty$, but *more slowly than* n, so that $p \to \infty$ and $p/n \to 0$: *e.g.* we might suppose p to be the integral part of $\sqrt{n}$.

Let ϵ be any positive number. However small ϵ may be, we can choose n_0 so that t_{p+1}, t_{p+2}, ..., t_n are all numerically less than $\frac{1}{2}\epsilon$ when $n \geqq n_0$, and so

$$|(t_{p+1} + t_{p+2} + \cdots + t_n)/n| < \tfrac{1}{2}\epsilon(n-p)/n < \tfrac{1}{2}\epsilon.$$

But, if A is the greatest of the moduli of all the numbers t_1, t_2, ..., we have

$$|(t_1 + t_2 + \cdots + t_p)/n| < pA/n,$$

and this also will be less than $\frac{1}{2}\epsilon$ when $n \geqq n_0$, if n_0 is large enough, since $p/n \to 0$ as $n \to \infty$. Thus

$$|(t_1 + t_2 + \cdots + t_n)/n| \leqq |(t_1 + t_2 + \cdots + t_p)/n| + |(t_{p+1} + \cdots + t_n)/n| < \epsilon$$

when $n \geqq n_0$; which proves the theorem.

The reader, if he desires to become expert in dealing with questions about limits, should study the argument above with great care. It is very often necessary, in proving the limit of some given expression to be zero, to split it into two parts which have to be proved to have the limit zero in slightly different ways. When this is the case the proof is never very easy.

The point of the proof is this: we have to prove that $(t_1 + t_2 + \cdots + t_n)/n$ is small when n is large, the t's being small when their suffixes are large. We split up the terms in the bracket into two groups. The terms in the first group are not all small, but their number is small compared with n. The number in the second group is *not* small compared with n, but the terms are all small, and their number at any rate less than n, so that their sum is small compared with n. Hence each of the parts into which $(t_1+t_2+\cdots+t_n)/n$ has been divided is small when n is large.]

28. If $\phi(n) - \phi(n-1) \to l$ as $n \to \infty$, then $\phi(n)/n \to l$.

[If $\phi(n) = s_1 + s_2 + \cdots + s_n$ then $\phi(n) - \phi(n-1) = s_n$, and the theorem reduces to that proved in the last example.]

29. If $s_n = \frac{1}{2}\{1 - (-1)^n\}$, so that s_n is equal to 1 or 0 according as n is odd or even, then $(s_1 + s_2 + \cdots + s_n)/n \to \frac{1}{2}$ as $n \to \infty$.

[This example proves that the converse of Ex. 27 is not true: for s_n oscillates as $n \to \infty$.]

30. If c_n, s_n denote the sums of the first n terms of the series

$$\tfrac{1}{2} + \cos\theta + \cos 2\theta + \ldots, \qquad \sin\theta + \sin 2\theta + \ldots,$$

then

$$\lim(c_1 + c_2 + \cdots + c_n)/n = 0, \quad \lim(s_1 + s_2 + \cdots + s_n)/n = \tfrac{1}{2}\cot\tfrac{1}{2}\theta.$$

CHAPTER V

LIMITS OF FUNCTIONS OF A CONTINUOUS VARIABLE. CONTINUOUS AND DISCONTINUOUS FUNCTIONS

89. Limits as x tends to ∞. We shall now return to functions of a continuous real variable. We shall confine ourselves entirely to *one-valued* functions,* and we shall denote such a function by $\phi(x)$. We suppose x to assume successively all values corresponding to points on our fundamental straight line Λ, starting from some definite point on the line and progressing always to the right. In these circumstances we say that x *tends to infinity*, or *to* ∞, and write $x \to \infty$. The only difference between the 'tending of n to ∞' discussed in the last chapter, and this 'tending of x to ∞', is that x assumes all values as it tends to ∞, *i.e.* that the point P which corresponds to x coincides in turn with every point of Λ to the right of its initial position, whereas n tended to ∞ by a series of jumps. We can express this distinction by saying that x tends *continuously* to ∞.

As we explained at the beginning of the last chapter, there is a very close correspondence between functions of x and functions of n. Every function of n may be regarded as a selection from the values of a function of x. In the last chapter we discussed the peculiarities which may characterise the behaviour of a function $\phi(n)$ as n tends to ∞. Now we are concerned with the same problem for a function $\phi(x)$; and the definitions and theorems to which we are led are practically repetitions of those of the last chapter. Thus corresponding to Def. 1 of § 58 we have:

Definition 1. *The function $\phi(x)$ is said to tend to the limit l as x tends to ∞ if, when any positive number ϵ, however small, is assigned, a number $x_0(\epsilon)$ can be chosen such that, for all values of x equal to or greater than $x_0(\epsilon)$, $\phi(x)$ differs from l by less than ϵ, i.e. if*

$$|\phi(x) - l| < \epsilon$$

when $x \geqq x_0(\epsilon)$.

*Thus $\sqrt{x}$ stands in this chapter for the one-valued function $+\sqrt{x}$ and not (as in § 26) for the two-valued function whose values are $+\sqrt{x}$ and $-\sqrt{x}$.

When this is the case we may write

$$\lim_{x\to\infty} \phi(x) = l,$$

or, when there is no risk of ambiguity, simply $\lim \phi(x) = l$, or $\phi(x) \to l$. Similarly we have:

Definition 2. *The function $\phi(x)$ is said to tend to ∞ with x if, when any number Δ, however large, is assigned, we can choose a number $x_0(\Delta)$ such that*

$$\phi(x) > \Delta$$

when $x \geqq x_0(\Delta)$.

We then write

$$\phi(x) \to \infty.$$

Similarly we define $\phi(x) \to -\infty$.* Finally we have:

Definition 3. *If the conditions of neither of the two preceding definitions are satisfied, then $\phi(x)$ is said to oscillate as x tends to ∞. If $|\phi(x)|$ is less than some constant K when $x \geqq x_0$,† then $\phi(x)$ is said to oscillate finitely, and otherwise infinitely.*

The reader will remember that in the last chapter we considered very carefully various less formal ways of expressing the facts represented by the formulae $\phi(n) \to l$, $\phi(n) \to \infty$. Similar modes of expression may of course be used in the present case. Thus we may say that $\phi(x)$ is small or nearly equal to l or large when x is large, using the words 'small', 'nearly', 'large' in a sense similar to that in which they were used in Ch. IV.

---

*We shall sometimes find it convenient to write $+\infty$, $x \to +\infty$, $\phi(x) \to +\infty$ instead of ∞, $x \to \infty$, $\phi(x) \to \infty$.

†In the corresponding definition of § 62, we postulated that $|\phi(n)| < K$ for *all* values of n, and not merely when $n \geqq n_0$. But then the two hypotheses would have been equivalent; for if $|\phi(n)| < K$ when $n \geqq n_0$, then $|\phi(n)| < K'$ for all values of n, where K' is the greatest of $|\phi(1)|$, $|\phi(2)|$, …, $|\phi(n_0 - 1)|$ and K. Here the matter is not quite so simple, as there are infinitely many values of x less than x_0.

Examples XXXIV. 1. Consider the behaviour of the following functions as $x \to \infty$: $1/x$, $1 + (1/x)$, x^2, x^k, $[x]$, $x - [x]$, $[x] + \sqrt{x - [x]}$.

The first four functions correspond exactly to functions of n fully discussed in Ch. IV. The graphs of the last three were constructed in Ch. II (Exs. XVI. 1, 2, 4), and the reader will see at once that $[x] \to \infty$, $x - [x]$ oscillates finitely, and $[x] + \sqrt{x - [x]} \to \infty$.

One simple remark may be inserted here. The function $\phi(x) = x - [x]$ oscillates between 0 and 1, as is obvious from the form of its graph. It is equal to zero whenever x is an integer, so that the function $\phi(n)$ derived from it is always zero and so tends to the limit zero. The same is true if

$$\phi(x) = \sin x\pi, \quad \phi(n) = \sin n\pi = 0.$$

It is evident that $\phi(x) \to l$ or $\phi(x) \to \infty$ or $\phi(x) \to -\infty$ involves the corresponding property for $\phi(n)$, but that the converse is by no means always true.

2. Consider in the same way the functions:

$$(\sin x\pi)/x, \quad x \sin x\pi, \quad (x \sin x\pi)^2, \quad \tan x\pi, \quad a\cos^2 x\pi + b\sin^2 x\pi,$$

illustrating your remarks by means of the graphs of the functions.

3. Give a geometrical explanation of Def. 1, analogous to the geometrical explanation of Ch. IV, § 59.

4. If $\phi(x) \to l$, and l is not zero, then $\phi(x)\cos x\pi$ and $\phi(x)\sin x\pi$ oscillate finitely. If $\phi(x) \to \infty$ or $\phi(x) \to -\infty$, then they oscillate infinitely. The graph of either function is a wavy curve oscillating between the curves $y = \phi(x)$ and $y = -\phi(x)$.

5. Discuss the behaviour, as $x \to \infty$, of the function

$$y = f(x)\cos^2 x\pi + F(x)\sin^2 x\pi,$$

where $f(x)$ and $F(x)$ are some pair of simple functions (*e.g.* x and x^2). [The graph of y is a curve oscillating between the curves $y = f(x)$, $y = F(x)$.]

90. Limits as x tends to $-\infty$. The reader will have no difficulty in framing for himself definitions of the meaning of the assertions 'x tends to $-\infty$', or '$x \to -\infty$' and

$$\lim_{x \to -\infty} \phi(x) = l, \quad \phi(x) \to \infty, \quad \phi(x) \to -\infty.$$

In fact, if $x = -y$ and $\phi(x) = \phi(-y) = \psi(y)$, then y tends to ∞ as x tends to $-\infty$, and the question of the behaviour of $\phi(x)$ as x tends to $-\infty$ is the same as that of the behaviour of $\psi(y)$ as y tends to ∞.

91. Theorems corresponding to those of Ch. IV, §§ 63–67. The theorems concerning the sums, products, and quotients of functions proved in Ch. IV are all true (with obvious verbal alterations which the reader will have no difficulty in supplying) for functions of the continuous variable x. Not only the enunciations but the proofs remain substantially the same.

92. Steadily increasing or decreasing functions. The definition which corresponds to that of § 69 is as follows: *the function $\phi(x)$ will be said to increase steadily with x if $\phi(x_2) \geqq \phi(x_1)$ whenever $x_2 > x_1$*. In many cases, of course, this condition is only satisfied from a definite value of x onwards, *i.e.* when $x_2 > x_1 \geqq x_0$. The theorem which follows in that section requires no alteration but that of n into x: and the proof is the same, except for obvious verbal changes.

If $\phi(x_2) > \phi(x_1)$, the possibility of equality being excluded, whenever $x_2 > x_1$, then $\phi(x)$ will be said to be *steadily increasing in the stricter sense*. We shall find that the distinction is often important (cf. §§ 108–109).

The reader should consider whether or no the following functions increase steadily with x (or at any rate increase steadily from a certain value of x onwards): $x^2 - x$, $x + \sin x$, $x + 2\sin x$, $x^2 + 2\sin x$, $[x]$, $[x] + \sin x$, $[x] + \sqrt{x - [x]}$. All these functions tend to ∞ as $x \to \infty$.

93. Limits as x tends to 0. Let $\phi(x)$ be such a function of x that $\lim\limits_{x\to\infty} \phi(x) = l$, and let $y = 1/x$. Then

$$\phi(x) = \phi(1/y) = \psi(y),$$

say. As x tends to ∞, y tends to the limit 0, and $\psi(y)$ tends to the limit l.

Let us now dismiss x and consider $\psi(y)$ simply as a function of y. We are for the moment concerned only with those values of y which correspond to large positive values of x, that is to say with small positive values of y. And $\psi(y)$ has the property that by making y sufficiently small we can make $\psi(y)$ differ by as little as we please from l. To put the matter more precisely, the statement expressed by $\lim \phi(x) = l$ means that, when any positive number ϵ, however small, is assigned, we can choose x_0 so that $|\phi(x) - l| < \epsilon$ for all values of x greater than or equal to x_0. But this is the same thing as saying that we can choose $y_0 = 1/x_0$ so that $|\psi(y) - l| < \epsilon$ for all positive values of y less than or equal to y_0.

We are thus led to the following definitions:

A. *If, when any positive number ϵ, however small, is assigned, we can choose $y_0(\epsilon)$ so that*

$$|\phi(y) - l| < \epsilon$$

when $0 < y \leqq y_0(\epsilon)$, then we say that $\phi(y)$ tends to the limit l as y tends to 0 by positive values, and we write

$$\lim_{y \to +0} \phi(y) = l.$$

B. *If, when any number Δ, however large, is assigned, we can choose $y_0(\Delta)$ so that*

$$\phi(y) > \Delta$$

when $0 < y \leqq y_0(\Delta)$, then we say that $\phi(y)$ tends to ∞ as y tends to 0 by positive values, and we write

$$\phi(y) \to \infty.$$

We define in a similar way the meaning of '$\phi(y)$ tends to the limit l as y tends to 0 by negative values', or '$\lim \phi(y) = l$ when $y \to -0$'. We have in fact only to alter $0 < y \leqq y_0(\epsilon)$ to $-y_0(\epsilon) \leqq y < 0$ in definition A. There is of course a corresponding analogue of definition B, and similar definitions in which

$$\phi(y) \to -\infty$$

as $y \to +0$ or $y \to -0$.

If $\lim_{y \to +0} \phi(y) = l$ *and* $\lim_{y \to -0} \phi(y) = l$, we write simply

$$\lim_{y \to 0} \phi(y) = l.$$

This case is so important that it is worth while to give a formal definition.

If, when any positive number ϵ, however small, is assigned, we can choose $y_0(\epsilon)$ so that, for all values of y different from zero but numerically less than or equal to $y_0(\epsilon)$, $\phi(y)$ differs from l by less than ϵ, then we say that $\phi(y)$ tends to the limit l as y tends to 0, and write

$$\lim_{y\to 0} \phi(y) = l.$$

So also, if $\phi(y) \to \infty$ as $y \to +0$ and also as $y \to -0$, we say that $\phi(y) \to \infty$ as $y \to 0$. We define in a similar manner the statement that $\phi(y) \to -\infty$ as $y \to 0$.

Finally, if $\phi(y)$ does not tend to a limit, or to ∞, or to $-\infty$, as $y \to +0$, we say that $\phi(y)$ oscillates as $y \to +0$, finitely or infinitely as the case may be; and we define oscillation as $y \to -0$ in a similar manner.

The preceding definitions have been stated in terms of a variable denoted by y: what letter is used is of course immaterial, and we may suppose x written instead of y throughout them.

94. Limits as x tends to a. Suppose that $\phi(y) \to l$ as $y \to 0$, and write

$$y = x - a, \quad \phi(y) = \phi(x - a) = \psi(x).$$

If $y \to 0$ then $x \to a$ and $\psi(x) \to l$, and we are naturally led to write

$$\lim_{x\to a} \psi(x) = l,$$

or simply $\lim \psi(x) = l$ or $\psi(x) \to l$, and to say that $\psi(x)$ *tends to the limit l as x tends to a*. The meaning of this equation may be formally and directly defined as follows: *if, given ϵ, we can always determine $\delta(\epsilon)$ so that*

$$|\phi(x) - l| < \epsilon$$

when $0 < |x - a| \leqq \delta(\epsilon)$, then

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

By restricting ourselves to values of x greater than a, *i.e.* by replacing $0 < |x - a| \leqq \delta(\epsilon)$ by $a < x \leqq a + \delta(\epsilon)$, we define '$\phi(x)$ tends to l when x approaches a from the right', which we may write as

$$\lim_{x \to a+0} \phi(x) = l.$$

In the same way we can define the meaning of

$$\lim_{x \to a-0} \phi(x) = l.$$

Thus $\lim\limits_{x \to a} \phi(x) = l$ is equivalent to the two assertions

$$\lim_{x \to a+0} \phi(x) = l, \quad \lim_{x \to a-0} \phi(x) = l.$$

We can give similar definitions referring to the cases in which $\phi(x) \to \infty$ or $\phi(x) \to -\infty$ as $x \to a$ through values greater or less than a; but it is probably unnecessary to dwell further on these definitions, since they are exactly similar to those stated above in the special case when $a = 0$, and we can always discuss the behaviour of $\phi(x)$ as $x \to a$ by putting $x - a = y$ and supposing that $y \to 0$.

95. Steadily increasing or decreasing functions. If there is a number δ such that $\phi(x') \leqq \phi(x'')$ whenever $a - \delta < x' < x'' < a + \delta$, then $\phi(x)$ will be said to *increase steadily in the neighbourhood of* $x = a$.

Suppose first that $x < a$, and put $y = 1/(a - x)$. Then $y \to \infty$ as $x \to a - 0$, and $\phi(x) = \psi(y)$ is a steadily increasing function of y, never greater than $\phi(a)$. It follows from § 92 that $\phi(x)$ tends to a limit not greater than $\phi(a)$. We shall write

$$\lim_{x \to a+0} \phi(x) = \phi(a + 0).$$

We can define $\phi(a - 0)$ in a similar manner; and it is clear that

$$\phi(a - 0) \leqq \phi(a) \leqq \phi(a + 0).$$

It is obvious that similar considerations may be applied to *decreasing* functions.

If $\phi(x') < \phi(x'')$, the possibility of equality being excluded, whenever $a - \delta < x' < x'' < a + \delta$, then $\phi(x)$ will be said to be *steadily increasing in the stricter sense.*

96. Limits of indetermination and the principle of convergence. All of the argument of §§ 80–84 may be applied to functions of a continuous variable x which tends to a limit a. In particular, if $\phi(x)$ is *bounded* in an interval including a (*i.e.* if we can find δ, H, and K so that $H < \phi(x) < K$ when $a - \delta \leqq x \leqq a + \delta$).* then we can define λ and Λ, the lower and upper limits of indetermination of $\phi(x)$ as $x \to a$, and prove that the necessary and sufficient condition that $\phi(x) \to l$ as $x \to a$ is that $\lambda = \Lambda = l$. We can also establish the analogue of the principle of convergence, *i.e.* prove that *the necessary and sufficient condition that $\phi(x)$ should tend to a limit as $x \to a$ is that, when ϵ is given, we can choose $\delta(\epsilon)$ so that $|\phi(x_2) - \phi(x_1)| < \epsilon$ when* $0 < |x_2 - a| < |x_1 - a| \leqq \delta(\epsilon)$.

Examples XXXV. 1. If

$$\phi(x) \to l, \quad \psi(x) \to l',$$

as $x \to a$, then $\phi(x) + \psi(x) \to l + l'$, $\phi(x)\psi(x) \to ll'$, and $\phi(x)/\psi(x) \to l/l'$, unless in the last case $l' = 0$.

[We saw in § 91 that the theorems of Ch. IV, §§ 63 *et seq.* hold also for functions of x when $x \to \infty$ or $x \to -\infty$. By putting $x = 1/y$ we may extend them to functions of y, when $y \to 0$, and by putting $y = z - a$ to functions of z, when $z \to a$.

The reader should however try to prove them directly from the formal definition given above. Thus, in order to obtain a strict direct proof of the first result he need only take the proof of Theorem I of § 63 and write throughout x for n, a for ∞ and $0 < |x - a| \leqq \delta$ for $n \geqq n_0$.]

2. If m is a positive integer then $x^m \to 0$ as $x \to 0$.

3. If m is a negative integer then $x^m \to +\infty$ as $x \to +0$, while $x^m \to -\infty$ or $x^m \to +\infty$ as $x \to -0$, according as m is odd or even. If $m = 0$ then $x^m = 1$ and $x^m \to 1$.

4. $\lim\limits_{x \to 0} (a + bx + cx^2 + \dots + kx^m) = a$.

5. $\lim\limits_{x \to 0} \{(a + bx + \dots + kx^m)/(\alpha + \beta x + \dots + \kappa x^\mu)\} = a/\alpha$, unless $\alpha = 0$. If $\alpha = 0$ and $a \neq 0$, $\beta \neq 0$, then the function tends to $+\infty$ or $-\infty$, as $x \to +0$,

*For some further discussion of the notion of *a function bounded in an interval* see § 102.

according as a and β have like or unlike signs; the case is reversed if $x \to -0$. The case in which both a and α vanish is considered in Ex. XXXVI. 5. Discuss the cases which arise when $a \neq 0$ and more than one of the first coefficients in the denominator vanish.

6. $\lim\limits_{x \to a} x^m = a^m$, if m is any positive or negative integer, except when $a = 0$ and m is negative. [If $m > 0$, put $x = y + a$ and apply Ex. 4. When $m < 0$, the result follows from Ex. 1 above. It follows at once that $\lim P(x) = P(a)$, if $P(x)$ is any polynomial.]

7. $\lim\limits_{x \to a} R(x) = R(a)$, if R denotes any rational function and a is not one of the roots of its denominator.

8. Show that $\lim\limits_{x \to a} x^m = a^m$ for all rational values of m, except when $a = 0$ and m is negative. [This follows at once, when a is positive, from the inequalities (9) or (10) of § 74. For $|x^m - a^m| < H|x - a|$, where H is the greater of the absolute values of mx^{m-1} and ma^{m-1} (cf. Ex. XXVIII. 4). If a is negative we write $x = -y$ and $a = -b$. Then

$$\lim x^m = \lim(-1)^m y^m = (-1)^m b^m = a^m.]$$

97. The reader will probably fail to see at first that any proof of such results as those of Exs. 4, 5, 6, 7, 8 above is necessary. He may ask 'why not simply put $x = 0$, or $x = a$? Of course we then get a, a/α, a^m, $P(a)$, $R(a)$'. It is very important that he should see exactly where he is wrong. We shall therefore consider this point carefully before passing on to any further examples.

The statement

$$\lim_{x \to 0} \phi(x) = l$$

is a statement about the values of $\phi(x)$ when x has any value *distinct from but differing by little from zero.** It is *not* a statement about the *value of* $\phi(x)$ *when* $x = 0$. When we make the statement we assert that, when x is *nearly* equal to zero, $\phi(x)$ is nearly equal to l. We assert nothing whatever about what happens when x is *actually* equal to 0. So far as we

*Thus in Def. A of § 93 we make a statement about values of y such that $0 < y \leqq y_0$, the first of these inequalities being inserted expressly in order to exclude the value $y = 0$.

know, $\phi(x)$ may not be defined at all for $x = 0$; or it may have some value other than l. For example, consider the function defined for all values of x by the equation $\phi(x) = 0$. It is obvious that

$$\lim \phi(x) = 0. \tag{1}$$

Now consider the function $\psi(x)$ which differs from $\phi(x)$ only in that $\psi(x) = 1$ when $x = 0$. Then

$$\lim \psi(x) = 0, \tag{2}$$

for, when x is nearly equal to zero, $\psi(x)$ is not only nearly but exactly equal to zero. But $\psi(0) = 1$. The graph of this function consists of the axis of x, with the point $x = 0$ left out, and one isolated point, viz. the point $(0, 1)$. The equation (2) expresses the fact that if we move along the graph towards the axis of y, from either side, then the ordinate of the curve, being always equal to zero, tends to the limit zero. This fact is in no way affected by the position of the isolated point $(0, 1)$.

The reader may object to this example on the score of artificiality: but it is easy to write down simple formulae representing functions which behave precisely like this near $x = 0$. One is

$$\psi(x) = [1 - x^2],$$

where $[1-x^2]$ denotes as usual the greatest integer not greater than $1-x^2$. For if $x = 0$ then $\psi(x) = [1] = 1$; while if $0 < x < 1$, or $-1 < x < 0$, then $0 < 1 - x^2 < 1$ and so $\psi(x) = [1 - x^2] = 0$.

Or again, let us consider the function

$$y = x/x$$

already discussed in Ch. II, § 24, (2). This function is equal to 1 for all values of x save $x = 0$. It is *not* equal to 1 when $x = 0$: it is in fact not defined at all for $x = 0$. For when we say that $\phi(x)$ is defined for $x = 0$ we mean (as we explained in Ch. II, *l.c.*) that we can calculate its value for $x = 0$ by putting $x = 0$ in the actual expression of $\phi(x)$. In this case we cannot. When we put $x = 0$ in $\phi(x)$ we obtain $0/0$, which is a meaningless expression. The reader may object 'divide numerator and denominator by x'. But he must admit that when $x = 0$ this is impossible. Thus $y = x/x$ is a function which differs from $y = 1$ solely in that it is not defined for $x = 0$. None the less

$$\lim(x/x) = 1,$$

for x/x is equal to 1 so long as x differs from zero, however small the difference may be.

Similarly $\phi(x) = \{(x+1)^2 - 1\}/x = x + 2$ so long as x is not equal to zero, but is undefined when $x = 0$. None the less $\lim \phi(x) = 2$.

On the other hand there is of course nothing to prevent the limit of $\phi(x)$ as x tends to zero from being equal to $\phi(0)$, the value of $\phi(x)$ for $x = 0$. Thus if $\phi(x) = x$ then $\phi(0) = 0$ and $\lim \phi(x) = 0$. This is in fact, from a practical point of view, *i.e.* from the point of view of what most frequently occurs in applications, the ordinary case.

Examples XXXVI. 1. $\lim\limits_{x\to a}(x^2 - a^2)/(x - a) = 2a$.

2. $\lim\limits_{x\to a}(x^m - a^m)/(x - a) = ma^{m-1}$, if m is any integer (zero included).

3. Show that the result of Ex. 2 remains true for all rational values of m, provided a is positive. [This follows at once from the inequalities (9) and (10) of § 74.]

4. $\lim\limits_{x\to 1}(x^7 - 2x^5 + 1)/(x^3 - 3x^2 + 2) = 1$. [Observe that $x - 1$ is a factor of both numerator and denominator.]

5. Discuss the behaviour of

$$\phi(x) = (a_0x^m + a_1x^{m+1} + \cdots + a_kx^{m+k})/(b_0x^n + b_1x^{n+1} + \cdots + b_lx^{n+l})$$

as x tends to 0 by positive or negative values.

[If $m > n$, $\lim \phi(x) = 0$. If $m = n$, $\lim \phi(x) = a_0/b_0$. If $m < n$ and $n - m$ is even, $\phi(x) \to +\infty$ or $\phi(x) \to -\infty$ according as $a_0/b_0 > 0$ or $a_0/b_0 < 0$. If $m < n$ and $n - m$ is odd, $\phi(x) \to +\infty$ as $x \to +0$ and $\phi(x) \to -\infty$ as $x \to -0$, or $\phi(x) \to -\infty$ as $x \to +0$ and $\phi(x) \to +\infty$ as $x \to -0$, according as $a_0/b_0 > 0$ or $a_0/b_0 < 0$.]

6. **Orders of smallness**. When x is small x^2 is very much smaller, x^3 much smaller still, and so on: in other words

$$\lim_{x\to 0}(x^2/x) = 0, \quad \lim_{x\to 0}(x^3/x^2) = 0, \; \dots .$$

Another way of stating the matter is to say that, when x tends to 0, x^2, x^3, ... all also tend to 0, but x^2 tends to 0 more rapidly than x, x^3 than x^2, and so on. It is convenient to have some scale by which to measure the rapidity with which a function, whose limit, as x tends to 0, is 0, diminishes with x, and it is natural to take the simple functions x, x^2, x^3, ... as the measures of our scale.

We say, therefore, that $\phi(x)$ *is of the first order of smallness* if $\phi(x)/x$ tends to a limit other than 0 as x tends to 0. Thus $2x + 3x^2 + x^7$ is of the first order of smallness, since $\lim(2x + 3x^2 + x^7)/x = 2$.

Similarly we define the second, third, fourth, ... orders of smallness. It must not be imagined that this scale of orders of smallness is in any way complete. If it were complete, then every function $\phi(x)$ which tends to zero with x would be of either the first or second or some higher order of smallness. This is obviously not the case. For example $\phi(x) = x^{7/5}$ tends to zero more rapidly than x and less rapidly than x^2.

The reader may not unnaturally think that our scale might be made complete by including in it *fractional* orders of smallness. Thus we might say that $x^{7/5}$ was of the $\frac{7}{5}$th order of smallness. We shall however see later on that such a scale of orders would still be altogether incomplete. And as a matter of fact the *integral* orders of smallness defined above are so much more important in applications than any others that it is hardly necessary to attempt to make our definitions more precise.

Orders of greatness. Similar definitions are at once suggested to meet the case in which $\phi(x)$ is large (positively or negatively) when x is small. We shall say that $\phi(x)$ is of the kth order of greatness when x is small if $\phi(x)/x^{-k} = x^k\phi(x)$ tends to a limit different from 0 as x tends to 0.

These definitions have reference to the case in which $x \to 0$. There are of course corresponding definitions relating to the cases in which $x \to \infty$ or $x \to a$. Thus if $x^k\phi(x)$ tends to a limit other than zero, as $x \to \infty$, then we say that $\phi(x)$ is of the kth order of smallness when x is large: while if $(x-a)^k\phi(x)$ tends to a limit other than zero, as $x \to a$, then we say that $\phi(x)$ is of the kth order of greatness when x is nearly equal to a.

7.* $\lim \sqrt{1+x} = \lim \sqrt{1-x} = 1$. [Put $1+x=y$ or $1-x=y$, and use Ex. XXXV. 8.]

8. $\lim\{\sqrt{1+x}-\sqrt{1-x}\}/x = 1$. [Multiply numerator and denominator by $\sqrt{1+x}+\sqrt{1-x}$.]

9. Consider the behaviour of $\{\sqrt{1+x^m}-\sqrt{1-x^m}\}/x^n$ as $x \to 0$, m and n being positive integers.

10. $\lim\{\sqrt{1+x+x^2}-1\}/x = \frac{1}{2}$.

11. $\lim \dfrac{\sqrt{1+x}-\sqrt{1+x^2}}{\sqrt{1-x^2}-\sqrt{1-x}} = 1$.

12. Draw a graph of the function

$$y = \left\{\frac{1}{x-1}+\frac{1}{x-\frac{1}{2}}+\frac{1}{x-\frac{1}{3}}+\frac{1}{x-\frac{1}{4}}\right\}\Bigg/\left\{\frac{1}{x-1}+\frac{1}{x-\frac{1}{2}}+\frac{1}{x-\frac{1}{3}}+\frac{1}{x-\frac{1}{4}}\right\}.$$

Has it a limit as $x \to 0$? [Here $y = 1$ except for $x = 1$, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, when y is not defined, and $y \to 1$ as $x \to 0$.]

13. $\lim \dfrac{\sin x}{x} = 1$.

[It may be deduced from the definitions of the trigonometrical ratios† that if x is positive and less than $\frac{1}{2}\pi$ then

$$\sin x < x < \tan x$$

or

$$\cos x < \frac{\sin x}{x} < 1$$

or

$$0 < 1 - \frac{\sin x}{x} < 1 - \cos x = 2\sin^2 \tfrac{1}{2}x.$$

But $2\sin^2\frac{1}{2}x < 2(\frac{1}{2}x)^2 = \frac{1}{2}x^2$. Hence $\lim\limits_{x\to+0}\left(1-\dfrac{\sin x}{x}\right) = 0$, and $\lim\limits_{x\to+0}\dfrac{\sin x}{x} = 1$. As $\dfrac{\sin x}{x}$ is an even function, the result follows.]

---

*In the examples which follow it is to be assumed that limits as $x \to 0$ are required, unless (as in Exs. 19, 22) the contrary is explicitly stated.

†The proofs of the inequalities which are used here depend on certain properties of the area of a sector of a circle which are usually taken as geometrically intuitive; for example, that the area of the sector is greater than that of the triangle inscribed in the sector. The justification of these assumptions must be postponed to Ch. VII.

14. $\lim \dfrac{1-\cos x}{x^2} = \frac{1}{2}$.

15. $\lim \dfrac{\sin \alpha x}{x} = \alpha$. Is this true if $\alpha = 0$?

16. $\lim \dfrac{\arcsin x}{x} = 1$. [Put $x = \sin y$.]

17. $\lim \dfrac{\tan \alpha x}{x} = \alpha, \quad \lim \dfrac{\arctan \alpha x}{x} = \alpha$.

18. $\lim \dfrac{\operatorname{cosec} x - \cot x}{x} = \frac{1}{2}$.

19. $\lim\limits_{x\to 1} \dfrac{1+\cos \pi x}{\tan^2 \pi x} = \frac{1}{2}$.

20. How do the functions $\sin(1/x)$, $(1/x)\sin(1/x)$, $x\sin(1/x)$ behave as $x \to 0$? [The first oscillates finitely, the second infinitely, the third tends to the limit 0. None is defined when $x = 0$. See Exs. XV. 6, 7, 8.]

21. Does the function

$$y = \left(\sin \frac{1}{x}\right) \Big/ \left(\sin \frac{1}{x}\right)$$

tend to a limit as x tends to 0? [*No.* The function is equal to 1 except when $\sin(1/x) = 0$; *i.e.* when $x = 1/\pi, 1/2\pi, \dots, -1/\pi, -1/2\pi, \dots$. For these values the formula for y assumes the meaningless form $0/0$, and y is therefore not defined for an infinity of values of x near $x = 0$.]

22. Prove that if m is any integer then $[x] \to m$ and $x-[x] \to 0$ as $x \to m+0$, and $[x] \to m-1$, $x-[x] \to 1$ as $x \to m-0$.

98. Continuous functions of a real variable. The reader has no doubt some idea as to what is meant by a *continuous curve*. Thus he would call the curve C in Fig. 29 continuous, the curve C' generally continuous but discontinuous for $x = \xi'$ and $x = \xi''$.

Either of these curves may be regarded as the graph of a function $\phi(x)$. It is natural to call a function *continuous* if its graph is a continuous curve, and otherwise discontinuous. Let us take this as a provisional definition and try to distinguish more precisely some of the properties which are involved in it.

In the first place it is evident that the property of the function $y = \phi(x)$ of which C is the graph may be analysed into some property possessed by

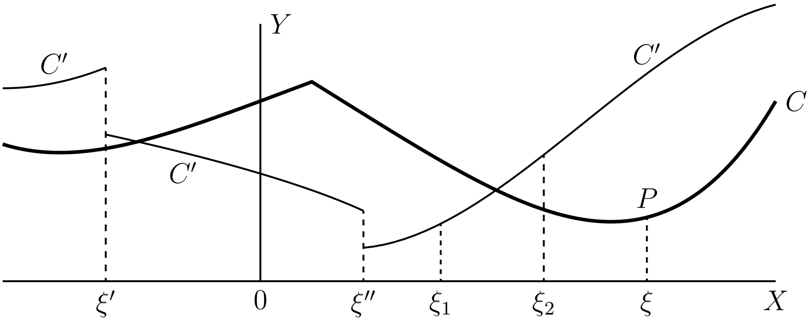

Fig. 29.

the curve at each of its points. To be able to define continuity *for all values of x* we must first define continuity *for any particular value of x*. Let us therefore fix on some particular value of x, say the value $x = \xi$ corresponding to the point P of the graph. What are the characteristic properties of $\phi(x)$ associated with this value of x?

In the first place $\phi(x)$ *is defined for $x = \xi$*. This is obviously essential. If $\phi(\xi)$ were not defined there would be a point missing from the curve.

Secondly $\phi(x)$ *is defined for all values of x near $x = \xi$*; *i.e.* we can find an interval, including $x = \xi$ in its interior, for all points of which $\phi(x)$ is defined.

Thirdly *if x approaches the value ξ from either side then $\phi(x)$ approaches the limit $\phi(\xi)$.*

The properties thus defined are far from exhausting those which are possessed by the curve as pictured by the eye of common sense. This picture of a curve is a generalisation from particular curves such as straight lines and circles. But they are the simplest and most fundamental properties: and the graph of any function which has these properties would, so far as drawing it is practically possible, satisfy our geometrical feeling of what a continuous curve should be. We therefore select these properties as embodying the mathematical notion of continuity. We are thus led to the following

DEFINITION. *The function $\phi(x)$ is said to be continuous for $x = \xi$ if it tends to a limit as x tends to ξ from either side, and each of these limits* *is equal to* $\phi(\xi)$.

We can now define *continuity throughout an interval.* The function $\phi(x)$ is said to be continuous throughout a certain interval of values of x if it is continuous for all values of x in that interval. It is said to be *continuous everywhere* if it is continuous for every value of x. Thus $[x]$ is continuous in the interval $[\delta, 1-\delta]$, where δ is any positive number less than $\frac{1}{2}$; and 1 and x are continuous everywhere.

If we recur to the definitions of a limit we see that our definition is equivalent to '$\phi(x)$ *is continuous for* $x = \xi$ *if, given* ϵ, *we can choose* $\delta(\epsilon)$ *so that* $|\phi(x) - \phi(\xi)| < \epsilon$ *if* $0 \leqq |x - \xi| \leqq \delta(\epsilon)$'.

We have often to consider functions defined only in an interval $[a, b]$. In this case it is convenient to make a slight and obvious change in our definition of continuity in so far as it concerns the particular points a and b. We shall then say that $\phi(x)$ is continuous for $x = a$ if $\phi(a + 0)$ exists and is equal to $\phi(a)$, and for $x = b$ if $\phi(b - 0)$ exists and is equal to $\phi(b)$.

99. The definition of continuity given in the last section may be illustrated geometrically as follows. Draw the two horizontal lines $y = \phi(\xi) - \epsilon$ and $y = \phi(\xi) + \epsilon$. Then $|\phi(x) - \phi(\xi)| < \epsilon$ expresses the fact that the point on the curve corresponding to x lies between these two lines. Similarly

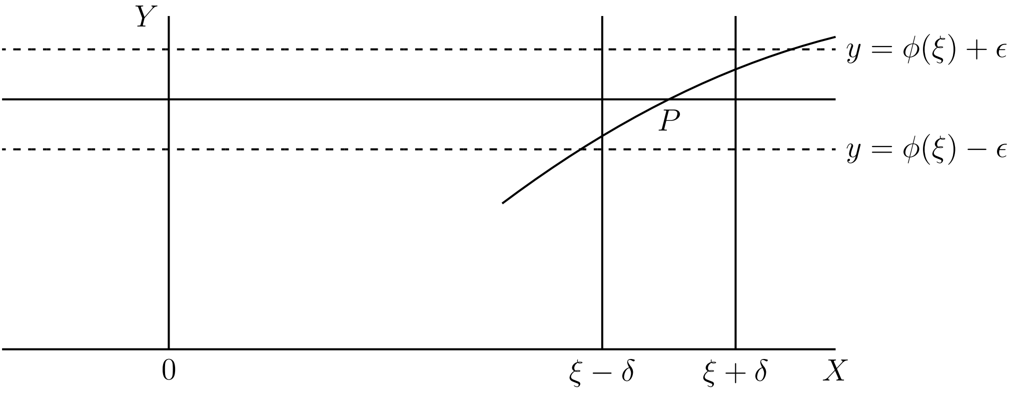

Fig. 30.

$|x - \xi| \leqq \delta$ expresses the fact that x lies in the interval $[\xi - \delta, \xi + \delta]$. Thus our definition asserts that if we draw two such horizontal lines, no matter how close together, we can always cut off a vertical strip of the plane by two vertical lines in such a way that all that part of the curve which is contained in the strip lies between the two horizontal lines. This is evidently true of the curve C (Fig. 29), whatever value ξ may have.

We shall now discuss the continuity of some special types of functions. Some of the results which follow were (as we pointed out at the time) tacitly assumed in Ch. II.

Examples XXXVII. 1. The sum or product of two functions continuous at a point is continuous at that point. The quotient is also continuous unless the denominator vanishes at the point. [This follows at once from Ex. XXXV. 1.]

2. Any polynomial is continuous for all values of x. Any rational fraction is continuous except for values of x for which the denominator vanishes. [This follows from Exs. XXXV. 6, 7.]

3. $\sqrt{x}$ is continuous for all positive values of x (Ex. XXXV. 8). It is not defined when $x < 0$, but is continuous for $x = 0$ in virtue of the remark made at the end of § 98. The same is true of $x^{m/n}$, where m and n are any positive integers of which n is even.

4. The function $x^{m/n}$, where n is odd, is continuous for all values of x.

5. $1/x$ is not continuous for $x = 0$. It has no value for $x = 0$, nor does it tend to a limit as $x \to 0$. In fact $1/x \to +\infty$ or $1/x \to -\infty$ according as $x \to 0$ by positive or negative values.

6. Discuss the continuity of $x^{-m/n}$, where m and n are positive integers, for $x = 0$.

7. The standard rational function $R(x) = P(x)/Q(x)$ is discontinuous for $x = a$, where a is any root of $Q(x) = 0$. Thus $(x^2 + 1)/(x^2 - 3x + 2)$ is discontinuous for $x = 1$. It will be noticed that in the case of rational functions a discontinuity is always associated with (a) a failure of the definition for a particular value of x and (b) a tending of the function to $+\infty$ or $-\infty$ as x approaches this value from either side. Such a particular kind of point of discontinuity is usually described as an **infinity** of the function. An 'infinity' is the kind of discontinuity of most common occurrence in ordinary work.

8. Discuss the continuity of

$$\sqrt{(x-a)(b-x)},\quad \sqrt[3]{(x-a)(b-x)},\quad \sqrt{(x-a)/(b-x)},\quad \sqrt[3]{(x-a)/(b-x)}.$$

9. $\sin x$ and $\cos x$ are continuous for all values of x.

[We have
$$\sin(x+h) - \sin x = 2\sin\tfrac{1}{2}h\cos(x+\tfrac{1}{2}h),$$
which is numerically less than the numerical value of h.]

10. For what values of x are $\tan x$, $\cot x$, $\sec x$, and $\operatorname{cosec} x$ continuous or discontinuous?

11. If $f(y)$ is continuous for $y = \eta$, and $\phi(x)$ is a continuous function of x which is equal to η when $x = \xi$, then $f\{\phi(x)\}$ is continuous for $x = \xi$.

12. If $\phi(x)$ is continuous for any particular value of x, then any polynomial in $\phi(x)$, such as $a\{\phi(x)\}^m + \dots$, is so too.

13. Discuss the continuity of
$$1/(a\cos^2 x + b\sin^2 x), \quad \sqrt{2+\cos x}, \quad \sqrt{1+\sin x}, \quad 1/\sqrt{1+\sin x}.$$

14. $\sin(1/x)$, $x\sin(1/x)$, and $x^2\sin(1/x)$ are continuous except for $x = 0$.

15. The function which is equal to $x\sin(1/x)$ except when $x = 0$, and to zero when $x = 0$, is continuous for all values of x.

16. $[x]$ and $x - [x]$ are discontinuous for all integral values of x.

17. For what (if any) values of x are the following functions discontinuous: $[x^2]$, $[\sqrt{x}\,]$, $\sqrt{x-[x]}$, $[x] + \sqrt{x-[x]}$, $[2x]$, $[x] + [-x]$?

18. **Classification of discontinuities.** Some of the preceding examples suggest a classification of different types of discontinuity.

(1) Suppose that $\phi(x)$ tends to a limit as $x \to a$ either by values less than or by values greater than a. Denote these limits, as in § 95, by $\phi(a-0)$ and $\phi(a+0)$ respectively. Then, for continuity, it is necessary and sufficient that $\phi(x)$ should be defined for $x = a$, and that $\phi(a-0) = \phi(a) = \phi(a+0)$. Discontinuity may arise in a variety of ways.

(α) $\phi(a-0)$ may be equal to $\phi(a+0)$, but $\phi(a)$ may not be defined, or may differ from $\phi(a-0)$ and $\phi(a+0)$. Thus if $\phi(x) = x\sin(1/x)$ and $a = 0$, $\phi(0-0) = \phi(0+0) = 0$, but $\phi(x)$ is not defined for $x = 0$. Or if $\phi(x) = [1-x^2]$ and $a = 0$, $\phi(0-0) = \phi(0+0) = 0$, but $\phi(0) = 1$.

(β) $\phi(a-0)$ and $\phi(a+0)$ may be unequal. In this case $\phi(a)$ may be equal to one or to neither, or be undefined. The first case is illustrated by $\phi(x) = [x]$, for which $\phi(0-0) = -1$, $\phi(0+0) = \phi(0) = 0$; the second by $\phi(x) = [x] - [-x]$, for which $\phi(0-0) = -1$, $\phi(0+0) = 1$, $\phi(0) = 0$; and the third by $\phi(x) = [x] + x\sin(1/x)$, for which $\phi(0-0) = -1$, $\phi(0+0) = 0$, and $\phi(0)$ is undefined.

In any of these cases we say that $\phi(x)$ has a **simple discontinuity** at $x = a$. And to these cases we may add those in which $\phi(x)$ is defined only on one side of $x = a$, and $\phi(a-0)$ or $\phi(a+0)$, as the case may be, exists, but $\phi(x)$ is either not defined when $x = a$ or has when $x = a$ a value different from $\phi(a-0)$ or $\phi(a+0)$.

It is plain from § 95 that *a function which increases or decreases steadily in the neighbourhood of $x = a$ can have at most a simple discontinuity for $x = a$.*

(2) It may be the case that only one (or neither) of $\phi(a-0)$ and $\phi(a+0)$ exists, but that, supposing for example $\phi(a+0)$ not to exist, $\phi(x) \to +\infty$ or $\phi(x) \to -\infty$ as $x \to a+0$, so that $\phi(x)$ tends to a limit or to $+\infty$ or to $-\infty$ as x approaches a from either side. Such is the case, for instance, if $\phi(x) = 1/x$ or $\phi(x) = 1/x^2$, and $a = 0$. In such cases we say (cf. Ex. 7) that $x = a$ is an **infinity** of $\phi(x)$. And again we may add to these cases those in which $\phi(x) \to +\infty$ or $\phi(x) \to -\infty$ as $x \to a$ from one side, but $\phi(x)$ is not defined at all on the other side of $x = a$.

(3) Any point of discontinuity which is not a point of simple discontinuity nor an infinity is called a point of **oscillatory discontinuity**. Such is the point $x = 0$ for the functions $\sin(1/x)$, $(1/x)\sin(1/x)$.

19. What is the nature of the discontinuities at $x = 0$ of the functions $(\sin x)/x$, $[x] + [-x]$, $\operatorname{cosec} x$, $\sqrt{1/x}$, $\sqrt[3]{1/x}$, $\operatorname{cosec}(1/x)$, $\sin(1/x)/\sin(1/x)$?

20. The function which is equal to 1 when x is rational and to 0 when x is irrational (Ch. II, Ex. XVI. 10) is discontinuous for all values of x. So too is any function which is defined only for rational or for irrational values of x.

21. The function which is equal to x when x is irrational and to $\sqrt{(1+p^2)/(1+q^2)}$ when x is a rational fraction p/q (Ch. II, Ex. XVI. 11) is discontinuous for all negative and for positive rational values of x, but continuous for positive irrational values.

22. For what points are the functions considered in Ch. IV, Exs. XXXI discontinuous, and what is the nature of their discontinuities? [Consider, *e.g.*, the function $y = \lim x^n$ (Ex. 5). Here y is only defined when $-1 < x \leqq 1$: it is equal to 0 when $-1 < x < 1$ and to 1 when $x = 1$. The points $x = 1$ and $x = -1$ are points of simple discontinuity.]

100. The fundamental property of a continuous function. It may perhaps be thought that the analysis of the idea of a continuous curve given in § 98 is not the simplest or most natural possible. Another method of analysing our idea of continuity is the following. Let A and B be two points on the graph of $\phi(x)$ whose coordinates are x_0, $\phi(x_0)$ and x_1, $\phi(x_1)$ respectively. Draw any straight line λ which passes between A and B. Then common sense certainly declares that if the graph of $\phi(x)$ is continuous it must cut λ.

If we consider this property as an intrinsic geometrical property of continuous curves it is clear that there is no real loss of generality in supposing λ to be parallel to the axis of x. In this case the ordinates of A and B cannot be equal: let us suppose, for definiteness, that $\phi(x_1) > \phi(x_0)$. And let λ be the line $y = \eta$, where $\phi(x_0) < \eta < \phi(x_1)$. Then to say that the graph of $\phi(x)$ must cut λ is the same thing as to say that there is a value of x between x_0 and x_1 for which $\phi(x) = \eta$.

We conclude then that a continuous function $\phi(x)$ must possess the following property: *if*

$$\phi(x_0) = y_0, \quad \phi(x_1) = y_1,$$

and $y_0 < \eta < y_1$, *then there is a value of* x *between* x_0 *and* x_1 *for which* $\phi(x) = \eta$. In other words *as* x *varies from* x_0 *to* x_1, y *must assume at least once every value between* y_0 *and* y_1.

We shall now prove that if $\phi(x)$ is a continuous function of x in the sense defined in § 98 then it does in fact possess this property. There is a certain range of values of x, to the right of x_0, for which $\phi(x) < \eta$. For $\phi(x_0) < \eta$, and so $\phi(x)$ is certainly less than η if $\phi(x) - \phi(x_0)$ is numerically less than $\eta - \phi(x_0)$. But since $\phi(x)$ is continuous for $x = x_0$, this condition is certainly satisfied if x is near enough to x_0. Similarly there is a certain range of values, to the left of x_1, for which $\phi(x) > \eta$.

Let us divide the values of x between x_0 and x_1 into two classes L, R as follows:

(1) in the class L we put all values ξ of x such that $\phi(x) < \eta$ when $x = \xi$ and for all values of x between x_0 and ξ;

(2) in the class R we put all the other values of x, *i.e.* all numbers ξ

such that either $\phi(\xi) \geqq \eta$ or there is a value of x between x_0 and ξ for which $\phi(x) \geqq \eta$.

Then it is evident that these two classes satisfy all the conditions imposed upon the classes L, R of § 17, and so constitute a section of the real numbers. Let ξ_0 be the number corresponding to the section.

First suppose $\phi(\xi_0) > \eta$, so that ξ_0 belongs to the upper class: and let $\phi(\xi_0) = \eta + k$, say. Then $\phi(\xi') < \eta$ and so

$$\phi(\xi_0) - \phi(\xi') > k,$$

for all values of ξ' less than ξ_0, which contradicts the condition of continuity for $x = \xi_0$.

Next suppose $\phi(\xi_0) = \eta - k < \eta$. Then, if ξ' is any number greater than ξ_0, either $\phi(\xi') \geqq \eta$ or we can find a number ξ'' between ξ_0 and ξ' such that $\phi(\xi'') \geqq \eta$. In either case we can find a number as near to ξ_0 as we please and such that the corresponding values of $\phi(x)$ differ by more than k. And this again contradicts the hypothesis that $\phi(x)$ is continuous for $x = \xi_0$.

Hence $\phi(\xi_0) = \eta$, and the theorem is established. It should be observed that we have proved more than is asserted explicitly in the theorem; we have proved in fact that ξ_0 is the *least* value of x for which $\phi(x) = \eta$. It is not obvious, or indeed generally true, that there is a least among the values of x for which a function assumes a given value, though this is true for continuous functions.

It is easy to see that the converse of the theorem just proved is not true. Thus such a function as the function $\phi(x)$ whose graph is represented by Fig. 31 obviously assumes at least once every value between $\phi(x_0)$ and $\phi(x_1)$: yet $\phi(x)$ is discontinuous. Indeed it is not even true that $\phi(x)$ must be continuous when it assumes each value *once and once only*. Thus let $\phi(x)$ be defined as follows from $x = 0$ to $x = 1$. If $x = 0$ let $\phi(x) = 0$; if $0 < x < 1$ let $\phi(x) = 1 - x$; and if $x = 1$ let $\phi(x) = 1$. The graph of the function is shown in Fig. 32; it includes the points O, C but *not* the points A, B. It is clear that, as x varies from 0 to 1, $\phi(x)$ assumes once and once only every value between $\phi(0) = 0$ and $\phi(1) = 1$; but $\phi(x)$ is discontinuous for $x = 0$ and $x = 1$.

As a matter of fact, however, the curves which usually occur in elementary mathematics are composed of *a finite number of pieces along which y always*

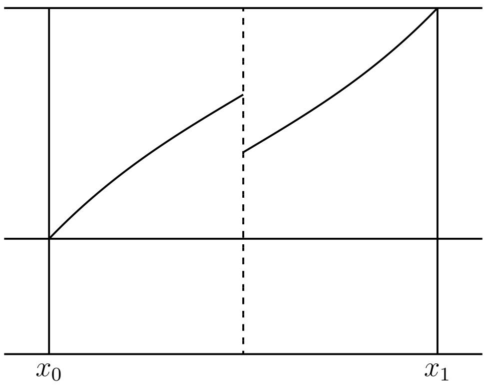

Fig. 31.

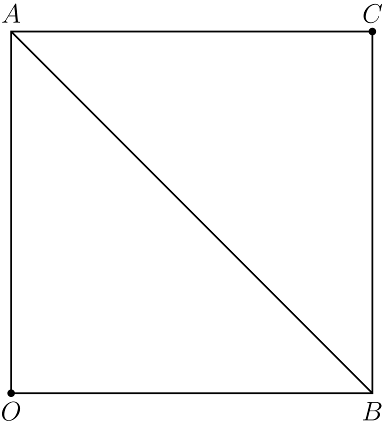

Fig. 32.

varies in the same direction. It is easy to show that if $y = \phi(x)$ always varies in the same direction, *i.e.* steadily increases or decreases, as x varies from x_0 to x_1, then the two notions of continuity are really equivalent, *i.e.* that if $\phi(x)$ takes every value between $\phi(x_0)$ and $\phi(x_1)$ then it must be a continuous function in the sense of § 98. For let ξ be any value of x between x_0 and x_1. As $x \to \xi$ through values less than ξ, $\phi(x)$ tends to the limit $\phi(\xi - 0)$ (§ 95). Similarly as $x \to \xi$ through values greater than ξ, $\phi(x)$ tends to the limit $\phi(\xi + 0)$. The function will be continuous for $x = \xi$ if and only if

$$\phi(\xi - 0) = \phi(\xi) = \phi(\xi + 0).$$

But if either of these equations is untrue, say the first, then it is evident that $\phi(x)$ never assumes any value which lies between $\phi(\xi - 0)$ and $\phi(\xi)$, which is contrary to our assumption. Thus $\phi(x)$ must be continuous. The net result of this and the last section is consequently to show that our common-sense notion of what we mean by continuity is substantially accurate, and capable of precise statement in mathematical terms.

101. In this and the following paragraphs we shall state and prove some general theorems concerning continuous functions.

Theorem 1. *Suppose that $\phi(x)$ is continuous for $x = \xi$, and that $\phi(\xi)$ is positive. Then we can determine a positive number δ such that $\phi(\xi)$ is positive throughout the interval $[\xi - \delta, \xi + \delta]$.*

For, taking $\epsilon = \frac{1}{2}\phi(\xi)$ in the fundamental inequality of p. 212, we can choose δ so that

$$|\phi(x) - \phi(\xi)| < \tfrac{1}{2}\phi(\xi)$$

throughout $[\xi - \delta, \xi + \delta]$, and then

$$\phi(x) \geqq \phi(\xi) - |\phi(x) - \phi(\xi)| > \tfrac{1}{2}\phi(\xi) > 0,$$

so that $\phi(x)$ is positive. There is plainly a corresponding theorem referring to negative values of $\phi(x)$.

Theorem 2. *If $\phi(x)$ is continuous for $x = \xi$, and $\phi(x)$ vanishes for values of x as near to ξ as we please, or assumes, for values of x as near to ξ as we please, both positive and negative values, then $\phi(\xi) = 0$.*

This is an obvious corollary of Theorem 1. If $\phi(\xi)$ is not zero, it must be positive or negative; and if it were, for example, positive, it would be positive for all values of x sufficiently near to ξ, which contradicts the hypotheses of the theorem.

102. The range of values of a continuous function. Let us consider a function $\phi(x)$ about which we shall only assume at present that it is defined for every value of x in an interval $[a, b]$.

The values assumed by $\phi(x)$ for values of x in $[a, b]$ form an aggregate S to which we can apply the arguments of § 80, as we applied them in § 81 to the aggregate of values of a function of n. If there is a number K such that $\phi(x) \leqq K$, for all values of x in question, we say that $\phi(x)$ is *bounded above*. In this case $\phi(x)$ possesses an *upper bound* M: no value of $\phi(x)$ exceeds M, but any number less than M is exceeded by at least one value of $\phi(x)$. Similarly we define '*bounded below*', '*lower bound*', '*bounded*', as applied to functions of a continuous variable x.

Theorem 1. *If $\phi(x)$ is continuous throughout $[a, b]$, then it is bounded in $[a, b]$.*

We can certainly determine an interval $[a, \xi]$, extending to the right from a, in which $\phi(x)$ is bounded. For since $\phi(x)$ is continuous for $x = a$, we can, given any positive number ϵ however small, determine an interval $[a, \xi]$ throughout which $\phi(x)$ lies between $\phi(a) - \epsilon$ and $\phi(a) + \epsilon$; and obviously $\phi(x)$ is bounded in this interval.

Now divide the points ξ of the interval $[a, b]$ into two classes L, R, putting ξ in L if $\phi(\xi)$ is bounded in $[a, \xi]$, and in R if this is not the case. It follows from what precedes that L certainly exists: what we propose to prove is that R does not. Suppose that R does exist, and let β be the number corresponding to the section whose lower and upper classes are L and R. Since $\phi(x)$ is continuous for $x = \beta$, we can, however small ϵ may be, determine an interval $[\beta - \eta, \beta + \eta]$* throughout which

$$\phi(\beta) - \epsilon < \phi(x) < \phi(\beta) + \epsilon.$$

Thus $\phi(x)$ is bounded in $[\beta - \eta, \beta + \eta]$. Now $\beta - \eta$ belongs to L. Therefore $\phi(x)$ is bounded in $[a, \beta - \eta]$: and therefore it is bounded in the whole interval $[a, \beta + \eta]$. But $\beta + \eta$ belongs to R and so $\phi(x)$ is *not* bounded in $[a, \beta + \eta]$. This contradiction shows that R does not exist. And so $\phi(x)$ is bounded in the whole interval $[a, b]$.

THEOREM 2. *If $\phi(x)$ is continuous throughout $[a, b]$, and M and m are its upper and lower bounds, then $\phi(x)$ assumes the values M and m at least once each in the interval.*

For, given any positive number ϵ, we can find a value of x for which $M - \phi(x) < \epsilon$ or $1/\{M - \phi(x)\} > 1/\epsilon$. Hence $1/\{M - \phi(x)\}$ is not bounded, and therefore, by Theorem 1, is not continuous. But $M - \phi(x)$ is a continuous function, and so $1/\{M - \phi(x)\}$ is continuous at any point at which its denominator does not vanish (Ex. XXXVII. 1). There must therefore be one point at which the denominator vanishes: at this point $\phi(x) = M$. Similarly it may be shown that there is a point at which $\phi(x) = m$.

*If $\beta = b$ we must replace this interval by $[\beta - \eta, \beta]$, and $\beta + \eta$ by β, throughout the argument which follows.

The proof just given is somewhat subtle and indirect, and it may be well, in view of the great importance of the theorem, to indicate alternative lines of proof. It will however be convenient to postpone these for a moment.*

Examples XXXVIII. 1. If $\phi(x) = 1/x$ except when $x = 0$, and $\phi(x) = 0$ when $x = 0$, then $\phi(x)$ has neither an upper nor a lower bound in any interval which includes $x = 0$ in its interior, as *e.g.* the interval $[-1, +1]$.

2. If $\phi(x) = 1/x^2$ except when $x = 0$, and $\phi(x) = 0$ when $x = 0$, then $\phi(x)$ has the lower bound 0, but no upper bound, in the interval $[-1, +1]$.

3. Let $\phi(x) = \sin(1/x)$ except when $x = 0$, and $\phi(x) = 0$ when $x = 0$. Then $\phi(x)$ is discontinuous for $x = 0$. In any interval $[-\epsilon, +\epsilon]$ the lower bound is -1 and the upper bound $+1$, and each of these values is assumed by $\phi(x)$ an infinity of times.

4. Let $\phi(x) = x - [x]$. This function is discontinuous for all integral values of x. In the interval $[0, 1]$ its lower bound is 0 and its upper bound 1. It is equal to 0 when $x = 0$ or $x = 1$, but it is never equal to 1. Thus $\phi(x)$ never assumes a value equal to its upper bound.

5. Let $\phi(x) = 0$ when x is irrational, and $\phi(x) = q$ when x is a rational fraction p/q. Then $\phi(x)$ has the lower bound 0, but no upper bound, in any interval $[a, b]$. But if $\phi(x) = (-1)^p q$ when $x = p/q$, then $\phi(x)$ has neither an upper nor a lower bound in any interval.

103. The oscillation of a function in an interval. Let $\phi(x)$ be any function bounded throughout $[a, b]$, and M and m its upper and lower bounds. We shall now use the notation $M(a, b)$, $m(a, b)$ for M, m, in order to exhibit explicitly the dependence of M and m on a and b, and we shall write

$$O(a, b) = M(a, b) - m(a, b).$$

This number $O(a, b)$, the difference between the upper and lower bounds of $\phi(x)$ in $[a, b]$, we shall call the **oscillation** *of* $\phi(x)$ *in* $[a, b]$. The simplest of the properties of the functions $M(a, b)$, $m(a, b)$, $O(a, b)$ are as follows.

*See § 104.

(1) *If* $a \leqq c \leqq b$ *then* $M(a, b)$ *is equal to the greater of* $M(a, c)$ *and* $M(c, b)$, *and* $m(a, b)$ *to the lesser of* $m(a, c)$ *and* $m(c, b)$.

(2) $M(a, b)$ *is an increasing,* $m(a, b)$ *a decreasing, and* $O(a, b)$ *an increasing function of* b.

(3) $O(a, b) \leqq O(a, c) + O(c, b)$.

The first two theorems are almost immediate consequences of our definitions. Let μ be the greater of $M(a, c)$ and $M(c, b)$, and let ϵ be any positive number. Then $\phi(x) \leqq \mu$ throughout $[a, c]$ and $[c, b]$, and therefore throughout $[a, b]$; and $\phi(x) > \mu - \epsilon$ somewhere in $[a, c]$ or in $[c, b]$, and therefore somewhere in $[a, b]$. Hence $M(a, b) = \mu$. The proposition concerning m may be proved similarly. Thus (1) is proved, and (2) is an obvious corollary.

Suppose now that M_1 is the greater and M_2 the less of $M(a, c)$ and $M(c, b)$, and that m_1 is the less and m_2 the greater of $m(a, c)$ and $m(c, b)$. Then, since c belongs to both intervals, $\phi(c)$ is not greater than M_2 nor less than m_2. Hence $M_2 \geqq m_2$, whether these numbers correspond to the same one of the intervals $[a, c]$ and $[c, b]$ or not, and

$$O(a, b) = M_1 - m_1 \leqq M_1 + M_2 - m_1 - m_2.$$

But

$$O(a, c) + O(c, b) = M_1 + M_2 - m_1 - m_2;$$

and (3) follows.

104. Alternative proofs of Theorem 2 of § 102. The most straightforward proof of Theorem 2 of § 102 is as follows. Let ξ be any number of the interval $[a, b]$. The function $M(a, \xi)$ increases steadily with ξ and never exceeds M. We can therefore construct a section of the numbers ξ by putting ξ in L or in R according as $M(a, \xi) < M$ or $M(a, \xi) = M$. Let β be the number corresponding to the section. If $a < \beta < b$, we have

$$M(a, \beta - \eta) < M, \quad M(a, \beta + \eta) = M$$

for all positive values of η, and so

$$M(\beta - \eta, \beta + \eta) = M,$$

by (1) of § 103. Hence $\phi(x)$ assumes, for values of x as near as we please to β, values as near as we please to M, and so, since $\phi(x)$ is continuous, $\phi(\beta)$ must be equal to M.

If $\beta = a$ then $M(a, a+\eta) = M$. And if $\beta = b$ then $M(a, b-\eta) < M$, and so $M(b-\eta, b) = M$. In either case the argument may be completed as before.

The theorem may also be proved by the method of repeated bisection used in § 71. If M is the upper bound of $\phi(x)$ in an interval PQ, and PQ is divided into two equal parts, then it is possible to find a half P_1Q_1 in which the upper bound of $\phi(x)$ is also M. Proceeding as in § 71, we construct a sequence of intervals PQ, P_1Q_1, P_2Q_2, ... in each of which the upper bound of $\phi(x)$ is M. These intervals, as in § 71, converge to a point T, and it is easily proved that the value of $\phi(x)$ at this point is M.

105. Sets of intervals on a line. The Heine-Borel Theorem. We shall now proceed to prove some theorems concerning the oscillation of a function which are of a somewhat abstract character but of very great importance, particularly, as we shall see later, in the theory of integration. These theorems depend upon a general theorem concerning intervals on a line.

Suppose that we are given a *set of intervals* in a straight line, that is to say an aggregate each of whose members is an interval $[\alpha, \beta]$. We make no restriction as to the nature of these intervals; they may be finite or infinite in number; they may or may not overlap;* and any number of them may be included in others.

It is worth while in passing to give a few examples of sets of intervals to which we shall have occasion to return later.

(i) If the interval $[0, 1]$ is divided into n equal parts then the n intervals thus formed define a finite set of non-overlapping intervals which just cover up the line.

(ii) We take every point ξ of the interval $[0, 1]$, and associate with ξ the interval $[\xi-\delta, \xi+\delta]$, where δ is a positive number less than 1, except that with 0 we associate $[0, \delta]$ and with 1 we associate $[1-\delta, 1]$, and in general we reject any

*The word *overlap* is used in its obvious sense: two intervals overlap if they have points in common which are not end points of either. Thus $[0, \frac{2}{3}]$ and $[\frac{1}{3}, 1]$ overlap. A pair of intervals such as $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ may be said to *abut*.

part of any interval which projects outside the interval $[0, 1]$. We thus define an infinite set of intervals, and it is obvious that many of them overlap with one another.

(iii) We take the rational points p/q of the interval $[0, 1]$, and associate with p/q the interval

$$\left[\frac{p}{q} - \frac{\delta}{q^3}, \frac{p}{q} + \frac{\delta}{q^3}\right],$$

where δ is positive and less than 1. We regard 0 as $0/1$ and 1 as $1/1$: in these two cases we reject the part of the interval which lies outside $[0, 1]$. We obtain thus an infinite set of intervals, which plainly overlap with one another, since there are an infinity of rational points, other than p/q, in the interval associated with p/q.

The Heine-Borel Theorem. *Suppose that we are given an interval $[a, b]$, and a set of intervals I each of whose members is included in $[a, b]$. Suppose further that I possesses the following properties:*

(i) *every point of $[a, b]$, other than a and b, lies inside* at least one interval of I;*

(ii) *a is the left-hand end point, and b the right-hand end point, of at least one interval of I.*

Then it is possible to choose **a finite number** *of intervals from the set I which form a set of intervals possessing the properties* (i) *and* (ii).

We know that a is the left-hand end point of at least one interval of I, say $[a, a_1]$. We know also that a_1 lies inside at least one interval of I, say $[a_1', a_2]$. Similarly a_2 lies inside an interval $[a_2', a_3]$ of I. It is plain that this argument may be repeated indefinitely, unless after a finite number of steps a_n coincides with b.

If a_n does coincide with b after a finite number of steps then there is nothing further to prove, for we have obtained a finite set of intervals, selected from the intervals of I, and possessing the properties required. If a_n never coincides with b, then the points a_1, a_2, a_3, ... must (since each lies to the right of its predecessor) tend to a limiting position, but this limiting position may, so far as we can tell, lie anywhere in $[a, b]$.

*That is to say 'in and not at an end of'.

Let us suppose now that the process just indicated, starting from a, is performed in all possible ways, so that we obtain all possible sequences of the type a_1, a_2, a_3, $\ldots$. Then we can prove that *there must be at least one such sequence which arrives at b after a finite number of steps.*

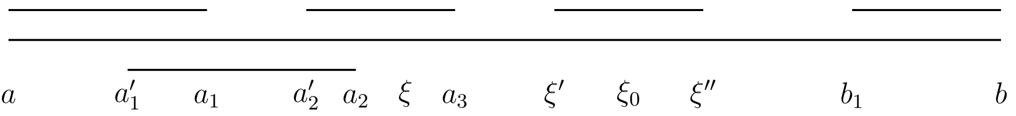

Fig. 33.

There are two possibilities with regard to any point ξ between a and b. Either (i) ξ lies to the left of *some* point a_n of *some* sequence or (ii) it does not. We divide the points ξ into two classes L and R according as to whether (i) or (ii) is true. The class L certainly exists, since all points of the interval $[a, a_1]$ belong to L. We shall now prove that R does not exist, so that every point ξ belongs to L.

If R exists then L lies entirely to the left of R, and the classes L, R form a section of the real numbers between a and b, to which corresponds a number ξ_0. The point ξ_0 lies inside an interval of I, say $[\xi', \xi'']$, and ξ' belongs to L, and so lies to the left of some term a_n of some sequence. But then we can take $[\xi', \xi'']$ as the interval $[a'_n, a_{n+1}]$ associated with a_n in our construction of the sequence a_1, a_2, a_3, $\ldots$; and all points to the left of ξ'' lie to the left of a_{n+1}. There are therefore points of L to the right of ξ_0, and this contradicts the definition of R. It is therefore impossible that R should exist.

Thus every point ξ belongs to L. Now b is the right-hand end point of an interval of I, say $[b_1, b]$, and b_1 belongs to L. Hence there is a member a_n of a sequence a_1, a_2, a_3, $\ldots$ such that $a_n > b_1$. But then we may take the interval $[a'_n, a_{n+1}]$ corresponding to a_n to be $[b_1, b]$, and so we obtain a sequence in which the term after the nth coincides with b, and therefore a finite set of intervals having the properties required. Thus the theorem is proved.

It is instructive to consider the examples of p. 223 in the light of this theorem.

(i) Here the conditions of the theorem are not satisfied: the points $1/n$, $2/n$, $3/n$, ... do not lie inside any interval of I.

(ii) Here the conditions of the theorem are satisfied. The set of intervals

$$[0, 2\delta], \quad [\delta, 3\delta], \quad [2\delta, 4\delta], \ \dots, \quad [1-2\delta, 1],$$

associated with the points δ, 2δ, 3δ, ..., $1-\delta$, possesses the properties required.

(iii) In this case we can prove, by using the theorem, that there are, if δ is small enough, points of $[0, 1]$ which do not lie in any interval of I.

If every point of $[0, 1]$ lay inside an interval of I (with the obvious reservation as to the end points), then we could find a finite number of intervals of I possessing the same property and having therefore a total length greater than 1. Now there are two intervals, of total length 2δ, for which $q = 1$, and $q - 1$ intervals, of total length $2\delta(q-1)/q^3$, associated with any other value of q. The sum of any finite number of intervals of I can therefore not be greater than 2δ times that of the series

$$1 + \frac{1}{2^3} + \frac{2}{3^3} + \frac{3}{4^3} + \dots,$$

which will be shown to be convergent in Ch. VIII. Hence it follows that, if δ is small enough, the supposition that every point of $[0, 1]$ lies inside an interval of I leads to a contradiction.

The reader may be tempted to think that this proof is needlessly elaborate, and that the existence of points of the interval, not in any interval of I, follows at once from the fact that the sum of all these intervals is less than 1. But the theorem to which he would be appealing is (when the set of intervals is infinite) far from obvious, and can only be proved rigorously by some such use of the Heine-Borel Theorem as is made in the text.

106. We shall now apply the Heine-Borel Theorem to the proof of two important theorems concerning the oscillation of a continuous function.

THEOREM I. *If $\phi(x)$ is continuous throughout the interval $[a, b]$, then we can divide $[a, b]$ into a finite number of sub-intervals $[a, x_1]$, $[x_1, x_2]$, ..., $[x_n, b]$, in each of which the oscillation of $\phi(x)$ is less than an assigned positive number ϵ.*

Let ξ be any number between a and b. Since $\phi(x)$ is continuous for $x = \xi$, we can determine an interval $[\xi - \delta, \xi + \delta]$ such that the oscillation of $\phi(x)$ in this interval is less than ϵ. It is indeed obvious that there are an infinity of such intervals corresponding to every ξ and every ϵ, for if the condition is satisfied for any particular value of δ, then it is satisfied *a fortiori* for any smaller value. What values of δ are admissible will naturally depend upon ξ; we have at present no reason for supposing that a value of δ admissible for one value of ξ will be admissible for another. We shall call the intervals thus associated with ξ *the ϵ-intervals of ξ*.

If $\xi = a$ then we can determine an interval $[a, a + \delta]$, and so an infinity of such intervals, having the same property. These we call the ϵ-intervals of a, and we can define in a similar manner the ϵ-intervals of b.

Consider now the set I of intervals formed by taking all the ϵ-intervals of all points of $[a, b]$. It is plain that this set satisfies the conditions of the Heine-Borel Theorem; every point interior to the interval is interior to at least one interval of I, and a and b are end points of at least one such interval. We can therefore determine a set I' which is formed by a finite number of intervals of I, and which possesses the same property as I itself.

The intervals which compose the set I' will in general overlap as in Fig. 34. But their end points obviously divide up $[a, b]$ into a finite set of

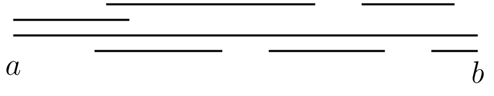

Fig. 34.

intervals I'' each of which is included in an interval of I', and in each of which the oscillation of $\phi(x)$ is less than ϵ. Thus Theorem I is proved.

THEOREM II. *Given any positive number ϵ, we can find a number η such that, if the interval $[a, b]$ is divided in any manner into sub-intervals of length less than η, then the oscillation of $\phi(x)$ in each of them will be less than ϵ.*

Take $\epsilon_1 < \frac{1}{2}\epsilon$, and construct, as in Theorem I, a finite set of sub-intervals j in each of which the oscillation of $\phi(x)$ is less than ϵ_1. Let η be the length of the least of these sub-intervals j. If now we divide $[a, b]$ into parts each of length less than η, then any such part must lie entirely within at most two successive sub-intervals j. Hence, in virtue of (3) of § 103, the oscillation of $\phi(x)$, in one of the parts of length less than η, cannot exceed twice the greatest oscillation of $\phi(x)$ in a sub-interval j, and is therefore less than $2\epsilon_1$, and therefore than ϵ.

This theorem is of fundamental importance in the theory of definite integrals (Ch. VII). It is impossible, without the use of this or some similar theorem, to prove that a function continuous throughout an interval necessarily possesses an integral over that interval.

107. Continuous functions of several variables. The notions of continuity and discontinuity may be extended to functions of several independent variables (Ch. II, §§ 31 *et seq.*). Their application to such functions however, raises questions much more complicated and difficult than those which we have considered in this chapter. It would be impossible for us to discuss these questions in any detail here; but we shall, in the sequel, require to know what is meant by a continuous function of two variables, and we accordingly give the following definition. It is a straightforward generalisation of the last form of the definition of § 98.

The function $\phi(x, y)$ of the two variables x and y is said to be **continuous** *for $x = \xi$, $y = \eta$ if, given any positive number ϵ, however small, we can choose $\delta(\epsilon)$ so that*

$$|\phi(x, y) - \phi(\xi, \eta)| < \epsilon$$

*when $0 \leqq |x - \xi| \leqq \delta(\epsilon)$ and $0 \leqq |y - \eta| \leqq \delta(\epsilon)$; that is to say if we can draw a square, whose sides are parallel to the axes of coordinates and of length $2\delta(\epsilon)$, whose centre is the point (ξ, η), and which is such that the value of $\phi(x, y)$ at any point inside it or on its boundary differs from $\phi(\xi, \eta)$ by less than ϵ.**

This definition of course presupposes that $\phi(x, y)$ is defined at all points of the square in question, and in particular at the point (ξ, η). Another method of stating the definition is this: *$\phi(x, y)$ is continuous for $x = \xi$, $y = \eta$ if $\phi(x, y) \to \phi(\xi, \eta)$ when $x \to \xi$, $y \to \eta$ in any manner.* This statement is apparently simpler; but it contains phrases the precise meaning of

*The reader should draw a figure to illustrate the definition.

which has not yet been explained and can only be explained by the help of inequalities like those which occur in our original statement.

It is easy to prove that the sums, the products, and in general the quotients of continuous functions of two variables are themselves continuous. A polynomial in two variables is continuous for all values of the variables; and the ordinary functions of x and y which occur in every-day analysis are *generally* continuous, *i.e.* are continuous except for pairs of values of x and y connected by special relations.

The reader should observe carefully that to assert the continuity of $\phi(x, y)$ with respect to the two variables x and y is to assert much more than its continuity with respect to each variable considered separately. It is plain that if $\phi(x, y)$ is continuous with respect to x and y then it is certainly continuous with respect to x (or y) when any fixed value is assigned to y (or x). But the converse is by no means true. Suppose, for example, that

$$\phi(x, y) = \frac{2xy}{x^2 + y^2}$$

when neither x nor y is zero, and $\phi(x, y) = 0$ when either x or y is zero. Then if y has any fixed value, zero or not, $\phi(x, y)$ is a continuous function of x, and in particular continuous for $x = 0$; for its value when $x = 0$ is zero, and it tends to the limit zero as $x \to 0$. In the same way it may be shown that $\phi(x, y)$ is a continuous function of y. But $\phi(x, y)$ is *not* a continuous function of x *and* y for $x = 0$, $y = 0$. Its value when $x = 0$, $y = 0$ is zero; but if x and y tend to zero along the straight line $y = ax$, then

$$\phi(x, y) = \frac{2a}{1 + a^2}, \quad \lim \phi(x, y) = \frac{2a}{1 + a^2},$$

which may have any value between -1 and 1.

108. Implicit functions. We have already, in Ch. II, met with the idea of an *implicit function*. Thus, if x and y are connected by the relation

$$y^5 - xy - y - x = 0, \tag{1}$$

then y is an 'implicit function' of x.

But it is far from obvious that such an equation as this does really define a function y of x, or several such functions. In Ch. II we were content to take this for granted. We are now in a position to consider whether the assumption we made then was justified.

We shall find the following terminology useful. Suppose that it is possible to surround a point (a, b), as in § 107, with a square throughout which a certain condition is satisfied. We shall call such a square a *neighbourhood* of (a, b), and say that the condition in question is satisfied *in the neighbourhood of* (a, b), or *near* (a, b), meaning by this simply that it is possible to find *some* square throughout which the condition is satisfied. It is obvious that similar language may be used when we are dealing with a single variable, the square being replaced by an interval on a line.

THEOREM. *If* (i) $f(x, y)$ *is a continuous function of x and y in the neighbourhood of* (a, b),

(ii) $f(a, b) = 0$,

(iii) $f(x, y)$ *is, for all values of x in the neighbourhood of a, a steadily increasing function of y, in the stricter sense of* § 95,

then (1) *there is a unique function* $y = \phi(x)$ *which, when substituted in the equation* $f(x, y) = 0$, *satisfies it identically for all values of x in the neighbourhood of a,*

(2) $\phi(x)$ *is continuous for all values of x in the neighbourhood of a.*

In the figure the square represents a 'neighbourhood' of (a, b) throughout which the conditions (i) and (iii) are satisfied, and P the point (a, b). If we take

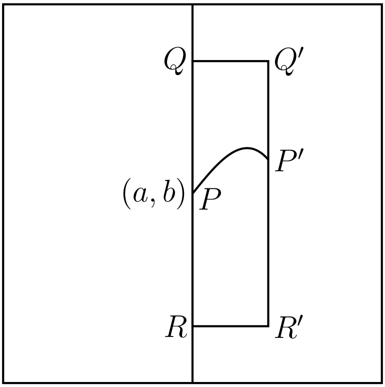

Fig. 35.

Q and R as in the figure, it follows from (iii) that $f(x, y)$ is positive at Q and negative at R. This being so, and $f(x, y)$ being continuous at Q and at R, we can draw lines QQ' and RR' parallel to OX, so that $R'Q'$ is parallel to OY and $f(x, y)$ is positive at all points of QQ' and negative at all points of RR'. In particular $f(x, y)$ is positive at Q' and negative at R', and therefore, in virtue of (iii) and § 100, vanishes once and only once at a point P' on $R'Q'$. The same construction gives us a unique point at which $f(x, y) = 0$ on each ordinate between RQ and $R'Q'$. It is obvious, moreover, that the same construction can be carried out to the left of RQ. The aggregate of points such as P' gives us the graph of the required function $y = \phi(x)$.

It remains to prove that $\phi(x)$ is continuous. This is most simply effected by using the idea of the 'limits of indetermination' of $\phi(x)$ as $x \to a$ (§ 96). Suppose that $x \to a$, and let λ and Λ be the limits of indetermination of $\phi(x)$ as $x \to a$. It is evident that the points (a, λ) and (a, Λ) lie on QR. Moreover, we can find a sequence of values of x such that $\phi(x) \to \lambda$ when $x \to a$ through the values of the sequence; and since $f\{x, \phi(x)\} = 0$, and $f(x, y)$ is a continuous function of x and y, we have

$$f(a, \lambda) = 0.$$

Hence $\lambda = b$; and similarly $\Lambda = b$. Thus $\phi(x)$ tends to the limit b as $x \to a$, and so $\phi(x)$ is continuous for $x = a$. It is evident that we can show in exactly the same way that $\phi(x)$ is continuous for any value of x in the neighbourhood of a.

It is clear that the truth of the theorem would not be affected if we were to change 'increasing' to 'decreasing' in condition (iii).

As an example, let us consider the equation (1), taking $a = 0$, $b = 0$. It is evident that the conditions (i) and (ii) are satisfied. Moreover

$$f(x, y) - f(x, y') = (y - y')(y^4 + y^3y' + y^2y'^2 + yy'^3 + y'^4 - x - 1)$$

has, when x, y, and y' are sufficiently small, the sign opposite to that of $y - y'$. Hence condition (iii) (with 'decreasing' for 'increasing') is satisfied. It follows that there is one and only one continuous function y which satisfies the equation (1) identically and vanishes with x.

The same conclusion would follow if the equation were

$$y^2 - xy - y - x = 0.$$

The function in question is in this case

$$y = \tfrac{1}{2}\{1 + x - \sqrt{1 + 6x + x^2}\},$$

where the square root is positive. The second root, in which the sign of the square root is changed, does not satisfy the condition of vanishing with x.

There is one point in the proof which the reader should be careful to observe. We supposed that the hypotheses of the theorem were satisfied 'in the neighbourhood of (a, b)', that is to say throughout a certain square $\xi - \delta \leqq x \leqq \xi + \delta$, $\eta - \delta \leqq y \leqq \eta + \delta$. The conclusion holds 'in the neighbourhood of $x = a$', that is to say throughout a certain interval $\xi - \delta_1 \leqq x \leqq \xi + \delta_1$. There is nothing to show that the δ_1 of the conclusion is the δ of the hypotheses, and indeed this is generally untrue.

109. Inverse Functions. Suppose in particular that $f(x, y)$ is of the form $F(y) - x$. We then obtain the following theorem.

If $F(y)$ is a function of y, continuous and steadily increasing (or decreasing), in the stricter sense of § 95, in the neighbourhood of $y = b$, and $F(b) = a$, then there is a unique continuous function $y = \phi(x)$ which is equal to b when $x = a$ and satisfies the equation $F(y) = x$ identically in the neighbourhood of $x = a$.

The function thus defined is called the *inverse function of $F(y)$*.

Suppose for example that $y^3 = x$, $a = 0$, $b = 0$. Then all the conditions of the theorem are satisfied. The inverse function is $x = \sqrt[3]{y}$.

If we had supposed that $y^2 = x$ then the conditions of the theorem would not have been satisfied, for y^2 is not a steadily increasing function of y in any interval which includes $y = 0$: it decreases when y is negative and increases when y is positive. And in this case the conclusion of the theorem does not hold, for $y^2 = x$ defines *two* functions of x, viz. $y = \sqrt{x}$ and $y = -\sqrt{x}$, both of which vanish when $x = 0$, and each of which is defined only for positive values of x, so that the equation has sometimes two solutions and sometimes none. The reader should consider the more general equations

$$y^{2n} = x, \quad y^{2n+1} = x,$$

in the same way. Another interesting example is given by the equation

$$y^5 - y - x = 0,$$

already considered in Ex. XIV. 7.

Similarly the equation

$$\sin y = x$$

has just one solution which vanishes with x, viz. the value of $\arcsin x$ which vanishes with x. There are of course an infinity of solutions, given by the other values of $\arcsin x$ (cf. Ex. XV. 10), which do not satisfy this condition.

So far we have considered only what happens in the neighbourhood of a particular value of x. Let us suppose now that $F(y)$ is positive and steadily increasing (or decreasing) throughout an interval $[a, b]$. Given any point ξ of $[a, b]$, we can determine an interval i including ξ, and a unique and continuous inverse function $\phi_i(x)$ defined throughout i.

From the set I of intervals i we can, in virtue of the Heine-Borel Theorem, pick out a finite sub-set covering up the whole interval $[a, b]$; and it is plain that the finite set of functions $\phi_i(x)$, corresponding to the sub-set of intervals i thus selected, define together a unique inverse function $\phi(x)$ continuous throughout $[a, b]$.

We thus obtain the theorem: *if $x = F(y)$, where $F(y)$ is continuous and increases steadily and strictly from A to B as x increases from a to b, then there is a unique inverse function $y = \phi(x)$ which is continuous and increases steadily and strictly from a to b as x increases from A to B.*

It is worth while to show how this theorem can be obtained directly without the help of the more difficult theorem of § 108. Suppose that $A < \xi < B$, and consider the class of values of y such that (i) $a < y < b$ and (ii) $F(y) \leqq \xi$. This class has an upper bound η, and plainly $F(\eta) \leqq \xi$. If $F(\eta)$ were less than ξ, we could find a value of y such that $y > \eta$ and $F(y) < \xi$, and η would not be the upper bound of the class considered. Hence $F(\eta) = \xi$. The equation $F(y) = \xi$ has therefore a unique solution $y = \eta = \phi(\xi)$, say; and plainly η increases steadily and continuously with ξ, which proves the theorem.

MISCELLANEOUS EXAMPLES ON CHAPTER V.

1. Show that, if neither a nor b is zero, then

$$ax^n + bx^{n-1} + \dots + k = ax^n(1 + \epsilon_x),$$

where ϵ_x is of the first order of smallness when x is large.

2. If $P(x) = ax^n + bx^{n-1} + \dots + k$, and a is not zero, then as x increases $P(x)$ has ultimately the sign of a; and so has $P(x + \lambda) - P(x)$, where λ is any constant.

3. Show that in general

$$(ax^n + bx^{n-1} + \cdots + k)/(Ax^n + Bx^{n-1} + \cdots + K) = \alpha + (\beta/x)(1 + \epsilon_x),$$

where $\alpha = a/A$, $\beta = (bA - aB)/A^2$, and ϵ_x is of the first order of smallness when x is large. Indicate any exceptional cases.

4. Express

$$(ax^2 + bx + c)/(Ax^2 + Bx + C)$$

in the form

$$\alpha + (\beta/x) + (\gamma/x^2)(1 + \epsilon_x),$$

where ϵ_x is of the first order of smallness when x is large.

5. Show that

$$\lim_{x\to\infty} \sqrt{x}\{\sqrt{x+a} - \sqrt{x}\} = \tfrac{1}{2}a.$$

[Use the formula $\sqrt{x+a} - \sqrt{x} = a/\{\sqrt{x+a} + \sqrt{x}\}$.]

6. Show that $\sqrt{x+a} = \sqrt{x} + \frac{1}{2}(a/\sqrt{x})(1+\epsilon_x)$, where ϵ_x is of the first order of smallness when x is large.

7. Find values of α and β such that $\sqrt{ax^2 + 2bx + c} - \alpha x - \beta$ has the limit zero as $x \to \infty$; and prove that $\lim x\{\sqrt{ax^2 + 2bx + c} - \alpha x - \beta\} = (ac - b^2)/2a$.

8. Evaluate

$$\lim_{x\to\infty} x\left\{\sqrt{x^2 + \sqrt{x^4+1}} - x\sqrt{2}\right\}.$$

9. Prove that $(\sec x - \tan x) \to 0$ as $x \to \frac{1}{2}\pi$.

10. Prove that $\phi(x) = 1 - \cos(1 - \cos x)$ is of the fourth order of smallness when x is small; and find the limit of $\phi(x)/x^4$ as $x \to 0$.

11. Prove that $\phi(x) = x\sin(\sin x) - \sin^2 x$ is of the sixth order of smallness when x is small; and find the limit of $\phi(x)/x^6$ as $x \to 0$.

12. From a point P on a radius OA of a circle, produced beyond the circle, a tangent PT is drawn to the circle, touching it in T, and TN is drawn perpendicular to OA. Show that $NA/AP \to 1$ as P moves up to A.

13. Tangents are drawn to a circular arc at its middle point and its extremities; Δ is the area of the triangle formed by the chord of the arc and the two tangents at the extremities, and Δ' the area of that formed by the three tangents. Show that $\Delta/\Delta' \to 4$ as the length of the arc tends to zero.

14. For what values of a does $\{a + \sin(1/x)\}/x$ tend to (1) ∞, (2) $-\infty$, as $x \to 0$? [To ∞ if $a > 1$, to $-\infty$ if $a < -1$: the function oscillates if $-1 \leqq a \leqq 1$.]

15. If $\phi(x) = 1/q$ when $x = p/q$, and $\phi(x) = 0$ when x is irrational, then $\phi(x)$ is continuous for all irrational and discontinuous for all rational values of x.

16. Show that the function whose graph is drawn in Fig. 32 may be represented by either of the formulae

$$1 - x + [x] - [1 - x], \quad 1 - x - \lim_{n\to\infty} (\cos^{2n+1} \pi x).$$

17. Show that the function $\phi(x)$ which is equal to 0 when $x = 0$, to $\frac{1}{2} - x$ when $0 < x < \frac{1}{2}$, to $\frac{1}{2}$ when $x = \frac{1}{2}$, to $\frac{3}{2} - x$ when $\frac{1}{2} < x < 1$, and to 1 when $x = 1$, assumes every value between 0 and 1 once and once only as x increases from 0 to 1, but is discontinuous for $x = 0$, $x = \frac{1}{2}$, and $x = 1$. Show also that the function may be represented by the formula

$$\tfrac{1}{2} - x - \tfrac{1}{2}[2x] - \tfrac{1}{2}[1 - 2x].$$

18. Let $\phi(x) = x$ when x is rational and $\phi(x) = 1 - x$ when x is irrational. Show that $\phi(x)$ assumes every value between 0 and 1 once and once only as x increases from 0 to 1, but is discontinuous for every value of x except $x = \frac{1}{2}$.

19. As x increases from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$, $y = \sin x$ is continuous and steadily increases, in the stricter sense, from -1 to 1. Deduce the existence of a function $x = \arcsin y$ which is a continuous and steadily increasing function of y from $y = -1$ to $y = 1$.

20. Show that the numerically least value of $\arctan y$ is continuous for all values of y and increases steadily from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$ as y varies through all real values.

21. Discuss, on the lines of §§ 108–109, the solution of the equations

$$y^2 - y - x = 0, \quad y^4 - y^2 - x^2 = 0, \quad y^4 - y^2 + x^2 = 0$$

in the neighbourhood of $x = 0$, $y = 0$.

22. If $ax^2 + 2bxy + cy^2 + 2dx + 2ey = 0$ and $\Delta = 2bde - ae^2 - cd^2$, then one value of y is given by $y = \alpha x + \beta x^2 + (\gamma + \epsilon_x)x^3$, where

$$\alpha = -d/e, \quad \beta = \Delta/2e^3, \quad \gamma = (cd - be)\Delta/2e^5,$$

and ϵ_x is of the first order of smallness when x is small.

[If $y - \alpha x = \eta$ then

$$-2e\eta = ax^2 + 2bx(\eta + \alpha x) + c(\eta + \alpha x)^2 = Ax^2 + 2Bx\eta + C\eta^2,$$

say. It is evident that η is of the second order of smallness, $x\eta$ of the third, and η^2 of the fourth; and $-2e\eta = Ax^2 - (AB/e)x^3$, the error being of the fourth order.]

23. If $x = ay + by^2 + cy^3$ then one value of y is given by

$$y = \alpha x + \beta x^2 + (\gamma + \epsilon_x)x^3,$$

where $\alpha = 1/a$, $\beta = -b/a^3$, $\gamma = (2b^2 - ac)/a^5$, and ϵ_x is of the first order of smallness when x is small.

24. If $x = ay + by^n$, where n is an integer greater than unity, then one value of y is given by $y = \alpha x + \beta x^n + (\gamma + \epsilon_x)x^{2n-1}$, where $\alpha = 1/a$, $\beta = -b/a^{n+1}$, $\gamma = nb^2/a^{2n+1}$, and ϵ_x is of the $(n-1)$th order of smallness when x is small.

25. Show that the least positive root of the equation $xy = \sin x$ is a continuous function of y throughout the interval $[0, 1]$, and decreases steadily from π to 0 as y increases from 0 to 1. [The function is the inverse of $(\sin x)/x$: apply § 109.]

26. The least positive root of $xy = \tan x$ is a continuous function of y throughout the interval $[1, \infty)$, and increases steadily from 0 to $\frac{1}{2}\pi$ as y increases from 1 towards ∞.

CHAPTER VI

DERIVATIVES AND INTEGRALS

110. Derivatives or Differential Coefficients. Let us return to the consideration of the properties which we naturally associate with the notion of a curve. The first and most obvious property is, as we saw in the last chapter, that which gives a curve its appearance of connectedness, and which we embodied in our definition of a continuous function.

The ordinary curves which occur in elementary geometry, such as straight lines, circles and conic sections, have of course many other properties of a general character. The simplest and most noteworthy of these is perhaps that they have a definite *direction* at every point, or what is the same thing, that at every point of the curve we can draw a *tangent* to it. The reader will probably remember that in elementary geometry the tangent to a curve at P is defined to be 'the limiting position of the chord PQ, when Q moves up towards coincidence with P'. Let us consider what is implied in the assumption of the existence of such a limiting position.

In the figure (Fig. 36) P is a fixed point on the curve, and Q a variable point; PM, QN are parallel to OY and PR to OX. We denote the coordinates of P by x, y and those of Q by $x+h$, $y+k$: h will be positive or negative according as N lies to the right or left of M.

We have assumed that there is a tangent to the curve at P, or that there is a definite 'limiting position' of the chord PQ. Suppose that PT, the tangent at P, makes an angle ψ with OX. Then to say that PT is the limiting position of PQ is equivalent to saying that the limit of the angle QPR is ψ, when Q approaches P along the curve from either side. We have now to distinguish two cases, a general case and an exceptional one.

The general case is that in which ψ is not equal to $\frac{1}{2}\pi$, so that PT is not parallel to OY. In this case RPQ tends to the limit ψ, and

$$RQ/PR = \tan RPQ$$

tends to the limit $\tan\psi$. Now

$$RQ/PR = (NQ - MP)/MN = \{\phi(x+h) - \phi(x)\}/h;$$

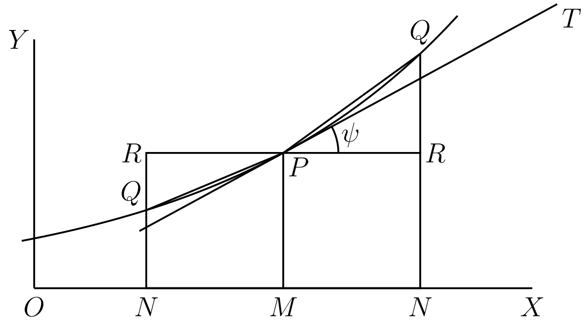

Fig. 36.

and so

$$\lim_{h \to 0} \frac{\phi(x+h) - \phi(x)}{h} = \tan \psi. \tag{1}$$

The reader should be careful to note that in all these equations all lengths are regarded as affected with the proper sign, so that (*e.g.*) RQ is negative in the figure when Q lies to the left of P; and that the convergence to the limit is unaffected by the sign of h.

Thus the assumption that the curve which is the graph of $\phi(x)$ has a tangent at P, which is not perpendicular to the axis of x, implies that $\phi(x)$ has, for the particular value of x corresponding to P, the property that $\{\phi(x+h) - \phi(x)\}/h$ *tends to a limit when* h *tends to zero.*

This of course implies that both of

$$\{\phi(x+h) - \phi(x)\}/h, \quad \{\phi(x-h) - \phi(x)\}/(-h)$$

tend to limits when $h \to 0$ by positive values only, and that the two limits are equal. If these limits exist but are not equal, then the curve $y = \phi(x)$ has an angle at the particular point considered, as in Fig. 37.

Now let us suppose that the curve has (like the circle or ellipse) a tangent at every point of its length, or at any rate every portion of its length which corresponds to a certain range of variation of x. Further let us suppose this tangent never perpendicular to the axis of x: in the case of a circle this would of course restrict us to considering an arc less than a semicircle. Then an equation such as (1) holds for all values of x which fall inside this range. To each such value of x corresponds a value of $\tan\psi$: $\tan\psi$ is a function of x, which is defined for all values of x in the range of values under consideration, and which may be calculated or *derived* from the original function $\phi(x)$. We shall call this function the **derivative** or *derived function* of $\phi(x)$, and we shall denote it by

$$\phi'(x).$$

Another name for the derived function of $\phi(x)$ is the **differential coefficient** of $\phi(x)$; and the operation of calculating $\phi'(x)$ from $\phi(x)$ is generally known as **differentiation**. This terminology is firmly established for historical reasons: see § 115.

Before we proceed to consider the special case mentioned above, in which $\psi = \frac{1}{2}\pi$, we shall illustrate our definition by some general remarks and particular illustrations.

111. Some general remarks. (1) The existence of a derived function $\phi'(x)$ for all values of x in the interval $a \leqq x \leqq b$ implies that $\phi(x)$ is continuous at every point of this interval. For it is evident that $\{\phi(x+h)-\phi(x)\}/h$ cannot tend to a limit unless $\lim\phi(x+h) = \phi(x)$, and it is this which is the property denoted by continuity.

(2) It is natural to ask whether the converse is true, *i.e.* whether every continuous curve has a definite tangent at every point, and every function

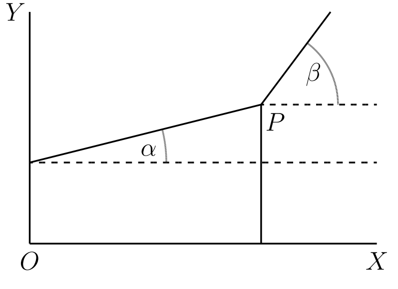

Fig. 37.

a differential coefficient for every value of x for which it is continuous.* The

*We leave out of account the exceptional case (which we have still to examine) in answer is obviously *No*: it is sufficient to consider the curve formed by two straight lines meeting to form an angle (Fig. 37). The reader will see at once that in this case $\{\phi(x+h)-\phi(x)\}/h$ has the limit $\tan\beta$ when $h \to 0$ by positive values and the limit $\tan\alpha$ when $h \to 0$ by negative values.

This is of course a case in which a curve might reasonably be said to have *two* directions at a point. But the following example, although a little more difficult, shows conclusively that there are cases in which a continuous curve cannot be said to have either one direction or several directions at one of its points. Draw the graph (Fig. 14, p. 61) of the function $x\sin(1/x)$. The function is not defined for $x = 0$, and so is discontinuous for $x = 0$. On the other hand the function defined by the equations

$$\phi(x) = x\sin(1/x) \quad (x \neq 0), \qquad \phi(x) = 0 \quad (x = 0)$$

is continuous for $x = 0$ (Exs. XXXVII. 14, 15), and the graph of this function is a continuous curve.

But $\phi(x)$ has no derivative for $x = 0$. For $\phi'(0)$ would be, by definition, $\lim\{\phi(h)-\phi(0)\}/h$ or $\lim\sin(1/h)$; and no such limit exists.

It has even been shown that a function of x may be continuous and yet have no derivative for *any* value of x, but the proof of this is much more difficult. The reader who is interested in the question may be referred to Bromwich's *Infinite Series*, pp. 490–1, or Hobson's *Theory of Functions of a Real Variable*, pp. 620–5.

(3) The notion of a derivative or differential coefficient was suggested to us by geometrical considerations. But there is nothing geometrical in the notion itself. The derivative $\phi'(x)$ of a function $\phi(x)$ may be defined, without any reference to any kind of geometrical representation of $\phi(x)$, by the equation

$$\phi'(x) = \lim_{h\to 0}\frac{\phi(x+h)-\phi(x)}{h};$$

and $\phi(x)$ has or has not a derivative, for any particular value of x, according as this limit does or does not exist. The geometry of curves is merely one

which the curve is supposed to have a tangent perpendicular to OX: apart from this possibility the two forms of the question stated above are equivalent.

of many departments of mathematics in which the idea of a derivative finds an application.

Another important application is in dynamics. Suppose that a particle is moving in a straight line in such a way that at time t its distance from a fixed point on the line is $s = \phi(t)$. Then the 'velocity of the particle at time t' is by definition the limit of

$$\frac{\phi(t+h) - \phi(t)}{h}$$

as $h \to 0$. The notion of 'velocity' is in fact merely a special case of that of the derivative of a function.

Examples XXXIX. 1. If $\phi(x)$ is a constant then $\phi'(x) = 0$. Interpret this result geometrically.

2. If $\phi(x) = ax + b$ then $\phi'(x) = a$. Prove this (i) from the formal definition and (ii) by geometrical considerations.

3. If $\phi(x) = x^m$, where m is a positive integer, then $\phi'(x) = mx^{m-1}$.

[For

$$\begin{aligned}\phi'(x) &= \lim \frac{(x+h)^m - x^m}{h} \\ &= \lim \left\{ mx^{m-1} + \frac{m(m-1)}{1 \cdot 2} x^{m-2} h + \cdots + h^{m-1} \right\}.\end{aligned}$$

The reader should observe that this method cannot be applied to $x^{p/q}$, where p/q is a rational fraction, as we have no means of expressing $(x+h)^{p/q}$ as a finite series of powers of h. We shall show later on (§ 118) that the result of this example holds for all rational values of m. Meanwhile the reader will find it instructive to determine $\phi'(x)$ when m has some special fractional value (*e.g.* $\frac{1}{2}$), by means of some special device.]

4. If $\phi(x) = \sin x$, then $\phi'(x) = \cos x$; and if $\phi(x) = \cos x$, then $\phi'(x) = -\sin x$.

[For example, if $\phi(x) = \sin x$, we have

$$\{\phi(x+h) - \phi(x)\}/h = \{2 \sin \tfrac{1}{2}h \cos(x + \tfrac{1}{2}h)\}/h,$$

the limit of which, when $h \to 0$, is $\cos x$, since $\lim \cos(x + \frac{1}{2}h) = \cos x$ (the cosine being a continuous function) and $\lim\{(\sin \frac{1}{2}h)/\frac{1}{2}h\} = 1$ (Ex. XXXVI. 13).]

5. **Equations of the tangent and normal to a curve** $y = \phi(x)$**.** The tangent to the curve at the point (x_0, y_0) is the line through (x_0, y_0) which makes with OX an angle ψ, where $\tan\psi = \phi'(x_0)$. Its equation is therefore

$$y - y_0 = (x - x_0)\phi'(x_0);$$

and the equation of the normal (the perpendicular to the tangent at the point of contact) is

$$(y - y_0)\phi'(x_0) + x - x_0 = 0.$$

We have assumed that the tangent is not parallel to the axis of y. In this special case it is obvious that the tangent and normal are $x = x_0$ and $y = y_0$ respectively.

6. Write down the equations of the tangent and normal at any point of the parabola $x^2 = 4ay$. Show that if $x_0 = 2a/m$, $y_0 = a/m^2$, then the tangent at (x_0, y_0) is $x = my + (a/m)$.

112. We have seen that if $\phi(x)$ is not continuous for a value of x then it cannot possibly have a derivative for that value of x. Thus such functions as $1/x$ or $\sin(1/x)$, which are not defined for $x = 0$, and so necessarily discontinuous for $x = 0$, cannot have derivatives for $x = 0$. Or again the function $[x]$, which is discontinuous for every integral value of x, has no derivative for any such value of x.

Example. Since $[x]$ is constant between every two integral values of x, its derivative, whenever it exists, has the value zero. Thus the derivative of $[x]$, which we may represent by $[x]'$, is a function equal to zero for all values of x save integral values and undefined for integral values. It is interesting to note that the function $1 - \dfrac{\sin\pi x}{\sin\pi x}$ has exactly the same properties.

We saw also in Ex. XXXVII. 7 that the types of discontinuity which occur most commonly, when we are dealing with the very simplest and most obvious kinds of functions, such as polynomials or rational or trigonometrical functions, are associated with a relation of the type

$$\phi(x) \to +\infty$$

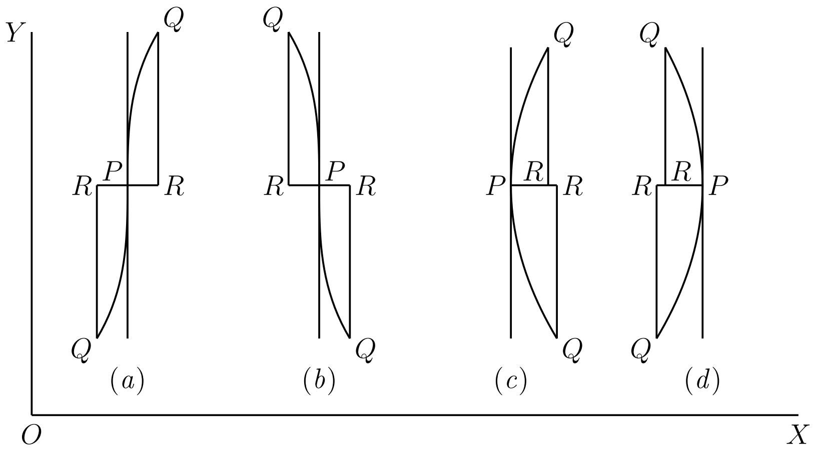

Fig. 38.

or $\phi(x) \to -\infty$. In all these cases, as in such cases as those considered above, there is no derivative for certain special values of x. In fact, as was pointed out in § 111, (1), *all discontinuities of* $\phi(x)$ *are also discontinuities of* $\phi'(x)$. But the converse is not true, as we may easily see if we return to the geometrical point of view of § 110 and consider the special case, hitherto left aside, in which the graph of $\phi(x)$ has a tangent parallel to OY. This case may be subdivided into a number of cases, of which the most typical are shown in Fig. 38. In cases (c) and (d) the function is two valued on one side of P and not defined on the other. In such cases we may consider the two sets of values of $\phi(x)$, which occur on one side of P or the other, as defining distinct functions $\phi_1(x)$ and $\phi_2(x)$, the upper part of the curve corresponding to $\phi_1(x)$.

The reader will easily convince himself that in (a)

$$\{\phi(x+h) - \phi(x)\}/h \to +\infty,$$

as $h \to 0$, and in (b)

$$\{\phi(x+h) - \phi(x)\}/h \to -\infty;$$

while in (c)

$$\{\phi_1(x+h) - \phi_1(x)\}/h \to +\infty, \quad \{\phi_2(x+h) - \phi_2(x)\}/h \to -\infty,$$

and in (d)

$$\{\phi_1(x+h)-\phi_1(x)\}/h \to -\infty, \quad \{\phi_2(x+h)-\phi_2(x)\}/h \to +\infty,$$

though of course in (c) only positive and in (d) only negative values of h can be considered, a fact which by itself would preclude the existence of a derivative.

We can obtain examples of these four cases by considering the functions defined by the equations

$$(a)\ y^3 = x, \quad (b)\ y^3 = -x, \quad (c)\ y^2 = x, \quad (d)\ y^2 = -x,$$

the special value of x under consideration being $x = 0$.

113. Some general rules for differentiation. Throughout the theorems which follow we assume that the functions $f(x)$ and $F(x)$ have derivatives $f'(x)$ and $F'(x)$ for the values of x considered.

(1) *If $\phi(x) = f(x) + F(x)$, then $\phi(x)$ has a derivative*

$$\phi'(x) = f'(x) + F'(x).$$

(2) *If $\phi(x) = kf(x)$, where k is a constant, then $\phi(x)$ has a derivative*

$$\phi'(x) = kf'(x).$$

We leave it as an exercise to the reader to deduce these results from the general theorems stated in Ex. XXXV. 1.

(3) *If $\phi(x) = f(x)F(x)$, then $\phi(x)$ has a derivative*

$$\phi'(x) = f(x)F'(x) + f'(x)F(x).$$

For

$$\begin{aligned}\phi'(x) &= \lim \frac{f(x+h)F(x+h) - f(x)F(x)}{h}\\ &= \lim\left\{f(x+h)\frac{F(x+h)-F(x)}{h} + F(x)\frac{f(x+h)-f(x)}{h}\right\}\\ &= f(x)F'(x) + F(x)f'(x).\end{aligned}$$

(4) *If* $\phi(x) = \dfrac{1}{f(x)}$, *then* $\phi(x)$ *has a derivative*

$$\phi'(x) = -\frac{f'(x)}{\{f(x)\}^2}.$$

In this theorem we of course suppose that $f(x)$ is not equal to zero for the particular value of x under consideration. Then

$$\phi'(x) = \lim \frac{1}{h}\left\{\frac{f(x) - f(x+h)}{f(x+h)f(x)}\right\} = -\frac{f'(x)}{\{f(x)\}^2}.$$

(5) *If* $\phi(x) = \dfrac{f(x)}{F(x)}$, *then* $\phi(x)$ *has a derivative*

$$\phi'(x) = \frac{f'(x)F(x) - f(x)F'(x)}{\{F(x)\}^2}.$$

This follows at once from (3) and (4).

(6) *If* $\phi(x) = F\{f(x)\}$, *then* $\phi(x)$ *has a derivative*

$$\phi'(x) = F'\{f(x)\}f'(x).$$

For let

$$f(x) = y, \quad f(x+h) = y + k.$$

Then $k \to 0$ as $h \to 0$, and $k/h \to f'(x)$. And

$$\begin{aligned}\phi'(x) &= \lim \frac{F\{f(x+h)\} - F\{f(x)\}}{h}\\ &= \lim\left\{\frac{F(y+k) - F(y)}{k}\right\} \times \lim\left(\frac{k}{h}\right)\\ &= F'(y)f'(x).\end{aligned}$$

This theorem includes (2) and (4) as special cases, as we see on taking $F(x) = kx$ or $F(x) = 1/x$. Another interesting special case is that in which

$f(x) = ax + b$: the theorem then shows that the derivative of $F(ax + b)$ is $aF'(ax + b)$.

Our last theorem requires a few words of preliminary explanation. Suppose that $x = \psi(y)$, where $\psi(y)$ is continuous and steadily increasing (or decreasing), in the stricter sense of § 95, in a certain interval of values of y. Then we may write $y = \phi(x)$, where ϕ is the 'inverse' function (§ 109) of ψ.

(7) *If $y = \phi(x)$, where ϕ is the inverse function of ψ, so that $x = \psi(y)$, and $\psi(y)$ has a derivative $\psi'(y)$ which is not equal to zero, then $\phi(x)$ has a derivative*

$$\phi'(x) = \frac{1}{\psi'(y)}.$$

For if $\phi(x + h) = y + k$, then $k \to 0$ as $h \to 0$, and

$$\phi'(x) = \lim_{h\to 0} \frac{\phi(x + h) - \phi(x)}{(x + h) - x} = \lim_{k\to 0} \frac{(y + k) - y}{\psi(y + k) - \psi(y)} = \frac{1}{\psi'(y)}.$$

The last function may now be expressed in terms of x by means of the relation $y = \phi(x)$, so that $\phi'(x)$ is the reciprocal of $\psi'\{\phi(x)\}$. This theorem enables us to differentiate any function if we know the derivative of the inverse function.

114. Derivatives of complex functions. So far we have supposed that $y = \phi(x)$ is a purely *real* function of x. If y is a complex function $\phi(x) + i\psi(x)$, then we define the derivative of y as being $\phi'(x) + i\psi'(x)$. The reader will have no difficulty in seeing that Theorems (1)–(5) above retain their validity when $\phi(x)$ is complex. Theorems (6) and (7) have also analogues for complex functions, but these depend upon the general notion of a 'function of a complex variable', a notion which we have encountered at present only in a few particular cases.

115. The notation of the differential calculus. We have already explained that what we call a *derivative* is often called a *differential coefficient*. Not only a different name but a different notation is often used; the derivative of the function $y = \phi(x)$ is often denoted by one or other of the expressions

$$D_x y, \quad \frac{dy}{dx}.$$

Of these the last is the most usual and convenient: the reader must however be careful to remember that dy/dx does not mean 'a certain number dy divided by another number dx': it means 'the result of a certain operation D_x or d/dx applied to $y = \phi(x)$', the operation being that of forming the quotient $\{\phi(x+h) - \phi(x)\}/h$ and making $h \to 0$.

Of course a notation at first sight so peculiar would not have been adopted without some reason, and the reason was as follows. The denominator h of the fraction $\{\phi(x+h)-\phi(x)\}/h$ is the difference of the values $x+h$, x of the independent variable x; similarly the numerator is the difference of the corresponding values $\phi(x+h)$, $\phi(x)$ of the dependent variable y. These differences may be called the *increments* of x and y respectively, and denoted by δx and δy. Then the fraction is $\delta y/\delta x$, and it is for many purposes convenient to denote the limit of the fraction, which is the same thing as $\phi'(x)$, by dy/dx. But this notation must for the present be regarded as purely symbolical. The dy and dx which occur in it cannot be separated, and standing by themselves they would mean nothing: in particular dy and dx do not mean $\lim \delta y$ and $\lim \delta x$, these limits being simply equal to zero. The reader will have to become familiar with this notation, but so long as it puzzles him he will be wise to avoid it by writing the differential coefficient in the form $D_x y$, or using the notation $\phi(x)$, $\phi'(x)$, as we have done in the preceding sections of this chapter.

In Ch. VII, however, we shall show how it is possible to define the symbols dx and dy in such a way that they have an independent meaning and that the derivative dy/dx is actually their quotient.

The theorems of § 113 may of course at once be translated into this notation. They may be stated as follows:

(1) *if* $y = y_1 + y_2$, *then*

$$\frac{dy}{dx} = \frac{dy_1}{dx} + \frac{dy_2}{dx};$$

(2) *if* $y = ky_1$, *then*

$$\frac{dy}{dx} = k\frac{dy_1}{dx};$$

(3) *if* $y = y_1 y_2$, *then*

$$\frac{dy}{dx} = y_1 \frac{dy_2}{dx} + y_2 \frac{dy_1}{dx};$$

(4) *if* $y = \dfrac{1}{y_1}$, *then*

$$\frac{dy}{dx} = -\frac{1}{y_1^2}\,\frac{dy_1}{dx};$$

(5) *if* $y = \dfrac{y_1}{y_2}$, *then*

$$\frac{dy}{dx} = \left(y_2 \frac{dy_1}{dx} - y_1 \frac{dy_2}{dx}\right) \Big/ y_2^2;$$

(6) *if y is a function of x, and z a function of y, then*

$$\frac{dz}{dx} = \frac{dz}{dy}\,\frac{dy}{dx};$$

(7) $$\frac{dy}{dx} = 1 \Big/ \left(\frac{dx}{dy}\right).$$

Examples XL. 1. If $y = y_1 y_2 y_3$ then

$$\frac{dy}{dx} = y_2 y_3 \frac{dy_1}{dx} + y_3 y_1 \frac{dy_2}{dx} + y_1 y_2 \frac{dy_3}{dx},$$

and if $y = y_1 y_2 \dots y_n$ then

$$\frac{dy}{dx} = \sum_{r=1}^{n} y_1 y_2 \dots y_{r-1} y_{r+1} \dots y_n \frac{dy_r}{dx}.$$

In particular, if $y = z^n$, then $dy/dx = nz^{n-1}(dz/dx)$; and if $y = x^n$, then $dy/dx = nx^{n-1}$, as was proved otherwise in Ex. xxxix. 3.

2. If $y = y_1 y_2 \dots y_n$ then

$$\frac{1}{y}\,\frac{dy}{dx} = \frac{1}{y_1}\,\frac{dy_1}{dx} + \frac{1}{y_2}\,\frac{dy_2}{dx} + \dots + \frac{1}{y_n}\,\frac{dy_n}{dx}.$$

In particular, if $y = z^n$, then $\dfrac{1}{y}\,\dfrac{dy}{dx} = \dfrac{n}{z}\,\dfrac{dz}{dx}$.

116. Standard forms. We shall now investigate more systematically the forms of the derivatives of a few of the the simplest types of functions.

A. **Polynomials.** If $\phi(x) = a_0x^n + a_1x^{n-1} + \dots + a_n$, then

$$\phi'(x) = na_0x^{n-1} + (n-1)a_1x^{n-2} + \dots + a_{n-1}.$$

It is sometimes more convenient to use for the standard form of a polynomial of degree n in x what is known as the *binomial form*, viz.

$$a_0x^n + \binom{n}{1}a_1x^{n-1} + \binom{n}{2}a_2x^{n-2} + \dots + a_n.$$

In this case

$$\phi'(x) = n\left\{a_0x^{n-1} + \binom{n-1}{1}a_1x^{n-2} + \binom{n-1}{2}a_2x^{n-3} + \dots + a_{n-1}\right\}.$$

The binomial form of $\phi(x)$ is often written symbolically as

$$(a_0, a_1, \dots, a_n \between x, 1)^n;$$

and then

$$\phi'(x) = n(a_0, a_1, \dots, a_{n-1} \between x, 1)^{n-1}.$$

We shall see later that $\phi(x)$ can always be expressed as the product of n factors in the form

$$\phi(x) = a_0(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n),$$

where the α's are real or complex numbers. Then

$$\phi'(x) = a_0\textstyle\sum(x - \alpha_2)(x - \alpha_3)\dots(x - \alpha_n),$$

the notation implying that we form all possible products of $n - 1$ factors, and add them all together. This form of the result holds even if several of the numbers α are equal; but of course then some of the terms on the right-hand side are repeated. The reader will easily verify that if

$$\phi(x) = a_0(x - \alpha_1)^{m_1}(x - \alpha_2)^{m_2}\dots(x - \alpha_\nu)^{m_\nu},$$

then

$$\phi'(x) = a_0\textstyle\sum m_1(x - \alpha_1)^{m_1-1}(x - \alpha_2)^{m_2}\dots(x - \alpha_\nu)^{m_\nu}.$$

Examples XLI. 1. Show that if $\phi(x)$ is a polynomial then $\phi'(x)$ is the coefficient of h in the expansion of $\phi(x+h)$ in powers of h.

2. If $\phi(x)$ is divisible by $(x-\alpha)^2$, then $\phi'(x)$ is divisible by $x-\alpha$: and generally, if $\phi(x)$ is divisible by $(x-\alpha)^m$, then $\phi'(x)$ is divisible by $(x-\alpha)^{m-1}$.

3. Conversely, if $\phi(x)$ and $\phi'(x)$ are *both* divisible by $x-\alpha$, then $\phi(x)$ is divisible by $(x-\alpha)^2$; and if $\phi(x)$ is divisible by $x-\alpha$ and $\phi'(x)$ by $(x-\alpha)^{m-1}$, then $\phi(x)$ is divisible by $(x-\alpha)^m$.

4. Show how to determine as completely as possible the multiple roots of $P(x)=0$, where $P(x)$ is a polynomial, with their degrees of multiplicity, by means of the elementary algebraical operations.

[If H_1 is the highest common factor of P and P', H_2 the highest common factor of H_1 and P'', H_3 that of H_2 and P''', and so on, then the roots of $H_1H_3/H_2^2=0$ are the *double* roots of $P=0$, the roots of $H_2H_4/H_3^2=0$ the *treble* roots, and so on. But it may not be possible to complete the solution of $H_1H_3/H_2^2=0$, $H_2H_4/H_3^2=0$, ... Thus if $P(x)=(x-1)^3(x^5-x-7)^2$ then $H_1H_3/H_2^2=x^5-x-7$ and $H_2H_4/H_3^2=x-1$; and we cannot solve the first equation.]

5. Find all the roots, with their degrees of multiplicity, of

$$x^4+3x^3-3x^2-11x-6=0, \quad x^6+2x^5-8x^4-14x^3+11x^2+28x+12=0.$$

6. If $ax^2+2bx+c$ has a double root, *i.e.* is of the form $a(x-\alpha)^2$, then $2(ax+b)$ must be divisible by $x-\alpha$, so that $\alpha=-b/a$. This value of x must satisfy $ax^2+2bx+c=0$. Verify that the condition thus arrived at is $ac-b^2=0$.

7. The equation $1/(x-a)+1/(x-b)+1/(x-c)=0$ can have a pair of equal roots only if $a=b=c$. (*Math. Trip.* 1905.)

8. Show that

$$ax^3+3bx^2+3cx+d=0$$

has a double root if $G^2+4H^3=0$, where $H=ac-b^2$, $G=a^2d-3abc+2b^3$.

[Put $ax+b=y$, when the equation reduces to $y^3+3Hy+G=0$. This must have a root in common with $y^2+H=0$.]

9. The reader may verify that if α, β, γ, δ are the roots of

$$ax^4+4bx^3+6cx^2+4dx+e=0,$$

then the equation whose roots are

$$\tfrac{1}{12}a\{(\alpha-\beta)(\gamma-\delta)-(\gamma-\alpha)(\beta-\delta)\},$$

and two similar expressions formed by permuting α, β, γ cyclically, is

$$4\theta^3 - g_2\theta - g_3 = 0,$$

where

$$g_2 = ae - 4bd + 3c^2, \quad g_3 = ace + 2bcd - ad^2 - eb^2 - c^3.$$

It is clear that if two of α, β, γ, δ are equal then two of the roots of this cubic will be equal. Using the result of Ex. 8 we deduce that $g_2^3 - 27g_3^2 = 0$.

10. **Rolle's Theorem for polynomials.** *If $\phi(x)$ is any polynomial, then between any pair of roots of $\phi(x) = 0$ lies a root of $\phi'(x) = 0$.*

A general proof of this theorem, applying not only to polynomials but to other classes of functions, will be given later. The following is an algebraical proof valid for polynomials only. We suppose that α, β are two successive roots, repeated respectively m and n times, so that

$$\phi(x) = (x - \alpha)^m (x - \beta)^n \theta(x),$$

where $\theta(x)$ is a polynomial which has the same sign, say the positive sign, for $\alpha \leqq x \leqq \beta$. Then

$$\begin{aligned}\phi'(x) &= (x-\alpha)^m(x-\beta)^n\theta'(x) + \{m(x-\alpha)^{m-1}(x-\beta)^n + n(x-\alpha)^m(x-\beta)^{n-1}\}\theta(x)\\ &= (x-\alpha)^{m-1}(x-\beta)^{n-1}[(x-\alpha)(x-\beta)\theta'(x) + \{m(x-\beta)+n(x-\alpha)\}\theta(x)]\\ &= (x-\alpha)^{m-1}(x-\beta)^{n-1}F(x),\end{aligned}$$

say. Now $F(\alpha) = m(\alpha - \beta)\theta(\alpha)$ and $F(\beta) = n(\beta - \alpha)\theta(\beta)$, which have opposite signs. Hence $F(x)$, and so $\phi'(x)$, vanishes for some value of x between α and β.

117. B. **Rational Functions.** If

$$R(x) = \frac{P(x)}{Q(x)},$$

where P and Q are polynomials, it follows at once from § 113, (5) that

$$R'(x) = \frac{P'(x)Q(x) - P(x)Q'(x)}{\{Q(x)\}^2},$$

and this formula enables us to write down the derivative of any rational function. The form in which we obtain it, however, may or may not be the simplest possible. It will be the simplest possible if $Q(x)$ and $Q'(x)$ have no common factor, *i.e.* if $Q(x)$ has no repeated factor. But if $Q(x)$ has a repeated factor then the expression which we obtain for $R'(x)$ will be capable of further reduction.

It is very often convenient, in differentiating a rational function, to employ the method of partial fractions. We shall suppose that $Q(x)$, as in § 116, is expressed in the form

$$a_0(x-\alpha_1)^{m_1}(x-\alpha_2)^{m_2}\dots(x-\alpha_\nu)^{m_\nu}.$$

Then it is proved in treatises on Algebra* that $R(x)$ can be expressed in the form

$$\begin{aligned}\Pi(x)+\frac{A_{1,1}}{x-\alpha_1}+\frac{A_{1,2}}{(x-\alpha_1)^2}+\dots+\frac{A_{1,m_1}}{(x-\alpha_1)^{m_1}}\\ +\frac{A_{2,1}}{x-\alpha_2}+\frac{A_{2,2}}{(x-\alpha_2)^2}+\dots+\frac{A_{2,m_2}}{(x-\alpha_2)^{m_2}}+\dots,\end{aligned}$$

where $\Pi(x)$ is a polynomial; *i.e.* as the sum of a polynomial and the sum of a number of terms of the type

$$\frac{A}{(x-\alpha)^p},$$

where α is a root of $Q(x)=0$. We know already how to find the derivative of the polynomial: and it follows at once from Theorem (4) of § 113, or, if α is complex, from its extension indicated in § 114, that the derivative of the rational function last written is

$$-\frac{pA(x-\alpha)^{p-1}}{(x-\alpha)^{2p}}=-\frac{pA}{(x-\alpha)^{p+1}}.$$

We are now able to write down the derivative of the general rational function $R(x)$, in the form

$$\Pi'(x)-\frac{A_{1,1}}{(x-\alpha_1)^2}-\frac{2A_{1,2}}{(x-\alpha_1)^3}-\dots-\frac{A_{2,1}}{(x-\alpha_2)^2}-\frac{2A_{2,2}}{(x-\alpha_2)^3}-\dots.$$

*See, *e.g.*, Chrystal's *Algebra*, vol. i, pp. 151 *et seq.*

Incidentally we have proved that *the derivative of x^m is mx^{m-1}, for all integral values of m positive or negative.*

The method explained in this section is particularly useful when we have to differentiate a rational function several times (see Exs. XLV).

Examples XLII. 1. Prove that

$$\frac{d}{dx}\left(\frac{x}{1+x^2}\right) = \frac{1-x^2}{(1+x^2)^2}, \quad \frac{d}{dx}\left(\frac{1-x^2}{1+x^2}\right) = -\frac{4x}{(1+x^2)^2}.$$

2. Prove that

$$\frac{d}{dx}\left(\frac{ax^2+2bx+c}{Ax^2+2Bx+C}\right) = \frac{(ax+b)(Bx+C)-(bx+c)(Ax+B)}{(Ax^2+2Bx+C)^2}.$$

3. If Q has a factor $(x-\alpha)^m$ then the denominator of R' (when R' is reduced to its lowest terms) is divisible by $(x-\alpha)^{m+1}$ but by no higher power of $x-\alpha$.

4. In no case can the denominator of R' have a *simple* factor $x-\alpha$. Hence no rational function (such as $1/x$) whose denominator contains any simple factor can be the derivative of another rational function.

118. C. **Algebraical Functions.** The results of the preceding sections, together with Theorem (6) of § 113, enable us to obtain the derivative of any explicit algebraical function whatsoever.

The most important such function is x^m, where m is a rational number. We have seen already (§ 117) that the derivative of this function is mx^{m-1} when m is an integer positive or negative; and we shall now prove that this result is true for all rational values of m. Suppose that $y = x^m = x^{p/q}$, where p and q are integers and q positive; and let $z = x^{1/q}$, so that $x = z^q$ and $y = z^p$. Then

$$\frac{dy}{dx} = \left(\frac{dy}{dz}\right)\Big/\left(\frac{dx}{dz}\right) = \frac{p}{q}z^{p-q} = mx^{m-1}.$$

This result may also be deduced as a corollary from Ex. XXXVI. 3. For, if $\phi(x) = x^m$, we have

$$\begin{aligned}\phi'(x) &= \lim_{h\to 0}\frac{(x+h)^m - x^m}{h}\\ &= \lim_{\xi\to x}\frac{\xi^m - x^m}{\xi - x} = mx^{m-1}.\end{aligned}$$

It is clear that the more general formula

$$\frac{d}{dx}(ax+b)^m = ma(ax+b)^{m-1}$$

holds also for all rational values of m.

The differentiation of *implicit* algebraical functions involves certain theoretical difficulties to which we shall return in Ch. VII. But there is no practical difficulty in the actual calculation of the derivative of such a function: the method to be adopted will be illustrated sufficiently by an example. Suppose that y is given by the equation

$$x^3 + y^3 - 3axy = 0.$$

Differentiating with respect to x we find

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

and so

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

Examples XLIII. 1. Find the derivatives of

$$\sqrt{\frac{1+x}{1-x}},\quad \sqrt{\frac{ax+b}{cx+d}},\quad \sqrt{\frac{ax^2+2bx+c}{Ax^2+2Bx+C}},\quad (ax+b)^m(cx+d)^n.$$

2. Prove that

$$\frac{d}{dx}\left\{\frac{x}{\sqrt{a^2+x^2}}\right\} = \frac{a^2}{(a^2+x^2)^{(3/2)}},\quad \frac{d}{dx}\left\{\frac{x}{\sqrt{a^2-x^2}}\right\} = \frac{a^2}{(a^2-x^2)^{3/2}}.$$

3. Find the differential coefficient of y when

(i) $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$, (ii) $x^5 + y^5 - 5ax^2y^2 = 0$.

119. D. **Transcendental Functions.** We have already proved (Ex. XXXIX. 4) that

$$D_x \sin x = \cos x, \quad D_x \cos x = -\sin x.$$

By means of Theorems (4) and (5) of § 113, the reader will easily verify that

$$D_x \tan x = \sec^2 x, \qquad D_x \cot x = -\operatorname{cosec}^2 x,$$
$$D_x \sec x = \tan x \sec x, \quad D_x \operatorname{cosec} x = -\cot x \operatorname{cosec} x.$$

And by means of Theorem (7) we can determine the derivatives of the ordinary inverse trigonometrical functions. The reader should verify the following formulae:

$$D_x \arcsin x = \pm 1/\sqrt{1-x^2}, \qquad D_x \arccos x = \mp 1/\sqrt{1-x^2},$$
$$D_x \arctan x = 1/(1+x^2), \qquad D_x \operatorname{arc\,cot} x = -1/(1+x^2),$$
$$D_x \operatorname{arc\,sec} x = \pm 1/\{x\sqrt{x^2-1}\}, \quad D_x \operatorname{arc\,cosec} x = \mp 1/\{x\sqrt{x^2-1}\}.$$

In the case of the inverse sine and cosecant the ambiguous sign is the same as that of $\cos(\arcsin x)$, in the case of the inverse cosine and secant the same as that of $\sin(\arccos x)$.

The more general formulae

$$D_x \arcsin(x/a) = \pm 1/\sqrt{a^2-x^2}, \quad D_x \arctan(x/a) = a/(x^2+a^2),$$

which are also easily derived from Theorem (7) of § 113, are also of considerable importance. In the first of them the ambiguous sign is the same as that of $a\cos\{\arcsin(x/a)\}$, since

$$a\sqrt{1-(x^2/a^2)} = \pm\sqrt{a^2-x^2}$$

according as a is positive or negative.

Finally, by means of Theorem (6) of § 113, we are enabled to differentiate composite functions involving symbols both of algebraical and trigonometrical functionality, and so to write down the derivative of any such function as occurs in the following examples.

Examples XLIV.* 1. Find the derivatives of

$$\cos^m x,\quad \sin^m x,\quad \cos x^m,\quad \sin x^m,\quad \cos(\sin x),\quad \sin(\cos x),$$
$$\sqrt{a^2\cos^2 x + b^2\sin^2 x},\quad \frac{\cos x \sin x}{\sqrt{a^2\cos^2 x + b^2\sin^2 x}},$$
$$x\arcsin x + \sqrt{1-x^2},\quad (1+x)\arctan\sqrt{x} - \sqrt{x}.$$

2. Verify by differentiation that $\arcsin x + \arccos x$ is constant for all values of x between 0 and 1, and $\arctan x + \operatorname{arc\,cot} x$ for all positive values of x.

3. Find the derivatives of

$$\arcsin\sqrt{1-x^2},\quad \arcsin\{2x\sqrt{1-x^2}\},\quad \arctan\left(\frac{a+x}{1-ax}\right).$$

How do you explain the simplicity of the results?

4. Differentiate

$$\frac{1}{\sqrt{ac-b^2}}\arctan\frac{ax+b}{\sqrt{ac-b^2}},\qquad -\frac{1}{\sqrt{-a}}\arcsin\frac{ax+b}{\sqrt{b^2-ac}}.$$

5. Show that each of the functions

$$2\arcsin\sqrt{\frac{x-\beta}{\alpha-\beta}},\quad 2\arctan\sqrt{\frac{x-\beta}{\alpha-x}},\quad \arcsin\frac{2\sqrt{(\alpha-x)(x-\beta)}}{\alpha-\beta}$$

has the derivative

$$\frac{1}{\sqrt{(\alpha-x)(x-\beta)}}.$$

*In these examples m is a rational number and a, b, …, α, β … have such values that the functions which involve them are real.

6. Prove that

$$\frac{d}{d\theta}\left\{\arccos\sqrt{\frac{\cos 3\theta}{\cos^3\theta}}\right\} = \sqrt{\frac{3}{\cos\theta\cos 3\theta}}.$$

(*Math. Trip.* 1904.)

7. Show that

$$\frac{1}{\sqrt{C(Ac-aC)}}\frac{d}{dx}\left[\arccos\sqrt{\frac{C(ax^2+c)}{c(Ax^2+C)}}\right] = \frac{1}{(Ax^2+C)\sqrt{ax^2+c}}.$$

8. Each of the functions

$$\frac{1}{\sqrt{a^2-b^2}}\arccos\left(\frac{a\cos x+b}{a+b\cos x}\right),\quad \frac{2}{\sqrt{a^2-b^2}}\arctan\left\{\sqrt{\frac{a-b}{a+b}}\tan\tfrac{1}{2}x\right\}$$

has the derivative $1/(a+b\cos x)$.

9. If $X = a + b\cos x + c\sin x$, and

$$y = \frac{1}{\sqrt{a^2-b^2-c^2}}\arccos\frac{aX-a^2+b^2+c^2}{X\sqrt{b^2+c^2}},$$

then $dy/dx = 1/X$.

10. Prove that the derivative of $F[f\{\phi(x)\}]$ is $F'[f\{\phi(x)\}]\,f'\{\phi(x)\}\phi'(x)$, and extend the result to still more complicated cases.

11. If u and v are functions of x, then

$$D_x \arctan(u/v) = (vD_x u - uD_x v)/(u^2+v^2).$$

12. The derivative of $y = (\tan x + \sec x)^m$ is $my\sec x$.

13. The derivative of $y = \cos x + i\sin x$ is iy.

14. Differentiate $x\cos x$, $(\sin x)/x$. Show that the values of x for which the tangents to the curves $y = x\cos x$, $y = (\sin x)/x$ are parallel to the axis of x are roots of $\cot x = x$, $\tan x = x$ respectively.

15. It is easy to see (cf. Ex. XVII. 5) that the equation $\sin x = ax$, where a is positive, has no real roots except $x = 0$ if $a \geqq 1$, and if $a < 1$ a finite number of roots which increases as a diminishes. Prove that the values of a for which the

number of roots changes are the values of $\cos\xi$, where ξ is a positive root of the equation $\tan\xi = \xi$. [The values required are the values of a for which $y = ax$ touches $y = \sin x$.]

16. If $\phi(x) = x^2 \sin(1/x)$ when $x \neq 0$, and $\phi(0) = 0$, then

$$\phi'(x) = 2x\sin(1/x) - \cos(1/x)$$

when $x \neq 0$, and $\phi'(0) = 0$. And $\phi'(x)$ is discontinuous for $x = 0$ (cf. § 111, (2)).

17. Find the equations of the tangent and normal at the point (x_0, y_0) of the circle $x^2 + y^2 = a^2$.

[Here $y = \sqrt{a^2 - x^2}$, $dy/dx = -x/\sqrt{a^2 - x^2}$, and the tangent is

$$y - y_0 = (x - x_0)\left\{-x_0/\sqrt{a^2 - x_0^2}\right\},$$

which may be reduced to the form $xx_0 + yy_0 = a^2$. The normal is $xy_0 - yx_0 = 0$, which of course passes through the origin.]

18. Find the equations of the tangent and normal at any point of the ellipse $(x/a)^2 + (y/b)^2 = 1$ and the hyperbola $(x/a)^2 - (y/b)^2 = 1$.

19. The equations of the tangent and normal to the curve $x = \phi(t)$, $y = \psi(t)$, at the point whose parameter is t, are

$$\frac{x - \phi(t)}{\phi'(t)} = \frac{y - \psi(t)}{\psi'(t)}, \quad \{x - \phi(t)\}\phi'(t) + \{y - \psi(t)\}\psi'(t) = 0.$$

120. Repeated differentiation. We may form a new function $\phi''(x)$ from $\phi'(x)$ just as we formed $\phi'(x)$ from $\phi(x)$. This function is called the *second derivative* or *second differential coefficient* of $\phi(x)$. The second derivative of $y = \phi(x)$ may also be written in any of the forms

$$D_x^2 y, \quad \left(\frac{d}{dx}\right)^2 y, \quad \frac{d^2y}{dx^2}.$$

In exactly the same way we may define the *nth derivative or nth differential coefficient of* $y = \phi(x)$, which may be written in any of the forms

$$\phi^{(n)}(x), \quad D_x^n y, \quad \left(\frac{d}{dx}\right)^n y, \quad \frac{d^n y}{dx^n}.$$

But it is only in a few cases that it is easy to write down a general formula for the nth differential coefficient of a given function. Some of these cases will be found in the examples which follow.

Examples XLV. 1. If $\phi(x) = x^m$ then

$$\phi^{(n)}(x) = m(m-1)\dots(m-n+1)x^{m-n}.$$

This result enables us to write down the nth derivative of any polynomial.

2. If $\phi(x) = (ax+b)^m$ then

$$\phi^{(n)}(x) = m(m-1)\dots(m-n+1)a^n(ax+b)^{m-n}.$$

In these two examples m may have any rational value. If m is a positive integer, and $n > m$, then $\phi^{(n)}(x) = 0$.

3. The formula

$$\left(\frac{d}{dx}\right)^n \frac{A}{(x-\alpha)^p} = (-1)^n \frac{p(p+1)\dots(p+n-1)A}{(x-\alpha)^{p+n}}$$

enables us to write down the nth derivative of any rational function expressed in the standard form as a sum of partial fractions.

4. Prove that the nth derivative of $1/(1-x^2)$ is

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

5. **Leibniz' Theorem.** If y is a product uv, and we can form the first n derivatives of u and v, then we can form the nth derivative of y by means of *Leibniz' Theorem*, which gives the rule

$$(uv)_n = u_n v + \binom{n}{1} u_{n-1}v_1 + \binom{n}{2} u_{n-2}v_2 + \dots + \binom{n}{r} u_{n-r}v_r + \dots + uv_n,$$

where suffixes indicate differentiations, so that u_n, for example, denotes the nth derivative of u. To prove the theorem we observe that

$$(uv)_1 = u_1 v + uv_1,$$
$$(uv)_2 = u_2 v + 2u_1 v_1 + uv_2,$$

and so on. It is obvious that by repeating this process we arrive at a formula of the type

$$(uv)_n = u_n v + a_{n,1} u_{n-1} v_1 + a_{n,2} u_{n-2} v_2 + \cdots + a_{n,r} u_{n-r} v_r + \cdots + u v_n.$$

Let us assume that $a_{n,r} = \binom{n}{r}$ for $r = 1, 2, \ldots, n-1$, and show that if this is so then $a_{n+1,r} = \binom{n+1}{r}$ for $r = 1, 2, \ldots n$. It will then follow by the principle of mathematical induction that $a_{n,r} = \binom{n}{r}$ for all values of n and r in question.

When we form $(uv)_{n+1}$ by differentiating $(uv)_n$ it is clear that the coefficient of $u_{n+1-r} v_r$ is

$$a_{n,r} + a_{n,r-1} = \binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}.$$

This establishes the theorem.

6. The nth derivative of $x^m f(x)$ is

$$\frac{m!}{(m-n)!} x^{m-n} f(x) + n \frac{m!}{(m-n+1)!} x^{m-n+1} f'(x) \\ + \frac{n(n-1)}{1 \cdot 2} \frac{m!}{(m-n+2)!} x^{m-n+2} f''(x) + \ldots,$$

the series being continued for $n+1$ terms or until it terminates.

7. Prove that $D_x^n \cos x = \cos(x + \frac{1}{2} n\pi)$, $D_x^n \sin x = \sin(x + \frac{1}{2} n\pi)$.

8. If $y = A \cos mx + B \sin mx$ then $D_x^2 y + m^2 y = 0$. And if

$$y = A \cos mx + B \sin mx + P_n(x),$$

where $P_n(x)$ is a polynomial of degree n, then $D_x^{n+3} y + m^2 D_x^{n+1} y = 0$.

9. If $x^2 D_x^2 y + x D_x y + y = 0$ then

$$x^2 D_x^{n+2} y + (2n+1) x D_x^{n+1} y + (n^2+1) D_x^n y = 0.$$

[Differentiate n times by Leibniz' Theorem.]

10. If U_n denotes the nth derivative of $(Lx + M)/(x^2 - 2Bx + C)$, then

$$\frac{x^2 - 2Bx + C}{(n+1)(n+2)} U_{n+2} + \frac{2(x-B)}{n+1} U_{n+1} + U_n = 0.$$

(*Math. Trip.* 1900.)

[First obtain the equation when $n = 0$; then differentiate n times by Leibniz' Theorem.]

11. **The nth derivatives of $a/(a^2 + x^2)$ and $x/(a^2 + x^2)$.** Since

$$\frac{a}{a^2 + x^2} = \frac{1}{2i}\left(\frac{1}{x - ai} - \frac{1}{x + ai}\right), \quad \frac{x}{a^2 + x^2} = \frac{1}{2}\left(\frac{1}{x - ai} + \frac{1}{x + ai}\right),$$

we have

$$D_x^n\left(\frac{a}{a^2 + x^2}\right) = \frac{(-1)^n n!}{2i}\left\{\frac{1}{(x - ai)^{n+1}} - \frac{1}{(x + ai)^{n+1}}\right\},$$

and a similar formula for $D_x^n\{x/(a^2 + x^2)\}$. If $\rho = \sqrt{x^2 + a^2}$, and θ is the numerically smallest angle whose cosine and sine are x/ρ and a/ρ, then $x + ai = \rho \operatorname{Cis} \theta$ and $x - ai = \rho \operatorname{Cis}(-\theta)$, and so

$$\begin{aligned} D_x^n\{a/(a^2 + x^2)\} &= \{(-1)^n n!/2i\}\rho^{-n-1}[\operatorname{Cis}\{(n+1)\theta\} - \operatorname{Cis}\{-(n+1)\theta\}] \\ &= (-1)^n n!\,(x^2 + a^2)^{-(n+1)/2} \sin\{(n+1)\arctan(a/x)\}. \end{aligned}$$

Similarly

$$D_x^n\{x/(a^2 + x^2)\} = (-1)^n n!\,(x^2 + a^2)^{-(n+1)/2} \cos\{(n+1)\arctan(a/x)\}.$$

12. Prove that

$$\begin{aligned} D_x^n\{(\cos x)/x\} &= \{P_n \cos(x + \tfrac{1}{2}n\pi) + Q_n \sin(x + \tfrac{1}{2}n\pi)\}/x^{n+1}, \\ D_x^n\{(\sin x)/x\} &= \{P_n \sin(x + \tfrac{1}{2}n\pi) - Q_n \cos(x + \tfrac{1}{2}n\pi)\}/x^{n+1}, \end{aligned}$$

where P_n and Q_n are polynomials in x of degree n and $n - 1$ respectively.

13. Establish the formulae

$$\frac{dx}{dy} = 1\bigg/\left(\frac{dy}{dx}\right), \quad \frac{d^2x}{dy^2} = -\frac{d^2y}{dx^2}\bigg/\left(\frac{dy}{dx}\right)^3,$$

$$\frac{d^3x}{dy^3} = -\left\{\frac{d^3y}{dx^3}\frac{dy}{dx} - 3\left(\frac{d^2y}{dx^2}\right)\right\}\bigg/\left(\frac{dy}{dx}\right)^5.$$

14. If $yz = 1$ and $y_r = (1/r!)D_x^r y$, $z_s = (1/s!)D_x^s z$, then

$$\frac{1}{z^3}\begin{vmatrix} z & z_1 & z_2 \\ z_1 & z_2 & z_3 \\ z_2 & z_3 & z_4 \end{vmatrix} = \frac{1}{y^2}\begin{vmatrix} y_2 & y_3 \\ y_3 & y_4 \end{vmatrix}.$$

(*Math. Trip.* 1905.)

15. If

$$W(y, z, u) = \begin{vmatrix} y & z & u \\ y' & z' & u' \\ y'' & z'' & u'' \end{vmatrix},$$

dashes denoting differentiations with respect to x, then

$$W(y, z, u) = y^3\, W\left(1, \frac{z}{y}, \frac{u}{y}\right).$$

16. If

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0,$$

then

$$dy/dx = -(ax + hy + g)/(hx + by + f)$$

and

$$d^2y/dx^2 = (abc + 2fgh - af^2 - bg^2 - ch^2)/(hx + by + f)^3.$$

121. Some general theorems concerning derived functions. In all that follows we suppose that $\phi(x)$ is a function of x which has a derivative $\phi'(x)$ for all values of x in question. This assumption of course involves the continuity of $\phi(x)$.

The meaning of the sign of $\phi'(x)$. THEOREM A. *If $\phi'(x_0) > 0$ then $\phi(x) < \phi(x_0)$ for all values of x less than x_0 but sufficiently near to x_0, and $\phi(x) > \phi(x_0)$ for all values of x greater than x_0 but sufficiently near to x_0.*

For $\{\phi(x_0+h) - \phi(x_0)\}/h$ converges to a positive limit $\phi'(x_0)$ as $h \to 0$. This can only be the case if $\phi(x_0 + h) - \phi(x_0)$ and h have the same sign for sufficiently small values of h, and this is precisely what the theorem states. Of course from a geometrical point of view the result is intuitive, the inequality $\phi'(x) > 0$ expressing the fact that the tangent to the curve $y = \phi(x)$ makes a positive acute angle with the axis of x. The reader should formulate for himself the corresponding theorem for the case in which $\phi'(x) < 0$.

An immediate deduction from Theorem A is the following important theorem, generally known as Rolle's Theorem. In view of the great importance of this theorem it may be well to repeat that its truth depends on the assumption of the existence of the derivative $\phi'(x)$ for all values of x in question.

Theorem B. *If $\phi(a) = 0$ and $\phi(b) = 0$, then there must be at least one value of x which lies between a and b and for which $\phi'(x) = 0$.*

There are two possibilities: the first is that $\phi(x)$ is equal to zero throughout the whole interval $[a, b]$. In this case $\phi'(x)$ is also equal to zero throughout the interval. If on the other hand $\phi(x)$ is not always equal to zero, then there must be values of x for which $\phi(x)$ is positive or negative. Let us suppose, for example, that $\phi(x)$ is sometimes positive. Then, by Theorem 2 of § 102, there is a value ξ of x, not equal to a or b, and such that $\phi(\xi)$ is at least as great as the value of $\phi(x)$ at any other point in the interval. And $\phi'(\xi)$ must be equal to zero. For if it were positive then $\phi(x)$ would, by Theorem A, be greater than $\phi(\xi)$ for values of x greater than ξ but sufficiently near to ξ, so that there would certainly be values of $\phi(x)$ greater than $\phi(\xi)$. Similarly we can show that $\phi'(\xi)$ cannot be negative.

Cor 1. *If $\phi(a) = \phi(b) = k$, then there must be a value of x between a and b such that $\phi'(x) = 0$.*

We have only to put $\phi(x) - k = \psi(x)$ and apply Theorem B to $\psi(x)$.

Cor 2. *If $\phi'(x) > 0$ for all values of x in a certain interval, then $\phi(x)$ is an increasing function of x, in the stricter sense of § 95, throughout that interval.*

Let x_1 and x_2 be two values of x in the interval in question, and $x_1 < x_2$. We have to show that $\phi(x_1) < \phi(x_2)$. In the first place $\phi(x_1)$ cannot be equal to $\phi(x_2)$; for, if this were so, there would, by Theorem B, be a value of x between x_1 and x_2 for which $\phi'(x) = 0$. Nor can $\phi(x_1)$ be greater than $\phi(x_2)$. For, since $\phi'(x_1)$ is positive, $\phi(x)$ is, by Theorem A, greater than $\phi(x_1)$ when x is greater than x_1 and sufficiently near to x_1. It follows that there is a value x_3 of x between x_1 and x_2 such that $\phi(x_3) = \phi(x_1)$; and so, by Theorem B, that there is a value of x between x_1 and x_3 for which $\phi'(x) = 0$.

Cor 3. *The conclusion of Cor. 2 still holds if the interval $[a, b]$ considered includes a finite number of exceptional values of x for which $\phi'(x)$ does not exist, or is not positive, provided $\phi(x)$ is continuous even for these exceptional values of x.*

It is plainly sufficient to consider the case in which there is one exceptional value of x only, and that corresponding to an end of the interval, say to a. If $a < x_1 < x_2 < b$, we can choose $a + \delta$ so that $a + \delta < x_1$, and $\phi'(x) > 0$ throughout $[a + \delta, b]$, so that $\phi(x_1) < \phi(x_2)$, by Cor. 2. All that remains is to prove that $\phi(a) < \phi(x_1)$. Now $\phi(x_1)$ decreases steadily, and in the stricter sense, as x_1 decreases towards a, and so

$$\phi(a) = \phi(a + 0) = \lim_{x_1 \to a+0} \phi(x_1) < \phi(x_1).$$

Cor 4. *If $\phi'(x) > 0$ throughout the interval $[a, b]$, and $\phi(a) \geqq 0$, then $\phi(x)$ is positive throughout the interval $[a, b]$.*

The reader should compare the second of these corollaries very carefully with Theorem A. If, as in Theorem A, we assume only that $\phi'(x)$ is positive *at a single point $x = x_0$*, then we can prove that $\phi(x_1) < \phi(x_2)$ when x_1 and x_2 are sufficiently near to x_0 and $x_1 < x_0 < x_2$. For $\phi(x_1) < \phi(x_0)$ and $\phi(x_2) > \phi(x_0)$, by Theorem A. But this does not prove that there is any interval including x_0 throughout which $\phi(x)$ is a steadily increasing function, for the assumption that x_1 and x_2 lie on opposite sides of x_0 is essential to our conclusion. We shall return to this point, and illustrate it by an actual example, in a moment (§ 124).

122. Maxima and Minima. We shall say that the value $\phi(\xi)$ assumed by $\phi(x)$ when $x = \xi$ is a *maximum* if $\phi(\xi)$ is greater than any other value assumed by $\phi(x)$ in the immediate neighbourhood of $x = \xi$, *i.e.* if we can find an interval $[\xi - \delta, \xi + \delta]$ of values of x such that $\phi(\xi) > \phi(x)$

when $\xi - \delta < x < \xi$ and when $\xi < x < \xi + \delta$; and we define a *minimum* in a similar manner. Thus in the figure the points A correspond to maxima, the points B to minima of the function whose graph is there shown. It is

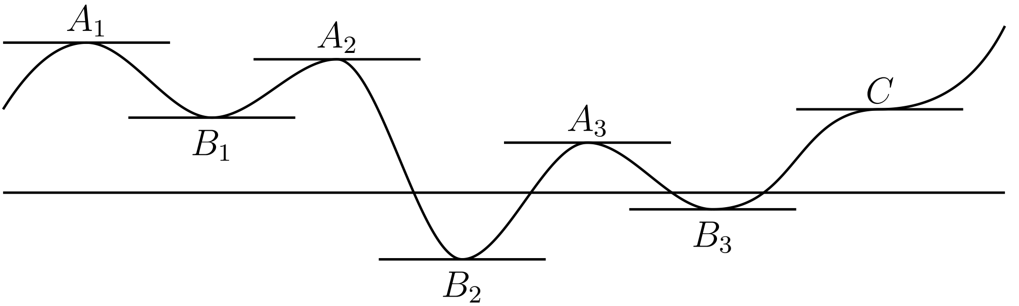

Fig. 39.

to be observed that the fact that A_3 corresponds to a maximum and B_1 to a minimum is in no way inconsistent with the fact that the value of the function is greater at B_1 than at A_3.

THEOREM C. *A* **necessary** *condition for a maximum or minimum value of $\phi(x)$ at $x = \xi$ is that $\phi'(\xi) = 0$.**

This follows at once from Theorem A. That the condition is not *sufficient* is evident from a glance at the point C in the figure. Thus if $y = x^3$ then $\phi'(x) = 3x^2$, which vanishes when $x = 0$. But $x = 0$ does not give either a maximum or a minimum of x^3, as is obvious from the form of the graph of x^3 (Fig. 10, p. 51).

But *there will certainly be a maximum at $x = \xi$ if $\phi'(\xi) = 0$, $\phi'(x) > 0$ for all values of x less than but near to ξ, and $\phi'(x) < 0$ for all values of x greater than but near to ξ*: and if the signs of these two inequalities are reversed there will certainly be a minimum. For then we can (by Cor. 3 of § 121) determine an interval $[\xi - \delta, \xi]$ throughout which $\phi(x)$ increases with x, and an interval $[\xi, \xi + \delta]$ throughout which it decreases as x increases: and obviously this ensures that $\phi(\xi)$ shall be a maximum.

---

*A function which is continuous but has no derivative may have maxima and minima. We are of course assuming the existence of the derivative.

This result may also be stated thus. If the sign of $\phi'(x)$ changes at $x = \xi$ from positive to negative, then $x = \xi$ gives a maximum of $\phi(x)$: and if the sign of $\phi'(x)$ changes in the opposite sense, then $x = \xi$ gives a minimum.

123. There is another way of stating the conditions for a maximum or minimum which is often useful. Let us assume that $\phi(x)$ has a second derivative $\phi''(x)$: this of course does not follow from the existence of $\phi'(x)$, any more than the existence of $\phi'(x)$ follows from that of $\phi(x)$. But in such cases as we are likely to meet with at present the condition is generally satisfied.

THEOREM D. *If $\phi'(\xi) = 0$ and $\phi''(\xi) \neq 0$, then $\phi(x)$ has a maximum or minimum at $x = \xi$, a maximum if $\phi''(\xi) < 0$, a minimum if $\phi''(\xi) > 0$.*

Suppose, *e.g.*, that $\phi''(\xi) < 0$. Then, by Theorem A, $\phi'(x)$ is negative when x is less than ξ but sufficiently near to ξ, and positive when x is greater than ξ but sufficiently near to ξ. Thus $x = \xi$ gives a maximum.

124. In what has preceded (apart from the last paragraph) we have assumed simply that $\phi(x)$ has a derivative for all values of x in the interval under consideration. If this condition is not fulfilled the theorems cease to be true. Thus Theorem B fails in the case of the function

$$y = 1 - \sqrt{x^2},$$

where the square root is to be taken positive. The graph of this function is shown in Fig. 40. Here $\phi(-1) = \phi(1) = 0$: but $\phi'(x)$, as is evident from the figure, is equal to 1 if x is negative and to -1 if x is positive, and never vanishes. There is no derivative for $x = 0$, and no tangent to the graph at P. And in this case $x = 0$ obviously gives a maximum of $\phi(x)$, but $\phi'(0)$, as it does not exist, cannot be equal to zero, so that the test for a maximum fails.

The bare existence of the derivative $\phi'(x)$, however, is all that we have assumed. And there is one assumption in particular that we have not made, and that is that $\phi'(x)$ *itself is a continuous function*. This raises a rather subtle but still a very interesting point. *Can* a function $\phi(x)$ have a derivative for all values of x which is not itself continuous? In other words can a curve have a tangent at every point, and yet the direction of the tangent not vary continuously? The

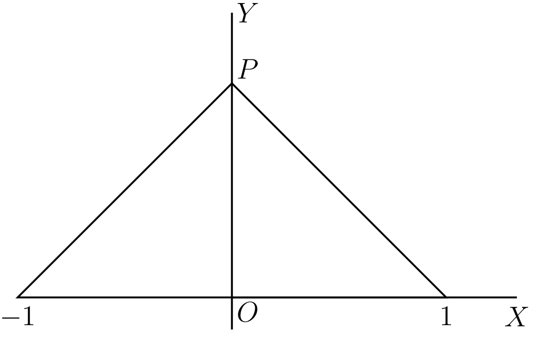

Fig. 40.

reader, if he considers what the question means and tries to answer it in the light of common sense, will probably incline to the answer *No.* It is, however, not difficult to see that this answer is wrong.

Consider the function $\phi(x)$ defined, when $x \neq 0$, by the equation

$$\phi(x) = x^2 \sin(1/x);$$

and suppose that $\phi(0) = 0$. Then $\phi(x)$ is continuous for all values of x. If $x \neq 0$ then

$$\phi'(x) = 2x \sin(1/x) - \cos(1/x);$$

while

$$\phi'(0) = \lim_{h \to 0} \frac{h^2 \sin(1/h)}{h} = 0.$$

Thus $\phi'(x)$ exists for all values of x. But $\phi'(x)$ is discontinuous for $x = 0$; for $2x \sin(1/x)$ tends to 0 as $x \to 0$, and $\cos(1/x)$ oscillates between the limits of indetermination -1 and 1, so that $\phi'(x)$ oscillates between the same limits.

What is practically the same example enables us also to illustrate the point referred to at the end of § 121. Let

$$\phi(x) = x^2 \sin(1/x) + ax,$$

where $0 < a < 1$, when $x \neq 0$, and $\phi(0) = 0$. Then $\phi'(0) = a > 0$. Thus the conditions of Theorem A of § 121 are satisfied. But if $x \neq 0$ then

$$\phi'(x) = 2x \sin(1/x) - \cos(1/x) + a,$$

which oscillates between the limits of indetermination $a-1$ and $a+1$ as $x \to 0$. As $a-1 < 0$, we can find values of x, as near to 0 as we like, for which $\phi'(x) < 0$; and it is therefore impossible to find any interval, including $x = 0$, throughout which $\phi(x)$ is a steadily increasing function of x.

It is, however, impossible that $\phi'(x)$ should have what was called in Ch. V (Ex. XXXVII. 18) a 'simple' discontinuity; *e.g.* that $\phi'(x) \to a$ when $x \to +0$, $\phi'(x) \to b$ when $x \to -0$, and $\phi'(0) = c$, unless $a = b = c$, in which case $\phi'(x)$ is continuous for $x = 0$. For a proof see § 125, Ex. XLVII. 3.

Examples XLVI. 1. Verify Theorem B when $\phi(x) = (x-a)^m(x-b)^n$ or $\phi(x) = (x-a)^m(x-b)^n(x-c)^p$, where m, n, p are positive integers and $a < b < c$.

[The first function vanishes for $x = a$ and $x = b$. And

$$\phi'(x) = (x-a)^{m-1}(x-b)^{n-1}\{(m+n)x - mb - na\}$$

vanishes for $x = (mb+na)/(m+n)$, which lies between a and b. In the second case we have to verify that the quadratic equation

$$(m+n+p)x^2 - \{m(b+c) + n(c+a) + p(a+b)\}x + mbc + nca + pab = 0$$

has roots between a and b and between b and c.]

2. Show that the polynomials

$$2x^3 + 3x^2 - 12x + 7, \quad 3x^4 + 8x^3 - 6x^2 - 24x + 19$$

are positive when $x > 1$.

3. Show that $x - \sin x$ is an increasing function throughout any interval of values of x, and that $\tan x - x$ increases as x increases from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$. For what values of a is $ax - \sin x$ a steadily increasing or decreasing function of x?

4. Show that $\tan x - x$ also increases from $x = \frac{1}{2}\pi$ to $x = \frac{3}{2}\pi$, from $x = \frac{3}{2}\pi$ to $x = \frac{5}{2}\pi$, and so on, and deduce that there is one and only one root of the equation $\tan x = x$ in each of these intervals (cf. Ex. XVII. 4).

5. Deduce from Ex. 3 that $\sin x - x < 0$ if $x > 0$, from this that $\cos x - 1 + \frac{1}{2}x^2 > 0$, and from this that $\sin x - x + \frac{1}{6}x^3 > 0$. And, generally,

prove that if

$$C_{2m} = \cos x - 1 + \frac{x^2}{2!} - \cdots - (-1)^m \frac{x^{2m}}{(2m)!},$$
$$S_{2m+1} = \sin x - x + \frac{x^3}{3!} - \cdots - (-1)^m \frac{x^{2m+1}}{(2m+1)!},$$

and $x > 0$, then C_{2m} and S_{2m+1} are positive or negative according as m is odd or even.

6. If $f(x)$ and $f''(x)$ are continuous and have the same sign at every point of an interval $[a, b]$, then this interval can include at most one root of either of the equations $f(x) = 0$, $f'(x) = 0$.

7. The functions u, v and their derivatives u', v' are continuous throughout a certain interval of values of x, and $uv' - u'v$ never vanishes at any point of the interval. Show that between any two roots of $u = 0$ lies one of $v = 0$, and conversely. Verify the theorem when $u = \cos x$, $v = \sin x$.

[If v does not vanish between two roots of $u = 0$, say α and β, then the function u/v is continuous throughout the interval $[\alpha, \beta]$ and vanishes at its extremities. Hence $(u/v)' = (u'v - uv')/v^2$ must vanish between α and β, which contradicts our hypothesis.]

8. Determine the maxima and minima (if any) of $(x-1)^2(x+2)$, $x^3 - 3x$, $2x^3 - 3x^2 - 36x + 10$, $4x^3 - 18x^2 + 27x - 7$, $3x^4 - 4x^3 + 1$, $x^5 - 15x^3 + 3$. In each case sketch the form of the graph of the function.

[Consider the last function, for example. Here $\phi'(x) = 5x^2(x^2 - 9)$, which vanishes for $x = -3$, $x = 0$, and $x = 3$. It is easy to see that $x = -3$ gives a maximum and $x = 3$ a minimum, while $x = 0$ gives neither, as $\phi'(x)$ is negative on both sides of $x = 0$.]

9. Discuss the maxima and minima of the function $(x-a)^m(x-b)^n$, where m and n are any positive integers, considering the different cases which occur according as m and n are odd or even. Sketch the graph of the function.

10. Discuss similarly the function $(x-a)(x-b)^2(x-c)^3$, distinguishing the different forms of the graph which correspond to different hypotheses as to the relative magnitudes of a, b, c.

11. Show that $(ax+b)/(cx+d)$ has no maxima or minima, whatever values a, b, c, d may have. Draw a graph of the function.

12. Discuss the maxima and minima of the function

$$y = (ax^2 + 2bx + c)/(Ax^2 + 2Bx + C),$$

when the denominator has complex roots.

[We may suppose a and A positive. The derivative vanishes if

$$(ax + b)(Bx + C) - (Ax + B)(bx + c) = 0. \tag{1}$$

This equation must have real roots. For if not the derivative would always have the same sign, and this is impossible, since y is continuous for all values of x, and $y \to a/A$ as $x \to +\infty$ or $x \to -\infty$. It is easy to verify that the curve cuts the line $y = a/A$ in one and only one point, and that it lies above this line for large positive values of x, and below it for large negative values, or *vice versa*, according as $b/a > B/A$ or $b/a < B/A$. Thus the algebraically greater root of (1) gives a maximum if $b/a > B/A$, a minimum in the contrary case.]

13. The maximum and minimum values themselves are the values of λ for which $ax^2+2bx+c-\lambda(Ax^2+2Bx+C)$ is a perfect square. [This is the condition that $y = \lambda$ should touch the curve.]

14. In general the maxima and maxima of $R(x) = P(x)/Q(x)$ are among the values of λ obtained by expressing the condition that $P(x) - \lambda Q(x) = 0$ should have a pair of equal roots.

15. If $Ax^2 + 2Bx + C = 0$ has real roots then it is convenient to proceed as follows. We have

$$y - (a/A) = (\lambda x + \mu)/\{A(Ax^2 + 2Bx + C)\},$$

where $\lambda = bA - aB$, $\mu = cA - aC$. Writing further ξ for $\lambda x + \mu$ and η for $(A/\lambda^2)(Ay - a)$, we obtain an equation of the form

$$\eta = \xi/\{(\xi - p)(\xi - q)\}.$$

This transformation from (x, y) to (ξ, η) amounts only to a shifting of the origin, keeping the axes parallel to themselves, a change of scale along each axis, and (if $\lambda < 0$) a reversal in direction of the axis of abscissae; and so a minimum of y, considered as a function of x, corresponds to a minimum of η considered as a function of ξ, and *vice versa*, and similarly for a maximum.

The derivative of η with respect to ξ vanishes if

$$(\xi - p)(\xi - q) - \xi(\xi - p) - \xi(\xi - q) = 0,$$

or if $\xi^2 = pq$. Thus there are two roots of the derivative if p and q have the same sign, none if they have opposite signs. In the latter case the form of the graph of η is as shown in Fig. 41a.

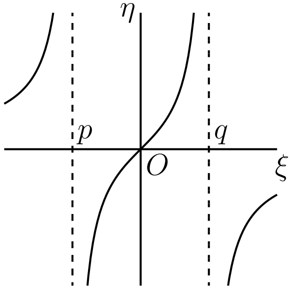

Fig. 41a.

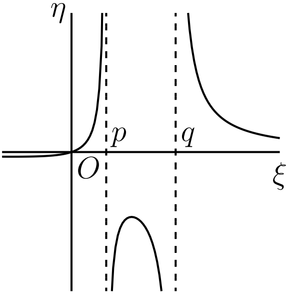

Fig. 41b.

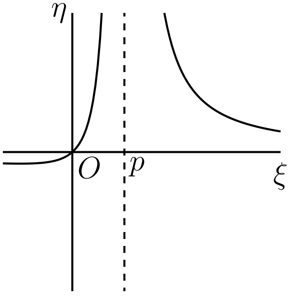

Fig. 41c.

When p and q are positive the general form of the graph is as shown in Fig 41b, and it is easy to see that $\xi = \sqrt{pq}$ gives a maximum and $\xi = -\sqrt{pq}$ a minimum.*

In the particular case in which $p = q$ the function is

$$\eta = \xi/(\xi - p)^2,$$

and its graph is of the form shown in Fig. 41c.

The preceding discussion fails if $\lambda = 0$, *i.e.* if $a/A = b/B$. But in this case we have

$$\begin{aligned} y - (a/A) &= \mu/\{A(Ax^2 + 2Bx + C)\} \\ &= \mu/\{A^2(x - x_1)(x - x_2)\}, \end{aligned}$$

say, and $dy/dx = 0$ gives the single value $x = \frac{1}{2}(x_1 + x_2)$. On drawing a graph it becomes clear that this value gives a maximum or minimum according as μ is positive or negative. The graph shown in Fig. 42 corresponds to the former case.

*The maximum is $-1/(\sqrt{p} - \sqrt{q})^2$, the minimum $-1/(\sqrt{p} + \sqrt{q})^2$, of which the latter is the greater.

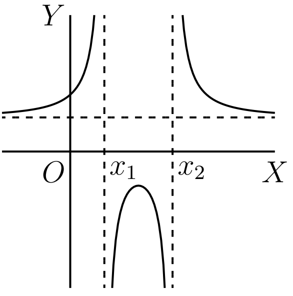

Fig. 42.

[A full discussion of the general function $y = (ax^2+2bx+c)/(Ax^2+2Bx+C)$, by purely algebraical methods, will be found in Chrystal's *Algebra*, vol i, pp. 464–7.]

16. Show that $(x-\alpha)(x-\beta)/(x-\gamma)$ assumes all real values as x varies, if γ lies between α and β, and otherwise assumes all values except those included in an interval of length $4\sqrt{|\alpha-\gamma||\beta-\gamma|}$.

17. Show that

$$y = \frac{x^2+2x+c}{x^2+4x+3c}$$

can assume any real value if $0 < c < 1$, and draw a graph of the function in this case. (*Math. Trip.* 1910.)

18. Determine the function of the form $(ax^2+2bx+c)/(Ax^2+2Bx+C)$ which has turning values (*i.e.* maxima or minima) 2 and 3 when $x = 1$ and $x = -1$ respectively, and has the value 2.5 when $x = 0$. (*Math. Trip.* 1908.)

19. The maximum and minimum of $(x+a)(x+b)/(x-a)(x-b)$, where a and b are positive, are

$$-\left(\frac{\sqrt{a}+\sqrt{b}}{\sqrt{a}-\sqrt{b}}\right)^2, \quad -\left(\frac{\sqrt{a}-\sqrt{b}}{\sqrt{a}+\sqrt{b}}\right)^2.$$

20. The maximum value of $(x-1)^2/(x+1)^3$ is $\frac{2}{27}$.

21. Discuss the maxima and minima of

$$x(x-1)/(x^2+3x+3), \quad x^4/(x-1)(x-3)^3,$$
$$(x-1)^2(3x^2-2x-37)/(x+5)^2(3x^2-14x-1).$$

(*Math. Trip.* 1898.)

[If the last function be denoted by $P(x)/Q(x)$, it will be found that

$$P'Q - PQ' = 72(x-7)(x-3)(x-1)(x+1)(x+2)(x+5).]$$

22. Find the maxima and minima of $a\cos x + b\sin x$. Verify the result by expressing the function in the form $A\cos(x-a)$.

23. Find the maxima and minima of

$$a^2\cos^2 x + b^2\sin^2 x, \quad A\cos^2 x + 2H\cos x\sin x + B\sin^2 x.$$

24. Show that $\sin(x+a)/\sin(x+b)$ has no maxima or minima. Draw a graph of the function.

25. Show that the function

$$\frac{\sin^2 x}{\sin(x+a)\sin(x+b)} \quad (0 < a < b < \pi)$$

has an infinity of minima equal to 0 and of maxima equal to

$$-4\sin a\sin b/\sin^2(a-b).$$

(*Math. Trip.* 1909.)

26. The least value of $a^2\sec^2 x + b^2\operatorname{cosec}^2 x$ is $(a+b)^2$.

27. Show that $\tan 3x\cot 2x$ cannot lie between $\frac{1}{9}$ and $\frac{3}{2}$.

28. Show that, if the sum of the lengths of the hypothenuse and another side of a right-angled triangle is given, then the area of the triangle is a maximum when the angle between those sides is 60°. (*Math. Trip.* 1909.)

29. A line is drawn through a fixed point (a, b) to meet the axes OX, OY in P and Q. Show that the minimum values of PQ, $OP+OQ$, and $OP\cdot OQ$ are respectively $(a^{2/3}+b^{2/3})^{3/2}$, $(\sqrt{a}+\sqrt{b})^2$, and $4ab$.

30. A tangent to an ellipse meets the axes in P and Q. Show that the least value of PQ is equal to the sum of the semiaxes of the ellipse.

31. Find the lengths and directions of the axes of the conic

$$ax^2 + 2hxy + by^2 = 1.$$

[The length r of the semi-diameter which makes an angle θ with the axis of x is given by

$$1/r^2 = a\cos^2\theta + 2h\cos\theta\sin\theta + b\sin^2\theta.$$

The condition for a maximum or minimum value of r is $\tan 2\theta = 2h/(a-b)$. Eliminating θ between these two equations we find

$$\{a - (1/r^2)\}\{b - (1/r^2)\} = h^2.]$$

32. The greatest value of $x^m y^n$, where x and y are positive and $x+y=k$, is

$$m^m n^n k^{m+n}/(m+n)^{m+n}.$$

33. The greatest value of $ax+by$, where x and y are positive and $x^2+xy+y^2 = 3\kappa^2$, is

$$2\kappa\sqrt{a^2 - ab + b^2}.$$

[If $ax+by$ is a maximum then $a + b(dy/dx) = 0$. The relation between x and y gives $(2x+y)+(x+2y)(dy/dx) = 0$. Equate the two values of dy/dx.]

34. If θ and ϕ are acute angles connected by the relation $a\sec\theta + b\sec\phi = c$, where a, b, c are positive, then $a\cos\theta + b\cos\phi$ is a minimum when $\theta = \phi$.

125. The Mean Value Theorem. We can proceed now to the proof of another general theorem of extreme importance, a theorem commonly known as '*The Mean Value Theorem*' or '*The Theorem of the Mean*'.

THEOREM. *If $\phi(x)$ has a derivative for all values of x in the interval $[a, b]$, then there is a value ξ of x between a and b, such that*

$$\phi(b) - \phi(a) = (b-a)\phi'(\xi).$$

Before we give a strict proof of this theorem, which is perhaps the most important theorem in the Differential Calculus, it will be well to point out its obvious geometrical meaning. This is simply (see Fig. 43) that if the curve APB has a tangent at all points of its length then there must be a point, such as P, where the tangent is parallel to AB. For $\phi'(\xi)$ is the tangent of the angle which the tangent at P makes with OX, and $\{\phi(b) - \phi(a)\}/(b-a)$ the tangent of the angle which AB makes with OX.

It is easy to give a strict analytical proof. Consider the function

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

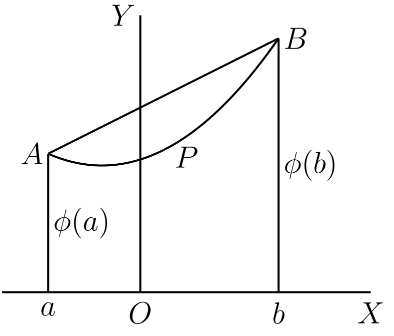

Fig. 43.

which vanishes when $x = a$ and $x = b$. It follows from Theorem B of § 121 that there is a value ξ for which its derivative vanishes. But this derivative is

$$\frac{\phi(b) - \phi(a)}{b - a} - \phi'(x);$$

which proves the theorem. It should be observed that it has not been assumed in this proof that $\phi'(x)$ is continuous.

It is often convenient to express the Mean Value Theorem in the form

$$\phi(b) = \phi(a) + (b - a)\phi'\{a + \theta(b - a)\},$$

where θ is a number lying between 0 and 1. Of course $a + \theta(b - a)$ is merely another way of writing 'some number ξ between a and b'. If we put $b = a + h$ we obtain

$$\phi(a + h) = \phi(a) + h\phi'(a + \theta h),$$

which is the form in which the theorem is most often quoted.

Examples XLVII. 1. Show that

$$\phi(b) - \phi(x) - \frac{b - x}{b - a}\{\phi(b) - \phi(a)\}$$

is the difference between the ordinates of a point on the curve and the corresponding point on the chord.

2. Verify the theorem when $\phi(x) = x^2$ and when $\phi(x) = x^3$.

[In the latter case we have to prove that $(b^3 - a^3)/(b - a) = 3\xi^2$, where $a < \xi < b$; *i.e.* that if $\frac{1}{3}(b^2 + ab + a^2) = \xi^2$ then ξ lies between a and b.]

3. Establish the theorem stated at the end of § 124 by means of the Mean Value Theorem.

[Since $\phi'(0) = c$, we can find a small positive value of x such that $\{\phi(x) - \phi(0)\}/x$ is nearly equal to c; and therefore, by the theorem, a small positive value of ξ such that $\phi'(\xi)$ is nearly equal to c, which is inconsistent with $\lim\limits_{x\to+0} \phi'(x) = a$, unless $a = c$. Similarly $b = c$.]

4. Use the Mean Value Theorem to prove Theorem (6) of § 113, assuming that the derivatives which occur are continuous.

[The derivative of $F\{f(x)\}$ is by definition

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

But, by the Mean Value Theorem, $f(x+h) = f(x) + hf'(\xi)$, where ξ is a number lying between x and $x + h$. And

$$F\{f(x) + hf'(\xi)\} = F\{f(x)\} + hf'(\xi)\, F'(\xi_1),$$

where ξ_1 is a number lying between $f(x)$ and $f(x) + hf'(\xi)$. Hence the derivative of $F\{f(x)\}$ is

$$\lim f'(\xi)\, F'(\xi_1) = f'(x)\, F'\{f(x)\},$$

since $\xi \to x$ and $\xi_1 \to f(x)$ as $h \to 0$.]

126. The Mean Value Theorem furnishes us with a proof of a result which is of great importance in what follows: *if $\phi'(x) = 0$, throughout a certain interval of values of x, then $\phi(x)$ is constant throughout that interval.*

For, if a and b are any two values of x in the interval, then

$$\phi(b) - \phi(a) = (b - a)\phi'\{a + \theta(b - a)\} = 0.$$

An immediate corollary is that if $\phi'(x) = \psi'(x)$, throughout a certain interval, then the functions $\phi(x)$ and $\psi(x)$ differ throughout that interval by a constant.

127. Integration. We have in this chapter seen how we can find the derivative of a given function $\phi(x)$ in a variety of cases, including all those of the commonest occurrence. It is natural to consider the converse question, that of *determining a function whose derivative is a given function.*

Suppose that $\psi(x)$ is the given function. Then we wish to determine a function such that $\phi'(x) = \psi(x)$. A little reflection shows us that this question may really be analysed into three parts.

(1) In the first place we want to know whether such a function as $\phi(x)$ *actually exists*. This question must be carefully distinguished from the question as to whether (supposing that there is such a function) we can find any simple formula to express it.

(2) We want to know whether it is possible that more than one such function should exist, *i.e.* we want to know whether our problem is one which admits of a *unique* solution or not; and if not, we want to know whether there is any simple relation between the different solutions which will enable us to express all of them in terms of any particular one.

(3) If there is a solution, we want to know *how to find an actual expression for it.*

It will throw light on the nature of these three distinct questions if we compare them with the three corresponding questions which arise with regard to the differentiation of functions.

(1) A function $\phi(x)$ may have a derivative for all values of x, like x^m, where m is a positive integer, or $\sin x$. It may generally, but not always have one, like $\sqrt[3]{x}$ or $\tan x$ or $\sec x$. Or again it may never have one: for example, the function considered in Ex. XXXVII. 20, which is nowhere continuous, has obviously no derivative for any value of x. Of course during this chapter we have confined ourselves to functions which are continuous except for some special values of x. The example of the function $\sqrt[3]{x}$, however, shows that a continuous function may not have a derivative for some special value of x, in this case $x = 0$. Whether there are continuous functions which *never* have derivatives, or continuous curves which never have tangents, is a further question which is at present beyond us. Common-sense says *No*: but, as we have already stated in § 111, this is one of the cases in which higher mathematics has proved common-sense to be mistaken.

But at any rate it is clear enough that the question 'has $\phi(x)$ a derivative $\phi'(x)$?' is one which has to be answered differently in different circumstances. And we may expect that the converse question 'is there a function $\phi(x)$ of which $\psi(x)$ is the derivative?' will have different answers too. We have already seen that there are cases in which the answer is *No*: thus if $\psi(x)$ is the function which is equal to a, b, or c according as x is less than, equal to, or greater than 0, then the answer is *No* (Ex. XLVII. 3), unless $a = b = c$.

This is a case in which the given function is discontinuous. In what follows, however, we shall always suppose $\psi(x)$ continuous. And then the answer is *Yes*: *if $\psi(x)$ is continuous then there is always a function $\phi(x)$ such that $\phi'(x) = \psi(x)$*. The proof of this will be given in Ch. VII.

(2) The second question presents no difficulties. In the case of differentiation we have a direct definition of the derivative which makes it clear from the beginning that there cannot possibly be more than one. In the case of the converse problem the answer is almost equally simple. It is that if $\phi(x)$ is one solution of the problem then $\phi(x) + C$ is another, for any value of the constant C, and that all possible solutions are comprised in the form $\phi(x) + C$. This follows at once from § 126.

(3) The practical problem of actually finding $\phi'(x)$ is a fairly simple one in the case of any function defined by some finite combination of the ordinary functional symbols. The converse problem is much more difficult. The nature of the difficulties will appear more clearly later on.

Definitions. *If $\psi(x)$ is the derivative of $\phi(x)$, then we call $\phi(x)$ an* **integral** *or* **integral function** *of $\psi(x)$. The operation of forming $\psi(x)$ from $\phi(x)$ we call* **integration**.

We shall use the notation

$$\phi(x) = \int \psi(x)\, dx.$$

It is hardly necessary to point out that $\int \dots dx$ like d/dx must, at present at any rate, be regarded purely as a symbol of operation: the $\int$ and the dx no more mean anything when taken by themselves than do the d and dx of the other operative symbol d/dx.

128. The practical problem of integration. The results of the earlier part of this chapter enable us to write down at once the integrals of some of the commonest functions. Thus

$$\int x^m\,dx = \frac{x^{m+1}}{m+1}, \qquad \int \cos x\,dx = \sin x, \qquad \int \sin x\,dx = -\cos x. \tag{1}$$

These formulae must be understood as meaning that the function on the right-hand side is *one* integral of that under the sign of integration. The *most general* integral is of course obtained by adding to the former a constant C, known as the **arbitrary constant** of integration.

There is however one case of exception to the first formula, that in which $m = -1$. In this case the formula becomes meaningless, as is only to be expected, since we have seen already (Ex. XLII. 4) that $1/x$ cannot be the derivative of any polynomial or rational fraction.

That there really is a function $F(x)$ such that $D_x F(x) = 1/x$ will be proved in the next chapter. For the present we shall be content to assume its existence. This function $F(x)$ is certainly not a polynomial or rational function; and it can be proved that it is not an algebraical function. It can indeed be proved that $F(x)$ is an essentially new function, independent of any of the classes of functions which we have considered yet, that is to say incapable of expression by means of any finite combination of the functional symbols corresponding to them. The proof of this is unfortunately too detailed and tedious to be inserted in this book; but some further discussion of the subject will be found in Ch. IX, where the properties of $F(x)$ are investigated systematically.

Suppose first that x is positive. Then we shall write

$$\int \frac{dx}{x} = \log x, \tag{2}$$

and we shall call the function on the right-hand side of this equation **the logarithmic function**: it is defined so far only for positive values of x.

Next suppose x negative. Then $-x$ is positive, and so $\log(-x)$ is defined by what precedes. Also

$$\frac{d}{dx}\log(-x) = \frac{-1}{-x} = \frac{1}{x},$$

so that, when x is negative,

$$\int \frac{dx}{x} = \log(-x). \tag{3}$$

The formulae (2) and (3) may be united in the formulae

$$\int \frac{dx}{x} = \log(\pm x) = \log|x|, \tag{4}$$

where the ambiguous sign is to be chosen so that $\pm x$ is positive: these formulae hold for all real values of x other than $x = 0$.

The most fundamental of the properties of $\log x$ which will be proved in Ch. IX are expressed by the equations

$$\log 1 = 0, \quad \log(1/x) = -\log x, \quad \log xy = \log x + \log y,$$

of which the second is an obvious deduction from the first and third. It is not really necessary, for the purposes of this chapter, to assume the truth of any of these formulae; but they sometimes enable us to write our formulae in a more compact form than would otherwise be possible.

It follows from the last of the formulae that $\log x^2$ is equal to $2\log x$ if $x > 0$ and to $2\log(-x)$ if $x < 0$, and in either case to $2\log|x|$. Either of the formulae (4) is therefore equivalent to the formula

$$\int \frac{dx}{x} = \tfrac{1}{2}\log x^2. \tag{5}$$

The five formulae (1)–(3) are the five most fundamental *standard forms* of the Integral Calculus. To them should be added two more, viz.

$$\int \frac{dx}{1+x^2} = \arctan x, \quad \int \frac{x}{\sqrt{1-x^2}} = \pm \arcsin x.^* \tag{6}$$

*See § 119 for the rule for determining the ambiguous sign.

129. Polynomials. All the general theorems of § 113 may of course also be stated as theorems in integration. Thus we have, to begin with, the formulae

$$\int \{f(x) + F(x)\}\,dx = \int f(x)dx + \int F(x)\,dx, \tag{1}$$

$$\int kf(x)\,dx = k\int f(x)\,dx. \tag{2}$$

Here it is assumed, of course, that the arbitrary constants are adjusted properly. Thus the formula (1) asserts that the sum of *any* integral of $f(x)$ and *any* integral of $F(x)$ is *an* integral of $f(x) + F(x)$.

These theorems enable us to write down at once the integral of any function of the form $\sum A_\nu f_\nu(x)$, the sum of a finite number of constant multiples of functions whose integrals are known. In particular we can write down the integral of any *polynomial*: thus

$$\int (a_0x^n + a_1x^{n-1} + \dots + a_n)\,dx = \frac{a_0x^{n+1}}{n+1} + \frac{a_1x^n}{n} + \dots + a_nx.$$

130. Rational Functions. After integrating polynomials it is natural to turn our attention next to *rational functions*. Let us suppose $R(x)$ to be any rational function expressed in the standard form of § 117, viz. as the sum of a polynomial $\Pi(x)$ and a number of terms of the form $A/(x-\alpha)^p$.

We can at once write down the integrals of the polynomial and of all the other terms except those for which $p = 1$, since

$$\int \frac{A}{(x-\alpha)^p}\,dx = -\frac{A}{p-1}\,\frac{1}{(x-\alpha)^{p-1}},$$

whether α be real or complex (§ 117).

The terms for which $p = 1$ present rather more difficulty. It follows immediately from Theorem (6) of § 113 that

$$\int F'\{f(x)\}\,f'(x)\,dx = F\{f(x)\}. \tag{3}$$

In particular, if we take $f(x) = ax + b$, where a and b are real, and write $\phi(x)$ for $F(x)$ and $\psi(x)$ for $F'(x)$, so that $\phi(x)$ is an integral of $\psi(x)$, we obtain

$$\int \psi(ax+b)\,dx = \frac{1}{a}\phi(ax+b). \tag{4}$$

Thus, for example,

$$\int \frac{dx}{ax+b} = \frac{1}{a}\log|ax+b|,$$

and in particular, if α is real,

$$\int \frac{dx}{x-\alpha} = \log|x-\alpha|.$$

We can therefore write down the integrals of all the terms in $R(x)$ for which $p = 1$ and α is real. There remain the terms for which $p = 1$ and α is complex.

In order to deal with these we shall introduce a restrictive hypothesis, viz. that all the coefficients in $R(x)$ are real. Then if $\alpha = \gamma + \delta i$ is a root of $Q(x) = 0$, of multiplicity m, so is its conjugate $\bar{\alpha} = \gamma - \delta i$; and if a partial fraction $A_p/(x-\alpha)^p$ occurs in the expression of $R(x)$, so does $\bar{A}_p/(x-\bar{\alpha})^p$, where $\bar{A}_p$ is conjugate to A_p. This follows from the nature of the algebraical processes by means of which the partial fractions can be found, and which are explained at length in treatises on Algebra.*

Thus, if a term $(\lambda+\mu i)/(x-\gamma-\delta i)$ occurs in the expression of $R(x)$ in partial fractions, so will a term $(\lambda-\mu i)/(x-\gamma+\delta i)$; and the sum of these two terms is

$$\frac{2\{\lambda(x-\gamma)-\mu\delta\}}{(x-\gamma)^2+\delta^2}.$$

This fraction is in reality the most general fraction of the form

$$\frac{Ax+B}{ax^2+2bx+c},$$

*See, for example, Chrystal's *Algebra*, vol. i, pp. 151–9.

where $b^2 < ac$. The reader will easily verify the equivalence of the two forms, the formulae which express λ, μ, γ, δ in terms of A, B, a, b, c being

$$\lambda = A/2a, \quad \mu = -D/(2a\sqrt{\Delta}), \quad \gamma = -b/a, \quad \delta = \sqrt{\Delta}/a,$$

where $\Delta = ac - b^2$, and $D = aB - bA$.

If in (3) we suppose $F\{f(x)\}$ to be $\log|f(x)|$, we obtain

$$\int \frac{f'(x)}{f(x)}\,dx = \log|f(x)|; \tag{5}$$

and if we further suppose that $f(x) = (x-\lambda)^2 + \mu^2$, we obtain

$$\int \frac{2(x-\lambda)}{(x-\lambda)^2+\mu^2}\,dx = \log\{(x-\lambda)^2+\mu^2\}.$$

And, in virtue of the equations (6) of § 128 and (4) above, we have

$$\int \frac{-2\delta\mu}{(x-\lambda)^2+\mu^2}\,dx = -2\delta \arctan\left(\frac{x-\lambda}{\mu}\right).$$

These two formulae enable us to integrate the sum of the two terms which we have been considering in the expression of $R(x)$; and we are thus enabled to write down the integral of any real rational function, if all the factors of its denominator can be determined. The integral of any such function is composed of *the sum of a polynomial, a number of rational functions of the type*

$$-\frac{A}{p-1}\,\frac{1}{(x-\alpha)^{p-1}},$$

a number of logarithmic functions, and a number of inverse tangents.

It only remains to add that if α is complex then the rational function just written always occurs in conjunction with another in which A and α are replaced by the complex numbers conjugate to them, and that the sum of the two functions is a real rational function.

Examples XLVIII. 1. Prove that

$$\int \frac{Ax+B}{ax^2+2bx+c}\,dx = \frac{A}{2a}\log|X| + \frac{D}{2a\sqrt{-\Delta}}\log\left|\frac{ax+b-\sqrt{-\Delta}}{ax+b+\sqrt{-\Delta}}\right|$$

(where $X = ax^2+bx+c$) if $\Delta < 0$, and

$$\int \frac{Ax+B}{ax^2+2bx+c}\,dx = \frac{A}{2a}\log|X| + \frac{D}{2a\sqrt{\Delta}}\arctan\left(\frac{ax+b}{\sqrt{\Delta}}\right)$$

if $\Delta > 0$, Δ and D having the same meanings as on p. 283.

2. In the particular case in which $ac = b^2$ the integral is

$$-\frac{D}{a(ax+b)} + \frac{A}{a}\log|ax+b|.$$

3. Show that if the roots of $Q(x) = 0$ are all real and distinct, and $P(x)$ is of lower degree than $Q(x)$, then

$$\int R(x)\,dx = \sum \frac{P(\alpha)}{Q'(\alpha)}\log|x-\alpha|,$$

the summation applying to all the roots α of $Q(x) = 0$.

[The form of the fraction corresponding to α may be deduced from the facts that

$$\frac{Q(x)}{x-\alpha} \to Q'(\alpha), \quad (x-\alpha)R(x) \to \frac{P(\alpha)}{Q'(\alpha)},$$

as $x \to \alpha$.]

4. If all the roots of $Q(x)$ are real and α is a double root, the other roots being simple roots, and $P(x)$ is of lower degree than $Q(x)$, then the integral is $A/(x-\alpha) + A'\log|x-\alpha| + \sum B\log|x-\beta|$, where

$$A = -\frac{2P(\alpha)}{Q''(\alpha)}, \quad A' = \frac{2\{3P'(\alpha)Q''(\alpha) - P(a)Q'''(\alpha)\}}{3\{Q''(\alpha)\}^2}, \quad B = \frac{P(\beta)}{Q'(\beta)},$$

and the summation applies to all roots β of $Q(x) = 0$ other than α.

5. Calculate

$$\int \frac{dx}{\{(x-1)(x^2+1)\}^2}.$$

[The expression in partial fractions is

$$\frac{1}{4(x-1)^2} - \frac{1}{2(x-1)} - \frac{i}{8(x-i)^2} + \frac{2-i}{8(x-i)} + \frac{i}{8(x+i)^2} + \frac{2+i}{8(x+i)},$$

and the integral is

$$-\frac{1}{4(x-1)} - \frac{1}{4(x^2+1)} - \tfrac{1}{2}\log|x-1| + \tfrac{1}{4}\log(x^2+1) + \tfrac{1}{4}\arctan x.]$$

6. Integrate

$$\frac{x}{(x-a)(x-b)(x-c)},\quad \frac{x}{(x-a)^2(x-b)},\quad \frac{x}{(x-a)^2(x-b)^2},\quad \frac{x}{(x-a)^3},$$
$$\frac{x}{(x^2+a^2)(x^2+b^2)},\quad \frac{x^2}{(x^2+a^2)(x^2+b)^2},\quad \frac{x^2-a^2}{x^2(x^2+a^2)},\quad \frac{x^2-a^2}{x(x^2+a^2)^2}.$$

7. Prove the formulae:

$$\int \frac{dx}{1+x^4} = \frac{1}{4\sqrt{2}}\left\{ \log\left(\frac{1+x\sqrt{2}+x^2}{1-x\sqrt{2}+x^2}\right) + 2\arctan\left(\frac{x\sqrt{2}}{1-x^2}\right)\right\},$$
$$\int \frac{x^2\,dx}{1+x^4} = \frac{1}{4\sqrt{2}}\left\{ -\log\left(\frac{1+x\sqrt{2}+x^2}{1-x\sqrt{2}+x^2}\right) + 2\arctan\left(\frac{x\sqrt{2}}{1-x^2}\right)\right\},$$
$$\int \frac{dx}{1+x^2+x^4} = \frac{1}{4\sqrt{3}}\left\{ \sqrt{3}\log\left(\frac{1+x+x^2}{1-x+x^2}\right) + 2\arctan\left(\frac{x\sqrt{3}}{1-x^2}\right)\right\}.$$

131. Note on the practical integration of rational functions. The analysis of § 130 gives us a general method by which we can find the integral of any real rational function $R(x)$, *provided we can solve the equation* $Q(x) = 0$. In simple cases (as in Ex. 5 above) the application of the method is fairly simple. In more complicated cases the labour involved is sometimes prohibitive, and other devices have to be used. It is not part of the purpose of this book to go into practical problems of integration in detail. The reader who desires fuller information may be referred to Goursat's *Cours d'Analyse*, second ed., vol. i, pp. 246 *et seq.*, Bertrand's *Calcul Intégral*, and Dr Bromwich's tract *Elementary Integrals* (Bowes and Bowes, 1911).

If the equation $Q(x) = 0$ cannot be solved algebraically, then the method of partial fractions naturally fails and recourse must be had to other methods.*

132. Algebraical Functions. We naturally pass on next to the question of the integration of *algebraical* functions. We have to consider the problem of integrating y, where y is an algebraical function of x. It is however convenient to consider an apparently more general integral, viz.

$$\int R(x, y)\, dx,$$

where $R(x, y)$ is any rational function of x and y. The greater generality of this form is only apparent, since (Ex. XIV. 6) the function $R(x, y)$ is itself an algebraical function of x. The choice of this form is in fact dictated simply by motives of convenience: such a function as

$$\frac{px + q + \sqrt{ax^2 + 2bx + c}}{px + q - \sqrt{ax^2 + 2bx + c}}$$

is far more conveniently regarded as a rational function of x and the simple algebraical function $\sqrt{ax^2 + 2bx + c}$, than directly as itself an algebraical function of x.

133. Integration by substitution and rationalisation. It follows from equation (3) of § 130 that if $\int \psi(x)\, dx = \phi(x)$ then

$$\int \psi\{f(t)\}\, f'(t)\, dt = \phi\{f(t)\}. \tag{1}$$

This equation supplies us with a method for determining the integral of $\psi(x)$ in a large number of cases in which the form of the integral is not directly obvious. It may be stated as a rule as follows: *put* $x = f(t)$,

*See the author's tract "The integration of functions of a single variable" (*Cambridge Tracts in Mathematics*, No. 2, second edition, 1915). This does not often happen in practice.

where $f(t)$ is any function of a new variable t which it may be convenient to choose; multiply by $f'(t)$, and determine (if possible) the integral of $\psi\{f(t)\} f'(t)$; express the result in terms of x. It will often be found that the function of t to which we are led by the application of this rule is one whose integral can easily be calculated. This is always so, for example, if it is a rational function, and it is very often possible to choose the relation between x and t so that this shall be the case. Thus the integral of $R(\sqrt{x})$, where R denotes a rational function, is reduced by the substitution $x = t^2$ to the integral of $2tR(t^2)$, *i.e.* to the integral of a rational function of t. This method of integration is called **integration by rationalisation**, and is of extremely wide application.

Its application to the problem immediately under consideration is obvious. *If we can find a variable t such that x and y are both rational functions of t, say $x = R_1(t)$, $y = R_2(t)$, then*

$$\int R(x, y)\, dx = \int R\{R_1(t), R_2(t)\}\, R_1'(t)\, dt,$$

and the latter integral, being that of a rational function of t, can be calculated by the methods of § 130.

It would carry us beyond our present range to enter upon any general discussion as to when it is and when it is not possible to find an auxiliary variable t connected with x and y in the manner indicated above. We shall consider only a few simple and interesting special cases.

134. Integrals connected with conics. Let us suppose that x and y are connected by an equation of the form

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0;$$

in other words that the graph of y, considered as a function of x is a conic. Suppose that (ξ, η) is any point on the conic, and let $x - \xi = X$, $y - \eta = Y$. If the relation between x and y is expressed in terms of X and Y, it assumes the form

$$aX^2 + 2hXY + bY^2 + 2GX + 2FY = 0,$$

where $F = h\xi+b\eta+f$, $G = a\xi+h\eta+g$. In this equation put $Y = tX$. It will then be found that X and Y can both be expressed as rational functions of t, and therefore x and y can be so expressed, the actual formulae being

$$x - \xi = -\frac{2(G+Ft)}{a+2ht+bt^2}, \quad y - \eta = -\frac{2t(G+Ft)}{a+2ht+bt^2}.$$

Hence the process of rationalisation described in the last section can be carried out.

The reader should verify that

$$hx + by + f = -\tfrac{1}{2}(a+2ht+bt^2)\frac{dx}{dt},$$

so that

$$\int \frac{dx}{hx+by+f} = -2\int \frac{dt}{a+2ht+bt^2}.$$

When $h^2 > ab$ it is in some ways advantageous to proceed as follows. The conic is a hyperbola whose asymptotes are parallel to the lines

$$ax^2 + 2hxy + by^2 = 0,$$

or

$$b(y - \mu x)(y - \mu' x) = 0,$$

say. If we put $y - \mu x = t$, we obtain

$$y - \mu x = t, \quad y - \mu' x = -\frac{2gx + 2fy + c}{bt},$$

and it is clear that x and y can be calculated from these equations as rational functions of t. We shall illustrate this process by an application to an important special case.