Department of Mathematics, Statistics
and Computing Science

Pure Mathematics 111-2/112-2

Introductory Calculus

H-Work (Optional)

Printed at The University of New England, January 1984.

THE UNIVERSITY OF NEW ENGLAND

Pure Mathematics 111-2/112-2

Preliminary Honours (H) Work

(INTRODUCTORY LETTER)

Preliminary honours (H) work is intended for students with a strong natural ability and interest in mathematics, particularly students preparing for the study of mathematics at a higher level (an Honours degree, Litt.B etc.) or who anticipate a similar course of study in a mathematically based discipline such as theoretical physics or chemistry, certain areas of economics and biology.

The higher examination grades (Distinction and High Distinction) are awarded to students who demonstrate outstanding ability. Normally such students will be expected to have attempted the H work. *In awarding these grades preference will be given to those who have displayed competence in the H work.*

In 1983 the H work supplements the pass course and will be available to both internal and external students. Change between the alternative versions PM111 and PM112 is not restricted to the first weeks of term. In fact, you would do best to take no action until it is clear which you want to do.

For internal students a regular lecture/tutorial devoted to the H work will be conducted at 11 a.m. each Tuesday, beginning in the third week of the academic year. To assist external students attempting the H work special sessions at the residential and weekend schools will be organised.

The enclosed notes cover the H-material for the first semester. You will see internal evidence that the course was begun by Dr. Sims!

Chapter I of the notes presents *Cantor's theory of infinite sets.* When first put out (in about 1870) this theory caused a revolution in Mathematics, with many eminent mathematicians of the time refusing to accept it. Subsequently it assumed a place of fundamental importance, being basic to much twentieth-century mathematics. This material is not directly related to the pass course. Largely, I have included it for its novelty. I hope you find it both challenging and interesting. Working through this material will enable you to assess your ability to handle the H-work. The material also provides a good opportunity for consolidating the mathematical concept of "function". As you will soon learn, the function concept is basic to modern mathematics. For this reason I have collected the material on functions into an appendix. Unlike the theory of infinite sets, the work on functions is directly relevant to later courses. Indeed, with the exception of Theorem 4 the material of the appendix is repeated, in a more specialised context in §1 of the Notes for Part II of the course on *Elementary Real Analysis.*

Chapter II, on *Mathematical Induction,* should join on to work you had at school but will, I hope, extend your previous knowledge.

At various points the Pass course contents itself with a fairly heuristic approach. Part II of the H-work will acquaint you with ideas and methods leading to a more rigorous development of the material, and provide a foundation for the subsequent study of mathematics at a higher level.

Notes for second-semester H-work will be sent later in the year to those who submit the First Semester assignment. These notes should cover:

§1 Real functions of a real variable;

§2 Sequences of real numbers and special properties of $\mathbb{R}$;

§3 Continuity;

§4 Differentiable functions;

and §5 Integrable functions

Part I of the H-work will be examined by means of an assignment (enclosed) due on Friday, 3rd June.

Part II of the H-work on Elementary Real Analysis will be examined by a separate 2-hour paper in November. (It will be necessary for external students to indicate their intention to sit for this paper early in third term.)

External students having any queries or problems with the H work please feel free to contact me. If you write out solutions to any of the exercises, please send them direct to me at the Mathematics Department. (They are recommended, but in no way compulsory.)

External students should submit the assignment through the Department of External Studies, to assist in record-keeping.

Wishing you success and enjoyment in your studies,

Professor R.L.T. Smith

PURE MATHEMATICS 112-2 (H)

ASSIGNMENT ON COUNTABILITY AND INDUCTION

(Due: Friday, 3rd June)

1. Let $\mathbb{Q}$ denote the set of rational numbers, that is, numbers expressible uniquely in the form p/q, where q is a strictly positive integer, p is an integer and p, q have no common factor.

   i) Show that the mapping $f:q \mapsto \frac{1}{q}$ is a one-to-one mapping of $\mathbb{N}$ into $\mathbb{Q}$.

   ii) Show that the mapping

$$g : \frac{p}{q} \to \begin{cases} 2^q\ 3^p & \text{if } p \geq 0 \\ 2^q\ 5^{-p} & \text{if } p < 0 \end{cases}$$

   is a one-to-one mapping from $\mathbb{Q}$ into $\mathbb{N}$.

   iii) Using the Schroeder-Bernstein Theorem deduce that $\mathbb{Q}$ is countably infinite.

2. Let $A_1, A_2, A_3, \ldots$ be a finite or countably infinite family of finite or countably infinite sets, show that their union $A = A_1 \cup A_2 \cup A_3 \cup \ldots$ is either finite or countably infinite.

3. Show that, for all $n \in \mathbb{N}$, each of $\cos nx$ and $\frac{\sin nx}{\sin x}$ can be expressed as a polynomial in cos x, that is, that we can write

$$\cos nx = C_n(\cos x),$$

$$\sin nx = \sin x \,.\, S_{n-1}(\cos x),$$

where each of C_n and S_{n-1} is a polynomial.

Show further that

(i) the subscripts give the degrees of the polynomials, and

(ii) $C_n(t)$ has leading term $2^{n-1}t^n$, $S_{n-1}(t)$ has leading term $2^{n-1}t^{n-1}$

[Depending on how you arrange your proof, you may find it natural to obtain the results in succession or to obtain them all at the same time.]

4. i) Let $f: X \to Y$ and $g: Z \to U$ be such that $g \circ f$ is defined. Show that $g \circ f$ is 1 to 1 if both f and g are.

   Is the converse true? (That is, if $g \circ f$ is 1 to 1 are both f and g necessarily 1 to 1 functions?)

   ii) If $f: X \to Y$ is invertible show that f^{-1} is both 1 to 1 and onto.

CHAPTER 1

COUNTABLE AND UNCOUNTABLE SETS

1. Introduction

Our ordinary use of the word "count" is to mean that we have a standard string of names

one, two, three, four, ... ,

and that we work through this string as we point at objects, as sheep go through a gate etc. Our report is made by giving the name that we stopped at. Obviously the string

un, deux, trois, quatre, ...

would do just as well, and there are plenty of others to use.

Obviously we could not apply this procedure to an infinite set since we should never stop. And so there is nothing we can say about infinite sets — or is there some rearrangement of what we did which will let us say something useful? In the 1870's the German mathematician Georg Cantor (1845-1918) recognised that having a standard string of names is not of dominating importance here: what we need is to be able to say of two given sets A and B that they have the "same number" of elements or that they do not. He recognised further that if we can set up what is often described as a "one-one correspondence" between A and B we can sensibly speak of them as having the "same number" of elements. (Here "sensibly" means "self-consistently", that is we do not generate contradictions.)

Cantor built a whole theory on this observation, a theory which has had a great influence on mathematics. Although he was the first to *exploit* the idea, it had been noticed before, but only as a curiosity or source of paradoxes. Thus Galileo* (1564-1642) noticed the correspondence

*There is no point in remembering dates exactly, but it does help your understanding of mathematics to have some idea of who lived before who. As an aid to your memory, observe that Galileo died in the year that Newton (1642-1727) was born.

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & \ldots & n & \ldots\ldots \\ 1 & 4 & 9 & 16 & 25 & \ldots & n^2 & \ldots\ldots \end{array}$$

and remarked that there are precisely as many perfect squares among the natural numbers as there are natural numbers. This struck him as strange, when we consider how sparsely the squares are strewn. Again the geometrical construction of Fig. 1 shows that there are as many points on a segment of length ℓ as on one of length 2ℓ.

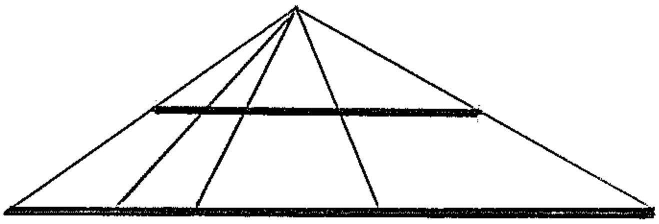


Fig. 1

Before we can start the study of Cantor's theory we need some preliminary general remarks and something about the language and symbols for sets.

2. Level of abstraction

In the sort of work that we are beginning, it is often natural to ask "Where should we start?" or "What may we assume?" It may be helpful to think of mathematics as a house, which we can easily enter at ground-floor level. It can be quite difficult to climb upwards, but it can be just as hard to penetrate to the basement or foundations. Going upwards means constructing more and more complex theories and probably developing more and more complicated formulae. Going downwards means analysing our basic concepts, introducing fine distinctions which may seem fussy for a long time and giving attention to making the weakest possible assumptions.

You should be aware, that just as a first-year course does not study the most complex topics, so also our analysis of our bases will not go very deep. For example, what we say about sets in §§3 to 5 can be further refined (in the downward direction: we could start from fewer concepts, weaker axioms). This does not mean that we proceed illogically. Rather we can say that we start from a redundant set of axioms and work carefully from there.

If the previous paragraph seems vague, you can find an example in what we shall do about the natural numbers. In this chapter we take them as familiar and do not try to analyse our assumptions. In Chapter II we introduce Peano's axioms, as a way of exhibiting mathematical induction as part of a logical development rather than a piece of black magic, and we could at that stage work through the steps by which that set of axioms leads to the propositions familiar in elementary arithmetic. Although we do not do this (it needs care rather than any very difficult proofs), the possibility is there. At a further stage we could set out to obtain Peano's axioms from more primitive ideas.

It is important to realise we must start from something and that the choice here is not the only one. It is time to get moving.

3. Sets

We shall take an intuitive approach to the concept of set. What is important to us is that when we are considering a set of numbers, of points, or any other sort of objects, we are able to determine whether an object does or does not belong to the set. (That is, we have a test. In some cases we may find that our test is indirect and is laborious to apply, but this makes no difference of principle.)

We want to introduce language and notation. Most often we use capital letters to denote sets and, quite often, small letters for their members. Thus we may have

$$x \text{ is a member of } A,$$

which is symbolised*

$$x \in A.$$

*It is nowadays almost universal to distinguish $\in$ and ε, the first used as here, the second for epsilon as a symbol to be employed in a calculation.

If we want to list the elements of a set we use curly brackets as in

$$A = \{x, y, z\} , \quad B = \{1, 3, 5, 7, \dots \},$$

and, by an extension of this,

$$C = \{2n : n \text{ an integer}\}$$

for the set of even integers,

$$D = \{x : 0 < x < 1\} ,$$

or, more generally,

$$H = \{t : t \text{ has property } P\} .$$

The colon which occurs here can be read "such that".

There is an important *algebra of sets* concerning, in particular, the symbols $\cup$ and $\cap$. We write

$$A \cup B \quad \text{for the } \textit{union}$$

and

$$A \cap B \quad \text{for the } \textit{intersection}$$

of the sets A and B, which we can define by

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

and

$$A \cap B = \{t : t \in A \text{ and } t \in B\}.$$

Note well that there is no question of counting points twice if they are common to A and B (that is, in $A \cap B$). Note well that here, and in all mathematical writing, "or" is interpreted in the inclusive sense to mean "$x \in A$ or $x \in B$ or possibly both".

The combination $A \cup B$ is often read "A cup B" and also often "A union B"; similarly $A \cap B$ is read "A cap B" or "A intersection B. A style of lettering that has become very popular for denoting sets is

$$\mathbb{N}, \ \mathbb{Z}, \ \mathbb{Q} \quad \text{etc.}$$

These three letters are just about standard as symbols:

$\mathbb{N}$ for the set of natural numbers,

$\mathbb{Z}$ for the set of integers,

$\mathbb{Q}$ for the set of rational numbers.

It is easy to see why $\mathbb{N}$ is selected; for $\mathbb{Z}$, the reference is to the German word Zahlen (= numbers) and, for $\mathbb{Q}$, the reference is to the word "quotient", since every rational number can be written in the form a/b, the quotient of two integers. (One of the maddening features of this style of lettering is that one cannot easily get a neat version of X or Y, which is so often wanted.)

We shall not be making serious algebraic calculations with sets, but you should notice that, in excellent analogy with + and ×, they satisfy

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A,$$

the two *commutative laws*,

$$(A \cup B) \cup C = A \cup (B \cup C), \qquad (A \cap B) \cap C = A \cap (B \cap C),$$

the two *associative laws* and, very clearly not analogously with + and × for numbers, the <u>two</u> *distributive laws*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

You have probably seen these, and other, relations illustrated with *Venn diagrams*. (Compare the figure on page 7 .)

4. Set inclusion

The symbols $\subset$ and $\subseteq$ are used much as $<$ and $\leq$ between numbers. Thus to say that

$$A \subseteq B$$

is to say that every member of A is a member of B, whilst to say that

$$C \subset D$$

is to say that $C \subseteq D$ but $C \neq D$, that is, there is at least one element x which is in D but not in C. As with inequalities between numbers, to make the distinction between $\subset$ and $\subseteq$ sometimes seems a piece of fussiness but is often to be recognised as a help in saying what we mean. Both for $<$ and $\subset$, experience shows that it pays to be strict with ourselves. Notice that we have, for all A and B,

$$A \cap B \subseteq A, \qquad A \subseteq A \cup B.$$

There is an important point about the ways of proving two sets equal. Suppose R and S have been obtained in roundabout ways and that it is not obvious that they are equal but we suspect this. Of course, we say R and S are equal if they have exactly the same members, and only if this holds. (This could have been said explicitly before now, but it didn't arise.) In symbols we want

$$R \subseteq S \tag{1}$$

and

$$S \subseteq R \tag{2}$$

It is nearly always the case that the only feasible way of showing $R = S$ is to establish first (1), which means that we show that

$$x \in R \Rightarrow x \in S\ ,$$

and then to establish (2). It often happens that one of (1) and (2) is very easy to prove and the other hard, but we still need to prove them separately.

5. Some other set-theoretic notations

If A and B are two sets, we write

$$A \times B$$

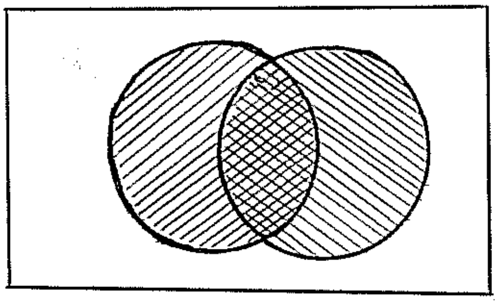
for their *Cartesian product*, the set of all couples formed with a member of A and a member of B or, in symbols,

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

If the name seems strange, notice that the plane with Cartesian coordinates can be regarded as the set of couples (x,y) where each of x and y is a real number. Here $A = B$, each being equal to $\mathbb{R}$, the set of real numbers. We could write

$$\mathbb{R} \times \mathbb{R} = \{(x,y) : x \in \mathbb{R} \quad \text{and} \quad y \in \mathbb{R}\} .$$

When $A = B$, it is quite common to write A^2 for $A \times A$; thus we could denote the real plane by $\mathbb{R}^2$.



A hatched ////

B hatched \\\\

A simple Venn diagram.

Here $A \cup B$ is the hatched set

$A \cap B$ is the doubly hatched set

$\mathscr{C}(A \cup B)$ is the set left unhatched

(Note that there it is in no way compulsory to use circular discs for the given sets but that you should plan your diagram for easy reading. Except in the simplest cases, give a clear statement of what your hatchings mean.)

If A is a set, we describe as the *complement* of A the set of elements not in A. Here it is essential to know what "largest" set we are working in. For example, the complement of the set of boys in the set of all males is different from the complement in the set of all humans.

The order of business is not as just described. Rather we decide what our "largest" set is; it is usually called the *universal set*, with I and U as frequent symbols in general discussions. (Of course, in a special discussion we could easily have $\mathbb{N}$ or $\mathbb{R}$ as our universal set.)

Then

$$\text{Complement of } A = \{x : x \in I \quad \text{and} \quad x \notin A\}.$$

Unhappily there is difficulty over symbolism here. Although $\bar{A}$ and A' will be often met with there are other important uses of these labels. In these notes we shall write

$$\mathscr{C}A \qquad \text{or} \qquad I - A.$$

The minus sign between sets is used as follows:

$$P - Q = \{x : x \in P \quad \text{and} \quad x \notin Q\}.$$

It is often convenient to write

$$P - Q = P \cap \mathscr{C}Q.$$

It is of great algebraic convenience to introduce as a symbol for the *empty set* a modified 0. Various modifications have been tried; the version which has finally been adopted is

$$\emptyset \; .$$

The essential property of $\emptyset$ is that for any element x we can assert

$$x \notin \emptyset.$$

Observe the consequences

$$\emptyset = \quad I, \qquad I = \quad \emptyset.$$

Observe the very convenient way of writing down "A and B are disjoint" as

$$A \cap B = \emptyset.$$

EXERCISES (Whenever possible, illustrate with Venn diagrams. You will probably find that you need two or more diagrams in some questions.)

1. Show that $A \cup A = A$ and $A \cap A = A$.

2. Prove the two distributive laws.

3. Show that

$$\mathscr{C}(A \cup B) = \mathscr{C}A \cap \mathscr{C}B, \qquad \mathscr{C}(A \cap B) = \mathscr{C}A \cup \mathscr{C}B.$$

6. Counting

Basic to our ideas about counting is the set $\mathbb{N}$ of *natural numbers*. The elements of $\mathbb{N}$ are ordered $1 < 2 < 3 < \ldots$, thus we can speak of "the first n natural numbers", meaning $\{1,2,3,\ldots,n\}$. To say that some set B has n elements means that we can imagine, at least in principle, selecting an element of B and "marking" it with the number 1, then selecting an unmarked element of B and marking it with 2, then marking another with 3 and so on. When there are no elements of B left unmarked we should find that we have used up each of the numbers 1,2,3,...,n and no others. For example the set

{Malphas, the dog; Mephisto, the cat; Aidan, the boy; and Sims, the lecturer}

has 4 elements, since we could "paint" 1 on Aidan, 2 on Malphas, 3 on Mephisto and 4 on Sims. More formally, this set has 4 elements since we have been able to find a *one-to-one function* from {1,2,3,4} *onto* it, viz

$1 \mapsto$ Aidan $2 \mapsto$ Malphas $3 \mapsto$ Mephisto $4 \mapsto$ Sims..

And in general we can say that *a set B has n elements if there exists a one-to-one function from $\{1,2,\ldots,n\}$ onto b.*

The number of elements in a set B is termed the Cardinal number of B and is sometimes denoted by #B or $|B|$. Thus, the cardinal number of B equals n if there exists a one-to-one and onto function $f:\{1,2,3,\ldots,n\} \to B$.

This is very similar to what we said about counting in §1. As was indicated there, our point of view is clearly suggested by the description "one-one correspondence" between two sets A, here $\{1,2,\ldots,n\}$, and B. However, since this terminology leaves quite vague whether we are going from A to B or from B to A, we shall avoid it, and talk always of functions, whose domains and target sets are clearly specified. (Compare Appendix 1.)

It may seem unnecessary and pedantic to point out that we need to check that #B is well defined. However, as far as we have seen, there might be a one-to-one function f from $\{1,2,3,\ldots,n\}$ onto B and another one-to-one function g from $\{1,2,3,\ldots,m\}$ onto B, where $m \neq n$, and then to say #B equals both m and n ! To see how we might exclude this, note that

g^{-1} exists and is a one-to-one and onto function from B to $\{1,2,\ldots,m\}$ (Cor. 3 of page A9), and so the composite $g^{-1} \circ f$ would be a one-to-one and onto function from $\{1,2,\ldots,n\}$ to $\{1,2,\ldots,m\}$(Cf. the Exercise on page A9).

$$\begin{tikzcd} \{1,2,3,\ldots,n\} \arrow[r, "f"] \arrow[rr, bend right, "g^{-1} \circ f"'] & B \arrow[r, "g^{-1}"] & \{1,2,3,\ldots,m\} \end{tikzcd}$$

Our feeling is that this is impossible unless $m = n$, and we shall construct a proof in Chapter 2. You should be able to get a feel for the sort of proof required (apart from the necessary induction) by trying to construct a one-to-one and onto function from $\{1,2,3,4\}$ to $\{1,2,3\}$.

We now consider two sets A and B, without the specialisation that A is $\{1,2,\ldots,n\}$, but in the case when the sets have the same cardinal number n. The following theorem gives us as foreshadowed in §1, a rearrangement which allows of an extension to infinite sets.

THEOREM 1: *Two finite sets* A *and* B *have the same number of elements if and only if there is a one-to-one and onto function* f *from* A *to* B.

REMARK

This is the first of many results we shall meet whose enunciation involves the phrase "if and only if". Let p stand for the statement "A and B have the same number of elements" and let q stand for "there is a one-to-one and onto function f from A to B". It is important to recognise that our theorem, which has the form "p if and only if q", is really two theorems:

(1) "p only if q",
(2) "p if q".

We start by rewriting (2) as

"If q , then p" or "$q \Rightarrow p$",

where the arrow-headed equals sign can be read "implies". Next we recognise that if (1) holds and p is true then q must be true, which allows us to rewrite (1) as

"$p \Rightarrow q$".

Thus to prove our theorem we must establish the two statements:

(1) $p \Rightarrow q$
and (2) $q \Rightarrow p$.

If as a purely notational point we rewrite (2) as

$$p \Leftarrow q$$

we get useful shorthand which we use here and in other places.

Proof of Theorem 1

($\Rightarrow$) If A and B have the same number of elements n then there are one-to-one and onto functions f: {1,2,...,n} → A and {g: 1,2,...n} → B.

$$\begin{tikzcd}
 & A \arrow[dd, "g \circ f^{-1}"] \arrow[dl, dashed, bend left, "f^{-1}"] \\
\{1,2,\ldots,n\} \arrow[ur, "f"] \arrow[dr, "g"] & \\
 & B \arrow[ul, dashed, bend left, "g^{-1}"]
\end{tikzcd}$$

Now, f^{-1}: A → {1,2,...,n} is one-to-one and onto and so the composite $g \circ f^{-1}$: A → B is one-to-one and onto, as required.

($\Leftarrow$) Let A have n elements and let f be a one-to-one and onto function from A to B. Then there is a one-to-one and onto functions g: {1,2,...,n} → A.

$$\begin{tikzcd}
\{1,2,\ldots,n\} \arrow[r, "g"] \arrow[rr, bend right, "f \circ g"'] & A \arrow[r, "f"] & B
\end{tikzcd}$$

The composite f o g is a one-to-one and onto function from {1,2,...,n} to B and so B has n elements, as required.

[The little square "□" is used to indicate the end of proofs. □

What we have done so far has been rather elementary (and perhaps boring — the basic ideas are often introduced, for better or more likely for worse, at a primary school level). However, it has afforded us the opportunity of introducing and revising some important material: functions (one-to-one, onto and composite); the logic of proofs.

7. Extended counting

We now turn to *infinite* sets, and join onto the result of Theorem 1 by introducing the

Definition

Two (infinite) sets A *and* B *have the 'same' number of elements if there exists a one-to-one and onto function from* A *to* B.

We are working towards the result that there can be different "infinite numbers" of elements in sets. Evidently the set of perfect squares (in Galileo's curious observation in §1) has, in the sense of our definition, the same number of elements as $\mathbb{N}$. More generally, we shall say that any set which has the same number of elements as $\mathbb{N}$ is *countably infinite.*

We use $\aleph_0$ as the symbol for the *cardinal number of any countably infinite set,* ($\aleph$, pronounced "aleph" is the first letter of the Hebrew alphabet.) Thus B has $\aleph_0$ elements ($\#B = \aleph_0$) if and only if there exists a one-to-one and onto function from $\mathbb{N}$ to B.

REMARKS. (i) In some works "denumerable" is used as a synonym for countably infinite.

(ii) We will say a set is countable if it either has a finite number of elements or is countably infinite.

Some other EXAMPLES of countably infinite sets are:

(1) The set of positive integers $\{0,1,2,\ldots,n,\ldots\}$, to see this consider the function

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 \ldots & n & \ldots. \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 0 & 1 & 2 & 3 \ldots & n-1 & \ldots. \end{array}$$

(2) The set of integers $\mathbb{Z} = \{\ldots,-n,\ldots,-2,-1,0,1,2,\ldots,n,\ldots\}$.

Consider the function

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 \ldots & 2n & 2n+1 \ldots. \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ 0 & 1 & -1 & 2 & -2 \ldots & n & -n \quad \ldots. \end{array}$$

(3) The set of even natural numbers

Consider

$$\begin{array}{cccccc} 1 & 2 & 3 & \dots & n & \dots \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \\ 2 & 4 & 6 & \dots & 2n & \dots . \end{array}$$

As an EXERCISE you should convince yourself that in each case the suggested mapping is both one-to-one and onto. The details for (3) are as follows.

We have the mapping $f: \mathbb{N} \to \{2,4,6,\dots\}: n \mapsto 2n$. Now if $f(n) = f(m)$ we have $2n = 2m$ or $n = m$, so f is one-to-one. To show f is onto we must show that every even number is the image of some element of $\mathbb{N}$ under f. If m is any even integer, then $m/2$ is a whole number, that is $m/2 \in \mathbb{N}$ and $f(m/2) = m$, so f is onto.

Convention

In (1) we classified 0 as positive. In doing this we adopted the convention that

"x is positive" means "$x \geq 0$".

Correspondingly, "$x > 0$" would be expressed in words as "x is strictly positive". We shall use this convention also in the case when x is a real number.

There is of course the possibility that we interpret "x is positive" to mean "$x > 0$". If we do this we can express "$x \geq 0$" in words as "x is non-negative" or "x is positive in the wide sense". You should be prepared to find either convention in use in books you consult. Of course, we can often be offhand about the distinction but not always.

8. Some countably infinite sets

If we consider a set whose definition is more complicated than (1),(2), (3) above and suspect it to be countably infinite, it may be difficult to find a convenient one-one and onto function. The following specialisation of the Schroeder-Bernstein theorem (Appendix 1, Theorem 4) is often useful.

THEOREM 2: *The set* B *is countably infinite if there exists a one-to-one function* f *from* $\mathbb{N}$ *into* B *and a one-to-one function* B *from* B *into* $\mathbb{N}$.

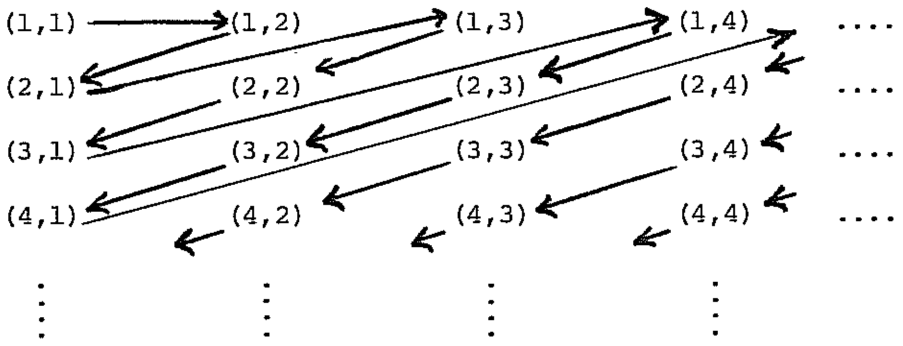
$$\mathbb{N} \xrightarrow[1-1]{f} B \xrightarrow[1-1]{g} \mathbb{N} .$$

Proof. If such an f and g exist, then by the Schroeder-Bernstein Theorem there is a one-to-one and onto function from $\mathbb{N}$ to B and so B has cardinality $\aleph_0$. □

Application of Theorem 2 is sometimes facilitated by the following lemma.
[The term 'lemma' is used to describe a mathematical result which is not of great significance in its own right and so does not merit the label "Theorem", but which is important for establishing subsequent results.]

LEMMA 3: *The set of ordered pairs of natural numbers is countably infinite.*
An ordered pair of natural numbers is (n,m) where n and $m \in \mathbb{N}$. For example $(5,3)$. The word 'ordered' means that the order in which the two numbers m and n appear in the pair is significant. Thus $(5,3)$ is not the same as $(3,5)$.

Proof. To construct a one-to-one and onto function from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$, imagine $\mathbb{N} \times \mathbb{N}$ arranged as follows



and proceed through the arrangement in the order indicated by the arrows, thus defining the function f by

$$\begin{array}{l}
1 \mapsto (1,1)\\
2 \mapsto (1,2)\\
3 \mapsto (2,1)\\
4 \mapsto (1,3)\\
5 \mapsto (2,2)\\
6 \mapsto (3,1)\\
7 \mapsto (1,4)\\
8 \mapsto (2,3)\\
\qquad\text{etc.}
\end{array}$$

Clearly f is one-to-one and, since we eventually "pass through" any given ordered pair (m,n), f is also onto.

[For an alternative proof see exercise (2), below.] □

EXERCISES:

1. (i) Show that the inverse of the function f constructed in the proof of lemma 3 is given by

   $f^{-1}: \mathbb{N} \times \mathbb{N} \to \mathbb{N}: (m,n) \mapsto \frac{1}{2}(m+n-2)(m+n-1) + m.$

   *(ii) (optional)

   Give an "algebraic" proof that the function f^{-1} (and hence f itself) is one-to-one and onto.

2. *An alternative proof of lemma 3*, which uses theorem 2.

   (i) Show that $f: \mathbb{N} \to \mathbb{N} \times \mathbb{N}: n \mapsto (n,1)$ is a one-to-one function into $\mathbb{N} \times \mathbb{N}$.

   (ii) For m, m', n, n' $\in \mathbb{N}$ show that $2^m 3^n = 2^{m'} 3^{n'}$ if and only if m = m' and n = n'.

   (iii) Using (ii) deduce that $g: \mathbb{N} \times \mathbb{N} \to \mathbb{N}: (m,n) \mapsto 2^m 3^n$ is one-to-one (but not onto).

2. (iv) Conclude from (i) and (iii) that $\#(\mathbb{N} \times \mathbb{N}) = \aleph_0$.

3. (i) Show that the set of ordered triplets of natural numbers $\mathbb{N}^3 = \mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{(m,n,p): m,n,p \in \mathbb{N}\}$ is countably infinite. [Hint: First show that the function $(m,n,p) \mapsto (f(m,n),p)$ is a one-to-one onto mapping of $\mathbb{N}^3$ to $\mathbb{N}^2$, where f is any one-to-one and onto function from $\mathbb{N}^2$ to $\mathbb{N}$.]

   *(ii) Use the principle of mathematical induction to show that the set of ordered n-tuples of natural numbers, $\mathbb{N}^n$, is countably infinite for any n $\in \mathbb{N}$.

By the set of rational numbers $\mathbb{Q}$ we mean those real numbers which may be written uniquely as p/q for some pair of integers p and q with $q > 0$ where p and q have no common factors (other than ± 1), that is the greatest common divisor of p and q is 1.

As a corollary to lemma 3 we have the following significant result.

THEOREM 4: *The set of positive rational numbers $\mathbb{Q}^+$ is countably infinite.*

Proof. Clearly the function $f: \mathbb{N} \to \mathbb{Q}^+ : n \mapsto \frac{1}{n}$ is one-to-one, as is the function

$$g : \mathbb{Q}^+ \to \mathbb{N} \times \mathbb{N} : \frac{p}{q} \mapsto (p,q).$$

Take any one-to-one and onto function h from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$. We have that $h \circ g$ is a one-to-one mapping from $\mathbb{Q}^+$ to $\mathbb{N}$.

$$\begin{tikzcd}
\mathbb{N} \arrow[r, "f"] & \mathbb{Q}^+ \arrow[r, "g"] \arrow[rr, bend right, "h^{o}g"'] & \mathbb{N} \times \mathbb{N} \arrow[r, "h"] & \mathbb{N}
\end{tikzcd}$$

The result now follows from theorem 2. □

EXERCISES:

1. Give an alternative proof to Theorem 4 along the lines of Exercise 2 above.
   [Hint: Consider $g: \mathbb{Q}^+ \to \mathbb{N}: \frac{p}{q} \mapsto 2^p 3^q$.]
2. Show that the set of all rational numbers is countably infinite.
   [Hint: Consider the type of function used in Example (2) on p.12.]

THEOREM 5: *Let* $A_1, A_2, \ldots, A_n, \ldots$ *be a countably infinite family of countably infinite sets, then their union*

$$A = A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots.$$

is countably infinite.

In Theorem 5, note that we are not given that the sets A_i are mutually disjoint. If they were we could simply apply Lemma 3, but if we avoid this assumption we must employ a proof (and notation) something like that given.

Proof. Since A_1 is countably infinite, there exists a one-to-one and onto function $f_1 : \mathbb{N} \to A_1$. Clearly f_1 provides a one-to-one function from $\mathbb{N}$ into A, since $A_1 \subseteq A$.

For each n there exists a one-to-one and onto function $f_n : \mathbb{N} \to A_n$, and thus, if $a \in A$ then $a \in A_n$ for some n and so there is a smallest n, call it n_a, with $a \in A_{n_a}$. Further since $f_{n_a} : \mathbb{N} \to A_{n_a}$ is one-to-one and onto there exists a unique natural number m_a such that $f_{n_a}(m_a) = a$ $(m_a = f_{n_a}^{-1}(a))$.

Intuitively the first appearance of a is the m_a'th element of A_{n_a}.

Define $F: A \to \mathbb{N} \times \mathbb{N}$ by $F(a) = (m_a, n_a)$, then F is a one-to-one function from A into $\mathbb{N} \times \mathbb{N}$. To see this, suppose $F(a) = F(b)$ i.e., $(m_a, n_a) = (m_b, n_b)$, then

$$n_a = n_b .$$

So $f_{n_b}(b) = f_{n_a}(b) = m_b = m_a = f_{n_a}(a)$, i.e. $f_{n_a}(a) = f_{n_a}(b)$ and since f_{n_a} is one-to-one we have $a = b$.

Now let ϕ be a one-to-one and onto mapping from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$ (lemma 3), then $g = \phi \circ F: A \to \mathbb{N}$ is a one-to-one mapping from A into $\mathbb{N}$.

The conclusion now follows by theorem 2. □

Corollary 6: *A finite union of countably infinite sets is countably infinite.*

Proof. If $A = A_1 \cup A_2 \cup \ldots \cup A_n$ where $A_1, A_2, \ldots, A_n$ are a finite number of countably infinite sets, then A may also be considered as the countable infinite union

$$A = A_1 \cup A_2 \cup \ldots \cup A_n \cup A_{n+1} \cup A_{n+2} \cup \ldots.$$

where

$$A_{n+1} = A_{n+2} = \ldots. = A_n$$

and so by theorem 5, A is countably infinite.

EXERCISES:

1. Show that the mapping $\frac{p}{q} \mapsto -\frac{p}{q}$ is a one-to-one and onto mapping from $\mathbb{Q}^+$ to $\mathbb{Q}^-$ the set of negative rational numbers. Hence, rededuce the conclusion of exercise (2) above, using Corollary 6.
2. Show that a countably infinite union of finite sets is either finite or countably infinite.

*3. A real number is said to be algebraic if it is the root of some polynomial with integer coefficients.

For example: any rational number p/q is the root of $qx - p = 0$ and so is algebraic; $\sqrt{2}$ is a root of $x^2 = 2$, so $\sqrt{2}$ is algebraic, though it is not rational.

A number which is not algebraic is said to be transcendental.

For each $m \in \mathbb{N}$ let A_m denote the set of algebraic numbers arising as the root of a polynomial of degree less than or equal to m with integer coefficients each of absolute value less than or equal to m.

i) Show that A_m is finite. [Indeed, there are at most $(2m+1)^{m+1}$ such polynomials each of which can have at most m roots (fundamental Theorem of Algebra), so $\# (A_m) \le m(2m+1)^{m+1}$]

ii) Show that $A_1 \cup A_2 \cup A_3 \cup \ldots$ is precisely the set of algebraic numbers. Hence conclude that *the set of algebraic numbers is countably infinite.*

9. Uncountable sets

DEFINITION: An *infinite* set which is not countably infinite is said to be *uncountable*. Thus an infinite set A is uncountable if and only if there does not exist a one-to-one function from $\mathbb{N}$ onto A.

(Intuitively, an uncountable set has "infinitely" more elements than there are natural numbers.)

We now establish the existence of uncountable sets. Once this is done the trivial observation that an uncountable set is never empty will be seen to have massive consequences.

THEOREM 7: *The set of real numbers between 0 and 1, that is the open interval* $(0,1) = \{x \in \mathbb{R}: \ 0 < x < 1\}$, *is uncountable.*

Proof. *Our proof is by contradiction, that is, we assume (0,1) is countably infinite and show that this leads to a contradiction, showing that our assumption must have been wrong and so (0,1) is not countably infinite and is therefore uncountable.*

Assume (0,1) is countably infinite, that is, there exists a one-to-one function f from $\mathbb{N}$ onto (0,1). Recalling that every real number in (0,1) has a unique decimal expansion provided we agree to disallow any infinite sequence of nines (otherwise, 0.5 = 0.4999... for example), for each $n \in \mathbb{N}$ let $0.a_{n_1}a_{n_2}a_{n_3}\ldots$ denote the unique decimal expansion of f(n).

Here, a_{n_1} represents the first digit in the decimal expansion of f(n)

a_{n_2} represents the second digit in the decimal expansion of f(n)

etc.,

thus $a_{n_i} \in \{0,1,2,3,4,5,6,7,8,9\}$ for $i = 1,2,3,\ldots$. We now construct a real number $r \in (0,1)$ with the property that $r \neq f(n)$ for any $n \in \mathbb{N}$.

Let $r = 0.b_1b_2b_3\ldots$

where $$b_i = \begin{cases} 3 & \text{if } a_{ii} = 5 \\ 5 & \text{if } a_{ii} \neq 5 \end{cases}$$

Thus, $$b_1 = \begin{cases} 3 & \text{if } a_{11} = 5 \\ 5 & \text{if } a_{11} \neq 5 \end{cases} , \quad \text{so } b_1 \neq a_{11} .$$

Since a_{11} is the first digit in the decimal expansion of f(1) we see that $r \neq f(1)$ as their decimal expansions differ in the first place. Similarly,

$$b_2 = \begin{cases} 3 & \text{if } a_{22} = 5 \\ 5 & \text{if } a_{22} \neq 5 \end{cases} ,$$ so $b_2 \neq a_{22}$ and we have $r \neq f(2)$ as their second digits are different.

In general, $r \neq f(n)$ for any n, as their decimal expansions differ in the n'th place.

Further, each of the digits in the decimal expansion of r are either a 3 or a 5 and so $r \in (0,1)$.

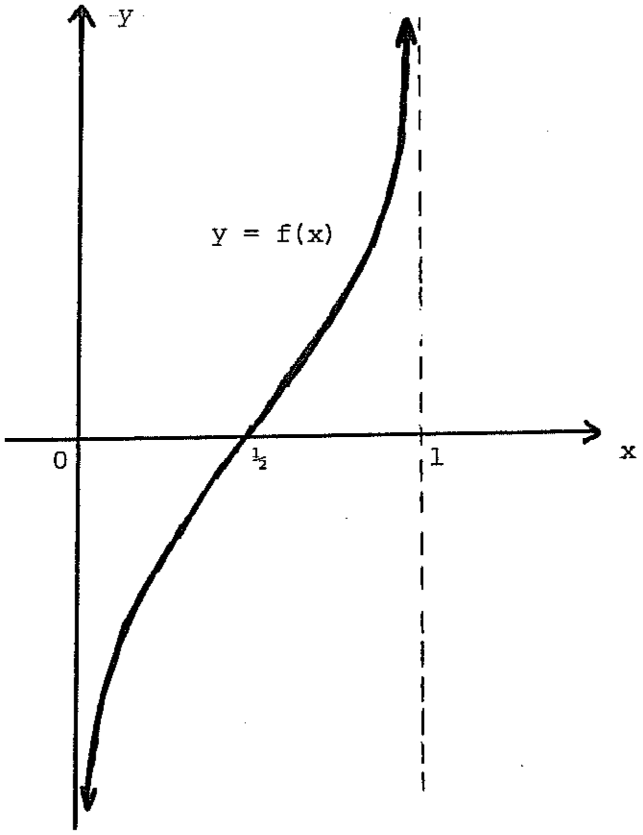
The proof is completed by observing that since $r \neq f(n)$ for any n, r is not in the range of f and so f is <u>not</u> onto, a contradiction. □

COROLLARY 8: *The set* $\mathbb{R}$ *of all real numbers is uncountable,* indeed it contains the same number of elements as the open interval (0,1).

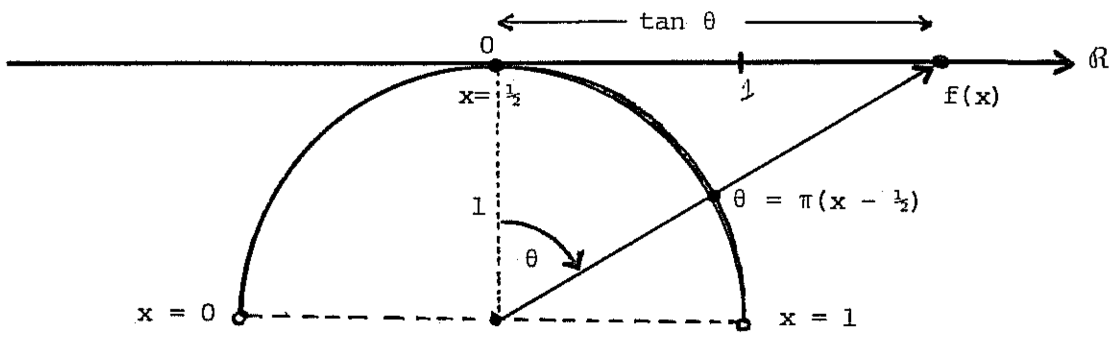
Proof. To see this we need only construct a one-to-one and onto function f from (0,1) to $\mathbb{R}$.

But, such a function is known from elementary calculus, namely

$$f(x) = \tan \pi(x - \frac{1}{2})$$



which may be interpreted geometrically according to the following diagram.



□

DEFINITION: A real number which is not a rational number (see p.6) is termed an irrational number. The existence of irrational numbers has long been known. The discovery that $\sqrt{2}$ is irrational is attributed to Pythagoras (540 B.C.), a modern proof being included in most secondary school courses. Other numbers are also irrational, for example: $\sqrt{n}$, where n is any natural number which is not a perfect square; e, the base of the natural logarithms; π.

None the less these constitute a mere handful of numbers and the proof of their irrationality becomes increasingly more difficult.

The next corollary gives a non-constructive proof of the existence of irrational numbers. That is to say, it does not identify a single number which is irrational but does show that there are "infinitely" more irrational numbers than there are rational ones. Intuitively, were we to pick a point at random on the real line it would be "infinitely more likely" to represent an irrational number than a rational one.

COROLLARY 9: *The set of irrational numbers is uncountable.*

Proof. Let I denote the set of irrational numbers, then the set of real numbers $\mathbb{R} = I \cup Q$, where Q is the countably infinite set of rational numbers (Exercise 2) on p.16 or 1) on p.18).

Now, if I were countably infinite, then by Corollary 6 we would have $\mathbb{R}$ is countably infinite, contradicting the last Corollary. Hence I is uncountable. □

A further Corollary is provided by the following result.

EXERCISE: Recalling the definition of transcendental number given in Exercise 3 on p.18, show that

The set of transcendental numbers is uncountable.

[Hint: Imitate the last proof.]

The above exercise provides a non-constructive proof of the existence of transcendental numbers and is one of the high points of Cantor's theory.

The proof that any particular number is transcendental is, as a rule, very difficult, even though our result show that "most" real numbers are indeed transcendental.

J. Liouville (in 1844) produced the first known transcendental number, namely $t = \sum_{n=0}^{\infty} 10^{-n!}$, and so proved that not all numbers are algebraic.

That e is transcendental was proved by Hermite in 1873.

Based on a considerable refinement of Hermite's argument Lindemann in 1882 was able to show that π is transcendental, and so answered one of the most famous problems in mathematics (see Appendix 2).

Gelfond established the transcendentality of e^{π} in 1929.

Today several classes of numbers have been shown to be transcendental (the work of Baker et al.), however the count of known transcendental numbers is still modest and many problems remain. For example, it appears to be unknown whether or not $e+\pi$ is transcendental (indeed, whether or not it is even irrational)!

II. MATHEMATICAL INDUCTION

10. Proofs step by step

Suppose we are easily able to prove some result when $n = 1$, and then, almost as easily, to prove it when $n = 2$, and then when $n = 3$, and so on. The occurrence of the words " and so on" is an almost certain indication that our proof, if carefully set out, would need to be by induction.

In that opening paragraph, I have intentionally been sloppy and said "to prove it when $n = 2$". In a simple case we may be sure what is meant by "it", but even in simple cases we help ourselves and readers by recognising that what we want to prove is of the form "For all $n \in \mathbb{N}$, P_n is true", and saying clearly what we mean by P_n. A typical case is that P_n denotes the proposition

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2} .$$

The *Principle of Mathematical Induction* says that if we have shown

(i) P_1 is true, and

(ii) $P_k \Rightarrow P_{k+1}$ for each $k \in \mathbb{N}$,

then P_n is true for all $n \in \mathbb{N}$. In (ii), we suppose that we can arrange a proof in which we carry k as a parameter; we are not thinking of getting (ii) by an induction. (The use of "each" $k \in \mathbb{N}$ and "all" $n \in \mathbb{N}$ above may emphasise this.) Our experience is, over and over again, that (i) is so easy as to be trivial, but that in different problems (ii) may be easy or quite hard. Nevertheless, a proof (which often amounts to a mere verification) of (i) should never be omitted, since it is a protection against errors. For instance, if P'_n is the proposition like P_n above but with the false right-hand side

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

then the step from P'_k to P'_{k+1} is just as easy as before, but P'_n is never true. Our step-by-step argument, starting from nowhere, has got nowhere.

11. The logical status of the P.M.I.

Set out as above, our principle forms a good working tool but it looks like magic, and indeed for many centuries it had been recognised as a method without being clearly understood. We can say now that during those centuries nobody had realised that what was being exploited was a property of $\mathbb{N}$, the set of natural numbers, and that this was at last recognised by Peano (1858-1932).

Suppose we want to use $\mathbb{N}$, and want the logic of our work to be clear. We shall need to pick out some simple properties from which all other properties of $\mathbb{N}$ can be deduced and either adopt these outright as axioms or deduce them from something else. For our purposes we shall adopt Peano's 5 properties as axioms, but the essential thing is to see what they are. (See the box.)

Axiom 1. 1 is a natural number.

Axiom 2. To every natural number n there corresponds just one natural number n', the successor of n.

Axiom 3. 1 is not the successor of any natural number, that is, we always have $n' \neq 1$.

Axiom 4. If $m' = n'$, then $m = n$.

Axiom 5. Suppose S is a set of natural numbers which satisfies:

(i) $1 \in S$,
(ii) $k \in S \Rightarrow k' \in S$.

Then S includes all natural numbers, that is, $S = \mathbb{N}$.

Peano's Five Axioms for $\mathbb{N}$

It is convenient, in Peano's arrangement, to have written n' here for what will later be called $n + 1$. Once we accept this, we see that Axiom 5 corresponds to the P.M.I.: we need only define S to be the subset of $\mathbb{N}$ consisting of those n for which P_n is true. (It might be good to mention here that, in Peano's arrangement, Axiom 5 is used in obtaining the standard results about $\mathbb{N}$: his Axioms 1 to 4 are not a sufficient basis for these results. A clear treatment will be found in Ch. 1 of Landau, E. *Foundations of Analysis*.)

12. Alternative forms of the P.M.I.

We should look at two variants of the P.M.I. as it was set out above. The first is almost trivial, but worth noticing; the second is more interesting and deserves attention for its use and for its logical status.

First, consider as P_n

$$\cos\alpha \cos 2\alpha \cos 4\alpha \ldots \cos 2^n\alpha = \frac{\sin 2^{n+1}\alpha}{2^{n+1}\sin\alpha} .$$

Obviously P_n is true when $n = 0$, and this would be a natural place to start from. If we insist on joining onto the form of the P.M.I. we had above, we need only write $Q_m = P_{m-1}$ and show Q_m true for all $m \in \mathbb{N}$. It is good to see how easy this is, but it is usually unnecessary even to comment about beginning our induction from some place other than 1. (This might be a good place to remark that there are two conventions about $\mathbb{N}$. With the setting-out of Peano's axioms we used, 0 does not belong to $\mathbb{N}$, but more recently, some books have adopted the convention that 0 is considered a natural number. The point is of no great importance but you should be aware of it.)

Secondly, consider as P_n the assertion that if u_n is defined by

$$u_n = u_{n-1} + u_{n-2} , \qquad n \geq 3,$$

$$u_1 = u_2 = 1,$$

then $u_n < \left(\frac{7}{4}\right)^n$. Evidently P_1 and P_2 are true, and it is easy to check that P_3 is true, but if we want to set out a clear inductive step we find we need to say

$$P_{k-1} \; \& \; P_k \Rightarrow P_{k+1} .$$

This is valid; indeed we can enunciate what looks like a greatly modified version of the P.M.I., namely

"If we have shown

(i') P_1 is true, and

(ii') $P_1 \; \& \; P_2 \; \& \ldots \& \; P_k \Rightarrow P_{k+1}$ for each $k \in \mathbb{N}$,

then P_n is true for all n."

That this is valid follows quickly from the standard version if we write

$$Q_1 = P_1 , \qquad Q_n = P_1 \; \& \; P_2 \; \& \ldots \& \; P_n .$$

We can now say that if (i') and (ii') hold for P, then (i) and (ii) hold for Q. This is because mere rewriting changes (i') and (ii') into

$$Q_1 \text{ is true} \qquad \text{and} \qquad Q_k \Rightarrow P_{k+1} ,$$

from the second of which, after the trivial remark $Q_k \Rightarrow Q_k$, we can deduce

$$Q_k \Rightarrow Q_k \;\&\; P_{k+1}, \qquad \text{that is,} \qquad Q_k \Rightarrow Q_{k+1} .$$

Our question about u_n above is quite typical of a class of problem for which we need to use the modified version of the P.M.I, but the most important applications come in algebra and number theory. There are many questions in those subjects where we need to make a step from P_h to P_{k+1}, not with h being less than $k + 1$ by 1 or 2 but with h a divisor of $k + 1$. For example, if we have defined a prime as an integer ≥ 2 which has no divisor except itself and 1, we may well hope to establish P_n for all $n \geq 2$, where P_n is the proposition

"n can be expressed as a product of primes".

We can deal with the above question as follows. Evidently

(i) P_2 is true.

Suppose $P_2,\ldots,P_k$ are true. If $k + 1$ is a prime, P_{k+1} is trivially true, whilst if

$$k + 1 = \alpha\beta ,$$

with $2 \leq \alpha < k + 1$ and $2 \leq \beta < k + 1$, we are free to appeal to P_α and P_β and say that each of α and β can be expressed as a product of primes. This of course gives us

$$k + 1 = \alpha\beta = p_1p_2\ldots p_r \cdot p_1'p_2'\ldots p_s':$$

(Remark. We have not shown that n can be expressed as a product in only one way. This requires rather different considerations.)

13. "Induction" and "Mathematical induction"

The word "induction" has been a popular one. To say nothing of its use in electromagnetism, we should notice that it is used as a philosophical term to describe the step from special to general as in: "Since the sun has risen every day for thousands of years, I conclude that it will rise to-morrow."

The reasons that "mathematical induction" is used to describe a method of deduction are of course to be found in the history of language. For us, as heirs to and sufferers from this history, it is necessary to regard "mathematical induction" as a compound technical term which, for our work here, we can often abbreviate to plain "induction". We should, though, remember that in any but a mathematical context this abbreviation might be dangerous.

14. Some applications of induction

The most impressive uses of this method come in algebra and number theory, and would require too much subsidiary material to be presented here. You have probably seen inductive proofs of the summation formulae in Exx.1 and 2 below; you may or may not have seen the results in Exx. 3,4,5.

Even if we have to forgo the best number-theoretic applications we can deal with such as:

Show that, for all $n \in \mathbb{N}$, $6^{2n-1} + 1$ *is divisible by* 7.

Write

$$u_n = 6^{2n-1} + 1,$$

and P_n for the proposition "u_n is divisible by 7". Then

(i) P_1 is true

and if P_k is true we can say

$$\begin{aligned} u_{k+1} &= u_k + (u_{k+1} - u_k) \\ &= u_k + 6^{2k-1}(6^2-1), \end{aligned}$$

which is evidently divisible by 7. Thus we have

(ii) $P_k \Rightarrow P_{k+1}$.

We deduce that P_n is true for all n. □

EXERCISES. Each of the statements is asserted for all $n \in \mathbb{N}$. Harder questions are starred.

1. $1^2 + 2^2 + 3^2 + \ldots + n^2 = \dfrac{n(n+1)(2n+1)}{3}$,

2. $\sum\limits_{r=1}^{n} r^3 = \dfrac{n^2(n+1)^2}{4}$

3. $1.2 + 2.3 + 3.4 + \ldots + n(n+1) = \dfrac{n(n+1)(n+2)}{3}$,

that is, if we write

$$r^{(2)} = r(r+1),$$

$$\sum_{r=1}^{n} r^{(2)} = \frac{n^{(3)}}{3} \quad .$$

4. Similarly, show that

$$1.2.3 + 2.3.4 + \ldots + n(n+1)(n+2) = \frac{n(n+1)(n+2)(n+3)}{4} .$$

5. More generally, if we write

$$r^{(t)} = r(r+1)(r+2)\ldots(r+t-1),$$

show that

$$\sum_{r=1}^{n} r^{(t)} = \frac{n^{(t+1)}}{t+1} .$$

[Compare the result $\int_0^x y^t dy = \frac{x^{t+1}}{t+1}$.]

6. $2^{4n-1} + 5^{n+1}$ is divisible by 11.

[Hint: In an obvious notation, express $u_{k+1} - u_k$ as the sum of multiples of 2^{4k-1} and 5^{k+1} .]

7. $11^{n+2} + 12^{2n+1}$ is divisible by 133.

8. In the U.S.S.R. the currency notes do not go 1,2,5 roubles but 1,3,5. Show that, if $n > 7$, an amount of n roubles can be paid using notes of denominations 3 and 5 only.
[Once you have thought around this question for a while you will probably be able to persuade yourself of its truth, but you might feel it would be difficult to persuade anybody else. One of the virtues of a formally set out inductive proof is that it is easy to explain to others.

Be prepared to find a case division.]

9* $\frac{4^n}{n+1} < \frac{(2n)!}{(n!)^2}$

10. $\frac{1}{1+x} + \frac{2}{1+x^2} + \frac{4}{1+x^4} + \ldots + \frac{2^n}{1+x^{2^n}} = \frac{1}{x-1} + \frac{2^{n+1}}{1-x^{2^{n+1}}} .$

11* From the 2n numbers 1,2,...,2n an arbitrary choice of n+1 numbers is made. Show that among these there is at least one pair, one of which is divisible by the other.
[Be prepared to find a division of cases.]

15. A deferred result

In Chapter 1, we said that it is impossible to have a one-one and onto mapping from $\{1,2,\ldots,n\}$ to $\{1,2,\ldots,m\}$ if $n \neq m$. We could hardly have given a proof at that stage since we had not brought properties of $\mathbb{N}$ clearly into our discussion. We have now done this. Remembering that the natural numbers are ordered

$$1 < 2 < 3 < \ldots.,$$

we can say that if $n \neq m$ then one is smaller than the other. We suppose the notation chosen so that $m < n$.

THEOREM 10

If $1 \leq m < n$, there cannot be a one-one function from $\{1,2,\ldots,n\}$ into $\{1,2,\ldots,m\}$.

Remark. We shall call the above proposition P_n. For each n, we make a finite number (in fact $n-1$) of assertions but we carry m as a parameter in our proof. There is no thought of induction on m.

LEMMA 11

If $n > 1$ there cannot be a one-one function from $\{1,2,\ldots,n\}$ into $\{1\}$.

There is only one possibility for a function into $\{1\}$, namely that with

$$f(1) = f(2) = \ldots = f(n) = 1,$$

and the equality $f(1) = f(2)$ excludes the possibility of having a one-one function. □

Proof (of Theorem 10)

(i) We start our induction from $n = 2$. There is only one assertion included in P_2, namely that with $m = 1$, and this is covered by Lemma 11. Hence P_2 is true.

(ii) Suppose P_k holds, and suppose if possible that g is a one-one function from

$$\{1,2,\ldots,k+1\} \quad \text{into} \quad \{1,2,\ldots, \ell\},$$

where $\ell < k + 1$ and, by Lemma 11, $1 < \ell$.

Case α. If $g(k + 1) = \ell$, write h for "g restricted to $\{1,2,\ldots,k\}$", that is, define $h : \{1,2,\ldots,k\} \to \{1,2,\ldots, \ell - 1\}$ by

$$h(r) = g(r) \quad \text{for} \quad 1 \leq r \leq k .$$

Then h is a one-one function and, evidently $1 \leq \ell - 1 < k$. But this is to say that h is a function having the property prohibited by P_k. It follows that there cannot be such a function as g, and that P_{k+1} holds.

Case β. If

$$g(k + 1) = r \neq \ell ,$$

define

$$\begin{aligned} \theta : \{1,2,\ldots,\ell\} &\to \{1,2,\ldots,\ell\} \text{ by} \\ \theta(r) &= \ell , \\ \theta(\ell) &= r , \\ \theta(i) &= i, \quad i \neq \ell, \quad i \neq r , \end{aligned}$$

that is, the function, evidently one-one and onto, which interchanges ℓ and r and leaves the other elements unshifted.

Define $g^*: \{1,2,\ldots,k+1\} \to \{1,2,\ldots,\ell\}$ by $h = \theta \circ g$. Then h has the properties used in Case (α), and here also we may assert that there cannot be such a function as g, and that P_{k+1} holds.

Thus, in both Case α and Case β ,

$$P_k \Rightarrow P_{k+1}$$

and the Principle of Mathematical Induction gives us that P_n holds for all n.

□

Appendix 1

FUNCTIONS

The modern definition

The word "function" is now used more generally than when it was first introduced (by Leibniz in the 17th century). But its meaning has not changed: we might better say that the word is now used without imposing distracting requirements.

The first examples that we meet are of functions defined by simple formulae such as

$$f(x) = x^2 , \qquad g(x) = \frac{x^2 + 7}{x^4 + 3}$$

where the notation suggests that we are concerned with real-valued functions of a real variable. Nothing we say about wider interpretations will change the fact that functions like this turn up often and are very important, but it will help us to take a wider view. Let me throw the modern definition at you and then make some comments.

By a function we understand two sets X, Y and a "rule" which associates with each element of X an element of Y. That is, for each element x of X the rule produces a unique* element y of Y, which we call the image of x under the function. It is usual to denote the rule by a single letter f, g, h etc., although this is by no means always the case, for example we use *sin* to denote the trigonometric function sine, *log* to denote the ordinary logarithmic function. The image of $x \in X$ under f is then denoted by f(x).

We indicate such a function by writing

$$f : X \to Y \quad (\textit{read: "f such that X goes to Y}),$$

and refer to it as a function from X to Y.

When it is necessary to specify the rule explicitly we may write

[*] In mathematical jargon "unique" is the opposite of "ambiguous". There is no suggestion of "special" or "remarkable".

$f : X \to Y : x \mapsto f(x)$ *(read: "f such that X goes to Y such that the element x is taken to the element $f(x)$.")*

For example, $f : \mathbb{R}^+ \to \mathbb{R} : x \mapsto \sqrt{x}$ indicates the function which takes positive* real numbers to real numbers by assigning to each positive real number x its positive square root.

The set X is referred to as the domain of f. The set Y is referred to as the codomain or target set of f.

When the domain and codomain are clearly understood from the context we may indicate the function by writing

$$x \mapsto f(x) \quad \text{or} \quad y = f(x).$$

REMARKS: 1) The specification of a function involves all three ingredients, the domain, the target set, and the "rule". If any one is changed a different function results. Thus, we say that

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2 ;$$

$$g : \mathbb{R} \to \mathbb{R}^+ : x \mapsto x^2 ;$$

and $$h : \mathbb{R}^+ \to \mathbb{R}^+ : x \mapsto x^2$$

are three different functions (with different properties, as we shall see shortly) even though they all share the same rule.

2) While it is frequently the case in very elementary mathematics, there is no requirement that the domain and target set be subsets of the real numbers. (A restatement of the remarks above!) For example:

(i) Let F be the set of all real valued functions of a real variable and let $\mathcal{D}$ denote the subset of differentiable functions, then the operation of differentiation.

$$D : f \mapsto f'$$

defines a function from $\mathcal{D}$ to F.

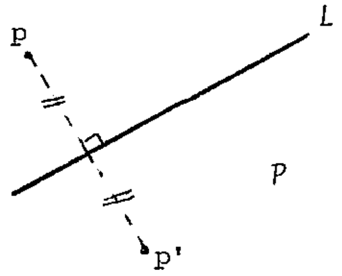
Similarly, for any fixed real number r_0,

*Remember that we are following the convention that "y is positive" means "$y \geq 0$" and "y is strictly positive" means "$y > 0$." (Compare Page 13)

$$E_{r_0}: \mathcal{F} \to \mathcal{R}: f \mapsto f(r_0)$$

defines a function from $\mathcal{F}$ to $\mathcal{R}$ which associates with each function $f \in \mathcal{F}$ its value at r_0.

(ii) Let $\mathcal{P}$ denote the set of all points in the plane and let L be a given line in the plane, then



$$R: \mathcal{P} \to \mathcal{P}: p \mapsto p'$$

is a function from $\mathcal{P}$ to $\mathcal{P}$ which takes each point to its "reflection" in L.

3) An important change in point of view is concealed in the statement that we "denote the rule by a single letter". If, as in many books on Calculus, there is talk of "the function $f(x)$" we should interpret this to mean "the function f defined on some domain whose typical member is called x". You may see the reasonableness of this point of view if you notice that the formulae

$$f(x) = \frac{x^4 + e^x}{1 + x^2} \quad \text{and} \quad f(t) = \frac{t^4 + e^t}{1 + t^2}$$

define the same rule, which we call f.

You might look again at the definition of E_{r_0} in 2)(i) above, and make sure you see what is going on.

4) Note the use of the ordinary and tailed arrows

$$\to \quad \text{and} \quad \mapsto$$

in the above. The first is used between the domain and the target set whilst the second is used between typical members of these two sets. You will not find them in all books, but they are being used more and more widely and you should regard them as the modern standard. (The untailed arrow also has the important meaning "tends to" but the two meanings do not cause confusion.)

Another useful word is the range (or image set) of a function $f : X \to Y$, the set of all images of points in X under f. We shall denote the range of f by $f(X)$. Obviously (being only an exercise on our notations)

$$f(X) \subseteq Y ,$$

but it is very likely that $f(X)$ is a proper subset of Y. For example, if as in Remark 1) we write

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2,$$

then $9 \in f(\mathbb{R})$, since 9 is the image of 3 (and of -3) under f. However, -1 is not in the range of f (although it is in the target set) since no real number has -1 as its square. Indeed, here

$$f(\mathbb{R}) = \mathbb{R}^+.$$

2. Properties of functions

We wanted to define the word "function" in the most general way. Having done so, we shall distinguish functions which have convenient special properties.

(i) A function $f : X \to Y$ is said to be onto if $f(X) = Y$.

For example, the function

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$$

is not onto but the function

$$g : \mathbb{R} \to \mathbb{R}^+ : x \mapsto x^2$$

has this property.

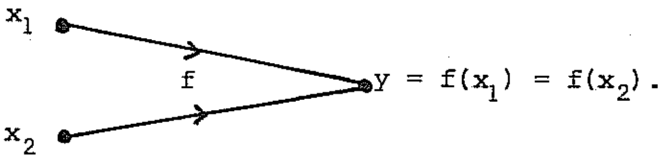
(ii) A function $f : X \to Y$ is said to be one-to-one (sometimes written 1-1) if no two distinct points in the domain have the same image under f. We can write this requirement as

$$(\alpha) \quad \text{if } x_1 \neq x_2 \text{, then } f(x_1) \neq f(x_2)$$

or, equivalently, as

$$(\beta) \quad \text{if } x_1 \text{ and } x_2 \text{ have } f(x_1) = f(x_2) \text{, then } x_1 = x_2 .$$

In diagrammatic form, the following situation does not occur:



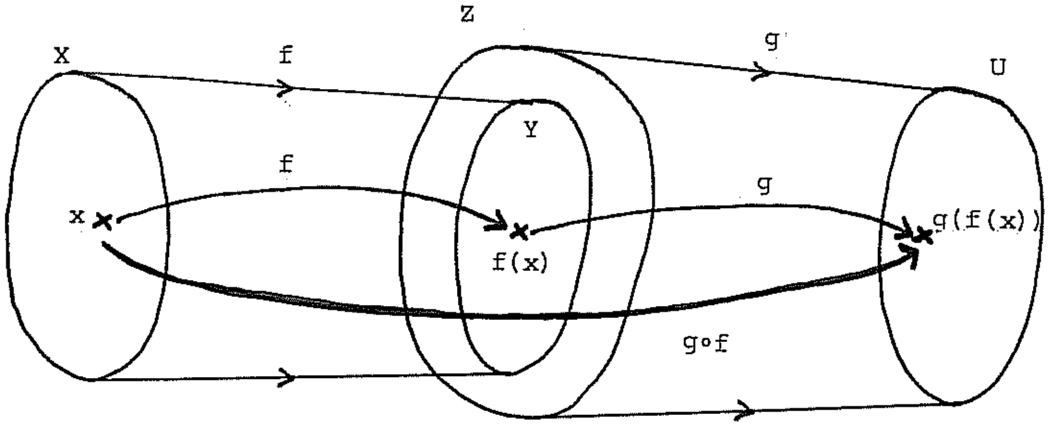
For example, $g: \mathbb{R} \to \mathbb{R}^+ : x \mapsto x^2$ is not one-to-one since $g(-1) = g(1)$. On the other hand $h: \mathbb{R}^+ \to \mathbb{R}^+ : x \mapsto x^2$ is one-to-one.(Why?)

(iii)For functions $f: X \to Y$ and $g: Z \to U$ where $Y \subseteq Z$ we may define the composite function $g \circ f$ (read "g of f") by

$$g \circ f: X \to U: x \mapsto g(f(x)).$$

(Note: $g \circ f$ is sometimes written as simply gf.)

Diagrammatically we have:



For example, let $f\colon \mathbb{R} \to \mathbb{R}^+\colon x \mapsto x^2$ and let $g\colon \mathbb{R}^+ \to \mathbb{R}\colon x \mapsto x+1$, then $g \circ f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2+1$. This same example shows that the operation of forming composites need not be commutative, that is $g \circ f$ need not in general equal $f \circ g$. In the above example, $f \circ g\colon \mathbb{R}^+ \to \mathbb{R}^+\colon x \mapsto (x+1)^2 = x^2 + 1 + 2x$ and so $f \circ g \neq g \circ f$ (indeed, not only are their rules different, so are their target sets.)

In fact it frequently happens that $f \circ g$ is undefined even when $g \circ f$ is defined. For example, let $f\colon \mathbb{R}^+ \to \mathbb{R}\colon x \mapsto \sqrt{x}$ and let

$g\colon \mathbb{R}^+ \to \mathbb{R}\colon x \mapsto -x$ then

$g \circ f\colon \mathbb{R}^+ \to \mathbb{R}\colon x \mapsto -\sqrt{x}$

while $f \circ g$ is undefined as for any $x \in \mathbb{R}^+$, $\sqrt{-x}$ is not a real number.

EXERCISE: Let $f\colon X \to Y$ and $g\colon Z \to U$ be such that $g \circ f$ is defined (that is $Y \subseteq Z$), show that

i) $g \circ f$ is onto if both f and g are onto and $Y = Z$.

and ii) $g \circ f$ is one-to-one if both f and g are one-to-one.

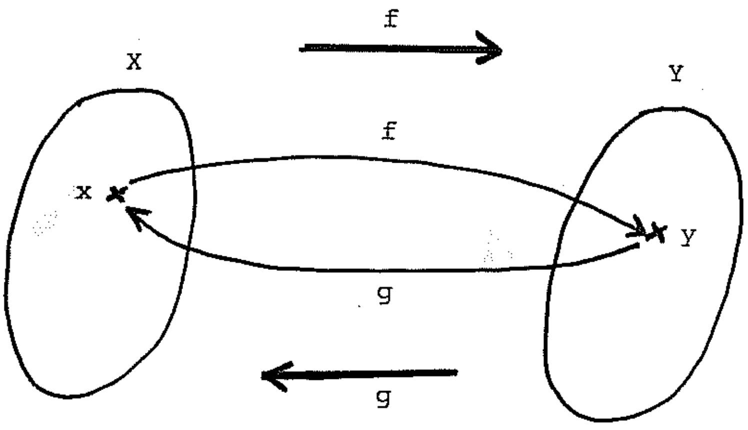
3. Invertible functions

A function $f\colon X \to Y$ is said to be <u>invertible</u> if there exists another function $g\colon Y \to X$ such that

$$g \circ f(x) = x \quad \text{for every } x \in X$$

and

$$f \circ g(y) = y \quad \text{for every } y \in Y.$$



For example, $f: \mathbb{R}^+ \to (0,1]: x \mapsto \frac{1}{x+1}$ is invertible (take $g: (0,1] \to \mathbb{R}^+: x \mapsto \frac{1}{x} - 1)$.

LEMMA 1: *If* $f: X \to Y$ *is invertible then there is a UNIQUE function* (that is, *there is only one function*) $g : Y \to X$ *with*

$$g \circ f(x) = x \quad \textit{for every } x \in X$$

and

$$f \circ g(y) = y \quad \textit{for every } y \in Y .$$

Proof. Assume that $g^*: Y \to X$ is also such that

$$g^* \circ f(x) = x \quad \text{for every } x \in X$$

and

$$f \circ g^*(y) = y \quad \text{for every } y \in Y,$$

then for every $y \in Y$ we have

$$\begin{aligned} g^*(y) &= g^*(f(g(y))), \quad \text{as } f \circ g(y) = y \text{ or } f(g(y)) = y. \\ &= g^* \circ f(g(y)), \quad \text{by the definition of "}\circ\text{".} \\ &= g(y), \quad \text{as } g^* \circ f(x) = x \quad \text{for every } x \in X \text{ and } g(y) \in X. \end{aligned}$$

Thus

$$g^*(y) = g(y) \quad \text{for every } y \in Y, \quad \text{or } g^* = g.$$

□

If $f: X \to Y$ is invertible the unique $g: Y \to X$ satisfying

$$g \circ f(x) = x \quad \text{for all } x \in X$$

and

$$f \circ g(x) = y \quad \text{for all } y \in Y$$

is termed the <u>inverse</u> of f and is usually denoted by f^{-1}. Thus $f^{-1}: Y \to X$ is the unique function such that

$$f^{-1} \circ f(x) = x \quad \text{for all } x \in X$$

and $f \circ f^{-1}(y) = y$ for all $y \in Y$.

EXERCISE: 1) If $f: X \to Y$ is invertible show that f^{-1} is invertible with $(f^{-1})^{-1} = f$.

2) If $f: X \to Y$ and $g: Y \to Z$ are invertible, noting that $g \circ f$ is defined, show that $g \circ f$ is invertible with $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Theorem 2: $f : X \to Y$ *is invertible if and only if* f *is both one-to-one and onto.*

NOTE that the occurrence of "if and only if" in the enunciation shows us that we must construct two proofs, that is, one in each direction. A convenient piece of shorthand is to write ($\Rightarrow$) and ($\Leftarrow$) respectively as labels on the two proofs. (Exercise. Check that this shorthand does make sense!)

Proof. ($\Rightarrow$) Since f^{-1} exists, if $x_1, x_2 \in X$ are such that $f(x_1) = f(x_2)$, we have

$$x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2 ,$$

so f is one-to-one.

To show f is onto, we must establish that each $y \in Y$ is the image of some $x \in X$ under f. Thus, given $y \in Y$, let $x = f^{-1}(y)$, then $f(x) = f(f^{-1}(y)) = y$, so f is onto.

($\Leftarrow$) Since f is onto, given any $y \in Y$ there exists an $x \in X$ with $f(x) = y$. Further this x is unique, otherwise there would be two points $x_1, x_2 \in X$ with $x_1 \neq x_2$ and $y = f(x_1) = f(x_2)$ which is impossible as f is one-to-one.

For each $y \in Y$ let us denote by x_y this unique point in X for which $y = f(x_y)$.

Define $g: Y \to X$ by $g(y) = x_y$ for each $y \in Y$.

Then, $f\circ g(y) = f(g(y)) = f(x_y) = y$ for all $y \in Y$. Further, for each $x \in X$ $g\circ f(x) = g(f(x)) = x_{f(x)}$ the <u>unique</u> point in X whose image under f is $f(x)$. Clearly x itself is such a point, hence, by the uniqueness, $x_{f(x)} = x$ and so $g\circ f(x) = x$ for every $x \in X$.

It therefore follows from the definition that f is invertible. □

COROLLARY 3: *If* f *is a one-to-one and onto function from* X *to* Y*, then* f^{-1} *exists and is a one-to-one and onto function from* Y *to* X.

EXERCISE: From Exercise 2) above deduce that if $f: X \to Y$ and $g: Y \to Z$ are both one-to-one and onto functions, then $g \circ f$ is a one-to-one and onto function.

Note: From this theorem we readily see that the double condition

$$g \circ f(x) = x \quad \text{for all } x \in X$$

and

$$f \circ g(y) = y \quad \text{for all } y \in Y$$

in the definition of invertibility is necessary. The function $f: \mathbb{R} \to \mathbb{R}^+: x \mapsto x^2$ is onto, but not one-to-one, hence f is not invertible. None the less the function $g: \mathbb{R}^+ \to \mathbb{R}: y \mapsto \sqrt{y}$ is such that $f \circ g(y) = y$ for all $y \in \mathbb{R}^+$ and so one of the above two conditions on f is satified!

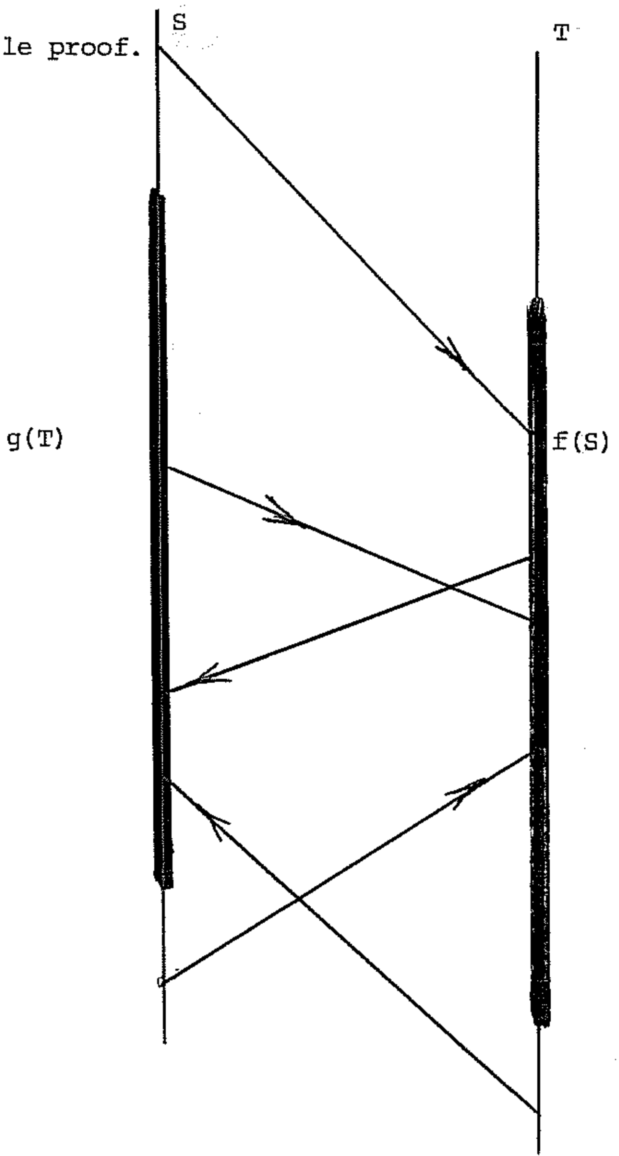
REMARK. The word *bijection* to mean a function which is both one-to-one and onto will be quite often met with. This is one of the new words introduced by Bourbaki, the pseudonym of a group of French authors. The words *injection* (for a one-to-one function) and *surjection* (for an onto function) come from the same source but have not been as widely adopted.

4. The theorem of Schroeder and Bernstein

The result of this section is important when we want to establish the existence of an invertible function between two given sets. Note that there is no suggestion that there is a unique F with the asserted properties, and this for two reasons. First, starting from f and g as here, we might think of some other procedure for constructing an F. Secondly, we might start from another pair of functions f_1 and g_1, which had the properties required of f and g.

When I talk of using f and g(or f_1 and g_1) I should point out a feature of the enunciation of the theorem, one that arises elsewhere. We suppose S and T to be such that there exist an f and a g with certain properties. In any application we start with the sets S and T, and look for two such functions. In fact, the rewording "Suppose that S and T are two sets and *that we have found* functions f and g with properties ...." might seem better. (But what has been used here is probably more typical of what you will find in books.)

The idea of the following proof is simple, and well indicated in Figs. 1,2,3. However, it is necessary to introduce notation which may give you trouble at first. If it does, work Exercises 1 and 2 before going through the whole proof.



The sets S and T are indicated by the vertical lines; the thickened parts denote g(T) and f(S) respectively.

The arrows going from left to right show the action of f on typical points; those from right to left the action of g. (Note that arrows start from every point of the lines; they can only end inside the thickened parts.)

Fig. 1. The effects of f and g

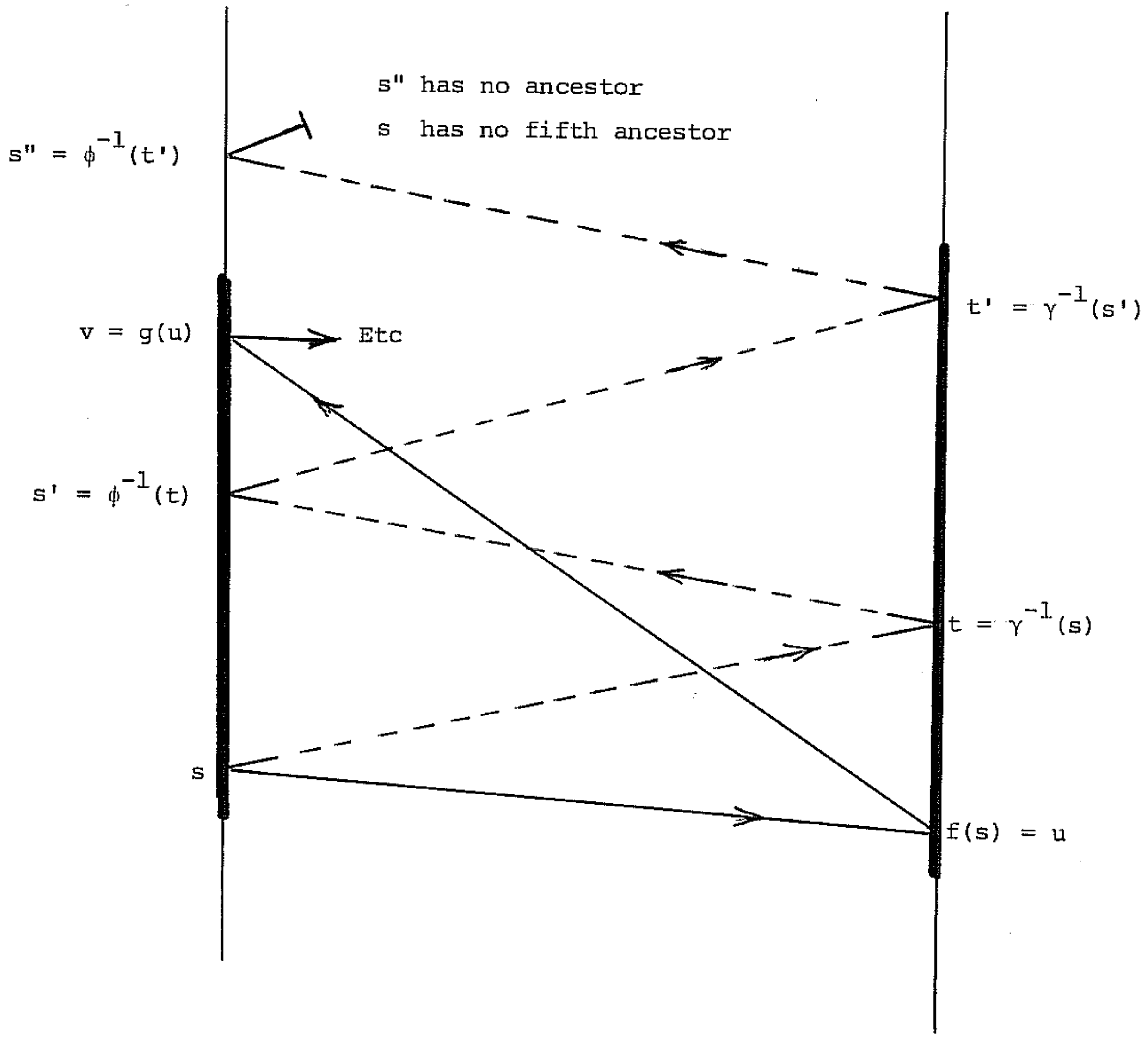
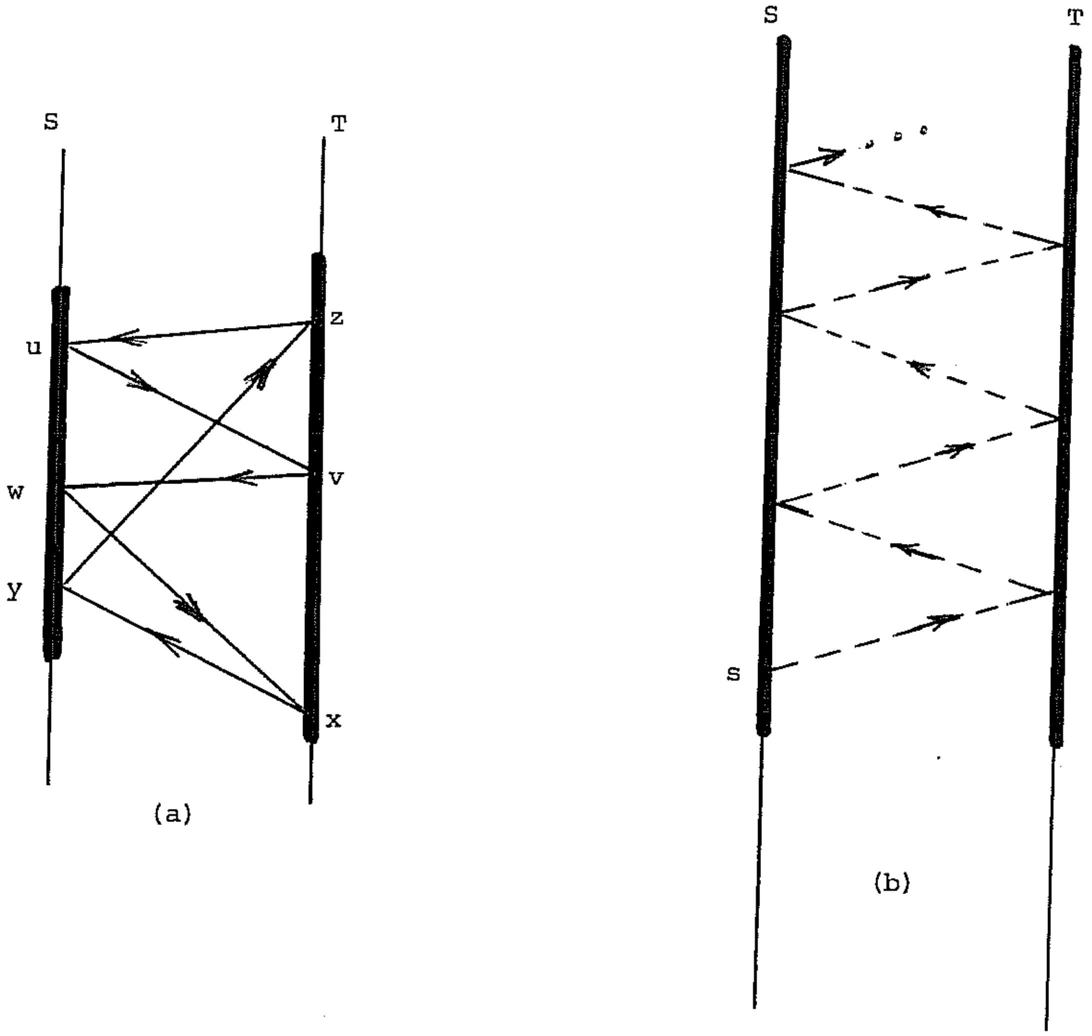


Fig. 2. The ancestors of a point s belonging to S_4

Dotted arrows from left to right show the action of γ^{-1}, those from right to left the action of ϕ^{-1}.

In this figure: $s \in S_4$, $t \in T_3$, $s' \in S_2$, $t' \in T_1$ and $s" \in S_0$. If we follow descendants of s, we have $u \in T_5$, $v \in S_6$, etc. (The "Etc." on the figure indicates these descendants.)



(a)

(b)

Fig. 3. Two ways in which points can belong to S_∞ (and corresponding points to T_∞)

In Fig. 3(a), the arrows u to v to w to x to y to z to u form a closed cycle. (The "descendants" of an element are also "ancestors"!) We have u, w, y in S_∞ and v,x,z in T_∞ .

In Fig. 3(b), the indication is that the thickened lines extend upward indefinitely and the arrows showing the actions of γ^{-1} and ϕ^{-1} continue upward indefinitely.

THEOREM 4. (Schroeder-Bernstein). *If* S *and* T *are two sets such that there exists a one-to-one function* $f : S \to T$ *and a one-to-one function* $g : T \to S$, *then there exists a one-to-one and onto function* $F : S \to T$.

Proof (due to Garrett Birkhoff*). We may assume that neither f nor g is onto (since if f is onto we need only take $F = f$ and if g is onto need only take $F = g^{-1}$).

We organise the work by talking about "ancestors" of points. For any $s \in S$ we ask whether

$$(\alpha) \quad s \in g(T) \quad \text{or} \quad (\beta) \quad s \notin g(T).$$

If (α) holds, there is a t, indeed, a unique t, such that

$$g(t) = s,$$

and we call t the first *ancestor* of s. If (β) holds, we can say s has no first ancestor.

We write

$$S_0 = \{s \in S : s \text{ has no first ancestor}\} ,$$

that is

$$S_0 = S - g(T).$$

Similarly we can define

$$T_0 = T - f(S),$$

as the set of points in T which have no first ancestor.

We are especially interested in points which have a first ancestor and among these we aim to pick out those which have higher ancestors. We can help ourselves by defining**

$$\phi : S \to f(S) : s \mapsto f(s)$$

and

$$\gamma : T \to g(T(: t \mapsto g(t),$$

* We need to give the Christian name to distinguish Garrett (1911-) from his father George David (1884-1944).

** The introduction of ϕ and γ may seem very fussy, and we might say that we could write f^{-1} and g^{-1} and remember where these were usable and where not. Indeed we could, but it is easy to get muddled. Experience this century (some of it with very subtle work) has shown that the fussiness pays.

two functions for each of which the target set and range agree. Since each is 1-1 (Check!) we are justified in using the notations ϕ^{-1} and γ^{-1} in labelling the figures. If $s \in g(T)$, so that it has a first ancestor t, we ask whether t has an ancestor. If so, we call it the *second ancestor* of s, and so on.

Now we define

$$S_1 = \{s \in S: \ s \text{ has a first ancestor but no second}\}$$

$$S_2 = \{s \in S: \ s \text{ has a second ancestor but no third}\},$$

and so on, and similarly we define T_1, T_2 etc. Also we define

$$S_\infty = \{s \in S : \ s \text{ has an infinity of ancestors}\}$$

and,similarly, T_∞. [Note that our metaphor may break down here: see Fig 3(a).]

This classification by number of ancestors leads us to the equations

$$S = S_\infty \cup S_0 \cup S_1 \cup S_2 \cup \ldots.. \qquad (1)$$

and

$$T = T_\infty \cup T_0 \cup T_1 \cup T_2 \cup \ldots.. , \qquad (2)$$

and to the result that in each of the equations the sets on its right-hand side are disjoint. Further, we can make the essential remarks that

$$f(S_\infty) = T_\infty , \qquad g(T_\infty) = S_\infty \qquad (3)$$

and, for $n = 0,1,2,\ldots$

$$f(S_n) = T_{n+1} \qquad g(T_n) = S_{n+1} . \qquad (4)$$

Now write

$$H = S_0 \cup S_2 \cup S_4 \cup \ldots. ,$$

the subset of S consisting of points having an even number of ancestors. (Note that although 0 is an even number, ∞ is not a number, even or odd, but a useful piece of shorthand.) Similarly, write

$$K = T_0 \cup T_2 \cup T_4 \cup \ldots. .$$

Evidently we obtain, from (4),

$$f(H) = T_1 \cup T_3 \cup T_5 \cup \ldots. .$$

and

$$g(K) = S_1 \cup S_3 \cup S_5 \cup \ldots\ .$$

We can now write (1) and (2) much more briefly as

$$S = S_\infty \cup H \cup g(K), \qquad T = T_\infty \cup f(H) \cup K\ ,$$

where in each of the equations the sets on its right-hand side are disjoint.

Finally, define $F : S \to T$ by

$$F(s) = \begin{cases} f(s) & \text{if} \quad s \in S_\infty \cup H, \\ \gamma^{-1}(s) & \text{if} \quad s \in g(K) \end{cases}$$

Then, as required, F is 1-1 and onto. (Check!) □

EXERCISES: (1) If

$$S = \{0, 1, 2, 4, 5, 6, 8, \ldots, 4n, 4n+1, 4n+2, 4n+4, \ldots.\}$$

and

$$T = \{0, 2, 3, 4, 6, 7, 8, \ldots, 4m, 4m+2, 4m+3, 4m+4, \ldots.\}\ ,$$

we take

$$f(s) = 2s, \qquad g(t) = 2t.$$

Find S_0, T_0, S_1, T_1 and show that

$$S_\infty = T_\infty = \{0\}.$$

Remarks (i) Obviously we do not need to appeal to the Schroeder-Bernstein theorem if we want to set up a bijection from S to T. The point of this exercise, and of Ex. 2., is to help you sort out the notation.

(ii) Recall that $\{0\}$ is the set with just one member, the number 0. It is not the empty set.

(2) If $S = T$, the set of all integers, and

$$f(s) = s + 1, \qquad g(t) = t + 2\ ,$$

show that $S_\infty = T_\infty = S$.

(3) Show that if

$$S = [0,1] \quad \text{and} \quad T = [0,1),$$

the closed and half-open intervals respectively, and

$$f(s) = \frac{1}{2}s, \quad g(t) = t$$

show that

$$S_0 = \{1\}, \qquad T_0 = (\tfrac{1}{2}, 1),$$

$$S_1 = (\tfrac{1}{2}, 1), \qquad T_1 = \{\tfrac{1}{2}\}$$

and that

$$S_\infty = T_\infty = \{0\}.$$

(4) In the notation of Ex. 3, determine F, as defined in our theorem. (You will need to use a broken definition, that is, you can define F by using formulae but there will be different formulae on different subsets.) Illustrate with a diagram.

(5) Show that if, in the notation of the theorem, we define $G : T \to S$ by

$$G(t) = \begin{cases} g(t) & \text{if } t \in T_\infty \cup K, \\ \phi^{-1}(t) & \text{if } t \in f(H), \end{cases}$$

then G is a bijection from T to S and that F^*, defined as G^{-1}, satisfies the requirements of the theorem.

Appendix 2

CONSTRUCTION PROBLEMS

1. The geometrical background

To understand what is being attempted here, and why, you need to know certain points of mathematical history.

Greek geometry, as typified by Euclid's textbook "The Elements", is concerned first and foremost with straight lines and circles. Correspondingly they preferred constructions which used only a straight edge and a compass, the straight edge being ungraduated. You have probably seen constructions, satisfying these requirements, for the bisection of a given angle and for the construction of perpendiculars. There are many others, one of the most interesting being that of a regular pentagon. But, in spite of many attempts in Greek times and later to find constructions (subject to the restrictions) for

(i) trisection of a given angle,

(ii) duplication of the cube,

(iii) squaring of the circle,

none was successful. The suspicion gradually arose that these constructions are impossible if we observe the restrictions. However, as you can see, to show that no construction is possible we must find a way of surveying all permissible constructions. This was achieved during the 19th century.

The key was to classify the "constructible numbers" and to show that they can all be found by solving chains of quadratic equations and that all are expressible in a form involving nested square root signs. (Cf. page A21 .) If now we can show that the solution of one of our problems would determine a length (in terms of a convenient unit) which is not of this form, we can assert that the problem cannot be solved.

What we can do here is to show that the constructible numbers can be expressed as finite sums of numbers using nested square root signs and how the transcendentality of π excludes the solution of (iii). But of course, we have not shown that π is transcendental.

It might be useful to make some comments on (i) and (ii). If we could solve these, our construction would determine a length given by a cubic equation and (except for the special angles 90°, 60°, 45° etc.) the cubic equation is one which cannot be solved by using a chain of quadratic equations. The proof of this requires a long discussion, of a quite different character from our work here. We must leave this for the future.

Another construction problem or, rather, class of problems which the Greeks (and their successors in modern times) looked at was the extension of the method of constructing a regular pentagon to the construction of a regular n-gon, for any value of n. This does not seem to have captivated imaginations as much as our (i), (ii), (iii) and since nobody had any success with the cases $n = 7, 9$, it was not much talked about. Hence, the discovery by Gauss that the case $n = 17$ is tractable came as a tremendous surprise. To achieve this positive result, it is necessary to show that the equation

$$\sum_{n=0}^{16} z^n = 0$$

can be solved by using a chain of quadratic equations. The calculations are not difficult, but it is impossible to motivate them unless we use some facts from number theory. Hence this too must be left for the future.

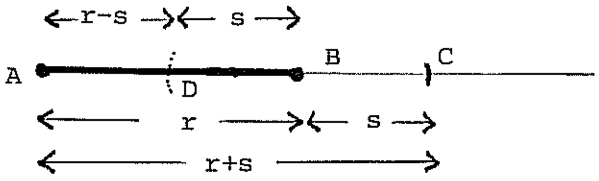
2. Constructible numbers

A number r is said to be constructible (in the Greek sense) if, starting with a given line segment, nominally of length one, it is possible to construct a line segment of length r in a finite number of precise steps using only a straight edge and a compass.

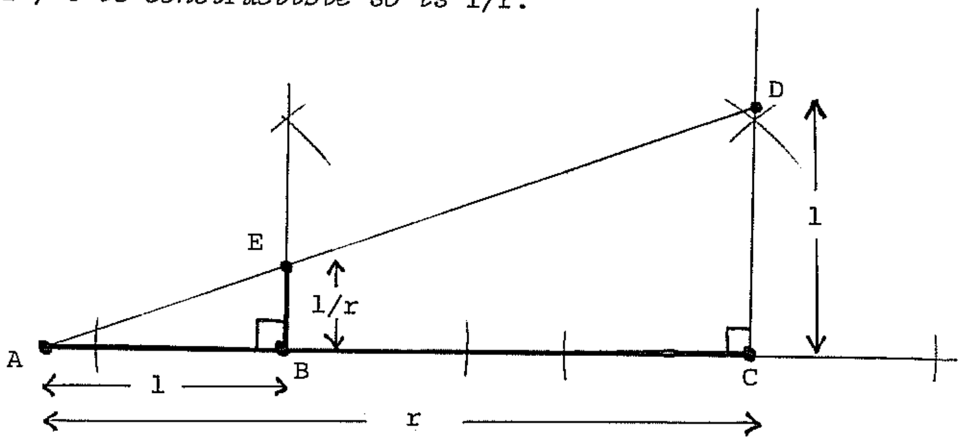
We have the following results.

1) *If* r *and* s *are constructible numbers so are* r+s *and* r-s *(provided* r > s*)* .

Let AB be a line segment of length r. Extend AB to the right. Open the compass to a width s and using it locate the point C on the extension of AB with BC = s. AC is of length r+s (the construction of r-s is similar).



2) *If* r ≠ 0 *is constructible so is* 1/r.



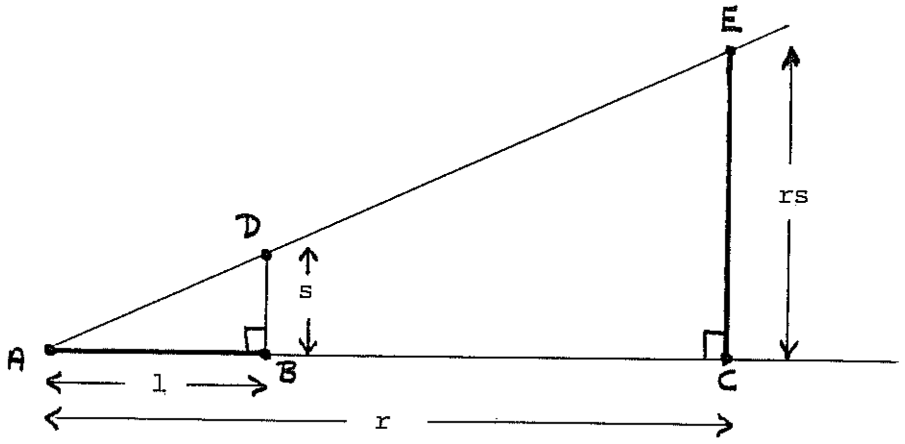
Let A, B, C be marked on a line as shown with AB = 1, AC = r. Construct perpendiculars to the line at B and C. Using the compass, open to width AB, locate the point D on the perpendicular at C such that CD = 1, join DA. Let E be the point of intersection of DA with the perpendicular at B, then triangle AEB and ADC are similar so AB:AC = EB:CD

or $1/r = EB/1$,

that is EB is the required segment of length 1/r.

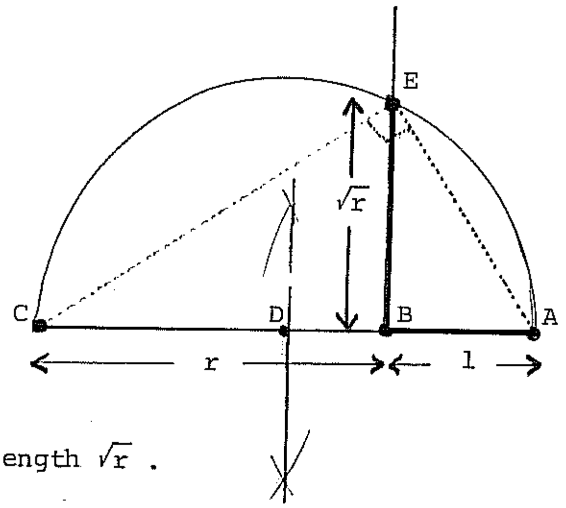
3) *If* r, s *are both constructible numbers, then their product* rs *is constructible.*

The construction is similar to that given in 2 and may be found in the following diagram.



4) *If* r *is constructible so is* $\sqrt{r}$.

Locate the three points A, B, C on a line such that AB = 1 and BC = r. Bisect the line to find D. With D as centre draw a circle of radius AD (= DC). Construct the perpendicular to AC at B and let it intersect the circle at E. Then BE is of length $\sqrt{r}$.



EXERCISE: Prove that this construction works.

COROLLARIES

5) *Any natural number* n *is constructible.*

Use the construction given in 1 a sufficient number of times to construct

$$\underbrace{1 + 1 + 1 + \ldots + 1}_{\text{n times}} = n .$$

6) *Any positive rational number* $\frac{p}{q}$ *is constructible.*

Construct p and q by 5). Using 2) construct $\frac{1}{q}$, then using 3) construct $p \times \frac{1}{q} = \frac{p}{q}$.

7) *Any number which is a finite sum of numbers of the form*

$$\frac{\pm n_1 \pm \sqrt{n_2 \pm \sqrt{n_3 \pm \sqrt{n_4 \pm \ldots \pm \sqrt{n_{k-1} \pm \sqrt{n_k}}}}}}{m}$$

is constructible, where m, n_1, n_2, ..., n_k *are positive integers, provided none of the subtractions leads to the square root of a negative no.*

The construction is done using successively: 4) to find $\sqrt{n_k}$; 1) to find $n_{k-1} \pm \sqrt{n_k}$; 4) to find $\sqrt{n_{k-1} \pm \sqrt{n_k}}$ etc; 2) and 3) to divide by m. After each term has been constructed in this way their sum is obtained from repeated application of 1.

The purpose of what follows is to show that:

8) *The only numbers which are constructible are of the form given in 7).*

9) *Any number of the form given in 7) is the root of some polynomial with integer coefficients.*

and hence

10) *Every constructible number is an algebraic number.*

From this and the Exercise on p.12 we then deduce

11) *The set of constructible numbers is countably infinite and so the set of non-constructible numbers is uncountable.* In particular, non-constructible numbers exist.

10) and 11) follow readily from 9) and so are left as EXERCISES.

9) itself is an easy EXERCISE. [As a hint consider the case where

$$r = \frac{n_1 - \sqrt{n_2 + \sqrt{n_3}}}{m} .$$

Then, $mr - n_1 = -\sqrt{n_2 + \sqrt{n_3}}$ squaring gives

$$(mr - n_1)^2 - n_2 = \sqrt{n_3} ,$$

squaring again yields

$$((mr - n_1)^2 - n_2)^2 - n_3 = 0$$

or

$$(m^4)r^4 - (4m^3n_1)r^3 + (6m^2n_1^2 - 2m^2n_2)r^2 - (4mn_1(n_1^2 - n_2))r + ((n_1^2 - n_2)^2 - n_3) = 0.$$

So we see r is the root of a fourth degree polynomial, whose coefficients are all integers as m, n_1, n_2 and n_3 are integers.]

We therefore concentrate on 8).

Let C denote the set of all real numbers of the form given in 7) - such numbers are sometimes called *quadratic surds*.

EXERCISE: A) For any r, s ϵ C show that

i) $r + s \in C$ and, provided $r > s$, $r-s \in C$,

ii) $rs \in C$,

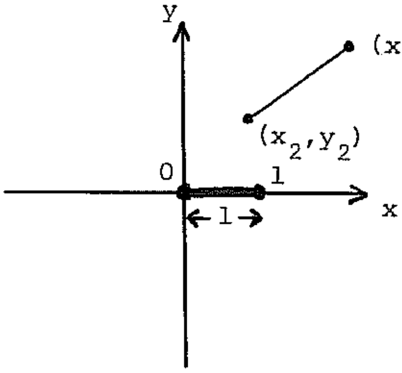
iii) provided $s \neq 0$ $\frac{r}{s} \in C$ [Hint: "rationalize" the denominator, and use ii)],

and iv) $\sqrt{r} \in C$.

That is C *is closed under the operations of* +, -, ×, ÷ *and* $\sqrt{\ }$.

3.Constructible points

Choose a set of rectangular coordinate axes in the plane such that the line between (0,0) and (1,0) is the given line of unit length.



where x_i, $y_i \in C$.

Say a point in the plane is *constructible* if it can be located, using straight-edge and compass, in a finite number of precise steps starting from the two points (0,0) and (1,0).

Clearly a point is constructible if and only if both its x and y coordinates are constructible numbers.

Let $\mathcal{P}_{\mathcal{C}}$ denote the set of points whose x and y coordinates are in $\mathcal{C}$. By 7) the points in $\mathcal{P}_{\mathcal{C}}$ are all constructible.

Further, since (0,0) and (1,0) $\in \mathcal{P}_{\mathcal{C}}$ it is possible to locate any constructible point, using straight-edge and compass, in a finite number of precise steps starting from the points in $\mathcal{P}_{\mathcal{C}}$.

The first step in such a construction will be the location of a new point as either

a) The intersection of two lines, both of which are constructed by joining points in $\mathcal{P}_{\mathcal{C}}$

or

b) The intersection of two circles whose centres are points in $\mathcal{P}_{\mathcal{C}}$ and whose radii are the distances between pairs of points in $\mathcal{P}_{\mathcal{C}}$

or

c) The intersection of a line of the form described in a) with a circle of the form described in b).

The purpose of the final exercise B) below is to show that the result of any of a), b) or c) is to produce another point in $\mathcal{P}_{\mathcal{C}}$.

Hence the first step in our construction produces nothing new, and so the second step, being of the same form and still starting from $\mathcal{P}_{\mathcal{C}}$, also produces no new point. The same is true of the third, fourth, fifth, etc. steps, and so any finite number of such steps will not produce a point outside of $\mathcal{P}_\mathcal{C}$. We therefore conclude that the constructible points are precisely the points in $\mathcal{P}_\mathcal{C}$ from which 8) follows immediately.

EXERCISE B) Using the results of exercise A) show that:

a) the line $ax + by = c$ passes through two points in $\mathcal{P}_\mathcal{C}$, if and only if a, b and c may be chosen from $\mathcal{C}$.

Hence conclude that the intersection of two such lines is a point in $\mathcal{P}_\mathcal{C}$.

b) i) the distance between two points in $\mathcal{P}_\mathcal{C}$ is a number in $\mathcal{C}$.

ii) if $x^2 + y^2 + 2ax + 2by + c = 0$ is a circle with centre in $\mathcal{P}_\mathcal{C}$ and radius the distance between two points in $\mathcal{P}_\mathcal{C}$, then a, b and c are in $\mathcal{C}$.

iii) the intersection of two circles

$$x^2 + y^2 + 2ax + 2by + c = 0$$

and

$$x^2 + y^2 + 2a'x + 2b'y + c \;= 0$$

is the intersection of the line

$$2(a - a')x + 2(b - b')y + (c - c') = 0$$

with either of the circles.

Hence conclude that the points of intersection of two circles with centres in $\mathcal{P}_\mathcal{C}$ and radii the distances between pairs of points in $\mathcal{P}_\mathcal{C}$ are also the points of intersection of a line passing through two points in $\mathcal{P}_\mathcal{C}$ and one of the circles (that is, case b) reduces to case c).)

c) Show that the points of intersection of a line passing through two points in $\mathcal{P}_\mathcal{C}$ with a circle centred on a point of $\mathcal{P}_\mathcal{C}$ and with radius the distance between a pair of points in $\mathcal{P}_\mathcal{C}$ are points in $\mathcal{P}_\mathcal{C}$. [Hint; use the characterizations of such a line and circle, found in a) and b) above.]

SQUARING THE CIRCLE. The classical Greek problem of "squaring the circle", that is, constructing a square of area equal to that of a given circle may be translated as follows:

Take the radius of the circle to be the unit of length, then the circle has area $\pi r^2 = \pi$, so the required square has side length $\sqrt{\pi}$. Since $\sqrt{\pi}$ is constructible if π is ((4) above) we have

The circle may be squared if and only if π is a constructible number.

We have not proved π is non-constructible, only that non-constructible numbers exist. None the less we have seen that any constructible number is algebraic and so we can now appreciate how Lindemann's result that π is transcendental led to the conclusion:

It is impossible to square the circle in a finite number of precise steps using only a straight edge and compass.