to have panted for a knowledge of the use of these symbols for years.' – John Perry.

ONE of the commonest words in modern life is 'speed'. It is therefore natural that mathematicians, who have had a hand in most of the scientific and industrial advances on which modern life is based, should have a special set of symbols to describe speed, and a special subject dealing with the use of these symbols. Like the sellers of patent medicine, mathematicians have not been able to resist the lure of high-sounding titles. The subject is known by the name of the Differential Calculus.

In any subject dealing with things that move, or grow, or change, you are likely to find the symbols of differential calculus. Even in subjects where nothing seems to be moving, these symbols turn up. We say that a road bends 'very suddenly'; we can discuss how 'quickly' the direction of a railway line changes. Neither the road nor the rail is moving at all. Yet we do mean something when we use such phrases. Words originally meant to describe motion – 'quickly', 'suddenly' – can be used to describe motionless objects. It is the same with the symbols which take the place of words in mathematical work. They too can be used to describe the curve of a road, or a railway, or of any similar object.

Differential calculus is therefore a subject which can be applied to anything which moves, or has a shape, or changes – and this does not leave much out! It is useful for the study of machinery of all kinds, for electric lighting and wireless, for economics and for life insurance. For two hundred years after the discovery of the differential calculus, the main advance of mathematics lay in applications of it. Very few really new ideas came into mathematics. Once the basic ideas of differential calculus have been grasped, a whole world of problems can be tackled without great difficulty. It is a subject well worth learning.

The Basic Problem

The basic problem of differential calculus is the following: we are given a rule for finding where an object is at any time, and are asked to find out how fast it is moving.

For instance, we might be given the following table, for a stone rolling down a hillside.

TABLE VII

Time in seconds	0	1	2	3	4	5	6
Distance gone (feet)	0	1	4	9	16	25	36

This rule, of course, is a very simple one: $y=x^2$, x meaning the time in seconds required to go y feet.

We are now asked: You know where the stone is at any time, find out how fast it is going. Let us try to find out how fast it is going after one second.

First of all, it is easy to see that the stone continually goes faster and faster. In the first second it goes only 1 foot; in the next 3 feet; in the third second 5 feet, and so on, 2 more feet for every second that passes.

But this does not tell us how fast it is going after 1 second, though it helps us to get an idea of the answer. In the first second the stone goes 1 foot. It is *averaging* a speed of 1 foot a second, during this second. This does not mean that its speed is 1 foot a second. A car which travels 30 miles in an hour does not travel at a speed of 30 miles an hour. If its owner lives in a big town, the car travels slowly while it is getting out of the town, and makes up for it by doing 50 on the arterial road in the country. The rolling stone is doing the same sort of thing – it starts slowly, but (so to speak) keeps its foot on the accelerator the whole time. As it covers 1 foot in the first second, its speed at the end of that second must be *more* than 1 foot a second, for it reaches its highest speed (for the first second) right at the end.

Its speed still goes on increasing during the second interval, in which it covers 3 feet. Accordingly, at the beginning of the second interval its speed must have been *less* than 3 feet a second.

Accordingly, after 1 second its speed is somewhere between 1 foot a second and 3 feet a second.

This is the best we can do, if we merely consider the figures for whole seconds. But there is no need to keep to whole seconds.

Our rule, $y=x^2$, applies equally well to fractions. If we work out the distances corresponding to 0·9 second and 1·1 seconds, we have the little table:

TABLE VIII

x	0·9	1	1·1
y	0·81	1	1·21

Just the same argument can now be applied again. In the tenth of a second between 0·9 and 1 the stone covers 0·19 of a foot. This represents an average speed of $\frac{0{\cdot}19}{0{\cdot}1}$ feet a second – that is, 1·9 feet a second. In the same way, the average speed in the tenth of a second just after 1 second is 2·1 feet a second. The number we want therefore lies between 1·9 and 2·1. In getting this result, we have not had to carry out any process more complicated than ordinary division.

But there is no limit to the accuracy we can obtain by this method. If we consider the hundredth of a second just before, and the hundredth just after, 1 second, we find that the speed lies between 1·99 and 2·01 feet a second. If we take a thousandth of a second, we find the speed lies between 1·999 and 2·001. And there is nothing to stop us considering a millionth or a billionth of a second if we want to. Only one speed will satisfy all these conditions – exactly 2 feet a second. And that is the answer to our question.

In exactly the same way you find the speed after 2 seconds. The little table may then be read:

TABLE IX

x	1·9	2	2·1
y	3·61	4	4·41

which shows that the speed after 2 seconds lies between 3·9 and 4·1. In fact, the speed is 4 feet a second.

So one may work out the speed after any time. The results of doing this are collected in the following table.

TABLE X

Time in seconds	0	1	2	3	4	5	6
Speed (in feet a second)	0	2	4	6	8	10	12

From this table it is easy to see the rule. After x seconds the speed is given by the number $2x$.

By this 'experimental' method it is fairly easy to find the rules for other formulae. First, one finds the speeds corresponding to 1, 2, 3, 4, 5, 6, etc.; then (by the method explained in Chapter 8) one tries to find a formula which will fit this set of numbers, and give the rule for the speed after x seconds. This enables one, at any rate, to *guess* the answer: to *prove* its correctness one has to use algebra.

You should be able to find for yourself the speeds corresponding to the formula $y=x^3$, the speeds corresponding to the formula $y=x^4$, and so on, with $y=x^5$, $y=x^6$, etc. When you have done $y=x^3$ and $y=x^4$, you will notice how simple the answers are. This suggests that the answers for $y=x^5$ and $y=x^6$ will also be simple, and helps you to shorten the work of guessing the rule. To work out the rule by the method of Chapter 8 would otherwise be a rather long piece of work. If you possibly can, work out for yourself the answers for the above cases, without looking at the results below. Anyone can do this work, and it makes all the difference to morale if you can find the results for yourself without ever looking at a text-book.

If you succeed in doing this, you will obtain the results set out in the table below.

TABLE XI

Formula for Distance Gone in x seconds.	Formula for Speed after x seconds.
$y=x^2$	$2x$
$y=x^3$	$3x^2$
$y=x^4$	$4x^3$
$y=x^5$	$5x^4$
$y=x^6$	$6x^5$

It is obvious that these results could be obtained by a simple rule. Where we have x^3 in the first column, we have something containing x^2 in the second; opposite x^4 stands a certain number of times x^3. The power of x in the second column is always one less than in the first. Opposite x^n there will be something containing x^{n-1}. Even simpler is the rule for the number which stands before x: it is the same as the number in the first column: it is n. Our rule is, 'If the formula for the distance is x^n, the formula for the speed is nx^{n-1}.'

Note what a lot of work is contained in this little result. To find the speed corresponding to x^2 we had to do a whole row of calculations; then we had to notice that the formula $2x$ would fit the results. We had to do this work also for x^3, x^4, x^5 and x^6. We then collected the formulae together in Table XI, and noticed that they could all be fused in one general rule.

Once having found the general rule, we can apply it immediately to any other case. To the formula $y=x^{17}$ will correspond the speed $17x^{16}$, the speed corresponding to x^{92} is $92x^{91}$.

In mechanics, and other applications, we often have to deal with formulae which contain several powers of x. For instance, if a ball is thrown straight upwards with a speed of 40 feet a second, its height after x seconds is given by $40x-16x^2$ feet. (We had this formula in Chapter 8 in a slightly different form. There it was given for the height after x *quarter*-seconds.) How is the speed after x seconds to be found?

The best way to deal with such a question is to split it up. We shall consider in turn the different parts which make up the problem.

(i) How quickly does the term $40x$ grow? $40x$ is the distance that a body would go in x seconds if it travelled with a *steady* speed of 40 feet a second. So the speed corresponding to $40x$ is obviously 40.

(ii) How quickly does the term $16x^2$ grow? We could get the table for $16x^2$ by multiplying all the numbers in the lower row of Table VII by 16. In other words, if a body travels according to the formula $16x^2$ after any number of seconds, it will have gone 16 times as far as one travelling according to the formula x^2. At any moment it must therefore be travelling 16 times as fast. But the speed corresponding to x^2 is $2x$. The speed corresponding to $16x^2$ must be 16 times as large: it must be $32x$.

(iii) We now know that $40x$ grows steadily at the rate 40, while $16x^2$ grows at the rate $32x$. How fast will $40x—16x^2$ grow? How are the two rates to be combined?

$40x—16x^2$ is obtained by subtracting $16x^2$ from $40x$. How can we picture this subtraction? We might think of $40x$ as representing a man's income at any moment, $16x^2$ as representing the expense of bringing up his family. Both income and expenditure are growing. $40x—16x^2$ represents the weekly balance which the man has, after meeting his expenses. It is obvious that this balance will increase at a rate given by the rate at which his income rises, *minus* the rate at which his expenses rise. (If the balance is decreasing, this rate will be less than nothing – it will have a minus sign.) The rate of increase of the income is 40; the rate of increase of expenditure is $32x$. The rate of increase of the balance is therefore $40—32x$. So we combine the rates by subtracting the second from the first.

We thus reach our grand conclusion: the speed corresponding to the formula $40x—16x^2$ is $40—32x$.

You will see that the arguments used could equally well be applied to any other expression of a similar type. For instance, the speed corresponding to $4x^3+x^2+3x+1$ is $12x^2+2x+3$. (The number 1 does not change at all: $y=1$ would mean that a body always stayed at a distance of 1 from some fixed point. Its speed would of course be nothing at all. So this 1 in the formula does not add anything to the answer. $4x^3+x^2+3x$ would lead to the same speed. This is quite reasonable. $4x^3+x^2+3x$ is always 1 less than $4x^3+x^2+3x+1$. The first formula neither catches up nor falls behind the second. So the speeds are naturally the same).

If you have any difficulty with this idea, reason out for yourself the speeds which correspond to $5x^2$ and to $2x$; then to $5x^2+2x$. Work out the speeds which correspond to x^2+x, $x^2—x$, $x+1$, x^2+x+1, and other examples which you can make up for yourself. Check your answers by working out the tables for these formulae, and seeing if your answers for the speeds are reasonable.

Signs for Speed

It is inconvenient to keep on saying 'the speed corresponding to the formula'. A sign is therefore used. If we have any formula giving y, the corresponding speed is represented by y'. This enables us to state a rule we had earlier in the shorter form, 'If $y=x^n$, $y'=nx^{n-1}$.' This means exactly the same as 'To the formula x^n corresponds the speed nx^{n-1}.' Just as $y(x)$ is used to represent the distance corresponding to x, so $y'(x)$ is used to represent the speed corresponding to x. Thus $y'(2)$ will mean the speed after 2 seconds.

It is sometimes convenient to use another sign, instead of y'. This other sign is $\frac{dy}{dx}$.

There is a reason why this sign is used. The d in it has a very special meaning, like the Δ used in Chapter 8. In fact, it is only through the sign Δ that you can see why d is used here.

What, after all, is a speed? If you are told that a train has covered 300 miles in 4 hours, you know that its speed has been (on the average) 75 miles an hour. The 75 is found by dividing 300 by 4. If you know that a train has travelled 150 miles by 7 a.m. and 270 miles by 10 a.m., how do you find its average speed between 7 a.m. and 10 a.m.? You find the time that has passed between 7 a.m. and 10 a.m., 3 hours. You find the change in the distance, 270—150=120. Then, dividing 120 by 3, you have your answer: 40 m.p.h.

If we call the time x hours, and the distance gone y miles, we have a table, rather like those in Chapter 8.

TABLE XII

Δx		3	
x	7		10
y	150		270
Δy		120	

As before, we have the values of x in one row, the corresponding values of y beneath them, and then a row giving Δy, the change in y. The only new feature is the row labelled Δx, giving the change in x. In Chapter 8, the change in x, between any number and the next, was always 1, so it would have been a waste of time, as well as a complication, to have had a row Δx. But for finding speeds Δx is absolutely essential. The speed of 40 m.p.h. was found by dividing 120 by 3: that is, by dividing Δy by Δx.

The rule for finding *average speed*, then, is to work out the change in distance divided by the change in time. In our symbols,

$$\text{average speed} = \frac{\Delta y}{\Delta x}.$$

But this only gives *average* speed. We are looking for the speed at any moment. If a car runs into you at 60 m.p.h. it is no comfort to your widow to be told that it had averaged only 10 miles over the last hour, because the driver had spent most of the hour in a pub. The thing that matters is not the average speed during the past hour: it is the actual speed at the exact moment when the car hits you that counts.

But the speed at the moment of the collision will not differ very much from the average speed during the previous tenth of a second. It will differ even less from the average speed for the previous thousandth of a second. In other words, if we take the average speed for smaller and smaller lengths of time, we shall get nearer and nearer – as near as we like – to the true speed. For most practical purposes, the average speed during a thousandth of a second may be regarded as the exact speed.

It is for this reason that mathematicians find it useful to represent *speed* by a sign similar to that for *average speed.* The sign Δ is used by the Greeks to represent the capital letter D. We cannot take over the sign $\frac{\Delta y}{\Delta x}$, unchanged, to represent the speed: for the average speed over a short interval, however near it may be to the exact speed at any moment, is never quite the same: it would lead to confusion to have the same sign for two different things. But, as it were for old times' sake, to remind us how the idea of average speed has helped us towards finding the exact speed, we replace the Greek Δ by an English d, and use $\frac{dy}{dx}$ to represent the speed. For the moment, do not ask what dy and dx mean separately. Just regard $\frac{dy}{dx}$ as a sign that can be used instead of y', to represent the speed.

Again, in mechanical problems the speed is usually called by the technical term 'velocity', and we may use v as an abbreviation.

The process of finding how quickly a quantity changes is known as *differentiation*. If we differentiate y, we obtain its *rate of change* (or *speed*), y', $\frac{dy}{dx}$, v. If we differentiate x^2 we obtain $2x$.

This process can be repeated. When we considered a stone rolling downhill, according to the formula $y=x^2$, we saw that the velocity v continually increased. One might easily ask, 'How fast does it increase?' There is no difficulty in answering this question. We have seen that the velocity v after x seconds is given by the formula $v=2x$. Thus we have a very simple formula for v, and it is easy to find v'. In fact, $v'=2$. The velocity increases steadily; it increases by 2 for every second that passes. (Check this result from the values of v given in Table X.)

Since v is the same thing as y', it is natural to represent v' by y''. There is nothing new involved in this sign y''. y' represents the rate at which y changes: y'' represents the rate at which y' changes. In Chapter 8 we found Δ^2y from Δy by just repeating the process we had already done to find Δy from y. It is the same here. We start with y. How quickly does y increase? The answer is y'. We now start again with a table (or a formula) for y'. How quickly does y' increase? The answer is y''. In many ways y'' resembles Δ^2y.

The Importance of y' and y''

The quantities y' and y'' have great importance in mechanics. It is obvious that y' (which simply means the speed, or velocity) is important. If anything, y'' is even more important. y'' measures how quickly the speed changes. If you are in a car travelling at 50 m.p.h. and the driver gradually brings the car to rest – say, within 10 minutes – you feel hardly anything at all. If, on the other hand, the car is brought to rest in one-hundredth of a second – by colliding with a stone wall – this is felt as a blow of tremendous force, sufficient to do serious damage. It does not hurt to travel at a high speed, such as 50 m.p.h. What does hurt is *a sudden change of speed.*

Usually, when our speed changes we feel pressure. If you are in a car, and the brakes are suddenly put on, you feel yourself thrown forward. What really happens is that you keep on moving at the same speed, but the car stops. You stop only when you strike the seat in front of you: you feel this pressing you back. In the same way, a bicycle cannot be slowed down if it has no brakes (unless perhaps there is a strong wind blowing against it, or it is going uphill, or it is badly oiled – any of these states can take the place of brakes).

Newton's Laws of Motion express this idea. According to Newton, if a body could get right away from all outside influences – far away from the pull of the earth and the sun, not being pressed or pulled by any other body, away from electric and magnetic forces – that body would keep on moving in a straight line at a steady speed. With a telescope one can observe little bits of matter, such as comets, and it is seen that the farther away they get from the earth and the sun, the more nearly they move in straight paths at a steady speed.

Whenever we find a body moving in a curved path, or with a changing speed, we therefore believe that something else is interfering with it, is acting upon it. We say that a *force* is acting on it and we try to discover what this force is. Is the body tied to a rope or string? Is it nailed to some other body? Is it being pulled by the earth or the sun or (in the case of tides) the moon? Has it been magnetized? Is it charged with electricity? Is it sliding on a rough surface, which is causing it to slow down? Is it moving in some liquid, such as water or treacle, which opposes its motion? Is it passing through the air, like a parachute, or a falling feather?

Next we ask how to measure the force. It is clear that the force needed to alter the speed of a body depends on how massive the body is. It is easy to stop a runaway pram; harder to stop a runaway wagon; very hard to stop a string of loaded trucks; almost impossible to stop an avalanche. Scientists therefore use the word *mass* to express this quality of a body. One cubic centimetre of water is chosen as the unit of mass; it is called a gramme. Anything which is just as easy to stop, or to get going, as a cubic centimetre of water is said to have a mass of 1 gramme. Anything which is as hard to stop as 100,000 cubic centimetres of water is said to have a mass of 100,000 grammes. And so on. For short, we shall call the mass of a body m grammes.

It is then found that the force which changes the speed of a body (mass m grammes) at the rate y'' is given by my''. When the body is going faster and faster, with y increasing, the force pulls the body forward. When the body is slowing down, y'' is negative: this means that the force is acting as a brake – it is pulling the body back.

In scientific work it is usual to measure the distance y, not in feet or inches, but in centimetres. By using the decimal system of measures we save all the complications of having 12 inches one foot, 3 feet one yard, 22 yards one chain, 30¼ square yards one pole, and so on. The English system of measurement has been handed down from very ancient times, long before science or modern engineering had been thought of. It is based on such things as the amount of land a team of oxen can plough in a day (furlong=a furrow long, etc.) or the average size of a part of the human body (*e.g.*, a foot). Such measures were convenient for their original uses. The French system, on the other hand, was introduced during the French Revolution of 1789, and was specially designed for modern trade and industry. Any difficulties that are found in changing from feet and tons to centimetres and grammes must be regarded as due to human history: they are not purely scientific problems.

English engineers also use a system of measurement, in which the standard mass is a one-pound weight, and the distance y is measured in feet. The force corresponding to m pounds and a speeding-up y'' (measured in feet and seconds) is then my'' *poundals*.

We have already considered the formula $y=40x-16x^2$, which gives the height in feet of a body x seconds after it has been thrown-up with a speed of 40 feet a second. What force is acting on this body? $y'=40-32x$. $y''=-32$. If the body has a mass m pounds, the force acting on it is my'', which equals $-32m$. This force does not depend on x. It is the same whatever x may be. The size of the force is $32m$: the sign in front of it is *minus* because the earth drags the weight *down*, as everyone knows. In the case of something which tended to rise (such as a balloon) the force would be $+$.

It is found by experiment that any heavy body thrown into the air moves in such a way that $y''=-32$. We suppose the body is sufficiently heavy for air resistance to be neglected. This law obviously does not hold for a feather, or for a parachute. The whole point of a parachute is that it falls in a way very different from a brick. The law just stated gives good results for a falling stone, cricket ball, or man. It does not work for feathers, rain-drops, or mice. Nor does it work for very high speeds. In the motion of a shell or a bullet, the force due to air resistance may well be greater than that due to gravity.

The figure 32 is of course not exact. The earth does not bother to pull us towards itself with a force that is a simple multiple of the length of our feet! But 32 is near enough for most purposes.

Since the pull of the earth causes y'' to have the value -32, the force exerted on a mass of m pounds by the earth must be $-32m$ poundals (found by putting $y''=-32$ in the expression my''). The *minus* sign means that this force acts *downwards*.

*Other Useful Subjects**

So far we have dealt with a very special case, that of a body moving in a straight line, and acted on by a single force only.

* The remainder of this chapter contains applications which may interest some readers, but are not necessary for an understanding of the rest of the book. A brief reference to this section is made in Chapter 13.

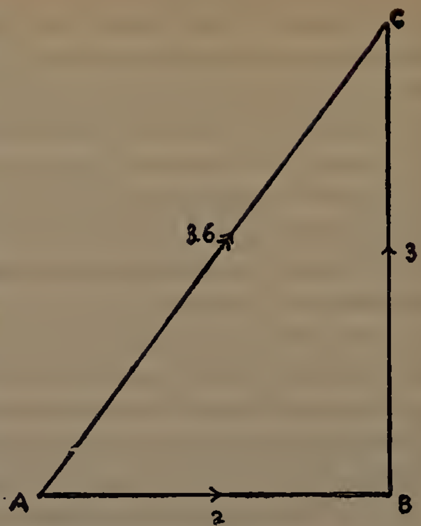
In nearly all practical studies the problem is more complicated. A lift moves up and down in a straight line, but two important forces act on it – the pull of the earth downwards, and the pull of the supporting rope upwards. We may also need to take into account any device used to stop the lift bumping against the walls of the shaft, friction, air-resistance, etc. – even if we neglect these, we still have *two* forces to consider. In other examples we may have to deal with bodies that do not move in straight lines: a train or motor-car going round a bend, a shell in the air, a piece of metal in a flywheel.

Statics deals with the combined effect of a number of forces. Its laws have to be discovered in the first place by experiment, and can then be used for reasoning.

When several forces act in the same direction, the result is what you might expect. If two men pull a sledge, each hauling with a force of 1000 poundals, the effect is the same as that of a single pull of 2000 poundals: the separate forces are simply added together.

When two forces act in opposite directions, the effect is easy to calculate. If a lift weighs 2500 pounds, the earth will pull it downwards with a force of 32×2500 poundals – i.e., 80,000 poundals. If the wire rope pulls the lift upwards with a force of 100,000 poundals, two forces are acting on the lift, 100,000 poundals upwards, 80,000 poundals downwards. These two forces combined have the same effect as a force of 20,000 poundals (obtained by subtracting 80,000 from 100,000), acting upwards. Note that the rope must be able to stand a strain which is *greater* than the weight of the lift. For it is only when the pull of the rope upwards is greater than the pull of the earth downwards that the total force is upwards. And one must be able to make the total force act upwards in order to start the lift for an upward journey and (equally important) to check it at the end of a downward journey. In a coalmine, the cage lifts and lowers the miners through hundreds of yards in an amazingly short space of time. Very great changes of speed take place, and it is a matter of vital importance that the rope be strong enough, not merely to bear the weight of the cage and the men inside it, but also to exert the extra force necessary for starting and braking. (It is possible to experiment with model lifts, using cotton thread instead of rope, to demonstrate the effect of a sudden jerk.)

An entirely new principle has to be learnt in order to deal with forces which do not act in the same straight line. Suppose we have a force of 2 poundals acting East, and a force of 3 poundals acting North: to what single force is this equivalent? It is impossible to answer this by argument: we can only try to see what happens when a small weight is dragged towards the East by one string attached to it, and towards the North by another. (For details of the experiment, see text-books on Statics.) The reader will be able to *feel* what sort of result is likely – the weight will be dragged in a direction somewhere between North and East. Experiments show that the following method gives the correct answer. Draw a line 2 inches long, towards the East. Call this line AB. From the Eastern end of this line (B) draw a line BC, 3 inches long, towards the North. The line AB is drawn 2 inches long, to represent a force of 2 poundals; the line BC is made 3 inches long, corresponding to the force of 3 poundals. If we measure AC, we find it to be 3·6 inches long. The length and direction of AC gives the answer to the question. The two forces, 2 poundals towards the East and 3 poundals towards the North, acting together, will drag the weight in the direction AC, with a force of 3·6 poundals. This principle is known as the Triangle of Forces, for an obvious reason. In the triangle ABC the two sides AB and BC represent the two forces given. The third side, AC, represents the single force produced by these two forces acting together.



TRIANGLE OF FORCES

In an ordinary catapult two pieces of elastic are fastened to a small piece of cloth. When the catapult is fired, the small piece of cloth moves in a direction that lies between the direction of the two pieces of elastic.

Co-ordinate Geometry

In dealing with graphs we used the idea of fixing the position of a point by measuring the distance across the paper ('to the East') and the distance 'to the North'. The same idea may be used to study the movement of any small weight, when it is not moving in a straight line. We suppose that, after x seconds, the small weight is y feet to the East and z feet to the North of some fixed landmark, O. The small weight might be part of some machine. The rules giving y and z in terms of x will depend on the way in which the machine is constructed. For instance, the weight might be some part of a locomotive. If we know the shape of the railway line, and the speed at which the train is travelling, we know *where* each part of the locomotive will be at any time. In other words, we know what values y and z will have after x seconds.

If no force is acting on a body, the body moves in a straight line. When a locomotive goes round a bend, it is not moving in a straight line, nor is any piece of the locomotive moving in a straight line. So there must be forces acting on each part of the locomotive. You will always notice that a locomotive, in going round a bend, presses against the outer rail, just as a motor-car going round a corner too fast tends to run into the outer edge of the road, if the road is not sufficiently banked. The rail presses back on the wheels, and makes these go round the bend, instead of going straight on, as they would prefer to do. Is it possible to find how large is the force acting on any part of the locomotive? It is possible, though it is not easy to describe the method in a few words.

First of all, in order not to make the problem too complicated, let us choose a part of the locomotive which does not move up and down. As it always stays at the same height, its motion can be described completely by giving a table, showing how far it has moved to the East of the landmark O, and how far it has moved to the North of O. Our first task, then, is to discover formulae, or to make tables, giving y and z corresponding to any time, x seconds. We suppose this part of the job completed.

It would be quite easy to study the motion of the locomotive (or the little part of it) towards the East, *if* there were no motion towards the North. The locomotive would then be moving due East, *in a straight line*. After x seconds its distance to the East would be given by y, and the force pushing the small part (of mass m pounds, say) towards the East would be (by our previous method) my''.

It would also be easy to find the answer, if the locomotive was moving due North. The force pushing the small part to the North would be mz'', by a very similar argument.

At this point we receive a free gift from Nature. *It turns out to be true* – and we had no reason to expect this – that the movement towards the East and the movement towards the North can be treated as if they were quite separate. The actual force pushing the small piece is obtained by combining (by the Triangle of Forces) a force my'' to the East and a force mz'' to the North.

The problem therefore can be completely solved. There is no essential difficulty brought in if we consider movements up and down, as well as to the East and to the North. The forces due to the fact that parts of a locomotive move up and down are very important. Old-fashioned locomotives, if driven fast, would jump into the air.

Of the design of modern locomotives, Kempe's *Engineer's Yearbook* writes, 'The horizontal forces are the most injurious, though American engineers consider the vertical forces to be so; but English practice is to take a medium course between excessive horizontal and vertical disturbing influences.'

The calculation of the forces brought into play by moving weights is a practical question in the design and balancing of machinery. In this short space it has not been possible to explain the method in a satisfactory way, but the fact that the method can be outlined, even vaguely, in so few words shows that the principles involved are both few and simple.

Conclusion

Statics and Dynamics will seem very unreal to you if you have had no experience of handling heavy weights. You can learn more dynamics in an afternoon starting and stopping a heavy (but well oiled) railway truck or garden roller than you can from whole books of dynamics. You can get the benefit of reading a book on dynamics only if such words as 'force' call up a vivid image in your mind. Once you have the necessary feel for the subject, the books can be most valuable, even interesting – but not before.

Calculus does not need the same experimental background. Almost everyone already knows what speed is. The job is rather to study sets of figures, until you realize what sort of motion they represent. Take any formula. Work out a table, showing the distance gone after various times. If you cannot see what the exact speed is, begin to ask questions. Silly ones are the best to begin with. Is the speed a million miles an hour? Or one inch a century? Somewhere between these limits. Good. We now know something about the speed. Begin to bring the limits in, and see how close together they can be brought. Study your own methods of thought. How do you know that the speed is less than a million miles an hour? What definite evidence does the table show to support this view? What method, in fact, are you unconsciously using to estimate speed? Can this method be applied to get closer estimates?

You know what speed is. You would not believe a man who claimed to walk at 5 miles an hour, but took 3 hours to walk 6 miles. You have only to apply the same common sense to stones rolling down hillsides, and the calculus is at your command.

EXAMPLES

It has already been urged that the reader should not try to reason about any problem until he has a perfectly clear picture of it in his mind, and has found some way of bringing the problem into touch with real life, so that he can see and handle the things of which it speaks. This is particularly important in the study of *speed*, which is by no means such a simple thing as we at first think it to be. The reader must find for himself some device by means of which he can observe movement. This may be a pencil rolling down a desklid, a bicycle on a hillside, or a weight hanging by a string. One particular device may be mentioned, on the lines of cinema cartoons. Most school-children are familiar with a way of drawing pictures on the leaves of a book, so that when the leaves are allowed to fall in rapid succession the figures seem to move. The same idea may be used to study the movement of a point. It has the advantage that one can study the movement 'frozen' – by looking at the points marked on the various pages of the book – as well as in action. In questions 1 and 2 it is assumed that the leaves of the book fall at the rate of ten each second.

1. On the first sheet of a book mark a point 0·1 inch from the bottom of the page, on the second 0·2 inch, etc., the point on the nth sheet being $\frac{n}{10}$ inches up the page. This illustrates the movement in which $y=x$ (y in inches, x in seconds). That the point moves with a steady speed is shown by the fact that the mark on any sheet is always 0·1 inch higher than that on the page before.

2. On the nth sheet mark a point at the height $\frac{n}{100}$. This corresponds to the movement $y=x^2$ discussed in this chapter. Notice how slowly the point moves in the first half-second (five sheets), how it gains speed as time passes.

3. A body moves according to the law $y=x$. Make a table for its motion and convince yourself: (i) that it is moving at a steady speed, (ii) that this speed is 1. In fact, when $y=x$, $y'=1$.

4. Similarly, show that when $y=2x$, $y'=2$.

5. And when $y=\frac{3}{4}x$, $y'=\frac{3}{4}$.

6. When $y=x+1$, $y'=1$.

7. And when $y=x+2$, $y'=1$.

8. When $y=\frac{3}{4}x+1$, $y'=\frac{3}{4}$.

9. When $y=\frac{3}{4}x+2$, $y'=\frac{3}{4}$.

10. A dog is chasing a cat. The cat moves according to the formula $y=30x+2$, the dog according to the formula $y=30x$. Is it true that (i) both animals are moving at 30 feet a second; (ii) the dog is always 2 feet behind the cat; (iii) $y'=30$ for both formulae?

11. If the cat moves according to $y=20x+10$ and the dog according to $y=25x$, is it true (i) that the dog starts 10 feet behind the cat; (ii) the dog moves faster than the cat; (iii) the dog will catch the cat within a short time? What is y' for the cat? for the dog? When will the dog overtake the cat?

12. Write down y' in the following cases:
(i) $y=x^2$. (ii) $y=2x^2$. (iii) $y=2x^2+1$. (iv) $y=\frac{1}{2}x^2$. (v) $y=x$. (vi) $y=x^2+x$. (vii) $y=x^2+x+1$. (viii) $y=\frac{1}{2}x^2-1$. (ix) $y=x-x^2$. (x) $y=1-x^2$. (xi) $y=x^3$. (xii) $y=2x^3$. (xiii) $y=2x^3+x$. (xiv) $y=2x^3+1$. (xv) $y=10x^3-20x^2+7x-3$.

13. We have seen that when $y=x^2$, $y'=2x$ and $y''=2$. Make tables showing y, y', y'', Δy and $\Delta^2 y$ for $x=0, 1, 2 \ldots 10$. Is it true that (i) the table for y' is rather like, but not quite the same as, the table for Δy; (ii) $\Delta^2 y$ is exactly the same as y''?

14. If $y=x^3$, $y'=3x^2$ and $y''=6x$. Make tables for y, y', y'', Δy and $\Delta^2 y$. Is it true (i) y' behaves rather like Δy; (ii) y'' behaves rather like $\Delta^2 y$?

15. If you had worked out a problem and found a formula for y' which behaved in a way quite different from Δy (for instance, y' getting steadily larger while Δy got steadily smaller), would you think you had made a slip in your work or not? What about y'' and $\Delta^2 y$? Do you expect them, as a rule, to behave in more or less the same way?

CHAPTER 11

FROM SPEED TO CURVES

'Our townsman, Dr Joule . . . instanced the porpoise, with its bluff figurehead, attaining a velocity of over thirteen miles an hour, whilst voracious fishes are so constructed that they can attain a much greater velocity. He advocated a study of natural proportions to those who wish to be successful shipbuilders.' – Bosdin Leech, *History of the Manchester Ship Canal.*

So far we have considered y' or $\frac{dy}{dx}$ simply as a sign for the speed of a moving point. For many important applications this is quite sufficient. But it is only one half of the story. There are many problems for which calculus can be used to describe shape. For instance, it is possible to find the curve in which a chain hangs when its ends are held, or the way in which the strain on a bridge is distributed. Movement does not come into either of these questions at all.

It is very easy to translate our discoveries about movement into statements about shape. Any type of movement can easily be

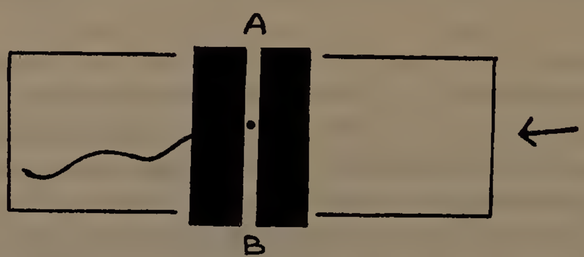


Fig. 9

represented by a curve. Consider the simple arrangement shown in Fig. 9. The point of a pencil is supposed to be moving in the slit *AB*. Underneath the slit is a sheet of paper, which is made to move at a steady pace towards the left. It is clear that the movement of the pencil will be recorded on the paper as a curve. If we want to see how the pencil was moving, we have only to pass the paper under the slit again. Through the slit we shall be able to see a very short piece of the curve, which will be seen as a point, and will seem to move up and down as the paper passes to the left.

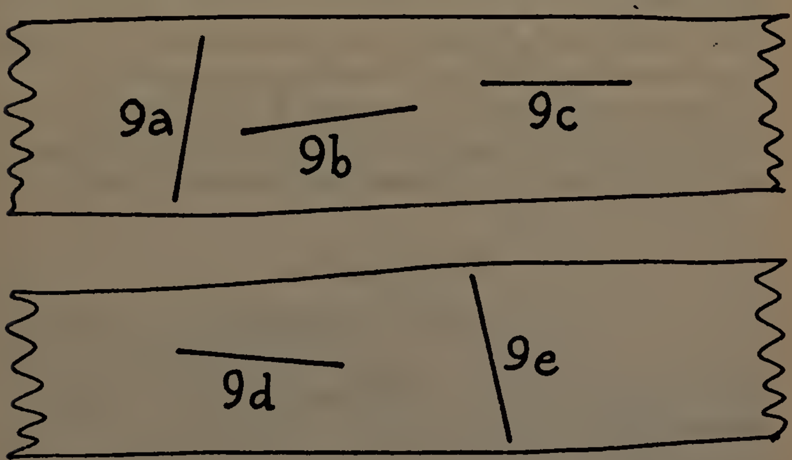
This arrangement is rather similar to a gramophone. The groove of the gramophone record is cut by a vibrating needle. When the record is played, the original vibrations are reproduced. All the peculiarities of the original motion, then, are somehow preserved in the shape of the groove; and any alteration in the shape of the groove would result in some difference when the record was played.

In the same way, the motion of the pencil point and the curve traced on the moving paper are closely connected. Anything that can be said about the movement of the pencil must tell us something about the shape of the curve. Anything that can be said about the curve tells us something about the movement of the pencil.

Now we know that y' tells us how fast the pencil is moving at any moment, and y'' tells us whether the pencil is speeding up or slowing down. It must be possible to find out what y' and y'' represent, what they tell us about the shape of the curve traced by the pencil. To do this is our next job.

The Case of Steady Movement

We will begin by considering the simplest case. Figures 9A to 9E show the traces left on the paper in five experiments. In each of these experiments the pencil moved at a steady speed. The strip of paper is 1 inch wide, and it moved through the slit,



towards the left, at the rate of 1 inch a second. You can, if you like, make an arrangement of the type illustrated in Fig. 9, and pass these strips through it, thus reproducing the original movements.

What happens when 9A passes through? It takes only one-fifth of a second, and during this time the moving point gets right across the paper, a distance of 1 inch. 9A is the track of a point moving at 5 inches a second. For it, $y'=5$.

9B is the track of a point that moves up the slit, but at a much slower rate. In one second the point has risen only one-tenth of an inch. Here $y'=\frac{1}{10}$.

Already, by comparing 9A and 9B, we can see that the line is *very steep* when the point has moved *very fast* (that is, for large y'), but is *nearly level* when the point is moving *slowly* (y' small). In fact, *y' measures how steep the graph is.*

9C is the track left when the pencil-point is *at rest*. The point stays at the same height. It has *no speed*, so $y'=0$.

Accordingly, when $y'=0$, the graph is *level.*

In 9D the pencil-point moves *down* the slit as the paper passes through. When a second has passed, the point has gone down one-tenth of an inch. The change in y during 1 second is therefore $-\frac{1}{10}$, from which it follows that $y'=-\frac{1}{10}$.

Again, in 9E the pencil-point descends 1 inch in one-fifth of a second; it is therefore descending at the rate of 5 inches a second, and $y'=-5$.

We notice that the graph slopes *downhill* when y' is minus (cases D and E) *uphill* when y' is plus (cases A and B).

To sum up: the *steepness* of the graph depends on how *big y'* is: whether the graph goes *uphill* or *downhill* depends on whether y' has a $+$ or a $-$ sign. $y'=0$ means that the graph is *level.*

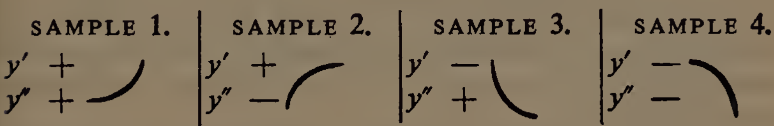
The General Case

So far we have been considering only what happens when the pencil-point moves at a steady speed. But as a rule this is not the case. We may often have to study things that move with different speeds at different times.

It is, however, still possible to use the conclusions to which we were led by a study of the simpler cases. You should test this for yourself, by means of some arrangement on the lines of Fig. 9. If you move a pencil up and down the slit, varying its speed, you will find that when the pencil is moving fast, the track it leaves is steep; when it is moving slowly, the track it leaves is not very steep. We can still say that *speed* corresponds to *steepness*. If the *speed varies*, then the *steepness* of the graph also *varies*. In this case the graph will be curved, instead of straight, as it was before.

This brings us to the question of y''. y'' tells us how quickly the speed, y', varies. We shall be mainly interested in the sign of y'', whether it is $+$ or $-$. If y'' is $+$, it means that y' is growing (i.e., that y' is changing by having something *added* to it). If y'' is $-$, it means that y' is decreasing (is having something *taken away* from it).

Look at the four sample tracks shown in the figure.



What are the signs of y' and y'' in Sample 1? This curve is rising, its slope is uphill; y' therefore must be $+$. The farther you go, the steeper this curve gets. Its steepness (measured by y') is getting bigger. So y' must be increasing. That means y'' must be $+$. It is easy to get a little confused between the meanings of y' and y''. Remember, y' measures how fast y is changing – that is, y' measures the speed of a moving point. y'' measures how quickly y' is changing – that is, how quickly the speed changes.

If this curve were part of a chart showing the course of a military campaign, it would mean (*a*) that the army was advancing, (*b*) that the speed of its advance was continually increasing. (*a*) corresponds to the mathematical statement that y' is $+$; (*b*) to the statement that y'' is $+$.

We have a different state of affairs in Sample 2. True, the slope of the curve is uphill, but the farther you go, the *less* steep it is. This corresponds to the military communiqué 'our advance continues, but is being slowed down by determined resistance'. Since it is an *advance*, y' is $+$. Since the rate of advance is being *slowed down*, y'' is $-$.

One has to be rather careful with Samples 3 and 4, owing to the fact that y' is minus. We have to remember that a change from $y'=-10$ to $y'=-1$ represents an increase in y', owing to the properties of negative numbers.

In Sample 3 the state of affairs at the beginning represents a *rapid descent* – in military terms, a rout. Later on the curve is still doing downhill, but not nearly so fast. The retreat is being checked. To this extent the situation is improving. The *improvement* is reflected by the fact that y'' is $+$. The *retreat* is shown by the downward slope of the curve: y' is $-$.

The reader may remember that the line 9E, going steeply downwards, had $y'=-5$, while 9D had $y'=-\frac{1}{10}$. In Sample 3 the early part of the curve slopes like the line 9E, while the end of the curve is more like 9D. In Sample 3, then, y' begins by being about -5, and ends by being about $-\frac{1}{10}$. You have to add something to -5 to make it into $-\frac{1}{10}$. That is why y'', the rate of change of y', is $+$.

In Sample 4, on the other hand, the situation is going to the dogs faster and faster. The curve is going downhill, becoming steeper and steeper. y' may be about $-\frac{1}{10}$ at the *beginning*, and -5 at the *end*. y' is therefore *changing for the worse* – that is, y'' is $-$.

In these four samples we have covered all the main possibilities. y' must be either $+$ or $-$, and y'' must be $+$ or $-$ (unless y' or y'' should happen to be 0). By combining our four Samples, we can see what the shape of any graph will be, provided we know how y' and y'' behave. Note that it is quite easy to give a simple meaning to y''. When y'' is $+$, the curve *bends upwards* (Samples 1 and 3); when y'' is $-$, the curve *bends downwards* (Samples 2 and 4).

An Example

Suppose you were asked. 'What is the general shape of the graph $y=x^3-3x$?' We will consider the shape of the graph between $x=-20$ and $x=+20$.

First of all we need to know y' and y''. Since $y = x^3 - 3x$, $y' = 3x^2 - 3$, $y'' = 6x$.

Clearly, y'' is $+$ when x is $+$, and $-$ when x is $-$. This tells us that the graph *bends upward* for all positive x, *downward* for all negative x.

If you try a few values for x, you will see that $3x^2 - 3$, the formula for y', is + when x lies between —20 and 1, and also when x lies between 1 and 20. $3x^2 - 3$ is — when x lies between —1 and 1.

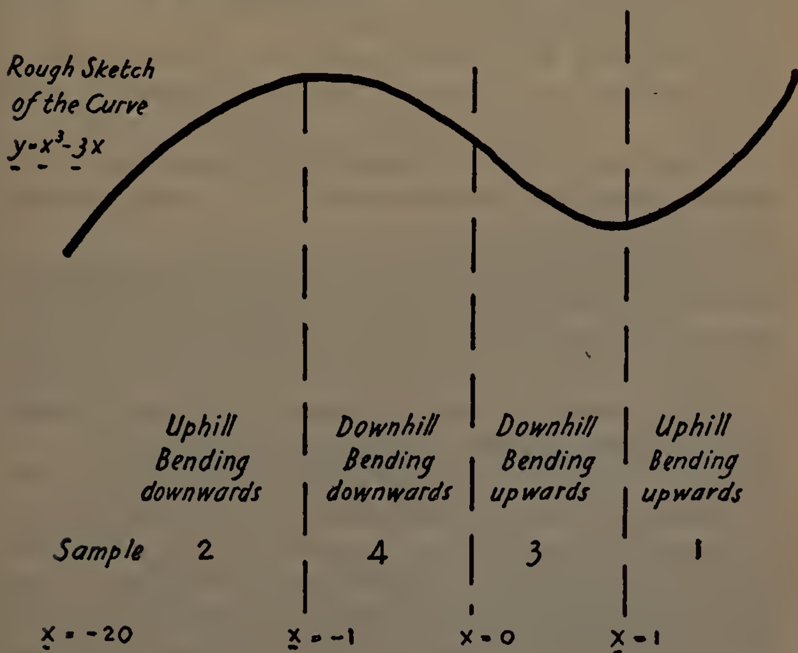
We can collect this information together in a diagram, thus:

x	—20	—1	......	0	......	1	 20
y'	+ + +	0	— —	—	— —	—0	+ + +
y''	— — —	—	— —	0	+ +	+	+ + +
Curve resembles sample	2 2 2		4 4		3 3		1 1 1

We thus conclude:

From $x =$ —20 to —1 the curve looks like SAMPLE 2.
—1 to 0 „ „ „ SAMPLE 4.
0 to 1 „ „ „ SAMPLE 3.
1 to 20 „ „ „ SAMPLE 1.

Fitting these together, we see that the general appearance of the graph must be as follows:



It would, of course, be possible to make an exact graph of the curve, by working out a table and plotting a large number of points on the curve. We should, of course, get the same curve in the end. But usually the method just explained, using y and y'', is shorter, more instructive and more artistic. Some of the examples at the end of this chapter are intended to illustrate this point.

We saw in Chapter 10 that y' measured the *speed*, y'' the *force acting* on a moving body. When y'' was $+$, it meant that the body was being pushed *upward* (or, in some cases, *forward*). When y'' was $-$, it meant that the body was being pulled *downward* (or, in some cases, *backward*).

But, by examining the graph that represents the motion of the body, we can see how y' and y'' are behaving. In this way, by looking at the graph, we are able to say (for example), 'Here the graph is rising very steeply. The body must be moving very fast. But the curve bends downwards. That means y'' is *minus*, some force is putting a brake on the motion.'

It is quite easy, with a little practice, to tell from a graph how the quantities y' and y'' behave, where y' is $+$, where y'' is minus, etc.

But suppose we are not content with a general description: suppose we want to measure the speed at a certain moment? How can we go about this?

We have already seen (in Chapter 10) that the true speed of a body does not differ much (as a rule) from the average speed over a short period of time. If we know how far a body goes in one tenth of a second, we can get *some* idea of how fast it is going.

If we are shown the graph representing an object's motion, can we tell how far it goes in one-tenth of a second?

In Fig. 10 part of a graph is shown, considerably magnified. The distance AB represents one-tenth of an inch, and corresponds to one-tenth of a second. At the beginning of this length of a second the pencil touched the paper at the point C. At the end of the tenth of a second it touched the paper at D. It must have moved upwards through a distance DE during this interval. In other words, DE represents the change in y: that is, Δy. The time that

ROLLING A CRICKET PITCH

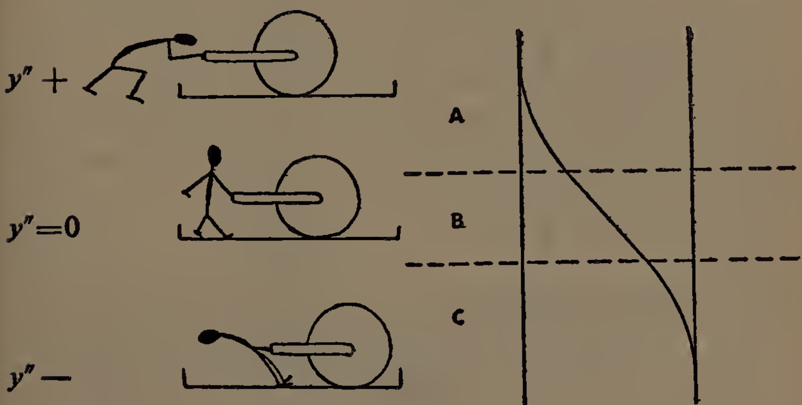
The curve on the right records the motion of the roller. To compare this curve with the arrangement of Fig. 9, it would be necessary to turn the page on its side; the roller would then seem to be moving *upwards* (like the pencil-point in the slit).

The motion may be divided into three sections, A, B, and C.

(a) The man has to push hard to get the roller moving. He is pushing *forwards* (y''+). But the roller is not yet moving fast (y' not large, curve not steep).

(b) The roller is now under way. The man walks beside it, but lets it roll without pushing or pulling ($y''=0$. *No* force).

(c) To stop the roller running into the wickets, the man has to *pull* it *back* (y''—).



Note that the man is working hardest in sections A and C, but the roller is moving fastest in section B. The *greatest force* (y'' large) does not occur at the same time as the *greatest speed* (y' large).

One can also test this with a bicycle. Use the highest gear, and note that (on a calm day) one works hard in *getting* the bicycle moving, not in *keeping* it moving.

has passed is represented by AB. Thus AB is Δx. The average speed, $\frac{\Delta y}{\Delta x}$, is therefore found by dividing the length DE by the length AB. AB is the same length as CE. DE divided by CE therefore represents the average speed. And it is clear that DE divided by CE gives us a rough measure for the steepness of the curve between C and D.

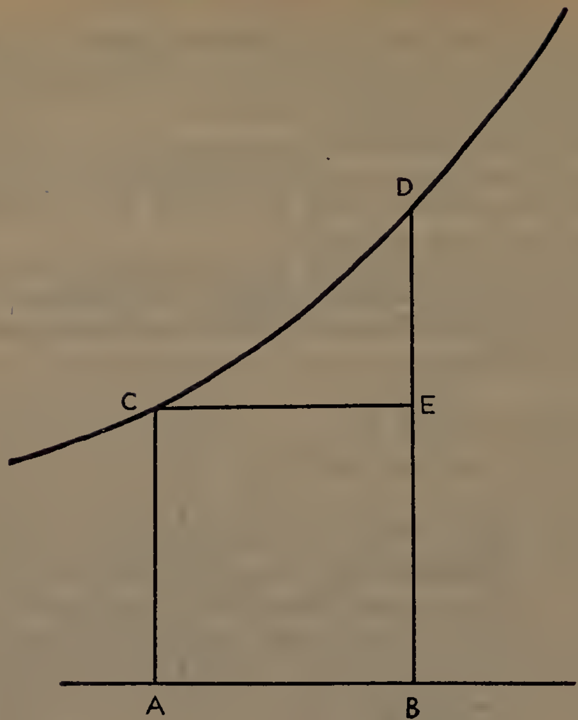


Fig. 10

If instead of a tenth of a second we had taken a hundredth or a thousandth, we should have found a better answer – one nearer to the exact value of y'. So that it is possible to explain what y' is without bringing in the idea of speed at all. Starting with the graph, one takes a point B near to A, draws the figure, and measures DE and CE. One then works out $\frac{DE}{CE}$. Then start again: take B still nearer to A, and work out DE divided by CE for the new figure. As B keeps on getting closer and closer to A, the answer keeps on getting closer and closer to some number. The number which it approaches is y': we may regard y' as measuring the steepness of the curve at the point C.

In this way we have given a meaning to y', quite apart from the idea of movement. The same can be done for y'', and these symbols (as was mentioned earlier) can be applied to problems about the shape of a hanging chain, or the arch of a bridge or the best curve for the tooth of a gear-wheel. It need not surprise you if you find still other applications of y' where neither shape nor speed seems to come in. For instance, x might represent the temperature of a body, and y might measure the amount of heat put into the body. It is then possible to give a meaning to y'. Whenever you meet the symbol y' or $\frac{dy}{dx}$, there must be in the problem two quantities, x and y, which are connected with each other, so that any change in x automatically produces a change in y.

So that our definition of y' as a *speed* may be regarded as a kind of scaffolding. It was helpful at the beginning, but no one keeps the scaffolding up when the house has been built. When you have had some experience of the different applications of y' you will probably be able to think of it simply as the number which $\frac{\Delta y}{\Delta x}$ approaches when Δx is made smaller and smaller. But do not be in a hurry to think of it in this way. Even when you are thoroughly familiar with the idea of y', it will often help you to think of it as a speed, or as the steepness of a graph.

The Use of Rough Ideas

It may not have struck you before what a subtle idea 'speed' is. It was quite easy to explain 'average speed'. A car has gone 30 miles in an hour: its average speed has been 30 m.p.h. This is a simple enough idea. But its speed at the precise moment of a collision? That is a much more difficult idea: it is a much harder thing to measure. We have to perform quite a complicated process. Work out the average speed for the minute before the collision, for the second before, for one-tenth of a second before, and so on. See if the answers approach closer and closer to some number; if so, this number represents the speed at the moment of collision.

Actually to carry out this process, we should have to measure very short periods of time – in turn, a hundredth, a thousandth, a millionth of a second – and the very short distances passed over in these times. And even then we should not be quite sure of getting an exact answer. It would always be possible that in the last billionth of a second the driver pressed the brake harder than before, so that the speed at the last instant was really less than we should have expected from the average speed over the last millionth of a second.

Engineers and pure mathematicians differ in their attitude to this question. The engineer thinks the discussion is a waste of time. He does not mind whether the speed was 50 m.p.h. or 50.00031 m.p.h. He cannot measure speed beyond a certain degree of accuracy, and he does not want to, anyway. Even if the brake is put on harder in the last billionth of a second, it will alter the speed by only a tiny amount. So far as an engineer is concerned, the average speed for the last millionth of a second *is* the speed at the moment of collision.

Why does the pure mathematician not agree with this point of view? It is not entirely due to the passion for hair-splitting which affects some mathematicians. One reason is historical. At first, calculus was treated rather in the spirit of the practical man. Very small periods of time were considered. In working out the average speed, this period of time was supposed to be some definite number, bigger than nothing. But in the answer certain parts came which were not wanted, and mathematicians then turned round and said that the little period was so small that it could be treated as nothing. So that at one time a millionth was a millionth, and at another time (when it gave a more convenient answer!) a millionth was treated as being nothing. Naturally students felt that this was a queer subject. Some mathematicians refused to believe that true results could follow from such a method. So mathematicians were forced to clear up the confusion, and to find a more exact and logical way of explaining what they meant by speed. In modern books on the calculus, written for professional mathematicians, you will therefore find very long and careful proofs, written in exact, lawyer-like fashion. It is good to understand these proofs, but not when starting calculus. First learn to use calculus, to see what can be done with it, to feel what it is about. In the course of this you will gradually become aware of a need for more exact ideas – then is the time to study the modern treatment, usually known by the name of *Analysis*.

There are other reasons for using the exact speed. For one thing, practical men are not agreed on the shortest amount that need be considered. The carpenter works in hundredths of inches, the engineer in thousandths, the scientist in millionths – microbes, atoms, rays of light. For a locomotive engineer one hundredth of a second is a short time; for the radio engineer, who thinks in terms of so many million cycles a second, a hundredth of a second is an eternity. The pure mathematician, whose results may be used by any of these men, can be sure of satisfying every possible demand only by giving the *exact* result.

Again, the exact result is often simpler than the inexact one. When we studied the speed of a body corresponding to the formula $y = x^2$, we found several rough results. For instance, we found that the speed after one second was between 1·99 and 2·01. Suppose we had stopped here, and said 2·01 is good enough as an answer. This is a more complicated result than the exact value 2. If we had throughout treated one-hundredth of a second as a sufficiently short time for our purposes, we should have been led to the formula $y' = 2x + 0{\cdot}01$, roughly. This is more complicated than the exact answer, $y' = 2x$. Even engineers use $2x$ as the speed corresponding to x^2. The simpler answer makes up for the more complicated definition.

There is, as you see, much practical justification for the exactness beloved of pure mathematicians. But this is only one side of the question. There are often cases in which the rough idea is very helpful. Often the rough idea of a problem will enable us to see what the problem means, to see the way to a solution. Our answer may be incorrect by a few millionths, but it will be sufficient to give us a general idea of the solution. We may then be able to go over our working, and polish each stage of the work, until the whole thing becomes exact. Or we may rest content with the rough solution of the problem. Many problems which are difficult by exact methods are studied by research workers, by rough methods. The answers are given true to two decimal places, or whatever it may be, sufficient for the purpose required.

Some Examples of Rough Ideas

Suppose, for instance, we were given the problem of finding $\frac{dy}{dx}$ (that is, y') corresponding to the formula $y = \log x$. This is a new problem. We know how to deal with an expression built up from powers of x, but $\log x$ does not belong to this simple type. What is to be done?

Students who have learnt merely to do simple problems by text-book methods are, of course, helpless in face of a new kind of problem. They sit before it, and do nothing at all. I hope readers will not find themselves in this situation, but that they will already see how to experiment with this problem, how to seek for a solution.

Do we know what $\log x$ is? If you had difficulty with Chapter 6 it is not the slightest use going on with this chapter. Obviously, if you have no clear picture of the meaning of $\log x$, it is absurd to expect to reason correctly about the speed of growth of $\log x$. If necessary, then, read Chapter 6 again. Get a table of logarithms, and draw a graph to illustrate the formula $y = \log x$. (By $\log x$ I mean the ordinary logarithm as given in the usual tables. $\text{Log}_{10}\, x$ is the complete sign. This means – to use the language of Chapter 6 – that 'one complete turn' corresponds to multiplication by 10.) Draw this graph for values of x between 1 and 10, plotting sufficient points to give an accurate graph. You will notice that the graph is most steep at $x = 1$. The larger x gets, the less steep the graph. We shall expect for y' some formula which makes y' get less as x gets larger.

We can obtain a *rough* idea of y' by taking a change of $0{\cdot}1$ in x, and seeing what change this produces in y. Part of the work is shown below. Instead of writing the values of x in rows across the page, it is more convenient to set the work out in columns.

TABLE XIII

x	$y = \log x$	Δy	$\dfrac{\Delta y}{\Delta x}$
1	0·0000	0·0414	0·414
1·1	0·0414	0·0378	0·378
1·2	0·0792	0·0347	0·347
1·3	0·1139	0·0322	0·322
1·4	0·1461	0·0300	0·300
1·5	0·1761	0·0280	0·280
1·6	0·2041		
....	....	....	....
2·0	0·3010	0·0212	0·212
2·1	0·3222		
....	....	....	....
10·0	1·0000	0·0043	0·043
10·1	1·0043		

In the first column are values of x. Each number exceeds the previous one by 0·1. So the change in x, Δx, is always 0·1. The second column gives the logarithms of the first column, found from a table of logarithms. The third column gives the changes in the second column, Δy. The fourth column gives a rough estimate of the speed, y'; the change in y, Δy, has been divided by the corresponding change in x, Δx. Dividing by 0·1 is the same as multiplying by 10. The numbers in the fourth column are thus ten times those in the third. The table given above is not complete: the numbers between 1·6 and 2·0 and those between 2·1 and 10 have been omitted, to save space. These gaps should be filled in by the reader.

The numbers in the fourth column give us a *rough* measure of the steepness of the graph of $y = \log x$. It is clear that the graph becomes less steep as x increases – a fact we have already noticed. The next problem is to find a formula for these numbers. Since the numbers are not exact, we shall be content with a formula that fits them reasonably well – we do not expect an exact fit.

Guessing a formula is often difficult. A new discovery usually requires years. No one should be discouraged if he takes a few weeks to solve a problem of this type. One simply has to try one idea after another until one hits on the right result. It helps if one makes a collection of graphs of different functions. One can draw the graph of the table given, and see which graph it most resembles.

One may find a clue, in our problem, by comparing the result for $x = 1$ with that for $x = 2$. Opposite $x = 1$, we have in the fourth column 0·414. Opposite x = 2 we have 0·212. Now 0·212 is roughly half of 0·414. This suggests that $x = 3$ will correspond to one-third of 0·414, $x = 4$ to one-quarter, and so on. $x = 10$ should correspond to one-tenth of 0·414 – that is, 0·0414. The table gives 0·043, which is not much different. To $x = 1{\cdot}5$ should correspond 0·414 divided by 1·5 – that is, 0·276. The table gives 0·280, which is quite close – as close, at any rate, as we can expect with such a rough method.

This work therefore suggests that y' corresponding to $y = \log x$ is *something like* $\dfrac{0{\cdot}414}{x}$.

This result should be regarded as a sort of hint. It suggests that we go back to the explanation of logarithms given in Chapter 6, and try to see if there is some obvious reason why $\log x$ should grow at a speed which is proportional to $\dfrac{1}{x}$. Actually, there is, and it can be shown that $\dfrac{0{\cdot}434294}{x}\cdots$ is the true answer to the problem. Our rough method has shown us the *form* of the answer, and has brought us reasonably near to the true value.

(In Chapter 6 we explained the meaning of 10^x. What is the formula for y', if $y = 10^x$?)

You may remember that it was mentioned in Chapter 6 that a slide-rule could be made to any scale we liked. If we mark the number 10 one inch from the end of the scale, the number x occurs at the distance $\log x$ inches from the end. But we could take 10 at any other distance, and we could still make a perfectly good slide-rule. Our result for y' corresponding to $y = \log x$ suggests that it might be worth while to alter the scale in a particular way. We found that y' was equal to $\dfrac{0{\cdot}434294\ldots}{x}$. If we took, instead of $y = \log x$, the formula $y = \dfrac{\log x}{0{\cdot}434294\ldots}$ we should have a simpler result: y' would then be simply $\frac{1}{x}$. This new expression y will do just as well for the distance at which the number x has to be marked. If we mark every number x at a distance $\frac{\log x}{0{\cdot}434294\ldots}$ inches from the end of the scale, we obtain a slide-rule that is on a larger scale, but is otherwise no different from the previous one. 10 now occurs at a distance equal to $\frac{\log 10}{0{\cdot}434294\ldots}$ As log 10 is 1, this can be worked out; it is equal to $2{\cdot}30258\ldots$ inches. The number which now occurs at a distance of one inch is $2{\cdot}71828\ldots$ This number is important in mathematics: it is given a special name, and is always spoken of as e. The distance at which any number occurs on this new slide-rule is called the *Natural Logarithm* of the number. The natural logarithm of x is written $\log x$.

When we first explained logarithms in terms of ropes wound on posts, we took the effect of one complete turn to be 10. The only reason for doing this was the accidental fact that we have ten fingers. If we had eight fingers, we would probably take one turn to correspond to 8, and we should get just as good a table of logarithms. For these we should use the sign $\log_8 x$. Any other number could be used. It need not be a whole number. $2\frac{5}{8}$ would do, for instance. All these different numbers would lead to perfectly good slide-rules, but of different sizes. We should always find $y' = \frac{a}{x}$, where a stands for 'some fixed number'. It is natural to prefer that system of logarithms for which $a = 1$. For this reason, in theoretical work we usually employ $\log_e x$, the 'natural logarithm'. If $y = \log_e x$, $y' = \frac{1}{x}$.

The Cartwheel Problem

We now consider another problem, in which a rough idea is helpful. If a wheel – for instance, a cartwheel – is rolling along a flat road, how fast do the various parts of it move? They certainly do not move all at the same speed. You will often notice, when a motor-car passes, that the lower spokes can be clearly seen, but the top spokes move so fast that they are invisible. How is this to be explained?

Many people find a rolling wheel difficult to imagine – not, of course, difficult to imagine in a vague way, but difficult to imagine

Fig. 11

so clearly that the speed of each part can be seen. Let us therefore replace this problem by a simpler one. It is fairly easy to imagine a square rolling, as, for instance, a large log of square section being pushed along the pavement. It starts with one side flat on the pavement. Then it turns about one corner, until the next side is flat on the pavement. Then it turns about the next corner, and so on. It is easy to see that the corner at the bottom of the square – the corner about which the whole thing turns – is at rest. The farther away a point is from this corner, the quicker it moves.

Now let us make our square log rather more like a circle, by four straight cuts with a saw, to remove the four corners of the square, as in Fig. 11. It is still quite easy to imagine the motion. As before, the point which touches the pavement, when the figure rolls, is (at any moment) at rest.

We can continue in this way, cutting off corners, and making the figure more and more like a circle. It will never *be* a circle, but it can be made *as near to* a circle as we like. The figure with 128 corners would make quite a good wheel for most practical purposes.

We are thus led to the conclusion – which every engineer has to know – that a rolling wheel turns about its lowest point, which is (for an instant) at rest.

The same method of approach would enable us to see what curve any point of the rolling wheel describes. The curves known as cycloids, epicycloids and hypocycloids arise in this way.

In the cutting of gear-wheels, a special curve has great advantages. This is the curve described by the end of a cotton thread, as the thread unwinds from a fixed reel. It may help you to see just what the thread does, if you think of the reel as being, not an exact circle, but a figure with corners such as we had above. The more corners you imagine, the nearer you come to the true state of affairs.

In the same way, the motion of a small body rolling down a curved hill can be simplified: we can replace the curve by a figure with corners. If we know all about the behaviour of a small body rolling down a straight line, we can thus build up a picture of a body rolling down a curve.

Again, we can study the behaviour of a hanging string by considering what happens to a hanging chain, made up of straight links like a bicycle chain. The more links we imagine, the nearer we come to a true idea of the string.

All these illustrations use what mathematicians call the idea of a Limit. Mathematicians speak of a circle as being the *limit* of the figures with corners described above: this simply means that you can get *as near as you like* to a circle by making *a sufficiently large number* of straight saw cuts. You can get *as near as you like* to the curve of a string by taking *sufficiently short* links for your chain.

You can get *as near as you like* to the speed $\frac{dy}{dx}$, by taking *sufficiently small* Δx and calculating $\frac{\Delta y}{\Delta x}$.

ILLUSTRATIONS AND EXPERIMENTS

1. The movements of trains on railways are recorded on a graph, when time-tables are being worked out, or emergency trains are fitted in (See J. W. Williamson, *Railways Today*, for a reproduction of such a graph.)

Draw graphs showing the position of the following trains:

(i) Leaves London at 12.00; travels at steady speed of 20 m.p.h.

(ii) Leaves London at 12.00; travels at 30 m.p.h., but stops for 5 minutes every half-hour.

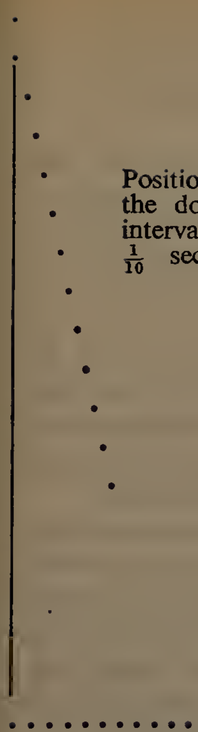
(iii) Leaves London at 12.48 and travels at 50 m.p.h.

In these graphs, have distance going *upwards*, the passage of time towards the *right*. See for yourself that the faster trains have steeper graphs, and that this makes it possible for the express (Number iii) to overtake the goods train (Number i).

2. A cricket ball is thrown straight up into the air. After x seconds its height is y feet, where $y = 40x - 16x^2$. Work out y' and y''. Show that y'' is always minus, and that y' starts by being $+$, but later becomes $-$. At what time is $y' = 0$? Draw the graph for the motion of the ball. When is the ball at its greatest height? What is this height?

3. A piece of wood is hurled downwards into the sea. After 3 seconds it again appears on the surface. A scientist states that during this period its graph had the equation $y = -30x + 10x^2$, where x stands for the time in seconds after it entered the water, and y for its height in feet (measured upwards, so that depths below sea-level are minus). How far is this a reasonable description of what the piece of wood does?

4. A body is shot from a catapult (it does not matter whether it is a pebble or a glider), and its motion is recorded by means of



Position of the dog at intervals of $\frac{1}{10}$ second.

Position of man at intervals of $\frac{1}{10}$ second.

a graph. How would you expect this graph to bend (i) while the body is still in the catapult and gathering speed; (ii) after it has left the catapult?

5. Find out the general appearance of the graphs corresponding to the formulae below, by working out y' and y'' and seeing what signs (+ or —) they have. Consider x from —10 to 10.

(i) $y = x^2 + x$. (ii) $y = x^3$. (iii) $y = x^3 - x$. (iv) $y = x - x^2$. (v) $y = x^3 - x^4$.

Be careful to note that where $y' = 0$ the curve is, for an instant, *level*. This applies particularly to (ii) and (v).

6. The following example can be dealt with by the method of Rough Ideas.

A man is walking along a straight path, which points due East, at a speed of 5 feet a second. The man whistles to his dog, which is at that moment 30 feet due North of the man. The man does not stop walking. The dog runs, at a speed of 20 feet a second, always facing towards his master. Draw roughly the curve which the dog describes. About how long does the dog take to reach the man? (It is clear that if the dog had given a little more thought to what it was doing it would have run in a straight line, towards a point somewhat ahead of its master, instead of behaving in this way.)

Method. Replace the steady movement of the man by a series of jerks. Suppose the man to remain still for one-tenth of a second, then suddenly to shoot forward six inches. In this way the man will cover 5 feet a second. The dog will run in a chain of straight lines, each 2 feet long, pointing in turn towards the various positions in which the man stands. Drawing this chain we can see roughly how the dog moves and how long he takes to reach the man.

CHAPTER 12

OTHER PROBLEMS OF CALCULUS*

'One of the hardest tasks that an expert in any subject can undertake is to try to explain to the layman what his subject is, and why he makes such a fuss about it.'

– G. C. Darwin, Introduction to *The Story of Mathematics*, by D. Larrett.

THE pioneers of every subject are amateurs. They start with the same knowledge and the same methods of thinking as any other uninstructed person. The early discoveries in any branch of science can be explained in everyday language, and often seem so obvious that the science appears hardly worth studying.

Later generations, building on the work of the pioneers, study more and more complicated problems, and in the course of this introduce new ideas, new words, technical terms which are Greek to the layman. The later discoveries of any science have to be stated in technical terms: they appear remote from ordinary life, and extremely hard to understand. The science now seems so difficult that it appears not worth while to try to master it.

In mathematics, as in other sciences, each generation builds upon the foundation provided by past workers, and adds another storey. By now the building is something of a skyscraper. There are many remarkable books on the eighteenth floor, but they are written in a language which is comprehensible only to those who are thoroughly familiar with the works on the seventeenth floor – and so on, floor by floor, until one reaches the ground floor and the multiplication table.

No living person is familiar with all the mathematical discoveries that are stored in the libraries of the various learned societies throughout the world. Every mathematician has to find out for himself just which parts of the subject are useful for his own purposes, which technical terms and ideas he has time to become familiar with, and to apply to his own problems.

* This chapter may be skipped by those who find it hard.

In this chapter a number of processes will be discussed which are useful to those engaged in the more exact sciences – physics, chemistry, engineering – and to those who use any kind of machine, from a drill to an aeroplane. This type of mathematics is also creeping in to subjects such as biology, economics, even psychology. If you are not interested in such applications, you will perhaps not find this chapter of interest, nor will you find much point in learning calculus at all, unless you belong to the type which is interested in mathematics for itself. If you neither like, nor need, calculus it is a sheer waste of time to study it.

In the course of scientific work it is frequently necessary to find $\frac{dy}{dx}$ for expressions $y(x)$ which are more complicated than any we have so far considered. One might, for instance, come across an expression such as $y = \left(\frac{x}{x^2 + 1}\right)^3$, and wish to know y' corresponding to this. Here a whole series of processes has been carried out. If we start with x, we have to work out x^2 and add 1 to this, giving $x^2 + 1$. Then x has to be divided by this result. The answer to this has to be raised to the power 3.

The problem is dealt with by splitting it up. We bring in new letters, and break up the chain of processes. In calculating y we were first led to calculate $x^2 + 1$. We will call this first result u. $u = x^2 + 1$. How quickly does u grow? We know this from our earlier work: $u' = 2x$. We then calculate $\frac{x}{x^2 + 1}$ ($x^2 + 1$ being the same thing as u). Call this result v, so that $v = \frac{x}{u}$. Now we know that u grows at the rate u', and that x grows at the rate 1. v is obtained by dividing x by u. Since we know x and u, and we know how fast each of them is growing, it ought not to be too difficult to find how fast v is growing. Suppose we solve this problem, and find the answer, v'. We now come to the final stage. y is obtained by raising v to the power 3; that is, $y = v^3$. y is v^3 and v is growing at the rate v'. How fast is y growing?

So far the problem has not been solved. We have simply shown that the one complicated problem can be split up into three simpler ones: I. To find u' when $u = x^2 + 1$. II. To find v' when $v = \frac{x}{u}$ and u' is known. III. To find y' when $y = v^3$ and v' is known.

It is because complicated problems can be split up in this way into simple ones that you find certain theorems in every calculus text-book, dealing with the differentiation of a Sum, Product, Quotient, and the Function of a Function. All these theorems have an object – to enable you to find y' corresponding to any formula, however complicated, by splitting the problem into simpler ones.

Differentiation of a Sum

Consider an example – rising prices. Let £y be the price of a watch after x days of war (rising at the rate y') and £z the price of a chain (rising at the rate z'). How quickly does the price of a watch and chain rise? Clearly $y' + z'$. As the price is £$(y + z)$, this shows how easy it is to find the rate of increase of the sum of two changing quantities.

Differentiation of a Product

Let n be the number of men in a town, and p the number of pints drunk daily by each man. Then np is the total number of pints drunk. If n is increasing at the rate n' and p at the rate p', how fast is np increasing? The answer is $p'n + n'p$.

Differentiation of a Quotient

If b barrels of beer are provided for n men, each will receive $\frac{b}{n}$ of a barrel. If the number of men increases at the rate n' and the number of barrels at the rate b', how fast does $\frac{b}{n}$ change? The answer turns out to be $\frac{b'n - n'b}{n^2}$. Notice how this answer compares with common sense. If $n' = 0$, it means that the number of men stays the same, and if b' is $+$, it means that the number of barrels is increasing. In that case, $\frac{b}{n}$ is increasing, and its rate of change should be $+$. This is so, for the formula above. If, on the other hand, the number of barrels stays the same, $b' = 0$, while the number of men increases, so that n' is $+$, the formula becomes $\frac{-n'b}{n^2}$, which has a minus sign, as it ought to do: the share per man is getting *less*, the change is for the worse, a *minus* sign is to be expected.

Function of a Function

It is well, at this stage, to look back to page 96 and to read again the sentences explaining what is meant by y being a function of x; namely, that y is connected with x by some rule. What now is 'a function of a function'? Consider the formula $y = \log_e (x^2 + x)$. We could make a table of y in the following way. In the first column we could enter the numbers, x. In the second column we could enter the corresponding numbers, $x^2 + x$. In the third column we could put the logarithms (to base e) of the numbers in the second column. This third column would then give the numbers $\log_e(x^2 + x)$. We have x in the first column, y in the third column. Let us call the numbers in the middle column z. The numbers in the second column are found from those in the first column by a definite rule. So z is a function of x. The numbers in the third column are found by a definite rule from those in the second. So y is a function of z. It is this process which gives rise to the name 'function of a function'. In fact, $z=x^2+x$ and $y = \log_e z$.

Now we know all about the rule connecting x with z, and we know all about the rule connecting z with y. It ought not to be too difficult to find how fast y increases.

It is possible to illustrate this double connexion by a machine. The relation $z = x^2 + x$ can be expressed by means of a graph. In Fig. 12 the curve OB represents a groove cut to the shape of this graph. OA and OC are straight grooves. A represents a small piece of metal sliding in the groove OA. In the same way B

slides in the groove OB, and C in the groove OC. A small ring is fixed to B, and through this ring pass the two rods, AB and CB, which are soldered to the sliding pieces A and C in such a way that AB is always upright, and BC is always level. If A moves, B is forced to move, and this in turn forces C to move. The distance OA represents x, the distance OC represents z. Any change in x produces a change in z, and the machine is so designed that $z=x^2+x$.

In the same way we can express the connexion, $y = \log_e z$. y is represented by the length OE. z is already shown by OC. The curved groove GD is the graph of $y = \log_e z$. The rod CD is always level, while ED is always upright. Both pass through a ring at D.

The whole chain of events can now be seen. If x (OA) changes, z (OC) cannot help changing, and because OC changes, OE (y) must also change.

How fast? We know that z increases $\frac{dz}{dx}$ times as fast as x. y increases $\frac{dy}{dz}$ times as fast as z. So y must increase $\frac{dy}{dz}.\frac{dz}{dx}$ times as fast as x. That is, $\frac{dy}{dx}=\frac{dy}{dz}.\frac{dz}{dx}$. This is the theorem about a Function of a Function and its rate of change.

In our particular example it is easy to find $\frac{dy}{dx}$. We have to find $\frac{dy}{dz}$ and $\frac{dz}{dx}$ and multiply these together. There is no difficulty with $\frac{dz}{dx}$. $z = x^2 + x$, so $\frac{dz}{dx} = 2x + 1$. Now for $\frac{dy}{dz}$. $y = \log_e z$. We saw in Chapter 11 that $\log_e x$ grew at the rate $\frac{1}{x}$. 'The natural logarithm of any number grows at a rate given by dividing one by that number' is this formula in words. It makes no difference if we call the number z instead of x. So $\frac{dy}{dz}=\frac{1}{z}$. Accordingly

$\frac{dy}{dx}=\frac{1}{z}(2x + 1)$. But z is short for the number in the second column, $x^2 + x$. This answer is therefore the same as $\frac{2x+1}{x^2+x}$ and this formula is the solution of the problem.

By combining the results stated above, it is possible to find $\frac{dy}{dx}$ for very complicated formulae.

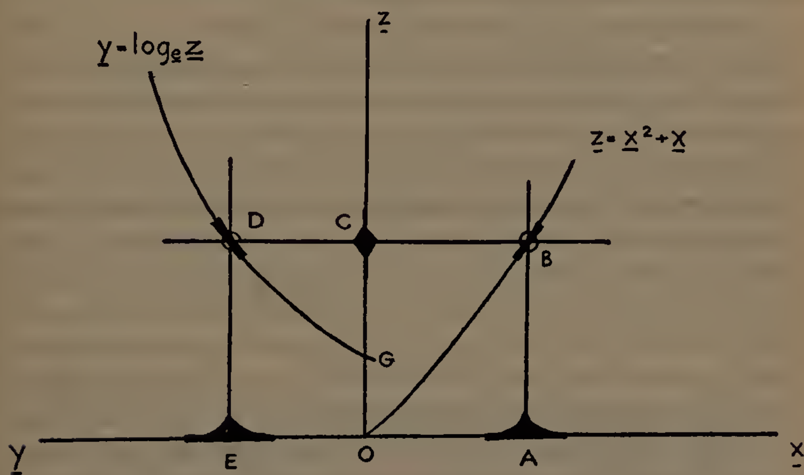


Fig. 12

Integration

We have already considered the problem of differentiation – that is, the problem of finding the speed y' of a body which moves in a manner described by means of a formula for y. The opposite problem frequently occurs: we know the speed at every moment, we have to find how far the body goes after any number of seconds. In other words, we are given a formula for y', and we are asked to find a formula for y. This is the problem of *integration.*

y' need not, of course, be considered as the speed of a moving body. It represents the rate of change of y, whatever y may be. For instance, it is easy to discover *how quickly* the pressure increases as a diver goes deeper into the sea. So y might represent the pressure per square foot on a diver's helmet, when he is at a depth x feet. It is easy to find the rate of increase, y'. In order to find y, the problem of integration (very easy in this particular case) has to be solved. Integration is also used in connexion with the question of air pressure at different heights, a question of interest to mountaineers, airmen, weather experts, and others. There are few, if any, branches of science and engineering in which problems of integration do not arise.

In dealing with any practical problem, a student has to do two things. First, the problem has to be put into mathematical form: then the mathematics necessary to solve the question has to be carried out. The second part is no use without the first. Our study of integration will therefore have two objects: (*a*) to understand the nature of integration so clearly that we immediately recognize any problem that can be solved by means of integration, (*b*) to master the mathematical method. The first part, (*a*), can be understood without any knowledge of the second part, (*b*). We are at present mainly concerned with (*a*), though some passing reference may be made to (*b*). We shall consider a very simple problem – one which can be solved mathematically in two lines – and look at it from all angles. To this simple problem we shall apply methods capable of solving far more difficult questions: we shall use steam hammers to crack a walnut, in fact. The object of this will not be to crack the walnut, but to demonstrate how the steam-hammers work. Our simple problem is the following: if $y' = x$, find a formula for y. This problem is not quite complete, as it stands. We are given y' which might represent the speed of a body after x seconds. Obviously, we must know where the body starts, if we are to work out its position. We will suppose, then, that we are also told that $y = 0$ when $x = 0$. The problem is quite definite. We can think of y as the distance of the body from a fixed point P. To begin with, the body is at P, since the distance y begins by being zero. Then the body begins to move. After 1 second its speed is 1 foot per second; after 2 seconds, its speed is 2 feet per second, and so on. The speed does not grow by jumps, but steadily. For $y' = x$ tells us that the speed after $1\frac{1}{8}$ seconds is

$1\frac{1}{8}$ feet per second; after $1\frac{1}{4}$ seconds, y' is $1\frac{1}{4}$, and so on. We have a complete picture of the motion.

The Method of Rough Ideas

Let us try, first of all, to get a rough idea of the distance which the body would go, for example, in the first second. We will split the second into ten equal parts, and see how much we can find out about the distance which the body travels in each tenth of a second. In the first tenth of a second the body moves with a speed which increases steadily from 0 at the beginning to 0·1 at the end. So the average speed lies between 0 and 0·1. So the distance gone must be more than 0 times 0·1, but less than 0·1 times 0·1. We can apply the same argument to each of the other parts. The distance gone in 0·1 second is more than 0·1 times the least speed, and less than 0·1 times the greatest speed in that part of the motion. We can put the argument in the form of a table.

TABLE XIV

Time	Least Speed.	Highest Speed.	Distance gone At least.	At most.	Difference.
0 to 0·1	0	0·1	0	0·01	0·01
0·1 to 0·2	0·1	0·2	0·01	0·02	0·01
0·2 to 0·3	0·2	0·3	0·02	0·03	0·01
0·3 to 0·4	0·3	0·4	0·03	0·04	0·01
0·4 to 0·5	0·4	0·5	0·04	0·05	0·01
0·5 to 0·6	0·5	0·6	0·05	0·06	0·01
0·6 to 0·7	0·6	0·7	0·06	0·07	0·01
0·7 to 0·8	0·7	0·8	0·07	0·08	0·01
0·8 to 0·9	0·8	0·9	0·08	0·09	0·01
0·9 to 1·0	0·9	1·0	0·09	0·1	0·01
		Total	0·45	0·55	0·10

In the first column we have the ten parts into which the first second is divided. Then follow the least speed and the greatest speed in each part of the motion. We then have two columns showing 0·1 times the least speed, and 0·1 times the greatest speed. In each tenth of a second the body must have gone more than the former, less than the latter. The last column shows the difference between the two previous ones. For instance, we know that, in the time between 0·6 and 0·7, the body goes at least 0·06, at most 0·07. The difference between these is 0·01, so we are left uncertain about the distance gone in this time, to the extent of one-hundredth. The total distance gone in the first second is found by adding up. It must be more than 0·45 foot, less than 0·55 foot.

We now have a rough idea how far the body goes in the first second. The difference between 0·45 and 0·55 is 0·1. This uncertainty of 0·1 is due to the uncertainty of 0·01 in each of the ten rows. If we want a more accurate answer, it will be necessary to carry through the same process, using smaller intervals. We might, for instance, divide one second into 100 parts and then carry out a similar calculation, though of course this would be rather long and boring to do. How close would such a method bring us to the true answer! The difference between the highest and the least speed in any small part would be 0·01, instead of 0·1. The distance gone in 0·01 second is found by multiplying the speed by 0·01. So the uncertainty in the distance gone, in any hundredth of a second, would be 0·01 times 0·01 – that is, 0·0001. But there would be a hundred rows in the table (instead of ten) and the uncertainty in the total distance would be 100 times 0·0001, that is 0·01. (Actually we should find that the distance was more than 0·495 and less than 0·505.) This result is ten times as good as the time before – we are repaid for having to do ten times as much work to obtain it. By taking still more intervals, we could get still better results.

This method is used only when a problem is so difficult that no other method will work. Even then, some shortening of the work would be arranged. The method is not given here as a good way of finding the actual answer, but rather to show what the problem means. The process above will help you to understand the sign used for integration. We have already used the sign Δx for the change in x. In the table given above, each row of the first column represents a change of 0·1 – e.g., 0·7 to 0·8. The next two columns tell us the least speed and the highest speed – that is, they help us to see how large y', the speed, is, during this part of the time. In the same way, the fourth and fifth columns give us a number rather less, and a number rather more, than the distance gone during any small part of the time. As distance gone is measured by the speed (y') times the time that passes (Δx), we may think of these columns as representing $y' \Delta x$. Of course there is some doubt about the meaning of y': for instance, as x goes from 0·6 to 0·7, y' also goes from 0·6 to 0·7, and it is not clear whether we should take y' as being 0·6 or 0·7 or some number in between. It is because of this uncertainty that we have the two columns, one headed 'At least', the other 'At most'. This must be borne in mind.

We then estimate the distance gone in the whole of the first second by adding up the fourth and fifth columns. So that the 'sum of $y'\Delta x$' is (at least) 0·45 and (at most) 0·55.

Here we have two estimates – one rather too small, one rather too large. But, fortunately, by taking smaller lengths of time – that is, taking Δx to be 0·01 or 0·001, etc. – these two estimates approach closer and closer to each other. In other words, if Δx is made very small, it matters *very little* whether we take y' to be the highest or the lowest speed that occurs in the interval of time, Δx. The answer will be the same, whichever is taken. If this were not so, we should have to bring in a new sign, such as $y'\Delta x(L)$, to mean 'the *least* speed, y', times the change in x, Δx'. But, as it has turned out, this would be a waste of time. The least speed and the highest speed give answers that draw closer and closer together as Δx gets smaller.

In Chapter 10 it was mentioned that $\frac{\Delta y}{\Delta x}$ came nearer and nearer to a certain number as Δx became smaller. The number thus approached was therefore christened $\frac{dy}{dx}$. In the same way, the number we have just been estimating – the number which is more than 0·45 and less than 0·55, more than 0·495 and less than 0·505, etc. – is represented by $\int_0^1 y'dx$. The sign $\int$ is an old-fashioned S, S for 'Sum'. The sign is meant to indicate that the number can be found by multiplying y' by Δx for each short part of the time, finding the sum of these, and seeing what happens when Δx gets very small. The numbers 0 and 1 are put in to show that we are interested in the distance gone during the first second – that is, between $x = 0$ and $x = 1$. In other words, the change of x from 0 to 1 has to be split up into little changes Δx, as in the first column of the table. $\int_2^5 y'dx$ would represent the distances gone in the period of time from 2 seconds after the start to 5 seconds after the start. $\int_0^n y'dx$ represents the distance gone in the first n seconds. Since we have supposed y' to be given by the formula $y' = x$, we may replace y' by x, and write $\int_0^1 x\,dx$. It seems likely, from the work above, that the number we are looking for is 0·5, and it can be proved that this is the correct answer. In symbols, $0{\cdot}5 = \int_0^1 x\,dx$. $\int$ is known as 'the integral sign'. 'To integrate' means 'to make whole': the name, I suppose, is chosen because the process consists in putting together a lot of little bits, all the little changes in y that occur in the brief passing moments of time.

Ways of Seeing Integration

It is useful to know different ways of getting the same result. Then, in any problem, one can imagine the illustration which is most convenient.

We may translate the symbol $\int_0^1 x\,dx$ as the distance gone, in the first second, by a body whose speed is always equal to the number of seconds that have passed (including fractions of a second).

To record such a motion we might use the device illustrated at the beginning of Chapter 11. It would be rather difficult to arrange for the speed y' always to be exactly equal to the number of seconds, x. Let us once more be content with a rough idea. During the first tenth of a second let the pencil-point stay at rest

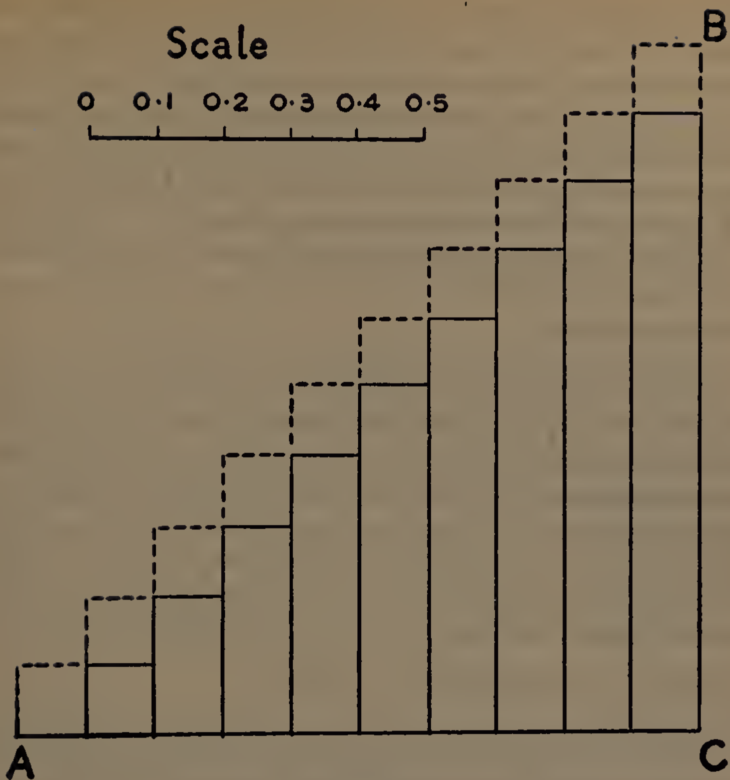


Fig. 13

in the slit *AB*. During the next tenth of a second let it move upwards with a speed of 0·1 foot a second: from $x = 0{\cdot}2$ to $x = 0{\cdot}3$, let the speed be 0·2 foot a second, and so on. In fact, let the speed during any interval given in the first column of Table XIV be the number given in the second column of that table. The graph that results will consist of straight lines, joined together like a chain. As we saw in Chapter 11, y' measures the steepness of these lines. As y' increases steadily, each line will be steeper than the previous one. So we could equally well have explained integration from the problem: You are told how steep a curve is at every point, draw the curve.

It is also possible to illustrate Table XIV directly. The numbers in the fourth column are obtained by multiplying those in the second column by 0·1. But multiplication – for instance, 0·9 times 0·1 – can be represented by the area of a rectangle, with sides 0·9 and 0·1. The ten numbers in the fourth column can be represented by the area of ten rectangles, as in Fig. 13. The sum of these numbers, 0·45, represents the total area, below the heavy line. In the same figure the area below the dotted line represents 0·55, the sum of the numbers in column five.

Between the dotted line and the heavy line are ten squares, each with an area equal to 0·01. These squares represent the numbers in the last column.

We know that $\int_0^1 x\,dx$ represents a number larger than the area below the heavy line, less than the area below the dotted line. By taking 100 instead of ten steps, we should get an even better idea of the number we want. But however many steps we have, the heavy line always lies below the straight line AB, the dotted line always lies above it. If we draw the line AB, the area of the triangle ABC is always more than the area below the heavy line, less than the area below the dotted line. In fact, the area ABC is equal to the number we are looking for, $\int_0^1 x\,dx$.

This result is quite general. If $f(x)$ is any function of x, then $\int_a^b f(x)\,dx$ will always represent the area below the graph of $f(x)$, between $x = a$ and $x = b$. *The problem of finding the area inside any curve is a problem in integration.* You should try for yourself to draw an area which represents $\int_0^1 x^2\,dx$.

The connexion between integrals and areas is useful in two ways. We can use an area to illustrate the meaning of an integral, and to help us to understand the behaviour of integrals. Secondly, we can find the actual size of a particular area by working out the value of an integral.

A Shorter Method

The integral, $\int_0^1 x\,dx$, can be found with very little work. We started the problem by trying to find y such that $y' = x$, and $y = 0$

when $x = 0$ (see page 166). But we know that to the formula $y = x^2$ there corresponds the speed $y' = 2x$. $2x$ is just twice the answer we want for y'. We can put this right by taking y half as large – that is, we consider the formula $y = \frac{1}{2}x^2$. This gives exactly the right answer, $y' = x$. Also, $\frac{1}{2}x^2$ equals 0 when $x = 0$, so the condition $y = 0$ when $x = 0$ has been met. So $y = \frac{1}{2}x^2$ is the formula we want. It gives the distance y corresponding to x seconds. Putting $x = 1$, we find $y = \frac{1}{2}$. So the distance gone in one second is $\frac{1}{2}$. This agrees with the result 0·5 which we found by the other method.

Many problems of integration can be solved by this method. The idea is simple enough. We have already learnt how to find y' corresponding to many different types of function, y. We are now asked to do the opposite problem: y' is given, we have to find y. It is natural to turn back to our records of the first problem. If in these we find y' of the type required, the problem is immediately solved. For instance: we have shown that to $y = \log_e x$ corresponds $y' = \frac{1}{x}$. If we are asked to find $\int \frac{1}{x}\,dx$, this is the same as saying: if $y' = \frac{1}{x}$, what is y? Obviously, $y = \log_e x$ gives *an* answer to this question. The complete answer will depend on the other condition: it is not enough to know how fast a body moves, one must also know where it is at some instant.

Differential Equations

Very many practical problems lead to what is known as a differential equation. The nature of a differential equation may best be seen by considering a definite example.

The light from an electric lamp spreads out equally in all directions. Often – as, for instance, in making a motor headlamp or a searchlight – this is inconvenient: we would prefer to have all the light coming out in one direction, and this is achieved by placing a reflector behind the lamp. If the reflected light is to come out in a perfect beam, what shape should the reflector be?

It is known how light behaves when it strikes a mirror. If we take a capital V and underline it, thus V, we have a rough picture of what occurs. The line represents the mirror: the left-hand arm of the V may represent the light striking the mirror, and the right-hand arm the light bouncing off the mirror. The two arms of the V must make the same angle with the line of the mirror. A billiard ball bounces off a cushion in more or less the same way, if it is free from spin.

We shall not get a proper beam if we simply put an ordinary straight mirror behind the lamp. The reflected light will scatter in different directions, as can be seen by drawing a figure.

We might tackle the problem by taking a large number of short pieces of mirror and trying to join them in a chain, in such a way as to get a proper beam. In Fig. 14 P represents the point where the electric lamp is placed. O is some other point, and we want to obtain a beam of light pointing in the direction OP. OA represents a short piece of mirror, so placed that the light from P which strikes the mirror at O is reflected back along the line OP. Of course the light from P which strikes the mirror between O and A will be reflected slightly upwards, slightly away from OP, but if the length OA is short, this will not be serious. When we reach A, we join on our next piece of mirror AB, in such a way that the ray of light PA is reflected in the proper direction – that is, parallel to OP. In the same way, we must turn the next piece of mirror, BC, in such a way that the ray PB is reflected parallel to OP. And so the process is continued: each mirror is added in such a way that its lowest point touches the highest point of the mirror before and it is then turned so as to reflect light in the proper direction.

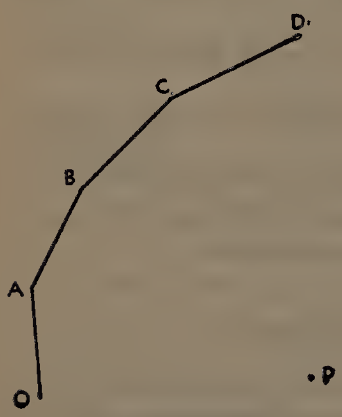


Fig. 14

In this way we could build up a mirror which very nearly produces a perfect beam. The shorter the pieces of mirror used, the better the beam would be. We can easily believe that there is a curve which reflects the light exactly in the correct direction. This curve is known as a *parabola*. The mirror built in this shape is called a *parabolic mirror*. Parabolic mirrors are used in some types of telescope, and in beam wireless upright wires arranged in the shape of a parabola are used.

Notice how we built up our chain of lines *OABCD* ... We started from *O*, and then, at each stage, we were told *in what direction to go*. Any problem which starts from some rule about the direction to follow at any moment will lead to a differential equation.

For instance, a ship at sea might steer straight towards a lighthouse. It could start wherever it liked, but once it had started the direction it had to follow would be fixed. The lighthouse might be spoken of as a magnet attracting the ship: in the language of magnetism the ship follows a 'line of force'. The problem would become more complicated if we had two magnets, each attracting a moving body. The path of the body would not then be obvious, as it was for the ship and the lighthouse. Differential equations therefore appear in the theory of electricity and magnetism.

What does a differential equation look like in algebraic symbols? We have some rule that gives us the direction of the curve at any point. We might equally well say we have a rule that gives the steepness of the curve at any point. Now, the steepness of the curve is measured by y' and the position of a point on a graph is measured by the two numbers x and y. To every point there corresponds a direction: we might imagine this by supposing the graph-paper to be covered with little arrows, signposts conveying the message, 'If you should arrive at this point, depart in this direction'. By continually following the signposts, one would follow out some curve. The signposts are arranged *according to some rule*: if we have any point (corresponding to any two numbers x and y) we have a rule giving the direction of the signpost, and the steepness of the arrow is measured by y'. So y' is given by some rule – that is, we have a formula giving y' when x and y are known.

For instance, if a lighthouse is placed at the point (0, 0), and all ships sail straight towards it, the formula is $y' = \frac{y}{x}$. For the steepness of the line joining any point (x, y) to the point $(0, 0)$ is $\frac{y}{x}$, and y' must be equal to this.

You will not be able to follow this argument if you are not familiar with Co-ordinate Geometry: you must master the earlier part of Co-ordinate Geometry (the plotting of points, the steepness of straight lines, the angles between straight lines, the distance between two points, the equation of a circle) before you try to learn the theory of differential equations.

EXAMPLES

The treatment of the subjects in this chapter is too sketchy to justify the setting of examples. Readers who have been able to follow the general ideas of this chapter will find examples in any text-book on Differential and Integral Calculus.

CHAPTER 13

TRIGONOMETRY, OR HOW TO MAKE TUNNELS AND MAPS

'*The utmost care must be taken to avoid errors, and that it is taken is proved by the wonderful accuracy with which the headings driven from opposite ends usually meet ... The Musconetcony Tunnel is about* 5,000 *feet long. When the headings met the error in alignment was found to be only half an inch, and the error in level only about one-sixth of an inch. In the Hoosac Tunnel,* 25,000 *feet long, the errors were even smaller.*'

A. Williams, *Victories of the Engineer*.

In this book I have tried to show (i) that mathematical problems can be stated in the language of everyday life; (ii) that any normal person can think about these problems for himself, by using common sense; and (iii) that the methods given in text-books simply represent the improvements on the original common-sense attack, which have gradually been built up by generations of mathematicians.

In no part of mathematics is this easier to show than in trigonometry. Trigonometry arises from very simple practical problems, such as the building of a railway tunnel, for instance. It may be necessary to make a tunnel which is to come out several miles away, on the other side of a range of mountains, at a point which cannot even be seen from this side. It may be necessary to bore the tunnel from both ends, and to meet somewhere far inside the mountain. How are we to find the correct direction in which to bore?

One method is explained in Chapter 4 of *The Railway*, by E. B. Schieldrop. It is illustrated in Fig. 15. The shaded part represents high ground. It is desired to connect the points A and D by a

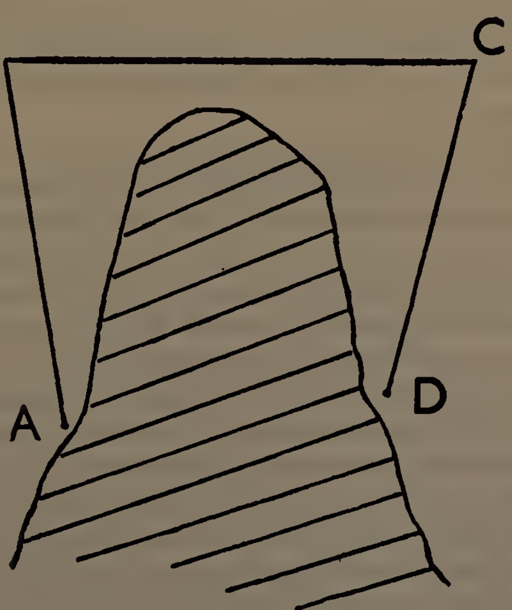


Fig. 15

tunnel. We may be able to find objects *B* and *C*, such that *B* can be seen from *A*, *C* from *B*, and *D* from *C*. We measure carefully the length and direction of the lines *AB*, *BC*, and *CD*.

This information is sufficient to fix the position of *D*. It would enable us to make a map of the district, showing *A*, *B*, *C*, and *D* on a scale, say, 1 foot to a mile. We start with *A*, and draw the lines *A* to *B*, *B* to *C*, *C* to *D*, in the proper directions, and to scale. This fixes *D*. On the map we can draw the line *AD*, and measure the angles which it makes with *AB* and *CD*. We now know in what directions to bore at *A* and *D*.

This method shows that the problem can be solved by a common-sense method, though not very accurately. Since we are working to the scale of 1 foot to a mile, an error of $\frac{1}{100}$ inch in our drawing will lead to an error of $1\frac{1}{2}$ yards in the actual work. And in drawing the figure we may easily make several mistakes of $\frac{1}{100}$ inch. Drawing a figure, then, is not sufficient to give a really good answer, but it gives us a general idea of what is needed. First attempts often turn out like this: they give us the germ of an idea: we have to work at this idea until it becomes practical. A mathematical invention goes through the same stages as a mechanical one: first an idea, then a toy, then a commercial proposition.

Trigonometry represents an attempt to improve on the method of drawing. The argument runs rather on the following lines. By drawing the map on a larger scale, we could get a more accurate answer to the tunnel problem. There seems to be no limit to the accuracy we could get by taking our plan large enough. Given the lengths and directions of the lines *AB*, *BC*, *CD* to a high degree of accuracy, we could find (by drawing on an immense scale) the length and direction of *AD* to a high degree of accuracy. It seems likely that some rule connects the answer with the facts given. We could collect information on the problem, taking *A*, *B*, *C*, and *D* in different positions, and trying to see how the length of *CD*, and its direction, depended on the other measurements given. The aim would be to find the rule: once we had this rule, we could work out *AD* to as many decimal places as we chose, without doing any drawing at all.

In trigonometry, then, we consider problems which could be solved by drawing, problems which therefore possess a definite answer (it is a waste of time to try any problem by trigonometry if you are not given sufficient facts to solve it by drawing: trigonometry is not magic): we try to discover *what rule* gives this answer, so that we may be able to find the answer by a formula, instead of by drawing. The aim is therefore to replace drawing by calculation.

Such a question can be tackled, in the first place, only by experiment. Lengths, directions – these are real things. They will not take orders from us: they follow the laws of their own nature. We can find what they do only by *observing* them.

But, of course, we shall not begin by experimenting with such a complicated problem as that of the tunnel and the four points A, B, C, D. It is rarely wise to attack a problem of a *new type* directly. It is better to make up a much simpler problem of the same type, experiment with that, and see if the method which solves the simple problem throws any light on the complicated one. In map-making, the simplest problems deal with three points only (hence the name trigonometry, i.e., three-line-ology). In particular, triangles containing a right-angle are easy to study.

The Measurement of Angles

It is easy enough to measure the length of a line. It is not so obvious how an angle is to be measured. Two methods are used.

The first method, measurement in degrees, has something in common with the markings on a clock-face. On a clock the numbers 1 to 12 are evenly spaced out, around the rim of the clock. If the hour-hand goes from 12 to 3, we know it has gone a quarter of the way round. To obtain degrees, the circle has to be divided, not into 12, but into 360 equal parts. Each part is called a degree. There is no deep reason for choosing the number 360. Turning through one quarter of the circle (a right angle) corresponds to 90 degrees, usually shortened to 90°. From 12 to 1 on the clock is 30°.

The second method is known as *radian* measure, and is particularly suitable for questions connected with speed. It may be explained as follows. Suppose we have a wheel, 1 foot in radius, fixed to an axle. A string passes round the rim of the wheel, one end being fixed to the wheel, rather like the rope on a capstan or a crane. By pulling the string, we may cause the wheel to revolve. It is clear that we can measure how much the wheel has turned by measuring the amount of string unrolled. When 1 foot of string has been unrolled, the wheel is said to have turned through *one radian*. When x feet have been unrolled, the wheel has turned through x radians.

It is easy to measure an angle in radians. We take a piece of wood, made in the shape of a circle of 1 foot radius. To measure a given angle, we place the centre of the circle. O, at the point of the angle, and mark the points A and B where the lines cross the rim (Fig. 16). We then wind a tape measure *around the rim* (*not* straight) and measure the distance from A to B. If this distance is $\frac{5}{8}$ foot, the angle is $\frac{5}{8}$ of a radian. If we are told that the hand of a clock turns through 10 radians, we measure 10 feet round the rim. We shall, of course, complete more than one turn: where we end up gives the angle 10 radians. If a wheel, of radius 1 foot, with a fixed centre, is turning at the rate 1 radian a second, any point

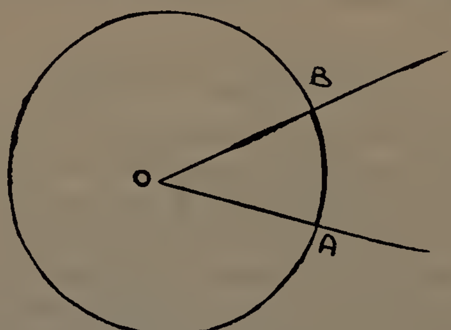


Fig. 16

on the rim is moving at 1 foot a second. Radians are therefore convenient for mechanical problems about ropes being wound on to wheels, for wheels rolling on the ground, and generally for theoretical purposes. If, anywhere in a book on mathematics, you see any statement about 'an angle x' or 'the angle 3·5', with nothing more said, you must expect this to mean '*x radians*' or '3·5 *radians*', *not* x degrees or 3·5 degrees. 3·5 degrees is always written as 3·5°. If nothing is said about degrees, the angle is in *radian measure*. For mathematicians radian measure is the most natural to use, since it gives the simplest results.

If you measure right round the rim of a circle with the radius 1 foot, you will find the length to be (roughly) 6·28 feet. So one complete turn, 360°, is the same thing as 6·28 radians. This number, 6·28, is not a pleasant number to meet, but we cannot do anything about it. The universe is so made that this number turns up: it is not the fault of mathematicians. We cannot get away from 6·28. If we measure in degrees, so as to make one complete turn a convenient number, 360°, we find that the complication turns up elsewhere. When a wheel turns through 360° a second, the speed of points on the rim (supposing the radius, as before, to be 1 foot) will be 6·28 feet a second. Accordingly, *radians* are used for most questions connected with speed, or with the rims of circles: degrees may be used when we are measuring the angles of things which do not move – fields, for instance.

If you are not familiar with radian measure, you may find it worth while to cut out a large circle, and to mark the two scales around the rim – degrees and radians. Whenever you meet an expression such as 202° or 2·78 radians, you will be able to look at your circle and see what angles these expressions represent. It will be best if you put 0° at the position '3 o'clock', and go anti-clockwise, so that 90° will come at the top (12 o'clock), 180° at 9 o'clock, 270° at 6 o'clock, and 360° back again at 3 o'clock. 0 radians will also be at 3 o'clock, 1·57 at 12 o'clock, 3·14 at 9 o'clock, 4·71 at 6 o'clock, 6·28 at 3 o'clock again. It has become a custom among mathematicians to think of angles in these positions (I do not know why), and it will save misunderstanding if you do the same.

Sines and Cosines

We can now proceed to our experiments on right-angled triangles. It is again to be emphasized that the beginnings of the subject must be experimental. I cannot imagine anyone making any progress who simply sat looking at a right-angled triangle, hoping to be inspired with a method of reasoning out the problem. We must begin with experiments, and then see how much these help us.

Problem: a railway line makes an angle of 5° with the level and is perfectly straight: if a train travels 10,000 feet along the line, how many feet does it rise? No use thinking about it – let us measure and see. We find the answer, correct to the nearest tenth of a foot, to be 871·6 feet (you must take my word for this, unless you are prepared to carry out the experiment for yourself). Nothing particularly simple about the answer: it does not suggest any way of calculating the result without measurement.

But this result does one important job for us: it means that we need not do any more measurements of this particular type, on railways rising at 5°. If we are asked, 'How much does the train rise if it travels 100 feet?' we know the answer immediately. Since the line is straight, the train climbs steadily. In 10,000 feet it will climb 100 times as much as in 100 feet. Therefore, in 100 feet of travel it rises 8·716 feet. In fact, for each foot the train travels, it rises 0·08716 feet (correct to five places of decimals). If it travels x feet, it rises $0{\cdot}08716x$ feet.

In the same way, to any angle (measured in radians or degrees) there corresponds a number. In travelling x feet along a slope of 13° we rise $0{\cdot}22495x$ feet: for 30° the formula is $0{\cdot}50000x$. (Note this, our first simple result, $\frac{1}{2}x$ corresponding to 30°.) It is convenient to have a short way of referring to the numbers that arise in this way. We therefore give them a name, *sines*. (The name goes back to the time when learned men of all countries wrote to each other in Latin: it means a 'bowstring' – the reason for this name may be guessed from Fig. 17.) We say that 0·08716 is the sine of 5° (usually shortened to *sin* 5°), that *sin* 13° = 0·22495 and *sin* 30° = 0·5. 30° is 0·52360 radians, 13° is 0·22689 radians, 5° is 0·08727 radians. (To get these results so accurately we should have to use a circle of 10,000 feet, and measure round the rim). So we may also write *sin* 0·53260 = 0·5, *sin* 0·22689 = 0·22495, *sin* 0·08727 = 0·08716.

Note, in passing, a fact which leaps to the eye: in radian measure, though not in degree measure, the sine of an angle and the angle itself, for fairly small angles, are nearly the same number. The smaller the angle, the nearer it is to its sine. 0·5 is *sin* 0·53260; the two numbers 0·5 and 0·53260 differ by 0·0326. But *sin* 0·08727 is 0·08716. The two numbers here, 0·08727 and 0·08716, differ only by 0·00011. (This fact we have discovered without any effort: you will usually find that, as soon as you start to collect evidence, the discoveries make themselves.) This result suggests that there is some simple law connecting an angle, in radian measure, and its sine. We shall not be surprised when, in Chapter 14, we find a series giving sin x in terms of x. It is important to remember that this series holds *only when the angle is measured in radians, not for degrees.* (Look for the series in Chapter 14, and write a note there, in the margin, to this effect.)

The *cosine* of an angle is defined in a similar way. If an aeroplane starts from an aerodrome and flies 10,000 feet in a straight line at 30° to the level, we know how high it is. It is 10,000 *sin* 30° feet above the ground. Directly underneath the aeroplane there is a certain point on the ground. How far is this point from the aerodrome? By measurement is it found to be 8660·3 feet. Every extra foot the aeroplane flies (still keeping in the original line), this point moves 0·86603 feet further from the aerodrome. If the plane flies x feet, the point moves 0·86603x feet. We call 0·86603 the cosine of 30°, and write, for short, 0·86603 = *cos* 30° = *cos* 0·52360. (The last figure gives 30° in radian measure.)

In short: if we move on a straight line, making any angle t with the level, each foot we travel increases our height in feet by a certain number, called *sin t*, and carries us sideways through a certain number of feet, the number being called *cos t*.

We can easily make a model to demonstrate the meaning of *sin t* and *cos t*. Draw a circle of 1 foot radius, mark around it scales both for degrees and radians. Pin it to a wall or blackboard. Take a strip of cardboard, something over a foot in length. Fasten one end of it by a tin-tack or drawing-pin through the centre of the circle, so that the strip is free to turn. One foot from the centre pierce a small hole in the strip, and hang a plumb-line from this

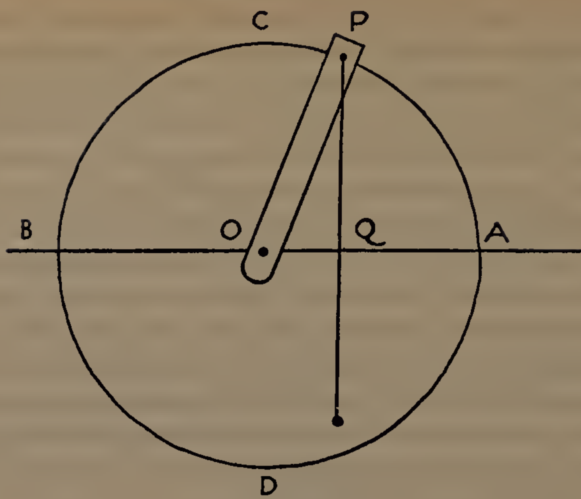


Fig. 17

hole. The arrangement is sketched in Fig. 17. The line BOA is level. The thread hanging from the small hole, P, crosses BOA at the point Q. P is higher than the line AOB by a height PQ, and it is a distance OQ to the right of O. As OP is 1 foot in length, the distance PQ, measured in feet, is equal to the sine of the angle AOP, and the distance OQ, also in feet, is equal to the cosine of AOP. To make a rough table of sines and cosines it would be better to have OP 1 metre long, then, by measuring OQ and PQ to the nearest millimetre, we should find results certainly true to two places of decimals, perhaps to three.

Actually doing this, or seeing it done, is helpful to the type of person who is good at games but has not a vivid imagination.

We have now explained what sines and cosines are. If you are given any statement about them, you can test for yourself whether it is true or not.

One point is worth mentioning. We have said that *sin t* represents the height of P above the line BOA. But if P is at the angle 270°, which is the same as 4·71 radians, P is 1 foot *below* BOA. Being 1 foot below BOA we may call being —1 foot above BOA. We therefore say that *sin* 270°, or *sin* 4·71, is —1. Similarly, the sines of all the angles between 180° and 360°, between 3·14 and

6·28 radians, have *minus* signs. Also, we take *cos t* to mean the distance *Q* is to the *right* of *O*. If *Q* lies to the left of *O*, as it does for angles between 90° and 270°, between 1·57 and 4·71 radians, the cosine has a *minus* sign.

Check for yourself the following table.

ANGLE (degrees)	0	90	180	270	360
(radians)	0	1·57	3·14	4·71	6·28
SINE	0	+1	0	—1	0
COSINE.. ..	+1	0	—1	0	+1

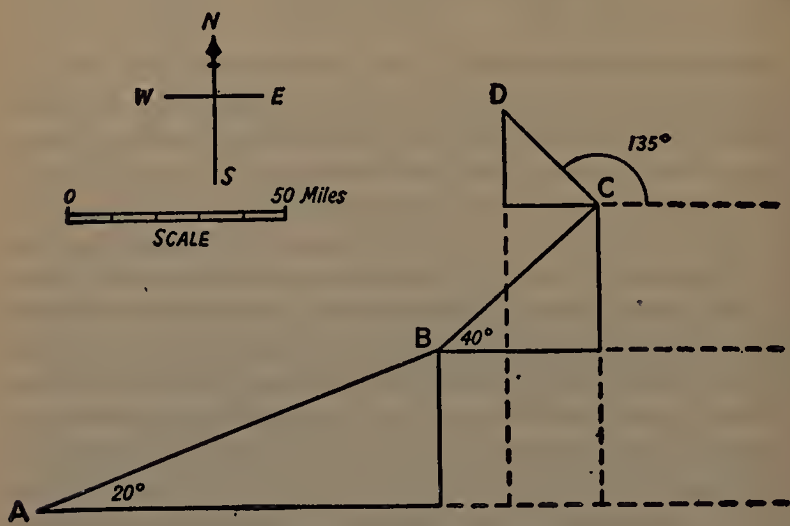
An explorer recording his journey will observe in what direction he travels, and for how many miles. If we regard East as corresponding to 0°, North will be 90°, and so on. On a map, North usually means *upwards*, East means *to the right*, so it is easy to apply our railway and aeroplane illustrations to map-making. If an explorer goes 100 miles in the direction 20°, then 50 miles in the direction 40°, what is his new position? 100 miles in the direction 20° is the same as going 100 *cos* 20° East, and then 100 *sin* 20° North. 50 miles in the direction 40° is the same as 50 *cos* 40° East, then 50 *sin* 40° North. If we have tables, we can work out these numbers, and then it is simple to add up, and find the total distance he had gone to the East, and the total distance to the North. Two things to note: (*a*) this method applies to the tunnel problem, Fig. 15, (*b*) it throws some light on the *minus* signs mentioned above. If the explorer, after doing the journey just mentioned, goes a further 30 miles in the direction 135° (i.e., North-West), this *increases* his distance to the North (*sin* 135° is +), but *decreases* his distance towards the East (*cos* 135° is —). In fact, by using the + and — signs in the definition of sine and cosine, we save ourselves the need for any further thought: we just have to write down *distance gone* times *sine*, for each part of the journey. The signs + and —, which then appear, automatically show whether the numbers have to be added on, or taken away.

The Formulae of Trigonometry

Certain other terms, besides sine and cosine, occur in trigonometry – namely, tangent, cotangent, secant, and cosecant.

These are, however, merely abbreviations, and do not bring in any essentially new idea: the subject could be mastered without using these terms at all. We shall therefore not deal with them here, but proceed to study the properties of the sines and cosines.

We shall, of course, try to discover properties of sines and cosines which are useful for our purposes. We have two particular problems in mind, and one rather general use as well.



THE EXPLORER'S JOURNEY

The explorer travels from *A* to *B*, from *B* to *C*, and then from *C* to *D*. *AB* is 100 miles, *BC* is 50 miles, *CD* is 30 miles. He records each part of his journey, and works out (by the method explained in the text) how far each part carries him towards the East, how far towards the North. Distances West and South appear with *minus* signs, since 10 miles farther West means 10 miles *less* to the East. The record appears as below:

	Distance.	Direction.	To East.	To North.
A to B.	100 miles	20°	94·0 miles	34·2 miles
B to C.	50 „	40°	38·3 „	32·1 „
C to D.	30 „	135°	—21·2 „	21·2 „
Whole journey A to D.			111·1 miles	87·5 miles

The first problem assumes that we have already in our possession satisfactory tables of sines and cosines, and is known as the *solution of triangles*. It is a problem which naturally arises in surveying. We are given certain information about a triangle, sufficient to enable us to draw the triangle, and are asked to find the remaining quantities. For instance, in any triangle *ABC* we might be told the length of *AB*, and the angles *ABC* and *BAC*, and asked to find the lengths *AC* and *BC*. This problem frequently arises in map-making, in the construction of range-finders, in determining the position of a ship at sea by taking the bearings of two lighthouses, in locating submarines, etc.

Surveyors and seamen are in a position to buy printed books of tables, containing sines and cosines and other information. But someone first had *to make these tables* – this is our second problem – and several of the properties of sines and cosines were discovered with this object in view. The interest which mathematicians of the sixteenth century showed in algebra was partly due to the fact that equations had to be solved before trigonometric tables could be made.*

Thirdly, it is desirable to know the properties of sines and cosines on quite general grounds. They arise in many problems, and the work can often be made shorter and simpler if the formulae are known. An example of this will be given later.

Pythagoras' Theorem

In Fig. 17 the sides *OQ* and *QP* have the lengths *cos t* and *sin t*, where *t* is short for the angle *QOP*. Students usually find this figure easy enough to grasp, but they do not always recognize it when it occurs in an unusual position, or on a different scale. For instance, in Fig. 18, *DF* makes an angle *t* with *DE*, and *DF* has the length 1. The line *EG* is drawn at right angles to *DF*. It is clear enough that the triangle *DEF* has the same shape as the triangle *OQP*. It may not be so obvious that there are two other

* See Zeuthen, *History of Mathematics in the Sixteenth and Seventeenth Centuries*, Chapter II, section 4. I am not sure if this work can be had in English.

triangles in the figure with the same shape. But this is so. If you cut out pieces of paper just large enough to cover the triangles *DEG* and *FEG*, you will find that it is possible (after turning the paper over) to lay these triangles in the positions *LVW* and *LTU*. The triangle *LMN* is exactly the same size and shape as *DEF*. It is now obvious that the three triangles have the same shape, and differ only in size.

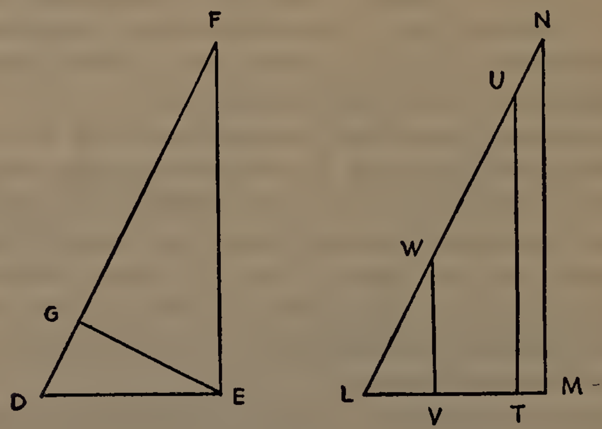


Fig. 18

How long are the lines *DG* and *GF*? The line *DG* can be put in the position *LV*, and is therefore *cos t* times *LW*. But *LW* has the same length as *DE*, which is equal to *cos t*. Accordingly, *DG* must be *cos t* times *cos t*, or $(\cos t)^2$. In exactly the same way, it may be shown that *GF* has the length $(\sin t)^2$. But *DG* and *GF* together make up *DF*, which we drew with the length 1 when we began. It follows that:

$$(\cos t)^2 + (\sin t)^2 = 1.$$

In Chapter 2 we met a triangle with the sides 3, 4, and 5, and a right-angle between the sides 3 and 4. If we draw this triangle on a scale one-fifth as large, we shall have sides $\frac{3}{5}$, $\frac{4}{5}$, and 1. For this triangle, then, $DE = \frac{3}{5}$ and $EF = \frac{4}{5}$, and $\cos t = \frac{3}{5}$, $\sin t = \frac{4}{5}$. The formula given above thus becomes:

$$(\tfrac{3}{5})^2 + (\tfrac{4}{5})^2 = 1, \text{ or } 3^2 + 4^2 = 5^2.$$

It is because of this relation between 3, 4, and 5 that the triangle is right-angled. Another such triangle is 5, 12, 13. $5^2 + 12^2 = 13^2$. If we draw an angle whose cosine is $\frac{5}{13}$, its sine will be $\frac{12}{13}$.

Here we have the answer to the question raised in Chapter 2: the proof given above is, in essentials, that given in Euclid. The result is usually known as Pythagoras' Theorem, and can be stated: if a, b, c are the lengths of the sides of a right-angled triangle, then $a^2 + b^2 = c^2$. This result is essentially the same as the result we have just found. For if t is the angle between the sides a and c, $a = c \cos t$ and $b = c \sin t$. (This result is obtained by enlarging the scale of our standard triangle, OPQ, c times.) So $a^2 + b^2 = (c \cos t)^2 + (c \sin t)^2$. This last expression is equal to c^2 multiplied by $(\cos t)^2 + (\sin t)^2$. By the result proved above, this is the same as c^2 multiplied by 1: that is, c^2. So $a^2 + b^2 = c^2$ follows from our earlier result, by simple algebra.

The Cosine Formula

So far we have considered only triangles containing a right-angle. We will now consider a more general problem. Suppose we have a triangle ABC, and we know the lengths AB and AC, and the size of the angle BAC. How long is BC? (It might be impossible to measure BC directly, owing to mountains, rivers, swamps, etc.)

In books on trigonometry it is usual to write a, b, c for the lengths of the sides BC, CA, AB and to write A, B, C for the three angles of the triangle. Thus, a is the side opposite the angle A, etc. Our problem is: given b, c, A, to find a.

Can this problem be solved at all? Are the facts given sufficient to allow us to draw a plan of ABC? They are. The problem can be solved by drawing: it is a reasonable problem to try.

Can we solve it with the help of tables of sines and cosines without drawing? What are the tables of sines and cosines? They are the result of experiments made on right-angled triangles. Sines and cosines therefore tell us nothing about a figure, unless that figure can be split up into right-angled triangles. Can we split ABC up into right-angled triangles? Very easily indeed. All we have to do is to draw CD at right-angles to AB (Fig. 19). We

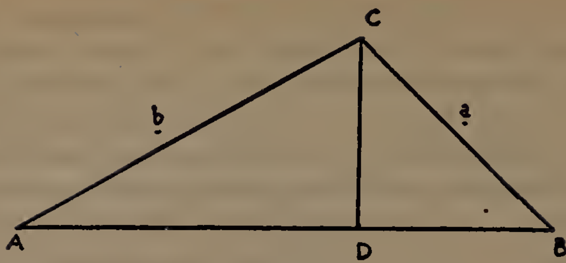


Fig. 19

now have two right-angled triangles, *ADC* and *BDC*. What do we know about these?

Triangle *BDC*: not much hope here. We want to find *BC*, but all we seem to know is that *BDC* is a right-angle.

Triangle *ADC*: quite a different story. We know $AC = b$, and we know the angle $CAD = A$. In fact, we know everything about this triangle: we have exactly the same information as we had in the railway problem, when we were told the angle the railway made with the level (A) and the distance the train travelled (b). The height *CD* is therefore $b \sin A$, and the distance sideways, *AD*, is $b \cos A$.

This new information helps us with triangle *BDC*. It tells us the length *CD*, and shows how *DB* can be found. For $AB = c$, and $AD = b \cos A$. *DB* is what is left when *AD* is taken away from *AB*. So *DB* must be equal to $c - b \cos A$.

We now know enough about the triangle *BDC* to fix it completely. We know *DC*, *BD*, and the angle *CDB* is a right-angle. *BC* can be found by Pythagoras' Theorem, for $BC^2 = DC^2 + DB^2$. Writing for *BC*, *DC*, and *DB* the lengths found for them, we have:

$$a^2 = (b \sin A)^2 + (c - b \cos A)^2.$$

This formula can be put in a more simple form. Before doing this, we may glance for a moment at the strategy by which we reached this point. The difficult thing in a mathematical problem is *to get started.* Before writing down any calculations at all, one should always prepare a plan of campaign. Otherwise one wanders around like a rudderless ship. While making this plan, forget all the difficulties that may come in the actual calculations. Try simply to build a framework connecting what we know with what we want to know. It is sometimes useful to draw a pencil figure, and to mark with ink those lines whose lengths are given, or angles whose size is known. Then mark with ink lines and angles which can be calculated from those already marked. And so go on, keeping a record of the steps.

For the present problem our plan would be as below.

Line AC and angle DAC given. (Ink these in.)
AD and DC can be calculated. (Ink these.)
AB is given. (Draw a line in ink, just below AB, so as not to blot out the line AD already drawn.)
So DB is found from AB minus AD.
BC is found from DC and DB by Pythagoras.

Do not worry if you have forgotten the formula $AD = b \cos A$, or the exact result of Pythagoras' Theorem. All you need to know in making this plan is that *a formula exists*: that the thing *can* be worked out. In real life (which is more important than examinations) you can always carry a book of formulae with you, and look them up. But no book will tell you on what lines to tackle a problem: that you must learn for yourself, by practice.

Now let us return to the formula we found for a^2. By simple algebra, we can work it out and obtain:

$$a^2 = b^2 \sin^2 A + c^2 - 2bc \cos A + b^2 \cos^2 A.$$

$\sin^2 A$ is the usual way of writing what, until now, we have written $(\sin A)^2$, and $\cos^2 A$ means the same as $(\cos A)^2$. Writing in this way saves a lot of brackets.

We notice that b^2 comes twice in this result. First we have b^2 multiplied by $\sin^2 A$, then b^2 multiplied by $\cos^2 A$. The total amount of b^2 that appears is therefore $\sin^2 A + \cos^2 A$, which is equal to 1. It follows that:

$$a^2 = b^2 + c^2 - 2bc \cos A$$

which is the usual formula given in text-books, and used in problems.

This is an example of the way in which formulae can be shortened by using the properties of sines and cosines. Earlier we promised that such an example would be given.

The Addition Formulae

Now, for some results which arise naturally in connexion with the problem of making tables, though they are also useful for general knowledge.

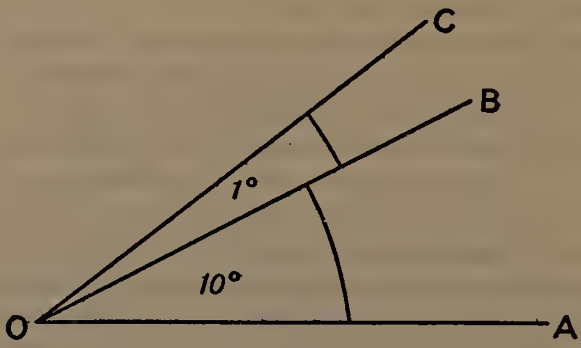


Fig. 20

Suppose that we are setting out to make a very accurate table of sines and cosines, and that (at great labour and expense) we have constructed large triangles and found accurate values for *sin* 1°, *cos* 1°, *sin* 10°, and *cos* 10°. It would be possible to keep on making fresh triangles, and to find by measurement *sin* 11°, *sin* 12°, etc. If carried out on a really large scale, this work would be very troublesome. It would be natural to think, 11° is 10° + 1°. Is it possible to use this fact in some way, and to find *sin* 11° by calculation, from what we know about 10° and 1°? If we can do this it will be very convenient, for the same method will give us information about 12°, since 12° = 11° + 1°, and we may continue thus as long as we care to do.

Our problem is: we have found, by measurement, that *sin* 1° = 0·01745, *cos* 1° = 0·99985, *sin* 10° = 0·17365, *cos* 10° = 0·98481. What are *sin* 11° and *cos* 11°?

The main difficulty in this problem is to draw a figure that brings out the facts clearly, It is easy enough to draw an angle of 10° with a further angle of 1° on top of it, as in Fig. 20. This illustrates the fact that 11° = 10° + 1° all right. But it does not

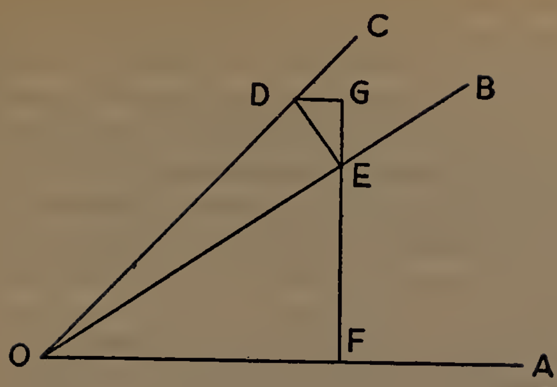


Fig. 21

tell us much about 10° and 1°. We have to take it on faith that the angles marked 10° and 1° *are* in fact 10° and 1°. There is nothing in the figure to show that they are: in particular, there is nothing to link them up with *sin* 1°, *sin* 10°, etc. (Actually, for the sake of clearness in the figure, it is necessary to draw the angles rather larger than they actually are.)

We want to bring out the fact that *BOC* is the angle, 1°, whose sine is 0·01745, and whose cosine is 0·99985. To do this we must bring in a right-angled triangle. Take *D* at a distance 1 from *O*, and draw *DE* at right angles to *OB* (Fig. 21). Then we know $OE = \cos 1° = 0{\cdot}99985$, and $DE = \sin 1° = 0{\cdot}01745$.

How is *sin* 11° to be brought into the picture? *OD* has the length 1 and makes the angle 11° with *OA*. The height of *D* above *OA* is therefore *sin* 11°. It is this height that we wish to find.

But this is easy to do: it is exactly the same problem as we had when the explorer went 100 miles in one direction, and then 50 miles in another. We can get from *O* to *D* by going first from *O* to *E*, then from *E* to *D*. We know the length and direction both of *OE* and of *ED*.

Draw an upright line *FEG* through *E*. *F* is on *OA*, and *G* is a point at the same height as *D*, so that *FG* equals the height of *D* above *OA*, that is $FG = \sin 11°$.

As $FG = FE + EG$, the problem is solved as soon as we can calculate FE and EG. FE presents no difficulties. $OE = 0{\cdot}99985$ and OE makes an angle $10°$ with OA, so

the height $FE = 0{\cdot}99985 \, sin \, 10° = 0{\cdot}99985 \times 0{\cdot}17365$.

EG can be found from the triangle EGD, which has a right-angle at G. The triangle GED could be obtained by turning the triangle EOF through a right-angle, and then making it shrink to a smaller scale. The angle DEG is, in fact, the same as the angle EOF – that is, $10°$. Accordingly, $EG = ED \, cos \, 10° = 0{\cdot}01745 \times 0{\cdot}98481$. Adding these two results together, we obtain the length of FG – that is, $sin \, 11°$.

The result we have just found may be written:

$$sin \, 11° = cos \, 1° \, sin \, 10° + sin \, 1° \, cos \, 10°.$$

There is nothing particular about the numbers 1 and 10. The same argument could be carried through for any two numbers, x and y, and we should find:

$$sin \, (x+y)° = cos \, x° \, sin \, y° + sin \, x° \, cos \, y°.$$

You should find no difficulty in working out $cos \, 11°$, the distance that D is to the right of O, and the corresponding general formula for $cos \, (x+y)°$.

Other Formulae

The formulae we have considered must be regarded as samples. There are other formulae in trigonometry, which, for the most part, can be found by arguments very similar to those given above. Some books contain vast masses of results. For most purposes, a few formulae and a few straightforward methods are quite sufficient. If you are studying trigonometry for some definite purpose – e.g., surveying, or navigation – you will do well to obtain a book on that subject, and see what formulae of trigonometry are actually used, and for what problems.

Differentiating Sines and Cosines

It often happens that sines and cosines occur in problems about the movement of machinery, the vibrations of some object, or the changes in electric currents. All these, being problems of change of speed, call for differentiation. It is therefore worth while to study the question: how fast do *sin t* and *cos t* change when *t* changes?

We shall study this problem by means of the model illustrated in Fig. 17. We suppose the point *P* begins at *A*, and travels round the circle at a steady speed of 1 foot per second. After *t* seconds it will have travelled *t* feet, and the angle *AOP* will therefore be *t* radians. (The results we shall obtain hold only when the angle is measured in radians.)

We know that *sin t* measures the height of *P* above *AOB* after *t* seconds. This we call, for short, *y* feet, so $y = sin\ t$. *cos t* measures the distance *P* lies to the right of *O* after *t* seconds. This we call *x* feet, so $x = cos\ t$. Of course, if *P* lies below *AOB*, *y* will be a number with a minus sign: *x* will be minus if *P* lies to the left of *O*. In the illustration, *x* equals the length of *OQ* in feet, *y* equals the length of *PQ* in feet.

Note that these signs, *x* and *y* have *no connexion at all with any signs x and y that may have been used in other chapters.* For instance, in Chapter 10 *x* stood for the number of seconds that had passed, and in Chapters 11 and 12 we discussed the expression $\frac{dy}{dx}$. In this section *t* is the sign used for 'the number of seconds': *x* and *y* have simply the meanings given to them in the last paragraph.

The speeds with which *x* and *y* change will be $\frac{dx}{dt}$ and $\frac{dy}{dt}$. We shall also write these as x' and y'. So x' is to mean the speed at which the length *OQ* changes, and y' the speed at which the length *PQ* changes. (If *P* lies below *AOB*, we shall have to continue the line of the plumb-line upwards, until it crosses *AOB*. This gives *Q*.) We have already explained carefully what is meant by speed, and how speed can be measured. The meaning of x' and y' should be clear.

There are four points on the circle at which it is particularly easy to see what is happening. These are the highest point, *C*, the lowest point, *D*, together with the two points, *A* and *B*. At *C* and *D* the track of *P* is level, at *A* and *B* it is upright.

As the track is level at C and D, the height of P cannot increase or decrease as it passes these points. As y' measures the speed with which the height of P changes, it follows that y' must be nothing when P passes C or D. It may be easier to see this result if you consider that P is moving upwards just before it reaches C (so y' is $+$), downwards just after it passes C (so y' is —). At C, y' is just at the moment of changing from $+$ to —, and must be 0. (Compare the remarks in Chapter 11, on the meaning of y'.)

The same argument shows that $x' = 0$ when P is at A or B.

What is x' when P is at C? At C the curve is level. Just for an instant the point P is neither rising nor falling, but is simply travelling towards the left, with a speed of 1 foot per second. In other words, at this instant x is decreasing at the rate 1 per second. That is, $x' = -1$.

At the point D, P is moving towards the right, with a speed of 1 foot a second. So $x' = 1$. In the same way, we may find y' for the points A and B. At A, P moves upwards, y is increasing, $y'=1$. At B, P is moving downward, $y' = -1$.

The points A, C, B, and D are reached after 0, 1·57, 3·14, 4·71 seconds (roughly). We may extend the table given earlier in this chapter, thus:

POSITION	A	C	B	D
t=time in seconds=angle in radians	0	1·57	3·14	4·71
$x=\cos t=OQ$	$+1$	0	-1	0
$y=\sin t=QP$	0	$+1$	0	-1
x'	0	-1	0	$+1$
y'	$+1$	0	-1	0

This table suggests something. The row y' is the same as the row x: the row x', except for a change of sign, is the same as y. This *suggests* that $y' = x$, and $x' = -y$. These results are, of course, not proved: they are not even made likely. We have taken only four points of the entire circle as evidence. (The law $y' = x^3$ would fit the table equally well.) But, as a bold guess, we might at least *investigate* the results.

It is left to the reader to prepare further evidence, using the points between A and C. A printed table of sines and cosines may be used. Remember to record the angle, t, in *radians*. (One degree $= 0{\cdot}01745$ radian.) In this way you can convince yourself that the guess was actually correct.

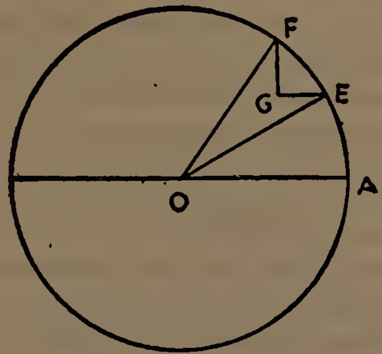


Fig. 22

It is also possible to see this result from a figure. In Fig. 22, E represents the position of P after any time, t seconds. That is, t feet of tape, wound round the circle, starting from A, would finish at E. A little later, P will have moved to a point, F, a little farther round the circle. The extra piece of tape, EF, will have the length Δt feet. If F is very close to E, the tape EF will be very nearly straight, and we shall not be making a serious error if we think of Δt as giving the length of the *straight line EF.*

The line EG is level, and the line GF is upright. So GF represents the increase in the height of P, as P goes from E to F – that is, $GF = \Delta y$. The angle GFE, you will find, is very nearly equal to the angle AOE, which is t radians. Accordingly, $GF = FE \cos t$, very nearly. (The smaller EF is, the nearer this equation comes to the truth.) That is, $\Delta y = \Delta t \cos t$, very nearly. When Δt becomes smaller and smaller, we find $\frac{dy}{dt} = \cos t$.

In the same way, GE is equal to the *decrease* in x, $-\Delta x$, and $GE = EF \sin t$, very nearly, which leads to the result $\frac{dx}{dt} = -\sin t$.

As x stands for $\cos t$, and y stands for $\sin t$, we may write these results as follows:

$$\frac{d\,(sin\ t)}{dt} = cos\ t \qquad \frac{d\,(cos\ t)}{dt} = -sin\ t.$$

This is shorter than saying, 'If $y = sin\ t$, $\frac{dy}{dt} = cos\ t$, etc.', and it means the same thing. We shall quote the result in this form in Chapter 14, when we find series for *cos t* and *sin t*, or, at any rate, for sine and cosine. I cannot promise that we shall always use the letter *t* when sine or cosine occurs.

Movement in a Circle

We saw in Chapter 10 that the force acting on a moving weight can be found if we know mx'' and my''. It often happens in machinery that a heavy weight goes round in a circle, as, for instance, any part of a flywheel, or the metal attached to a locomotive wheel (though this also travels along the line). An aeroplane looping the loop or a motor-car going round a corner raises a similar problem.

We may therefore consider a weight attached to the point P of Fig. 17, and see what force would be required, to make it move in the desired way. Since $y = sin\ t$, $y' = cos\ t$, and y'' (the rate at which y' changes) is therefore $-sin\ t$. In the same manner, we find $x'' = -cos\ t$. There is no difficulty in finding x'' and y'', and the total force acting on the weight at P can be found easily by anyone who has had some practice in the elementary problems of statics and dynamics.

EXERCISES

1. Cut out a circular disc of cardboard, and mark a scale for measuring angles in radians around the edges, by the method explained in this chapter. On the same disc mark a scale for measuring angles in degrees.

Draw on a piece of paper an angle of $\frac{3}{8}$ radian, 1 radian, $2\frac{1}{2}$ radians, 5 radians, 10 radians.

How many radians are 10°, 50°, 95°, 184°?

2. Make an actual model from the design of Fig. 17. From this

make a table giving the sines and cosines of 5°, 10°, 15°, etc. (as far as 90°), to two places of decimals. Check your results for sines from printed tables.

3. Write down, from your results for question 2, *sin* 10°, *sin* 20°, etc. up to *sin* 80°. Write down the cosines in the opposite order: *cos* 80°, *cos* 70° ... *cos* 10°. What do you notice about the two lists? What can you say about *sin* $x°$ and *cos* $(90—x)°$? Can you see any reason for your result?

4. From your model (question 2) find to two places of decimals *sin* 100°, *sin* 110° ... *sin* 170°. Compare these with *sin* 10°, *sin* 20° ... *sin* 80°. What do you notice about the two sets? What formula connects *sin* $(180—x)°$, and *sin* $x°$?

5. Find from your model *cos* 100°, *cos* 110° ... *cos* 170°. (For all of these Q is to the left of O, so the cosines all have a minus sign.) Compare these with *cos* 10°, *cos* 20° ... *cos* 80°. What formula connects *cos* $(180—x)°$ and *cos* $x°$?

Also compare the set just found, *cos* 100° ... *cos* 170° with *sin* 10° ... *sin* 80°. Is there a formula connecting *cos* $(90+x)°$ with *sin* $x°$? If so, what is it?

6. The printed tables give the sines of angles between 0° and 90°. To find the sines of angles between 90° and 180°, and to find cosines, we have to use the results of questions 3, 4, and 5. For example, the tables tell us *sin* 37° = 0·6018. What are the values of *cos* 53°, *sin* 143°, *cos* 127°?

7. An airman flies 200 miles in the direction 37° North of East. How many miles to the East, and how many miles to the North does his position change? (*Note* – For long flights it is necessary to take account of the fact that the earth is shaped like a ball. All questions in this chapter refer to short journeys, for which the earth may be supposed flat.)

8. Find how many miles East and how many miles North of A is the point C given by the explorer's log below:

A to B 30 miles, in the direction 40° N of E.
B to C 10 miles, due West.

9. Do the same for the journey below:

A to B 40 miles, in direction 70°.
B to C 20 miles, in direction 110°.

10. Also for:

A to B 100 miles in direction 315° (i.e., South-East).

B to C 150 miles in direction 80°.

11. An aeroplane has to fly to a town 100 miles away. By mistake it flies in a direction which is 2° off the correct course. When it has flown 100 miles, how far will it be from the town?

12. Ipswich is 65 miles from London, in the direction 36°. Peterborough is 75 miles from London, in the direction 95°. Birmingham is 105 miles from London, in the direction 139°.

How far is Ipswich from Peterborough, Peterborough from Birmingham, and Birmingham from Ipswich?

(This can be done by means of the formula $a^2=b^2+c^2-2bc. \cos A$. To find the distance from Ipswich to Peterborough, we may take London as A, Ipswich as B, Peterborough as C. A similar calculation gives the other two distances. Remember that $\cos$ 103°, which comes in the formula for the distance from Birmingham to Ipswich, has a minus sign. Check your calculations by drawing.)

CHAPTER 14

ON BACKGROUNDS

'*Recent workers in the sociology of science have stressed that experimental science arose from the theorists' taking account of the crafts. On the other hand, the crafts have often failed to learn from the theorists almost down to our own day.*' – H. T. Pledge, *Science Since 1500.*

STUDENTS of mathematics often have the experience of understanding the proof of some results, but not being able to see what it is all about. The subject remains 'up in the air', a disconnected piece of knowledge. As memory depends on connexions, the result is hard to remember. In ordinary life we remember familiar objects well, because other things continually remind us of them, and thus refresh their images in our mind. Students rightly feel uneasy when they are asked to remember something unconnected with the rest of life: the mind cannot work efficiently unless it is properly treated.

In elementary algebra this 'unearthly' feeling is easily aroused. Many text-books, for instance, explain quite accurately what is meant by an arithmetical progression, or a geometrical progression. The teacher (who may be passionately interested in some other subject, and forced to teach mathematics without understanding it) follows the text-book and teaches A.P.s and G.P.s simply because they are in the book.

We have already (without noticing it) had two examples of arithmetical progressions. The man falling off a house travels 1 foot in the first quarter-second, 3 feet in the next quarter-second, 5 feet in the third quarter-second, 7 feet in the fourth, and so on. The total distance gone in one second is $1+3+5+7$ feet. In the set of numbers, 1, 3, 5, 7, etc., each number is 2 more than the previous one. A set of numbers in which each number is bigger (or smaller) than the number before it by a fixed amount is called an arithmetical progression (A.P.).

The second example was in Chapter 12, when we added together the numbers, 0, 0·01, 0·02, etc., up to 0·09. These numbers also form an A.P. Later in the same section we saw that we could have found a more accurate estimate of $\int_0^1 x\,dx$ if we had divided the first second, not into ten, but into 100 parts. We should then have had to work out the sum of 100 numbers, beginning with 0, 0·0001, 0·0002 ... and ending with 0·0098, 0·0099. Can we shorten the work, so that we shall not actually have to carry out the addition? Yes, this is possible. The first number, 0, and the last number, 0·0099, add up to 0·0099. The second number, 0·0001, and the next to last, 0·0098, also add up to the same amount. Going on in this way, we can arrange all the numbers in pairs, each pair adding up to 0·0099. There will be 50 such pairs. The total will be 50 times 0·0099, that is, 0·495. This result was quoted, without proof, in Chapter 12.

Geometrical Progressions

A geometrical progression consists of a series of numbers, each of which is obtained by multiplying the previous one by a fixed number – e.g., 1, 2, 4, 8, 16 ... or 3, $1\frac{1}{2}$, $\frac{3}{4}$, $\frac{3}{8}$ ... Such series can arise in a number of ways.

For instance, there is the well-known question: at what time between 3 o'clock and 4 o'clock is the minute hand of a clock over the hour hand? It is quite natural to begin thinking in the following way. At 3 o'clock the minute hand is 15 minutes behind the hour hand: the hour hand moves slowly, so that in 15 minutes' time the minute hand will nearly have caught up with the hour hand. The hour hand moves through 5 minutes in each hour – one-twelfth as fast as the minute hand. By 3.15 the hour hand will have moved $\frac{15}{12}$ minutes – this is the amount the minute hand still has to catch up. The minute hand will reach this position after an extra $\frac{15}{12}$ minutes. But meanwhile the hour hand will have moved through another $\frac{15}{12^2}$ minutes. In this way we keep revising our original guess of 15 minutes, by adding to it in turn, $\frac{15}{12}$, then $\frac{15}{12^3}$, and so on – each correction being one-twelfth the size of the previous one. In this way we obtain the result $15+\frac{15}{12}+\frac{15}{12^2}+\frac{15}{12^3}+$ etc., the sum of a geometrical progression. By taking sufficiently many terms of this series, we can get an answer correct to any degree of accuracy: for instance, the four terms actually written above give an answer that differs from the correct answer by less than 0·001.

It is possible to see what the sum of the series is. The hour hand goes only 5 minutes while the minute hand goes 60 – that is, the minute hand gains on the hour hand at the rate of 55 minutes in each hour, or $\frac{55}{60}$ minute in every minute. $\frac{55}{60}$ is the same thing as $\frac{11}{12}$. Accordingly, to catch up 15 minutes on the hour hand, the minute hand will require 15 divided by $\frac{11}{12}$ minutes – that is, $16\frac{4}{11}$ minutes. So the sum of the series must be $16\frac{4}{11}$.

Another problem: if a ton of seed potatoes will produce a crop of 3 tons, which can either be consumed or used again as seed, how much must a gardener buy, if his family want to consume a ton of potatoes every year for ever?

First of all, he must buy a ton to meet his needs for this year. To get a ton for next year, it will be sufficient to plant $\frac{1}{3}$ ton now. To meet the needs of the year after next, $\frac{1}{9}$ ton will be sufficient: for it will yield $\frac{1}{3}$ ton next year, and this planted again will yield 1 ton the year after. And so on. To meet the needs of his family for ever, the farmer must plant $1+\frac{1}{3}+\frac{1}{9}+\frac{1}{27}+\frac{1}{81}+\ldots$ tons.

How much does this series add up to? Let us call the number it adds up to x. We can find x by a simple trick – namely, by working out $3x$ and comparing it with x.

Thus $$x=1+\tfrac{1}{3}+\tfrac{1}{9}+\tfrac{1}{27}+\ldots$$

so $$3x=3+1+\tfrac{1}{3}+\tfrac{1}{9}+\ldots$$

We notice that $3x$ is the same series as x, except that 3 has been added at the front, so $3x=3+x$. It follows that $2x=3$, so x must be $1\frac{1}{2}$.

We can easily see that this is the right answer. If the gardener buys $1\frac{1}{2}$ tons, he needs 1 ton to eat this year, and has $\frac{1}{2}$ ton to plant. The crop will be three times as much as what is planted – this is, it will be $1\frac{1}{2}$ tons – and again he has 1 ton to eat and $\frac{1}{2}$ ton to plant. In this way he and his descendants can continue as long as they care to do.

The same type of series arises in connexion with annuities, compound interest, discount, stocks and shares, etc. Compound interest is one of the main historical reasons why geometrical progressions first came to be studied. To a man making a fortune by money-lending, it is, no doubt, an absorbing subject: for most other people, and particularly for children at school, compound interest is likely to prove deadly dull.

Another application of geometrical series is for the study of air resistance. A body moving through the air is like a man rushing through a crowd. The faster he runs, the more people he knocks into: in other words, the resistance to his progress is proportional to his speed. The same is true for a body moving through the air (provided its speed is not too great): the faster it goes, the more air it has to brush out of its way each second. The result is that it loses a definite fraction of its speed each second. If two-thirds of the speed is lost each second, one-third remains: thus a body might move 1 foot in the first second, $\frac{1}{3}$ foot in the following second, $\frac{1}{9}$ foot in the third second, and so on. The distance gone altogether would be $1+\frac{1}{3}+\frac{1}{9}+\frac{1}{27}+\ldots$, the same series as we had before. It is, of course, taken for granted that no force is acting on the body, apart from the resistance of the air. Thus, the body might be a propeller; if it is properly balanced, and is not connected to an engine, there is no force to make it turn round; if it is given a push, it will start spinning, but will gradually slow down, in the way described.

The connexion between a moving body and geometrical progressions was known already in the seventeenth century. A more modern application of the same idea is to an electric current in a wire: an electron moving inside the wire collides with the atoms composing the wire, just like a man moving in a crowd. If the wire is connected to an electric battery, the problem is altered: there is then a force dragging the electron forward. In the same way, a falling raindrop is subject to the pull of gravitation; for this reason, it is not brought to rest by the resistance of the air. The problem of the way in which a raindrop falls is therefore slightly more complicated, but it too was solved, by means of geometrical progressions, in the seventeenth century.

If x stands for 'any number', we can show (by the method used in the potato problem) that the series $1+x+x^2+x^3+\ldots$ is equal to $\frac{1}{1-x}$, provided x is not bigger than 1.

Other Series

We have just seen that $\frac{1}{1-x}$ can be expressed in the form of a series, containing the various powers of x. This is not a peculiarity of $\frac{1}{1-x}$: almost any function of x you are likely to meet can be expressed in this way. For instance, $\sqrt{1+x}$ is equal to a series which begins with $1+\frac{1}{2}x-\frac{1}{8}x^2\ldots$ and $-\log_e(1-x)$ is equal to the series $x+\frac{1}{2}x^2+\frac{1}{3}x^3+\frac{1}{4}x^4+\ldots$ For both these series we suppose x to be less than 1. It is obvious that the series are not true when x is bigger than 1. Of course, we have not *proved* the series to be equal to the functions stated: for proofs you will need to consult text-books.

It is often very convenient to express a function in the form of a series. For instance, while we know from Chapter 12 what is meant by $\log_e 2$, it may not be easy to say just what number this is. By means of the series we can find it. For $\log_e 2$ can be found from $\log_e \frac{1}{2}$. In fact, 2 times $\frac{1}{2}$ equals 1. Taking logarithms, it follows that $\log_e 2+\log_e \frac{1}{2}=\log_e 1$. But $\log_e 1=0$. So $\log_e 2=-\log_e \frac{1}{2}$. But putting $x=\frac{1}{2}$ in the series given, it follows that $-\log_e \frac{1}{2}$ is equal to $\frac{1}{2}+\frac{1}{8}+\frac{1}{24}+\frac{1}{64}+ \ldots$ The terms of this series get small quite rapidly: we do not have to take very many terms to get quite a good value for $\log_e 2$.

Another advantage of such series is that they are easy to differentiate or integrate, since we know how to deal with powers of x. If you differentiate the series for $-\log_e (1-x)$, what series do you get? Is this result reasonable?

Later in this chapter we shall find a series for e^x, and we are now going to find series for $\cos x$ and $\sin x$, in order to show how such a question can be tackled.

In Chapter 13 we showed $\sin 0$ to be 0, and $\cos 0$ to be 1, also that $\frac{d \sin x}{dx}=\cos x$ and $\frac{d \cos x}{dx}=-\sin x$. It is rather surprising that from this information alone we can find the series we want.

If $\cos x$ is expressed as a series containing the powers of x, certain numbers will occur in the various terms (as the numbers $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$, etc., did in the series for $-\log_e (1-x)$): these numbers we shall call for short $a,b,c,f,g,h,j,k \ldots$ (The numbers d and e are left out, because d is used with a special meaning in $\frac{dy}{dx}$ and e also has a special meaning.) So the series will be:

$$\cos x=a+bx+cx^2+fx^3+gx^4+hx^5+jx^6+kx^7+ \ldots$$

Our job is to find out what particular numbers a, b, c, etc., are.

a we can find straight away. If we put $x=0$, $\cos 0=1$, while the series becomes simply a. It follows that $a=1$.

If we differentiate the equation above, we find (since the differential of $\cos x$ is $-\sin x$):

$$-\sin x=b+2cx+3fx^2+4gx^3+5hx^4+6jx^5+7kx^6+ \ldots$$

b can now be found by putting $x=0$. $\sin 0=0$, so it follows that $b=0$.

We now differentiate the series for $-\sin x$. The differential of $\sin x$ is $\cos s$, so we have:

$$-\cos x=2c+6fx+12gx^2+20hx^3+30jx^4+42kx^5+\dots$$

c is now found by exactly the same method. Put $x=0$. This leads to the equation $-1=2c$, so $c=-\frac{1}{2}$.

It is clear that there is nothing to stop us continuing with this as long as we like, and finding as many of the numbers $f, g, h, \dots$ as we care to do. The results are (as you can check for yourself):

$$a=1,\ b=0,\ c=-\tfrac{1}{2},\ f=0,\ g=\tfrac{1}{24},\ h=0,\ j=-\tfrac{1}{720},\ k=0:$$

so: $$\cos x=1-\tfrac{1}{2}x^2+\tfrac{1}{24}x^4-\tfrac{1}{720}x^6\dots$$

The rule which gives the numbers 1, 2, 24, 720, etc., which appear here, is the following. We start with 1. We multiply this by 1 times 2; this gives the second number, 2. We multiply the second number by 3 times 4, this gives the third number, which again is multiplied by 5 times 6 to give the fourth number: and so on. Differentiating the above series, we find the series for $\sin x$:

$$\sin x=x-\tfrac{1}{6}x^3+\tfrac{1}{120}x^5-\dots$$

These are good series for the purpose of calculation, since the terms get smaller very rapidly, and the first few terms of the series give quite accurate answers. These series therefore give an answer to the problem put in Chapter 13, to find a way of making a table of sines and cosines without drawing any figures.

The Dangers of Series

Series played an important part in the early days of the calculus, particularly in the years following 1660. This was a period of great practical activity: men were interested in the new developments of science, and were faced with a great variety of practical problems – the construction of clocks and of telescopes, of maps and ships. If a mathematical method gave the correct answer to a practical problem, people did not bother much whether it was logical or not. In dealing with small changes, Δx, mathematicians followed their own convenience: at one moment they said, 'Δx

is very small, it will be convenient to regard Δx as being equal to 0.' A little later they wanted to divide by Δx, so they said, 'If Δx is 0 we cannot divide by it: we will suppose Δx to be small, but not quite 0.' Whichever was more convenient, that they supposed to be true. If the answer turned out to be wrong, they scrapped their work. As the results were always compared with practice, this rough-and-ready method worked quite well.

Series were treated in this way, too. If it looked reasonable to make a certain step, that step was made. If it gave a ridiculous answer, one soon recognized that something was wrong.

After about 150 years of carefree mathematics, difficulties began to be felt. For instance, in the calculation of logarithms, the series $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}\ldots$ arises. Half the terms of this series have a $+$ sign; we will call the sum of these terms a, so that $a=1+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}\ldots$ We will call b the sum of the other numbers in the series – that is, $b=\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\ldots$ We notice that every number that occurs in b is even. If we double b we thus have $2b=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\ldots$; in the series for $2b$ we have all the terms of the series for a, together with those for b. So it seems that the series for $2b$ should equal $a+b$, so that $2b=a+b$. It follows from this that $b=a$. But b cannot equal a, for every term in a is bigger than the corresponding term in b; 1 is bigger than $\frac{1}{2}$, $\frac{1}{3}$ is bigger than $\frac{1}{4}$, $\frac{1}{5}$ is bigger than $\frac{1}{6}$, and so on. If a equals b, our original series $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}\ldots=a-b=0$. But the sum of the original series is, in fact, just less than 0·7.

In other words, by doing things that *look* reasonable, we have been led to an untrue result. On the other hand, in many cases true and useful results have been obtained by the use of series. It is therefore natural that mathematicians should have begun to inquire more carefully just what is meant by a series, and just what operations may be carried out with series. During the nineteenth century mathematicians carried out such an inquiry. There was a reaction from the carefree attitude of earlier times, and a spirit of caution spread. Mathematicians became rather like lawyers, very concerned about the exact use of words, very suspicious of arguments which merely 'looked reasonable'. Terms such as 'convergence' and 'uniform convergence' were invented, designed to distinguish series which were reliable from those which led to wrong conclusions.

Mathematicians did not only investigate the logic of series: they became uneasy about all the words they were using, and did not become comfortable again until they had given very exact explanations of all the terms they used. Modern books on mathematics are often much longer than the old books, because they spend time explaining and justifying things, which at first sight seem obvious.

There is a fable about a centipede which was asked in what order it moved its feet, and became so puzzled by the question that it was unable to walk at all. Students who begin by studying modern mathematics often suffer from a similar disorder: they spend so much time learning how to criticize, that they never understand how to create. The best policy is to follow the course of history: first to learn to *see* results as the old pioneers were able to see them, only later to examine the weaker points in the natural approach. If there had been no risks taken by the creative mathematicians in the seventeenth and eighteenth centuries, there would have been nothing for the pure mathematicians to criticize in the nineteenth.

The Background of e^x.

Many text-books give explanations of the number e which in themselves are excellent and logical, but leave the reader with the feeling that everything has come 'out of the blue': the argument is logical, but how was it discovered? What is it all about?

We have already had to deal with expressions such as 10^x, a^x, $\log_e x$. We shall now try to collect the facts about these expressions, and to show the connexions between them.

The idea of an exponential function springs immediately from the practice of money-lending. The way in which debt mounts and strangles its victim is an old story, both in fact and fiction. If a money-lender advances £100 now in return for £110 in a month's time, and in a month's time the borrower cannot pay, he finds himself obliged to contract a new loan on the same terms for another month – but the new loan is for £110, not for £100. In a year the debt will become something over £313. The debt for each successive month contains an extra multiplication by $1\frac{1}{10}$. (Compare this with the section 'How Logarithms were invented' in Chapter 6.) 10% a month is the same as 213% a year, much more than 12 times 10%.

We could reverse this and ask: what rate per month is equal to 5% per year? We could try different rates per month, until we found one which worked out (to a sufficient degree of accuracy) to 5% a year. We could ask what rate a week, what rate a day corresponds to 5% a year. If we liked, we could find the rate per hour, or per minute or per second. There would be only one correct answer to any of these questions. By fixing the rate of interest for a year, we automatically fix the rate of interest for any other length of time.

Suppose, for instance, the rate for a whole year to be 100%, and that an inexperienced money-lender charged 40% for six months. Then nobody would borrow money by the year. It would be cheaper to borrow for two periods of six months. £100 now would mean paying £140 after six months. This debt could be met by starting a new loan, for £140. The interest for the remaining six months, at 40%, would be £56, so that £196 would have to be paid at the end of the full year. Borrowing for a year at a time, £200 would have to be paid. In the same way, if the rate for six months were fixed at 50%, it would pay people to borrow money for a year, and lend it out again for two periods of six months: in the first six months, £100 would become £150: in the second six months, £150 would become £225; after paying back £200, a profit of £25 would be left. From practical necessity, then, the rate for six months would be something more than 40%, but less than 50%.

We could make a table showing what £1 would become after any length of time – weeks, days, hours, minutes – once the rate of interest for a year was given. If £1 becomes £a in one year, in n years (n standing for any whole number) it will become £a^n. It is therefore natural to suppose that £$a^{\frac{1}{2}}$ represents the amount that £1 would become after $\frac{1}{2}$ year. A sign such as $a^{\frac{1}{2}}$ has no meaning in itself: in the South of England 'stack' usually means a haystack, in the North a chimney – it is a waste of time to argue which is 'right'. A word is a label tied to a real thing for convenience. It does not matter whether the label is pink or green. A rose by any other name would smell as sweet. If we like to say that £$a^{\frac{1}{2}}$ is going to mean what £1 becomes in $\frac{1}{2}$ year under certain conditions, we have a perfect right to do so. (This definition agrees with that given in Chapter 6, although it uses a different picture.) £a^x will mean what £1 becomes after any number, x, years: x may be a fraction.

As we saw in Chapter 6, the most convenient way of making a table is to start with a small change, and build up from this to larger changes. At whatever rate of interest £1 grows, there will be a time in which it increases by one-thousandth, say, in k years (k may be a small fraction). Every k years that pass, the amount due will be multiplied once again by $1\frac{1}{1000}$. We can thus draw a graph, showing the growth of the debt, by taking upright lines, each separated from the next by a distance k inches. Each line must be longer than the previous line by one part in a thousand. Corresponding to $x=0$ we must have a height of 1 inch: since the sum of money, to begin with, is £1.

Minus Numbers

It is clear that we could extend our graph to minus values of x. Each line is $\frac{1000}{1001}$ of the line to the right of it. So, k inches to the left of $x=0$, we could put a line having the height $\frac{1000}{1001}$, and continuing in this way, we could continue the graph, and find a height corresponding to any distance to the left – that is, to minus values of x. We now have a graph extending as far as we like, both to the right and to the left.

Change of Scale

The distance k depends on the rate of interest. By a suitable choice of k we can get any rate we wish. The distances between our upright lines must remain *equal*, but by altering their size we change the rate of interest. Mechanically this can be done by the arrangement known as 'lazy tongs'. In Fig. 23 we indicate this by a lattice-work of diamond shapes. We suppose a model made to this design, with loosely jointed pieces of wood. Some device (not shown in the figure) will be needed to keep the upright rods in the proper direction. By pulling or pushing at the points A and B, the upright rods can be spaced out, or brought more closely together. In this way, we have one model which represents a^x for *any* number a (within a certain range of values). (In the

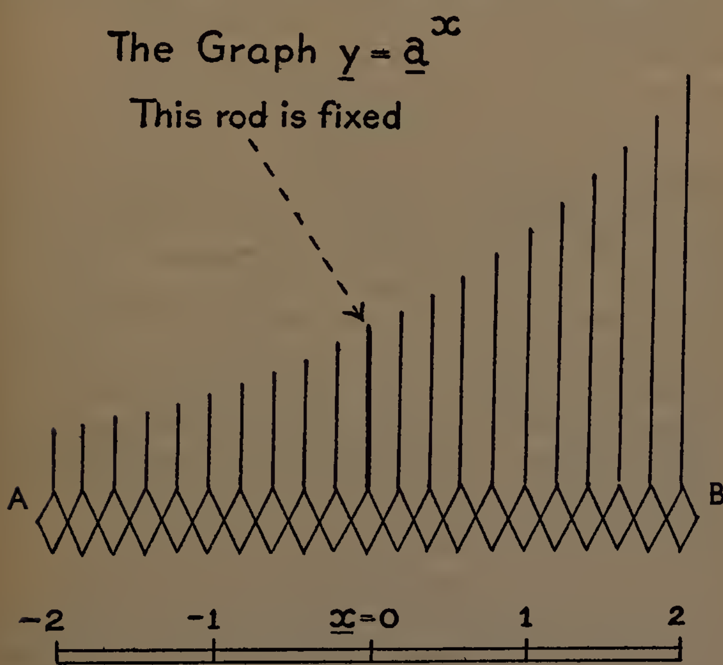


Fig. 23

figure, the rate of change has been exaggerated – each upright stick is actually one-tenth part greater than its neighbour, instead of one-thousandth.) x is throughout measured in inches. When $x=1$, $a^x=a$. Thus a is the length of the rod standing at a distance 1 inch to the right of $x=0$. For example, to obtain the graph of 2^x we must press the points A and B until the rod 2 inches long stands above the point marked 1 on the scale for x. When the model is set in this way, we shall find that 1 inch to the right of any rod there stands a rod twice as high.

The model actually drawn – that is, one in which each rod is $1\frac{1}{10}$ as long as its neighbour on the left – we shall call the *crude* model: the one described in the text, using the ratio $1\frac{1}{1000}$, we shall call the *fine* model. The crude model is suitable for actual construction, and for class work: the fine model should stay in the imagination, for the purpose of argument. As we saw in Chapter 6, the ratio $1 \cdot 1$ is not sufficiently near to 1 to give accurate values for logarithms.

Logarithms

In Chapter 6 we defined the logarithm as 'the length of rope' needed to multiply one's strength by a given number. In our graph, the distance x corresponds to the length of rope, the height of the rod standing there (y inches, say) measures the multiplying effect. In the crude model we use, in fact, the numbers given in the table on page 75. To get logarithms to base 10, we would need to take $a=10$. But for the moment we are not particularly interested in 10, and we suppose the model set for any number a. Then $y=a^x$, or, to put the same thing the other way round, $x=\log_a y$.

The Number e

If $y=a^x$, what is y'? Consider this with the fine model in your mind. As we go from one rod to the next, x increases by k – that is, $\Delta x=k$. Each rod is $1 \cdot 001$ times as long as the rod before it; the change in length, Δy, will therefore be $0 \cdot 001$ times y. So $\frac{\Delta y}{\Delta x}$ will be $\frac{0 \cdot 001}{k}y$. This gives us some idea of y': it suggests (what is actually true) that y' is proportional to y. If we take $k=0 \cdot 001$, we shall have a particularly simple result. $\frac{0 \cdot 001}{k}y$ will then be simply y, and we shall have, very nearly, $y'=y$. (We have to write 'very nearly', because $\frac{\Delta y}{\Delta x}$ is 'very nearly', but not quite, y' when $\Delta x=0{\cdot}001$.)

Taking $k=0{\cdot}001$ means that the rods are spaced with one-thousandth of an inch between them. In going from $x=0$ to $x=1$, the length of the upright rod will have been multiplied by $1{\cdot}001$ a thousand times. So a, the length of the rod at $x=1$, will be $(1{\cdot}001)^{1000}$.

If instead of the fine model we had argued from the crude model, we should have been led to the result $(1\frac{1}{10})^{10}$. The fine model gives the better answer, $(1\frac{1}{1000})^{1000}$. By taking still larger numbers of rods we could get still better answers. Our result would always be of the form $\left(1+\frac{1}{n}\right)^n$. The larger n is, the more nearly would y' be equal to y. As n is made larger and larger, $\left(1+\frac{1}{n}\right)^n$ gets closer and closer to the number $2{\cdot}71828\ldots$, which was mentioned in Chapter 11, and is named e. If $y=e^x$, $y'=y$ exactly.

In Chapter 11, e was found by another method – namely, by choosing a number a which gave the simplest result for the differentiation of the logarithm to base a. As $y=a^x$ means the same thing as $x=\log_a y$, it is not surprising that the same number e should give the simplest answer in both cases. The reader may be able to show that the two methods are really the same. The only complication is due to the fact that the signs x and y have changed places: in Chapter 11, we assumed $y=\log_a x$: here, $x=\log_a y$.

The Series for e^x

We now have enough information about e^x to find a series for it. When $x=0$, $e^x=1$. If $y=e^x$, $y'=e^x$. The method that was used to find a series for *cos x* works equally well for e^x. We find, in fact:

$$e^x=1+x+\tfrac{1}{2}x^2+\tfrac{1}{6}x^3+\tfrac{1}{24}x^4+\tfrac{1}{120}x^5+\tfrac{1}{720}x^6+\ldots$$

Differentiate this series for yourself, and verify that the series for y' is the same as for y, so that $y'=y$.

This method works also for many other functions, and is associated with the names of Taylor and Maclaurin.

a^x has simple properties, similar to those of e^x, for, as we have seen, the graph of a^x can be got from that of e^x simply by changing the scale of x (in the model, by pushing or pulling the points A and B).

The Applications of e^x

The importance of e^x is due to the property $y'=y$: that is to say, the rate at which it grows is equal to its size. If instead of e^x we take e^{bx}, where b is any fixed number, we have $y'=by$; that is, the rate of growth is proportional to the size.

There are many things which grow in this way. We have already mentioned the example of money-lending. £1,000 grows a thousand times as fast as £1.

Much the same thing holds in business. Within certain limits, the more shops a company owns, the more rapidly can it extend its business.

If a country desires to build up its industry, and starts with very little equipment, it finds that the rate at which it can install new factories is very slow; but the more factories it gets, the quicker it can equip new factories. In a country suffering foreign conquest, the reverse holds: the more factories it loses, the less it is able to replace its losses.

The population of a country, under settled conditions, may grow according to the law e^{bx}. The more people there are in the country, the more children are likely to be born. The population of the U.S.A. between 1790 and 1890 roughly corresponded to the formula $y=3\cdot 9\times 10^{0\cdot 012x}$, where y is the population in millions x years after 1790.* The formula, of course, ceases to work when a

* *Introduction to Mathematics* by Cooley, Gans, Kline, and Wahlert, page 363.

country reaches the stage where it cannot support any more people. Rather similar considerations apply to the rate at which microbes multiply in a glass of milk that is going sour, the spread of rabbits in Australia, and other forms of living growth.

There are also conditions in which a new religion or political creed grows by an exponential law. If there are large numbers of people in the mood to accept some new doctrine, once it is put to them, the spread of that doctrine will largely depend on the number of men and women who act as missionaries for it. So long as Mahomet is a lonely man, he can only influence those in his own district. With every convert his power to make himself heard increases. It is possible to find cases where statistics show that a movement has grown according to an exponential law, subject of course to slight variations, due to other causes and particular events which helped or hindered the movement. The fact that a movement grows in this way *during a certain period* tells us nothing at all about its future prospects: it may be smashed by bad leadership, or by disillusion, or by superior force, or by sheer bad luck. When human events show a mathematical law, it means that one or two simple causes were decisive at a certain time: the more different causes are at work, the more complicated will the graph of the movement become.

But the place in which an exponential function really feels at home is far from the complications of human or animal life. In the sciences of non-living matter, exponential functions abound; the speed of a body moving against air-resistance, the pressure of the atmosphere at different heights, the vibrations of an electric circuit, the passage of electrons through a gas, the decay of radium, the speed of a chemical reaction, the current in an electro-magnet, the dying away of any vibration – in these and in countless other problems some quantity grows or shrinks at a rate proportional to its size. It is indeed remarkable how much of the physical world, amid the conflicting action of a great variety of unconnected forces, can be described by the simplest mathematical functions, x^n and e^x.

CHAPTER 15

THE SQUARE ROOT OF MINUS ONE

'*The prevalent idea of mathematical works is that you must understand the reason why first, before you proceed to practise. That is fudge and fiddlesticks. I know mathematical processes that I have used with success for a very long time, of which neither I nor anyone else understands the scholastic logic. I have grown into them, and so understand them that way.*' – Oliver Heaviside.

At the end of Chapter 5 we saw that the square of every number had a + sign, so that no number could exist with a square equal to —1. One would naturally expect the matter to rest here, and mathematicians to admit that any problem which led to the equation $x^2 = -1$ was meaningless and had no solution.

But a strange thing happened. From time to time mathematicians noticed that their work could be much shortened, and the correct answer obtained, if in the middle of the working they used a sign i, assumed i^2 to be —1, and in all other respects treated i just as if it were an ordinary number. This was first done about 1572. The people who did it were very doubtful about the method, but it kept on giving the correct answers. Nobody knew why it should, but the new sign i proved so useful that for two centuries mathematicians used it, without any justification other than success. It was not until 1800 that a logical explanation of the meaning of i was given. (The whole story will be found in Dantzig's *Number, the Language of Science*.)

If, for the moment, we allow i to be treated as an ordinary number, we can see the type of result obtained by eighteenth-century mathematicians.

In Chapter 14 we found series for e^x *cos x*, and *sin x*. You may have noticed that the same numbers came into these series. In fact, if you take the series for e^x:

$$1+x+\tfrac{1}{2}x^2+\tfrac{1}{6}x^3+\tfrac{1}{24}x^4+\tfrac{1}{120}x^5+\tfrac{1}{720}x^6 \ldots$$

and leave out every other term:

$$1+\tfrac{1}{2}x^2+\tfrac{1}{24}x^4+\tfrac{1}{720}x^6+\ldots$$

and then make the signs alternately + and —

$$1-\tfrac{1}{2}x^2+\tfrac{1}{24}x^4-\tfrac{1}{720}x^6\ldots$$

you obtain the series for *cos x*. In the same way, *sin x* corresponds to the other half of the terms.

By making use of the sign *i* we can express the relation between the three series in a short formula.

Let us suppose x to have the value ia, where a may be any number. Putting $x=ia$ in the series for e^x we have:

$$e^{ia}=1+ia+\tfrac{1}{2}i^2a^2+\tfrac{1}{6}i^3a^3+\tfrac{1}{24}i^4a^4+\ldots$$

i^2 we know is -1, $i^3=i\times i^2=-i$. $i^4=i^2\times i^2=(-1)\times(-1)=+1$. In the same way, all the higher powers of i turn out to be 1, i, -1, and $-i$ in turn. Accordingly,

$$e^{ia}=1+ia-\tfrac{1}{2}a^2-\tfrac{1}{6}ia^3+\tfrac{1}{24}a^4+\ldots$$

If we sort out the terms which contain i from those which do not, we see that the terms without i give the series for *cos a*, while the terms containing i are equal to i times the series for *sin a*. In short:

$$e^{ia}=\cos a+i\sin a$$

This is rather a surprising result. e^x is a simple type of function, like 2^x or 3^x. We meet functions of this type very early – in the arithmetic course, for compound interest. e is just a number between 2 and 3, round about 2·7.

cos a and *sin a* are quite different. We meet them first of all in connexion with geometry, as the sides of a right-angled triangle. We have no reason at all to expect that they will be connected with any of the simpler formulae of algebra: in fact, it is a mystery to most people how the tables for *sin a* are worked out at all.

The formula above shows that *cos a* and *sin a* are in fact closely connected with the simplest type of function. e^x has many simple properties; for instance, $e^p\times e^q=e^{p+q}$, p and q standing for any two numbers. If we take $p=ia$ and $q=ib$ we obtain the results below:

$$e^{ia}\,e^{ib}=e^{i(a+b)}$$

that is $(\cos a+i \sin a)(\cos b+i \sin b)=\cos(a+b)+i \sin(a+b)$. If we multiply out the expression on the left-hand side, and compare the two sides, we see that $\cos(a+b)$, the part free from i, must be the same as $(\cos a \cos b-\sin a \sin b)$, while the number that appears with i, $\sin(a+b)$, must be the same as the number that appears with i on the other side, $(\cos a \sin b+\sin a \cos b)$.

All the formulae given for sines and cosines in books on trigonometry can be obtained, usually without very much work, from the properties of e^x. This fact can be used to take a burden off the memory. Instead of learning the formulae, one can work them out whenever they are needed, by using e^{ia}.

It is easy to find formulae giving $\cos a$ and $\sin a$ in terms of e^{ia}. $\cos a$, in fact, equals $\frac{1}{2}(e^{ia}+e^{-ia})$. We can turn any problem about cosines into a problem about exponentials. For instance, we can immediately find $\int(\cos x)^6 dx$ by this means, for exponential functions are easy to integrate.

We can regard all problems about sines and cosines as problems about exponentials, thus saving ourselves the trouble of learning special methods for dealing with sines and cosines. i is therefore a very helpful device: as was mentioned in Chapter 5, electrical engineers make great use of it.

What is i?

It seems at first sight very strange that the square root of minus one – something which no one has ever seen, and which seems in its own nature to be impossible – should be useful for such material tasks as the design of dynamos, electric motors, electric lighting, and wireless apparatus.

When some natural fact strikes us as strange, it means that we are looking at it from the wrong point of view. If we find the universe mysterious, it is because we have some idea about what the universe ought to be like, and are then surprised to find it is something different. The fault lies with our original idea – not with the universe.

When we find i mysterious, it is because we are thinking of i as an ordinary number. But there is no *number* x for which $x^2 = -1$. We convinced ourselves of this in Chapter 5.

We have also seen that the sign i, for which $i^2 = -1$ is assumed, leads to perfectly correct results, and clearly has *some* meaning. It is impossible that i should be a *number*; but there is no contradiction at all if we suppose i to be something else. i, in fact, can be interpreted as an operator.

An operation means doing something: 'Turn the piano upside down', 'Move two paces to the right', 'Throw Mr Jones out' are examples of operations applied to a piano, a soldier, and Mr Jones. If we use $\mathbf{U}$ as an abbreviation for 'turn upside down', and p for 'the piano', $\mathbf{U}p$ has the same meaning as the first sentence given above. $\mathbf{U}$ is called an *operator*. The operation can be repeated. $\mathbf{UU}p$ would mean 'Turn the piano upside down, and then turn it upside down again', which would, of course, bring the piano back to its original position. Usually $\mathbf{UU}$ is represented by $\mathbf{U}^2$, $\mathbf{UUU}$ by $\mathbf{U}^3$, etc. Since turning the piano upside down twice leaves it in its original position, $\mathbf{UU}p = p$, or $\mathbf{U}^2 p = p$. It is convenient to use the sign $\mathbf{1}$ for the operation of leaving something alone. Thus $\mathbf{1}p$ means the result of leaving the piano alone, which is just the piano in its original position, p. So $\mathbf{U}^2 p = \mathbf{1}p$. This sort of result is not only true for a piano; the result of turning any solid body (not a glass of water, however!) upside down twice is the same as that of leaving it alone. We express this by the equation $\mathbf{U}^2 = \mathbf{1}$.

You will see that it is perfectly possible to argue about operations, and to get results about them, the truth of which can be seen purely by common sense. You will also see that these results, when stated in the shorthand of algebra, *look* like equations for numbers, and might easily be mistaken for statements about numbers. And in fact it is precisely this mistake which has been made in connexion with the equation $\mathbf{i}^2 = -\mathbf{1}$. To avoid any such misunderstanding, we shall print all signs representing operations in heavy type. From now on, in particular, we shall write $\mathbf{i}$ instead of i, so as to make it clear that we are *not* dealing with the sign for a number.

While operators are not numbers, they are often closely connected with numbers. In an adding machine, for instance, we have a number of gear wheels arranged in the same way as the wheels in the mileage recorder of a motor-car. Every time a motor-car goes a mile, the wheel representing units turns through one division, and one is added to the mileage. The turning of the wheels is an operation, and this operation corresponds to adding one to the mileage. It is because of this correspondence between numbers and certain mechanical operations that it is possible to make calculating machines at all.

We shall now find a set of operators **1**, **2**, **3** ... and **+** which correspond very closely indeed to 1, 2, 3 ... and + in ordinary arithmetic. **1**, **2**, **3** ... are not numbers, but there are many relations between these operators which correspond to those between ordinary numbers: they have a pattern in common with ordinary numbers, just as a family of father, mother, son, and daughter in a colony of chimpanzees have a pattern in common with a family of father, mother, son, and daughter in Birmingham. This is not to say that everyone in Birmingham is a chimpanzee.

We shall then find it quite natural to bring in an operation, **i**, such that $\mathbf{i}^2 = -\mathbf{1}$.

The Operators **1**, **2**, **3** ....

To define the operators **1**, **2**, **3** ... we begin by imagining a long lath of wood, on which a scale has been marked. O is any fixed point and the figures 1, 2, 3, etc., are marked at distances of 1, 2, and 3 inches to the right. Going to the left from O, we find the figures —1, —2, —3, etc., at distances 1, 2, 3 inches. This is an ordinary scale, such as might be marked on a thermometer.

One end of a wire is fastened to O, and on this slides a bead, A. The wire may point either to the right or the left. The operations we shall consider will consist either in turning the wire from one direction to the other, or in sliding the bead A along the wire.

The operation **2** may now be defined. It consists in moving the bead A to a point on the wire twice as far from O as the former position of A. The operation **2** may be put into words, 'Double the distance OA'. In the same way, the operation **3** means

'Make OA 3 times as long.' $\mathbf{2\frac{7}{8}}$ means, 'Make OA $2\frac{7}{8}$ times as long.' $\mathbf{x}$ means 'Make OA x times as long', where x stands for any positive number. **1** will mean, 'Leave A where it is.'

Several operations may be carried out one after the other. For instance, the operation **(4) (3) (2)** means that the length OA has to be doubled, then trebled, then again made four times as large: in short, OA has to be made 24 times its original length. The three operations, applied in turn, are equivalent to the single operation **24**. **(4) (3) (2)=24**. So there is a close correspondence between the doing in turn of various operations and the multiplication of ordinary numbers. We might say that the operations have the same multiplication table as ordinary numbers.

By the operation **—1** we understand that the direction of the wire is reversed, but the distance OA is left unchanged. Thus, if A was originally above the mark 3, the operation **—1** would cause it to come over the mark —3. If A was originally at —3, the operation **—1** would bring it to 3.

By **—x** we understand that the distance OA is made x times as large, and its direction reversed.

Check for yourself that **(—2) (3)=—6** and **(—4) (—5)=20**. The rules for multiplying **+** and **—** operators are the same as those for ordinary numbers.

Addition

How shall we find a meaning for **2+3** or **2+ (—3)**? We *might* say straight away that **2+3** is to be **5** and that **2+ (—3)** is to be **—1**; that is, we could use ordinary arithmetic as a way of defining addition for operators. But this would have the disadvantage that when we come to consider **i**, which does not correspond to any ordinary number, we should not know what to take for **1+i**.

It is therefore more satisfactory to make use of another method which will apply equally well to operations such as **i**, while agreeing with the first method for operators that correspond to ordinary numbers.

We suppose the bead A to be at any point P, to begin with. Let

Q be the point to which the operation **2** sends A, R the point to which the operation **3** sends A, and S the point to which A is sent by the operation **5**. Then $OQ=2.\ OP$; $OR=3.\ OP$; $OS=5.\ OP$. (OP stands for the distance from O to P, 2, 3, 5 for the ordinary numbers: there are no operators in these equations.) it is obvious that $OS=OQ+OR$, so that we could find the position of S by putting lengths equal to OQ and OR end to end.

In the same way we could find the effect of the operation **2+(−3)**. We must remember that **2** and **−3** will cause OA to point in opposite directions: when we put the two lengths end to end, it must be in such a way that the second length points in the opposite direction to the first. Fig. 24 may help to make this clear.

2+3=5

2+(−3)=−1

−2+3=1

−2+(−3)=−5

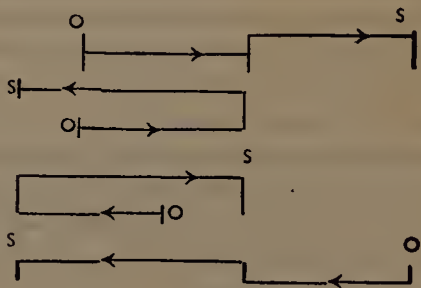


Fig. 24

Accordingly we are led to define addition as follows: if the operation **x** sends A from P to Q, and the operation **y** sends A from P to R, **x+y** is defined as the operation which sends A to S, where S is the point obtained by putting OQ and OR end to end.

We may write **2+ (−3)** more shortly as **2−3**. Be careful to distinguish between **2−3** and **(2)·(−3)**. **(2) (−3)** means that the operation **2** has to be applied to the result of **−3** acting on A.

We have now found a set of operators which correspond completely to the ordinary numbers: they can be multiplied and added, and the answers look exactly like the answers for ordinary numbers, except that they are in heavy type. If you picked up a page of calculations, dealing with these operators, you might mistake it for examples in elementary arithmetic, written by someone who pressed very hard on the pen: there would be no way of distinguishing between the two.

The Operator i

The operation **—1** has the effect of reversing the direction of OA, without altering its length – i.e., it rotates OA through 180°.

Can we find an operation **i** such that $\mathbf{i}^2 = \mathbf{-1}$? $\mathbf{i}^2$ means that the operation **i** is carried out twice. The question asks us to find an operation which, performed twice, turns OA through 180°.

The question is now ridiculously simple. The mysterious operation **i** consists simply in turning OA through 90°! In Fig. 25 A is supposed to start by being at P. The operation **i** sends A to E, $\mathbf{i}^2$ sends A to F, $\mathbf{i}^3$ sends A to G, and $\mathbf{i}^4$ brings A back again to P. **—1** sends A from P to F, and $\mathbf{i}^2 = \mathbf{-1}$, as we hoped.

Before we brought **i** into the picture, the bead A moved on a straight line. It could lie to the right or the left of O, but always

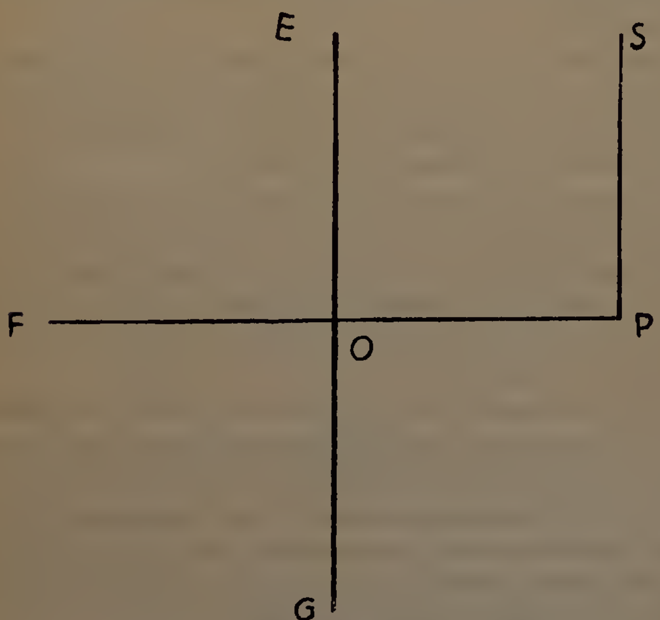


Fig. 25

at the same level as O. Now that **i** has been brought in, it is possible for A to lie above or below O, and in fact we shall soon have A wandering over the whole surface of the paper.

Addition

The 'end-to-end' method of addition can now be used to give a meaning to expressions such as **1+i** and **2+3i**.

Suppose, as before, that the bead A is at P, and is then acted upon by the operation **1+i** (Fig. 25). Where will **1+i** send it? We have to find the points Q and R to which **1** and **i** send A, and then put OQ and OR end to end. The operation **1** leaves A at P. The operation **i** sends A to E. So Q is P and R is E. We have to put OP and OE end to end. At P we draw PS equal to OE, and having the same direction as OE. This gives us the point S, which we require. The operation **1+i** sends A from P to S. If A starts at any other point, we can find where the operation **1+i** sends it. (For instance, if the bead A were at E to begin with, where would **1+i** send it?)

In the same way, the operation **2+3i** can be studied. The bead A may start at any point of the paper, say at K (Fig. 26). We have to study the operations **2** and **3i** separately, and then to combine them by 'end-to-end' addition.

The operation **2** would send A from K to L. **3i** acting on A means that OA has to turn through 90° and become 3 times as long. Thus **3i** would send A from K to M. Now OM has to be put on to the end of OL. We draw LN equal to OM, and in the same direction as OM. N is the point we are looking for. The operation **2+3i** sends A from K to N.

K may be chosen anywhere. We could choose K in different positions, and notice where the corresponding N came. You may notice that the angle KON is always the same, and that ON is always in the same proportions to OK, wherever K may be chosen, In other words, wherever the bead A may be, the operation **2+3i** turns OA through the same angle and stretches the length OA the same number of times.

We may picture any operation **a+bi** as being a turn through

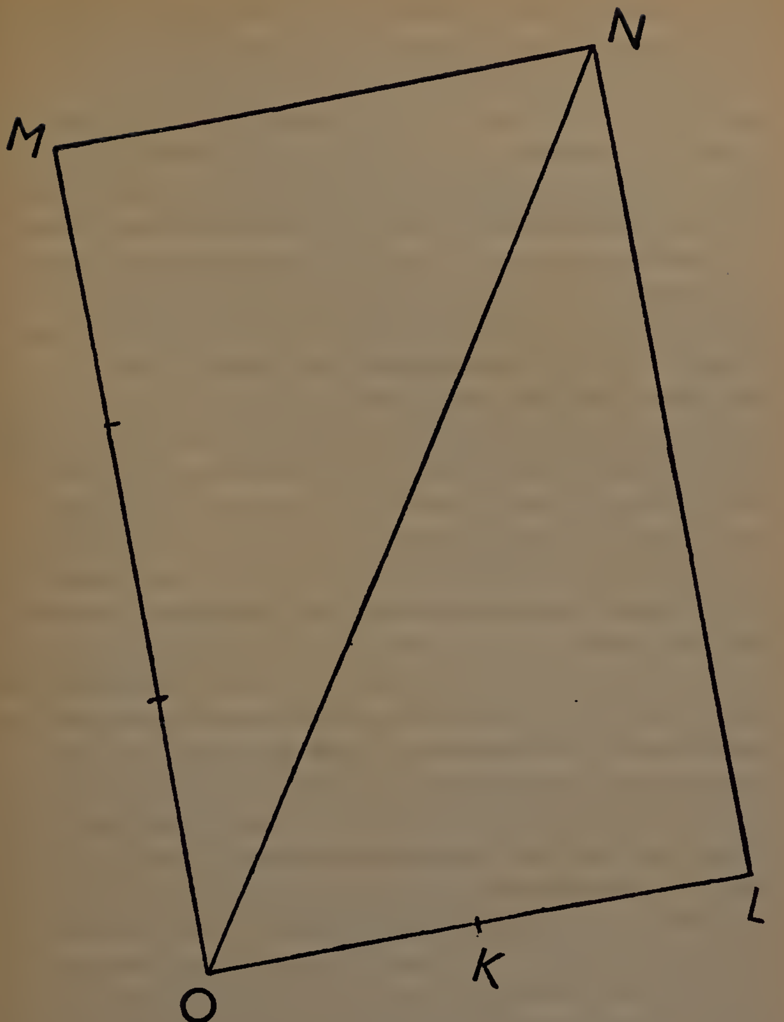


Fig. 26

some angle followed by a stretch. If **a+bi** corresponds to a turn through the angle θ and a stretch r times, it is easy to see that $a = r \cos \theta$ and $b = r \sin \theta$. If we are given a and b, we can find r and θ graphically by drawing a right-angled triangle, with sides a and b. r is known as the *modulus*, θ as the *argument* (sometimes the *amplitude*) of the operation **a+bi**. Special names are given to these two quantities, because they arise naturally and often occur in formulae: we therefore save time by giving them names.

You are now in a position to think about these operations for yourself. We have shown, by the examples $1+i$ and $2+3i$, how to find the operation represented by any symbol of the type $a+bi$. You now understand what the symbols represent, and it is up to you to make yourself familiar with the actual operations, by experimenting with them. What do you understand by the operations $1+2i$, $1-i$, $-3i$? If two operations are carried out in turn, does it matter in which order this is done? Is $(2)(i)$ the same as $(i)(2)$? Is $(1+i)(1+2i)$ the same as $(1+2i)(1+i)$? By carrying out the actual geometrical operations, find a single operation which has the same effect as $(1+2i)(1+i)$. What is the modulus of i? Of $3+4i$? What operation is $-i$? Does $(i)(i)=(-i)(-i)$? What is $(i)(-i)$? $-i$ means the operation $(-1)(i)$.

Once we have given a definite meaning to the symbols, we lose all control over them. We can decide what name to give any operation, but once we have chosen the name, we have to observe what that operator actually does. We have reached that stage. We have given the names $1, 2, 3 \ldots$ and i to certain operators, and have explained what we mean when two operators are written side by side, or are linked by the signs $+$ and $-$. Division is to mean the opposite of multiplication. The only sign which has not yet been given a meaning is e^x, which we shall come to later. But so far as addition, subtraction, multiplication, and division are concerned, everything is settled. We must not *assume* that these new signs obey the same rules as ordinary numbers: for they are not ordinary numbers. We must try to see whether they do.

For example, we must not *assume* that $(2)(i)$ is the same as $(i)(2)$. Actually $(2)(i)$ is equal to $(i)(2)$, but we must convince ourselves that this is so by trial. There are operators for which the order in which they are performed alters the answer. The effect of being beaten and then beheaded is different from being beheaded and then beaten.

The interesting thing about the operators we are now dealing with is that they behave just like ordinary numbers. If you take any formula which is true for ordinary numbers, it will be true for these operators. For instance, $(x+1)(x-1)=x^2-1$, when x is an ordinary number. If we put any operator $\mathbf{a+bi}$ in the place of x, we find the result (in heavy type) still holds true. For instance, putting $\mathbf{i}$ in place of x, it is true that $\mathbf{(i+1)(i-1)=i^2-1}$. $\mathbf{i^2}$ is $\mathbf{-1}$, so $\mathbf{i^2-1=-2}$. You will find that the result of carrying out the operations $\mathbf{i-1}$ and $\mathbf{i+1}$, one after the other, is to double the length OA, and to turn it through 180°; that is, to do what the operation $\mathbf{-2}$ does.

You will find, too, that it does not matter in what order multiplication is done, or in what order signs are added to each other. $\mathbf{i2}$ and $\mathbf{2i}$ have exactly the same meaning (multiplication meaning that the operations are carried out, one after the other) and $\mathbf{i+1}$ has the same meaning as $\mathbf{1+i}$ (it does not matter which line is put to the end of the other, in 'end-to-end' addition). In short, any rule which is true for ordinary numbers is true for these operators.

This fact is extremely convenient. Often, when we begin to study a new type of operation, we find laws which are entirely fresh to us. Each type of operation has its own particular way of behaving, which we have to get used to. But we do not have to learn any new rules for the operators $\mathbf{a+bi}$. They behave exactly *as if* they were numbers: while they are not, in fact, numbers, they yet have so much in common with numbers that, *for most purposes*, they can be thought of as numbers. Mathematicians usually refer to them as *complex numbers*, to show that they are close relations of the ordinary numbers. If, in making some calculation, you treat $\mathbf{i}$ as if it were an ordinary number, you will obtain the correct result.

On the other hand, by thinking of $\mathbf{i}$ as an operator, you can often see results more quickly than by using the methods of ordinary arithmetic. For instance, you may be asked to solve the equation $\mathbf{x^2=i}$. We know that $\mathbf{i}$ represents a rotation through a right-angle. We are asked: what operation $\mathbf{x}$, carried out twice, has the same effect as $\mathbf{i}$? The answer is obvious: turn through half a right-angle. This operation does not involve any stretching, so its modulus $r=1$. (*No stretching* does *not* mean $r=0$. The length OA is *multiplied* by r. If OA is unchanged, this means $r=1$.) Since the angle θ is half a right-angle, it is easily seen that the numbers a and b must be 0·707 and 0·707 (from a table of sines and cosines), and the operation $\mathbf{a+ib}$, which represents a turn through half a right angle, is $\mathbf{0{\cdot}707+0{\cdot}707i}$. (This is an outline only. Draw the figure for yourself. Also check the result by ordinary arithmetic, treating **i** as if it were an ordinary number.)

Complex Numbers and Electricians

It is now easier to see why electricians make such frequent use of the operator **i**. Every generator of electric current contains parts which are rotating, and which every minute pass through many right angles – that is, have the operation **i** applied to them, again and again.

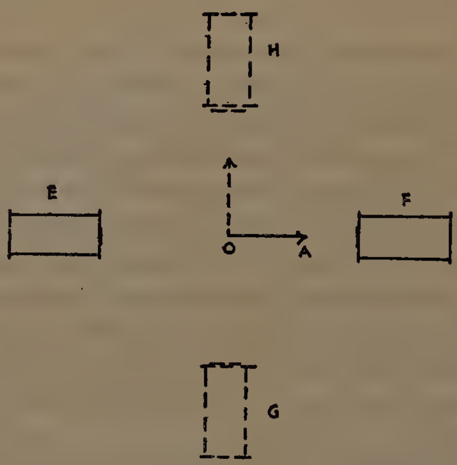


Fig. 27

It would be possible, and for electrical students interesting, to explain **i** entirely in terms of a simple generator of alternating current. For mathematical simplicity, it is best to consider a design of generator very different from that actually used in engineering practice.

We suppose a small coil be be rotating in a magnetic field. The direction of the magnetic field may be represented by means of an arrow, and the strength of the magnetic field may be represented by the length of this arrow. In Fig. 27, OA is an arrow, representing the magnetic field. This arrow, OA, takes the place of the line OA (joining a fixed point O to a bead A). 'Turning OA' means that we change the direction of the magnetic field. 'Stretching OA' means that we make the field stronger. Both operations can easily be done if the magnetic field is produced by electro-magnets, mounted on a bar which can rotate about the fixed point O.

We may take as a standard situation that in which the electro-magnets are at E and F and a current of 1 ampere is flowing through them. The operation **a** is to mean that the current in the electro-magnets is increased until the magnetic field at the centre, O, is a times as strong as before.

The operation **ib** would mean that we start from the standard position, make the field b times its standard strength, and then turn through a right angle. The coils would then be in the positions, shown dotted, at G and H, and the magnetic field would be represented by the dotted arrow.

The operation **a+ib** can be interpreted by supposing we combine the two arrangements. We have coils at E and F, through which there flows a current sufficient to produce a magnetic field of a units at O, and *at the same time* we have coils at G and H with a current strong enough to produce a field of b units at the centre. The *combined effect* will be to produce a magnetic field in a direction which lies between the directions OF and OH. The precise position of the arrow representing the combined effect is given by exactly the same rule as before – 'end-to-end addition', more commonly known as the Parallelogram Law, or the Triangle of Forces.

It will be clear to electricians that the arrow representing the magnetic field can be brought to *any* desired position by suitable choice of the size (and direction) of the currents in the circuits $E-F$ and $G-H$. That is, any operation consisting of 'a turn and a stretch' can be put in the form **a+ib**.

To avoid complicating the figure, the small coil rotating about O has not been drawn. The changes in the direction and strength of the magnetic field will, of course, produce corresponding changes in the phase and amplitude of the alternating current generated. It is natural that the operator **i** should be used in connexion with alternating currents, to show the changes produced by including extra resistance, inductance, etc. What we really do, in using the sumbol **i**, is to compare the effect of such changes in the circuit with the effect of certain changes (represented by signs such as $\mathbf{a+bi}$) made *inside* the generator which produced the current.

The Further Study of i

One question which has been raised in this chapter, but not yet answered, is how to define $\mathbf{e^x}$, when $\mathbf{x}$ is a complex number. $\mathbf{e^x}$, in heavy type, is a new label, and we could (if we chose) attach this label to any operation whatever. But this would be very misleading: we should always have to remember that the operation chosen had nothing whatever to do with the ordinary e^x, and we should always be liable to make mistakes through forgetting the difference. On practical grounds it is clearly best not to use the label $\mathbf{e^x}$ at all, unless we can find some operation which has properties very similar to those of e^x.

We have already found operations whose claim to the labels of $\mathbf{x}$, $\mathbf{x^2}$, $\mathbf{x^3}$, etc., has been admitted, and we know that e^x can be expressed by means of a series containing x, x^2, x^3, etc. It would therefore be natural to form the corresponding series, in heavy type, and to define $\mathbf{e^x}$ by saying:

$$\mathbf{e^x=1+x+\tfrac{1}{2}x^2+\tfrac{1}{6}x^3+\tfrac{1}{24}x^4+\tfrac{1}{120}x^5+\ldots}$$

This definition is, in fact, quite satisfactory. It gives us a meaning for $\mathbf{e^x}$ which can be proved to have all the properties of the ordinary e^x.

It is important to understand what is implied in this definition. If, for example, we wished to find $\mathbf{e^{2+3i}}$ by means of this series, we should have to replace $\mathbf{x}$ by $\mathbf{2+3i}$, giving:

$$\mathbf{e^{2+3i}=1+(2+3i)+\tfrac{1}{2}(2+3i)^2+\tfrac{1}{6}(2+3i)^3+}\text{ etc.}$$

We should then have to work out $\mathbf{(2+3i)^2}$, $\mathbf{(2+3i)^3}$, etc., and put the answers in. $(\mathbf{2+3i})^2$ turns out to be $\mathbf{-5+12i}$, $(\mathbf{2+3i})^3$ is $\mathbf{-46+9i}$, and so on. Taking into account the numbers $\frac{1}{2}$, $\frac{1}{6}$, etc. which occur in the series, we thus find:

$$\mathbf{e}^{\mathbf{2+3i}}=\mathbf{1+(2+3i)+(-2\tfrac{1}{2}+6i)+(-7\tfrac{2}{3}+1\tfrac{1}{2}i)+}\ldots$$

We may now collect together the terms which contain $\mathbf{i}$, and those which are free from $\mathbf{i}$, so that:

$$\begin{aligned}\mathbf{e}^{\mathbf{2+3i}}=\mathbf{1+2-2\tfrac{1}{2}-7\tfrac{2}{3}}\ldots\\ \mathbf{+i(3+6+1\tfrac{1}{2})}\ldots\end{aligned}$$

This step is justified, because $\mathbf{i}$ can be treated just like x in ordinary algebra, as we noted earlier.

But this result will be no use *unless* we find that the series $\mathbf{1+2-2\tfrac{1}{2}-7\tfrac{2}{3}}$ ... and the series $\mathbf{3+6+1\tfrac{1}{2}}$ ... settle down to steady values, when sufficiently many terms are taken. Nor will these series be any use if they turn out to be 'dangerous series' of the type described in Chapter 14.

Actually, these two series are very tame and reliable. The later terms in the series for e^x contain numbers such as $\frac{1}{24}$, $\frac{1}{120}$, $\frac{1}{720}$, which rapidly become very small; and at a certain point, the later terms make hardly any difference to the sum of the series. The rule by which the numbers in e^x are formed is the following: $6=1\times2\times3$, $24=1\times2\times3\times4$, and so on. 120 is 5 times 24. 720 is 6 times 120. The farther we go, the more rapidly do the terms of the series decrease.

It can be proved that the series for $\mathbf{e}^{\mathbf{x}}$ is all right (in professional language, is *convergent*) whatever $\mathbf{x}$ may be. That is, if $\mathbf{x=a+ib}$, it does not matter how large the numbers a and b may be: the series will still converge. If a and b are large numbers, we may have to take a large number of terms before we get a good estimate of $\mathbf{e}^{\mathbf{a+ib}}$: nevertheless, as the series *does* define $\mathbf{e}^{\mathbf{x}}$, we have a logical foundation on which to build.

Actually, to find $\mathbf{e}^{\mathbf{a+ib}}$, it is better to proceed as follows. $\mathbf{e}^{\mathbf{a+ib}}=\mathbf{e}^{\mathbf{a}}.\ \mathbf{e}^{\mathbf{ib}}=\mathbf{e}^{\mathbf{a}}$ (*cos* $\mathbf{b}+\mathbf{i}$ *sin* $\mathbf{b}$). $\mathbf{a}$ and $\mathbf{b}$ correspond to ordinary numbers a and b. We can look up e^a, *cos b* and *sin b* in tables. But this procedure is possible only *after* we have proved (by means of the series for $\mathbf{e}^{\mathbf{x}}$) that $\mathbf{e}^{\mathbf{x}}$ has all the properties of e^x, so that the steps taken above are justified. If our definition of $\mathbf{e}^{\mathbf{x}}$, by means of the series, is not water-tight, then we cannot trust results obtained from this definition.

Mathematicians are therefore forced to study the convergence of series in which complex numbers occur. They have also studied what meaning can be attached to the sign $\frac{dy}{dx}$ when **x** and **y** are complex numbers.

We have already seen that the use of **i** allows us to show a close connexion between e^x, *sin x*, and *cos x*, a connexion that is surprising, since e^x appears at first sight very different from *sin x* and *cos x*. We have also seen that this connexion is of practical use, as it helps us to see the solution of many problems about sines and cosines.

In the same way, the further study of complex numbers throws light on many problems about ordinary numbers. In fact, the subject of complex numbers is one of the most beautiful and instructive departments of mathematics. It gives one the feeling of having been taken behind the scenes: one sees easily and quickly the reasons for results which had previously seemed quite accidental. It is a subject in which calculation plays a small part: its results frequently take forms which one can *see*, and remember, as one remembers a striking poster. By enabling one to see the inner significance of many practical problems, it is therefore of great value for applied mathematicians.

No one could have foreseen that the study of **i** would lead to such welcome results, any more than the first men who played with magnets and silk could have foreseen the application of electro-magnetic theory to the invention of wireless. In both cases, it just turned out to be so.

When you first learn to use **i**, you will suffer from a feeling of strangeness. The subject will seem unreal to you. That is inevitable. Any new subject feels strange at first. When wireless first became popular, people felt it to be strange. But the children who are born today take wireless for granted. If, as a war economy, all wireless were to stop, people would say, 'How strange it is, having no wireless!' But no one had wireless in 1914–18, and no one felt it to be strange. Nothing is either strange or familiar in itself. Anything is strange the first time you meet it: anything is familiar when you have known it five years. The more you use **i**, the more you will come to feel that **i** is a natural and reasonable thing. But this feeling can come only gradually.

Complex numbers show pure mathematics at its best. Pure mathematics is the study of method. Given any problem, we want to know the best way of attacking it. Many problems, at first sight hard, become simple only if one can look at the problem from the proper angle, if one can see the problem in its proper setting. It is the job of pure mathematicians to classify problems, to suggest that this problem is essentially similar to that, and likely to yield to a certain type of attack. It may not be in the least obvious that the problems are connected: it is not in the least obvious that the equation $\mathbf{x}^2=-\mathbf{1}$ is going to throw light on the question of electric lighting. The less obvious the connexion is the more credit must go to the pure mathematician for discovering it; the harder the problem appears, the greater is the credit due for showing that it is connected with some simpler problem.

Engineers do not need to know more than the most elementary results about complex numbers. The more advanced results are chiefly of interest to professional mathematicians, who are inventing and perfecting new methods, which, when complete, can be used by scientists and practical men. Anyone with a taste for mathematics should try, when as young as possible, to gain some knowledge of complex number theory. Books on the subject have such titles as *The Theory of Complex Variables*, *The Theory of Functions*, etc. Too often, boys at school fail to realize how much mathematics there is to know. Talented boys find themselves ahead of their fellows, and begin to think they have a mastery of mathematics. As a result, they waste their last year at school. The first year at college (for those who are able to go there) they meet the best boys from other schools, and experience a tremendous shock.

INDEX

Introducing Mathematics 3

THE SEARCH FOR PATTERN

Anyone who has read *Mathematician's Delight* or *Prelude to Mathematics* knows W. W. Sawyer as a mathematical lion-tamer. Figures do not merely come to life for him; they eat out of his hand.

Introducing Mathematics is designed to give readers the essentials of mathematical literacy. The first volume, *Vision in Elementary Mathematics*, set out the basic processes and routines of arithmetic and algebra; having mastered these, readers might well ask: 'Right! what can we do with them?'

In *The Search for Pattern* the author illustrates such themes as the Binomial Theorem or quadratic equations by showing examples of their use in scientific work or in those books which make some use of measurements, graphs, or statistics. The reproductive pattern among blowflies, the incidence of human twins, the design of the Morse Code, and the random behaviour of molecules, dice, and football teams are among the lively examples by which he shows mathematics in action.

Those who (without being specialists) need a quantum of mathematical knowledge in their work or their reading will gain it here more informally and painlessly than from standard textbooks. And W. W. Sawyer constantly has the teacher in mind in suggesting ways of bringing figures to life.

Prelude to Mathematics

W. W. SAWYER

In *Mathematician's Delight* the topics were selected mainly for their practical utility – the traditional mathematics of the engineer and scientist. In *Prelude to Mathematics* the emphasis is not on those branches of mathematics which have great practical utility but on those which are exciting in themselves: mathematics which is strange, novel, apparently impossible; for instance, an arithmetic in which no number is bigger than four. These topics are preceded by an analysis of that enviable attribute 'the mathematical mind'.

Mr Sawyer not only shows what mathematicians get out of mathematics, but also what they are trying to do, how mathematics grows, and why there are new mathematical discoveries still to be made. His aim is to give an all-round picture of his subject, and he therefore begins by describing the relationship between pleasure-giving mathematics and that which is the servant of technical and social advance.

published by Penguin Books

This volume is designed to convince the general reader that mathematics is not a forbidding science but an attractive mental exercise. Its success in this intention is confirmed by some of the reviews it received on its first appearance:

'It may be recommended with confidence for the light it throws upon the discovery and application of many common mathematical operations' – *The Times Literary Supplement*

'It jumps to life from the start, and sets the reader off with his mind working intelligently and with interest. It relates mathematics to life and thought and points out the value of the practical approach by reminding us that the Pyramids were built on Euclid's principles three thousand years before Euclid thought of them' – *John O'London's*

'The writer clearly not only loves his subject but has unusual gifts as a teacher . . . from start to finish the reader, whose own interests and training may lie in very different fields, can follow the thread' – *Financial News*

Cover design by John and Kenneth Astrop

Unite
Canad
U.S.A