SEVENTEEN EQUATIONS

THAT CHANGED THE WORLD

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
00Seventeen_Prelims.3d Page 1 of 10

SEVENTEEN EQUATIONS

THAT CHANGED THE WORLD

IAN STEWART

PROFILE BOOKS

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
00Seventeen_Prelims.3d Page 3 of 10

First published in Great Britain in 2012 by
PROFILE BOOKS LTD
3A Exmouth House
Pine Street
London EC1R 0JH
www.profilebooks.com

1 3 5 7 9 10 8 6 4 2

Printed and bound in Great Britain by
Clays, Bungay, Suffolk

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

ISBN 978 1 84668 531 6
eISBN 978 184765 769 5

FSC
Mixed Sources
Product group from well-managed forests and other controlled sources
Cert no. SGS-COC-2061
www.fsc.org
© 1996 Forest Stewardship Council

The paper this book is printed on is certified by the © 1996 Forest Stewardship Council A.C. (FSC). It is ancient-forest friendly. The printer holds FSC chain of custody SGS-COC-2061

Contents

Profile Books - Seventeen Equations Data Standards Ltd, Frome, Somerset – 5/12/2011
00Seventeen_Prelims.3d Page 6 of 10

To avoide the tediouse repetition of these woordes: is equalle to: I will sette as I doe often in woorke use, a paire of paralleles, or gemowe lines of one lengthe: ═══, bicause noe .2. thynges, can be moare equalle.

Robert Recorde, *The Whetstone of Witte*, 1557

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
00Seventeen_Prelims.3d Page 7 of 10

Why Equations?

Equations are the lifeblood of mathematics, science, and technology. Without them, our world would not exist in its present form. However, equations have a reputation for being scary: Stephen Hawking's publishers told him that every equation would halve the sales of *A Brief History of Time*, but then they ignored their own advice and allowed him to include $E = mc^2$ when cutting it out would allegedly have sold another 10 million copies. I'm on Hawking's side. Equations are too important to be hidden away. But his publishers had a point too: equations are formal and austere, they look complicated, and even those of us who love equations can be put off if we are bombarded with them.

In this book, I have an excuse. Since it's *about* equations, I can no more avoid including them than I could write a book about mountaineering without using the word 'mountain'. I want to convince you that equations have played a vital part in creating today's world, from mapmaking to satnav, from music to television, from discovering America to exploring the moons of Jupiter. Fortunately, you don't need to be a rocket scientist to appreciate the poetry and beauty of a good, significant equation.

There are two kinds of equations in mathematics, which on the surface look very similar. One kind presents relations between various mathematical quantities: the task is to prove the equation is true. The other kind provides information about an unknown quantity, and the mathematician's task is to *solve* it – to make the unknown known. The distinction is not clear-cut, because sometimes the same equation can be used in both ways, but it's a useful guideline. You will find both kinds here.

Equations in pure mathematics are generally of the first kind: they reveal deep and beautiful patterns and regularities. They are valid because, given our basic assumptions about the logical structure of mathematics, there is no alternative. Pythagoras's theorem, which is an equation expressed in the language of geometry, is an example. If you accept Euclid's basic assumptions about geometry, then Pythagoras's theorem is *true*.

Equations in applied mathematics and mathematical physics are usually of the second kind. They encode information about the real

Profile Books - Seventeen Equations Data Standards Ltd, Frome, Somerset – 5/12/2011
00Seventeen_Prelims.3d Page 8 of 10

world; they express properties of the universe that could in principle have been very different. Newton's law of gravity is a good example. It tells us how the attractive force between two bodies depends on their masses, and how far apart they are. Solving the resulting equations tells us how the planets orbit the Sun, or how to design a trajectory for a space probe. But Newton's law isn't a mathematical theorem; it's true for physical reasons, it fits observations. The law of gravity might have been different. Indeed, it *is* different: Einstein's general theory of relativity improves on Newton by fitting some observations better, while not messing up those where we already know Newton's law does a good job.

The course of human history has been redirected, time and time again, by an equation. Equations have hidden powers. They reveal the innermost secrets of nature. This is not the traditional way for historians to organise the rise and fall of civilisations. Kings and queens and wars and natural disasters abound in the history books, but equations are thin on the ground. This is unfair. In Victorian times, Michael Faraday was demonstrating connections between magnetism and electricity to audiences at the Royal Institution in London. Allegedly, Prime Minister William Gladstone asked whether anything of practical consequence would come from it. It is said (on the basis of very little actual evidence, but why ruin a nice story?) that Faraday replied: 'Yes, sir. One day you will tax it.' If he did say that, he was right. James Clerk Maxwell transformed early experimental observations and empirical laws about magnetism and electricity into a system of equations for electromagnetism. Among the many consequences were radio, radar, and television.

An equation derives its power from a simple source. It tells us that two calculations, which appear different, have the same answer. The key symbol is the equals sign, =. The origins of most mathematical symbols are either lost in the mists of antiquity, or are so recent that there is no doubt where they came from. The equals sign is unusual because it dates back more than 450 years, yet we not only know who invented it, we even know *why*. The inventor was Robert Recorde, in 1557, in *The Whetstone of Witte*. He used two parallel lines (he used an obsolete word *gemowe*, meaning 'twin') to avoid tedious repetition of the words 'is equal to'. He chose that symbol because 'no two things can be more equal'. Recorde chose well. His symbol has remained in use for 450 years.

The power of equations lies in the philosophically difficult correspondence between mathematics, a collective creation of human minds, and an external physical reality. Equations model deep patterns in the outside world. By learning to value equations, and to read the stories

Profile Books - Seventeen Equations Data Standards Ltd, Frome, Somerset – 5/12/2011
00Seventeen_Prelims.3d Page 9 of 10

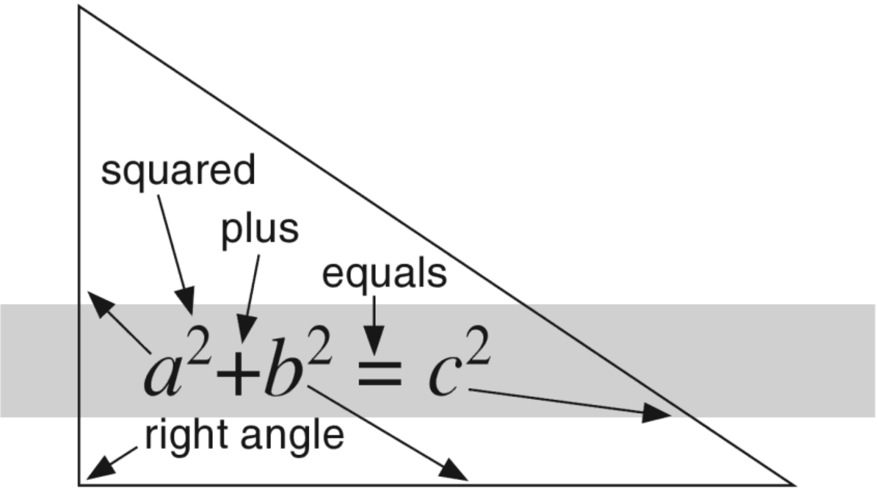
they tell, we can uncover vital features of the world around us. In principle, there might be other ways to achieve the same result. Many people prefer words to symbols; language, too, gives us power over our surroundings. But the verdict of science and technology is that words are too imprecise, and too limited, to provide an effective route to the deeper aspects of reality. They are too coloured by human-level assumptions. Words alone can't provide the essential insights.

Equations can. They have been a prime mover in human civilisation for thousands of years. Throughout history, equations have been pulling the strings of society. Tucked away behind the scenes, to be sure – but the influence was there, whether it was noticed or not. This is the story of the ascent of humanity, told through 17 equations.

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
00Seventeen_Prelims.3d Page 10 of 10

1 The squaw on the hippopotamus

Pythagoras's Theorem



What does it tell us?

How the three sides of a right-angled triangle are related.

Why is that important?

It provides a vital link between geometry and algebra, allowing us to calculate distances in terms of coordinates. It also inspired trigonometry.

What did it lead to?

Surveying, navigation, and more recently special and general relativity – the best current theories of space, time, and gravity.

Ask any school student to name a famous mathematician, and, assuming they can think of one, more often than not they will opt for Pythagoras. If not, Archimedes might spring to mind. Even the illustrious Isaac Newton has to play third fiddle to these two superstars of the ancient world. Archimedes was an intellectual giant, and Pythagoras probably wasn't, but he deserves more credit than he is often given. Not for what he achieved, but for what he set in motion.

Pythagoras was born on the Greek island of Samos, in the eastern Aegean, around 570 BC. He was a philosopher and a geometer. What little we know about his life comes from much later writers and its historical accuracy is questionable, but the key events are probably correct. Around 530 BC he moved to Croton, a Greek colony in what is now Italy. There he founded a philosophico-religious cult, the Pythagoreans, who believed that the universe is based on number. Their founder's present-day fame rests on the theorem that bears his name. It has been taught for more than 2000 years, and has entered popular culture. The 1958 movie *Merry Andrew*, starring Danny Kaye, includes a song whose lyrics begin:

The square on the hypotenuse
of a right triangle
is equal to
the sum of the squares
on the two adjacent sides.

The song goes on with some *double entendre* about not letting your participle dangle, and associates Einstein, Newton, and the Wright brothers with the famous theorem. The first two exclaim 'Eureka!'; no, that was Archimedes. You will deduce that the lyrics are not hot on historical accuracy, but that's Hollywood for you. However, in Chapter 13 we will see that the lyricist (Johnny Mercer) was spot on with Einstein, probably more so than he realised.

Pythagoras's theorem appears in a well-known joke, with terrible puns about the squaw on the hippopotamus. The joke can be found all over the internet, but it's much harder to discover where it came from.[1] There are Pythagoras cartoons, T-shirts, and a Greek stamp, Figure 1.

Fig 1 Greek stamp showing Pythagoras's theorem.

All this fuss notwithstanding, we have no idea whether Pythagoras actually *proved* his theorem. In fact, we don't know whether it was his theorem at all. It could well have been discovered by one of Pythagoras's minions, or some Babylonian or Sumerian scribe. But Pythagoras got the credit, and his name stuck. Whatever its origins, the theorem and its consequences have had a gigantic impact on human history. They literally opened up our world.

The Greeks did not express Pythagoras's theorem as an equation in the modern symbolic sense. That came later with the development of algebra. In ancient times, the theorem was expressed verbally and geometrically. It attained its most polished form, and its first recorded proof, in the writings of Euclid of Alexandria. Around 250 BC Euclid became the first modern mathematician when he wrote his famous *Elements*, the most influential mathematical textbook ever. Euclid turned geometry into logic by making his basic assumptions explicit and invoking them to give systematic proofs for all of his theorems. He built a conceptual tower whose foundations were points, lines, and circles, and whose pinnacle was the existence of precisely five regular solids.

One of the jewels in Euclid's crown was what we now call Pythagoras's theorem: Proposition 47 of Book I of the *Elements*. In the famous translation by Sir Thomas Heath this proposition reads: 'In right-angled triangles the square on the side subtending the right angle is equal to the squares on the sides containing the right angle.'

No hippopotamus, then. No hypotenuse. Not even an explicit 'sum' or 'add'. Just that funny word 'subtend', which basically means 'be opposite to'. However, Pythagoras's theorem clearly expresses an equation, because it contains that vital word: *equal*.

For the purposes of higher mathematics, the Greeks worked with lines and areas instead of numbers. So Pythagoras and his Greek successors would decode the theorem as an equality of areas: 'The area of a square constructed using the longest side of a right-angled triangle is the sum of the areas of the squares formed from the other two sides.' The longest side is the famous hypotenuse, which means 'to stretch under', which it does if you draw the diagram in the appropriate orientation, as in Figure 2 (left).

Within a mere 2000 years, Pythagoras's theorem had been recast as the algebraic equation

$$a^2 + b^2 = c^2$$

where c is the length of the hypotenuse, a and b are the lengths of the other two sides, and the small raised 2 means 'square'. Algebraically, the square of any number is that number multiplied by itself, and we all know that the area of any square is the square of the length of its side. So Pythagoras's equation, as I shall rename it, says the same thing that Euclid said – except for various psychological baggage to do with how the ancients thought about basic mathematical concepts like numbers and areas, which I won't go into.

Pythagoras's equation has many uses and implications. Most directly, it lets you calculate the length of the hypotenuse, given the other two sides. For instance, suppose that $a = 3$ and $b = 4$. Then $c^2 = a^2 + b^2 = 3^2 + 4^2 = 9 + 16 = 25$. Therefore $c = 5$. This is the famous 3–4–5 triangle, ubiquitous in school mathematics, and the simplest example of a Pythagorean triple: a list of three whole numbers that satisfies Pythagoras's equation. The next simplest, other than scaled versions such as 6–8–10, is the 5–12–13 triangle. There are infinitely many such triples, and the Greeks knew how to construct them all. They still retain some interest in number theory, and even in the last decade new features have been discovered.

Instead of using a and b to work out c, you can proceed indirectly, and solve the equation to obtain a provided that you know b and c. You can also answer more subtle questions, as we will shortly see.

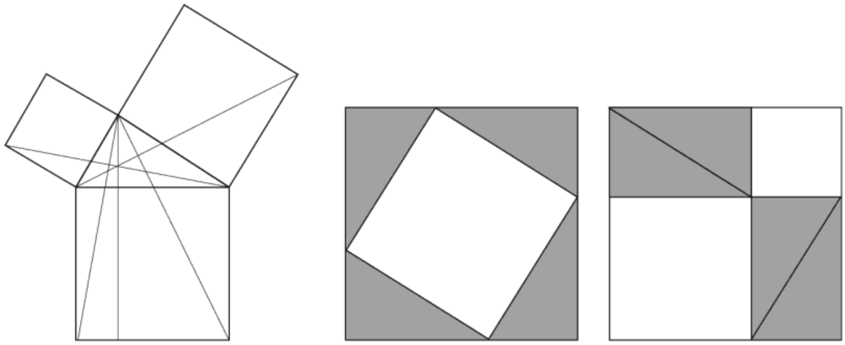


Fig 2 *Left*: Construction lines for Euclid's proof of Pythagoras. *Middle* and *right*: Alternative proof of the theorem. The outer squares have equal areas, and the shaded triangles all have equal areas. Therefore the tilted white square has the same area as the other two white squares combined.

Why is the theorem true? Euclid's proof is quite complicated, and it involves drawing five extra lines on the diagram, Figure 2 (left), and appealing to several previously proved theorems. Victorian schoolboys (few girls did geometry in those days) referred to it irreverently as Pythagoras's pants. A straightforward and intuitive proof, though not the most elegant, uses four copies of the triangle to relate two solutions of the same mathematical jigsaw puzzle, Figure 2 (right). The picture is compelling, but filling in the logical details requires some thought. For instance: how do we know that the tilted white region in the middle picture is a square?

There is tantalising evidence that Pythagoras's theorem was known long before Pythagoras. A Babylonian clay tablet[2] in the British Museum contains, in cuneiform script, a mathematical problem and answer that can be paraphrased as:

4 is the length and 5 the diagonal. What is the breadth?
4 times 4 is 16.
5 times 5 is 25.
Take 16 from 25 to obtain 9.
What times what must I take to get 9?
3 times 3 is 9.
Therefore 3 is the breadth.

So the Babylonians certainly knew about the 3–4–5 triangle, a thousand years before Pythagoras.

Another tablet, YBC 7289 from the Babylonian collection of Yale University, is shown in Figure 3 (*left*). It shows a diagram of a square of side 30, whose diagonal is marked with two lists of numbers: 1, 24, 51, 10 and 42, 25, 35. The Babylonians employed base-60 notation for numbers, so the first list actually refers to $1+24/60+51/60^2+10/60^3$, which in decimals is 1.4142129. The square root of 2 is 1.4142135. The second list is 30 times this. So the Babylonians knew that the diagonal of a square is its side multiplied by the square root of 2. Since $1^2+1^2=2=(\sqrt{2})^2$, this too is an instance of Pythagoras's theorem.

Fig 3 *Left*: YBC 7289. *Right*: Plimpton 322.

Even more remarkable, though more enigmatic, is the tablet Plimpton 322 from George Arthur Plimpton's collection at Columbia University, Figure 3 (right). It is a table of numbers, with four columns and 15 rows. The final column just lists the row number, from 1 to 15. In 1945 historians of science Otto Neugebauer and Abraham Sachs[3] noticed that in each row, the square of the number (say c) in the third column, minus the square of the number (say b) in the second column, is itself a square (say a). It follows that $a^2+b^2=c^2$, so the table appears to record Pythagorean triples. At least, this is the case provided four apparent errors are corrected. However, it is not absolutely certain that Plimpton 322 has anything to do with Pythagorean triples, and even if it does, it might just have been a convenient list of triangles whose areas were easy to calculate. These could then be assembled to yield good approximations to other triangles and other shapes, perhaps for land measurement.

Another iconic ancient civilisation is that of Egypt. There is some evidence that Pythagoras may have visited Egypt as a young man, and some have conjectured that this is where he learned his theorem. The surviving records of Egyptian mathematics offer scant support for this idea, but they are few and specialised. It is often stated, typically in the context of pyramids, that the Egyptians laid out right angles using a 3–4–5 triangle, formed from a length of string with knots at 12 equal intervals, and that archaeologists have found strings of that kind. However, neither claim makes much sense. Such a technique would not be very reliable, because strings can stretch and the knots would have to be very accurately spaced. The precision with which the pyramids at Giza are built is superior to anything that could be achieved with such a string. Far more practical tools, similar to a carpenter's set square, have been found. Egyptologists specialising in ancient Egyptian mathematics know of no records of string being employed to form a 3–4–5 triangle, and no examples of such strings exist. So this story, charming though it may be, is almost certainly a myth.

If Pythagoras could be transplanted into today's world he would notice many differences. In his day, medical knowledge was rudimentary, lighting came from candles and burning torches, and the fastest forms of communication were a messenger on horseback or a lighted beacon on a hilltop. The known world encompassed much of Europe, Asia, and Africa – but not the Americas, Australia, the Arctic, or the Antarctic. Many cultures considered the world to be flat: a circular disc or even a square aligned with the four cardinal points. Despite the discoveries of classical Greece this belief was still widespread in medieval times, in the form of *orbis terrae* maps, Figure 4.

Who first realised the world is round? According to Diogenes Laertius, a third-century Greek biographer, it was Pythagoras. In his book *Lives and Opinions of Eminent Philosophers*, a collection of sayings and biographical notes that is one of our main historical sources for the private lives of the philosophers of ancient Greece, he wrote: 'Pythagoras was the first who called the Earth round, though Theophrastus attributes this to Parmenides and Zeno to Hesiod.' The ancient Greeks often claimed that major discoveries had been made by their famous forebears, irrespective of historical fact, so we can't take the statement at face value, but it is not in dispute that from the fifth century BC all reputable Greek philosophers and mathematicians considered the Earth to be round. The idea does seem to have originated around the time of Pythagoras, and it might have come from one of his followers. Or it might have been common currency, based

Fig 4 Map of the world made around 1100 by the Moroccan cartographer al-Idrisi for King Roger of Sicily.

on evidence such as the round shadow of the Earth on the Moon during an eclipse, or the analogy with an obviously round Moon.

Even for the Greeks, though, the Earth was the centre of the universe and everything else revolved around it. Navigation was carried out by dead reckoning: looking at the stars and following the coastline. Pythagoras's equation changed all that. It set humanity on the path to today's understanding of the geography of our planet and its place in the Solar System. It was a vital first step towards the geometric techniques needed for mapmaking, navigation, and surveying. It also provided the key to a vitally important relation between geometry and algebra. This line of development leads from ancient times right through to general relativity and modern cosmology, see Chapter 13. Pythagoras's equation opened up entirely new directions for human exploration, both metaphorically and literally. It revealed the shape of our world and its place in the universe.

Many of the triangles encountered in real life are not right-angled, so the equation's direct applications may seem limited. However, any triangle can be cut into two right-angled ones as in Figure 6 (page 11), and any polygonal shape can be cut into triangles. So right-angled triangles are the key: they prove that there is a useful relation between the shape of a triangle and the lengths of its sides. The subject that developed from this insight is trigonometry: 'triangle measurement'.

The right-angled triangle is fundamental to trigonometry, and in particular it determines the basic trigonometric functions: sine, cosine, and tangent. The names are Arabic in origin, and the history of these functions and their many predecessors shows the complicated route by which today's version of the topic arose. I'll cut to the chase and explain the eventual outcome. A right-angled triangle has, of course, a right angle, but its other two angles are arbitrary, apart from adding to 90°. Associated with any angle are three functions, that is, rules for calculating an associated number. For the angle marked A in Figure 5, using the traditional a, b, c for the three sides, we define the sine (sin), cosine (cos), and tangent (tan) like this:

$$\sin A = a/c \quad \cos A = b/c \quad \tan A = a/b$$

These quantities depend only on the angle A, because all right-angled triangles with a given angle A are identical except for scale.

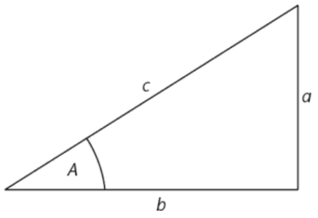


Fig 5 Trigonometry is based on a right-angle triangle.

In consequence, it is possible to draw up a table of the values of sin, cos, and tan, for a range of angles, and then use them to calculate features of right-angled triangles. A typical application, which goes back to ancient times, is to calculate the height of a tall column using only measurements made on the ground. Suppose that, from a distance of 100 metres, the angle to the top of the column is 22°. Take $A = 22°$ in Figure 5, so that a is the height of the column. Then the definition of the tangent function tells us that

$$\tan 22° = a/100$$

so that

$$a = 100 \tan 22°.$$

Since tan 22° is 0.404, to three decimal places, we deduce that $a = 40.4$ metres.

Profile Books - Seventeen Equations Data Standards Ltd, Frome, Somerset – 5/12/2011
01Seventeen_Chap1.3d Page 10 of 20

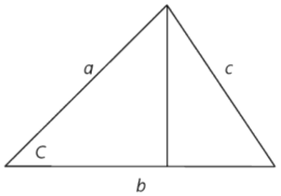


Fig 6 Splitting a triangle into two with right angles.

Once in possession of trigonometric functions, it is straightforward to extend Pythagoras's equation to triangles that do not have a right angle. Figure 6 shows a triangle with an angle *C* and sides *a*, *b*, *c*. Split the triangle into two right-angled ones as shown. Then two applications of Pythagoras and some algebra[4] prove that

$$a^2 + b^2 - 2ab\cos C = c^2$$

which is similar to Pythagoras's equation, except for the extra term $-2ab \cos C$. This 'cosine rule' does the same job as Pythagoras, relating *c* to *a* and *b*, but now we have to include information about the angle *C*.

The cosine rule is one of the mainstays of trigonometry. If we know two sides of a triangle and the angle between them, we can use it to calculate the third side. Other equations then tell us the remaining angles. All of these equations can ultimately be traced back to right-angled triangles.

Armed with trigonometric equations and suitable measuring apparatus, we can carry out surveys and make accurate maps. This is not a new idea. It appears in the Rhind Papyrus, a collection of ancient Egyptian mathematical techniques dating from 1650 BC. The Greek philosopher Thales used the geometry of triangles to estimate the heights of the Giza pyramids in about 600 BC. Hero of Alexandria described the same technique in 50 AD. Around 240 BC Greek mathematician, Eratosthenes, calculated the size of the Earth by observing the angle of the Sun at noon in two different places: Alexandria and Syene (now Aswan) in Egypt. A succession of Arabian scholars preserved and developed these methods, applying them in particular to astronomical measurements such as the size of the Earth.

Surveying began to take off in 1533 when the Dutch mapmaker Gemma Frisius explained how to use trigonometry to produce accurate maps, in *Libellus de Locorum Describendorum Ratione* ('Booklet Concerning a

Way of Describing Places'). Word of the method spread across Europe, reaching the ears of the Danish nobleman and astronomer Tycho Brahe. In 1579 Tycho used it to make an accurate map of Hven, the island where his observatory was located. By 1615 the Dutch mathematician Willebrord Snellius (Snel van Royen) had developed the method into essentially its modern form: *triangulation*. The area being surveyed is covered with a network of triangles. By measuring one initial length very carefully, and many angles, the locations of the corners of the triangle, and hence any interesting features within them, can be calculated. Snellius worked out the distance between two Dutch towns, Alkmaar and Bergen op Zoom, using a network of 33 triangles. He chose these towns because they lay on the same line of longitude and were exactly one degree of arc apart. Knowing the distance between them, he could work out the size of the Earth, which he published in his *Eratosthenes Batavus* ('The Dutch Eratosthenes') in 1617. His result is accurate to within 4%. He also modified the equations of trigonometry to reflect the spherical nature of the Earth's surface, an important step towards effective navigation.

Triangulation is an indirect method for calculating distances using angles. When surveying a stretch of land, be it a building site or a country, the main practical consideration is that it is much easier to measure angles than it is to measure distances. Triangulation lets us measure a few distances and lots of angles; then everything else follows from the trigonometric equations. The method begins by setting out one line between two points, called the baseline, and measuring its length directly to very high accuracy. Then we choose a prominent point in the landscape that is visible from both ends of the baseline, and measure the angle from the baseline to that point, at both ends of the baseline. Now we have a triangle, and we know one side of it and two angles, which fix its shape and size. We can then use trigonometry to work out the other two sides.

In effect, we now have two more baselines: the newly calculated sides of the triangle. From those, we can measure angles to other, more distant points. Continue this process to create a network of triangles that covers the area being surveyed. Within each triangle, observe the angles to all noteworthy features – church towers, crossroads, and so on. The same trigonometric trick pinpoints their precise locations. As a final twist, the accuracy of the entire survey can be checked by measuring one of the final sides directly.

By the late eighteenth century, triangulation was being employed routinely in surveys. The Ordnance Survey of Great Britain began in 1783, taking 70 years to complete the task. The Great Trigonometric Survey of

Profile Books - Seventeen Equations Data Standards Ltd, Frome, Somerset – 5/12/2011
01Seventeen_Chap1.3d Page 12 of 20

India, which among other things mapped the Himalayas and determined the height of Mount Everest, started in 1801. In the twenty-first century, most large-scale surveying is done using satellite photographs and GPS (the Global Positioning System). Explicit triangulation is no longer employed. But it is still there, behind the scenes, in the methods used to deduce locations from the satellite data.

Pythagoras's theorem was also vital to the invention of coordinate geometry. This is a way to represent geometric figures in terms of numbers, using a system of lines, known as axes, labelled with numbers. The most familiar version is known as Cartesian coordinates in the plane, in honour of the French mathematician and philosopher René Descartes, who was one of the great pioneers in this area – though not the first. Draw two lines: a horizontal one labelled x and a vertical one labelled y. These lines are known as axes (plural of axis), and they cross at a point called the origin. Mark points along these two axes according to their distance from the origin, like the markings on a ruler: positive numbers to the right and up, negative to the left and down. Now we can determine any point in the plane in terms of two numbers x and y, its coordinates, by connecting the point to the two axes as in Figure 7. The pair of numbers (x, y) completely specifies the location of the point.

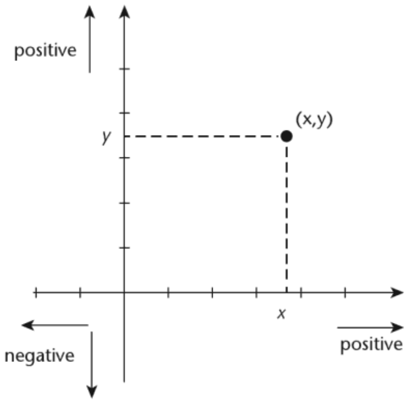


Fig 7 The two axes and the coordinates of a point.

The great mathematicians of seventeenth-century Europe realised that, in this context, a line or curve in the plane corresponds to the set of solutions (x, y) of some equation in x and y. For instance, $y = x$ determines a diagonal line sloping from lower left to top right, because (x, y) lies on that line if and only if $y = x$. In general, a linear equation – of the form $ax + by = c$ for constants a, b, c – corresponds to a straight line, and vice versa.

What equation corresponds to a circle? This is where Pythagoras's equation comes in. It implies that the distance r from the origin to the point (x, y) satisfies

$$r^2 = x^2 + y^2$$

and we can solve this for r to obtain

$$r = \sqrt{x^2 + y^2}$$

Since the set of all points that lie at distance r from the origin is a circle of radius r, whose centre is the origin, so the same equation defines a circle. More generally, the circle of radius r with centre at (a, b) corresponds to the equation

$$(x - a)^2 + (y - b)^2 = r^2$$

and the same equation determines the distance r between the two points (a, b) and (x, y). So Pythagoras's theorem tells us two vital things: which equations yield circles, and how to calculate distances from coordinates.

Pythagoras's theorem, then, is important in its own right, but it exerts even more influence through its generalisations. Here I will pursue just one strand of these later developments to bring out the connection with relativity, to which we return in Chapter 13.

The proof of Pythagoras's theorem in Euclid's *Elements* places the theorem firmly within the realm of Euclidean geometry. There was a time when that phrase could have been replaced by just 'geometry', because it was generally assumed that Euclid's geometry was the true geometry of physical space. It was obvious. Like most things assumed to be obvious, it turned out to be false.

Euclid derived all of his theorems from a small number of basic assumptions, which he classified as definitions, axioms, and common notions. His set-up was elegant, intuitive, and concise, with one glaring exception, his fifth axiom: 'If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.' It's a bit of a mouthful: Figure 8 may help.

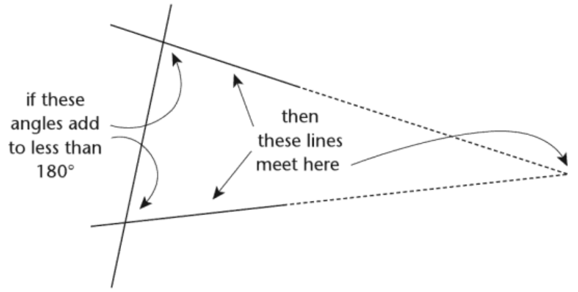


Fig 8 Euclid's parallel axiom.

For well over a thousand years, mathematicians tried to repair what they saw as a flaw. They weren't just looking for something simpler and more intuitive that would achieve the same end, although several of them found such things. They wanted to get rid of the awkward axiom altogether, by proving it. After several centuries, mathematicians finally realised that there were alternative 'non-Euclidean' geometries, implying that no such proof existed. These new geometries were just as logically consistent as Euclid's, and they obeyed all of his axioms except the parallel axiom. They could be interpreted as the geometry of geodesics – shortest paths – on curved surfaces, Figure 9. This focused attention on the meaning of curvature.

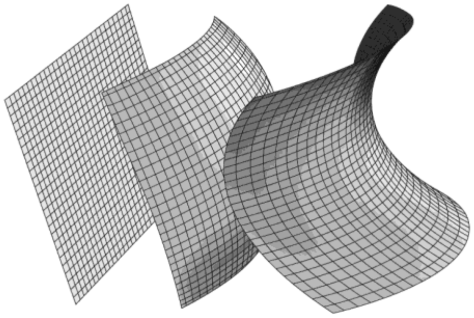


Fig 9 Curvature of a surface. *Left*: zero curvature. *Middle*: positive curvature. *Right*: negative curvature.

The plane of Euclid is flat, curvature zero. A sphere has the same curvature everywhere, and it is positive: near any point it looks like a dome. (As a technical fine point: great circles meet in two points, not one as Euclid's axioms require, so spherical geometry is modified by identifying antipodal points on the sphere – considering them to be identical. The surface becomes a so-called projective plane and the geometry is called elliptic.) A surface of constant negative curvature also exists: near any point, it looks like a saddle. This surface is called the hyperbolic plane, and it can be represented in several entirely prosaic ways. Perhaps the simplest is to consider it as the interior of a circular disc, and to define 'line' as an arc of a circle meeting the edge of the disc at right angles (Figure 10).

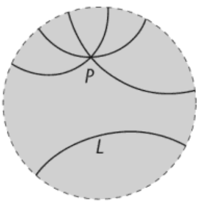


Fig 10 Disc model of the hyperbolic plane. All three lines through *P* fail to meet line *L*.

It might seem that, while plane geometry might be non-Euclidean, this must be impossible for the geometry of space. You can bend a surface by pushing it into a third dimension, but you can't bend *space* because there's no room for an extra dimension along which to push it. However, this is a rather naive view. For example, we can model three-dimensional hyperbolic space using the interior of a sphere. Lines are modelled as arcs of circles that meet the boundary at right angles, and planes are modelled as parts of spheres that meet the boundary at right angles. This geometry is three-dimensional, satisfies all of Euclid's axioms except the Fifth, and in a sense that can be pinned down it defines a curved three-dimensional space. But it's not curved round anything, or in any new direction.

It's just curved.

With all these new geometries available, a new point of view began to occupy centre stage – but as physics, not mathematics. Since space doesn't *have* to be Euclidean, what shape *is* it? Scientists realised that they didn't actually know. In 1813, Gauss, knowing that in a curved space the angles of a triangle do not add to 180°, measured the angles of a triangle formed by three mountains – the Brocken, the Hohehagen, and the Inselberg. He obtained a sum 15 seconds of arc greater than 180°. If correct, this indicated that space (in that region, at least) was positively curved. But you'd need a far larger triangle, and far more accurate measurements, to eliminate observational errors. So Gauss's observations were inconclusive. Space might be Euclidean, and then again, it might not be.

My remark that three-dimensional hyperbolic space is 'just curved' depends on a new point of view about curvature, which also goes back to Gauss. The sphere has constant positive curvature, and the hyperbolic plane has constant negative curvature. But the curvature of a surface doesn't have to be constant. It might be sharply curved in some places, less sharply curved in others. Indeed, it might be positive in some regions but negative in others. The curvature could vary continuously from place to place. If the surface looks like a dog's bone, then the blobs at the ends are positively curved but the part that joins them is negatively curved.

Gauss searched for a formula to characterise the curvature of a surface at any point. When he eventually found it, and published it in his *Disquisitiones Generales Circa Superficies Curva* ('General Research on Curved Surfaces') of 1828, he named it the 'remarkable theorem'. What was so remarkable? Gauss had started from the naive view of curvature: embed the surface in three-dimensional space and calculate how bent it is. But the answer told him that this surrounding space didn't matter. It didn't enter into the formula. He wrote: 'The formula ... leads itself to the remarkable theorem: If a curved surface is developed upon any other surface whatever, the measure of curvature in each point remains unchanged.' By 'developed' he meant 'wrapped round'.

Take a flat sheet of paper, zero curvature. Now wrap it round a bottle. If the bottle is cylindrical the paper fits perfectly, without being folded, stretched, or torn. It is bent as far as visual appearance goes, but it's a trivial kind of bending, because it hasn't changed geometry on the paper in any way. It's just changed how the paper relates to the surrounding space. Draw a right-angled triangle on the flat paper, measure its sides, check Pythagoras. Now wrap the diagram round a bottle. The lengths of sides, *measured along the paper*, don't change. Pythagoras is still true.

The surface of a sphere, however, has nonzero curvature. So it is not possible to wrap a sheet of paper so that it fits snugly against a sphere, without folding it, stretching it, or tearing it. Geometry on a sphere is intrinsically different from geometry on a plane. For example, the Earth's equator and the lines of longitude for 0° and 90° to its north determine a triangle that has three right angles and three equal sides (assuming the Earth to be a sphere). So Pythagoras's equation is false.

Today we call curvature in its intrinsic sense 'Gaussian curvature'. Gauss explained why it is important using a vivid analogy, still current. Imagine an ant confined to the surface. How can it work out whether the surface is curved? It can't step outside the surface to see whether it looks bent. But it can use Gauss's formula by making suitable measurements purely within the surface. We are in the same position as the ant when we try to figure out the true geometry of our space. We can't step outside it. Before we can emulate the ant by taking measurements, however, we need a formula for the curvature of a space of three dimensions. Gauss didn't have one. But one of his students, in a fit of recklessness, claimed that he did.

The student was Georg Bernhard Riemann, and he was trying to achieve what German universities call Habilitation, the next step after a PhD. In Riemann's day this meant that you could charge students a fee for your lectures. Then and now, gaining Habilitation requires presenting your research in a public lecture that is also an examination. The candidate offers several topics, and the examiner, which in Riemann's case was Gauss, chooses one. Riemann, a brilliant mathematical talent, listed several orthodox topics that he knew backwards, but in a rush of blood to the brain he also suggested 'On the hypotheses which lie at the foundation of geometry'. Gauss had long been interested in just that, and he naturally selected it for Riemann's examination.

Riemann instantly regretted offering something so challenging. He had a hearty dislike of public speaking, and he hadn't thought the mathematics through in detail. He just had some vague, though fascinating, ideas about curved space. In *any* number of dimensions. What Gauss had done for two dimensions, with his remarkable theorem, Riemann wanted to do in as many dimensions as you like. Now he had to perform, and fast. The lecture was looming. The pressure nearly gave him a nervous breakdown, and his day job helping Gauss's collaborator Wilhelm Weber with experiments in electricity didn't help. Well, maybe it did, because while Riemann was thinking about the relation between electrical and magnetic forces in the day job, he realised that force can be related to curvature. Working backwards, he could use the mathematics of forces to define curvature, as required for his examination.

In 1854 Riemann delivered his lecture, which was warmly received, and no wonder. He began by defining what he called a 'manifold', in the sense of many-foldedness. Formally, a 'manifold', is specified by a system of many coordinates, together with a formula for the distance between nearby points, now called a Riemannian metric. Informally, a manifold is a multidimensional space in all its glory. The climax of Riemann's lecture was a formula that generalised Gauss's remarkable theorem: it defined the curvature of the manifold solely in terms of its metric. And it is here that the tale comes full circle like the snake Orobouros and swallows its own tail, because the metric contains visible remnants of Pythagoras.

Suppose, for example, that the manifold has three dimensions. Let the coordinates of a point be (x, y, z), and let $(x+\mathrm{d}x, y+\mathrm{d}y, z+\mathrm{d}z)$ be a nearby point, where the d means 'a little bit of'. If the space is Euclidean, with zero curvature, the distance $\mathrm{d}s$ between these two points satisfies the equation

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

and this is just Pythagoras, restricted to points that are close together. If the space is curved, with variable curvature from point to point, the analogous formula, the metric, looks like this:

$$\mathrm{d}s^2 = X\,\mathrm{d}x^2 + Y\,\mathrm{d}y^2 + Z\,\mathrm{d}z^2 + 2U\,\mathrm{d}x\,\mathrm{d}y + 2V\,\mathrm{d}x\,\mathrm{d}z + 2W\,\mathrm{d}y\,\mathrm{d}z$$

Here X, Y, Z, U, V, W can depend on x, y and z. It may seem a bit of a mouthful, but like Pythagoras's equation it involves sums of squares (and closely related products of two quantities like $\mathrm{d}x\,\mathrm{d}y$) plus a few bells and whistles. The 2s occur because the formula can be packaged as a 3×3 table, or matrix:

$$\begin{bmatrix} X & U & V \\ U & Y & W \\ V & W & Z \end{bmatrix}$$

where X, Y, Z appear once, but U, V, W appear twice. The table is symmetric about its diagonal; in the language of differential geometry it is a symmetric tensor. Riemann's generalisation of Gauss's remarkable theorem is a formula for the curvature of the manifold, at any given point, in terms of this tensor. In the special case when Pythagoras applies, the curvature turns out to be zero. So the validity of Pythagoras's equation is a test for the absence of curvature.

Like Gauss's formula, Riemann's expression for curvature depends only on the manifold's metric. An ant confined to the manifold could observe the metric by measuring tiny triangles and computing the curvature. Curvature is an intrinsic property of a manifold, independent of any surrounding space. Indeed, the metric already determines the geometry, so no surrounding space is required. In particular, we human ants can ask what shape our vast and mysterious universe is, and hope to answer it by making observations that do not require us to step outside the universe. Which is just as well, because we can't.

Riemann found his formula by using forces to define geometry. Fifty years later, Einstein turned Riemann's idea on its head, using geometry to define the force of gravity in his general theory of relativity, and inspiring new ideas about the shape of the universe: see Chapter 13. It's an astonishing progression of events. Pythagoras's equation first came into being around 3500 years ago to measure a farmer's land. Its extension to triangles without right angles, and triangles on a sphere, allowed us to map our continents and measure our planet. And a remarkable generalisation lets us measure the shape of the universe. Big ideas have small beginnings.

2 Shortening the proceedings

Logarithms

multiply   add

$$\log xy = \log x + \log y$$

logarithm

What does it tell us?

How to multiply numbers by adding related numbers instead.

Why is that important?

Addition is much simpler than multiplication.

What did it lead to?

Efficient methods for calculating astronomical phenomena such as eclipses and planetary orbits. Quick ways to perform scientific calculations. The engineers' faithful companion, the slide rule. Radioactive decay and the psychophysics of human perception.

Numbers originated in practical problems: recording property, such as animals or land, and financial transactions, such as taxation and keeping accounts. The earliest known number notation, aside from simple tallying marks like ||||, is found on the outside of clay envelopes. In 8000 BC Mesopotamian accountants kept records using small clay tokens of various shapes. The archaeologist Denise Schmandt-Besserat realised that each shape represented a basic commodity – a sphere for grain, an egg for a jar of oil, and so on. For security, the tokens were sealed in clay wrappings. But it was a nuisance to break a clay envelope open to find out how many tokens were inside, so the ancient accountants scratched symbols on the outside to show what was inside. Eventually they realised that once you had these symbols, you could scrap the tokens. The result was a series of written symbols for numbers – the origin of all later number symbols, and perhaps of writing too.

Along with numbers came arithmetic: methods for adding, subtracting, multiplying, and dividing numbers. Devices like the abacus were used to do the sums; then the results could be recorded in symbols. After a time, ways were found to use the symbols to perform the calculations without mechanical assistance, although the abacus is still widely used in many parts of the world, while electronic calculators have supplanted pen and paper calculations in most other countries.

Arithmetic proved essential in other ways, too, especially in astronomy and surveying. As the basic outlines of the physical sciences began to emerge, the fledgeling scientists needed to perform ever more elaborate calculations, by hand. Often this took up much of their time, sometimes months or years, getting in the way of more creative activities. Eventually it became essential to speed up the process. Innumerable mechanical devices were invented, but the most important breakthrough was a conceptual one: think first, calculate later. Using clever mathematics, you could make difficult calculations much easier.

The new mathematics quickly developed a life of its own, turning out to have deep theoretical implications as well as practical ones. Today, those early ideas have become an indispensable tool throughout science, But in Napier's day it all had to be done by hand. Wouldn't it be great if there were some mathematical trick that would convert those nasty multiplications into nice, quick addition sums? It sounds too good to be true, but Napier realised that it was possible. The trick was to work with powers of a fixed number.

In algebra, powers of an unknown x are indicated by a small raised number. That is, $xx = x^2$, $xxx = x^3$, $xxxx = x^4$, and so on, where as usual in algebra placing two letters next to each other means you should multiply them together. So, for instance, $10^4 = 10 \times 10 \times 10 \times 10 = 10{,}000$. You don't need to play around with such expressions for long before you discover an easy way to work out, say, $10^4 \times 10^3$. Just write down

$$\begin{aligned} 10{,}000 \times 1{,}000 &= (10 \times 10 \times 10 \times 10) \times (10 \times 10 \times 10) \\ &= 10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10 \\ &= 10{,}000{,}000 \end{aligned}$$

The number of 0s in the answer is 7, which equals $4 + 3$. The first step in the calculation shows *why* it is $4 + 3$: we stick four 10s and three 10s next to each other. In short,

$$10^4 \times 10^3 = 10^{4+3} = 10^7$$

In the same way, whatever the value of x might be, if we multiply its ath power by its bth power, where a and b are whole numbers, then we get the $(a + b)$th power:

$$x^a x^b = x^{a+b}$$

This may seem an innocuous formula, but on the left it multiplies two quantities together, while on the right the main step is to add a and b, which is simpler.

Suppose you wanted to multiply, say, 2.67 by 3.51. By long multiplication you get 9.3717, which to two decimal places is 9.37. What if you try to use the previous formula? The trick lies in the choice of x. If we take x to be 1.001, then a bit of arithmetic reveals that

$$\begin{aligned} (1.001)^{983} &= 2.67 \\ (1.001)^{1256} &= 3.51 \end{aligned}$$

correct to two decimal places. The formula then tells us that 2.87×3.41 is

$$(1.001)^{983+1256} = (1.001)^{2239}$$

which, to two decimal places, is 9.37.

The core of the calculation is an easy addition: $983 + 1256 = 2239$. However, if you try to check my arithmetic you will quickly realise that if anything I've made the problem harder, not easier. To work out $(1.001)^{983}$ you have to multiply 1.001 by itself 983 times. And to discover that 983 is the right power to use, you have to do even more work. So at first sight this seems like a pretty useless idea.

Napier's great insight was that this objection is wrong. But to overcome it, some hardy soul has to calculate lots of powers of 1.001, starting with $(1.001)^2$ and going up to something like $(1.001)^{10,000}$. Then they can publish a table of all these powers. After that, most of the work has been done. You just have to run your fingers down the successive powers until you see 2.67 next to 983; you similarly locate 3.51 next to 1256. Then you add those two numbers to get 2239. The corresponding row of the table tells you that this power of 1.001 is 9.37. Job done.

Really accurate results require powers of something a lot closer to 1, such as 1.000001. This makes the table far bigger, with a million or so powers. Doing the calculations for that table is a huge enterprise. *But it has to be done only once*. If some self-sacrificing benefactor makes the effort up front, succeeding generations will be saved a gigantic amount of arithmetic.

In the context of this example, we can say that the powers 983 and 1256 are the *logarithms* of the numbers 2.67 and 3.51 that we wish to multiply. Similarly 2239 is the logarithm of their product 9.38. Writing log as an abbreviation, what we have done amounts to the equation

$$\log ab = \log a + \log b$$

which is valid for any numbers a and b. The rather arbitrary choice of 1.001 is called the *base*. If we use a different base, the logarithms that we calculate are also different, but for any fixed base everything works the same way.

This is what Napier should have done. But for reasons that we can only guess at, he did something slightly different. Briggs, approaching the technique from a fresh perspective, spotted two ways to improve on Napier's idea.

When Napier started thinking about powers of numbers, in the late sixteenth century, the idea of reducing multiplication to addition was already circulating among mathematicians. A rather complicated method known as 'prosthapheiresis', based on a formula involving trigonometric functions, was in use in Denmark.[3] Napier, intrigued, was smart enough to realise that powers of a fixed number could do the same job more simply. The necessary tables didn't exist – but that was easily remedied. Some public-spirited soul must carry out the work. Napier volunteered himself for the task, but he made a strategic error. Instead of using a base that was slightly bigger than 1, he used a base slightly smaller than 1. In consequence, the sequence of powers started out with big numbers, which got successively smaller. This made the calculations slightly more clumsy.

Briggs spotted this problem, and saw how to deal with it: use a base slightly larger than 1. He also spotted a subtler problem, and dealt with that as well. If Napier's method were modified to work with powers of something like 1.0000000001, there would be no straightforward relation between the logarithms of, say, 12.3456 and 1.23456. So it wasn't entirely clear when the table could *stop*. The source of the problem was the value of log 10, because

$$\log 10x = \log 10 + \log x$$

Unfortunately log 10 was messy: with the base 1.0000000001 the logarithm of 10 was 23,025,850,929. Briggs thought it would be much nicer if the base could be chosen so that $\log 10 = 1$. Then $\log 10x = 1 + \log x$, so that whatever log 1.23456 might be, you just had to add 1 to it to get log 12.3456. Now tables of logarithms need only run from 1 to 10. If bigger numbers turned up, you just added the appropriate whole number.

To make $\log 10 = 1$, you do what Napier did, using a base of 1.0000000001, but then you divide every logarithm by that curious number 23,025,850,929. The resulting table consists of logarithms to base 10, which I'll write as $\log_{10} x$. They satisfy

$$\log_{10} xy = \log_{10} x + \log_{10} y$$

as before, but also

$$\log_{10} 10x = \log_{10} x + 1$$

Within two years Napier was dead, so Briggs started work on a table of base-10 logarithms. In 1617 he published *Logarithmorum Chilias Prima* ('Logarithms of the First Chiliad'), the logarithms of the integers from 1 to 1000 accurate to 14 decimal places. In 1624 he followed it up with *Arithmetic Logarithmica* ('Arithmetic of Logarithms'), a table of base-10 logarithms of numbers from 1 to 20,000 and from 90,000 to 100,000, to the same accuracy. Others rapidly followed Briggs's lead, filling in the large gap and developing auxiliary tables such as logarithms of trigonometric functions like log sin x.

The same ideas that inspired logarithms allow us to define powers x^a of a positive variable x for values of a that are not positive whole numbers. All we have to do is insist that our definitions must be consistent with the equation $x^a x^b = x^{a+b}$, and follow our noses. To avoid nasty complications, it is best to assume x is positive, and to define x^a so that this is also positive. (For negative x, it's best to introduce complex numbers, as in Chapter 5.)

For example, what is x^0? Bearing in mind that $x^1 = x$, the formula says that x^0 must satisfy $x^0 x = x^{0+1} = x$. Dividing by x we find that $x^0 = 1$. Now what about x^{-1}? Well, the formula says that $x^{-1} x = x^{-1+1} = x^0 = 1$. Dividing by x, we get $x^{-1} = 1/x$. Similarly $x^{-2} = 1/x^2$, $x^{-3} = 1/x^3$, and so on.

It starts to get more interesting, and potentially very useful, when we think about $x^{1/2}$. This has to satisfy $x^{1/2}\ x^{1/2} = x^{1/2+1/2} = x^1 = x$. So $x^{1/2}$, multiplied by itself, is x. The only number with this property is the square root of x. So $x^{1/2} = \sqrt{x}$. Similarly, $x^{1/3} = \sqrt[3]{x}$, the cube root. Continuing in this manner we can define $x^{p/q}$ for any fraction p/q. Then, using fractions to approximate real numbers, we can define x^a for any real a. And the equation $x^a x^b = x^{a+b}$ still holds.

It also follows that $\log \sqrt{x} = \frac{1}{2} \log x$ and $\log \sqrt[3]{x} = \frac{1}{3} \log x$, so we can calculate square roots and cube roots easily using a table of logarithms. For example, to find the square root of a number we form its logarithm, divide by 2, and then work out which number has the result as its logarithm. For cube roots, do the same but divide by 3. Traditional methods for these problems were tedious and complicated. You can see why Napier showcased square and cube roots in the preface to his book.

As soon as complete tables of logarithms were available, they became an indispensable tool for scientists, engineers, surveyors, and navigators. They saved time, they saved effort, and they increased the likelihood that the answer was correct. Early on, astronomy was a major beneficiary, because astronomers routinely needed to perform long and difficult calculations. The French mathematician and astronomer Pierre Simon de Laplace said that the invention of logarithms 'reduces to a few days the labour of many months, doubles the life of the astronomer, and spares him the errors and disgust'. As the use of machinery in manufacturing grew, engineers started to make more and more use of mathematics – to design complex gears, analyse the stability of bridges and buildings, and construct cars, lorries, ships, and aeroplanes. Logarithms were a firm part of the school mathematics curriculum a few decades ago. And engineers carried what was in effect an analogue calculator for logarithms in their pockets, a physical representation of the basic equation for logarithms for on-the-spot use. They called it a slide rule, and they used it routinely in applications ranging from architecture to aircraft design.

The first slide rule was constructed by an English mathematician, William Oughtred, in 1630, using circular scales. He modified the design in 1632, by making the two rulers straight. This was the first slide rule. The idea is simple: when you place two rods end to end, their lengths add. If the rods are marked using a logarithmic scale, in which numbers are spaced according to their logarithms, then the corresponding numbers *multiply*. For instance, set the 1 on one rod against the 2 on the other. Then against any number x on the first rod, we find $2x$ on the second. So opposite 3 we find 6, and so on, see Figure 11. If the numbers are more complicated, say 2.67 and 3.51, we place 1 opposite 2.67 and read off whatever is opposite 3.59, namely 9.37. It's just as easy.

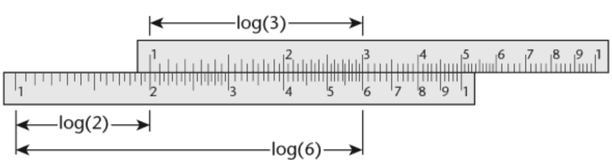


Fig 11 Multiplying 2 by 3 on a slide rule.

Engineers quickly developed fancy slide rules with trigonometric functions, square roots, log-log scales (logarithms of logarithms) to calculate powers, and so on. Eventually logarithms took a back seat to digital computers, but even now the logarithm still plays a huge role in science and technology, alongside its inseparable companion, the exponential function. For base-10 logarithms, this is the function 10^x; for natural logarithms, the function e^x, where $e = 2.71828$, approximately. In each pair, the two functions are inverse to each other. If you take a number, form its logarithm, and then form the exponential of that, you get back the number you started with.

Why do we need logarithms now that we have computers?

In 2011 a magnitude 9.0 earthquake just off the east coast of Japan

Radioactive decay is just one area of many in which Napier's and Briggs's logarithms continue to serve science and humanity. If you thumb through later chapters you will find them popping up in thermodynamics and information theory, for example. Even though fast computers have now made logarithms redundant for their original purpose, rapid calculations, they remain central to science for conceptual rather than computational reasons.

Another application of logarithms comes from studies of human perception: how we sense the world around us. The early pioneers of the psychophysics of perception made extensive studies of vision, hearing, and touch, and they turned up some intriguing mathematical regularities.

In the 1840s a German doctor, Ernst Weber, carried out experiments to determine how sensitive human perception is. His subjects were given weights to hold in their hands, and asked when they could tell that one weight felt heavier than another. Weber could then work out what the smallest detectable difference in weight was. Perhaps surprisingly, this difference (for a given experimental subject) was not a fixed amount. It depended on how heavy the weights being compared were. People didn't sense an absolute minimum difference – 50 grams, say. They sensed a *relative* minimum difference – 1% of the weights under comparison, say. That is, the smallest difference that the human senses can detect is proportional to the stimulus, the actual physical quantity.

In the 1850s Gustav Fechner rediscovered the same law, and recast it mathematically. This led him to an equation, which he called Weber's law, but nowadays it is usually called Fechner's law (or the Weber–Fechner law if you're a purist). It states that the perceived sensation is proportional to the *logarithm* of the stimulus. Experiments suggested that this law applies not only to our sense of weight but to vision and hearing as well. If we look at a light, the brightness that we perceive varies as the logarithm of the actual energy output. If one source is ten times as bright as another, then the difference we perceive is constant, however bright the two sources really are. The same goes for the loudness of sounds: a bang with ten times as much energy sounds a fixed amount louder.

The Weber–Fechner law is not totally accurate, but it's a good approximation. Evolution pretty much had to come up with something like a logarithmic scale, because the external world presents our senses with stimuli over a huge range of sizes. A noise may be little more than a mouse scuttling in the hedgerow, or it may be a clap of thunder; we need to be able to hear both. But the range of sound levels is so vast that no biological sensory device can respond in proportion to the energy generated by the sound. If an ear that could hear the mouse did that, then a thunderclap would destroy it. If it tuned the sound levels down so that the thunderclap produced a comfortable signal, it wouldn't be able to hear the mouse. The solution is to compress the energy levels into a comfortable range, and the logarithm does exactly that. Being sensitive to proportions rather than absolutes makes excellent sense, and makes for excellent senses.

Our standard unit for noise, the decibel, encapsulates the Weber–Fechner law in a definition. It measures not absolute noise, but relative noise. A mouse in the grass produces about 10 decibels. Normal conversation between people a metre apart takes place at 40–60 decibels. An electric mixer directs about 60 decibels at the person using it. The noise in a car, caused by engine and tyres, is 60–80 decibels. A jet airliner a hundred metres away produces 110–140 decibels, rising to 150 at thirty metres. A vuvuzela (the annoying plastic trumpet-like instrument widely heard during the football World Cup in 2010 and brought home as souvenirs by misguided fans) generates 120 decibels at one metre; a military stun grenade produces up to 180 decibels.

Scales like these are widely encountered because they have a safety aspect. The level at which sound can potentially cause hearing damage is about 120 decibels. Please throw away your vuvuzela.

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
02Seventeen_Chap2.3d Page 34 of 34

3 Ghosts of departed quantities
Calculus

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}$$

change in the value of quantity
rate of change of quantity
limit
new value
minus
old value
with respect to
time
time interval
tends to zero
(becomes very small)
divided by
time interval

What does it say?

To find the instantaneous rate of change of a quantity that varies with (say) time, calculate how its value changes over a short time interval and divide by the time concerned. Then let that interval become arbitrarily small.

Why is that important?

It provides a rigorous basis for calculus, the main way scientists model the natural world.

What did it lead to?

Calculation of tangents and areas. Formulas for volumes of solids and lengths of curves. Newton's laws of motion, differential equations. The laws of conservation of energy and momentum. Most of mathematical physics.

In 1665 Charles II was king of England and his capital city, London, was a sprawling metropolis of half a million people. The arts flourished, and science was in the early stages of an ever-accelerating ascendancy. The Royal Society, perhaps the oldest scientific society now in existence, had been founded five years earlier, and Charles had granted it a royal charter. The rich lived in impressive houses, and commerce was thriving, but the poor were crammed into narrow streets overshadowed by ramshackle buildings that jutted out ever further as they rose, storey by storey. Sanitation was inadequate; rats and other vermin were everywhere. By the end of 1666, one fifth of London's population had been killed by bubonic plague, spread first by rats and then by people. It was the worst disaster in the capital's history, and the same tragedy played out all over Europe and North Africa. The king departed in haste for the more sanitary countryside of Oxfordshire, returning early in 1666. No one knew what caused plague, and the city authorities tried everything – burning fires continually to cleanse the air, burning anything that gave off a strong smell, burying the dead quickly in pits. They killed many dogs and cats, which ironically removed two controls on the rat population.

During those two years, an obscure and unassuming undergraduate at Trinity College, Cambridge, completed his studies. Hoping to avoid the plague, he returned to the house of his birth, from which his mother managed a farm. His father had died shortly before he was born, and he had been brought up by his maternal grandmother. Perhaps inspired by rural peace and quiet, or lacking anything better to do with his time, the young man thought about science and mathematics. Later he wrote: 'In those days I was in the prime of my life for invention, and minded mathematics and [natural] philosophy more than at any other time since.' His researches led him to understand the importance of the inverse square law of gravity, an idea that had been hanging around ineffectually for at least 50 years. He worked out a practical method for solving problems in calculus, another concept that was in the air but had not been formulated in any generality. And he discovered that white sunlight is composed of many different colours – all the colours of the rainbow.

increases; when we stamp on the brakes, the car decelerates – negative acceleration.

If the car is moving at a fixed speed, it's easy to work out what that speed is. The abbreviation mph gives it away: miles per hour. If the car travels 50 miles in 1 hour, we divide the distance by the time, and that's the speed. We don't need to drive for an hour: if the car goes 5 miles in 6 minutes, both distance and time are divided by 10, and their ratio is still 50 mph. In short,

speed = distance travelled divided by time taken.

In the same way, a fixed rate of acceleration is given by

acceleration = change in speed divided by time taken.

This all seems straightforward, but conceptual difficulties arise when the speed or acceleration is not fixed. And they can't both be constant, because constant (and nonzero) acceleration implies a changing speed. Suppose you drive along a country lane, speeding up on the straights, slowing for the corners. Your speed keeps changing, and so does your acceleration. How can we work them out at any given instant of time? The pragmatic answer is to take a short interval of time, say a second. Then your instantaneous speed at (say) 11.30 am is the distance you travel between that moment and one second later, divided by one second. The same goes for instantaneous acceleration.

Except … that's not quite your *instantaneous* speed. It's really an average speed, over a one-second interval of time. There are circumstances in which one second is a *huge* length of time – a guitar string playing middle C vibrates 440 times every second; average its motion over an entire second and you'll think it's standing still. The answer is to consider a shorter interval of time – one ten thousandth of a second, perhaps. But this still doesn't capture instantaneous speed. Visible light vibrates one quadrillion (10^{15}) times every second, so the appropriate time interval is less than one quadrillionth of a second. And even then … well, to be pedantic, that's still not an *instant*. Pursuing this line of thought, it seems to be necessary to use an interval of time that is shorter than any other interval. But the only number like that is 0, and that's useless, because now the distance travelled is also 0, and 0/0 is meaningless.

Early pioneers ignored these issues and took a pragmatic view. Once the probable error in your measurements exceeds the increased precision you would theoretically get by using smaller intervals of time, there's no point in doing so. The clocks in Galileo's day were very inaccurate, so he measured time by humming tunes to himself – a trained musician can subdivide a note into very short intervals. Even then, timing a falling body is tricky, so Galileo hit on the trick of slowing the motion down by rolling balls down an inclined slope. Then he observed the position of the ball at successive intervals of time. What he found (I'm simplifying the numbers to make the pattern clear, but it's the same pattern) is that for times 0, 1, 2, 3, 4, 5, 6, … these positions were

0 1 4 9 16 25 36

The distance was (proportional to) the square of the time. What about the speeds? Averaged over successive intervals, these were the differences

1 3 5 7 9 11

between the successive squares. In each interval, other than the first, the average speed increased by 2 units. It's a striking pattern – all the more so to Galileo when he dug something very similar out of dozens of measurements with balls of many different masses on slopes with many different inclinations.

From these experiments and the observed pattern, Galileo deduced something wonderful. The path of a falling body, or one thrown into the air, such as a cannonball, is a parabola. This is a U-shaped curve, known to the ancient Greeks. (The U is upside down in this case. I'm ignoring air resistance, which changes the shape: it didn't have much effect on Galileo's rolling balls.) Kepler encountered a related curve, the ellipse, in his analysis of planetary orbits: this must have seemed significant to Newton too, but that story must wait until the next chapter.

With only this particular series of experiments to go on, it's not clear what general principles underlie Galileo's pattern. Newton realised that the source of the pattern is rates of change. Velocity is the rate at which position changes with respect to time; acceleration is the rate at which velocity changes with respect to time. In Galileo's observations, position varied according to the square of time, velocity varied linearly, and acceleration didn't vary at all. Newton realised that in order to gain a deeper understanding of Galileo's patterns, and what they meant for our view of nature, he had to come to grips with instantaneous rates of change. When he did, out popped calculus.

You might expect an idea as important as calculus to be announced with a fanfare of trumpets and parades through the streets. However, it takes time

Profile Books - Seventeen Equations *Data Standards Ltd, Frome, Somerset – 5/12/2011*
03Seventeen_Chap3.3d Page 41 of 52

for the significance of novel ideas to sink in and to be appreciated, and so it was with calculus. Newton's work on the topic dates from 1671 or earlier, when he wrote *The Method of Fluxions and Infinite Series*. We are unsure of the date because the book was not published until 1736, nearly a decade after his death. Several other manuscripts by Newton also refer to ideas that we now recognise as differential and integral calculus, the two main branches of the subject. Leibniz's notebooks show that he obtained his first significant results in calculus in 1675, but he published nothing on the topic until 1684.

After Newton had risen to scientific prominence, long after both men had worked out the basics of calculus, some of Newton's friends sparked a largely pointless but heated controversy about priority, accusing Leibniz of plagiarising Newton's unpublished manuscripts. A few mathematicians from continental Europe responded with counter-claims of plagiarism by Newton. English and continental mathematicians were scarcely on speaking terms for a century, which caused huge damage to English mathematicians, but none whatsoever to the continental ones. They developed calculus into a central tool of mathematical physics while their English counterparts were seething about insults to Newton instead of exploiting insights from Newton. The story is tangled and still subject to scholarly disputation by historians of science, but broadly speaking it seems that Newton and Leibniz discovered the basic ideas of calculus independently – at least, as independently as their common mathematical and scientific culture permitted.

Leibniz's notation differs from Newton's, but the underlying ideas are more or less identical. The intuition behind them, however, is different. Leibniz's approach was a formal one, manipulating algebraic symbols. Newton had a physical model at the back of his mind, in which the function under consideration was a physical quantity that varies with time. This is where his curious term 'fluxion' comes from – something that flows as time passes.

Newton's method can be illustrated using an example: a quantity y that is the square x^2 of another quantity x. (This is the pattern that Galileo found for a rolling ball: its position is proportional to the square of the time that has elapsed. So there y would be position and x time. The usual symbol for time is t, but the standard coordinate system in the plane uses x and y.) Start by introducing a new quantity o, denoting a small change in x. The corresponding change in y is the difference

$$(x+o)^2 - x^2$$