THE MOST BEAUTIFUL
THEOREM IN MATHEMATICS

$e^{i\pi} + 1 = 0$

EULER'S PIONEERING EQUATION

ROBIN WILSON

EULER'S PIONEERING EQUATION

Leonhard Euler (1707–83), in an engraving by Bartholomaeus Hübner, 1786.

EULER'S piOneerinG EQUATION

The most beautiful theorem in mathematics

ROBIN WILSON

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2018
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2017948015

ISBN 978–0–19–879492–9

Printed in Great Britain by
Clays Ltd, St Ives plc

PREFACE

This book celebrates the appeal and beauty of mathematics, as illustrated by *Euler's equation*, 'the most beautiful theorem in mathematics', and its near-relative, *Euler's identity*. Named after the 18th-century Swiss mathematician Leonhard Euler, this famous equation connects five of the most important constants in mathematics:

$$1, 0, \pi, e, \text{ and } i.$$

These numbers appear in the Introduction and are then featured one at a time in the succeeding chapters before being brought together in the final one.

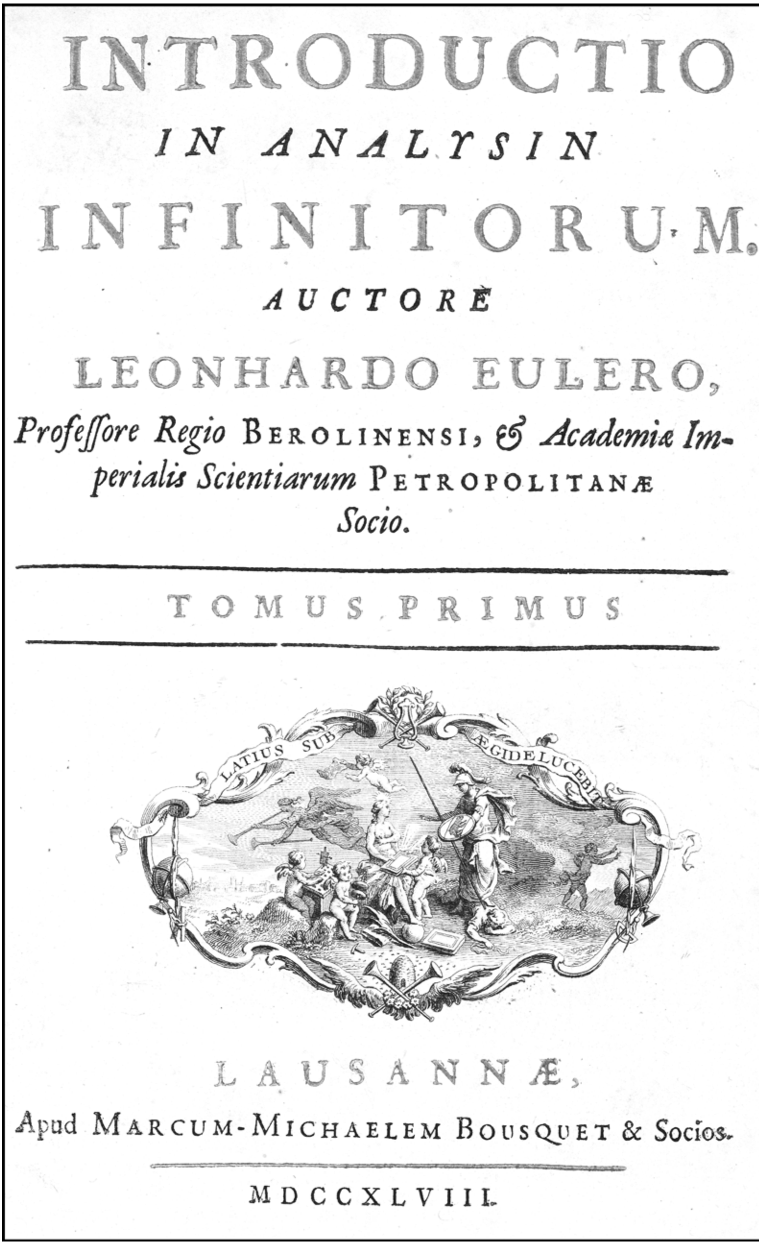
We've tried to keep the treatment as simple as possible throughout, with each topic introduced from the beginning and with many historical remarks to spice up the story. More advanced topics are included in a number of 'boxes' along the way.

Aimed primarily at the interested lay reader, this book should also be a 'good read' for school and university mathematics students interested in the history of their subject and for physicists, engineers, and other scientists.

My grateful thanks are due to Lars Andersen, Rob Bradley, Chris Budd, Bill Dunham, Howard Emmens, Florence Fasanelli, Raymond Flood, Adrian Rice, Ed Sandifer, and Andrey Umerski for their helpful comments on the manuscript, and to Latha Menon, Fiona Orbell, Jenny Nugee, and Daniel Gill of Oxford University Press for their encouragement and support.

Robin Wilson
September 2017

CONTENTS


INTRODUCTIO

IN ANALYSIN

INFINITORUM.

AUCTORE

LEONHARDO EULERO,

Professore Regio BEROLINENSI, *& Academiæ Imperialis Scientiarum* PETROPOLITANÆ *Socio.*

TOMUS PRIMUS

LATIUS SUB ÆGIDE LUCEBIT

LAUSANNÆ,

Apud MARCUM-MICHAELEM BOUSQUET & Socios.

MDCCXLVIII.

Fig. 1. Euler's *Introductio in Analysin Infinitorum,* 1748

INTRODUCTION

The most beautiful theorem in mathematics

Mathematics, rightly viewed,
possesses not only truth,
but supreme beauty – a beauty cold
and austere, like that of sculpture

Bertrand Russell

In mathematics a *theorem* is a result, derived from more basic principles, that's been proved to be true: examples include the well-known Pythagorean theorem on right-angled triangles, and Euclid's theorem that the list of prime numbers continues for ever. Many theorems are relatively simple to prove; others, such as 'Fermat's last theorem', may take many years, or even centuries. Some theorems have short and straightforward proofs; others may involve much tedious and lengthy analysis of special cases. In general, mathematicians tend to prefer proofs that are efficient, ingenious, surprising, or elegant – even *beautiful*.

So what's the most beautiful theorem in mathematics? Over many years mathematicians have conducted surveys to answer this question. *The Mathematical Intelligencer*, a quarterly mathematics journal, carried out such a poll in 1988 in which twenty-four theorems were listed and readers were invited to award each a 'score for beauty'. While there were many worthy competitors, the winner was 'Euler's equation'.

An *equation* is a statement with an equals sign in it, such as $1 + 1 = 2$, Einstein's equation $E = mc^2$, or an algebraic equation such as $x^2 - 3x + 2 = 0$. In 2004 the popular monthly magazine *Physics World* polled its readers to find *The greatest equations ever*, and even among physicists Euler's equation came a close second to the winning entry, Maxwell's equations for electromagnetism. Lagging far behind these two front-runners were such classical equations as the Pythagorean theorem, Einstein's equation, Newton's second law of motion ($F = ma$) linking force and acceleration, and Boltzmann's law of entropy ($S = k \ln W$). In their responses to the poll's findings the participants described Euler's equation as 'mind-blowing', 'filled with cosmic beauty', and 'the most profound mathematical statement ever written'.

In 1933, at the age of only 14, the future Nobel Prize-winning physicist Richard P. Feynman described Euler's equation in similar terms as 'The most remarkable formula in math', and in later years he referred to its close relative 'Euler's identity' in the same way, adding 'This is our jewel'. Waxing even more enthusiastically, the Stanford mathematician Keith Devlin was moved to claim that

> Like a Shakespearian sonnet that captures the very essence of love, or a painting that brings out the beauty of the human form that is far more than just skin deep, Euler's equation reaches down into the very depths of existence.

Can this beauty be tested? In 2014, in what was described as an 'equation beauty contest', two neuroscientists used MRI scanning to test the brain activity of several mathematicians while viewing each of fifteen equations that they'd earlier described as beautiful, indifferent, or ugly. They found, as BBC News put it, that the same emotional brain centres used to appreciate art were being activated by 'beautiful' maths.

Unsurprisingly the equation most consistently rated as 'beautiful' in this experiment was Euler's equation. Sir Michael Atiyah, a former winner of the Abel prize and Fields medal (regarded as the mathematical equivalents of the Nobel Prize) who collaborated with the neuroscientists, may have hit the nail on the head when he described the winning equation as the mathematical equivalent of Hamlet's 'To be, or not to be' – 'very short, very succinct, but at the same time very deep'.

Few equations reach beyond the realms of academia, but Euler's equation has even featured in two episodes (MoneyBART and Homer[3]) of the popular American television series *The Simpsons*. The same equation was also crucial in a criminal court conviction when Billy Cottrell, a talented physics graduate student at the California Institute of Technology (Caltech), was sentenced to eight years in a federal prison for causing over $2 million worth of property damage. A self-confessed environmental extremist, he was convicted of spray-painting slogans onto more than 100 sports utility vehicles which he criticized as 'gas guzzlers, killers, and smog machines'. He was identified after writing Euler's equation (which had 'just popped into my head') on a Mitsubishi Montero. As he proclaimed in his trial:

> I think I've known Euler's theorem since I was five… Everyone should know Euler's theorem.

So what is this result that 'Everyone should know'? And who was Euler?

Euler, his equation, and his identity

> *Old MacDonald had a farm,*
> *e, i, e, i, 0,*
> *And on that farm he had 1 π-g,*
> *e, i, e, i, 0.*
>
> Nursery rhyme

Leonhard Euler was an 18th-century mathematical genius who grew up in Basel in Switzerland but spent most of his life in the Imperial courts of St Petersburg and Berlin. The most prolific mathematician of all time, and one of the four greatest (the other three being Archimedes, Newton, and Gauss), Euler published 866 books and papers in over 70 volumes. These range across almost all branches of mathematics and physics that were then being investigated, and much of today's research in these subjects can be traced back to Euler's pioneering work.

Euler's equation usually appears in one of two equivalent forms:

$$e^{i\pi} + 1 = 0 \quad \text{or} \quad e^{i\pi} = -1$$

It also appears in the forms $e^{\pi i} + 1 = 0$ and $e^{\pi i} = -1$, with the π appearing before the i (as in $2i$, $3i$, etc.).

Why is Euler's equation so popular, and yet so profound? Perhaps it's because the first form above connects five of the most important numbers in mathematics:

1 – the basis of our counting system
0 – the number that expresses 'nothingness'
π – the basis of circle measurement
e – the number linked to exponential growth
i – an 'imaginary' number, the square root of –1.

It also involves the three fundamental mathematical operations of addition (+), multiplication (×), and taking powers, and the notion of equality (=). As one participant in the *Physics World* poll was moved to remark:

> What could be more mystical than an imaginary number interacting with real numbers to produce nothing?

As we'll see, Euler's equation is a special case of a general result that he published in 1748 in his *Introductio in Analysin Infinitorum* (Introduction to the Analysis of Infinites), one of the most important mathematical books ever written (see Figure 1). This general result is *Euler's identity*,

$$e^{ix} = \cos x + i \sin x$$

which simply and beautifully connects the exponential function and the trigonometric functions. Euler's identity has even featured on a Swiss postage stamp (see Figure 2).

But why should the exponential function e^x, which goes 'shooting off to infinity' as x becomes large, have anything to do with the trigonometric functions $\cos x$ and $\sin x$, which forever oscillate between the values 1 and –1 (see Figure 3)? And why should the 'imaginary' square root of –1 be the cause of this connection? Answering these questions is the aim of this book.

Although they may appear rather abstract at first sight, Euler's equation and identity are of fundamental importance to physicists and engineers. As we'll see, exponentials of the form e^{kt} describe things that grow (if $k > 0$) or decay (if $k < 0$) with time t, and those also involving the number i, such as e^{ikt}, describe circular motion (see Figure 4). But by Euler's identity,

Fig. 2. Euler and his identity, $e^{i\varphi} = \cos\varphi + i\sin\varphi$

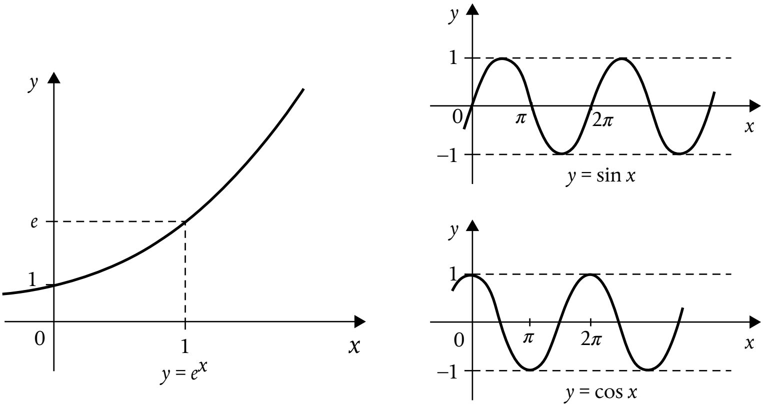


Fig. 3. The graphs of $y = e^x$, $y = \sin x$, and $y = \cos x$

e^{ikt} is made up from cos kt and sin kt and can therefore be used to represent things that oscillate; for example, $e^{i\omega t}$ refers to alternating electric current with angular frequency ω (although electrical engineers tend to use j instead of i, reserving i for electric current), and $e^{i\omega(x-ct)}$ represents the form of radio and sound waves. These 'imaginary exponentials', rather than the related cosines and sines, greatly facilitate mathematical calculations – indeed, for certain more advanced topics in physics and engineering, such as quantum mechanics, signal analysis, and image processing, many calculations cannot be carried out without them.

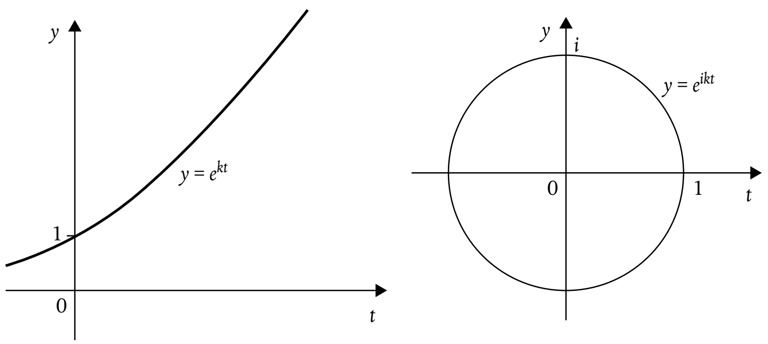


Fig. 4. The graphs of $y = e^{kt}\ (k > 0)$ and $y = e^{ikt}$

A word of warning! The terminology used here is not universal because authors have differing views on what are meant by such terms as 'equation', 'identity', and 'formula'. What I call 'Euler's equation' also appears in the literature as 'Euler's formula' and 'Euler's identity', while 'Euler's identity' is sometimes called 'Euler's equation' or 'Euler's formula'! For me, as mentioned above, an *equation* is a statement with an equals sign in it, whereas a *formula* is a general expression into which we can substitute numbers to obtain a particular answer; examples are πr^2 for the area of a circle with radius r (so substituting $r = 5$ gives 25π for the area of a circle with radius 5), and $\{-b \pm \sqrt{(b^2 - 4ac)}\}/2a$ for the solutions of the quadratic equation $ax^2 + bx + c = 0$. An *identity* is a mathematical equation that holds for a range of values of the variable, such as

$$\sin 2x = 2 \sin x \cos x,$$

which is true for all numbers x, or

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - \dots,$$

which holds for all values of x between −1 and 1.

In summary, Euler's equation is remarkable in combining five entities, each with deep mathematical significance and each with its own story. In this book I'll look at each one in turn. I begin in Chapter 1 with the number 1, in which I look at counting systems from around the world. Chapter 2 then explores the gradual emergence of 0 and of the negative numbers, leading to fractions and irrational numbers and culminating in the 'real numbers' that we all take for granted. Chapters 3, 4, and 5, on the numbers π, e, and i, form the core of this book, and their very different stories arise (respectively) from ideas in geometry, analysis, and algebra. Finally, in Chapter 6, I'll show how Euler and his contemporaries combined all these numbers into the 'most beautiful theorem in mathematics'.

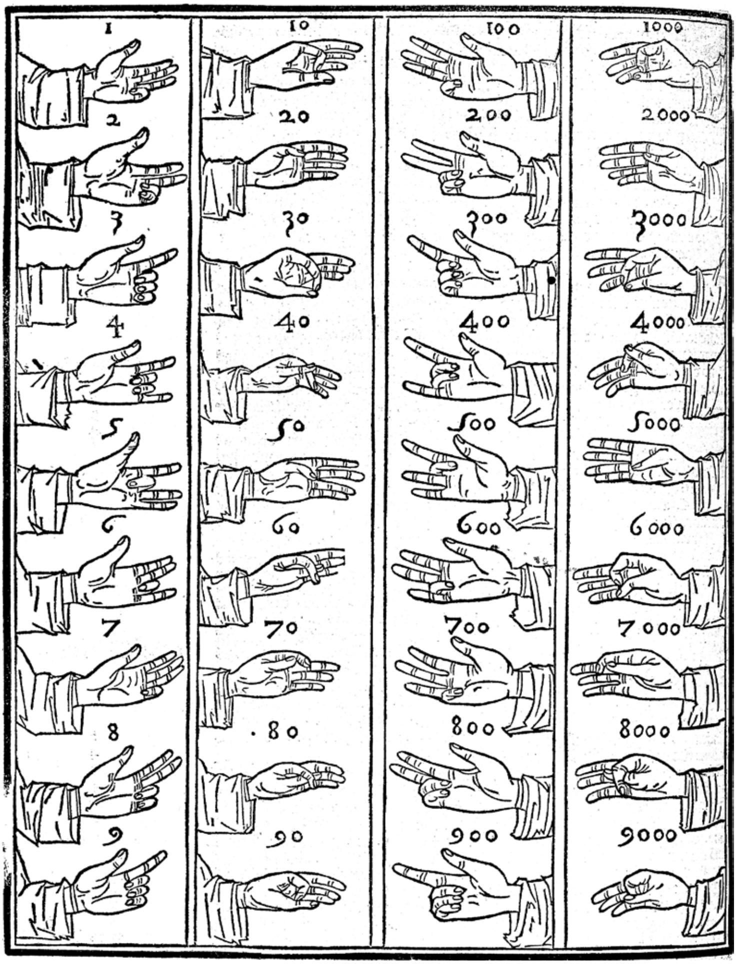


Fig. 5. Finger counting, from Luca Pacioli's *Summa* of 1494

CHAPTER 1

1

The counting numbers

I'll sing you one, O
Green grow the rushes, O
What is your one, O?
One is one and all alone
And evermore shall be so.

English folk song

From earliest times people have needed to count – to keep track of their sheep, to measure their land, to settle financial matters, and much else besides. To do so, they made notches on tally sticks, formed piles of stones, and counted on their fingers, and it is the last of these that undoubtedly led to our decimal system, based on 10, that was adopted by most human civilizations.

However, not every community has used a decimal system for its numbers and its weights and measures. Some have chosen a number system based on 5 (the fingers of one hand) or on 20 (fingers and toes), while a few others have favoured a duodecimal system based on 12, a number with more factors (2, 3, 4, 6) than our decimal system (2 and 5). And as we'll see, the ancient Mesopotamians used a number system based on 60.

Most nations have adopted the metric system for their weights and measures and their coinage. But some still measure lengths in yards, feet, and inches, with 12 inches in a foot and 3 feet in a yard, and until 1971 British money was based on 12 pennies in a shilling and 20 shillings in a pound. Although most countries now use the Celsius (or centigrade) scale for measuring temperatures, with the freezing and boiling points of water as 0° and 100°, the USA still uses the Fahrenheit scale with these points fixed at 32° and 212°. In the UK weights are often given in stones, with 14 pounds in a stone, and in pounds, with 16 ounces in a pound.

With numbers featuring so prominently in our daily existence, it's not surprising that several counting rhymes, used by children as they learn their numbers, are based on our decimal number system, as we see in Box 1.

Number systems

Don't panic!
Base 8 is just like base 10 really
– if you're missing two fingers!

Tom Lehrer's 'New Math'

Number is the basis of counting, and in this section we tour the ancient world and look at its number systems. But first, let's review our familiar decimal system.

Decimal numbers

There are three types of people:
those who can count and those who can't.

Anon

In our decimal system, based on the number 10, we write our numerals with separate columns for units (1), tens (10), hundreds (100), thousands (1000),…, as we move from right to left. Such a system is called a *place-value system* because the placing of each number determines its value. Each place-value is ten times the next one: for example, the decimal number 475 means

$$(4 \times 100) + (7 \times 10) + (5 \times 1) = (4 \times 10^2) + (7 \times 10^1) + (5 \times 10^0),$$

where $10^0 = 1$. In such a representation the same symbol may stand for different numbers, depending on its position; for example, 555 means

$$(5 \times 100) + (5 \times 10) + (5 \times 1),$$

Box 1: Nursery rhymes

Numbers play a role in several nursery rhymes and children's counting games.

Here's a small selection from some numerically based rhymes:

One, two, three, four, five,
Once I caught a fish alive,
Six, seven, eight, nine, ten,
Then I let it go again.

One for sorrow, Two for joy,
Three for a girl, Four for a boy,
Five for silver, Six for gold,
Seven for a secret, Never to be told!
Eight for a wish, Nine for a kiss,
Ten for a bird, You must not miss.

The next children's rhyme provided the title of a detective novel by Agatha Christie:

One, two, Buckle my shoe,
Three, four, Open the door,
Five, six, Pick up sticks,
Seven, eight, Lay them straight,
Nine, ten, A good fat hen.

Finally, the following song can be continued for as long as one wishes:

This old man, he played one,
He played knick-knack on my thumb,
With a knick-knack paddy whack, Give the dog a bone,
This old man came rolling home.

This old man, he played two,
He played knick-knack on my shoe,
With a knick-knack paddy whack…

This old man, he played three,
He played knick-knack on my knee,
With a knick-knack paddy whack…

where the first 5 (representing 500) has ten times the value of the second 5 (representing 50), which in turn has ten times that of the third 5. It follows that we can write down any whole number, however large, using just the ten symbols 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0. Moreover, unlike some of the number systems we consider below, there's just one symbol in each position; for example, we write a single 5 rather than five 1s.

Binary numbers

There are 10 *types of people:*
those who can count in binary
and those who can't.

Anon

Another number system, used throughout computer science, is the binary place-value system. Based on the number 2, the counting numbers are formed from 0s and 1s only, beginning

decimal:	1	2	3	4	5	6	7	8	9	10
binary:	1	10	11	100	101	110	111	1000	1001	1010

To transform a decimal number to its binary form we write it as a sum of powers of 2; for example, the decimal number 13 corresponds to the binary number 1101, since

$$13 = 8 + 4 + 1 = (1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0)$$
$$= \text{the binary number } 1101.$$

Conversely, to turn a binary number such as 1101 into its decimal form we simply reverse the rows above. It's as easy as 1, 10, 11.

A number system based on 2 has been in use by islanders of the Torres Strait, between Papua New Guinea and Australia. It begins

1 = *urapun*, 2 = *okosa*, 3 = *okosa-urapun*, 4 = *okosa-okosa*, 5 = *okosa-okosa-urapun*.

Roman numbers

We're also familiar with Roman numerals, which appear on many clock faces and are used for names of monarchs (Henry VIII), chapter numbers in books, and dates. This is a decimal system, with capital letters used for numbers, but it's not a place-value system because different letters are used for 1, 10, 100, and 1000, and for 5, 50, and 500:

number:	1	5	10	50	100	500	1000
letter:	I	V	X	L	C	D	M

In this system letters are repeated as necessary – for example,

$$287 = 200+50+30+5+2 \text{ is written as } \text{CCLXXXVII}.$$

Numbers that were once written as IIII and XXXX are usually written as IV and XL, meaning 'one before five' and 'ten before fifty', so following this pattern we usually write the year 1944 in Roman numerals as MCMXLIV.

With numerals written in this way arithmetical calculations are impractical, and the Romans generally used an abacus or similar instrument for counting and trading.

Egyptian numbers

Around 2600 BC the Egyptians built the magnificent pyramids, which attest to their extremely accurate measuring abilities. But their writings on papyrus reeds have rarely survived the ravages of several millennia, and most of our knowledge of their mathematical activities comes from just two large but fragile papyri from about 1850–1650 BC – the Rhind mathematical papyrus (named after its purchaser, the Victorian explorer Henry Rhind) in the British Museum and a similar one in Moscow. Both present tables of numbers and dozens of worked problems written in hieratic script (see Figure 6).

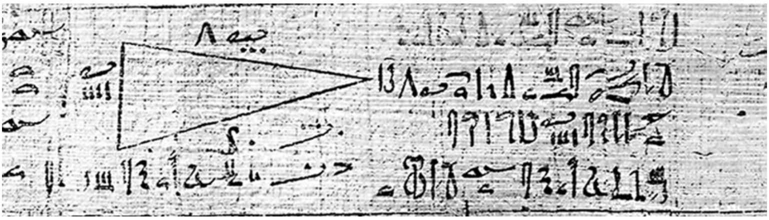


Fig. 6. Part of the Rhind papyrus

It emerges from these primary sources that the Egyptians employed a decimal system, but it was not a place-value system because it used different symbols for the powers of 10: a vertical rod for 1, a heel bone for 10, a coiled rope for 100, a lotus flower for 1000, and so on (see Figure 7).

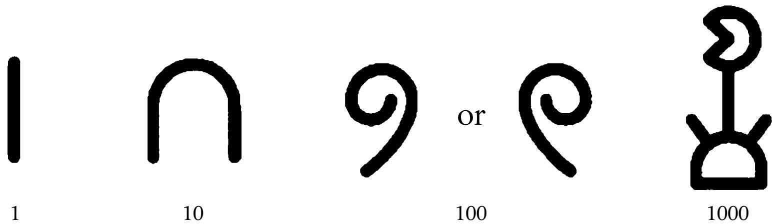


Fig. 7. Egyptian powers of 10

To represent a number the Egyptians used appropriate repetitions of each symbol, usually written from right to left. For example, the numbers 367 and 756 are as shown in Figure 8, and to add them we combine similar symbols, replacing any group of ten that arises by the next symbol (10 rods = 1 heel bone, 10 coiled ropes = 1 lotus flower, etc.).

Fig. 8. Adding Egyptian numbers

The Egyptians had an interesting way of writing fractions which we'll look at in Chapter 2, and in Chapter 3 we'll see how they found the area of a circle with a given diameter.

Mesopotamian numbers

At around the same time, mathematics was flourishing in the region between the rivers Tigris and Euphrates known as Mesopotamia or Old Babylonia. The Mesopotamians, or Babylonians, imprinted their symbols with a stylus on wet clay (cuneiform writing) which was then left to

Fig. 9. A Mesopotamian clay tablet

dry in the sun (see Figure 9). Many thousands of mathematical clay tablets have survived, giving us much valuable information about the Mesopotamians' mathematical activities.

Unlike the Egyptian decimal system the Mesopotamians used a 'sexagesimal system', based on 60. Moreover, it was a place-value system in which each position has 60 times the value of the next. There were only two written symbols, corresponding to 1 and 10 and repeated as necessary; for example, Figure 10 shows the sexagesimal number 1,12,37, representing the decimal number

$$(1 \times 60^2) + (12 \times 60^1) + (37 \times 60^0) = 3600 + 720 + 37 = 4357.$$

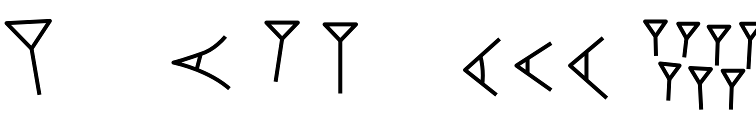


Fig. 10. A Mesopotamian number

There were essentially three types of mathematical clay tablet – *table texts* with lists of numbers for use in calculations, *problem texts* in which questions are posed and answered, and *rough-work tablets,* used by children for their calculations. Several table texts contain multiplication tables, such as the 9-times table in Figure 11.

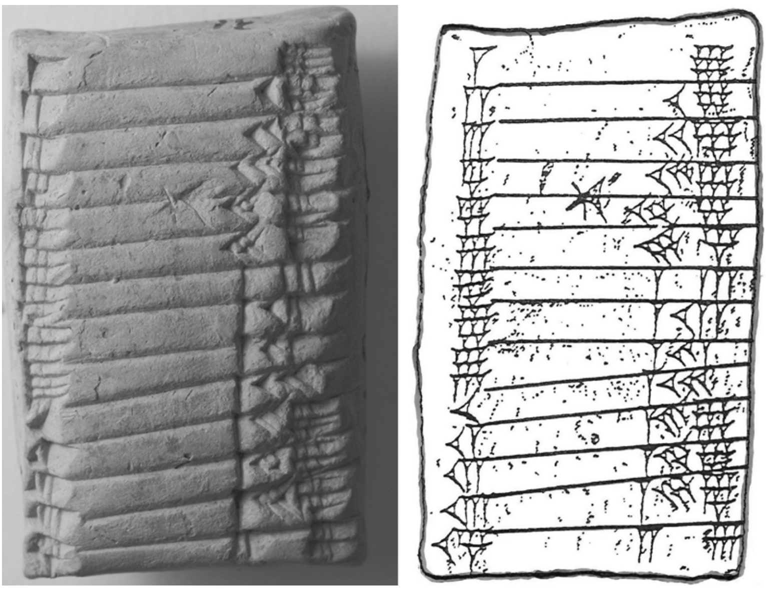


Fig. 11. A Mesopotamian 9-times table

Remnants of the Mesopotamian sexagesimal system survive in our measurements of time (60 seconds in a minute and 60 minutes in an hour) and of angles (such as 60° and 360°).

Greek numbers

The Greek word ἀριθμὸς (*arithmos*) means 'number', and for the Pythagorean brotherhood of around 550 BC 'arithmetic' originally referred to calculating with whole numbers, and by extension to what we now call 'number theory'.

From around 400 BC the Greeks adopted a decimal counting system in which the twenty-four letters of the Greek alphabet, together with three archaic letters (digamma, koppa, and sampi), represented the numbers 1, 2, 3,…, 9, the numbers 10, 20,…, 90, and the numbers 100, 200,…, 900 (see Figure 12); for example, $\phi\nu\varepsilon$ represents the number 555. For larger numbers they added accents or dots; for example, $\nearrow\!\varepsilon$ meant 5000.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
α	β	γ	δ	ε	ς	ζ	η	θ
10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**
ι	κ	λ	μ	ν	ξ	o	π	ϱ
100	**200**	**300**	**400**	**500**	**600**	**700**	**800**	**900**
ρ	σ	τ	υ	ϕ	χ	ψ	ω	ϡ

Fig. 12. The Greek number system

In Euclid's *Elements*, dating from the 3rd century BC, Book VII on arithmetic opens with the following definitions of number:

I. Unity is that, by which everything that is, is called One.
II. Number is a multitude composed of unities.

Before we leave 3rd-century Greece, we mention Archimedes' proficiency with extremely large numbers. In *The Sand Reckoner* he showed how to construct numbers larger than the number of grains of sand in the universe. To this end he began with a myriad (10,000), then a myriad myriad (100,000,000), then $100{,}000{,}000^2$, $100{,}000{,}000^3$,…, up to $P = 100{,}000{,}000^{100000000}$. He then formed powers of P, eventually stopping when he reached the desired number: its size was about 1 followed by 80,000,000,000,000,000 zeros!

Chinese numbers

Although the ancient Chinese employed different symbols for 1, 10, 100,…when writing their numbers, they used a decimal place-value system for their counting boards. These had separate compartments for units, tens, hundreds,…into which small bamboo rods were placed (see Figure 13). Each number from 1 to 9 had two different forms (horizontal and vertical) to enable the calculator to distinguish more easily between adjacent compartments; Figure 14 shows the counting-board representations of 6736 and 2888.

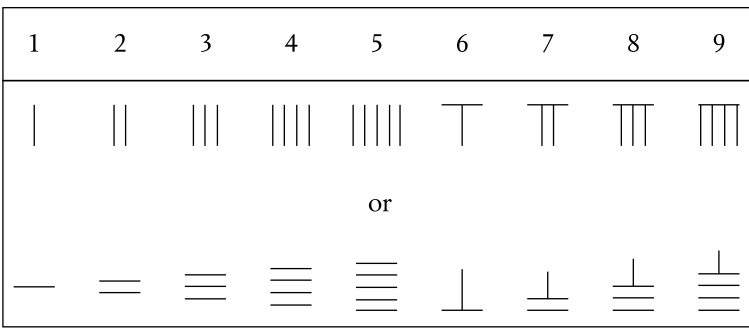


Fig. 13. Chinese counting-board numbers

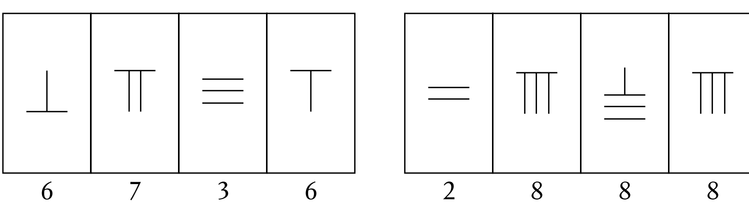


Fig. 14. The counting-board numbers 6736 and 2888

Mayan numbers

The Classic Mayan period lasted from about 250 to 900. Situated over a large area extending from present-day Guatemala, Belize, and Honduras to the Yucatan peninsula of Mexico, the Mayans carried out extensive calendar calculations. Primary sources survive in the form of stone columns called *stelae* and a handful of codices, drawn on tree bark and folded. Part of a Mayan codex (the 'Dresden codex') is shown in Figure 15.

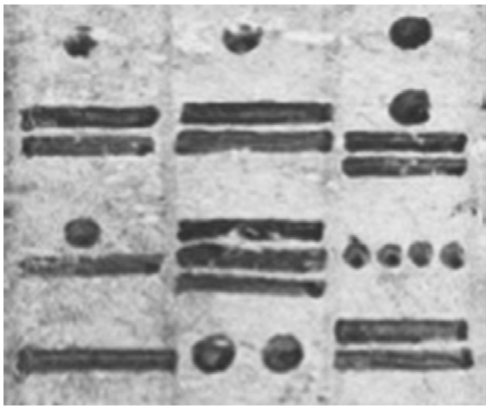


Fig. 15. A Mayan codex

The Mayan number system was mainly based on 20, with a dot to represent 1 and a line to represent 5; Figure 16 shows the numbers from 1 to 19. An attractive feature of the Mayan counting system was that each number also had a pictorial *head-form* representing the head of a man, bird, animal, or deity. Some of these are shown in Figure 17 overleaf.

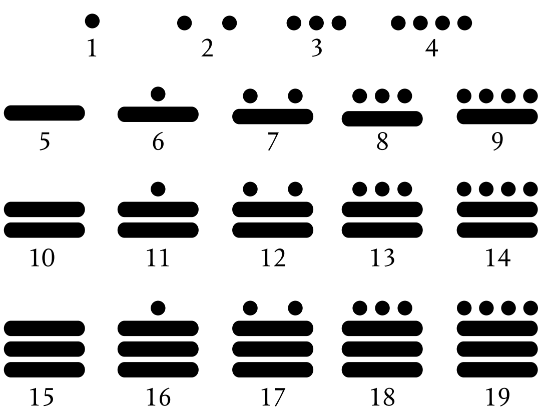


Fig. 16. The Mayan numerals from 1 to 19

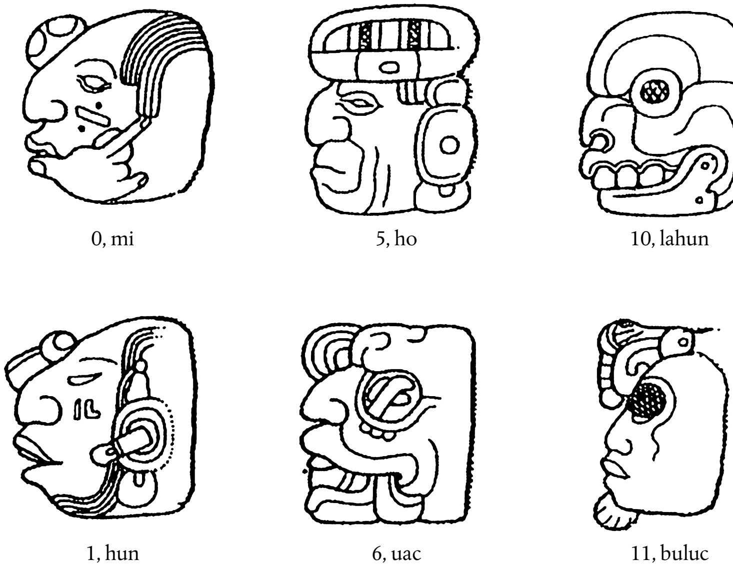


Fig. 17. Head-forms of some Mayan numerals

The Mayans used two basic types of calendar: a ritual calendar of 260 days, known as the *tzolkin* and consisting of thirteen months of 20 days, and a calendar of 360 days with eighteen months of 20 days (plus an extra five 'inauspicious' days to make up the usual 365). These two calendars operated independently but were sometimes combined to give a long-count or *calendar round* with 18,980 days (52 calendar years). These periods of 52 years were then packaged into even longer time periods.

In order to accommodate the 360-day calendar their number system was based on 20 and 18:

1 kin = 1 day
20 kins = 1 uinal = 20 days
18 uinals = 1 tun = 360 days
20 tuns = 1 katun = 7200 days
20 katuns = 1 baktun = 144,000 days
20 baktuns = 1 pictun = 2,880,000 days.

The Mayans had no problem calculating with such large numbers. Indeed, the longest time period found on a codex is 1 kin, 15 uinals, 13 tuns, 14 katuns, 6 baktuns, and 4 pictuns, giving a grand total of 12,489,781 days, or over 34,000 years.

The Hindu–Arabic numbers

Around 250 BC in India the edicts of Emperor Aśoka, the first Buddhist monarch, were carved on pillars around his kingdom. These pillars contained early examples of Indian base-10 numerals, but it was not a place-value system because there were different symbols for 1 to 9, 10 to 90, 100, and 1000. A place-value system with separate columns for units, tens, hundreds,... started to emerge around the year AD 250, and led eventually to the introduction of zero and negative numbers, as we'll see in the next chapter.

In Mesopotamia the period from 750 to 1400 experienced a reawakening of interest in European and Eastern culture. Inspired by the teachings of the prophet Muhammad, Islamic scholars seized on the ancient Greek and Indian texts, translated them into Arabic, and extended and commented on them.

Baghdad, in particular, was well placed to receive these writings, being on the east–west trade routes for silks and spices. There the caliphs actively promoted mathematics, and in the early 9th century Caliph Hārūn al-Rashīd and his son al-Ma'mūn established the 'House of Wisdom', a scientific research academy with its own astronomical observatory and an extensive library of manuscripts. Here the Islamic scholars developed the Indian decimal place-value counting system into what are now our *Hindu–Arabic numerals*. Gradually the numerals diverged into the three separate types shown in Figure 18: the modern Hindu script, the East Arabic numerals written from right to left and still found today in the countries of the Middle East, and the West Arabic numerals that became the number system used throughout Western Europe.

One of the earliest scholars at the House of Wisdom was the Persian scholar Muhammad ibn Mūsā al-Khwārizmī, who is remembered by mathematicians primarily for two books, one on arithmetic and the other on algebra. The former work contained no results of great originality, but

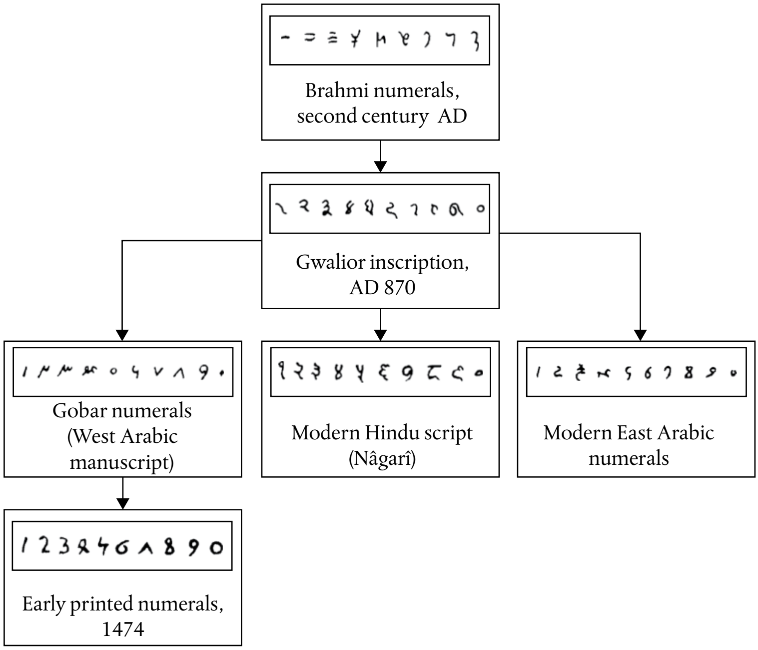


Fig. 18. The origins of our number system

was important for introducing the Indian number system to the Islamic world, and later for helping to spread the decimal counting system throughout Christian Europe. Indeed, al-Khwārizmī's name, transmuted into 'algorism', came to be used in Europe to mean 'arithmetic', and we still use the word *algorithm* to refer to a step-by-step procedure for solving problems.

It took many centuries for the Western form of the Hindu–Arabic numerals to become fully established. These were more convenient to calculate with than Roman numerals, but for practical use many people continued to use an abacus. Figure 19 shows a 16th-century picture representing 'Arithmetic' which contrasts the modern algorist (represented by the 6th-century mathematician Boethius and his Hindu–Arabic numerals) with the old-fashioned abacist (represented by Pythagoras and his counting board).

Fig. 19. The algorist and the abacist, from Gregor Reisch's encyclopedia *Margarita Philosophica*, 1503

As time progressed the situation improved, with the publication of influential arithmetic books that promoted the Hindu–Arabic numerals. These included Leonardo Fibonacci's *Liber Abbaci* (Book of Calculation) in Latin (1202), Luca Pacioli's *Summa in Arithmetica, Geometrica, Proportioni et Proportionalita* (Summary of Arithmetic, Geometry, Proportion, and

Proportionality) in Italian (1494), and Robert Recorde's *The Grounde of Artes* in English (1543). Pacioli's compilation, in particular, gave illustrations for how to calculate on one's fingers (see Figure 5 which opens this chapter). By the 16th century, when such printed books became widely available, the Hindu–Arabic numerals were already in widespread use.

Number names

Up to now we've concentrated on numerals, the *written forms* of numbers. But what about the *spoken forms* – the names 'one', 'two', 'three',...? Although some primitive tribes had no need for numbers beyond this, counting 'one, two, many', it became increasingly necessary to use names for larger numbers such as 'ten', 'twenty', and 'one hundred'.

The use of a decimal system is certainly evident in such English-language number names as 'seventeen' (seven and ten) and 'sixty-seven' (six tens and seven), as it is in many other languages. Figure 20 illustrates the similarities and differences among the number names of some Western languages, both ancient and modern.

Among the similarities here we note those between the forms of *six*, of *three*, and of *eight*. Among the differences we see that the Latin *quattuor*

English	*Gothic*	*Latin*	*Greek*	*French*	*German*	*Basque*
one	ains	unus	heis	un	eins	bat
two	twai	duo	duo	deux	zwei	biga
three	threis	tres	treis	trois	drei	hirur
four	fidwor	quattuor	tettares	quatre	vier	laur
five	fimf	quinque	pente	cinq	fünf	bortz
six	saihs	sex	hex	six	sechs	sei
seven	sibun	septem	hepta	sept	sieben	zazpi
eight	ahtau	octo	okto	huit	acht	zortzi
nine	niun	novem	ennea	neuf	neun	bederatzi
ten	taihun	decem	deka	dix	zehn	hamar
eleven	ainlif	undecim	hendeka	onze	elf	hamaika
twelve	twalif	duodecim	dodeka	douze	zwölf	hamabi
seventeen	sibuntaihun	septendecim	dekaepta	dix-sept	siebzehn	hamasei
twenty	twaitigjus	viginti	eikosi	vingt	zwanzig	hogoi

Fig. 20. Number names in the Western world

and the French *quatre* have a different root from *four, fidwor,* and *vier*. The names for 12 sometimes have the form 2 + 10, as in *duodecim* and *dodeka,* and sometimes seem to be new words, as in *douze* and *zwölf*. The English word *twelve* has the same root as the Gothic *twalif,* which may mean 'two left' (when we've already counted up to ten).

Remnants of counting systems based on 20 survive in the biblical reference (in Psalm 90) to our life span of seventy years as *three-score years and ten,* where a *score* is 20. French also has the remnants of a base-20 system in its use of *quatre-vingt* (four twenties) for 80 and *quatre-vingt-dix* (four twenties and ten) for 90, and similarly the Basque words for 60 and 90 are *hirur-hogoi* (three twenties) and *laur-hogoi-ta-hamar* (four twenties and ten). Even more intriguingly, the full Danish names for 60 and 80 are *tre-sinds-tyve* (three times twenty) and *fir-sinds-tyve* (four times twenty), while those for 50 and 70 are *halv-tred-sindstyve* (half-before-three times twenty) and *halv-fjerd-sindstyve* (half-before-four times twenty), now usually abbreviated to *halvtreds* and *halvfjerds*. Here the appearance of 'half' is reminiscent of the German use of *halb vier* for half-before-four o'clock, or half-past-three.

Many of the words we use in daily life are derived from number names. For example,

the Latin prefix 'uni' (one) appears in 'unicorn', 'universe', and 'unilateral', while its Greek equivalent 'mono' gives us 'monologue' and 'monocle'

the prefix 'twi' (two) appears in 'twin', 'twilight', 'entwine', and 'between', while the versions 'bi' and 'di' appear in 'bicycle', 'biscuit' (originally something twice-cooked), and 'diploma' (originally a document folded in two)

the prefix 'tri' (three) appears in 'trio', 'tricycle', and 'triangle', while the three parts of Yorkshire known as 'Ridings' were originally 'thrithings' or 'thirdings', meaning thirds

a 'siesta' was originally a rest period taken at the *sixth* hour (noon)

and your 'duodenum' gets its name from its length of *twelve* finger-widths!

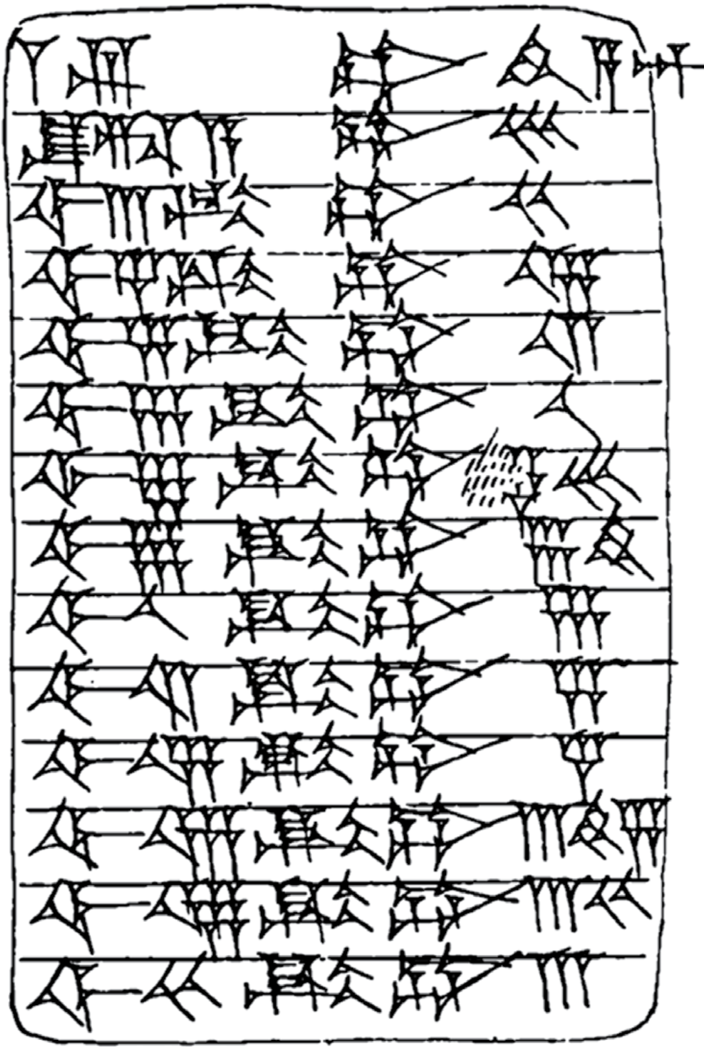


Fig. 21. A Mesopotamian table of reciprocals

CHAPTER 2

0

The nothingness number

Now thou art an O without a figure.
I am better than thou art now.
I am a fool, thou art nothing.
William Shakespeare (*King Lear*)

The civilizations we met in Chapter 1 counted the objects around them: 5 cows, 12 people, etc. – but if there were no things there, there was no need to count them or to introduce a symbol for 0. Even less did they need to use negative numbers: –2 cows would have been meaningless.

But to see how natural the concepts of zero and the negative numbers can be, think of the temperatures on a weather forecast chart. With the Celsius (centigrade) scale the temperatures above freezing are represented by positive numbers (2°C, 10°C, etc.), while the freezing point is 0°C and the temperatures below freezing appear as negative numbers (–2°C). So even though the concepts of zero and the negative numbers took thousands of years to emerge, they appear here very naturally.

The earliest appearances of zero and negative numbers probably arose in trading. Here, profits were recorded as positive numbers, while negative numbers corresponded to debts and zero represented a balance.

Where did the AD/BC numbering system for dates originate? A few years ago there was much discussion as to whether the new millennium should begin in the year 2000 (which it did) or in 2001 (since the 20th century began in 1901). The trouble arose because there was no year 0. The terms

AD and BC had been introduced in the year 531 by Dionysus Exiguus, who calculated that Christ's birth had taken place 531 years earlier. But he couldn't call it year 0, because such a number hadn't yet been introduced – so the year we now call AD 1 immediately followed the year 1 BC. A more natural system was proposed in 1740 by the astronomer Jacques Cassini. Choosing 0 as the year of Christ's birth, and keeping the AD years as they were, he shifted all the BC years by 1, so that the millennium would certainly have started in 2000.

Before proceeding, we must distinguish clearly between 'zero' and 'nothing'. Zero (0) is a number like any other, while 'nothing' means the absence of anything. For example, your age before you reached your first birthday was 0 years – it wasn't nothing. Another example, more a linguistic matter than a misunderstanding, concerns football scores: in the UK the score may be 2–nil, with 'nil' meaning zero, whereas US commentators often say '2 to nothing' – but it isn't nothing, it's something: a score of zero. In tennis we say '30–love', with 'love' meaning zero.

From now on, we shall refer to the positive whole numbers 1, 2, 3,…, the negative whole numbers –1, –2, –3,…, and 0 collectively as *integers* (from the Latin word for 'whole'). Note that 0 is the 'additive identity' for our number system – that is, adding 0 leaves any number unchanged ($x+0=x$ for all x) – just as 1 is the 'multiplicative identity' ($x\times1=x$).

Much ado about nothing

In our decimal place-value system a number like 35 means 3 tens and 5 ones, whereas the number 305 means 3 hundreds and 5 ones. The zero between 3 and 5 registers that there are no tens; if we'd simply left a space instead of writing 0, we might confuse it with 35 or with the two separate numbers 3 and 5. The place-holder 0 shows what's intended.

But how did the early civilizations cope with 'nothingness'? And when did the symbol 0 become widely adopted for the number zero? Let's tour the ancient world again.

As you may remember from Chapter 1, the Egyptians developed a decimal counting system with separate symbols for 1, 10, 100,…, each repeated as often as necessary. Their number system had no need of a special symbol for zero.

The ancient Greeks used separate letters to represent 1 to 9, 10 to 90, and 100 to 900, but likewise had no special symbol for zero. In their geometrical style of arithmetic, numbers were often represented by lines, so 0 (with zero length) would have been difficult to represent.

But the situation was rather different for the Mesopotamians who used a place-value system based on 60. In their early tablets from around 1800 BC the Mesopotamians sometimes left a gap to indicate no entry in that position, so 3, 5 would mean $(3\times60)+5$ (= 185 in our decimal system), whereas 3, ,5 would mean $(3\times60^2)+5$ (=10,805). Having no zero the Mesopotamians expected the intended value to be clear from the context – for example, when distinguishing between 3 and 3×60. Later, possibly around 600 BC, they began to introduce a special place-holder symbol to mark the space (see Figure 22), but it wasn't a number like 0 that they could use in calculations.

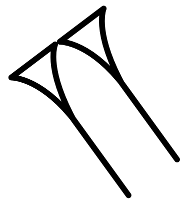


Fig. 22. A Mesopotamian symbol for zero

For the Mayans of Central America the idea of zero was firmly established. As we saw in Chapter 1, their number system was a place-value system based mainly on 20. To keep track of the different powers of 20, they used a special symbol for zero, rather like an eye or a shell (see Figure 23).

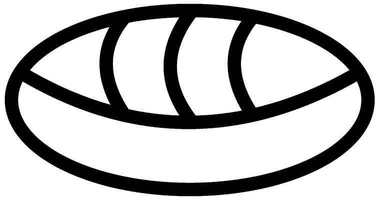


Fig. 23. A Mayan symbol for zero

In China the counting boards had separate compartments for units, tens, hundreds,…, and leaving one empty indicated a zero; for example, Figure 24 shows the number 2301. Although it would have been natural to introduce a zero symbol to represent this, the Chinese didn't do so.

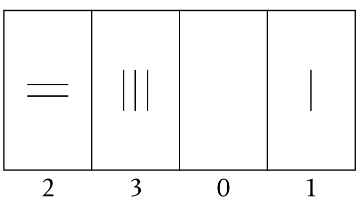


Fig. 24. The number 2301 on a Chinese counting board

The Indians did, however. Whether they were familiar with Chinese counting boards is unknown, though the Chinese did visit India and their boards were transportable (like laptops). In any case, the Indian number system developed as a decimal place-value system based on the numbers 1 to 9, and later including the number 0. With these ten digits they could represent any whole number, however large.

The origin of the round form of the numeral for zero is unclear. It may have arisen from the use of round counters as place-holders for numbers written in the sand, or from the circular impression left by a stone. Or it may have arisen from the astronomer Claudius Ptolemy's use of *o* to represent zero in his influential astronomical work, *The Almagest*; *o* is the Greek letter omicron, denoting *οὐδὲν* (ouden) and meaning 'nothing'. The oldest known occurrences of zero, from around the 3rd or 4th century, appear in a fragmentary Indian text known as the 'Bakhshali manuscript'; inscribed on birch bark it may have been a training text for merchants (see Figure 25). Later, zero was represented by a small dot in a Cambodian inscription from the year 603, and by a small circle in Indonesia in 686.

Fig. 25. Part of the Bakhshali manuscript: zero appears as a large dot in the bottom row

As for the word 'zero', this originated in the Arabic word *sifr,* meaning 'empty', which was later transcribed into Latin as *cifra,* or as *cefirum* or *zefirum*. The word *cifra* eventually became our word *cipher,* while the Italians changed *zefirum* to *zefiro,* and then to *zevero,* which was eventually shortened to *zero.*

Calculating with zero and negative numbers

The other use of 0 is as a number to use in calculations, where the difference between 3 and 2 (= 1) and the difference between 3 and 3 (= 0) are awarded the same mathematical status. For negative numbers a similar remark can be made for the difference between 2 and 3 (= −1).

In a famous passage from the 7th century the Indian mathematician Brahmagupta laid down rules for such calculations; here the words ‘cipher’ and ‘nought’ can be regarded as interchangeable:

The sum of cipher and negative is negative; of positive and nought, positive; of two ciphers, cipher.

[For example, $0+(-2)=-2;\ 3+0=3;\ 0+0=0$.]

Negative taken from cipher becomes positive, and positive from cipher is negative; cipher taken from cipher is nought.

[For example, $0-(-2)=2;\ 0-3=-3;\ 0-0=0$.]

The product of cipher and positive, or of cipher and negative, is nought; of two ciphers is cipher.

[For example, $0\times 3=0;\ 0\times(-2)=0;\ 0\times 0=0$.]

From this time onwards zero could be used in arithmetical calculations – or almost…

‘Thou shalt not divide by zero’

Having dealt with addition, subtraction, and multiplication, Brahmagupta then turned his attention to division:

Cipher divided by cipher is nought. Positive or negative divided by cipher is a fraction with that as denominator.

As we’ll see, the first of these statements ($0/0 = 0$) is nonsense, whereas the second is vacuous.

The trouble arises when we break the mathematicians’ eleventh commandment, ‘Thou shalt not divide by zero’. For if we take two numbers, such as 4 and 9, we can write

$$4\times 0=9\times 0$$

(since both products are zero). If we now divide both sides of this equation by 0 we get $4 = 9$, which is nonsense. We could also rearrange the equation to give $0/0 = 4/9$, which disagrees with Brahmagupta’s answer of 0. And we could similarly use the equation

$$a\times 0=b\times 0$$

to prove that *any* two numbers a and b (such as your house number and telephone number) are equal. Rearranging this equation gives $0/0 = a/b$, so that $0/0$ can take any value we choose.

We might try to get around this problem by saying that $a/0$ is 'infinity' (denoted by ∞), but infinity isn't a number and doesn't obey the rules of arithmetic – for example,

$$\infty + 2 = \infty, \quad \infty - 2 = \infty, \quad \text{and } \infty \times 2 = \infty.$$

Also, since $a + \infty = b + \infty$ for any two numbers a and b, we can subtract infinity from both sides to deduce that $a = b$, which again is nonsense.

We conclude this section with a letter written by the Oxford mathematician Charles L. Dodgson (better known as Lewis Carroll, author of *Alice's Adventures in Wonderland*) to Wilton Rix, a young lad of 14. It uses the algebraic fact that $x^2 - y^2 = (x + y)(x - y)$.

Honoured Sir,

Understanding you to be a distinguished algebraist (i.e. distinguished from other algebraists by different face, different height, etc.), I beg to submit to you a difficulty which distresses me much.

If x and y are each equal to "1," it is plain that $2 \times (x^2 - y^2) = 0$,
and also that $5 \times (x - y) = 0$.
Hence $2 \times (x^2 - y^2) = 5 \times (x - y)$.
Now divide each side of this equation by $(x - y)$.
Then $2 \times (x + y) = 5$.
But $(x + y) = (1 + 1)$, i.e. $= 2$.
So that $2 \times 2 = 5$.

Ever since this painful fact has been forced on me, I have not slept more than 8 hours a night, and have not been able to eat more than 3 meals a day.

I trust that you will pity me and will kindly explain the difficulty to

Your obliged, Lewis Carroll.

The difficulty arises when he divides both sides of the equation by $x - y$, which equals $1 - 1 = 0$. This method of 'proof' has also been adapted to other situations; for example, it's even been used to prove that 'Winston Churchill is a carrot'!

0

From integers to real numbers

The dear Lord made the integers,
Everything else is the work of man.

Leopold Kronecker,
German mathematician

Now that we have all the integers (positive, negative, and zero), let's look at some other types of numbers – fractions, irrational numbers, and real numbers – to set the scene for the next two chapters on π and e.

Fractions

A *fraction* (meaning 'broken number') is a ratio of whole numbers – for example, $\frac{3}{4}$ and $-\frac{11}{7}$ are fractions; we do not allow division by 0. To multiply fractions we just multiply their numerators (top) and denominators (bottom), and to add or subtract them we first replace them by fractions on a common denominator – for example:

$$\frac{3}{4}\times\frac{2}{7}=\frac{6}{28}=\frac{3}{14} \quad \text{and} \quad \frac{1}{4}+\frac{2}{7}=\frac{7}{28}+\frac{8}{28}=\frac{15}{28}.$$

Fractions were already fully established by the time that zero appeared on the scene, particularly in problems on the sharing of objects among people. Several of these appear on the Rhind papyrus – for example:

Problem 65: Example of dividing 100 loaves among 10 men, including a boatman, a foreman, and a doorkeeper, who receive double shares. What is the share of each?

Here we're essentially sharing the loaves among 13 people, so that those receiving single portions receive $\frac{100}{13}=7\frac{9}{13}$ loaves and those receiving double portions receive twice this number, which is $15\frac{5}{13}$ loaves. But the answer given on the papyrus was $15\ \frac{1}{3}\,\frac{1}{26}\,\frac{1}{78}$ for those receiving double portions, and $7\ \frac{2}{3}\,\frac{1}{39}$ for the rest; these answers agree with our earlier ones, since

$$\frac{1}{3}+\frac{1}{26}+\frac{1}{78}=\frac{26+3+1}{78}=\frac{30}{78}=\frac{5}{13} \quad \text{and} \quad \frac{2}{3}+\frac{1}{39}=\frac{26+1}{39}=\frac{27}{39}=\frac{9}{13}.$$

So Egyptian fractions were very different from ours, being mainly *unit fractions* or reciprocals $1/n$, together with the fraction $\frac{2}{3}$. In order to calculate with these reciprocals the Egyptians used fraction tables – for example, the Rhind papyrus includes the unit fraction equivalents of $2/n$ for all odd numbers *n* from 5 to 101, beginning with

$$\frac{2}{11}=\frac{1}{6}=\frac{1}{66} \text{ and } \frac{2}{13}=\frac{1}{8}\,\frac{1}{52}\,\frac{1}{104}.$$

The Egyptians' ability to calculate with these unit fractions is evident from the following problem on the Rhind papyrus:

Problem 31: A quantity, its $\frac{2}{3}$, its $\frac{1}{2}$, and its $\frac{1}{7}$, added together become 33. What is the quantity?

To answer this, we'd now use algebraic notation, solving the equation $x+\frac{2}{3}+x\frac{1}{2}x+\frac{1}{7}x=33$ and obtaining the answer $14\frac{28}{97}$. The answer that appears on the papyrus is

$$14\ \frac{1}{4}\ \frac{1}{56}\ \frac{1}{97}\ \frac{1}{194}\ \frac{1}{388}\ \frac{1}{679}\ \frac{1}{776}$$

– a truly impressive feat of calculation.

The Mesopotamians also calculated with fractions, with a sexagesimal number such as 0;06,40 representing the decimal number

$$0+(6\times 60^{-1})+(40\times 60^{-2})=0+\frac{6}{60}+\frac{40}{3600}=0+\frac{1}{10}+\frac{1}{90}=\frac{1}{9}.$$

Of particular importance was their use of reciprocals to facilitate division – to divide by a number, they multiplied by its reciprocal. The Mesopotamian table which opens this chapter (Figure 21) lists the reciprocals of those numbers that are made up from 2, 3, and 5 – these are the only reciprocals with a terminating sequence of sexagesimal digits. For example, the reciprocal of 2 is $0;30$ because

$$2\times 0;30=2\times\frac{30}{60}=1,$$

and the reciprocal of 9 is $\frac{1}{9}=0;06,40$ by the calculation above.

The Greeks' use of fractions arose through their interest in 'commensurability'. The Pythagoreans called two numbers *commensurable* if each can be 'measured' a whole number of times by the same ruler. For example, the numbers 21 and 9 are commensurable because each can be measured an exact number of times by a ruler of length 3 (or of length 1), and similarly 7π and 3π are commensurable because each can be measured by a ruler of length π. In general, two numbers are commensurable if we can write their ratio as a fraction – for example, each of the two commensurable pairs above has ratio $\frac{7}{3}$. However, as the Greeks discovered, the ratio of the diagonal and side of a square are not commensurable; by the Pythagorean theorem, this ratio is $\sqrt{2}/1 = \sqrt{2}$ (see Figure 26). Plato remarked that 'he who knows not this, is not a Man, but a Beast'.

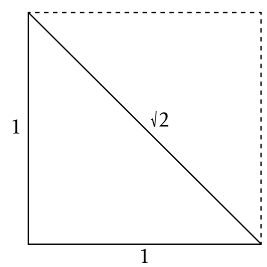


Fig. 26. The diagonal and side of a square

Irrational numbers

A *rational number* is one that can be represented by a ratio or fraction; for example, the fractions $\frac{2}{4}$, $\frac{10}{20}$, and $\frac{-35}{-70}$ all represent the same rational number which we usually denote by $\frac{1}{2}$, the fraction in its 'lowest terms'. Note that every integer, such as 5 (= $\frac{5}{1}$), is rational. A number that isn't rational, such as $\sqrt{2}$, is called *irrational*. Other important examples of irrational numbers include π and e, as we'll see in Chapters 3 and 4.

To explain why $\sqrt{2}$ is irrational we must show that this number cannot be written as a fraction a/b, where a and b are integers. Our proof, which is typical of the Greek approach, is by contradiction and is given in Box 2, but where the Greeks would have couched everything in geometrical terms, we'll use algebraic notation. We use the fact that the square of an even number is even, and the square of an odd number is odd.

Box 2: $\sqrt{2}$ is irrational

We need to prove that $\sqrt{2}$ cannot be written as a fraction. To do so *we'll assume that $\sqrt{2}$ can be written as a/b*, and obtain a contradiction. We may assume that a/b is written in its lowest terms, so a and b have no common factor.

By squaring, we can rewrite the equation $\sqrt{2} = a/b$ as $a^2 = 2b^2$, so a^2 (being twice an integer) must be an even number. But if a^2 is even, then a must also be even.

Since a is even, we can write $a = 2k$, for some integer k. So $a^2 = 2b^2 = 4k^2$, which gives $b^2 = 2k^2$. It follows that b^2 is even, and so b is even.

This gives us our contradiction: a and b are both even, and so both are divisible by 2, contradicting the fact that a and b have no common factor.

This contradiction arises from our original assumption that $\sqrt{2}$ can be written as a fraction, so this assumption is wrong: $\sqrt{2}$ *cannot* be written as a fraction, and so is irrational.

We can use similar contradiction arguments to prove that $\sqrt{3}$, $\sqrt{5}$, and $\sqrt[3]{2}$ are all irrational.

Another interesting irrational number, also involving a square root, is the *golden ratio*

$$\varphi = \tfrac{1}{2}(1+\sqrt{5}) = 1.61803399\ldots$$

where φ is the Greek letter phi. This number has some remarkable arithmetic properties – for example,

$$\varphi^2 = 2.61803399\ldots = \varphi + 1 \quad \text{and} \quad 1/\varphi = 0.61803399\ldots = \varphi - 1.$$

The golden ratio also has some interesting geometrical properties. If we take a 'golden rectangle' whose sides are φ to 1, and remove a square with side 1, we obtain a rectangle whose sides (of lengths 1 and $\varphi - 1$) are in the same ratio as before (see Figure 27). This is because $\varphi^2 = \varphi + 1$, from which we can deduce that $\varphi/1 = 1/(\varphi - 1)$. The golden rectangle has often been regarded as perfectly shaped – neither too fat nor too thin.

0

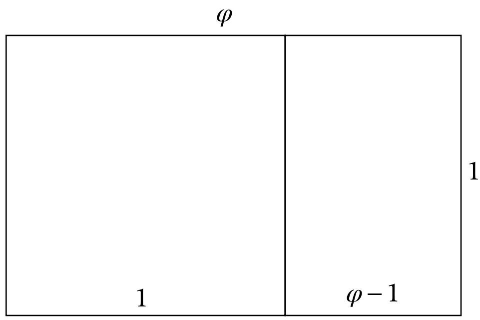


Fig. 27. Removing squares from golden rectangles

If we now continue to remove squares, as in Figure 28, we eventually obtain a spiral pattern that appears throughout nature – for example, on certain shells and in the pattern of seeds on the head of a sunflower.

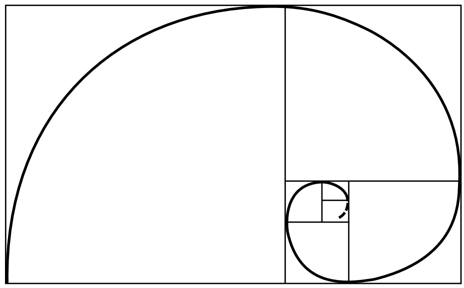


Fig. 28. Removing squares from golden rectangles

Real numbers

We're all familiar with the *real number line,* the line with all the numbers on it, both rational and irrational (see Figure 29). But what exactly is a 'real number'?

We might try to define real numbers by saying that the number line consists of all the rational and all the irrational numbers – but this is

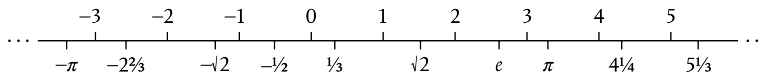


Fig. 29. The real number line

unsatisfactory because we've already defined the irrational numbers to be all those real numbers that aren't rational!

It is not difficult to prove that:

Every rational number can be written as a finite or recurring decimal fraction.

For example,

$$\frac{3}{8} = 0.375, \ \frac{2}{9} = 0.2222\ldots, \text{ and } \frac{1}{7} = 0.14285714285714\ldots$$

We can also go the other way and prove that

Every finite or recurring decimal can be written as a fraction.

For example, $0.375 = \frac{375}{1000} = \frac{3}{8}$, and if $x = 0.2222\ldots$, then $10x = 2.2222\ldots$, and so (by subtraction) $9x = 2$, giving $x = \frac{2}{9}$. We can similarly prove that $0.9999\ldots = 1$: if you don't believe this just multiply the equation $0.3333\ldots = \frac{1}{3}$ by 3.

So we might try to define a real number to be a finite or infinite decimal: if this decimal is finite or recurs, then the number's rational, but if it's an infinite decimal that doesn't recur, then the number's irrational. Unfortunately, although this approach based on decimals may seem to define the real numbers, we run into trouble when we try to do arithmetic. For example, if we agree to define $\sqrt{2}$ in this way, as the non-recurring infinite decimal $1.41421356237\ldots$, how do we then prove that $\sqrt{2} \times \sqrt{2} = 2$? (Try multiplying two infinite decimals together!)

Much time was spent in the second half of the 19th century in addressing such fundamental difficulties. A formal definition of the real numbers was eventually provided by the German mathematician Richard Dedekind, using the idea of what is now called a *Dedekind cut* (see Box 3).

Box 3: Defining a real number

Dedekind noticed that the set of real numbers (which we're trying to define) differs from the set of rational numbers (which we already know about), because the latter has 'gaps' (at $\sqrt{2}$, π, and e, for example). We can fill these gaps with numbers corresponding to the irrational numbers. Indeed, for Dedekind, each gap *is* a number, which we can specify by the sets L (left), consisting of all the rational numbers less than it, and R (right), consisting of all the rational numbers greater than it. For example, Figure 30 shows the gap at $\sqrt{2}$, the set L consisting of all rational numbers less than $\sqrt{2}$ (such as -3, 0, and 1.4142), and the set R consisting of all rational numbers greater than $\sqrt{2}$ (such as 1.4143 and 57).

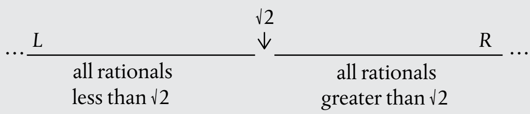


Fig. 30. A Dedekind cut

As Dedekind wrote in 1872:

> If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.

He then defined the set of real numbers to be the set of all these cuts, and showed how arithmetic can be carried out when we use this definition.

Algebraic and transcendental numbers

We've already classified real numbers as rational or irrational, and we'll conclude this chapter with a different classification that we'll need in Chapter 6. It separates numbers into those that are 'algebraic' and those that are 'transcendental'.

A *polynomial* is an expression made up from sums and differences of powers of a variable, such as

$$x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60.$$

Algebraic numbers (generalizing the rational numbers) are numbers that are solutions to polynomial equations with integers as coefficients. For example,
$\frac{2}{9}$ is algebraic – it's the solution of the equation $9x = 2$.
$\sqrt{2}$ is algebraic – it's a solution of the equation $x^2 = 2$.
$\sqrt[3]{7}$ is algebraic – it's a solution of the equation $x^3 = 7$.
$\sqrt{2} + \sqrt{3}$ is algebraic – it's a solution of the equation $x^4 - 10x^2 + 1 = 0$.
the imaginary number $i = \sqrt{-1}$ is algebraic – it's a solution of the equation $x^2 = -1$.
Numbers that are not algebraic are called *transcendental* – as Euler remarked, they 'transcend the power of algebraic methods'.

The first number that was proved to be transcendental was discovered by the French mathematician Joseph Liouville in 1848. He started with the *factorial numbers n*!, defined by

$$1! = 1,\ 2! = 2\times 1 = 2,\ 3! = 3\times 2\times 1 = 6,\ 4! = 4\times 3\times 2\times 1 = 24, \ldots,$$

and proved that the number

$$1/10^{1!} + 1/10^{2!} + 1/10^{3!} + 1/10^{4!} + \ldots = 1/10^{1} + 1/10^{2} + 1/10^{6} + 1/10^{24} + \ldots$$

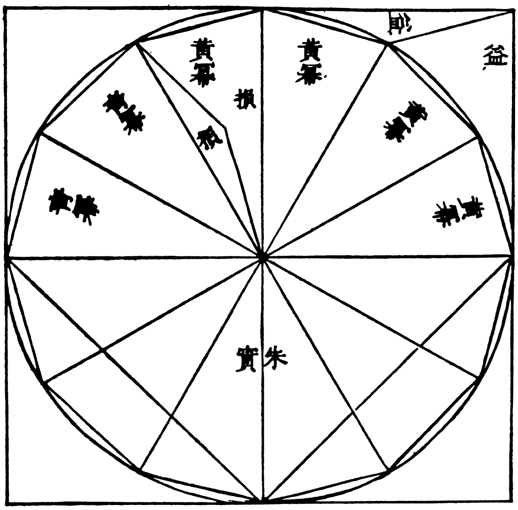
$$= 0.110001000000000000000001000\ldots$$

is transcendental.

But classifying numbers in this way is extremely difficult. In 1873 the French mathematician Charles Hermite proved that e is transcendental, and the corresponding result for π was proved nine years later by Ferdinand Lindemann of Germany (see Chapter 6). Later, in a famous lecture on unsolved mathematical problems in 1900, David Hilbert asked whether numbers like $2^{\sqrt{2}}$ are transcendental, a result that was eventually proved in the 1930s. Yet even now we know very little: e^{π} has been proved transcendental, but we still don't know whether $e + \pi$, πe, and π^{e} are algebraic or transcendental.

得徑七周二十二乃祖氏之約率非密率也淳風等以爲密率失其實矣徽率與祖氏之約率相較則徽率密於約率

弧田圖



據注意取半圓驗之黃冪令損益相補適滿大方四分之一則青冪適八分之一也合青黃冪爲半外方四分之三朱實與黃冪相等舊以十觚之冪爲圓冪又以半圓論弧矢立法之疎顯然

訂訛補圖

微波榭

Fig. 31. Liu Hui's study of a dodecagon

CHAPTER 3

π

The circle number

'Tis a favorite project of mine
A new value of π to assign.
I would fix it at 3
For it's simpler, you see,
Than 3 *point* 14159.

Harvey L. Carter

This chapter concerns the irrational number

$$\pi = 3.14159265358979323846264338327 9\ldots.$$

It is the ratio of the circumference C of a circle of radius r to its diameter d (= $2r$), as shown in Figure 32 – that is,

$$\pi = C/d, \quad \text{so} \quad C = \pi d = 2\pi r.$$

This ratio is the same for circles of any size – from a pizza to the Moon.

It is also the ratio of the area A of a circle to the square of its radius r,

$$\pi = A/r^2, \quad \text{so} \quad A = \pi r^2 = \pi d^2/4.$$

This ratio is also the same for all circles. Indeed, as Euclid proved in Book XII of his *Elements* in the 3rd century BC,

[The areas of] circles are to one another as the squares on their diameters,

so that the area of a circle of radius r is proportional to d^2, and so also to r^2.

But why does the same number π appear in the formulas for both the circumference and the area? One way to answer this question is to take

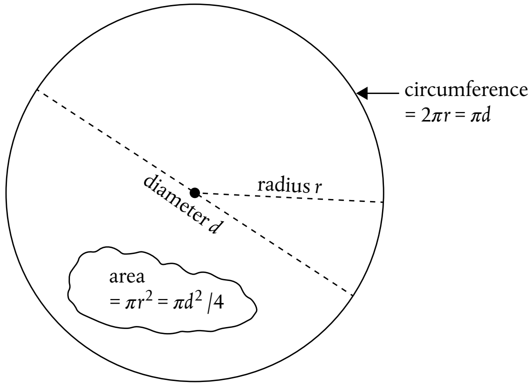


Fig. 32. The circle number π

two circles of radius r, one shaded and the other unshaded, and assume that each has circumference $2\pi r$. We then divide each circle into a number of sectors and rearrange these sectors into a shape that resembles a parallelogram, as in Figure 33. If we now let the number of sectors in each circle become larger and larger, then this parallelogram increasingly resembles a rectangle with sides of length $2\pi r$ and r, which has area $2\pi r \times r = 2\pi r^2$. So the combined area of the two original circles must also be $2\pi r^2$, and consequently each one has area πr^2, as expected.

We can also reverse the previous paragraph to show that if the area is πr^2, then the circumference must be $2\pi r$. Such ideas appear in Sato Moshun's *Tengen Shinan*, a Japanese treatise from 1698 (see Figure 34).

Another link between these two formulas, for those familiar with the integral calculus, is shown in Box 4.

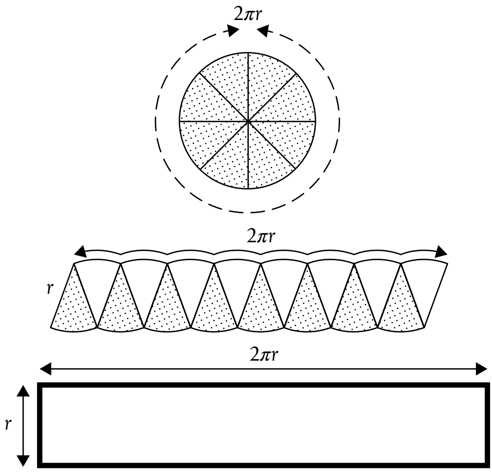


Fig. 33. From $2\pi r$ to πr^2

Box 4: Finding the area of a circle

Consider a circle of radius r. If we assume that the area of each ring of radius x and width dx is $2\pi x\,dx$ (see Figure 35), we can then integrate over all radii from 0 to r to give the area A of the circle as

$$A = \int_0^r 2\pi x\ dx = \pi r^2.$$

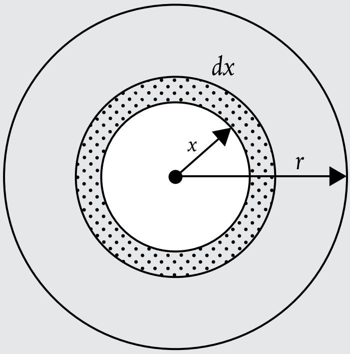


Fig. 35. Finding the area of a circle

與寄左相消
得開方式

三乘方開之得圓徑
推前術得各合問

圖解

如圖上ノ圖形三十二ニ割 此三十二ハ唯見ヨキ
タメナリ幾ツニ割テモ同意ナリ分別スベシ

變形

半円周

半円徑

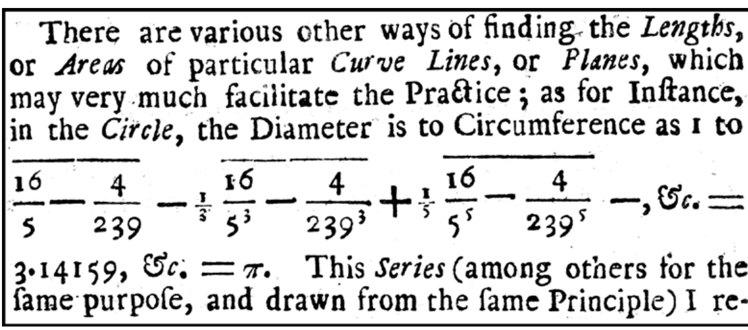
上ノ圓積ノ内ニ黒積十六ト白積十六ト
共ニ三十二アリ 下ノ變形ノ内ニモ黒積十六ト白積十六ト
共ニ三十二アリ 此故ニ上下トモニ積相同ジ
又變形ヲ見ルニ半円徑ト半円周ト相乘ナリ 故円徑ト円周ト相
乘スレバ變形ノ積四段ナリ則円積四段ト同ジ

今有方錐積三十六歩 只云方面長於竪二倍
方面竪各幾何

Fig. 34. A Japanese manuscript

Why π?

The Greek letter π (pi), corresponding to our letter *p*, was first used for the ratio of a circle's circumference to its diameter by a Welsh mathematics teacher called William Jones in 1706; it appeared in his *Synopsis Palmariorum Matheseos* or *A New Introduction to the Mathematicks* (see Figure 36). Earlier, in 1647, William Oughtred had already written π/δ, where π is the circle's $\pi\epsilon\rho\iota\varphi\acute{\epsilon}\rho\epsilon\iota\alpha$ (periphereia, or periphery) and δ is its diameter. But it was Leonhard Euler who popularized the use of the letter π in his 1737 *Variae Observationes Circa Series Infinitas* (Various Observations on Infinite Series) and in many later writings, so that it soon came to be used universally.



There are various other ways of finding the *Lengths*, or *Areas* of particular *Curve Lines*, or *Planes*, which may very much facilitate the Practice; as for Inſtance, in the *Circle*, the Diameter is to Circumference as 1 to $\overline{\frac{16}{5}-\frac{4}{239}}-\frac{1}{3}\,\overline{\frac{16}{5^3}-\frac{4}{239^3}}+\frac{1}{5}\,\overline{\frac{16}{5^5}-\frac{4}{239^5}}-$, &c. $=$ 3.14159, &c. $=\pi$. This *Series* (among others for the ſame purpoſe, and drawn from the ſame Principle) I re-

Fig. 36. William Jones introduces the symbol π

Many mnemonics have been devised for memorizing the first few digits of π. To get the digits, count the letters in the words of these sentences:

How I wish I could calculate pi! (3.141592)
May I have a large container of coffee? (3.1415926)
How I need a drink, alcoholic of course, after the heavy lectures involving quantum mechanics. (3.14159265358979)

and in 1906 the *Literary Digest* published the following poem in praise of Archimedes of Syracuse by A. C. Orr; it yields the first thirty digits of π, as listed at the beginning of this chapter:

Now I, even I, would celebrate
In rhymes inapt, the great
Immortal Syracusan, rivaled nevermore,
Who in his wondrous lore,
Passed on before,
Left men his guidance how to circles mensurate.

These digits, and many more (478 in total), are displayed at the Karlsplatz stop of the Vienna Metro (see Figure 37).

Some people have a remarkable facility for memorizing the digits of π. Those who can recite a thousand digits are not particularly rare, but the record is currently held by a Japanese retired engineer, Akira Haraguchi, who recited 100,000 digits in 16½ hours in 2006, and who later claimed to be able to increase this number to 111,700.

Fig. 37. The Vienna Metro

Early values

Several of the early civilizations we met in Chapter 1 obtained estimates for the areas and circumferences of circles. Such communities had no concept of π as a number, but their results yield lower and upper estimates for its value. We shall frequently write these as decimal numbers, even though such notation was not available to the peoples we discuss.

Let's begin with the Mesopotamians who, we may recall, wrote their mathematical calculations on clay tablets, using a number system based on 60. One of these tablets, dating from around 1800 BC, gives the ratio of the perimeter of a regular hexagon to the circumference of its circumscribed circle as the sexagesimal number 0;57,36. If the radius of the circle is *r*, then each side of the hexagon also has length *r* (see Figure 38), and so

$$\frac{6r}{2\pi r} = \frac{57}{60} + \frac{36}{3600}.$$

This leads, after some calculation, to a value for π of $3\frac{1}{8} = 3.125$, a lower estimate that's within one per cent of its true value.

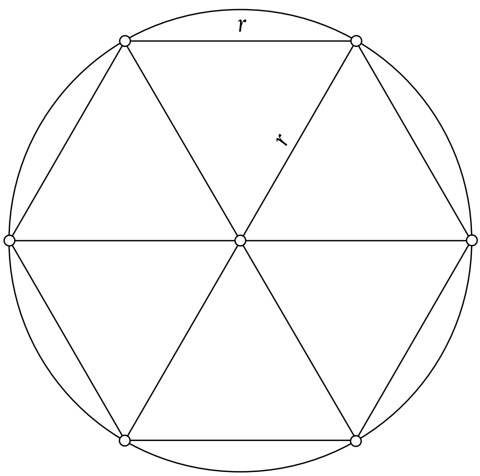


Fig. 38. The Mesopotamian estimate for π

Around the same time as the Mesopotamian clay tablet, the Egyptian Rhind papyrus included the following problem:

Problem 50: Example of a round field of diameter 9 khet. What is its area?
Answer: Take away $\frac{1}{9}$ of the diameter, namely 1. The remainder is 8.
Multiply 8 times 8; it makes 64.
Therefore it contains 64 setat of land.

From this calculation it seems as though the Egyptians approximated the area of a circle of diameter d by reducing d by one-ninth and squaring the result – that is,

$$\text{area} = \left(d - \frac{1}{9}d\right)^2 = \left(\frac{8}{9}d\right)^2 = \frac{64}{81}d^2.$$

It's likely that this method was discovered by experience: various other explanations have been proposed, including one involving a related octagon, but none seems to be supported by historical evidence. In terms of the radius this area is $\frac{256}{81}r^2$, which corresponds to a value for π of about 3.160, an upper estimate that's also within one per cent of the true value.

Both of the above estimates are better than the biblical value given some 1000 years later. In 1 Kings 7:23 and 2 Chronicles 4:2, we read:

> Also, he made a molten sea [a large basin] of ten cubits from brim to brim, round in compass…and a line of thirty cubits did compass it round about.

This corresponds to a value of $\pi = \frac{30}{10} = 3$, an inaccurate but easy-to-use approximation that later reappeared in India, China, and several other lands over many centuries.

Indian mathematicians also developed methods for estimating the area of a circle. Vedic manuscripts of the first millennium BC give various geometrical constructions that yield approximations to π. For example,

> to convert a circle into a square with the same area, take $\frac{13}{15}$ of the diameter and construct the square with this length as its side:

this yields the rather poor value of $\pi = 3.004$. More accurate, though more obscure, was

> to take a square and construct a circle whose radius is half the side of the square plus one-third of the difference of half the diagonal of the square and half the side:

this gives the approximation $\pi = 3.088$.

In the following millennium the Jains discovered a simpler approximation for π – this was $\pi = \sqrt{10}$, which is about 3.162, also within one per cent of the true value. Like the inferior biblical value of 3, it also had widespread use; for example, the value $\sqrt{10}$ was also proposed in China around the year AD 125 by Zhang Heng, chief astrologer and inventor of the seismograph for measuring earthquakes.

Using polygons

An important new method for estimating π was introduced by the Greeks. Although often attributed to Archimedes, the method of estimating π by approximating a circle with polygons can be traced further back to the 5th century BC, to the Greek sophists Antiphon and Bryson. Their aim was to obtain better and better estimates for π by repeatedly doubling the number of sides of a regular polygon until the polygons 'became' the circle. This approach for estimating π would be used for almost two millennia.

Antiphon began by drawing a regular polygon inside the circle and finding its area, giving a lower estimate for π. For example, if we inscribe a square in a circle of radius r, its area is $2r^2$, giving the very poor lower estimate of $\pi = 2$ (see Figure 39). But if we now double the number of sides, giving an octagon, we obtain the better estimate of $\pi = 2\sqrt{2} \approx 2.828$. Bryson's approach was the same, except that he also considered circumscribed polygons. This yields the upper estimates of $\pi = 4$ for the square and $\pi = 8(\sqrt{2}-1) \approx 3.318$ for the octagon.

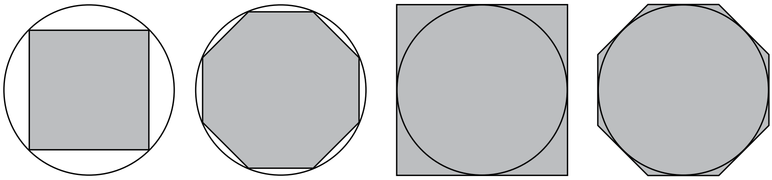


Fig. 39. Approximating the area by squares and octagons

Around 250 BC Archimedes also became interested in circular measurement, proving in his *Measurement of the Circle* that the area of a circle of radius r is πr^2, and in *On the Sphere and Cylinder* that a sphere of radius r has surface area $4\pi r^2$ and volume $\frac{4}{3}\pi r^3$.

Adapting Antiphon and Bryson's earlier ideas, he attempted to estimate π by approximating the circumference of a circle by the perimeters (rather than the areas) of regular polygons drawn inside and outside the circle. By repeatedly doubling the number of sides of the polygons, he could then obtain better and better approximations for π.

Archimedes began by drawing regular hexagons inside and outside a circle, and compared their perimeters with the circumference of the circle (see Figure 40). This gave a lower estimate for π of 3 and an upper estimate of $2\sqrt{3}$, so that, in decimal form,

$$3.000 < \pi < 3.464.$$

Doubling the number of sides, replacing the hexagons by regular dodecagons (12-sided polygons), gives the estimates

$$6\sqrt{\left(2-\sqrt{3}\right)} < \pi < 12\left(2-\sqrt{3}\right) \text{ or, in decimal form, } 3.106 < \pi < 3.215.$$

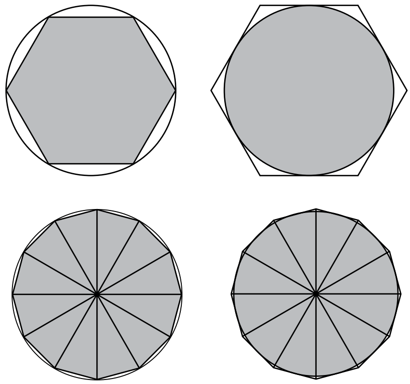


Fig. 40. Approximating the circumference by hexagons and dodecagons

Three more doublings of the number of sides to polygons with 24, 48, and 96 sides give the following ever-closer estimates:

24 sides: $3.133 < \pi < 3.160$
48 sides: $3.139 < \pi < 3.146$
96 sides: $3.141 < \pi < 3.143$.

Archimedes didn't have decimal notation, but gave the results of his calculations for polygons with 96 sides as

$$3\frac{10}{71} < \pi < 3\frac{1}{7}.$$

Four hundred years later Claudius Ptolemy considered the perimeter of a 360-sided regular polygon, obtaining the sexagesimal value $3;8,30 = 3 + \frac{8}{60} + \frac{30}{3600}$, which is about 3.14167.

Around the year AD 263, in his commentary on the Chinese classic *Nine Chapters on the Mathematical Art*, Liu Hui also used inscribed regular polygons to approximate π. Working with areas and starting with regular hexagons and dodecagons (see Figure 31 which opens this chapter), he developed a simple method for calculating the successive areas and perimeters when one doubles the number of sides, and for polygons with 192 $(= 2 \times 96)$ sides he obtained the estimates

$$3.141024 < \pi < 3.142704.$$

His calculations extended to the area of a polygon with 3072 sides (four more doublings), giving the estimate $\pi = 3.14159$.

Even more impressively, around the year 500 Zu Chongzhi and his son Zu Gengzhi doubled the number of sides three more times, extending their calculations to polygons with 24,576 sides and obtaining the estimates

$$3.1415926 < \pi < 3.1415927,$$

which give π to six decimal places. They also replaced Archimedes' fractional approximation of $\frac{22}{7}$ by the more accurate one of $\frac{355}{113}$, which also gives π to six decimal places. As we'll see, this latter approximation wasn't rediscovered in Europe for another thousand years.

Also around the year 500 the Indian mathematician and astronomer Āryabhata proposed the following recipe, derived from the perimeter of an inscribed regular polygon with 384 sides:

> Add 4 to 100, multiply by 8, and add 62,000. The result is approximately the circumference of a circle with diameter 20,000.

This gives $\pi = \frac{62832}{20000} = 3.1416$. Later, around the year 1150, Bhāskaracharya (or Bhāskara II) contrasted Āryabhata's 'accurate' value with the more 'practical' value of $\frac{22}{7}$.

In 1424 the Persian mathematician and astronomer Jamshīd al-Kāshī (or al-Kāshānī) was working in Ulugh Beg's observatory in Samarkand, now in Uzbekistan. Using polygons with $3 \times 2^{28} = 805{,}306{,}368$ sides he improved Zu Chongzhi's estimates, approximating π to a remarkable nine sexagesimal (= sixteen decimal) places. This remained the best value of π for almost two hundred years.

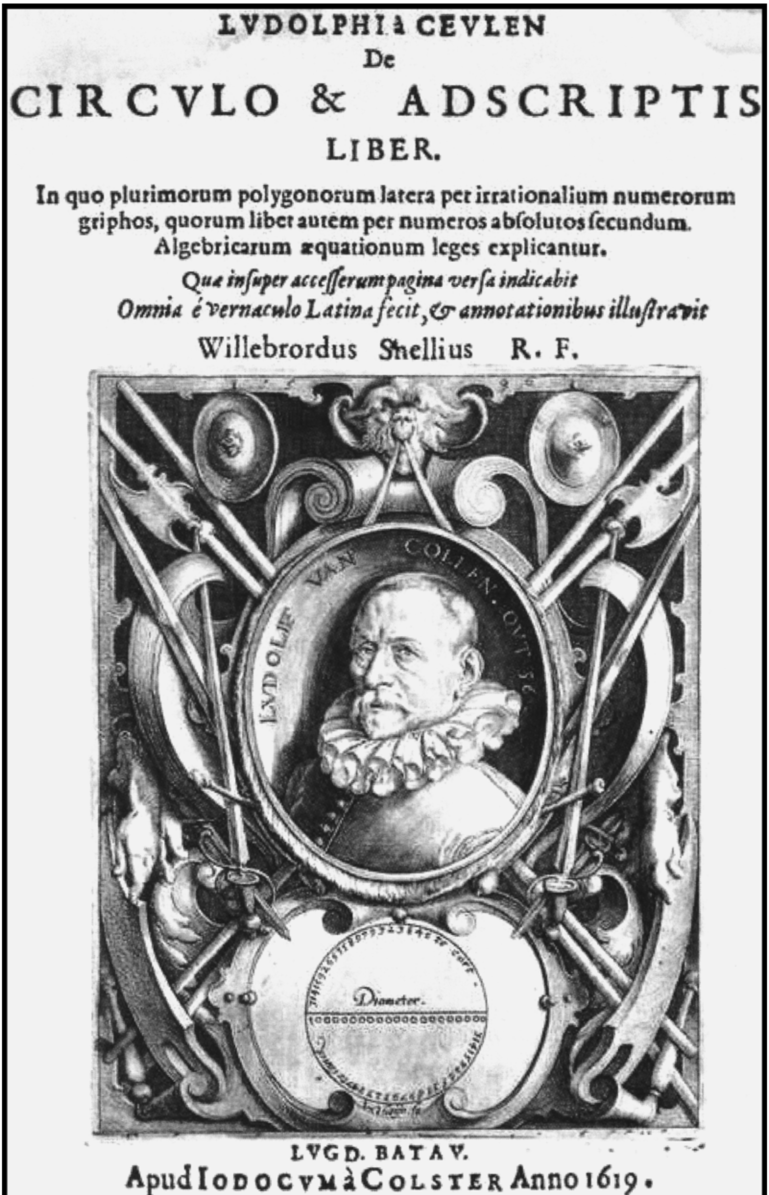
Meanwhile, mathematicians from several European countries were using similar methods to find better estimates for π – by continually doubling the number of sides of the polygons and calculating the appropriate perimeters or areas.

In Italy, in his *Practica Geometriae* of 1220, Leonardo of Pisa (better known to us as Fibonacci) cited earlier calculations and used polygons with 96 sides to find the approximation $\pi = 3.141818$.

In 1579 the French mathematician and lawyer François Viète used a polygon with $3 \times 2^{17} = 393{,}216$ sides to approximate π to nine decimal places.

Six years earlier, the German mathematician and astronomer Valentin Otto (or Otho) proposed the fraction $\frac{355}{113}$; as we saw earlier, this approximates π to six decimal places and had already been known to Zu Chongzhi and his son a thousand years previously. The Dutch cartographer Adriaan Anthonisz coincidentally obtained the same value in 1585, having found the lower and upper estimates $\frac{355}{106}$ and $\frac{377}{120}$ and averaged their numerators and denominators.

In the Netherlands in 1593 Adriaan van Roomen used polygons with 2^{30} = 1,073,741,824 sides to approximate π to 15 decimal places. Subsequent efforts by Ludolph van Ceulen resulted in his approximating π to 20 decimal places (see Figure 41) and later, using polygons with $2^{62} = 4{,}611{,}686{,}018{,}427{,}387{,}904$ sides, to 35 decimal places. He apparently asked for this latter value to appear on his tombstone in the St Pieterskerk in Leiden (a replica can still be seen there), and for many years π was known in Germany as the *Ludolphian number*.



LVDOLPHI à CEVLEN
De
CIRCVLO & ADSCRIPTIS
LIBER.
In quo plurimorum polygonorum latera per irrationalium numerorum griphos, quorum liber autem per numeros absolutos secundum. Algebricarum æquationum leges explicantur.
Quæ insuper accesserum pagina versa indicabit
Omnia é vernaculo Latina fecit, & annotationibus illustravit
Willebrordus Snellius R. F.

LVGD. BATAV.
Apud IODOCVM à COLSTER Anno 1619.

Fig. 41. Ludolph van Ceulen, with his estimates for π to 20 decimal places

One of the last European attempts to use inscribed and circumscribed polygons was in 1630, when the Austrian astronomer Christoph Grienberger approximated π to 38 decimal places.

Radian measure

The number π also provides the basis for a natural way of measuring angles. Rather than measuring them in degrees, mathematicians usually prefer to use radians, where a *radian* is the angle θ (about 57.3°) subtended at the centre of a circle by a circular arc whose length is the radius of the circle (see Figure 42).

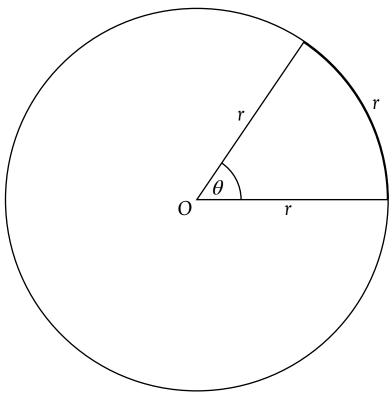


Fig. 42. Radian measure

More generally, for any angle the number of radians is the ratio of the length of the arc to the radius of the circle. So an angle of 180°, subtended by a semicircular arc, is π radians, and we can draw up a list of radian equivalents for certain well-known angles:

degrees	0°	30°	45°	60°	90°	120°	135°	180°	240°	270°	360°
radians	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$3\pi/4$	π	$4\pi/3$	$3\pi/2$	2π

The credit for introducing radian measure is usually ascribed to the English mathematician Roger Cotes, the first Plumian Professor of Astronomy and Experimental Philosophy in the University of Cambridge. Born in 1682 he worked closely with Isaac Newton on the second edition of the *Principia Mathematica*. After his early death from a violent fever in 1716, his somewhat disorganized results on a range of mathematical topics were collected together and edited with commentary by his cousin and Plumian successor Robert Smith in the form of a book, *Harmonia Mensuarum* (Harmony of Measures). Smith's commentary includes the impressive calculation

$$1 \text{ radian} = 57.2957795130 \text{ degrees.}$$

We'll meet Cotes's work again in Chapter 6 in connection with Euler's equation.

Why use radians? One advantage is that a measure that's defined as the ratio of the length of a circular arc to the radius is less arbitrary than choosing a particular number like 360 as the number of degrees in a full circle. Another is that many mathematical results become much simpler to state when we use them, as we'll see.

Infinite expressions

Up to now, most results involving π have been estimates of its value. A new approach was taken by François Viète, who obtained an exact expression for π involving the product of infinitely many terms. Shortly after this, another infinite product was discovered by John Wallis, the Savilian Professor of Geometry at the University of Oxford.

Viète's infinite product

In 1579, and again in 1593 in his *Variorum de Rebus Mathematicis Reponsorum, Liber VII* (Book 7 of the Varied Mathematical Responses), Viète showed that

$$\begin{aligned} 2/\pi &= \cos \pi/4 \times \cos \pi/8 \times \cos \pi/16 \times \cos \pi/32 \times \ldots \\ &= \frac{1}{2}\sqrt{2} \times \frac{1}{2}\sqrt{(2+\sqrt{2})} \times \frac{1}{2}\sqrt{(2+\sqrt{(2+\sqrt{2})})} \times \frac{1}{2}\sqrt{(2+\sqrt{(2+\sqrt{(2+\sqrt{2})})})} \times \ldots . \end{aligned}$$

By taking the first few terms in this product we can obtain successively better approximations for π, although Viète's result is awkward to use because of all the square roots. Details of Viète's infinite product are given in Box 5 for those with the appropriate mathematical background.

Box 5: Viète's infinite product

We use the result that the area of a triangle with sides a and b that enclose an angle α is $\frac{1}{2}ab\sin\alpha$, and we consider regular polygons inscribed in a circle of radius r.

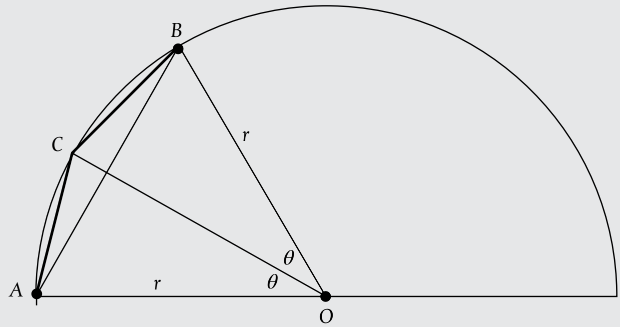


Fig. 43.

If $A(n)$ denotes the area inside a polygon with n sides (see Figure 43), then

$$A(n) = n \times \text{area } AOB = n \times \tfrac{1}{2} r^2 \sin 2\theta = nr^2 \sin\theta \cos\theta,$$

because $\sin 2\theta = 2\sin\theta\cos\theta$.

Also,

$$A(2n) = 2n \times \text{area } AOC = 2n \times \tfrac{1}{2} r^2 \sin\theta = nr^2 \sin\theta,$$

and dividing these two results gives

$$A(n)/A(2n) = \cos\theta.$$

Similarly, $A(2n)/A(4n) = \cos\theta/2$ and $A(4n)/A(8n) = \cos\theta/4$.

Multiplying these together gives

$$A(n)/A(8n) = [A(n)/A(2n)] \times [A(2n)/A(4n)] \times [A(4n)/A(8n)]$$
$$= \cos\theta \times \cos\theta/2 \times \cos\theta/4.$$

Similarly, after k steps we have

$$A(n)/A(2^k n) = \cos\theta \times \cos\theta/2 \times \ldots \times \cos\theta/2^k.$$

But as k becomes large, the area $A(2^k n)$ approaches πr^2, the area of the circle, so

$$A(n) = \pi r^2 \times \cos\theta \times \cos\theta/2 \times \cos\theta/4 \times \cos\theta/8 \times \ldots.$$

If we now let $n = 4$, so $A(4) = 2r^2$, and let $\theta = \pi/4$ radians (45°), then

$$2r^2 = \pi r^2 \times \cos\pi/4 \times \cos\pi/8 \times \cos\pi/16 \times \cos\pi/32 \times \ldots.$$

It follows that $2/\pi = \cos\pi/4 \times \cos\pi/8 \times \cos\pi/16 \times \cos\pi/32 \ldots$, as required.

The result with all the √2s arises because

$$\cos\pi/4 = \sqrt{2},\ \cos\pi/8 = \sqrt{(2+\sqrt{2})}, \ldots.$$

Wallis's infinite product

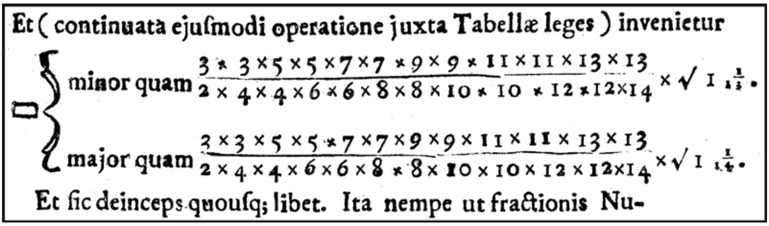
In 1656 Wallis obtained a different type of infinite product, which appeared in his book *Arithmetica Infinitorum* and which seems to have nothing to do with circular measure. We can write it as

$$\frac{4}{\pi} = \frac{3\times3\times5\times5\times7\times7\times9\times9\times\ldots}{2\times4\times4\times6\times6\times8\times8\times10\times\ldots}.$$

Again, we can find approximations for π by stopping after a few terms; for example, taking just the first eight terms shown above, we get the approximation

$$\frac{4}{\pi} = \frac{3\times3\times5\times5\times7\times7\times9\times9}{2\times4\times4\times6\times6\times8\times8\times10} \approx 1.21122,$$

giving $\pi \approx 3.302$. Wallis gave his result as upper and lower estimates for $4/\pi$, which he wrote as □ (see Figure 44). But although such approximations are of great theoretical significance, they converge very slowly to π and have no practical value.



Et (continuata ejusmodi operatione juxta Tabellæ leges) invenietur

$$\square \begin{cases} \text{minor quam } \dfrac{3\times3\times5\times5\times7\times7\times9\times9\times11\times11\times13\times13}{2\times4\times4\times6\times6\times8\times8\times10\times10\times12\times12\times14}\times\sqrt{1\tfrac{1}{13}}. \\[2ex] \text{major quam } \dfrac{3\times3\times5\times5\times7\times7\times9\times9\times11\times11\times13\times13}{2\times4\times4\times6\times6\times8\times8\times10\times10\times12\times12\times14}\times\sqrt{1\tfrac{1}{14}}. \end{cases}$$

Et sic deinceps quousq; libet. Ita nempe ut fractionis Nu-

Fig. 44. Wallis's estimates for π

Continued fractions

Consider the following fractions, which are all approximations to π:

$$\frac{22}{7}=3.1428571\ldots,\quad \frac{355}{113}=3.1415929\ldots,\quad \frac{103993}{33102}=3.1415926\ldots.$$

We can rewrite these in the following forms:

$$\frac{22}{7}=3+\frac{1}{7};\ \frac{355}{113}=3+\frac{16}{113}=3+\frac{1}{7+1/16}$$

and

$$\frac{103993}{33102}=3+\cfrac{1}{7+\cfrac{1}{15+\cfrac{1}{1+1/292}}}$$

Conversely, to simplify such expressions we start at the end and work backwards.

These expressions are called *continued fractions*, a term coined by Wallis who introduced them in 1655. Wallis showed his infinite product for $4/\pi$ to his friend William Brouncker, later the first President of the Royal Society, and Brouncker apparently used it (though no one knows how) to obtain the following infinite continued fraction:

$$\frac{4}{\pi} = 1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{7^2}{2+\cfrac{9^2}{2+\dots}}}}}$$

The numbers in this expression form a regular pattern, with 2s appearing as denominators and the odd perfect squares $1^2 = 1$, $3^2 = 9$, $5^2 = 25$, $7^2 = 49, \dots$, appearing as numerators.

By breaking off Brouncker's continued fraction at various places, we can obtain estimates for $4/\pi$, and hence for π, as follows:

the first estimate is $4/\pi \approx 1+1^2/2 = 1+\frac{1}{2} = \frac{3}{2}$, giving $\pi \approx 2.667$;

the second is $4/\pi \approx 1+1/\left(2+\frac{9}{2}\right) = 1+1/\left(\frac{13}{2}\right) = 1+\left(\frac{2}{13}\right) = \left(\frac{15}{13}\right)$, giving

$\pi \approx 3.467$;

the third is $4/\pi \approx 1+1/\left(2+3^2/\left(2+\frac{25}{2}\right)\right) = 1+1/\left(2+9/\frac{29}{2}\right)$

$= 1+1/\left(2+\frac{18}{19}\right) = 1+1/\frac{76}{29} = 1+\frac{29}{76} = \frac{105}{76}$, giving $\pi \approx 2.895$;

and the next two turn out to be $4/\pi \approx \frac{945}{789}$ and $\frac{30936}{10395}$,

giving $\pi \approx 3.340$ and $\pi \approx 2.976$.

As with Wallis's result, these approximations to π are of theoretical interest, but they converge very slowly and have no practical use.

Somewhat better was another continued fraction expression with a similar format, included by Wallis in his *Arithmetica Infinitorum* but possibly due to Brouncker. It later reappeared in a paper 'On the continued fractions of Wallis' by Leonhard Euler:

$$\pi = 3+\cfrac{1}{6+\cfrac{3^2}{6+\cfrac{5^2}{6+\cfrac{7^2}{6+\dots}}}}$$

Here the first few estimates for π are

$\frac{19}{6} = 3.167$, $\frac{47}{15} = 3.133$, $\frac{1321}{420} = 3.1452$, and $\frac{910}{289} = 3.1488$, which converge more quickly to π.

A much improved continued fraction for π was given by the Swiss mathematician Johann Heinrich Lambert in 1767:

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292 + \cfrac{1}{1 + \cfrac{1}{1 + \dots}}}}}}$$

Here the numerators are all 1, but the denominators don't seem to follow a regular pattern. However, the approximations converge much more quickly than before, and some of the resulting estimates for π are ones that we've seen earlier:

the first is $3 + \frac{1}{7}$ or $\frac{22}{7}$ (Archimedes' value, giving the first two decimal places),

the second is $3 + 1/\left(7 + \frac{1}{15}\right) = 3 + 1/\frac{106}{15} = \frac{333}{106}$ (giving 3.1415),

the third is $\frac{355}{113}$ (Zu Chongzhi and Otho's value, giving the first six decimal places),

and the fourth is $\frac{103993}{33102}$ (giving the first ten decimal places).

Arctan formulas

A new and highly productive method for estimating π, which came to be used extensively throughout the 18th and 19th centuries, involved the *inverse tangent function*, usually denoted by $\arctan x$ or $\tan^{-1} x$. If $y = \tan x$, then $x = \arctan y$; for example,

$$\tan \pi/4 = 1, \text{ so } \arctan 1 = \pi/4 \quad \text{and} \quad \tan \pi/6 = 1/\sqrt{3}, \text{ so } \arctan 1/\sqrt{3} = \pi/6.$$

(Note that $\tan 9\pi/4 = 1$ also, so $\arctan y$ has many different values, all differing by multiples of 2π; this will be important later, but for now we'll stick to the principal value.)

Arising from the addition formula for the tangent function is a useful result for arctan, obtained by writing $u = \tan x$ and $v = \tan y$:

$$\tan(x+y) = \frac{\tan x + \tan y}{1 - \tan x \tan y} \text{ and so } \arctan u + \arctan v = \arctan\frac{u+v}{1-uv}.$$

In particular, the following result is among many that we can prove:

$$\arctan \tfrac{1}{2} + \arctan \tfrac{1}{3} = \arctan 1 = \pi/4.$$

This is because the angle α in Figure 45 is $\arctan\frac{1}{2}$ (from the triangle *CDE*), the angle β is $\arctan\frac{1}{3}$ (from a classical result that the medians *CE* and *DF* of the triangle *ACD* divide each other in the ratio 1:3), and the angle $\alpha + \beta$ is $45° = \pi/4$ (from the triangle *BCD*).

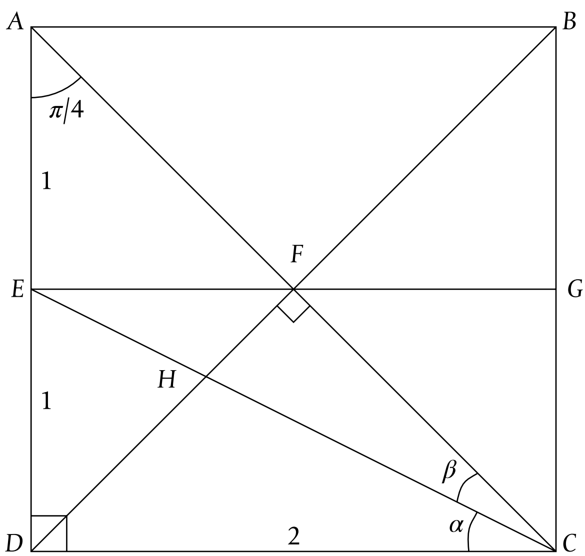


Fig. 45. Combining two arctans

Many functions can be written as infinite series. For example, because

$$(1+x^2) \times (1 - x^2 + x^4 - x^6 + x^8 - \ldots) = 1,$$

as we can check by multiplying out and cancelling terms, leaving just 1, we can write

$$\left(1+x^2\right)^{-1}=1-x^2+x^4-x^6+x^8-\ldots.$$

We can similarly express $\arctan x$ as an infinite series:

$$\arctan x = x-\frac{1}{3}x^3+\frac{1}{5}x^5-\frac{1}{7}x^7+\frac{1}{9}x^9-\ldots;$$

this result was already known to Mādhava in 14th-century India, but is usually named after the Scotsman James Gregory, who rediscovered it 300 years later. (Those familiar with the integral calculus can obtain this series for $\arctan x$ by integrating the series for $\left(1+x^2\right)^{-1}$ term by term from 0 to 1.)

If we now substitute $x = 1$ into the series for $\arctan x$, we get

$$\pi/4=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\ldots,$$

a result also due to Mādhava, but usually credited to Gottfried Leibniz. This last result is quite remarkable: by simply adding and subtracting reciprocals of the form $1/n$ we get a result involving the circular number π.

Unfortunately the Leibniz series converges exceedingly slowly, and so cannot be used to find the value of π in practice; for example, the first 300 terms of the series give π correct to only two decimal places, while the first 500,000 terms give only five correct decimal places. But fortunately a remedy is not hard to find, as was discovered by the Yorkshireman Abraham Sharp, a friend of Isaac Newton and of the first Astronomer Royal John Flamsteed. In 1699, instead of substituting $x = 1$ in the above series, he put $x=1/\sqrt{3}$, to give

$$\begin{aligned}\pi/6&=\arctan(1/\sqrt{3})=1/\sqrt{3}-\tfrac{1}{3}(1/\sqrt{3})^3+\tfrac{1}{5}(1\sqrt{3})^5-\tfrac{1}{7}(1/\sqrt{3})^7+\ldots\\&=1/\sqrt{3}\times\left\{1-1/(3\times3)+1/\left(3^2\times5\right)-1/\left(3^3\times7\right)+1/\left(3^4\times9\right)-\ldots\right\}.\end{aligned}$$

This series converges more quickly than before because of the increasing powers of 3 in the denominator, and Sharp was able to find π to no fewer than 72 decimal places, a dramatic improvement on earlier estimates.

We can also obtain estimates for π by using arctan addition formulas, such as the one illustrated in Figure 45:

$$\pi/4 = \arctan \tfrac{1}{2} + \arctan \tfrac{1}{3}.$$

If we now substitute $x = \frac{1}{2}$ and $x = \frac{1}{3}$ into the series for $\arctan x$, we have

$$\begin{aligned}\pi/4 &= \left\{\frac{1}{2} - \frac{1}{3}\left(\frac{1}{2}\right)^3 + \frac{1}{5}\left(\frac{1}{2}\right)^5 - \frac{1}{7}\left(\frac{1}{2}\right)^7 + \ldots\right\} + \left\{\frac{1}{3} - \left(\frac{1}{3}\right)^3 + \frac{1}{5}\left(\frac{1}{3}\right)^5 - \frac{1}{7}\left(\frac{1}{3}\right)^7 + \ldots\right\} \\ &= \left\{\frac{1}{2} - 1/\left(2^3 \times 3\right) + 1/\left(2^5 \times 5\right) - 1/\left(2^7 \times 7\right) + \ldots\right\} \\ &\quad + \left\{\frac{1}{3} - 1/\left(3^3 \times 3\right) + 1/\left(3^5 \times 5\right) - 1/\left(3^7 \times 7\right) + \ldots\right\}.\end{aligned}$$

These two series converge reasonably quickly because of the increasing powers of 2 and 3 in the denominators and yield good estimates for π. Indeed, in 1861 a certain W. Lehmann of Potsdam used these very series to find π to 261 decimal places.

From here the search was on to find new arctan identities where the series converge even faster. In 1706 the Englishman John Machin used the addition formula several times over to prove that

$$\pi/4 = 4 \arctan \tfrac{1}{5} - \arctan \tfrac{1}{239}.$$

Writing out the series for arctan x with $x = \frac{1}{5}$ and $x = \frac{1}{239}$, he obtained

$$\begin{aligned}\pi/4 = 4&\left\{\frac{1}{5} - \frac{1}{3}\left(\frac{1}{5}\right)^3 + \frac{1}{5}\left(\frac{1}{5}\right)^5 - \frac{1}{7}\left(\frac{1}{5}\right)^7 + \ldots\right\} \\ &- \left\{\frac{1}{239} - \frac{1}{3}\left(\frac{1}{239}\right)^3 + \frac{1}{5}\left(\frac{1}{239}\right)^5 - \frac{1}{7}\left(\frac{1}{239}\right)^7 + \ldots\right\}.\end{aligned}$$

These arctan series converge rapidly because of the powers of 5 and 239 in the denominators – for example, using only the first three terms in each bracket already gives the approximation 3.14. Furthermore, 5 is an easy number to work with and Machin was able to calculate π by hand to 100 decimal places, an improvement on anything that had gone before. William Jones, who had introduced the symbol π in the same year, was highly impressed by 'the Truly Ingenious Mr. John Machin' and reproduced the 100 digits of π in his book (see Figure 46).

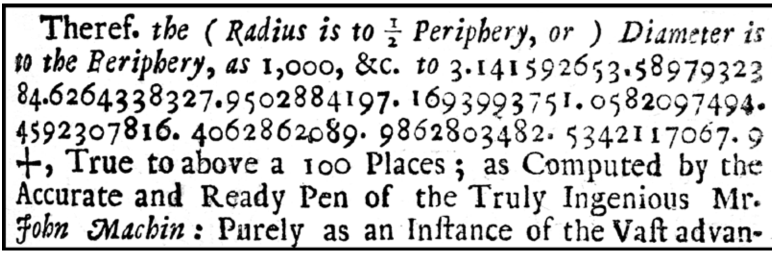

Theref. *the (Radius is to* ½ *Periphery, or) Diameter is to the Periphery, as* 1,000, &c. *to* 3.141592653.58979323
84.6264338327.9502884197.1693993751.0582097494.
4592307816. 4062862089. 9862803482. 5342117067. 9
+, True to above a 100 Places; as Computed by the
Accurate and Ready Pen of the Truly Ingenious Mr.
John Machin : Purely as an Inſtance of the Vaſt advan-

Fig. 46. William Jones praises Machin's achievement

This arctan approach to finding improved estimates for π was developed further by several others. In 1755 Euler used the equation

$$\pi = 20 \arctan \tfrac{1}{7} + 8 \arctan \tfrac{3}{79}$$

to calculate 20 decimal places of π in one hour. He also found many other arctan results, including the equation

$$\pi / 4 = 4 \arctan \tfrac{1}{5} - \arctan \tfrac{1}{70} + \arctan \tfrac{1}{99},$$

which was subsequently used by the Englishman William Rutherford in 1841 to calculate π to 152 decimal places.

In 1794 the Slovenian Jurij (or Georg) Vega used Euler's equation,

$$\pi = 20 \arctan \tfrac{1}{7} + 8 \arctan \tfrac{3}{79},$$

to find π to 136 decimal places, and for many years this was the most accurate value known. But there were persistent references in the literature to an earlier and more accurate value that the Hungarian Baron Franz Xaver von Zach had noticed while visiting Oxford's Bodleian Library in the 1780s. This reference was eventually located in 2014 by Benjamin Wardhaugh and confirmed that in 1721 a Philadelphia resident had used Sharp's arctan series in the form

$$\pi = 6 \arctan\left(1/\sqrt{3}\right) = \sqrt{12} - \frac{\sqrt{12}}{3\times3} + \frac{\sqrt{12}}{9\times5} - \frac{\sqrt{12}}{27\times7} + \ldots.$$

The unknown author, who never published the result, determined $\sqrt{12}$ to 154 decimal places and calculated no fewer than 314 terms of this arctan series, obtaining π to 154 decimal places with only the last two incorrect. This was indeed the world's most accurate value of π for over 100 years, even though it had been largely unknown.

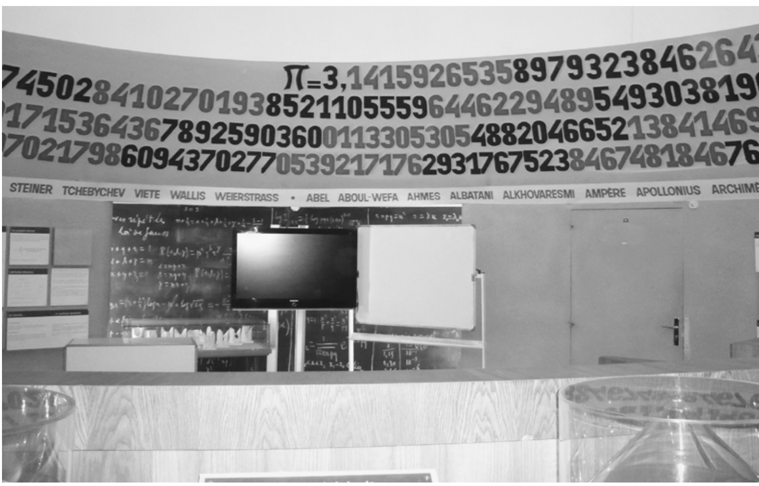


Fig. 47. The Palais de la Découverte in Paris

Most notorious of all, in 1873 William Shanks used Machin's formula to calculate π to an impressive 707 decimal places; these were later inscribed in a ceiling frieze in the Palais de la Découverte in Paris where they can still be seen (see Figure 47). Unfortunately for him, and for the Palais, it was subsequently discovered that only the first 527 of these decimal places (shown here) were correct:

3.14159 26535 89793 23846 26433 83279 50288 41971 69399 37510 58209
74944 59230 78164 06286 20899 86280 34825 34211 70679 82148 08651
32823 06647 09384 46095 50582 23172 53594 08128 48111 74502 84102
70193 85211 05559 64462 29489 54930 38196 44288 10975 66593 34461

28475 64823 37867 83165 27120 19091 45648 56692 34603 48610 45432 66482 13393 60726 02491 41273 72458 70066 06315 58817 48815 20920 96282 92540 91715 36436 78925 90360 01133 05305 48820 46652 13841 46951 94151 16094 33057 27036 57595 91953 09218 61173 81932 61179 31051 18548 07446 23799 62749 56735 18857 52724 89122 79381 83011 94912 98336 73362 44065 66430 86021 39...

A miscellany of results

The number π turns up in many unexpected places. Here are a few of them.

Some results of Euler

We've seen how Leonhard Euler was the first to popularize the use of the letter π for circle measurement and how he used an arctan series to estimate it. We've also met his continued fraction expansion. But π features in many other places in his writings.

Earlier we saw that π turns up as the sum of various infinite series, such as the Leibniz series, which can be written as

$$\pi = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \ldots.$$

But this isn't a unique phenomenon – many more infinite series have sums that involve π for no apparent reason.

One of the best-known of these is Euler's celebrated solution of the so-called *Basel problem,* which asked for the exact sum of the infinite series

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \ldots,$$

where the denominators are the perfect squares $1^2 = 1$, $2^2 = 4$, $3^2 = 9, \ldots$. To his great surprise Euler discovered that the required sum is $\pi^2/6$, exclaiming:

> Quite unexpectedly I have found an elegant formula involving the quadrature of the circle.

He next summed the series whose denominators are the fourth powers, giving

$$\frac{1}{1}+\frac{1}{16}+\frac{1}{81}+\frac{1}{256}+\frac{1}{625}+\frac{1}{1296}+\ \ldots = \pi^4/90,$$

and the sixth powers, giving

$$\frac{1}{1}+\frac{1}{64}+\frac{1}{729}+\frac{1}{256}\ \frac{1}{4096}+\frac{1}{15625}+\frac{1}{16656}+\ \ldots = \pi^6/945.$$

Continuing in this way, he correctly summed the reciprocals of all the even powers up to the 26th power. In each case, the answer involves the corresponding power of π. (This is not true for the odd powers.)

Another result of Euler's, related to the Basel problem, involves the prime numbers 2, 3, 5, 7, 11,.... If, for each prime number p, we calculate the quotient $p^2/(p^2-1)$ and then multiply the results together, then π again appears out of the woodwork:

$$2^2/\left(2^2-1\right)\times 3^2/\left(3^2-1\right)\times 5^2/\left(5^2-1\right)\times 7^2/\left(7^2-1\right)\times 11^2/\left(11^2-1\right)\times\ldots$$
$$=\frac{4}{3}\times\frac{9}{8}\times\frac{25}{24}\times\frac{49}{48}\times\frac{121}{120}\times\ldots=\pi^2/6.$$

Yet another extraordinary product of Euler's involving π and the primes is the following:

$$\pi/2=\frac{3}{2}\times\frac{5}{6}\times\frac{7}{6}\times\frac{11}{10}\times\frac{13}{14}\times\frac{17}{18}\times\frac{19}{18}\times\ldots;$$

here the numerators are the odd primes and the denominators are the even numbers that are not multiples of 4.

Probabilistic results

Some results in probability involve π. Two celebrated examples are the following.

If we choose a whole number at random, then it may have a repeated prime factor; for example, $63 = 3 \times 3 \times 7$, with a repeated prime factor of 3. Or it may have no repeated prime factor, such as $105 = 3 \times 5 \times 7$. What's the probability that a number chosen at random has no repeated prime factor? The answer is $6/\pi^2$, which is about 0.608.

If we now choose two numbers at random, then they may have a factor in common; for example, 63 and 91 share a common factor of 7, since $63 = 7 \times 9$ and $91 = 7 \times 13$. Or they may have no common factor other than 1, such as 64 and 91. What's the probability that two randomly chosen whole numbers have no factors in common? Again, the answer is $6/\pi^2$.

Buffon's needle experiment

In 1777 an experimental method for finding π was introduced by Georges-Louis Leclerc, the Comte de Buffon. Suppose that you throw a large number of needles (or matchsticks) of length L onto a grid of parallel lines at a distance D apart, where $L < D$, and record the proportion of the needles that cross a line of the grid. The probability that a needle crosses a line can be shown to equal $2/\pi \times L/D$, from which a value for π can then be calculated. For example, in Figure 48, $L/D = 4/5$ and five of the ten needles cross lines; so $2/\pi \times 4/5 = 5/10$, giving an experimental value for π of $80/25 = 3.2$.

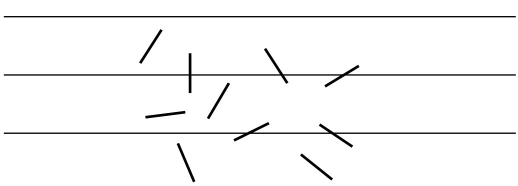


Fig. 48. Buffon's needle experiment

In 1901 an Italian mathematician called Mario Lazzerini carried out such a needle experiment in which $L/D = 5/6$, performing 3408 trials and claiming 1808 crossings. This leads to the estimate $\pi = 355/113$ which, as we have seen, gives π to six decimal places. He was lucky. If only one of his needles had landed differently, then his value for π would have been correct to only two decimal places.

Gauss's circle problem

Another problem in which π makes an unexpected appearance arises from number theory and is credited to Carl Friedrich Gauss. Let $s(n)$ be the number of different ways in which the whole number n can be written as the sum of two perfect squares: here, both positive and negative squares are allowed, and it matters in which order the squares appear. For example, $s(5) = 8$, corresponding to the eight sums

$$5 = 2^2 + 1^2 = 1^2 + 2^2 = (-2)^2 + 1^2 = 1^2 + (-2)^2$$
$$= 2^2 + (-1)^2 = (-1)^2 + 2^2 = (-2)^2 + (-1)^2 = (-1)^2 + (-2)^2.$$

Notice that $s(n)$ is the number of points with integer coordinates that lie on the circle with equation $x^2 + y^2 = n$ (see Figure 49).

The behaviour of $s(n)$ is very erratic. For example, $s(7) = 0$ and $s(8) = 4$, while $s(250) = 16$ and $s(251) = 0$. To smooth out the variation, we can look at the average of the values over the first n numbers – namely,

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

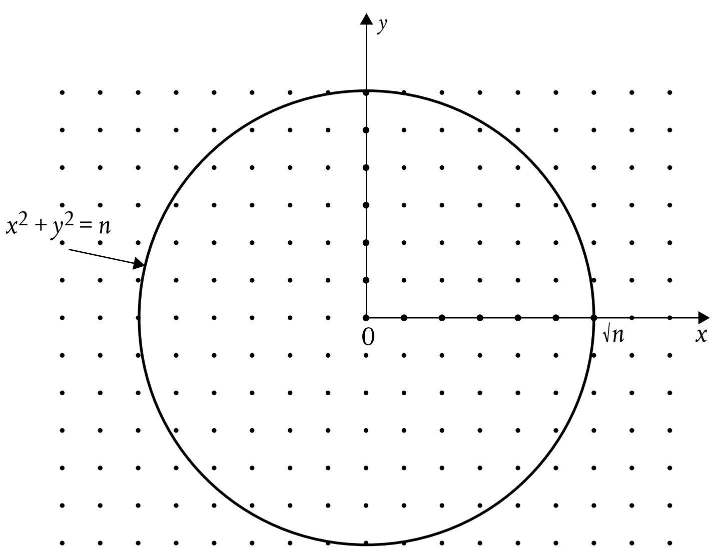


Fig. 49. Gauss's circle problem

What is the behaviour of this average as n becomes arbitrarily large? The answer is that it tends to a limit, and this limit is π.

π *is irrational*

Is π irrational, or can it be written as a fraction a/b, where a and b are whole numbers? Although everyone expected the former outcome, finding a proof turned out to be very difficult. It was not until 1767 that this was successfully achieved by Johann Heinrich Lambert, who showed that

If x is a rational number (other than 0), then tan x must be irrational.

Turning this around, it follows that if tan x is a rational number, then x must be irrational or 0. But $\tan \pi/4 = 1$, which is rational. So $\pi/4$, and hence π, must be irrational.

Sometimes a number is irrational but its square is rational, such as $\sqrt{2}$ whose square is 2. But this doesn't happen here – in 1794 the French mathematician Adrien-Marie Legendre proved that π^2 is also irrational.

In Chapter 6 we'll visit the even harder problem of determining whether π is transcendental.

Legislating for π

In 1897 a bizarre event took place in the State of Indiana, USA, where the House of Representatives considered and unanimously passed 'A bill introducing a new Mathematical Truth'. This House Bill No. 246 attempted to legislate an incorrect value for π provided by a local physician, Edwin J. Goodwin, M.D. of Solitude, Posey County, who would then allow the State to use his value free of charge but would expect royalties to be paid to him by anyone else who employed it:

> A bill for an act introducing a new mathematical truth and offered as a contribution to education to be used only in the State of Indiana free of cost by paying any royalties whatever on the same.

According to Dr Goodwin,

The ratio of the diameter and circumference is as five-fourths to four.

This yields a value for π of 3.2.

For some reason the bill was passed on to the House Committee on Canals or Swamp Lands, who then passed it on to the Committee on Education.

Be it enacted by the General Assembly of the State of Indiana: it has been found that a circular area is to the quadrant of the circumference, as the area of an equilateral rectangle is to the square on one side. The diameter employed as the linear unit according to the present rule in computing the circle's area is entirely wrong....

This clearly makes little sense, but even so it then went on to scrutiny by the Committee on Temperance (!), who recommended its passage. Fortunately, a local professor of mathematics, C. A. Waldo of Purdue University, happened to be visiting the statehouse when the bill was about to be finally ratified, and he managed to persuade the senators to stop it just in time. As far as we know, it is still with the Committee on Temperance....

Some weird results

The 20th century saw a number of surprising discoveries about π. Here's a small selection.

Around 1913 the Indian mathematician Srinivasa Ramanujan found several bizarre approximations to π, including

$$\pi \approx \tfrac{63}{25}\left(17+15\sqrt{5}\right)/\left(7+15\sqrt{5}\right),$$

which is correct to nine decimal places, and

$$\pi \approx \sqrt[4]{\left(9^2+19^2/22\right)},$$

which is correct to eleven decimal places.

In the following year Ramanujan wrote a remarkable paper, 'Modular equations and approximations to π', in which he presented many extraordinary exact formulas for $1/\pi$, such as

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n)!}{(n!)^4} \times \frac{1103 + 26390n}{396^{4n}}.$$

This series converges extremely rapidly and forms the basis of some of today's fastest algorithms for calculating π. Many years later, in 1989, David and Gregory Chudnovsky produced a similar, but even more complicated, result:

$$\frac{1}{\pi} = 12 \sum_{n=0}^{\infty} (-1)^n \frac{(6n)!}{(n!)^3 (3n)!} \times \frac{13591409 + 545140134n}{(640320)^{n+1/2}}$$

A different type of result was discovered in 1995 by David Bailey, Peter Borwein, and Simon Plouffe, and caused a great deal of surprise:

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16^n} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right).$$

The importance of this result is that, if we work in a base-16 number system, rather than in base 10, we can successively calculate each successive digit of π without having to calculate all the preceding ones first.

Earlier we saw that the Leibniz series

$$\pi / 4 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots$$

converges extremely slowly. If we calculate the first 500,000 terms of this series to forty decimal places and ignore the rest, then we get the following approximation to π, in which the first six digits, 3.14159, are correct:

$$\pi \approx 3.14159\underline{0}6535897932\underline{40}4626433832\underline{6}9502884197.$$

Because the next digit is incorrect, we might expect most of the later digits to be erroneous as well but this is not the case. Surprisingly, as Roy North of Colorado noticed in 1989, all but these four underlined digits turn out to be correct, and similar results hold when more digits of π are taken.

Enter the computer

As soon as desk calculators and computers entered the scene, it became possible to use an arctan series to calculate π to a much greater accuracy.

The first advance was in 1949 when John Wrench and L. R. Smith, using Machin's arctan result, put the 18,800 electron tubes of ENIAC (the Electronic Numerical Integrator And Computer at the U.S. Army's Ballistic Research Laboratories in Maryland) to good use to calculate π to 2037 places in 70 hours. Machin's result was also used by S. C. Nicholson and J. Jeenel in 1955 to find π to 3089 decimal places in just 13 minutes on the Naval Ordnance Research Calculator NORC. Meanwhile, progress was being made in England: in 1957 George E. Felton used a different arctan series to calculate π to 10,021 decimal places in 33 hours on the Ferranti PEGASUS computer, though not all were correct.

At this stage IBM (International Business Machines Corporation) entered the scene. In 1958 François Genuys used an IBM 704 computer in Paris to produce 10,000 decimal places in 100 minutes, and in 1959 they improved this to 16,167 places, calculated in 4.3 hours. Then, in 1961, using IBM 7090 computers, J. M. Gerard of London found 20,000 places in 39 minutes, while Daniel Shanks and John Wrench in New York obtained 100,000 places in 8.7 hours.

The first people to calculate one million decimal places were Jean Guillard and Martin Bouyer, who in 1973 achieved this target on a CDC 7600 machine in 23.3 hours. The scene then moved to Japan, where Yoshiaki Tamura, Yasumada Kanada, and others gradually pushed the number of places up to 10 million in 1983, 100 million in 1987, and 538 million in 1989. Using very sophisticated arctan formulas, Kanada and his team carried out their calculations in base 16, where the individual digits of π could be calculated one at a time (see the result of Bailey, Borwein, and Plouffe above), before translating their conclusions to base 10.

Meanwhile, in New York City, the Chudnovsky brothers were developing algorithms to use on their home-built supercomputers to push the numbers even higher, and in 1989 they were the first to exceed one billion places. There was then a frantic race between them and the Japanese group, with a trillion places being achieved in 2002 and 10 trillion places in 2011. Since then the number of calculated places has increased to over 20 trillion.

Why bother?

Why do we need to calculate so many digits? After all, just a handful of decimal places is sufficient for most everyday uses – for example, 3.14 is correct to about 0.5 per cent, and 3.1416 is within 0.0002 per cent of its true value. NASA's Jet Propulsion Laboratory in California uses only 15 places: even the calculations on their furthest spacecraft (some 12.5 billion miles away) are then correct to within a couple of inches. And just 39 decimal places (already known back in the 17th century) are sufficient to determine the circumference of a circle surrounding the visible universe (about 289 billion light years) to within the diameter of a hydrogen atom.

One reason for calculating so many digits is that the results can be used to test the speed and accuracy of a new computer. If you have a new machine you can ask it to print out the millionth (or the billionth or the trillionth) digit of π and see whether, and how quickly, it produces the correct answer.

Another reason is that the availability of such extensive lists enables us to search for patterns (if there are any) among the digits – for example, do all ten digits from 0 to 9 occur equally often in the expansion of π? The following list shows remarkable similarities in the occurrences of these digits among the first trillion digits of π and provides evidence for the answers to such questions:

0	99,999,485,134	5	99,999,671,008
1	99,999,945,664	6	99,999,807,503
2	100,000,480,057	7	99,999,818,723
3	99,999,787,805	8	100,000,791,469
4	100,000,357,857	9	99,999,854,780

π

Measuring the Earth

Let's end this chapter with a simple puzzle that appeared in 1702 in *The Elements of Euclid* by the Cambridge mathematician William Whiston. You may find its answer surprising.

The circumference of the Earth is about 25,000 miles (= 132 million feet). Assuming the Earth to be a perfect sphere, suppose that we tie a piece of string of this great length tightly around its equator. We then extend this string by just $2\pi(\approx 6.3)$ feet, and prop up the string equally all around the equator (see Figure 50). How high above the ground will the string be?

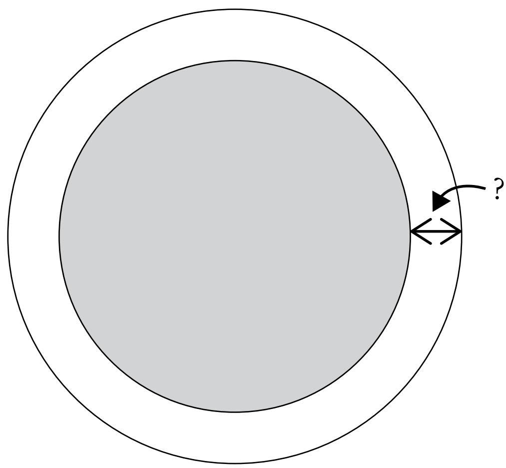
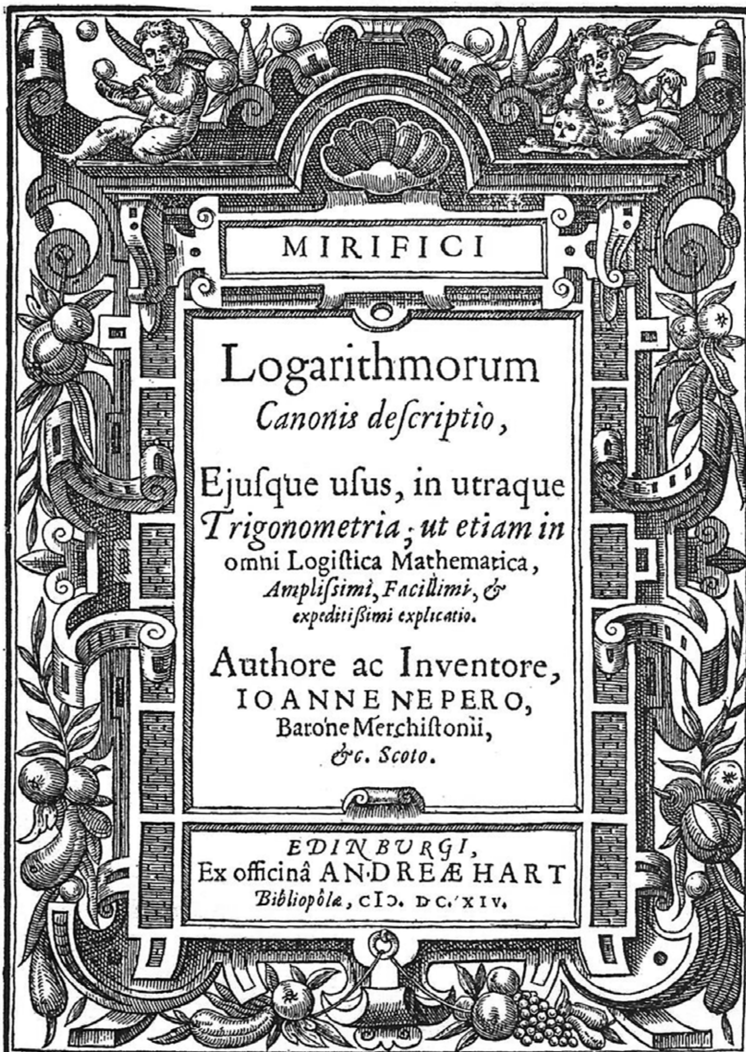


Fig. 50. Tying a string around the Earth

Most people think that the gap will be extremely small – perhaps a tiny fraction of an inch – but the answer is *one foot*! (In metric terms, the Earth's circumference is about 40,000 kilometres, and adding just 2 metres of string produces a gap of about 0.3 metres.)

In fact, we get the same answer whether we tie the string around the Earth, a tennis ball, or any other sphere. For if the sphere has radius r feet, then the original string has length $2\pi r$. When we extend the string by 2π feet, the new circumference is $2\pi r + 2\pi = 2\pi(r+1)$, so that the new radius is $r + 1$: that is, one foot more than before.


MIRIFICI

Logarithmorum
Canonis deſcriptio,

Ejuſque uſus, in utraque
Trigonometria; ut etiam in
omni Logiſtica Mathematica,
Amplißimi, Facillimi, &
expeditißimi explicatio.

Authore ac Inventore,
IOANNE NEPERO,
Barone Merchiſtonii,
&c. Scoto.

EDINBURGI,
Ex officinâ ANDREÆ HART
Bibliopôlæ, CIↃ. DC.XIV.

Fig. 51. The title page of John Napier's *Mirifici Logarithmorum,* 1614

CHAPTER 4

e

The exponential number

How fast do things grow? We often use the phrase 'exponential growth' to indicate something that grows very fast, but how quickly is this?

This chapter concerns the irrational number

$$e = 2.718281828459045235360287471352\ldots.$$

The symbol *e* was first used for this number by Leonhard Euler in an unpublished paper from around 1727, and in a letter of 1731. It first appeared in print in 1736 in Volume 1 of his two-part work *Mechanica* on the mathematics of motion.

As with π, various mnemonics have been devised for remembering the first few digits of *e*. To get the digits count the letters in the words of these sentences:

In Glasgow I lectured on geometry. (2.71828)
To disrupt a playroom is commonly a practice of children. (2.718281828)
We require a mnemonic to remember e whenever we scribble math. (2.7182818284)

and, if we allow the letter O to represent zero,

In showing a painting to probably a critical or venomous lady, anger dominates. O take guard, or she raves and shouts. (2.71828182845904523536)

Polynomial and exponential growth

To illustrate what is meant by 'exponential growth', let's start with a story concerning the invention of the game of chess (see Figure 52).

The wealthy king of a certain country was so impressed by the newly introduced game of chess that he offered the wise man who'd invented it any reward he wished. The wise man replied:

My prize is for you to give me 1 grain of wheat for the first square of the chessboard, 2 grains for the second square, 4 grains for the third square, and so on, doubling the number of grains on each successive square until the chessboard is filled.

The king was amazed to have been asked for such an insignificant reward (or so he believed), until his treasurers calculated the total number of grains of wheat to be

$$1 + 2 + 2^2 + 2^3 + 2^4 + \ldots + 2^{63}.$$

This works out at $2^{64} - 1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$ grains, enough wheat to form a pile the size of Mount Everest. Placed end to end they would reach to the nearest star, Alpha Centauri, and back again!

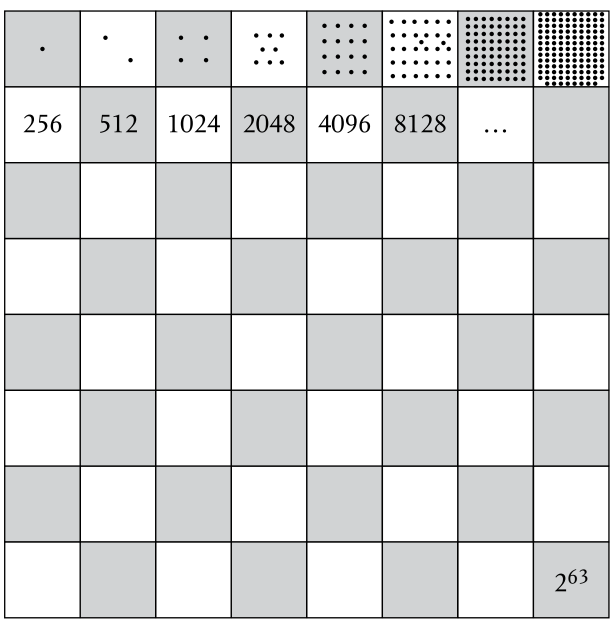


Fig. 52. Grains of wheat on a chessboard

Comparing types of growth

Let's see how quickly various sequences can grow.

A simple form of growth is *linear growth,* illustrated by the sequence of counting numbers

$$1, 2, 3, 4, 5, 6, \ldots;$$

–we refer to this sequence by its nth term, n.

Somewhat faster is *quadratic growth,* involving the perfect squares n^2:

$$1^2, 2^2, 3^2, 4^2, 5^2, 6^2, \ldots \text{ or } 1, 4, 9, 16, 25, 36, \ldots.$$

Even more rapid is *cubic growth,* involving the cubes n^3:

$$1^3, 2^3, 3^3, 4^3, 5^3, 6^3, \ldots \text{ or } 1, 8, 27, 64, 125, 216, \ldots$$

These are all examples of *polynomial growth,* since they involve powers of n.

Alternatively, we could look at powers of 2, or of any other number. As we saw in the chessboard story, the sequence 2^n of powers of 2 starts off fairly slowly –

$$1, 2, 4, 8, 16, 32, \ldots$$

– but it soon gathers pace because each successive term is twice the previous one.

The sequence 3^n of powers of 3 takes off more quickly:

$$1, 3, 9, 81, 243, 729, \ldots.$$

These are examples of *exponential growth,* where n is the exponent.

This distinction between polynomial growth and exponential growth was already recognized by Thomas Malthus in 1798. In his *Essay on Population* he contrasted the steady linear growth of food supplies with the exponential growth in population. He concluded that, however we may cope in the short term, the exponential growth would win in the long term, and that there would be severe food shortages – a conclusion that was borne out in practice.

To compare polynomial growth and exponential growth in greater detail, let's calculate the running times of various polynomials and exponentials when $n = 10$, 30, and 50, for a computer performing one million operations per second:

	$n = 10$	$n = 30$	$n = 50$
polynomial			
n	0.00001 seconds	0.00003 seconds	0.00005 seconds
n^2	0.0001 seconds	0.0009 seconds	0.0025 seconds
n^3	0.001 seconds	0.027 seconds	0.125 seconds
n^5	0.1 seconds	24.3 seconds	5.2 minutes
exponential			
2^n	0.001 seconds	17.9 minutes	35.7 years
3^n	0.059 seconds	6.5 years	2.3×10^{10} years

Clearly, exponential growth is generally much greater than polynomial growth: algorithms that run in polynomial time are generally thought to be 'efficient', whereas those that run in exponential time normally take much longer to implement as the input size increases, and are considered 'inefficient'.

Introducing logarithms

The exponential number *e* is intimately linked with logarithms, so before going further let's delve into the nature and history of these.

Logarithms to base 2

Early ideas of logarithms (the word means 'ratio-numbers') appeared around the year 1500, when Nicolas Chuquet of France and Michael Stifel of Germany explained how to turn certain calculations involving multiplication and division into simpler ones involving addition and subtraction. To illustrate this they listed the first few powers of 2,

n	1	2	3	4	5	6	7	8	9	10	11	12
2^n	2	4	8	16	32	64	128	256	512	1024	2048	4096

and observed that

To multiply powers of 2*, we add their exponents.*

For example, to multiply $16 = 2^4$ and $128 = 2^7$ we write

$$16 \times 128 = 2^4 \times 2^7 = 2^{4+7} = 2^{11} = 2048.$$

We now introduce logarithms: for $x = 2^n$ we'll write $\log_2 x = n$ (pronounced 'log-to-the-base-2-of-*x*'). Then

$$\log_2 16 = \log_2 2^4 = 4, \log_2 128 = \log_2 2^7 = 7, \log_2 2048 = \log_2 2^{11} = 11,$$

and

$$\log_2(16 \times 128) = \log_2 2048 = 11 = 4 + 7 = \log_2 16 + \log_2 128.$$

These calculations illustrate the general rule that

The logarithm of a product is the sum of the logarithms of the separate terms.

In symbols:

$$\log_2(a \times b) = \log_2 a + \log_2 b.$$

So far we've defined these logarithms only for powers of 2, but they are also defined for numbers other than integers. For example, it can be shown that $\log_2 3 = 1.585\ldots$, $\log_2 5 = 2.322\ldots$, $\log_2 15 = 3.907\ldots$, and

$$\log_2(3 \times 5) = \log_2 15 = 3.907\ldots = 1.585\ldots + 2.322\ldots = \log_2 3 + \log_2 5.$$

In general, to multiply two (or more) numbers we look up their logarithms, add, and then locate the number whose logarithm is their sum.

For division we can take a similar approach:

To divide powers of 2 *we subtract their exponents.*
For example, to divide $4096 = 2^{12}$ by $512 = 2^9$ we write

$$4096 \div 512 = 2^{12} \div 2^9 = 2^{12-9} = 2^3 = 8.$$

Because $\log_2 4096 = 12$, $\log_2 512 = 9$, and $\log_2 8 = 3$, we can then write

$$\log_2 (4096 \div 512) = \log_2 8 = 3 = 12 - 9 = \log_2 4096 - \log_2 512.$$

In general, the logarithm of a quotient is the difference of the logarithms of the separate terms.

In symbols:

$$\log_2 (a \div b) = \log_2 a - \log_2 b.$$

So to divide two numbers we look up their logarithms, subtract, and then locate the number whose logarithm is their difference.

Up to now we've considered only logarithms based on powers of 2, but we can carry out similar processes with powers of other numbers, and relations of the form

$$\log_n (a \times b) = \log_n a + \log_n b, \quad \log_n (a \div b) = \log_n a - \log_n b$$

hold for logarithms to any other number base *n*.

The logarithms of Napier and Briggs

Since the Middle Ages calculations that turn multiplications into additions or subtractions have been known in trigonometry. For example, using the addition formulas for cosine, we have

$$\begin{aligned}\cos (x + y) &= \cos x \cos y - \sin x \sin y \\ \cos (x - y) &= \cos x \cos y + \sin x \sin y,\end{aligned}$$

and so

$$\begin{aligned}\cos x \cos y &= \tfrac{1}{2} \{\cos(x + y) + \cos (x - y)\} \\ \sin x \sin y &= \tfrac{1}{2} \{\cos(x - y) - \cos (x + y)\}.\end{aligned}$$

So the product of two cosines or sines can be written as the sum or difference of two cosines.

In the 16th century, following an idea suggested in Michael Stifel's *Arithmetica Integra* of 1544, mathematicians developed another method for replacing multiplication by addition: this was to turn geometric progressions whose successive terms have a common ratio r,

$$a,\ a\times r,\ a\times r^2,\ a\times r^3,\ a\times r^4,\ \ldots,$$

into arithmetic progressions whose successive terms have a common difference d,

$$a,\ a+d,\ a+2d,\ a+3d,\ a+4d,\ \ldots.$$

This process was called *prosthaphairesis*, from the Greek words for 'addition' and 'subtraction'.

In 1614 the Scotsman John Napier or Neper, Eighth Laird of Merchiston (to the south-west of Edinburgh), produced his *Mirifici Logarithmorum Canonis Descriptio* (Description of the Wonderful Canon of Logarithms) (see Figure 51 which opens this chapter). This work contains extensive tables of logarithms of the sines and tangents of all the angles from 0 to 90 degrees in steps of 1 minute of arc; his emphasis on the 'circular functions' of trigonometry, arising from his interest in spherical geometry, was so that his tables could be used by navigators and astronomers. Napier justified his work as follows:

> Seeing there is nothing, (right well beloved students in the mathematics) that is so troublesome to Mathematicall practise, nor that doth molest and hinder Calculators, than the Multiplications, Divisions, square and cubical Extractions of great numbers, which besides the tedious expence of time, are, for the most part subject to many slippery errors. I began therefore to consider in my minde, by what certaine and ready Art I might remove those hindrances, And having thought upon many things to this purpose, I found at length some excellent briefe rules to be treated of (perhaps) hereafter.

Napier's 'excellent briefe rules' were not based on 2 and are not the ones that we use now, but originated from the above idea of prosthaphairesis.

He considered two points moving along straight lines – one (PQ) of finite length and the other (L_0L) of infinite length (see Figure 53) – as follows:

the first point moves from P along PQ in such a way that its speed at each point is proportional to the distance that it still has to travel to Q.
the second point, representing its 'Napierian logarithm' (which we'll write as $\log_N$), starts from L_0 and travels at constant speed towards L for ever.

So, in successive periods of time, the distances still to be travelled by the first point form a geometric progression, and the distances already travelled by the second point form an arithmetic progression.

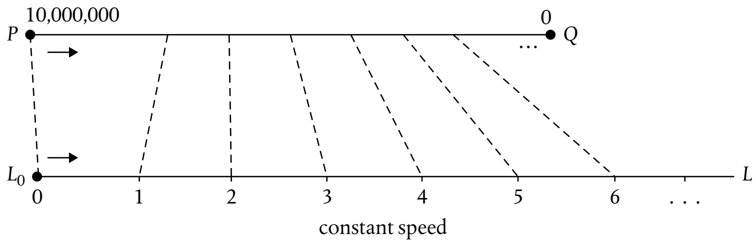


Fig. 53. Constructing Napier's logarithms

Napier took 10^{-7} as his successive time intervals and then multiplied his results by 10^7, in order to avoid the use of decimal fractions which were still largely unfamiliar at the time. It follows from his construction that the logarithm of 10,000,000 is 0, and that as n decreases its logarithm $\log_N n$ increases. It also follows that

$$\log_N (a \times b) = \log_N a + \log_N b - \log_N 1,$$

so that for each calculation he had to subtract the cumbersome term $\log_N 1 = 161{,}180{,}956$.

Napier was not the first to introduce logarithms. A few years earlier a similar approach had been taken by the Liechtensteinian clockmaker Joost Bürgi who worked in Prague. But Bürgi didn't publish his results until 1620, by which time Napier's logarithms were already widely known.

In 1615 Henry Briggs, the first professor of geometry at Gresham College in London, heard about Napier's logarithms and was wildly excited by them. He included them in his lectures, enthusing that Napier had

> set my Head and hands a Work with his new and remarkable logarithms… I never saw a Book which pleased me better or made me more wonder.

But Napier's logarithms were cumbersome to use, and Briggs wished to redefine them so as to avoid having to subtract $\log_N 1$ in every calculation:

> I myself, when expounding this doctrine to my auditors in Gresham College, remarked that it would be much more convenient that 0 should be kept for the logarithm of the whole sine [namely, 1].

Briggs went up to Edinburgh for two summers to stay with Napier, and it is recorded that when they first met they spent the first quarter-hour looking at each other in admiration without speaking a word. The outcome of their meetings was that Briggs started to construct 'logarithms to base 10' in which

$$\log_{10} 1 = 0,\ \log_{10} 10 = 1,\ \log_{10} 100 = 2,\ \text{and if } x = 10^n,\ \text{then } \log_{10} x = n.$$

Other values he found by interpolation – for example,

$$10^{1/2} = 3.162\ldots \text{ so } \log_{10} 3.162\ldots = 0.5.$$

In order to find all these intermediate values and to obtain the necessary accuracy in his tables, he calculated $\sqrt{10}$, $\sqrt{\sqrt{10}}$, $\sqrt{\sqrt{\sqrt{10}}}, \ldots$, eventually taking the square root fifty-four times, all to thirty decimal places! Since $\log_{10} 1 = 0$, as he had demanded, Briggs's logarithms satisfied the simpler fundamental rule:

$$\log_{10}(a \times b) = \log_{10} a + \log_{10} b.$$

In 1617 Briggs produced his *Logarithmorum Chilias Prima* (The First Thousand Logarithms), a small pamphlet of sixteen pages containing his calculations (see Figure 54). Seven years later, after he had left London to become the first Savilian Professor of Geometry at Oxford University, he followed this with his *Arithmetica Logarithmica*, an extensive collection of

2		Logarithmi.			Logarithmi.
1	0	0000,00000,00000	34	1	5314,78917,04226
2	0	3010,29995,66398	35	1	5440,68044,35028
3	0	4771,21254,71966	36	1	5563,02500,76729
4	0	6020,59991,32796	37	1	5682,01724,06700
5	0	6989,70004,33602	38	1	5797,83596,61681
6	0	7781,51250,38364	39	1	5910,64607,02650
7	0	8450,98040,01426	40	1	6020,59991,32796
8	0	9030,89986,99194	41	1	6127,83856,71974
9	0	9542,42509,43932	2	1	6232,49290,39790
10	1	0000,00000,00000	43	1	6334,68455,57959

Fig. 54. Some of Briggs's 1617 logarithms

logarithms to base 10 of the integers from 1 to 20,000 and from 90,000 to 100,000, all calculated by hand to fourteen decimal places. The gap in these tables between 20,000 and 90,000 was filled in by the Dutch mathematician Adriaan Vlacq and published in 1628.

In the 1630s a number of mechanical instruments based on logarithmic scales were created. Designed to be used for complicated calculations, particularly by astronomers and navigators, these included the *slide rule* which was used for over 300 years until the advent of the pocket calculator in the 1970s.

Logarithms were soon recognized as being of immense value to those needing to carry out extensive calculations. In the 1733 English edition of his *Item de Natura et Arithmetica Logarithmorum Tractatus Brevis* (A Short Treatise of the Nature and Arithmetick of Logarithms) the Oxford mathematician and astronomer John Keill observed that

> By their assistance the Mariner steers his Vessel, the Geometrician investigates the Nature of the higher Curves, the Astronomer determines the Places of the Stars, the Philosopher accounts for other Phenomena of Nature; and lastly, the Usurer computes the Interest of his Money.

Indeed, according to the French mathematician and physicist Pierre-Simon Laplace, by 'shortening the labours' involved in calculations, Napier's logarithms 'doubled the life of the astronomer'.

Enter the calculus

The 17th century saw the development of the calculus, culminating in the achievements of Isaac Newton and Gottfried Wilhelm Leibniz. The subject was made up from two seemingly unrelated strands, now called *differentiation* and *integration*. Differentiation is concerned with how things move or change, and is used to find velocities and the slopes of tangents to curves; for example, if the curve is $y = x^n$, then its slope (denoted by dy/dx) is nx^{n-1}. Integration is used to find areas of shapes – in particular, the area under a curve. Of particular interest to us is the area under a hyperbola.

In the 1640s the Flemish Jesuit and mathematician Gregory of Saint-Vincent investigated the area under the rectangular hyperbola $y = 1/x$, and in his *Opus Geometricum* of 1647 he showed that the area between $x = a$ and $x = b$ is the same as the area between $x = c$ and $x = d$ whenever the ratios b/a and d/c are equal (see Figure 55).

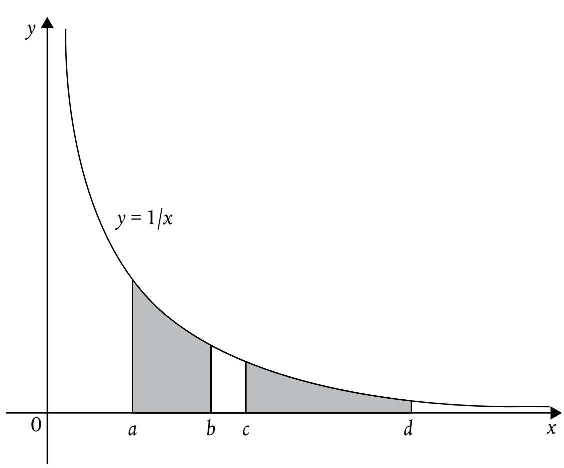


Fig. 55. The area under the hyperbola $y = 1/x$

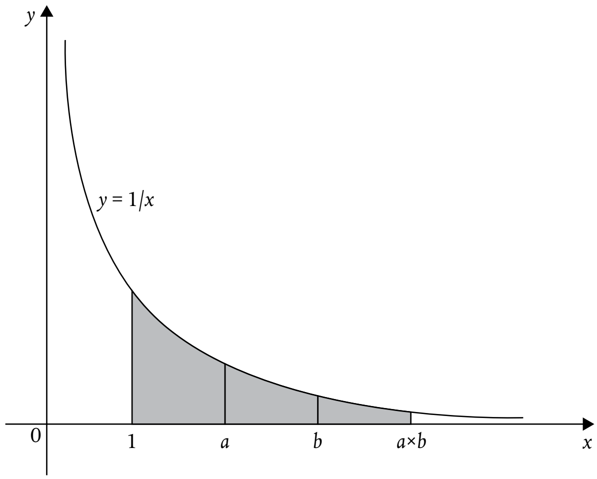


Fig. 56. The area under the hyperbola $y = 1/x$

On perusing Saint-Vincent's work, his student Alfonso de Sarasa realized that the area under the curve $y = 1/x$ satisfies the basic equation satisfied by the logarithm function – namely, if $A(t)$ is the area between $x = 1$ and $x = t$, then

$$A(a \times b) = A(a) + A(b)$$

(see Figure 56). Details of the calculation are given in Box 6 for those familiar with integration.

At this time the connections between this logarithm function and the exponential number *e* were unrecognized. As we shall see, however, Napier's logarithms were 'nearly' logarithms to base $1/e$, and the area under a hyperbola is a logarithm to base *e*, but the nature and properties of the number *e* had not yet been clarified.

In the 1660s Nicolaus Mercator, and independently Isaac Newton and James Gregory, were investigating the area under the hyperbola $y = 1/(1 + x)$; this is similar to the previous hyperbola, but translated one unit to the left (see Figure 57).

Box 6: The area under a hyperbola

For any $t \geq 1$, let $A(t)$ be the area under the hyperbola $y = 1/x$ between $x = 1$ and $x = t$. Then, on splitting up the range of integration $[1, ab]$ into $[1, a]$ and $[a, ab]$, we have

$$A(a \times b) = \int_1^{ab} \frac{1}{x}\,dx = \int_1^{a} \frac{1}{x}\,dx + \int_a^{ab} \frac{1}{x}\,dx.$$

The first integral on the right is simply $A(a)$.
For the second integral we substitute $u = x/a$, so that $x = au$, $dx = a\,du$, and the limits of integration become 1 and b. Then the second integral becomes

$$\int_1^{b} \frac{1}{au}\,a\,du = \int_1^{b} \frac{1}{u}\,du = A(b).$$

It follows that $A(a \times b) = A(a) + A(b)$.

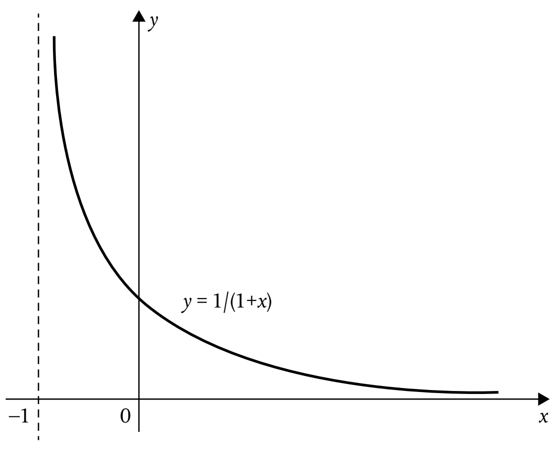


Fig. 57. The hyperbola $y = 1/(1 + x)$

They began with the identity

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - \ldots,$$

which can be verified by multiplying both sides by $1 + x$ and noticing that most of the terms on the right cancel out, leaving just 1, or by summing the geometric series on the right. They then integrated this infinite series term by term between 0 and 1, using the fact that the integral of x^k is $x^{k+1}/(k+1)$ for each number k, and obtained the new series

$$\log(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \tfrac{1}{5}x^5 - \ldots.$$

(We suggested a similar method in Chapter 3 to find the infinite series for arctan x.)

This series for log $(1+x)$ appeared in print for the first time in 1668, in Mercator's *Logarithmotechnica*, but was already known to Newton, who hadn't bothered to publish it. It is valid for all values of x between -1 and 1, and also for $x = 1$ when it gives

$$\log 2 = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \ldots.$$

Again, these are 'logarithms to base *e*', but at the time this was neither specified nor fully understood. Newton also used such a series to calculate a logarithm to no fewer than 55 decimal places (see Figure 58).

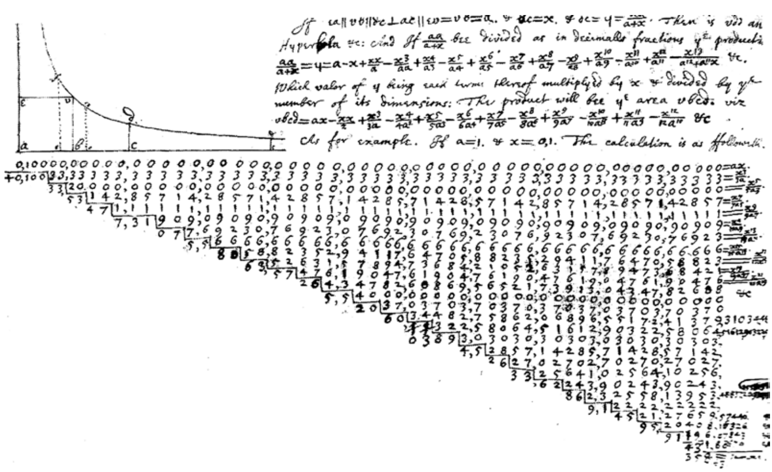


Fig. 58. Part of Newton's calculation of the area under a hyperbola

A problem of interest

So what exactly is this number *e*, and how did it arise?

In 1683 the Swiss mathematician Jakob Bernoulli was concerned with problems of calculating interest. Given a sum of money to invest at a given rate of interest over a number of years, how fast will it grow? The answer depends on whether we use simple or compound interest, and on how often we calculate the interest.

As an example, suppose that we invest £100 at an annual rate of 10 per cent over a number of years. With simple interest the amount increases linearly – to £110 after one year, £120 after two years, and so on. After k years our £100 has risen to £$100 + 10k$.

What happens with compound interest? After one year the amount rises to £110, as before. But after two years we've added a further 10 per cent, not of £100 but of £110, giving us £110 + £11 = £121. After three years we've added a further 10 per cent of £121, giving us £133.10, and so on. After k years our £100 has risen to £100×1.1^k.

Let's now change the problem. Bernoulli wanted to find out what would happen if we calculate the interest more frequently – say, n times per year, or even continuously.

Suppose we calculate the interest after every period of six months ($n = 2$). Then after the first period the amount increases by $\left(\frac{1}{2} \times 10\right) = 5$ per cent to £$100 \times 1.05 =$ £105, and after the second period it increases by a further 5 per cent, not of £100 but of £105, giving us £105 + £5.25 = £110.25 – that is, £100×1.05^2.

Suppose we next calculate the interest every three months ($n = 4$). Then after the first period the amount increases by $\left(\frac{1}{4} \times 10\right) = 2\frac{1}{2}$ per cent to £$100 \times 1.025 =$ £102.50, after the second period it becomes £$100 \times 1.025^2 \approx$ £105.06, after the third period it becomes £$100 \times 1.025^3 \approx$ £107.69, and by the end of the year it has become £$100 \times 1.025^4 \approx$ £110.38.

In a similar way, if we calculate the interest n times per year, then after each period the amount increases by $10/n$ per cent – that is, it is multiplied by $1 + 0.1/n$ – and at the end of the year it has become £$100 \times (1 + 0.1/n)^n$. For example, if we calculate the interest every month, then the final amount is £100 × $(1+0.1/12)^{12}$ = £110.47, and if we calculate it every day, then the final amount is £100 × $(1+0.1/365)^{365}$ = £110.51.

As the year is further subdivided, what happens to these amounts? Do they increase without bound or do they settle down to a limiting value? And what happens if the interest is calculated continuously? It turns out that, in either case they approach a limit of just under £110.52, which is obtained by multiplying £100 by $e^{0.1}$, where e is the exponential number. So what is this number e?

To find out, we'll repeat the above process but we'll start with just £1 and increase the rate of interest to the unlikely annual rate of 100 per cent. What happens when the year is then divided into shorter periods? We obtain the following list of final amounts in pounds, calculated to five decimal places:

period:	year	half-year	quarter-year	two months	month
final amount:	2.00000	2.25000	2.44141	2.52153	2.61304
period:	week	day	hour	minute	second
final amount:	2.69260	2.71417	2.71813	2.71828	2.71828

What we've done to obtain these results is to note that if the year is divided into n periods, then after each period the amount is multiplied by $1+1/n$, so that the final amount is $(1+1/n)^n$. We also see that, as n increases indefinitely, these numbers tend to a limiting value that corresponds to when the interest is calculated continuously. This limiting value is the exponential number that Euler called e. (Bernoulli had called it b, but there's no suggestion that either of them deliberately chose the first letter of his name for this constant.)

In the same way, if the rate of interest is x, then the final amount is obtained by multiplying the original sum by the limiting value of the expression $(1+x/n)^n$, which turns out to be equal to e^x. For example, as we saw earlier when the rate of interest was 10 per cent or 0.1, the limiting value is obtained by multiplying the original sum of £100 by $e^{0.1} \approx 1.1052$, giving £110.52.

Properties of *e*

The greatest advances in understanding logarithms, the exponential function, and the connections between them, were made in the early 18th century. The main figure in this story was Leonhard Euler, who investigated the main properties of the exponential number e and the function $y = e^x$, and who placed e at the centre of discussions of the logarithmic function. In 1748 his celebrated *Introductio in Analysin Infinitorum*, mentioned in the Introduction, brought together many results from his earlier works. Here are some of his main findings.

e as a limit

In the previous section we saw that e is the limit of the numbers $(1 + x/n)^n$ as n increases indefinitely, and that e^x is the limit of $(1 + x/n)^n$ for any number x. To summarize, using the notation of limits, we can write:

$$\lim_{n\to\infty}(1+1/n)^n = e \text{ and } \lim_{n\to\infty}(1+x/n)^n = e^x.$$

e as an infinite series

As Isaac Newton had already discovered, the number e is also the sum of the infinite series

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots,$$

where the numbers in the denominators are the factorials $n!$ that we introduced in Chapter 2 in connection with Liouville's transcendental number.

More generally we have, for any x,

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \ldots.$$

This series converges for all values of x.

In fact, these series converge quickly because the factorials increase very rapidly; for example, taking just the first ten terms of the series for e yields the approximation $e \approx 2.7182787\ldots$, which is accurate to five decimal places.

For those familiar with the binomial theorem, Box 7 shows the connection between these two expressions for e^x – as a limit and as an infinite series.

Box 7: Linking two expressions for e^x

We start with the binomial expansion for $(1 + a)^n$:

$$(1+a)^n = 1 + \frac{n}{1!}a + \frac{n(n-1)}{2!}a^2 + \frac{n(n-1)(n-2)}{3!}a^3 + \frac{n(n-1)(n-2)(n-3)}{4!}a^4 + \ldots.$$

Putting $a = x/n$ gives

$$\left(1+\frac{x}{n}\right)^n = 1 + \frac{n}{1!}\left(\frac{x}{n}\right) + \frac{n(n-1)}{2!}\left(\frac{x}{n}\right)^2 + \frac{n(n-1)(n-2)}{3!}\left(\frac{x}{n}\right)^3 + \frac{n(n-1)(n-2)(n-3)}{4!}\left(\frac{x}{n}\right)^4 + \ldots,$$

which we can rewrite as

$$\left(1+\frac{x}{n}\right)^n = 1 + \frac{1}{1!}x + \frac{1}{2!}\left(1-\frac{1}{n}\right)x^2 + \frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)x^3 + \frac{1}{4!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)x^4 + \ldots$$

We now take the limit as n increases indefinitely.

Then the left-hand side converges to e^x and each bracket of the form $(1-k/n)$ on the right converges to 1, leaving

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \ldots.$$

The multiplication rule

Earlier we saw that to multiply powers of 2 we add their exponents. A similar multiplication rule holds for powers of e:

$$e^x \times e^y = e^{x+y}, \text{ for any numbers } x \text{ and } y.$$

This basic rule can be proved by multiplying together the infinite series for e^x and e^y.

The slope of the graph of $y = e^x$

The graph of the function $y = e^x$ is shown in Figure 59. One of its most important features is that the slope at each point x of the graph is also e^x

– that is, the slope at any point is the y-value; a reason for this is given in Box 8, for those familiar with calculus. It follows that the curve becomes steeper and steeper as x increases.

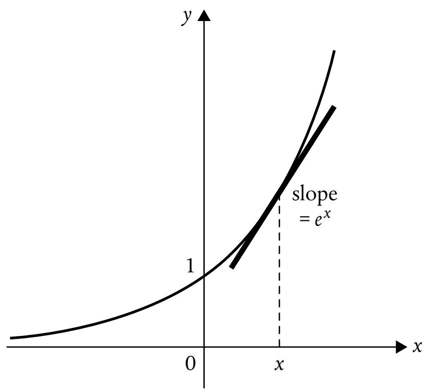


Fig. 59. The slope at each point of the curve $y = e^x$ is e^x

Box 8: The slope at each point x of the graph of $y = e^x$ is e^x

If we take the series for e^x,

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \ldots,$$

and differentiate the right-hand side one term at a time, we get

$$0 + \frac{1}{1!}(1) + \frac{1}{2!}(2x) + \frac{1}{3!}(3x^2) + \frac{1}{4!}(4x^2) + \ldots = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \ldots = e^x.$$

It follows that the slope at each point of the graph of $y = e^x$ is e^x.

In the language of differential equations, it follows that $y = e^x$ is a solution of the equation $dy/dx = y$. In fact, the *only* solutions of this differential equation are $y = e^x$ and its multiples $y = Ce^x$, where C is a constant.

We can also draw the graphs of other exponential functions such as $y = 2^x$, or in general $y = k^x$ for any number $k > 1$. Here the shape of the curve is similar to that of $y = e^x$, and the slope of the graph at the point x is $\log_e k \times k^x$. If $k = e$, then $\log_e k = 1$ and the slope is e^x as before.

Exponentials and logarithms are inverse functions

Earlier we saw that if $x = 2^n$ then $n = \log_2 x$; similarly, if $x = e^n$ then $n = \log_e x$. These connections link exponentials and logarithms and suggest the fundamental connection between the functions $y = e^x$ and $y = \log_e x$ (where e is the base of the logarithms): they are 'inverses' of each other. In symbols,

$$\log_e e^x = x \text{ and } e^{\log_e x} = x.$$

It follows that:

if we take x, calculate e^x, and take the logarithm (to base e) of the result, we get back to x.

if we take x, calculate $\log_e x$, and take the exponential of the result, we get back to x.

This inverse relationship had been observed by John Wallis in 1685 and was developed by Euler in his *Introductio* of 1748.

We often simplify $\log_e x$ to $\ln x$, where 'ln' means 'natural logarithm'. With this notation,

$$y = e^x \text{ if and only if } x = \ln y.$$

We shall use ln from now on.

Since $y = e^x$ and $y = \ln x$ are inverses of each other, their graphs can be obtained from each other by reflection in the line $y = x$ (see Figure 60).

We can also use this inverse relationship to show that the multiplicative rule for exponentials and the basic rule for logarithms are essentially the same result.

For, the multiplicative rule for exponentials is $e^x \times e^y = e^{x+y}$.
Writing $x = \ln a$ and $y = \ln b$, we have $a \times b = e^x \times e^y = e^{x+y} = e^{(\ln a + \ln b)}$,
and so, on taking logarithms of both sides, $\ln (a \times b) = \ln a + \ln b$,
which is the basic rule for logarithms.
Conversely, the basic rule for logarithms is $\ln (a \times b) = \ln a + \ln b$.
Writing $a = e^x$ and $b = e^y$, we have $\ln (e^x \times e^y) = \ln e^x + \ln e^y = x + y$,

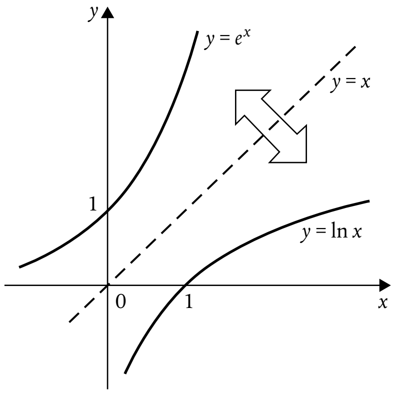


Fig. 60. The graphs of $y = e^x$ and $y = \ln x$

and so, on taking exponentials of both sides, $e^x \times e^y = e^{x+y}$, which is the multiplicative rule for exponentials.

e is irrational

In 1737 Euler proved that e is an irrational number – that is, it cannot be written as a/b, where a and b are integers. Noting that finite continued fractions always correspond to rational numbers, he proved that infinite continued fractions (such as the continued fractions for π in Chapter 3) must correspond to irrational numbers. He then showed that e can be written as an infinite continued fraction, as follows (here the counting numbers 1, 2, 3, 4, . . . appear as both numerators and denominators):

$$2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{6+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{8+\dots}}}}}}}}}}}$$

Because this continued fraction for e continues for ever, e must be irrational.

Another proof that e is irrational uses the infinite series for e. It is due to Joseph Fourier, best known for his work on *Fourier series*, and is given in Box 9.

Box 9: The number e is irrational

We shall assume that $e = a/b$ is a rational number, where a and b are positive integers, and obtain a contradiction. We shall need the series for e:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{b!} + \frac{1}{(b+1)!} + \frac{1}{(b+2)!} + \frac{1}{(b+3)!} + \ldots.$$

Consider the number

$$\begin{aligned} N &= b! \times \left\{ e - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{b!} \right) \right\} \\ &= b! \times \left\{ a/b - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{b!} \right) \right\} \\ &= (b-1)!\, a - \left(b! + \frac{b!}{1!} + \frac{b!}{2!} + \frac{b!}{3!} + \frac{b!}{4!} + \ldots + \frac{b!}{b!} \right). \end{aligned}$$

But the first term is an integer, and so is each of the terms in the second bracket, so N is an integer.

Moreover, $N > 0$, because $e > 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{b!}$.

But

$$e - \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{b!} \right) = \frac{1}{(b+1)!} + \frac{1}{(b+2)!} + \frac{1}{(b+3)!} + \ldots,$$

and so

$$\begin{aligned} N &= b! \times \left(\frac{1}{b+1!} + \frac{1}{b+2!} + \frac{1}{b+3!} + \ldots \right) \\ &= \frac{1}{b+1} + \frac{1}{(b+1)(b+2)} + \frac{1}{(b+1)(b+2)(b+3)} + \ldots \\ &< \frac{1}{b+1} + \frac{1}{(b+1)^2} + \frac{1}{(b+1)^3} + \ldots = 1/b, \end{aligned}$$

on summing the geometric progression.

Now $b > 1$ since e is not an integer, and so $1/b < 1$.

It follows that N is an integer lying strictly between 0 and 1, which is impossible. This contradiction shows that e is irrational.

In Chapter 6 we'll visit the harder problem of determining whether e is transcendental.

Napier's definition of the logarithm

We conclude this section by returning to our earlier claim that Napierian logarithms are 'nearly' logarithms to base $1/e$. We recall that Napier compared the movements of two points over successive time intervals of length 10^{-7} in order to avoid the use of decimal fractions, and his definition amounts to the following:

$$\text{if } y = 10^7 \left(1 - 10^{-7}\right)^x, \text{ then } \log_N y = x.$$

If we now re-scale y and x by a factor of 10^7, by writing $Y = y/10^7$ and $X = x/10^7$, then

$$Y = \left(1 - 10^{-7}\right)^{10^7 X}.$$

But as n increases indefinitely, e^{-1} is the limiting value of $(1 - 1/n)^n$, and so $\left(1 - 10^{-7}\right)^{10^7}$ is very close to e^{-1}. It follows that Y is very close to $(1/e)^X$, and so $\log_{1/e} Y$ is very close to X – that is, Napierian logarithms are 'nearly' logarithms to base $1/e$, as we claimed.

Hanging chains and derangements

Let's consider two contrasting situations where exponential functions arise.

Hanging chains

What U-shaped curve is taken by a hanging chain (see Figure 61)? Galileo Galilei considered this problem in his *Two New Sciences* of 1638, and claimed that it approximated a parabola. That this is not the correct curve was proved by Joachim Jungius in a posthumous publication of 1669.

Fig. 61. A hanging chain

The curve's properties were studied by Robert Hooke in the 1670s, and its equation was later obtained by Gottfried Wilhelm Leibniz, Christiaan Huygens, and Johann Bernoulli, who showed its connection with the exponential functions $y = e^x$ and $y = e^{-x}$. If we add and subtract these functions, we obtain the so-called *hyperbolic functions*,

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}) \text{ and } \sinh x = \tfrac{1}{2}(e^x - e^{-x}),$$

whose graphs are shown in Figure 62. The curve $y = \cosh x$ is the shape taken by a hanging chain and is called a *catenary* after the Latin word 'catena' for a chain; the use of the word 'catenary' for such a curve is attributed to Thomas Jefferson.

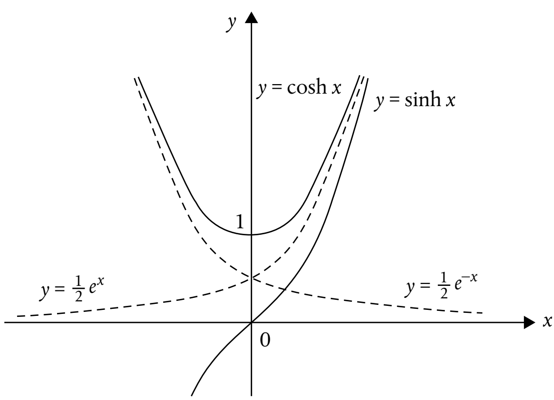


Fig. 62. The graphs of $y = \cosh x$ and $y = \sinh x$

Although hyperbolic functions are defined in terms of exponentials, their properties are remarkably similar to those of the trigonometric functions. For example, compare

$$\sin(x+y) = \sin x \cos y + \cos x \sin y$$
$$\text{and } \sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y,$$
$$\sin 2x = 2\sin x \cos x \text{ and } \sinh 2x = 2\sinh x \cosh x,$$
$$\text{and } \cos^2 x + \sin^2 x = 1 \text{ and } \cosh^2 x - \sinh^2 x = 1.$$

Even their power series are very similar:

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \ldots \text{ and } \sinh x = x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \frac{1}{7!}x^7 + \ldots$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \ldots \text{ and } \cosh x = 1 + \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \frac{1}{6!}x^6 + \ldots$$

The trigonometric and exponential functions seem very different, so what is the reason for these similarities? After all, as we saw in the Introduction, the graphs of $y = \sin x$ and $y = \cos x$ oscillate indefinitely, whereas the graph of the exponential function $y = e^x$ goes shooting off to infinity when x becomes large. All will be explained in Chapter 6.

Derangements

In 1708 Pierre Rémond de Montmort posed the following problem in his *Essay on the Analysis of Games of Chance*:

How many permutations of the numbers 1, 2,…, n leave no number in its original position?

Such rearrangements are now known as *derangements*. For example, when $n = 4$ there are $4! = 24$ possible permutations of the four numbers 1, 2, 3, 4, but only nine of these are derangements, with no number in its usual ordering of 1 2 3 4:

2 1 4 3, 2 3 4 1, 2 4 1 3, 3 1 4 2, 3 4 1 2, 3 4 2 1, 4 1 2 3, 4 3 1 2, 4 3 2 1.

The derangement problem is sometimes expressed in a more popular form:

If we randomly place a number of messages into addressed envelopes, what is the probability that no message ends up in its correct envelope?

To answer this question, we'll let d_n denote the number of derangements of n letters (for example, $d_4 = 9$). Then the table overleaf gives the values of n, $n!$, d_n, and the corresponding probability $d_n/n!$, for $n \leq 8$.

As n increases, it seems as though the probability $d_n/n!$ approaches a fixed value that is close to 0.3679. But what is this value exactly?

n	1	2	3	4	5	6	7	8
$n!$	1	2	6	24	120	720	5040	40320
d_n	0	1	2	9	44	265	1854	14833
$d_n/n!$	0	0.5	0.3333	0.375	0.3667	0.3681	0.3678	0.3679

Around 1779 Euler became interested in the derangement problem, and used a counting argument to show that

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2}.$$

For example, when $n = 4$ we have $d_4 = 3d_3 + 3d_2 = (3 \times 2) + (3 \times 1) = 9$.

He then solved this equation to give

$$d_n = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots \pm \frac{1}{n!}\right).$$

For example,

$$d_4 = 4!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}\right) = 24\left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24}\right)$$

$$= 24 - 24 + 12 - 4 + 1 = 9, \text{ as expected.}$$

Unfortunately, the formula for d_n can be time-consuming to evaluate for all but very small values of n – but there's a quicker way. Because

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots,$$

d_n is very close to $n! \times e^{-1}$. Indeed, it turns out that

For every n, the number of derangements of n symbols is the integer closest to $n!/e$.

For example, when $n = 8$, $n!/e = 14832.9\ldots$, and so $d_n = 14833$.

Exponential growth and decay

We conclude this chapter by returning to the subject of exponential growth, and to the related topic of exponential decay. For this section we assume some knowledge of differential equations: the results will not be needed in later chapters, and the section can be omitted if desired.

Population growth

How fast does a population grow?

If $N(t)$ is the size of a population at time t, and if the population grows at a fixed rate k proportional to its size, then it satisfies the differential equation

$$dN/dt = kN.$$

It follows that $dN/N = k\,dt$, and we can integrate this equation to give

$$\ln N = kt + \text{constant, or (in terms of exponentials) } N(t) = Ce^{kt},$$

where C is another constant.

If the initial population is N_0 when $t = 0$, then $N_0 = Ce^{0k} = C$, so $C = N_0$ and the population at time t is

$$N(t) = N_0 e^{kt}.$$

This is an example of exponential growth (see Figure 63).

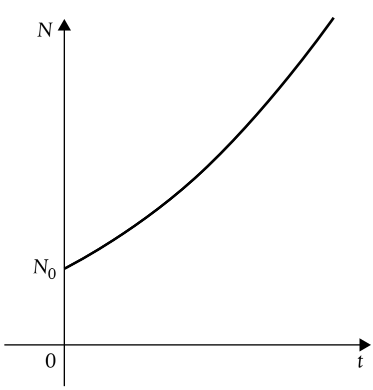


Fig. 63. Population growth

Cooling of a cup of tea

How fast does a cup of tea cool?

By Newton's law of cooling, the rate at which tea cools is proportional to the difference in temperature between the tea and the surrounding room: if $T(t)$ is the temperature of the tea at time t and T_0 is the temperature of the room, then the rate of cooling is given by

$$dT/dt = -K(T - T_0),$$

where K is a constant; the negative sign arises since the temperature is decreasing. This equation can then be written as $dT/(T - T_0) = -K\,dt$, which can then be solved to give the temperature of the tea at time t as

$$T(t) = T_0 + Ce^{-Kt},$$

where C is a constant which we can find if we know the initial temperature of the tea.

This is an example of exponential decay (see Figure 64).

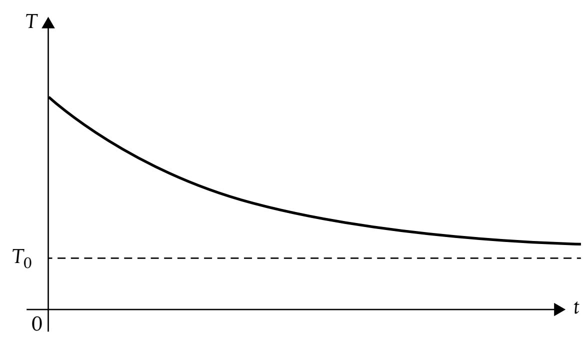


Fig. 64. The cooling of a cup of tea

The half-life of radium

Another example of exponential decay concerns radioactive material such as radium (see Figure 65). Here the radium decays according to a similar formula,

$$dm/dt = -Km,$$

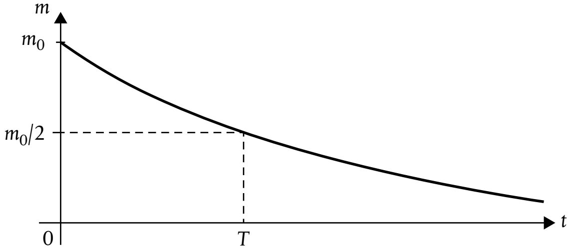


Fig. 65. The decay of radium

where $m(t)$ is the mass of the radium at time t, and K is a constant.

As above, we can rewrite this equation as $dm/m = -K\,dt$, and we can then solve it to find $m(t)$. The answer is

$$m(t) = m_0 e^{-Kt},$$

where m_0 is the amount of radium at time $t = 0$.

To find the constant K, we usually introduce the 'half-life', the time T taken for the radium to reduce to half of its original size. Then, after time T,

$$m_0/2 = m_0 e^{-KT}, \text{ so } e^{KT} = 2 \text{ and } KT = \ln 2.$$

Substituting this back into the original solution gives the mass of the radium at time t as

$$m(t) = m_0 e^{-(\ln 2)t/T}.$$

For example, if the original mass of the radium is 1 mg, and if the half-life is 1590 years (as it is for radium-226), then the mass in milligrams after t years is given by

$$m(t) = e^{-(\ln 2)t/1590}.$$

Fig. 66. Sir William Rowan Hamilton scratches his quaternions on a bridge in Dublin

CHAPTER 5

i

The imaginary number

We have shown the symbol $\sqrt{-a}$ to be void of meaning,
or rather self-contradictory and absurd.

Augustus De Morgan

The story of *i* concerns the square root of minus 1 and the so-called *imaginary numbers*. But is there such a thing as $\sqrt{-1}$? After all, if you square either 1 or −1 you get 1, so what can you possibly square to get −1?

The answer began to emerge in the 16th century, but even three centuries later there was still much confusion about the subject. In 1831 Augustus De Morgan, Professor of Mathematics at University College, London, made the above remark in his book *On the Study and Difficulties of Mathematics,* and the Victorian Astronomer Royal, George Airy, commented:

I have not the smallest confidence in any result which is essentially obtained by the use of imaginary symbols.

In 1854 George Boole, founder of 'Boolean algebra', described the square root of −1 as 'an uninterpretable symbol' in his celebrated logic book *Laws of Thought.*

Earlier, Gottfried Wilhelm Leibniz had been more encouraging, claiming that

The imaginary numbers are a wonderful flight of God's spirit; they are almost an amphibian between being and not being.

But the great Euler, who made so many contributions to the development and use of these imaginary numbers, seemed to take an uncharacteristically different view:

> All such expressions as $\sqrt{-1}, \sqrt{-2}$, etc., are consequently impossible or imaginary numbers, since they represent roots of negative quantities; and of such numbers we may truly assert that they are neither nothing, nor greater than nothing, nor less than nothing, which necessarily constitutes them imaginary or impossible.

It was René Descartes who first called these numbers *imaginary*, in his *Discourse on Method* of 1637. They're now called *complex numbers*, a name given to them in 1831 by Carl Friedrich Gauss. But how did they arise? And why did they cause so much confusion for several centuries?

Different types of numbers

In Chapters 1 and 2 we looked at how our number system is built up. Starting with the counting numbers, 1, 2, 3, . . . , we obtained all the *integers* – positive, negative, and zero. This was a non-trivial process, extending over thousands of years, and negative numbers were initially treated with the same ridicule that the imaginary numbers would later have to face. These days, when we have no difficulty understanding negative temperatures in our weather forecasts, it seems hard to see why negative numbers caused so much suspicion.

Another crucial step was to divide one integer by another to give fractions, or *rational numbers*. All we need to remember is not to divide by 0, and that different fractions can represent the same rational number; for example, $\frac{1}{2}$ is the same as $\frac{2}{4}$ or $\frac{-35}{-70}$. But many numbers cannot be written as fractions – for example, $\sqrt{2}, \sqrt[3]{7}$, and the numbers π and e that we met in Chapters 3 and 4. These are the *irrational numbers*, which when combined with the rational numbers form the *real numbers*.

For many purposes the real numbers are all we need. But suppose that we now agree to allow this mysterious object called '$\sqrt{-1}$'. We can then form many more 'numbers', such as $3 + 4\sqrt{-1}$. Ignoring for the moment what they actually mean, we can carry out simple calculations with these objects.

Addition is easy:

$$(2+3\sqrt{-1})+(4+5\sqrt{-1})=(2+4)+(3+5)\sqrt{-1}=6+8\sqrt{-1}.$$

So is multiplication (after we've replaced $\sqrt{-1}\times\sqrt{-1}$ whenever it appears by -1):

$$\begin{aligned}(2+3\sqrt{-1})\times(4+5\sqrt{-1}) &= (2\times4)+(3\sqrt{-1}\times4)+(2\times5\sqrt{-1})\\ &\quad+(15\times\sqrt{-1}\times\sqrt{-1})\\ &=8+12\sqrt{-1}+10\sqrt{-1}-15\\ &=(8-15)+(12+10)\sqrt{-1}=-7+22\sqrt{-1}.\end{aligned}$$

For convenience, from now on we'll usually follow Euler who in 1777 replaced the cumbersome symbol $\sqrt{-1}$ by the letter *i* (the first letter of 'imaginary'), so that $i^2=-1$; for example, the result of this last calculation would be written more clearly as

$$(2+3i)\times(4+5i)=-7+22i.$$

We'll need some terminology in what follows. Given a complex number of the form $a+bi$, we say that a is its *real part* and that b is its *imaginary part*: if $b=0$ we get the real number a, and if $a=0$ we get the 'imaginary number' bi. We also say that the *conjugate* of $z=a+bi$ is the complex number $\overline{z}=a-bi$, its *modulus* or *absolute value* $|z|=|a+bi|$ is the real number $\sqrt{(a^2+b^2)}$, and if $z\neq0$ its *argument* $\arg z=\arg(a+bi)$ is arctan b/a. For example, the real part of the complex number $3+4i$ is 3, its imaginary part is 4, its conjugate is $3-4i$, its modulus $|3+4i|$ is $\sqrt{(3^2+4^2)}=\sqrt{25}=5$, and its argument is arctan $\frac{4}{3}$ (which is about 53° or 0.93 radians). The geometrical meanings of these terms will become clear later on.

We can carry out all the usual arithmetic operations on complex numbers, as follows:

Addition: $(a+bi)+(c+di)=(a+c)+(b+d)i$
Subtraction: $(a+bi)-(c+di)=(a-c)+(b-d)i$
Multiplication: $(a+bi)\times(c+di)=(ac-bd)+(ad+bc)i$
(after replacing i^2 by -1).

To divide $a+bi$ by $c+di$ we first multiply the numerator and denominator by the conjugate of $c+di$, as follows:

$$\frac{a+bi}{c+di}=\frac{(a+bi)(c-di)}{(c+di)(c-di)}=\frac{(ac+bd)+(bc-ad)i}{(c^2+d^2)}=\frac{ac+bd}{c^2+d^2}+\left(\frac{bc-ad}{c^2+d^2}\right)i.$$

Solving equations

Let's now return to our various types of numbers and look at them from a different point of view.

If we're restricted to the counting numbers, then we can solve certain algebraic equations. For example, the equation that we now write as $x+3=7$ has the solution $x=4$. But to solve the equation $x+7=3$ (a task considered impossible for many centuries), we need to expand our number system to include the negative integers, and the solution is $x=-4$. We can now solve all equations of the form $x+a=b$, where a and b are integers.

The next stage is to bring in fractions. Using these we can solve an equation such as $7x=5$: the solution is $x=\frac{5}{7}$. We can now solve any *linear* equation – those of the forms $ax=b$ or $ax+b=c$, where a, b, and c are integers or rational numbers with $a\neq 0$.

Once we've introduced irrational numbers, we can go beyond linear equations and look at equations such as $x^2=2$ (with its solutions $x=\sqrt{2}$ and $x=-\sqrt{2}$), and $x^4-10x^2+1=0$ (with its solutions $x=\sqrt{2}+\sqrt{3}$, $x=\sqrt{2}-\sqrt{3}$, $x=-\sqrt{2}+\sqrt{3}$, and $x=-\sqrt{2}-\sqrt{3}$). But we still can't solve *all* quadratic equations, because to solve the equation $x^2=-1$ we need to introduce another type of number, the square root of –1. Once we've done so, we can then solve any quadratic equation.

To illustrate this, let's consider three particular quadratic equations. For the quadratic equation $x^2-4x+3=0$, we can factorize directly:

$$x^2-4x+3=(x-3)(x-1)=0,$$

so there are two real solutions: $x=3$ and $x=1$.

For the quadratic equation $x^2 - 4x + 4 = 0$, we can again factorize directly:

$$x^2 - 4x + 4 = (x-2)(x-2) = 0,$$

so we have a repeated real solution: $x = 2$.

But to factorize the quadratic equation $x^2 - 4x + 5 = 0$, we need to bring in $i = \sqrt{-1}$:

$$x^2 - 4x + 5 = (x-2-i)(x-2+i) = 0,$$

so there are two complex solutions: $x = 2+i$ and $x = 2-i$.

To explore the differences among these solutions, let's recall the quadratic equation formula:

$$\text{if } ax^2 + bx + c = 0, \text{ then } x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Taking $a = 1$ and $b = -4$, as in the above three equations, we deduce that the equation $x^2 - 4x + c = 0$ has the solutions

$$x = \{4 \pm \sqrt{(16-4c)}\}/2 = 2 \pm \sqrt{(4-c)}.$$

When $c = 3$, we have $x = 2 \pm \sqrt{1}$, giving the two solutions $x = 3$ or $x = 1$, as before.

When $c = 4$, we have $x = 2 \pm \sqrt{0}$, giving the single solution $x = 2$.

When $c = 5$, we have $x = 2 \pm \sqrt{-1}$, giving the two solutions $x = 2+i$ and $x = 2-i$.

If we now draw the graphs of these three quadratic equations (see Figure 67 overleaf), we find that

the curve $y = x^2 - 4x + 3$ crosses the x-axis twice (when $x = 3$ and $x = 1$);
the curve $y = x^2 - 4x + 4$ just touches the x-axis (when $x = 2$);
the curve $y = x^2 - 4x + 5$ (with complex solutions) misses the x-axis altogether, so the solutions are truly imaginary: they're there, but you can't see them!

The fundamental theorem of algebra

What happens for higher-degree equations, such as

$$x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60 = 0?$$

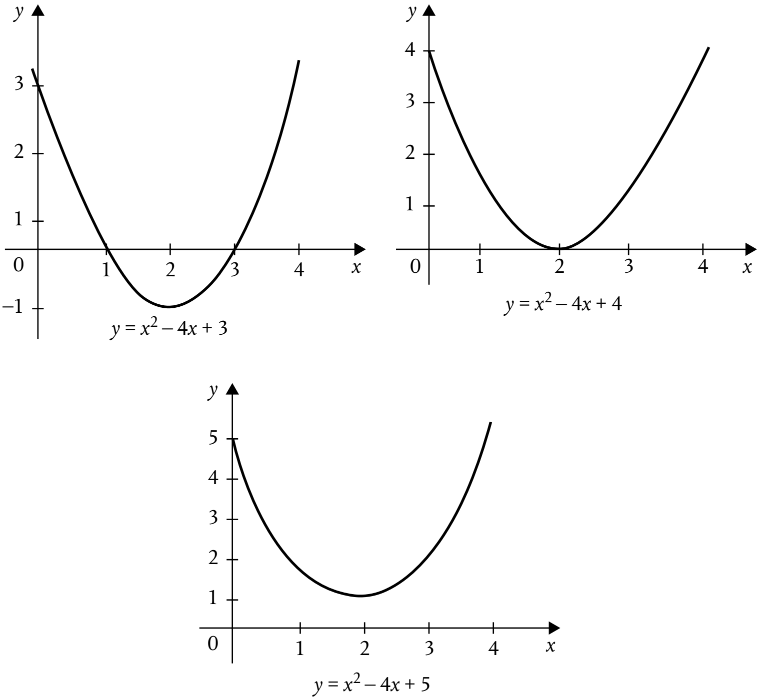


Fig. 67. The graphs of three quadratic equations

Can this be solved? If so, can we solve it with just real and complex numbers, or do we need to introduce yet another type of number?

Around 1700 there was some discussion about what forms the solutions of these more complicated equations might take. By this time, polynomial equations of degrees 1, 2, 3, and 4 (those involving terms up to x^4) had been solved, as we'll see, but what about equations of degree 5 or more, which no one could solve in general? There seemed to be several scenarios:

One can solve *all* equations using only real and complex numbers.

One may need to introduce new 'hyper-complex' numbers to solve some equations.

Some equations might have solutions that aren't numbers and don't behave like them.

Some equations might not have solutions of any kind.

To get a feeling for which of these is actually the case, let's first try to take the square root of i – that is, we'll solve the equation $x^2 = i$. Do we need to introduce further 'numbers', or are our existing complex numbers sufficient? If the latter, then we can write $x = a + bi$, and so

$$x^2 = (a+bi)^2 = i, \text{ giving } (a^2 - b^2) + 2abi = i.$$

On comparing real and imaginary parts we see that

$$a^2 - b^2 = 0 \text{ and } 2ab = 1, \text{ giving } a = b = \sqrt{\tfrac{1}{2}} \text{ or } a = b = -\sqrt{\tfrac{1}{2}}.$$

So the solutions are $x = (1+i)/\sqrt{2}$ and $x = -(1+i)/\sqrt{2}$, and complex numbers are all we need in this case.

In fact, complex numbers are *always* enough to solve any polynomial equation. For example,

$$\begin{aligned} &x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60 \\ &\quad = (x^2 - 4x + 3)(x^2 - 4x + 4)(x^2 - 4x + 5) \\ &\quad = (x-1)(x-3)(x-2)^2(x^2 - 4x + 5) \\ &\quad = (x-1)(x-3)(x-2)^2(x-2-i)(x-2+i), \end{aligned}$$

so the solutions of the polynomial equation

$$x^6 - 12x^5 + 60x^4 - 160x^3 + 239x^2 - 188x + 60 = 0$$

are $x = 1, 3, 2$ (twice), $2+i$, and $2-i$.

This is a special case of what came to be known as the *fundamental theorem of algebra*. It can be stated in various ways:

Every polynomial $p(x)$ with real coefficients can be factorized into linear and quadratic polynomials with real coefficients.

Every polynomial $p(x)$ with real coefficients can be factorized completely into linear factors with complex coefficients.

Every polynomial equation $p(x) = 0$ has at least one real or complex solution.

Every polynomial equation of the form $p(x) = 0$ of degree n has exactly n real or complex solutions (as long as we count them appropriately).

For a long time these observations were folklore, but they seem to have been first stated formally in 1629 by the Flemish mathematician Albert Girard:

Every equation of algebra has as many solutions as the exponent of the highest term indicates.

But others also seemed aware of it. For example, René Descartes stated the result in his *Discourse on Method* in 1637.

It was not until the 18th century that the matter received any serious discussion – notably by Jean le Rond d'Alembert, Leonhard Euler, Joseph-Louis Lagrange, and Pierre-Simon Laplace. Carl Friedrich Gauss dismissed these earlier efforts and gave the first 'proof' of the fundamental theorem in his doctoral dissertation of 1799, but it too was deficient and not easy to patch up. In 1814 an attempt was made by the Swiss mathematician Jean-Robert Argand, whom we meet again later, but it was also incomplete. Gauss subsequently provided three corrected proofs – but the waters surrounding all these attempts are very murky and it is difficult to be sure who gave the first 'rigorous' proof.

The origins of *i*

Let's look briefly at some early attempts to solve equations.

The following problem in sexagesimal notation appeared on a Mesopotamian clay tablet dating from around 1800 BC:

I have subtracted the side of my square from the area: 14,30.
You write down 1, the coefficient.
You break off half of 1. 0;30 and 0;30 you multiply. You add 0;15 to 14,30.
Result 14,30;15.
This is the square of 29;30. You add 0;30, which you multiplied, to 29;30.
Result: 30, the side of the square.

Writing this in modern algebraic notation, interpreting 'the side of a square' as x and 'the area' as x^2, and rewriting all the sexagesimal numbers in decimal form (for example, $14{,}30 = (14 \times 60) + 30 = 870$ and $0;30 = \frac{1}{2}$), we obtain the quadratic equation

$$x^2 - x = 870.$$

The above steps then give us successively:

$$1, \frac{1}{2}, \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}, \frac{1}{4} + 870 = 870\frac{1}{4}, \sqrt{870\frac{1}{4}} = 29\frac{1}{2}, \frac{1}{2} + 29\frac{1}{2} = 30.$$

The solution $x = 30$ is correct, since $30^2 - 30 = 870$.

This is just one of a dozen or more similar problems on the same clay tablet, which may therefore have been used for teaching purposes. It turns out that if we apply the same operations to the general equation $x^2 - bx = c$, we get

$$b, \frac{1}{2}b, \frac{1}{2}b \times \frac{1}{2}b = \frac{1}{4}b^2, \frac{1}{4}b^2 + c, \sqrt{\frac{1}{4}(b^2 + c)},$$
$$\frac{1}{2}b + \sqrt{\frac{1}{4}(b^2 + c)} = \frac{1}{2}\left\{b + \sqrt{\left(b^2 + 4c\right)}\right\}.$$

This is the result obtained by the quadratic equation formula.

So the Mesopotamians of 4000 years ago could solve particular instances of what we now call 'quadratic equations', and they used essentially the same sequence of operations that we use today – but there were differences. For a start, their idea of subtracting a side of a square from the area makes no sense to us geometrically. Moreover, they seemed to be satisfied with finding just one solution: any suggestion that there might be others didn't arise, and here the other solution (–29) would have been meaningless to them anyway, because it's a negative number.

Somewhat later, in the Greek world of the first century AD, a certain amount of fudging was used when the square root of a negative quantity unexpectedly turned up. In his *Stereometria*, Heron of Alexandria was attempting to find the height h of a frustum of a pyramid (that is, a pyramid with its top chopped off), where the sides a and b of the base and the top and the slant edge-length c are given numbers (see Figure 68).

In one of his examples Heron took the lengths a and b to be 28 and 4 units and the slant edge-length c to be 15, presumably unaware that such a frustum is not physically possible. Here the appropriate formula turns out to be

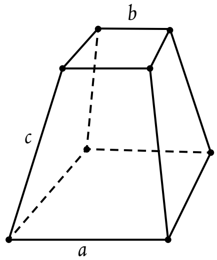


Fig. 68. Finding the height of a chopped-off pyramid

$$h = \sqrt{\{c^2 - \tfrac{1}{2}(a-b)^2\}},$$

which in this particular case is

$$\sqrt{\{15^2 - \tfrac{1}{2}(28-4)^2\}} = \sqrt{(225-288)} = \sqrt{-63}.$$

This answer was clearly far too dangerous to contemplate and it appeared in the *Stereometria* simply as $\sqrt{63}$.

In the 9th century AD Islamic scholars in Baghdad became interested in solving equations. In his book *Kitāb al-jabr w'al muqābalah,* from whose title we derive our word *algebra,* the Persian mathematician al-Khwārizmī presented a lengthy account of how to solve quadratic equations. Since negative numbers were still not considered meaningful, he split the equations into six types, corresponding (in modern notation) to the forms

$$ax^2 = bx,\ ax^2 = b,\ ax = b,\ ax^2 + bx = c,\ ax^2 + c = bx, \text{ and } ax^2 = bx + c,$$

where a, b, and c are *positive* constants. He then proceeded to solve particular instances of each type using a geometrical form of 'completing the square'; an example was the equation $x^2 + 10x = 39$, for which he found the solution $x = 3$. There was no discussion of the negative solution –13, and even a simpler equation such as $x + 1 = 0$ would have been considered as having no solutions.

Later, around the year 1100, the Persian poet and mathematician Omar Khayyám, best remembered for his classic collection of poems called the *Rubaiyat*, carried out a similar classification of cubic equations (those involving x^3) in which all the constants are positive; in this case there are fourteen different types.

In the early 16th century the algebraic scene moved to Italy. In 1545 Girolamo Cardano of Milan published an important algebra book, his *Ars Magna* (The Great Art) (see Figure 69), in which he explained how to solve a number of problems that give rise to algebraic equations. One of these problems asked how one can divide 10 into two parts whose product is 40. Taking the parts to be x and $10 - x$, he tried to solve the quadratic equation

$$x \times (10 - x) = 40 \ \text{ – that is, } x^2 - 10x + 40 = 0.$$

He obtained the solutions $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$, and could see no meaning to these. But after remarking

Nevertheless we will operate, putting aside the mental tortures involved,

he found that everything works out correctly:

$$(5 + \sqrt{-15}) + (5 - \sqrt{-15}) = 10$$

$$(5 + \sqrt{-15}) \times (5 - \sqrt{-15}) = 5^2 - (\sqrt{-15})^2 = 25 - (-15) = 40.$$

In view of the 'mental tortures' involved, Cardano was led to complain:

So progresses arithmetic subtlety the end of which is as refined as it is useless.

In his *Ars Magna* Cardano also showed how to solve cubic and quartic equations – equations of degrees 3 and 4. Following Scipione del Ferro, Niccolò Tartaglia, and other Italians, the method he described for solving a cubic equation of the form $x^3 + cx = d$ involved finding two other numbers u and v satisfying the equations

$$u - v = d \ \text{ and } \ uv = \left(\tfrac{1}{3}c\right)^3,$$

leading to a solution of the form $x = \sqrt[3]{u} - \sqrt[3]{v}$.

HIERONYMI CAR
DANI, PRÆSTANTISSIMI MATHE-
MATICI, PHILOSOPHI, AC MEDICI,

ARTIS MAGNÆ,

SIVE DE REGVLIS ALGEBRAICIS,
Lib. unus. Qui & totius operis de Arithmetica, quod
OPVS PERFECTVM
inſcripſit, eſt in ordine Decimus.

HAbes in hoc libro, ſtudioſe Lector, Regulas Algebraicas (Itali, de la Coſ ſa uocant) nouis adinuentionibus, ac demonſtrationibus ab Authore ita locupletatas, ut pro pauculis antea uulgò tritis, iam ſeptuaginta euaſerint. Neq; ſolum, ubi unus numerus alteri, aut duo uni, uerum etiam, ubi duo duobus, aut tres uni æquales fuerint, nodum explicant. Hunc aũt librum ideo ſeorſim edere placuit, ut hoc abſtruſiſſimo, & planè inexhauſto totius Arithmeticæ theſauro in lucem eruto, & quaſi in theatro quodam omnibus ad ſpectandum expoſito, Lectores incitarẽtur, ut reliquos Operis Perfecti libros, qui per Tomos edentur, tanto auidius amplectantur, ac minore faſtidio perdiſcant.

Fig. 69. Cardano's *Ars Magna*

For example, to solve the cubic equation $x^3 + 6x = 20$, he sought numbers u and v satisfying

$$u - v = 20 \text{ and } uv = \left(\tfrac{1}{3} \times 6\right)^3 = 8.$$

Because $v = u - 20$ we have

$$uv = u(u - 20) = u^2 - 20u = 8.$$

This is a quadratic equation which he easily solved to give

$$u = 10 + \sqrt{108}, \text{ so that } v = -10 + \sqrt{108}.$$

The solution for x then has the form $\sqrt[3]{u} - \sqrt[3]{v}$, which is

$$x = \sqrt[3]{\left(10 + \sqrt{108}\right)} - \sqrt[3]{\left(-10 + \sqrt{108}\right)}.$$

This is cumbersome, but if you now work this number out on a calculator, you'll get the much simpler answer 2, which clearly satisfies the original equation. But Cardano (having no calculator) seemed unable to perform the necessary simplifications.

The situation was clarified by Rafael Bombelli, an engineer from Bologna who was an expert in draining swampy marshes. In his *Algebra* of 1572 he applied Cardano's method to the cubic equation $x^3 = 15x + 4$ and found that

$$x = \sqrt[3]{\left(2 + \sqrt{-121}\right)} + \sqrt[3]{\left(2 - \sqrt{-121}\right)} = \sqrt[3]{(2 + 11i)} + \sqrt[3]{(2 - 11i)}.$$

This involves imaginary numbers – but this equation actually has three *real* roots, $4, -2 + \sqrt{3}$, and $-2 - \sqrt{3}$, with no imaginary numbers in sight! This seemed paradoxical; indeed Leibniz, who was dissatisfied with Bombelli's explanations, later asked:

> How can it be that a real quantity, a root of the proposed equation, is expressed by the intervention of an imaginary?

But after some investigation Bombelli noticed that

$$2 + 11i = (2 + i)^3 \text{ and } 2 - 11i = (2 - i)^3,$$

and so, on taking the cube roots, he found that

$$x=(2+i)+(2-i)=4,$$

as expected. He was then able to calculate the other two roots. As the French mathematician Jacques Hadamard was later to observe:

> The shortest path between two truths in the real domain passes through the complex domain.

But Bombelli went even further, and proposed the following general rules for dealing with these complex numbers:

> Plus times plus of minus, makes plus of minus.
> Minus times minus of minus, makes plus of minus.
> Plus of minus times plus of minus, makes minus.
> Plus of minus times minus of minus, makes plus.

We can interpret these by writing +1 for 'plus', –1 for 'minus', + *i* for 'plus of minus', and –*i* for 'minus of minus', giving us the rules

$$(+1)\times(+i)=+i;\;\; (-1)\times(-i)=+i;\;\; (+i)\times(+i)=-1;\;\; (+i)\times(-i)=+1.$$

Picturing complex numbers

How can we visualize complex numbers? Before answering this we first look at some geometrical constructions of René Descartes and John Wallis.

Constructing square roots

In his *Discourse on Method* of 1637 Descartes presented geometrical constructions for various algebraic operations. For example, following Euclid, he gave the following ruler-and-compasses construction for finding the square root of a positive quantity (see Figure 70):

> To find the square root of *GH*, draw *FG* with length 1 and bisect the line *FH* at the point *K*.
> Draw the circle with centre *K* and radius *FK*.

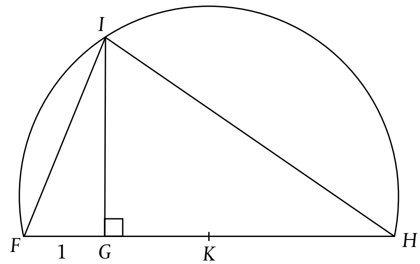


Fig. 70. Constructing a square root

Draw the line at G perpendicular to FH, and let I be the point where this perpendicular line meets the circle.
Then GI is the required square root.
For, by the similar triangles ΔFGI and ΔGIH, we have $GI/FG = GH/GI$, and so $GI^2 = FG \times GH = GH$.

Descartes also gave the following construction for finding the positive solution of the quadratic equation $x^2 = ax + b^2$ (see Figure 71):

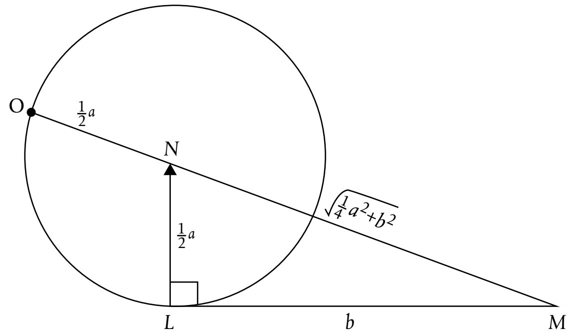


Fig. 71. Solving a quadratic equation

Draw a line LM of length b and draw the line LN perpendicular to LM and of length $\frac{1}{2}a$.

Draw the circle with centre N and radius LN.
Draw the line MN and extend it to the point O on the circle.

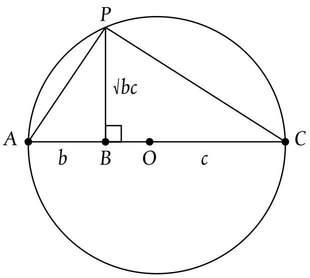
Then OM is the required solution.
For, by the Pythagorean theorem, $OM = ON + NM = \frac{1}{2}a + \sqrt{\left(\frac{1}{4}a^2 + b^2\right)}$.

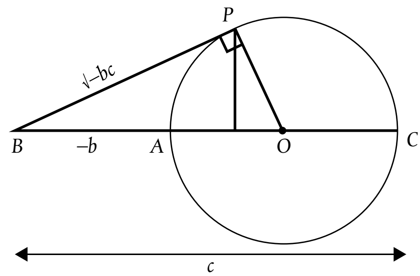
Influenced by Descartes, other 17th-century mathematicians attempted to picture algebraic ideas. In particular, Wallis gave a similar construction to that of Descartes for constructing the square root of a product *bc*, when *b* and *c* are both positive. As he wrote in his *Treatise on Algebra* of 1685:

Forward from A, I take $AB = +b$; and Forward from thence, $BC = +c$;
(making $AC = +AB + BC = +b + c$, the Diameter of a Circle:)
Then is the Sine, or Mean Proportional $BP = \sqrt{+bc}$.

This corresponds to drawing a circle with diameter AC of length $b + c$ and constructing a perpendicular from the point at distance b from A (see Figure 72); the length of this perpendicular is then the required square root $\sqrt{bc}$.

Wallis then tried to modify this process in order to construct the square root of *bc* when *b* is negative and *c* is positive (see Figure 73):





Figs. 72, 73. Finding the square root of *bc* when *b* is positive and when *b* is negative

But if Backward from A, I take $AB = -b$; and then Forward from that B, $BC = +c$;
(making $AC = -AB + BC = -b + c$, the Diameter of the Circle:)
Then is the Tangent or Mean proportional $BP = \sqrt{-bc}$.

This time *b* is measured to the left of *A* and the length of the tangent *BP* is the required square root $\sqrt{-bc}$.

Wallis also attempted a construction that hinted at the idea of an imaginary number being at right angles to a real one, but didn't quite get there.

The complex plane

Pictorial representations of complex numbers were first introduced by the self-taught Norwegian–Danish surveyor Caspar Wessel. Unfortunately, his article (in Danish) was overlooked for a hundred years and his ideas had no influence on the development of the subject. Similar representations may also have been obtained by the Frenchman Henri Dominique Truel and by Gauss, but neither published his results at this time.

In 1797 Wessel presented a paper to the Royal Danish Academy of Sciences 'On the analytic representation of direction', in which he outlined the idea of what we now call the *complex plane*. He considered each complex number $a + bi$ as the point in the plane with Cartesian coordinates (a, b), or as the vector from the origin $(0, 0)$ to this point. Figure 74 shows the four points (1, 2), (3, 1), (–2, 1), and (3, –2) that correspond to the complex numbers $1+2i$, $3+i$, $-2+i$, and $3-2i$. The x-axis of real numbers $a=(a, 0)$ is called the *real axis*, and the y-axis of imaginary numbers $bi=(0, b)$ is called the *imaginary axis*.

To add two complex numbers, we add the corresponding vectors using the parallelogram law. This corresponds to the *addition rule*:

$$(a+bi)+(c+di)=(a+c)+(b+d)i.$$

For example, as illustrated in Figure 75,

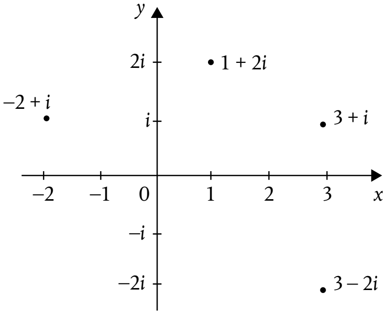


Fig. 74. Points in the complex plane

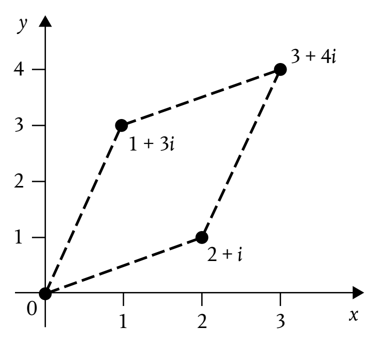


Fig. 75. The addition rule

$$(1+3i)+(2+i)=3+4i.$$

We can also write each non-zero complex number $z=a+bi$ in *polar form* as $[r, \theta]$, where

$$r=|z|=|a+bi|=\sqrt{(a^2+b^2)}$$

is the modulus of z, the length of the line segment from the origin O to the number $a+bi$, and

$$\theta=\arctan b/a$$

is the argument of z, the angle (in radians) between the line segment from O to $a+bi$ and the positive x-axis (see Figure 76).

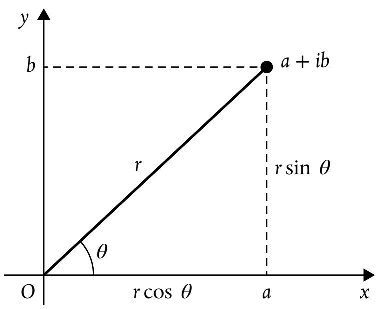


Fig. 76. The polar form of a complex number

We note that:

For a non-zero point $a=(a, 0)$ on the real axis, $r=a$ and $\theta=0$ (or 2π, or -2π, or 100π, or any other integer multiple of 2π);

for a non-zero point $bi=(0, b)$ on the imaginary axis, $r=b$ and $\theta=\pi/2$ (or $\pi/2$ + any integer multiple of 2π).

We also see that the polar point $[r, \theta]$ corresponds to the complex number $r(\cos\theta+i\sin\theta)$ with Cartesian coordinates $(r\cos\theta, r\sin\theta)$.

We'll need this result in Chapter 6, noting in particular that there are infinitely many possible values for the angle θ, the argument of z, all differing by integer multiples of 2π.

Using the polar form we can easily multiply two complex numbers together: we simply multiply the corresponding moduli r and add the corresponding angles θ. This corresponds to the *multiplication rule*:

i

$$[r,\theta]\times[s,\varphi]=[rs,\theta+\varphi].$$

This is because, by the addition formulas for cosine and sine,

$$\begin{aligned}r(\cos\theta+i\sin\theta)\times s(\cos\varphi+i\sin\varphi)&=rs(\cos\theta+i\sin\theta)(\cos\varphi+i\sin\varphi)\\&=rs\{(\cos\theta\cos\varphi-\sin\theta\sin\varphi)+i(\sin\theta\cos\varphi+\cos\theta\sin\varphi)\}\\&=rs\{\cos\,(\theta+\varphi)+i\sin\,(\theta+\varphi)\}.\end{aligned}$$

It follows from the multiplication rule that multiplying a complex number $z=[r,\theta]$ by $i=[1,\pi/2]$ gives $[r,\theta+\pi/2]$, which corresponds to rotating z anticlockwise through a right angle. For example, multiplying the complex number $3+2i$ by i to give $-2+3i$ corresponds to such a rotation (see Figure 77). (As the telephone operator said: 'The number you've dialled is purely imaginary: please rotate your phone through 90° and try again'.)

Similarly, multiplying by i twice gives a rotation through two right angles – that is, a rotation of the plane through $\pi=180°$. This rotation sends each complex number $a+bi$ to its negative $-a-bi$, corresponding to the rule $i\times i=-1$ (see Figure 78).

Using the multiplicative rule, Wessel also calculated the powers of complex numbers. For example, on taking $r=s=1$ and $\varphi=\theta$, we have

$$[1,\theta]\times[1,\theta]=[1,2\theta]\ \text{ or }\ (\cos\theta+i\sin\theta)^2=\cos2\theta+i\sin2\theta.$$

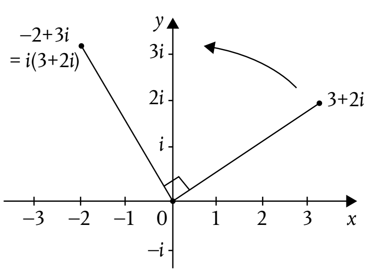


Fig. 77. Multiplying by i

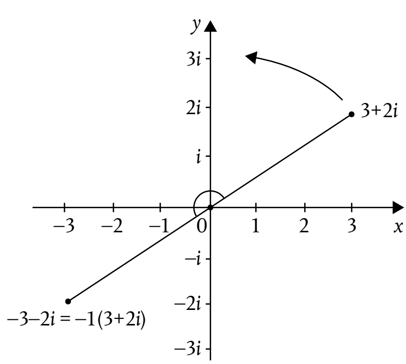


Fig. 78. Multiplying by $i^2=-1$

On replacing 2θ by θ and taking square roots, we then have

$$\cos \tfrac{1}{2}\theta + i \sin \tfrac{1}{2}\theta = (\cos \theta + i \sin \theta)^{1/2}.$$

These are special cases of an important result known as *De Moivre's theorem*, one form of which Abraham De Moivre discovered in the early 18th century:

For any number n, $(\cos \theta + i \sin \theta)^n = \cos n\theta + i \sin n\theta$.

Note that, when we replace θ by $-\theta$, then $\cos(-\theta) = \cos \theta$ and $\sin(-\theta) = -\sin \theta$, and so:

For any number n, $(\cos \theta - i \sin \theta)^n = \cos n\theta - i \sin n\theta$.

We shall need these results in Chapter 6.

We'll conclude this section by finding the complex nth roots of the number 1 for each positive integer n. For example,

$n = 2$: the square roots satisfy $z^2 - 1 = (z-1)(z+1) = 0$,
and are $z = 1$ and $z = -1$.

Here are some further complex roots of 1:

$n = 3$: the cube roots satisfy $z^3 - 1 = (z-1)(z^2 + z + 1) = 0$,
and are $z = 1$, $z = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i$, and $z = -\frac{1}{2} - \frac{1}{2}\sqrt{3}i$.

$n = 4$: the fourth roots satisfy $z^4 - 1 = (z-1)(z+1)(z^2 + 1) = 0$,
and are $z = 1$, $z = -1$, $z = i$, and $z = -i$.

$n = 6$: the sixth roots satisfy

$$z^6 - 1 = (z-1)(z+1)(z^2 - z + 1)(z^2 + z + 1) = 0,$$

and are $z = 1$, $z = -1$, $z = \frac{1}{2} + \frac{1}{2}\sqrt{3}i$, $z = \frac{1}{2} - \frac{1}{2}\sqrt{3}i$,

$$z = -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{3}i, \text{ and } z = -\tfrac{1}{2} - \tfrac{1}{2}\sqrt{3}i.$$

By the Pythagorean theorem we can represent the unit circle with centre 0 and radius 1 by the simple equation $|z| = 1$. If we plot these complex roots of 1, we find that they always lie on the unit circle at the corners of a regular polygon (see Figure 79).

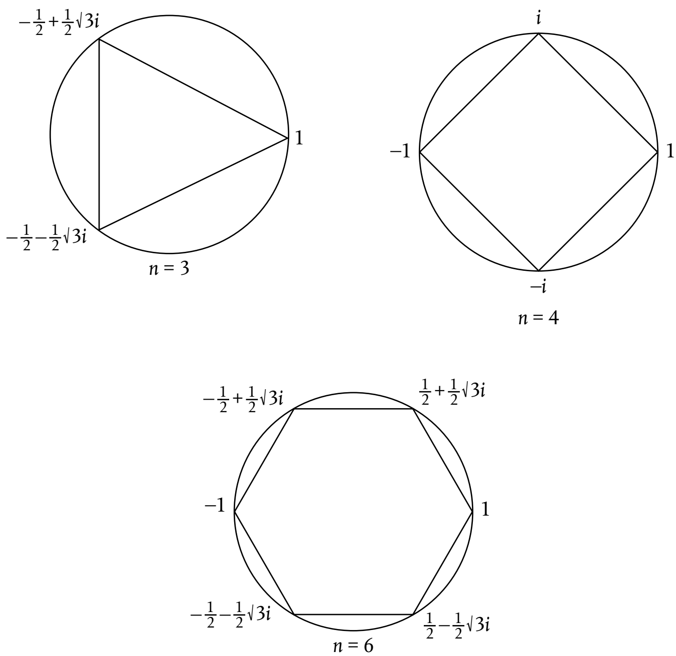


Fig. 79. The complex *n*th roots of 1, for $n = 3$, 4, and 6

To find the complex *n*th roots of 1 for an arbitrary value of *n*, we note that, by De Moivre's theorem with $\theta = 2\pi / n$,

$$(\cos\ 2\pi/n + i\sin\ 2\pi/n)^n = \cos 2\pi + i\sin\ 2\pi = 1 + 0i = 1,$$

and more generally, for any integer *k*,

$$(\cos\ 2k\pi/n + i\sin\ 2k\pi/n)^n = \cos\ 2k\pi + i\sin\ 2k\pi = 1 + 0i = 1.$$

So the *n*th roots of 1 are the complex numbers

$$\cos 2k\pi/n + i\ \sin 2k\pi/n, \text{ for } k = 0, 1, 2, \ldots, n-1.$$

For example, taking $n = 3$ and $k = 0, 1$, and 2, we find the cube roots of 1 to be

$$\cos\ 0 + i\ \sin 0,\ \cos\ 2\pi/3 + i\ \sin 2\pi/3,\ \text{ and }\ \cos\ 4\pi/3 + i\ \sin\ 4\pi/3,$$

which turn out to be 1, $\frac{1}{2}\left(-1+\sqrt{3}i\right)$, and $\frac{1}{2}\left(-1-\sqrt{3}i\right)$, as above.

Argand and Gauss

The complex plane is often called the *Argand diagram*, although this is not historically accurate, as we've seen. But the ideas were already 'in the air' and were rediscovered more than once – in particular by the Swiss-born bookshop owner Jean-Robert Argand.

In 1806 Argand wrote an *Essay on the Geometrical Interpretation of Imaginary Quantities*, which he printed privately for his friends without his name on the title page. He sent a copy to the famous French mathematician Adrien-Marie Legendre, who in turn sent it to another mathematician called François Français, who then died shortly after. Fortunately, his brother Jacques, also a mathematician, was looking through François's papers. Intrigued by the results that they contained he published his own paper on the subject, mentioning Legendre's letter and inviting the originator of the ideas to make himself known. Again fortunately, Argand learned about this request and did so.

In Germany the complex plane is often called the *Gaussian plane*. Gauss had already been working on related ideas for some years but never told anyone, claiming in 1812 that

> I have in my papers many things for which I could perhaps lose the priority of publication, but you know, I prefer to let things ripen.

Gauss finally committed himself on the subject in 1831, and such was his reputation that complex numbers received a great boost. As he observed:

> That this subject has hitherto been considered from the wrong point of view and surrounded by a mysterious obscurity, is to be attributed largely to an ill-adapted notation. If, for instance, $+1, -1, \sqrt{-1}$ had been called direct, inverse, and lateral units, instead of positive, negative, and imaginary (or even impossible), such an obscurity would have been out of the question.

In particular, he studied the mathematical properties of what are now called the *Gaussian integers*. These are complex numbers of the form $a+bi$ where a and b are both integers, and they behave surprisingly like the ordinary integers; for example, we can factorize them into 'primes' in only one way, just like factorizing an ordinary whole number into its prime factors.

Generalizing complex numbers

The idea of representing each complex number $x+iy$ as a point (x, y) in the plane was developed by Sir William Rowan Hamilton, the Astronomer Royal of Ireland.

Hamilton's quaternions

Even as late as the 1830s there was still a great deal of confusion about complex numbers. This situation had persisted for several centuries and it fell to Hamilton to diffuse much of the suspicion.

Hamilton proposed that the complex numbers $a+bi$ should be thought of more concretely as *pairs of real numbers* (a, b), which we combine by using certain specified rules. These rules he took to be

$$(a, b)+(c, d)=(a+c, b+d) \text{ and } (a, b)\times(c, d)=(ac-bd, ad+bc),$$

corresponding to the equations

$$(a+bi)+(c+di)=(a+c)+(b+d)i,$$
$$(a+bi)\times(c+di)=(ac-bd)+(ad+bc)i.$$

Here the pair $(a, 0)$ corresponds to the real number a, the pair $(0, b)$ corresponds to the imaginary number bi, and the equation $(0,1)\times(0,1)=(-1,0)$ a restatement of the equation $i\times i=-1$. Hamilton's approach was completely successful, and the complex numbers at last became almost universally accepted.

Hamilton then tried to extend his ideas to three dimensions. If the points (a, b) of the plane correspond to complex numbers of the form $a+bi$, where $i^2=-1$, then the points (a, b, c) of three-dimensional space should surely correspond to objects of the form $a+bi+cj$, where i and j are both taken to be imaginary square roots of -1, so that $i^2=j^2=-1$. Certainly, addition works well:

$$(a+bi+cj)+(d+ei+fj)=(a+d)+(b+e)i+(c+f)j.$$

But he couldn't make multiplication work:

$$(a+bi+cj)\times(d+ei+fj)=(ad-be-cf)+(ae+bd)i+(af+cd)j+(bf+ce)ij.$$

This gives four terms, rather than three, and the problem was to get rid of the last term, involving the product ij. We cannot let $ij=0$, because then

$$0=(ij)^2=i^2\times j^2=(-1)\times(-1)=1.$$

Hamilton tried everything, such as writing $ij=1$ or $ij=-1$, but nothing seemed to work, and in a letter to one of his sons he later recalled:

> Every morning, on my coming down to breakfast, your little brother William Edwin and yourself used to ask me, 'Well Papa, can you multiply triples?' Whereto I was obliged to reply, with a shake of the head: 'No, I can only add and subtract them'.

Hamilton struggled with his triples for several years, until one day he took a walk along the Royal Canal in Dublin:

> As I was walking with Lady Hamilton to Dublin, and came up to Brougham Bridge, I then and there felt the galvanic circuit of thought close; and the sparks which fell from it were the fundamental equations exactly as I have used them ever since.
>
> I pulled out on the spot a pocket book and made an entry…it is fair to say that this was because I felt a problem to have been at that moment solved – an intellectual want relieved which had haunted me for at least fifteen years since.

Hamilton was so excited that he carved his fundamental equations on the bridge (see Figure 66 which opens this chapter).

What Hamilton had eventually come up with were his *quaternions*: these are objects of the form $a+bi+cj+dk$, where a, b, c, and d are real numbers and i, j, and k are all imaginary square roots of -1, so $i^2=j^2=k^2=-1$. Addition worked as expected, but in order to make multiplication work he had to abandon the familiar 'commutative law' of arithmetic, according to which $X \times Y = Y \times X$, for all X and Y; for example, $3 \times 4 = 4 \times 3$.

Instead, Hamilton needed to impose the following non-commutative multiplication rules:

whenever they occur:

replace $i \times j$ by k, but replace $j \times i$ by $-k$,

replace $j \times k$ by i, but replace $k \times j$ by $-i$,

replace $k \times i$ by j, but replace $i \times k$ by $-j$.

We can depict these rules diagrammatically (see Figure 80): when we travel clockwise around the circle, the results are positive; when we go counter-clockwise, they're negative. We can also express the rules more concisely as

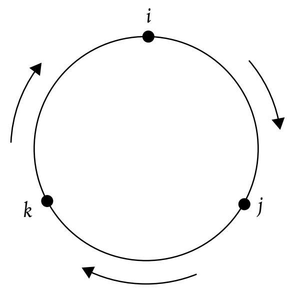


Fig. 80.

$$i^2 = j^2 = k^2 = ijk = -1.$$

There's now a plaque on Brougham Bridge that commemorates the discovery of these equations (see Figure 81), and over the years the Irish Post Office has issued a number of stamps featuring them (see Figure 82).

Hamilton's quaternions have many applications in geometry and physics. They can be used to represent rotations in three and four dimensions and so arise in many contexts, such as the theory of relativity, film animation, and the tracking of satellites.

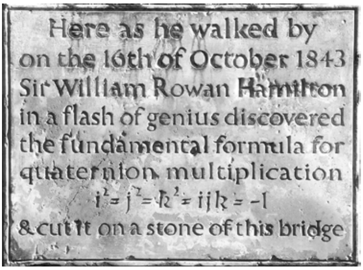


Fig. 81. The plaque at Brougham Bridge

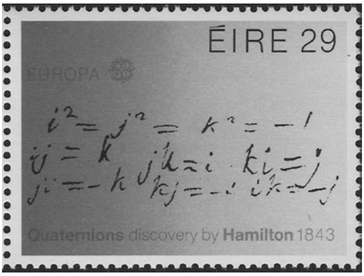


Fig. 82. An Irish postage stamp of 1983

But initial reactions to them were mixed. William Thomson (later Lord Kelvin) was unenthusiastic:

Quaternions came from Hamilton after his really good work had been done; and though beautifully ingenious, have been an unmixed evil to those who have touched them in any way, including Clerk Maxwell.

And yet James Clerk Maxwell had written some years earlier:

The invention of the calculus of quaternions is a step towards the knowledge of quantities related to space which can only be compared, for its importance, with the invention of triple coordinates by Descartes. The ideas of this calculus, as distinguished from its operations and symbols, are fitted to be of the greatest use in all parts of science.

Octonions

Can we go any further? We've described number systems with one term (the real numbers), two terms (the complex numbers), and four terms (the quaternions). Are there similar systems with a higher number of terms? Augustus De Morgan was unsure:

I think the time may come when double algebra [the algebra of pairs of numbers (complex numbers)] will be the beginner's tool; and quaternions will be where double algebra is now. The Lord only knows what will come above the quaternions.

It turns out that just one further system 'comes above the quaternions' – but only if we agree to abandon yet another arithmetical law. These new numbers are the *octonions*, or *octaves*, introduced independently in the 1840s by John Graves (a friend of Hamilton) and the English mathematician Arthur Cayley. Each octonion consists of eight terms of the form

$$a + bi + cj + dk + el + fm + gn + ho,$$

where a, b, c, d, e, f, g, and h are real numbers and i, j, k, l, m, n, and o are square roots of -1:

$$i^2 = j^2 = k^2 = l^2 = m^2 = n^2 = o^2 = -1.$$

i

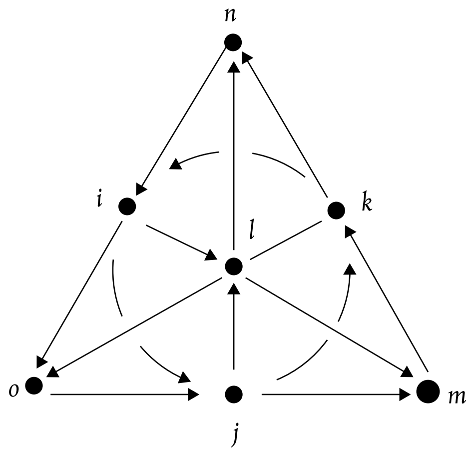


Fig. 83. Multiplying octonions

As before, such objects can be added term by term, but multiplication is more complicated, being defined by seven sets of equations, such as:

$$i = jk = lm = on = -kj = -ml = -no.$$

The full set of equations is rather indigestible, but we can use the diagram in Figure 83 and 'follow the arrows' to derive them all. For example,

whenever we come across $l \times m$, we replace it by i.

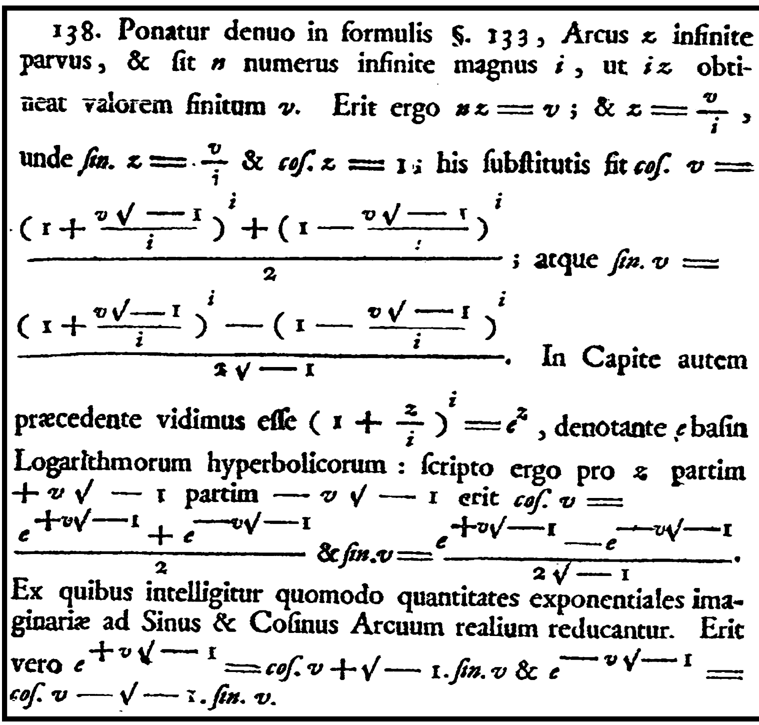
But when the arrows are reversed, we insert a minus sign:

whenever we come across $m \times l$, we replace it by $-i$,

and so on.

All multiplications can then be successfully carried out, but we have to abandon the commutative law as before, and we also have to lose the 'associative law', according to which $(X \times Y) \times Z = X \times (Y \times Z)$, for all X, Y, and Z; for example, $(3 \times 4) \times 5 = 3 \times (4 \times 5)$.

Can we go yet further? We've produced number systems with one, two, four, and eight terms, so might there be a similar system with sixteen terms? With no more arithmetical laws to abandon, the answer must surely be 'no', and this was eventually confirmed in 1898 by the German mathematician Adolf Hurwitz. We have indeed come to the end of this particular road.



138. Ponatur denuo in formulis §. 133, Arcus z infinite parvus, & ſit n numerus infinite magnus i, ut iz obtineat valorem finitum v. Erit ergo $nz = v$; & $z = \frac{v}{i}$, unde $\mathit{ſin.}\, z = \frac{v}{i}$ & $\mathit{coſ.}\, z = 1$; his ſubſtitutis fit $\mathit{coſ.}\, v =$

$$\frac{\left(1 + \frac{v\sqrt{-1}}{i}\right)^i + \left(1 - \frac{v\sqrt{-1}}{i}\right)^i}{2};$$

atque $\mathit{ſin.}\, v =$

$$\frac{\left(1 + \frac{v\sqrt{-1}}{i}\right)^i - \left(1 - \frac{v\sqrt{-1}}{i}\right)^i}{2\sqrt{-1}}.$$

In Capite autem præcedente vidimus eſſe $\left(1 + \frac{z}{i}\right)^i = e^z$, denotante e baſin Logarithmorum hyperbolicorum: ſcripto ergo pro z partim $+v\sqrt{-1}$ partim $-v\sqrt{-1}$ erit $\mathit{coſ.}\, v =$

$$\frac{e^{+v\sqrt{-1}} + e^{-v\sqrt{-1}}}{2} \;\&\; \mathit{ſin.}\, v = \frac{e^{+v\sqrt{-1}} - e^{-v\sqrt{-1}}}{2\sqrt{-1}}.$$

Ex quibus intelligitur quomodo quantitates exponentiales imaginariæ ad Sinus & Coſinus Arcuum realium reducantur. Erit vero $e^{+v\sqrt{-1}} = \mathit{coſ.}\, v + \sqrt{-1}.\, \mathit{ſin.}\, v$ & $e^{-v\sqrt{-1}} = \mathit{coſ.}\, v - \sqrt{-1}.\, \mathit{ſin.}\, v$.

Fig. 84. Euler's identity, from his *Introductio in Analysin Infinitorum*

CHAPTER 6

$e^{i\pi} + 1 = 0$

Euler's equation

We've looked in turn at each of the numbers 1, 0, π, e, and i, each with its rich associations and its own story. Now is the time to bring them together. An equation that combines such varied entities must be profound indeed.

To motivate Euler's equation, we note the following interconnections:

the trigonometric functions $y = \sin x$ and $y = \cos x$ are related to a circle;

the exponential and logarithmic functions $y = e^x$ and $y = \ln x$ are related to a hyperbola;

the hyperbola and circle are related to each other (because both are conic sections).

So are there any direct relationships between the exponential and logarithmic functions and the trigonometric ones?

There are indeed no *real* reasons why there should be any such relationship, but there are *complex* reasons! Introducing complex numbers leads to such connections, and realizing this was one of Euler's greatest achievements. In his *Introductio* of 1748 he presented a proof of *Euler's identity*:

$$e^{ix} = \cos x + i \sin x.$$

But by then, several other results linking these functions were already 'in the air'. These included the following 'near misses' to the discoveries of Euler's identity and Euler's equation.

Two near misses

In 1702 Johann Bernoulli presented a formula for the area of a sector of a circle which implies the following equation linking π, i, and the logarithm of a negative number:

$$\pi = \frac{1}{i}\ \ln(-1).$$

He also obtained the identity

$$\arctan x = \frac{i}{2}\ \ln\frac{i+x}{i-x},$$

which indicates that inverse tangents and logarithms of complex numbers are, in some sense, the same. And around 1712 Roger Cotes was investigating the surface areas of ellipsoids and discovered that, for any angle φ,

$$\ln\left(\cos\varphi + i\sin\varphi\right) = i\varphi.$$

In this section we look at these near misses of Johann Bernoulli and Roger Cotes, before turning to the fundamental contributions of Leonhard Euler.

Johann Bernoulli

As we saw in Chapter 4, the logarithm function $y=\ln\ x$ is defined for all positive values of x. But can it be defined when x is negative? This question caused much disagreement between Gottfried Leibniz, who believed the logarithm of a negative number to be 'impossible', and Johann Bernoulli, who used the basic equation

$$\ln a + \ln b = \ln\ (a \times b)$$

to prove that, for any number x,

$$\begin{aligned} 2\ \ln(-x) &= \ln(-x) + \ln(-x) = \ln(-x \times -x) = \ln(x^2) \\ &= \ln(x \times x) = \ln\ x + \ln\ x = 2\ \ln\ x, \end{aligned}$$

and so, for all x, $\ln(-x) = \ln x$. In particular, $\ln(-1) = \ln 1 = 0$.

In 1702 Bernoulli was investigating the area of a sector of a circle of radius a with its centre at the origin O – the shaded area in Figure 85 bounded by the x-axis and the line joining O to the point (x, y) on the circle: he found this area to be

$$\frac{a^2}{4i}\ln\frac{x+iy}{x-iy}.$$

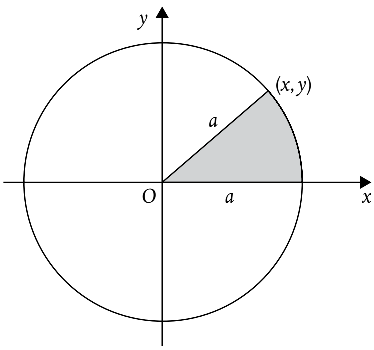


Fig. 85. The area of a sector

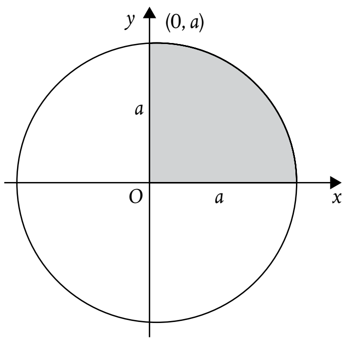


Fig. 86. The area of a sector

Leaving aside for the time being the meaning of the logarithm of a complex number, Euler later observed that this formula simplifies to $a^2/4i\ln(-1)$ when $x=0$. Because such a sector clearly has a non-zero area (see Figure 86), he deduced that the logarithm of –1 cannot be zero, contradicting Bernoulli's result above. Moreover, since this sector is a quarter-circle with area $\pi a^2/4$,

$$\frac{\pi a^2}{4}=\frac{a^2}{4i}\ln(-1), \text{ and so } \ln(-1)=i\pi.$$

Although Euler wrote down this last result explicitly, he doesn't seem to have taken exponentials to deduce that $e^{i\pi}=-1$, which is Euler's equation. Indeed, Euler often credited Bernoulli with discovering this value for ln (–1), but Bernoulli didn't include it in his 1702 paper or in any later work, continuing to insist that $\ln(-1)=0$.

$$e^{i\pi}+1=0$$

Bernoulli's 1702 paper also contained an unfinished calculation involving an arctan integral, leading him to assert that 'imaginary logarithms express real circular functions'. The details are given in Box 10 for those familiar with calculus.

Box 10: Bernoulli's arctan result

A standard result in calculus is that

$$\arctan x=\int\frac{1}{1+x^2}\,dx.$$

If we now allow the use of complex numbers, we can split $1/(1+x^2)$ into partial fractions:

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

because $1+x^2=(1-ix)(1+ix)$. We therefore have:

$$\begin{aligned}\arctan x&=\frac{1}{2}\int\frac{1}{1+x^2}\,dx=\frac{1}{2}\int\frac{1}{1-ix}\,dx+\frac{1}{2}\int\frac{1}{1+ix}\,dx\\&=\frac{1}{2}\int\frac{i}{i+x}\,dx+\frac{1}{2}\int\frac{i}{i-x}\,dx\\&=\frac{i}{2}\ln\,(i+x)-\frac{i}{2}\ln\,(i-x)=\frac{i}{2}\ln\left(\frac{i+x}{i-x}\right).\end{aligned}$$

Roger Cotes

In Chapter 3 we described Roger Cotes's introduction of radian measure. During his short lifetime he produced only one published work, a paper entitled *Logometria*, which appeared in 1714 and included a detailed discussion of the logarithmic spiral; this curve with polar equation $r=e^{k\theta}$ (where k is a constant) occurs in nature and had previously been studied by Jakob Bernoulli, Johann's elder brother. Cotes died at the age of 33 and, as we saw in Chapter 3, his mathematical works, including his findings

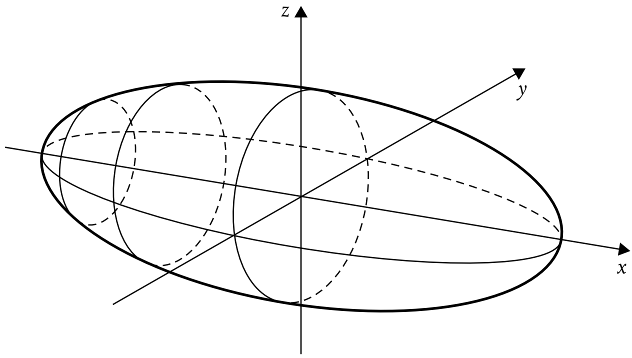


Fig. 87. Cotes's ellipsoid

on logarithms and geometric curves, were published by his cousin in the book *Harmonia Mensuarum*.

Cotes was attempting to find the surface area of the ellipsoid obtained by rotating an ellipse around the x-axis (see Figure 87). The details are somewhat complicated, but he managed to find two mathematical expressions for the required area – one involving logarithms and the other involving the inverse sine function. Both of these expressions involved an angle φ, where $\cos\varphi = b/a$ and a and b are the half-lengths of the ellipse's axes. He first proved that the surface area is a certain multiple of $\ln(\cos\varphi + i\sin\varphi)$, and then proved it to be the same multiple of $i\varphi$. Equating these results he deduced the identity

$$\ln(\cos\varphi + i\sin\varphi) = i\varphi,$$

which gives a connection between logarithms and trigonometric functions. If he'd then taken exponentials (which he didn't), he'd have discovered Euler's identity in the form

$$e^{i\varphi} = \cos\varphi + i\sin\varphi.$$

Another near miss!

Euler's identity

We now come to Euler's most celebrated result, relating the exponential function $y=e^x$ and the trigonometric functions $y=\cos x$ and $y=\sin x$. Recall that these functions can be expanded as power series, valid for all values of x:

$$e^x=1+\frac{1}{1!}x+\frac{1}{2!}x^2+\frac{1}{3!}x^3+\frac{1}{4!}x^4+\frac{1}{5!}x^5+\frac{1}{6!}x^6+\frac{1}{7!}x^7+\ldots,$$
$$\cos x=1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\frac{1}{6!}x^6+\ldots,\quad \text{and}$$
$$\sin x=\frac{1}{1!}x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\frac{1}{7!}x^7+\ldots.$$

Because the exponential function shoots off to infinity as x becomes large, whereas the cosine and sine functions forever oscillate between 1 and -1, there seems to be no relationship between these functions.

But as Euler discovered in 1737, there is indeed a fundamental connection if we allow ourselves to introduce the complex number i, the square root of -1. One way of seeing this, as Euler showed, is to start with the power series for e^x, and to replace x by ix:

$$e^{ix}=1+\frac{1}{1!}(ix)+\frac{1}{2!}(ix)^2+\frac{1}{3!}(ix)^3+\frac{1}{4!}(ix)^4+\frac{1}{5!}(ix)^5+\frac{1}{6!}(ix)^6+\frac{1}{7!}(ix)^7+\ldots.$$

Since $i^2=-1$, it follows that $i^3=-i$, $i^4=1$, etc., and so

$$\begin{aligned}e^{ix}&=1+\frac{i}{1!}x-\frac{i}{2!}x^2-\frac{i}{3!}x^3+\frac{i}{4!}x^4+\frac{i}{5!}x^5-\frac{i}{6!}x^6-\frac{i}{7!}x^7+\ldots\\&=\left\{1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\frac{1}{6!}x^6+\ldots\right\}+i\left\{\frac{1}{1!}x-\frac{1}{3!}x^3+\frac{1}{5!}x^5-\frac{1}{7!}x^7+\ldots\right\}\\&=\cos x+i\sin x.\end{aligned}$$

As we've seen, this result,

$$e^{ix}=\cos x+i\sin x,$$

Euler's identity, is one of the most remarkable equations in the whole of mathematics.

$$e^{i\pi}+1=0$$

Euler gave more than one proof of his identity. Box 11 presents a different approach in which he made use of 'infinitesimals'. It appears in his *Introductio in Analysin Infinitorum* of 1748 and opens with De Moivre's results, which Euler seems to have discovered independently.

Box 11: Another proof of Euler's identity

By adding and subtracting the identities

$$\cos nx + i \sin nx = (\cos x + i \sin x)^n \quad \text{and} \quad \cos nx - i \sin nx = (\cos x - i \sin x)^n,$$

Euler deduced that

$$\cos nx = \frac{1}{2}\{(\cos x + i \sin x)^n + (\cos x - i \sin x)^n\}$$

and

$$\sin nx = \frac{1}{2i}\{(\cos x + i \sin x)^n - (\cos x - i \sin x)^n\}.$$

He then took x to be infinitely small and n to be infinitely large, in such a way that nx has the finite value v. But $x = v/n$ is small, and so the power series for $\sin x$ and $\cos x$ tell us that, to a first approximation, $\sin x = x = v/n$ and $\cos x = 1$ (ignoring terms in $x^2, x^3, x^4, \ldots$).

This gives

$$\cos v = \frac{1}{2}\{(1+iv/n)^n + (1-iv/n)^n\} \text{ and}$$

$$\sin v = \frac{1}{2i}\{(1+iv/n)^n - (1-iv/n)^n\}.$$

Euler now let n increase indefinitely. Then, for any z, $(1+z/n)^n$ can be replaced by its limiting value e^z (see Chapter 4). So $(1+iv/n)^n$ is replaced by its limiting value e^{iv}, and $(1-iv/n)^n$ is replaced by its limiting value e^{-iv}. This gives

$$\cos v = \frac{1}{2}\{e^{iv} + e^{-iv}\} \quad \text{and} \quad \sin v = \frac{1}{2i}\{e^{iv} - e^{-iv}\},$$

and rearranging these gives

$$e^{iv} = \cos v + i \sin v \quad \text{and} \quad e^{-iv} = \cos v - i \sin v.$$

This passage from Euler's *Introductio* is shown in Figure 84 which opens this chapter. As Euler himself commented:

From these equations we can understand how complex exponentials can be expressed by real sines and cosines.

$$e^{i\pi}+1=0$$

Some consequences

Euler's identity has many simple, yet profound, consequences.

Euler's equation

The most important consequence of Euler's identity follows when we substitute $x=\pi$ (the radian form of 180°) to obtain Euler's equation:

$$e^{i\pi}=\cos\pi+i\sin\pi=-1+0i=-1,\ \text{so}\ e^{i\pi}+1=0.$$

Although Euler must surely have made this deduction, it doesn't appear explicitly in any of his published works.

In 1959 the English schoolteacher L. W. H. Hull illustrated Euler's equation pictorially. Putting $x=i\pi$ into the power series for e^x, he obtained

$$e^{i\pi}=1+i\pi-\frac{1}{2}\pi^2-\frac{i}{6}\pi^3+\frac{1}{24}\pi^4+\frac{i}{120}\pi^5-\frac{1}{720}\pi^6-\ldots.$$

He then started at the point 1 on the complex plane, added $i\pi$, subtracted $\frac{1}{2}\pi^2$ and $\frac{i}{6}\pi^3$, added $\frac{1}{24}\pi^4$ and $\frac{i}{120}\pi^5$, subtracted $\frac{1}{720}\pi^6$, and so on. This produced a spiral path that converges to the sum of the series, which is $e^{i\pi}=-1$ (see Figure 88).

De Moivre's theorem

We saw in Box 11 how Euler used De Moivre's results to derive his identity. Conversely, Euler's identity gives us a very simple proof of De Moivre's theorem:

$$\text{for any number } n,\ (\cos\theta+i\sin\theta)^n=(e^{i\theta})^n=e^{i(n\theta)}=\cos n\theta+i\sin n\theta.$$

So, in some sense, De Moivre's theorem and Euler's identity are equivalent results.

Multiplying complex numbers

In Chapter 5 we saw that if $[r,\theta]$ and $[s,\varphi]$ are complex numbers written in polar form, then

$$[r,\theta]\times[s,\varphi]=[rs,\theta+\varphi].$$

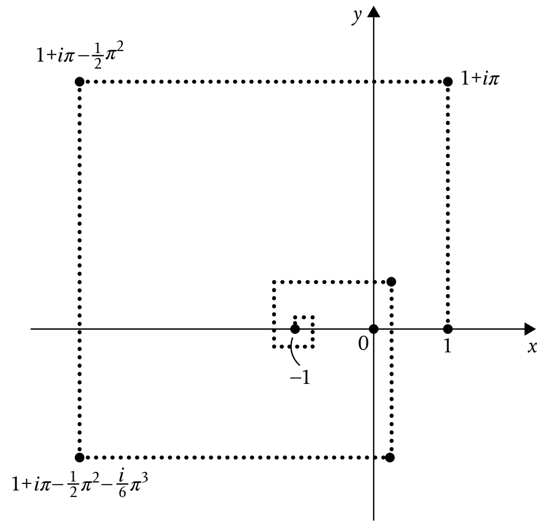


Fig. 88. Hull's spiral diagram

We proved this by using the addition formulas for cosine and sine, but it also follows easily from Euler's identity, because

$$\begin{aligned}[r,\theta]\times[s,\varphi] &= r\,(\cos\theta+i\sin\theta)\times s\,(\cos\varphi+i\sin\varphi)\\ &= re^{i\theta}\times se^{i\varphi}=rse^{i\theta+i\varphi}=rse^{i(\theta+\varphi)}=[rs,\theta+\varphi].\end{aligned}$$

To explore this connection further, we recall from Chapter 4 that the exponential function satisfies the basic identity

$$e^{a+b}=e^a\times e^b.$$

If we now replace a by $i\theta$ and b by $i\varphi$, we have $e^{i\theta+i\varphi}=e^{i\theta}\times e^{i\varphi}$, which we can rewrite as

$$e^{i(\theta+\varphi)}=e^{i\theta}\times e^{i\varphi}.$$

Applying Euler's identity to each term of this equation and rearranging the result, we have:

$$\begin{aligned}\cos(\theta+\varphi)+i\sin(\theta+\varphi) &= (\cos\theta+i\sin\theta)\times(\cos\varphi+i\sin\varphi)\\ &= (\cos\theta\cos\varphi-\sin\theta\sin\varphi)+i\,(\sin\theta\cos\varphi+\cos\theta\sin\varphi).\end{aligned}$$

$$e^{i\pi}+1=0$$

Equating real and imaginary parts now gives us

$$\cos(\theta+\varphi)=\cos\theta\cos\varphi-\sin\theta\sin\varphi$$

and

$$\sin(\theta+\varphi)=\sin\theta\cos\varphi+\cos\theta\sin\varphi,$$

which are the addition formulas for cosine and sine. Once again, mathematical results that originally looked very different are seen to be essentially the same.

Relating the trigonometric and hyperbolic functions

Euler's identity expresses the exponential function in terms of cos x and sin x. Let's now reverse the process.

It follows from the equation $e^{ix}=\cos x+i\sin x$, on replacing x by $-x$, that

$$e^{-ix}=\cos(-x)+i\sin(-x)=\cos x-i\sin x.$$

Adding and subtracting these two equations gives

$$\cos x=\frac{1}{2}(e^{ix}+e^{-ix})\quad\text{and}\quad\sin x=\frac{1}{2i}(e^{ix}-e^{-ix}).$$

These remarkable results, which appeared in Box 11, show how by allowing complex numbers we can express the standard trigonometric functions in terms of the exponential function.

If we now replace x by ix in these expressions for cos x and sin x, we get

$$\cos ix=\frac{1}{2}(e^{-x}+e^{x})=\cosh x\quad\text{and}$$

$$\sin ix=\frac{1}{2i}(e^{-x}-e^{x})=\frac{1}{i}(-\sinh x)=i\sinh x,$$

where cosh x and sinh x are the hyperbolic functions introduced in Chapter 4. So

$$\cosh x=\cos ix\quad\text{and}\quad\sinh x=-i\sin ix,\ \text{ for all } x.$$

By using complex numbers we've obtained simple relationships between the trigonometric functions and the hyperbolic ones. These connections explain why they share such similar properties, as we saw in

Chapter 4. For example, we can deduce from the identity $\cos^2 z+\sin^2 z=1$ that $\cosh^2 x-\sinh^2 x=1$, because

$$\cosh^2 x-\sinh^2 x=\cos^2 ix-(-i\sin ix)^2=\cos^2 ix+\sin^2 ix=1.$$

Roots of 1

In Chapter 5 we found the following complex *n*th roots of 1, for $n=2, 3, 4$, and 6:

$n=2$: 1 and -1

$n=3$: $1, \frac{1}{2}(-1+\sqrt{3}i)$, and $\frac{1}{2}(-1-\sqrt{3}i)$

$n=4$: $1, -1, i$, and $-i$

$n=6$: $1, -1, \frac{1}{2}(1+\sqrt{3}i), \frac{1}{2}(1-\sqrt{3}i), \frac{1}{2}(-1+\sqrt{3}i)$, and $\frac{1}{2}(-1-\sqrt{3}i)$.

If, for each value of *n*, we add up these roots we get the sum 0 in every case. But does this happen for *all* values of *n*?

To answer this, recall from Chapter 5 that the *n*th roots of 1 are the complex numbers

$$\cos 2k\pi/n+i\sin 2k\pi/n, \text{ for } k=0, 1, 2, \ldots, n-1.$$

By Euler's identity, these are

$$e^{2k\pi i/n}, \text{ for } k=0, 1, 2, \ldots, n-1.$$

We want to show that the sum of all these values is 0. But if $z=e^{2\pi i/n}$, then this sum is

$$1+z+z^2+z^3+\ldots+z^{n-1}.$$

Summing this geometric progression gives $(z^n-1)/(z-1)$, which is 0, because $z^n=1$. So

For any *n*, the complex *n*th roots of 1 have sum 0.

$$e^{i\pi}+1=0$$

For example, when $n=2$, we have

$$e^0+e^{\pi i}=1+(-1)=0,$$

so this result can be thought of as a generalization of Euler's equation.

The golden ratio

What has the golden ratio to do with any of these results?

Recall from Chapter 2 that if $\varphi=\frac{1}{2}(1+\sqrt{5})$ is the golden ratio, then

$$\varphi^{-1}=\frac{1}{2}\left(\sqrt{5}-1\right) \quad \text{and} \quad \varphi^{-1}-\varphi=-1.$$

We also saw above that, for all *x*,

$$\sin x=\frac{1}{2i}\left(e^{ix}-e^{-ix}\right).$$

If we now choose $x=i\ln\varphi$, so that $ix=-\ln\varphi$ and $-ix=\ln\varphi$, then

$$\sin\left(i\ln\varphi\right)=\frac{1}{2i}\left(e^{-\ln\varphi}-e^{\ln\varphi}\right)=\frac{1}{2i}\left(\varphi^{-1}-\varphi\right)=\frac{1}{2i}(-1)=\frac{i}{2}.$$

On multiplying by 2π and taking exponentials, we have

$$e^{2\pi\sin(i\ln\varphi)}=e^{2\pi i/2}=e^{\pi i}=-1, \text{ or } e^{2\pi\sin(i\ln\varphi)}+1=0.$$

This last equation connects 1, 0, π, e, i, and φ, which are *six* of the most important numbers in mathematics (or *seven*, if we include the 2)!

e and π are transcendental

In Chapter 2 we saw that the numbers $\sqrt{2}$, $\sqrt[3]{7}$, $\sqrt{2}+\sqrt{3}$, and i are *algebraic numbers* – they're solutions of polynomial equations with integer coefficients. We also stated that the numbers e and π are *transcendental* – neither is the solution of any such equations. We can deduce this from a result of Ferdinand Lindemann who discovered it around 1882 while investigating the transcendence of π. A simple form of it is:

If r is a non-zero algebraic number (real or complex), then e^r is transcendental.

On taking $r=1$ we deduce immediately that e is transcendental.

$$e^{i\pi}+1=0$$

Now suppose, for a contradiction, that π were algebraic. Then $i\pi$ would also be algebraic, since i is algebraic $(i^2=-1)$ and the product of algebraic numbers is algebraic. So, on taking $r=i\pi$ in Lindemann's result, we'd deduce that $e^{i\pi}$ is transcendental. But, by Euler's equation, $e^{i\pi}$ has the value $-i$, which is not transcendental. This contradiction shows that π is not algebraic, and is therefore transcendental.

By proving that π is transcendental, Lindemann had finally answered a long-standing problem of the ancient Greeks, who had asked whether one can 'square the circle':

Using only a ruler and compasses, can one construct a square with the same area as a given circle?

Lindemann's negative answer to this question proved, once and for all, that it is impossible to square the circle.

What are ln i, i^i, and $\sqrt[i]{i}$?

Three of the most unexpected results in this subject concern the logarithm of i, the ith power i^i, and the ith root $\sqrt[i]{i}$. None of these has just one value – in fact, they all have infinitely many – while the last two (which look highly complex) take only real values!

Earlier we saw some of the difficulties involved with defining the logarithm of a negative number. Euler brilliantly clarified the whole issue by defining the logarithm of a complex number.

What is ln i?

We've seen that any non-zero complex number can be written in polar form as

$$z=r\left(\cos\theta+i\sin\theta\right)=re^{i\theta}.$$

It follows, on taking logarithms, that

$$\ln z=\ln re^{i\theta}=\ln r+\ln e^{i\theta}=\ln r+i\theta=\ln|z|+i\arg z.$$

This identity gives the logarithm of any non-zero complex number z. But, as we saw in Chapter 5, arg z has infinitely many values, all differing by multiples of 2π. It follows, as Euler discovered, that the logarithm of a non-zero complex number has infinitely many values, all differing by multiples of $2\pi i$. As he explained in 1748:

In the same way that to one sine there correspond an infinite number of different angles I have found that it is the same with logarithms. Each number has an infinity of different logarithms, all of them imaginary unless the number is real and positive. Then there is only one logarithm which is real, and we regard it as its unique logarithm.

Two special cases are particularly important.
When $z=-1$, $r=|-1|=1$, and $\theta=\arg(-1)=\pi$ (or $\pi+2k\pi$, for any integer k). So one value of ln (–1), corresponding to $\theta=\pi$, is

$$\ln(-1)=\ln 1+\pi i=\pi i,$$

as we saw earlier, and the full list of possible values is

$$\ldots, -5\pi i, -3\pi i, -\pi i, \pi i, 3\pi i, 5\pi i, 7\pi i, \ldots$$

When $z=i$, $r=|i|=1$, and $\theta=\arg i=\pi/2$ (or $\pi/2+2k\pi$, for any integer k). So one value of ln i, corresponding to $\theta=\pi/2$, is

$$\ln i=\ln 1+\pi i/2=\pi i/2,$$

and the full list of possible values is

$$\ldots, -11\pi i/2, -7\pi i/2, -3\pi i/2, \pi i/2, 5\pi i/2, 9\pi i/2, 13\pi i/2, \ldots.$$

What are i^i and $\sqrt[i]{i}$?

To find the ith power i^i we use this last result, that

$$\ln i=\pi i/2\left(\text{or } \pi i/2+2k\pi i\right).$$

Then one value of i^i is

$$i^i = e^{i\ln i} = e^{i(i\pi/2)} = e^{-\pi/2} = 1/\sqrt{e^{\pi}},$$

a real number that is approximately 0.2078795763. The other values of i^i all have the form $e^{-\pi/2+2k\pi}$, so i^i has infinitely many values, all of them real:

$$e^{i\pi}+1=0$$

$$\ldots, e^{-9\pi/2}, e^{-5\pi/2}, e^{-\pi/2}, e^{3\pi/2}, e^{7\pi/2}, e^{11\pi/2}, \ldots.$$

Similarly, one value of the *i*th root $\sqrt[i]{i}$ is

$$\sqrt[i]{i}=i^{1/i}=e^{(1/i)\ln i}=e^{(1/i)(i\pi/2)}=e^{\pi/2}=\sqrt{e^{\pi}},$$

a real number that is approximately 4.8104773821. The other values of $\sqrt[i]{i}$ all have the form $e^{\pi/2+2k\pi}$, so $\sqrt[i]{i}$ also has infinitely many values, all of them real:

$$\ldots, e^{-7\pi/2}, e^{-3\pi/2}, e^{\pi/2}, e^{5\pi/2}, e^{9\pi/2}, e^{13\pi/2}, \ldots.$$

These results are very surprising. Even Euler himself, in a letter to his colleague Christian Goldbach, wrote that this last result 'seems to me to be very odd'. And in one of his lectures Benjamin Peirce, the distinguished professor of mathematics at Harvard University from 1831 to 1880, was so taken with proving that $\sqrt[i]{i}=e^{\pi/2}$ that, according to one of his students,

> after contemplating the result for a few minutes he turned to his class and said very slowly and impressively, 'Gentlemen, that is surely true, it is absolutely paradoxical, we can't understand it, and we don't know what it means, but we have proved it, and therefore we know it must be the truth'.

Finally, we note that if we take the equation $\sqrt[i]{i}=e^{\pi/2}$ and raise each side to the 2*i*th power, then

$$(\sqrt[i]{i})^{2i}=i^2=-1 \quad \text{and} \quad (e^{\pi/2})^{2i}=e^{i\pi},$$

so that $e^{i\pi}=-1$, and we have obtained Euler's equation yet again!

Who discovered Euler's equation?

What should we call the equation $e^{i\pi}+1=0$?

We've seen how it can easily be deduced from results of Johann Bernoulli and Roger Cotes, but that neither of them seems to have done so. Even Euler seems not to have written it down explicitly – and certainly it doesn't appear in any of his publications – though he must surely have realized that it follows immediately from his identity, $e^{ix}=\cos x+i\sin x$.

Moreover, it seems to be unknown who first stated the result explicitly, although Jacques Français, who appeared briefly in Chapter 5 in connection with the Argand diagram, certainly wrote about it in 1813–14 in a French mathematical journal.

But almost everybody nowadays attributes the result to Leonhard Euler. For this reason we're surely justified in naming it 'Euler's equation', to honour the achievements of this truly great mathematical *pioneer* – a word that describes him so well, and which appropriately includes among its letters our five constants *pi, i, o, one,* and *e*.

FURTHER READING

The following book on Euler's equation contains much information, but at a more advanced level:

Paul J. Nahin, *Dr. Euler's Fabulous Formula: Cures Many Mathematical Ills,* Princeton University Press (2006).

Euler

There are many books on Leonhard Euler. Among the biographies are:

Ronald Calinger, *Leonhard Euler: Mathematical Genius in the Enlightenment,* Princeton University Press (2016).
Emil A. Fellmann, *Leonhard Euler,* Birkhäuser (2007).

A more mathematical treatment is:

William Dunham, *Euler: The Master of Us All,* Mathematical Association of America (1999).

Euler's *Introductio in Analysin Infinitorum* has been published in English:

L. Euler, *Introduction to Analysis of the Infinite* (trans. John Blanton), Springer-Verlag (1988).

In 2007, as part of the MAA Tercentenary Euler Celebration, the Mathematical Association of America published a set of five books on various aspects of Euler and his works.

All of Euler's writings, many with commentary and translations, can be found on the Euler archive, *www.eulerarchive@maa.org*.

The most beautiful equation

The following articles describe the various polls taken to find the world's greatest equation:

Robert P. Crease, 'The greatest equations ever', *Physics World* 17 (5) (May 2004), 19, and 17 (10) (October 2004), 14–15.

Keith Devlin, 'The most beautiful equation', *Wabash Magazine* (Winter/Spring 2002).
David Wells, 'Which is the most beautiful?' and 'Are these the most beautiful?', *Mathematical Intelligencer* 10 (4) (Fall 1988), 30–1, and 12 (3) (Summer 1990), 37–41.

Introduction

Richard Feynman's description of Euler's equation at age 14 appears as the frontispiece to Nahin's book (listed above), and in:

Matthew Sands, Richard Feynman, and Robert B. Leighton (eds), *The Feynman Lectures on Physics*, Vol. 1, Pearson (2006), 22-10.

The 'equation beauty contest' is described in:

Semir Zeki, John Paul Romaya, Dionigi M. T. Benincasa, and Michael F. Atiyah, 'The experience of mathematical beauty and its neural correlates', *Frontiers in Human Neuroscience* (13 February 2014); reported by Peter Rowlett as 'The neuroscience of mathematical beauty, or, Equation beauty contest!', News, *Phil. Trans. Aperiodic, The Aperiodical* (13 February 2014).

Number systems

Discussions of many number systems from around the world can be found in:

Graham Flegg (ed.), *Numbers through the Ages*, Macmillan (1989).
Georges Ifrah, *The Universal History of Numbers*, Harvill Press, London (1998).
John D. Barrow, *The Book of Nothing*, Vintage Books (2000).
Robert Kaplan, *The Nothing that is: A Natural History of Zero*, Oxford University Press, Oxford (2000).

Information about many numbers and their origins is given in:

David Wells, *The Penguin Dictionary of Curious and Interesting Numbers*, Penguin (1986).

Entertaining facts about the use of number words in our everyday language appear in:

Tony Augarde, Every word counts, *Oxfordshire Limited Edition* (December 2013), 29.

π, e, and i

A number of useful books present the history and properties of π:

Petr Beckmann, *A History of Pi*, Dorset Press, New York (1989).

J. L. Berggren, Jonathan Borwein, and Peter Borwein, *Pi: A Source Book*, 3rd edn, Springer-Verlag (2004).
David Blatner, *The Joy of π*, Walker Books (1999).
Alfred S. Posamentier and Ingmar Lehmann, *Pi: A Biography of the World's Most Mysterious Number*, Prometheus Books (2004).

At a lighter level, but with much interesting information, is:

Simon Singh, *The Simpsons and their Mathematical Secrets*, Bloomsbury (2013).

The Bodleian Library discovery is described in

Benjamin Wardhaugh, 'Filling a gap in the history of π: An exciting discovery', *Mathematical Intelligencer* 38 (1) (March 2016), 6–7.

The standard book on *e*, which contains a wealth of further information, is:

Eli Maor, *e: The Story of a Number*, Princeton University Press (1994).

The origins of logarithms are discussed in:

Ronald Calinger, *A Contextual History of Mathematics*, Prentice-Hall (1999), 485–92.

The following book is the standard work on *i*:

Paul J. Nahin, *An Imaginary Tale: The Story of* $\sqrt{-1}$, Princeton University Press (1998).

Euler's equation

Further information about Euler's equation can be found in:

R. E. Bradley, 'Euler, d'Alembert and the logarithm function', *Leonhard Euler: Life, Work and Legacy* (ed. Robert E. Bradley and C. Edward Sandifer), Studies in the History and Philosophy of Mathematics, Vol. 5, Elsevier (2007), 255–77.
Jacques Français, 'Nouveaux principes de géométrie de position, et interprétation géométrique des symboles imaginaires', *Annales de Mathématiques Pures et Appliquées* (Gergonne's Journal) 4 (1813–14), 64.
L. W. H. Hull, 'Convergence on the Argand diagram', *Mathematical Gazette* 43, no 345 (October 1959), 205–7.
C. Edward Sandifer, '*e*, π and *i*: Why is "Euler" in the Euler Identity?', *How Euler did Even More*, Mathematical Association of America (2016), Chapter 12.
J. Stillwell, *Mathematics and its History*, 2nd edn, Springer (2002), 95, 262–3, 294–5.

IMAGE CREDITS

Frontispiece: Leonhard Euler (1707–83): An engraving by Bartholomaeus Hübner, 1786; in Euler's *Opera Omnia*, Vol. I, Book 1.

Figure 1: Title page of Euler's *Introductio in Analysin Infinitorum*, Vol. 1, Lausanne, 1748; Courtesy of The Bodleian Libraries, The University of Oxford (Vet. D.4.d.48, title page).

Figure 2: 'Children's fund' postage stamp from Switzerland, 1957.

Figure 5: 'Finger counting', from Luca Pacioli, *Summa de Arithmetica, Geometrica, Proportioni et Proportionalita*, 1494, f. 36v Courtesy of The Bodleian Libraries, The University of Oxford (Saville P.8, fol. e4v).

Figure 6: From the Rhind papyrus, British Museum. © The Trustees of the British Museum.

Figure 9: Plimpton 322 cuneiform tablet, Rare Book and Manuscript Library, Columbia University.

Figure 11: A Mesopotamian 9-times table. Courtesy of the Jena Collection, Germany.

Figure 15: Part of a Mayan codex. SLUB Dresden, digital.slub-dresden.de/id30409217Z/100 (CC-BY-SA 4.0).

Figure 19: 'Arithmetica', from Gregor Reisch, *Margarita Philosophica*, 1503. Houghton Library, Harvard University (Typ 520.03.736).

Figure 21: A Mesopotamian table of reciprocals, MLC 1670, after A. T. Clay, Babylonian Record in the Library of J. Pierpoint Morgan, Vol. 4, Yale University Press, New Haven (1923), plate 37. Courtesy of Yale Babylonian Collection.

Figure 25: Bakhshali manuscript; Courtesy of The Bodleian Libraries, The University of Oxford (MS. +Sansk_d.14_16v closeup).

Figure 31: 16th-century edition Ming dynasty edition of Liu Hui's commentary on the *Jiuzhang suanshu* (Nine Chapters on the Mathematical Art). The Warden and Fellows of All Souls College, Oxford.

Figure 34: Sato Moshun, *Tengen Shinan*, 1686 Courtesy of The Bodleian Libraries, The University of Oxford (Hist. QA27.J3 SMI 1914, p. 131).

Figure 36: William Jones, *Synopsis Palmariorum Matheseos or A New Introduction to the Mathematics* (1706), 263. Courtesy of The Bodleian Libraries, The University of Oxford (8° B. 573 Linc., p. 263).

Figure 41: Title page of Ludolph van Ceulen, *De Circolo & Adscriptis Liber*, 1619; courtesy of the Edward Worth Library, Dublin.

Figure 44: John Wallis, *Arithmetica Infinitorum*, Oxford (1656), 179.
Figure 46: William Jones, *Synopsis Palmariorum Matheseos or A New Introduction to the Mathematics*, London (1706), 243 Courtesy of The Bodleian Libraries, The University of Oxford (HB4/5.b.4.21(1), p. 179).
Figure 47: Pi room in the Palais de la Découverte, Avenue Franklin D. Roosevelt, Paris VIII, France. © Stéphanie Garnier.
Figure 52: Title page of John Napier, *Mirifici Logarithmorum Canonis Descriptio*, Edinburgh, 1614 Courtesy of The Bodleian Libraries, The University of Oxford (4° N 8 Art. Seld., title page).
Figure 54: Henry Briggs, *Logarithmorum Chilias Prima*, London, 1617.
Figure 58: Newton manuscript from 1665; by permission of the Syndics of Cambridge University Library.
Figure 66: Charcoal Sketch by Sean Keating; courtesy of Dunsink Observatory.
Figure 69: Title page of Girolamo Cardano, *Ars Magna*, 1545 Courtesy of The Bodleian Libraries, The University of Oxford (HB4/6.d.5.4(1), title page).
Figure 81: The plaque at Brougham Bridge, Royal Canal, Dublin. Brian Dolan.
Figure 82: 'Europa' postage stamp from Ireland, 1983. Dale Hathaway.
Figure 84: From Leonhard Euler, *Introductio in Analysin Infinitorum*, Vol. 1, Lausanne (1648), 147–8; the Euler Archive, http://eulerarchive.maa.org/.

PUBLISHER'S ACKNOWLEDGEMENTS

We are grateful for permission to include the following copyright material in this book.

Extract from 'New Math' © Tom Lehrer. Used by permission.

The publisher and author have made every effort to trace and contact all copyright holders before publication. If notified, the publisher will be pleased to rectify any errors or omissions at the earliest opportunity.

INDEX