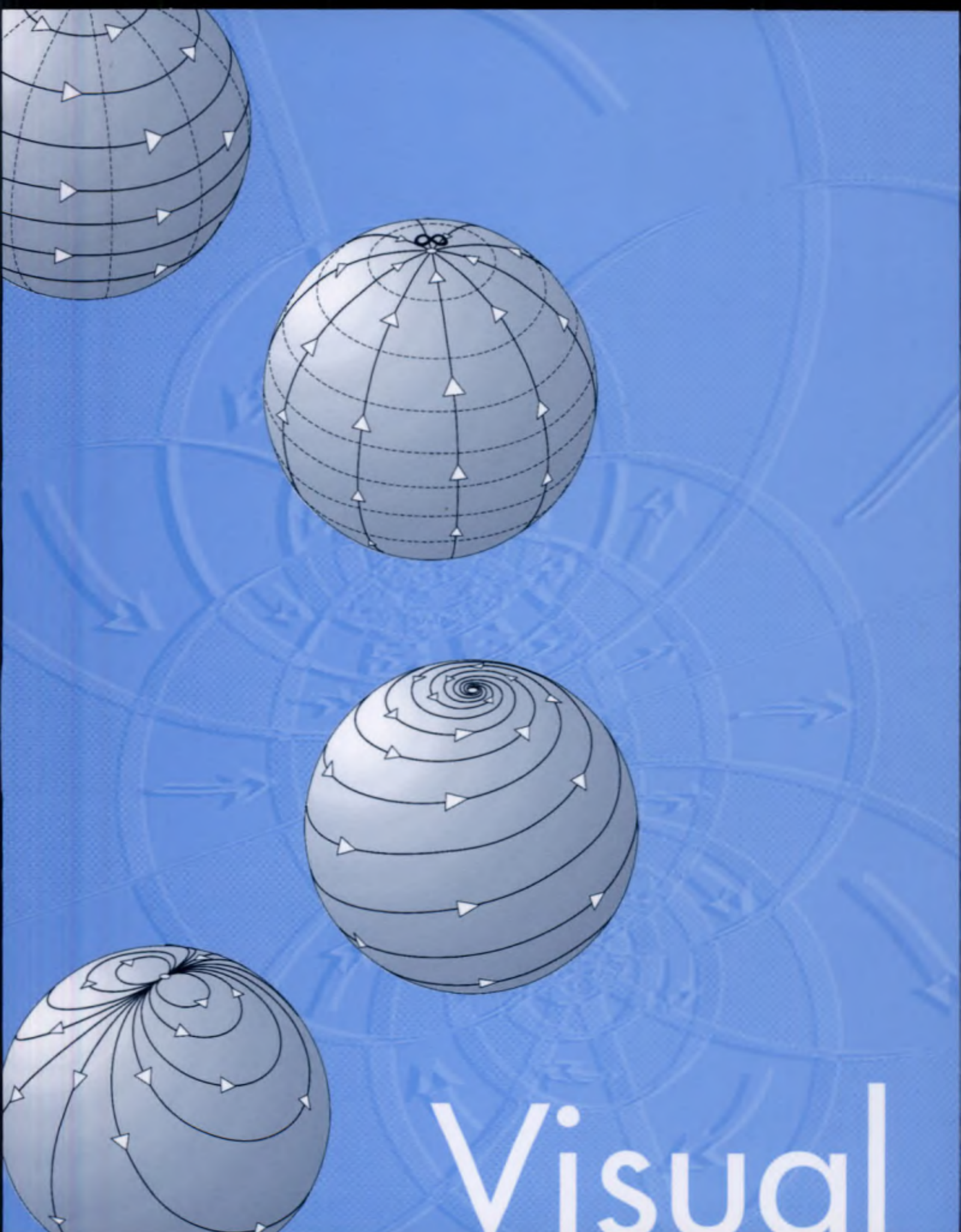


Visual Complex Analysis

Tristan Needham

Visual Complex Analysis

Tristan Needham

Department of Mathematics
University of San Francisco

CLARENDON PRESS • OXFORD

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

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 in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Published in the United States
by Oxford University Press Inc., New York

First published 1997
Reprinted 1997, 1998
First published in paperback 1998 (with corrections)
Reprinted 2000 (twice with corrections), 2001, 2002 (twice), 2003, 2004

A catalogue record for this book is available from the British Library
(Data available)

ISBN 0 19 853446 9 (Pbk)

Printed in Great Britain
on acid-free paper by
Biddles Ltd, King's Lynn, Norfolk

Contents

1

Geometry and Complex Arithmetic

I Introduction

1 Historical Sketch

Four and a half centuries have elapsed since complex numbers were first discovered. Here, as the reader is probably already aware, the term *complex number* refers to an entity of the form $a + ib$, where a and b are ordinary real numbers and, unlike any ordinary number, i has the property that $i^2 = -1$. This discovery would ultimately have a profound impact on the whole of mathematics, unifying much that had previously seemed disparate, and explaining much that had previously seemed inexplicable. Despite this happy ending—in reality the story continues to unfold to this day—progress following the initial discovery of complex numbers was painfully slow. Indeed, relative to the advances made in the nineteenth century, *little was achieved during the first 250 years of the life of the complex numbers.*

How is it possible that complex numbers lay dormant through ages that saw the coming and the passing of such great minds as Descartes, Fermat, Leibniz, and even the visionary genius of Newton? The answer appears to lie in the fact that, far from being embraced, complex numbers were initially greeted with suspicion, confusion, and even hostility.

Girolamo Cardano's *Ars Magna*, which appeared in 1545, is conventionally taken to be the birth certificate of the complex numbers. Yet in that work Cardano introduced such numbers only to immediately dismiss them as "subtle as they are useless". As we shall discuss, the first substantial *calculations* with complex numbers were carried out by Rafael Bombelli, appearing in his *L'Algebra* of 1572. Yet here too we find the innovator seemingly disowning his discoveries (at least initially), saying that "the whole matter seems to rest on sophistry rather than truth". As late as 1702, Leibniz described i, the square root of -1, as "that amphibian between existence and nonexistence". Such sentiments were echoed in the terminology of the period. To the extent that they were discussed at all, complex numbers were called "impossible" or "imaginary", the latter term having (unfortunately) lingered to the present day[1]. Even in 1770 the situation was still sufficiently confused that it was possible for so great a mathematician as Euler to mistakenly argue that $\sqrt{-2}\,\sqrt{-3} = \sqrt{6}$.

[1]However, an "imaginary number" now refers to a real multiple of i, rather than to a general complex number. Incidentally, the term "real number" was introduced precisely to distinguish such a number from an "imaginary number".

The root cause of all this trouble seems to have been a psychological or philosophical block. How could one investigate these matters with enthusiasm or confidence when nobody felt they knew the answer to the question, "What *is* a complex number?"

A satisfactory answer to this question was only found at the end of the eighteenth century[2]. Independently, and in rapid succession, Wessel, Argand, and Gauss all recognized that complex numbers could be given a simple, concrete, *geometric interpretation* as points (or vectors) in the plane: The mystical quantity $a + ib$ should be viewed simply as the point in the xy-plane having Cartesian coordinates (a, b), or equivalently as the vector connecting the origin to that point. See [1]. When thought of in this way, the plane is denoted $\mathbb{C}$ and is called the *complex plane*[3].

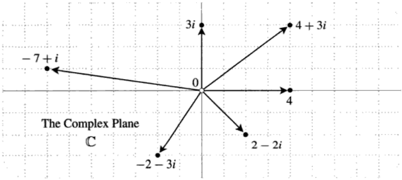


Figure [1]

The operations of adding or multiplying two complex numbers could now be given equally definite meanings as geometric operations on the two corresponding points (or vectors) in the plane. The rule for addition is illustrated in [2a]:

The sum $A+B$ of two complex numbers is given by the parallelogram rule of ordinary vector addition. (1)

Note that this is consistent with [1], in the sense that $4+3i$ (for example) is indeed the sum of 4 and $3i$.

Figure [2b] illustrates the much less obvious rule for multiplication:

The length of AB is the product of the lengths of A and B, and the angle of AB is the sum of the angles of A and B. (2)

This rule is not forced on us in any obvious way by [1], but note that it is at least consistent with it, in the sense that $3i$ (for example) is indeed the product of 3 and

[2]Wallis almost hit on the answer in 1673; see Stillwell [1989, p. 191] for an account of this interesting near miss.

[3]Also known as the "Gauss plane" or the "Argand plane".

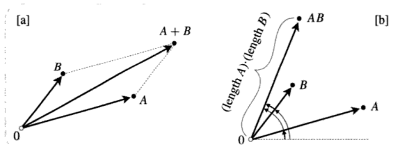


Figure [2]

i. Check this for yourself. As a more exciting example, consider the product of i with itself. Since i has unit length and angle $(\pi/2)$, i^2 has unit length and angle π. Thus $i^2 = -1$.

The publication of the geometric interpretation by Wessel and by Argand went all but unnoticed, but the reputation of Gauss (as great then as it is now) ensured wide dissemination and acceptance of complex numbers as points in the plane. Perhaps less important than the details of this new interpretation (at least initially) was the mere fact that there now existed *some* way of making sense of these numbers—that they were now *legitimate* objects of investigation. In any event, the floodgates of invention were about to open.

It had taken more than two and a half centuries to come to terms with complex numbers, but the development of a beautiful new theory of how to do *calculus* with such numbers (what we now call *complex analysis*) was astonishingly rapid. Most of the fundamental results were obtained (by Cauchy, Riemann, and others) between 1814 and 1851—a span of less than forty years!

Other views of the history of the subject are certainly possible. For example, Stewart and Tall [1983, p. 7] suggest that the geometric interpretation[4] was somewhat incidental to the explosive development of complex analysis. However, it should be noted that Riemann's ideas, in particular, would simply not have been possible without prior knowledge of the geometry of the complex plane.

2 Bombelli's "Wild Thought"

The power and beauty of complex analysis ultimately springs from the multiplication rule (2) in conjunction with the addition rule (1). These rules were first discovered by Bombelli in *symbolic* form; more than two centuries passed before the complex plane revealed figure [2]. Since we merely plucked the rules out of thin air, let us return to the sixteenth century in order to understand their algebraic origins.

Many texts seek to introduce complex numbers with a convenient historical fiction based on solving quadratic equations,

[4]We must protest one piece of their evidence: Wallis did *not* possess the geometric interpretation in 1673; see footnote 2.

$$x^2 = mx + c. \tag{3}$$

Two thousand years BC, it was already known that such equations could be solved using a method that is equivalent to the modern formula,

$$x = \tfrac{1}{2}\left[m \pm \sqrt{m^2 + 4c}\right].$$

But what if $m^2 + 4c$ is negative? This was the very problem that led Cardano to consider square roots of negative numbers. Thus far the textbook is being historically accurate, but next we read that the need for (3) to always have a solution *forces* us to take complex numbers seriously. This argument carries almost as little weight now as it did in the sixteenth century. Indeed, we have already pointed out that Cardano did not hesitate to discard such "solutions" as useless.

It was not that Cardano lacked the imagination to pursue the matter further, rather he had a fairly compelling reason *not to*. For the ancient Greeks mathematics was synonymous with geometry, and this conception still held sway in the sixteenth century. Thus an algebraic relation such as (3) was not so much thought of as a problem in its own right, but rather as a mere vehicle for solving a genuine problem of geometry. For example, (3) may be considered to represent the problem of finding the intersection points of the parabola $y = x^2$ and the line $y = mx + c$. See [3a].

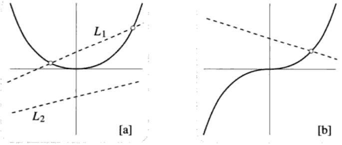


Figure [3]

In the case of L_1 the problem has a solution; algebraically, $(m^2 + 4c) > 0$ and the two intersection points are given by the formula above. In the case of L_2 the problem clearly does *not* have a solution; algebraically, $(m^2 + 4c) < 0$ and the absence of solutions is correctly manifested by the occurrence of "impossible" numbers in the formula.

It was not the quadratic that forced complex numbers to be taken seriously, it was the *cubic*,

$$x^3 = 3px + 2q.$$

[Ex. 1 shows that a general cubic can always be reduced to this form.] This equation represents the problem of finding the intersection points of the cubic curve $y = x^3$ and the line $y = 3px + 2q$. See [3b]. Building on the work of del Ferro and

Tartaglia, Cardano's *Ars Magna* showed that this equation could be solved by means of a remarkable formula [see Ex. 2]:

$$x = \sqrt[3]{q + \sqrt{q^2 - p^3}} + \sqrt[3]{q - \sqrt{q^2 - p^3}}. \tag{4}$$

Try it yourself on $x^3 = 6x + 6$.

Some thirty years after this formula appeared, Bombelli recognized that there was something strange and paradoxical about it. First note that if the line $y = 3px + 2q$ is such that $p^3 > q^2$ then the formula involves complex numbers. For example, Bombelli considered $x^3 = 15x + 4$, which yields

$$x = \sqrt[3]{2 + 11i} + \sqrt[3]{2 - 11i}.$$

In the previous case of [3a] this merely signalled that the geometric problem had no solution, but in [3b] it is clear that the line will *always* hit the curve! In fact inspection of Bombelli's example yields the solution $x = 4$.

As he struggled to resolve this paradox, Bombelli had what he called a "wild thought": perhaps the solution $x = 4$ could be recovered from the above expression if $\sqrt[3]{2 + 11i} = 2 + ni$ and $\sqrt[3]{2 - 11i} = 2 - ni$. Of course for this to work he would have to assume that the addition of two complex numbers $A = a + i\widetilde{a}$ and $B = b + i\,\widetilde{b}$ obeyed the plausible rule,

$$A + B = (a + i\,\widetilde{a}) + (b + i\,\widetilde{b}) = (a + b) + i\,(\widetilde{a} + \widetilde{b}). \tag{5}$$

Next, to see if there was indeed a value of n for which $\sqrt[3]{2 + 11i} = 2 + in$, he needed to calculate $(2 + in)^3$. To do so he assumed that he could multiply out brackets as in ordinary algebra, so that

$$(a + i\,\widetilde{a})\;(b + i\,\widetilde{b}) = ab + i\,(a\,\widetilde{b} + \widetilde{a}\,b) + i^2\,\widetilde{a}\,\widetilde{b}.$$

Using $i^2 = -1$, he concluded that the product of two complex numbers would be given by

$$AB = (a + i\,\widetilde{a})\;(b + i\,\widetilde{b}) = (ab - \widetilde{a}\,\widetilde{b}) + i\,(a\,\widetilde{b} + \widetilde{a}\,b). \tag{6}$$

This rule vindicated his "wild thought", for he was now able to show that $(2 \pm i)^3 = 2 \pm 11i$. Check this for yourself.

While complex numbers themselves remained mysterious, Bombelli's work on cubic equations thus established that perfectly real problems required complex arithmetic for their solution.

Just as with its birth, the subsequent development of the theory of complex numbers was inextricably bound up with progress in other areas of mathematics (and also physics). Sadly, we can only touch on these matters in this book; for a full and fascinating account of these interconnections, the reader is instead referred to Stillwell [1989]. Repeating what was said in the Preface, we cannot overstate the value of reading Stillwell's book alongside this one.

3 Some Terminology and Notation

Leaving history behind us, we now introduce the modern terminology and notation used to describe complex numbers. The information is summarized in the table below, and is illustrated in [4].

Name	Meaning	Notation
modulus of z	length r of z	$\lvert z\rvert$
argument of z	angle θ of z	$\arg(z)$
real part of z	x coordinate of z	$\mathrm{Re}(z)$
imaginary part of z	y coordinate of z	$\mathrm{Im}(z)$
imaginary number	real multiple of i	
real axis	set of real numbers	
imaginary axis	set of imaginary numbers	
complex conjugate of z	reflection of z in the real axis	$\overline{z}$

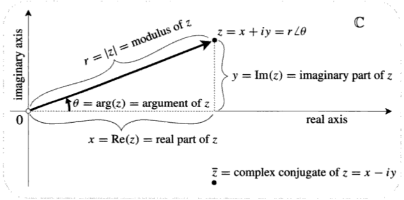


Figure [4]

It is valuable to grasp from the outset that (according to the geometric view) a complex number is a single, indivisible entity—a point in the plane. Only when we choose to describe such a point with numerical coordinates does a complex number appear to be compound or "complex". More precisely, $\mathbb{C}$ is said to be *two dimensional*, meaning that *two* real numbers (coordinates) are needed to label a point within it, but exactly *how* the labelling is done is entirely up to us.

One way to label the points is with Cartesian coordinates (the real part x and the imaginary part y), the complex number being written as $z = x + iy$. This is the natural labelling when we are dealing with the addition of two complex numbers, because (5) says that the real and imaginary parts of $A + B$ are obtained by adding the real and imaginary parts of A and B.

In the case of multiplication, the Cartesian labelling no longer appears natural, for it leads to the messy and unenlightening rule (6). The much simpler geometric rule (2) makes it clear that we should instead label a typical point z with its *polar* coordinates, $r = |z|$ and $\theta = \arg z$. In place of $z = x + iy$ we may now write $z = r\angle\theta$, where the symbol $\angle$ serves to remind us that θ is the *angle* of z. [Although this notation is still used by some, we shall only employ it briefly; later in this chapter we will discover a much better notation (the standard one) which will then be used throughout the remainder of the book.] The geometric multiplication rule (2) now takes the simple form,

$$(R\angle\phi)\,(r\angle\theta) = (Rr)\angle(\phi + \theta). \qquad (7)$$

In common with the Cartesian label $x + iy$, a given polar label $r\angle\theta$ specifies a unique point, but (unlike the Cartesian case) a given point does not have a unique polar label. Since any two angles that differ by a multiple of 2π correspond to the same direction, a given point has infinitely many different labels:

$$\ldots = r\angle(\theta - 4\pi) = r\angle(\theta - 2\pi) = r\angle\theta = r\angle(\theta + 2\pi) = r\angle(\theta + 4\pi) = \ldots$$

This simple fact about angles will become increasingly important as our subject unfolds.

The Cartesian and polar coordinates are the most common ways of labelling complex numbers, but they are not the only ways. In Chapter 3 we will meet another particularly useful method, called "stereographic" coordinates.

4 Practice

Before continuing, we strongly suggest that you make yourself comfortable with the concepts, terminology, and notation introduced thus far. To do so, try to convince yourself geometrically (*and/or* algebraically) of each of the following facts:

$$\mathrm{Re}(z) = \tfrac{1}{2}[z + \bar{z}] \qquad \mathrm{Im}(z) = \tfrac{1}{2i}[z - \bar{z}] \qquad |z| = \sqrt{x^2 + y^2}$$

$$\tan[\arg z] = \frac{\mathrm{Im}(z)}{\mathrm{Re}(z)} \qquad z\,\bar{z} = |z|^2 \qquad r\angle\theta = r\,(\cos\theta + i\sin\theta)$$

$$\text{Defining } \tfrac{1}{z} \text{ by } (1/z)\,z = 1, \text{ it follows that } \tfrac{1}{z} = \frac{1}{r\angle\theta} = \tfrac{1}{r}\angle(-\theta)\,.$$

$$\frac{R\angle\phi}{r\angle\theta} = \frac{R}{r}\angle(\phi - \theta) \qquad \frac{1}{(x + iy)} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}$$

$$(1 + i)^4 = -4 \qquad (1 + i)^{13} = -2^6(1 + i) \qquad (1 + i\sqrt{3})^6 = 2^6$$

$$\frac{(1 + i\sqrt{3})^3}{(1 - i)^2} = -4i \qquad \frac{(1 + i)^5}{(\sqrt{3} + i)^2} = -\sqrt{2}\angle - (\pi/12) \qquad \overline{r\angle\theta} = r\angle(-\theta)$$

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \qquad \overline{z_1\,z_2} = \overline{z_1}\,\overline{z_2} \qquad \overline{z_1/z_2} = \overline{z_1}/\overline{z_2}.$$

Lastly, establish the so-called *generalized triangle inequality*:

$$|z_1 + z_2 + \cdots + z_n| \leq |z_1| + |z_2| + \cdots + |z_n|. \tag{8}$$

When does equality hold?

5 Equivalence of Symbolic and Geometric Arithmetic

We have been using the symbolic rules (5) and (6) interchangeably with the geometric rules (1) and (2), and we now justify this by showing that they are indeed equivalent. The equivalence of the addition rules (1) and (5) will be familiar to those who have studied vectors; in any event, the verification is sufficiently straightforward that we may safely leave it to the reader. We therefore only address the equivalence of the multiplication rules (2) and (6).

First we will show how the symbolic rule may be derived from the geometric rule. To do so we shall rephrase the geometric rule (7) in a particularly useful and important way. Let z denote a general point in $\mathbb{C}$, and consider what happens to it—where it moves to—when it is multiplied by a fixed complex number $A = R\angle\phi$. According to (7), the length of z is magnified by R, while the angle of z is increased by ϕ. Now imagine that this is done simultaneously to *every* point of the plane:

> *Geometrically, multiplication by a complex number $A = R\angle\phi$ is a rotation of the plane through angle ϕ, and an expansion of the plane by factor R.* (9)

A few comments are in order:

- Both the rotation and the expansion are centred at the origin.
- It makes no difference whether we do the rotation followed by the expansion, or the expansion followed by the rotation.
- If $R < 1$ then the "expansion" is in reality a contraction.

Figure [5] illustrates the effect of such a transformation, the lightly shaded shapes being transformed into the darkly shaded shapes. Check for yourself that in this example $A = 1 + i\sqrt{3} = 2\angle\frac{\pi}{3}$.

It is now a simple matter to deduce the symbolic rule from the geometric rule. Recall the essential steps taken by Bombelli in deriving (6): (i) $i^2 = -1$; (ii) brackets can be multiplied out, i.e., if A, B, C, are complex numbers then $A(B + C) = AB + AC$. We have already seen that the geometric rule gives us (i), and figure [5] now reveals that (ii) is also true, for the simple reason that *rotations and expansions preserve parallelograms*. By the geometric definition of addition, $B + C$ is the fourth vertex of the parallelogram with vertices 0, B, C. To establish (ii), we merely observe that multiplication by A rotates and expands this parallelogram into another parallelogram with vertices 0, AB, AC and $A(B + C)$. This completes the derivation of (6).

Conversely, we now show how the geometric rule may be derived from the

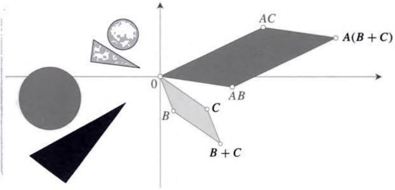


Figure [5]

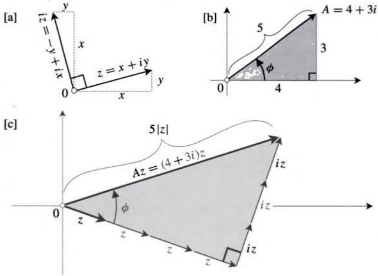


Figure [6]

symbolic rule[5]. We begin by considering the transformation $z \mapsto iz$. According to the symbolic rule, this means that $(x + iy) \mapsto (-y + ix)$, and [6a] reveals that *iz is z rotated through a right angle.* We now use this fact to interpret the transformation $z \mapsto Az$, where A is a general complex number. How this is done may be grasped sufficiently well using the example $A = 4 + 3i = 5\angle\phi$, where

[5]In every text we have examined this is done using trigonometric identities. We believe that the present argument supports the view that such identities are merely complicated manifestations of the simple rule for complex multiplication.

$\phi = \tan^{-1}(3/4)$. See [6b]. The symbolic rule says that brackets can be multiplied out, so our transformation may be rewritten as follows:

$$\begin{aligned} z \mapsto A\,z &= (4+3i)z \\ &= 4z + 3(iz) \\ &= 4z + 3\left(z \text{ rotated by } \tfrac{\pi}{2}\right). \end{aligned}$$

This is visualized in [6c]. We can now see that the shaded triangles in [6c] and [6b] are *similar*, so multiplication by $5\angle\phi$ does indeed rotate the plane by ϕ, and expand it by 5. Done.

II Euler's Formula

1 Introduction

It is time to replace the $r\angle\theta$ notation with a much better one that depends on the following miraculous fact:

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta}\ ! \qquad (10)$$

This result was discovered by Leonhard Euler around 1740, and it is called *Euler's formula* in his honour.

Before attempting to explain this result, let us say something of its meaning and utility. As illustrated in [7a], the formula says that $e^{i\theta}$ is the point on the unit circle at angle θ. Instead of writing a general complex number as $z = r\angle\theta$, we can now write $z = r\,e^{i\theta}$. Concretely, this says that to reach z we must take the unit vector $e^{i\theta}$ that points at z, then stretch it by the length of z. Part of the beauty of this representation is that the geometric rule (7) for multiplying complex numbers now looks almost obvious:

$$\left(R\,e^{i\phi}\right)\left(r\,e^{i\theta}\right) = Rr\,e^{i(\phi+\theta)}.$$

Put differently, algebraically manipulating $e^{i\theta}$ in the same way as the real function e^x yields true facts about complex numbers.

In order to explain Euler's formula we must first address the more basic question, "What does $e^{i\theta}$ *mean*?" Surprisingly, many authors answer this by defining $e^{i\theta}$, out of the blue, *to be* $(\cos\theta + i\sin\theta)$! This gambit is logically unimpeachable, but it is also a low blow to Euler, reducing one of his greatest achievements to a mere tautology. We will therefore give two heuristic arguments in support of (10); deeper arguments will emerge in later chapters.

2 Moving Particle Argument

Recall the basic fact that e^x is its own derivative: $\frac{d}{dx}e^x = e^x$. This is actually a *defining* property, that is, if $\frac{d}{dx}f(x) = f(x)$, and $f(0) = 1$, then $f(x) = e^x$. Similarly, if k is a real constant, then e^{kx} may be defined by the property

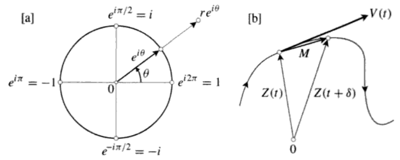


Figure [7]

$\frac{d}{dx} f(x) = k\, f(x)$. To extend the action of the ordinary exponential function e^x from real values of x to imaginary ones, let us cling to this property by insisting that it remain true if $k = i$, so that

$$\frac{d}{dt} e^{it} = i e^{it}. \tag{11}$$

We have used the letter t instead of x because we will now think of the variable as being *time*. We are used to thinking of the derivative of a real function as the slope of the tangent to the graph of the function, but how are we to understand the derivative in the above equation?

To make sense of this, imagine a particle moving along a curve in $\mathbb{C}$. See [7b]. The motion of the particle can be described *parametrically* by saying that at time t its position is the complex number $Z(t)$. Next, recall from physics that the *velocity* $V(t)$ is the vector—now thought of as a complex number—whose length and direction are given by the instantaneous speed, and the instantaneous direction of motion (tangent to the trajectory), of the moving particle. The figure shows the movement M of the particle between time t and $t + \delta$, and this should make it clear that

$$\frac{d}{dt} Z(t) = \lim_{\delta \to 0} \frac{Z(t+\delta) - Z(t)}{\delta} = \lim_{\delta \to 0} \frac{M}{\delta} = V(t).$$

Thus, given a complex function $Z(t)$ of a real variable t, we can always visualize Z as the position of a moving particle, and $\frac{dZ}{dt}$ as its velocity.

We can now use this idea to find the trajectory in the case $Z(t) = e^{it}$. See [8]. According to (11),

velocity $= V = iZ =$ *position, rotated through a right angle.*

Since the initial position of the particle is $Z(0) = e^0 = 1$, its initial velocity is i, and so it is moving vertically upwards. A split second later the particle will have moved very slightly in this direction, and its *new* velocity will be at right angles to its *new* position vector. Continuing to construct the motion in this way, it is clear that the particle will travel round the unit circle.

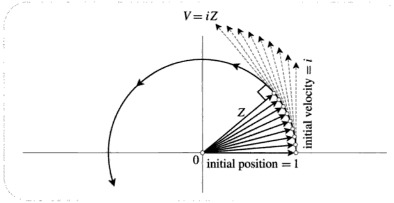


Figure [8]

Since we now know that $|Z(t)|$ remains equal to 1 throughout the motion, it follows that the particle's speed $|V(t)|$ also remains equal to 1. Thus after time $t = \theta$ the particle will have travelled a distance θ round the unit circle, and so the angle of $Z(\theta) = e^{i\theta}$ will be θ. This is the geometric statement of Euler's formula.

3 Power Series Argument

For our second argument, we begin by re-expressing the defining property $\frac{d}{dx} f(x) = f(x)$ in terms of power series. Assuming that $f(x)$ can be expressed in the form $a_0 + a_1 x + a_2 x^2 + \cdots$, a simple calculation shows that

$$e^x = f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

and further investigation shows that this series converges for all (real) values of x.

Putting x equal to a real value θ, this infinite sum of horizontal real numbers is visualized in [9]. To make sense of $e^{i\theta}$, we now cling to the power series and put $x = i\theta$:

$$e^{i\theta} = 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \cdots.$$

As illustrated in [9], this series is just as meaningful as the series for e^{θ}, but instead of the terms all having the same direction, here each term makes a right angle with the previous one, producing a kind of spiral.

This picture makes it clear that the known convergence of the series for e^{θ} guarantees that the spiral series for $e^{i\theta}$ converges to a definite point in $\mathbb{C}$. However, it is certainly *not* clear that it will converge to the point on the unit circle at angle θ. To see this, we split the spiral into its real and imaginary parts:

$$e^{i\theta} = C(\theta) + iS(\theta),$$

where

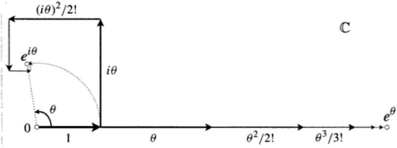


Figure [9]

$$C(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots, \quad \text{and} \quad S(\theta) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots.$$

At this point we could obtain Euler's formula by appealing to Taylor's Theorem, which shows that $C(\theta)$ and $S(\theta)$ are the power series for $\cos\theta$ and $\sin\theta$. However, we can also get the result by means of the following elementary argument that does not require Taylor's Theorem.

We wish to show two things about $e^{i\theta} = C(\theta) + iS(\theta)$: (i) it has unit length, and (ii) it has angle θ. To do this, first note that differentiation of the power series C and S yields

$$C' = -S \quad \text{and} \quad S' = C,$$

where a prime denotes differentiation with respect to θ.

To establish (i), observe that

$$\frac{d}{d\theta}|e^{i\theta}|^2 = (C^2 + S^2)' = 2(CC' + SS') = 0,$$

which means that the length of $e^{i\theta}$ is independent of θ. Since $e^{i0} = 1$, we deduce that $|e^{i\theta}| = 1$ for all θ.

To establish (ii) we must show that $\Theta(\theta) = \theta$, where $\Theta(\theta)$ denotes the angle of $e^{i\theta}$, so that

$$\tan\Theta(\theta) = \frac{S(\theta)}{C(\theta)}.$$

Since we already know that $C^2 + S^2 = 1$, we find that the derivative of the LHS of the above equation is

$$[\tan\Theta(\theta)]' = (1 + \tan^2\Theta)\,\Theta' = \left(1 + \frac{S^2}{C^2}\right)\Theta' = \frac{\Theta'}{C^2},$$

and that the derivative of the RHS is

$$\left[\frac{S}{C}\right]' = \frac{S'C - C'S}{C^2} = \frac{1}{C^2}.$$

Thus

$$\frac{d\Theta}{d\theta} = \Theta' = 1,$$

which implies that $\Theta(\theta) = \theta + const$. Taking the angle of $e^{i0} = 1$ to be 0 [would it make any geometric difference if we took it to be 2π?], we find that $\Theta = \theta$.

Although it is incidental to our purpose, note that we can now conclude (without Taylor's Theorem) that $C(\theta)$ and $S(\theta)$ are the power series of $\cos\theta$ and $\sin\theta$.

4 Sine and Cosine in Terms of Euler's Formula

A simple but important consequence of Euler's formula is that sine and cosine can be constructed from the exponential function. More precisely, inspection of [10] yields

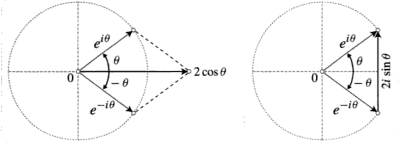


Figure [10]

$$e^{i\theta} + e^{-i\theta} = 2\cos\theta \quad \text{and} \quad e^{i\theta} - e^{-i\theta} = 2i\sin\theta,$$

or equivalently,

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad \text{and} \quad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}. \tag{12}$$

III Some Applications

1 Introduction

Often problems that do not appear to involve complex numbers are nevertheless solved most elegantly by viewing them through complex spectacles. In this section we will illustrate this point with a variety of examples taken from diverse areas of mathematics. Further examples may be found in the exercises at the end of the chapter.

The first example [trigonometry] merely illustrates the power of the concepts already developed, but the remaining examples develop important new ideas.

2 Trigonometry

All trigonometric identities may be viewed as arising from the rule for complex multiplication. In the following examples we will reduce clutter by using the following shorthand: $C \equiv \cos\theta$, $S \equiv \sin\theta$, and similarly, $c \equiv \cos\phi$, $s \equiv \sin\phi$.

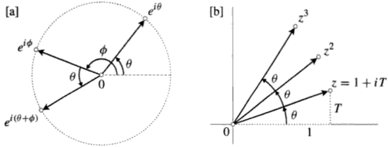


Figure [11]

To find an identity for $\cos(\theta+\phi)$, view it as a component of $e^{i(\theta+\phi)}$. See [11a]. Since

$$\begin{aligned}\cos(\theta+\phi)+i\sin(\theta+\phi) &= e^{i(\theta+\phi)}\\ &= e^{i\theta}e^{i\phi}\\ &= (C+iS)(c+is)\\ &= [Cc-Ss]+i[Sc+Cs],\end{aligned}$$

we obtain not only an identity for $\cos(\theta+\phi)$, but also one for $\sin(\theta+\phi)$:

$$\cos(\theta+\phi) = Cc - Ss \quad \text{and} \quad \sin(\theta+\phi) = Sc + Cs.$$

This illustrates another powerful feature of using complex numbers: every complex equation says two things at once.

To simultaneously find identities for $\cos 3\theta$ and $\sin 3\theta$, consider $e^{i3\theta}$:

$$\cos 3\theta + i\sin 3\theta = e^{i3\theta} = (e^{i\theta})^3 = (C+iS)^3 = \left[C^3 - 3CS^2\right] + i\left[3C^2S - S^3\right].$$

Using $C^2 + S^2 = 1$, these identities may be rewritten in the more familiar forms,

$$\cos 3\theta = 4C^3 - 3C \quad \text{and} \quad \sin 3\theta = -4S^3 + 3S.$$

We have just seen how to express trig functions of multiples of θ in terms of powers of trig functions of θ, but we can also go in the opposite direction. For example, suppose we want an identity for $\cos^4\theta$ in terms of multiples of θ. Since $2\cos\theta = e^{i\theta} + e^{-i\theta}$,

$$\begin{aligned}2^4\cos^4\theta &= \left(e^{i\theta}+e^{-i\theta}\right)^4\\ &= \left(e^{i4\theta}+e^{-i4\theta}\right)+4\left(e^{i2\theta}+e^{-i2\theta}\right)+6\\ &= 2\cos 4\theta + 8\cos 2\theta + 6\end{aligned}$$

$$\Longrightarrow \cos^4\theta = \tfrac{1}{8}\left[\cos 4\theta + 4\cos 2\theta + 3\right].$$

Although Euler's formula is extremely convenient for doing such calculations, it is not essential: all we are really using is the equivalence of the geometric and symbolic forms of complex multiplication. To stress this point, let us do an example without Euler's formula.

To find an identity for $\tan 3\theta$ in terms of $T = \tan\theta$, consider $z = 1 + iT$. See [11b]. Since z is at angle θ, z^3 will be at angle 3θ, so $\tan 3\theta = \mathrm{Im}(z^3)/\mathrm{Re}(z^3)$. Thus,

$$z^3 = (1+iT)^3 = (1-3T^2) + i(3T - T^3) \quad \Longrightarrow \quad \tan 3\theta = \frac{3T - T^3}{1 - 3T^2}.$$

3 Geometry

We shall base our discussion of geometric applications on a single example. In [12a] we have constructed squares on the sides of an arbitrary quadrilateral. Let

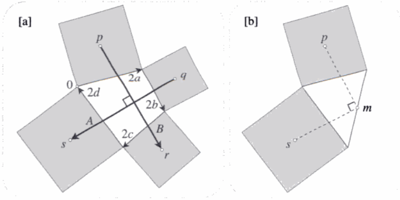


Figure [12]

us prove what this picture strongly suggests: *the line-segments joining the centres of opposite squares are perpendicular and of equal length.* It would require a great deal of ingenuity to find a purely geometric proof of this surprising result, so instead of relying on our own intelligence, let us invoke the intelligence of the complex numbers!

Introducing a factor of 2 for convenience, let $2a$, $2b$, $2c$, and $2d$ represent complex numbers running along the edges of the quadrilateral. The only condition is that the quadrilateral close up, i.e.,

$$a + b + c + d = 0.$$

As illustrated, choose the origin of $\mathbb{C}$ to be at the vertex where $2a$ begins. To reach the centre p of the square constructed on that side, we go along a, then an equal distance at right angles to a. Thus, since ia is a rotated through a right angle, $p = a + ia = (1+i)a$. Likewise,

$$q = 2a + (1+i)b, \quad r = 2a + 2b + (1+i)c, \quad s = 2a + 2b + 2c + (1+i)d.$$

The complex numbers $A = s - q$ (from q to s) and $B = r - p$ (from p to r) are therefore given by

$$A = (b + 2c + d) + i(d - b) \quad \text{and} \quad B = (a + 2b + c) + i(c - a).$$

We wish to show that A and B are perpendicular and of equal length. These two statements can be combined into the single complex statement $B = iA$, which says that B is A rotated by $(\pi/2)$. To finish the proof, note that this is the same thing as $A + iB = 0$, the verification of which is a routine calculation:

$$A + iB = (a + b + c + d) + i\,(a + b + c + d) = 0.$$

As a first step towards a purely geometric explanation of the result in [12a], consider [12b]. Here squares have been constructed on two sides of an arbitrary triangle, and, as the picture suggests, *the line-segments from their centres to the midpoint m of the remaining side are perpendicular and of equal length.* As is shown in Ex. 21, [12a] can be quickly deduced[6] from [12b]. The latter result can, of course, be proved in the same manner as above, but let us instead try to find a purely geometric argument.

To do so we will take an interesting detour, investigating translations and rotations of the plane in terms of complex functions. In reality, this "detour" is much more important than the geometric puzzle to which our results will be applied.

Let $\mathcal{T}_v$ denote a translation of the plane by v, so that a general point z is mapped to $\mathcal{T}_v(z) = z + v$. See [13a], which also illustrates the effect of the translation on a triangle. The *inverse* of $\mathcal{T}_v$, written $\mathcal{T}_v^{-1}$, is the transformation that undoes it; more formally, $\mathcal{T}_v^{-1}$ is defined by $\mathcal{T}_v^{-1} \circ \mathcal{T}_v = \mathcal{E} = \mathcal{T}_v \circ \mathcal{T}_v^{-1}$, where $\mathcal{E}$ is the "do nothing" transformation (called the *identity*) that maps each point to itself: $\mathcal{E}(z) = z$. Clearly, $\mathcal{T}_v^{-1} = \mathcal{T}_{-v}$.

If we perform $\mathcal{T}_v$, followed by another translation $\mathcal{T}_w$, then the composite mapping $\mathcal{T}_w \circ \mathcal{T}_v$ of the plane is another translation:

$$\mathcal{T}_w \circ \mathcal{T}_v(z) = \mathcal{T}_w(z + v) = z + (w + v) = \mathcal{T}_{w+v}(z).$$

This gives us an interesting way of motivating addition itself. If we had introduced a complex number v *as being* the translation $\mathcal{T}_v$, then we could have defined the "sum" of two complex numbers $\mathcal{T}_v$ and $\mathcal{T}_w$ to be the net effect of performing these translations in succession (in either order). Of course this would have been equivalent to the definition of addition that we actually gave.

[6]This approach is based on a paper of Finney [1970].

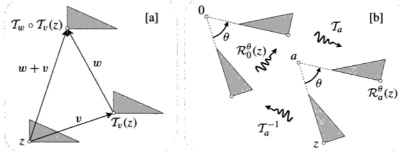


Figure [13]

Let $\mathcal{R}_a^\theta$ denote a rotation of the plane through angle θ about the point a. For example, $\mathcal{R}_a^\phi \circ \mathcal{R}_a^\theta = \mathcal{R}_a^{\theta+\phi}$, and $\left(\mathcal{R}_a^\theta\right)^{-1} = \mathcal{R}_a^{-\theta}$. As a first step towards expressing rotations as complex functions, note that (9) says that a rotation about the origin can be written as $\mathcal{R}_0^\theta(z) = e^{i\theta}z$.

As illustrated in [13b], the general rotation $\mathcal{R}_a^\theta$ can be performed by translating a to 0, rotating θ about 0, then translating 0 back to a:

$$\mathcal{R}_a^\theta(z) = \left(\mathcal{T}_a \circ \mathcal{R}_0^\theta \circ \mathcal{T}_a^{-1}\right)(z) = e^{i\theta}(z-a) + a = e^{i\theta}z + k,$$

where $k = a(1 - e^{i\theta})$. Thus we find that a rotation about any point can instead be expressed as an equal rotation about the origin, followed by a translation: $\mathcal{R}_a^\theta = \left(\mathcal{T}_k \circ \mathcal{R}_0^\theta\right)$. Conversely, a rotation of α about the origin followed by a translation of v can always be reduced to a single rotation:

$$\mathcal{T}_v \circ \mathcal{R}_0^\alpha = \mathcal{R}_c^\alpha, \quad \text{where} \quad c = v/(1 - e^{i\alpha}).$$

In the same way, you can easily check that if we perform the translation before the rotation, the net transformation can again be accomplished with a single rotation: $\mathcal{R}_0^\theta \circ \mathcal{T}_v = \mathcal{R}_p^\theta$. What is p?

The results just obtained are certainly not obvious geometrically [try them], and they serve to illustrate the power of thinking of translations and rotations as complex functions. As a further illustration, consider the net effect of performing two rotations about different points. Representing the rotations as complex functions, an easy calculation [exercise] yields

$$\left(\mathcal{R}_b^\phi \circ \mathcal{R}_a^\theta\right)(z) = e^{i(\theta+\phi)}z + v, \quad \text{where} \quad v = ae^{i\phi}(1 - e^{i\theta}) + b(1 - e^{i\phi}).$$

Unless $(\theta + \phi)$ is a multiple of 2π, the previous paragraph therefore tells us that

$$\mathcal{R}_b^\phi \circ \mathcal{R}_a^\theta = \mathcal{R}_c^{(\theta+\phi)}, \quad \text{where} \quad c = \frac{v}{1 - e^{i(\theta+\phi)}} = \frac{ae^{i\phi}(1 - e^{i\theta}) + b(1 - e^{i\phi})}{1 - e^{i(\theta+\phi)}}.$$

[What should c equal if $b = a$ or $\phi = 0$? Check the formula.] This result is

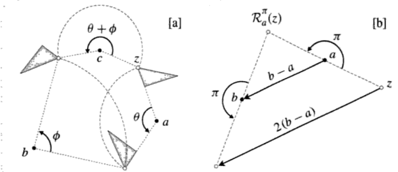


Figure [14]

illustrated in [14a]. Later we shall find a purely geometric explanation of this result, and, in the process, a very simple geometric construction of the point c given by the complicated formula above.

If, on the other hand, $(\theta + \phi)$ is a multiple of 2π, then $e^{i(\theta+\phi)} = 1$, and

$$\mathcal{R}_b^{\phi} \circ \mathcal{R}_a^{\theta} = \mathcal{T}_v, \quad \text{where} \quad v = (1 - e^{i\phi})(b - a).$$

For example, putting $\theta = \phi = \pi$, this predicts that $\mathcal{R}_b^{\pi} \circ \mathcal{R}_a^{\pi} = \mathcal{T}_{2(b-a)}$ is a translation by twice the complex number connecting the first centre of rotation to the second. That this is indeed true can be deduced directly from [14b].

The above result on the composition of two rotations implies [exercise] the following:

> *Let $\mathcal{M} = \mathcal{R}_{a_n}^{\theta_n} \circ \cdots \circ \mathcal{R}_{a_2}^{\theta_2} \circ \mathcal{R}_{a_1}^{\theta_1}$ be the composition of n rotations, and let $\Theta = \theta_1 + \theta_2 + \cdots + \theta_n$ be the total amount of rotation. In general, $\mathcal{M} = \mathcal{R}_c^{\Theta}$ (for some c), but if Θ is a multiple of 2π then $\mathcal{M} = \mathcal{T}_v$, for some v.*

Returning to our original problem, we can now give an elegant geometric explanation of the result in [12b]. Referring to [15a], let $\mathcal{M} = \mathcal{R}_m^{\pi} \circ \mathcal{R}_p^{(\pi/2)} \circ \mathcal{R}_s^{(\pi/2)}$. According to the result just obtained, $\mathcal{M}$ is a translation. To find out *what* translation, we need only discover the effect of $\mathcal{M}$ on a single point. Clearly, $\mathcal{M}(k) = k$, so $\mathcal{M}$ is the zero translation, i.e., the identity transformation $\mathcal{E}$. Thus

$$\mathcal{R}_p^{(\pi/2)} \circ \mathcal{R}_s^{(\pi/2)} = \left(\mathcal{R}_m^{\pi}\right)^{-1} \circ \mathcal{M} = \mathcal{R}_m^{\pi}.$$

If we define $s' = \mathcal{R}_m^{\pi}(s)$ then m is the midpoint of ss'. But, on the other hand,

$$s' = \left(\mathcal{R}_p^{(\pi/2)} \circ \mathcal{R}_s^{(\pi/2)}\right)(s) = \mathcal{R}_p^{(\pi/2)}(s).$$

Thus the triangle sps' is isosceles and has a right angle at p, so sm and pm are perpendicular and of equal length. Done.

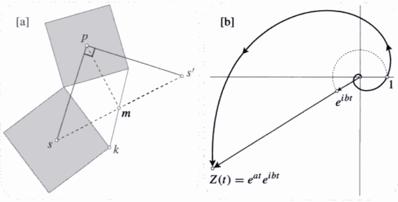


Figure [15]

4 Calculus

For our calculus example, consider the problem of finding the 100^{th} derivative of $e^x \sin x$. More generally, we will show how complex numbers may be used to find the n^{th} derivative of $e^{ax} \sin bx$.

In discussing Euler's formula we saw that e^{it} may be thought of as the location at time t of a particle travelling around the unit circle at unit speed. In the same way, e^{ibt} may be thought of as a unit complex number rotating about the origin with (angular) speed b. If we stretch this unit complex number by e^{at} as it turns, then its tip describes the motion of a particle that is spiralling away from the origin. See [15b].

The relevance of this to the opening problem is that the location of the particle at time t is

$$Z(t) = e^{at}e^{ibt} = e^{at}\cos bt + i\, e^{at}\sin bt.$$

Thus the derivative of $e^{at}\sin bt$ is simply the vertical (imaginary) component of the velocity V of Z.

We could find V simply by differentiating the components of Z in the above expression, but we shall instead use this example to introduce the geometric approach that will be used throughout this book. In [16], consider the movement $M = Z(t+\delta) - Z(t)$ of the particle between time t and $(t+\delta)$.

Recall that V is defined to be the limit of (M/δ) as δ tends to zero. Thus V and (M/δ) are very nearly equal if δ is very small. This suggests two intuitive ways of speaking, both of which will be used in this book: (i) we shall say that "$V = (M/\delta)$ when δ is *infinitesimal*" or (ii) that "V and (M/δ) are *ultimately equal*" (as δ tends to zero).

We stress that here the words "ultimately equal" and "infinitesimal" are being used in definite, technical senses; in particular, "infinitesimal" does not refer to some mystical, infinitely small quantity[7]. More precisely, if two quantities X and

[7]For more on this distinction, see the discussion in Chandrasekhar [1995].

Y depend on a third quantity δ, then

$$\lim_{\delta \to 0} \frac{X}{Y} = 1 \iff \text{“}X = Y \text{ for infinitesimal } \delta\text{”.}$$
$$\iff \text{“}X \text{ and } Y \text{ are ultimately equal as } \delta \text{ tends to zero”.}$$

It follows from the basic theorems on limits that "ultimate equality" inherits many of the properties of ordinary equality. For example, since V and (M/δ) are ultimately equal, so are $V\delta$ and M.

We now return to the problem of finding the velocity of the spiralling particle. As illustrated in [16], draw rays from 0 through $Z(t)$ and $Z(t+\delta)$, together with circular arcs (centred at 0) through those points. Now let A and B be the complex numbers connecting $Z(t)$ to the illustrated intersection points of these rays and arcs. If δ is infinitesimal, then B is at right angles to A and Z, and $M = A + B$.

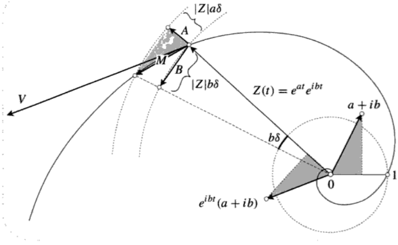


Figure [16]

Let us find the ultimate lengths of A and B. During the time interval δ, the angle of Z increases by $b\delta$, so the two rays cut off an arc of length $b\delta$ on the unit circle, and an arc of length $|Z|b\delta$ on the circle through Z. Thus $|B|$ is ultimately equal to $|Z|b\delta$. Next, note that $|A|$ is the increase in $|Z(t)|$ occurring in the time interval δ. Thus, since

$$\frac{d}{dt}|Z(t)| = \frac{d}{dt}e^{at} = a|Z|,$$

$|A|$ is ultimately equal to $|Z|a\delta$.

The shaded triangle at Z is therefore ultimately similar to the shaded right triangle with hypotenuse $a + ib$. Rotating the latter triangle by the angle of Z, you should now be able to see that if δ is infinitesimal then

$$\begin{aligned} M &= (a+ib) \text{ rotated by the angle of } Z, \text{ and expanded by } |Z|\delta \\ &= (a+ib)Z\delta \\ \Longrightarrow V &= \frac{d}{dt}Z = (a+ib)Z. \end{aligned} \tag{13}$$

Thus all rays from the origin cut the spiral at the same angle [the angle of $(a+ib)$], and the speed of the particle is proportional to its distance from the origin.

Note that although we have not yet given meaning to e^z (where z is a general complex number), it is certainly tempting to write $Z(t) = e^{at}e^{ibt} = e^{(a+ib)t}$. This makes the result (13) look very natural. Conversely, this suggests that we should *define* $e^z = e^{(x+iy)}$ to be $e^x e^{iy}$; another justification for this step will emerge in the next chapter.

Using (13), it is now easy to take further derivatives. For example, the acceleration of the particle is

$$\frac{d^2}{dt^2}Z = \frac{d}{dt}V = (a+ib)^2\,Z = (a+ib)\,V.$$

Continuing in this way, each new derivative is obtained by multiplying the previous one by $(a+ib)$. [Try sketching these successive derivatives in [16].] Writing $(a+ib) = R\,e^{i\phi}$, where $R = \sqrt{a^2+b^2}$ and ϕ is the appropriate value of $\tan^{-1}(b/a)$, we therefore find that

$$\frac{d^n}{dt^n}Z = (a+ib)^n\,Z = R^n\,e^{in\phi}\,e^{at}\,e^{ibt} = R^n e^{at} e^{i(bt+n\phi)}.$$

Thus

$$\frac{d^n}{dt^n}\left[e^{at}\sin bt\right] = (a^2+b^2)^{\frac{n}{2}}\,e^{at}\sin\left[bt + n\tan^{-1}(b/a)\right]. \tag{14}$$

5 Algebra

In the final year of his life (1716) Roger Cotes made a remarkable discovery that enabled him (in principle) to evaluate the family of integrals,

$$\int \frac{dx}{x^n-1},$$

where $n = 1, 2, 3, \ldots$. To see the connection with algebra, consider the case $n = 2$. The key observations are that the denominator (x^2-1) can be *factorized* into $(x-1)(x+1)$, and that the integrand can then be split into *partial fractions*:

$$\int \frac{dx}{x^2-1} = \tfrac{1}{2}\int \left[\frac{1}{x-1} - \frac{1}{x+1}\right] dx = \tfrac{1}{2}\ln\left[\frac{x-1}{x+1}\right].$$

As we shall see, for higher values of n one cannot completely factorize (x^n-1) into linear factors without employing complex numbers—a scarce and dubious commodity in 1716! However, Cotes was aware that if he could break down $(x^n - 1)$ into real *linear and quadratic* factors, then he would be able to evaluate the integral. Here, a "real quadratic" refers to a quadratic whose coefficients are all real numbers.

For example, $(x^4 - 1)$ can be broken down into $(x-1)(x+1)(x^2+1)$, yielding a partial fraction expression of the form

$$\frac{1}{x^4-1} = \frac{A}{x-1} + \frac{B}{x+1} + \frac{Cx}{x^2+1} + \frac{D}{x^2+1},$$

and hence an integral that can be evaluated in terms of ln and $\tan^{-1}$. More generally, even if the factorization involves more complicated quadratics than $(x^2 + 1)$, it is easy to show that only ln and $\tan^{-1}$ are needed to evaluate the resulting integrals.

In order to set Cotes' work on $(x^n - 1)$ in a wider context, we shall investigate the general connection between the roots of a polynomial and its factorization. This connection can be explained by considering the geometric series,

$$G_{m-1} = c^{m-1} + c^{m-2}z + c^{m-3}z^2 + \cdots + cz^{m-2} + z^{m-1},$$

in which c and z are complex. Just as in real algebra, this series may be summed by noting that zG_{m-1} and cG_{m-1} contain almost the same terms—try an example, say $m = 4$, if you have trouble seeing this. Subtracting these two expressions yields

$$(z - c)G_{m-1} = z^m - c^m, \tag{15}$$

and thus

$$G_{m-1} = \frac{z^m - c^m}{z - c}.$$

If we think of c as fixed and z as variable, then $(z^m - c^m)$ is an m^{th}-degree polynomial in z, and $z = c$ is a root. The result (15) says that this m^{th}-degree polynomial can be factored into the product of the linear term $(z - c)$ and the $(m-1)^{\text{th}}$-degree polynomial G_{m-1}.

In 1637 Descartes published an important generalization of this result. Let $P_n(z)$ denote a general polynomial of degree n:

$$P_n(z) = z^n + Az^{n-1} + \cdots + Dz + E,$$

where the coefficients $A, \ldots, E$ may be complex. Since (15) implies

$$P_n(z) - P_n(c) = (z - c)\left[G_{n-1} + AG_{n-2} + \cdots + D\right],$$

we obtain *Descartes' Factor Theorem* linking the existence of roots to factorizability:

> *If c is a solution of $P_n(z) = 0$ then $P_n(z) = (z - c)\,P_{n-1}$, where P_{n-1} is of degree $(n - 1)$.*

If we could in turn find a root c' of P_{n-1}, then the same reasoning would yield $P_n = (z-c)(z-c')\,P_{n-2}$. Continuing in this way, Descartes' theorem therefore holds out the promise of factoring P_n into precisely n linear factors:

$$P_n(z) = (z-c_1)(z-c_2)\cdots(z-c_n). \tag{16}$$

If we do not acknowledge the existence of complex roots (as in the early 18th century) then this factorization will be possible in some cases (e.g., z^2-1), and impossible in others (e.g., z^2+1). But, in splendid contrast to this, if one admits complex numbers then it can be shown that P_n *always has n roots in* $\mathbb{C}$, *and the factorization (16) is always possible.* This is called the *Fundamental Theorem of Algebra*, and we shall explain its truth in Chapter 7.

Each factor $(z-c_k)$ in (16) represents a complex number connecting the root c_k to the variable point z. Figure [17a] illustrates this for a general cubic polynomial. Writing each of these complex numbers in the form $R_k\,e^{i\phi_k}$, (16) takes the more vivid form

$$P_n(z) = R_1 R_2 \cdots R_n\, e^{i(\phi_1+\phi_2+\cdots+\phi_n)}.$$

Although the Fundamental Theorem of Algebra was not available to Cotes, let us see how it guarantees that he would succeed in his quest to decompose x^n-1 into real linear and quadratic factors. Cotes' polynomial has real coefficients, and, quite generally, we can show that

> *If a polynomial has real coefficients then its complex roots occur in complex conjugate pairs, and it can be factorized into real linear and quadratic factors.*

For if the coefficients $A, \ldots, E$ of $P_n(z)$ are all real then $P_n(c) = 0$ implies [exercise] $P_n(\overline{c}) = 0$, and the factorization (16) contains

$$(z-c)(z-\overline{c}) = z^2 - (c+\overline{c})\,z + c\overline{c} = z^2 - 2\,\mathrm{Re}(c)\,z + |c|^2,$$

which is a real quadratic.

Let us now discuss how Cotes was able to factorize x^n-1 into real linear and quadratic factors *by appealing to the geometry of the regular n-gon*. [An "n-gon" is an n-sided polygon.] To appreciate the following, place yourself in his 18th century shoes and forget all you have just learnt concerning the Fundamental Theorem of Algebra; even forget about complex numbers and the complex plane!

For the first few values of n, the desired factorizations of $U_n(x) = x^n - 1$ are not too hard to find:

$$\begin{aligned}
U_2(x) &= (x-1)(x+1), && (17)\\
U_3(x) &= (x-1)(x^2+x+1), && (18)\\
U_4(x) &= (x-1)(x+1)(x^2+1), && (19)\\
U_5(x) &= (x-1)\left(x^2+\left[\tfrac{1+\sqrt{5}}{2}\right]x+1\right)\left(x^2+\left[\tfrac{1-\sqrt{5}}{2}\right]x+1\right), &&
\end{aligned}$$

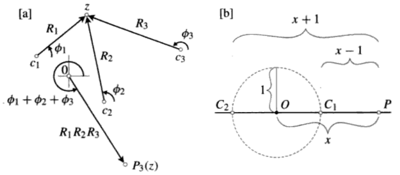


Figure [17]

but the general pattern seems elusive.

To find such a pattern, let us try to *visualize* the simplest case, (17). See [17b]. Let O be a fixed point, and P a variable point, on a line in the plane (which we are *not* thinking of as $\mathbb{C}$), and let x denote the distance OP. If we now draw a circle of unit radius centred at O, and let C_1 and C_2 be its intersection points with the line, then clearly[8] $U_2(x) = PC_1 \cdot PC_2$.

To understand quadratic factors in this spirit, let us skip over (18) to the simpler quadratic in (19). This factorization of $U_4(x)$ is the best we could do without complex numbers, but ideally we would have liked to have decomposed $U_4(x)$ into *four* linear factors. This suggests that we rewrite (19) as

$$U_4(x) = (x-1)(x+1)\sqrt{x^2+1}\sqrt{x^2+1},$$

the last two "factors" being analogous to genuine linear factors. If we are to interpret this expression (by analogy with the previous case) as the product of the distances of P from four fixed points, then the points corresponding to the last two "factors" must be *off the line*. More precisely, Pythagoras' Theorem tells us that a point whose distance from P is $\sqrt{x^2+1^2}$ must lie at unit distance from O in a direction at right angles to the line OP. Referring to [18a], we can now see that $U_4(x) = PC_1 \cdot PC_2 \cdot PC_3 \cdot PC_4$, where $C_1C_2C_3C_4$ is the illustrated square inscribed in the circle.

Since we have factorized $U_4(x)$ with the regular 4-gon (the square), perhaps we can factorize $U_3(x)$ with the regular 3-gon (the equilateral triangle). See [18b]. Applying Pythagoras' Theorem to this figure,

$$\begin{aligned} PC_1 \cdot PC_2 \cdot PC_3 &= PC_1 \cdot (PC_2)^2 = (x-1)\left(\left[x+\tfrac{1}{2}\right]^2 + \left[\tfrac{\sqrt{3}}{2}\right]^2\right) \\ &= (x-1)(x^2+x+1), \end{aligned}$$

[8]Here, and in what follows, we shall suppose for convenience that $x > 1$, so that $U_n(x)$ is positive.

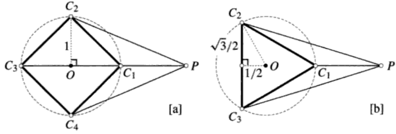


Figure [18]

which is indeed the desired factorization (18) of $U_3(x)$!

A plausible generalization for U_n now presents itself:

> *If $C_1 C_2 C_3 \cdots C_n$ is a regular n-gon inscribed in a circle of unit radius centred at O, and P is the point on OC_1 at distance x from O, then $U_n(x) = PC_1 \cdot PC_2 \cdots PC_n$.*

This is Cotes' result. Unfortunately, he stated it without proof, and he left no clue as to how he discovered it. Thus we can only speculate that he may have been guided by an argument like the one we have just supplied[9].

Since the vertices of the regular n-gon will always come in symmetric pairs that are equidistant from P, the examples in [18] make it clear that Cotes' result is indeed equivalent to factorizing $U_n(x)$ into real linear and quadratic factors.

Recovering from our feigned bout of amnesia concerning complex numbers and their geometric interpretation, Cotes' result becomes simple to understand and to prove. Taking O to be the origin of the complex plane, and C_1 to be 1, the vertices of Cotes' n-gon are given by $C_{k+1} = e^{ik(2\pi/n)}$. See [19], which illustrates the case $n = 12$. Since $(C_{k+1})^n = e^{ik2\pi} = 1$, all is suddenly clear: *The vertices of the regular n-gon are the n complex roots of $U_n(z) = z^n - 1$.* Because the solutions of $z^n - 1 = 0$ may be written formally as $z = \sqrt[n]{1}$, the vertices of the n-gon are called the n^{th} *roots of unity*.

By Descartes' Factor Theorem, the complete factorization of $(z^n - 1)$ is therefore

$$z^n - 1 = U_n(z) = (z - C_1)(z - C_2) \cdots (z - C_n),$$

with each conjugate pair of roots yielding a real quadratic factor,

$$\left(z - e^{ik(2\pi/n)}\right)\left(z - e^{-ik(2\pi/n)}\right) = z^2 - 2z\cos\left[\tfrac{2k\pi}{n}\right] + 1.$$

Each factor $(z - C_k) = R_k\, e^{i\phi_k}$ may be viewed (cf. [17a]) as a complex number connecting a vertex of the n-gon to z. Thus, if P is an arbitrary point in the plane

9 Stillwell [1989, p. 195] has instead speculated that Cotes used complex numbers (as we are about to), but then deliberately stated his findings in a form that did not require them.

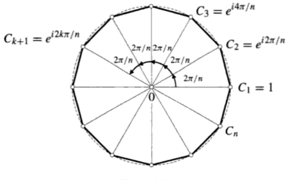


Figure [19]

(not merely a point on the real axis), then we obtain the following generalized form of Cotes' result:

$$U_n(P) = [PC_1 \cdot PC_2 \cdots PC_n]\, e^{i\Phi},$$

where $\Phi = (\phi_1 + \phi_2 + \cdots + \phi_n)$. If P happens to be a real number (again supposed greater than 1) then $\Phi = 0$ [make sure you see this], and we recover Cotes' result.

We did not immediately state and prove Cotes' result in terms of complex numbers because we felt there was something rather fascinating about our first, direct approach. Viewed in hindsight, it shows that even if we attempt to avoid complex *numbers*, we cannot avoid the geometry of the complex *plane!*

6 Vectorial Operations

Not only is complex addition the same as vector addition, but we will now show that the familiar vectorial operations of dot and cross products (also called scalar and vector products) are both subsumed by complex multiplication. Since these vectorial operations are extremely important in physics—they were discovered by physicists!—their connection with complex multiplication will prove valuable both in applying complex analysis to the physical world, and in using physics to understand complex analysis.

When a complex number $z = x + iy$ is being thought of merely as a vector, we shall write it in bold type, with its components in a column:

$$z = x + iy \iff \mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Although the dot and cross product are meaningful for arbitrary vectors in space, we shall assume in the following that our vectors all lie in a single plane—the complex plane.

Given two vectors **a** and **b**, figure [20a] recalls the definition of the dot product as the length of one vector, times the projection onto that vector of the other vector:

$$\mathbf{a} \cdot \mathbf{b} = |a|\,|b| \cos\theta = \mathbf{b} \cdot \mathbf{a},$$

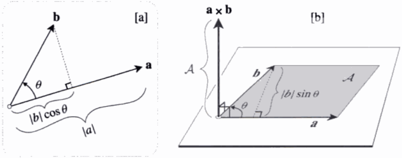


Figure [20]

where θ is the angle between **a** and **b**.

Figure [20b] recalls the definition of the cross product: $\mathbf{a} \times \mathbf{b}$ is the vector perpendicular to the plane of **a** and **b** whose length is equal to the area $\mathcal{A}$ of the parallelogram spanned by **a** and **b**. But wait, there are *two* (opposite) directions perpendicular to $\mathbb{C}$; which should we choose?

Writing $\mathcal{A} = |a|\,|b| \sin\theta$, the area $\mathcal{A}$ has a *sign* attached to it. An easy way to see this sign is to think of the angle θ *from* **a** *to* **b** as lying in range $-\pi$ to π; the sign of $\mathcal{A}$ is then the same as θ. If $\mathcal{A} > 0$, as in [20b], then we define $\mathbf{a} \times \mathbf{b}$ to point upwards from the plane, and if $\mathcal{A} < 0$ we define it to point downwards. It follows that $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$.

This conventional definition of $\mathbf{a} \times \mathbf{b}$ is intrinsically three-dimensional, and it therefore presents a problem: if **a** and **b** are thought of as complex numbers, $\mathbf{a} \times \mathbf{b}$ *cannot be*, for it does not lie in the (complex) plane of **a** and **b**. No such problem exists with the dot product because $\mathbf{a} \cdot \mathbf{b}$ is simply a real number, and this suggests a way out.

Since all our vectors will be lying in the same plane, their cross products will all have equal (or opposite) directions, so the only distinction between one cross product and another will be the value of $\mathcal{A}$. For the purposes of this book we will therefore *redefine the cross product to be the (signed) area $\mathcal{A}$ of the parallelogram spanned by* **a** *and* **b**:

$$\mathbf{a} \times \mathbf{b} = |a|\,|b| \sin\theta = -(\mathbf{b} \times \mathbf{a}).$$

Figure [21] shows two complex numbers $a = |a|\,e^{i\alpha}$ and $b = |b|\,e^{i\beta}$, the angle from a to b being $\theta = (\beta - \alpha)$. To see how their dot and cross products are related to complex multiplication, consider the effect of multiplying each point in $\mathbb{C}$ by $\overline{a}$. This is a rotation of $-\alpha$ and an expansion of $|a|$, and if we look at the image under this transformation of the shaded right triangle with hypotenuse b, then we immediately see that

$$\overline{a}\, b = \mathbf{a} \cdot \mathbf{b} + i\,(\mathbf{a} \times \mathbf{b}). \tag{20}$$

Of course we could also have got this by simple calculation:

$$\overline{a}\, b = (|a|\,e^{-i\alpha})(|b|\,e^{i\beta}) = |a|\,|b|\,e^{i(\beta-\alpha)} = |a|\,|b|\,e^{i\theta} = |a|\,|b|(\cos\theta + i\,\sin\theta).$$

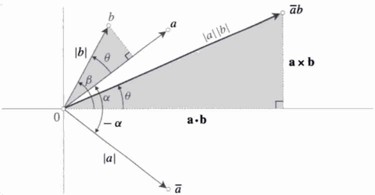


Figure [21]

When we refer to the dot and cross products as "vectorial operations" we mean that they are defined *geometrically*, independently of any particular choice of coordinate axes. However, once such a choice has been made, (20) makes it easy to *express* these operations in terms of Cartesian coordinates. Writing $a = x + iy$ and $b = x' + iy'$,

$$\bar{a}\, b = (x - iy)(x' + iy') = (xx' + yy') + i\,(xy' - yx'),$$

so

$$\begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} x' \\ y' \end{pmatrix} = xx' + yy' \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} \times \begin{pmatrix} x' \\ y' \end{pmatrix} = xy' - yx'.$$

We end with an example that illustrates the importance of the sign of the area $(\mathbf{a} \times \mathbf{b})$. Consider the problem of finding the area $\mathcal{A}$ of the quadrilateral in [22a] whose vertices are, *in counterclockwise order*, a, b, c, and d. Clearly this is just the sum of the ordinary, unsigned areas of the four triangles formed by joining the vertices of the quadrilateral to the origin. Thus, since the area of each triangle is

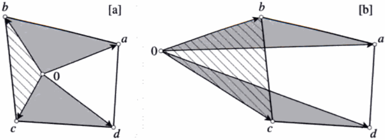


Figure [22]

simply half the area of the corresponding parallelogram,

$$\begin{aligned} \mathcal{A} &= \tfrac{1}{2}\,[(\mathbf{a}\times\mathbf{b})+(\mathbf{b}\times\mathbf{c})+(\mathbf{c}\times\mathbf{d})+(\mathbf{d}\times\mathbf{a})] \\ &= \tfrac{1}{2}\,\mathrm{Im}\big[\,\overline{a}\,b+\overline{b}\,c+\overline{c}\,d+\overline{d}\,a\big]. \end{aligned} \tag{21}$$

Obviously this formula could easily be generalized to polygons with more than four sides.

But what if 0 is outside the quadrilateral? In [22b], $\mathcal{A}$ is clearly the sum of the ordinary areas of three of the triangles, *minus* the ordinary area of the striped triangle. Since the angle from **b** to **c** is negative, $\frac{1}{2}(\mathbf{b}\times\mathbf{c})$ is automatically the negative of the striped area, and $\mathcal{A}$ *is therefore given by exactly the same formula as before!*

Can you find a location for 0 that makes two of the signed areas negative? Check that the formula still works. Exercise 35 shows that (21) *always* works.

IV Transformations and Euclidean Geometry*

1 Geometry Through the Eyes of Felix Klein

Even with the benefit of enormous hindsight, it is hard to introduce complex numbers in a compelling manner. Historically, we have seen how cubic equations forced them upon us algebraically, and in discussing Cotes' work we saw something of the inevitability of their geometric interpretation. In this section we will attempt to show how complex numbers arise very naturally, almost inevitably, from a careful re-examination of *plane Euclidean geometry*[10].

As the * following the title of this section indicates, the material it contains may be omitted. However, in addition to "explaining" complex numbers, these ideas are very interesting in their own right, and they will also be needed for an understanding of other optional sections of the book.

Although the ancient Greeks made many beautiful and remarkable discoveries in geometry, it was two thousand years later that Felix Klein first asked and answered the question, "What *is* geometry?"

Let us restrict ourselves from the outset to *plane* geometry. One might begin by saying that this is the study of geometric properties of geometric figures in the plane, but what are (i) "geometric properties", and (ii) "geometric figures"? We will concentrate on (i), swiftly passing over (ii) by interpreting "geometric figure" as anything we might choose to draw on an infinitely large piece of flat paper with an infinitely fine pen.

As for (i), we begin by noting that if two figures (e.g., two triangles) have the same geometric properties, then (from the point of view of geometry) they must be the "same", "equal", or, as one usually says, *congruent*. Thus if we had a clear definition of congruence ("geometric equality") then we could reverse this

[10]The excellent book by Nikulin and Shafarevich [1987] is the only other work we know of in which a similar attempt is made.

observation and *define geometric properties as those properties that are common to all congruent figures.* How, then, can we tell if two figures are geometrically equal?

Consider the triangles in [23], and imagine that they are pieces of paper that you could pick up in your hand. To see if T is congruent to T', you could pick up T and check whether it could be placed on top of T'. Note that it is essential that we be allowed to move T in space: in order to place T on top of $\widetilde{T}$ we must first flip it over; we can't just slide T around within the plane. Tentatively generalizing, this suggests that *a figure F is congruent to another figure F' if there exists a motion of F through space that makes it coincide with F'.* Note that the discussion suggests that there are two fundamentally different types of motion: those that involve flipping the figure over, and those that do not. Later, we shall return to this important point.

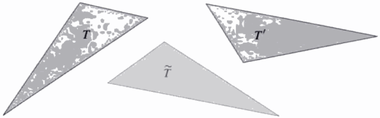


Figure [23]

It is clearly somewhat unsatisfactory that in attempting to define geometry in the *plane* we have appealed to the idea of motion through *space*. We now rectify this. Returning to [23], imagine that T and T' are drawn on separate, transparent sheets of plastic. Instead of picking up just the triangle T, we now pick up the *entire sheet* on which it is drawn, then try to place it on the second sheet so as to make T coincide with T'. At the end of this motion, each point A on T's sheet lies over a point A' of T''s sheet, and we can now define the motion $\mathcal{M}$ to be this mapping $A \mapsto A' = \mathcal{M}(A)$ of the plane to itself.

However, not any old mapping qualifies as a motion, for we must also capture the (previously implicit) idea of the sheet remaining *rigid* while it moves, so that distances between points remain constant during the motion. Here, then, is our definition:

> *A motion $\mathcal{M}$ is a mapping of the plane to itself such that the distance between any two points A and B is equal to the distance between their images $A' = \mathcal{M}(A)$ and $B' = \mathcal{M}(B)$.* (22)

Note that what we have called a motion is often termed a "rigid motion", or an "isometry".

Armed with this precise concept of motion, our final definition of geometric equality becomes

F is congruent to F', written $F \cong F'$, if there exists a motion $\mathcal{M}$ such that $F' = \mathcal{M}(F)$. (23)

Next, as a consequence of our earlier discussion, *a geometric property of a figure is one that is unaltered by all possible motions of the figure.* Finally, in answer to the opening question of "What is geometry?", Klein would answer that it is the study of these so-called *invariants* of the set of motions.

One of the most remarkable discoveries of the last century was that Euclidean geometry is not the *only* possible geometry. Two of these so-called *non-Euclidean* geometries will be studied in Chapter 6, but for the moment we wish only to explain how Klein was able to generalize the above ideas so as to embrace such new geometries.

The aim in (23) was to use a family of transformations to introduce a concept of geometric equality. But will this $\cong$-type of equality behave in the way we would like and expect? To answer this we must first make these expectations explicit. So as not to confuse this general discussion with the particular concept of congruence in (23), let us denote geometric equality by $\sim$.

(i) A figure should equal itself: $F \sim F$, for all F.

(ii) If F equals F', then F' should equal F: $F \sim F' \Rightarrow F' \sim F$.

(iii) If F and F' are equal, and F' and F'' are equal, then F and F'' should also be equal: $F \sim F'$ & $F' \sim F'' \Rightarrow F \sim F''$.

Any relation satisfying these expectations is called an *equivalence relation*.

Now suppose that we retain the definition (23) of geometric equality, but that we generalize the definition of "motion" given in (22) by replacing the family of distance-preserving transformations with some other family G of transformations. It should be clear that not any old G will be compatible with our aim of defining geometric equality. Indeed, (i), (ii), and (iii) imply that G must have the following very special structure, which is illustrated[11] in [24].

(i) The family G must contain a transformation $\mathcal{E}$ (called the *identity*) that maps each point to itself.

(ii) If G contains a transformation $\mathcal{M}$, then it must also contain a transformation $\mathcal{M}^{-1}$ (called the *inverse*) that undoes $\mathcal{M}$. [Check for yourself that for $\mathcal{M}^{-1}$ to exist (let alone be a member of G) $\mathcal{M}$ must have the special properties of being (a) *onto* and (b) *one-to-one*, i.e., (a) every point must be the image of some point, and (b) distinct points must have distinct images.]

(iii) If $\mathcal{M}$ and $\mathcal{N}$ are members of G then so is the composite transformation $\mathcal{N} \circ \mathcal{M} = (\mathcal{M}$ followed by $\mathcal{N})$. This property of G is called *closure*.

We have thus arrived, very naturally, at a concept of fundamental importance in the

[11] Here G is the group of *projections*. If we do a perspective drawing of figures in the plane, then the mapping from that plane to the "canvas" plane is called a *perspectivity*. A projection is then defined to be any sequence of perspectivities. Can you see why the set of projections should form a group?

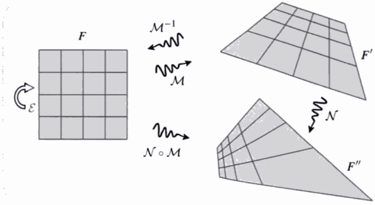


Figure [24]

whole of mathematics: a family G of transformations that satisfies these three[12] requirements is called a *group*.

Let us check that the motions defined in (22) do indeed form a group: (i) Since the identity transformation preserves distances, it is a motion. (ii) Provided it exists, the inverse of a motion will preserve distances and hence will be a motion itself. As for existence, (a) it is certainly *plausible* that when we apply a motion to the entire plane then the image is the entire plane—we will prove this later—and (b) the non-zero distance between distinct points is preserved by a motion, so their images are again distinct. (iii) If two transformations do not alter distances, then applying them in succession will not alter distances either, so the composition of two motions is another motion.

Klein's idea was that we could first select a group G at will, then define a corresponding "geometry" as the study of the invariants of that G. [Klein first announced this idea in 1872—when he was 23 years old!—at the University of Erlangen, and it has thus come to be known as his *Erlangen Program*.] For example, if we choose G to be the group of motions, we recover the familiar Euclidean geometry of the plane. But this is far from being the *only* geometry of the plane, as the so-called *projective geometry* of [24] illustrates.

Klein's vision of geometry was broader still. We have been concerned with what geometries are possible *when figures are drawn anywhere in the plane*, but suppose for example that we are only allowed to draw within some disc D. It should be clear that we can construct "geometries of D" in exactly the same way that we constructed geometries of the plane: given a group H of transformations of D to itself, the corresponding geometry is the study of the invariants of H. If you doubt that any such groups exist, consider the set of all rotations around the centre of D.

[12]In more abstract settings it is necessary to add a fourth requirement of *associativity*, namely, $\mathcal{A} \circ (\mathcal{B} \circ \mathcal{C}) = (\mathcal{A} \circ \mathcal{B}) \circ \mathcal{C}$. Of course for transformations this is automatically true.

The reader may well feel that the above discussion is a chronic case of mathematical generalization running amuck—that the resulting conception of geometry is (to coin a phrase) "as subtle as it is useless". Nothing could be further from the truth! In Chapter 3 we shall be led, very naturally, to consider a particularly interesting group of transformations of a disc to itself. The resulting non-Euclidean geometry is called *hyperbolic* or *Lobachevskian* geometry, and it is the subject of Chapter 6. Far from being useless, this geometry has proved to be an immensely powerful tool in diverse areas of mathematics, and the insights it continues to provide lie on the cutting edge of contemporary research.

2 Classifying Motions

To understand the foundations of Euclidean geometry, it seems we must study its group of motions. At the moment, this group is defined rather abstractly as the set of distance-preserving mappings of the plane to itself. However, it is easy enough to think of concrete examples of motions: a rotation of the plane about an arbitrary point, a translation of the plane, or a reflection of the plane in some line. Our aim is to understand the most general possible motions in equally vivid terms.

We begin by stating a key fact:

A motion is uniquely determined by its effect on any triangle (i.e., on any three non-collinear points). (24)

By this we mean that knowing what happens to the three points tells us what must happen to *every* point in the plane. To see this, first look at [25]. This shows that

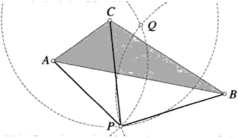


Figure [25]

each point P is uniquely determined by its distances from the vertices A, B, C of such a triangle[13]. The distances from A and B yield two circles which (in general) intersect in two points, P and Q. The third distance (from C) then picks out P.

To obtain the result (24), now look at [26]. This illustrates a motion $\mathcal{M}$ mapping A, B, C to A', B', C'. By the very definition of a motion, $\mathcal{M}$ must map an arbitrary

[13]This is how earthquakes are located. Two types of wave are emitted by the quake as it begins: fast-moving "P-waves" of compression, and slower-moving "S-waves" of destructive shear. Thus the P-waves will arrive at a seismic station before the S-waves, and the time-lag between these events may be used to calculate the distance of the quake from that station. Repeating this calculation at two more seismic stations, the quake may be located.

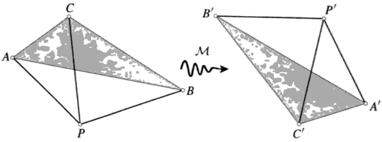


Figure [26]

point P to a point P' whose distances from A', B', C' are equal to the original distances of P from A, B, C. Thus, as shown, P' is uniquely determined. Done.

A big step towards classification is the realization that there are two fundamentally different kinds of motions. In terms of our earlier conception of motion through space, the distinction is whether or not a figure must be flipped over before it can be placed on top of a congruent figure. To see how this dichotomy arises in terms of the new definition (22), suppose that a motion sends two points A and B to A' and B'. See [27]. According to (24), the motion is not yet determined: we need to know the image of any (non-collinear) third point C, such as the one shown in [27]. Since motions preserve the distances of C from A and B, there are just two possibilities for the image of C, namely, C' and its reflection $\widetilde{C}$ in the line L through A' and B'. Thus there are precisely two motions ($\mathcal{M}$ and $\widetilde{\mathcal{M}}$, say) that map A, B to A', B': $\mathcal{M}$ sends C to C', and $\widetilde{\mathcal{M}}$ sends C to $\widetilde{C}$.

A distinction can be made between $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ by looking at how they affect angles. All motions preserve the magnitude of angles, but we see that $\mathcal{M}$ also preserves the *sense* of the angle θ, while $\widetilde{\mathcal{M}}$ reverses it. The fundamental nature of this distinction can be seen from the fact that $\mathcal{M}$ must in fact preserve *all* angles, while $\widetilde{\mathcal{M}}$ must reverse all angles.

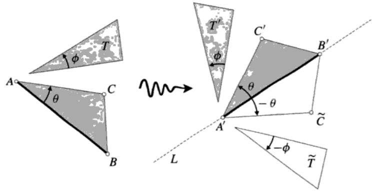


Figure [27]

To see this, consider the fate of the angle ϕ in the triangle T. If C goes to C' (i.e., if the motion is $\mathcal{M}$) then, carrying out the construction indicated in [26], the image of T is T', and the angle is preserved. If, on the other hand, C goes to $\widetilde{C}$ (i.e., if the motion is $\widetilde{\mathcal{M}}$) then the image of T is the reflection $\widetilde{T}$ of T' in L, and the angle is reversed. Motions that preserve angles are called *direct*, and those that reverse angles are called *opposite*. Thus rotations and translations are direct, while reflections are opposite. Summarizing what we have found,

There is exactly one direct motion $\mathcal{M}$ (and exactly one opposite motion $\widetilde{\mathcal{M}}$) that maps a given line-segment AB to another line-segment $A'B'$ of equal length. Furthermore, $\widetilde{\mathcal{M}} = (\mathcal{M}$ followed by reflection in the line $A'B')$. (25)

To understand motions we may thus consider two randomly drawn segments AB and $A'B'$ of equal length, then find *the* direct motion (and *the* opposite motion) that maps one to the other. It is now easy to show that

Every direct motion is a rotation, or else (exceptionally) a translation. (26)

Note that this result gives us greater insight into our earlier calculations on the composition of rotations and translations: since the composition of any two direct motions is another direct motion [why?], it can only be a rotation or a translation. Conversely, those calculations allow us to restate (26) in a very neat way:

Every direct motion can be expressed as a complex function of the form $\mathcal{M}(z) = e^{i\theta}z + v$. (27)

We now establish (26). If the line-segment $A'B'$ is parallel to AB then the vectors $\overrightarrow{AB}$ and $\overrightarrow{A'B'}$ are either equal or opposite. If they are equal, as in [28a], the motion is a translation; if they are opposite, as in [28b], the motion is a rotation of π about the intersection point of the lines AA' and BB'.

If the segments are not parallel, produce them (if necessary) till they meet at M, and let θ be the angle between the directions of $\overrightarrow{AB}$ and $\overrightarrow{A'B'}$. See [28c]. First recall an elementary property of circles: the chord AA' subtends the same angle θ at every point of the circular arc AMA'. Next, let O denote the intersection point of this arc with the perpendicular bisector of AA'. We now see that the direct motion carrying AB to $A'B'$ is a rotation of θ about O, for clearly A is rotated to A', and the direction of $\overrightarrow{AB}$ is rotated into the direction of $\overrightarrow{A'B'}$. Done.

The sense in which translations are "exceptional" is that if the two segments are drawn at random then it is very unlikely that they will be parallel. Indeed, given AB, a translation is only needed for one possible direction of $A'B'$ out of infinitely many, so the mathematical probability that a random direct motion is a translation is actually *zero!*

Direct transformations will be more important to us than opposite ones, so we relegate the investigation of opposite motions to Exs. 39, 40, 41. The reason for

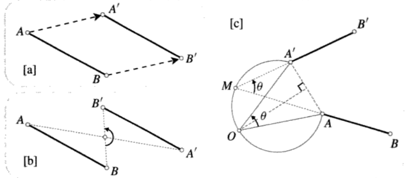


Figure [28]

the greater emphasis on direct motions stems from the fact that they form a group (a *subgroup* of the full group of motions), while the opposite motions do not. Can you see why?

3 Three Reflections Theorem

In chemistry one is concerned with the interactions of atoms, but to gain deeper insights one must study the electrons, protons, and neutrons from which atoms are built. Likewise, though our concern is with direct motions, we will gain deeper insights by studying the opposite motions from which direct motions are built. More precisely,

Every direct motion is the composition of two reflections. (28)

Note that the second sentence of (25) then implies that *every opposite motion is the composition of three reflections*. See Ex. 39. In brief, every motion is the composition of either two or three reflections, a result that is called the *Three Reflections Theorem*[14].

Earlier we tried to show that the set of motions forms a group, but it was not clear that every motion had an inverse. The Three Reflections Theorem settles this neatly and explicitly, for the inverse of a sequence of reflections is obtained by reversing the order in which the reflections are performed.

In what follows, let $\mathfrak{R}_L$ denote reflection in a line L. Thus reflection in L_1 followed by reflection in L_2 is written $\mathfrak{R}_{L_2} \circ \mathfrak{R}_{L_1}$. According to (26), proving (28) amounts to showing that every rotation (and every translation) is of the form $\mathfrak{R}_{L_2} \circ \mathfrak{R}_{L_1}$. This is an immediate consequence of the following:

If L_1 and L_2 intersect at O, and the angle from L_1 to L_2 is ϕ, then $\mathfrak{R}_{L_2} \circ \mathfrak{R}_{L_1}$ is a rotation of 2ϕ about O,

and

[14]Results such as (26) may instead be viewed as consequences of this theorem; see Stillwell [1992] for an elegant and elementary exposition of this approach.

If L_1 and L_2 are parallel, and $\mathbf{V}$ *is the perpendicular connecting vector from L_1 to L_2, then $\Re_{L_2} \circ \Re_{L_1}$ is a translation of* $2\mathbf{V}$.

Both these results are easy enough to prove directly [try it!], but the following is perhaps more elegant.

First, since $\Re_{L_2} \circ \Re_{L_1}$ is a direct motion (because it reverses angles twice), it is either a rotation or a translation. Second, note that rotations and translations may be distinguished by their *invariant curves*, that is, curves that are mapped into themselves. For a rotation about a point O, the invariant curves are circles centred at O, while for a translation they are lines parallel to the translation.

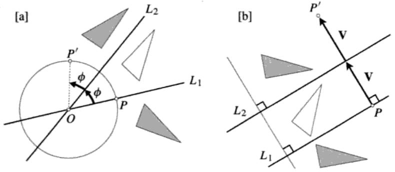


Figure [29]

Now look at [29a]. Clearly $\Re_{L_2} \circ \Re_{L_1}$ leaves invariant any circle centred at O, so it is a rotation about O. To see that the angle of the rotation is 2ϕ, consider the image P' of any point P on L_1. Done.

Now look at [29b]. Clearly $\Re_{L_2} \circ \Re_{L_1}$ leaves invariant any line perpendicular to L_1 and L_2, so it is a translation parallel to such lines. To see that the translation is $2\mathbf{V}$, consider the image P' of any point P on L_1. Done.

Note that a rotation of θ can be represented as $\Re_{L_2} \circ \Re_{L_1}$, where L_1, L_2 is *any* pair of lines that pass through the centre of the rotation and that contain an angle $(\theta/2)$. Likewise, a translation of $\mathbf{T}$ corresponds to *any* pair of parallel lines separated by $\mathbf{T}/2$. This circumstance yields a very elegant method for composing rotations and translations.

For example, see [30a]. Here a rotation about a through θ is being represented as $\Re_{L_2} \circ \Re_{L_1}$, and a rotation about b through ϕ is being represented as $\Re_{L_2'} \circ \Re_{L_1'}$. To find the net effect of rotating about a and then about b, choose $L_2 = L_1'$ to be the line through a and b. If $\theta + \phi \neq 2\pi$, then L_1 and L_2' will intersect at some point c, as in [30b]. Thus the composition of the two rotations is given by

$$(\Re_{L_2'} \circ \Re_{L_1'}) \circ (\Re_{L_2} \circ \Re_{L_1}) = \Re_{L_2'} \circ \Re_{L_1},$$

which is a rotation about c through $(\theta + \phi)$! That this construction agrees with our calculation on p. 18 is demonstrated in Ex. 36.

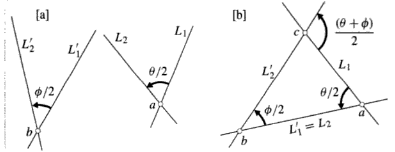


Figure [30]

Further examples of this method may be found in Ex. 42 and Ex. 43.

4 Similarities and Complex Arithmetic

Let us take a closer look at the role of distance in Euclidean geometry. Suppose we have two right triangles T and $\widetilde{T}$ drawn in the same plane, and suppose that Jack measures T while Jill measures $\widetilde{T}$. If Jack and Jill both report that their triangles have sides 3, 4, and 5, then it is tempting to say that the two triangles are the same, in the sense that there exists a motion $\mathcal{M}$ such that $\widetilde{T} = \mathcal{M}(T)$. But wait! Suppose that Jack's ruler is marked in centimetres, while Jill's is marked in inches. The two triangles are *similar*, but they are *not* congruent. Which is the "true" 3, 4, 5 triangle? Of course they both are.

The point is that *whenever we talk about distances numerically, we are presupposing a unit of measurement.* This may be pictured as a certain line-segment U, and when we say that some other segment has a length of 5, for example, we mean that precisely 5 copies of U can be fitted into it. But on our flat[15] plane any choice of U is as good as any other—there is no *absolute* unit of measurement, and our geometric theorems should reflect that fact.

Meditating on this, we recognize that Euclidean theorems do not in fact depend on this (arbitrary) choice of U, for they only deal with *ratios* of lengths, which are independent of U. For example, Jack can verify that his triangle T satisfies Pythagoras' Theorem in the form

$$(3\text{cm})^2 + (4\text{cm})^2 = (5\text{cm})^2,$$

but, dividing both sides by $(5\text{cm})^2$, this can be rewritten in terms of the ratios of the sides, which are pure numbers:

$$(3/5)^2 + (4/5)^2 = 1.$$

Try thinking of another theorem, and check that it too deals only with ratios of lengths.

[15]In the non-Euclidean geometries of Chapter 6 we will be drawing on *curved* surfaces, and the amount of curvature in the surface *will* dictate an absolute unit of length.

Since the theorems of Euclidean geometry do not concern themselves with the actual sizes of figures, our earlier definition of geometric equality in terms of motions is clearly too restrictive: two figures should be considered the same if they are *similar*. More precisely, we now consider two figures to be the same if there exists a *similarity* mapping one to the other, where

A similarity $\mathcal{S}$ is a mapping of the plane to itself that preserves ratios of distances.

It is easy to see [exercise] that a given similarity $\mathcal{S}$ expands every distance by the same (non-zero) factor r, which we will call the *expansion* of $\mathcal{S}$. We can therefore refine our notation by including the expansion as a superscript, so that a general similarity of expansion r is written $\mathcal{S}^r$. Clearly, the identity transformation is a similarity, $\mathcal{S}^k \circ \mathcal{S}^r = \mathcal{S}^{kr}$, and $(\mathcal{S}^r)^{-1} = \mathcal{S}^{(1/r)}$, so it is fairly clear that the set of all similarities forms a group. We thus arrive at the definition of Euclidean geometry that Klein gave in his Erlangen address:

Euclidean geometry is the study of those properties of geometric figures that are invariant under the group of similarities. (29)

Since the motions are just the similarities $\mathcal{S}^1$ of unit expansion, the group of motions is a subgroup of the group of similarities; our previous attempt at defining Euclidean geometry therefore yields a "subgeometry" of (29).

A simple example of an $\mathcal{S}^r$ is a *central dilation* $\mathcal{D}_o^r$. As illustrated in [31a], this leaves o fixed and radially stretches each segment oA by r. Note that the inverse of a central dilation is another central dilation with the same centre: $(\mathcal{D}_o^r)^{-1} = \mathcal{D}_o^{(1/r)}$. If this central dilation is followed by (or preceded by) a rotation $\mathcal{R}_o^\theta$ with the same centre, then we obtain the *dilative rotation*

$$\mathcal{D}_o^{r,\theta} \equiv \mathcal{R}_o^\theta \circ \mathcal{D}_o^r = \mathcal{D}_o^r \circ \mathcal{R}_o^\theta,$$

shown in [31b]. Note that a central dilation may be viewed as a special case of a dilative rotation: $\mathcal{D}_o^r = \mathcal{D}_o^{r,0}$.

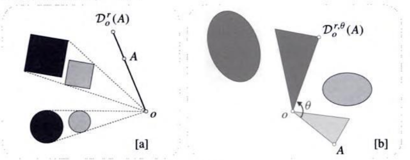


Figure [31]

This figure should be ringing loud bells. Taking o to be the origin of $\mathbb{C}$, (9) says that $\mathcal{D}_o^{r,\theta}$ corresponds to multiplication by $r\,e^{i\theta}$:

$$\mathcal{D}_o^{r,\theta}(z) = \left(r\,e^{i\theta}\right) z.$$

Conversely, and this is the key point, *the rule for complex multiplication may be viewed as a consequence of the behaviour of dilative rotations.*

Concentrate on the set of dilative rotations with a common, fixed centre o, which will be thought of as the origin of the complex plane. Each $\mathcal{D}_o^{r,\theta}$ is uniquely determined by its expansion r and rotation θ, and so it can be represented by a vector of length r at angle θ. Likewise, $\mathcal{D}_o^{R,\phi}$ can be represented by a vector of length R at angle ϕ. What vector will represent the composition of these dilative rotations? Geometrically it is clear that

$$\mathcal{D}_o^{R,\phi} \circ \mathcal{D}_o^{r,\theta} = \mathcal{D}_o^{r,\theta} \circ \mathcal{D}_o^{R,\phi} = \mathcal{D}_o^{Rr,(\theta+\phi)},$$

so the new vector is obtained from the original vectors by multiplying their lengths and adding their angles—complex multiplication!

On page 17 we saw that if complex numbers are viewed as translations then composition yields complex addition. We now see that if they are instead viewed as dilative rotations then composition yields complex multiplication. To complete our "explanation" of complex numbers in terms of geometry, we will show that these translations and dilative rotations are fundamental to Euclidean geometry as defined in (29).

To understand the general similarity $\mathcal{S}^r$ involved in (29), note that if p is an arbitrary point, $\mathcal{M} \equiv \mathcal{S}^r \circ \mathcal{D}_p^{(1/r)}$ is a *motion*. Thus *any similarity is the composition of a dilation and a motion:*

$$\mathcal{S}^r = \mathcal{M} \circ \mathcal{D}_p^r. \tag{30}$$

Our classification of motions therefore implies that similarities come in two kinds: if $\mathcal{M}$ preserves angles then so will $\mathcal{S}^r$ [a *direct similarity*]; if $\mathcal{M}$ reverses angles then so will $\mathcal{S}^r$ [an *opposite similarity*].

Just as we concentrated on the group of direct motions, so we will now concentrate on the group of direct similarities. The fundamental role of translations and dilative rotations in Euclidean geometry finally emerges in the following surprising theorem:

Every direct similarity is a dilative rotation or (exceptionally) a translation. (31)

For us this fact constitutes one satisfying "explanation" of complex numbers; as mentioned in the Preface, other equally compelling explanations may be found in the laws of physics.

To begin to understand (31), observe that (25) and (30) imply that a direct similarity is determined by the image $A'B'$ of any line-segment AB. First consider the exceptional case in which $A'B'$ are of equal length AB. We then have the three

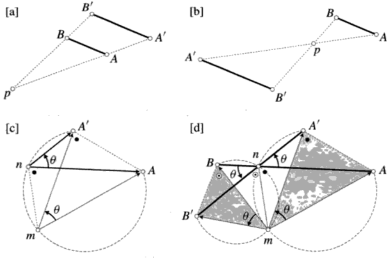


Figure [32]

cases in [28], all of which are consistent with (31). If $A'B'$ and AB are parallel but not of equal length, then we have the two cases shown in [32a] and [32b], in both of which we have drawn the lines AA' and BB' intersecting in p. By appealing to the similar triangles in these figures, we see that in [32a] the similarity is $\mathcal{D}_p^{r,0}$, while in [32b] it is $\mathcal{D}_p^{r,\pi}$, where in both cases $r = (pA'/pA) = (pB'/pB)$.

Now consider the much more interesting general case where $A'B'$ and AB are neither the same length, nor parallel. Take a peek at [32d], which illustrates this. Here n is the intersection point of the two segments (produced if necessary), and θ is the angle between them. To establish (31), we must show that we can carry AB to $A'B'$ with a single dilative rotation. For the time being, simply note that if AB is to end up having the same direction as $A'B'$ then it must be rotated by θ, so the claim is really this: *There exists a point q, and an expansion factor r, such that $\mathcal{D}_q^{r,\theta}$ carries A to A' and B to B'.*

Consider the part of [32d] that is reproduced in [32c]. Clearly, by choosing $r = (nA'/nA)$, $\mathcal{D}_n^{r,\theta}$ will map A to A'. More generally, you see that we can map A to A' with $\mathcal{D}_q^{r,\theta}$ if and only if AA' subtends angle θ at q. Thus, with the appropriate value of r, *$\mathcal{D}_q^{r,\theta}$ maps A to A' if and only if q lies on the circular arc AnA'.* The figure illustrates one such position, $q = m$. Before returning to [32d], we need to notice one more thing: mA subtends the same angle (marked •) at n and A'.

Let us return to [32d]. We want $\mathcal{D}_q^{r,\theta}$ to map A to A' and B to B'. According to the argument above, q must lie on the circular arc AnA' *and* on the circular arc BnB'. Thus there are just two possibilities: $q = n$ or $q = m$ (the other intersection point of the two arcs). If you think about it, this is a moment of high drama. We have narrowed down the possibilities for q to just two points by consideration of angles *alone*; for either of these two points we can choose the value of the expansion r so as to make A go to A', but, once this choice has been made, either B will map to B' or it won't! Furthermore, it is clear from the figure that if $q = n$ then B does *not* map to B', so $q = m$ is the only possibility left.

In order for $\mathcal{D}_m^{r,\theta}$ to simultaneously map A to A' and B to B', we need to have $r = (mA'/mA) = (mB'/mB)$; in other words, the two shaded triangles need to be similar. That they are indeed similar is surely something of a miracle. Looking at the angles formed at n, we see that $\theta + \odot + \bullet = \pi$, and the result follows immediately by thinking of the RHS as the angle-sum of each of the two shaded triangles. This completes our proof[16] of (31).

The reader may feel that it is unsatisfactory that (31) calls for dilative rotations about arbitrary points, while complex numbers represent dilative rotations about a *fixed* point o (the origin). This may be answered by noting that the images of AB under $\mathcal{D}_q^{r,\theta}$ and $\mathcal{D}_o^{r,\theta}$ will be parallel and of equal length, so there will exist a translation [see Ex. 44 for details] $\mathcal{T}_v$ mapping one onto the other. In other words, a general dilative rotation differs from an origin-centred dilative rotation by a mere translation: $\mathcal{D}_q^{r,\theta} = \mathcal{T}_v \circ \mathcal{D}_o^{r,\theta}$. To sum up,

> *Every direct similarity $\mathcal{S}^r$ can be expressed as a complex function of the form $\mathcal{S}^r(z) = re^{i\theta}z + v$.*

5 Spatial Complex Numbers?

Let us briefly attempt to generalize the above ideas to *three*-dimensional space. Firstly, a central dilation of space (centred at O) is defined exactly as before, and a dilative rotation with the same centre is then the composition of such a dilation with a rotation of space about an axis passing through O. Once again taking (29) as the definition of Euclidean geometry, we get off to a flying start, because the key result (31) generalizes: *Every direct similarity of space is a dilative rotation, a translation, or the composition of a dilative rotation and a translation along its rotation axis.* See Coxeter [1969, p. 103] for details.

It is therefore natural to ask if there might exist "spatial complex numbers" for which addition would be composition of translations, and for which multiplication would be composition of dilative rotations. With addition all goes well: the position vector of each point in space may be viewed as a translation, and composition of these translations yields ordinary vector addition in space. Note that this vector addition makes equally good sense in *four*-dimensional space, or n-dimensional space for that matter.

Now consider the set Q of dilative rotations with a common, fixed centre O. Initially, the definition of multiplication goes smoothly, for the "product" $Q_1 \circ Q_2$ of two such dilative rotations is easily seen to be another dilative rotation (Q_3, say) of the same kind. This follows from the above classification of direct similarities by noting that $Q_1 \circ Q_2$ leaves O fixed. If the expansions of Q_1 and Q_2 are r_1 and

[16] The present argument has the advantage of proceeding in steps, rather than having to be discovered all at once. For other proofs, see Coxeter and Greitzer [1967, p. 97], Coxeter [1969, p. 73], and Eves [1992, p. 71]. Also, see Ex. 45 for a simple proof using complex functions.

r_2 then the expansion of Q_3 is clearly $r_3 = r_1 r_2$, and in Chapter 6 we shall give a simple geometric construction for the rotation of Q_3 from the rotations of Q_1 and Q_2. However, unlike rotations in the plane, it makes a difference in what order we perform two rotations in space, so our multiplication rule is *not commutative*:

$$Q_1 \circ Q_2 \neq Q_2 \circ Q_1. \tag{32}$$

We are certainly *accustomed* to multiplication being commutative, but there is nothing inconsistent about (32), so this cannot be considered a decisive obstacle to an algebra of "spatial complex numbers".

However, a fundamental problem does arise when we try to represent these dilative rotations as points (or vectors) in space. By analogy with complex multiplication, we wish to interpret the equation $Q_1 \circ Q_2 = Q_3$ as saying that the dilative rotation Q_1 maps the point Q_2 to the point Q_3. But this interpretation is impossible! The specification of a point in space requires *three* numbers, but the specification of a dilative rotation requires *four:* one for the expansion, one for the angle of rotation, and two[17] for the direction of the axis of the rotation.

Although we have failed to find a three-dimensional analogue of complex numbers, we have discovered the *four*-dimensional space Q of dilative rotations (centred at O) of three-dimensional space. Members of Q are called *quaternions*, and they may be pictured as points or vectors in four dimensions, but the details of how to do this will have to wait till Chapter 6. Quaternions can be added by ordinary vector addition, and they can be multiplied using the non-commutative rule above (composition of the corresponding dilative rotations).

The discoveries of the rules for multiplying complex numbers and for multiplying quaternions have some interesting parallels. As is well known, the quaternion rule was discovered in *algebraic* form by Sir William Rowan Hamilton in 1843. It is less well known that three years earlier Olinde Rodrigues had published an elegant geometric investigation of the composition of rotations in space that contained essentially the same result; only much later[18] was it recognized that Rodrigues' geometry was equivalent to Hamilton's algebra.

Hamilton and Rodrigues are just two examples of hapless mathematicians who would have been dismayed to examine the unpublished notebooks of the great Karl Friedrich Gauss. There, like just another log entry in the chronicle of his private mathematical voyages, Gauss recorded his discovery of the quaternion rule in 1819.

In Chapter 6 we shall investigate quaternion multiplication in detail and find that it has elegant applications. However, the immediate benefit of this discussion is that we can now see what a remarkable property it is of *two*-dimensional space that it is possible to interpret points *within it* as the fundamental Euclidean transformations *acting on it*.

[17]To see this, imagine a sphere centred at O. The direction of the axis can be specified by its intersection with the sphere, and this point can be specified with two coordinates, e.g., longitude and latitude.

[18]See Altmann [1989] for the intriguing details of how this was unravelled.

V Exercises

1 The roots of a general cubic equation in X may be viewed (in the XY-plane) as the intersections of the X-axis with the graph of a cubic of the form,

$$Y = X^3 + AX^2 + BX + C.$$

(i) Show that the point of inflection of the graph occurs at $X = -\frac{A}{3}$.

(ii) Deduce (geometrically) that the substitution $X = \left(x - \frac{A}{3}\right)$ will reduce the above equation to the form $Y = x^3 + bx + c$.

(iii) Verify this by calculation.

2 In order to solve the cubic equation $x^3 = 3px + 2q$, do the following:

(i) Make the inspired substitution $x = s + t$, and deduce that x solves the cubic if $st = p$ and $s^3 + t^3 = 2q$.

(ii) Eliminate t between these two equations, thereby obtaining a quadratic equation in s^3.

(iii) Solve this quadratic to obtain the two possible values of s^3. By symmetry, what are the possible values of t^3?

(iv) Given that we know that $s^3 + t^3 = 2q$, deduce the formula (4).

3 In 1591, more than forty years after the appearance of (4), François Viète published another method of solving cubics. The method is based on the identity (see p. 15) $\cos 3\theta = 4C^3 - 3C$, where $C = \cos\theta$.

(i) Substitute $x = 2\sqrt{p}\,C$ into the (reduced) general cubic $x^3 = 3px + 2q$ to obtain $4C^3 - 3C = \frac{q}{p\sqrt{p}}$.

(ii) Provided that $q^2 \le p^3$, deduce that the solutions of the original equation are

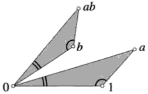
$$x = 2\sqrt{p}\cos\left[\tfrac{1}{3}(\phi + 2m\pi)\right],$$

where m is an integer and $\phi = \cos^{-1}(q/p\sqrt{p})$.

(iii) Check that this formula gives the correct solutions of $x^3 = 3x$, namely, $x = 0, \pm\sqrt{3}$.

4 Here is a basic fact about integers that has many uses in number theory: *If two integers can be expressed as the sum of two squares, then so can their product.* With the understanding that each symbol denotes an integer, this says that if $M = a^2 + b^2$ and $N = c^2 + d^2$, then $MN = p^2 + q^2$. Prove this result by considering $|(a + ib)(c + id)|^2$.

5 The figure below shows how two similar triangles may be used to construct the product of two complex numbers. Explain this.



6 (i) If c is a fixed complex number, and R is a fixed real number, explain with a picture why $|z - c| = R$ is the equation of a circle.

(ii) Given that z satisfies the equation $|z + 3 - 4i| = 2$, find the minimum and maximum values of $|z|$, and the corresponding positions of z.

7 Use a picture to show that if a and b are fixed complex numbers then $|z - a| = |z - b|$ is the equation of a line.

8 Let L be a straight line in $\mathbb{C}$ making an angle ϕ with the real axis, and let d be its distance from the origin. Show geometrically that if z is any point on L then

$$d = \left|\operatorname{Im}[e^{-i\phi}\, z]\right|.$$

[*Hint:* Interpret $e^{-i\phi}$ using (9).]

9 Let A, B, C, D be four points on the unit circle. If $A + B + C + D = 0$, show that the points must form a rectangle.

10 Show geometrically that if $|z| = 1$ then

$$\operatorname{Im}\left[\frac{z}{(z+1)^2}\right] = 0.$$

Apart from the unit circle, what other points satisfy this equation?

11 Explain geometrically why the locus of z such that

$$\arg\left(\frac{z-a}{z-b}\right) = const.$$

is an arc of a certain circle passing through the fixed points a and b.

12 By using pictures, find the locus of z for each of the following equations:

$$\operatorname{Re}\left(\frac{z-1-i}{z+1+i}\right) = 0, \quad \text{and} \quad \operatorname{Im}\left(\frac{z-1-i}{z+1+i}\right) = 0.$$

[*Hints:* What does $\operatorname{Re}(W) = 0$ imply about the angle of W? Now use the previous exercise.]

13 Find the geometric configuration of the points a, b, and c if

$$\left(\frac{b-a}{c-a}\right) = \left(\frac{a-c}{b-c}\right).$$

[*Hint:* Separately equate the lengths and angles of the two sides.]

14 By considering the product $(2+i)(3+i)$, show that

$$\frac{\pi}{4} = \tan^{-1}\frac{1}{2} + \tan^{-1}\frac{1}{3}.$$

15 Draw $e^{i\pi/4}$, $e^{i\pi/2}$, and their sum. By expressing each of these numbers in the form $(x+iy)$, deduce that

$$\tan\frac{3\pi}{8} = 1+\sqrt{2}.$$

16 Starting from the origin, go one unit east, then the same length north, then $(1/2)$ of the previous length west, then $(1/3)$ of the previous length south, then $(1/4)$ of the previous length east, and so on. What point does this "spiral" converge to?

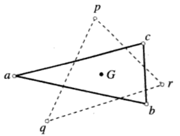
17 If $z = e^{i\theta} \neq -1$, then $(z-1) = (i\ \tan\frac{\theta}{2})(z+1)$. Prove this (i) by calculation, (ii) with a picture.

18 Prove that

$$e^{i\theta} + e^{i\phi} = 2\cos\left[\tfrac{\theta-\phi}{2}\right] e^{\frac{i(\theta+\phi)}{2}} \quad \text{and} \quad e^{i\theta} - e^{i\phi} = 2i\sin\left[\tfrac{\theta-\phi}{2}\right] e^{\frac{i(\theta+\phi)}{2}}$$

(i) by calculation, and (ii) with a picture.

19 The "centroid" G of a triangle T is the intersection of its medians. If the vertices are the complex numbers a, b, and c, then you may assume that



$$G = \tfrac{1}{3}(a+b+c).$$

On the sides of T we have constructed three *similar* triangles [dotted] of arbitrary shape, so producing a new triangle [dashed] with vertices p, q, r. Using complex algebra, show that the centroid of the new triangle is in exactly the same place as the centroid of the old triangle!

20 *Gaussian integers* are complex numbers of the form $m + in$, where m and n are integers—they are the grid points in [1]. Show that it is impossible to draw an equilateral triangle such that all three vertices are Gaussian integers. [*Hints*: You may assume that one of the vertices is at the origin; try a proof by contradiction; if a triangle is equilateral, you can rotate one side into another; remember that $\sqrt{3}$ is irrational.]

21 Make a copy of [12a], draw in the diagonal of the quadrilateral shown in [12b], and mark its midpoint m. As in [12b], draw the line-segments connecting m to p, q, r, and s. According to the result in [12b], what happens to p and to r under a rotation of $(\pi/2)$ about m? So what happens to the line-segment pr? Deduce the result shown in [12a].

22 Will the result in [12a] survive if the squares are instead constructed on the inside of the quadrilateral?

23 Draw an arbitrary triangle, and on each side draw an equilateral triangle lying outside the given triangle. What do you suspect is special about the new triangle formed by joining the centroids (cf. Ex. 19) of the equilateral triangles? Use complex algebra to prove that you are right. What happens if the equilateral triangles are instead drawn on the inside of the given triangle?

24 From (15), we know that

$$1 + z + z^2 + \cdots + z^{n-1} = \frac{z^n - 1}{z - 1}.$$

(i) In what region of $\mathbb{C}$ must z lie in order that the *infinite* series $1 + z + z^2 + \cdots$ converges?

(ii) If z lies in this region, to which point in the plane does the infinite series converge?

(iii) In the spirit of figure [9], draw a large, accurate picture of the infinite series in the case $z = \frac{1}{2}(1 + i)$, and check that it does indeed converge to the point predicted by part (ii).

25 Let $S = \cos\theta + \cos 3\theta + \cos 5\theta + \cdots + \cos(2n - 1)\theta$. Show that

$$S = \frac{\sin 2n\theta}{2\sin\theta} \quad \text{or equivalently} \quad S = \frac{\sin n\theta \, \cos n\theta}{\sin\theta}.$$

[*Hint:* Use Ex. 24, then Ex. 18 to simplify the result.]

26 (i) By considering $(a+ib)(\cos\theta + i\sin\theta)$, show that

$$b\cos\theta + a\sin\theta = \sqrt{a^2+b^2}\,\sin\left[\theta + \tan^{-1}(b/a)\right].$$

(ii) Use this result to prove (14) by the method of induction.

27 Show that the polar equation of the spiral $Z(t) = e^{at}e^{ibt}$ in [15b] is $r = e^{(a/b)\theta}$.

28 Reconsider the spiral $Z(t) = e^{at}e^{ibt}$ in [15b], where a and b are fixed real numbers. Let τ be a variable real number. According to (9), $z \mapsto \mathcal{F}_\tau(z) = \left(e^{a\tau}e^{ib\tau}\right) z$ is an expansion of the plane by factor $e^{a\tau}$, combined with a rotation of the plane through angle $b\tau$.

(i) Show that $\mathcal{F}_\tau[Z(t)] = Z(t+\tau)$, and deduce that the spiral is an *invariant curve* (cf. p. 38) of the transformations $\mathcal{F}_\tau$.

(ii) Use this to give a calculus-free demonstration that all rays from the origin cut the spiral at the same angle.

(iii) Show that if the spiral is rotated about the origin through an arbitrary angle, the new spiral is again an invariant curve of each $\mathcal{F}_\tau$.

(iv) Argue that the spirals in the previous part are the *only* invariant curves of $\mathcal{F}_\tau$.

29 (i) If $V(t)$ is the complex velocity of a particle whose orbit is $Z(t)$, and dt is an infinitesimal moment of time, then $V(t)\,dt$ is a complex number along the orbit. Thinking of the integral as the (vector) sum of these movements, what is the geometric interpretation of $\int_{t_1}^{t_2} V(t)\,dt$?

(ii) Referring to [15b], sketch the curve $Z(t) = \frac{1}{a+ib}\, e^{at}e^{ibt}$.

(iii) Given the result (13), what is the velocity of the particle in the previous part.

(iv) Combine the previous parts to deduce that $\int_0^1 e^{at}e^{ibt}\,dt = \left[\frac{1}{a+ib}\, e^{at}e^{ibt}\right]_0^1$, and draw in this complex number in your sketch for part (ii).

(v) Use this to deduce that

$$\int_0^1 e^{at}\cos bt\,dt = \frac{a\,(e^a\cos b - 1) + b\,e^a\sin b}{a^2+b^2},$$

and

$$\int_0^1 e^{at}\sin bt\,dt = \frac{b\,(1 - e^a\cos b) + a\,e^a\sin b}{a^2+b^2}.$$

30 Given two starting numbers S_1, S_2, let us build up an infinite sequence S_1, S_2, S_3, S_4, ... with this rule: *each new number is twice the difference of the previous two*. For example, if $S_1 = 1$ and $S_2 = 4$, we obtain $1, 4, 6, 4, -4, -16, -24, \ldots$. Our aim is to find a formula for the n^{th} number S_n.

(i) Our generating rule can be written succinctly as $S_{n+2} = 2(S_{n+1} - S_n)$. Show that $S_n = z^n$ will solve this *recurrence relation* if $z^2 - 2z + 2 = 0$.

(ii) Use the quadratic formula to obtain $z = 1 \pm i$, and show that if A and B are arbitrary complex numbers, $S_n = A(1+i)^n + B(1-i)^n$ is a solution of the recurrence relation.

(iii) If we want only real solutions of the recurrence relation, show that $B = \overline{A}$, and deduce that $S_n = 2\,\mathrm{Re}[A(1+i)^n]$.

(iv) Show that for the above example $A = -(1/2) - i$, and by writing this in polar form deduce that $S_n = 2^{n/2}\sqrt{5}\cos\left[\frac{(n+4)\pi}{4} + \tan^{-1} 2\right]$.

(v) Check that this formula predicts $S_{34} = 262144$, and use a computer to verify this.

[Note that this method can be applied to any recurrence relation of the form $S_{n+2} = pS_{n+1} + qS_n$.]

31 With the same recurrence relation as in the previous exercise, use a computer to generate the first 30 members of the sequence given by $S_1 = 2$ and $S_2 = 4$. Note the repeating pattern of zeros.

(i) With the same notation as before, show that this sequence corresponds to $A = -i$, so that $S_n = 2\,\mathrm{Re}[-i(1+i)^n]$.

(ii) Draw a sketch showing the locations of $-i(1+i)^n$ for $n = 1$ to $n = 8$, and hence explain the pattern of zeros.

(iii) Writing $A = a + ib$, our example corresponds to $a = 0$. More generally, explain geometrically why such a repeating pattern of zeros will occur if and only if $(a/b) = 0, \pm 1$ or $b = 0$.

(iv) Show that $\frac{S_1}{S_2} = \frac{1}{2}\left[1 - \frac{a}{b}\right]$, and deduce that a repeating pattern of zeros will occur if and only if $S_2 = 2S_1$ (as in our example), $S_1 = S_2$, $S_1 = 0$, or $S_2 = 0$.

(v) Use a computer to verify these predictions.

32 The Binomial Theorem says that if n is a positive integer,

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r, \quad \text{where} \quad \binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

are the binomial coefficients [*not vectors!*]. The algebraic reasoning leading to this result is equally valid if a and b are *complex* numbers. Use this fact to show that if $n = 2m$ is even then

$$\binom{2m}{1} - \binom{2m}{3} + \binom{2m}{5} - \cdots + (-1)^{m+1}\binom{2m}{2m-1} = 2^m \sin\left(\tfrac{m\pi}{2}\right).$$

33 Consider the equation $(z-1)^{10} = z^{10}$.

(i) Without attempting to solve the equation, show geometrically that all 9 solutions [why not 10?] must lie on the vertical line, $\text{Re}(z) = \frac{1}{2}$. [*Hint:* Ex. 7.]

(ii) Dividing both sides by z^{10}, the equation takes the form $w^{10} = 1$, where $w = (z-1)/z$. Hence solve the original equation.

(iii) Express these solutions in the form $z = x + iy$, and thereby verify the result in (i). [*Hint:* To do this neatly, use Ex. 18.]

34 Let S denote the set of 12^{th} roots of unity shown in [19], one of which is $\xi = e^{i(\pi/6)}$. Note that ξ is a *primitive* 12^{th} root of unity, meaning that its powers yield *all* the 12^{th} roots of unity: $S = \{\xi, \xi^2, \xi^3, \ldots, \xi^{12}\}$.

(i) Find all the primitive 12^{th} roots of unity, and mark them on a copy of [19].

(ii) Write down, in the form of (16), the factorization of the polynomial $\Phi_{12}(z)$ whose roots are the primitive 12^{th} roots of unity. [In general, $\Phi_n(z)$ is the polynomial (with the coefficient of the highest power of z equal to 1) whose roots are the primitive n^{th} roots of unity; it is called the n^{th} *cyclotomic polynomial*.]

(iii) By first multiplying out pairs of factors corresponding to conjugate roots, show that $\Phi_{12}(z) = z^4 - z^2 + 1$.

(iv) By repeating the above steps, show that $\Phi_8(z) = z^4 + 1$.

(v) For a general value of n, explain the fact that if ζ is a primitive n^{th} root of unity, then so is $\overline{\zeta}$. Deduce that if $n > 2$ then $\Phi_n(z)$ always has even degree and real coefficients.

(vi) Show that if p is a prime number then $\Phi_p(z) = 1 + z + z^2 + \cdots + z^{p-1}$. [*Hint:* Ex. 24.]

[In these examples it is striking that $\Phi_n(z)$ has integer coefficients. In fact it can be shown that this is true for *every* $\Phi_n(z)$! For more on these fascinating polynomials, see Stillwell [1994].]

35 Show algebraically that the formula (21) is invariant under a translation by k, i.e., its value does not change if a becomes $a+k$, b becomes $b+k$, etc. Deduce from [22a] that the formula always gives the area of the quadrilateral. [*Hint:* Remember, $(z + \overline{z})$ is always real.]

36 According to the calculation on p. 18, $\mathcal{R}_b^{\phi} \circ \mathcal{R}_a^{\theta} = \mathcal{R}_c^{(\theta+\phi)}$, where

$$c = \frac{ae^{i\phi}(1 - e^{i\theta}) + b(1 - e^{i\phi})}{1 - e^{i(\theta+\phi)}}.$$

Let us check that this c is the same as the one given by the geometric construction in [30b].

(i) Explain why the geometric construction is equivalent to saying that c satisfies the two conditions

$$\arg\left[\frac{c-b}{a-b}\right] = \tfrac{1}{2}\phi \quad \text{and} \quad \arg\left[\frac{c-a}{b-a}\right] = -\tfrac{1}{2}\theta.$$

(ii) Verify that the calculated value of c (given above) satisfies the first of these conditions by showing that

$$\frac{c-b}{a-b} = \left[\frac{\sin\frac{\theta}{2}}{\sin\frac{(\theta+\phi)}{2}}\right] e^{i\phi/2}. \tag{33}$$

[*Hint*: Use $\left(1 - e^{i\alpha}\right) = -2i\,\sin(\alpha/2)\,e^{i\alpha/2}$.]

(iii) In the same way, verify that the second condition is also satisfied.

37 Deduce (33) directly from [30b]. [*Hint*: Draw in the altitude through b of the triangle abc, and express its length first in terms of $\sin\frac{\theta}{2}$, then in terms of $\sin\frac{(\theta+\phi)}{2}$.]

38 On page 18 we calculated that for any non-zero α, $\mathcal{T}_v \circ \mathcal{R}_0^{\alpha}$ is a rotation:

$$\mathcal{T}_v \circ \mathcal{R}_0^{\alpha} = \mathcal{R}_c^{\alpha}, \quad \text{where} \quad c = v/(1 - e^{i\alpha}).$$

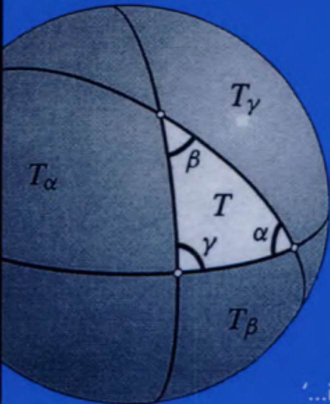
However, if $\alpha = 0$ then $\mathcal{T}_v \circ \mathcal{R}_0^{\alpha} = \mathcal{T}_v$ is a translation. Try to reconcile these facts by considering the behaviour of $\mathcal{R}_c^{\alpha}$ in the limit that α tends to zero.

39 A *glide reflection* is the composition $\mathcal{T}_v \circ \Re_L = \Re_L \circ \mathcal{T}_v$ of reflection in a line L and a translation v in the direction of L. For example, if you walk at a steady pace in the snow, your tracks can be obtained by repeatedly applying the same glide reflection to a single footprint. Clearly, a glide reflection is an opposite motion.

(i) Draw a line L, a line-segment AB, the image $\widetilde{A}\widetilde{B}$ of the segment under $\Re_L$, and the image $A'B'$ of AB under the glide reflection $\mathcal{T}_v \circ \Re_L$.

(ii) Suppose you erased L from your picture; by considering the line-segments AA' and BB', show that you can reconstruct L.

(iii) Given any two segments AB and $A'B'$ of equal length, use the previous part to construct the glide reflection that maps the former to the latter.



'*Visual Complex Analysis* is a delight, and a book after my own heart. By his innovative and exclusive use of the geometrical perspective, Tristan Needham uncovers many surprising and largely unappreciated aspects of the beauty of complex analysis.'

Sir Roger Penrose

'...it is comparable with Feynman's *Lectures on Physics*. At every point it asks 'why' and finds a beautiful visual answer.'

Newsletter of the European Mathematical Society

'If your budget limits you to buying only one mathematics book in a year then make sure that this is the one that you buy.'

Mathematical Gazette

'It is...a special pleasure to see Tristan Needham's *Visual Complex Analysis* with its elegantly illustrated visual approach.'

Ian Stewart, ***New Scientist***

'A truly unusual and notably creative look at a classical subject.'

Paul Zorn, ***American Mathematical Monthly***

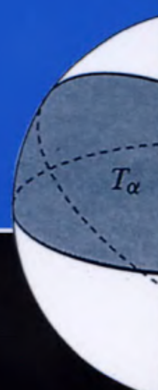
'This informal style is excellently judged and works extremely well...Many of the arguments presented will be new even to experts, and the book will be of great interest to professionals working in either complex analysis or in some field where complex analysis is used.'

David Armitage, ***Mathematical Reviews***

'Newton would have approved...a fascinating and refreshing look at a familiar subject...It is essential reading for anybody with any interest at all in this absorbing area of mathematics.'

Times Higher Education Supplement

In 1995 the MAA presented the author with the Carl B. Allendoerfer Award for part of his work on this book.



OXFORD
UNIVERSITY PRESS

ISBN 0-19-853446-9

9 780198 534464