The Complex Plane and The Euclidean Plane

Algebraic Approach to Plane Geometry

Michael Taylor

Math. Dept., UNC

E-mail address: met@math.unc.edu

2010 *Mathematics Subject Classification.* 22-01

Key words and phrases. complex numbers, Euclidean geometry, Pythagorean theorem, inner products, isometries, triangles, angle measurements, congruence, similarity, squares, rectangles, polygons, circles, calculus, curves, arclength, exponential and trigonometric functions, Euler formula, real measurement of angles, π, area, Euclidean numbers, linear fractional transformations, fundamental theorem of algebra, planetary orbits, ellipse, parabola, hyperbola, inner product spaces, volume, area of spheres

Contents

Preface

The set $\mathbb{C}$ of complex numbers arose as an extension of the set $\mathbb{R}$ of real numbers in which one could solve the equation $z^2 = -1$. This was obtained by adding an imaginary unit i, satisfying $i^2 = -1$. A complex number $z \in \mathbb{C}$ has the form

$$z = x + iy, \quad x, y \in \mathbb{R}.$$

We say x is the real part of z and y is the imaginary part. Algebraic operations of addition and multiplication on $\mathbb{R}$ extend to $\mathbb{C}$ in a natural fashion, as we detail in §1.1.

People working with this enlarged number system realized that it is convenient to visualize $\mathbb{C}$ as the *complex plane*, and to utilize the insights of Euclidean geometry. Turning this around, one can perceive that the algebraic structure of $\mathbb{C}$ provides an ideal set of tools with which to obtain basic geometric results set out in the books of Euclid. We describe how this works here.

This treatment facilitates a transition between what students learn about Euclidean geometry in high school and what is needed for more advanced studies. It should be useful to students who have completed a three-semester calculus sequence, and have taken or are currently taking an analysis course that gives a rigorous treatment of the real numbers, the fundamental theorems of calculus, and concepts involving infinite sequences and series of functions.

Our goal in Chapter 1 is to show that the use of the complex plane $\mathbb{C}$, with its algebraic structure, provides a very pleasant way to develop plane geometry. We have a notion of distance between two points in $\mathbb{C}$, given by $d(z, w) = |z - w|$, leading to the notion of an isometry, $F : \mathbb{C} \to \mathbb{C}$,

preserving the distance, and having a simple algebraic characterization. We proceed to lines, the notion of parallelism and orthogonality, aided by the special formula

$$|z+w|^2 = |z|^2 + |w|^2 + 2\langle z, w\rangle,$$

involving the inner product, $\langle z, w\rangle = \operatorname{Re} z\overline{w}$. This in turn leads to the Pythagorean theorem, the cornerstone of Euclidean geometry, and we are on our way, with treatments of triangles, congruences, polygons, and circles.

In Chapter 2 the text makes contact with calculus, applies this to a study of arclength, and develops trigonometry in a logical fashion, emphasizing Euler's formula as a key to that subject. We show how this throws light on classical results, including computations of Archimedes and Ptolemy of perimeters and areas of regular polygons, approximations to π, and trigonometric tables. Other topics include results on elements of $\mathbb{C}$ constructible via compass and straightedge, linear fractional transformations and their application to a three-circle problem of Apollonius, and Argand's proof of the fundamental theorem of algebra. We include an approach using the complex plane to the planetary orbit problem, whose orbits are given by various conic sections. We close with a brief description of Euclidean geometry in higher dimensions. A more detailed sketch of the material we cover is given in the chapter introductions.

Some basic notation

This list of some basic notation will be used throughout the text.

$\mathbb{C}$ is the set of complex numbers.

$\mathbb{R}$ is the set of real numbers.

$\mathbb{Z}$ is the set of integers.

$\mathbb{Z}^+$ is the set of integers ≥ 0.

$\mathbb{N}$ is the set of integers ≥ 1 (the natural numbers).

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

$z \in \mathbb{C}$ means z is an element of $\mathbb{C}$.

$|z|$ denotes the absolute value of z.

$\{z \in \mathbb{C} : |z| \leq 1\}$ denotes the set of $z \in \mathbb{C}$ such that $|z| \leq 1$.

$f : A \to B$ means that the function f maps points in the set A to points in B. One also says f maps A to B.

Chapter 1

Basic algebraic results on Euclidean geometry

This first chapter introduces the complex plane $\mathbb{C}$ and shows how use of its structure leads to nice proofs of basic results of Euclidean plane geometry. Section 1.1 presents the algebraic operations of addition and multiplication on $\mathbb{C}$, introduces complex conjugation and the absolute value $|z|$ of a complex number z, and derives a formula for division.

Section 1.2 discusses metric properties of $\mathbb{C}$, equipped with the distance function $d(z, w) = |z - w|$. A key tool is the identity

$$|z + w|^2 = |z|^2 + |w|^2 + 2\langle z, w\rangle, \tag{1.0.1}$$

bringing in the inner product $\langle z, w\rangle = \operatorname{Re} z\overline{w}$. As we will see, this contains the essence of the Pythagorean theorem.

Section 1.3 introduces the class $\mathrm{Isom}(\mathbb{C})$ of maps $F : \mathbb{C} \to \mathbb{C}$ that preserve distance, and gives this class an algebraic characterization. We show that $\mathrm{Isom}(\mathbb{C}) = \mathrm{Isom}^+(\mathbb{C}) \cup \mathrm{Isom}^-(\mathbb{C})$, whose elements have the form

$$F(z) = az + b, \quad a\overline{z} + b, \quad a \in S^1,\ b \in \mathbb{C}, \tag{1.0.2}$$

respectively, where $S^1 = \{z \in \mathbb{C} : |z| = 1\}$.

Section 1.4 introduces the concepts of lines and line segments, perpendicular lines, and parallel lines. We show that if L is a line, $p \in \mathbb{C}$, then through p there is a unique line perpendicular to L, and, if $p \notin L$, a unique line parallel to L.

In §1.5 we discuss rays and wedges, and define the angle measurement of a wedge as an element ω of the unit circle S^1. A related concept, measuring

angles in "radians," requires more work, and will appear in §2.2, with a study of arclength.

In §1.6 we consider triangles, $\mathcal{T} = \triangle(A, B, C)$, with vertices $A, B, C \in \mathbb{C}$. We define angle measurements at each vertex, e.g., $\Omega_{\mathcal{T}}(A) \in S^1$, and show that

$$\Omega_{\mathcal{T}}(A)\Omega_{\mathcal{T}}(B)\Omega_{\mathcal{T}}(C) = -1. \tag{1.0.3}$$

We also formally state the Pythagorean theorem, and derive it from (1.0.1).

In §1.7 we study congruence of triangles, which occurs if $\mathcal{T}' = F(\mathcal{T})$ for some $F \in \mathrm{Isom}(\mathbb{C})$. We establish various classical criteria for two triangles to be congruent, stated in terms involving side lengths and angle measurements. We also characterize which triples of positive numbers can be sidelengths of a triangle.

Section 1.8 considers the family of isosceles triangles and the subfamily of equilateral triangles. We see equivalent characterizations in terms of sidelengths and in terms of angle measurements.

Section 1.9 deals with similarity, arising via maps $F : \mathbb{C} \to \mathbb{C}$ whose application scales distances by some fixed factor. We see that two triangles with identical angle measurements are similar. We display three similar triangles arising from a right triangle, divided by a perpendicular dropped from the right angle to the opposite side, and relate this to a proof of the Pythagorean theorem given in Book VI of Euclid.

Section 1.10 studies squares, rectangles, and more general figures, including parallelograms, as well as other quadrilaterals (4-gons), and further figures with n vertices (n-gons).

Section 1.11 introduces circles, of the form

$$S_R(a) = \{z \in \mathbb{C} : |z - a| = R\}, \tag{1.0.4}$$

given $R > 0$, $a \in \mathbb{C}$. The case $S_1(0)$ is S^1, encountered earlier. We characterize when two circles $S_R(a)$ and $S_r(b)$ intersect, and investigate other issues, such as a family of right triangles inscribed in S^1, and a geometrical construction of square roots. As we discuss in the introduction to Chapter 2, further study of circles demands more sophisticated tools.

To close this chapter introduction, let us note how Euclid's postulates for plane geometry are seen to hold for geometry on $\mathbb{C}$. The five postulates can be stated as follows.

1. A unique straight line segment can be drawn between any two points.
2. Any straight line segment can be extended indefinitely in a straight line.
3. A unique circle can be drawn with any given center and radius.

4. All right angles are congruent.
5. Through a point not on a line, exactly one parallel line can be drawn.

The fifth is the famous parallel postulate. The statement of the parallel postulate in the Elements is more elaborate, but this simpler statement has been shown to suffice. The original version is given in §1.7.

In the setting of the geometry of $\mathbb{C}$, all these are provable assertions. Postulates 1, 2, and 5 are shown to hold in Propositions 1.4.1–1.4.3, with more on the fifth postulate in §1.7. Postulate 4 is a consequence of results of §1.5; see (1.5.20). Postulate 3, on circles, follows from the defining formula (1.0.4), which starts §1.11. Further elaboration, treating a circle as a curve, is given in §2.2 of Chapter 2.

1.1. The set $\mathbb{C}$ of complex numbers

As mentioned in the Preface, the set $\mathbb{C}$ of complex numbers was created to extend the set $\mathbb{R}$ of real numbers so as to allow one to solve the equation $z^2 = -1$. This was obtained by adding an imaginary unit i, defined to satisfy

$$i^2 = -1. \tag{1.1.1}$$

A complex number $z \in \mathbb{C}$ has the form

$$z = x + iy, \quad x, y \in \mathbb{R}. \tag{1.1.2}$$

We have $\mathbb{R} \hookrightarrow \mathbb{C}$ via $x \mapsto x + i0$. Given z as in (1.1.2), we define $\operatorname{Re} z$, $\operatorname{Im} z \in \mathbb{R}$ by

$$\operatorname{Re} z = x, \quad \operatorname{Im} z = y, \quad \text{so} \quad z = \operatorname{Re} z + i \operatorname{Im} z. \tag{1.1.3}$$

We say an element $z \in \mathbb{C}$ is real ($z \in \mathbb{R}$) provided $\operatorname{Im} z = 0$, and that z is purely imaginary ($z \in i\mathbb{R}$) provided $\operatorname{Re} z = 0$.

To define the algebraic structure of $\mathbb{C}$, we start with addition. If z is as in (1.1.2), and also $w = u + iv$, $u, v \in \mathbb{R}$, we define

$$z + w = (x + u) + i(y + v). \tag{1.1.4}$$

Given that we want multiplication in $\mathbb{C}$ to satisfy

$$(x + i0)(u + i0) = xu, \quad (0 + iy)(0 + iv) = -yv, \tag{1.1.5}$$

and expect to obtain the usual commutative, associative, and distributive laws, we are led to the following definition of multiplication on $\mathbb{C}$:

$$(x + iy)(u + iv) = (xu - yv) + i(xv + yu). \tag{1.1.6}$$

One readily verifies the standard commutative, associative, and distributive laws, given these definitions of addition and multiplication on $\mathbb{C}$, as a consequence of their counterparts on $\mathbb{R}$. (A systematic development of $\mathbb{R}$ is given in Chapter 1 of [**15**].)

The set $\mathbb{C}$ also possesses an important map called *complex conjugation*: for $x, y \in \mathbb{R}$,

$$z = x + iy \Longrightarrow \overline{z} = x - iy. \tag{1.1.7}$$

Note that

$$2 \operatorname{Re} z = z + \overline{z}, \quad 2 \operatorname{Im} z = -i(z - \overline{z}). \tag{1.1.8}$$

One readily verifies that, for $z, w \in \mathbb{C}$,

$$\overline{z + w} = \overline{z} + \overline{w}, \quad \overline{zw} = \overline{z}\,\overline{w}. \tag{1.1.9}$$

Another important identity is

$$z\overline{z} = x^2 + y^2. \tag{1.1.10}$$

With this in mind, we define the absolute value of z as

$$|z| = (z\overline{z})^{1/2} = \sqrt{x^2 + y^2}. \tag{1.1.11}$$

Note that

$$z \neq 0 \Longrightarrow |z| > 0. \tag{1.1.12}$$

Also, by (1.1.9), together with the commutative and associative laws of multiplication,

$$|zw| = |z|\,|w|. \tag{1.1.13}$$

Furthermore,

$$z \neq 0 \Rightarrow z\Big(\frac{1}{|z|^2}\overline{z}\Big) = 1 \Rightarrow \frac{1}{z} = \frac{1}{|z|^2}\overline{z}. \tag{1.1.14}$$

In such a case, we can divide:

$$\frac{w}{z} = w\frac{1}{z} = \frac{1}{|z|^2}w\overline{z}. \tag{1.1.15}$$

The algebraic properties established above make $\mathbb{C}$ an object known as a field. The identity (1.1.13) makes it a normed field.

Exercises

1. The following laws governing addition and multiplication on $\mathbb{C}$ were mentioned after the definitions of these operations in (1.1.4)–(1.1.6):

commutative and associative laws of addition:

$$z + w = w + z, \quad z + (w + u) = (z + w) + u,$$

commutative and associative laws of multiplication:

$$zw = wz, \quad z(wu) = (zw)u,$$

distributive law:

$$z(w + u) = zw + zu.$$

These hold for all $z, w, u \in \mathbb{C}$. Verify these identities, using their counterparts for $\mathbb{R}$.

Example. In (1.1.4), note that the commutative law of addition on $\mathbb{R}$ yields

$$x + u = u + x \text{ and } y + v = v + y.$$

2. Verify the identities involving complex conjugation stated in (1.1.8)–(1.1.10).

3. Verify the identity (1.1.13) by taking

$$|z|^2 = z\overline{z}, \quad |w|^2 = w\overline{w}, \quad |zw|^2 = (zw)(\overline{zw}),$$

and using results established in Exercises 1–2.

4. Given $x \in \mathbb{R}$, write the quotient

$$\frac{x + i}{x - i}$$

in the form given in (1.1.15).

1.2. Metric properties of $\mathbb{C}$

A key to the metric structure of $\mathbb{C}$ is provided by the following computation of the absolute value of a sum. Given $z, w \in \mathbb{C}$, we have

$$\begin{aligned} |z+w|^2 &= (z+w)(\overline{z}+\overline{w}) \\ &= |z|^2 + |w|^2 + z\overline{w} + w\overline{z} \\ &= |z|^2 + |w|^2 + 2\,\mathrm{Re}\, z\overline{w}. \end{aligned} \tag{1.2.1}$$

The last term has special importance. We set

$$(z,w) = z\overline{w}, \quad \langle z,w\rangle = \mathrm{Re}(z,w) = \mathrm{Re}\, z\overline{w}, \tag{1.2.2}$$

the latter quantity called the inner product of z and w. Then (1.2.1) says

$$|z+w|^2 = |z|^2 + |w|^2 + 2\langle z,w\rangle. \tag{1.2.3}$$

Note that $(z,w) = \overline{(w,z)}$ and $\langle z,w\rangle = \langle w,z\rangle$. Also, if $z = x+iy,\ w = u+iv,\ x,y,u,v \in \mathbb{R}$, then

$$z\overline{w} = xu + yv + i(yu - xv), \tag{1.2.4}$$

hence

$$\langle z,w\rangle = xu + yv. \tag{1.2.5}$$

Note that the quantity on the right side is simply the dot product on $\mathbb{R}^2$.

Definition. Given $z, w \in \mathbb{C}$,

$$z \perp w \Longleftrightarrow \langle z,w\rangle = 0. \tag{1.2.6}$$

When (1.2.6) holds, we say z is orthogonal (or perpendicular) to w. Note that $z \perp w \Leftrightarrow w \perp z$. The following consequence of (1.2.3) merits special attention.

Proposition 1.2.1. *Given $z, w \in \mathbb{C}$,*

$$|z+w|^2 = |z|^2 + |w|^2 \Longleftrightarrow z \perp w. \tag{1.2.7}$$

Our next result is a special case of the *triangle inequality*. This terminology will be explained a little later.

Proposition 1.2.2. *Given $z, w \in \mathbb{C}$,*

$$|z+w| \le |z| + |w|. \tag{1.2.8}$$

Proof. We compare the formula (1.2.3) for $|z+w|^2$ with the computation

$$(|z| + |w|)^2 = |z|^2 + |w|^2 + 2|z|\,|w|. \tag{1.2.9}$$

The fact that (1.2.8) follows from (1.2.3) and (1.2.9) is the content of the following result, a special case of *Cauchy's inequality*. □

Proposition 1.2.3. *Given $z, w \in \mathbb{C}$,*

$$\langle z, w\rangle \le |z|\,|w|. \tag{1.2.10}$$

Proof. Generally, for $\zeta \in \mathbb{C}$, the analogue of (1.1.10) implies $\operatorname{Re}\zeta \le |\zeta|$. Hence

$$\operatorname{Re} z\overline{w} \le |z\overline{w}| = |z|\,|w|, \tag{1.2.11}$$

the last identity by (1.1.13), plus the fact that $|w| = |\overline{w}|$. □

NOTE. Replacing z by $-z$ in (1.2.10) yields $-\langle z, w\rangle \le |z|, |w|$, hence

$$|\langle z, w\rangle| \le |z|\,|w|. \tag{1.2.12}$$

Having these results, we define the *distance* between points in $\mathbb{C}$:

Definition. Given $p, q \in \mathbb{C}$, the distance between p and q is

$$d(p, q) = |p - q|. \tag{1.2.13}$$

Note that $d(p, q) = d(q, p)$. Also

$$d(p + z, q + z) = d(p, q), \quad \forall\, z \in \mathbb{C}. \tag{1.2.14}$$

The following consequence of Proposition 1.2.2 is a more general form of the triangle inequality.

Proposition 1.2.4. *Given $p, q, r \in \mathbb{C}$,*

$$d(p, r) \le d(p, q) + d(q, r). \tag{1.2.15}$$

Proof. Set $z = p - q$, $w = q - r$, so $z + w = p - r$. Then (1.2.14) is equivalent to the statement that

$$|z + w| \le |z| + |w|, \tag{1.2.16}$$

which is the content of (1.2.8). □

Returning to the concept of orthogonality in (1.2.6), we have the following useful result.

Proposition 1.2.5. *Two nonzero elements $z, w \in \mathbb{C}$ are orthogonal if and only if their quotient is purely imaginary, i.e.,*

$$z \perp w \Longleftrightarrow \frac{z}{w} \in i\mathbb{R}, \tag{1.2.17}$$

where

$$i\mathbb{R} = \{iy : y \in \mathbb{R}\}. \tag{1.2.18}$$

Proof. Indeed,

$$\operatorname{Re} z\overline{w} = \operatorname{Re} |w|^2 \frac{z}{w} = |w|^2 \operatorname{Re} \frac{z}{w}. \tag{1.2.19}$$

□

Exercises

1. Show that

$$|1+i| = \sqrt{2}, \quad |\sqrt{2}+i| = \sqrt{3}, \ \ldots, \ |\sqrt{n}+i| = \sqrt{n+1}.$$

2. Given nonzero $z, w \in \mathbb{C}$, show that

$$|z+w| = |z| + |w| \iff \frac{z}{w} \in \mathbb{R}^+.$$

Hint. Following the proof of Proposition 1.2.3, reduce the identity on the left to

$$\operatorname{Re} z\overline{w} = |z\overline{w}|.$$

Generally, for $\xi \in \mathbb{C}$,

$$\operatorname{Re} \xi = |\xi| \iff \xi = |\xi|.$$

3. To complement the inner product on $\mathbb{C}$, we have the cross product,

$$z \times w = \operatorname{Im} z\overline{w}.$$

Thus

$$z\overline{w} = \langle z, w \rangle + (z \times w)i.$$

Deduce that, for $z, w \in \mathbb{C}$,

$$|z|^2|w|^2 = \langle z, w \rangle^2 + (z \times w)^2.$$

Note how this refines Cauchy's inequality, (1.2.10). Show also that the cross product is skew-symmetric:

$$z \times w = -w \times z, \quad \text{hence} \ \ z \times z = 0.$$

4. Show that

$$z \times (iw) = -\langle z, w \rangle, \quad \text{hence} \ \ z \times (iz) = -|z|^2.$$

1.3. Rigid motions on $\mathbb{C}$

We say a map $F : \mathbb{C} \to \mathbb{C}$ is a *rigid motion* (or an *isometry*) provided

$$d(F(z), F(w)) = d(z, w), \quad \forall\, z, w \in \mathbb{C}. \tag{1.3.1}$$

We denote the set of all such maps by $\mathrm{Isom}(\mathbb{C})$. We have

$$F_1, F_2 \in \mathrm{Isom}(\mathbb{C}) \Longrightarrow F_1 \circ F_2 \in \mathrm{Isom}(\mathbb{C}), \tag{1.3.2}$$

where the composition $F_1 \circ F_2$ is defined by

$$F_1 \circ F_2(z) = F_1(F_2(z)). \tag{1.3.3}$$

Here are some examples of isometries:

$$\tau_a(z) = z + a, \quad a \in \mathbb{C}, \tag{1.3.4}$$

$$\rho_\omega(z) = \omega z, \quad \omega \in \mathbb{C},\ |\omega| = 1, \tag{1.3.5}$$

$$C(z) = \overline{z}. \tag{1.3.6}$$

These maps are known, respectively, as translations, rotations, and reflection about the real axis. Note that each of these maps is invertible, and the inverses are isometries:

$$\tau_a^{-1} = \tau_{-a}, \quad \rho_\omega^{-1} = \rho_{1/\omega}, \quad C^{-1} = C. \tag{1.3.7}$$

To formulate our first main result, we set

$$\begin{aligned} \mathrm{O}(\mathbb{C}) &= \{F \in \mathrm{Isom}(\mathbb{C}) : F(0) = 0\}, \\ \mathrm{O}_+(\mathbb{C}) &= \{\rho_\omega : \omega \in S^1\}, \\ \mathrm{O}_-(\mathbb{C}) &= \{C \circ \rho_\omega : \omega \in S^1\}, \end{aligned} \tag{1.3.8}$$

where

$$S^1 = \{\omega \in \mathbb{C} : |\omega| = 1\} \tag{1.3.9}$$

is the unit circle, centered at the origin. Here is our first result.

Proposition 1.3.1. *We have*

$$\mathrm{O}(\mathbb{C}) = \mathrm{O}_+(\mathbb{C}) \cup \mathrm{O}_-(\mathbb{C}). \tag{1.3.10}$$

As a preliminary, we establish:

Lemma 1.3.2. *If $F \in \mathrm{O}(\mathbb{C})$, $z, w \in \mathbb{C}$,*

$$\langle F(z), F(w)\rangle = \langle z, w\rangle. \tag{1.3.11}$$

Proof. By (1.2.3),

$$\begin{aligned} -2\langle z, w\rangle &= |z - w|^2 - |z|^2 - |w|^2 \\ &= d(z, w)^2 - d(z, 0)^2 - d(w, 0)^2, \end{aligned} \tag{1.3.12}$$

and similarly

$$-2\langle F(z), F(w)\rangle = d(F(z), F(w))^2 - d(F(z), 0)^2 - d(F(w), 0)^2. \tag{1.3.13}$$

Since the right sides of (1.3.12) and (1.3.13) coincide, we have (1.3.11). □

Corollary 1.3.3. *If $F \in \mathrm{O}(\mathbb{C})$, $z, w \in \mathbb{C}$,*

$$z \perp w \Longrightarrow F(z) \perp F(w), \tag{1.3.14}$$

hence

$$F(iz) = \pm iF(z). \tag{1.3.15}$$

Proof. The result (1.3.14) is immediate from (1.3.11). Now, if $z \neq 0$, since $z \perp iz$, $F(z) \perp F(iz)$, so by Proposition 1.2.5,

$$\frac{F(iz)}{F(z)} = ir, \quad r \in \mathbb{R}, \ |r| = \frac{|F(iz)|}{|F(z)|} = 1, \tag{1.3.16}$$

so $r = \pm 1$. □

Proof of Proposition 1.3.1. Given $F \in \mathrm{O}(\mathbb{C})$, suppose $F(1) = \omega$. Then $\omega \in S^1$, and

$$G = \rho_\omega^{-1} \circ F \in \mathrm{O}(\mathbb{C}), \quad G(1) = 1. \tag{1.3.17}$$

By Corollary 1.3.3, $G(i) = \pm i$. We claim that

$$\begin{aligned} G(i) = i &\Longrightarrow G(z) \equiv z, \\ G(i) = -i &\Longrightarrow G(z) \equiv \overline{z}. \end{aligned} \tag{1.3.18}$$

This result will yield (1.3.10). To start, suppose

$$G \in \mathrm{O}(\mathbb{C}), \quad G(1) = 1, \ G(i) = i. \tag{1.3.19}$$

Pick $z = x + iy$ and set $G(z) = u + iv$, $x, y, u, v \in \mathbb{R}$. Then

$$\begin{aligned} \langle z, 1\rangle = x, &\qquad \langle z, i\rangle = y, \\ \langle G(z), 1\rangle = u, &\quad \langle G(z), i\rangle = v, \end{aligned} \tag{1.3.20}$$

so Lemma 1.3.2 implies $u = x$, $v = y$, and we have the first half of (1.3.18). The proof of the second half is similar.

Corollary 1.3.4. *If $F \in \mathrm{Isom}(\mathbb{C})$, then either*

$$F(z) = \omega z + a, \quad \forall\, z \in \mathbb{C}, \tag{1.3.21}$$

for some $a \in \mathbb{C}$, $\omega \in S^1$, or

$$F(z) = \omega \overline{z} + a, \quad \forall\, z \in \mathbb{C}, \tag{1.3.22}$$

for some such a, ω.

Corollary 1.3.5. *If $F \in \mathrm{Isom}(\mathbb{C})$, then $F : \mathbb{C} \to \mathbb{C}$ is one-to-one and onto, with inverse $F^{-1} \in \mathrm{Isom}(\mathbb{C})$.*

REMARK. Given $F \in \mathrm{Isom}(\mathbb{C})$, we can write

$$F(z) = F_0(z) + a, \quad F_0 \in \mathrm{O}(\mathbb{C}),\ a \in \mathbb{C}, \tag{1.3.23}$$

and then, for $z, w \in \mathbb{C}$,

$$F(z) - F(w) = F_0(z) - F_0(w) = F_0(z - w). \tag{1.3.24}$$

We denote by $\mathrm{Isom}^+(\mathbb{C})$ the set of isometries of the form (1.3.21) and by $\mathrm{Isom}^-(\mathbb{C})$ the set of isometries of the form (1.3.22). Note that if $F^\pm, G^\pm \in \mathrm{Isom}^\pm(\mathbb{C})$, then

$$\begin{aligned} &F^+ \circ G^+, F^- \circ G^- \in \mathrm{Isom}^+(\mathbb{C}), \\ &F^+ \circ G^-, F^- \circ G^+ \in \mathrm{Isom}^-(\mathbb{C}), \end{aligned} \tag{1.3.25}$$

and

$$(F^\pm)^{-1} \in \mathrm{Isom}^\pm(\mathbb{C}). \tag{1.3.26}$$

We now introduce the concept of *congruence*. Given two subsets of $\mathbb{C}$, $S_1, S_2 \subset \mathbb{C}$, we say S_1 and S_2 are congruent provided there exists $F \in \mathrm{Isom}(\mathbb{C})$ such that

$$F(S_1) = S_2, \tag{1.3.27}$$

where $F(S) = \{F(z) : z \in S\}$. If (1.3.27) holds for some $F \in \mathrm{Isom}^+(\mathbb{C})$, we say F is a rotational congruence. If it holds with $F \in \mathrm{Isom}^-(\mathbb{C})$, we say F is a reflection congruence.

Exercises

1. Given $z = x + iy,\ x, y \in \mathbb{R}$, set

$$\widetilde{C}(z) = z^\dagger = -x + iy.$$

Show that $\widetilde{C} \in \mathrm{Isom}^-(\mathbb{C})$. Express $\widetilde{C}$ in the form $C \circ \rho_\omega$.
Hint. Compute $\overline{z}^\dagger$.

2. Recall the cross product $z \times w$ on $\mathbb{C}$, introduced in Exercise 3 of §1.2. Show that, for $z, w \in \mathbb{C}$,

$$F \in \mathrm{O}_+(\mathbb{C}) \Longrightarrow F(z) \times F(w) = z \times w,$$
$$F \in \mathrm{O}_-(\mathbb{C}) \Longrightarrow F(z) \times F(w) = -z \times w.$$

3. Let p_1, p_2 be distinct points in $\mathbb{C}$, $a = \mathrm{dist}(p_1, p_2)$. Show that there are a unique $F_+ \in \mathrm{Isom}^+(\mathbb{C})$ and a unique $F_- \in \mathrm{Isom}^-(\mathbb{C})$ such that

$$F_\pm(p_1) = 0, \quad F_\pm(p_2) = a.$$

Hint. Consider $F_1 = \tau_{-p_1}$. Then $F_1(p_1) = 0,\ F_2(p_2) = p_2 - p_1 = \tilde{p},\ |\tilde{p}| = a$. Next, take $\omega = \tilde{p}/a \in S^1$, $F_2 = \rho_{1/\omega} \circ F_1$. Then $F_2(p_1) = 0,\ F_2(p_2) = a$. Show that $F_+ = F_2,\ F_- = C \circ F_2$.

4. In the setting of Exercise 3, suppose also $q_1, q_2 \in \mathbb{C}$, $d(q_1, q_2) = d(p_1, p_2)$. Show that there are a unique $G_+ \in \mathrm{Isom}^+(\mathbb{C})$ and a unique $G_- \in \mathrm{Isom}^-(\mathbb{C})$ such that

$$G_\pm(p_1) = q_1, \quad G_\pm(p_2) = q_2.$$

5. Given $p \in \mathbb{C},\ \omega \in S^1$, set

$$\rho_{p,\omega}(z) = \omega(z - p) + p.$$

Show that

$$\rho_{p,\omega} \in \mathrm{Isom}^+(\mathbb{C}), \quad \rho_{p,\omega}(p) = p, \ \ \forall\, \omega \in S^1.$$

We say $\rho_{p,\omega}$ is rotation by ω about the point p.

1.4. Lines and line segments

A line in $\mathbb{C}$ is a set of points of the form

$$L_{u,v} = \{ut + v : t \in \mathbb{R}\} = \{\ell_{u,v}(t) : t \in \mathbb{R}\}, \tag{1.4.1}$$

given $u, v \in \mathbb{C},\ u \neq 0$. Note that, by scaling $\mathbb{R}$, we have

$$L_{u,v} = L_{ru,v}, \quad \forall\, r \in \mathbb{R} \setminus 0. \tag{1.4.2}$$

In particular,

$$L_{u,v} = L_{\omega,v}, \quad \omega = \frac{u}{|u|} \in S^1, \tag{1.4.3}$$

where

$$S^1 = \{z \in \mathbb{C} : |z| = 1\}. \tag{1.4.4}$$

Note that if L is a line and $F \in \mathrm{Isom}(\mathbb{C})$, then $F(L)$ is also a line. Indeed,

$$F(L_{u,v}) = \begin{array}{ll} L_{\omega u, \omega v + a}, & \text{for } F \text{ in (1.3.21)}, \\ L_{\omega \overline{u}, \omega \overline{v} + a}, & \text{for } F \text{ in (1.3.22)}. \end{array} \tag{1.4.5}$$

In particular,

$$F(\mathbb{R}) = L_{\omega,a}, \quad \text{for } F \text{ in (1.3.21) or (1.3.22)}. \tag{1.4.6}$$

Equivalently, for such F,

$$\mathbb{R} = F^{-1}(L_{\omega,a}). \tag{1.4.7}$$

The following result verifies Euclid's Postulates 1 and 2, in the setting of geometry on $\mathbb{C}$.

Proposition 1.4.1. *Given two distinct points $p, q \in \mathbb{C}$, there is a line through p and q.*

Proof. Just take the line

$$\ell(t) = \ell_{q-p,p}(t) = p + (q - p)t. \tag{1.4.8}$$

□

There is also a uniqueness result, but first we treat the following basic proposition.

Proposition 1.4.2. *Consider the lines*

$$L_{\omega,a},\ L_{u,v}, \quad \omega, u \in S^1. \tag{1.4.9}$$

If $\omega \neq \pm u$, these lines intersect in exactly one point. If $\omega = \pm u$, these lines either coincide or have empty intersection.

Proof. Take $F \in \mathrm{Isom}(\mathbb{C})$ such that (1.4.6) holds. Then

$$F^{-1}(L_{u,v}) = L_{\alpha,\beta}, \quad |\alpha| = 1,\ \beta \in \mathbb{C}, \tag{1.4.10}$$

and

$$\omega = \pm u \Longleftrightarrow \alpha = \pm 1. \tag{1.4.11}$$

Thus it suffices to consider the lines $\mathbb{R}$ and

$$L_{\alpha,\beta} = \{\alpha t + \beta : t \in \mathbb{R}\}. \tag{1.4.12}$$

Note that, for $t \in \mathbb{R}$,

$$\mathrm{Im}(\alpha t + \beta) = at + b, \quad a = \mathrm{Im}\,\alpha,\ b = \mathrm{Im}\,\beta, \tag{1.4.13}$$

and, given $\alpha \in S^1$,

$$\alpha \neq \pm 1 \Longleftrightarrow a \neq 0. \tag{1.4.14}$$

If $a \neq 0$, then $at + b = 0$ if and only if $t = -b/a$, and the lines $L_{\alpha,\beta}$ and $\mathbb{R}$ have exactly one point of intersection. If $a = 0$, then $at + b \equiv b$, and this vanishes for all t if and only if $b = 0$, and vanishes for no t if and only if $b \neq 0$. This completes the proof. □

Remark. Given the lines in (1.4.9), if $\omega/u \in \mathbb{R}$ and the lines do not coincide, we say they are *parallel*. If $\omega \perp u$, we say they are *orthogonal*.

Given Proposition 1.4.2, and given a line L through distinct points p and q, if L' is another line through p, and L' does not coincide with L, it follows that L' intersects L only at p. This gives the uniqueness of the line arising in Proposition 1.4.1. As another consequence of Proposition 1.4.2, we verify the Parallel Postulate, for geometry on $\mathbb{C}$.

Proposition 1.4.3. *If L is a line in $\mathbb{C}$ and $p \in \mathbb{C}$, $p \notin L$, there is a unique line L' through p that does not intersect L, and it is parallel to L.*

Proof. We can write $L = L_{\omega,v}$, as in (1.4.3). Then $L' = L_{\omega,p}$ contains p and is parallel to L. If L'' is another line through p, it is expressible as $L'' = L_{\alpha,p}$, for some $\alpha \neq \pm\omega$. By Proposition 1.4.2, it must intersect L. □

The previous result on drawing a parallel has the following useful complement.

Proposition 1.4.4. *Let L be a line in $\mathbb{C}$, $p \in \mathbb{C}$. Then there is a unique line $L_p^\perp$ through p that intersects L orthogonally.*

Proof. Write $L = L_{\omega,v}$, as in (1.4.3). Then $L_p^\perp$ is given by

$$L_p^\perp = L_{i\omega,p}. \tag{1.4.15}$$

□

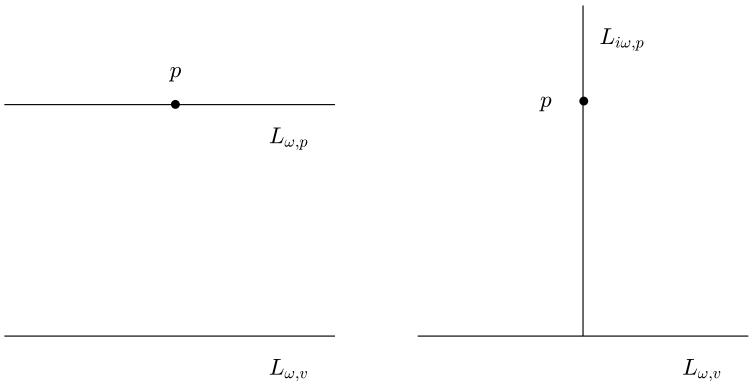

Figure 1.4.1. Drawing a parallel and dropping a perpendicular

REMARK. One says we drop a perpendicular from p through L. See Figure 1.4.1 for an illustration of Propositions 1.4.3–1.4.4.

We now define the *reflection* across a line.

Proposition 1.4.5. *Let $L = L_{\omega,a}$ be a line. There is a unique*

$$R_L \in \mathrm{Isom}^-(\mathbb{C}), \text{ such that } R_L(z) = z, \ \forall\, z \in L. \tag{1.4.16}$$

Proof. By (1.4.6), we can take $F \in \mathrm{Isom}^+(\mathbb{C})$ such that $L_{\omega,a} = F(\mathbb{R})$, so it suffices to treat the case $L = \mathbb{R}$. In this case, the map satisfying (1.4.16) is simply

$$C(z) = \overline{z}. \tag{1.4.17}$$

Thus, for L itself, R_L in (1.4.16) is $R_L = F \circ C \circ F^{-1}$. □

Associated to the notion (1.4.1) of a line in $\mathbb{C}$ is the notion of a *line segment*, which has the form

$$L_{u,v,I} = \{ut + v : t \in I\}, \quad I = [a, b], \tag{1.4.18}$$

where $a, b \in I,\ a < b$. Such a line segment has endpoints

$$p = ua + v, \quad q = ub + v. \tag{1.4.19}$$

We say the line segment $L_{u,v,I}$ connects p and q.

We have the following analogue of Proposition 1.4.1

Proposition 1.4.6. *Given two distinct point $p, q \in \mathbb{C}$, there is a unique line segment connecting p and q.*

Proof. For existence, take the line

$$\ell(t) = p + (q - p)t, \quad t \in [0, 1]. \tag{1.4.20}$$

Uniqueness again follows from Proposition 1.4.2. □

We denote by $\mathcal{L}_{[p,q]}$ the line segment given by (1.4.20). Thus

$$\mathcal{L}_{[p,q]} = L_{q-p,p,[0,1]}. \tag{1.4.21}$$

Given such points p and q, we have

$$\frac{1}{2}(p + q) = \ell(1/2). \tag{1.4.22}$$

We call this the *midpoint* of the line segment $\mathcal{L}_{[p,q]}$.

Exercises

1. In the setting of Proposition 1.4.6, give a formula for the line through the midpoint of $\mathcal{L}_{[p,q]}$ that is orthogonal to this line segment. We call such a line the perpendicular bisector of $\mathcal{L}_{[p,q]}$.

2. Retaining this setting, show that there are unique maps $F^{\pm} \in \mathrm{Isom}^{\pm}(\mathbb{C})$ such that

$$F^{\pm} : \mathcal{L}_{[p,q]} \to \mathcal{L}_{[p,q]}, \quad F^{\pm}(p) = q, \quad F^{\pm}(q) = p.$$

Show that each map $F^{\pm}$ fixes the midpoint of $\mathcal{L}_{[p,q]}$.

3. Let L and $\widetilde{L}$ be lines that intersect at one point, and take R_L as in Proposition 1.4.5. Show that L and $\widetilde{L}$ intersect orthogonally if and only if $R_L(\widetilde{L}) = \widetilde{L}$.

We say three distinct points p_j in $\mathbb{C}$ are non collinear if p_3 does not lie on the line through p_1 and p_2.

4. Assume p_1, p_2, p_3 are not collinear. Assume $F \in \mathrm{Isom}(\mathbb{C})$ and $F(p_j) = p_j$ for each j. Show that F is the identity map, i.e., $F(z) = z$ for all $z \in \mathbb{C}$.
Hint. Build on the results in Exercises 3–4 of §1.3.

1.5. Rays, wedges, and angles

A *ray* in $\mathbb{C}$ is a half-line. Given $p \in \mathbb{C}$, a ray emanating from p has the form

$$R_{p,u} = \{p + tu : t \in [0, \infty)\}, \tag{1.5.1}$$

with $u \in \mathbb{C},\ u \neq 0$. Scaling, we can take $u \in S^1$, defined by (1.4.4).

A *wedge* in $\mathbb{C}$ is an ordered pair of rays issuing from a common point p:

$$W_{p,u,v} = \{R_{p,u}, R_{p,v}\}, \quad u, v \in \mathbb{C} \setminus 0. \tag{1.5.2}$$

Again, we could assume $u, v \in S^1$. We say p is the vertex of $W_{p,u,v}$. We define the *angle measurement* of this wedge as

$$\Omega(W_{p,u,v}) = \Omega(u, v) = \frac{u}{v}\frac{|v|}{|u|} = \frac{u\overline{v}}{|u|\,|v|}. \tag{1.5.3}$$

Equivalently, $\Omega(W_{p,u,v}) = \omega$, where $\omega \in S^1$ satisfies

$$\frac{u}{|u|} = \omega\frac{v}{|v|}, \tag{1.5.4}$$

that is,

$$\omega v = ru, \quad r \in (0, \infty). \tag{1.5.5}$$

In other words, the angle measurement ω of the wedge $W_{p,u,v}$ is characterized by the fact that rotation by ω about p moves the ray $R_{p,v}$ to line up with $R_{p,u}$. See Figure 1.5.1 for an illustration of

$$\Omega(W_{0,\omega,1}) = \omega, \tag{1.5.6}$$

given $\omega \in S^1$. Note that

$$\Omega(u, v) = \overline{\Omega(v, u)}, \ \text{ hence } \ \Omega(W_{p,u,v}) = \overline{\Omega(W_{p,v,u})}. \tag{1.5.7}$$

It is frequently convenient to bring in the sets

$$S^1_+ = \{z \in S^1 : \operatorname{Im} z > 0\}, \quad S^1_- = \{z \in S^1 : \operatorname{Im} z < 0\}, \tag{1.5.8}$$

so we have a disjoint union

$$S^1 = S^1_+ \cup S^1_- \cup \{1, -1\}. \tag{1.5.9}$$

The following result will allow us to *bisect* an angle.

Proposition 1.5.1. *Given $\omega \in S^1_+$, there is a unique $\alpha \in S^1_+$ such that*

$$\alpha^2 = \omega, \quad \textit{namely } \ \alpha = \frac{1 + \omega}{|1 + \omega|}. \tag{1.5.10}$$

Similarly, given $\omega \in S^1_-$, there is a unique $\alpha \in S^1_-$ satisfying (1.5.10).

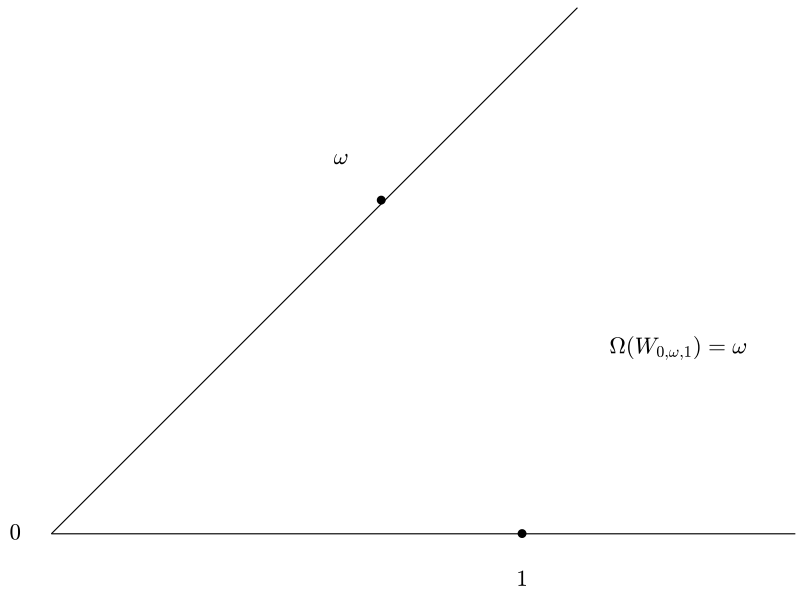

Figure 1.5.1. Wedge with angle measurement ω

Proof. Set $r = |1 + \omega|$, so

$$r^2 = (1 + \omega)(1 + \overline{\omega}) = 2 + \omega + \overline{\omega}. \tag{1.5.11}$$

We claim that

$$\alpha = \frac{1}{r}(1 + \omega) \Longrightarrow \alpha^2 = \omega. \tag{1.5.12}$$

Indeed,

$$\begin{aligned} r^2\alpha^2 &= (1 + \omega)^2 = 1 + 2\omega + \omega^2, \\ r^2\omega &= (2 + \omega + \overline{\omega})\omega = 2\omega + \omega^2 + 1. \end{aligned} \tag{1.5.13}$$

This covers existence. For uniqueness, if also $\beta \in S^1$,

$$\alpha^2 = \beta^2 = \omega \Rightarrow \left(\frac{\alpha}{\beta}\right)^2 = 1 \Rightarrow \frac{\alpha}{\beta} = \pm 1. \tag{1.5.14}$$

□

REMARK. More generally, given $\omega \in S^1_+$ (resp., S^1_-) and $n \in \mathbb{N},\ n \geq 2$, there exists $\alpha \in S^1_+$ (resp., S^1_-) such that $\alpha^n = \omega$. For $n \geq 3$, there is a unique such α with maximal real part. This result is more subtle when $n \geq 3$ than for $n = 2$. We will see a neat derivation in §2.5.

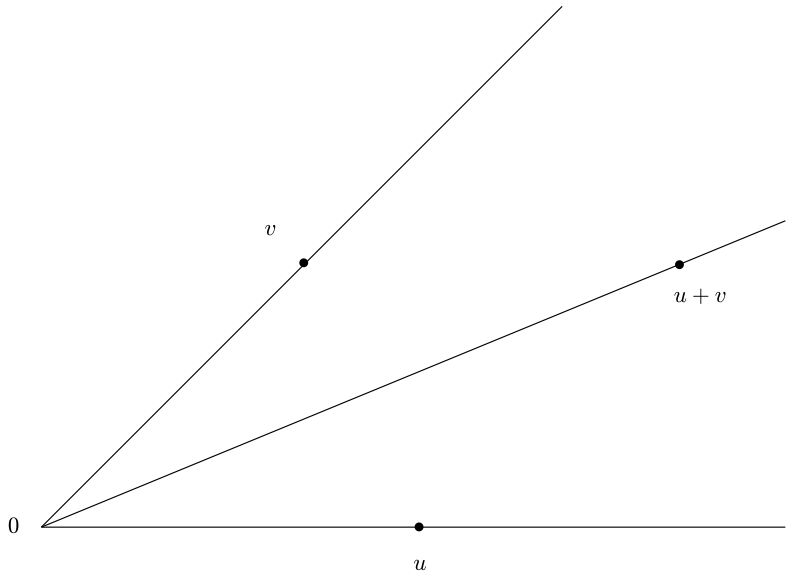

Figure 1.5.2. Bisecting a wedge

Proposition 1.5.1 yields the following result on bisecting a wedge.

Proposition 1.5.2. *Let $W_{p,u,v}$ be a wedge. Assume*

$$|u| = |v|, \quad \Omega(W_{p,u,v}) = \omega \in S^1_+. \tag{1.5.15}$$

Take $\alpha \in S^1_+$ satisfying (1.5.10)*. Then*

$$W_{p,u,\alpha v}, \quad W_{p,\alpha v,v} \tag{1.5.16}$$

are wedges, each with angle measurement α. We say the ray

$$R_{p,\alpha v} \tag{1.5.17}$$

bisects the wedge $W_{p,u,v}$. Furthermore,

$$R_{p,\alpha v} = R_{p,u+v}. \tag{1.5.18}$$

Proof. We have $u = \omega v = \alpha(\alpha v)$. □

Similar results hold if, in (1.5.15), $\omega \in S^1_-$. See Figure 1.5.2 for an illustration of angle bisection. We also note the following.

Corollary 1.5.3. *In the setting of Proposition 1.5.2, reflection R_L across the line $L = L_{u+v,p}$ gives $R_L \in \mathrm{Isom}^-(\mathbb{C})$ such that*

$$R_L(u) = v, \quad R_L(v) = u, \quad R_L(z) = z, \ \forall\, z \in L_{u+v,p}. \tag{1.5.19}$$

Exercises

1. Given $u, v, w \in \mathbb{C} \setminus 0,\ p \in \mathbb{C}$, show that

$$\Omega(W_{p,u,v})\Omega(W_{p,v,w}) = \Omega(W_{p,u,w}).$$

2. In the setting of Exercise 1, show that

$$\begin{aligned} F \in \mathrm{Isom}^+(\mathbb{C}) &\Rightarrow \Omega(F(W_{p,u,v})) = \Omega(W_{p,u,v}), \\ F \in \mathrm{Isom}^-(\mathbb{C}) &\Rightarrow \Omega(F(W_{p,u,v})) = \overline{\Omega(W_{p,u,v})}. \end{aligned}$$

3. Assume $u_j, v_j \in S^1,\ p_j \in \mathbb{C}$. Show that

$$\begin{aligned} &\Omega(W_{p_1,u_1,v_1}) = \Omega(W_{p_2,u_2,v_2}) \\ &\Rightarrow \exists\, F \in \mathrm{Isom}^+(\mathbb{C}) \text{ such that } F(W_{p_1,u_1,v_1}) = W_{p_2,u_2,v_2}. \end{aligned} \tag{1.5.20}$$

4. In the setting of Proposition 1.5.2, show that the wedge bisector $R_{p,\alpha v}$ passes through the midpoint of the line segment from u to v.

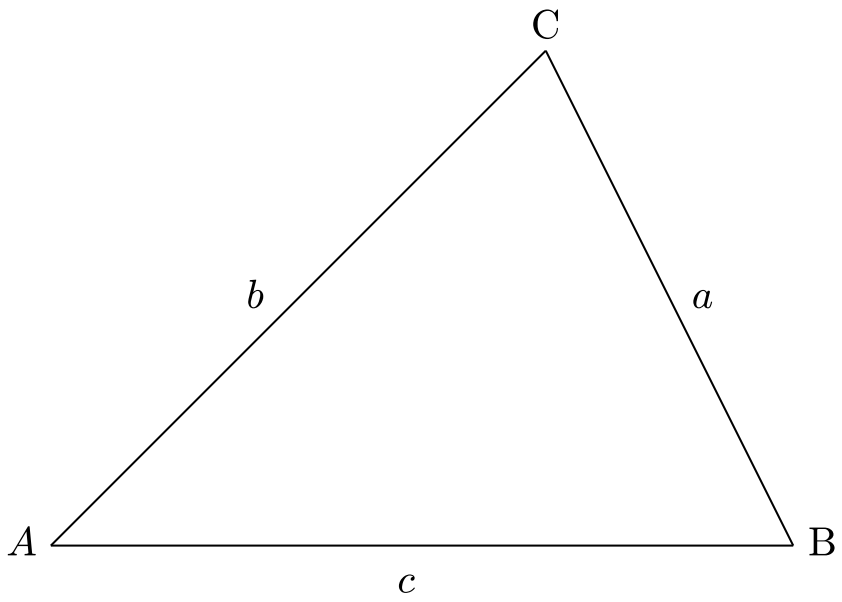

Figure 1.6.1. Triangle

1.6. Triangles

Given three points $A, B, C \in \mathbb{C}$, not collinear, we associate a triangle $\mathcal{T} = \triangle(A, B, C)$. The points A, B, and C are the vertices of $\mathcal{T}$. The sides of $\mathcal{T}$ are the line segments, a, b, and c, opposite these vertices. See Figure 1.6.1.

With slight abuse of notation, we also set

$$c = B - A, \quad a = C - B, \quad b = A - C. \tag{1.6.1}$$

Note that

$$a + b + c = 0. \tag{1.6.2}$$

Recall from §1.5 that, for nonzero $z, w \in \mathbb{C}$, we define the angle measurement $\Omega(z, w) \in S^1$ to be the element $\omega \in S^1$ such that $\omega w = rz,\ r > 0$. Equivalently,

$$\Omega(z, w) = \mathcal{U}\Big(\frac{z}{w}\Big), \tag{1.6.3}$$

where $\mathcal{U} : \mathbb{C} \setminus 0 \to S^1$ is given by

$$\mathcal{U}(z) = \frac{z}{|z|}. \tag{1.6.4}$$

We now define the angle measurements of the three angles of $\mathcal{T}$ as

$$\begin{aligned} \Omega_{\mathcal{T}}(A) &= \Omega(-b, c) = \Omega(C - A, B - A), \\ \Omega_{\mathcal{T}}(B) &= \Omega(-c, a) = \Omega(A - B, C - B), \\ \Omega_{\mathcal{T}}(C) &= \Omega(-a, b) = \Omega(B - C, A - C). \end{aligned} \tag{1.6.5}$$

Equivalently,

$$\Omega_{\mathcal{T}}(A) = \mathcal{U}\Big(-\frac{b}{c}\Big), \quad \Omega_{\mathcal{T}}(B) = \mathcal{U}\Big(-\frac{c}{a}\Big), \quad \Omega_{\mathcal{T}}(C) = \mathcal{U}\Big(-\frac{a}{b}\Big). \tag{1.6.6}$$

Here is one more perspective on the angular measurements for the triangle $\mathcal{T} = \triangle(A, B, C)$. Let us cyclically order the vertices of $\mathcal{T}$, as follows:

$$A, B, C, A, B, C, A, B, \ldots$$

If V is one of these vertices, we take the ray R'_V from V to the vertex right after V, and the ray R''_V from V to the next vertex after that. Then $\Omega_{\mathcal{T}}(V) = \omega$ where rotation by ω about the vertex V moves the ray R'_V to line up with R''_V. For example, $\Omega_{\mathcal{T}}(A) = \omega$ satisfies $\omega(B - A)$ is a positive multiple of $C - A$.

The formulas (1.6.6) lead to the following result on how the angles of a triangle combine.

Proposition 1.6.1. *If $\mathcal{T} = \triangle(A, B, C)$ is a triangle with vertices A, B, C, then*

$$\Omega_{\mathcal{T}}(A)\Omega_{\mathcal{T}}(B)\Omega_{\mathcal{T}}(C) = -1. \tag{1.6.7}$$

Proof. From (1.6.4) we have $\mathcal{U}(z)\mathcal{U}(w) = \mathcal{U}(zw)$. Hence (1.6.6) implies that the left side of (1.6.7) is equal to $\mathcal{U}(-1) = -1$. □

A refinement of this is the sum of angles formula presented in Proposition 2.3.3. The following is an important complement to Proposition 1.6.1.

Proposition 1.6.2. *In the setting of Proposition 1.6.1, the angle measurements* (1.6.5) *are either all elements of S^1_+ or all elements of S^1_-, where*

$$S^1_+ = \{\omega \in S^1 : \operatorname{Im}\omega > 0\}, \quad S^1_- = \{\omega \in S^1 : \operatorname{Im}\omega < 0\}. \tag{1.6.8}$$

In light of the formula (1.6.6), this is a consequence of the following result.

Lemma 1.6.3. *If $a, b, c \in \mathbb{C} \setminus 0$ satisfy $a + b + c = 0$ and all the fractions*

$$\frac{b}{c}, \quad \frac{c}{a}, \quad \frac{a}{b} \tag{1.6.9}$$

belong to $\mathbb{C} \setminus \mathbb{R}$, then their imaginary parts are either all positive or all negative.

Proof. It suffices to compare the imaginary parts of

$$1 + \frac{a}{b}, \quad -1 - \frac{b}{a}, \quad \frac{a}{b}, \tag{1.6.10}$$

since the first two quantities in (1.6.10) are equal, respectively, to $-c/b$ and c/a, and, for $z \in \mathbb{C} \setminus \mathbb{R}$, $\operatorname{sgn} \operatorname{Im}(1/z) = \operatorname{sgn} \operatorname{Im}(-z)$. □

We have that all the angle measurements (1.6.5) belong to S^1_+ if and only if all the ratios in (1.6.9) have negative imaginary part.

Here is another useful formula for angle measurements. By (1.5.3), for nonzero $z, w \in \mathbb{C}$,

$$\langle z, w \rangle = |z| \, |w| \operatorname{Re} \Omega(z, w). \tag{1.6.11}$$

It follows that, in the setting of Proposition 1.6.1,

$$\begin{aligned} \operatorname{Re} \Omega_{\mathcal{T}}(A) &= -\frac{\langle b, c \rangle}{|b| \, |c|}, \\ \operatorname{Re} \Omega_{\mathcal{T}}(B) &= -\frac{\langle c, a \rangle}{|c| \, |a|}, \\ \operatorname{Re} \Omega_{\mathcal{T}}(C) &= -\frac{\langle a, b \rangle}{|a| \, |b|}. \end{aligned} \tag{1.6.12}$$

On the other hand, since $a + b + c = 0$, we have from (1.2.3) that

$$\begin{aligned} 2\langle a, b \rangle &= +|a|^2 + |b|^2 - |c|^2, \\ 2\langle b, c \rangle &= -|a|^2 + |b|^2 + |c|^2, \\ 2\langle c, a \rangle &= +|a|^2 - |b|^2 + |c|^2. \end{aligned} \tag{1.6.13}$$

Here is a classification of the angles of a triangle $\mathcal{T} = \mathcal{T}(A, B, C)$. We say that the angle of $\mathcal{T}$ at A is

$$\begin{aligned} \text{an acute angle} &\iff \operatorname{Re} \Omega_{\mathcal{T}}(A) > 0, \\ \text{a right angle} &\iff \operatorname{Re} \Omega_{\mathcal{T}}(A) = 0, \\ \text{an obtuse angle} &\iff \operatorname{Re} \Omega_{\mathcal{T}}(A) < 0. \end{aligned} \tag{1.6.14}$$

A similar characterization applies to the angles at B and C. Note from (1.6.12) that

$$A \text{ is a right angle} \iff b \perp c. \tag{1.6.15}$$

Here is a key property of triangles.

Proposition 1.6.4. *Each triangle has at least two acute angles.*

In light of Propositions 1.6.1–1.6.2, this is a consequence of the following result.

Lemma 1.6.5. *If* $\omega_1, \omega_2 \in S^1_+$ *and* $\operatorname{Re}\omega_1, \operatorname{Re}\omega_2 < 0$*, then*

$$\operatorname{Im}\omega_1\omega_2 < 0. \tag{1.6.16}$$

Hence if also $\omega_3 \in S^1$ *and* $\omega_1\omega_2\omega_3 = -1$*, then* $\operatorname{Im}\omega_3 < 0$*.*

Proof. Suppose that for $j = 1, 2$,

$$\omega_j = -c_j + is_j, \quad c_j, s_j > 0. \tag{1.6.17}$$

Then

$$\omega_2\omega_2 = c_1c_2 - s_1s_2 - i(c_1s_2 + s_1c_2), \tag{1.6.18}$$

and we have (1.6.16). Then if $\omega_3 \in S^1$,

$$\begin{aligned} (\omega_1\omega_2)\omega_3 = -1 &\Rightarrow \omega_3 = -\overline{(\omega_1\omega_2)} \\ &\Rightarrow \operatorname{Im}\omega_3 < 0, \end{aligned} \tag{1.6.19}$$

as asserted. □

If one of the angles of a triangle $\mathcal{T}$ is a right angle, we say $\mathcal{T}$ is a right triangle.

The following result is the Pythagorean theorem. It was essentially established in Proposition 1.2.1, but at that point we had not introduced the terminology for its classical formulation. Here it is.

Proposition 1.6.6. *Assume* $\mathcal{T} = \triangle(A, B, C)$ *is a right triangle, with right angle at* C*. Then*

$$|c|^2 = |a|^2 + |b|^2. \tag{1.6.20}$$

Proof. Parallel to (1.6.15), C is a right angle if and only if $a \perp b$. Then the identity (1.6.20) is immediate from the first identity in (1.6.13). □

Remark. In the setting of Proposition 1.6.6, the side opposite the right angle C is called the *hypotenuse* of $\mathcal{T}$. Thus the content of the Pythagorean theorem for a right triangle $\mathcal{T}$ is that the square of the hypotenuse of $\mathcal{T}$ is equal to the sum of the squares of the other sides.

Exercises

1. Let $\mathcal{T} = \triangle(A, B, C)$ be a triangle. Show that the line segment from A to the midpoint $(B + C)/2$ of the opposite side contains the point

$$\frac{1}{3}(A + B + C).$$

Show that the three line segments from the vertices of $\mathcal{T}$ to the midpoints of their opposite sides all meet at this point. We call this point the center of $\mathcal{T}$.

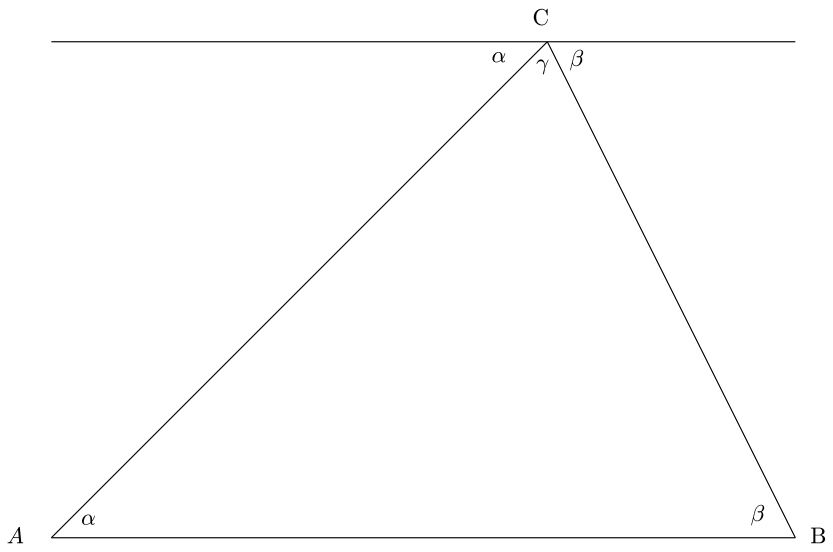

Figure 1.6.2. Commentary on Proposition 1.6.1

2. Let $\mathcal{T} = \triangle(A, B, C)$ be a triangle. Also consider

$$\mathcal{T}' = \triangle(A, C, B), \quad \mathcal{T}'' = \triangle(C, A, B).$$

Show that

$$\Omega_{\mathcal{T}}(A) = \overline{\Omega_{\mathcal{T}'}(A)}, \quad \Omega_{\mathcal{T}}(B) = \overline{\Omega_{\mathcal{T}'}(B)}, \quad \Omega_{\mathcal{T}}(C) = \overline{\Omega_{\mathcal{T}'}(C)},$$

while

$$\Omega_{\mathcal{T}}(A) = \Omega_{\mathcal{T}''}(A), \quad \Omega_{\mathcal{T}}(B) = \Omega_{\mathcal{T}''}(B), \quad \Omega_{\mathcal{T}}(C) = \Omega_{\mathcal{T}''}(C).$$

3. Relate Figure 1.6.2 to the proof of Proposition 1.6.1. The top line through C is parallel to the line through A and B.

1.7. Congruence of triangles

Let $\mathcal{T} = \triangle(A, B, C)$ be a triangle, with associated $a, b, c \in \mathbb{C}$ given by (1.6.1). Suppose $F \in \mathrm{Isom}(\mathbb{C})$, and consider

$$\mathcal{T}' = F(\mathcal{T}), \tag{1.7.1}$$

a triangle with vertices

$$A' = F(A), B' = F(B), C' = F(C), \tag{1.7.2}$$

and associated a', b', c', given by

$$c' = B' - A' = F_0(c), \text{ etc.}, \tag{1.7.3}$$

where, as in (1.3.23), we have

$$F(z) = F_0(z) + p, \quad F_0 \in \mathrm{O}(\mathbb{C}),\ p \in \mathbb{C}. \tag{1.7.4}$$

We say $\mathcal{T}$ and $\mathcal{T}'$ are congruent. Since F preserves distances, we have, of course,

$$|a'| = |a|,\ |b'| = |b|,\ |c'| = |c|. \tag{1.7.5}$$

Let us compare angles, e.g.,

$$\begin{aligned} \Omega_{\mathcal{T}'}(A') &= \Omega(-b', c') = \mathcal{U}\Big(-\frac{b'}{c'}\Big), \\ \Omega_{\mathcal{T}}(A) &= \Omega(-b, c) = \mathcal{U}\Big(-\frac{b}{c}\Big). \end{aligned} \tag{1.7.6}$$

There are two cases to consider:

(i) $F \in \mathrm{Isom}^+(\mathbb{C}) \Rightarrow F_0(z) = \alpha z$,
(ii) $F \in \mathrm{Isom}^-(\mathbb{C}) \Rightarrow F_0(z) = \alpha \overline{z}$,
for some $\alpha \in S^1$.

$$\begin{aligned} &\text{Case i.} \qquad \frac{b'}{c'} = \frac{b}{c}, \quad \text{so } \Omega_{\mathcal{T}'}(A') = \Omega_{\mathcal{T}}(A), \\ &\text{Case ii.} \qquad \frac{b'}{c'} = \frac{\overline{b}}{\overline{c}}, \quad \text{so } \Omega_{\mathcal{T}'}(A') = \overline{\Omega_{\mathcal{T}}(A)}. \end{aligned}$$

Similar calculations hold for the other angles. We summarize.

Proposition 1.7.1. *If $F \in \mathrm{Isom}^+(\mathbb{C})$ and $F(\mathcal{T}) = \mathcal{T}'$, then*

$$\Omega_{\mathcal{T}'}(A') = \Omega_{\mathcal{T}}(A), \quad \Omega_{\mathcal{T}'}(B') = \Omega_{\mathcal{T}}(B), \quad \Omega_{\mathcal{T}'}(C') = \Omega_{\mathcal{T}}(C). \tag{1.7.7}$$

If $F \in \mathrm{Isom}^-(\mathbb{C})$, then

$$\Omega_{\mathcal{T}'}(A') = \overline{\Omega_{\mathcal{T}}(A)}, \quad \Omega_{\mathcal{T}'}(B') = \overline{\Omega_{\mathcal{T}}(B)}, \quad \Omega_{\mathcal{T}'}(C') = \overline{\Omega_{\mathcal{T}}(C)}. \tag{1.7.8}$$

We now consider the existence and uniqueness (up to a rigid motion) of a triangle, given partial geometric data. To begin, we take

$$\alpha \in S^1_+, \quad |b|,\ |c| \in \mathbb{R}^+, \tag{1.7.9}$$

and seek a triangle $\mathcal{T} = \triangle(A, B, C)$ such that

$$|B - A| = |c|, \quad |C - A| = |b|, \quad \Omega_{\mathcal{T}}(A) = \alpha. \tag{1.7.10}$$

Since these data are invariant under application of elements of $\mathrm{Isom}^+(\mathbb{C})$, we can translate and rotate and see that it suffices to treat the case

$$A = 0, \quad B = |c|. \tag{1.7.11}$$

Then the conditions (1.7.10) impose the requirement on $C \in \mathbb{C}$ that

$$|C| = |b|, \quad \mathcal{U}\Big(\frac{C}{B}\Big) = \alpha, \tag{1.7.12}$$

hence

$$C = |b|\alpha. \tag{1.7.13}$$

This establishes the following.

Proposition 1.7.2. *Given the data* (1.7.9)*, there exists a triangle* $\mathcal{T} = \triangle(A, B, C)$ *satisfying* (1.7.10)*. If* $\mathcal{T}'$ *is another such triangle, then* $\mathcal{T}' = F(\mathcal{T})$ *for some* $F \in \mathrm{Isom}^+(\mathbb{C})$*.*

Remarks. The same analysis works if one takes $\alpha \in S^1_-$ instead of $\alpha \in S^1_+$. Also, if $\mathcal{T}'$ satisfies (1.7.10) except that α is replaced by $\overline{\alpha}$, then there exists $F \in \mathrm{Isom}^-(\mathbb{C})$ such that $F(\mathcal{T}') = \mathcal{T}$.

In classical Euclidean geometry, the uniqueness result given above is stated as the use of

$$\text{side-angle-side} \tag{1.7.14}$$

to establish congruence of two triangles.

In the situation described above, we were given one angle of a triangle and (the lengths of) two adjacent sides. We move on to the scenario where we are given (the length of) one side, and (the angle measurements of) two adjacent angles. To be specific, we take

$$\alpha, \beta \in S^1_+, \quad |c| \in \mathbb{R}^+, \tag{1.7.15}$$

and seek a triangle $\mathcal{T} = \triangle(A, B, C)$ such that

$$|B - A| = |c|, \quad \Omega_{\mathcal{T}}(A) = \alpha, \quad \Omega_{\mathcal{T}}(B) = \beta. \tag{1.7.16}$$

In light of Propositions 1.6.1–1.6.2, we must impose the following constraint on α and β:

$$\exists\, \gamma \in S^1_+ \text{ such that } \alpha\beta\gamma = -1. \tag{1.7.17}$$

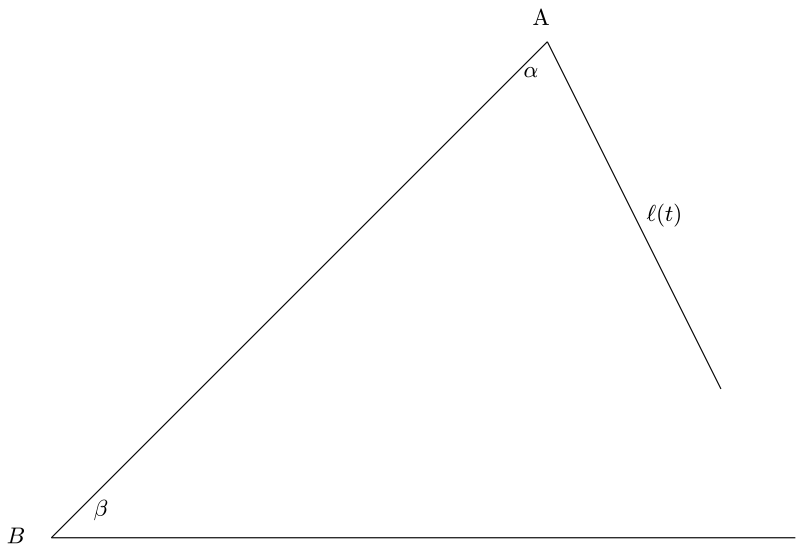

Figure 1.7.1. Illustration of angle-side-angle construction

Given (1.7.15), this holds if and only if

$$\alpha\beta \in S^1_+. \tag{1.7.18}$$

To proceed, this time it is convenient to take a translation and rotation such that

$$B = 0, \quad A = |A|\beta = |c|\beta. \tag{1.7.19}$$

See Figure 1.7.1. We will then find $C \in \mathbb{R}^+$ such that $\triangle(A, B, C)$ satisfies (1.7.16). To find C, we look at the ray,

$$\ell(t) = A + t\overline{\gamma}, \quad t \geq 0. \tag{1.7.20}$$

We have

$$\operatorname{Im} \ell(t) = \operatorname{Im} A - t \operatorname{Im} \gamma. \tag{1.7.21}$$

Since $\operatorname{Im} \gamma > 0$,

$$\operatorname{Im} \ell(t_0) = 0 \text{ for } t_0 = \frac{\operatorname{Im} A}{\operatorname{Im} \gamma} > 0. \tag{1.7.22}$$

This ray intersects $\mathbb{R}^+$ at

$$C = \ell(t_0) = \operatorname{Re}\ell(t_0) = \operatorname{Re} A + (\operatorname{Im} A)\frac{\operatorname{Re}\gamma}{\operatorname{Im}\gamma}. \tag{1.7.23}$$

Now the conditions $\alpha, \beta, \gamma \in S^1_+,\ \alpha\beta\gamma = -1$ imply that $\beta\gamma \in S^1_+$, i.e.,

$$\operatorname{Im}\beta\gamma = \operatorname{Re}\beta\ \operatorname{Im}\gamma + \operatorname{Im}\beta\ \operatorname{Re}\gamma > 0, \tag{1.7.24}$$

so

$$\begin{aligned} C &= |c|\Big(\operatorname{Re}\beta + \operatorname{Im}\beta\frac{\operatorname{Re}\gamma}{\operatorname{Im}\gamma}\Big) \\ &= \frac{|c|}{\operatorname{Im}\gamma}\ \operatorname{Im}(\beta\gamma) > 0. \end{aligned} \tag{1.7.25}$$

We have the following.

Proposition 1.7.3. *Given the data* (1.7.15)*, and given the constraint* (1.7.17) *on* α *and* β*, there exists a triangle* $\mathcal{T} = \triangle(A, B, C)$ *satisfying* (1.7.16)*. In such a case,*

$$\Omega_{\mathcal{T}}(C) = \gamma. \tag{1.7.26}$$

If $\mathcal{T}'$ *is another such triangle, then* $\mathcal{T}' = F(\mathcal{T})$ *for some* $F \in \operatorname{Isom}^+(\mathbb{C})$*.*

Remark. A similar analysis works if one takes $\alpha, \beta \in S^1_-$ instead of $\alpha, \beta \in S^1_+$. Then the constraint (1.7.15) is modified to say $\gamma \in S^1_-$. Also, if $\mathcal{T}'$ satisfies (1.7.15) except that α, β are replaced by $\overline{\alpha}, \overline{\beta}$, then there exists $F \in \operatorname{Isom}^-(\mathbb{C})$ such that $F(\mathcal{T}') = \mathcal{T}$.

In classical Euclidean geometry, the uniqueness result given above is stated as the use of

$$\text{angle-side-angle} \tag{1.7.27}$$

to establish congruence of the triangles.

We move to what is called the

$$\text{side-side-side} \tag{1.7.28}$$

argument to establish congruence of two triangles. Let $\mathcal{T} = \triangle(A, B, C)$ and $\mathcal{T}' = \triangle(A', B', C')$ be triangles, with associated sides a, b, c and a', b', c', respectively. This time, we assume

$$|a| = |a'|, \quad |b| = |b'|, \quad |c| = |c'|. \tag{1.7.29}$$

Recall from (1.6.12)–(1.6.13) that

$$2\langle b, c\rangle = |b|^2 + |c|^2 - |a|^2, \tag{1.7.30}$$

and hence

$$\operatorname{Re}\Omega_{\mathcal{T}}(A) = -\frac{1}{2|b|\,|c|}(|b|^2 + |c|^2 - |a|^2). \tag{1.7.31}$$

Similarly,

$$\operatorname{Re}\Omega_{\mathcal{T}'}(A') = -\frac{1}{2|b'|\,|c'|}(|b'|^2 + |c'|^2 - |a'|^2). \tag{1.7.32}$$

Hence, if (1.7.29) holds, we have

$$\alpha = \Omega_{\mathcal{T}}(A),\ \alpha' = \Omega_{\mathcal{T}'}(B') \Rightarrow \operatorname{Re}\alpha = \operatorname{Re}\alpha'. \tag{1.7.33}$$

Since $\alpha, \alpha' \in S^1$, it follows that

$$\text{either } \alpha' = \alpha \text{ or } \alpha' = \overline{\alpha}. \tag{1.7.34}$$

We pair this identity with the last two identities of (1.7.29) and see that the side-angle-side argument, establishing Proposition 1.7.2 and subsequent comments, applies. This yields the following.

Proposition 1.7.4. *Given triangles $\mathcal{T} = \triangle(A, B, C)$ and $\mathcal{T}' = \triangle(A', B', C')$ such that* (1.7.29) *holds, there exists $F \in \operatorname{Isom}(\mathbb{C})$ such that $\mathcal{T}' = F(\mathcal{T})$.*

Proposition 1.7.4 is the uniqueness result for triangles, given side lengths $|a|, |b|$, and $|c|$. We now investigate the existence question. In light of the triangle inequality, this requires the constraints

$$|a| \le |b| + |c|, \quad |b| \le |c| + |a|, \quad |c| \le |a| + |b|. \tag{1.7.35}$$

Furthermore, strict inequality must hold for the vertices not to be collinear. In the next result, we use the notation a, b, c in place of $|a|, |b|, |c|$, for convenience.

Proposition 1.7.5. *Let a, b, c be positive numbers, satisfying*

$$a < b + c, \quad b < c + a, \quad c < a + b. \tag{1.7.36}$$

Then there is a triangle $\mathcal{T} = \triangle(A, B, C)$ with such sidelengths.

Proof. We take $A = 0,\ B = c \in \mathbb{R}^+$, and need to find $C \in \mathbb{C}$ of the form

$$C = c + az, \quad z \in S^1 \setminus \mathbb{R}. \tag{1.7.37}$$

The triangle $\triangle(A, B, C)$ will have the desired sidelengths provided

$$|c + az| = b. \tag{1.7.38}$$

Since

$$|c + az|^2 = (c + az)(c + a\overline{z}) = c^2 + a^2 + ac(z + \overline{z}), \tag{1.7.39}$$

the requirement is

$$c^2 + a^2 - b^2 = -ac(z + \overline{z}) = -2acx, \tag{1.7.40}$$

for some $z = x + iy \in S^1,\ x, y \in \mathbb{R},\ y \neq 0$. We can find such z satisfying (1.7.40) if and only if

$$-x = \frac{c^2 + a^2 - b^2}{2ac} \in (-1, 1). \tag{1.7.41}$$

If this holds, then (1.7.40) holds, with

$$y = \pm\sqrt{1 - x^2}. \tag{1.7.42}$$

We claim that (1.7.36) implies (1.7.41).

Indeed, given $a, b, c > 0$, (1.7.36) is equivalent to

$$|c - a| < b < c + a, \tag{1.7.43}$$

and this holds

$$\begin{aligned} &\Leftrightarrow a^2 + c^2 - 2ac < b^2 < a^2 + c^2 + 2ac \\ &\Leftrightarrow a^2 + c^2 - b^2 < 2ac \text{ and} \\ &\quad\ b^2 - a^2 - c^2 < 2ac \\ &\Leftrightarrow \frac{|a^2 + c^2 - b^2|}{2ac} < 1. \end{aligned} \tag{1.7.44}$$

This completes the proof. □

Exercises

1. Formulate and prove an analogue of the criteria (1.7.14), (1.7.27), and (1.7.28), of the form

angle-angle-side,

guaranteeing congruence of two triangles.
Hint. The equality of any two sets of angles of two triangles implies the equality of the third. Then (1.7.27) applies.

2. Justify the remarks made below Proposition 1.7.2.

3. Justify the remarks made below Proposition 1.7.3.

4. As stated in the introduction to this chapter, the statement of the parallel postulate given in the Elements is more elaborate than the one given there. It runs as follows:

5′. If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines must intersect each other on that side if extended far enough.

Relate this angle constraint to that of (1.7.18) (in the setting of (1.7.15)), and show that such a result follows from the first part of Proposition 1.7.3.

1.8. Isosceles triangles and equilateral triangles

A triangle is said to be isosceles provided two of its sides have equal lengths. For example, take $\mathcal{T} = \triangle(A, B, C)$ and assume

$$|A - C| = |B - C|, \quad \text{i.e., } |b| = |a|. \tag{1.8.1}$$

We say $\mathcal{T}$ is isosceles at the vertex C. In such a case, let $\mathcal{T}' = \triangle(A', B', C')$, where

$$A' = B, \quad B' = A, \quad C' = C. \tag{1.8.2}$$

We see that

$$|a'| = |b'| = |a| = |b|, \tag{1.8.3}$$

and in particular $|a'| = |a|$, $|b'| = |b|$. Also

$$\Omega_{\mathcal{T}'}(C') = \overline{\Omega_{\mathcal{T}}(C)}. \tag{1.8.4}$$

The side-angle-side result, Proposition 1.7.2, or rather the comments following it, apply. Hence we have:

Proposition 1.8.1. *If $\mathcal{T}$ is isosceles, satisfying* (1.8.1)*, and if $\mathcal{T}'$ has vertices given by* (1.8.2)*, then*

$$\exists\, F \in \mathrm{Isom}^-(\mathbb{C}) \text{ such that } F(\mathcal{T}) = \mathcal{T}'. \tag{1.8.5}$$

In other words, $F \in \mathrm{Isom}^-(\mathbb{C})$ has the property

$$F(A) = B, \quad F(B) = A, \quad F(C) = C. \tag{1.8.6}$$

It follows from Proposition 1.7.1 that

$$\Omega_{\mathcal{T}'}(A') = \overline{\Omega_{\mathcal{T}}(A)}, \quad \Omega_{\mathcal{T}'}(B') = \overline{\Omega_{\mathcal{T}}(B)}, \tag{1.8.7}$$

and since $\Omega_{\mathcal{T}'}(A) = \overline{\Omega_{\mathcal{T}}(A)}$ and $\Omega_{\mathcal{T}'}(B) = \overline{\Omega_{\mathcal{T}}(B)}$, we have:

Corollary 1.8.2. *If $\mathcal{T} = \triangle(A, B, C)$ is isosceles with vertex at C, then*

$$\Omega_{\mathcal{T}}(A) = \Omega_{\mathcal{T}}(B). \tag{1.8.8}$$

We can obtain a converse result as follows. Suppose $\mathcal{T} = \triangle(A, B, C)$ is a triangle, and replace the hypothesis (1.8.1) by the hypothesis that (1.8.8) holds. Again define the triangle $\mathcal{T}' = \triangle(A', B', C')$ by (1.8.2). Since $|c| = |c'|$, we can use the angle-side-angle result, Proposition 1.7.3, or rather the comments following it, to obtain:

Proposition 1.8.3. *If $\mathcal{T} = \triangle(A, B, C)$ and the equi-angularity result* (1.8.8) *holds, then there exists $F \in \mathrm{Isom}^-(\mathbb{C})$ such that* (1.8.6) *holds. Hence $\mathcal{T}$ is isosceles.*

If $\mathcal{T} = \triangle(A, B, C)$ is a triangle all of whose sides have the same length, so

$$|A - C| = |B - C| = |A - B|, \text{ i.e., } |a| = |b| = |c|, \tag{1.8.9}$$

we say $\mathcal{T}$ is an equilateral triangle. Such $\mathcal{T}$ is isosceles in three ways, and repeated applications of Proposition 1.8.1 and Corollary 1.8.2 yield the following.

Proposition 1.8.4. *If $\mathcal{T} = \triangle(A, B, C)$ is an equilateral triangle, then there exist $F_j \in \mathrm{Isom}^-(\mathbb{C})$ such that*

$$\begin{aligned} F_1 &: \triangle(A, B, C) \longrightarrow \triangle(B, A, C), \\ F_2 &: \triangle(A, B, C) \longrightarrow \triangle(A, C, B), \\ F_3 &: \triangle(A, B, C) \longrightarrow \triangle(C, B, A). \end{aligned} \tag{1.8.10}$$

Also, there exists $G \in \mathrm{Isom}^+(\mathbb{C})$ such that

$$\begin{aligned} G &: \triangle(A, B, C) \longrightarrow \triangle(B, C, A), \\ G &: \triangle(B, C, A) \longrightarrow \triangle(C, A, B), \\ G &: \triangle(C, A, B) \longrightarrow \triangle(A, B, C). \end{aligned} \tag{1.8.11}$$

Furthermore,

$$\Omega_{\mathcal{T}}(A) = \Omega_{\mathcal{T}}(B) = \Omega_{\mathcal{T}}(C). \tag{1.8.12}$$

As with Proposition 1.8.3, we also have the converse result:

Proposition 1.8.5. *If $\mathcal{T} = \triangle(A, B, C)$ is a triangle and* (1.8.12) *holds, then $\mathcal{T}$ is an equilateral triangle.*

Exercises

1. Let $\mathcal{T} = \triangle(A, B, C)$ be isosceles at the vertex C. Let

$$M = \frac{1}{2}(A + B)$$

be the midpoint of the opposite side c. Show that $\triangle(C, M, A)$ and $\triangle(C, M, B)$ are congruent, using side-side-side. From this deduce another proof of Corollary 1.8.2.

2. Assume $\mathcal{T} = \triangle(A, B, C)$ has the property (1.8.8), i.e., $\Omega_{\mathcal{T}}(A) = \Omega_{\mathcal{T}}(B)$. Let the angle bisector of $\mathcal{T}$ at C meet the line through A and B at X.

(a) Use Proposition 1.6.1 to show that $\triangle(C, X, A)$ and $\triangle(C, X, B)$ both have right angles at X.

(b) Then use angle-side-angle to show that these two triangles are congruent.

(c) From this, obtain another proof of Proposition 1.8.3.

3. Show that

$$\alpha_{\pm} = \pm\frac{1}{2} + \frac{i}{2}\sqrt{3} \Longrightarrow \alpha_{\pm} \in S^1,\ \alpha_{+}^2 = \alpha_{-},\ \alpha_{\pm}^3 = \mp 1.$$

Show that whenever (1.8.12) holds, their common value is α_{+} (or $\overline{\alpha}_{+}$).
Now let $\mathcal{T} = \triangle(A, B, C)$ and assume $A + B + C = 0$. Show that $\mathcal{T}$ is an equilateral triangle if and only if, with $F(z) = \alpha_{-}z$, α_{-} as above, $F(\mathcal{T}) = \mathcal{T}$.

1.9. Similarity

Given $r \in (0, \infty)$, we have the scaling transformation

$$\delta_r : \mathbb{C} \longrightarrow \mathbb{C}, \quad \delta_r(z) = rz. \tag{1.9.1}$$

We say $G : \mathbb{C} \to \mathbb{C}$ is a similarity transformation if

$$G(z) = F \circ \delta_r(z), \quad F \in \mathrm{Isom}(\mathbb{C}), \tag{1.9.2}$$

and we write $G \in \mathrm{Sim}(\mathbb{C})$. If (1.9.2) holds with $F \in \mathrm{Isom}^{\pm}(\mathbb{C})$, we write $G \in \mathrm{Sim}^{\pm}(\mathbb{C})$. In light of Corollary 1.3.4, $G \in \mathrm{Sim}^{+}(\mathbb{C})$ if and only if

$$G(z) = uz + p, \quad \forall\, z \in \mathbb{C}, \tag{1.9.3}$$

for some $u, p \in \mathbb{C}$ $(u \neq 0)$, and $G \in \mathrm{Sim}^{-}(\mathbb{C})$ if and only if

$$G(z) = u\overline{z} + p, \quad \forall\, z \in \mathbb{C}, \tag{1.9.4}$$

for some such u, p.

If $\mathcal{T} = \triangle(A, B, C)$ and $\mathcal{T}' = \triangle(A', B', C')$, and

$$G \in \mathrm{Sim}(\mathbb{C}), \quad G(\mathcal{T}) = \mathcal{T}', \tag{1.9.5}$$

we say $\mathcal{T}$ and $\mathcal{T}'$ are similar, and write $\mathcal{T} \sim \mathcal{T}'$ (or $\mathcal{T}' \sim \mathcal{T}$). Note that if a, b, c are as in (1.6.1) and a', b', c' are similarly defined, then

$$|a'| = r|a|, \quad |b'| = r|b|, \quad |c'| = r|c|, \tag{1.9.6}$$

if G has the form (1.9.2). Thus (1.9.6) holds with $r = |u|$ if G has the form (1.9.3) or (1.9.4).

It is elementary that if $\mathcal{T}' = \delta_r(\mathcal{T})$, then

$$\Omega_{\mathcal{T}'}(A') = \Omega_{\mathcal{T}}(A), \quad \Omega_{\mathcal{T}'}(B') = \Omega_{\mathcal{T}}(B), \quad \Omega_{\mathcal{T}'}(C') = \Omega_{\mathcal{T}}(C). \tag{1.9.7}$$

Hence we can apply Proposition 1.7.1 to conclude:

Proposition 1.9.1. *Suppose* (1.9.5) *holds. If* $G \in \mathrm{Sim}^{+}(\mathbb{C})$*, then* (1.9.7) *holds. If* $G \in \mathrm{Sim}^{-}(\mathbb{C})$*, then*

$$\Omega_{\mathcal{T}'}(A') = \overline{\Omega_{\mathcal{T}}(A)}, \quad \Omega_{\mathcal{T}'}(B') = \overline{\Omega_{\mathcal{T}}(B)}, \quad \Omega_{\mathcal{T}'}(C') = \overline{\Omega_{\mathcal{T}}(C)}. \tag{1.9.8}$$

Here is an important converse.

Proposition 1.9.2. *Let* $\mathcal{T} = \triangle(A, B, C)$ *and* $\mathcal{T}' = \triangle(A', B', C')$*. If* (1.9.7) *holds, then* $\mathcal{T}' = G(\mathcal{T})$ *for some* $G \in \mathrm{Sim}^{+}(\mathbb{C})$*. If* (1.9.8) *holds, then* $\mathcal{T}' = G(\mathcal{T})$ *for some* $G \in \mathrm{Sim}^{-}(\mathbb{C})$*. In either case,* (1.9.6) *holds for some* $r \in \mathbb{R}^{+}$*.*

Proof. Say $|a'| = s|a|$, $s \in \mathbb{R}^{+}$. Then consider $\mathcal{T}_1 = \delta_s \mathcal{T}$. If (1.9.7) or (1.9.8) holds, it also holds with $\mathcal{T}$ replaced by $\mathcal{T}_1$. But then the angle-side-angle result, Proposition 1.7.3 (and subsequent comments) give $F \in \mathrm{Isom}^{\pm}(\mathbb{C})$ such that $\mathcal{T}' = F(\mathcal{T}_1)$. This establishes the desired result. □

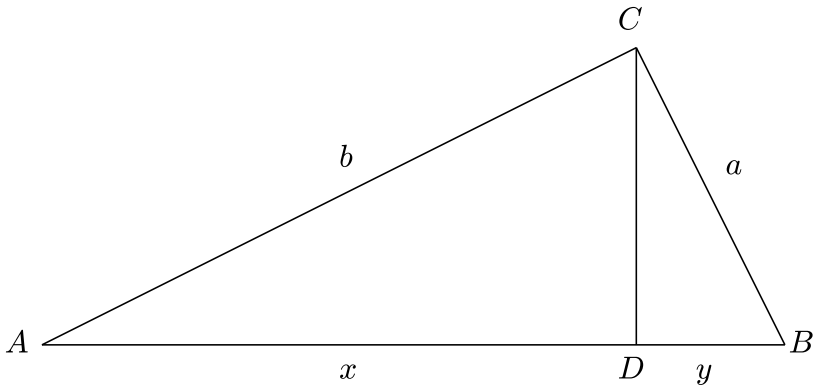

Figure 1.9.1. Three similar triangles

REMARK. In light of Proposition 1.6.1, we see that for the three identities in (1.9.7) to hold, it suffices for any two of them to hold. Ditto for (1.9.8).

We now discuss an interesting collection of three similar triangles, depicted in Figure 1.9.1. We have a right triangle $\mathcal{T} = \triangle(A, B, C)$, with right angle at C. We drop a perpendicular from C to the line segment from A to B, intersecting it at D (cf. Proposition 1.4.4). Here, to simplify notation, we let a, b, c stand for $|a|, |b|, |c|$. The point D divides the interval from A to B into two subintervals, of lengths x and y, satisfying

$$x + y = c. \tag{1.9.9}$$

The other triangles are

$$\mathcal{T}_2 = \triangle(A, C, D), \quad \mathcal{T}_3 = \triangle(C, B, D). \tag{1.9.10}$$

Note that $\mathcal{T}$ and $\mathcal{T}_2$ have the vertex A in common, and $\mathcal{T}$ and $\mathcal{T}_3$ have the vertex B in common. We have

$$\Omega_{\mathcal{T}}(A) = \overline{\Omega_{\mathcal{T}_2}(A)}, \quad \Omega_{\mathcal{T}}(B) = \overline{\Omega_{\mathcal{T}_3}(B)}. \tag{1.9.11}$$

All these triangles have right angles, $\mathcal{T}$ at C, $\mathcal{T}_1$ and $\mathcal{T}_2$ at D, Therefore we can invoke Proposition 1.9.2 (keeping in mind its subsequent remark). We have

$$\mathcal{T} \sim \mathcal{T}_2 \sim \mathcal{T}_3. \tag{1.9.12}$$

We conclude from the analogues of (1.9.6) that

$$\mathcal{T} \sim \mathcal{T}_2 \Rightarrow \frac{c}{b} = \frac{b}{x}, \quad \mathcal{T} \sim \mathcal{T}_3 \Rightarrow \frac{c}{a} = \frac{a}{y}. \tag{1.9.13}$$

These identities imply

$$b^2 = cx, \quad a^2 = cy, \tag{1.9.14}$$

and adding these, keeping in mind that $x + y = c$, yields

$$a^2 + b^2 = c^2, \tag{1.9.15}$$

The identity (1.9.15) states the Pythagorean theorem for the right triangle $\mathcal{T}$. Now the Pythagorean theorem has been baked into our infrastructure since Proposition 1.2.1, so we cannot really regard this as a second proof of this theorem. On the other hand, the argument presented above does give the gist of the second proof of the Pythagorean theorem, in Book VI of Euclid. (Euclid's first proof, in Book I, involves area comparisons. For more on this, see the discussion following Proposition 2.6.5.) Anyway, the argument above might provide insight into how people derived the Pythagorean theorem in the old days. Of course, it does not completely show how they thought. Euclid and his colleagues did not use algebra the way one does in modern times to obtain their geometrical results. See the remarks at the end of §2.9.

Exercises

1. Consider the triangle $\mathcal{T} = \triangle(A, B, C)$. Let ℓ be a line parallel to $c = \mathcal{L}_{[A,B]}$. Assume ℓ intersects $b = \mathcal{L}_{[C,A]}$ at A' and that ℓ intersects $a = \mathcal{L}_{[C,B]}$ at B'. Show that

$$\mathcal{T}' = \triangle(A', B', C) \text{ is similar to } \mathcal{T}.$$

Hint. Show that $\Omega_{\mathcal{T}}(A) = \Omega_{\mathcal{T}'}(A')$.

2. In the setting of Figure 1.9.1, show that (1.9.14) implies

$$\frac{a}{b} = \sqrt{\frac{y}{x}}.$$

1.10. Squares, rectangles, and other quadrilaterals

Here we look at various classes of four-sided figures, known as quadrilaterals, or 4-gons. First, let us make some general comments about n-gons, for $n \geq 3$. An n-gon $\mathcal{P} = G_n(A_0, A_1, \ldots, A_{n-1})$ is specified by n distinct points $A_j \in \mathbb{C},\ 0 \leq j \leq n-1$, which are joined by the sides L_j, line segments from A_j to A_{j+1}, for $0 \leq j \leq n-1$. We make the convention here and below that

$$A_n = A_0, \quad A_{-1} = A_{n-1}. \tag{1.10.1}$$

We impose the requirement that

(1.10.2)

these line segments L_j do not intersect except at their endpoints.

We define the angle measure of the angle at A_j in $\mathcal{P}$ as

$$\Omega_{\mathcal{P}}(A_j) = \Omega_{\mathcal{T}_j}(A_j), \quad \mathcal{T}_j = \triangle(A_{j-1}, A_j, A_{j+1}), \tag{1.10.3}$$

using the convention (1.10.1) for $j = 0$ and $j = n-1$. A straightforward extension of Proposition 1.6.1 is that, if $\mathcal{P} = G_n(A_0, \ldots, A_{n-1})$,

$$\Omega_{\mathcal{P}}(A_0)\Omega_{\mathcal{P}}(A_1)\cdots\Omega_{\mathcal{P}}(A_{n-1}) = (-1)^n. \tag{1.10.4}$$

Specializing to $n = 4$, we first describe rectangles. A rectangle $R = G_4(A, B, C, D)$ is a 4-gon all of whose angles are right angles. See Figure 1.10.1. To see better what R looks like, we can find $F \in \mathrm{Isom}(\mathbb{C})$ such that

$$F(R) = R' = G_4(A', B', C', D') = G_4(0, a, C', ib), \quad a, b > 0, \tag{1.10.5}$$

whose sides include L_0', from 0 to a, and L_3', from ib to 0. The line segment L_1' connects a to C' and L_2' connects C' to ib. For the angles at $B' = a$ and at $D' = ib$ to be right angles, we require

$$L_1' \perp L_0' \text{ and } L_2' \perp L_3', \tag{1.10.6}$$

i.e.,

$$C' - a \in i\mathbb{R} \text{ and } C' - ib \in \mathbb{R}. \tag{1.10.7}$$

This forces

$$\mathrm{Re}\, C' = a,\ \mathrm{Im}\, C' = b, \quad \text{hence } C' = a + ib. \tag{1.10.8}$$

We have specified R'. Note that the calculations (1.10.6)–(1.10.8) only require the angles at A', B', and D' to be right angles. This forces C' to be a right angle. This is consistent with the case $n = 4$ of (1.10.4).

One consequence of (1.10.8) is the following.

Proposition 1.10.1. *Opposite sides of a rectangle $R = G_4(A, B, C, D)$ have equal lengths, i.e.,*

$$|A - B| = |D - C| \text{ and } |A - D| = |B - C|. \tag{1.10.9}$$

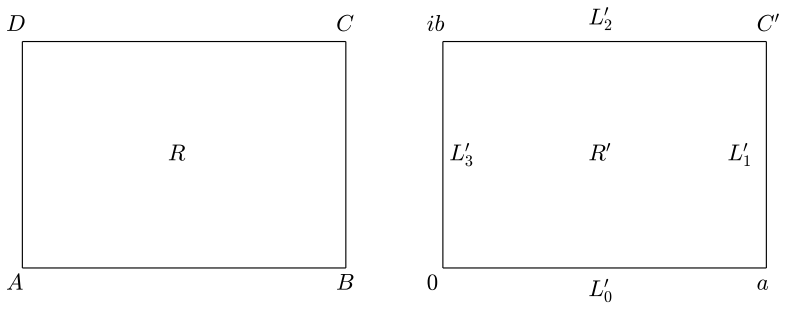

Figure 1.10.1. Rectangles $R = G_4(A, B, C, D)$, and R'

With a and b as in (1.10.5), we say the rectangle R has sidelengths a and b.

A square is a rectangle S whose sidelengths are all equal, say to c. In such a case, calculations parallel to (1.10.5)–(1.10.8) yield a transformation $F \in \mathrm{Isom}(\mathbb{C})$ such that

$$F(S) = S' = G_4(0, c, c + ic, ic). \tag{1.10.10}$$

Translating by $-c(1 + i)/2$ yields a square S'', with vertices

$$-\frac{c}{2}(1 + i), \quad \frac{c}{2}(1 - i), \quad \frac{c}{2}(1 + i), \quad \frac{c}{2}(-1 + i), \tag{1.10.11}$$

which is clearly invariant under $\rho_i(z) = iz$, yielding for the original square $S = G_4(A, B, C, D)$ a rotational symmetry about its center $(A + B + C + D)/4 = p$, i.e., S is invariant under $F_p(z) = i(z - p) + p$.

We next introduce a family of 4-gons more general than the class of rectangles. We say $P = G_4(A, B, C, D)$ is a parallelogram provided its opposite sides are parallel, i.e.,

$$D - A \,\|\, C - B, \quad B - A \,\|\, C - D. \tag{1.10.12}$$

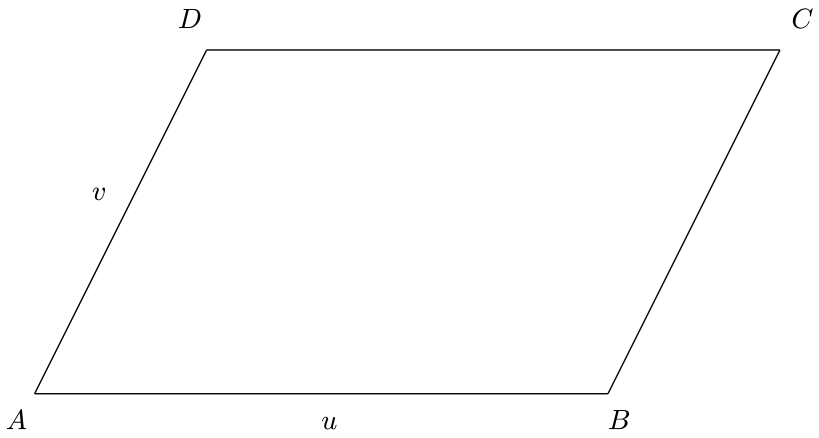

Figure 1.10.2. Parallelogram

See Figure 1.10.2. In such a case, through D there is a unique line L_2, parallel to the line through A and B, and through B there is a unique line L_1, parallel to the line through A and D. Equivalently, if we set

$$u = B - A, \quad v = D - A, \tag{1.10.13}$$

then

$$L_2 = \{D + tu : t \in \mathbb{R}\}, \quad L_1 = \{B + sv : s \in \mathbb{R}\}. \tag{1.10.14}$$

We assume A, B, and D are not collinear, hence

$$\frac{u}{v} \notin \mathbb{R}. \tag{1.10.15}$$

Lemma 1.10.2. *In the setting described above, the lines L_1 and L_2 meet at the point $A + u + v$.*

Proof. Where L_1 and L_2 meet, we have

$$A + v + tu = A + u + sv, \tag{1.10.16}$$

hence

$$(t-1)u = (s-1)v. \tag{1.10.17}$$

Now (1.10.15) implies that $t = 1$ and $s = 1$, as asserted. □

In summary, we have

Proposition 1.10.3. *If $P = G_4(A, B, C, D)$ is a parallelogram, with u, v given by* (1.10.13)*, then*

$$C = A + u + v. \tag{1.10.18}$$

As a consequence, opposite sides of P have the same length.

Corollary 1.10.4. *In the setting of Proposition 1.10.3, the parallelogram P is invariant under the map $F \in \mathrm{Isom}(\mathbb{C})$ given by*

$$F(z) = -(z-p) + p, \quad p = \frac{1}{4}(A + B + C + D). \tag{1.10.19}$$

Consequently, opposite vertices of P have the same angular measurements:

$$\Omega_P(A) = \Omega_P(C), \quad \Omega_P(B) = \Omega_P(D). \tag{1.10.20}$$

Remark. If $|u| = |v|$ in (1.10.13), then all four sides of P have equal length. In such a case, one says P is a rhombus.

There is a result, called the parallelogram law, equating the sum of the squares of the sidelengths of a parallelogram P and the sum of the squares of the lengths of the diagonals of P, i.e., the line segments from A to C and from B to D. As for the latter, note that

$$|A - C| = |u + v|, \quad |B - D| = |u - v|. \tag{1.10.21}$$

To derive the asserted identity, start with

$$\begin{aligned} |u+v|^2 &= |u|^2 + |v|^2 + 2\langle u, v\rangle, \\ |u-v|^2 &= |u|^2 + |v|^2 - 2\langle u, v\rangle, \end{aligned} \tag{1.10.22}$$

and add, to get

$$|u+v|^2 + |u-v|^2 = 2|u|^2 + 2|v|^2. \tag{1.10.23}$$

This is the parallelogram law.

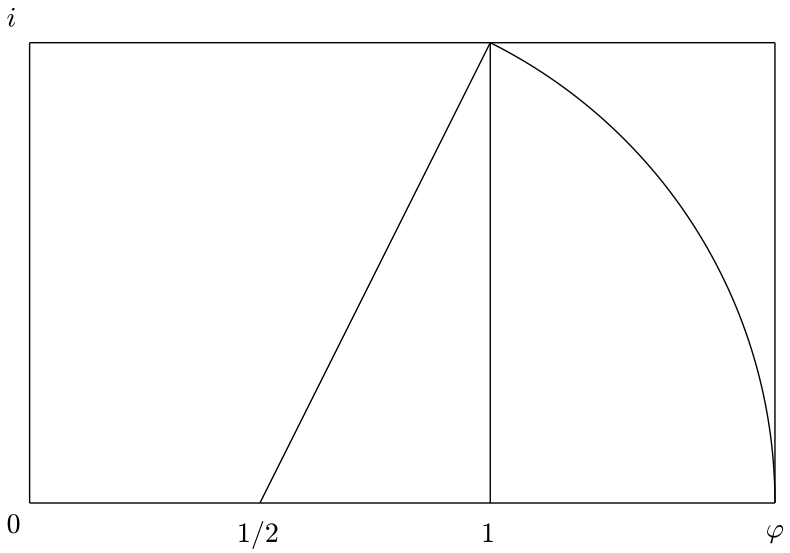

Figure 1.10.3. Golden rectangle

Exercises

1. Let R_j be rectangles, with sidelengths a_j, b_j $(a_j \leq b_j)$. Show that R_1 is similar to R_2 if and only if

$$\frac{a_1}{b_1} = \frac{a_2}{b_2}.$$

2. Let R be a rectangle of sidelengths x and $x + y$ $(x, y > 0)$. We can partition R into two pieces, a square of sidelengths x, and a rectangle $\widetilde{R}$ of sidelengths x and y. We say R is a *golden rectangle* if R is similar to $\widetilde{R}$, i.e.,

$$\frac{x}{x+y} = \frac{y}{x}.$$

Show that, if this identity holds,

$$x = 1 \Rightarrow y^2 + y = 1 \Rightarrow y = \frac{\sqrt{5}-1}{2}$$
$$\Rightarrow x + y = \frac{\sqrt{5}+1}{2}.$$

See Figure 1.10.3. The number

$$\varphi = \frac{\sqrt{5}+1}{2}$$

is called the *golden ratio*. Note that

$$\frac{\sqrt{5}}{2} = \left|\frac{1}{2} + i\right|.$$

See (2.5.8)–(2.5.14) for how the golden ratio arises in the construction of a regular pentagon. See [**10**] for further results on the golden ratio.

3. After looking at the next section, decipher the meaning of the circular arc in Figure 1.10.3.

4. Find the points of intersection of the circle and parabola

$$x^2 + y^2 = 1, \quad y = x^2.$$

Relate the y-coordinate of the intersection points with the golden ratio.

1.11. Circles

A circle in $\mathbb{C}$, centered at a and of radius r, is a set of the form

$$S_r(a) = \{z \in \mathbb{C} : |z - a| = r\}. \tag{1.11.1}$$

The unit circle S^1 we have encountered before is $S^1 = S_1(0)$. Note that

$$z \in \mathbb{C},\ z \neq 0 \Rightarrow \omega = \frac{z}{|z|} \in S^1 \Rightarrow a + r\omega \in S_r(a). \tag{1.11.2}$$

The circle $S_r(a)$ is the boundary of the *disk*

$$D_r(a) = \{z \in \mathbb{C} : |z - a| \leq r\}. \tag{1.11.3}$$

It is of great interest to know when two circles, say $S_R(a)$ and $S_r(b)$, intersect. See Figure 1.11.1 for an illustration. We have the following result.

Proposition 1.11.1. *Given $a \neq b \in \mathbb{C},\ R, r > 0$, the following are equivalent.*

(a) $S_R(a) \cap S_r(b) \neq \emptyset$,

(b) $|R - r| \leq |a - b| \leq R + r$.

If one has strict inequality in (b), the intersection in (a) has two points. If one of the inequalities in (b) is not strict, the intersection in (a) has one point.

Proof. The demonstration is essentially a repeat of the proof of Proposition 1.7.5, with different notation (cf. (1.7.43)). If we bring in the isometry

$$F(z) = \overline{\omega}(z - a), \quad \omega = \frac{b - a}{|b - a|}, \tag{1.11.4}$$

the image circles are $S_R(0)$ and $S_r(d)$, $d = |b - a|$, and, provided $|R - r| \leq d \leq R + r$, they intersect at points $x \pm iy$, given by computations parallel to those done in the proof of Proposition 1.7.5. Then $S_R(a)$ and $S_r(b)$ intersect at $F^{-1}(x \pm iy)$. If one of the inequalities in (b) is not strict, $y = 0$. □

Remark. Constructions involving using points where two circles (drawn with a compass) intersect play a fundamental role in classical Euclidean geometry, though this aspect is not particularly emphasized here, except in §2.9 on Euclidean numbers.

The Euclideans do not seem to have proved a result like Proposition 1.11.1. As we see from the proof of Proposition 1.7.5, a key ingredient in the proof is the ability to take the square root of a positive number and

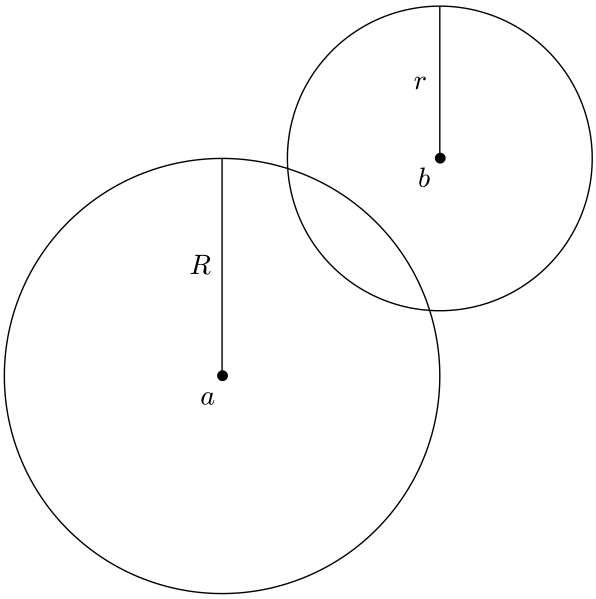

Figure 1.11.1. Intersecting circles

know that it defines a positive real number. Such a foundational issue is treated in Chapter 1 of [**15**], in the development of the real number system. See also Proposition 2.11.5 and subsequent results in §2.11 of this text.

There is a parallel result for the intersection of a circle $S_R(a)$ and a line L. Given such objects, let $L_a^\perp$ denote the line through a that intersects L orthogonally, given by Proposition 1.4.4. Say L and $L_a^\perp$ intersect at b. Then

$$\begin{aligned} &|a-b| < R \Rightarrow S_R(a) \cap L \text{ has two points,} \\ &|a-b| = R \Rightarrow S_R(a) \cap L \text{ has one point,} \\ &|a-b| > R \Rightarrow S_R(a) \cap L \text{ is empty.} \end{aligned} \tag{1.11.5}$$

We leave the demonstration to the reader.

We next describe an interesting family of right triangles, inscribed in a circle, illustrated in Figure 1.11.2.

Proposition 1.11.2. *Pick $v \in S^1$, $v \neq \pm 1$. Then $\mathcal{T} = \triangle(-1, 1, v)$ is a right triangle, with right angle at v.*

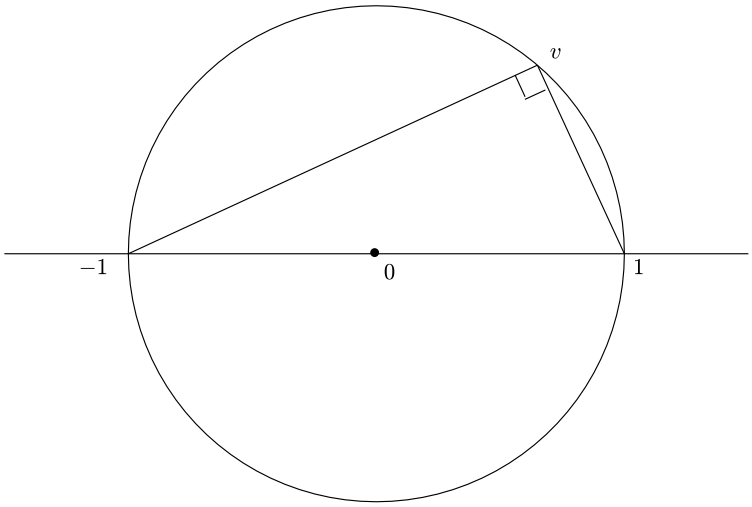

Figure 1.11.2. Right triangle in a circle

Proof. It is equivalent to show that $v+1 \perp v-1$, if $|v|=1$. Indeed,

$$\begin{aligned} \langle v+1, v-1\rangle &= |v|^2 + \langle 1, v\rangle - \langle v, 1\rangle - 1 \\ &= |v|^2 - 1 = 0. \end{aligned} \tag{1.11.6}$$

□

The computation (1.11.6) shows that if $v \in S^1, v \neq 1$, then $(v+1)/(v-1) \in i\mathbb{R}$ (cf. Proposition 1.2.5). Equivalently, if we set

$$\psi(v) = \frac{v+1}{v-1}\, i, \tag{1.11.7}$$

then

$$\psi : S^1 \setminus \{1\} \longrightarrow \mathbb{R}. \tag{1.11.8}$$

For another check of this, note that if $|v|=1,\ v \neq 1$,

$$\frac{v+1}{v-1} = \frac{(v+1)(\overline{v}-1)}{(v-1)(\overline{v}-1)} = -2i\,\frac{\operatorname{Im} v}{|v-1|^2}. \tag{1.11.9}$$

Furthermore, if we set

$$\varphi(x) = \frac{x+i}{x-i}, \tag{1.11.10}$$

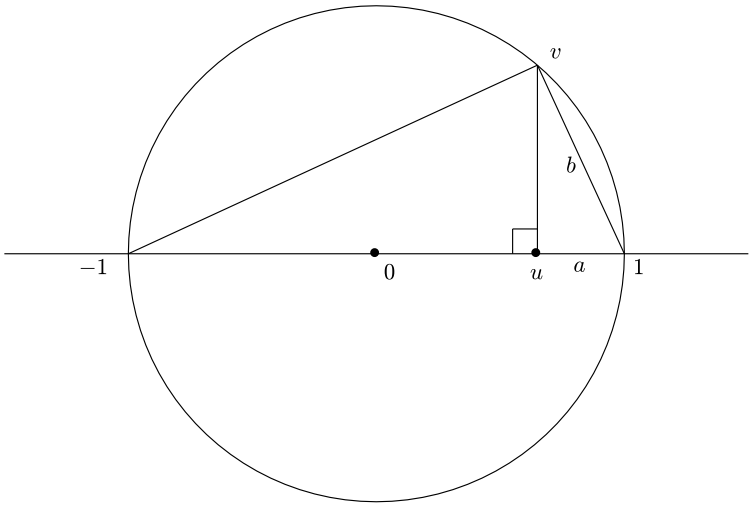

Figure 1.11.3. Geometric construction of $b = \sqrt{2a}$

then

$$\varphi : \mathbb{R} \longrightarrow S^1 \setminus \{1\} \text{ is one-to-one and onto,} \tag{1.11.11}$$

with inverse ψ in (1.11.7)–(1.11.8).

We next describe a geometrical approach to the construction of the square root of a given positive number, illustrated in Figure 1.11.3. To start, assume $a \in (0, 1)$, and set $u = 1 - a \in (0, 1)$. Let v denote the point in S^1_+ where the line orthogonal to $\mathbb{R}$ through u intersects S^1_+. Explicitly,

$$v = u + iy,\ y > 0, \quad u^2 + y^2 = 1, \tag{1.11.12}$$

so

$$y = \sqrt{1 - u^2}. \tag{1.11.13}$$

Proposition 1.11.3. *In the setting described above,*

$$b = |v - 1| \Longrightarrow b = \sqrt{2a}. \tag{1.11.14}$$

Proof. Since $\triangle(v, u, 1)$ is a right triangle with right angle at u, we have

$$\begin{aligned} b^2 &= y^2 + a^2 \\ &= 1 - u^2 + a^2 \\ &= 1 - (1-a)^2 + a^2 \\ &= 2a. \end{aligned} \tag{1.11.15}$$

Second proof. The triangles $\triangle(v, u, 1)$ and $\triangle(-1, v, 1)$ are both right triangles (thanks to Proposition 1.11.2), and they have a common angle at 1. Hence they are similar. Similarity implies

$$\frac{b}{2} = \frac{a}{b}, \tag{1.11.16}$$

which in turn yields the conclusion in (1.11.14). □

More generally, given $r > 0$, pick $n \in \mathbb{N}$ such that

$$a = \frac{r}{2n^2} \in (0, 1). \tag{1.11.17}$$

Then the construction above yields $b = \sqrt{2a}$, giving $b = \sqrt{r}/n$, or

$$\sqrt{r} = nb. \tag{1.11.18}$$

We return to the functions ψ and φ, arising in (1.11.7)–(1.11.11), and derive a result with a number-theoretic flavor. Namely, let $\mathbb{Q} \subset \mathbb{R}$ denote the field of rational numbers, and set

$$\begin{aligned} \mathbb{Q}[i] &= \{x + iy : x, y \in \mathbb{Q}\}, \\ S^1_{\mathbb{Q}} &= S^1 \cap \mathbb{Q}[i] = \{x + iy : x, y \in \mathbb{Q}, x^2 + y^2 = 1\}. \end{aligned} \tag{1.11.19}$$

It follows from the formulas (1.11.7) and (1.11.10), and the result (1.11.11) that

$$\varphi : \mathbb{Q} \longrightarrow S^1_{\mathbb{Q}} \setminus \{1\} \text{ is one-to-one and onto,} \tag{1.11.20}$$

with inverse given by

$$\psi : S^1_{\mathbb{Q}} \setminus \{1\} \longrightarrow \mathbb{Q}. \tag{1.11.21}$$

Note also that

$$x > 0 \Longleftrightarrow \varphi(x) \in S^1_+. \tag{1.11.22}$$

These results apply to the study of Pythagorean triples, which consist of triples (j, k, ℓ) of positive integers satisfying

$$j^2 + k^2 = \ell^2. \tag{1.11.23}$$

We have the following.

Proposition 1.11.4. *For each Pythagorean triple (j,k,ℓ), there is a unique $m/n \in \mathbb{Q}^+$ such that*

$$\varphi\Big(\frac{m}{n}\Big) = \frac{j}{\ell} + \frac{k}{\ell}i. \tag{1.11.24}$$

Conversely, if $m/n \in \mathbb{Q}^+$, then (1.11.24) *yields a Pythagorean triple (up to a sign on j).*

Since

$$\varphi(x) = \frac{x^2-1}{x^2+1} + \frac{2x}{x^2+1}i, \tag{1.11.25}$$

we have

$$\varphi\Big(\frac{m}{n}\Big) = \frac{m^2-n^2}{m^2+n^2} + \frac{2mn}{m^2+n^2}i, \tag{1.11.26}$$

yielding

$$j = m^2 - n^2, \quad k = 2mn, \quad \ell = m^2 + n^2. \tag{1.11.27}$$

Circles and triangles – inscribed and circumscribed

Let $A, B, C \in \mathbb{C}$ be non-collinear points, hence defining a triangle $\mathcal{T} = \triangle(A,B,C)$. As we will see, there is a unique circle $\mathcal{C}$ through these three points, which is hence said to circumscribe $\mathcal{T}$; alternatively, we say $\mathcal{T}$ is inscribed in $\mathcal{C}$. To find $\mathcal{C}$, we start by looking for all circles that pass through A and B. This family is described as follows.

Proposition 1.11.5. *Let c denote the midpoint of the line segment L_{AB} from A to B, and let $L_c^\perp$ denote the line through c orthogonal to L_{AB}. Pick $p \in L_c^\perp$ and set*

$$r = |p - A| = |p - B|. \tag{1.11.28}$$

Then the circle $S_r(p)$ passes through A and B. All circles containing A and B arise in this fashion.

Proof. Exercise. Use symmetry to show that $|p - A| = |p - B|$. □

To proceed, let b denote the midpoint of the line segment L_{AC} from A to C, and let $L_b^\perp$ denote the line through b orthogonal to L_{AC}. Since L_{AB} and L_{AC} are not parallel, neither are $L_c^\perp$ and $L_b^\perp$. Say

$$p = L_c^\perp \cap L_b^\perp. \tag{1.11.29}$$

Then r in (1.11.28) also satisfies

$$r = |p - A| = |p - C|. \tag{1.11.30}$$

Hence $S_r(p)$ passes through A, B, and C.

Doing the construction above yields a bonus. Namely, we have

$$|p - A| = |p - B| = |p - C| = r, \tag{1.11.31}$$

and the second identity implies

$$p \in L_a^\perp, \tag{1.11.32}$$

where a is the midpoint of the line segment L_{BC} and $L_a^\perp$ is the line through a orthogonal to L_{BC}.

Proposition 1.11.6. *Given a triangle $\mathcal{T} = \triangle(A, B, C)$, the perpendicular bisectors $L_a^\perp, L_b^\perp$, and $L_c^\perp$ of the sides of $\mathcal{T}$ all meet at a point, namely at the center of the circle that circumscribes $\mathcal{T}$.*

We next show that, given a triangle $\mathcal{T} = \triangle(A, B, C)$, there is a unique circle $\mathcal{C}$ that touches each side of $\mathcal{T}$ at just one point; we say $\mathcal{C}$ is tangent to each side of $\mathcal{T}$, and that $\mathcal{C}$ is inscribed in $\mathcal{T}$.

To start, we look at the problem of finding the circle $\mathcal{C}$, centered at p, tangent to a line L, given $p \notin L$. To produce $\mathcal{C}$, simply drop a perpendicular $L_p^\perp$ from p to L, and say $L_p^\perp$ intersects L at q. Then

$$\mathcal{C} = S_r(p), \quad r = |p - q|. \tag{1.11.33}$$

To continue, we look for those circles tangent to the two sides L_{AB} and L_{AC} of $\mathcal{T}$. Symmetry considerations imply that the center of such a circle lies on the angle bisector ℓ_A of the angle of $\mathcal{T}$ at A. Similarly, if ℓ_B is the angle bisector of $\mathcal{T}$ at B, and if

$$p = \ell_A \cap \ell_B, \tag{1.11.34}$$

then there is a circle centered at p, tangent to L_{AB} and to L_{AC}, and there is also a circle, centered at p, tangent to L_{AB} and to L_{BC}. Since both circles are tangent to L_{AB}, they must coincide. Thus we have a circle, centered at p, given by (1.11.34), that is tangent to each of the three sides of $\mathcal{T} = \triangle(A, B, C)$. This is the circle inscribed in $\mathcal{T}$.

Such a construction of inscribed circles also yields a bonus. Namely, if this inscribed circle is $S_r(p)$, then we must have

$$p \in \ell_C, \tag{1.11.35}$$

where ℓ_C denotes the angle bisector of $\mathcal{T}$ at C.

Proposition 1.11.7. *Given a triangle $\mathcal{T} = \triangle(A, B, C)$, the angle bisectors of $\mathcal{T}$ at its three vertices all meet at a point, namely at the center of the circle inscribed in $\mathcal{T}$.*

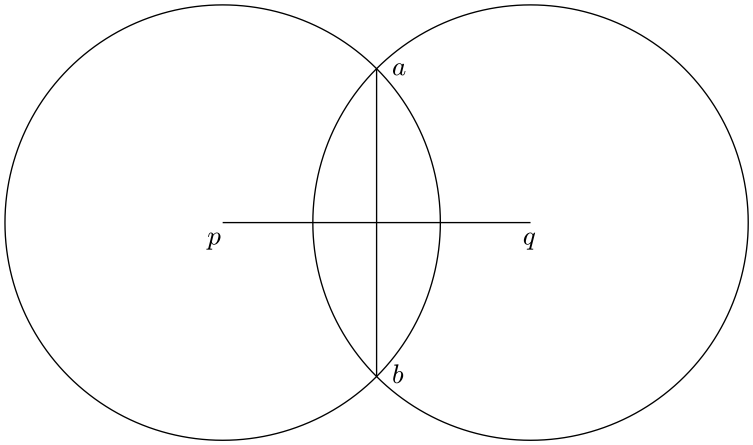

Figure 1.11.4. Perpendicular bisector of $\mathcal{L}_{[p,q]}$

Exercises

1. Given a line segment $\mathcal{L}_{[p,q]}$, take $r > |p-q|/2$ and consider the circles $S_r(p)$ and $S_r(q)$. Show that

$$S_r(p) \cap S_r(q) = \{a, b\},$$

a set of two distinct points in $\mathbb{C}$. Show that

$$\mathcal{L}_{[a,b]} \cap \mathcal{L}_{[p,q]} = \frac{1}{2}(p+q) = \frac{1}{2}(a+b).$$

Show also that

$$q - p \perp a - b.$$

See Figure 1.11.4.
Hint. Apply an element of $\mathrm{Isom}(\mathbb{C})$ to show that it suffices to assume $p, q \in \mathbb{R}$ and $p = -q$. Use symmetries.

2. Give a demonstration of the assertion in (1.11.5), regarding when a circle and a line intersect.

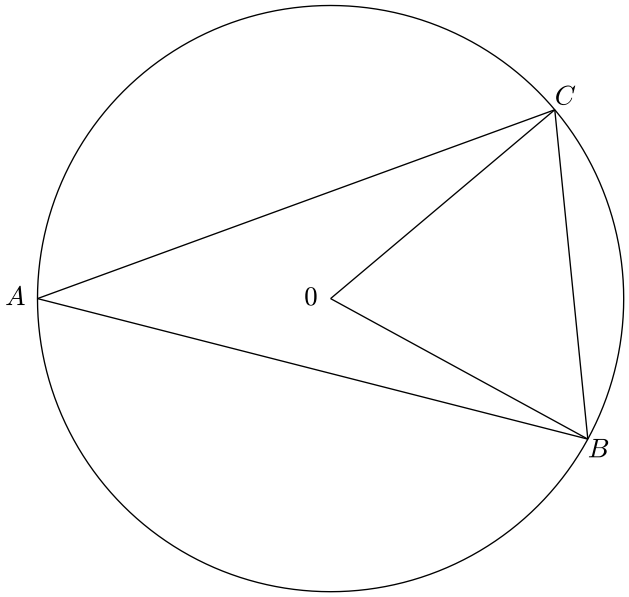

Figure 1.11.5. Another angle bisection

3. Consider, as in Proposition 1.11.2, the triangle $\mathcal{T} = \triangle(-1, 1, v)$. Assume $v \in S^1_+$. Show that

$$\Omega_{\mathcal{T}}(-1) = \frac{v+1}{|v+1|},$$

and deduce from Proposition 1.5.1 that

$$\Omega_{\mathcal{T}}(-1) = v^{1/2}.$$

4. Let A, B, C be distinct points on S^1. Set

$$\mathcal{T} = \triangle(A, B, C), \quad \mathcal{T}_0 = \triangle(0, B, C), \text{ if also } B \neq -C.$$

Show that

$$\Omega_{\mathcal{T}}(A)^2 = \Omega_{\mathcal{T}_0}(0).$$

See Figure 1.11.5 for an illustration. In case $B = -C$, show that

$$\Omega_{\mathcal{T}}(A)^2 = -1.$$

Show that this case yields another proof of Proposition 1.11.2. For another perspective on this exercise, see Exercise 9 of §2.4.

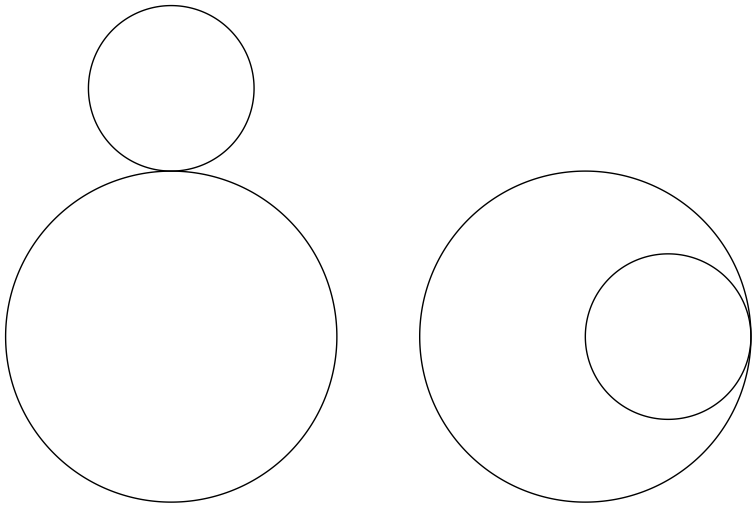

Figure 1.11.6. Two pairs of tangent circles

5. If two circles, $S_R(a)$ and $S_r(b)$, intersect at exactly one point, say p, we say these circles are tangent at p. Show that if this happens, p lies on the line through a and b. See Figure 1.11.6 for an illustration of tangent circles.

Chapter 2

Bringing in calculus

Once circles enter the picture, a proper treatment of Euclidean geometry requires substantially more sophistication. The thing about circles, as opposed to lines, triangles, and other polygons, is that they are, well, *curved*. This makes the computation of various geometric quantities associated to circles more subtle. For example, the length of a line segment is simply the distance between its endpoints. The total length of the perimeter of a polygon is the sum of the lengths of the line segments from one vertex to the next. How about the length of a circle? One can ask similar questions about areas.

These problems were tackled by Archimedes, who approached the computations as limits of a sequence of approximations obtained by replacing a circle by a polygon with many sides. This was the sort of brilliant insight that had the potential to lead to a revolution in mathematics. However, Archimedes was too far ahead of his time, and the revolution had to wait for centuries.[1] The revolution was calculus.

Section 2.1 gives a short presentation of basic calculus. We define the derivative and the integral, and discuss the fundamental theorem of calculus, connecting these two concepts. We also treat such key results as the mean value theorem, the inverse function theorem, and the change of variable formula for the integral, as well as functions given by power series and their derivatives. This section is in the way of a review, since we assume the reader has had a course in calculus.

[1]Perhaps one should say his time was not right. Archimedes lived near the end of a golden age of Mediterranean civilization. Under the boot of the Roman empire, higher learning withered in that part of the world. Archimedes' work itself was almost lost.

Section 2.2 uses methods of calculus to treat curves and arclength. We derive the arclength formula

$$\ell(\gamma) = \int_a^b |\gamma'(t)|\, dt, \tag{2.0.1}$$

for an appropriate class of curves $\gamma : [a, b] \to \mathbb{C}$. We discuss parametrizing a curve by arclength. When this is applied to the unit circle, one obtains the parametrization

$$\operatorname{cis} : \mathbb{R} \longrightarrow \mathbb{C}, \quad \operatorname{cis}(t) = \cos t + i \sin t, \tag{2.0.2}$$

bringing in the classical trigonometric functions cos and sin. We also define the real measurement of angles, in radians. Recall from §1.6 the definition of the angle measurement $\Omega_{\mathcal{T}}(A) = \omega \in S^1$, when $\mathcal{T} = \triangle(A, B, C)$ is a triangle with vertices A, B, and C. This construction is purely algebraic. The real angle measurement $\angle_{\mathcal{T}}(A)$ is the real number in $(-\pi, \pi)$ such that

$$\Omega_{\mathcal{T}}(A) = \operatorname{cis} \angle_{\mathcal{T}}(A). \tag{2.0.3}$$

In §2.3 we introduce the exponential function e^z, defined by a power series, valid for all $z \in \mathbb{C}$, so as to solve a basic differential equation, $(d/dt)e^{at} = ae^{at}$, given $a \in \mathbb{C}$. We show that, for $t \in \mathbb{R}$,

$$e^{it} = \operatorname{cis}(t), \tag{2.0.4}$$

which is the famous Euler formula. This leads to a characterization of $\pi = \ell(S^1)/2$ as the smallest positive number such that

$$e^{\pi i} = -1. \tag{2.0.5}$$

We show that, for a triangle, the angle sum is π. This is a more sophisticated version of the identity (1.0.3).

Section 2.4 pursues further results on Euler's formula, trigonometric functions, and π, including numerical evaluation of π.

Section 2.5 studies regular polygons, specifying the vertices of a regular n-gon inscribed in a circle using the function e^{it}. We also derive an algebraic formula for the vertices of a regular pentagon.

Section 2.6 introduces the concept of area of a set $S \subset \mathbb{C}$. We define this in terms of taking tilings of $\mathbb{C}$, counting how many tiles are contained in S, and how many tiles it takes to contain S, and passing to the limit as the size of the tiles vanishes. We show that Area S is invariant under the application of $F \in \operatorname{Isom}(\mathbb{C})$. The analysis brings in a popular approach to a proof of the Pythagorean theorem and turns it on its head. We obtain an integral formula for the area of a set of points in $\mathbb{C}$ bounded by the graphs of two continuous functions, and show that if D_1 is the unit disk, then

$$\operatorname{Area} D_1 = \pi. \tag{2.0.6}$$

Going further, we consider the pie slices $\Pi_t = \{re^{is} : 0 \le r \le 1,\ 0 \le s \le t\}$, and show that, for $0 < t \le 2\pi$,

$$\text{Area}\,\Pi_t = \frac{1}{2}t. \tag{2.0.7}$$

We give two proofs of this, and relate the second proof to methods used by Archimedes to evaluate π.

Section 2.7 derives a formula for the area of a triangle, strictly in terms of its sidelengths, known as Heron's formula.

In §2.8 we see how calculations in Sections 2.5 and 2.6 can be used to produce a trig table.

Section 2.9 discusses the set $\mathbb{E} \subset \mathbb{C}$ of "Euclidean numbers," designed to consist of those points in $\mathbb{C}$ that can be constructed via compass and straightedge. It is seen that $\mathbb{E}$ is a field, satisfying

$$z \in \mathbb{E},\ w \in \mathbb{C},\ w^2 = z \Longrightarrow w \in \mathbb{E}. \tag{2.0.8}$$

Also, $\mathbb{E}$ is countable.

In §2.10, we study the class of linear fractional transformations. We apply this to an approach to Apollonius' three circles problem: given three disjoint circles $C_j \subset \mathbb{C}$, no one of which separates the other two, find a fourth circle S tangent to each C_j. In this section we bring in the concept of a 2×2 matrix, $A \in M(2, \mathbb{C})$, which acts on a 2-vector:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad v = \begin{pmatrix} z \\ w \end{pmatrix}, \quad Av = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} az + bw \\ cz + dw \end{pmatrix}. \tag{2.0.9}$$

We have a matrix multiplication: $AB = C$, defined so that $Cv = A(Bv)$.

In §2.11, we establish the fundamental theorem of algebra, which says that if $p(z)$ is a non-constant polynomial, then $p(z)$ vanishes for some $z \in \mathbb{C}$. We use a proof of Argand, who was one of the first to represent $\mathbb{C}$ using the Euclidean plane.

Section 2.12 studies Newton's equation for planetary motion, as an ODE of the form

$$\frac{d^2z}{dt^2} = -\frac{K}{|z|^3}z, \tag{2.0.10}$$

for a curve in $\mathbb{C}$. We treat this by setting

$$z(t) = r(t)e^{i\theta(t)}, \tag{2.0.11}$$

and studying the resulting pair of equations for r and θ. We obtain a reduction via conservation of angular momentum and then show that each solution satisfies an identity

$$r(1 + e\cos(\theta - \theta_0)) = p. \tag{2.0.12}$$

We introduce the classes of curves known as ellipses, parabolas, and hyperbolas, and show that (2.0.12) specifies one of these classes of curves, depending on e.

We proceed to explore properties of these classes of curves. We see that after applying a rigid motion each such curve can be given in the respective form

$$\begin{aligned} &\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \\ &ax^2 - y = 0, \\ &\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \end{aligned} \tag{2.0.13}$$

Amongst other things, we consider arclength of such curves. For parabolas one gets elementary integrals, but not for ellipses (except circles) nor for hyperbolas. For an additional perspective on hyperbolas, in particular the case

$$x^2 - y^2 = 1, \tag{2.0.14}$$

we introduce the Minkowski plane $\mathbb{M}$, whose elements have the form

$$z = x + \iota y, \quad x, y \in \mathbb{R}, \tag{2.0.15}$$

where now $\iota^2 = 1$. This is seen to be a ring, equipped with an inner product $\langle z, w\rangle_M$, satisfying

$$\langle z, z\rangle_M = x^2 - y^2. \tag{2.0.16}$$

We see that

$$\gamma(t) = \cosh t + \iota \sinh t \tag{2.0.17}$$

is a curve tracing out the right half of the hyperbola (2.0.14), and

$$\langle \gamma'(t), \gamma'(t)\rangle_M \equiv -1. \tag{2.0.18}$$

We say this is a unit-speed timelike curve.

Section 2.13 gives a brief taste of the algebraic approach to Euclidean geometry in higher dimensions. Topics include the notion of volume of sets in $\mathbb{R}^n$, and areas of $(n-1)$-dimensional surfaces in $\mathbb{R}^n$, with applications to the area of spheres S^{n-1} and volumes of balls in $\mathbb{R}^n$. We contrast the algebraic approaches taken in dimension 2 and in dimension $n > 2$. On $\mathbb{R}^n$, we bring in a matrix product,

$$Ax = y, \tag{2.0.19}$$

where $x, y \in \mathbb{R}^n$ and $A \in M(n, \mathbb{R})$ is an $n \times n$ matrix. This contrasts with the use of the product in $\mathbb{C}$ in case of dimension 2. We see a synthesis in case $n = 4$, involving the algebra of quaternions.

This chapter ends with Appendix 2.A, presenting basic background material on matrices, such as arise in Sections 2.10 and 2.13.

2.1. Outline of calculus

Calculus has two facets, differential calculus and integral calculus, tied together by the fundamental theorem of calculus. Differential calculus studies the rate of change of a function $f(x)$, given in terms of the derivative,

$$f'(x) = \lim_{h\to 0} \frac{1}{h}(f(x+h) - f(x)). \tag{2.1.1}$$

We say f is differentiable at x if this limit exists. We also write $df/dx = f'(x)$.

Integral calculus studies the accumulated quantity over an interval $[a, b]$ associated to a function g, given in terms of the integral,

$$\int_a^b g(x)\,dx = \lim_{N\to\infty} \sum_{k=1}^{N} g(x_k)(x_k - x_{k-1}), \tag{2.1.2}$$

where $x_k = a + (k/N)(b-a)$, given that $g : [a, b] \to \mathbb{C}$ is continuous. To say g is continuous at x is to say that

$$x_k \to x \Longrightarrow g(x_k) \to g(x). \tag{2.1.3}$$

Here we describe some basic results on these concepts of calculus. For details, one can see Chapter 4 of [**15**].

First, a quick application of the definition (2.1.1) shows that, for

$$p_1(x) = x, \quad p_2(x) = x^2, \tag{2.1.4}$$

we have

$$p_1'(x) = 1, \quad p_2'(x) = 2x. \tag{2.1.5}$$

Next, one verifies from the definition that if f and g are differentiable,

$$\frac{d}{dx}(fg)(x) = f'(x)g(x) + f(x)g'(x). \tag{2.1.6}$$

Applying this to (2.1.4)–(2.1.5) yields inductively that

$$\frac{d}{dx}x^n = nx^{n-1}, \tag{2.1.7}$$

for $n \in \mathbb{N}$. In addition, a direct check of (2.1.1) yields

$$\frac{d}{dx}x^{-1} = -x^{-2}, \quad \text{for } x \neq 0, \tag{2.1.8}$$

and an argument using (2.1.6) then extends (2.1.7) to $n \in \mathbb{Z}$ (assuming $x \neq 0$ if $n < 0$). Another important general result is the *chain rule*:

$$h(x) = f(g(x)) \Rightarrow h'(x) = f'(g(x))g'(x). \tag{2.1.9}$$

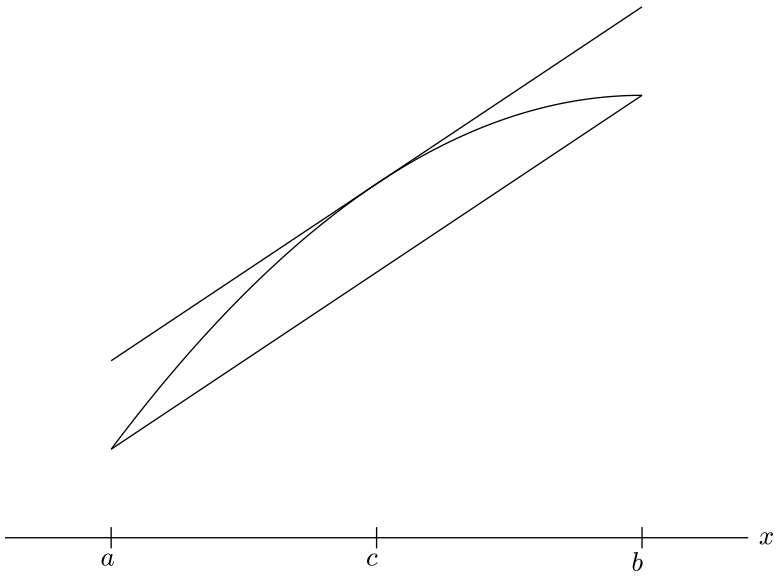

Figure 2.1.1. Mean value theorem

One major theoretical result of differential calculus is the following, known as the mean value theorem.

Proposition 2.1.1. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Assume f is differentiable on (a, b). Then there exists $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{2.1.10}$$

This result is illustrated in Figure 2.1.1, showing the line $y = f'(c)(x - c) + f(c)$, through the point on the graph of $y = f(x)$ at $x = c$, parallel to the line through the points $(a, f(a))$ and $(b, f(b))$.

The mean value theorem plays a major role in establishing the following result, known as the inverse function theorem.

Proposition 2.1.2. *Let $f : [a, b] \to \mathbb{R}$ be continuous, and differentiable on (a, b). Say $f(a) = \alpha$, $f(b) = \beta$. Assume that for $x \in (a, b)$, $f'(x) > c > 0$. Then $f : [a, b] \to [\alpha, \beta]$ is one-to-one and onto, with inverse*

$$\varphi : [\alpha, \beta] \longrightarrow [a, b], \tag{2.1.11}$$

and φ is differentiable at each point of (α, β).

Under the hypotheses given above, the fact that $f(x_1) < f(x_2)$ for $x_1 < x_2$ follows directly from the mean value theorem, hence f is one-to-one. The fact that f is onto is discussed further in Proposition 2.11.5 and its corollary.

Combining Proposition 2.1.2 with the chain rule gives

$$\varphi'(f(x))f'(x) = 1, \quad \forall\, x \in (a,b). \tag{2.1.12}$$

One consequence is that we have

$$p_r : (0,\infty) \longrightarrow (0,\infty), \tag{2.1.13}$$

with $r = 1/n$, inverting $p_n(x) = x^n$. We write

$$p_r(x) = x^r, \tag{2.1.14}$$

with $r = 1/n$. Use of (2.1.12) gives

$$\frac{d}{dx}x^r = rx^{r-1}, \quad x > 0, \tag{2.1.15}$$

for $r = 1/n$. Then we can define (2.1.14) for $r = m/n \in \mathbb{Q}$, and again obtain (2.1.15). Further extension to $r \in \mathbb{R}$ is conveniently accomplished using the exponential function, which will arise in §2.3.

We move to integral calculus, starting with the fundamental theorem of calculus. This has two parts.

Proposition 2.1.3. *Let $f : [a,b] \to \mathbb{C}$ be continuous, and set*

$$g(x) = \int_a^x f(y)\,dy. \tag{2.1.16}$$

Then g is differentiable for $x \in (a,b)$, and

$$g'(x) = f(x). \tag{2.1.17}$$

The proof consists of writing

$$\frac{1}{h}(g(x+h) - g(x)) = \frac{1}{h}\int_x^{x+h} f(y)\,dy, \tag{2.1.18}$$

and taking $h \to 0$.

Proposition 2.1.4. *Let $g : [a,b] \to \mathbb{C}$ and assume g and g' are continuous on $[a,b]$. Then*

$$\int_a^b g'(x)\,dx = g(b) - g(a). \tag{2.1.19}$$

To see this, set

$$G(x) = \int_a^x g'(y)\,dy, \tag{2.1.20}$$

and use Proposition 2.1.3 to show that $G'(x) = g'(x)$. Hence $F(x) = G(x) - g(x)$ satisfies $F'(x) \equiv 0$. The mean value theorem (applied separately to $\operatorname{Re} F$ and $\operatorname{Im} F$) implies F is constant, and one gets (2.1.19).

Another important tool is the following change of variable formula.

Proposition 2.1.5. *Take f as in Proposition 2.1.2, and assume f' is continuous. Then, if $u : [\alpha, \beta] \to \mathbb{C}$ is continuous and $y \in [a, b]$,*

$$\int_a^y u(f(x))f'(x)\,dx = \int_\alpha^{f(y)} u(x)\,dx. \tag{2.1.21}$$

To see this, denote the two sides of (2.1.21) by $L(y)$ and $R(y)$. By Proposition 2.1.3,

$$L'(y) = u(f(y))f'(y) = R'(y), \tag{2.1.22}$$

the second identity also using the chain rule. Hence $L(y) - R(y)$ is constant, but $L(a) = R(a) = 0$, so we have (2.1.21).

The integral plays an essential role in the study of lengths of curves (including circles) and the areas of planar regions bounded by curves (including circles). We go into these matters in Sections 2.2 and 2.6.

One further tool from calculus we will need involves power series, of the form

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \tag{2.1.23}$$

with $a_k, z \in \mathbb{C}$. Now for some classes of coefficients a_k (e.g., $a_k = k!$) this might not converge for any nonzero z. Suppose, however, that it does converge for $z = z_1 \neq 0$. Then the terms in the series are bounded, so one has

$$|a_k R^k| \leq A < \infty, \quad \forall\, k \in \mathbb{Z}^+, \text{ with } R = |z_1|. \tag{2.1.24}$$

Hence, given $\alpha \in (0, 1)$,

$$\begin{aligned} |z| \leq \alpha A &\Rightarrow |a_k z^k| \leq \alpha^k A \\ &\Rightarrow \sum_k |a_n z^k| \leq A \sum_k \alpha^k = \frac{A}{1-\alpha}, \end{aligned} \tag{2.1.25}$$

and we deduce that the power series in (2.1.23) converges for all $z \in D_R(0)$, uniformly on $D_{\alpha R}(0)$ for each $\alpha < 1$, to a limit f that is continuous on $D_R(0)$. The following result will be essential in the treatment of the exponential function.

Proposition 2.1.6. *Assume the power series* (2.1.23) *converges for* $|z| < R$. *Then* $f(x)$ *is differentiable for* $x \in (-R, R)$, *and* $f'(x)$ *is given by*

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}. \tag{2.1.26}$$

For a demonstration, it is convenient first to show that the power series given in (2.1.26) converges for $|x| < R$, defining

$$g(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}, \tag{2.1.27}$$

continuous on $(-R, R)$. It is relatively straightforward to show that one can integrate a convergent power series term by term, obtaining

$$\begin{aligned} \int_0^x g(t)\, dt &= \sum_{k=1}^{\infty} k a_k \int_0^x t^{k-1}\, dt \\ &= \sum_{k=1}^{\infty} a_k x^k = f(x) - f(0), \end{aligned} \tag{2.1.28}$$

and then Proposition 2.1.3 yields (2.1.26).

Remark. One can define the notion of complex differentiability and extend Proposition 2.1.6 to a result on $f'(z)$, for $z \in D_R(0)$. Such matters are treated in Chapter 1 of [**17**], but they will not play a role here.

Once we have Proposition 2.1.6, we have inductively that, for $n \in \mathbb{N}$,

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1) a_k x^{k-n}, \tag{2.1.29}$$

and in particular

$$f^{(n)}(0) = n! a_n, \tag{2.1.30}$$

hence

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\, x^k. \tag{2.1.31}$$

For example, take

$$f_r(x) = (1-x)^{-r}, \quad |x| < 1. \tag{2.1.32}$$

Then inductive use of (2.1.15) and the chain rule gives

$$f_r^{(k)}(x) = r(r+1)\cdots(r+k-1)(1-x)^{-r-k}, \tag{2.1.33}$$

so, granted that this function has a power series of the form (2.1.23), we have

$$(1-x)^{-r} = \sum_{k=0}^{\infty} a_k x^k, \quad a_k = \frac{r(r-1)\cdots(r+k-1)}{k!}, \tag{2.1.34}$$

for $|x| < 1$. One can use the ratio test (see (2.1.43) below) to show that the right side converges for $|x| < 1$, defining a continuous function g_r on $(-1, 1)$. To show that $f_r = g_r$ takes an additional argument. Here is one way to do it. For another, see §4.3 of [**15**].

Lemma 2.1.7. *With $f_r(x)$ as in* (2.1.32) *and $g_r(x)$ given by the power series on the right side of* (2.1.34)*, we have*

$$f_r(x) = g_r(x), \quad \textit{for} \ \ |x| < 1. \tag{2.1.35}$$

Proof. Applying (2.1.15) and the chain rule to compute $f_r'(x)$ and applying Proposition 2.1.6 to compute $g_r'(x)$ gives

$$(1-x)f_r'(x) = rf_r(x), \quad \text{and} \ \ (1-x)g_r'(x) = rg_r(x). \tag{2.1.36}$$

From here we get

$$\frac{d}{dx}(1-x)^r g_r(x) = 0, \tag{2.1.37}$$

which in turn yields (2.1.35). □

This power series, with $r = 1/2$, will appear in calculations for arclength of a circle.

The case $r = 1$ of (2.1.34) also merits special attention. This is the geometric series

$$\frac{1}{1-x} = \sum_{k=1}^{\infty} x^k = 1 + x + x^2 + \cdots, \quad |x| < 1. \tag{2.1.38}$$

This can be established without calculus, as follows. Write

$$\begin{aligned} S_n &= 1 + x + x^2 + \cdots + x^n, \\ xS_n &= x + x^2 + \cdots + x^n + x^{n+1}, \end{aligned} \tag{2.1.39}$$

and subtract. We have $(1-x)S_n = 1 - x^{n+1}$, hence

$$S_n = \frac{1 - x^{n+1}}{1-x}, \tag{2.1.40}$$

for $x \neq 1$. If $|x| < 1$, taking the limit $n \to \infty$ yields (2.1.38).

We also obtain a useful sufficient condition for convergence of an infinite series

$$A = \sum_{k=0}^{\infty} a_k. \tag{2.1.41}$$

If we have $C < \infty$, $x \in (0,1)$, and $N < \infty$ such that

$$|a_k| \le C x^k, \quad \forall\, k \ge N, \tag{2.1.42}$$

then (2.1.41) converges, via comparison with (2.1.38). The condition (2.1.41) holds provided

$$\left|\frac{a_{k+1}}{a_k}\right| \le x, \quad \forall\, k \ge N, \tag{2.1.43}$$

for some $x \in (0,1)$. This test for convergence of (2.1.41) is known as the ratio test. One use of this test was mentioned right below (2.1.34), and is also relevant for (2.1.25) and (2.1.27). We will see other uses in subsequent sections.

Exercises

1. Use (2.1.15) and the fundamental theorem of calculus to show that, for $x > 0$,

$$\int_1^x y^r\, dy = \frac{x^{r+1} - 1}{r+1},$$

for $r \in \mathbb{Q}$, $r \neq -1$.

2. Use the substitution $t = s^2 + 1$ to evaluate

$$\int_0^1 s\sqrt{s^2+1}\, ds.$$

3. Let $f, g : (a,b) \to \mathbb{C}$ be differentiable, and assume g does not vanish on this interval. Show that, for $x \in (a,b)$,

$$\frac{d}{dx}\frac{f}{g}(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

This is called the quotient rule for derivatives. It complements the product rule, given in (2.1.6).

4. Let $f, g : [a,b] \to \mathbb{C}$ be differentiable, and assume f', g' are continuous.

Show that

$$\int_a^b f(x)g'(x)\,dx = -\int_a^b f'(x)g(x)\,dx + [f(b)g(b) - f(a)g(a)].$$

This identity is known as integration by parts. (*Hint.* Consider the derivative of $u(x) = f(x)g(x)$.) Peek at §2.2 and note the special case

$$\int_0^b x\cos x\,dx = -\int_0^b \sin x\,dx + b\sin b.$$

5. Let $f_k, f : [a,b] \to \mathbb{C}$. We say f_k converges to f uniformly on $[a,b]$ provided there exist $\varepsilon_k \to 0$ such that $|f_k(x) - f(x)| \leq \varepsilon_k$, for all $x \in [a,b]$. Show that, if this happens and each f_k is continuous on $[a,b]$, then the limit f is continuous on $[a,b]$.

6. In the setting of Exercise 5, with f_k continuous on $[a,b]$ and $f_k \to f$ uniformly, show that

$$\int_a^b f_k(x)\,dx \longrightarrow \int_a^b f(x)\,dx.$$

Hint. Show that

$$\int_a^b |f_k(x) - f(x)|\,dx \leq (b-a)\varepsilon_k.$$

2.2. Curves, arclength, and the real measurement of angles

A curve in $\mathbb{C}$ is given by a continuous function from an interval $I \subset \mathbb{R}$ to $\mathbb{C}$,

$$\gamma : I \longrightarrow \mathbb{C}, \quad \gamma(t) = x(t) + iy(t). \tag{2.2.1}$$

Typically we work with differentiable functions; $\gamma'(t)$ is the velocity of γ, at "time" t, and its "speed" is the magnitude of $\gamma'(t)$:

$$|\gamma'(t)| = \sqrt{x'(t)^2 + y'(t)^2}. \tag{2.2.2}$$

One also calls the image of I under the map γ a curve in $\mathbb{C}$. If $u : J \to I$ is continuous, one-to-one, and onto, the map

$$\sigma : J \longrightarrow \mathbb{C}, \quad \sigma(t) = \gamma(u(t)) \tag{2.2.3}$$

has the same image as γ. We say σ is a reparametrization of γ. We typically require that u be C^1, with C^1 inverse. If also γ is C^1, the chain rule gives

$$\sigma'(t) = u'(t)\gamma'(u(t)). \tag{2.2.4}$$

Assume $I = [a, b]$ and γ is C^1. We want to define the length of this curve. We follow §4.4 of [**15**], which treats curves in n-dimensional Euclidean space $\mathbb{R}^n$.

To start, we take a partition $\mathcal{P}$ of $[a, b]$, given by

$$a = t_0 < t_1 < \cdots < t_N = b, \tag{2.2.5}$$

and set

$$\ell_{\mathcal{P}}(\gamma) = \sum_{j=1}^{N} |\gamma(t_j) - \gamma(t_{j-1})|. \tag{2.2.6}$$

See Figure 2.2.1. We convert this into something that looks like a Riemann sum for $\int_a^b |\gamma'(t)|\, dt$. We have

$$\begin{aligned} \gamma(t_j) - \gamma(t_{j-1}) &= \int_{t_{j-1}}^{t_j} \gamma'(t)\, dt \\ &= \int_{t_{j-1}}^{t_j} [\gamma'(t_j) + \gamma'(t) - \gamma'(t_j)]\, dt \\ &= (t_j - t_{j-1})\gamma'(t_j) + \int_{t_{j-1}}^{t_j} [\gamma'(t) - \gamma'(t_j)]\, dt. \end{aligned} \tag{2.2.7}$$

We deduce that

$$|\gamma(t_j) - \gamma(t_{j-1})| = (t_j - t_{j-1})|\gamma'(t_j)| + r_j, \tag{2.2.8}$$

with

$$|r_j| \le \int_{t_{j-1}}^{t_j} |\gamma'(t) - \gamma'(t_j)|\, dt, \tag{2.2.9}$$

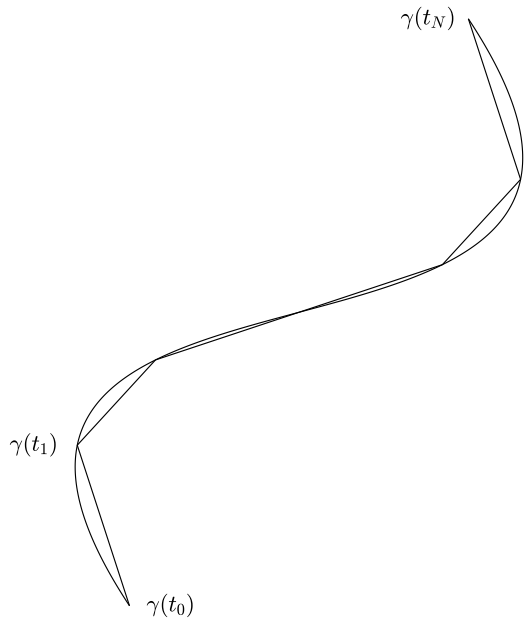

Figure 2.2.1. Approximating $\ell(\gamma)$ by $\ell_{\mathcal{P}}(\gamma)$

If γ' is continuous on $[a, b]$, so is $|\gamma'|$, and in fact both are uniformly continuous on $[a, b]$. Hence

$$s, t \in [a, b],\ |s - t| \le h \Longrightarrow |\gamma'(t) - \gamma'(s)| \le \omega(h), \tag{2.2.10}$$

where $\omega(h) \to 0$ as $h \to 0$. Summing (2.2.8) over j, we get

$$\ell_{\mathcal{P}}(\gamma) = \sum_{j=1}^{N} |\gamma'(t_j)|(t_j - t_{j-1}) + R_{\mathcal{P}}, \tag{2.2.11}$$

with

$$|R_{\mathcal{P}}| \le (b - a)\omega(h)), \quad \text{if each } t_j - t_{j-1} \le h. \tag{2.2.12}$$

Since the sum on the right side of (2.2.11) is a Riemann sum, we have:

Proposition 2.2.1. *Assume $\gamma : [a, b] \to \mathbb{C}$ is a C^1 curve. Then*

$$\ell_{\mathcal{P}}(\gamma) \longrightarrow \int_a^b |\gamma'(t)|\, dt, \tag{2.2.13}$$

as the maximal lengths of the intervals in the partition tend to 0.

We call the limit the length of the curve γ, and write

$$\ell(\gamma) = \int_a^b |\gamma'(t)|\, dt. \tag{2.2.14}$$

Note that if $u : [\alpha, \beta] \to [a, b]$ is a C^1 map with C^1 inverse, and $\sigma = \gamma \circ u$, as in (2.2.3), we have from (2.2.4) that $|\sigma'(t)| = |u'(t)| \cdot |\gamma'(t)|$, and the change of variable formula for integrals gives

$$\int_\alpha^\beta |\sigma'(t)|\, dt = \int_a^b |\gamma'(t)|\, dt, \tag{2.2.15}$$

verifying the geometrically natural result that

$$\ell(\sigma) = \ell(\gamma). \tag{2.2.16}$$

Given such a C^1 curve γ, it is natural to consider the length function

$$\ell_\gamma(t) = \int_a^t |\gamma'(s)|\, ds, \quad \ell_\gamma'(t) = |\gamma'(t)|. \tag{2.2.17}$$

If we assume that γ' is nowhere vanishing on $[a, b]$, the inverse function theorem, Proposition 2.1.2, implies that $\ell_\gamma : [a, b] \to [0, \ell(\gamma)]$ has a C^1 inverse

$$u : [0, \ell(\gamma)] \longrightarrow [a, b], \tag{2.2.18}$$

and then $\sigma = \gamma \circ u : [0, \ell(\gamma)] \to \mathbb{C}$ satisfies

$$\begin{aligned} \sigma'(t) &= u'(t)\gamma'(u(t)) \\ &= \frac{1}{\ell_\gamma'(s)}\gamma'(u(t)), \quad \text{for } t = \ell_\gamma(s),\ s = u(t), \end{aligned} \tag{2.2.19}$$

since the chain rule applied to $u(\ell_\gamma(t)) = t$ yields $u'(\ell_\gamma(t))\ell_\gamma'(t) = 1$. Also, by (2.2.17), $\ell_\gamma'(s) = |\gamma'(s)| = |\gamma'(u(t))|$, so

$$|\sigma'(t)| \equiv 1. \tag{2.2.20}$$

Then σ is a reparametrization of γ, and σ has unit speed. We say σ is a reparametrization by arc length.

We apply these considerations to the circle

$$S^1 = \{z \in \mathbb{C} : |z| = 1\}. \tag{2.2.21}$$

We can parametrize S^1 away from $z = \pm 1$ by

$$\gamma_+(t) = t + i\sqrt{1 - t^2}, \quad \gamma_-(t) = t - i\sqrt{1 - t^2}, \tag{2.2.22}$$

on the intersection of S^1 with $\{z : \operatorname{Im} z > 0\}$ and $\{z : \operatorname{Im} z < 0\}$, respectively. Here $\gamma_\pm : (-1, 1) \to \mathbb{C}$, and both maps are smooth. In fact, we can take

$\gamma_\pm : [-1,1] \to \mathbb{C}$, but these functions are not differentiable at ± 1. We can also parametrize S^1 away from $z = \pm i$, by

$$\gamma_\ell(t) = -\sqrt{1-t^2} + it, \quad \gamma_r(t) = \sqrt{1-t^2} + it, \tag{2.2.23}$$

again with $t \in (-1,1)$. A calculation gives

$$|\gamma_+'(t)|^2 = 1 + \frac{t^2}{1-t^2} = \frac{1}{1-t^2}. \tag{2.2.24}$$

Hence, if $\ell(t)$ is the length of the image $\gamma_+([0,t])$, we have

$$\ell(t) = \int_0^t \frac{1}{\sqrt{1-s^2}}\, ds, \quad \text{for } 0 < t < 1. \tag{2.2.25}$$

The same formula holds with γ_+ replaced by γ_-, γ_ℓ, or γ_r.

These formulas imply that S^1 can be parametrized by arc length. We take $C : \mathbb{R} \to \mathbb{C}$ to be such a parametrization, satisfying

$$C(0) = 1, \quad C'(0) = i. \tag{2.2.26}$$

We are led to the standard trigonometrical functions $\cos t$ and $\sin t$, real-valued functions of t defined by

$$C(t) = \operatorname{cis} t = \cos t + i \sin t. \tag{2.2.27}$$

Also, it is standard to identify the arclength t with the *real angle measurement* (in radians) of the angle made by the wedge $W_{0,1,\omega}$:

$$\Omega(W_{0,1,\omega}) = \omega = \operatorname{cis} t, \quad \angle(W_{0,1,\omega}) = t. \tag{2.2.28}$$

See Figure 2.2.2. Here $\Omega(W_{0,1,\omega})$ is the angle measurement defined in (1.5.3).

We turn to an evaluation of the derivative $C'(t)$. Applying d/dt to the identity

$$\langle C(t), C(t) \rangle = 1 \tag{2.2.29}$$

and using the product formula gives

$$\langle C'(t), C(t) \rangle = 0, \tag{2.2.30}$$

hence $C'(t) \perp C(t)$, or equivalently $C'(t) = \pm i C(t)$ (since $C'(t)$ and $C(t)$ both have unit length). Since $C'(0) = i$, there is only one possibility:

$$C'(t) = iC(t), \quad \text{i.e.,} \quad \frac{d}{dt} \operatorname{cis} t = i \operatorname{cis} t. \tag{2.2.31}$$

In light of (2.2.27), this gives the classical identities

$$\frac{d}{dt} \cos t = -\sin t, \quad \frac{d}{dt} \sin t = \cos t. \tag{2.2.32}$$

In the next section we will provide another path to a direct unit speed parametrization of S^1, connecting trigonometric functions to the exponential function, in the classical Euler formula.

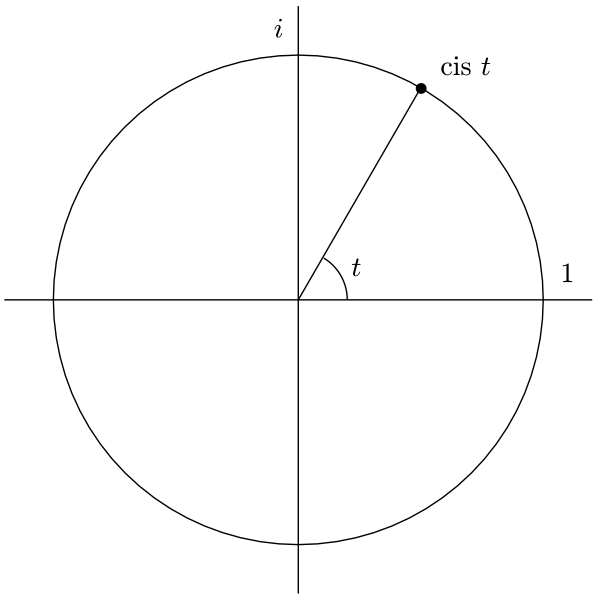

Figure 2.2.2. The circle $C(t) = \operatorname{cis} t$

The famous number π is defined as follows:

$$\pi = \ell(S^1_+). \tag{2.2.33}$$

Equivalently,

$$\frac{\pi}{2} = \ell(S^1_{++}), \tag{2.2.34}$$

where

$$S^1_+ = \{z \in S^1 : \operatorname{Im} z > 0\}, \quad S^1_{++} = \{z \in S^1_+ : \operatorname{Re} z > 0\}. \tag{2.2.35}$$

Going further, the bisector of $W_{0,1,i}$ intersects S^1_{++} at

$$\omega = \frac{1+i}{\sqrt{2}}, \tag{2.2.36}$$

and we have, by (2.2.25),

$$\frac{\pi}{4} = \int_0^{1/\sqrt{2}} \frac{ds}{\sqrt{1-s^2}}. \tag{2.2.37}$$

Behind the results (2.2.34)–(2.2.37) is the following observation.

Proposition 2.2.2. *If $\gamma : [a, b] \to \mathbb{C}$ is a smooth curve and $F \in \mathrm{Isom}(\mathbb{C})$, then*

$$\ell(\gamma) = \ell(F(\gamma)). \tag{2.2.38}$$

Further results bearing on formulas for π and its numerical evaluation are given in §2.4.

To end this section, we mention the following *law of cosines*. Given nonzero $u, v \in \mathbb{C}$, we have the angle measurement defined in (1.5.3),

$$\Omega(u, v) = \frac{u}{v}\,\frac{|v|}{|u|} = \frac{u\overline{v}}{|u|\,|v|} \in S^1. \tag{2.2.39}$$

As in (2.2.28), write

$$\Omega(u, v) = \operatorname{cis} t, \quad -\pi < t \leq \pi. \tag{2.2.40}$$

Then (cf. (1.6.11)),

$$\langle u, v\rangle = \operatorname{Re} u\overline{v} = |u|\,|v| \operatorname{Re}\Omega(u, v), \tag{2.2.41}$$

which yields the law of cosines:

$$\langle u, v\rangle = |u|\,|v| \cos t. \tag{2.2.42}$$

Exercises

1. Consider the parabola $y = x^2/2$, parametrized by $\gamma(t) = t + (t^2/2)i$. Show that the length $\ell(t)$ of $\gamma([0, t])$ is given by

$$\ell(t) = \int_0^t \sqrt{1 + s^2}\, ds.$$

2. Consider the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2},$$

parametrized by $\gamma(t) = a\cos t + (b \sin t)i$. Produce an integral formula for the circumference of this ellipse.

2.3. Exponential and trigonometric functions, and Euler's formula

Here we present the major insight of Euler relating the exponential and trigonometric functions. We start with a definition of the exponential function, both for real and complex arguments.

We construct the exponential function to solve the differential equation

$$\frac{dx}{dt} = x, \quad x(0) = 1. \tag{2.3.1}$$

We seek a solution as a power series

$$x(t) = \sum_{k=0}^{\infty} a_k t^k. \tag{2.3.2}$$

In such a case, if this series converges for $|t| < R$, then, as stated in §2.1,

$$\begin{aligned} x'(t) &= \sum_{k=1}^{\infty} k a_k t^{k-1} \\ &= \sum_{\ell=0}^{\infty} (\ell+1) a_{\ell+1} t^{\ell}, \end{aligned} \tag{2.3.3}$$

so for (2.3.1) to hold we need

$$a_0 = 1, \quad a_{k+1} = \frac{a_k}{k+1}, \tag{2.3.4}$$

i.e., $a_k = 1/k!$, where $k! = k(k-1)\cdots 2 \cdot 1$. Thus (2.3.1) ia solved by

$$x(t) = e^t = \sum_{k=0}^{\infty} \frac{1}{k!} t^k, \quad t \in \mathbb{R}. \tag{2.3.5}$$

This defines the exponential function e^t, for $t \in \mathbb{R}$.

More generally, we can define

$$e^z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k, \quad z \in \mathbb{C}. \tag{2.3.6}$$

The ratio test then shows that the series (2.3.6) is absolutely convergent for all $z \in \mathbb{C}$, and uniformly convergent for $|z| \leq R$, for each $R < \infty$. Results given in §2.1 also imply that

$$e^{at} = \sum_{k=0}^{\infty} \frac{a^k}{k!} t^k \tag{2.3.7}$$

solves

$$\frac{d}{dt} e^{at} = a e^{at}, \tag{2.3.8}$$

and this works for each $a \in \mathbb{C}$,

We claim that e^{at} is the unique solution to

$$\frac{dy}{dt} = ay, \quad y(0) = 1. \tag{2.3.9}$$

To see this, compute the derivative of $e^{-at}y(t)$:

$$\frac{d}{dt}(e^{-at}y(t)) = -ae^{-at}y(t) + e^{-at}ay(t) = 0, \tag{2.3.10}$$

where we use the product rule, (2.3.8) (with a replaced by $-a$) and (2.3.9). Thus $e^{-at}y(t)$ is independent of t. Evaluating at $t = 0$ gives

$$e^{-at}y(t) = 1, \quad \forall\, t \in \mathbb{R}, \tag{2.3.11}$$

whenever $y(t)$ solves (2.3.9). Since e^{at} solves (2.3.9), we have $e^{-at}e^{at} = 1$, hence

$$e^{-at} = \frac{1}{e^{at}}. \quad \forall\, t \in \mathbb{R},\ a \in \mathbb{C}. \tag{2.3.12}$$

Thus multiplying both sides of (2.3.11) by e^{at} gives the desired uniqueness:

$$y(t) = e^{at}, \quad \forall\, t \in \mathbb{R}. \tag{2.3.13}$$

We can draw further useful conclusions from applying d/dt to products of exponential functions. In fact, let $a, b \in \mathbb{C}$. Then

$$\begin{aligned} &\frac{d}{dt}\Big(e^{-at}e^{-bt}e^{(a+b)t}\Big) \\ &= -ae^{-at}e^{-bt}e^{(a+b)t} - be^{-at}e^{-bt}e^{(a+b)t} + (a+b)e^{-at}e^{-bt}e^{(a+b)t} \\ &= 0, \end{aligned} \tag{2.3.14}$$

so again we are differentiating a function that is independent of t. Evaluation at $t = 0$ gives

$$e^{-at}e^{-bt}e^{(a+b)t} = 1, \quad \forall\, t \in \mathbb{R}. \tag{2.3.15}$$

Again using (2.3.12), we get

$$e^{(a+b)t} = e^{at}e^{bt}, \quad \forall\, t \in \mathbb{R}, a, b \in \mathbb{C}, \tag{2.3.16}$$

or, setting $t = 1$,

$$e^{a+b} = e^a e^b, \quad \forall\, a, b \in \mathbb{C}. \tag{2.3.17}$$

We next record some properties of $\exp(t) = e^t$ for real t. The power series (2.3.5) clearly gives $e^t > 0$ for $t \geq 0$. Since $e^{-1} = 1/e^t$, we see that $e^t > 0$ for all $t \in \mathbb{R}$. Since $de^t/dt = e^t > 0$, the function is monotone increasing in t. Note that, for $t > 0$,

$$e^t = 1 + t + \frac{t^2}{2} + \cdots > 1 + t \nearrow +\infty, \tag{2.3.18}$$

as $t \to \infty$. Hence

$$\lim_{t \to +\infty} e^t = +\infty. \tag{2.3.19}$$

Since $e^{-t} = 1/e^t$,

$$\lim_{t \to -\infty} e^t = 0. \tag{2.3.20}$$

As a consequence,

$$\exp : \mathbb{R} \longrightarrow (0, \infty) \tag{2.3.21}$$

is one-to-one and onto, with positive derivative, so there is a smooth inverse:

$$L : (0, \infty) \longrightarrow \mathbb{R}. \tag{2.3.22}$$

We call this inverse the natural logarithm:

$$\log x = L(x). \tag{2.3.23}$$

Applying d/dt to

$$L(e^t) = t \tag{2.3.24}$$

gives $L'(e^t)e^t = 1$, hence $L'(e^t) = 1/e^t$, so

$$\frac{d}{dx} \log x = \frac{1}{x}. \tag{2.3.25}$$

Since $\log 1 = 0$, we get

$$\log x = \int_1^x \frac{dy}{y}. \tag{2.3.26}$$

We move on to a study of e^z for purely imaginary z, i.e., of

$$\gamma(t) = e^{it}, \quad t \in \mathbb{R}. \tag{2.3.27}$$

This traces out a curve in $\mathbb{C}$, and we want to understand which curve it is. First we calculate $|e^{it}|$. Recall that $|z|^2 = z\overline{z}$. The basic results (1.1.9) on complex conjugates of sums and products yield

$$\overline{e^z} = \sum_{k=0}^{\infty} \frac{\overline{z}^k}{k!} = e^{\overline{z}}, \tag{2.3.28}$$

for $z \in \mathbb{C}$. In particular,

$$t \in \mathbb{R} \Longrightarrow |e^{it}|^2 = e^{it}e^{-it} = 1. \tag{2.3.29}$$

Hence $\gamma(t) = e^{it}$ lies on the unit circle centered at the origin in $\mathbb{C}$. Also

$$\gamma'(t) = ie^{it} \Longrightarrow |\gamma'(t)| \equiv 1, \tag{2.3.30}$$

so $\gamma(t)$ moves at unit speed on the unit circle. Also $\gamma(0) = 1$, $\gamma'(0) = i$.

It follows that $\gamma(t) = e^{it}$ is the parametrization of the unit circle by arc length introduced in (2.2.27). This establishes the following result, known as Euler's formula.

Proposition 2.3.1. *For $t \in \mathbb{R}$,*

$$e^{it} = \operatorname{cis} t = \cos t + i \sin t. \tag{2.3.31}$$

Given Proposition 2.3.1, the identity

$$\frac{d}{dt} = ie^{it}, \tag{2.3.32}$$

hence gives a second proof of (2.2.31), i.e.,

$$\frac{d}{dt} \operatorname{cis} t = i \operatorname{cis} t, \tag{2.3.33}$$

or equivalently (2.2.32), i.e.,

$$\frac{d}{dt} \cos t = -\sin t, \quad \frac{d}{dt} \sin t = \cos t. \tag{2.3.34}$$

Given Euler's formula, we can use (2.3.17) to derive formulas for sin and cos of the sum of two angles. Indeed, comparing

$$e^{i(s+t)} = \cos(s+t) + i \sin(s+t) \tag{2.3.35}$$

with

$$e^{is}e^{it} = (\cos s + i \sin s)(\cos t + i \sin t) \tag{2.3.36}$$

gives

$$\begin{aligned} \cos(s+t) &= (\cos s)(\cos t) - (\sin s)(\sin t), \\ \sin(s+t) &= (\sin s)(\cos t) + (\cos s)(\sin t). \end{aligned} \tag{2.3.37}$$

Another compact presentation of these sum formulas is simply

$$\operatorname{cis}(s+t) = (\operatorname{cis} s)(\operatorname{cis} t). \tag{2.3.38}$$

Remark. An alternative approach to Euler's formula (2.3.31) is to use the formula for the derivative of $\operatorname{cis} t$, (2.3.33), relying on the proof in (2.2.31), together with the initial condition $\operatorname{cis}(0) = 1$. Then the uniqueness result (2.3.9)–(2.3.13) implies that $\operatorname{cis} t = e^{it}$.

Recall from (2.2.33) that the number π is defined by the identity

$$\pi = \ell(S^1_+). \tag{2.3.39}$$

Equivalently, π is the smallest positive number such that

$$\operatorname{cis} \pi = -1, \tag{2.3.40}$$

or alternatively, by (2.3.31), the smallest positive number such that

$$e^{\pi i} = -1. \tag{2.3.41}$$

Note also that

$$e^{-\pi i} = -1, \quad e^{\pm 2\pi i} = 1. \tag{2.3.42}$$

In fact, we have the following.

Proposition 2.3.2. *The maps*

$$\mathrm{cis} : (0,\pi) \to S^1_+, \quad \mathrm{cis} : (0,2\pi) \to S^1 \setminus \{1\} \tag{2.3.43}$$

are one-to-one and onto. Similarly,

$$\mathrm{cis}^-(t) = \overline{\mathrm{cis}\, t} = \mathrm{cis}(-t) \tag{2.3.44}$$

defines

$$\mathrm{cis}^- : (0,\pi) \to S^1_-, \quad \mathrm{cis}^- : (0,2\pi) \to S^1 \setminus \{1\}, \tag{2.3.45}$$

and these maps are one-to-one and onto.

Another important role played by π is in the angle sum formula for triangles.

Proposition 2.3.3. *Let $\mathcal{T} = \triangle(A,B,C)$ be a triangle. Then*

$$\Omega_{\mathcal{T}}(A) = \alpha = e^{is}, \ \Omega_{\mathcal{T}}(B) = \beta = e^{it}, \ \Omega_{\mathcal{T}}(C) = \gamma = e^{iu}, \tag{2.3.46}$$

where either

$$\alpha, \beta, \gamma \in S^1_+, \quad s,t,u \in (0,\pi), \tag{2.3.47}$$

or

$$\alpha, \beta, \gamma \in S^1_-, \quad s,t,u \in (-\pi,0). \tag{2.3.48}$$

Then

$$s + t + u = \pi \quad \text{or} \quad -\pi, \tag{2.3.49}$$

in these two respective cases. We use the notation

$$\angle_{\mathcal{T}}(A) = s, \quad \angle_{\mathcal{T}}(B) = t, \quad \angle_{\mathcal{T}}(C) = u. \tag{2.3.50}$$

Proof. The cases (2.3.47) and (2.3.48) hold by Propositions 1.6.2 and 2.3.2. Proposition 1.6.1 says

$$\alpha\beta\gamma = -1, \tag{2.3.51}$$

hence

$$e^{i(s+t+u)} = -1. \tag{2.3.52}$$

Consider case (2.3.47). Since $\alpha\beta = -\gamma^{-1}$, we also have

$$\alpha\beta \in S^1_+, \tag{2.3.53}$$

hence $s+t \in (0,\pi)$. Thus $s+t+u < 2\pi$, and then (2.3.52) forces $s+t+u = \pi$. The analysis of the case (2.3.48) is similar. □

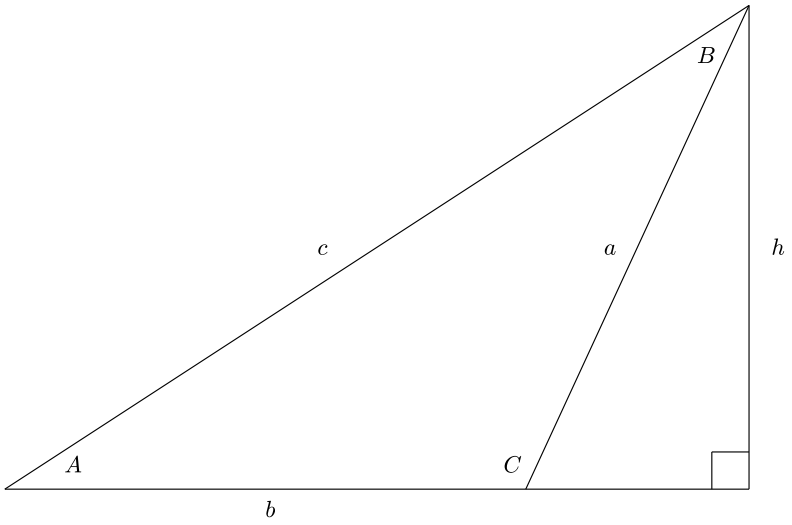

Figure 2.3.1. The Law of Sines

REMARK. We could skip the use of (2.3.53). In fact, given $\theta \in \mathbb{R}$,

$$e^{i\theta} = -1 \Longleftrightarrow \theta = \pi + 2k\pi, \quad \text{for some} \quad k \in \mathbb{Z}, \tag{2.3.54}$$

and since $s, t, u \in (0, \pi) \Rightarrow s + t + u \in (0, 3\pi)$, (2.3.52) forces this sum to be π.

Exercise

1. Let $\mathcal{T} = \triangle(A, B, C)$ be a triangle. Use the notation

$$\angle_{\mathcal{T}}(A), \quad \angle_{\mathcal{T}}(B), \quad \angle_{\mathcal{T}}(C),$$

as in (2.3.50), and let a, b, c denote the lengths of the sides opposite these vertices. Show that

$$\frac{\sin \angle_{\mathcal{T}}(A)}{a} = \frac{\sin \angle_{\mathcal{T}}(B)}{b} = \frac{\sin \angle_{\mathcal{T}}(C)}{c}.$$

This is called the law of sines.

Hint. Consider Figure 2.3.1. Show that

$$h = c \sin \angle_{\mathcal{T}}(A), \quad \text{and} \quad h = a \sin \angle_{\mathcal{T}}(C).$$

2. Deduce from Euler's formula (2.3.31) that

$$\cos t = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} t^{2k} = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \cdots,$$
$$\sin t = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} t^{2k+1} = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots.$$

3. Deduce from (2.3.34) that

$$\int_0^t \sin s \, ds = 1 - \cos t, \quad \int_0^t \cos s \, ds = \sin t.$$

4. Deduce from (2.3.37) that

$$\cos 2t = \cos^2 t - \sin^2 t = 2\cos^2 t - 1 = 1 - 2\sin^2 t.$$

Given this and the result of Exercise 3, evaluate

$$\int_0^t \cos^2 s \, ds, \quad \int_0^t \sin^2 s \, ds.$$

5. Let $f : (a, b) \to \mathbb{R}$ be differentiable. Deduce from (2.1.9) that

$$\frac{d}{dt} \cos f(t) = -f'(t) \sin f(t), \quad \frac{d}{dt} \sin f(t) = f'(t) \cos f(t).$$

Then deduce that

$$\frac{d}{dt} e^{if(t)} = if'(t) e^{if(t)}.$$

More generally, using $e^{(a+ib)f} = e^{af} e^{ibf}$, show that, for $c \in \mathbb{C}$,

$$\frac{d}{dt} e^{cf(x)} = cf'(x) e^{cf(x)}.$$

6. Given smooth $r, \theta : [a, b] \to \mathbb{R}$, show that

$$\begin{aligned} &\gamma(t) = r(t) e^{i\theta(t)} \\ &\Longrightarrow |\gamma'(t)|^2 = r'(t)^2 + r(t)^2 \theta'(t)^2 \\ &\Longrightarrow \int_a^b |\gamma'(t)| \, dt = \int_a^b \sqrt{r'(t)^2 + r(t)^2 \theta'(t)^2} \, dt. \end{aligned}$$

7. Given $r \in \mathbb{C}$, define x^r for $x > 0$ by

$$x^r = e^{r \log x}.$$

Show that

$$\frac{d}{dx} x^r = r x^{r-1}, \quad \forall \, x > 0.$$

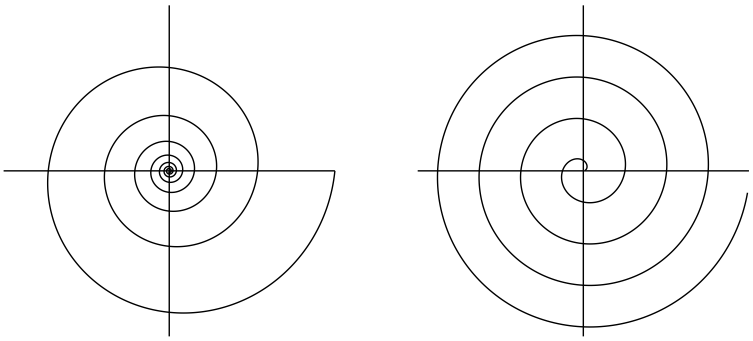

Figure 2.3.2. Logarithmic spiral and spiral of Archimedes

If also $s \in \mathbb{C}$, show that

$$x^{r+s} = x^r x^s, \quad \forall\, x > 0.$$

Also show

$$(x^r)^s = x^{rs}, \quad \forall\, x > 0,\ r \in \mathbb{R},\ s \in \mathbb{C}.$$

8. Two spirals are presented in Figure 2.3.2:

$$\sigma(t) = t^{1+ai}, \quad \text{logarithmic spiral,}$$
$$\gamma(t) = te^{ibt}, \quad \text{spiral of Archimedes.}$$

In these figures, $a = 10,\ b = 25$. Compute $|\sigma'(t)|$ and $|\gamma'(t)|$ and write down integral formulas for the arclengths of σ and γ over $t \in [0, s]$.

Note the similarity of the formula for the arclength of γ with that for the parabola in Exercise 1 of §2.2. For two approaches to the evaluation of this integral, see Exercise 2 of §2.4.

2.4. More on Euler's formula, trigonometric functions, and π

We continue examining some consequences of the interaction of Euler's formula and π, initiated in (2.3.39)–(2.3.42), which gave

$$e^{\pm\pi i} = -1, \quad e^{\pm 2\pi i} = 1, \quad e^{\pm\pi i/2} = \pm i. \tag{2.4.1}$$

For one, as seen in §1.5,

$$\omega \in S^1_{\pm},\ \alpha = \frac{\omega + 1}{|\omega + 1|} \Rightarrow \alpha \in S^1_{\pm} \text{ and } \alpha^2 = \omega. \tag{2.4.2}$$

Consequently,

$$e^{\pi i/4} = \frac{1+i}{\sqrt{2}}, \quad e^{-\pi i/4} = \frac{1-i}{\sqrt{2}}. \tag{2.4.3}$$

We next examine

$$e^{\pi i/3} = z, \quad z^3 = -1, \quad z^3 + 1 = 0. \tag{2.4.4}$$

We have

$$z^3 + 1 = (z+1)(z^2 - z + 1), \tag{2.4.5}$$

so $e^{\pi i/3} = z$ is a solution to

$$z^2 - z + 1 = 0. \tag{2.4.6}$$

The roots of this equation are given by the quadratic formula:

$$z = \frac{1}{2} \pm \frac{i}{2}\sqrt{3}. \tag{2.4.7}$$

Since $e^{\pi i/3} \in S^1_+$, we have

$$e^{\pi i/3} = \frac{1}{2} + \frac{i}{2}\sqrt{3}. \tag{2.4.8}$$

It follows that

$$\begin{aligned} e^{2\pi i/3} &= \Big(\frac{1}{2} + \frac{i}{2}\sqrt{3}\Big)^2 \\ &= -\frac{1}{2} + \frac{i}{2}\sqrt{3}, \end{aligned} \tag{2.4.9}$$

and

$$\begin{aligned} e^{\pi i/6} &= e^{\pi i/2 - \pi i/3} = e^{\pi i/2} e^{-\pi i/3} \\ &= i\Big(\frac{1}{2} - \frac{i}{2}\sqrt{3}\Big) \\ &= \frac{\sqrt{3}}{2} + \frac{i}{2}. \end{aligned} \tag{2.4.10}$$

The formula (2.4.8) in concert with (2.2.25) leads to the following variant of (2.2.37).

$$\frac{\pi}{6} = \int_0^{1/2} \frac{ds}{\sqrt{1-s^2}}. \tag{2.4.11}$$

This is the value at $t = 1/2$ of the integral (2.2.25) for $\ell(t)$. We can evaluate $\ell(t)$ as a power series in t, in the following way. As noted in §2.1,

$$(1-x)^{-1/2} = \sum_{k=0}^{\infty} \frac{a_k}{k!} x^k, \quad \text{for } |x| < 1, \tag{2.4.12}$$

where

$$a_0 = 1,\ a_1 = \frac{1}{2},\ a_j = \Big(\frac{1}{2}\Big)\Big(\frac{3}{2}\Big)\cdots\Big(k - \frac{1}{2}\Big). \tag{2.4.13}$$

It follows that

$$(1-s^2)^{-1/2} = \sum_{k=0}^{\infty} \frac{a_k}{k!} s^{2k}, \quad |s| < 1, \tag{2.4.14}$$

with uniform convergence on $[-a, a]$ for each $a \in (0, 1)$. Then we can integrate (2.4.14) term-by-term to get

$$\ell(t) = \sum_{k=0}^{\infty} \frac{a_k}{k!} \frac{t^{2k+1}}{2k+1}, \quad 0 \le t < 1. \tag{2.4.15}$$

In particular,

$$\frac{\pi}{6} = \sum_{k=0}^{\infty} \frac{a_k}{k!(2k+1)}\, 2^{-(2k+1)}. \tag{2.4.16}$$

Using (2.4.13) (see (2.6.96)), one can show that summing (2.4.16) over $0 \le k \le 20$ yields π to within an error $< 10^{-13}$. One gets

$$\pi \approx 3.141592653589\cdots. \tag{2.4.17}$$

Here is another take on the integral

$$\ell(t) = \int_0^t \frac{ds}{\sqrt{1-s^2}}, \quad 0 \le t < 1. \tag{2.4.18}$$

We make a change of variable

$$s = \sin\theta. \tag{2.4.19}$$

Note that, since

$$\sin\theta = \operatorname{Im} e^{i\theta}, \quad \frac{d}{d\theta}\sin\theta = \cos\theta = \operatorname{Re} e^{i\theta}, \tag{2.4.20}$$

the function $\sin\theta$ has positive derivative for $-\pi/2 < \theta < \pi/2$, hence is monotone increasing on $[-\pi/2, \pi/2]$, with inverse

$$\sin^{-1} : [-1,1] \longrightarrow \Big[-\frac{\pi}{2}, \frac{\pi}{2}\Big], \tag{2.4.21}$$

smooth on $(-1,1)$. Hence the change of variable (2.4.19) leads to $ds = \cos\theta\, d\theta$, and then to

$$\ell(t) = \int_0^{\varphi} \frac{\cos\theta\, d\theta}{\cos\theta} = \varphi, \quad \sin\varphi = t, \tag{2.4.22}$$

i.e.,

$$\ell(t) = \sin^{-1} t, \quad 0 \leq t < 1. \tag{2.4.23}$$

In particular,

$$\ell\Big(\frac{1}{2}\Big) = \sin^{-1}\frac{1}{2} = \frac{\pi}{6}, \tag{2.4.24}$$

by (2.4.10).

We can obtain other formulas for π by bringing in the trigonometric function

$$\tan t = \frac{\sin t}{\cos t}, \quad |t| < \frac{\pi}{2}. \tag{2.4.25}$$

The quotient rule for derivatives gives

$$\begin{aligned} \frac{d}{dt}\tan t &= \frac{\cos^2 t + \sin^2 t}{\cos^2 t} \\ &= 1 + \tan^2 t. \end{aligned} \tag{2.4.26}$$

Thus, to evaluate the integral

$$I(y) = \int_0^y \frac{dx}{1+x^2}, \tag{2.4.27}$$

we can use the substitution $x = \tan t$, with $dx = (1 + \tan^2 t)\, dt$, to get

$$I(y) = \int_0^u dt = u, \quad \tan u = y, \tag{2.4.28}$$

i.e.,

$$I(y) = \tan^{-1} y. \tag{2.4.29}$$

On the other hand, for $|y| < 1$, the integral (2.4.27) has a convenient power series, derived as follows. We have the geometrical series

$$\frac{1}{1-r} = \sum_{k=0}^{\infty} r^k, \quad |r| < 1, \tag{2.4.30}$$

and setting $r = -x^2$ yields

$$\frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k}, \quad |x| < 1. \tag{2.4.31}$$

Hence integrating term-by-term gives

$$\tan^{-1} y = \sum_{k=0}^{\infty} (-1)^k \frac{y^{2k+1}}{2k+1}, \quad |y| < 1. \tag{2.4.32}$$

To get formulas for π, note that computations in (2.4.3)–(2.4.10) give

$$\tan \frac{\pi}{4} = 1, \quad \tan \frac{\pi}{3} = \sqrt{3}, \quad \tan \frac{\pi}{6} = \frac{1}{\sqrt{3}}. \tag{2.4.33}$$

Only the last of these identities is of use for (2.4.32). It gives

$$\frac{\pi}{6} = \tan^{-1} \frac{1}{\sqrt{3}} = \frac{1}{\sqrt{3}} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \Big(\frac{1}{3}\Big)^k. \tag{2.4.34}$$

This has a simpler form than (2.4.16) (except for the irrational factor $1/\sqrt{3}$). On the other hand, convergence is somewhat slower. To get accuracy as in (2.4.17), one needs to sum over $0 \le k \le 30$.

We can do better by expressing π as a finite linear combination of terms $\tan^{-1} x_j$ for certain fairly small numbers x_j. The key to this is the following formula for $\tan(a+b)$. Using (2.3.37), we have

$$\begin{aligned} \tan(a+b) &= \frac{\sin(a+b)}{\cos(a+b)} = \frac{\sin a \, \cos b + \cos a \, \sin b}{\cos a \, \cos b - \sin a \, \sin b} \\ &= \frac{\tan a + \tan b}{1 - \tan a \, \tan b}. \end{aligned} \tag{2.4.35}$$

Since $\tan \pi/4 = 1$, we have, for $a, b, a+b \in (-\pi/2, \pi/2)$,

$$\frac{\pi}{4} = a + b \Longleftarrow \frac{\tan a + \tan b}{1 - \tan a \, \tan b} = 1. \tag{2.4.36}$$

Taking $a = \tan^{-1} x, \ b = \tan^{-1} y$ gives

$$\begin{aligned} \frac{\pi}{4} = \tan^{-1} x + \tan^{-1} y &\Longleftarrow x + y = 1 - xy \\ &\Longleftarrow x = \frac{1-y}{1+y}. \end{aligned} \tag{2.4.37}$$

If we set $y = 1/2$, we get $x = 1/3$, so

$$\frac{\pi}{4} = \tan^{-1} \frac{1}{3} + \tan^{-1} \frac{1}{2}. \tag{2.4.38}$$

The power series (2.4.32) for $\tan^{-1}(1/3)$ and $\tan^{-1}(1/2)$ both converge faster than (2.4.34), but that for $\tan^{-1}(1/2)$ converges at essentially the same rate as (2.4.16). We might optimize by taking $x = y$ in (2.4.37), but that yields

$x = y = \sqrt{2} - 1$, and we do not want to plug this irrational number into (2.4.32). Taking a cue from $\sqrt{2} - 1 \approx 0.414$, we set $y = 2/5$, which yields $x = 3/7$, so

$$\frac{\pi}{4} = \tan^{-1}\frac{2}{5} + \tan^{-1}\frac{3}{7}. \tag{2.4.39}$$

Both resulting power series converge a little faster than (2.4.16).

To do better, one can bring in a formula for $\tan(a + 2b)$. Computations somewhat parallel to (2.4.35)–(2.4.37) give rise to the following analogues of (2.4.38)–(2.4.39):

$$\begin{aligned} \frac{\pi}{4} &= \tan^{-1}\frac{1}{7} + 2\tan^{-1}\frac{1}{3}, \\ \frac{\pi}{4} &= \tan^{-1}\frac{7}{23} + 2\tan^{-1}\frac{1}{4}. \end{aligned} \tag{2.4.40}$$

Details can be found in [**15**], Appendix A.5. All the power series here converge significantly faster than the series in (2.4.16).

Going further, one can bring in a formula for $\tan(a+4b)$ and run through similar computations, producing a number of identities, including the following, known as Machin's formula:

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}. \tag{2.4.41}$$

We now discuss another approach to approximating π. Namely, we try to solve

$$\sin x = 0, \tag{2.4.42}$$

by an iterative procedure starting from an initial guess that is known to be a rough approximation to π. The iteration is the following: given an approximation x_k, we set

$$x_{k+1} = x_k + \sin x_k. \tag{2.4.43}$$

This approximation is predicated on the fact that a rapid calculation of $\sin x_k$ is feasible. To analyze how much x_{k+1} is an improvement over x_k, set

$$x_k = \pi + \delta_k. \tag{2.4.44}$$

We only know x_k, not π and not δ_k, but note that

$$\sin x_k = \sin(\pi + \delta_k) = -\sin\delta_k, \tag{2.4.45}$$

so

$$x_{k+1} = \pi + \delta_k - \sin\delta_k. \tag{2.4.46}$$

Now the power series (2.3.7), applied to $e^{it} = \cos t + i\sin t$, implies

$$\sin\delta_k = \delta_k - \frac{1}{3!}\delta_k^3 + O(\delta_k^5), \tag{2.4.47}$$

so we get

$$x_{k+1} = \pi + \delta_{k+1}, \quad \delta_{k+1} = \delta_k - \sin\delta_k = O(\delta_k^3). \tag{2.4.48}$$

This implies very rapid convergence of the sequence (x_k) to its limit (which is π), provided the initial guess is sufficiently close.

To implement this iteration, we start with

$$x_0 = 3, \tag{2.4.49}$$

and implement (2.4.43), obtaining (to 15 digits),

$$\begin{aligned} x_1 &= 3.14112000805987, \\ x_2 &= 3.14159265357220, \\ x_3 &= 3.14159265358979. \end{aligned} \tag{2.4.50}$$

The error $\pi - x_2$ is $< 2 \cdot 10^{-11}$, and all the printed digits of x_3 are accurate. Indeed, complementing (2.4.48), we can deduce from the fact that $\sin\delta_k$ has alternating series that

$$|\delta_k| \le 1 \Longrightarrow |\delta_{k+1}| \le \frac{1}{6}|\delta_k^3|. \tag{2.4.51}$$

Consequently

$$\begin{aligned} |\delta_2| < 2 \cdot 10^{-11} &\Longrightarrow |\delta_3| < \frac{4}{3} \cdot 10^{-33} \\ &\Longrightarrow |\delta_4| < 4 \cdot 10^{-100}. \end{aligned} \tag{2.4.52}$$

For another implementation of (2.4.43), we start with the classical approximation $x_0 = 3\frac{1}{7}$:

$$\begin{aligned} x_0 = 3\tfrac{1}{7} &\Rightarrow \sin x_0 = -0.001264488930377 \\ &\Rightarrow x_1 = 3.141592653926765. \end{aligned} \tag{2.4.53}$$

In this case, $\pi - x_1 < 4 \cdot 10^{-10}$. For even more accuracy, add to x_0, not $\varepsilon_0 = \sin x_0$, but rather the first 2 terms in the power series of $\sin^{-1}\varepsilon_0$, yielding

$$x_1' = x_1 + \frac{1}{6}\varepsilon_0^3 = 3.141592653589793\cdots. \tag{2.4.54}$$

Regarding the ability to compute e^{ix} rapidly and accurately for $x \approx 3$, note that the power series (2.3.7) for e^{it} converges very fast, due the the $k!$ in the denominator, especially when $|t| < 1/2$. We can also write

$$e^{ix} = \alpha^8, \quad \alpha = e^{ix/8}, \quad \alpha^8 = ((\alpha^2)^2)^2. \tag{2.4.55}$$

Such devices are routinely used in the software for a computer program that has a command to compute $\sin x$ and $\cos x$.

Here is another wrinkle, once you have done the first iteration. If you have just computed x_k and already have $e^{ix_{k-1}}$, then write

$$e^{ix_k} = e^{ix_{k-1}} e^{i\gamma_k}, \quad \gamma_k = x_k - x_{k-1}, \tag{2.4.56}$$

and use the rapidly convergent power series for $e^{i\gamma_k}$.

Exercises

1. As a complement to the trigonometric functions sin and cos, we have the following hyperbolic functions:

$$\cosh t = \tfrac{1}{2}(e^t + e^{-t}), \quad \sinh t = \tfrac{1}{2}(e^t - e^{-t}).$$

show that

$$\cosh^2 t - \sinh^2 t = 1, \quad \frac{d}{dt}\cosh t = \sinh t, \quad \frac{d}{dt}\sinh t = \cosh t.$$

2. Exercise 1 of §2.2 produced the integral

$$\ell(t) = \int_0^t \sqrt{1+s^2}\, ds,$$

for the length of a segment of a parabola.

(a) Use the substitution $s = \tan u$, together with (2.4.26), to show that

$$\int_0^t \sqrt{1+s^2}\, ds = \int_0^v \sec^3 u\, du, \quad t = \tan v,$$

where $\sec u = (\cos u)^{-1}$.

(b) Use the substitution $s = \sinh x$ to show that

$$\int_0^t \sqrt{1+s^2}\, ds = \int_0^y \cosh^2 x\, dx, \quad t = \sinh y.$$

(c) Think about doing the integrals described in parts (a) and (b).
Hint for (b): $\cosh^2 x = \frac{1}{4}(e^{2x} + 2 + e^{-2x})$.

3. Parallel to Exercise 2, consider

$$I(t) = \int_0^t \frac{ds}{\sqrt{1+s^2}}.$$

Use the substitutions $s = \tan u$ and $s = \sinh x$ to obtain

$$I(t) = \int_0^v \sec u\, du, \quad t = \tan v,$$

and

$$I(t) = \int_0^y dx = y, \quad t = \sinh y.$$

Comparing these two, show that

$$\int_0^v \sec u\, du = \sinh^{-1}(\tan v).$$

4. Consider the second-order differential equation

$$u'' + u = 0. \tag{2.4.57}$$

(a) Show that two solutions are given by $u_+(t) = e^{it},\ u_-(t) = e^{-it}$.

(b) Show that two other solutions are given by $u_c(t) = \cos t,\ u_s(t) = \sin t$.

(c) Consider solutions to (2.4.57) with initial data

$$u(0) = a, \quad u'(0) = b. \tag{2.4.58}$$

Show that a solution is given by

$$v(t) = a\cos t + b\sin t.$$

5. Given a smooth function $u(t)$, consider

$$E(t) = |u(t)|^2 + |u'(t)|^2.$$

Show that

$$u'' + u = 0 \Longrightarrow E'(t) \equiv 0.$$

6. Deduce from Exercise 5 that if $u'' + u = 0$ and $v'' + v = 0$,

$$u(0) = v(0),\ u'(0) = v'(0) \Longrightarrow u(t) \equiv v(t).$$

7. Consider

$$u(t) = A\cos(t - \alpha), \quad A, \alpha \in \mathbb{R}. \tag{2.4.59}$$

Show that $u'' + u = 0$ and $u(0) = A\cos\alpha,\ u'(0) = A\sin\alpha$, hence Exercise 4(c) applies with $a + ib = Ae^{i\alpha}$. Going further, show that if (2.4.57)–(2.4.58) hold with $a, b \in \mathbb{R}$, then u has the form (2.4.59).

8. Define $\sin z$ and $\sinh z$ for $z \in \mathbb{C}$, and show that

$$\sinh iz = i\sin z.$$

Similarly treat $\cos z$ and $\cosh z$.

9. Take another look at Exercise 4 of §1.11, with

$$B = 1, \quad C = e^{i\theta}, \quad A = e^{i\varphi}.$$

Deduce the conclusion $\Omega_{\mathcal{T}}(A)^2 = \Omega_{\mathcal{T}_0}(0)$ stated there from the calculation

$$\frac{e^{i\varphi} - e^{i\theta}}{e^{i\varphi} - 1} = e^{i\theta/2}\,\frac{\sin(\varphi - \theta)/2}{\sin\varphi/2}.$$

2.5. Regular polygons

The regular polygon $\mathcal{P}_n$ is specified as having n vertices, in S^1, one vertex being 1, and as being invariant under the rotation ρ_n, given by

$$\rho_n(z) = \omega_n z, \quad \omega_n = e^{2\pi i/n}. \tag{2.5.1}$$

Then the complete set of vertices of $\mathcal{P}_n$ is

$$\omega_n^k = e^{2k\pi i/n}, \quad 0 \le k \le n-1. \tag{2.5.2}$$

These quantities are precisely the roots of the polynomial

$$\mathcal{P}_n(z) = z^n - 1. \tag{2.5.3}$$

Once one has the point ω_n, the remaining vertices can be constructed with a compass, using the observation that

$$|\omega_n^{k+1} - \omega_n^k| = |\omega_n - 1|. \tag{2.5.4}$$

One can start with the points $\omega_n^0 = 1$, $\omega_n^1 = \omega_n$, and inductively produce ω_n^{k+1} from ω_n^k as a point where the circle $S_{r_n}(\omega_n^k)$ intersects S^1, where $r_n = |\omega_n - 1|$.

Some of the basic vertices ω_n have been computed in §2.4. As seen there, we have

$$\begin{aligned} \omega_3 &= e^{2\pi i/3} = -\frac{1}{2} + \frac{i}{2}\sqrt{3}, \\ \omega_4 &= e^{\pi i/2} = i, \\ \omega_6 &= e^{\pi i/3} = \frac{1}{2} + \frac{i}{2}\sqrt{3}, \\ \omega_8 &= e^{\pi i/4} = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}. \end{aligned} \tag{2.5.5}$$

In the first three cases, one can find $x_n = \operatorname{Re}\omega_n$, an element of $\{0, \pm 1/2\}$, and read off $y_n = \operatorname{Im}\omega_n$ as the point where the line through x_n orthogonal to $\mathbb{R}$ intersects S^1_+. The number ω_8 can be obtained from ω_4 by angle bisection. Alternatively, we have $1/\sqrt{2} = \sqrt{2}/2$, and $\sqrt{2}$ can be constructed as the length of the hypothenuse of a right triangle with base and height 1.

The regular hexagon ($n = 6$) is illustrated in Figure 2.5.1. In this case we have

$$\omega_6 - 1 = \omega_3, \quad \text{hence} \quad |\omega_6 - 1| = 1, \tag{2.5.6}$$

and ω_6 can be constructed as the intersection of S^1_+ with $S_1(1)$. Note incidentally that

$$\omega_3 + 1 = \omega_6, \quad \text{hence} \quad |\omega_3 + 1| = 1, \tag{2.5.7}$$

so ω_3 can be constructed as the intersection of S^1_+ with $S_1(-1)$.

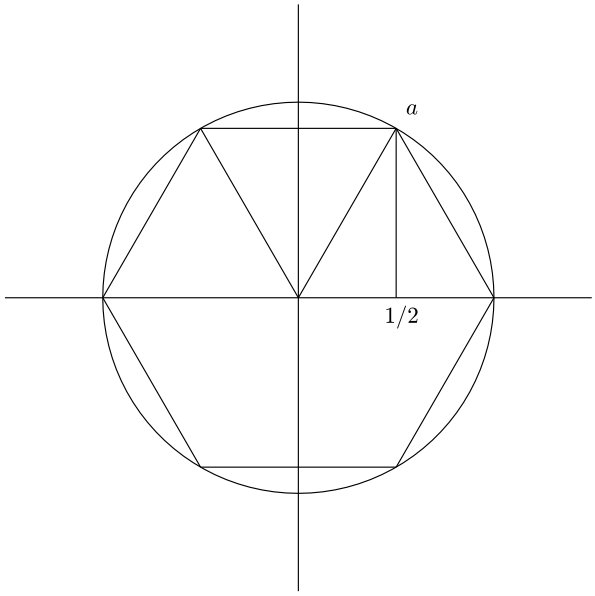

Figure 2.5.1. Regular hexagon, $a = e^{\pi i/3}$

There are two gaps in the sequence of n-gons to which the calculations in (2.5.5) apply, at $n = 5$ and $n = 7$. We discuss each of these in turn.

We first treat ω_5, which is a root of $p_5(z) = z^5 - 1$, leading to the construction of a regular pentagon. To start, write

$$z^5 - 1 = (z - 1)(z^4 + z^3 + z^2 + z + 1). \tag{2.5.8}$$

Our task is to find solutions to

$$z^4 + z^3 + z^2 + z + 1 = 0. \tag{2.5.9}$$

It is useful to divide by z^2, obtaining

$$z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} = 0. \tag{2.5.10}$$

Now a clever trick enters. Take

$$w = z + \frac{1}{z}, \quad \text{so} \quad w^2 = z^2 + 2 + \frac{1}{z^2}. \tag{2.5.11}$$

We substitute this into (2.5.10), obtaining

$$w^2 + w - 1 = 0. \tag{2.5.12}$$

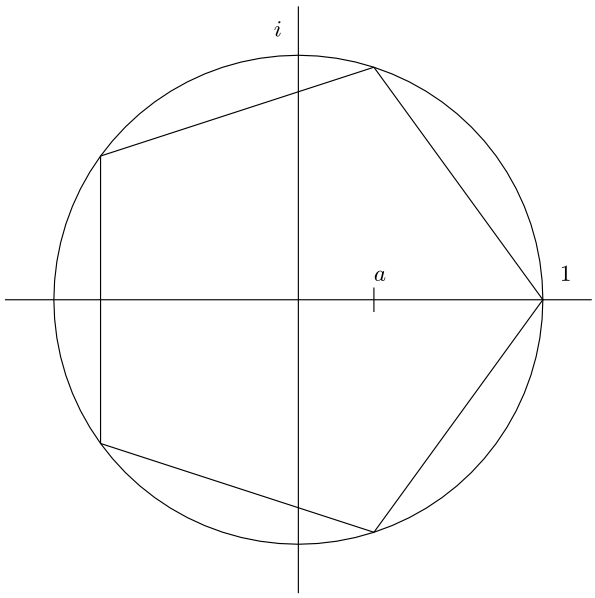

Figure 2.5.2. Regular pentagon, $a = (\sqrt{5}-1)/4$

The quadratic formula presents the roots as

$$w = -\frac{1}{2} \pm \frac{1}{2}\sqrt{5}. \tag{2.5.13}$$

Note that since $|z| = 1$, $w = z + \overline{z} = 2\operatorname{Re} z$. We deduce that

$$\operatorname{Re}\omega_5 = \cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}. \tag{2.5.14}$$

Going further, we can write $w = z + 1/z$ as

$$z^2 - wz + 1 = 0, \tag{2.5.15}$$

and solve for z. We leave this to the reader. Note that ω_5 can be obtained from (2.5.14) as the point where the line through $\operatorname{Re}\omega_5$ orthogonal to $\mathbb{R}$ intersects S^1_+. See Figure 2.5.2.

Let us see where such an approach leads when applied to ω_7, which is a root of $p_7(z) = z^7 - 1$, hence a solution to

$$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0. \tag{2.5.16}$$

Dividing by z^3 gives

$$z^3 + z^2 + z + 1 + z^{-1} + z^{-2} + z^{-3} = 0, \tag{2.5.17}$$

and again we take w as in (2.5.11). We complement (2.5.11) with

$$w^3 = z^3 + 3z + 3z^{-1} + z^{-3}, \tag{2.5.18}$$

and substitute into (2.5.17), obtaining

$$w^3 + w^2 - 2w - 1 = 0. \tag{2.5.19}$$

As opposed to (2.5.12), this is a cubic equation for w. Now there is a formula for the roots of such an equation (cf. [**17**], Appendix A.6), but of course it involves taking cube roots.

As seen in §1.11, square roots can be constructed via compass and straightedge. But it has been shown that in general cube roots cannot be so constructed, and in particular that ω_7 cannot be.

Numbers that can be constructed via compass and straightedge constitute a class that we consider further in §2.9, where the set of such numbers is denoted $\mathbb{E}$. The numbers $\omega_3, \omega_4, \omega_5$, and ω_6 belong to $\mathbb{E}$, but ω_7 does not. In fact, it has been shown that the regular n-gon $\mathcal{P}_n$ is constructible via compass and straightedge if and only if n has the form $n = 2^k$, or more generally,

$$n = 2^k \times \text{ a product of distinct Fermat primes}, \tag{2.5.20}$$

where a Fermat prime is a prime p of the form

$$p = 2^{2^k} + 1. \tag{2.5.21}$$

Examples include

$$2^1 + 1 = 3, \quad 2^2 + 1 = 5, \quad 2^4 + 1 = 17. \tag{2.5.22}$$

Such results are typically demonstrated using an area of algebra known as Galois theory. See [**9**].

A related problem that cannot be solved via compass and straightedge is that of trisecting an angle, that is, solving for $\alpha \in S^1$

$$\alpha^3 = \omega, \quad \omega \in S^1. \tag{2.5.23}$$

Of course, we can solve this using the complex exponential function. Given $\omega \in S^1$, we can write

$$\omega = e^{i\theta}, \quad -\pi < \theta \leq \pi, \tag{2.5.24}$$

and then take

$$\alpha = e^{i\theta/3}. \tag{2.5.25}$$

More generally, for each $n \in \mathbb{N}$,

$$\alpha = e^{i\theta/n} \text{ solves } \alpha^n = \omega. \tag{2.5.26}$$

The general solution to $\alpha^n = \omega$ is then

$$e^{i\theta/n + 2\pi i k/n}, \quad 0 \leq k \leq n - 1. \tag{2.5.27}$$

Here is an iterative procedure to solve

$$z^n = w \in \mathbb{C}. \tag{2.5.28}$$

Suppose you have an approximate solution z_k, such that

$$z_k^n = w + r_k = w(1 + \delta_k), \tag{2.5.29}$$

with $|\delta_k|$ sufficiently small. Then set

$$z_{k+1} = z_k\Big(1 + \frac{\delta_k}{n}\Big)^{-1}. \tag{2.5.30}$$

The binomial theorem gives

$$(1 + u)^n = 1 + nu + R_n(u), \quad |R_n(u)| \le C_n|u|^2, \text{ for } |u| \le 1, \tag{2.5.31}$$

hence

$$\Big(1 + \frac{\delta_k}{n}\Big)^n = 1 + \delta_k + R_n\Big(\frac{\delta_k}{n}\Big), \tag{2.5.32}$$

Consequently

$$z_{k+1}^n = w + r_{k+1} = w(1 + \delta_{k+1}), \quad \delta_{k+1} = O(|\delta_k|^2). \tag{2.5.33}$$

Proceed, to get a sequence converging rapidly to $w^{1/n}$, if $|\delta_k|$ is sufficiently small.

Example. Solve $z^3 = 2$, with initial guess $z_0 = 1$. You get (with $w_k = z_k^3$)

$$\begin{aligned} z_0 &= 1.0 && w_0 = 1.0 \\ z_1 &= 1.2 && w_1 = 1.728 \\ z_2 &= 1.25698\cdots && w_2 = 1.98603 \\ z_3 &= 1.2599141\cdots && w_3 = 1.9999673 \\ z_4 &= 1.259921049857\cdots && w_4 \approx 2 - 2\cdot 10^{-10} \\ z_5 &= 1.259921049894873 && |w_5 - 2| < 10^{-15}. \end{aligned} \tag{2.5.34}$$

In the next section, Proposition 2.6.9 describes a method to compute the area of a regular $2n$-gon $\mathcal{P}_{2n}$ from the area of $\mathcal{P}_n$, which involves taking square roots. The method (2.5.29)–(2.5.30), with $n = 2$, works well for computing such square roots.

Exercises

1. Relate the computation (2.5.14) of $\operatorname{Re}\omega_5$ to the golden ratio introduced in Exercise 2 of §1.10.

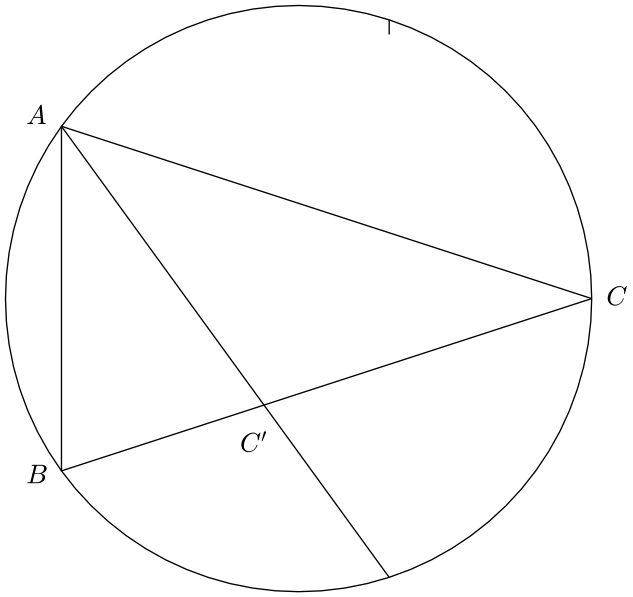

Figure 2.5.3. Golden triangle and similar companion

2. We introduce the concept of a golden triangle. Let the triangle $\mathcal{T} = \triangle(A, B, C)$ be isosceles at vertex C. We say $\mathcal{T}$ is a golden triangle provided

$$\angle_{\mathcal{T}}(A) = \angle_{\mathcal{T}}(B) = 2\angle_{\mathcal{T}}(C).$$

Show that this happens if and only if $\angle_{\mathcal{T}}(C) = \pm\pi/5$. Deduce that all golden triangles are similar.

3. Let $\omega = e^{2\pi i/5}$, so $\{1, \omega, \omega^2, \omega^3, \omega^4\}$ form the vertices of a regular pentagon. Show that

$$\mathcal{T} = \triangle(\omega^2, \omega^3, 1)$$

is a golden triangle.

4. Let $\mathcal{T} = \triangle(A, B, C)$ be a golden triangle (isosceles at vertex C). Let the angle bisector of A intersect the line segment $\mathcal{L}_{[B,C]}$ at C', and consider

$$\mathcal{T}' = \triangle(B, C', A).$$

See Figure 2.5.3. Show that $\mathcal{T}'$ is a golden triangle, similar to $\mathcal{T}$.
Bonus: find the regular pentagon in Figure 2.5.3.

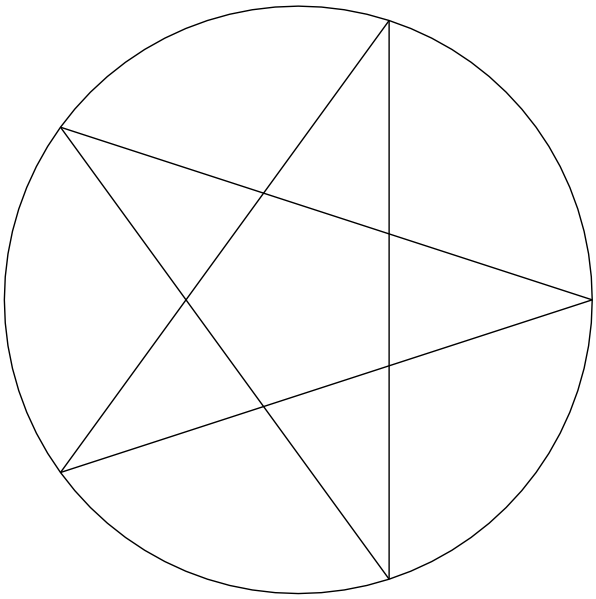

Figure 2.5.4. Pentagram

5. Figure 2.5.4 shows a regular pentagram. Find five golden triangles in this figure.

6. Show that, in the construction (2.5.29)–(2.5.33) of approximate solutions to $z^n = w$, we have

$$(1+\delta_{k+1})^{-1} = 1 + R_n\Big(\frac{\delta_k}{n}\Big)(1+\delta_k)^{-1},$$

hence

$$\frac{\delta_{k+1}}{1+\delta_{k+1}} = -R_n\Big(\frac{\delta_k}{n}\Big)(1+\delta_k)^{-1}.$$

Show that the estimate on $|R_n(u)|$ in (2.5.31) holds with $C_n = 2^n$.
Hint. $(1+1)^n = 2^n$.
Look for sharper estimates, valid for $|u| \leq 1/n$.
Hint. Show that

$$\Big|R_n\Big(\frac{\delta}{n}\Big)\Big| \leq |\delta|^2\Big(1+\frac{1}{n}\Big)^n, \quad \text{if } \ |\delta| \leq 1.$$

Then use

$$\left(1+\frac{1}{n}\right)^n = e^{n\log(1+1/n)}, \quad \log(1+x) = \int_0^x \frac{dt}{1+t}.$$

7. Solve $z^3 = 1 + i/2$, with initial guess $z_0 = 1 + i/6$. Find the solution to 6 decimal places. How might this relate to drawing a regular 9-gon?

2.6. Area

Here we present material on the area of various sets $S \subset \mathbb{C}$. To start, given $a > 0$, we say S is a standard square of sidelength a provided it has the form

$$S = \mathrm{Sq}_{z_0}(a) = \{x + iy : x_0 \le x \le x_0 + a,\ y_0 \le y \le y_0 + a\}, \tag{2.6.1}$$

for some $z_0 = x_0 + iy_0 \in \mathbb{C}$. If S is such a set, we say

$$\mathrm{Area}\, S = a^2. \tag{2.6.2}$$

To proceed, given $\varepsilon > 0$, we tile the plane $\mathbb{C}$ with squares of sidelength ε, given by

$$S_{jk\varepsilon} = \mathrm{Sq}_{(j+ki)\varepsilon}(\varepsilon), \quad j, k \in \mathbb{Z}. \tag{2.6.3}$$

Given a bounded $S \subset \mathbb{C}$, we say

$$\mathrm{Cont}^+(S) \le M\varepsilon^2 \tag{2.6.4}$$

if S is contained in a union of M squares of the form (2.6.3), and

$$\mathrm{Cont}^-(S) \ge N\varepsilon^2 \tag{2.6.5}$$

if S contains a union of N squares of the form (2.6.3).

We then set $\mathrm{Cont}^+(S)$ to be the infimum of the quantities arising in (2.6.4) and we set $\mathrm{Cont}^-(S)$ to be the supremum of the quantities arising in (2.6.5), as ε runs over $(0, 1]$. Clearly

$$\mathrm{Cont}^-(S) \le \mathrm{Cont}^+(S). \tag{2.6.6}$$

If these two quantities coincide, we say S is *contented*, and write

$$\mathrm{Area}\, S = \mathrm{Cont}^+(S) = \mathrm{Cont}^-(S). \tag{2.6.7}$$

One verifies that if S is a standard square with sidelength a, then the definition via (2.6.4)–(2.6.7) agrees with (2.6.2). Another easy check is that if R is a standard rectangle of sidelengths a and b, of the form

$$R = R_{z_0}(a, b) = \{x + iy : x_0 \le x \le x_0 + a, y_0 \le y \le y_0 + b\}, \tag{2.6.8}$$

then

$$\mathrm{Area}\, R = ab. \tag{2.6.9}$$

Here are some useful properties of the concepts defined above. Given a set S, let $\overline{S}$ denote the closure of S, consisting of all points $p \in \mathbb{C}$ such that for some $q_k \in S$, $q_k \to p$, and let $\overset{\circ}{S}$ denote the interior of S, consisting of points $p \in S$ such that for some $\varepsilon > 0$,

$$z \in \mathbb{C},\ |z - p| < \varepsilon \Longrightarrow z \in S. \tag{2.6.10}$$

Then

$$\mathrm{Cont}^+(S) = \mathrm{Cont}^+(\overline{S}), \quad \mathrm{Cont}^-(S) = \mathrm{Cont}^-(\overset{\circ}{S}). \tag{2.6.11}$$

Also, if R is a rectangle, as in (2.6.8), and

$$S \subset R, \quad T = R \setminus S, \tag{2.6.12}$$

then

$$\mathrm{Cont}^+(S) + \mathrm{Cont}^-(T) = \mathrm{Area}\, R. \tag{2.6.13}$$

Similarly $\mathrm{Cont}^-(S) + \mathrm{Cont}^+(T) = \mathrm{Area}\, R$, so

$$S \text{ contented} \Rightarrow T \text{ contented and } \mathrm{Area}\, S + \mathrm{Area}\, T = \mathrm{Area}\, R. \tag{2.6.14}$$

Another useful observation is that, if $S \subset \mathbb{C}$ is a bounded set,

$$\mathrm{Cont}^-(S) + \mathrm{Cont}^+(bS) \geq \mathrm{Cont}^+(S), \tag{2.6.15}$$

where $bS = \overline{S} \setminus \overset{\circ}{S}$. In particular,

$$\mathrm{Cont}^+(bS) = 0 \Longrightarrow S \text{ is contented.} \tag{2.6.16}$$

And another:

Proposition 2.6.1. *Given $S, T \subset \mathbb{C}$ bounded and contented,*

$$\mathrm{Cont}^+(S \cap T) = 0 \Longrightarrow \mathrm{Area}(S \cup T) = \mathrm{Area}\, S + \mathrm{Area}\, T. \tag{2.6.17}$$

We say a bounded set $S \subset \mathbb{C}$ is a nil set if $\mathrm{Cont}^+(S) = 0$. The union of any two nil sets is a nil set. This follows from the fact that, if S and $T \subset \mathbb{C}$ are bounded,

$$\mathrm{Cont}^+(S \cup T) \leq \mathrm{Cont}^+(S) + \mathrm{Cont}^+(T). \tag{2.6.18}$$

Here is a preliminary result on invariance of area under isometries.

Proposition 2.6.2. *Let $F(z) = az+b$ or $a\overline{z}+b$, and let $S \subset \mathbb{C}$ be a bounded set. Then*

$$\mathrm{Cont}^{\pm}(F(S)) = \mathrm{Cont}^{\pm}(S), \tag{2.6.19}$$

provided

$$a \in \{\pm 1, \pm i\}. \tag{2.6.20}$$

Proof. In case $b = 0$, such F preserves the tiling described in (2.6.3). It remains to treat $F(z) = z + b$, which we leave to the reader. □

We move to another class of figures for which we can compute areas. We call a triangle $\mathcal{T} = \triangle(A, B, C)$ a standard right triangle if it has a right angle at C and the sides meeting at C are parallel to the real and imaginary axes, i.e., $A - C \in i\mathbb{R}$, $B - C \in \mathbb{R}$.

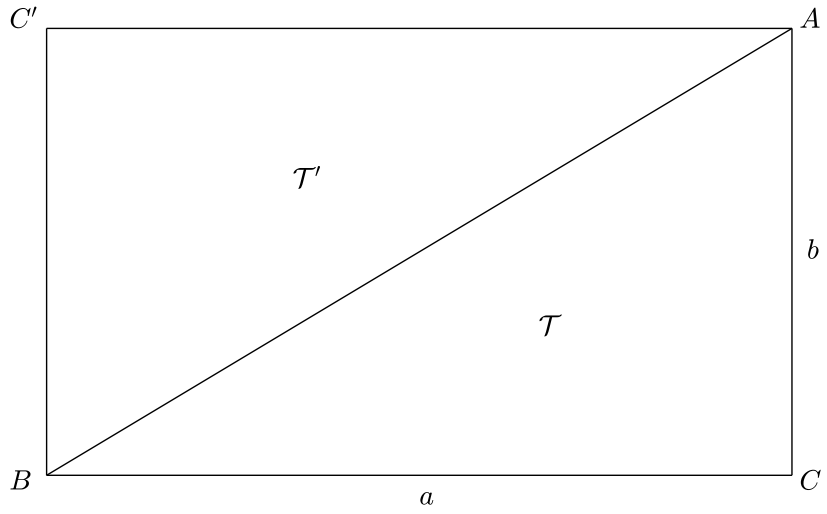

Figure 2.6.1. Finding the area of a standard right triangle

Proposition 2.6.3. *Let $\mathcal{T} = \triangle(A, B, C)$ be a standard right triangle, with sidelengths*

$$a = |B - C|, \quad b = |A - C|. \tag{2.6.21}$$

Then

$$\operatorname{Area} \mathcal{T} = \frac{1}{2}ab. \tag{2.6.22}$$

Proof. The triangle $\mathcal{T}$ has a twin, $\mathcal{T}' = \triangle(A, B, C')$, where

$$C' = C + (A - C) + (B - C). \tag{2.6.23}$$

See Figure 2.6.1. These two triangles are related by an isometry $F : \mathcal{T} \to \mathcal{T}'$ of the form considered in Proposition 2.6.2, with $a = -1$. Hence $\operatorname{Area} \mathcal{T}' = \operatorname{Area} \mathcal{T}$, and, by Proposition 2.6.1, $\operatorname{Area} \mathcal{T} + \operatorname{Area} \mathcal{T}'$ equals the area of a rectangle R with sidelengths a and b. This gives (2.6.22). □

As a preliminary to extending Proposition 2.6.2, we have the following.

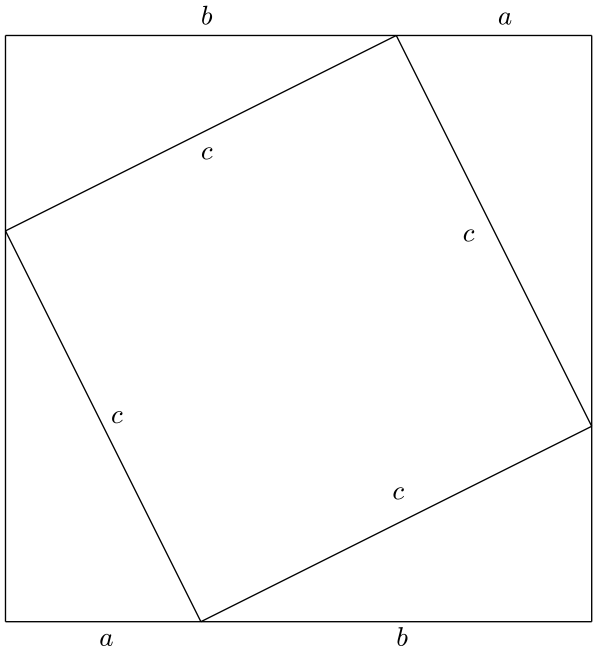

Figure 2.6.2. Area of an arbitrary square

Lemma 2.6.4. *If S_c is an arbitrary square of sidelength c (not necessarily standard), then*

$$\text{Area}\, S_c = c^2. \tag{2.6.24}$$

Proof. Translating S_c, we can assume S_c is obtained by applying a rotation to the standard square $S_c' = \{x + iy : |x|, |y| \le c/2\}$. If S_c itself is not standard, consider the horizontal lines through the top and bottom vertices of S_c, and the vertical lines through the left and right vertices of S_c. They form a rectangle R, in which S_c is inscribed. See Figure 2.6.2. This figure is invariant under the rotation $z \mapsto iz$, so R is in fact a standard square of sidelength $a + b$. R is divided into five figures; one is S_c, and the other four are standard right triangles, of sidelengths a and b. Therefore we have

$$\text{Area}\, R = \text{Area}\, S_c + 4 \cdot \frac{ab}{2}, \tag{2.6.25}$$

hence

$$(a + b)^2 = \text{Area}\, S_c + 2ab, \tag{2.6.26}$$

hence

$$a^2 + b^2 = \text{Area}\, S_c. \tag{2.6.27}$$

Hence (2.6.24) is equivalent to the Pythagorean identity $a^2 + b^2 = c^2$. □

Now for our extension of Proposition 2.6.2.

Proposition 2.6.5. *If $S \subset \mathbb{C}$ is bounded, then*

$$\text{Cont}^{\pm}(F(S)) = \text{Cont}^{\pm}(S), \quad \forall F \in \text{Isom}(\mathbb{C}). \tag{2.6.28}$$

Proof. It suffices to establish this when $F(z) = az$, $|a| = 1$, i.e., F is a rotation. In such a case, if we have a tiling $\{S_{jk\varepsilon} : j, k \in \mathbb{Z}\}$ of the form (2.6.3) and S is contained in the union of M of the squares in this tiling, and S contains N of the squares in this tiling, so (2.6.4) and (2.6.5) hold, it follows that $F(S)$ is contained in the union of M squares of the rotated tiling and contains N squares of this rotated tiling. Thanks to Lemma 2.6.4 (and also Proposition 2.6.1), we also have

$$\text{Cont}^{+}(F(S)) \leq M\varepsilon^2, \quad \text{Cont}^{-}(F(S)) \geq N\varepsilon^2. \tag{2.6.29}$$

In the limit we get

$$\text{Cont}^{+}(F(S)) \leq \text{Cont}^{+}(S), \quad \text{Cont}^{-}(F(S)) \geq \text{Cont}^{-}(S), \tag{2.6.30}$$

for all bounded S and $F \in \text{Isom}(\mathbb{C})$. Replacing S by $F(S)$ and F by F^{-1} gives the converse inequalities, and we have (2.6.28). □

Remark. Figure 2.6.2 plays a role in some popular proofs of the Pythagorean theorem, in which the flow of the argument is reversed, and one works under the hypothesis that (2.6.24) holds.

As a related matter, we mention that the proof of the Pythagorean theorem in Book I of Euclid uses as a key ingredient the proposition that congruent triangles have equal area. For us, the order of arguments is different. The Pythagorean theorem comes first.

We now bring calculus into the study of area, with the following key result.

Proposition 2.6.6. *Let $\varphi, \psi : [a, b] \to \mathbb{R}$ be continuous, and assume $\varphi \leq \psi$. Set*

$$\mathcal{O} = \{z = x + iy : a \leq x \leq b, \varphi(x) \leq y \leq \psi(x)\}. \tag{2.6.31}$$

Then $\mathcal{O}$ is contented, and

$$\text{Area}\, \mathcal{O} = \int_a^b [\psi(x) - \varphi(x)]\, dx. \tag{2.6.32}$$

Proof. Translating, we can assume $a = 0$. For $n \in \mathbb{N}$, tile $\mathbb{C}$ as in (2.6.3) with tiles of edgelength $\varepsilon = (b-a)/n$. Then considering collections of tiles that are contained in $\mathcal{O}$ but, with the addition of a minimum number of extra tiles, contain $\mathcal{O}$, we obtain two-sided bounds of the form

$$N\varepsilon^2 \leq \mathrm{Cont}^-(\mathcal{O}) \leq \mathrm{Cont}^+(\mathcal{O}) \leq M\varepsilon^2, \tag{2.6.33}$$

where $N\varepsilon^2$ and $M\varepsilon^2$ are, for ε sufficiently small, close to the Riemann sums of the form (2.1.2), which approximate

$$\int_a^b g(x)\,dx, \quad g(x) = \psi(x) - \varphi(x). \tag{2.6.34}$$

Passing to the limit gives (2.6.32) □

We can present the unit disk

$$D_1 = \{z \in \mathbb{C} : |z| \leq 1\} \tag{2.6.35}$$

in the form (2.6.31), with $a = -1$, $b = 1$,

$$\varphi(x) = -\sqrt{1-x^2}, \quad \psi(x) = \sqrt{1-x^2}, \tag{2.6.36}$$

to obtain

$$\mathrm{Area}\, D_1 = 2\int_{-1}^{1} \sqrt{1-x^2}\,dx. \tag{2.6.37}$$

To evaluate this integral, we can make a change of variable, $x = \sin t$, $-\pi/2 \leq t \leq \pi/2$, giving $dx = \cos t\,dt$, hence, by Proposition 2.1.5,

$$\mathrm{Area}\, D_1 = 2\int_{-\pi/2}^{\pi/2} \cos^2 t\,dt. \tag{2.6.38}$$

Using

$$\cos 2t = \cos^2 t - \sin^2 t = 2\cos^2 t - 1, \tag{2.6.39}$$

which follows from (2.3.37), we get

$$\mathrm{Area}\, D_1 = \int_{-\pi/2}^{\pi/2} (\cos 2t + 1)\,dt. \tag{2.6.40}$$

Furthermore, via (2.2.32),

$$\int_{-\pi/2}^{\pi/2} \cos 2t\,dt = \frac{\sin 2t}{2}\Big|_{-\pi/2}^{\pi/2} = 0. \tag{2.6.41}$$

Hence our conclusion:

$$\mathrm{Area}\, D_1 = \pi. \tag{2.6.42}$$

We aim to extend this result to a computation of the area of a "pie slice,"

$$\Pi_t = \{re^{is} : 0 \le r \le 1,\ 0 \le s \le t\}, \tag{2.6.43}$$

given $t \in [0, 2\pi]$. Here is the result.

Proposition 2.6.7. *For Π_t given by* (2.6.43), $0 \le t \le 2\pi$,

$$\operatorname{Area} \Pi_t = \frac{t}{2}. \tag{2.6.44}$$

Proof. Set

$$A(t) = \operatorname{Area} \Pi_t. \tag{2.6.45}$$

We have

$$A(s+t) = A(s) + A(t), \tag{2.6.46}$$

for $s, t \ge 0,\ s + t \le 2\pi$, as a consequence of Propositions 2.6.1 and 2.6.5. Hence it suffices to prove (2.6.44) for $t \in (0, \pi/2)$. Following Figure 2.6.3, consider the region

$$\mathcal{O}_\xi = \{x + iy : 0 \le x \le \xi,\ 0 \le y \le \sqrt{1 - x^2}\}, \quad \xi = \sin t. \tag{2.6.47}$$

The ray from 0 to $e^{it} \in S^1$ divides $\mathcal{O}_\xi$ into two parts, one a "pie slice," obtained from Π_t via a rotation, which, by Proposition 2.6.5 has the same area as Π_t. The other is a standard right triangle, with sidelengths $\sin t$ and $\cos t$. By Proposition 2.6.6,

$$\operatorname{Area} \mathcal{O}_\xi = B(\xi) = \int_0^\xi \sqrt{1 - x^2}\, dx, \quad B'(\xi) = \sqrt{1 - \xi^2}. \tag{2.6.48}$$

The division of $\mathcal{O}_\xi$ described above implies

$$B(\sin t) = A(t) + \frac{1}{2} \cos t \sin t. \tag{2.6.49}$$

Applying d/dt gives

$$\begin{aligned} A'(t) &= B'(\sin t) \cos t - \frac{1}{2}(\cos^2 t - \sin^2 t) \\ &= \frac{1}{2} \cos^2 t + \frac{1}{2} \sin^2 t \\ &= \frac{1}{2}. \end{aligned} \tag{2.6.50}$$

Since $A(0) = 0$, we have (2.6.44) □

Second proof. Given (2.6.46), we have

$$A'(s) = \lim_{t \to 0} \frac{A(t)}{t}, \tag{2.6.51}$$

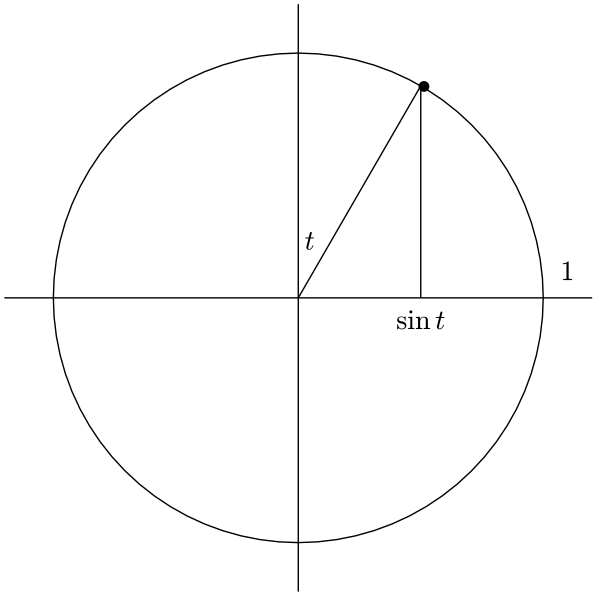

Figure 2.6.3. Area under a circle

if the limit on the right side exists. To investigate this, note that, if $0 < t < \pi/2$,

$$T_1(t) \subset \Pi_t \subset T_2(t), \tag{2.6.52}$$

where $T_1(t)$ is the triangle with vertices at $0,\ \cos t$, and $\cos t + i \sin t$, and $T_2(t)$ is the triangle with vertices at $0,\ 1$, and $1 + i \tan t$. Computing the areas of $T_j(t)$, we see that, for this range of t,

$$\frac{1}{2} \cos t \sin t \le A(t) \le \frac{1}{2} \tan t, \tag{2.6.53}$$

hence

$$\frac{1}{2}\, \cos t\, \frac{\sin t}{t} \le \frac{A(t)}{t} \le \frac{1}{2}\, \frac{1}{\cos t}\, \frac{\sin t}{t}. \tag{2.6.54}$$

By (2.2.32), $\sin t/t \to 1$ as $t \to 0$, so we have

$$\frac{A(t)}{t} \longrightarrow \frac{1}{2} \text{ as } t \to 0, \quad \text{hence} \quad A'(s) \equiv \frac{1}{2}. \tag{2.6.55}$$

This gives (2.6.44).

While the first proof of Proposition 2.6.7 used arguments that arose to show in (2.6.42) that $\operatorname{Area} D_1 = \pi$, the second proof used different arguments, providing an alternative approach to proving (2.6.42). We want to press these arguments further, and make contact with the approach of Archimedes to the numerical evaluation of π, in terms of evaluating the areas of regular polygons $\mathcal{P}_n$, defined in §2.5.

Areas of regular polygons and Archimedes' approximation to π

To start, note that $\mathcal{P}_n$ can be partitioned into n isosceles triangles, with two sides of length 1, meeting at a vertex of angle $2\pi/n$. Each of these in turn is composed of a union of two triangles congruent to $T_1(\pi/n)$. Consequently,

$$\begin{aligned} \operatorname{Area} \mathcal{P}_n &= 2n \operatorname{Area} T_1(\pi/n) \\ &= n \cos\frac{\pi}{n} \sin\frac{\pi}{n} \\ &= \frac{n}{2} \sin\frac{2\pi}{n}. \end{aligned} \tag{2.6.56}$$

The polygon $\mathcal{P}_n$ is inscribed in the circle S^1. Conversely, S^1 is inscribed in the regular n-gon $\widetilde{\mathcal{P}}_n$, consisting of $2n$ triangles congruent to $T_2(\pi/n)$, so

$$\begin{aligned} \operatorname{Area} \widetilde{\mathcal{P}}_n &= 2n \operatorname{Area} T_2(\pi/n) \\ &= n \tan\frac{\pi}{n} \\ &= \left(\cos\frac{\pi}{n}\right)^{-2} \operatorname{Area} \mathcal{P}_n. \end{aligned} \tag{2.6.57}$$

Since $\mathcal{P}_n \subset D_1 \subset \widetilde{\mathcal{P}}_n$, we deduce that

$$\begin{aligned} \operatorname{Area} \mathcal{P}_n < \pi < \left(\cos\frac{\pi}{n}\right)^{-2} \operatorname{Area} \mathcal{P}_n, \\ \operatorname{Area} \mathcal{P}_n = \frac{n}{2} \sin\frac{2\pi}{n}. \end{aligned} \tag{2.6.58}$$

As an aside, the conclusion that

$$\pi = \frac{n}{2} \sin\frac{2\pi}{n} + O(n^{-2}) \tag{2.6.59}$$

can be established directly from the power series

$$\sin\delta = \delta - \frac{\delta^3}{3!} + \frac{\delta^5}{5!} - \cdots. \tag{2.6.60}$$

Recall that we have algebraic expressions for $e^{2\pi i/n}$, hence for $\sin(2\pi/n)$, for various values of n, given in (2.5.5), which lead to formulas for $\operatorname{Area}\mathcal{P}_n$:

$$
\begin{aligned}
e^{2\pi i/3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i &\Rightarrow \operatorname{Area}\mathcal{P}_3 = \frac{3}{2}\sin\frac{2\pi}{3} = \frac{3}{4}\sqrt{3},\\
e^{\pi i/2} = i &\Rightarrow \operatorname{Area}\mathcal{P}_4 = 2\sin\frac{\pi}{2} = 2,\\
e^{\pi i/3} = \frac{1}{2} + \frac{\sqrt{3}}{2}i &\Rightarrow \operatorname{Area}\mathcal{P}_6 = 3\sin\frac{\pi}{3} = \frac{3}{2}\sqrt{3},\\
e^{\pi i/4} = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}} &\Rightarrow \operatorname{Area}\mathcal{P}_8 = 4\sin\frac{\pi}{4} = 2\sqrt{2},\\
e^{\pi i/6} = \frac{\sqrt{3}}{2} + \frac{i}{2} &\Rightarrow \operatorname{Area}\mathcal{P}_{12} = 6\sin\frac{\pi}{6} = 3.
\end{aligned}
\tag{2.6.61}
$$

Note also that

$$
\begin{aligned}
e^{\pi i/12} &= e^{\pi i/3}e^{-\pi i/4}\\
&= \Big(\frac{1}{2} + \frac{\sqrt{3}}{2}i\Big)\Big(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\Big)\\
&= \frac{1}{2\sqrt{2}}\Big[(\sqrt{3}+1) + (\sqrt{3}-1)i\Big],
\end{aligned}
\tag{2.6.62}
$$

hence

$$
\operatorname{Area}\mathcal{P}_{24} = 12\sin\frac{\pi}{12} = 3\sqrt{2}(\sqrt{3}-1) \approx 3.10582854\cdots . \tag{2.6.63}
$$

Rather than continue with special cases, we now seek a formula that allows us to pass from a computation of $\operatorname{Area}\mathcal{P}_n$ to that of $\operatorname{Area}\mathcal{P}_{2n}$. Assume we have

$$
S_n = \sin\frac{2\pi}{n}, \quad \operatorname{Area}\mathcal{P}_n = \frac{n}{2}S_n. \tag{2.6.64}
$$

We know that

$$
\operatorname{Area}\mathcal{P}_{2n} = n\sin\frac{\pi}{n}. \tag{2.6.65}
$$

The following formula is helpful.

Lemma 2.6.8. *For $|\alpha| < \pi$,*

$$
\sin\frac{\alpha}{2} = \frac{\sin\alpha}{\sqrt{2+2\cos\alpha}}. \tag{2.6.66}
$$

Proof. By Proposition 1.5.1,

$$
e^{i\alpha/2} = \frac{1+e^{i\alpha}}{|1+e^{i\alpha}|}, \tag{2.6.67}
$$

and taking imaginary parts yields (2.6.66). □

This leads to the following conclusion.

Proposition 2.6.9. *The areas of $\mathcal{P}_n$ and $\mathcal{P}_{2n}$ are related as follows:*

$$\begin{aligned}\operatorname{Area}\mathcal{P}_{2n} &= n\,\frac{\sin\frac{2\pi}{n}}{\sqrt{2+2\cos\frac{2\pi}{n}}}\\ &= \Big(\operatorname{Area}\mathcal{P}_n\Big)\frac{2}{\sqrt{2+2C_n}},\end{aligned}\tag{2.6.68}$$

with

$$\begin{aligned}C_n &= \cos\frac{2\pi}{n} = \sqrt{1-S_n^2}\\ &= \sqrt{1-\Big(\frac{2}{n}\operatorname{Area}\mathcal{P}_n\Big)^2}.\end{aligned}\tag{2.6.69}$$

This gives an iterative method to evaluate $\operatorname{Area}\mathcal{P}_n$ with $n = m\cdot 2^k$, starting with an evaluation of $\operatorname{Area}\mathcal{P}_m$. For example, proceeding from the evaluation of $\operatorname{Area}\mathcal{P}_{24}$ in (2.6.63), the iteration yields the following:

$$\begin{aligned}\operatorname{Area}\mathcal{P}_{48} &= 3.13262861\cdots\\ \operatorname{Area}\mathcal{P}_{96} &= 3.13935020\cdots\\ \operatorname{Area}\mathcal{P}_{192} &= 3.14103195\cdots\\ \operatorname{Area}\mathcal{P}_{384} &= 3.14145247\cdots.\end{aligned}\tag{2.6.70}$$

Archimedes pursued such an iteration and evaluated $\operatorname{Area}\mathcal{P}_n$ for $n = 6\cdot 2^5 = 192$, presenting his findings in the form

$$3\tfrac{10}{71} < \pi < 3\tfrac{1}{7}.\tag{2.6.71}$$

Actually, he considered $\mathcal{P}_{96}$ and evaluated the length $\ell(\mathcal{P}_{96})$ of its perimeter, but one readily verifies that

$$\ell(\mathcal{P}_n) = n\,|e^{\pi i/n} - e^{-\pi i/n}| = 2\operatorname{Area}\mathcal{P}_{2n},\tag{2.6.72}$$

and in particular $\ell(\mathcal{P}_{96}) = 2\operatorname{Area}\mathcal{P}_{192}$. Noting that

$$3\tfrac{10}{71} = 3.14084507\cdots,\tag{2.6.73}$$

we see that the figure for $\operatorname{Area}\mathcal{P}_{192}$ in (2.6.70) matches $3\frac{10}{71}$ quite closely.

We look more closely at the result (2.6.71), which we write in decimal form as

$$3.14084507 < \pi < 3.14285714,\tag{2.6.74}$$

and compare with what we get from the formula for $\operatorname{Area}\mathcal{P}_{192}$ in (2.6.70), in conjunction with the estimate in (2.6.58):

$$3.14103195 < \pi < 3.14187305.\tag{2.6.75}$$

As already mentioned, the lower estimates are quite close. However, the upper estimate in (2.6.75) is significantly better than Archimedes' upper estimate in (2.6.74). This is explained because Archimedes used

$$\ell(\mathcal{P}_n) < 2\pi < \ell(\widetilde{\mathcal{P}}_n) = \Big(\cos\frac{\pi}{n}\Big)^{-1}\ell(\mathcal{P}_n). \tag{2.6.76}$$

Indeed, use of (2.6.76), with $n = 96$, gives

$$\pi < 3.14271459, \tag{2.6.77}$$

which is quite close to Archimedes' upper estimate in (2.6.74). On the other hand, the identity (2.6.72), in conjunction with (2.6.58), yields

$$\ell(\mathcal{P}_n) < 2\pi < \Big(\cos\frac{\pi}{2n}\Big)^{-2}\ell(\mathcal{P}_n), \tag{2.6.78}$$

which is sharper than (2.6.76), since

$$\cos\frac{\pi}{n} = \cos^2\frac{\pi}{2n} - \sin^2\frac{\pi}{2n} < \cos^2\frac{\pi}{2n}. \tag{2.6.79}$$

We derive further estimates on the difference $\pi - \operatorname{Area}\mathcal{P}_n$. From (2.6.58) we obtain

$$\begin{aligned} 0 < \pi - \operatorname{Area}\mathcal{P}_n &< \Big[\Big(\cos\frac{\pi}{n}\Big)^{-2} - 1\Big]\operatorname{Area}\mathcal{P}_n \\ &= \tan^2\frac{\pi}{n}\,\operatorname{Area}\mathcal{P}_n \\ &< \pi\tan^2\frac{\pi}{n}. \end{aligned} \tag{2.6.80}$$

On the other hand, we also have

$$\pi - \operatorname{Area}\mathcal{P}_n = \frac{n}{2}\Big(\frac{2\pi}{n} - \sin\frac{2\pi}{n}\Big), \tag{2.6.81}$$

and the power series (2.6.60), with $\delta = 2\pi/n$, yields

$$\pi - \operatorname{Area}\mathcal{P}_n = \frac{n}{2}\Big[\frac{1}{3!}\Big(\frac{2\pi}{n}\Big)^3 - \frac{1}{5!}\Big(\frac{2\pi}{n}\Big)^5 + \cdots\Big] < \frac{2}{3}\,\frac{\pi^3}{n^2}. \tag{2.6.82}$$

the last estimate holding for $n \geq 2$. We see that the estimate (2.6.82) is stronger than (2.6.80). We have

$$\begin{aligned} A_k = \operatorname{Area}\mathcal{P}_{2^k} &\Rightarrow 0 < \pi - A_k < \frac{2\pi^3}{3}\cdot 2^{-2k}, \\ A_k^{\#} = \operatorname{Area}\mathcal{P}_{3\cdot 2^k} &\Rightarrow 0 < \pi - A_k^{\#} < \frac{2\pi^3}{27}\cdot 2^{-2k}. \end{aligned} \tag{2.6.83}$$

We can show that

$$\frac{2\pi^3}{3} < 21, \quad \frac{2\pi^3}{27} < 2.3. \tag{2.6.84}$$

We test these error estimates with further specific calculations of Area $\mathcal{P}_n$. Starting with $A_2 = \text{Area}\,\mathcal{P}_4 = 2$, we can use Proposition 2.6.9 repeatedly to compute A_k for $k \geq 3$. Results include

$$\begin{aligned} A_5 &= 3.124\cdots \\ A_{10} &= 3.14157\cdots \\ A_{15} &= 3.14159263\cdots \\ A_{20} &= 3.14159265357\cdots \\ A_{25} &= 3.14159265358978\cdots . \end{aligned} \tag{2.6.85}$$

From both (2.6.70) and (2.6.85), we see that the error estimates from (2.6.83)–(2.6.84) are quite sharp. Consistent with this, we see that a more accurate approximation to π than Area $\mathcal{P}_n$ can arise from writing

$$\pi - \text{Area}\,\mathcal{P}_n = \frac{2}{3}\,\frac{\pi^3}{n^2} - R_n, \tag{2.6.86}$$

with

$$R_n = \frac{n}{2}\Big[\frac{1}{5!}\Big(\frac{2\pi}{n}\Big)^5 - \frac{1}{7!}\Big(\frac{2\pi}{n}\Big)^7 + \cdots\Big] < \frac{1}{15}\,\frac{\pi^5}{n^4}, \tag{2.6.87}$$

and use $\text{Area}\,\mathcal{P}_n = \pi + O(n^{-2})$ to write

$$\pi = \text{Area}\,\mathcal{P}_n + \frac{2}{3n^2}\Big(\text{Area}\,\mathcal{P}_n\Big)^3 + O(n^{-4}). \tag{2.6.88}$$

If we denote the main part of the right side of (2.6.88) as $B(\mathcal{P}_n)$, we see that

$$\begin{aligned} &\text{Area}\,\mathcal{P}_{12} = 3 \Rightarrow B(\mathcal{P}_{12}) = 3.125, \\ &\text{Area}\,\mathcal{P}_{24} = 3\sqrt{2}(\sqrt{3}-1) \Rightarrow B(\mathcal{P}_{24}) = 3.1405\cdots . \end{aligned} \tag{2.6.89}$$

The quantity $B(\mathcal{P}_{24})$ is approximately as accurate as Archimedes' approximation (2.6.71). Going further, we have

$$B(\mathcal{P}_{192}) = 3.14159238\cdots . \tag{2.6.90}$$

Having just tweaked the approximation $\pi \approx \text{Area}\,\mathcal{P}_n$, we can hardly resist going all the way, writing (2.6.56) as

$$\sin\frac{2\pi}{n} = \frac{2}{n}\,\text{Area}\,\mathcal{P}_n, \tag{2.6.91}$$

hence

$$\pi = \frac{n}{2}\sin^{-1}\Big(\frac{2}{n}\,\text{Area}\,\mathcal{P}_n\Big), \tag{2.6.92}$$

and bringing in the power series (2.4.15):

$$\sin^{-1} t = \sum_{k=0}^{\infty} \frac{a_k}{k!}\,\frac{t^{2k+1}}{2k+1}, \tag{2.6.93}$$

with a_k given by (2.4.13). Consequently, for each $n \geq 3$,

$$\pi = \Big(\operatorname{Area} \mathcal{P}_n\Big) \sum_{k=0}^{\infty} \frac{a_k}{k!(2k+1)} \Big(\frac{2}{n} \operatorname{Area} \mathcal{P}_n\Big)^{2k}. \tag{2.6.94}$$

For $n = 12$, $\operatorname{Area} \mathcal{P}_{12} = 3$, this specializes to (2.4.16).

Note that all the terms in the series (2.6.94) are positive. To get a good upper bound on π from a partial sum, say over $0 \leq k \leq N-1$, note that

$$b_k = \frac{a_k}{k!} \Rightarrow b_{k+1} = \frac{k+1/2}{k+1} b_k \Rightarrow b_k \searrow, \tag{2.6.95}$$

hence, for $0 < t < 1$,

$$\begin{aligned} \sum_{k=N}^{\infty} \frac{a_k}{k!} \frac{t^{2k}}{2k+1} &\leq \frac{a_N}{N!(2N+1)} \sum_{k=N}^{\infty} t^{2k} \\ &= \frac{a_N}{N!(2N+1)} \frac{t^{2N}}{1-t^2}. \end{aligned} \tag{2.6.96}$$

For example, summing over $k \in \{0, 1\}$ yields the quantity $B(\mathcal{P}_n)$ given in (2.6.88), and we have

$$\begin{aligned} 0 < \pi - B(\mathcal{P}_n) &< \frac{a_2}{10} \Big(\operatorname{Area} \mathcal{P}_n\Big) \frac{t^4}{1-t^2}, \\ t = \frac{2}{n} \operatorname{Area} \mathcal{P}_n, &\quad a_2 = \frac{3}{4}. \end{aligned} \tag{2.6.97}$$

Remark. For $n = 4$, we have $t = 1$, the series (2.6.94) converges very slowly, and the upper bounds in (2.6.96)–(2.6.97) have no content. One might seek viable upper bounds in this case.
Hint. Use (2.6.95) to get

$$\log \frac{1}{b_k} = \sum_{\ell=0}^{k-1} \log\Big(1 + \frac{1}{2\ell+1}\Big) \approx \sum_{\ell=0}^{k-1} \frac{1}{2\ell+1} \approx \frac{1}{2} \log k. \tag{2.6.98}$$

As a parting comment on the use of (2.6.94), we write out the first three terms in the series arising there,

$$p_3(t) = 1 + \frac{1}{6} t^2 + \frac{3}{40} t^4, \tag{2.6.99}$$

and evaluate the associated partial sum of (2.6.94), in case $n = 24$, with $\operatorname{Area} \mathcal{P}_{24}$ given by (2.6.63), obtaining

$$\Big(\operatorname{Area} \mathcal{P}_{24}\Big) p_3\Big(\frac{1}{12} \operatorname{Area} \mathcal{P}_{24}\Big) = 3.14154897\cdots, \tag{2.6.100}$$

a somewhat better approximation to π than those given in (2.6.70).

Returning to generalities, we provide the following complement to Proposition 2.6.5.

Proposition 2.6.10. *Given $a, b > 0$, let $F_{a,b} : \mathbb{C} \to \mathbb{C}$ be given by*

$$F_{a,b}(x + iy) = ax + iby. \tag{2.6.101}$$

Than, if $S \subset \mathbb{C}$ is bounded,

$$\operatorname{Cont}^{\pm}(F_{a,b}(S)) = ab\, \operatorname{Cont}^{\pm}(S). \tag{2.6.102}$$

Proof. Following the proof of Proposition 2.6.5, if we have a tiling $\{S_{jk\varepsilon} : j, k \in \mathbb{Z}\}$ of $\mathbb{C}$ of the form (2.6.3) and S is contained in the union of M of the squares in this tiling and S contains N of the squares in this tiling, so (2.6.4) and (2.6.5) hold, it follows that $F_{a,b}(S)$ is contained in the union of M rectangles of the image tiling, and contains N of such rectangles, each of which has area $ab\varepsilon^2$, by (2.6.9). Hence, parallel to (2.6.29), we have

$$\operatorname{Cont}^{+}(F_{a,b}(S)) \le abM\varepsilon^2, \quad \operatorname{Cont}^{-}(F_{a,b}(S)) \ge abN\varepsilon^2. \tag{2.6.103}$$

In the limit we get

$$\operatorname{Cont}^{+}(F_{a,b}(S)) \le ab \operatorname{Cont}^{+}(S), \quad \operatorname{Cont}^{-}(F_{a,b}(S)) \ge ab \operatorname{Cont}^{-}(S), \tag{2.6.104}$$

for all bounded S, and all $F_{a,b}$ of the form (2.6.101). Replacing S by $S_{a,b} = F_{a,b}(S)$ and $F_{a,b}$ by $F_{a,b}^{-1} = F_{1/a,1/b}$, we have (2.6.102). □

As an example, we note that if $a, b > 0$, the region bounded by the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{2.6.105}$$

is

$$E_{a,b} = F_{a,b}(D_1), \tag{2.6.106}$$

where D_1 is the unit disk (2.6.35). Hence, from (2.6.42) and Proposition 2.6.10 we get

$$\operatorname{Area} E_{a,b} = \pi ab. \tag{2.6.107}$$

Exercises

1. How many terms in (2.6.94) are needed to approximate π to 10 digits, given $n = 24$? Compare with the figure for $n = 12$.

2. Compute the area of the region bounded from above by the arc $y = \sqrt{1 - x^2}$ of the unit circle and from below by the parabola $y = x^2$.

3. Verify the identity (2.6.72): $\ell(\mathcal{P}_n) = 2\,\mathrm{Area}(\mathcal{P}_{2n})$.

4. Establish the following variant of (2.6.68).

$$\frac{\mathrm{Area}\,\mathcal{P}_n}{\mathrm{Area}\,\mathcal{P}_{2n}} = \cos\frac{\pi}{n}. \tag{2.6.108}$$

Using

$$\mathrm{Area}\,\mathcal{P}_{2^k} = \mathrm{Area}\,\mathcal{P}_4 \cdot \frac{\mathrm{Area}\,\mathcal{P}_8}{\mathrm{Area}\,\mathcal{P}_4} \cdots \frac{\mathrm{Area}\,\mathcal{P}_{2^k}}{\mathrm{Area}\,\mathcal{P}_{2^{k-1}}},$$

deduce that

$$\frac{2}{\pi} = \lim_{k\to\infty} \cos\frac{\pi}{4} \cdot \cos\frac{\pi}{8} \cdot \cos\frac{\pi}{16} \cdots \cos\frac{\pi}{2^k}.$$

This is known as the formula of Vieta.

5. Deduce from the identity $\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1$ that

$$\cos\frac{\alpha}{2} = \frac{1}{2}\sqrt{2 + 2\cos\alpha}.$$

Relate this to the Vieta formula.

6. Rewrite the Vieta formula as

$$\frac{4}{\pi^2} = \prod_{k=2}^{\infty}\Big(1 - \sin^2\frac{\pi}{2^k}\Big).$$

Note that

$$\sin^2\frac{\alpha}{2} = \frac{\sin^2\alpha}{2 + 2\cos\alpha}, \quad \cos\alpha = \sqrt{1 - \sin^2\alpha}.$$

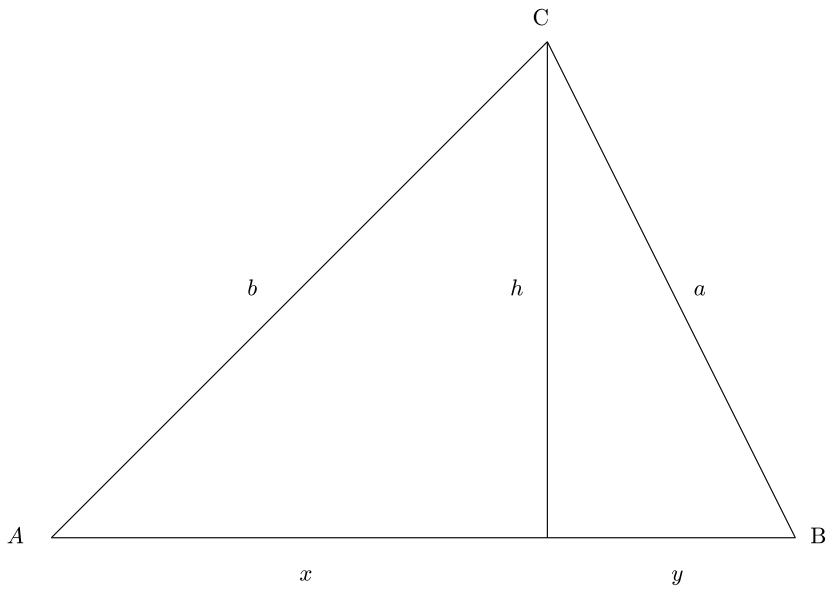

Figure 2.7.1. Setup for computing Area $\mathcal{T}$

2.7. Heron's formula

Let $\mathcal{T} \subset \mathbb{C}$ be a triangle, with vertices A, B, C, having opposing sides of length a, b, c, respectively. Set

$$s = \frac{1}{2}(a + b + c). \tag{2.7.1}$$

Heron's formula for the area of $\mathcal{T}$ is the following

$$(\text{Area}\,\mathcal{T})^2 = s(s-a)(s-b)(s-c). \tag{2.7.2}$$

To set up a proof, recall from Proposition 1.6.4 that at most one angle of $\mathcal{T}$ is obtuse. We can assume A and B are acute. Drop a perpendicular from C to the line segment from A to B, dividing this segment into two pieces, of length x and y, so

$$x + y = c. \tag{2.7.3}$$

See Figure 2.7.1. Let h denote the length of this perpendicular line segment. Hence

$$\text{Area}\,\mathcal{T} = \frac{1}{2}ch. \tag{2.7.4}$$

Meanwhile, the Pythagorean theorem yields

$$h^2 + x^2 = b^2, \quad h^2 + y^2 = a^2. \tag{2.7.5}$$

We aim to deduce (2.7.2) from (2.7.3)–(2.7.5).

Subtracting in (2.7.5) gives $x^2 - y^2 = b^2 - a^2$, and since

$$x^2 - y^2 = (x + y)(x - y) = c(x - y),$$

we have

$$x - y = \frac{b^2 - a^2}{c}. \tag{2.7.6}$$

Adding (2.7.3) and (2.7.6) gives

$$x = \frac{c^2 + b^2 - a^2}{2c}, \tag{2.7.7}$$

and hence from (2.7.5) we get

$$h^2 = b^2 - x^2 = b^2 - \frac{(c^2 + b^2 - a^2)^2}{4c^2}. \tag{2.7.8}$$

Therefore,

$$\begin{aligned} 4c^2h^2 &= (2bc)^2 - (c^2 + b^2 - a^2)^2 \\ &= (2bc + c^2 + b^2 - a^2)(2bc - c^2 - b^2 + a^2) \\ &= [(b + c)^2 - a^2] \cdot [a^2 - (b - c)^2] \\ &= (a + b + c)(-a + b + c)(a + b - c)(a - b + c) \\ &= 16s(s - a)(s - b)(s - c). \end{aligned} \tag{2.7.9}$$

By (2.7.4),

$$(\text{Area}\,\mathcal{T})^2 = \frac{c^2h^2}{4}, \tag{2.7.10}$$

and we have the desired identity (2.7.2).

Exercises

1. Test Heron's formula on the computation of the area of an equilateral triangle, say of sidelengths $a = b = c = 1$.

2.8. Making a trig table

Work of Archimedes discussed in §2.6 led to the production of trigonometric tables, including tabulations of $\sin \ell^\circ$ and $\cos \ell^\circ$ for integers ℓ. Here ℓ° stands for $\ell/360$ parts of a circle, i.e., $(\pi/180)\ell$ radians, so

$$\cos \ell^\circ + i \sin \ell^\circ = e^{i\theta_\ell}, \quad \theta_\ell = \frac{\pi \ell}{180}. \tag{2.8.1}$$

We do not want to just plug θ_ℓ into the power series for $e^{i\theta}$. We want to compute these quantities without bringing in a computed value of π. Archimedes himself might have worked on such a table, in his *Book of Circles*, but that work is mostly lost. What is known is later work of Ptolemy (see [**23**], pp. 18–22).

We discuss formulas for (2.8.1), starting with the identities

$$e^{\pi i/3} = \frac{1}{2}(1 + i\sqrt{3}), \quad e^{\pi i/4} = \frac{\sqrt{2}}{2}(1 + i), \tag{2.8.2}$$

obtained in (2.4.3) and (2.4.8), supplemented by

$$e^{2\pi i/5} = c_5 + i s_5, \quad c_5 = \frac{1}{4}(\sqrt{5} - 1), \tag{2.8.3}$$

obtained in (2.5.14), which in turn yields

$$s_5 = \sqrt{1 - c_5^2} = \frac{1}{4}\sqrt{10 + 2\sqrt{5}}. \tag{2.8.4}$$

Numerical evaluation of the real and imaginary parts of the quantities in (2.8.3)–(2.8.4) hence involves numerical evaluation of

$$\sqrt{2}, \quad \sqrt{3}, \quad \sqrt{5}, \quad \sqrt{10 + 2\sqrt{5}}. \tag{2.8.5}$$

One approach to the evaluation of $\sqrt{a}$ (indeed, of $a^{1/n}$ for general $n \in \mathbb{N}$) was given in (2.5.28)–(2.5.30). Another goes as follows. Suppose you have an approximation x_k to $\sqrt{a}$:

$$x_k - \sqrt{a} = \delta_k. \tag{2.8.6}$$

Square this to obtain $x_k^2 + a - 2x_k\sqrt{a} = \delta_k^2$, hence

$$\sqrt{a} = x_{k+1} - \frac{\delta_k^2}{2x_k}, \quad x_{k+1} = \frac{x_k^2 + a}{2x_k} = \frac{1}{2}\Big(x_k + \frac{a}{x_k}\Big). \tag{2.8.7}$$

Then x_{k+1} is an improved approximation, as long as $|\delta_k| < 2x_k$, and iteration produces a rapidly convergent sequence. In connection with this, we note that if $a, x_k > 0$, then $\sqrt{a}$ lies between x_k and a/x_k, hence

$$|\delta_k| \le \Big|x_k - \frac{a}{x_k}\Big| = \frac{|x_k^2 - a|}{x_k}. \tag{2.8.8}$$

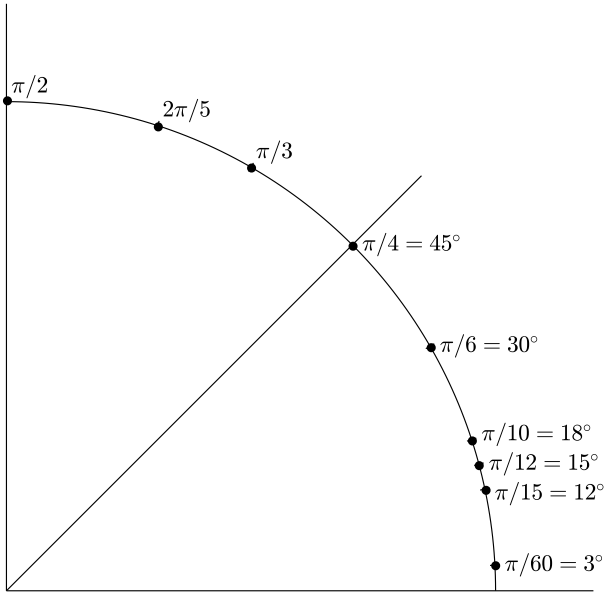

Figure 2.8.1. Special angles θ, at which to evaluate $\sin\theta$ and $\cos\theta$

To get started on (2.8.5), one can use

$$\sqrt{2} \approx \frac{7}{5}, \quad \sqrt{3} \approx \frac{7}{4}, \quad \sqrt{5} \approx \frac{9}{4}, \tag{2.8.9}$$

which one can validate by squaring both sides. To start an approximation to the last quantity in (2.8.5), take $10 + 2\sqrt{5} \approx 14.5$, hence

$$\sqrt{10 + 2\sqrt{5}} \approx 3.8. \tag{2.8.10}$$

Remark. The iteration (2.8.7) is a bit different from (2.5.30), which for $n = 2$ can be written

$$\tilde{x}_{k+1} = \frac{2ax_k}{a + x_k^2}, \quad \Big(\tilde{x}_{k+1} = \frac{a}{x_{k+1}}\Big). \tag{2.8.11}$$

Both do essentially equally good jobs of approximating $\sqrt{a}$. Note that

$$\tilde{x}_{k+1} \leq \sqrt{a} \leq x_{k+1}, \tag{2.8.12}$$

so we get reliable two-sided bounds. Note also that

$$\frac{1}{2}(x_{k+1} + \tilde{x}_{k+1}) = x_{k+2}. \tag{2.8.13}$$

Now, algebraic operations yield the following identities:

$$\begin{aligned} e^{\pi i/2}e^{-\pi i/3} &= e^{\pi i/6}, \\ e^{\pi i/2}e^{-2\pi i/5} &= e^{\pi i/10}, \\ e^{\pi i/3}e^{-\pi i/4} &= e^{\pi i/12}, \\ e^{2\pi i/5}e^{-\pi i/3} &= e^{\pi i/15}, \\ e^{\pi i/12}e^{-\pi i/15} &= e^{\pi i/60}. \end{aligned} \tag{2.8.14}$$

See Figure 2.8.1 for representations of the relevant angles. Regarding the last quantity, we also have

$$\begin{aligned} e^{\pi i/60} &= e^{2\pi i/3}e^{-\pi i/4}e^{-2\pi i/5} \\ &= \frac{\sqrt{2}}{4}(-1+i\sqrt{3})(1-i)(c_5 - is_5). \end{aligned} \tag{2.8.15}$$

It follows that

$$2\sqrt{2}\sin\frac{\pi}{60} = (\sqrt{3}+1)c_5 - (\sqrt{3}-1)s_5. \tag{2.8.16}$$

From evaluations of (2.8.5), we get

$$\begin{aligned} \sin 3^\circ = \sin\frac{\pi}{60} &= 0.052335956242\cdots, \\ \cos 3^\circ = \cos\frac{\pi}{60} &= 0.998629534754\cdots. \end{aligned} \tag{2.8.17}$$

Using the half-angle formula (2.6.66), we get numerical evaluations of $\sin\frac{3}{2}^\circ$ and $\sin\frac{3}{4}^\circ$. We display the results as follows:

$$\begin{aligned} 120\sin\frac{\pi}{120} &= 3.14123379\cdots, \\ 240\sin\frac{\pi}{240} &= 3.14150293\cdots. \end{aligned} \tag{2.8.18}$$

We turn to the evaluation of $\sin 1^\circ$ and $\cos 1^\circ$. We will use the identity

$$(\cos 1^\circ + i\sin 1^\circ)^3 = \cos 3^\circ + i\sin 3^\circ. \tag{2.8.19}$$

But before getting to this, we follow [**23**] and discuss how Ptolemy used (2.8.18) to give an approximate evaluation of $\sin 1^\circ$, based on the following result.

Proposition 2.8.1. *The function*

$$f(t) = \frac{\sin t}{t} \tag{2.8.20}$$

is monotone decreasing, for $0 < t < \pi$.

Proof. We evaluate the derivative:

$$f'(t) = \frac{t\cos t - \sin t}{t^2}, \tag{2.8.21}$$

so we claim

$$t\cos t < \sin t, \quad \text{for } \ 0 < t < \pi. \tag{2.8.22}$$

Since $\cos t \le 0$ for $\pi/2 \le t < \pi$, it suffices to check $t \in (0, \pi/2)$, so we claim

$$t < \tan t, \quad \text{for } \ 0 < t < \frac{\pi}{2}, \tag{2.8.23}$$

which follows from

$$\frac{d}{dt}(\tan t - t) = \tan^2 t > 0. \tag{2.8.24}$$

□

Remark. According to [**23**], Archimedes called the result

$$0 < s < t < \pi \Longrightarrow \frac{\sin t}{t} < \frac{\sin s}{s} \tag{2.8.25}$$

Aristarchus' formula. See the third graph in Figure 2.8.2 for an illustration.

As a consequence of Proposition 2.8.1, we have that

$$\frac{2}{3}\sin\frac{3^\circ}{2} < \sin 1^\circ < \frac{4}{3}\sin\frac{3^\circ}{4}, \tag{2.8.26}$$

and bringing in (2.8.18) gives

$$0.01745129 < \sin 1^\circ < 0.01745279, \tag{2.8.27}$$

as Ptolemy observed (and as Archimedes probably also knew).

For a more accurate formula, we have from (2.8.19) that

$$\begin{aligned} z = \cos 1^\circ + i\sin 1^\circ \text{ satisfies } z^3 = w,\\ w = \cos 3^\circ + i\sin 3^\circ, \end{aligned} \tag{2.8.28}$$

the latter given by (2.8.17). We can use the iterative procedure described in (2.5.28)–(2.5.30), with $n = 3$, to solve (2.8.28) for z. Since (2.8.17) implies $w \approx 1$, a good initial guess is $z_0 = 1$. Application of this iteration gives

$$\begin{aligned} \sin 1^\circ &= 0.017452406437\cdots,\\ \cos 1^\circ &= 0.999847695156\cdots. \end{aligned} \tag{2.8.29}$$

Remark. Indications are Archimedes had ways to numerically approximate cube roots of positive real numbers. Complex numbers? Probably not.

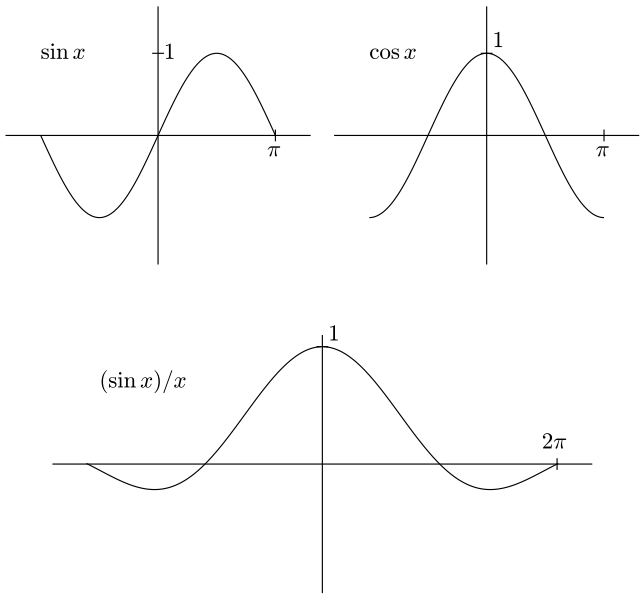

Figure 2.8.2. Graphs of $y = \sin x$, $y = \cos x$, and $y = (\sin x)/x$.

Having these results, to prepare that trig table one can use the calculations set out in (2.8.2)–(2.8.5) together with the identities in (2.8.14) to compute $\operatorname{cis}\ell^\circ = \cos\ell^\circ + i\sin\ell^\circ$ for all the angles depicted in Figure 2.8.1, bring in the formula (2.8.29) for $\operatorname{cis}1^\circ$, and then use the addition formula

$$\operatorname{cis}(k^\circ + \ell^\circ) = (\operatorname{cis}k^\circ)(\operatorname{cis}\ell^\circ) \tag{2.8.30}$$

to fill in the table for $\operatorname{cis}\ell^\circ$, for arbitrary $\ell \in \{0, \dots, 360\}$. Further use of the half-angle formula (2.6.66) leads to formulas for $\operatorname{cis}\mu^\circ$, with μ an integral multiple of $1/2$ (or $1/2^j$, if one persists). See also Exercises 1–2 below.

Of course, given that we do have an accurate evaluation of π, once we have a table of $\operatorname{cis}\theta$ for $\theta = k\pi/180$, we can use power series to quickly evaluate $\operatorname{cis}\delta$ for $|\delta| < \pi/180$, and then use $\operatorname{cis}(\theta+\delta) = (\operatorname{cis}\theta)(\operatorname{cis}\delta)$ to fill in the trig table. In fact, it suffices to compute $\operatorname{cis}\delta$ for $|\delta| \le \pi/360$.

Remark. After Archimedes and before Ptolemy, there was Hipparchus, an eminent astronomer who is also regarded as the founder of trigonometry. He is said to have made a table of values of $\sin\mu^\circ$ for μ an integral multiple of $15/4$, i.e., $\sin\theta$ for $\theta = k\pi/48$. On the other hand, Archimedes' calculation of $\ell(\mathcal{P}_{96})$ was equivalent to the evaluation of $\sin\pi/96$.

We also mention that in those times geometers used the notion of chord(θ) to denote the distance between the two points where a pair of rays, from the origin, making an angle θ, intersect S^1. We have, for $0 \le \theta \le \pi$,

$$\text{chord}(\theta) = |e^{i\theta} - 1| = |e^{i\theta/2} - e^{-i\theta/2}| = 2\sin\frac{\theta}{2}. \tag{2.8.31}$$

From their perspective, they were actually making tables of chords. The identity (2.8.31) shows how to convert these to tables of sines.

For another approach to computing $\sin\theta/3$, we can bring in a formula for $\sin 3\varphi$, starting with

$$\sin(\varphi + 2\varphi) = \sin\varphi\cos 2\varphi + \cos\varphi\sin 2\varphi, \tag{2.8.32}$$

to deduce that

$$\sin 3\varphi = -4\sin^3\varphi + 3\sin\varphi, \tag{2.8.33}$$

hence

$$\sin\theta = -4\sin^3\frac{\theta}{3} + 3\sin\frac{\theta}{3}. \tag{2.8.34}$$

Consequently,

$$x = \sin\frac{\theta}{3} \ \text{ solves } \ 4x^3 - 3x + a = 0, \ a = \sin\theta. \tag{2.8.35}$$

A related treatment of such an approach to angle trisection is given by Descartes in [**5**], p. 396. Fast numerical approximation to a solution to such a cubic equation can be given by Newton's method. Rather than giving details, we refer to calculus texts, such as [**15**], Chapter 5, §5.5.

Exercises

1. Use various symmetries, such as

$$\operatorname{cis}\left(\tfrac{\pi}{2} - \alpha\right) = i\,\overline{\operatorname{cis}\,\alpha},$$

to reduce tabulation of $\cos\ell^\circ + i\sin\ell^\circ$ from $0 \le \ell \le 360$ to

$$0 \le \ell \le 45.$$

2. Proceed from the formula (2.8.29), yielding a 12 digit approximation to $\alpha = \operatorname{cis} 1^\circ$, to make a table of

$$\operatorname{cis} k' = \operatorname{cis}\left(\frac{k}{60}\right)^\circ, \quad 0 \le k \le 60,$$

as follows.

(a) Use the method of (2.5.28)–(2.5.30) to compute approximations to

$$\alpha^{1/2}, \quad \alpha^{1/3}, \quad \alpha^{1/4}, \quad \alpha^{1/5}.$$

(b) Implement the computations

$$\alpha^{1/2}\alpha^{-1/3} = \alpha^{1/6}, \quad \alpha^{1/3}\alpha^{-1/4} = \alpha^{1/12}, \quad \alpha^{1/5}\alpha^{-1/6} = \alpha^{1/30},$$
$$\alpha^{1/2}\alpha^{-2/5} = \alpha^{1/10}, \quad \alpha^{1/10}\alpha^{-1/12} = \alpha^{1/60}.$$

(c) Fill in the table using these computations.

3. Explain how to use the table produced via Exercise 2 together with a table for $\operatorname{cis} \ell^\circ$ to compute

$$\operatorname{cis} \ell^\circ k'.$$

4. Regarding the iterative method (2.8.6)–(2.8.7) for approximating $\sqrt{a}$ ($a > 0$), given an initial guess $x_0 > 0$, with error $\delta_0 = x_0 - \sqrt{a}$, write (2.8.7) as

$$x_{k+1} - \sqrt{a} = \delta_{k+1}, \quad \delta_{k+1} = \frac{\delta_k^2}{2x_k}.$$

Deduce that (if $\delta_0 \neq 0$), $\delta_k > 0$, hence $x_k > \sqrt{a}$, for all $k \geq 1$, and furthermore

$$\frac{a}{x_k} < \sqrt{a} < x_{k+1} < x_k, \quad \forall\, k \geq 1.$$

2.9. Euclidean numbers

Here we describe collections of numbers (in $\mathbb{C}$), lines, and circles that we call Euclidean, denoted $\mathbb{E}$, $\mathcal{E}_L$, and $\mathcal{E}_C$, respectively, which can be constructed using compass and straightedge. We start with some rules for obtaining objects in these classes.

(A) $0, 1 \in \mathbb{E}$,

(B) If $a, b \in \mathbb{E}$, $a \neq b$, the line ℓ_{ab} through a and b belongs to $\mathcal{E}_L$. In particular,

$$\mathbb{R} \in \mathcal{E}_L.$$

(C) If $a, b \in \mathbb{E}$, $a \neq b$, the circle $C_{a,b}$ centered at a, passing through b, belongs to $\mathcal{E}_C$.

(D) If $a \in \mathbb{E}$, $r > 0$, $r \in \mathbb{E}$, then $S_r(a)$, the circle with center a, radius r, belongs to $\mathcal{E}_C$. In particular,

$$S^1 \in \mathcal{E}_C.$$

Rules (B)–(D) state how to produce elements of $\mathcal{E}_L$ and $\mathcal{E}_C$, given elements of $\mathbb{E}$. We also have the following rules for generating elements of $\mathbb{E}$, given elements of $\mathcal{E}_L$ and $\mathcal{E}_C$.

(a) If $\ell, \ell' \in \mathcal{E}_L$, $\ell \neq \ell'$, then $\ell \cap \ell' \subset \mathbb{E}$.

(b) If $C, C' \in \mathcal{E}_C$, $C \neq C'$, then $C \cap C' \subset \mathbb{E}$.

(c) If $\ell \in \mathcal{E}_L$, $C \in \mathcal{E}_C$, then $\ell \cap C \subset \mathbb{E}$.

We define $\mathbb{E}$, $\mathcal{E}_L$, and $\mathcal{E}_C$ to be the smallest collections of numbers, lines, and circles satisfying (A)–(D) and (a)–(c). We now describe other elements of $\mathbb{E}$, $\mathcal{E}_L$, and $\mathcal{E}_C$, applying these rules. We leave the demonstrations of these results as exercises, with some hints.

1. $\mathbb{Z} \subset \mathbb{E}$. Start with (A) and (B), and use (D) and (c) repeatedly.

2. If $\ell \in \mathcal{E}_L$ and $p \in \mathbb{E}$, then $\ell_p^\perp$, the line through p, perpendicular to ℓ, lies in $\mathcal{E}_L$.

Start with a circle $S_r(p)$, with $r \in \mathbb{N}$ large enough that $S_r(p) \cap \ell$ consists of two points, say a, b. If $p \notin \ell$, then $\triangle(a, p, b)$ is isosceles. Take circles

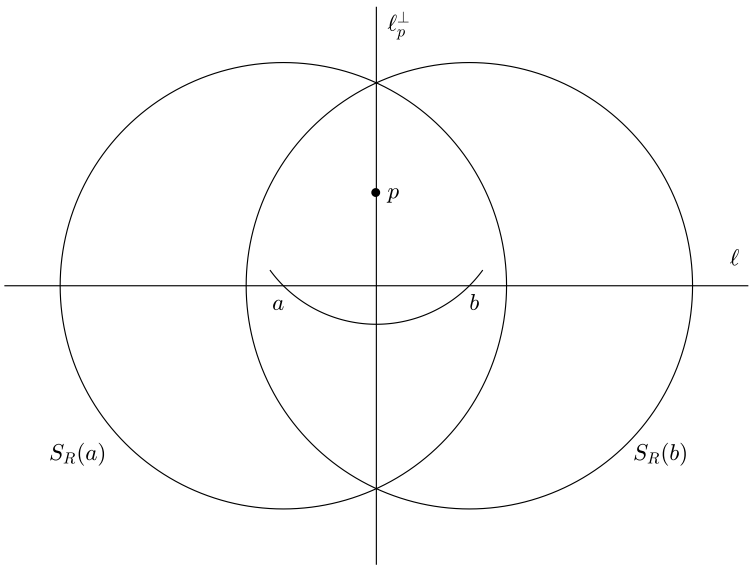

Figure 2.9.1. Showing $\ell \in \mathcal{E}_L,\ p \in \mathbb{E} \Rightarrow \ell_p^\perp \in \mathcal{E}_L$

$S_R(a)$ and $S_R(b)$, with $R \in \mathbb{N},\ R > |a-b|/2$. Then $\ell_p^\perp$ is the line through the two points in $S_R(a) \cap S_R(b)$. See Figure 2.9.1.
Consequences:

$$i\mathbb{R} \in \mathcal{E}_L, \quad i \in \mathbb{E}, \quad ki \in \mathbb{E},\ \forall\, k \in \mathbb{Z}.$$

3. If $\ell \in \mathcal{E}_L,\ q \in \mathbb{E},\ q \notin \ell$, and ℓ'_q is the line through q parallel to ℓ, then $\ell'_q \in \mathcal{E}_L$. Indeed, take $\ell'_q = (\ell_q^\perp)_q^\perp$.

4. $a \in \mathbb{E} \Rightarrow \operatorname{Re} a,\ \operatorname{Im} a,\ |a| \in \mathbb{E}$. Drop perpendiculars from a to $\mathbb{R}$ and to $i\mathbb{R}$. Also, C_{0a} intersects $\mathbb{R}$ at $\pm|a|$.

5. If $x, y \in \mathbb{R} \cap \mathbb{E}$, then $-x, x+y, |y| \in \mathbb{R} \cap \mathbb{E}$. Furthermore, $x + iy \in \mathbb{E}$.
For the latter part, we have $\ell_x^\perp \in \mathcal{E}_L$, and $x + iy \in \ell_x^\perp \cap S_{|y|}(x)$.

6. If $a, b \in \mathbb{E}$, then $a+b \in \mathbb{E}$. Also $ia \in \mathbb{E}$. Use Exercises 4–5. Consequence:

$$\mathbb{Z}[i] \subset \mathbb{E}.$$

7. Assume $a, b \in \mathbb{E},\ a, b > 0$. Then

$$ab,\ \frac{1}{a} \in \mathbb{E}.$$

The line through 0 and $1 + ib$ intersects $\ell_a^\perp$ at $a + iab = z$, and $ab = \operatorname{Im} z$. The line through 0 and $a+i$ intersects $\ell_1^\perp$ at $1+i(1/a) = w$, and $1/a = \operatorname{Im} w$. Corollary:

$$\mathbb{Q}[i] \subset \mathbb{E}.$$

8. More generally,

$$a, b \in \mathbb{E} \Longrightarrow ab,\ \frac{a}{b} \in \mathbb{E},$$

the latter holding provided $b \neq 0$.

For ab, express the product in terms of $\operatorname{Re} a, b$ and $\operatorname{Im} a, b$. Next, write $a/b = a\bar{b}/|b|^2$ and apply Exercise 7 to get $1/|b|^2 \in \mathbb{E}$.
REMARK. Exercises 6 and 8 imply $\mathbb{E}$ is a field.

We next consider square roots.

9. Given $\omega, \alpha \in S^1,\ \alpha^2 = \omega,\ \omega \in \mathbb{E} \Longrightarrow \alpha \in \mathbb{E}$.

Indeed, if $\omega \neq -1$, we have

$$\alpha = \pm \frac{\omega + 1}{|\omega + 1|}.$$

10. If $x \in \mathbb{E},\ x > 0$, then $\sqrt{x} \in \mathbb{E}$.

For the geometrical construction yielding this result, see §1.11.

11. Given $z \in \mathbb{E}$, if $w \in \mathbb{C}$ and $w^2 = z$, then $w \in \mathbb{E}$.

Indeed, if $z \neq 0$, set

$$z = r\omega, \quad r = |z|,\ \omega = \frac{z}{|z|}.$$

Then $r, \omega \in \mathbb{E}$ and

$$w = \pm\sqrt{r}\alpha,$$

with α as in Exercise 9.

12. Let $p(z) = a_2 z^2 + a_1 z + a_0,\ a_j \in \mathbb{E},\ a_2 \neq 0$. Then its roots, i.e., the solutions to $p(z) = 0$, belong to $\mathbb{E}$.

Use the quadratic formula: $r_\pm = -a_1/2a_2 \pm (1/2a_2)\sqrt{a_1^2 - 4a_2 a_0}$.

13. $\mathbb{E}$ is the smallest subfield of $\mathbb{C}$ that is closed under taking square roots.

Results stated above give half of this. For the other half, one needs that elements of $\mathbb{E}$ produced in (b)–(c) arise as solutions to quadratic equations.

14. All the numbers of the form $e^{i\theta}$ in (2.8.14) belong to $\mathbb{E}$.

15. $\mathbb{E}$ is countable. Each application of (B)–(D) produces one element of $\mathcal{E}_L$ or one element of $\mathcal{E}_C$. Each application of (a)–(c) produces no more than one or two elements of $\mathbb{E}$.

16. We say an element of $\mathrm{Isom}(\mathbb{C})$, resp. $\mathrm{Isom}^{\pm}(\mathbb{C})$ belongs to $\mathrm{Isom}(\mathbb{E})$, resp. $\mathrm{Isom}^{\pm}(\mathbb{E})$, provided $F : \mathbb{E} \to \mathbb{E}$. Then one has $F \in \mathrm{Isom}^{+}(\mathbb{E})$ if and only if

$$F(z) = az + p, \quad a \in S^1 \cap \mathbb{E},\ p \in \mathbb{E},$$

and $F \in \mathrm{Isom}^{-}(\mathbb{E})$ if and only if

$$F(z) = a\overline{z} + p, \quad a \in S^1 \cap \mathbb{E},\ p \in \mathbb{E}.$$

REMARK. One might question the label "Euclidean numbers" for the objects treated in this section. The Pythagoreans loved numbers, but between them and Euclid a tradition arose to develop geometry in a fashion that avoided the use of numbers. Apparently this approach was motivated by the discovery that the sidelength of a square and its diagonal are incommensurable. (Modern translation: $\sqrt{2}$ is an irrational number.[2])

The Euclideans did not reject numbers outright. There are several books of Euclid's Elements that deal with number theory. They just kept the concept away from geometry. The work of Archimedes seems to stand as a serious effort to bring numbers into the game. Subsequently, Diophantus developed algebra as a tool for number theory, but not for geometry.

As the lights went out in the Roman empire, works of Euclid, Archimedes, Diophantus and others were preserved, translated, studied, and built upon in other parts of the world. A notable algebra text was produced by M. al-Khwarizmi, [**1**]. Sometime into the second millenium, works of Euclid et al., and also al-Khwarizmi's algebra, were introduced into an awakening Europe. Cf. [**8**], Chapters 9–10.

The use of algebra as a tool for Euclidean geometry took a great leap forward with the work of Descartes, [**5**], regarded as a monumental advance, which set up analytic geometry and paved the way for the introduction of calculus. Though there was a whiff in the air of complex numbers, they were regarded with great suspicion, and played little role in [**5**]. Most workers were not willing to regard $\sqrt{-1}$ as an actual number, and they were definitely not ready to deal with the field $\mathbb{E}$. Even into the time of Euler, complex

[2]See Proposition 2.11.5 and the ensuing discussion for more on this.

numbers were not largely in favor, but under the influence of Gauss they became widely accepted (cf. [**12**]). Given the accomplishments of Gauss, it seems reasonable to say he knew about $\mathbb{E}$, perhaps as early as 1796.

2.10. Linear fractional transformations

Here we consider an important extension of the family of maps $z \mapsto az + b$, making up $\mathrm{Isom}^+(\mathbb{C})$ and more generally $\mathrm{Sim}^+(\mathbb{C})$. This larger class consists of linear fractional transformations. They have the form

$$L_A(z) = \frac{az+b}{cz+d}, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{2.10.1}$$

Here A is an invertible 2×2 matrix; we say $A \in G\ell(2, \mathbb{C})$. (See Appendix 2.A for background material on matrices.) We do not want the denominator in the fraction written above to be $\equiv 0$, and we do not want the numerator to be a constant multiple of the denominator. This leads to the requirement that $\det A \neq 0$. If $c \neq 0$, L_A is well defined on $\mathbb{C} \setminus \{-d/c\}$. We extend L_A to

$$L_A : \widehat{\mathbb{C}} \longrightarrow \widehat{\mathbb{C}}, \quad \widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}, \tag{2.10.2}$$

by setting

$$L_A(-d/c) = \infty, \quad \text{if } \ c \neq 0. \tag{2.10.3}$$

and

$$L_A(\infty) = \begin{matrix} \dfrac{a}{c} & \text{if } \ c \neq 0, \\ \infty & \text{if } \ c = 0. \end{matrix} \tag{2.10.4}$$

If also $B \in G\ell(2, \mathbb{C})$, a calculation gives

$$L_A \circ L_B = L_{AB}. \tag{2.10.5}$$

In particular L_A is bijective on $\widehat{\mathbb{C}}$, with inverse $L_{A^{-1}}$.

The set $G\ell(2, \mathbb{C})$ is a group, i.e., a nonempty set of matrices G with the property that

$$A, B \in G \Longrightarrow AB, \ A^{-1} \in G. \tag{2.10.6}$$

The result (2.10.5) says the map $A \mapsto L_A$ is a group homomorphism.

Note that $L_{sA} = L_A$ for each nonzero $s \in \mathbb{C}$. In particular, $L_A = L_{A_1}$ for some A_1 of determinant 1; we say $A_1 \in S\ell(2, \mathbb{C})$. Given $A_j \in S\ell(2, \mathbb{C})$, $L_{A_1} = L_{A_2}$ if and only if $A_1 = \pm A_2$.

Note that if a, b, c, d are all real, then L_A in (2.10.1) preserves $\mathbb{R} \cup \{\infty\}$. In this case we have $A \in G\ell(2, \mathbb{R})$. We still have $L_{sA} = L_A$ for all nonzero s, but we need $s \in \mathbb{R}$ to get $sA \in G\ell(2, \mathbb{R})$. We can write $L_A = L_{A_1}$ for $A_1 \in S\ell(2, \mathbb{R})$ if $A \in G\ell(2, \mathbb{R})$ and $\det A > 0$. We can also verify that

$$A \in S\ell(2, \mathbb{R}) \Longrightarrow L_A : \mathcal{U} \to \mathcal{U}, \tag{2.10.7}$$

where

$$\mathcal{U} = \{z : \mathrm{Im}\, z > 0\} \tag{2.10.8}$$

is the upper half-plane. In more detail, for $a, b, c, d \in \mathbb{R}$, $z = x + iy$,

$$\frac{az+b}{cz+d} = \frac{(az+b)(c\overline{z}+d)}{(cz+d)(c\overline{z}+d)} = \frac{R}{P} + iy\frac{ad-bc}{P}, \tag{2.10.9}$$

with

$$R = ac|z|^2 + bd + (ad+bc)x \in \mathbb{R}, \quad P = |cz+d|^2 > 0, \text{ if } y \neq 0, \tag{2.10.10}$$

which gives (2.10.7).

We now single out for attention the following linear fractional transformation:

$$\varphi(z) = \frac{z-i}{z+i}, \quad \varphi(z) = L_{A_0}(z), \ A_0 = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}. \tag{2.10.11}$$

Note that

$$\varphi(x+iy) = \frac{x+i(y-1)}{x+i(y+1)} \Longrightarrow |\varphi(x+iy)|^2 = \frac{x^2+(y-1)^2}{x^2+(y+1)^2}. \tag{2.10.12}$$

In particular, $|\varphi(x+iy)| < 1$ if and only if $y > 0$. We have

$$\varphi : \mathcal{U} \to \mathcal{D}, \quad \varphi : \mathbb{R} \cup \{\infty\} \to S^1 = \partial\mathcal{D}, \tag{2.10.13}$$

where

$$\mathcal{D} = \{z : |z| < 1\} \tag{2.10.14}$$

is the unit disk. The bijectivity of φ on $\widehat{\mathbb{C}}$ implies that φ is bijective in (2.10.13).

Conjugating the $S\ell(2,\mathbb{R})$ action on $\mathcal{U}$ by φ yields the mappings

$$M_A = L_{A_0 A A_0^{-1}} : \mathcal{D} \longrightarrow \mathcal{D}. \tag{2.10.15}$$

In detail, if A is as in (2.10.1), with a, b, c, d real, and if A_0 is as in (2.10.11),

$$\begin{aligned} A_0 A A_0^{-1} &= \frac{1}{2i}\begin{pmatrix} (a+d)i - b + c & (a-d)i + b + c \\ (a-d)i - b - c & (a+d)i + b - c \end{pmatrix} \\ &= \begin{pmatrix} \alpha & \beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix}. \end{aligned} \tag{2.10.16}$$

Note that

$$|\alpha|^2 - |\beta|^2 = \det A = ad - bc. \tag{2.10.17}$$

It follows that

$$\begin{aligned} A_0 S\ell(2,\mathbb{R}) A_0^{-1} &= \left\{ \begin{pmatrix} \alpha & \beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} \in G\ell(2,\mathbb{C}) : |\alpha|^2 - |\beta|^2 = 1 \right\} \\ &= SU(1,1), \end{aligned} \tag{2.10.18}$$

the latter identity defining the group $SU(1,1)$. Hence we have linear fractional transformations

$$\begin{aligned} &L_B : \mathcal{D} \to \mathcal{D}, \quad L_B(z) = \frac{\alpha z + \beta}{\overline{\beta} z + \overline{\alpha}}, \\ &B = \begin{pmatrix} \alpha & \beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} \in SU(1,1), \qquad |\alpha|^2 - |\beta|^2 = 1. \end{aligned} \tag{2.10.19}$$

Note that for such B as in (2.10.19),

$$L_B(e^{i\theta}) = \frac{\alpha e^{i\theta} + \beta}{\overline{\beta} e^{i\theta} + \overline{\alpha}} = e^{i\theta} \frac{\alpha + \beta e^{-i\theta}}{\overline{\alpha} + \overline{\beta} e^{i\theta}}, \tag{2.10.20}$$

and in the last fraction the numerator is the complex conjugate of the denominator. This directly implies the result $L_B : S^1 \to S^1$ for such B.

We have the following important transitivity properties.

Proposition 2.10.1. *Given the groups $S\ell(2,\mathbb{R})$ and $SU(1,1)$ defined above,*

$$S\ell(2,\mathbb{R}) \text{ acts transitively on } \mathcal{U}, \text{ via } (2.10.1), \tag{2.10.21}$$

and

$$SU(1,1) \text{ acts transitively on } \mathcal{D}, \text{ via } (2.10.19). \tag{2.10.22}$$

Proof. To demonstrate (2.10.21), take $p = a + ib \in \mathcal{U}$ ($a \in \mathbb{R}$, $b > 0$). Then $L_p(z) = bz + a = (b^{1/2}z + b^{-1/2}a)/b^{-1/2}$ maps i to p. Given another $q \in \mathcal{U}$, we see that $L_p L_q^{-1}$ maps q to p, so (2.10.21) holds. The conjugation (2.10.18) implies that (2.10.21) and (2.10.22) are equivalent.

We can also demonstrate (2.10.22) directly. Given $p \in \mathcal{D}$, we can pick $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 - |\beta|^2 = 1$ and $\beta/\overline{\alpha} = p$. Then $L_B(0) = p$, so (2.10.22) holds. □

We list some building blocks for the group of linear fractional transformations, namely (with $a \neq 0$)

$$\delta_a(z) = az, \quad \tau_b(z) = z + b, \quad \iota(z) = \frac{1}{z}. \tag{2.10.23}$$

We call these respectively (complex) dilations, translations, and inversion about the unit circle $S^1 = \{z : |z| = 1\}$. These have the form (2.10.1) with A given respectively by

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.10.24}$$

We can produce the inversion ι_D about the boundary of a disk $D = D_r(p)$ as

$$\iota_D = \tau_p \circ \delta_r \circ \iota \circ \delta_{1/r} \circ \tau_{-p}. \tag{2.10.25}$$

The reader can work out the explicit form of this linear fractional transformation. Note that ι_D leaves ∂D invariant and interchanges p and ∞.

Recall that the linear fractional transformation φ in (2.10.11) was seen to map $\mathbb{R} \cup \{\infty\}$ to S^1. Similarly its inverse, given by $-i\psi(z)$ with

$$\psi(z) = \frac{z+1}{z-1}, \tag{2.10.26}$$

maps S^1 to $\mathbb{R} \cup \{\infty\}$; equivalently ψ maps S^1 to $i\mathbb{R} \cup \{\infty\}$. To see this directly, write

$$\psi(e^{i\theta}) = \frac{e^{i\theta}+1}{e^{i\theta}-1} = \frac{-i}{\tan\theta/2}. \tag{2.10.27}$$

These are special cases of an important general property of linear fractional transformations. To state it, let us say that en extended line is a set $\ell \cup \{\infty\}$, where ℓ is a line in $\mathbb{C}$.

Proposition 2.10.2. *If L is a linear fractional transformation, then L maps each circle to a circle or an extended line, and L maps each extended line to a circle or an extended line.*

To begin the proof, suppose $D \subset \mathbb{C}$ is a disk. We investigate where L maps ∂D.

Claim 1. If L has a pole at $p \in \partial D$ (i.e., if $L(p) = \infty$), then L maps ∂D to an extended line.

Proof. Making use of the transformations (2.10.23), we have $L(\partial D) = L'(S^1)$ for some linear fractional transformation L', so we need to check only the case $D = \{z : |z| < 1\}$, with L having a pole on S^1, and indeed we can take the pole to be at $z = 1$. Thus we look at

$$\begin{aligned} L(e^{i\theta}) &= \frac{ae^{i\theta}+b}{e^{i\theta}-1} \\ &= -\frac{a+b}{2}\,\frac{1}{\tan\theta/2} + \frac{a-b}{2}, \end{aligned} \tag{2.10.28}$$

whose image is clearly an extended line.

Claim 2. If L has no pole on ∂D, then L maps ∂D to a circle.

Proof. One possibility is that L has no pole in $\mathbb{C}$. Then $c = 0$ in (2.10.1). This case is elementary.

Next, suppose L has a pole at $p \in D$. Composing (on the right) with various linear fractional transformations, we can reduce to the case $D = \{z : |z| < 1\}$, and making further compositions (via Proposition 2.10.1), we need only deal with the case $p = 0$. So we are looking at

$$L(z) = \frac{az + b}{z}, \quad L(e^{i\theta}) = a + be^{-i\theta}. \tag{2.10.29}$$

Clearly the image $L(S^1)$ is a circle.

If L has a pole at $p \in \mathbb{C} \setminus \overline{D}$, we can use an inversion about ∂D to reduce the study to that done in the previous paragraph. This finishes Claim 2.

To finish the proof of Proposition 2.10.2, there are two more claims to establish.

Claim 3. If $\ell \subset \mathbb{C}$ is a line and L has a pole on ℓ, or if L has no pole in $\mathbb{C}$, then L maps $\ell \cup \{\infty\}$ to an extended line.

Claim 4. If $\ell \subset \mathbb{C}$ is a line and L has a pole in $\mathbb{C} \setminus \ell$, then L maps $\ell \cup \{\infty\}$ to a circle.

We leave Claims 3–4 as exercises for the reader.

We now take a look at the group $G_{\mathcal{HU}}$ of linear fractional transformations that preserve both the upper half plane $\mathcal{U}$ and the unit disk $\mathcal{D}$. These have the form

$$L_A, \quad A \in S\ell(2, \mathbb{R}) \cap SU(1, 1), \tag{2.10.30}$$

hence

$$A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad a, b \in \mathbb{R},\ a^2 - b^2 = 1. \tag{2.10.31}$$

Since $L_A = L_{-A}$, we can take $a > 0$. Note that

$$A^{-1} = \begin{pmatrix} a & -b \\ -b & a \end{pmatrix}. \tag{2.10.32}$$

It is convenient to parametrize the group $G_{\mathcal{HU}}$ by setting

$$a = \cosh t, \quad b = \sinh t, \quad t \in \mathbb{R}, \tag{2.10.33}$$

where we bring in the hyperbolic functions

$$\cosh t = \frac{1}{2}(e^t + e^{-t}), \quad \sinh t = \frac{1}{2}(e^t - e^{-t}). \tag{2.10.34}$$

A computation gives $\cosh^2 t - \sinh^2 t = 1$. In such a case, we can write A in (2.10.31) as

$$H(t) = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} = (\cosh t)I + (\sinh t)K, \tag{2.10.35}$$

at least for $a > 0$, where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.10.36}$$

As usual, I denotes the identity matrix.

The casual resemblance of the right side of (2.10.35) to the right side of Euler's formula (2.3.31) is worth pursuing. Let's start with a derivative formula for $H(t)$. By (2.3.8), we have the derivative formulas

$$\frac{d}{dt}\cosh t = \sinh t, \quad \frac{d}{dt}\sinh t = \cosh t. \tag{2.10.37}$$

Applying this to (2.10.35) gives

$$H'(t) = (\sinh t)I + (\cosh t)K = KH(t), \tag{2.10.38}$$

using the fact that $K^2 = I$. This invites comparison with (2.3.32), or more generally (2.3.8). Indeed, we can form the matrix exponential

$$e^{tK} = \sum_{n=0}^{\infty} \frac{t^n}{n!} K^n, \tag{2.10.39}$$

and a computation parallel to (2.3.3) yields

$$\frac{d}{dt} e^{tK} = K e^{tK}. \tag{2.10.40}$$

Now we compare e^{tK} with $H(t)$ via a computation parallel to (2.3.10):

$$\frac{d}{dt}\Big(e^{-tK} H(t)\Big) = -K e^{-tK} H(t) + e^{-tK} K H(t) = 0, \tag{2.10.41}$$

hence

$$e^{-tK} H(t) = I, \quad \forall\, t \in \mathbb{R}. \tag{2.10.42}$$

A similar computation gives

$$e^{-tK} e^{tK} = I, \quad \forall\, t \in \mathbb{R}, \ \text{ hence } \ e^{-tK} = \big(e^{tK}\big)^{-1}, \tag{2.10.43}$$

so

$$e^{tK} = H(t). \tag{2.10.44}$$

Also, somewhat parallel to (2.3.14)–(2.3.16), we have, for $s, t \in \mathbb{R}$,

$$\frac{d}{dt} e^{(s+t)K} e^{-tK} = 0, \tag{2.10.45}$$

hence $e^{(s+t)K}e^{-tK} = e^{sK}$, so

$$e^{(s+t)K} = e^{sK}e^{tK}, \quad \forall\, s,t \in \mathbb{R}, \tag{2.10.46}$$

or, equivalently,

$$H(s+t) = H(s)H(t), \quad \forall\, s,t \in \mathbb{R}. \tag{2.10.47}$$

Returning to the action of $H(t)$ as a linear fractional transformation, we have

$$\Phi_t(z) = L_{H(t)}(z) = \frac{(\cosh t)z + \sinh t}{(\sinh t)z + \cosh t}, \tag{2.10.48}$$

and (2.10.47), together with (2.10.5), yields

$$\Phi_{s+t} = \Phi_s \circ \Phi_t, \quad s,t \in \mathbb{R}. \tag{2.10.49}$$

Note that

$$\Phi_t(0) = \frac{\sinh t}{\cosh t} = \tanh t, \tag{2.10.50}$$

where the last identity defines $\tanh t$. We have

$$\tanh : \mathbb{R} \longrightarrow (-1,1), \quad \tanh t \to \pm 1 \text{ as } t \to \pm\infty. \tag{2.10.51}$$

Also, by (2.10.34) and basic identities discussed in §2.1,

$$\frac{d}{dt}\tanh t = \frac{\cosh^2 t - \sinh^2 t}{\cosh^2 t} = 1 - \tanh^2 t = \frac{1}{\cosh^2 t}. \tag{2.10.52}$$

In particular the derivative is > 0 for all t, so the inverse function theorem (Proposition 2.1.2) implies the map (2.10.51) is bijective. Hence each $x \in (-1,1)$ can be uniquely written as $x = \tanh t$ for $t \in \mathbb{R}$. Combining (2.10.52) with Proposition 2.1.5 gives

$$\int_0^x \frac{dy}{1-y^2} = \tanh^{-1} x, \quad |x| < 1, \tag{2.10.53}$$

via the substitution $y = \tanh t$.

Note that applying (2.10.49) to $z = 0$ gives

$$\Phi_s(\tanh t) = \tanh(s+t), \quad s,t \in \mathbb{R}. \tag{2.10.54}$$

As an alternative derivation, note that (2.10.48) gives

$$\Phi_s(z) = \frac{z + \tanh s}{(\tanh s)z + 1}, \tag{2.10.55}$$

and then (2.10.54) is equivalent to the identity

$$\tanh(s+t) = \frac{\tanh s + \tanh t}{1 + (\tanh s)(\tanh t)}, \tag{2.10.56}$$

which can also be established by calculations parallel to (2.4.35).

Here is one significant consequences of these calculations regarding Φ_t.

Proposition 2.10.3. *Let $S \subset \mathcal{D}$ be a circle centered at a point $p \in (-1,1)$. Then we can find $t \in \mathbb{R}$ such that $\Phi_t(S) \subset \mathcal{D}$ is a circle centered at 0.*

Proof. Say S intersects $(-1,1)$ at $\tanh s_1$ and $\tanh s_2$, $s_1 < s_2$. Then $\Phi_t(S) \subset \mathcal{D}$ is a circle, centered at a point in $(-1,1)$, that intersects $(-1,1)$ at $\tanh(s_1 + t)$ and $\tanh(s_2 + t)$. Now

$$s_2 + t = -(s_1 + t) \Longleftrightarrow t = -\frac{s_1 + s_2}{2}. \tag{2.10.57}$$

Thus, if t satisfies (2.10.57), it follows that $\Phi_t(S)$ intersects $(-1,1)$ at

$$\tanh \frac{s_1 - s_2}{2} \quad \text{and at} \quad \tanh \frac{s_2 - s_1}{2}. \tag{2.10.58}$$

Symmetry considerations imply $\Phi_t(S)$ is centered at 0. □

Remark. Using (2.10.56), we have

$$\tanh(s_1 - s_2) = \frac{\tanh s_1 - \tanh s_2}{1 - (\tanh s_1)(\tanh s_2)}, \tag{2.10.59}$$

and

$$\begin{aligned} &x = \tanh \frac{s}{2}, \quad y = \tanh s \\ &\Longrightarrow y = \frac{2x}{1 + x^2} \\ &\Longrightarrow x = \frac{1 - \sqrt{1 - y^2}}{y}. \end{aligned} \tag{2.10.60}$$

We can readily extend Proposition 2.10.3 to a broader setting.

Proposition 2.10.4. *If $S \subset \mathcal{D}$ is a circle, we can find $A \in SU(1,1)$ such that*

$$L_A(S) \subset \mathcal{D} \quad \textit{is centered at } 0. \tag{2.10.61}$$

Proof. Rotate $\mathcal{D}$ so that S is centered at a point in $(-1,1)$. □

We can use Proposition 2.10.4 to tackle the following classical problem.

Apollonius' three circles problem. Let C_1, C_2, C_3 be three disjoint circles in $\mathbb{C}$. Assume no one circle separates the other two. In this situation, find a fourth circle S that is tangent to each C_j.

Solution. Using a translation and dilation, we can assume

$$C_1 = S^1 = \{z : |z| = 1\}. \tag{2.10.62}$$

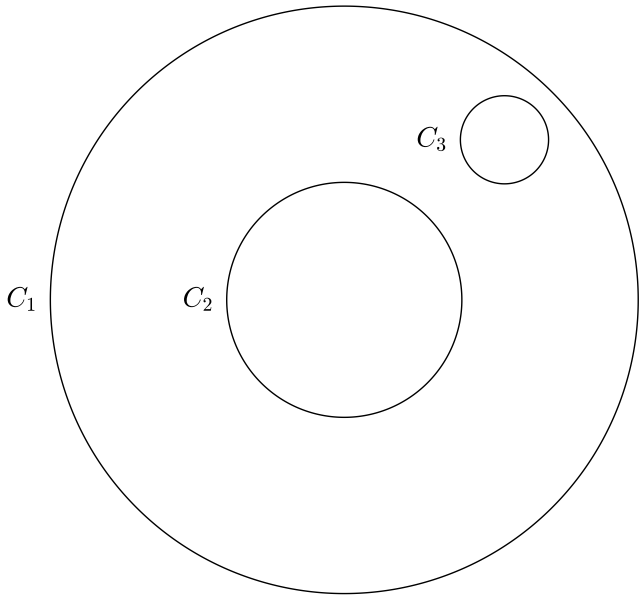

Figure 2.10.1. Setup for solution to three circles problem

Then either $C_2, C_3 \subset \mathcal{D}$ or $C_2, C_3 \subset \mathbb{C} \setminus \overline{\mathcal{D}}$. In the latter case, use the inversion $z \mapsto 1/z$ to arrange that

$$C_2,\ C_3 \subset \mathcal{D}. \tag{2.10.63}$$

Now, using Proposition 2.10.4, apply a linear fractional transformation, preserving $\mathcal{D}$, to arrange that C_2 is centered at 0, so

$$C_2 = \{z : |z| = a\}, \quad 0 < a < 1. \tag{2.10.64}$$

See Figure 2.10.1. Let

$$d = \frac{1-a}{2}, \quad \Gamma = \left\{z : |z| = a + d = \frac{1+a}{2}\right\}. \tag{2.10.65}$$

Say C_3 has center $p \in \mathcal{D}$ and radius r, i.e.,

$$C_3 = \{z : |z - p| = r\}, \tag{2.10.66}$$

and set

$$\Sigma = \{z : |z - p| = r + d\}. \tag{2.10.67}$$

Then Σ intersects Γ at two points. Pick one, call it q. Note that q has distance d from both C_1 and C_2. Furthermore, it has distance $r + d$ from p, hence distance d from C_3. Hence the circle

$$S = \{z : |z - q| = d\} \tag{2.10.68}$$

is tangent to C_1, C_2, and C_3. In this way, we get two circles S, each tangent to all C_j, having the property that both C_2 and C_3 are exterior to the disk D bounded by S. Similarly we have two such circles S, bounding D, with each of the properties:

$$\begin{aligned} &D \text{ contains } C_3 \text{ but is exterior to } C_2,\\ &D \text{ contains } C_2 \text{ but is exterior to } C_3,\\ &D \text{ contains both } C_2 \text{ and } C_3, \end{aligned}$$

for a total of eight circles, each tangent to all C_j.

Hyperbolic and spherical geometry

One can put distance functions $d_{\mathcal{U}}$ and $d_{\mathcal{D}}$ on the upper half-plane $\mathcal{U}$ and the disk $\mathcal{D}$ that have the properties

$$\begin{aligned} \mathrm{Isom}^+(\mathcal{U}) &= \{L_A : A \in S\ell(2, \mathbb{R})\},\\ \mathrm{Isom}^+(\mathcal{D}) &= \{L_A : A \in \mathrm{SU}(1,1)\}. \end{aligned} \tag{2.10.69}$$

Such spaces are called the Poincaré upper half-plane and the Poincaré disc. These are non-Euclidean geometries, known as hyperbolic geometries. The map $\varphi : \mathcal{U} \to \mathcal{D}$ given in (2.10.11) is an isometry between the two spaces. One can see how this is done in §5.12 of [**17**], and also see applications to deep results in complex analysis, including the Riemann mapping theorem and Picard's theorems.

In another direction, one can use stereographic projection to produce a bijective map of $\widehat{\mathbb{C}}$ onto S^2 (the unit sphere in $\mathbb{R}^3$, discussed in the next section). This yields maps

$$\widetilde{L}_A : S^2 \longrightarrow S^2, \quad A \in S\ell(2, \mathbb{C}), \tag{2.10.70}$$

satisfying $\widetilde{L}_{AB} = \widetilde{L}_A \circ \widetilde{L}_B$, and preserving angles. One has

$$\mathrm{Isom}^+(S^2) = \{\widetilde{L}_A : A \in \mathrm{SU}(2)\}, \tag{2.10.71}$$

where $A \in \mathrm{SU}(2) \Leftrightarrow A \in S\ell(2, \mathbb{C})$ and $A^*A = I$. (Note that $L_A = L_{-A} \Rightarrow \widetilde{L}_A = \widetilde{L}_{-A}$.) See §5.3 of [**17**] for more on this. The sphere S^2 also has a non-Euclidean geometry, known as spherical geometry.

Exercises

1. Write down demonstrations of Claims 3 and 4, arising in the proof of Proposition 2.10.2.

2.11. Argand and the fundamental theorem of algebra

J.-R. Argand published in 1813 a paper on representing complex numbers as points in the Euclidean plane.[3] The following year he published a proof of the fundamental theorem of algebra, regarded as the first complete proof of this result, not suffering from gaps such as were present in the 1799 proof of Gauss. We present this proof here.

Theorem 2.11.1. *If $p(z)$ is a nonconstant polynomial (with complex coefficients), then $p(z)$ must have a complex root.*

Proof. We have, for some $n \geq 1,\ a_n \neq 0$,

$$\begin{aligned} p(z) &= a_n z^n + \cdots + a_1 z + a_0 \\ &= a_n z^n \big(1 + O(z^{-1})\big), \quad |z| \to \infty, \end{aligned} \tag{2.11.1}$$

which implies

$$\lim_{|z|\to\infty} |p(z)| = \infty. \tag{2.11.2}$$

Picking $R \in (0, \infty)$ such that

$$\inf_{|z|\geq R} |p(z)| > |p(0)|, \tag{2.11.3}$$

we deduce that

$$\inf_{|z|\leq R} |p(z)| = \inf_{z\in\mathbb{C}} |p(z)|. \tag{2.11.4}$$

Since $\overline{D}_R = \{z : |z| \leq R\}$ is closed and bounded and p is continuous, there exists $z_0 \in \overline{D}_R$ such that

$$|p(z_0)| = \inf_{z\in\mathbb{C}} |p(z)|. \tag{2.11.5}$$

The theorem hence follows from: □

Lemma 2.11.2. *If $p(z)$ is a nonconstant polynomial and* (2.11.5) *holds, then $p(z_0) = 0$.*

Proof. Suppose to the contrary that

$$p(z_0) = a \neq 0. \tag{2.11.6}$$

We can write

$$p(z_0 + \zeta) = a + q(\zeta), \tag{2.11.7}$$

[3] This work was done in 1806 and circulated informally. Argand published it after learning that others were publishing related results. By this time, the geometrical representation of $\mathbb{C}$ was in the air. There was also unpublished 1796 work of Gauss. For more on such early work, see Chapter 3 of [**12**].

where $q(\zeta)$ is a nonconstant polynomial in ζ, satisfying $q(0) = 0$. Hence, for some $k \geq 1$ and $b \neq 0$, we have $q(\zeta) = b\zeta^k + \cdots + b_n\zeta^n$, i.e.,

$$q(\zeta) = b\zeta^k + O(\zeta^{k+1}), \quad \zeta \to 0, \tag{2.11.8}$$

so, uniformly on $S^1 = \{\omega : |\omega| = 1\}$

$$p(z_0 + \varepsilon\omega) = a + b\omega^k\varepsilon^k + O(\varepsilon^{k+1}), \quad \varepsilon \searrow 0. \tag{2.11.9}$$

Pick $\omega \in S^1$ such that

$$\frac{b}{|b|}\omega^k = -\frac{a}{|a|}, \tag{2.11.10}$$

which is possible since $a \neq 0$ and $b \neq 0$. In more detail, since $-(a/|a|)(|b|/b) \in S^1$, Euler's formula implies

$$-\frac{a}{|a|}\,\frac{|b|}{b} = e^{i\theta},$$

for some $\theta \in \mathbb{R}$, so we can take

$$\omega = e^{i\theta/k}.$$

Given (2.11.10),

$$p(z_0 + \varepsilon\omega) = a\Big(1 - \Big|\frac{b}{a}\Big|\varepsilon^k\Big) + O(\varepsilon^{k+1}), \tag{2.11.11}$$

which contradicts (2.11.5) for $\varepsilon > 0$ small enough. Thus (2.11.6) is impossible. This proves Lemma 2.11.2, hence Theorem 2.11.1. □

Now that we have shown that $p(z)$ in (2.11.1) must have one root, we can show it has n roots (counting multiplicity).

Proposition 2.11.3. *For a polynomial $p(z)$ of degree n, as in* (2.11.1)*, there exist $r_1, \ldots, r_n \in \mathbb{C}$ such that*

$$p(z) = a_n(z - r_1)\cdots(z - r_n). \tag{2.11.12}$$

Proof. We have shown that $p(z)$ has one root; call it r_1. Dividing $p(z)$ by $z - r_1$, we have

$$p(z) = (z - r_1)\tilde{p}(z) + q, \tag{2.11.13}$$

where $\tilde{p}(z) = a_n z^{n-1} + \cdots + \tilde{a}_0$ and q is a polynomial of degree < 1, i.e., a constant. Setting $z = r_1$ in (2.11.13) yields $q = 0$, so

$$p(z) = (z - r_1)\tilde{p}(z). \tag{2.11.14}$$

Since $\tilde{p}(z)$ is a polynomial of degree $n - 1$, the result (2.11.12) follows by induction on n. □

REMARK. A key step in the proof given above is the passage from (2.11.4) to (2.11.5). This is predicated on the assertion that a continuous real-valued function ($f(z) = |p(z)|$) on $\overline{D}_R$ must achieve a minimum. A proof of this uses the fact that a sequence in $\overline{D}_R$ must have a convergent subsequence. It was perceived later that such a result needs proof, and a foundation for this was established by Cauchy and Weierstrass. One can read about this in Chapter 1 of [**15**].

Polynomials with real coefficients are of particular interest, and in that connection we point out the following, which is more elementary than Theorem 2.11.1.

Proposition 2.11.4. *Let $p(z)$ have the form (2.11.1), with $a_n \neq 0$. Assume each coefficient $a_j \in \mathbb{R}$. If n is odd, then $p(x) = 0$ for some $x \in \mathbb{R}$.*

Proof. If $a_n > 0$, then $p(x) \to +\infty$ as $x \to +\infty$, and $p(x) \to -\infty$ as $x \to -\infty$. The conclusion that $p(x) = 0$ for some $x \in \mathbb{R}$ is then a special case of the following result, known as the intermediate value theorem. □

Proposition 2.11.5. *Let $a, b \in \mathbb{R}$, $a < b$. Assume $f : [a, b] \to \mathbb{R}$ is continuous, $f(a) < 0$, and $f(b) > 0$. Then $f(x) = 0$ for some $x \in \mathbb{R}$ between a and b.*

Proof. Divide $I_0 = [a, b]$ into two equal subintervals I_ℓ and I_r, meeting at $(a + b)/2 = p$. Then either $f(p) = 0$ or $f(p)$ has a different sign from $f(a)$ or $f(b)$. If $f(p) = 0$, stop. Otherwise choose I_1 to be the subinterval (I_ℓ or I_r) for which f has different signs at the endpoints. Repeat this process, obtaining successive subintervals I_k, of length $2^{-k}|a - b|$, on which f changes sign (or with an endpoint at which f vanishes). The structure of $\mathbb{R}$ (presented in Chapter 1 of [**15**]) shows that the resulting sequence of endpoints converges to a limit $x \in I_0$, at which f vanishes. □

This result is behind the proof of the inverse function theorem, Proposition 2.1.2. We restate the part that follows quickly from Proposition 2.11.5.

Corollary 2.11.6. *Assume $f : [a, b] \to \mathbb{R}$ is continuous, $f(a) = \alpha$, $f(b) = \beta$. Assume also that f is differentiable on (a, b) and $f'(x) > 0$ for all $x \in (a, b)$. Then*

$$f : [a, b] \longrightarrow [\alpha, \beta] \quad \textit{is one-to-one and onto.} \tag{2.11.15}$$

Corollary 2.11.7. *Let $n \in \mathbb{N}$, $f(x) = x^n$. Then*

$$f : [0, \infty) \longrightarrow [0, \infty) \quad \textit{is one-to-one and onto.} \tag{2.11.16}$$

Hence, given $a \in \mathbb{R}$, $a > 0$, there is a unique positive $x \in \mathbb{R}$ such that $x^n = a$. We write $x = a^{1/n}$. For $n = 2$, we write $x = \sqrt{a}$.

Here is a result that caused a stir among the Pythagoreans.

Proposition 2.11.8. *There is no rational number $x = m/n$ that solves $x^2 = 2$.*

Proof. Suppose there were such a rational number, so

$$\left(\frac{m}{n}\right)^2 = 2. \tag{2.11.17}$$

We can assume m and n are positive integers with no common factor. Now (2.11.17) yields

$$m^2 = 2n^2, \tag{2.11.18}$$

hence m is even (the square of an odd integer is odd), so $m = 2k$, so $4k^2 = 2n^2$, so

$$n^2 = 2k^2. \tag{2.11.19}$$

Hence n must also be even. This is a contradiction. □

Corollary 2.11.9. *The number $\sqrt{2}$ is irrational.*

The following classical result yields an extensive list of irrational numbers.

Proposition 2.11.10. *Suppose $x \in \mathbb{R}$ satisfies*

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad a_j \in \mathbb{Z}. \tag{2.11.20}$$

Then, if x is rational, it must be an integer. Equivalently, if $x \notin \mathbb{Z}$, then x must be irrational.

A proof, using the fundamental theorem of arithmetic, is given in §1.7 of [**15**].

Exercises

1. If $p(z)$ is as in (2.11.1), with all $a_j \in \mathbb{R}$, show that z is a root of p if and only if $\overline{z}$ is a root. Hence the non-real roots of p occur in complex-conjugate pairs.

2. Take

$$p(z) = z^4 + 2z^3 + 3z^2 + 4z + 5.$$

Show that all 4 roots of p lie in

$$D_4 = \{z \in \mathbb{C} : |z| < 4\}.$$

Hint. Show that, with

$$R(z) = \frac{2}{z} + \frac{3}{z^2} + \frac{4}{z^3} + \frac{5}{z^4},$$
$$|z| \geq 4 \Longrightarrow |R(z)| < 1.$$

3. For p as in Exercise 2, show that

$$p(-4) > 0, \quad p(-3) < 0, \quad p(x) > 0 \text{ for } x \geq 0.$$

Deduce that p has at least two real roots. Note that, for $x \in \mathbb{R}$,

$$p''(x) = 12(x^2 + x + \tfrac{1}{2}) > 0, \quad \forall\, x \in \mathbb{R}.$$

Hence conclude that p has exactly two real roots, and two non-real roots (forming a complex-conjugate pair).

4. Deduce the following from Proposition 2.11.10. Take $k, n \in \mathbb{N}$. Assume $x \in \mathbb{R}$ and

$$x^n = k.$$

Then x is either an integer or an irrational number. In particular, the following numbers are irrational:

$$\sqrt{3},\ \sqrt{5},\ \sqrt{6},\ \sqrt{7},\ \sqrt{8}.$$

5. Show that $\sqrt{2} + \sqrt{3}$ is irrational.
Hint. Square it.

2.12. Planetary orbits: ellipses, parabolas, hyperbolas

If a body of mass m is moving in $\mathbb{C}$ under the influence of a force F, Newton's law "$F = ma$" yields the differential equation

$$m\frac{d^2z}{dt^2} = F(z), \tag{2.12.1}$$

for the trajectory $z(t)$ of the body. Many such forces are central forces, taking the form

$$F(z) = f(|z|)z, \tag{2.12.2}$$

with $f : [0, \infty) \to \mathbb{R}$. In particular, Newton's law of gravitation yields

$$F(z) = -\frac{Km}{|z|^3}z. \tag{2.12.3}$$

When Newton formulated and solved such an equation, he obtained a theoretical understanding of Kepler's observations about the orbits of planets about the sun.

It would have been natural for Newton to use the work of Descartes to present the analysis of such an equation, but Newton chose instead to present his work in the formalism of classical Euclidean geometry.[4] Standard modern approaches use the Cartesian framework. Here we present the analysis in the framework of the algebraic and geometric structure of $\mathbb{C}$.

We start with results on the general central force problem,

$$m\frac{d^2z}{dt^2} = f(|z|)z. \tag{2.12.4}$$

Associated with the trajectory $z(t)$ is the angular momentum

$$\alpha(t) = mz(t) \times z'(t), \tag{2.12.5}$$

where, as in §1.2, $z \times w = \operatorname{Im} z\overline{w}$. Let us recall the identities

$$z \times z = 0, \quad z \times (iz) = -|z|^2. \tag{2.12.6}$$

Differentiating (2.12.5) yields

$$\begin{aligned} \alpha'(t) &= mz'(t) \times z'(t) + mz(t) \times z''(t) \\ &= z(t) \times f(|t|)z(t) \\ &= 0, \end{aligned} \tag{2.12.7}$$

where we have used the first identity in (2.12.6) twice. Equation (2.12.7) implies $\alpha(t)$ is independent of t, and hence states the conservation of angular momentum, for motion in a central force.

[4] Some say Newton did this because of his conservative tendencies. Others say he did this because he did not like Descartes. In any event, it seems an upshot was for the torch of calculus to pass from Britain to the European continent (cf. [**8**], Chapter 17).

To proceed, it is convenient to write

$$z(t) = r(t)e^{i\theta(t)} \tag{2.12.8}$$

and translate (2.12.4) into a system of differential equations for $r(t)$ and $\theta(t)$. A computation gives

$$\begin{aligned} z' &= (r' + ir\theta')e^{i\theta}, \\ z'' &= [r'' - r(\theta')^2 + i(2r'\theta' + r\theta'')]e^{i\theta}. \end{aligned} \tag{2.12.9}$$

Since the right side of (2.12.4) becomes $f(r)re^{i\theta}$, cancelling factors of $e^{i\theta}$ and separating out real and imaginary parts yields the system

$$\begin{aligned} r'' - r(\theta')^2 &= \frac{f(r)}{m}r, \\ 2r'\theta' + r\theta'' &= 0. \end{aligned} \tag{2.12.10}$$

The second of the equations (2.12.10) begs for comparison with (2.12.7). Indeed, we have

$$\begin{aligned} z \times z' &= re^{i\theta} \times (r' + ir\theta')e^{i\theta} \\ &= rr'e^{i\theta} \times e^{i\theta} + r^2\theta' e^{i\theta} \times ie^{i\theta} \\ &= -r^2\theta', \end{aligned} \tag{2.12.11}$$

the last identity using both parts of (2.12.6). Meanwhile,

$$\frac{d}{dt}(r^2\theta') = r(2r'\theta' + r\theta''), \tag{2.12.12}$$

so the second part of (2.12.10) says

$$\frac{d}{dt}(r^2\theta') = 0, \tag{2.12.13}$$

recovering (2.12.7). Therefore we have

$$r^2(t)\theta'(t) = L, \tag{2.12.14}$$

for some constant L. Plugging this into the first equation of the system (2.12.10) yields the following second order differential equation for r:

$$\frac{d^2r}{dt^2} = \frac{f(r)r}{m} + \frac{L^2}{r^3}. \tag{2.12.15}$$

Before tackling (2.12.15), we present the following geometrical interpretation of the conservation law (2.12.14). By an argument parallel to the second proof of Proposition 2.6.7, the area $A(t)$ swept out by the ray from 0 to $z(s)$, as s runs from t_0 to t is given by

$$A(t) = \frac{1}{2}\int_{\theta(t_0)}^{\theta(t)} r^2\, d\theta = \frac{1}{2}\int_{t_0}^{t} r(s)^2\theta'(s)\, ds, \tag{2.12.16}$$

so

$$A'(t) = \frac{1}{2}r^2\theta' = \frac{L}{2}. \tag{2.12.17}$$

This says

(2.12.18) Equal areas are swept out in equal times,

which, as we will explain below, is Kepler's second law.

We return to the equation (2.12.15). This has the form

$$\frac{d^2r}{dt^2} = g(r), \tag{2.12.19}$$

a much studied class of differential equations, whose solution we sketch. Take $w(r)$ such that $g(r) = -w'(r)$, so (2.12.19) becomes

$$\frac{d^2r}{dt^2} = -w'(r). \tag{2.12.20}$$

Then form the "energy"

$$E(t) = \frac{1}{2}\Big(\frac{dr}{dt}\Big)^2 + w(r), \tag{2.12.21}$$

and compute that if $r(t)$ solves (2.12.20) then

$$\frac{dE}{dt} = \frac{d^2r}{dt^2}\,\frac{dr}{dt} + w'(r)\frac{dr}{dt} = 0, \tag{2.12.22}$$

so for each solution to (2.12.20), there is a constant E such that

$$\frac{dr}{dt} = \pm\sqrt{2E - 2w(r)}, \tag{2.12.23}$$

or equivalently

$$\int \frac{dr}{\sqrt{2E - 2w(r)}} = \pm t + C. \tag{2.12.24}$$

Note that dividing (2.12.23) by (2.12.14) yields a differential equation for r as a function of θ:

$$\frac{dr}{d\theta} = \pm\frac{r^2}{L}\sqrt{2E - 2w(r)}, \tag{2.12.25}$$

yielding in turn

$$L\int \frac{dr}{r^2\sqrt{2E - 2w(r)}} = \pm\theta + C. \tag{2.12.26}$$

Both the integrals in (2.12.24) and (2.12.26) can be quite messy.

Specializing to the case (2.12.3) (the Kepler problem), we have

$$\frac{f(r)r}{m} = -\frac{K}{r^2}, \tag{2.12.27}$$

and (2.12.19) becomes

$$\frac{d^2r}{dt^2} = -\frac{K}{r^2} + \frac{L^2}{r^3}. \tag{2.12.28}$$

Then the formulas (2.12.20)–(2.12.26) apply, with

$$w(r) = -\frac{K}{r} + \frac{L^2}{r^3}. \tag{2.12.29}$$

However, rather than implementing this, we make use of a neat trick that has been discovered to facilitate the treatment of (2.12.28). We describe how this works.

What we do is produce a differential equation for u, defined by

$$u = \frac{1}{r}. \tag{2.12.30}$$

To start, note that by the chain rule,

$$\frac{dr}{dt} = -r^2\frac{du}{dt} = -r^2\frac{du}{d\theta}\,\frac{d\theta}{dt} = -L\frac{du}{d\theta}, \tag{2.12.31}$$

the last identity by (2.12.14). Taking another t-derivative gives

$$\frac{d^2r}{dt^2} = -L\frac{d}{dt}\,\frac{du}{d\theta} = -L\frac{d^2u}{d\theta^2}\,\frac{d\theta}{dt} = -L^2u^2\frac{d^2u}{d\theta^2}, \tag{2.12.32}$$

again using (2.12.14). Comparing this with (2.12.28), we get

$$-L^2u^2\frac{d^2u}{d\theta^2} = L^2u^3 - Ku^2, \tag{2.12.33}$$

or equivalently,

$$\frac{d^2u}{d\theta^2} + u = \frac{K}{L^2}. \tag{2.12.34}$$

Via this device, we have obtained a linear equation, whose general real-valued solution (see Exercise 1 at the end of this section) can be written

$$u(\theta) = A\cos(\theta - \theta_0) + \frac{K}{L^2}, \tag{2.12.35}$$

which by (2.12.30) gives

$$r\Big[A\cos(\theta - \theta_0) + \frac{K}{L^2}\Big] = 1. \tag{2.12.36}$$

This is equivalent to

$$r[1 + e\cos(\theta - \theta_0)] = p, \quad p = \frac{L^2}{K},\ e = A\frac{L^2}{K}. \tag{2.12.37}$$

The quantity e in (2.12.37) is called the eccentricity of the curve described there. If $e = 0$, (2.12.37) is the equation of a circle. As we will see, if $0 < e < 1$ it is the equation of an ellipse, if $e = 1$, it is the equation of a parabola, and if $e > 1$ it is the equation of a hyperbola.

An ellipse is a curve in $\mathbb{C}$ that can be mapped by a rigid motion to one of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{2.12.38}$$

for some $a, b > 0$. (The case $a = b$ gives a circle.) A hyperbola is a curve in $\mathbb{C}$ that can be mapped by a rigid motion to one of the form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{2.12.39}$$

A parabola is a curve in $\mathbb{C}$ that can be mapped by a rigid motion to one of the form

$$y = ax^2, \tag{2.12.40}$$

with $a > 0$ (or, alternatively, via rotation, $x = ay^2$).

Let us check out the various possibilities for curves of the form (2.12.37). Applying a rotation, we can arrange $\theta_0 = 0$, and write

$$r(1 + e\cos\theta) = p. \tag{2.12.41}$$

Since $r\cos\theta = x$, we have

$$r + ex = p, \quad \text{hence} \quad r = p - ex, \tag{2.12.42}$$

hence

$$x^2 + y^2 = (p - ex)^2 = p^2 - 2epx + e^2x^2, \tag{2.12.43}$$

hence

$$(1 - e^2)x^2 + 2epx + y^2 = p^2. \tag{2.12.44}$$

From this, the implications

$$\begin{aligned} 0 < e < 1 &\Longrightarrow \text{ellipse}, \\ e = 1 &\Longrightarrow \text{parabola}, \\ e > 1 &\Longrightarrow \text{hyperbola} \end{aligned} \tag{2.12.45}$$

are elementary.

Ellipses, hyperbolas, and parabolas, known collectively as conic sections, were studied in ancient times. Euclid write four books on conics, which were incorporated in a larger work of Apollonius. See [**7**], Chapter 4. For more on conic sections, see Exercise 1 of §2.13.

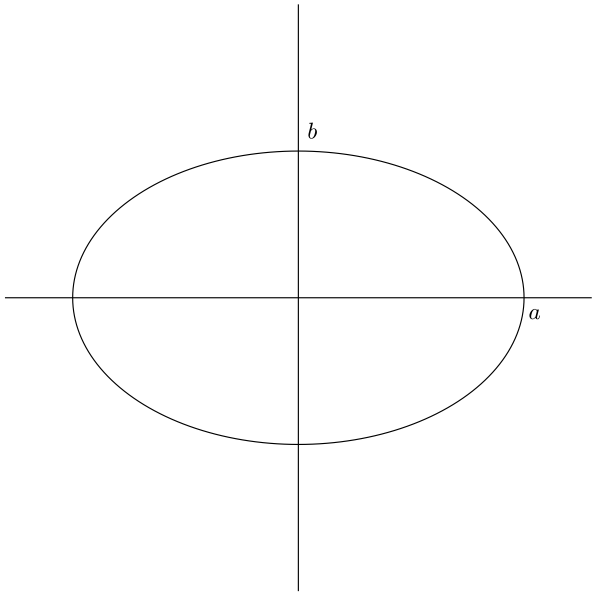

Figure 2.12.1. Ellipse $x^2/a^2 + y^2/b^2 = 1$

Ellipses

Let us assume $0 < e < 1$, so our curves are ellipses, and work on (2.12.44) a bit more. Completing the square gives

$$\left(\sqrt{1-e^2}\,x + \frac{ep}{\sqrt{1-e^2}}\right)^2 + y^2 = \frac{p^2}{1-e^2}, \tag{2.12.46}$$

and then implementing a translation

$$x \mapsto x + \frac{ep}{1-e^2} \tag{2.12.47}$$

gives the curve

$$(1-e^2)x^2 + y^2 = \frac{p^2}{1-e^2}, \tag{2.12.48}$$

which then takes the form (2.12.38), with

$$a = \frac{p}{1-e^2}, \quad b = \frac{p}{\sqrt{1-e^2}} = p^{1/2}a^{1/2}. \tag{2.12.49}$$

See Figure 2.12.1. Note also that $b = a\sqrt{1-e^2}$.

At this point we have enough information to work out the period T of such an elliptical orbit. Indeed, by (2.12.16)–(2.12.17),

$$\begin{aligned} \frac{L}{2}T &= \text{area enclosed by this ellipse} \\ &= \pi ab, \end{aligned} \tag{2.12.50}$$

the latter identity by (2.6.107). Then we bring in (2.12.49) to get

$$T = \frac{2\pi ab}{L} = 2\pi \frac{\sqrt{p}}{L} a^{3/2} = \frac{2\pi}{\sqrt{K}} a^{3/2}, \tag{2.12.51}$$

the last identity by $p = L^2/K$.

We now have Kepler's three laws:

(1) Each planetary orbit lies on a curve that is either an ellipse, a parabola, or a hyperbola, given by (2.12.37).

(2) Each orbit sweeps out an area at a constant rate, given by (2.12.17).

(3) The period of each elliptical orbit is given by (2.12.51).

As for the detailed motion of the planet about the ellipse (2.12.41), note that (2.12.14) plus (2.12.27) (with $\theta = \theta_0$) give

$$\theta'(t) = \frac{K^2}{L^3}(1 + e\cos\theta)^2, \tag{2.12.52}$$

hence

$$\int_0^{\varphi} \frac{d\theta}{(1 + e\cos\theta)^2} = \frac{K^2}{L^3}\tau, \quad \text{if } \theta(0) = 0,\ \theta(\tau) = \varphi. \tag{2.12.53}$$

In particular,

$$\int_0^{2\pi} \frac{d\theta}{(1 + e\cos\theta)^2} = \frac{K^2}{L^3}T. \tag{2.12.54}$$

The integral in (2.12.53) can be evaluated by elementary means, but rather than dwelling on this, we refer to [**15**], p. 166.

We look at arclength of an ellipse. An ellipse of the form (2.12.38) can be parametrized by

$$\gamma(t) = a\cos t + (b\sin t)i. \tag{2.12.55}$$

We have

$$\begin{aligned} |\gamma'(t)|^2 &= a^2\cos^2 t + b^2\sin^2 t \\ &= b^2 + \eta\sin^2 t, \quad \eta = a^2 - b^2, \end{aligned} \tag{2.12.56}$$

so the length of $\gamma([0,t])$ is

$$\ell_\gamma(t) = b\int_0^t \sqrt{1+\sigma\sin^2 s}\,ds, \quad \sigma = \frac{\eta}{b^2}. \tag{2.12.57}$$

If $a \neq b$, this is not an elementary integral. It is known as an elliptic integral. Material on such integrals, and the related theory of elliptic functions, can be found in Chapter 6 of [**17**].

Parabolas

Taking $e = 1$ in (2.12.44) gives

$$2px + y^2 = p^2. \tag{2.12.58}$$

Then taking $x \mapsto x + p/2$ gives $y^2 = -2px$, and then taking $x \mapsto y, y \mapsto -x$ yields

$$y = ax^2, \quad a = \frac{1}{2p}. \tag{2.12.59}$$

The parabola (2.12.58) is described as in (2.12.41), with $e = 1$, i.e.,

$$r(1+\cos\theta) = p = \frac{L^2}{K}. \tag{2.12.60}$$

Parallel to (2.12.52), we have

$$\theta'(t) = \frac{K^2}{L^3}(1+\cos\theta)^2 = \frac{4K^2}{L^3}\cos^4\frac{\theta}{2}. \tag{2.12.61}$$

Hence, parallel to (2.12.53), we have

$$\int_0^\varphi \sec^4\frac{\theta}{2}\,d\theta = \frac{4K^2}{L^3}\tau, \quad \text{if } \theta(0) = 0,\ \theta(\tau) = \varphi. \tag{2.12.62}$$

Clearly $\tau \to \pm\infty$ as $\varphi \to \pm\pi$. Material on p. 166 of [**15**] also provides an elementary evaluation of this integral.

We look at the arclength of a parabola. A parabola of the form (2.12.59) can be parametrized by

$$\gamma(t) = t + at^2 i. \tag{2.12.63}$$

We have

$$|\gamma'(t)|^2 = 1 + 4a^2t^2, \tag{2.12.64}$$

so the length of $\gamma([0,t])$ is

$$\begin{aligned}\ell_\gamma(t) &= \int_0^t \sqrt{1+4a^2s^2}\,ds \\ &= \frac{1}{2a}\int_0^{2at}\sqrt{1+\sigma^2}\,d\sigma.\end{aligned} \tag{2.12.65}$$

This integral has an elementary evaluation. A treatment is given in Exercise 2 of §2.4.

Hyperbolas

Taking $e > 1$ in (2.12.44) gives hyperbolas. Parallel to (2.12.46), we can complete the square and write

$$-\Big(\sqrt{e^2-1}\,x - \frac{ep}{\sqrt{e^2-1}}\Big)^2 + y^2 = -\frac{p^2}{e^2-1}, \tag{2.12.66}$$

and then implementing a translation

$$x \mapsto x - \frac{ep}{e^2-1} \tag{2.12.67}$$

gives the curve

$$-(e^2-1)x^2 - y^2 = -\frac{p^2}{e^2-1}. \tag{2.12.68}$$

This in turn leads to (2.12.39), with

$$a = \frac{p}{e^2-1}, \quad b = \frac{p}{\sqrt{e^2-1}} = p^{1/2}a^{1/2}. \tag{2.12.69}$$

See Figure 2.12.2. Note also that $b = a\sqrt{e^2-1}$.

For the hyperbola, as described in (2.12.41), we continue to have the planetary motion described by (2.12.52), this time with $e > 1$. Hence, as in (2.12.53),

$$\int_0^\varphi \frac{d\theta}{(1+e\cos\theta)^2} = \frac{K^2}{L^3}\tau, \quad \text{if } \ \theta(0) = 0,\ \theta(\tau) = \varphi, \tag{2.12.70}$$

this time under the restriction that

$$|\varphi| < \pi \ \text{ and } \ e\cos\varphi > -1, \tag{2.12.71}$$

and clearly $\tau \to \pm\infty$ as $\varphi \to \pm\cos^{-1}(-1/e)$. As in (2.12.59), the integral (2.12.70) can be evaluated by means described in [**15**], p. 166.

Note that if we have θ run over $[-\pi, \pi]$, then $r = r(\theta)$ in (2.12.41) is undefined at $\theta = \pm\cos^{-1}(-1/e)$, and when θ crosses these points, $r(\theta)$ changes sign and the curve $r(\theta)e^{i\theta}$ jumps from one branch of the hyperbola to the other.

We turn to a study of arclength of a hyperbola, presented in the form (2.12.39). This curve has two components. The right component is parametrized by

$$\gamma(t) = a\cosh t + ib\sinh t, \tag{2.12.72}$$

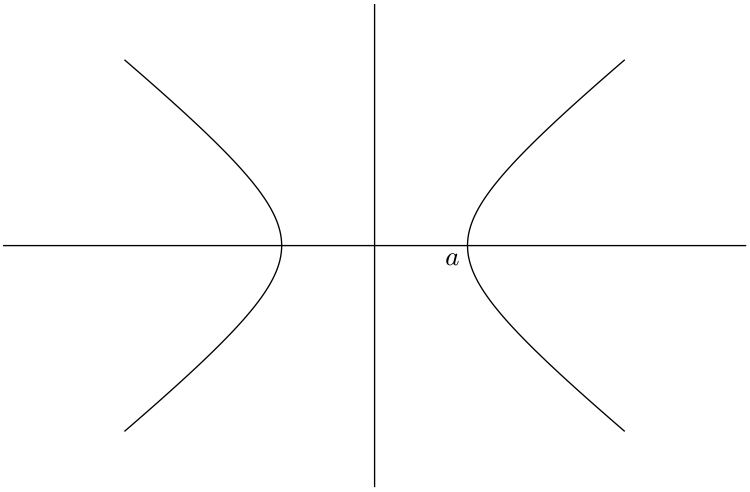

Figure 2.12.2. Hyperbola $x^2/a^2 - y^2/b^2 = 1$

assuming $a > 0$. (For the left component, replace a by $-a$.) In such a case,

$$\gamma'(t) = a \sinh t + ib \cosh t, \tag{2.12.73}$$

and

$$|\gamma'(t)|^2 = a^2 \sinh^2 t + b^2 \cosh^2 t. \tag{2.12.74}$$

Hence the length of $\gamma([0,t])$ is

$$\ell_\gamma(t) = \int_0^t \sqrt{a^2 \sinh^2 t + b^2 \cosh^2 t}\, dt. \tag{2.12.75}$$

As with (2.12.57), this is not an elementary integral. Unlike the case of ellipses, this is the case even for $a = b$ (for which $e = \sqrt{2}$). For $a = b = 1$, we have

$$\begin{aligned} \ell_\gamma(t) &= \int_0^t \sqrt{\sinh^2 t + \cosh^2 t}\, dt \\ &= \int_0^t \sqrt{1 + 2\sinh^2 t}\, dt. \end{aligned} \tag{2.12.76}$$

Note that for $a = b = 1$ we have the hyperbola

$$x^2 - y^2 = 1, \tag{2.12.77}$$

whose right half is parametrized by

$$\gamma(t) = \cosh t + i \sinh t. \tag{2.12.78}$$

It seems a pity that this does not share with

$$e^{it} = \cos t + i \sin t \tag{2.12.79}$$

the property of being a unit speed parametrization of its curve (S^1 in case (2.12.79)). We can rectify this, upon replacing the complex plane $\mathbb{C}$, with its Euclidean (positive definite) inner product, by the *Minkowski plane* $\mathbb{M}$, equipped with an indefinite inner product, which we proceed to describe.

The Minkowski plane

Recall that the complex plane $\mathbb{C}$ was formed by adjoining to $\mathbb{R}$ an element i, satisfying $i^2 = -1$. To form $\mathbb{M}$, adjoin to $\mathbb{R}$ a different element, ι, satisfying $\iota^2 = 1$. Thus a general element of $\mathbb{M}$ has the form

$$z = x + \iota y, \quad x, y \in \mathbb{R}. \tag{2.12.80}$$

Parallel to (1.1.3), we set

$$\operatorname{Re} z = x, \quad \operatorname{Im} z = y. \tag{2.12.81}$$

If also $w = u + \iota v$, we define

$$\begin{aligned} z + w &= (x + u) + \iota(y + v), \\ zw &= (xu + yv) + \iota(xv + yu). \end{aligned} \tag{2.12.82}$$

As for $\mathbb{C}$, addition and multiplication on $\mathbb{M}$ satisfy the usual commutative, associative, and distributive laws. This structure makes $\mathbb{M}$ what is called a commutative ring.

Parallel to (1.1.7) we define a conjugation,

$$z = x + \iota y \Longrightarrow \overline{z} = x - \iota y. \tag{2.12.83}$$

This time we have

$$z\overline{z} = x^2 - y^2. \tag{2.12.84}$$

Next we define the $\mathbb{R}$-bilinear inner product on $\mathbb{M}$:

$$\langle z, w \rangle_M = \operatorname{Re} z\overline{w}. \tag{2.12.85}$$

A calculation parallel to (2.12.82) gives

$$\langle z, w \rangle_M = xu - yv, \tag{2.12.86}$$

and

$$\langle z, z \rangle_M = z\overline{z}. \tag{2.12.87}$$

We also use the notation

$$Q_M(z) = \langle z, z\rangle_M. \tag{2.12.88}$$

It follows from (2.12.87), together with the commutative and associative laws for multiplication on $\mathbb{M}$, that

$$Q_M(zw) = Q_M(z)Q_M(w). \tag{2.12.89}$$

It follows from (2.12.89) that if $Q_M(z) = 0$, then z does not have a multiplicative inverse, so $\mathbb{M}$ is not a field. On the other hand,

$$Q_M(z) \neq 0 \Longrightarrow \frac{1}{z} = \frac{\overline{z}}{Q_M(z)}. \tag{2.12.90}$$

Given nonzero $z \in \mathbb{M}$, we say

$$\begin{aligned} z \text{ is spacelike} &\Longleftrightarrow Q_M(z) > 0, \\ z \text{ is timelike} &\Longleftrightarrow Q_M(z) < 0, \\ z \text{ is lightlike} &\Longleftrightarrow Q_M(z) = 0. \end{aligned} \tag{2.12.91}$$

This terminology celebrates the role of the Minkowski plane (and higher dimensional variants) in the theory of relativity. We will not go into the connection with relativity here; see Chapter 18 of [**22**]. In any case, we set

$$|z| = |Q_M(z)|^{1/2}. \tag{2.12.92}$$

To proceed, if $\gamma : (a, b) \to \mathbb{M}$ is a smooth curve with nonvanishing derivative, we say

$$\begin{aligned} \gamma \text{ is spacelike} &\Longleftrightarrow \gamma'(t) \text{ is spacelike}, \\ \gamma \text{ is timelike} &\Longleftrightarrow \gamma'(t) \text{ is timelike}, \\ \gamma \text{ is lightlike} &\Longleftrightarrow \gamma'(t) \text{ is lightlike}, \end{aligned} \tag{2.12.93}$$

for all $t \in (a, b)$. If γ satisfies one of these conditions, we set

$$\ell_\gamma(t) = \int_a^t |\gamma'(t)|\, dt. \tag{2.12.94}$$

We say γ is a unit speed curve $\Leftrightarrow |\gamma'(t)| = 1$ for all $t \in (a, b)$.

We return to the hyperbola (2.12.77), and note the following property of a natural parametrization of its right half.

Proposition 2.12.1. *The curve*

$$\gamma(t) = \cosh t + \iota \sinh t \tag{2.12.95}$$

is a unit-speed, timelike curve in $\mathbb{M}$.

Proof. We have

$$\gamma'(t) = \sinh t + \iota \cosh t, \tag{2.12.96}$$

hence

$$\langle \gamma'(t), \gamma'(t) \rangle_M = \sinh^2 t - \cosh^2 t \equiv -1. \tag{2.12.97}$$

□

Here is the Minkowski variant of (2.12.79).

Proposition 2.12.2. *We have, for all* $t \in \mathbb{R}$,

$$e^{\iota t} = \cosh t + \iota \sinh t. \tag{2.12.98}$$

Proof. Note that, for $\gamma(t)$ in (2.12.95),

$$\gamma'(t) = \iota \gamma(t). \tag{2.12.99}$$

An argument parallel to that used in (2.10.35)–(2.10.44) then yields (2.12.98).

Alternatively, using $\iota^2 = 1$, write

$$\begin{aligned} e^{\iota t} &= \sum_{k=0}^{\infty} \frac{\iota^k}{k!} t^k \\ &= \sum_{\ell=0}^{\infty} \frac{t^{2\ell}}{(2\ell)!} + \iota \sum_{\ell=0}^{\infty} \frac{t^{2\ell+1}}{(2\ell+1)!} \\ &= \cosh t + \iota \sinh t, \end{aligned} \tag{2.12.100}$$

the last identity following from the definitions

$$\cosh t = \tfrac{1}{2}(e^t + e^{-t}), \quad \sinh t = \tfrac{1}{2}(e^t - e^{-t}), \tag{2.12.101}$$

and the power series for e^t and e^{-t}. □

Remark. The complex plane $\mathbb{C}$ and the Minkowski plane $\mathbb{M}$ are the two Clifford algebras generated by $\mathbb{R}$. For material on Clifford algebras generated by $\mathbb{R}^n$, see Chapter 8 of [**14**], or Chapter 7 of [**19**].

Exercises

1. Suppose u satisfies (2.12.34). Show that you can write

$$u = v + \frac{K}{L^2}, \quad \text{with} \quad \frac{d^2 v}{d\theta^2} + v = 0.$$

Review Exercises 4–7 of §2.4 to show that the general real-valued solution to (2.12.34) is given by (2.12.35).

2. Consider the hyperbola

$$xy = 1.$$

Produce a rotation that puts this in the form (2.12.39). What are a and b?

3. Given the hyperbola of Exercise 2, parametrize its right side by

$$\gamma(t) = t + it^{-1}.$$

Write down the integral formula for the length of $\gamma([1, t])$.

4. Let $A(t) = \text{Area}\{x + it : 0 \le x \le t,\ 0 \le y \le \sqrt{1 + x^2}\}$, the area of a region under the hyperbola $y^2 - x^2 = 1$ and above the x-axis. Show that

$$A(t) = \int_0^t \sqrt{1 + x^2}\, dx.$$

Compare this with the arclength of a parabola considered in Exercise 1 of §2.2. Note the treatment of this integral in Exercise 2 of §2.4.

5. Use the parametrization

$$\gamma(\theta) = r(\theta)e^{i\theta}, \quad r(\theta) = p(1 + e\cos\theta)^{-1},$$

from (2.12.41), to produce integrals for arclength of the ellipses, parabolas, and hyperbolas studied here. Start with the identity

$$|\gamma'(\theta)|^2 = r'(\theta)^2 + r(\theta)^2.$$

6. Verify that the conjugation $z \mapsto \overline{z}$ on $\mathbb{M}$ given by (2.12.83) satisfies

$$\overline{z + w} = \overline{z} + \overline{w}, \quad \overline{zw} = \overline{z}\,\overline{w}.$$

See how this figures in the verification of (2.12.89).

7. Show that, for $z, w, a \in \mathbb{M}$,

$$\langle az, aw\rangle_M = Q_M(a)\langle z, w\rangle_M.$$

8. We consider how solutions to a cubic equation like (2.8.35) can be specified as points in the intersection of a hyperbola and a parabola,

$$(x - a)^2 - y^2 = a^2, \quad y = bx^2.$$

Given $a, b \in \mathbb{R} \setminus 0$, show that there are intersection points $0 + i0$ and $x + ibx^2$, where x solves

$$b^2x^3 - x + 2a = 0.$$

2.13. Higher dimensions

Beyond plane geometry, we have n-dimensional Euclidean geometry, for each $n \in \mathbb{N}$. We sketch some results on this subject here, referring to Chapter 3 of [**14**] for a systematic development of the algebraic aspects, and to [**16**] for the very rich theory of calculus of several variables.

We start with $\mathbb{R}^n$, which consists of n-tuples of real numbers,

$$x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad x_j \in \mathbb{R},\ 1 \le j \le n. \tag{2.13.1}$$

The number x_j is called the jth component of x. We discuss some algebraic and metric structures on $\mathbb{R}^n$. First, there is addition. If x is as in (2.13.1) and also $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we have

$$x + y = (x_1 + y_1, \dots, x_n + y_n) \in \mathbb{R}^n. \tag{2.13.2}$$

Also, given $a \in \mathbb{R}^n$, we have

$$ax = (ax_1, \dots, ax_n) \in \mathbb{R}^n. \tag{2.13.3}$$

This is scalar multiplication. In (2.13.1), we represent x as a row vector. Sometimes it is convenient to represent x as a column vector,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{2.13.4}$$

Then (2.13.2)–(2.13.3) are converted to

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad ax = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}. \tag{2.13.5}$$

As seen in Section 1.2 and subsequent material, the inner product played a central role in 2D geometry. In the setting of $\mathbb{R}^n$, we have the inner product

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j y_j = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}, \tag{2.13.6}$$

given $x, y \in \mathbb{R}^n$. This inner product has the properties

$$\begin{aligned} &\langle x, y\rangle = \langle y, x\rangle, \\ &\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle, \\ &\langle x, x\rangle > 0 \ \text{ unless } \ x = 0. \end{aligned} \tag{2.13.7}$$

Note that

$$\langle x, x\rangle = x_1^2 + \cdots + x_n^2. \tag{2.13.8}$$

We set

$$|x| = \sqrt{\langle x, x\rangle}, \tag{2.13.9}$$

which we call the norm of x. Note that

$$\langle ax, ax\rangle = a^2\langle x, x\rangle, \tag{2.13.10}$$

hence

$$|ax| = |a| \cdot |x|, \quad \text{for } \ a \in \mathbb{R},\ x \in \mathbb{R}^n. \tag{2.13.11}$$

Extending the notion of distance in (1.2.13), we say that the distance from x to y in $\mathbb{R}^n$ is

$$d(x, y) = |x - y|. \tag{2.13.12}$$

We claim that the Euclidean distance, defined by (2.13.12), satisfies the "triangle inequality,"

$$d(x, y) \le d(x, z) + d(z, y), \quad \forall\, x, y, z \in \mathbb{R}^n \tag{2.13.13}$$

This in turn is a consequence of the following, also called the triangle inequality.

Proposition 2.13.1. *The norm* (2.13.9) *on* $\mathbb{R}^n$ *has the property*

$$|x + y| \le |x| + |y|, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.13.14}$$

Proof. We compare the squares of the two sides of (2.13.14). first,

$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y\rangle \\ &= |x|^2 + |y|^2 + 2\langle x, y\rangle. \end{aligned} \tag{2.13.15}$$

Next,

$$(|x| + |y|)^2 = |x|^2 + |y|^2 + 2|x| \cdot |y|. \tag{2.13.16}$$

We see that (2.13.14) holds if and only if $\langle x, y\rangle \le |x| \cdot |y|$. Thus the proof of Proposition 2.13.1 is finished by the following result, known as Cauchy's inequality. □

Proposition 2.13.2. *For all* $x, y \in \mathbb{R}^n$,

$$|\langle x, y\rangle| \le |x| \cdot |y|. \tag{2.13.17}$$

Proof. We start with the chain

$$0 \le |x - y|^2 = \langle x - y, x - y\rangle = |x|^2 + |y|^2 - 2\langle x, y\rangle, \tag{2.13.18}$$

which implies

$$2\langle x, y\rangle \le |x|^2 + |y|^2, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.13.19}$$

If we replace x by tx and y by $t^{-1}y$, with $t > 0$, the left side of (2.13.19) is unchanged, so we have

$$2\langle x, y\rangle \le t^2|x|^2 + t^{-2}|y|^2, \quad \forall\, t > 0. \tag{2.13.20}$$

We pick t so that the two terms on the right side of (2.13.20) are equal, namely,

$$t^2 = \frac{|y|}{|x|}, \quad t^{-2} = \frac{|x|}{|y|}. \tag{2.13.21}$$

Note that (2.13.17) is obvious if $x = 0$ or $y = 0$, so we will assume that $x \neq 0$ and $y \neq 0$. Plugging (2.13.21) into (2.13.20) gives

$$\langle x, y\rangle \leq |x| \cdot |y|, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.13.22}$$

This is almost (2.13.17). To finish, we can replace x in (2.13.22) by $-x = (-1)x$, getting

$$-\langle x, y\rangle \leq |x| \cdot |y|, \tag{2.13.23}$$

and together (2.13.22) and (2.13.23) give (2.13.17). □

Let us emphasize the conclusion of (2.13.15) as a very significant extension of (1.2.3):

$$|x + y|^2 = |x|^2 + |y|^2 + 2\langle x, y\rangle, \tag{2.13.24}$$

for all $x, y \in \mathbb{R}^n$. Parallel to (1.2.6), we write

$$x \perp y \Longleftrightarrow \langle x, y\rangle = 0, \tag{2.13.25}$$

and say x is orthogonal (or perpendicular) to y in $\mathbb{R}^n$. Parallel to Proposition 1.2.1 we have:

Proposition 2.13.3. *Given $x, y \in \mathbb{R}^n$,*

$$|x + y|^2 = |x|^2 + |y|^2 \Longleftrightarrow x \perp y. \tag{2.13.26}$$

To proceed, it is desirable to extend our scope, and consider more general inner product spaces V. First, V is a vector space. That is, given $x, y \in V$, $a \in \mathbb{R}$, we have defined vector addition, and multiplication by a scalar,

$$x, y \in V,\ a \in \mathbb{R} \Longrightarrow x + y,\ ax \in V. \tag{2.13.27}$$

These operations verify a standard set of commutative, associative, and distributive laws. See Chapter 1 of [**14**]. We mention that a map $T : V \to V$ is called a linear transformation provided T preserves these vector operations, i.e.,

$$T(ax + by) = aTx + bTy, \quad \forall\, x, y \in V,\ a, b \in \mathbb{R}. \tag{2.13.28}$$

As shown in Chapter 1 of [**14**], in case $V = \mathbb{R}^n$, matrix multiplication associates each $n \times n$ matrix $A \in M(n, \mathbb{R})$ to a linear transformation on $\mathbb{R}^n$.

An inner product space is a vector space V, equipped with an inner product:

$$x, y \in V \Longrightarrow \langle x, y\rangle \in \mathbb{R}, \tag{2.13.29}$$

satisfying (2.13.7).

Example. The integral formula

$$\langle f, g\rangle = \int_a^b f(t)g(t)\,dt \tag{2.13.30}$$

defines an inner product on the space $C([a,b])$ of continuous functions on $[a,b]$. It also defines an inner product on $\mathcal{P}$, the space of polynomials in t.

For a general inner product space V, (2.13.9) defines a norm on V. Generally, a norm on V is a function $x \mapsto |x|$ satisfying

$$\begin{aligned} |ax| &= |a|\cdot|x|, \quad \forall\, a \in \mathbb{R},\ x \in V, \\ |x| &> 0 \ \text{ unless } \ x = 0, \\ |x+y| &\le |x| + |y|. \end{aligned} \tag{2.13.31}$$

As before, the last of these results is called the triangle inequality. The proof that this holds for a general inner product space is identical to that of Proposition 2.13.1 (for $\mathbb{R}^n$). In particular, Cauchy's inequality (2.13.17) arises.

We specialize to finite-dimensional vector spaces V, i.e., to the case where there is a finite set $S = \{v_1, \dots, v_n\} \subset V$ such that $\operatorname{Span} S = V$, where

$$\operatorname{Span}\{v_1, \dots, v_n\} = \Big\{\sum_{j=1}^n a_j v_j : a_j \in \mathbb{R}\Big\}. \tag{2.13.32}$$

A minimal spanning set of V is called a basis. It is a fundamental fact that any two bases of V have the same number of elements (see [**14**], Corollary 1.3.3). This number is called the dimension of V, and denoted $\dim V$. In particular, of course, $\dim \mathbb{R}^n = n$.

If V is a finite-dimensional inner product space, a basis $\{u_1, \dots, u_n\}$ of V is called an orthonormal basis of V provided

$$\langle u_j, u_k\rangle = \delta_{jk}, \quad 1 \le j, k \le n, \tag{2.13.33}$$

i.e.,

$$|u_j| = 1, \quad j \neq k \Rightarrow \langle u_j, u_k\rangle = 0. \tag{2.13.34}$$

When (2.13.33) holds, we have

$$v = \sum_{j=1}^n a_j u_j,\ w = \sum_{j=1}^n b_j u_j \Longrightarrow \langle u, v\rangle = \sum_{j=1}^n a_j b_j. \tag{2.13.35}$$

It is very useful to be able to construct orthonormal bases. The construction we now describe is called the Gram-Schmidt construction.

Proposition 2.13.4. *Let $\{v_1, \dots v_n\}$ be a basis of V, an inner product space. Then there is an orthonormal basis $\{u_1, \dots, u_n\}$ of V such that*

$$\text{Span}\{u_j : 1 \le j \le \ell\} = \text{Span}\{v_j : 1 \le j \le \ell\}, \quad 1 \le \ell \le n. \tag{2.13.36}$$

Proof. To begin, take

$$u_1 = \frac{1}{|v_1|} v_1. \tag{2.13.37}$$

Now define the linear transformation $P_1 : V \to V$ by $P_1 v = \langle v, u_1\rangle u_1$ and set

$$\tilde{v}_2 = v_2 - P_1 v_2 = v_2 - \langle v_2, u_1\rangle u_1. \tag{2.13.38}$$

We see that $\langle \tilde{v}_2, u_1\rangle = \langle v_2, u_1\rangle - \langle v_2, u_1\rangle = 0$. Also $\tilde{v}_2 \neq 0$ since $v_2 \notin \text{Span}(u_1)$. Hence we can set

$$u_2 = \frac{1}{|\tilde{v}_2|} \tilde{v}_2. \tag{2.13.39}$$

Inductively, suppose we have an orthonormal set $\{u_1, \dots, u_m\}$ with $m < n$ and (2.13.36) holding for $1 \le \ell \le m$. Then define $P_m : V \to V$ (the orthogonal projection of V onto $\text{Span}(u_1, \dots, u_m)$) by

$$P_m v = \sum_{j=1}^{m} \langle v, u_j\rangle u_j, \tag{2.13.40}$$

and set

$$\begin{aligned} \tilde{v}_{m+1} &= v_{m+1} - P_m v_{m+1} \\ &= v_{m+1} - \sum_{j=1}^{m} \langle v_{m+1}, u_j\rangle u_j. \end{aligned} \tag{2.13.41}$$

We see that

$$j \le n \Rightarrow \langle \tilde{v}_{m+1}, u_j\rangle = \langle v_{m+1}, u_j\rangle - \langle v_{m+1}, u_j\rangle = 0. \tag{2.13.42}$$

Also, since $v_{m+1} \notin \text{Span}(v_1, \dots, v_m) = \text{Span}(u_1, \dots, u_m)$, it follows that $\tilde{v}_{m+1} \neq 0$. Hence we can set

$$u_{m+1} = \frac{1}{|\tilde{v}_{m+1}|} \tilde{v}_{m+1}. \tag{2.13.43}$$

This completes the construction. □

The following is a useful extension. Let V be an n-dimensional inner product space, and let $W \subset V$ be a linear subspace, i.e., a subset satisfying

$$w_j \in W,\ a_j \in \mathbb{R} \Longrightarrow \sum a_j w_j \in W. \tag{2.13.44}$$

Then W has an orthonormal basis $\{w_1, \dots, w_m\}$ ($m = \dim W$), and

$$P_W : V \longrightarrow W, \quad P_W v = \sum_{j=1}^{m} \langle v, w_j \rangle w_j \tag{2.13.45}$$

is the orthogonal projection of V onto W. Furthermore,

$$I - P_W : V \longrightarrow W^\perp \tag{2.13.46}$$

is the orthogonal projection of V onto

$$W^\perp = \{v \in V : v \perp w, \, \forall \, w \in W\}. \tag{2.13.47}$$

Given an n-dimensional inner product space V, with orthonormal basis $S = \{u_1, \dots, u_n\}$, we can define a linear map

$$\mathcal{J}_S : \mathbb{R}^n \to V, \quad \mathcal{J}_S \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \sum_{j=1}^{n} a_j u_j. \tag{2.13.48}$$

Equivalently,

$$\mathcal{J}_S e_j = u_j, \quad 1 \le j \le n, \tag{2.13.49}$$

where $\{e_1, \dots, e_n\}$ is the standard orthonormal basis of $\mathbb{R}^n$:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \; e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{2.13.50}$$

The map $\mathcal{J}_S$ is bijective, with inverse

$$\mathcal{J}_S^{-1} : V \longrightarrow \mathbb{R}^n, \quad \mathcal{J}_S^{-1} \sum a_j u_j = \sum a_j e_j. \tag{2.13.51}$$

The map $\mathcal{J}_S$ also preserves inner products:

$$\langle \mathcal{J}_S x, \mathcal{J}_S y \rangle = \langle x, y \rangle, \quad \forall \, x, y \in \mathbb{R}^n, \tag{2.13.52}$$

where the inner product on the right is given by (2.13.6). The maps $\mathcal{J}_S$ and $\mathcal{J}_S^{-1}$ provide what are known as Cartesian coordinates on the space V, celebrating the work of R. Descartes, who initiated the algebraic approach to Euclidean geometry.

For a related concept, if V is an inner product space, and $T : V \to V$ is a linear transformation, we say

$$T \in \mathrm{O}(V) \Longleftrightarrow \langle Tu, Tv \rangle = \langle u, v \rangle, \quad \forall \, u, v, \in V. \tag{2.13.53}$$

We denote $\mathrm{O}(\mathbb{R}^n)$ by $\mathrm{O}(n)$. If $A \in M(n, \mathbb{R})$ defines a linear transformation,

$$A : \mathbb{R}^n \to \mathbb{R}^n, \quad (Ax)_j = \sum_k a_{jk} x_k, \; A = (a_{jk}), \tag{2.13.54}$$

then

$$A \in \mathrm{O}(n) \Longleftrightarrow A^t A = I,\ A^t = (a_{kj}). \tag{2.13.55}$$

Given an n-dimensional inner product space V, with distance $d(u,v) = |u-v|$, and a map $F : V \to V$, we say

$$F \in \mathrm{Isom}(V) \Longleftrightarrow d(F(u),F(v)) = d(u,v), \quad \forall\, u, v \in V. \tag{2.13.56}$$

We have the following extension of Corollary 1.3.4.

Proposition 2.13.5. *Given $F \in \mathrm{Isom}(V)$, then F has the form*

$$F(v) = Av + y, \quad A \in \mathrm{O}(V),\ y \in V. \tag{2.13.57}$$

In fact, as in §1.3, there is a decomposition $\mathrm{Isom}(V) = \mathrm{Isom}^+(V) \cup \mathrm{Isom}^-(V)$. To describe this, it suffices to take $V = \mathbb{R}^n$, so (2.13.57) becomes

$$F(v) = Av + y, \quad A \in \mathrm{O}(n),\ y \in \mathbb{R}^n, \tag{2.13.58}$$

and to see that

$$\mathrm{O}(n) = \mathrm{O}^+(n) \cup \mathrm{O}^-(n). \tag{2.13.59}$$

To do this, we bring in the determinant,

$$\det : M(n,\mathbb{R}) \longrightarrow \mathbb{R}. \tag{2.13.60}$$

To characterize $\det A$ for $A \in M(n,\mathbb{R})$, we write $A = (a_1, \dots, a_n)$, where each $a_j \in \mathbb{R}^n$ is a column vector. The following result is the key to determinants.

Proposition 2.13.6. *There is a unique function $\vartheta : M(n,\mathbb{R}) \to \mathbb{R}$ such that*

(a) $\vartheta(A)$ *is linear in each column of* A,
(b) $\vartheta(\widetilde{A}) = -\vartheta(A)$ *if* $\widetilde{A}$ *is obtained by switching two columns of* A,
(c) $\vartheta(I) = 1$.

We denote this function by $\det A$. *If (c) is changed to*

(c′) $\vartheta(I) = r$,

then $\vartheta(A) = r \det A$.

For a proof, see [**14**], §1.5. The analysis there gives also

$$\det AB = (\det A)(\det B), \quad \det A^t = \det A. \tag{2.13.61}$$

Then (2.13.55) yields

$$A \in \mathrm{O}(n) \Rightarrow (\det A)^2 = 1 \Rightarrow \det A = \pm 1. \tag{2.13.62}$$

To define $\mathrm{O}^{\pm}(n)$, we set

$$\mathrm{O}^{\pm}(n) = \{A \in \mathrm{O}(n) : \det A = \pm 1\}. \tag{2.13.63}$$

We also use the notation

$$\mathrm{SO}(n) = \mathrm{O}^{+}(n). \tag{2.13.64}$$

Returning to orthogonal complements, we mention an alternative way to find $W^\perp$ when W is a 2-dimensional subspace of $\mathbb{R}^3$, using the cross product, an $\mathbb{R}$-bilinear map $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$i \times j = k, \quad j \times k = i, \quad k \times i = j, \quad u \times v = -v \times u. \tag{2.13.65}$$

Here i, j, k is a typical notation for the standard basis e_1, e_2, e_3 of $\mathbb{R}^3$. A convenient alternative characterization is the following.

Proposition 2.13.7. *Given $u, v, w \in \mathbb{R}^3$,*

$$\langle w, u \times v\rangle = \det \begin{pmatrix} w_1 & u_1 & v_1 \\ w_2 & u_2 & v_2 \\ w_3 & u_3 & v_3 \end{pmatrix}. \tag{2.13.66}$$

Proof. The linearity of the right side in u, v, w and the antisymmetry follow from Proposition 2.13.6. It is straightforward to check the identities $i \times j = k$, etc., listed in (2.13.65). □

The antisymmetry of the determinant also implies

$$w \in \mathrm{Span}(u, v) \Longrightarrow \langle w, u \times v\rangle = 0, \tag{2.13.67}$$

hence

$$u \times v \perp \mathrm{Span}(u, v). \tag{2.13.68}$$

Consequently, if $\{u, v\}$ span the 2-dimensional space $W \subset \mathbb{R}^3$, then

$$N = \frac{1}{|u \times v|}(u \times v) \tag{2.13.69}$$

is a unit vector spanning $W^\perp$. Thus

$$P_{W^\perp} w = \langle w, N\rangle N. \tag{2.13.70}$$

We mention another useful fact about the cross product.

Proposition 2.13.8. *Given $u, v \in \mathbb{R}^3$,*

$$T \in \mathrm{SO}(3) \Longrightarrow Tu \times Tv = T(u \times v). \tag{2.13.71}$$

Proof. Denote the matrix on the right side of (2.13.66) by M. Then, if $T \in \mathrm{SO}(3)$,

$$\langle Tw, Tu \times Tv\rangle = \det TM = (\det T)(\det M) = \det M, \tag{2.13.72}$$

hence $T^t(Tu \times Tv) = u \times v$, yielding (2.13.71). □

We briefly mention the regular solids in $\mathbb{R}^3$, also called the Platonic solids. In contrast to the infinite family of regular polygons, there are only five:

tetrahedron,
cube,
octahedron,
dodecahedron,
icosahedron,

having, respectively, 4,6,8,12, and 20 faces. Constructions of the first three are straightforward. See Exercises 4–5 below. The last two are discussed in [**20**]. See also [**2**] for a detailed treatment.

We move on to a discussion of volume. There is a notion of n-dimensional volume for certain bounded sets $S \subset \mathbb{R}^n$, defined in a fashion parallel to the definition of area for the planar case done in §2.6. To start, we define a standard n-dimensional cube Q of side-length ε parallel to (2.6.2),

$$\mathrm{Vol}\, Q = \varepsilon^n. \tag{2.13.73}$$

and then define $\mathrm{Cont}^{\pm}(S)$ and $\mathrm{Vol}(S)$ in a fashion parallel to (2.6.3)–(2.6.7). Parallel to Proposition 2.6.5, we have

Proposition 2.13.9. *If $S \subset \mathbb{R}^n$ is bounded, then*

$$\mathrm{Cont}^{\pm}(F(S)) = \mathrm{Cont}^{\pm}(S), \quad \forall\, F \in \mathrm{Isom}(\mathbb{R}^n). \tag{2.13.74}$$

However, we need a different proof of this result, since it is not so convenient to extend Lemma 2.6.4. In fact, it is convenient to take a different approach, yielding the following more general result.

Proposition 2.13.10. *Let $S \subset \mathbb{R}^n$ be bounded and assume $A \in M(n, \mathbb{R})$ is invertible. Then*

$$\mathrm{Cont}^{\pm}(A(S)) = |\det A|\, \mathrm{Cont}^{\pm}(S). \tag{2.13.75}$$

Here is an application of Proposition 2.13.10 to the volume of a parallelopiped in $\mathbb{R}^3$. To set it up, consider the standard unit cube in $\mathbb{R}^3$,

$$Q = \{x \in \mathbb{R}^3 : 0 \le x_j \le 1,\ 1 \le j \le 3\}, \quad \mathrm{Vol}\, Q = 1. \tag{2.13.76}$$

Let M denote the matrix appearing on the right side of (2.13.66), so $w = Mi,\ u = Mj,\ v = Mk$. The image

$$P = M(Q) \tag{2.13.77}$$

is a parallelopiped, with vertices $0, u, v, w, u+v, u+w, v+w$, and $u+v+w$. If we combine Proposition 2.13.10 with (2.13.66), we obtain the volume calculation

$$\operatorname{Vol} P = |\langle w, u \times v\rangle|. \tag{2.13.78}$$

A proof of Proposition 2.13.10 is sketched in Proposition 1.6.5 of [**14**]. A detailed proof is given in Proposition 3.1.12 of [**16**]. This proof is done in the setting of Riemann integrals of functions on $\mathbb{R}^n$. This result is extended in Proposition 3.1.14 of [**16**] to the identity

$$\int_\Omega f(y)\,dy = \int_{\mathcal{O}} f(G(x))|\det DG(x)|\,dx, \tag{2.13.79}$$

where $G : \mathcal{O} \to \Omega$ is a C^1 map with C^1 inverse and $DG(x)$ is the $n \times n$ matrix of first order partial derivatives of the components of $G(x)$. This is an n-dimensional extension of Proposition 2.1.5.

Extending the notion of arc-length of a planar curve, we have the notion of $(n-1)$-dimensional area of an $(n-1)$-dimensional surface $M \subset \mathbb{R}^n$. This material is developed in Chapter 3 of [**16**] (along with the k-dimensional area of a k-dimensional surface in $\mathbb{R}^n$, for $1 \le k \le n-1$). We restrict ourselves to a description of a selection of results established there, with a focus on the computation of volumes of balls and areas of spheres in $\mathbb{R}^n$,

$$B^n = \{x \in \mathbb{R}^n : |x| \le 1\}, \quad S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}. \tag{2.13.80}$$

To start, suppose M is the image of a C^1 map

$$\varphi : \mathcal{O} \longrightarrow \mathbb{R}^n, \tag{2.13.81}$$

where $\mathcal{O} \subset \mathbb{R}^{n-1}$ is smoothly bounded. Let $D\varphi(x)$ denote the $n \times (n-1)$ matrix of partial derivatives of the components of φ, and set

$$G(x) = D\varphi(x)^t D\varphi(x), \quad g(x) = \det G(x). \tag{2.13.82}$$

Then

$$\operatorname{Area} M = \int_{\mathcal{O}} \sqrt{g(x)}\,dx. \tag{2.13.83}$$

More generally, if $f : M \to \mathbb{R}$ is continuous,

$$\int_M f\,dS = \int_{\mathcal{O}} f(\varphi(x))\sqrt{g(x)}\,dx. \tag{2.13.84}$$

See (3.2.12) of [**16**]. In case $n = 3$, this becomes

$$\int_M f\,dS = \int_{\mathcal{O}} f(\varphi(x,y))|\partial_x\varphi \times \partial_y\varphi|\,dx\,dy. \tag{2.13.85}$$

See (3.2.18) of [**16**].

A surface such as S^{n-1} can be presented as a union of a family of images of C^1 maps (called coordinate patches)

$$\varphi_j : \mathcal{O}_j \longrightarrow \mathbb{R}^n, \quad M = \bigcup_{j=1}^N \varphi_j(\mathcal{O}_j). \tag{2.13.86}$$

A continuous function $f : M \to \mathbb{R}$ can be written

$$f = \sum_{j=1}^N f_j, \quad f_j = 0 \text{ outside } \varphi_j(\mathcal{O}_j). \tag{2.13.87}$$

Then we can set

$$\int_M f\,dS = \sum_{j=1}^N \int_M f_j\,dS = \sum_j \int_{\mathcal{O}_j} f_j(\varphi_j(x))\sqrt{g_j(x)}\,dx. \tag{2.13.88}$$

One shows that $\int_M f\,dS$ defined this way is independent of the choice of such coordinate charts and of the decomposition of f. See [**16**], §3.2.

We turn to calculations involving spherical polar coordinates, which will result in calculations of

$$V_n = \operatorname{Vol} B^n, \quad A_{n-1} = \operatorname{Area} S^{n-1}. \tag{2.13.89}$$

As seen in (3.2.28) of [**16**], we can use (2.13.79) to show that if f is a continuous function on $\mathbb{R}^n$ that decreases sufficiently rapidly at infinity, then

$$\int_{\mathbb{R}^n} f(x)\,dx = \int_{S^{n-1}} \left\{ \int_0^\infty f(r\omega) r^{n-1}\,dr \right\} dS(\omega). \tag{2.13.90}$$

In particular, for $f(x) = g(|x|)$, we have

$$\int_{\mathbb{R}^n} g(|x|)\,dx = A_{n-1} \int_0^\infty g(r) r^{n-1}\,dr. \tag{2.13.91}$$

By a limiting argument, (2.13.90) applies to $g(r) = 1$ for $0 \le r \le 1$, 0 otherwise, yielding

$$V_n = A_{n-1} \int_0^1 r^{n-1}\,dr = \frac{1}{n} A_{n-1}. \tag{2.13.92}$$

Recall we already have this for $n = 2$:

$$A_1 = \ell(S^1) = 2\pi, \quad V_2 = \operatorname{Area} \mathcal{D} = \pi. \tag{2.13.93}$$

We desire to compute A_{n-1} for $n \geq 3$. Along the way we pick up some further marvelous identities by applying (2.13.90) to the Gaussian, $f(x) = e^{-|x|^2}$. We are looking at

$$I_n = \int_{\mathbb{R}^n} e^{-|x|^2}\, dx. \tag{2.13.94}$$

Since $e^{-|x|^2} = e^{-x_1^2} \cdots e^{-x_n^2}$, we have

$$I_n = I_1^n. \tag{2.13.95}$$

Meanwhile, (2.13.91) implies

$$\begin{aligned} I_n &= A_{n-1} \int_0^\infty e^{-r^2} r^{n-1}\, dr \\ &= \frac{1}{2} A_{n-1} \int_0^\infty e^{-s} s^{n/2-1}\, ds. \end{aligned} \tag{2.13.96}$$

The case $n = 2$ yields an elementary integral:

$$I_2 = \frac{1}{2} A_1 \int_0^\infty e^{-s}\, ds = \frac{1}{2} A_1 = \pi, \tag{2.13.97}$$

the last identity thanks to (2.13.93). Hence, by (2.13.95),

$$I_1 = \pi^{1/2}, \quad I_n = \pi^{n/2}. \tag{2.13.98}$$

Therefore (2.13.96) yields

$$A_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}, \tag{2.13.99}$$

where

$$\Gamma(z) = \int_0^\infty e^{-s} s^{z-1}\, ds, \quad z > 0, \tag{2.13.100}$$

defines Euler's gamma function. We record our result.

Proposition 2.13.11. *For $n \geq 1$, the area of S^{n-1} is given by* (2.13.99).

Note that

$$A_0 = 2 \Longrightarrow \Gamma\Big(\frac{1}{2}\Big) = \pi^{1/2}, \tag{2.13.101}$$

and (2.13.97) yields

$$\Gamma(1) = 1. \tag{2.13.102}$$

We can evaluate $\Gamma(n/2)$ for other $n \in \mathbb{N}$ via the following result.

Proposition 2.13.12. *For $z > 0$,*

$$\Gamma(z+1) = z\Gamma(z). \tag{2.13.103}$$

Proof. Indeed,

$$\begin{aligned}\Gamma(z+1) &= \int_0^\infty e^{-s}s^z\,ds = -\int_0^\infty \Big(\frac{d}{ds}e^{-s}\Big)s^z\,ds \\ &= \int_0^\infty e^{-s}\frac{d}{ds}s^z\,ds \\ &= \int_0^\infty e^{-s}zs^{z-1}\,ds,\end{aligned} \tag{2.13.104}$$

the third identity involving integration by parts. □

For $n = 3$, we have

$$\begin{aligned}A_2 &= \frac{2\pi^{3/2}}{\Gamma(3/2)} = \frac{2\pi^{3/2}}{(1/2)\Gamma(1/2)} = 4\pi, \\ V_3 &= \frac{1}{3}A_2 = \frac{4\pi}{3},\end{aligned} \tag{2.13.105}$$

results obtained by Archimedes. Moving up a dimension, for $n = 4$ we have

$$\begin{aligned}A_3 &= \frac{2\pi^2}{\Gamma(2)} = 2\pi^2, \\ V_4 &= \frac{1}{4}A_3 = \frac{\pi^2}{2}.\end{aligned} \tag{2.13.106}$$

Having discussed various aspects of Euclidean geometry on $\mathbb{R}^n$, with emphasis on $n > 2$ in this section, we connect with the planar case, which was the focus of the previous sections. One clear difference is that in the plane we used multiplication in $\mathbb{C}$. For dimension $n \geq 3$, this role of multiplication shifted to multiplication of a matrix $A \in M(n, \mathbb{R})$ by a vector $x \in \mathbb{R}^n$. Indeed, $\mathbb{R}^n$ does not generally have the structure of an algebra with such nice properties that we exploited on $\mathbb{C}$, such as multiplicative inverses and a multiplicative norm. W. Hamilton spent some time trying to find such a structure on $\mathbb{R}^3$. Eventually he succeeded, not for $\mathbb{R}^3$, but for $\mathbb{R}^4$, with the quaternions, which we briefly discuss. Detailed treatments are given in Chapter 8 of [**14**] and Chapter 10 of [**19**].

The algebra $\mathbb{H}$ of quaternions is a 4-dimensional real vector space, with elements of the form

$$\xi = a + bi + cj + dk, \quad a, b, c, d \in \mathbb{R}. \tag{2.13.107}$$

Here, i, j, and k play a role similar to the standard basis of $\mathbb{R}^3$ that arose above. The product on $\mathbb{H}$ is an $\mathbb{R}$-bilinear map $\mathbb{H} \times \mathbb{H} \to \mathbb{H}$, with $1 \in \mathbb{H}$ acting as the multiplicative identity. Products involving i, j, and k have some features in common with the cross product:

$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik. \tag{2.13.108}$$

On the other hand, the squares of these objects follow a new rule:

$$i^2 = j^2 = k^2 = -1. \tag{2.13.109}$$

Otherwise stated, if we write

$$\xi = a + u, \quad a \in \mathbb{R},\ u \in \mathbb{R}^3, \tag{2.13.110}$$

and similarly write $\eta = b + v,\ b \in \mathbb{R},\ v \in \mathbb{R}^3$, the product is given by

$$\xi\eta = (a+u)(b+v) = (ab - u \cdot v) + av + bu + u \times v. \tag{2.13.111}$$

Here $u \cdot v = \langle u, v\rangle$ is the inner product on $\mathbb{R}^3$, and $u \times v$ is the cross product. The quantity $ab - u \cdot v$ is the real part of $\xi\eta$ and $av + bu + u \times v$ is the vector part.

Multiplication on $\mathbb{H}$ is not commutative, but we have the following basic result.

Proposition 2.13.13. *Multiplication on $\mathbb{H}$ is associative, i.e.,*

$$\zeta(\xi\eta) = (\zeta\xi)\eta, \quad \forall\, \zeta, \xi, \eta \in \mathbb{H}. \tag{2.13.112}$$

The proof is a straightforward check. See Proposition 8.1.2 in [**14**].

In addition to the product, we have a conjugation operator:

$$\overline{\xi} = a - bi - cj - dk = a - u. \tag{2.13.113}$$

A calculation gives

$$\xi\overline{\eta} = (ab + u \cdot v) - av + bu - u \times v. \tag{2.13.114}$$

In particular,

$$\mathrm{Re}(\xi\overline{\eta}) = \mathrm{Re}(\overline{\eta}\xi) = \langle \xi, \eta\rangle, \tag{2.13.115}$$

the right side denoting the inner product on $\mathbb{R}^4$. Setting $\eta = \xi$ in (2.13.114) gives

$$\xi\overline{\xi} = |\xi|^2, \tag{2.13.116}$$

the Euclidean square-norm of ξ. Hence, whenever $\xi \in \mathbb{H}$ is nonzero, it has a multiplicative inverse,

$$\xi^{-1} = |\xi|^{-2}\overline{\xi}. \tag{2.13.117}$$

This structure makes $\mathbb{H}$ what is called a division ring.

To continue with products and conjugates, a routine calculation gives

$$\overline{\xi\eta} = \overline{\eta}\,\overline{\xi}. \tag{2.13.118}$$

Hence, via the associative law,

$$\begin{aligned} |\xi\eta|^2 &= (\xi\eta)(\overline{\xi\eta}) = \xi\eta\overline{\eta}\,\overline{\xi} \\ &= |\eta|^2\xi\overline{\xi} = |\xi|^2|\eta|^2, \end{aligned} \tag{2.13.119}$$

so

$$|\xi\eta| = |\xi|\,|\eta|. \tag{2.13.120}$$

This makes $\mathbb{H}$ a normed division ring.

Let us examine (2.13.120) when $\xi = u$ and $\eta = v$ are purely vectorial. We have

$$uv = -u \cdot v + u \times v. \tag{2.13.121}$$

Hence, directly,

$$|uv|^2 = (u \cdot v)^2 + |u \times v|^2, \tag{2.13.122}$$

while (2.13.120) implies

$$|uv|^2 = |u|^2|v|^2. \tag{2.13.123}$$

Now the law of cosines gives

$$u \cdot v = |u|\,|v| \cos\theta, \tag{2.13.124}$$

where θ is the angle between u and v in the plane in $\mathbb{R}^3$ spanned by these two vectors. See (2.2.42). Hence (2.13.122) implies

$$|u \times v|^2 = |u|^2|v|^2 \sin^2\theta. \tag{2.13.125}$$

We note that the set of unit-norm elements of $\mathbb{H}$,

$$\mathrm{Sp}(1) = \{\xi \in \mathbb{H} : |\xi| = 1\} = S^3, \tag{2.13.126}$$

forms a multiplicative group. Furthermore, by (2.13.120), we have the map

$$\lambda : \mathrm{Sp}(1) \times \mathrm{Sp}(1) \to \mathrm{SO}(4), \quad \lambda(\xi, \eta)\zeta = \xi\zeta\overline{\eta}, \tag{2.13.127}$$

which is a two-to-one homomorphism of $\mathrm{Sp}(1) \times \mathrm{Sp}(1)$ onto $\mathrm{SO}(4)$, and

$$\pi : \mathrm{Sp}(1) \to \mathrm{SO}(3), \quad \pi(\xi)v = \xi v\overline{\xi}, \tag{2.13.128}$$

which is a two-to-one homomorphism of $\mathrm{Sp}(1)$ onto $\mathrm{SO}(3)$. We also have the following variant of Euler's formula. Assume $u \in \mathbb{R}^3$, $|u| = 1$, and regard $u \in \mathbb{H}$. Then, for $t \in \mathbb{R}$,

$$e^{tu} = \cos t + (\sin t)u, \quad e^{tu} \in \mathrm{Sp}(1). \tag{2.13.129}$$

Note that, with π as in (2.13.128), u as in (2.13.129), and $v \in \mathbb{R}^3$,

$$\pi(e^{tu})v = e^{tu}ve^{-tu}. \tag{2.13.130}$$

Clearly $e^{tu}ue^{-tu} \equiv u$. This yields:

Proposition 2.13.14. *With π and u as in* (2.13.128)–(2.13.129),

$$\{\pi(e^{tu}) : t \in \mathbb{R}\} \subset \mathrm{SO}(3) \tag{2.13.131}$$

is the group of rotations about the axis generated by u.

Together, the formulas (2.13.129)–(2.13.130) provide a convenient representation of such rotations. This has led to the use of quaternions as a tool in robotics.

Further results on the group Sp(1) can be found in Chapter 8 of [**14**] and in Chapter 10 of [**19**].

Exercises

1. Conic sections, mentioned below (2.12.45) as planar curves consisting of ellipses, parabolas, and hyperbolas, were treated in classical times as the locus of intersection of a 2-plane in $\mathbb{R}^3$ with the cone $C = \{(u, v, w) \in \mathbb{R}^3 : u^2 + v^2 = w^2\}$. We explore this concept here. Set

$$Q(u, v, w)^t = u^2 + v^2 - w^2.$$

Then, given $A \in \mathrm{O}(3),\ a \in \mathbb{R} \setminus 0$, set

$$q_{A,a}(x, y) = QA(x, y, a)^t.$$

A conic section is defined by $q_{A,a}(x, y) = 0$. Show that precisely the ellipses, parabolas, and hyperbolas considered in §2.12 arise as conic sections.

2. With $A_{n-1} = \mathrm{Area}(S^{n-1})$, as in (2.13.89), use (2.13.99) to show that

$$\sum_{n=0}^{\infty} A_{2n+1} x^n = 2\pi e^{\pi x}.$$

Hint. Deduce from Proposition 2.13.12 that $\Gamma(n + 1) = n!$.

3. Show that

$$\frac{A_{2n+2}}{A_{2n}} = \frac{\pi}{n + \frac{1}{2}}.$$

4. Let $\mathcal{T}$ be an equilateral triangle in the (x, y)-plane, with center at the origin, sidelength 1. Consider the following point on the z-axis of $\mathbb{R}^3$:

$$p = (0, 0, \sqrt{\tfrac{2}{3}}).$$

Show that p together with the three vertices of $\mathcal{T}$ form the four vertices of a regular tetrahedron. Compute the volume of this tetrahedron.

5. Let $Q \subset \mathbb{R}^3$ be the cube with the 8 vertices $(\pm 1, \pm 1, \pm 1)$. Note that Q has 6 faces and $\mathrm{Vol}\, Q = 8$. Let P denote the figure whose 6 vertices lie at the center of the 6 faces of Q. Show that P has 8 faces. Note that all the symmetries of Q are also symmetries of P. We call P a regular octahedron. Compute $\mathrm{Vol}\, P$.

6. Let $\xi \in \mathrm{Sp}(1)$, i.e., $\xi \in \mathbb{H},\ |\xi| = 1$. Assume $\xi \neq -1$. Show that

$$\eta = \frac{\xi + 1}{|\xi + 1|} \Longrightarrow \eta \in \mathrm{Sp}(1),\ \eta^2 = (-\eta)^2 = \xi.$$

Show that, given $\eta \in \mathrm{Sp}(1)$,

$$\eta^2 = -1 \Longleftrightarrow \eta = v \in \mathbb{R}^3,\ |v| = 1.$$

2.A. Matrices and their action

Material in Sections 2.10 and 2.13 led to the introduction of matrices, first to 2×2 complex matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{2.A.1}$$

in (2.10.1), then to $n \times n$ real matrices

$$A = (a_{jk}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \tag{2.A.2}$$

in (2.13.54). In case (2.A.1), we write $A \in M(2, \mathbb{C})$, and in case (2.A.2) we write $A \in M(n, \mathbb{R})$. Here we describe some of the properties of these matrices, alluded to in those sections.

A 2×2 matrix of the form (2.A.1) acts on a 2-vector as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} az + bw \\ cz + dw \end{pmatrix}. \tag{2.A.3}$$

If B is another 2×2 matrix, then the matrix product $C = AB$ is defined so that, for a 2-vector ξ,

$$C\xi = A(B\xi). \tag{2.A.4}$$

This leads to the matrix product

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z & u \\ w & v \end{pmatrix} = \begin{pmatrix} az + bw & au + bv \\ cz + dw & cu + dv \end{pmatrix}. \tag{2.A.5}$$

The columns of AB are given by the action of A on the columns of B.

Given $A \in M(2, \mathbb{C})$, we say $B = A^{-1}$ (the inverse of A), provided

$$AB = BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I, \tag{2.A.6}$$

the identity matrix. If such B exists, we say A is invertible, and write $A \in G\ell(2, \mathbb{C})$. In such a case, we can solve

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}, \tag{2.A.7}$$

uniquely for $\binom{z}{w}$, given $\binom{u}{v}$. This is equivalent to solving

$$\begin{aligned} az + bw &= u, \\ cz + dw &= v. \end{aligned} \tag{2.A.8}$$

This system can be treated by multiplying these equations by d and b, respectively, and subtracting, and then multiplying them by c and a, respectively, and subtracting, obtaining

$$\begin{aligned} (ad - bc)z &= du - bu, \\ (ad - bc)w &= av - cu. \end{aligned} \tag{2.A.9}$$

The factor on the left is

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{2.A.10}$$

called the determinant of A. Solving (2.A.9) for z and w leads to

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Longrightarrow A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{2.A.11}$$

provided $\det A \neq 0$. We have

$$G\ell(2, \mathbb{C}) = \{A \in M(2, \mathbb{C}) : \det A \neq 0\}. \tag{2.A.12}$$

In general, for $A, B \in M(2, \mathbb{C})$,

$$\det AB = (\det A)(\det B). \tag{2.A.13}$$

As a consequence, if

$$S\ell(2, \mathbb{C}) = \{A \in M(2, \mathbb{C}) : \det A = 1\}, \tag{2.A.14}$$

then

$$A, B \in S\ell(2, \mathbb{C}) \Longrightarrow AB, A^{-1} \in S\ell(2, \mathbb{C}). \tag{2.A.15}$$

We move to $n \times n$ matrices, $A \in M(n, \mathbb{R})$, of the form (2.A.2), with entries $a_{jk} \in \mathbb{R}$. (If we took $a_{jk} \in \mathbb{C}$, we would write $A \in M(n, \mathbb{C})$.) Such A acts on a column vector

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{2.A.16}$$

(we also write $x = (x_1, \dots, x_n)^t$), to produce $Ax = y = (y_1, \dots, y_n)^t$, via the following analogue of (2.A.7):

$$y_j = \sum_{k=1}^{n} a_{jk} x_k = a_{j1}x_1 + \cdots + a_{jn}x_n. \tag{2.A.17}$$

If also $B \in M(n, \mathbb{R})$, we define the matrix product $C = AB$ so that for $x \in \mathbb{R}^n$,

$$Cx = A(Bx). \tag{2.A.18}$$

This works out to

$$C = (c_{jk}), \quad c_{jk} = \sum_{\ell=1}^{n} a_{j\ell} b_{\ell k}. \tag{2.A.19}$$

The columns of C are given by the action of A on the columns of B.

Multiplication in $M(n, \mathbb{R})$ is associative, i.e.,

$$A_j \in M(n, \mathbb{R}) \Longrightarrow A_1(A_2A_3) = (A_1A_2)A_3, \tag{2.A.20}$$

but one does not necessarily have

$$A_1A_2 = A_2A_1. \tag{2.A.21}$$

If (2.A.21) holds, we say the matrices A_1 and A_2 commute. One matrix that commutes with all elements of $M(n, \mathbb{R})$ is the identity matrix,

$$I = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}, \tag{2.A.22}$$

whose entries δ_{jk} are equal to 1 for $j = k$ and to 0 for $j \neq k$. Indeed,

$$AI = IA = A, \tag{2.A.23}$$

for all $A \in M(n, \mathbb{R})$.

Parallel to (2.A.6), it is of interest to consider when $A \in M(n, \mathbb{R})$ has an inverse, i.e., a matrix $B = A^{-1}$, satisfying $AB = BA = I$. When such B exists (we say A is invertible and write $A \in G\ell(n, \mathbb{R})$) it is unique. Indeed, for $B_j \in M(n, \mathbb{R})$,

$$B_1A = AB_2 = I \Rightarrow B_1 = B_1(AB_2) = (B_1A)B_2 = B_2, \tag{2.A.24}$$

the second identity on the right by associativity, (2.A.20). Matrices B_1 and B_2 satisfying the hypothesis of (2.A.24) are said to be a left inverse of A and a right inverse of A. It is a fact that a matrix $A \in M(n, \mathbb{R})$ has a left inverse if and only if it has a right inverse. A proof can be found in §1.3 of [**14**]. Again there is a determinant function,

$$\det : M(n, \mathbb{R}) \longrightarrow \mathbb{R}, \tag{2.A.25}$$

satisfying (2.A.13), and such that $A \in M(n, \mathbb{R})$ is invertible if and only if $\det A \neq 0$, so

$$G\ell(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) : \det A \neq 0\}. \tag{2.A.26}$$

A defining characterization of the determinant is given in Proposition 2.13.6, and details can be found in §1.5 of [**14**]. We mention that there is a formula for $\det A$, extending (2.A.10), involving an alternating sum of n-fold products of matrix entries of A, and a formula for A^{-1} (somewhat more elaborate than (2.A.11)), given in §1.5 of [**14**].

Bibliography

[1] M. al-Khwarizmi, *The Algebra of Mohamed Ben Musa (Al-Jabr)*, Edited and translated by F. Rosen, Univ. Press of the Pacific, Stockton CA, 2003.

[2] A. Beardon, *Algebra and Geometry*, Cambridge Univ. Press, Cambridge UK, 2005.

[3] P. Beckmann, *A History of* π, The Golem Press, Boulder CO, 1971.

[4] J. Conway and D. Smith, *On Quaternions and Octonions: Their Geometry, Arithmetic, and Symmetry*, A. K. Peters, Natik MA, 2003.

[5] R. Descartes, *The Geometry*, Translated by D. Smith and M. Latham, Dover NY, 1954.

[6] Euclid, *The Thirteen Books of the Elements*, Vols. 1–3, Translated and with commentary by T. L. Heath, Dover NY, 2007.

[7] L. Gårding, *Encounter with Mathematics*, Springer-Verlag, New York, 1977.

[8] M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford Univ. Press, Oxford UK, 1972.

[9] S. Lang, *Algebra*, Addison-Wesley, Reading MA, 1965.

[10] M. Livio, *The Golden Ratio*, Broadway Books, New York, 2002.

[11] E. Maor, *The Pythagorean Theorem – A 4,000-Year History*, Princeton Univ. Press, Princeton NJ, 2007.

[12] P. Nahin, *An Imaginary Tale – The Story of* $\sqrt{-1}$, Princeton Univ. Press, Princeton NJ, 1998.

[13] G. Phillips, *Two Millenia of Mathematics – From Archimedes to Gauss*, Springer, New York, 2000.

[14] M. Taylor, *Linear Algebra*, Undergraduate Texts #45, Amer. Math. Soc., Providence RI, 2020.

[15] M. Taylor, *Introduction to Analysis in One Variable*, Undergraduate Texts #47, Amer. Math. Soc., Providence RI, 2020.

[16] M. Taylor, *Introduction to Analysis in Several Variables*, Undergraduate Texts #46, Amer. Math. Soc., Providence RI, 2020.

[17] M. Taylor, *Introduction to Complex Analysis*, GSM #202, Amer. Math. Soc., Providence RI, 2019.

[18] M. Taylor, *Measure Theory and Integration*, GSM #76, Amer. Math. Soc., Providence RI, 2006.

[19] M. Taylor, *Introduction to Lie Groups*, Open Math Notes, Amer. Math. Soc., Providence RI, 2021.

[20] M. Taylor, *The Pentagon, the Dodecahedron, and the Icosahedron*, Notes, available at https://mtaylor.web.unc.edu/notes, Item 18, Elementary geometry notes.

[21] M. Taylor, *Introduction to Differential Equations*, Undergraduate Texts #52, Amer. Math. Soc., Providence RI, 2011 (2nd ed., 2020).

[22] M. Taylor, *Partial Differential Equations*, Vols. 1–3, Springer, New York, 1996 (3rd ed., 2023).

[23] O. Toeplitz, *The Calculus – A Genetic Approach*, Univ. of Chicago Press, Chicago, 1963.

Index