

And to what else do the point pairs $\pm(a, b, c, d)$ correspond? Well, remember from Section 5.4 how we described the real projective space \mathbb{RP}^3 . Its “points” are lines through O in \mathbb{R}^4 . But a line through O in \mathbb{R}^4 meets \mathbb{S}^3 in a pair of antipodal points $\pm(a, b, c, d)$. Thus, *it is also valid to view the point pairs $\pm(a, b, c, d)$ on \mathbb{S}^3 as single “points” of \mathbb{RP}^3* . Hence, rotations of \mathbb{S}^2 correspond to points of \mathbb{RP}^3 , and the group of all rotations of \mathbb{S}^2 is in some sense the “same” as the geometric object \mathbb{RP}^3 .

To explain what we mean by “sameness” here, we have to say something about groups in general, what it means for two groups to be the “same”, and what it means for a geometric object to acquire the structure of a group.

Abstract groups and isomorphisms

We began this chapter with the idea of a *group of transformations*: a collection G of functions on a space S with the properties that

- if $f, g \in G$, then $fg \in G$,
- if $f \in G$, then $f^{-1} \in G$.

The “product” fg of f and g here is the *composite function*, which is defined by $fg(x) = f(g(x))$. However, we have found it convenient to *represent* certain functions, such as rotations, by algebraic objects, such as matrices, whose “product” is defined algebraically.

It is therefore desirable to have a more general concept of group, which does not presuppose that the product operation is function composition. We define an *abstract group* to be a set G , which contains a special element 1 and for each g an element g^{-1} , with a “product” operation satisfying the following axioms:

$$g_1(g_2g_3) = (g_1g_2)g_3 \quad (\text{associativity})$$

$$g1 = g \quad (\text{identity})$$

$$gg^{-1} = 1 \quad (\text{inverse})$$

The associative axiom is automatically satisfied for function composition, because if g_1, g_2, g_3 are functions, then $g_1(g_2g_3)$ and $(g_1g_2)g_3$ both mean the same thing, namely, the function $g_1(g_2(g_3(x)))$. It is also satisfied when the group consists of numbers, because the product of numbers is well known to be associative.

In other cases, the easiest way to prove associativity is to show, if possible, that the group operation corresponds to function composition. For example, the matrix product operation is associative because matrices behave like linear transformations under composition. It follows in turn that the quaternion product operation is associative, because quaternions can be viewed as matrices.

When we say that the elements of a certain group G “correspond to” or “behave like” or “can be viewed as” elements of another group G' , we have in mind a precise relationship called an *isomorphism* of G onto G' . The word “isomorphism” comes from the Greek for “same form,” and it means that *there is a one-to-one correspondence between G and G' that preserves products*. That is, an isomorphism is a function

$$\varphi : G \rightarrow G' \quad \text{such that} \quad \varphi(g_1g_2) = \varphi(g_1)\varphi(g_2).$$

For example, the group G of rotations of the circle, under composition of isometries, is isomorphic to group G' of complex numbers of the form

$$\cos \theta + i \sin \theta, \quad \text{under multiplication.}$$

If r_θ denotes the rotation through angle θ , then the isomorphism φ is defined by

$$\varphi(r_\theta) = \cos \theta + i \sin \theta.$$

Sometimes there is a natural one-to-one correspondence φ between a group G and a set S . In that case, we can use φ to *transfer the group structure from G to S* . That is, we *define* the product of elements $\varphi(g_1)$ and $\varphi(g_2)$ to be $\varphi(g_1g_2)$. Here are some examples.

- The complex numbers $\cos \theta + i \sin \theta$ form a group, and they correspond to the points $(\cos \theta, \sin \theta)$ of the unit circle \mathbb{S}^1 . Therefore, we can define the product of points $(\cos \theta_1, \sin \theta_1)$ and $(\cos \theta_2, \sin \theta_2)$ on \mathbb{S}^1 to be the point corresponding to the product of the corresponding complex numbers. This point is $(\cos(\theta_1 + \theta_2), \sin(\theta_1 + \theta_2))$.
- Likewise, the quaternions $\mathbf{q} = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ with $|\mathbf{q}| = 1$ form a group, and they correspond to the points (a, b, c, d) of the 3-sphere \mathbb{S}^3 . Hence, we can define the product of points (a_1, b_1, c_1, d_1) and (a_2, b_2, c_2, d_2) corresponding to the quaternions \mathbf{q}_1 and \mathbf{q}_2 , say, to be the point corresponding to the quaternion $\mathbf{q}_1\mathbf{q}_2$. Under this product operation, \mathbb{S}^3 is a group.

- Finally, pairs of opposite quaternions $\pm \mathbf{q}$ with $|\mathbf{q}| = 1$ form a group under the operation defined by

$$(\pm \mathbf{q}_1)(\pm \mathbf{q}_2) = \pm \mathbf{q}_1 \mathbf{q}_2.$$

We know that these pairs are in one-to-one correspondence with the points of \mathbb{RP}^3 . Hence, we can transfer the group structure of these quaternion pairs (which is also the group structure of the rotations of \mathbb{S}^2) to \mathbb{RP}^3 . Under the transferred operation, \mathbb{RP}^3 is a group.

The group structures on \mathbb{S}^1 , \mathbb{S}^3 , and \mathbb{RP}^3 obtained in this way are particularly interesting because they are *continuous*. That is, if g'_1 is near to g_1 and g'_2 is near to g_2 , then $g'_1 g'_2$ is near to $g_1 g_2$. It is known that \mathbb{S}^2 does not have a continuous group structure, and in fact \mathbb{S}^1 and \mathbb{S}^3 are the *only* spheres with continuous group structures on them.

7.9 Discussion

The word “geometry” comes from the Greek for “earth measurement,” and legend has it that the subject grew from the land measurement concerns of farmers whose land was periodically flooded by the river Nile. As recently as the 18th century, one finds carpenters and other artisans listed among the subscribers to geometry books, so there is no doubt that Euclidean geometry is the geometry of down-to-earth measurement. It continues to be a *tactile* subject today, when one talks about “translating,” “rotating,” and “moving objects rigidly.”

The most *visual* branch of geometry is projective geometry, because it is more concerned with how objects *look* than with what they actually are. It is no surprise that projective geometry originated from the concerns of artists, and that many of its practitioners today work in the fields of video games and computer graphics.

Affine geometry occupies a position in the middle. It also originates from an artistic tradition, but from one less radical than that of Renaissance Italy—the classical art of China and Japan. Chinese and Japanese drawings often adopt unusual viewpoints, where one might expect perspective, but they generally preserve parallels. Typically, the picture is a “projection from infinity,” which is an affine map. Figure 7.9 shows an example, a woodblock print by the Japanese artist Suzuki Harunobu from around 1760.



Figure 7.9: Harunobu's *Courtesan on veranda*

Notice that all the parallel lines are shown as parallel, with the result that the (obviously rectangular) panels on the screen appear as identical parallelograms. Likewise, the planks on the veranda appear with parallel edges and equal widths, which creates a certain “flatness” because all parts of the picture seem to be the same distance away from us. Speaking mathematically, they are—because the view is what one would see from infinity with infinite magnification. A similar effect occurs in photographs of distant buildings taken with a large amount of zoom.

Affine maps are also popular in engineering drawing, in which the so-called “axonometric projection” is often used to depict an object in three dimensions while retaining correct proportions in a given direction. The

affine picture gives a good compromise between a realistic view and an accurate plan. See Figure 7.10, which shows an axonometric projection of a cube.

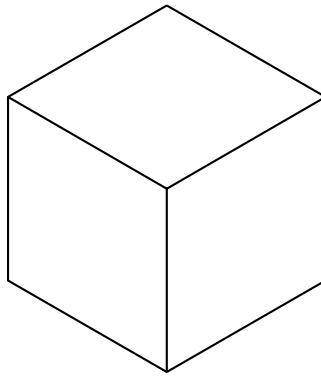


Figure 7.10: Affine view of the cube

The fourth dimension

The discovery of quaternions in 1843 was the first of a series of discoveries that drew attention to spaces of more than three dimensions and to the remarkable properties of \mathbb{R}^4 in particular.

From around 1830, the Irish mathematician William Rowan Hamilton had been searching in vain for “ n -dimensional number systems” analogous to the real numbers \mathbb{R} and the complex numbers \mathbb{C} . Because \mathbb{C} can be viewed as \mathbb{R}^2 under vector addition

$$(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2)$$

and the multiplication operation

$$(u_1, u_2)(v_1, v_2) = (u_1 v_1 - u_2 v_2, u_1 v_2 + u_2 v_1),$$

Hamilton thought that \mathbb{R}^3 could also be viewed as a number system by some clever choice of multiplication rule. He took a “number system” to be what we now call a field, together with an absolute value

$$|\mathbf{u}| = |(u_1, u_2, u_3)| = \sqrt{u_1^2 + u_2^2 + u_3^2}$$

which is multiplicative: $|\mathbf{uv}| = |\mathbf{u}||\mathbf{v}|$.

We now know (for example, by Exercise 7.6.7) that such a system is impossible in any \mathbb{R}^n with $n \geq 3$. But, luckily for Hamilton, it is *almost* possible in \mathbb{R}^4 . The quaternions satisfy all the field axioms except commutative multiplication, and their absolute value is multiplicative. The only other \mathbb{R}^n that comes close is \mathbb{R}^8 , where the octonions \mathbb{O} satisfy all the field properties except the commutative and associative laws. (Recall from Section 6.8 that the quaternions and octonions also play an important role in projective geometry.)

Hamilton knew that quaternions give a nice representation of rotations in \mathbb{R}^3 , but the first to work out the quaternions for the symmetries of regular polyhedra was Cayley in 1863. Cayley's enumeration of these quaternions may be found in his *Mathematical Papers*, volume 5, p. 529. The five regular polyhedra actually exhibit only three types of symmetry—because the cube and octahedron have the same symmetry type, as do the dodecahedron and the icosahedron—which therefore correspond to three highly symmetric sets of quadruples in \mathbb{R}^4 .

Cayley did not investigate the geometric properties of these point sets in \mathbb{R}^4 , but in fact they were already known to the Swiss geometer Ludwig Schläfli in 1852. As we have seen, the 12 rotations of the tetrahedron correspond to the 24 vertices of a figure called the 24-cell. It gets this name because it is bounded by 24 identical regular octahedra. It is one of six regular figures in \mathbb{R}^4 , analogous to the regular polyhedra in \mathbb{R}^3 , called the *regular polytopes*. They were discovered by Schläfli, who also proved that there are regular figures analogous to the tetrahedron, cube, and octahedron in each \mathbb{R}^n , but that \mathbb{R}^3 and \mathbb{R}^4 are the only \mathbb{R}^n containing other regular figures.

The 24-cell is the simplest of the exceptional regular figures in \mathbb{R}^4 ; the other two are the *120-cell* (bounded by 120 regular dodecahedra) and the *600-cell* (bounded by 600 regular tetrahedra). The 120-cell has 600 vertices, which correspond to the cell centers of the 600-cell, and vice versa, so the two are related “dually” like the dodecahedron and the icosahedron.

Moreover, the 600-cell arises from the icosahedron in the same way that the 24-cell arises from the tetrahedron. Its 120 vertices correspond to 60 pairs of opposite quaternions, each representing a rotational symmetry of the icosahedron. For more on these amazing objects, see my article *The story of the 120-cell*, which can be read online at

<http://www.ams.org/notices/200201/fea-stillwell.pdf>

8

Non-Euclidean geometry

PREVIEW

In previous chapters, we have seen several reasons why there is such a subject as “foundations of geometry.” Geometry is fundamentally visual; yet it can be communicated by nonvisual means: by logic, linear algebra, or group theory, for example. The several ways to communicate geometry give several foundations.

But also, *there is more than one geometry*. Section 7.4 gave a hint of this when we briefly discussed the geometry of the sphere in the language of “points” and “lines.” It seems reasonable to call great circles “lines” because they are the straightest curves on the sphere; but they certainly do not have all of the properties of Euclid’s lines.

This characteristic makes geometry on a sphere a *non-Euclidean* geometry—one that has been known since ancient times. But it was never seen as a challenge to Euclid, probably because the geometry of the sphere is simply a part of three-dimensional Euclidean geometry, where great circles coexist with genuine straight lines.

The real challenge to Euclid emerged from disquiet over the parallel axiom. Many people found it inelegant and wished that it was a consequence of Euclid’s other axioms. It is not, because *there is a geometry that satisfies all of Euclid’s axioms except the parallel axiom*. This is the geometry of a surface called the *non-Euclidean plane*.

The non-Euclidean plane is *not* an artificial construct built only to show that the parallel axiom cannot be proved. It arises in many places, and today one can hardly discuss differential geometry, the theory of complex numbers, and projective geometry without it. In this chapter, we will see how it arises from the real projective line.

8.1 Extending the projective line to a plane

In this book, we have been concerned mainly with the geometry of lines, partly because lines are the foundation of geometry and partly because lines are remarkably interesting. From the regular polygons to the Pappus and Desargues configurations, figures built from lines reveal beautiful connections between geometry and other parts of mathematics. We have seen some of these connections but have barely begun to explore them in depth.

In fact we have not yet gone far toward understanding the geometry of even *one* line—the real projective line \mathbb{RP}^1 . In Chapter 5, we arrived at an algebraic summary of \mathbb{RP}^1 by representing its transformations as linear fractional functions

$$f(x) = \frac{ax+b}{cx+d}, \quad \text{where } a, b, c, d \in \mathbb{R} \quad \text{and} \quad ad - bc \neq 0, \quad (*)$$

and by uncovering the *cross-ratio*, a “ratio of ratios” left invariant by all linear fractional transformations. But this summary is not as geometric as one would like. It is hard (although perhaps not impossible) to “see” the cross-ratio, and indeed it is hard to see geometric phenomena on the line at all. If only we could extend the projective line in another dimension so that we could see it as a plane! Amazingly, this is possible, and the present chapter shows how.

The idea is to let \mathbb{RP}^1 be the boundary (“at infinity”) of a plane whose transformations extend the linear fractional transformations of \mathbb{RP}^1 in a natural way. Algebra suggests how this should be done. It suggests replacing the real variable x in the linear fractional transformations (*) by a complex variable z and interpreting

$$f(z) = \frac{az+b}{cz+d}$$

as a transformation of the plane \mathbb{C} of complex numbers.

This idea needs a little modification. We should really use transformations of the *upper half plane* of complex numbers $z = x + iy$ with $y > 0$, because the line of real numbers divides \mathbb{C} into two halves. Either half can be taken as the “plane” bounded by the real line, but, when we want to transform one particular half, the extension from x to z is not always the obvious one. For example, the transformation $x \mapsto -x$ of the line should *not* be extended to the transformation $z \mapsto -z$, because the latter maps the

upper half plane onto the lower. The correct extension is described in Section 8.2.

Nevertheless, the extension from real to complex numbers works amazingly well. The extended transformations leave invariant a geometric quantity that is clearly visible, namely *angle*. The cross-ratio is also invariant (well, almost), for much the same algebraic reasons as before. And it gives us an invariant *length* in the half plane—something we certainly do not have in the projective line \mathbb{RP}^1 .

The concept of length that emerges in this way is a little subtle, and it is not as easily visible as angle. It is a *non-Euclidean* measure of length, and it gives rise to *non-Euclidean lines*, which turn out to be the ordinary lines of the form $x = \text{constant}$ and the semicircles with their centers on the x -axis. These “lines” are the curves of shortest non-Euclidean length between given points, and they have all the properties of “lines” in Euclid’s geometry *except the parallel property*. That is, if \mathcal{L} is any non-Euclidean line and P is a point of the upper half plane outside \mathcal{L} then there is more than one non-Euclidean line through P that does not meet \mathcal{L} . Figure 8.1 shows an example.

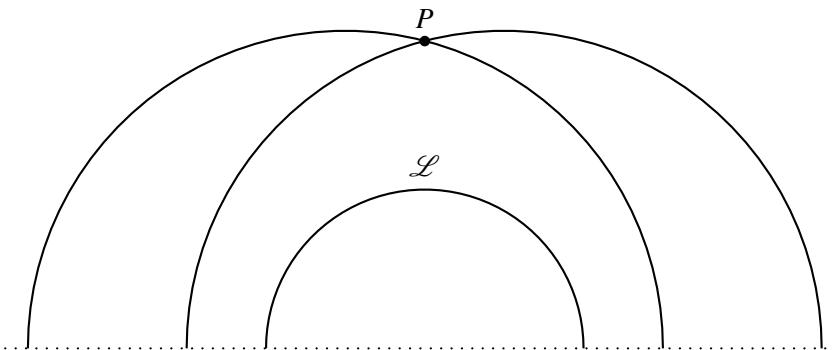


Figure 8.1: Failure of the parallel axiom for non-Euclidean “lines”

The dotted line represents the real line $y = 0$ so the half plane $y > 0$ consists of the points strictly above it. The semicircle \mathcal{L} is one “line” in the half plane, and the two semicircles passing through the point P clearly do not meet \mathcal{L} , so they are two “parallels” of \mathcal{L} . In the remainder of this chapter we explain in more detail why these semicircles should be regarded as “lines,” and why they satisfy all of Euclid’s axioms for lines except the parallel axiom.

Thus, the complex half plane not only allows us to visualize the geometry of the projective line; it also answers a fundamental question in the foundations of geometry by showing that *the parallel axiom does not follow from Euclid's other axioms*.

The non-Euclidean “line” through two points

One property of non-Euclidean “lines” can be established immediately. *There is a unique non-Euclidean “line” through any two points* (and hence non-Euclidean “lines” satisfy the first of Euclid’s axioms).

- If the two points lie on the same vertical line $x = l$, then $x = l$ is a non-Euclidean “line” containing them. And it is the only one, because a semicircle with its endpoints on the x -axis has at most one point on each line $x = l$.
- If the two points P and Q do not lie on the same vertical line, there is a unique point R on the x -axis equidistant for both of them, namely, where the equidistant line of P and Q meets the x -axis. Then the semicircle with center R through P and Q is the unique non-Euclidean “line” through P and Q .

Exercises

One can begin to understand the geometric significance of linear fractional transformations of the half plane by studying the simplest ones, $z \mapsto z + l$ and $z \mapsto kz$ for real k and l .

- 8.1.1** Show that the transformations $z \mapsto z + l$ and $z \mapsto kz$ (for $k > 0$) map the upper half plane onto itself and that they map “lines” to “lines.”
- 8.1.2** Explain why this is *not* the case when k and l are not real.
- 8.1.3** Show how to map the semicircle $x^2 + y^2 = 1$, $y > 0$, onto the semicircle $(x - 1)^2 + y^2 = 4$, $y > 0$, by a combination of transformations $z \mapsto z + l$ and $z \mapsto kz$.
- 8.1.4** More generally, explain why any semicircle with center on the x -axis can be mapped onto any other by a combination of transformations $z \mapsto z + l$ and $z \mapsto kz$.

What is not yet clear is why semicircles should be regarded as “lines.” Their “linelike” behavior stems from the transformation $z \mapsto 1/\bar{z}$, which we study in Section 8.2.

8.2 Complex conjugation

We know from Section 5.6 that all linear fractional transformations of \mathbb{RP}^1 are products of the transformations $x \mapsto x + l$, $x \mapsto kx$, and $x \mapsto 1/x$ for real constants $k \neq 0$ and l . We called these the *generating transformations* of \mathbb{RP}^1 . The transformation $x \mapsto x + l$ obviously extends to the transformation $z \mapsto z + l$, which maps the upper half plane onto itself for any real l , but the appropriate extension of $x \mapsto kx$ is $z \mapsto kz$ only for $k > 0$, because $z \mapsto kz$ does not map the upper half plane onto itself when $k < 0$. In particular, *what is the appropriate extension of $x \mapsto -x$ to a map of the upper half plane?*

Geometrically, the answer is obvious. The transformation $x \mapsto -x$ is reflection of the line in O , so its most appropriate extension is *reflection of the half plane in the y-axis*, that is, the transformation

$$x + iy \mapsto -x + iy.$$

This transformation can be expressed more simply with the help of the *complex conjugate* \bar{z} , which is defined as follows. If $z = x + iy$, then $\bar{z} = x - iy$. Then the reflection of z in the y-axis is $-\bar{z}$ because

$$-\bar{z} = -\overline{(x + iy)} = -(x - iy) = -x + iy.$$

Thus, the appropriate extension of $x \mapsto -x$ is $z \mapsto -\bar{z}$ (Figure 8.2).

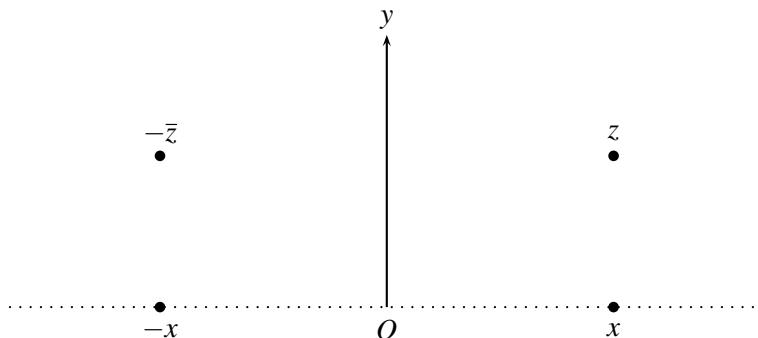


Figure 8.2: Extending reflection from the line to the half plane

More generally, the appropriate extension of $x \mapsto kx$ when $k < 0$ is $z \mapsto k\bar{z}$, the product of the reflection $z \mapsto -\bar{z}$ with the map $z \mapsto |k|z$ (dilation by factor $|k|$).