the functions $\frac{az+b}{\bar{b}z+\bar{a}}$, differing from those of the sphere only by one $-$ sign. This is just one of many ways in which the non-Euclidean plane is "opposite" to the sphere. It has many parallel lines and the sphere has none, its triangles have angle sum $< \pi$ and those of the sphere have angle sum $> \pi$, and so on.

Poincaré (1883) generalized the half plane model to a "half space model" of *non-Euclidean space* by considering the upper half of three-dimensional space, the half above the (x, y)-plane say. He defined the isometries of this space to be products of reflections in spheres with centers on the (x, y)-plane. Reflection in a sphere is defined analogously to reflection in a circle (Section 3.9*) and gives a geometry of the half space analogous to the non-Euclidean geometry of the half plane.

Despite the third dimension, the isometries of non-Euclidean space can be represented by functions of the complex variable $z = x + iy$. This is because the spheres of reflection are determined by the circles in which they cut the (x, y)-plane, so reflections in the latter circles determine the reflections in the spheres above them. Products of these reflections in circles then determine all isometries of the half space, and in this way Poincaré found that the orientation-preserving isometries of non-Euclidean space correspond to all the complex functions $\frac{az+b}{cz+d}$ with $ad - bc \neq 0$.

Notice that in all cases the isometries are represented by *linear fractional* functions of z, that is, quotients of linear functions $az + b$ and $cz + d$.

Poincaré's space explains why isometries are linear fractional functions in all three geometries of surfaces—Euclidean plane, sphere, and non-Euclidean plane—*these surfaces inherit their isometries from non-Euclidean space.* All three geometries actually occur in non-Euclidean space: Euclidean planes as planes parallel to the (x, y)-plane, spheres as ordinary spheres lying completely in the upper half space, and non-Euclidean planes as vertical half planes and hemispheres with their centers on the (x, y)-plane. This amazing unification of geometry was discovered by Eugenio Beltrami (1868), though it was Poincaré (1883) who first linked the geometries by linear fractional functions. For more details, see Stillwell (1992) or the papers of Beltrami and Poincaré in Stillwell (1996).

Quadratic Forms

The story of sums of squares, which started another thread in the history of complex numbers, also took a turn toward non-Euclidean geometry in the 19th century. To see how this came about, we have to say more about the work of Fermat and Lagrange.

Fermat's two squares theorem, describing the primes of the form x^2+y^2, was the first of several such theorems. Fermat also described the primes of the form $x^2 + 2y^2$ (they are the primes $p \equiv 1$ or $3 \pmod 8$) and the form $x^2 + 3y^2$ (they are the primes $p \equiv 1 \pmod 3$). However, he failed to find any such description of the primes of the form $x^2 + 5y^2$. The reasons for this did not become completely clear for another two centuries, but Lagrange made important progress in 1773.

Lagrange decided to develop a general theory of *binary quadratic forms*, that is, functions of the form $ax^2 + bxy + cy^2$, where a, b, c are integer constants and x, y are integer variables. Problems of Fermat's type then fall under the general problem of finding the possible values of a binary quadratic form, and in particular, finding the possible prime values. Lagrange noticed that many forms are *equivalent* in the sense that they are related by a change of variables. For example, if we substitute

$$\begin{aligned} x &= x' + y', \\ y &= y' \end{aligned}$$

in $x^2 + y^2$ we get the form $x'^2 + 2x'y' + 2y'^2$, which takes exactly the same values. Why? Because as x' and y' run through all pairs of integers, so do $x = x' + y'$ and $y = y'$. Forms related by such a change of variables are called *equivalent* because they have the same sets of values.

It is not hard to work out that the transformations

$$\begin{aligned} x &= ax' + by', \\ y &= cx' + dy' \end{aligned}$$

relating equivalent forms are those for which a, b, c, d are integers and $ad - bc = \pm 1$. Such transformations are now called *unimodular*. Lagrange used them to find the "simplest" form in a class of equiva-

lent forms, and in this way found more efficient proofs of Fermat's theorems and many others.

The story turned geometric when Gauss noticed that the equivalents of a given form could be viewed as points in the upper half plane, related by functions

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

corresponding to the unimodular transformations. Because f is a linear fractional function with real coefficients, it is an isometry if the half plane is interpreted as the non-Euclidean plane. Gauss didn't realize what kind of geometry he was looking at here (though in fact he had speculated about non-Euclidean geometry in an abstract way), but Poincaré did, and he used geometric insights to help understand quadratic forms.

Poincaré realized, in fact, that the real problem was to understand unimodular transformations, which form a nonabelian group. This was the first nonabelian group encountered in number theory, and the first time the group concept was used to make a bridge from number theory to geometry, where the problem could be more easily understood.

Quadratic Integers and Lattices

In this chapter we have given several impressive results that follow from unique prime factorization in the rings of quadratic integers $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$. Nevertheless, some readers may feel that unique prime factorization is a trivial property, which doesn't deserve the credit for Fermat's two squares theorem (say) or for showing that there is only one positive integer solution of $y^3 = x^2 + 2$.

In fact, unique prime factorization cannot be taken for granted, because it is sometimes *false*, and it is worth taking a closer look at the conditions that make it possible.

The proofs of unique prime factorization in $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$ depend on finding a division property, like the one for $\mathbb{Z}$, which depends in turn on the "shape" of $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$ in the plane $\mathbb{C}$. For example, to establish the division property of $\mathbb{Z}[i]$ we used the

fact that its members lie at the corners of a square grid, and hence so do the Gaussian integer multiples of any number $\beta \neq 0$.

It is not clear that this process will always work, and it will certainly fail for $\mathbb{Z}[\sqrt{-5}]$, because $\mathbb{Z}[\sqrt{-5}]$ does *not* have unique prime factorization. An example that shows this is

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

The factors 2,3 and $1 + \sqrt{-5}$, $1 - \sqrt{-5}$ are all primes in $\mathbb{Z}[\sqrt{-5}]$, as can be seen most easily by using the norm

$$N(a + b\sqrt{-5}) = |a + b\sqrt{-5}|^2 = a^2 + 5b^2.$$

The norms of 2, 3, $1 + \sqrt{-5}$, $1 - \sqrt{-5}$ are 2^2, 3^2, 6, 6, respectively, and the proper divisors 2 and 3 of these norms are *not* norms of any numbers in $\mathbb{Z}[\sqrt{-5}]$. Hence 2, 3, $1 + \sqrt{-5}$, $1 - \sqrt{-5}$ have no proper divisors, and therefore they are primes.

The failure of unique prime factorization in $\mathbb{Z}[\sqrt{-5}]$ can also be explained by a geometric property, which neatly distinguishes $\mathbb{Z}[\sqrt{-5}]$ from $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$. All of these rings are abelian groups under addition of complex numbers, and they and their subgroups are called *lattices* (because that is what they look like if their neighboring members are joined by lines; look again at Figure 7.4).

Prime factorization is related to certain subgroups called *ideals*. An ideal I in a ring R is a subgroup with the additional property that

$$(\text{a member of } I) \times (\text{a member of } R) = (\text{a member of } I).$$

It turns out that each ideal in $\mathbb{Z}[i]$ is of a specially simple type called *principal*; it consists of all the multiples of some nonzero member β. The same is true of ideals in $\mathbb{Z}[\sqrt{-2}]$. In fact, this explains algebraically why prime factorization is unique in $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$, because unique prime factorization is true of any ring in which all ideals are principal (the proof is basically the proof of the prime divisor property in Section 1.6). It follows that in a ring such as $\mathbb{Z}[\sqrt{-5}]$, where prime factorization is not unique, there will be ideals that are not principal, and hence not the same shape as the ring itself.

This allows us to "see" the failure of unique prime factorization in $\mathbb{Z}[\sqrt{-5}]$, in the shape of a nonprincipal ideal. Figure 7.7 shows $\mathbb{Z}[\sqrt{-5}]$, with stars marking the members of the ideal consisting of sums of multiples of 2 and $1 + \sqrt{-5}$. It is clear that the lattice of

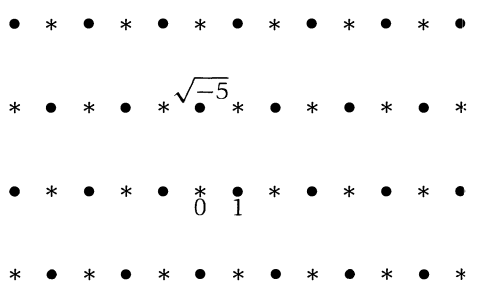

FIGURE 7.7 A nonprincipal ideal in $\mathbb{Z}[\sqrt{-5}]$.

stars is not rectangular, whereas $\mathbb{Z}[\sqrt{-5}]$ itself is. Hence the lattice is a nonprincipal ideal.

These examples give only a glimpse of the fascinating structure behind unique prime factorization in the quadratic integers. Readers are urged to consult Artin (1991) for details and Dedekind (1877) for the history of the subject. However, I cannot resist making one more tantalizing remark, because it unites the current train of thought with the previous one. Fermat's trouble with $x^2 + 5y^2$ is due to the failure of unique prime factorization in $\mathbb{Z}[\sqrt{-5}]$ and in fact *quadratic forms and quadratic integers are really the same subject*. Both depend on the study of lattice shapes, and lattice shapes are most naturally located in the non-Euclidean plane. Thus all roads lead to non-Euclidean geometry!