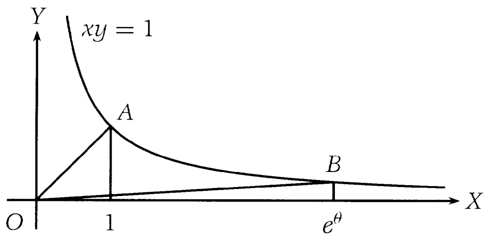

FIGURE 9.9 Areas associated with the hyperbola.

consider the hyperbola $xy = 1$ and the region under it from $x = 1$ to $x = e^\theta$ (Figure 9.9). This region has area θ. (Why?)

9.5.3. Show by geometry that the region bounded by the curve and the lines OA and OB also has area θ.

Now we use the result of exercise 8.3.3 that rotation of $xy = 1$ through angle $\pi/4$ gives the curve $x^2 - y^2 = 2$.

9.5.4. Show that this rotation sends the point $(e^\theta, e^{-\theta})$ on $xy = 1$ to the point $(\sqrt{2}\cosh\theta, -\sqrt{2}\sinh\theta)$ on $x^2 - y^2 = 2$.

9.5.5. By scaling down the curve $x^2 - y^2 = 2$, deduce that the arc of the hyperbola $x^2 - y^2 = 1$ between $(0, 1)$ and $(\cosh\theta, \sinh\theta)$, together with the lines connecting its endpoints to O, bounds a sector of area $\theta/2$.

9.6 The Pell Equation Revisited

The parametric equations $x = \cosh\theta$, $y = \sinh\theta$ for $x^2 - y^2 = 1$ generalize easily to other hyperbolas. The most interesting, from the perspective of this book, are the hyperbolas given by the Pell equation

$$x^2 - dy^2 = 1, \quad \text{where } d \text{ is a nonsquare positive integer.}$$

This equation is satisfied by

$$x = \cosh\theta, \quad y = \frac{1}{\sqrt{d}}\sinh\theta,$$

and in fact the latter equations define a one-to-one correspondence between the real numbers and the points on the positive branch of $x^2 - dy^2 = 1$, as one sees by checking the range and sign of $\cosh\theta$ and $\frac{1}{\sqrt{d}}\sinh\theta$.

Generalizing another idea from the previous section, we let

$$P_\theta = \left(\cosh\theta, \frac{1}{\sqrt{d}}\sinh\theta\right)$$

be the point with parameter value θ, and we define the *sum of points* P_θ and P_ϕ by

$$P_\theta + P_\phi = P_{\theta+\phi}.$$

We can similarly define the difference of points by $P_\theta - P_\phi = P_{\theta-\phi}$.

Now we can work out a rule for computing $P_{\theta+\phi}$ from P_θ and P_ϕ, using the addition formulas for cosh and sinh:

$$\begin{aligned}\cosh(\theta+\phi) &= \cosh\theta\cosh\phi + \sinh\theta\sinh\phi,\\ \sinh(\theta+\phi) &= \sinh\theta\cosh\phi + \cosh\theta\sinh\phi.\end{aligned}$$

As mentioned in the previous exercise set, these addition formulas are very similar to those for cos and sin, but they are easier to prove, because it is only necessary to expand each side in terms of exponentials. It follows from the addition formulas that

$$\begin{aligned}P_{\theta+\phi} &= \left(\cosh(\theta+\phi), \frac{1}{\sqrt{d}}\sinh(\theta+\phi)\right)\\ &= \Big(\cosh\theta\cosh\phi + \sinh\theta\sinh\phi,\\ &\qquad \frac{1}{\sqrt{d}}(\sinh\theta\cosh\phi + \cosh\theta\sinh\phi)\Big).\end{aligned}$$

This gives a rule to compute the coordinates of $P_{\theta+\phi}$ from the coordinates of P_θ and P_ϕ. The rule can be stated concisely as follows.

Rule for adding points on $x^2 - dy^2 = 1$.
If

$$P_\theta = \left(\cosh\theta, \frac{1}{\sqrt{d}}\sinh\theta\right) = (x_\theta, y_\theta)$$

and

$$P_\phi = \left(\cosh\phi, \frac{1}{\sqrt{d}}\sinh\phi\right) = (x_\phi, y_\phi),$$

then

$$P_{\theta+\phi} = (x_{\theta+\phi}, y_{\theta+\phi}),$$

where

$$x_{\theta+\phi} = x_\theta x_\phi + dy_\theta y_\phi \quad \text{and} \quad y_{\theta+\phi} = x_\theta y_\phi + y_\theta x_\phi. \qquad \square$$

In the special case where $P_\theta = (x_\theta, y_\theta)$ and $P_\phi = (x_\phi, y_\phi)$ are integer points, we notice that $P_{\theta+\phi}$ is also an integer point, and its coordinates $x_{\theta+\phi}$ *and* $y_{\theta+\phi}$ *are the integers defined by*

$$x_{\theta+\phi} + y_{\theta+\phi}\sqrt{d} = (x_\theta + y_\theta\sqrt{d})(x_\phi + y_\phi\sqrt{d}).$$

This follows immediately by expanding the right-hand side to

$$x_\theta x_\phi + dy_\theta y_\phi + (x_\theta y_\phi + y_\theta x_\phi)\sqrt{d}$$

and comparing with the rule for adding points.

Thus *the rule previously used to generate integer points on* $x^2 - dy^2 = 1$ (Section 8.5) *is the same as the rule for adding points by their parameter values.*

This wonderful pre-established harmony between arithmetic and geometry gives a new and useful view of the integer points on the positive branch of $x^2 - dy^2 = 1$. It shows that they form a *subgroup* of the abelian group of all points on the positive branch. The points P_θ on the positive branch are an abelian group because of the way they correspond to real numbers θ: they inherit their $+$ operation from ordinary $+$ on $\mathbb{R}$, and with it the abelian group properties of ordinary $+$ we observed in Section 6.10. And the integer points on the positive branch are also a group because:

- the sum of integer points is an integer point, by the rule for adding points,
- the inverse of an integer point (x, y) is an integer point, because

$$(x + y\sqrt{d})^{-1} = \frac{x - y\sqrt{d}}{x^2 - dy^2} = x - y\sqrt{d},$$

  because $x^2 - dy^2 = 1$ by hypothesis.

In fact, the integer points form an *infinite cyclic* group by the following result.

Generation of integer points on $x^2 - dy^2 = 1$. *The integer points on $x^2 - dy^2 = 1$ all result from* $(1, 0)$ *and the integer point nearest to it by addition and subtraction of points.*

Proof Consider the group of integer points P_ϕ on $x^2 - dy^2 = 1$, and the corresponding group of real numbers ϕ. It will suffice to show that the latter group consists of the integer multiples of θ, because the multiples of θ are precisely the numbers resulting from θ and 0 by addition and subtraction, and hence the corresponding points are those resulting from P_θ and $P_0 = (1, 0)$ by addition and subtraction of points.

The crucial property of the group of integer points P_ϕ is that it has a member nearest to $(1, 0)$, because integer points cannot approach arbitrarily close to $(1, 0)$, and hence the corresponding group of real numbers has a member θ closest to 0. Such a group of reals consists of the integer multiples $n\theta$ of θ. Why? Because if ϕ were a member strictly between, say, $k\theta$ and $(k+1)\theta$ then $\phi - k\theta$ would be a member of the group strictly between 0 and θ, contrary to the choice of θ. □

This theorem implies that the solutions of $x^2 - 2y^2 = 1$ we found in Section 8.5, by adding the point $(3, 2)$ to itself, are in fact *all* the integer solutions with x and y positive, because $(3, 2)$ is the nearest integer point to $(1, 0)$. Similarly, all the solutions of $x^2 - 3y^2 = 1$ with x and y positive are found by adding the point $(2, 1)$ to itself. A similar result holds on $x^2 - dy^2 = 1$ for any nonsquare positive integer d, thanks to the result of Section 8.7*.

Exercises

The idea of this section can be extended to other hyperbolas, even those that result from rotation of axes. An interesting example is the hyperbola $x^2 + xy - y^2 = 1$, which Vajda (1989) p. 34 showed to contain the integer points (F_{2n-1}, F_{2n}), where the F_k are the *Fibonacci numbers* defined by

$$F_0 = 0, \quad F_1 = 1, \quad F_{k+1} = F_k + F_{k-1}.$$

These numbers are linked in many ways to the roots $\tau = (1+\sqrt{5})/2$ and $\tau^* = (1-\sqrt{5})/2$ of the equation $t^2 = t + 1$, and so is the curve $x^2 + xy - y^2 = 1$.

9.6.1. Show that $x^2 + xy - y^2 = (x + y\tau)(x + y\tau^*)$.

The irrational number $1 + \tau$ can be used to generate the pairs (F_{2n-1}, F_{2n}) of consecutive Fibonacci numbers, much as $1 + \sqrt{2}$ was used to generate side and diagonal numbers in Exercise 8.5.3.

9.6.2. Deduce from the definition of Fibonacci numbers that $F_{k+3} = 2F_{k+1} + F_k$. Use this formula to prove by induction that $(1+\tau)^n = F_{2n-1} + F_{2n}\tau$ for all natural numbers n.

Now we take the hint from the factorization in Exercise 9.6.1 by defining the *sum of integer points* (x_1, y_1) *and* (x_2, y_2) *on* $x^2 + xy - y^2 = 1$ to be (x_3, y_3), where x_3 and y_3 are the integers satisfying

$$x_3 + y_3\tau = (x_1 + y_1\tau)(x_2 + y_2\tau).$$

9.6.3. Check that if (x_1, y_1) and (x_2, y_2) are points on $x^2 + xy - y^2 = 1$ then so is their sum (x_3, y_3).

9.6.4. Use addition of points to show that the hyperbola $x^2 + xy - y^2 = 1$ contains all the points (F_{2n-1}, F_{2n}) for positive integers n.

The equation $x^2 + xy - y^2 = 1$ can be rewritten $\left(x + \frac{y}{2}\right)^2 - \left(\frac{\sqrt{5}}{2}y\right)^2 = 1$ by completing the square, which suggests the parametric equations

$$x + \frac{y}{2} = \cosh\theta, \quad \frac{\sqrt{5}}{2}y = \sinh\theta.$$

9.6.5. Deduce from these parametric equations that $x + y\tau = e^\theta$, and hence conclude that the rule for adding points on $x^2 + xy - y^2 = 1$ amounts to adding their parameter values θ.

But adding parameter values makes sense for *any* points on the parameterized branch of the hyperbola, so we can extend addition of integer points to addition of any points on this branch.

As with the Pell equation, this extended addition operation allows us to find all the integer points on the branch. First we throw in some new ones—the inverses (with respect to addition of points) of the points previously found.

9.6.6. Show that $(1+\tau)^{-1} = 2 - \tau$, and hence show by induction that $(1+\tau)^{-n} = F_{2n+1} - F_{2n}\tau$ for all positive integers n.

9.6.7. Deduce from Exercises 9.6.6 and 9.6.3 that the points $(F_{2n+1}, -F_{2n})$ are on the curve $x^2 + xy - y^2 = 1$ for all positive integers n.

Finally, we see that the integer points found so far are the only ones on the branch, by relating them to a subgroup of the real numbers.

9.6.8. Show that the integer points (x_n, y_n) on $x^2 + xy - y^2 = 1$ defined by $x_n + y_n\tau = (1 + \tau)^n$ for all integers n form a subgroup of all the points on $x^2 + xy - y^2 = 1$ under addition of points, and deduce that they are *all* the integer points on the branch containing them.

9.7 Discussion

There is no last word on numbers and geometry, because these themes have infinite depth and variety. Nevertheless, this last chapter is a high point of sorts, from which it is possible to survey the ideas we have developed so far and give them some order and direction. I shall therefore discuss the ideas of Chapter 9 against the background of the whole book, reviewing some trains of thought that have led us to this point and suggesting how they might be pursued further.

From Natural Numbers to Complex Numbers

The long march of the number concept from $\mathbb{N}$ to $\mathbb{C}$, from counting to geometry, is one of the great sagas of mathematics. It was a struggle against almost insuperable obstacles, and every learner relives the struggle, to some extent, even when following the marked trail. Unfortunately, those already in the know tend to forget this. Once overcome, an obstacle is no longer an obstacle, only a step up to the next level. But making sense of negative, irrational, and imaginary numbers once seemed *impossible* tasks, so it is worth reflecting on the power of the ideas that made them possible.

The step from $\mathbb{N}$ to the integers $\mathbb{Z}$ or the rationals $\mathbb{Q}$ is technically a small one today, now that we are happy to accept infinite sets as mathematical objects. An integer can be defined as a set of pairs of natural numbers with constant difference, for example,

$$+3 = \{(0,3), (1,4), (2,5), (3,6), \dots\}$$
$$-3 = \{(3,0), (4,1), (5,2), (6,3), \dots\}.$$

We can similarly use sets of pairs of integers with constant quotient to define rational numbers. The real point of expanding $\mathbb{N}$ to $\mathbb{Z}$ and $\mathbb{Q}$ is the simplification in structure, allowing unlimited subtraction and division (except division by 0). The structure obtained—a ring in the case of $\mathbb{Z}$, a field in the case of $\mathbb{Q}$—turns out to be one of the most fruitful ideas in mathematics. It allows our intuition about numbers to be used in the study of congruence classes, polynomials, and even more abstract objects. The field structure also guides further extensions of the number concept; it becomes precisely the thing we want to preserve.

The step from $\mathbb{Q}$ to $\mathbb{R}$ is the most profound, because it creates a model of the real line and hence bridges the gap between discrete and continuous. It is this step that commits us to set theory irrevocably; we cannot do without infinite sets and an uncountable number of them. However, this step also brings a huge gain in understanding. It not only shows that rational numbers are "rare," in the sense that there are only countably many of them, it shows that the same is true of the *algebraic* numbers; the numbers (such as $\sqrt{2}$) that are roots of polynomial equations with integer coefficients.

When Cantor proved that $\mathbb{R}$ is uncountable in 1874 he actually began by proving that there are only countably many algebraic numbers (which is not hard, because it amounts to listing the polynomial equations with integer coefficients). His uncountability proof then enabled him to conclude immediately that *transcendental numbers exist,* a result previously obtained with great difficulty by proving specific numbers transcendental.

Dedekind's idea of creating $\mathbb{R}$ by completing $\mathbb{Q}$, or "filling its gaps," also plugged many holes in the logic of calculus and geometry. In particular, it is crucial in the so-called *fundamental theorem of algebra*, which states that any polynomial equation has a root in $\mathbb{C}$. This turns out to depend on properties of continuous functions on the plane, which were not properly proved until the completeness of $\mathbb{R}$ was understood. The details may be found for example in Stillwell (1994).

Extending $\mathbb{R}$ to $\mathbb{C}$ is technically a small step. As mentioned in Section 7.1, it suffices to define complex numbers as pairs of real numbers and define $+$ and $\times$ appropriately. The unexpected power of $\mathbb{C}$ derives partly from the completeness of $\mathbb{R}$ and partly from the harmony between $+$ and $\times$ on $\mathbb{C}$ and the geometry of the plane. It is a kind of miracle that so much algebraic and geometric structure can coexist in the same object, and in fact this miracle is not repeated. There is no way to define $+$ and $\times$ in a space of three or more dimensions so that the resulting structure is a field, though there are "near misses" in dimension 4 (the multiplication is noncommutative) and dimension 8 (the multiplication is nonassociative and noncommutative). More information on this interesting question may be found in Artmann (1988) and Ebbinghaus et al. (1991).

The Exponential Function

The exponential function, like $\mathbb{C}$ itself, is an amazing confluence of arithmetic and geometric ideas.

In arithmetic, the idea of exponentiation first arises in $\mathbb{N}$. As we know from Section 1.7, powers of 2 occur in Euclid's theorem on perfect numbers. The fundamental property of powers of 2 is that when powers are multiplied, their exponents add:

$$2^m \times 2^n = 2^{m+n}.$$

In Section 6.10 we observed that in $\mathbb{Q}$, where we also have negative powers of 2, there is in fact an *isomorphism* between the group of powers of 2 under $\times$ and the group of integers under $+$. One of the advantages of $\mathbb{R}$ is that it allows 2^a to be defined for any real numbers a, extending the exponent addition property to all real a and b:

$$2^a \times 2^b = 2^{a+b}.$$

This gives an isomorphism between the group of real powers of 2, under $\times$, and the group $\mathbb{R}$ under $+$.

This isomorphism allows us to multiply positive reals x and y by the simpler operation of addition as soon as we know *logarithms* and *antilogarithms*. The logarithm (to base 2) of $x > 0$ is the number a such that $x = 2^a$, so if we also know the logarithm b of y we can find