the origin, so if necessary one can test all integer points within this radius to see whether they satisfy the equation.

The parabola $y = ax^2$ has infinitely many integer points, which are easily computed from a, assuming a is rational. We write a as a fraction in lowest terms, m/n, so the equation becomes $ny = mx^2$. Then each integer x makes mx^2, and hence ny, an integer, and so it gives an integer point just in case n divides mx^2. This in fact happens precisely for the multiples of the least positive x divisible by n.

The hyperbola is interesting even in the special case $x^2 - dy^2 = 1$, where d is a positive integer. In fact, we shall first study $x^2 - 2y^2 = 1$, and concentrate on its "positive branch," for which $x > 0$. There is one obvious integer point on this hyperbola, namely, $(1, 0)$, and the next is found by trial to be $(3, 2)$.

The point $(3, 2)$ is a "seed" that produces all the integer points on $x^2 - 2y^2 = 1$. Here we shall generate infinitely many integer points from it, and in the next chapter we'll show that they are *all* the integer points on the positive branch. The process of generation is surprising, because it uses the irrational number $\sqrt{2}$. We use $\sqrt{2}$ to define a "product" of integer points on $x^2 - 2y^2 = 1$ as follows.

Generation of integer points on $x^2 - 2y^2 = 1$. *If (m_1, n_1) and (m_2, n_2) are integer points on the hyperbola $x^2 - 2y^2 = 1$, then so is the point (m_3, n_3), where m_3 and n_3 are defined by*

$$m_3 + n_3\sqrt{2} = (m_1 + n_1\sqrt{2})(m_2 + n_2\sqrt{2}).$$

Proof First we should make sure that m_3 and n_3 really are *defined* by

$$m_3 + n_3\sqrt{2} = (m_1 + n_1\sqrt{2})(m_2 + n_2\sqrt{2}).$$

The intention is to expand the right-hand side as

$$m_1m_2 + 2n_1n_2 + \sqrt{2}(n_1m_2 + m_1n_2)$$

and set $m_3 = m_1m_2 + 2n_1n_2$ and $n_3 = n_1m_2 + m_1n_2$ by "equating rational and irrational parts." But why is this valid? The reason is that $r + s\sqrt{2} = u + v\sqrt{2}$ for rational r, s, u, v only if $r = u$ and $s = v$;

if not, we have $\sqrt{2} = (r-u)/(v-s)$, contrary to the irrationality of $\sqrt{2}$.

It is clear from the definition of m_3 and n_3 that they are integers. Now because (m_1, n_1) and (m_2, n_2) are points on $x^2 - 2y^2 = 1$, we have

$$m_1^2 - 2n_1^2 = 1 \qquad \text{and} \qquad m_2^2 - 2n_2^2 = 1.$$

Factorizing these equations we get

$$(m_1 + n_1\sqrt{2})(m_1 - n_1\sqrt{2}) = 1 \quad \text{and} \quad (m_2 + n_2\sqrt{2})(m_2 - n_2\sqrt{2}) = 1,$$

and taking their product,

$$(m_1 + n_1\sqrt{2})(m_1 - n_1\sqrt{2})(m_2 + n_2\sqrt{2})(m_2 - n_2\sqrt{2}) = 1.$$

Rearranging the factors gives

$$(m_1 + n_1\sqrt{2})(m_2 + n_2\sqrt{2})(m_1 - n_1\sqrt{2})(m_2 - n_2\sqrt{2}) = 1,$$

which is in fact

$$(m_3 + n_3\sqrt{2})(m_3 - n_3\sqrt{2}) = 1.$$

The first factor comes from the definition of m_3 and n_3, and the second because changing + to − signs in the definition still gives a valid identity. Expanding the last equation, we get

$$m_3^2 - 2n_3^2 = 1,$$

so (m_3, n_3) is an integer point on $x^2 - 2y^2 = 1$, as required. □

It follows from this result that the points (m_k, n_k) defined by

$$m_k + n_k\sqrt{2} = (3 + 2\sqrt{2})^k$$

are infinitely many integer points on the hyperbola $x^2 - 2y^2 = 1$. The result also has an obvious generalization for any nonsquare positive integer d, using the fact that $\sqrt{d}$ is irrational for such a d.

Generation of integer points on $x^2 - dy^2 = 1$. *If (m_1, n_1) and (m_2, n_2) are integer points on the hyperbola $x^2 - dy^2 = 1$, then so is the point (m_3, n_3), where m_3 and n_3 are defined by*

$$m_3 + n_3\sqrt{d} = (m_1 + n_1\sqrt{d})(m_2 + n_2\sqrt{d}).$$ □

Exercises

In case you are wondering why we started with the hyperbola $x^2 - 2y^2 = 1$ instead of $x^2 - y^2 = 1$ ....

8.5.1. Observe that $x^2 - y^2 = (x + y)(x - y)$, and hence show that $(\pm 1, 0)$ are the only integer points on $x^2 - y^2 = 1$. What can you say about integer points on $x^2 - dy^2 = 1$ when d is an integer square?

The integer points (m_k, n_k) on $x^2 - 2y^2 = 1$ were known to the Greeks under the name of "side and diagonal numbers," because the ratios n_k/m_k approximate the ratio $\sqrt{2}$ between the diagonal and side of the square.

8.5.2. Check that the first few values of (m_k, n_k) are $(3, 2)$, $(17, 12)$, $(99, 70)$, and show that $n_k/m_k \to \sqrt{2}$ as $k \to \infty$. Give a geometric interpretation of this fact in terms of the asymptote $x = \sqrt{2}y$ of the hyperbola $x^2 - 2y^2 = 1$.

The pairs (m_k, n_k) are actually not all the side and diagonal number pairs. The Greeks discovered the sequence

$$(1, 1), \quad (3, 2), \quad (7, 5), \quad (17, 12), \quad (41, 29), \quad (99, 70), \quad \ldots$$

of pairs (x_k, y_k) that alternately satisfy $x^2 - 2y^2 = -1$ and $x^2 - 2y^2 = 1$. We are not sure how they discovered the sequence, but they computed it from $(x_1, y_1) = (1, 1)$ and the following *recurrence relations* giving (x_{k+1}, y_{k+1}) in terms of (x_k, y_k):

$$x_{k+1} = x_k + 2y_k,$$
$$y_{k+1} = x_k + y_k.$$

8.5.3. Show by induction that the sequence (x_k, y_k) defined by

$$x_k + y_k\sqrt{2} = (1 + \sqrt{2})^k$$

satisfies the recurrence relations and hence agrees with the sequence of pairs of side and diagonal numbers.

8.5.4. Deduce from Exercise 8.5.3 that $(m_k, n_k) = (x_{2k}, y_{2k})$.

8.5.5. Show also that the pairs (x_{2k+1}, y_{2k+1}) satisfy the equation $x^2 - 2y^2 = -1$.

It should be noted that the *negative* integer powers of $3 + 2\sqrt{2}$ also give integer points on $x^2 - 2y^2 = 1$, for the following reason.

8.5.6. Show that if $m_k+n_k\sqrt{2} = (3+2\sqrt{2})^k$ then $m_k-n_k\sqrt{2} = (3-2\sqrt{2})^k = (3+2\sqrt{2})^{-k}$.

Notice also that $(3+2\sqrt{2})^0 = 1+0\sqrt{2}$ gives the point $(1,0)$ on $x^2-2y^2=1$. Thus the points we have found so far correspond to members of the infinite cyclic group of numbers $(3+2\sqrt{2})^k$ for integers k. (See Section 6.10 for the definitions of abelian and cyclic groups.) There is in fact a way to treat the whole positive branch of the curve as a group, as we shall see in Chapter 9, and it then becomes clear that the points we have found so far are the subgroup of all integer points.

The deduction of $m_3^2-2n_3^2=1$ from $m_1^2-2n_1^2=1$ and $m_2^2-2n_2^2=1$ hints at another instance of a multiplicative norm, like the norm $N(a+bi)$ on $\mathbb{Z}[i]$ we studied in Section 7.4. Here we are dealing with the ring

$$\mathbb{Z}[\sqrt{2}] = \{a+b\sqrt{2} : a,b \in \mathbb{Z}\}$$

with the norm defined by

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

We are interested in the members $a+b\sqrt{2}$ with norm 1, because they correspond to integer points on $x^2-2y^2=1$, but the norm is in fact multiplicative for all members of $\mathbb{Z}[\sqrt{2}]$.

8.5.7. Show that $N((a_1+b_1\sqrt{2})(a_2+b_2\sqrt{2})) = N(a_1+b_1\sqrt{2})N(a_2+b_2\sqrt{2})$.

8.5.8. Suggest a norm for $\mathbb{Z}[\sqrt{d}]$ and show that it is multiplicative.
(*Hint*: It may help to recall Brahmagupta's identity from the exercises to Section 7.1.)

8.6* Square Roots and the Euclidean Algorithm

The irrationality of $\sqrt{2}$ and other numbers tormented Greek mathematicians for hundreds of years and provoked many attempts to relate irrationals to integers in a comprehensible way. The most interesting, as far as square roots are concerned, is a generalization of the Euclidean algorithm. As Euclid himself described it, the Euclidean algorithm "continually subtracts the lesser number from the greater." Such a process can also be applied to a pair of numbers

whose ratio is irrational, such as $\sqrt{2}$ and 1. Of course the algorithm will not terminate in this case, but if there is some pattern to the numbers produced it surely gives some new understanding of the nature of $\sqrt{2}$. And for $\sqrt{2}$ we get a pattern that is the next best thing to termination, namely, *periodicity*.

The pattern can be seen most easily by applying the Euclidean algorithm to the pair $(\sqrt{2}+1, 1)$. The lesser number 1 can be subtracted twice from the greater, $\sqrt{2}+1$, producing the pair $(1, \sqrt{2}-1)$. It so happens that the new lesser number $\sqrt{2}-1$ can also be subtracted twice from the new greater number 1, and the same thing happens again and again; it appears that the lesser number can *always* be subtracted twice from the greater number. But how can we be sure?

The fog clears miraculously if we view each number pair as adjacent sides of a rectangle and subtract the lesser number from the greater by cutting off the square on the lesser side. In particular, the first two subtractions are interpreted as cutting off unit squares from the rectangle with sides $\sqrt{2}+1$ and 1, as shown in Figure 8.7.

This produces a rectangle with sides 1 and $\sqrt{2}-1$ and the new rectangle is the same shape as the original. We can confirm this by computing the ratio of the sides:

$$\frac{1}{\sqrt{2}-1} = \frac{\sqrt{2}+1}{(\sqrt{2}-1)(\sqrt{2}+1)} = \frac{\sqrt{2}+1}{1}.$$

It follows that subtracting the lesser side twice from the greater will produce rectangles of the same shape indefinitely. Thus the Euclidean algorithm is *periodic* on the pair $(\sqrt{2}+1, 1)$ in the sense that it continually subtracts the lesser number twice from the greater.

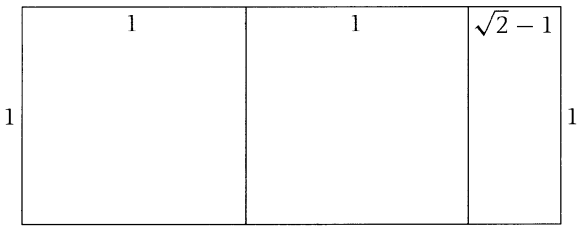

FIGURE 8.7 Periodicity and $\sqrt{2}$.

This means that if we use the algorithm in division-with-remainder form, the quotient at each step is 2.

As for the pair $(\sqrt{2}, 1)$, at first the lesser number is subtracted once from the greater, producing the pair $(1, \sqrt{2} - 1)$. But this is the pair we have just seen, so after the first subtraction the lesser number is always subtracted twice from the greater. Equivalently, the sequence of quotients in the division-with-remainder algorithm is $1, 2, 2, 2, 2, 2, \ldots$. We say that the Euclidean algorithm is *ultimately periodic* on the pair $(\sqrt{2}, 1)$.

The geometric explanation of the periodicity of the Euclidean algorithm also shows that each new pair is obtained from the previous pair by multiplying its members by $\sqrt{2} - 1$. Hence the $(k+1)$th pair is $(\sqrt{2}-1)^k(\sqrt{2}+1, 1)$. This links the periodicity of $\sqrt{2}$ with the process used in the previous section to generate integer points on the hyperbola $x^2 - 2y^2 = 1$.

Perhaps the most attractive way to display the periodicity of $\sqrt{2}$ is to work out its *continued fraction*. Recycling some of the facts found previously, we find

$$\begin{aligned}
\sqrt{2}+1 &= 2 + \sqrt{2} - 1 \\
&= 2 + \frac{1}{\sqrt{2}+1} \quad \text{as we already know} \\
&= 2 + \frac{1}{2 + \sqrt{2} - 1} \quad \text{by the first line} \\
&= 2 + \frac{1}{2 + \frac{1}{\sqrt{2}+1}} \quad \text{by the second line} \\
&= 2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{\sqrt{2}+1}}} \quad \text{similarly,}
\end{aligned}$$

and so on. The limit of these fractions exists and is called the *continued fraction for* $\sqrt{2}+1$. We write it

$$\sqrt{2}+1 = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \ddots}}},$$

and subtracting 1 from both sides gives the continued fraction for $\sqrt{2}$:

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}.$$

The sequence $1, 2, 2, 2, 2, 2, 2, 2, \ldots$ of natural numbers to the left of the $+$ signs is the sequence of quotients occurring in the running of the Euclidean algorithm on $(\sqrt{2}, 1)$.

Exercises

The Euclidean algorithm on $(\sqrt{2}, 1)$ is linked not only with the integer points on the hyperbola $x^2 - 2y^2 = 1$ but also with the side and diagonal numbers mentioned in the previous exercise set. The Greeks may very well have discovered the side and diagonal numbers as the coefficients of $\sqrt{2}$ and 1 occurring in successive terms produced by the Euclidean algorithm.

8.6.1. Show that $(\sqrt{2} - 1)^k = (-1)^k(y_k\sqrt{2} - x_k)$, where y_k and x_k are the side and diagonal numbers defined in the previous exercise set.

8.6.2. Deduce from Exercise 8.6.1 the coefficients of each term produced from $(\sqrt{2}, 1)$ by the Euclidean algorithm form a side and diagonal number pair.

There is a similar relationship between $\sqrt{3}$, the Euclidean algorithm and integer points on $x^2 - 3y^2 = 1$. Again we find that $\sqrt{3} + 1$ has slightly simpler behavior than $\sqrt{3}$, by considering a rectangle of width $\sqrt{3}+1$ and height 1.

8.6.3. Show that the large and small rectangles in Figure 8.8 are the same shape.

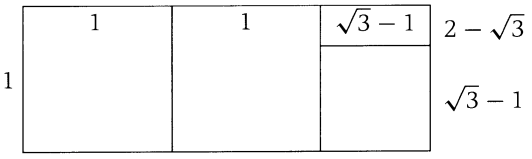

FIGURE 8.8 Periodicity and $\sqrt{3}$.

8.6.4. Deduce from Exercise 8.6.3 that

- the Euclidean algorithm is periodic on $(\sqrt{3}+1, 1)$, with successive quotients $2, 1, 2, 1, 2, 1, 2, 1, \ldots$,
- $\sqrt{3}+1 = 2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{\ddots}}}}}}$,
- each period of the Euclidean algorithm on $(\sqrt{3}+1, 1)$ reduces the size of the pair by a factor $2-\sqrt{3}$.

Now define integers a_k, b_k by $(2+\sqrt{3})^k = a_k + b_k\sqrt{3}$, or equivalently by $(2-\sqrt{3})^k = a_k - b_k\sqrt{3}$.

8.6.5. Show that all the points (a_k, b_k) lie on the hyperbola $x^2 - 3y^2 = 1$.

As in the previous exercise set, these results can be interpreted as finding quadratic integers of norm 1 as powers of a "seed" quadratic integer of norm 1.

8.6.6. Interpret the previous result in terms of the norm $N(a+b\sqrt{3}) = a^2 - 3b^3$ on $\mathbb{Z}[\sqrt{3}]$, observing that $2+\sqrt{3}$ and $2-\sqrt{3}$ have norm 1.

8.7* Pell's Equation

The equation $x^2 - dy^2 = 1$, where d is a nonsquare integer, is known as *Pell's equation*. John Pell was a 17th-century mathematician who had little or nothing to do with the equation, but Euler attached his name to it by mistake, and it stuck. The equation would be better named after Brahmagupta or Fermat, who solved it for particular values of d, or after Lagrange, who first gave the complete solution. We already know solutions for $d = 2$ and $d = 3$, where small solutions can be found by trial, and we have seen how these small solutions generate infinitely many others. For larger values of d, however, it is hard to find even one solution, apart from the trivial one $x = 1$, $y = 0$. The smallest nontrivial solution appears to vary wildly with d, and can be alarmingly large. Brahmagupta said "whoever can solve $x^2 - 92y^2 = 1$ in less than a year is a mathematician," and the equation $x^2 - 61y^2 = 1$ (posed by Fermat) is tougher still.