The smallest nontrivial solution of $x^2 - 92x^2 = 1$ is $x = 1151$, $y = 120$, while the smallest nontrivial solution of $x^2 - 61y^2 = 1$ is $x = 1766319049$, $y = 226153980$.

In such a situation, proving the *existence* of a nontrivial solution is easier than finding it. We shall in fact find infinitely many candidates for the smallest nontrivial solution, and show that one of them must be correct. The method of proof was invented by Dirichlet, and he called it the *pigeonhole principle*. The finite form of the principle says that if $n + 1$ pigeons are in n boxes then at least one box contains two pigeons. The infinite form of the principle says that if infinitely many pigeons are in finitely many boxes, then at least one box contains infinitely many pigeons. Both forms of the pigeonhole priciple are involved in the proof; we shall use the finite one first.

Dirichlet's approximation theorem. *For any real number α and any integer $Q > 1$ there are integers p, q with $0 < q < Q$ and $|q\alpha - p| \leq 1/Q$.*

Proof Consider the $Q + 1$ numbers

$$0, \quad 1, \quad \alpha - p_1, \quad 2\alpha - p_2, \quad \ldots, \quad (Q-1)\alpha - p_{Q-1},$$

where $p_1, p_2, \ldots, p_{Q-1}$ are integers chosen so that all the numbers lie in the interval from 0 to 1. If we divide this interval into subintervals of length $1/Q$, then we have Q subintervals containing $Q + 1$ numbers. Hence at least two numbers are in the same subinterval; that is, they are distance $\leq 1/Q$ apart. Because the difference between any two of the numbers is of the form $q\alpha - p$, for integers p and q with $0 < q < Q$, this means $|q\alpha - p| \leq 1/Q$ as required. □

This theorem says that $q\alpha - p$ can be made at least as small as $1/q$, by suitable choice of p and q. It is particularly useful when α is irrational, because $q\alpha - p$ is never zero in that case, and hence we get infinitely many numbers $q\alpha - p$, each no larger than the corresponding $1/q$.

Here is how Dirichlet used his approximation theorem to show there are integers x and y such that $x^2 - dy^2 = 1$. The strategy is to make $p - q\sqrt{d}$ small enough that

$$p^2 - dq^2 = (p - q\sqrt{d})(p + q\sqrt{d}) \leq 3\sqrt{d}.$$

Thanks to the irrationality of $\sqrt{d}$, this gives infinitely many integers p and q for which $p^2 - dq^2 \leq 3\sqrt{d}$, and the infinite pigeonhole principle can then be used to show that some of them give $p^2 - dq^2 = 1$.

The other tool in the proof is the norm $N(p - q\sqrt{d}) = p^2 - dq^2$ and its multiplicative property, which can be verified by multiplying out both sides: $N((p_1 - q_1\sqrt{d})(p_2 - q_2\sqrt{d})) = N(p_1 - q_1\sqrt{d})N(p_2 - q_2\sqrt{d})$.

Existence of nontrivial solutions of $x^2 - dy^2 = 1$. *If $\sqrt{d}$ is irrational there are positive integers x and y such that $x^2 - dy^2 = 1$.*

Proof Applying the Dirichlet approximation theorem to $\alpha = \sqrt{d}$, for any integer $Q > 1$ we have positive integers p, q with

$$|p - q\sqrt{d}| \leq 1/Q \quad \text{and} \quad 0 < q < Q.$$

Because $\sqrt{d}$ is irrational, $p - q\sqrt{d} \neq 0$. By letting $Q \to \infty$ we therefore get infinitely many pairs of positive integers p, q with $|p - q\sqrt{d}| \leq 1/q$.

Now for any such pair

$$|p + q\sqrt{d}| \leq |p - q\sqrt{d} + 2q\sqrt{d}| \leq |p - q\sqrt{d}| + |2q\sqrt{d}| \leq 3q\sqrt{d},$$

hence

$$|p^2 - q^2 d| = |p + q\sqrt{d}||p - q\sqrt{d}| \leq 3q\sqrt{d}/q = 3\sqrt{d}.$$

Thus we have infinitely many pairs of positive integers p, q with $N(p - q\sqrt{d}) \leq 3\sqrt{d}$.

We now apply the infinite pigeonhole principle to obtain infinitely many pairs p, q with even more special properties.

1. Because there are only finitely many natural numbers $\leq 3\sqrt{d}$, infinitely many of the numbers $p - q\sqrt{d}$ have the same norm, say N.
2. Because there are only finitely many congruence classes mod N, infinitely many of the numbers $p - q\sqrt{d}$ with norm N have p in the same congruence class, and infinitely many of the latter numbers have q in the same congruence class.

To sum up, there is an infinite set of numbers $p - q\sqrt{d}$ with the same norm $N(p - q\sqrt{d}) = N$, all p in the same congruence class mod N, and all q in the same congruence class mod N.

Now take two numbers $p_1 - q_1\sqrt{d}$, $p_2 - q_2\sqrt{d}$ from this set and consider their quotient $x + y\sqrt{d}$, which has norm 1 by the multiplicative

property of norm. I claim that x and y are integers. Because

$$\frac{p_1 - q_1\sqrt{d}}{p_2 - q_2\sqrt{d}} = \frac{(p_1 - q_1\sqrt{d})(p_2 + q_2\sqrt{d})}{p_2^2 - q_2^2 d}$$
$$= \frac{p_1p_2 - q_1q_2d}{N} + \frac{p_1q_2 - q_1p_2}{N}\sqrt{d},$$

we have to prove N divides $p_1p_2 - q_1q_2d$ and $p_1q_2 - q_1p_2$. By hypothesis,

$$p_1 \equiv p_2 \pmod{N} \quad \text{and} \quad q_1 \equiv q_2 \pmod{N},$$

hence

$$p_1p_2 - q_1q_2d \equiv p_1^2 - q_1^2 d \equiv 0 \pmod{N},$$

because

$$p_1^2 - q_1^2 d = N(p_1 - q_1 d) = N.$$

Thus N divides $p_1p_2 - q_1q_2d$. It also follows, by multiplying the congruences $p_1 \equiv p_2 \pmod{N}$ and $q_2 \equiv q_1 \pmod{N}$, that $p_1q_2 \equiv q_1p_2 \pmod{N}$, and hence

$$p_1q_2 - q_1p_2 \equiv 0 \pmod{N}.$$

Thus N divides $p_1q_2 - q_1p_2$. This proves the claim that

$$\frac{p_1 - q_1\sqrt{d}}{p_2 - q_2\sqrt{d}} = x + y\sqrt{d} \quad \text{for some integers } x \text{ and } y,$$

and $1 = N(x + y\sqrt{d}) = x^2 - dy^2$. Finally, $y \neq 0$ because $p_1 - q_1\sqrt{d} \neq p_2 - q_2\sqrt{d}$, so this is a nontrivial solution. □

Exercises

Dirichlet's approximation theorem says there are rational numbers p/q "very close" to any irrational number α. For each $q > 1$ there are fewer than q^2 rationals with denominator $\leq q$ between successive integers, nevertheless ....

8.7.1. Deduce from Dirichlet's approximation theorem that there are infinitely many values of q for which there is a rational p/q at distance no more than $1/q^2$ from α.

Two instances of this close approximation phenomenon are the approximation 22/7 to π and the even more remarkable approximation 355/113 discovered by the Chinese mathematician Zǔ Chōngzhi (429–500 A.D.).

8.7.2. Using the numerical value $\pi = 3.14159265\ldots$, show that 22/7 approximates π within $1/7^2$ and 355/113 approximates π within $1/113^2$ (in fact much more closely).

The existence of a nontrivial solution to $x^2 - dy^2 = 1$ is connected with the periodicity of the continued fraction for $\sqrt{d}$, as one would imagine from the examples $d = 2$ and $d = 3$ studied in the previous section and its exercises. In fact the traditional method for finding a nontrivial solution of $x^2 - dy^2 = 1$ was to derive it from the ultimate periodicity of the continued fraction for $\sqrt{d}$. However, the periodicity result is somewhat harder, and for proofs we refer the reader to Stark (1978) or Baker (1984).

8.8 Discussion

The Projective View of Conic Sections

An interesting alternative to the process of cutting a cone by a plane is the process of *projecting a circle*. These two processes are much the same, but the concept of projection is worth a closer look, because it brings new ideas to the fore and actually leads to a whole new branch of mathematics: *projective geometry*. To grasp the idea of projection, imagine a vertical pane of glass with a circle C drawn on it, illuminated by light from a point P (Figure 8.9).

The shadow $\mathcal{K}$ of C on a horizontal plane is a conic section:

- an ellipse if P is above the top of C,
- a parabola if P is level with the top of C,
- a hyperbola if P is below the top of C (but above the bottom of C).

So far, this is just a way to produce a cone (the cone of rays through P and C) and cut it by a plane (the horizontal plane), and hence obtain a conic section (the shadow of C). But it gets more interesting when

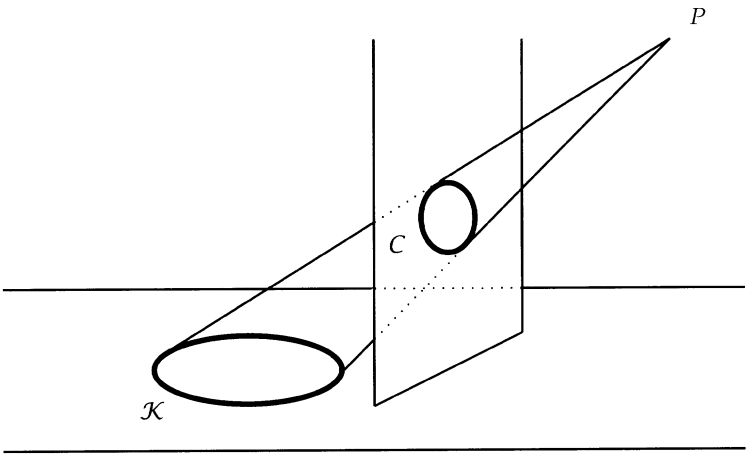

FIGURE 8.9 A conic section as the projection of a circle.

the light rays are reversed. Imagine that your eye is at the point P, receiving light rays emitted by a conic section $\mathcal{K}$ in the horizontal plane. The view of $\mathcal{K}$ seen when looking straight ahead can be captured by drawing, on a vertical window, a curve C that appears to cover $\mathcal{K}$. This in fact is the method used by Renaissance artists to draw three-dimensional scenes in correct perspective. The window was called "Alberti's veil," and Figure 8.10 is a woodcut by Albrecht Dürer showing how to use it.

Because any conic section is the projection of a circle, it follows that *any conic section looks like a circle, when suitably viewed.* But in that case, how do we identify which kind of conic section $\mathcal{K}$ we are viewing from a point P? We do so by observing the position of the circle, C, relative to the *horizon*, which is:

- above C when $\mathcal{K}$ is an ellipse, because in this case P is above C,
- tangential to the top of C when $\mathcal{K}$ is a parabola, because in this case P is level with the top of C,
- through two points of C when $\mathcal{K}$ is a hyperbola, because in this case P is below the top of C.