Diophantus' success can be partly explained by his innovations in notation, which enabled him to carry out more complex algebraic manipulations than his predecessors. He used a symbol for the unknown, and abbreviations for the arithmetic operations, which were sufficient to solve certain polynomial equations and compute with complicated fractions. The limitation of his notation is that there is only *one* symbol for an unknown, so problems with several unknowns are solved by choosing particular values for all but one of them. This is why he restricts himself to particular problems—and why the restriction is not severe, as Euler realized.

The distinctive feature of Diophantus' work is an interest in *rational* solutions of equations. In some ways, rational numbers are easier to work with than integers, so Diophantus had the advantage of being first into a field his predecessors were not equipped to explore. However, he brought to this field exceptional ingenuity and insight. His ideas were not fully understood, let alone extended, until Fermat reconsidered them in the 17th century.

Diophantus' subject matter is now called *Diophantine equations*, a rather misleading term that replaces the equally misleading "indeterminate equations" found in older books. It would be better described as *finding rational solutions of equations*. Typically, the equations considered have infinitely many solutions (hence the term *indeterminate*) and the challenge is to find the rational solutions, if any. Since the time of Fermat, it has been recognized that finding integer solutions is an even more challenging problem, and the term *Diophantine equations* is sometimes reserved for the subject with this narrower aim. Today, mathematicians have come to view these subjects geometrically, and they are often described as *finding rational points on curves* and *finding integer points on curves*. Is this really what Diophantus was doing? He did not say so, but his solutions are open to both algebraic and geometric interpretations.

The classic source of Diophantus in English is the translation and commentary by Heath (1910). This book is still the most complete and informative, and incidentally it's also a superb introduction to the number theory of Fermat and Euler. However, Heath views Diophantus purely as an algebraist, and to see the geometric side of the story, it is also advisable to read Weil (1984).

Several interesting problems are conspicuous by their absence from the *Arithmetica*. Diophantus sometimes skirts around a problem, answering several questions but not the one that seems most central. It looks like he has been stumped, then (like a student faced with a similar situation on an exam) decided to tell what he knows about something else. He is answering related questions, but with extra conditions that make them easier to solve. The missing questions were eventually raised by readers of the *Arithmetica*, particularly Fermat, and it became clear that new ideas were needed to answer them. Fermat claimed solutions, but divulged very few; most of the published solutions were by Euler and Lagrange. We shall study some of their innovations later, but it is appropriate to mention the questions here.

The first two arise from the study of rational right-angled triangles, as we saw in Section 4.7*.

1. *Can the area of a rational right-angled triangle be a square?*
   All of Book VI in the *Arithmetica* is concerned with rational right-angled triangles. As mentioned in Section 4.7*, Diophantus finds examples whose area is a square $\pm$ a given number, a square $\pm$ the sum of the perpendiculars, a square minus the hypotenuse, and a square minus the perimeter—virtually everything *except* a square.

   Fermat proved that the latter is impossible, by an argument similar to that given in Section 4.7*.

2. *Can the sum of two fourth powers be a square?*
   In Book V, Problem 29, Diophantus gave an example of three numbers, 144/25, 9 and 16, whose fourth powers sum to a square.

   Fermat asked about the sum of two fourth powers, and showed that the answer is no. It is remarkable that the answer comes from his proof that the area of a rational right-angled triangle is not a square.

3. *Is every positive integer the sum of four squares?*
   This question was raised in 1621 by Bachet, whose edition of Diophantus was the one used by Fermat. Bachet was prompted by Diophantus' Problem 29 of Book IV, which answers a more complicated question about sums of squares.

Fermat claimed he could prove that every positive integer is the sum of four squares, Euler attacked the problem without complete success, and the first solution was published by Joseph Louis Lagrange in 1770.

4. *Is every prime of the form* $4n + 1$ *the sum of two squares?*
This question arises from Problem 19 of Book III, which is discussed further in Section 7.1.

Fermat claimed a proof in 1640, but the first published proof was by Euler in 1749. It was followed by many other proofs, one of which is presented in Section 7.6.

5. *Is* $x = 5$, $y = 3$ *the only positive integer solution of* $y^3 = x^2 + 2$?
Diophantus gives this solution in Book VI, Problem 17, though without claiming it is unique.

Fermat in 1657 claimed that it was the only positive integer solution, and a remarkable proof was given by Euler (1770). Euler's proof (with some necessary slight corrections) is in the exercises to Section 7.6.

Fermat's Last Theorem

The formula for rational Pythagorean triples credited to Diophantus in Section 4.3 is not exactly what he wrote. However, this formula can be read between the lines of Problem 8 in Book II of the *Arithmetica*: splitting a square into two squares. Because this problem is important for other reasons, it is worth studying Diophantus' solution, given here in the translation of Heath (1910).

8. To divide a given square number into two squares.

Given square number 16.

x^2 one of the required squares. Therefore $16 - x^2$ must be equal to a square.

Take a square of the form $(mx - 4)^2$, m being any integer and 4 the number which is the square root of 16, e.g. take $(2x - 4)$, and equate it to $16 - x^2$.

Therefore $4x^2 - 16x + 16 = 16 - x^2$,
or $5x^2 = 16x$, and $x = 16/5$.
The required squares are therefore $\frac{256}{25}$, $\frac{144}{25}$.

We see from this why Diophantus seeks *rational* solutions of equations. There are no positive integer squares x^2 and y^2 that sum to 16, so rational solutions are the interesting ones in this problem. But why try to express $16 - x^2$ as a square of the form $(mx - 4)^2$? It works algebraically because the constant term in $(mx - 4)^2$ cancels the 16, but $mx - 4$ also has a *geometric meaning*, which makes the solution easier to understand and generalize.

The pairs of numbers (x, y) such that $x^2 + y^2 = 16$ form a circle in the (x, y) plane, so Diophantus' problem is equivalent to finding *rational points* on this circle. Because $16 = 4^2$, there are some obvious rational points, for example, $x = 0$, $y = -4$. And $y = mx - 4$ *is a line through the "obvious" rational point* $(0, 4)$. Diophantus is simply finding the *other* intersection of this line with the circle, in this case $m = 2$. He could choose any rational value of m and still find the other intersection to be rational.

By implication, Diophantus allows any rational value of m, so he can actually find *all* rational points on the circle, simply because the line through any rational point (s, t) and $(0, -4)$ has rational slope $m = \frac{t+4}{s}$. There is also nothing special about the radius 4. The rational points on a circle of any rational radius r can be found by multiplying those on the circle of radius 4 by $r/4$. Thus Diophantus has really solved the problem of finding all rational points on a circle of rational radius, as we claimed in Section 4.3.

He has also solved the equivalent problem: to divide a given (rational) square into two (rational) squares. It was this solution that inspired Fermat to make a note in the margin next to Problem 8 of Book II in his copy of Diophantus:

> It is impossible to separate a cube into two cubes, or a biquadrate into two biquadrates, or in general any power higher than second into powers of like degree: I have discovered a truly marvellous proof of this which however this margin is too small to contain.

More concisely, Fermat's claim is that the equation $x^n + y^n = z^n$ has no solution in positive integers x, y, z when n is an integer > 2. This

became known as *Fermat's last theorem*, not because Fermat proved it, but because it was the last of Fermat's claims to be settled. In fact, Fermat was almost certainly mistaken to think he had a proof, though he could prove the case of biquadrates (fourth powers), as we saw in Section 4.7*.

Elliptic Curves

Fermat's last theorem was not proved until 1994, and then only through the work of several mathematicians: Gerhardt Frey, Jean-Pierre Serre, Ken Ribet, Richard Taylor, and especially Andrew Wiles. The proof involves some of the most abstract and difficult techniques of modern mathematics, but they are used to make a connection between the nth-degree equation $x^n+y^n = z^n$ and something relatively simple: a *cubic* equation of the form $y^2 = x(x-\alpha)(x-\beta)$. In 1984, Frey had the wild idea to suppose (contrary to Fermat's last theorem) that there are positive integers a, b, c with $a^n + b^n = c^n$, and to see what this implied about the curve with equation $y^2 = x(x-a^n)(x+c^n)$. He suspected, but could not prove, that the unlikely numbers a^n and c^n would give the curve an unlikely property, known as *nonmodularity*.

To cut a long story short, Fermat's last theorem was proved by showing that a counterexample (a, b, c) to Fermat's last theorem *does* imply nonmodularity (Serre and Ribet), but that nonmodularity is impossible for the curves $y^2 = x(x-\alpha)(x-\beta)$ (Taylor and Wiles). Consequently, there is no counterexample to Fermat's last theorem! It is way beyond the scope of this book to explain what *nonmodularity* is, but it is worth saying a few words about the cubic curves $y^2 = x(x-\alpha)(x-\beta)$, as they also go back to Diophantus and Fermat.

There is an important difference between quadratic and higher-degree curves, as we know from the exercises in Sections 4.3, 4.5, and 4.7*. Any quadratic curve can be parameterized by rational functions, but a higher degree curve generally can not. The simplest functions that can parameterize the cubic curve $y^2 = x(x-\alpha)(x-\beta)$ when $0 \neq \alpha \neq \beta$ are called *elliptic functions*, and for this reason we call $y^2 = x(x-\alpha)(x-\beta)$ an *elliptic curve*. (The elliptic curves also include some fourth-degree curves, such as $y^2 = 1 + x^4$. This curve

can be parameterized by elliptic functions but, as we know from Exercise 4.7.5*, not by rational functions.)

Despite this, it is not that hard to find rational *points* on a cubic curve $\mathcal{K}$, provided the equation of $\mathcal{K}$ has rational coefficients. A simple argument, like the one given for quadratic curves in Section 4.5, shows that a line through two rational points on a cubic $\mathcal{K}$ with rational coefficients meets $\mathcal{K}$ in a third rational point. It is not even necessary to find two rational points to get started; one is enough, because the tangent at one rational point P effectively "meets $\mathcal{K}$ twice" at P, and hence its other intersection with $\mathcal{K}$ is also rational.

The algebraic equivalent of the tangent construction was actually used by Diophantus. In his Problem 18 of Book VI he uses the obvious solution $x = 0$, $y = 1$ of the equation $y^2 = x^3 - 3x^2 + 3x + 1$ to find the nonobvious solution $x = 21/4$, $y = 71/8$, by substituting $y = 3x/2 + 1$. The latter equation represents the tangent at $(0, 1)$.

Fermat took up Diophantus' tangent method to find rational solutions of cubic equations, and Newton pointed out the related method (the "chord construction") of drawing a line through two rational points to find a third. Finally, in 1922 Louis Mordell proved that these two methods suffice to find *all* rational points on a cubic curve, provided finitely many rational points are given. Mordell's theorem is difficult and deeply dependent on elliptic functions; nevertheless it shows that elliptic curves are near relatives of quadratic curves when it comes to finding rational points.

For this reason and because they are related to many classical problems, elliptic curves have been intensely studied over recent decades. The proof of Fermat's last theorem is the most spectacular result of this study so far, but others can be expected. The book of Koblitz (1985) is an attractive introduction to the subject, organized around an ancient problem that is still not solved: which integers are the areas of rational right-angled triangles?