CHAPTER 1

INTRODUCTION

As long as algebra and geometry have been separated, their progress have been slow and their uses limited; but when these two sciences have been united, they have lent each mutual forces, and have marched together towards perfection.

Joseph-Louis Lagrange, 1795

The main goal of this book is to study $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$, i.e., the natural numbers, the integers, and the rational numbers:

$$\begin{aligned}\mathbb{N} &= \{1, 2, 3, \ldots\},\\ \mathbb{Z} &= \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\},\\ \mathbb{Q} &= \left\{\frac{m}{n} : m, n \in \mathbb{Z},\ n \neq 0\right\}.\end{aligned}$$

In the next chapter, we will be much more careful defining these sets using *axioms*, but, for now, we appeal to our intuition of the properties that these numbers satisfy. One can study these sets of numbers from their intrinsic properties, and much can be gained from such an endeavor, but in this book we study these sets from the point of view of their interaction with geometric objects (graphs of polynomials, lines in the plane, conics, elliptic curves, etc.).

Our generic approach will be as follows: we will define a geometric object G and then we will try to find all the points in the geometric object with coordinates in $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{Q}$, which we will denote by $G(\mathbb{N})$, $G(\mathbb{Z})$, and $G(\mathbb{Q})$, respectively. As we attempt to find the natural, integral, or rational points, we will develop the theory that is usually called "elementary number theory". Our approach will use the problem of finding arithmetic points on a geometric object as the motivation for the definitions and techniques of elementary number theory. Let us begin with our first example.

1.1. Roots of Polynomials

We begin this section with a discussion about polynomials and, in particular, which polynomials have roots in a given number system. Roots of polynomials will be treated in more detail in Part 1, and in particular in Section 2.8. We will also discuss polynomials (as a *ring*) in Section 5.5.

A polynomial $p(x)$ is an expression of the form

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $n \geq 0$ is a non-negative integer and $a_0, a_1, \ldots, a_n$ are constants (in $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$, for example). By a *polynomial equation*, we mean an equation that can be expressed in the form $p(x) = 0$, for some polynomial $p(x)$. A *root* of the polynomial equation $p(x) = 0$ is a number α such that $p(\alpha) = 0$.

For humans, it is natural to work with the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ as we often need to count things in our daily routine. However, as soon as we try to solve the simplest linear polynomial equations using only natural numbers, we run into problems. An equation of the form

$$3 + x = 5 \tag{1.1}$$

has a (unique!) solution in $\mathbb{N}$, namely $x = 2$. But the similar equation

$$5 + x = 3 \tag{1.2}$$

has no solutions in $\mathbb{N}$, since $5 + x > 5 > 3$, for any $x \in \mathbb{N}$. Indeed, if a and b are natural numbers, then an equation $a + x = b$ has a solution in $\mathbb{N}$ if and only if $a < b$. Thus, in order to solve (1.2), we need to augment $\mathbb{N}$ to include all numbers of the form $-n$, where $n \in \mathbb{N}$. Notice that we also need to include 0 to be able to solve an equation of the form $5 + x = 5$. Thus, we define $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ and every equation of the form

$$a + x = b, \tag{1.3}$$

where $a, b \in \mathbb{Z}$, has a (unique!) solution $x = b - a$ in $\mathbb{Z}$. The integers, however, are not enough to solve an equation of the form

$$5x = 3 \tag{1.4}$$

as there is no integer n such that $5n = 3$ (indeed, the number 3 is prime and its only positive divisors are 1 and 3). More generally, an equation of the form $ax + b = 0$, with $a, b \in \mathbb{Z}$ and $a \neq 0$, has solutions in $\mathbb{Z}$ if and only if a is a divisor of b. In order to be able to solve all equations of the form $ax + b = 0$, we need to augment $\mathbb{Z}$ to be a number system such that every non-zero number has a *multiplicative inverse*. And so, we define $\mathbb{Q} = \{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\}$. Now every linear equation $ax + b = c$, with $a, b, c \in \mathbb{Q}$ and $a \neq 0$, has a unique solution $x = \frac{c-b}{a} \in \mathbb{Q}$.

How about quadratic polynomials? Do they all have roots in $\mathbb{Q}$? Of course not. For instance, the polynomial $x^2 - 2 = 0$ does not have any rational roots because $\sqrt{2}$ is not a rational number. (In order to rigorously prove that $\sqrt{2} \notin \mathbb{Q}$ we will first need to prove the fundamental theorem of arithmetic! See Theorems 2.10.2 and 2.10.6 and Section 2.10.1.) We usually represent numbers such as $\sqrt{2}$ by their decimal expansion, i.e., $\sqrt{2} = 1.41421356237309\ldots$. The decimal expansion of a

rational number has a period, i.e., the expansion eventually repeats a given pattern of finitely many digits (why?). For example,

$$\frac{13}{17} = 0.764705882352941176470588235294117647058823529411\ldots 7647058823529411\ldots.$$

Conversely, any decimal expansion that has a period represents a rational number (see Section 8.9.2). The expansion of $\sqrt{2}$ has no period, as we have already mentioned that $\sqrt{2} \notin \mathbb{Q}$. In order to be able to solve quadratic equations (and other higher-degree polynomial equations), one can augment $\mathbb{Q}$ to include all decimal expansions and not only those that are periodic. This leads to an informal definition of the real numbers:

$$\mathbb{R} = \{\text{set of all decimal expansions}\},$$

with the usual identification of decimals with "trailing nines"; e.g., the expansion $0.9999\ldots$, with infinitely many nines, is equal to the decimal expansion $1 = 1.0000\ldots$ (see Exercise 1.8.1).

Unfortunately, not all quadratic polynomial equations $ax^2 + bx + c = 0$, with $a, b, c \in \mathbb{R}$ and $a \neq 0$, have a solution in $\mathbb{R}$. In fact, $ax^2 + bx + c = 0$, with $a, b, c \in \mathbb{R}$ and $a \neq 0$, has a solution in $\mathbb{R}$ if and only if $b^2 - 4ac \geq 0$. Similarly, there are higher-degree polynomials with no roots in $\mathbb{R}$. For instance, the polynomial equation $x^4 + x^3 + x^2 + x + 1 = 0$ has no real roots.

In order to ameliorate the "shortcomings" of $\mathbb{R}$, we would like to augment $\mathbb{R}$ so that, at least, all quadratic polynomials have a root. In order to accomplish this, it is sufficient to add a square root of -1 to $\mathbb{R}$, which we shall denote by i, an *imaginary number* such that $i^2 = -1$. Indeed, a polynomial $p(x) = ax^2 + bx + c = 0$, with $a, b, c \in \mathbb{R}$ and $a \neq 0$, with $b^2 - 4ac \geq 0$ has real roots

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

and if $b^2 - 4ac < 0$, then $p(x) = 0$ has roots

$$x = \frac{-b \pm i\sqrt{|b^2 - 4ac|}}{2a}.$$

Therefore, if we define the complex numbers as

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R},\ i^2 = -1\},$$

then all linear and quadratic polynomials with coefficients in $\mathbb{C}$ have roots in $\mathbb{C}$ (the reader needs to verify that every complex number $\alpha \in \mathbb{C}$ has a square root $\sqrt{\alpha}$ also in $\mathbb{C}$; see Exercise 1.8.6). Perhaps one of the most surprising and beautiful theorems in algebra is that, in fact, *every* non-constant polynomial (of arbitrary degree ≥ 1) with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$. This is known as the fundamental theorem of algebra.

Theorem 1.1.1 (Fundamental theorem of algebra)**.** *Let $p(x)$ be a polynomial of degree ≥ 1 with coefficients in $\mathbb{C}$. Then, there is $\alpha \in \mathbb{C}$ such that $p(\alpha) = 0$.*

For example, let $p(x) = x^4 + x^3 + x^2 + x + 1$. As we mentioned above, $p(x)$ is a polynomial that has no real roots. The number $\alpha = \cos(\frac{2\pi}{5}) + i\sin(\frac{2\pi}{5})$ is a

complex root of $p(x)$. Indeed, by Euler's formula

$$e^{ix} = \cos x + i \cdot \sin x,$$

we have that $\alpha = e^{2\pi i/5}$. Moreover,

$$x^4 + x^3 + x^2 + x + 1 = \frac{x^5 - 1}{x - 1},$$

and $\alpha^5 - 1 = (e^{2\pi i/5})^5 - 1 = e^{2\pi i} - 1 = 1 - 1 = 0$. Thus, $p(\alpha) = 0$ as well.

Complex numbers are fascinating in their own right, and there is a whole area of mathematics dedicated to the study of $\mathbb{C}$ and complex-valued functions, namely the area known as complex analysis. Here, however, we are (mostly) interested in, and shall concentrate on, the study of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. Let us try to find out when a polynomial with integer coefficients has a rational root.

Example 1.1.2. Let $p(x) = 3x^3 - 44x^2 - 257x + 190$ be a polynomial. We would like to find the natural ($\mathbb{N}$), integral ($\mathbb{Z}$), or rational ($\mathbb{Q}$) roots of $p(x)$; i.e., we want to find those natural, integral, or rational numbers x that satisfy $p(x) = 0$. Suppose that the natural number $n \in \mathbb{N}$ is a root of $p(x)$. Then,

$$p(n) = 3n^3 - 44n^2 - 257n + 190 = 0,$$

and we may rewrite this expression as $n(3n^2 - 44n - 257) = -190$. Since n is a natural number, the number $3n^2 - 44n - 257$ is an integer (not necessarily in $\mathbb{N}$) and we may conclude that n would necessarily be a divisor of -190. The list of natural divisors of -190 is $L = \{1, 2, 5, 10, 19, 38, 95, 190\}$. Thus, we can try to see whether any of these numbers $n \in L$ is a root of $p(x)$ by calculating $p(n)$. After carrying this out, we find that the only natural number that is a root of $p(x)$ is $n = 19$.

Are there any integral roots of $p(x)$ that are not natural numbers? If $n \in \mathbb{Z}$ and $p(n) = 0$, the expression $n(3n^2 - 44n - 257) = -190$ is still valid, and we may also conclude that n must be a divisor of -190. The *integer* divisors of -190 are those in the list $L' = \{\pm 1, \pm 2, \pm 5, \pm 10, \pm 19, \pm 38, \pm 95, \pm 190\}$. Since we have already checked that the only natural root is 19, we only need to check whether any of the negative divisors is a root. In this manner, we find that $n = 19$ and $n = -5$ are the only integral roots of $p(x)$.

Finally, we wish to find out whether $p(x)$ has any rational roots. Since we know that 19 and -5 are roots, we deduce that $f(x) = (x + 5)(x - 19)$ is a factor of $p(x)$ as polynomials (here we are using the so-called root theorem, Corollary 5.5.15). We may divide $p(x)$ by $f(x)$ to find a third linear factor, and therefore the value of the third root of $p(x)$. Instead, we shall approach this using a divisibility method that works more generally. Suppose $\frac{m}{n} \in \mathbb{Q}$ is a reduced fraction (i.e., m and n share no common divisors) and it is a root of $p(x)$. Then,

$$p\left(\frac{m}{n}\right) = 3\left(\frac{m}{n}\right)^3 - 44\left(\frac{m}{n}\right)^2 - 257\frac{m}{n} + 190 = 0.$$

If we multiply this expression by n^3, we obtain

$$3m^3 - 44m^2n - 257mn^2 + 190n^3 = 0.$$

This expression can be rewritten as $m(3m^2 - 44mn - 257n^2) = -190n^3$. This is an equality of integer numbers and we may deduce that m is a divisor of $-190n^3$.

Since m and n share no common divisors, it follows that m is a divisor of -190; i.e., $m \in L'$ with L' as defined above. The same displayed expression can be rewritten as $3m^3 = n(44m^2 + 257mn - 190n^2)$ and, once again, we may deduce a divisibility property. In this case, we deduce that n is a divisor of $3m^3$. Since m and n share no common divisors, we conclude that n is an integer divisor of 3 and so $n \in \{\pm 1, \pm 3\}$. Therefore, if $m/n \in \mathbb{Q}$ is a rational root of $p(x)$, we have shown that $m \in L'$ and $n \in \{\pm 1, \pm 3\}$. Now it is a matter of checking whether any of these rational numbers are actually roots, and we find that $\frac{m}{n} = \frac{2}{3}$ is indeed the third root we were looking for. Hence, the roots of $p(x)$ are $19 \in \mathbb{N}$, $-5 \in \mathbb{Z}$, and $\frac{2}{3} \in \mathbb{Q}$.

The previous example motivates some of our first definitions and theorems in the book (in Part 1). In the course of finding the roots of a polynomial, we have relied heavily upon the theory of divisibility of natural and integer numbers (and we alluded to divisibility of polynomials too). It is likely that the reader is perfectly comfortable with many of the steps in the example, but one needs to carefully prove some of them. For instance, at some point we used the following fact:

- If m, n, a, b are integers such that $ma = nb$ and m and n share no common factors (i.e., $\gcd(m, n) = 1$), then m is a divisor of b and n is a divisor of a.

Although this fact may be intuitively true, we need a proof! In order to provide a proof, we will need to establish first a number of basic facts about divisibility (see Corollary 2.7.6). But, for now, let us see how our next two examples motivate the study of the greatest common divisor of two integers.

1.2. Lines

In this section we discuss examples of the most basic 1-dimensional object: a line in the plane. We will come back to studying points on a line in detail in Section 2.9.

Example 1.2.1. Let $L : 5x + 17y = 1$ be a line in the plane. See Figure 1.1. There are infinitely many rational points in this line, and they can be found by solving for one of the variables. For example, we may write

$$y = \frac{1 - 5x}{17},$$

and it follows that $Q = \left(x_0, \frac{1-5x_0}{17}\right)$ is a point in L with rational coordinates, for each rational number x_0. In fact, every rational point Q in L is of this form. For instance, the points $(0, 1/17)$ and $(1, -4/17)$ are in L. Are there any points $(x_0, y_0) \in L$ with integer coefficients, with $x_0, y_0 \in \mathbb{Z}$? A quick search for points (using trial and error) reveals at least one point: $(7, -2)$.

Are there more? Yes, in fact, there are infinitely many integral points of the form $P_k = (7 + 17k, -2 - 5k)$ where $k \in \mathbb{Z}$. Let us check that these points belong to L:

$$5(7 + 17k) + 17(-2 - 5k) = 35 + 5 \cdot 17k - 34 - 5 \cdot 17k = 35 - 34 = 1.$$

Interestingly, the points $\{P_k : k \in \mathbb{Z}\}$ are *all* the integral points on L, but this is not so easy to prove (try!). This will be shown in Theorem 2.9.4.

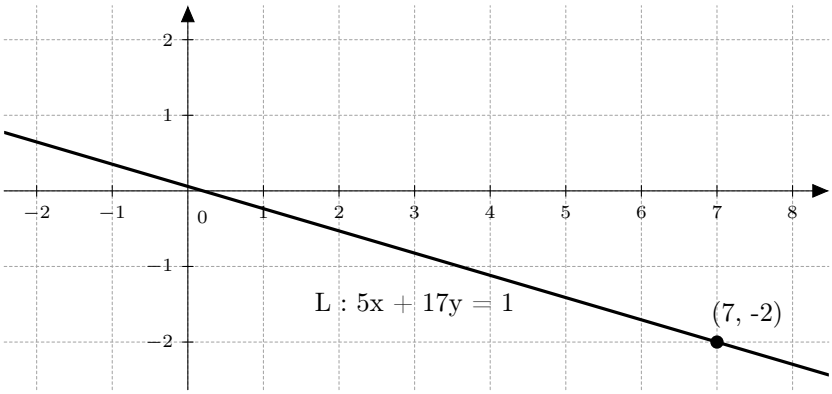

Figure 1.1. The line $5x + 17y = 1$ passes through infinitely many integral points.

Example 1.2.2. Let L' be the line in the plane with equation $5x + 15y = 1$ (see Figure 1.2).

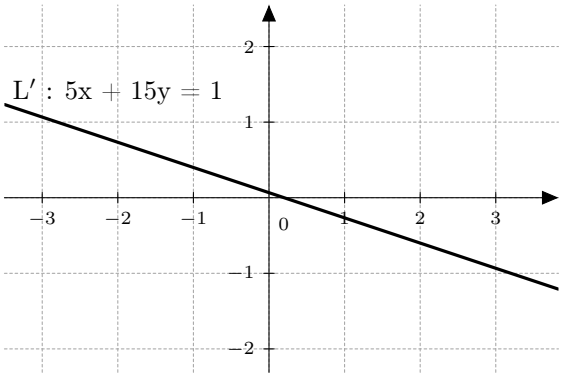

Figure 1.2. The line $5x + 15y = 1$ does not pass through any integral point.

As in our previous example, there are infinitely many rational points on L' given by $(x_0, \frac{1-5x_0}{15})$ for any $x_0 \in \mathbb{Q}$. Are there any integral points on L'? It turns out that there are none. Suppose m and n are integers with $5m + 15n = 1$. Then, $5(m + 3n) = 1$ and we have reached a contradiction because this equation implies that 1 has a non-trivial factorization into integers (other than $1 = 1 \cdot 1 = (-1)(-1)$). Another way to see this is that, in the integers, 5 is not a divisor of 1 (however, the number 5 is a divisor of 1 in the rational numbers: $1 = 5 \cdot \frac{1}{5}$).

Examples 1.2.1 and 1.2.2 show two lines L and L' that behave very differently when we look for integral points on them. Why is their behavior so different? The reason, as we shall see, is that $\gcd(5, 17) = 1$ while $\gcd(5, 15) = 5$. Using an argument similar to that in Example 1.2.2, one can show that a line $L'' : ax+by = c$, with $\gcd(a, b) = d$ and d not a divisor of c, will have no integral points. Indeed, if $m, n \in \mathbb{Z}$ satisfy $am + bn = c$, then d would be a divisor of c and that is a

contradiction to one of our assumptions. However, if $\gcd(a, b) = 1$, why should there be integral points on L''? For example, consider $L'' : 1234x + 5007y = 1$. Are there integral points on L''? The greatest common divisor of 1234 and 5007 is 1 and, as we shall see, this implies the existence of integral points and, moreover, we will describe an efficient algorithm to find these points (see Sections 2.6 and 2.7). Here is one such point $P = (-1481, 365)$:

$$1234(-1481) + 365(5007) = -1827554 + 1827555 = 1.$$

All other integral points are of the form $(-1481+5007k, 365-1234k)$ for any $k \in \mathbb{Z}$.

1.3. Quadratic Equations and Conic Sections

In this section we discuss several examples of rational and integral points on quadratic equations in two variables, i.e., equations of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

where a, b, c, d, e, f are integers and a, b, or c is non-zero. When the graph of a quadratic equation is *smooth* (non-singular; see Section 15.1.5), we call them conic sections (because they arise as sections of cones; see Figure 9.5). We will discuss quadratic equations and conic sections at length in Part 2.

Figure 1.3. Muhammad ibn Musa al-Khwarizmi (c. 780 – c. 850) was a Persian mathematician, astronomer, and geographer. His treaty on algebra contained the first systematic treatment of linear and quadratic equations, including the first demonstration of the "completing the square" method. Image source: Wikimedia Commons.

In our next example, we find rational points on a conic section (a hyperbola, in this case).

Example 1.3.1. Let $C : x^2 - 7y^2 = 1$ be a hyperbola in the plane. Can we find all the rational points on C? Yes, and we will do so using a little bit of geometry. First, notice that there are two (integral) points which are easily found, namely $(\pm 1, 0)$. If we trace a line L that goes through $P = (1, 0)$, it will intersect the hyperbola

at exactly two points: the point P and a second point Q (since C is given by a quadratic equation, the intersection with a line is formed by either no points or two points). Let us find the second point of intersection, Q, in terms of the slope of L. The equation of L is given by

$$L : y - 0 = m(x - 1),$$

where m is the slope of L.

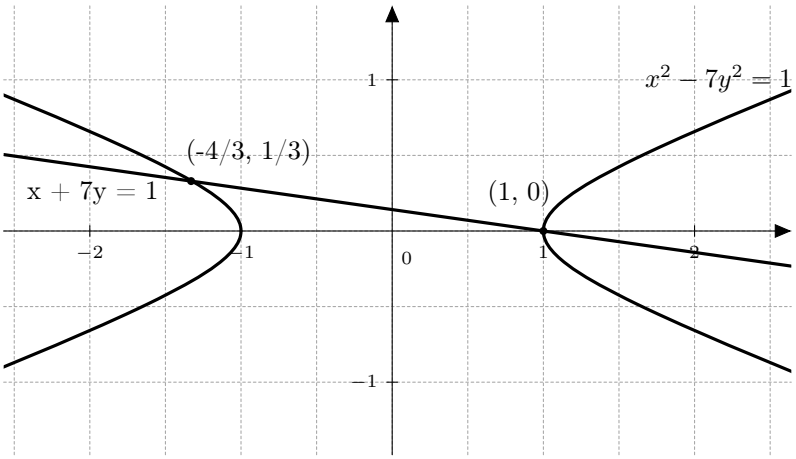

Figure 1.4. The hyperbola $x^2 - 7y^2 = 1$ passes through infinitely many rational points and also through infinitely many integral points.

Notice that L passes through $P = (1, 0)$, as desired. Now we may find the intersection of L and C by solving the system:

$$\begin{cases} x^2 - 7y^2 = 1, \\ y = m(x - 1). \end{cases}$$

Plugging the equation of L into the equation for C, we obtain

$$1 = x^2 - 7(m(x - 1))^2 = (1 - 7m^2)x^2 + 14m^2x - 7m^2,$$

or, equivalently, $(1-7m^2)x^2+14m^2x-(1+7m^2) = 0$. Now we can use the quadratic formula to solve for x:

$$\begin{aligned} x &= \frac{-14m^2 \pm \sqrt{14^2m^4 - 4(1 - 7m^2)(-(1 + 7m^2))}}{2(1 - 7m^2)} \\ &= \frac{-14m^2 \pm \sqrt{14^2m^4 + 4(1 - 7^2m^4)}}{2(1 - 7m^2)} \\ &= \frac{-14m^2 \pm \sqrt{4}}{2(1 - 7m^2)} = \frac{-14m^2 \pm 2}{2(1 - 7m^2)} \\ &= \frac{-7m^2 \pm 1}{1 - 7m^2} = \begin{cases} 1 & \text{or} \\ \frac{7m^2+1}{7m^2-1}. \end{cases} \end{aligned}$$

As we expected, $x = 1$ is a solution (since L passes through $P = (1, 0)$). The second point of intersection, Q_m, has x-coordinate $x = (7m^2 + 1)/(7m^2 - 1)$. The y-coordinate of Q_m is given by

$$y = m(x - 1) = m\left(\frac{7m^2 + 1}{7m^2 - 1} - 1\right) = \frac{2m}{7m^2 - 1}.$$

Thus, the point $Q_m = \left(\frac{7m^2+1}{7m^2-1}, \frac{2m}{7m^2-1}\right)$ is a rational point on C for every rational slope $m \in \mathbb{Q}$. For instance, when $m = -1/7$ (see Figure 1.4), the point $Q_1 = (-\frac{4}{3}, \frac{1}{3})$ is in C:

$$\left(-\frac{4}{3}\right)^2 - 7\left(\frac{1}{3}\right)^2 = \frac{16 - 7}{9} = \frac{9}{9} = 1.$$

It is not difficult to see that this construction yields *all* the rational points on C; i.e., $C(\mathbb{Q}) = \{Q_m : m \in \mathbb{Q}\}$. Indeed, if Q' is a rational point on C and the line $\overline{PQ'}$ has slope m, then $Q' = Q_m$ (notice that the slope cannot be infinite, as there is only one point on C with $x = 1$, namely P).

We have found all the rational points on $C : x^2 - 7y^2 = 1$. Are there integral points on C? If so, how many? It turns out that, in this particular case, there are infinitely many integral points $(m, n) \in C(\mathbb{Z})$ and these points are intimately related with the rational approximations of $\sqrt{7}$. More concretely, if (m, n) is an integral point on C, then $\frac{m}{n}$ is a (very) good rational approximation of $\sqrt{7}$. Indeed,

$$m^2 - 7n^2 = 1$$

implies that

$$7 = \frac{m^2}{n^2} - \frac{1}{n^2} = \left(\frac{m}{n}\right)^2 - \frac{1}{n^2},$$

so that $|7 - (\frac{m}{n})^2| = \frac{1}{n^2}$. For instance, $(m, n) = (8, 3)$ is an integral point on C, and $\frac{8}{3} = 2.666\ldots$ while $\sqrt{7} = 2.645751\ldots$. We will explain later on that, once we have one rational solution (m, n), there is a method to find infinitely many solutions, by squaring the number $m + n\sqrt{7}$ (see Section 14.3.1). More concretely, if (m, n) is an integral point on C and $(m + n\sqrt{7})^2 = a + b\sqrt{7}$, then (a, b) is another integral point on C. In our case,

$$(8 + 3\sqrt{7})^2 = 64 + 48\sqrt{7} + 63 = 127 + 48\sqrt{7},$$

and we can verify that $(127, 48)$ is another point on C. Also, $\frac{127}{48} = 2.6458333\ldots$ is another approximation of $\sqrt{7}$ (see Chapter 13 and Theorem 14.2.3).

Example 1.3.2. Let us now consider the hyperbola $C' : x^2 - 7y^2 = 3$. Are there any integral points? We will show that, in fact, this hyperbola does not have any integral points. Let us assume, for a contradiction, that (m, n) is an integral point on C. It follows that $m^2 = 3+7n^2$ and, in particular, the remainder when we divide m^2 by 7 is 3. This is impossible, as the only remainders that occur when we divide a *perfect square* by 7 are 0, 1, 2, or 4, and 3, 5 and 6 never occur as remainders. Let us prove this last claim.

Indeed, every number $m \in \mathbb{Z}$ has a remainder of $0, 1, 2, 3, 4, 5$, or 6 when we divide by 7. In other words, we can always write $m = 7k + r$, where $k \in \mathbb{Z}$ and

$r = 0, 1, 2, 3, 4, 5$, or 6. Let us see what happens when we square $m = 7k + r$, for each possible remainder r:

$$(7k+0)^2 = 49k^2 = 7(7k^2) + 0,$$
$$(7k+1)^2 = 49k^2 + 14k + 1 = 7(7k^2 + 2k) + 1,$$
$$(7k+2)^2 = 49k^2 + 28k + 4 = 7(7k^2 + 4k) + 4,$$
$$(7k+3)^2 = 49k^2 + 42k + 9 = 7(7k^2 + 6k) + 9 = 7(7k^2 + 6k + 1) + 2,$$
$$(7k+4)^2 = 49k^2 + 56k + 16 = 7(7k^2 + 8k) + 16 = 7(7k^2 + 8k + 2) + 2,$$
$$(7k+5)^2 = 49k^2 + 70k + 25 = 7(7k^2 + 10k) + 25 = 7(7k^2 + 10k + 3) + 4,$$
$$(7k+6)^2 = 49k^2 + 84k + 36 = 7(7k^2 + 12k) + 36 = 7(7k^2 + 12k + 5) + 1.$$

Thus, we have just shown that the remainder of $m^2 = (7k + r)^2$ when we divide by 7 is 0, 1, 2, or 4, and never 3, 5, or 6. Hence, $m^2 = 3 + 7n^2$ is impossible and C does not have any integral points. Similarly, C does not have any rational points either. Suppose $(\frac{m}{a}, \frac{n}{b})$ is a rational point. Then $(\frac{m}{a})^2 - 7(\frac{n}{b})^2 = 3$ and it follows that $(mb)^2 - 7(na)^2 = 3(ab)^2$, or, equivalently, $(mb)^2 = 3(ab)^2 + 7(na)^2$. Suppose that the remainder of dividing $(ab)^2$ by 7 is r; i.e., there is a $k \in \mathbb{Z}$ such that $(ab)^2 = 7k + r$. Then, as before, $r = 0, 1, 2$, or 4 and

$$(mb)^2 = 3(7k + r) + 7(na)^2 = 3r + 7((na)^2 + 3k).$$

In particular, $3r = 0, 3, 6$, or 12. If $3r = 12$, then we may write $(mb)^2 = 5 + 7((na)^2 + 3k + 1)$. Hence, the remainder of dividing $(mb)^2$ by 7 is 0, 3, 6, or 5. We have shown above that it cannot be $3, 5$, or 6, so it must be 0 (so that $3r = 0$. Hence, $(mb)^2 = 7((na)^2 + 3k)$, or $A^2 = 7B^2$, where $A = mb$ and $B = (na)^2 + 3k$ are integers. However, the equation $A^2 = 7B^2$ has no solutions in the integers as the left-hand side is a perfect square but the right-hand side is not a square (a consequence of the fundamental theorem of arithmetic, Theorem 2.10.6; see two paragraphs below).

In Example 1.3.1 we have seen that a little bit of geometry can go a long way. The trick of intersecting a curve with a line passing through a known point is very useful. We will see similar tricks in the examples that follow. Another theme that Example 1.3.1 has introduced is that of the approximation of irrational numbers by rationals (e.g., $\sqrt{7} \approx \frac{127}{48}$), usually referred to as diophantine approximation.

In Example 1.3.2 we have claimed that $A^2 = 7B^2$ is impossible, for $A, B \in \mathbb{N}$. This fact relies on the so-called fundamental theorem of arithmetic: every natural number has a unique factorization as a product of prime numbers. Also in Example 1.3.2, we have seen for the first time that working with the remainders of long division can be a very effective technique. This tool will lead us to the study of *congruences* modulo an integer (see Chapter 4). In particular, we were interested in the remainder left out when dividing a perfect square n^2 by a fixed number m. This will lead us to the study of *quadratic residues* and Gauss's law of quadratic reciprocity—one of the theorems in all of mathematics with the largest number of known distinct proofs (Gauss alone published six different proofs; there are now over 200 published proofs). Quadratic congruences and quadratic reciprocity will be dealt with in Chapter 10.

1.4. Cubic Equations and Elliptic Curves

In this section we discuss our first examples of cubic equations in the plane, i.e., equations in two variables of the form

$$ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + jy + k = 0,$$

for some integers $a, b, c, \ldots, k \in \mathbb{Z}$ such that not all of a, b, c, d are zero. Cubic equations will be studied in Part 3.

An *elliptic curve* is a smooth cubic curve in the plane, which is also smooth "at infinity", and such that it contains at least one rational point (we will discuss elliptic curves in Chapter 16). Smooth means that the curve has a well-defined tangent line at every point. We will not describe in detail here what is the meaning of the smoothness at infinity condition (see Section 15.1.5 instead). It is sufficient to say that, after an appropriate change of variables, every elliptic curve (defined over $\mathbb{Q}$) can be written in the simpler form

$$y^2 = x^3 + Ax + B$$

where $A, B \in \mathbb{Z}$ and the polynomial $x^3 + Ax + B$ has no repeated roots (which, in turn, is equivalent to $4A^3 + 27B^2 \neq 0$). For example, we have drawn the graph of the elliptic curve $y^2 = x^3 + 1$ in Figure 1.5.

Example 1.4.1. Let E be the elliptic curve given by the equation $y^2 = x^3 + 1$. A quick inspection for points reveals two integral points $P = (-1, 0)$ and $Q = (0, 1)$. By symmetry, there is one additional point $Q' = (0, -1)$. Now, in order to find new points, we may use a trick we have already seen. Let L be the line that goes through P and Q. With a little bit of basic plane geometry, we find an equation for $L : y = x - 1$ (see Exercise 1.8.8). In order to find the intersection points of E and L, we need to solve the system

$$\begin{cases} y^2 = x^3 + 1, \\ y = x + 1. \end{cases}$$

Thus, we plug the equation for L into the equation for E and obtain a polynomial of degree 3 whose roots are the x-coordinates of the points of intersection of E and L:

$$(x + 1)^2 = x^3 + 1, \quad \text{or} \quad x^3 - x^2 - 2x = x(x^2 - x - 2) = 0.$$

The roots are $x = -1, 0, 2$, and the corresponding y-coordinates are $0, 1, 3$, respectively. Hence, we have found a new point $R = (2, 3)$ on E, with natural coordinates. By symmetry, there is an additional point $(2, -3)$ on E.

So far, we have found one natural point, $(2, 3)$, and four additional integral points, $(-1, 0)$, $(0, \pm 1)$, and $(2, -3)$. It turns out that these are all the *rational* points on E, but this is fairly hard to prove.

Example 1.4.2. Let E' be the elliptic curve given by the equation $y^2 = x^3 - 2$. A quick inspection reveals one integral point, $P = (3, 5)$, but no other integral point is easily found. We can modify our previous geometric trick by finding the line L that is *tangent* to E' at P. A little bit of calculus (e.g., implicit differentiation; see

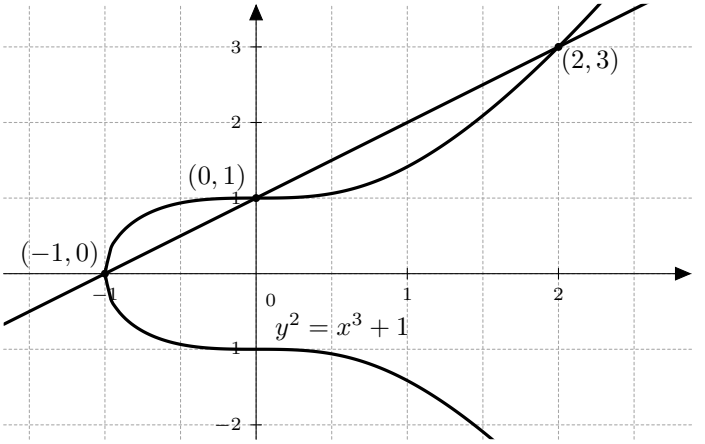

Figure 1.5. The elliptic curve $y^2 = x^3 + 1$.

Exercise 1.8.20) yields that the slope of a tangent line to E at a point $(x, y) \in E$ is given by

$$\frac{dy}{dx} = \frac{3x^2}{2y}.$$

In our case, L has slope $27/10$ and passes through $P = (3, 5)$, so it is given by the equation $L : y = \frac{27}{10}(x - 3) + 5$. Now we can find the intersection points of L and E' by solving the system

$$\begin{cases} y^2 = x^3 - 2, \\ y = \frac{27}{10}(x - 3) + 5. \end{cases}$$

Plugging the equation for L into the equation for E' yields a polynomial equation

$$\left(\frac{27}{10}(x - 3) + 5\right)^2 = x^3 - 2,$$

or, equivalently,

$$x^3 - \frac{729}{100}x^2 + \frac{837}{50}x - \frac{1161}{100} = 0.$$

We know that $x = 3$ is a root of this polynomial (and, in fact, it must be a double-root, because L is *tangent* to E' at P). Thus, this polynomial factors as $(x - 3)^2(x - \alpha) = 0$. Hence, we can find the value of α and this turns out to be $\alpha = \frac{129}{100}$. In particular, the x-coordinates of the points of intersection of L and E' are 3 and $\frac{129}{100}$, and their y-coordinates are 5 and $-\frac{383}{100}$, respectively. Hence, we have found a new *rational* point on E', namely $Q = (\frac{129}{100}, -\frac{383}{100})$. By symmetry of the graph of E' with respect to the y-axis, there is an additional point $Q' = (\frac{129}{100}, \frac{383}{100})$.

This construction of a rational point can be repeated to find other points. For instance, we can trace the line L' that goes through P and Q'. This line will intersect E' at a third rational point. We leave it to the reader to verify that the points of intersection of E' and L' are P, Q' and

$$Q'' = \left(\frac{164323}{29241}, \frac{66234835}{5000211}\right).$$

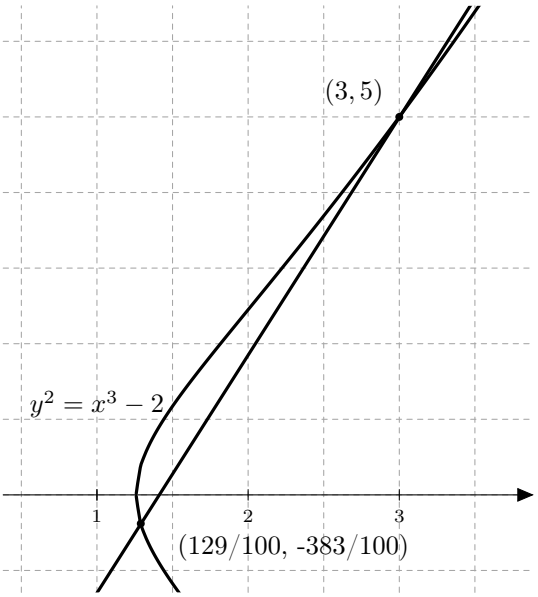

Figure 1.6. The elliptic curve $y^2 = x^3 - 2$.

The previous examples illustrate the method of *chords and tangents* that can be used on cubic curves to find new rational points (see Section 16.3). The most important theorem in the theory of elliptic curves is the following result, proved by Louis Mordell, and vastly generalized by André Weil (see Figure 1.7). The so-called Mordell–Weil theorem says that there is a finite set of rational points S such that every other rational point can be obtained from the points in S, using the method of chords and tangents.

Theorem 1.4.3 (Mordell–Weil theorem)**.** *Let E be an elliptic curve (a smooth cubic curve, together with a given rational point $\mathcal{O}$). Then, there is a set formed by finitely many rational points $P_1, \ldots, P_n$ on E such that if R is any other rational point on E, then R can be obtained from $P_1, \ldots, P_n$ using the method of chords and tangents.*

Example 1.4.4. All the rational points on $E : y^2 = x^3 + 1$ can be generated from the point $P_1 = (2, 3)$ using chords and tangents. In this case, there are only five rational points on E (plus a point at "infinity").

The rational points on $E' : y^2 = x^3 - 2$ are generated from the point $P_1 = (3, 5)$ using chords and tangents. In this case, however, the curve has infinitely many distinct rational points.

The rational points on the curve $E'' : y^2 + y = x^3 - 7x + 6$ are generated using three points $P_1 = (1, 0)$, $P_2 = (2, 0)$, and $P_3 = (0, -3)$. These three points generate infinitely many distinct rational points on E''.

Figure 1.7. Louis Mordell (1888–1972) and André Weil (1906–1998). Images author: Konrad Jacobs (Erlangen). Source: Archives of the Mathematisches Forschungsinstitut Oberwolfach.

1.5. Curves of Higher Degree

By a curve of *higher* degree we refer to equations of the form

$$f(x, y) = 0,$$

where $f(x, y)$ is a polynomial in two variables, with integer coefficients, and such that the largest degree of a monomial is greater than or equal to 4 (here we define the degree of a monomial $x^a y^b$ as $a + b$). For instance, the curves of the form

$$y^2 = p(x),$$

with $p(x)$ a polynomial of degree ≥ 4, are called hyperelliptic curves.

Example 1.5.1. The following is problem 17 in Book VI of Diophantus's *Arithmetica*:

> *Find three squares which when added give a square, and such that the first one is the side of the second, and the second is the side of the third.*

Let A, B, C be integers, and let A^2, B^2, C^2 be the squares mentioned in the problem. Then, $Y^2 = A^2 + B^2 + C^2$, for some $Y \in \mathbb{Z}$, and the first one is the side of the second (so $A^2 = B$) and the second one is the side of the third (so $B^2 = C$). It follows that if $A = x$, then $B = x^2$ and $C = x^4$. Therefore, we are trying to find x and Y integers such that

$$Y^2 = x^2 + x^4 + x^8.$$

If we exclude the unique solution with $x = 0$, i.e., $(0, 0)$, then we can write $Y = xy$, and therefore we are looking for a rational point on the curve

$$x^2y^2 = x^2 + x^4 + x^8,$$

with $x \neq 0$. Thus, we can divide through by x and simplify the equation to

$$\mathcal{C} : y^2 = 1 + x^2 + x^6,$$

and we are looking for all the rational solutions with $x \neq 0$. In his work, Diophantus finds one rational solution of $\mathcal{C}$, namely $(x, y) = (1/2, 9/8)$, which corresponds to $(x, Y) = (1/2, 9/16)$, and therefore

$$A = \frac{1}{2},\ B = \frac{1}{4},\ \text{ and }\ C = \frac{1}{16},$$

so that

$$A^2 + B^2 + C^2 = \frac{1}{4} + \frac{1}{16} + \frac{1}{256} = \frac{81}{256} = \left(\frac{9}{16}\right)^2.$$

A natural question arises: are there any other solutions to the problem? In other words, are there any other rational points in $\mathcal{C}$? In 1998, Joseph Wetherell showed in his Ph.D. thesis [**Wet98**] that the only rational points on $\mathcal{C}$ are precisely $(0, \pm 1)$, $(1/2, \pm 9/8)$, and $(-1/2, \pm 9/8)$. Hence, the solution (A, B, C) equal to $(1/2, 1/4, 1/16)$ is the only solution with positive rational numbers to the original problem posed by Diophantus.

Example 1.5.2. The curve $C : y^2 = x^6 - 6x^5 + 11x^4 - 8x^3 + 11x^2 - 6x + 1$ has exactly four rational points, namely $(0, \pm 1)$ and $(1, \pm 2)$. The proof of this fact was the subject of a Ph.D. thesis by Harris Daniels [**Dan13**]. The rational points on C classify elliptic curves that satisfy a certain property, and the fact that $C(\mathbb{Q})$ is finite implies that only finitely many such elliptic curves exist.

In general, when studying rational points, the degree is not the most relevant invariant to classify curves. Instead, we classify curves according to their *genus* (from now on we assume that every curve is smooth). A curve defined over $\mathbb{Q}$ may be regarded as a curve defined over $\mathbb{C}$ and the graph of C in $\mathbb{C} \times \mathbb{C}$ is a 1-complex-dimensional curve (a Riemann surface), which can be viewed as a 2-real-dimensional surface (compact and orientable). Loosely speaking, the genus of C is the number of "holes" in this surface.

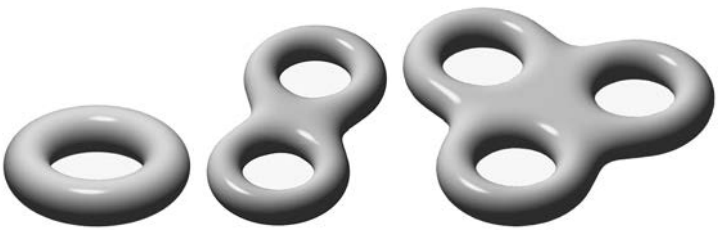

Figure 1.8. Curves of genus 1, 2, and 3, defined over $\mathbb{C}$. Images source: Wikipedia Commons.

If $C : f(x, y) = 0$ is a smooth equation for the curve C defined over $\mathbb{Q}$ and the highest degree of a monomial in the polynomial $f(x, y)$ is d, then the genus of C is given by the formula

$$\text{genus}(C) = \frac{(d-1)(d-2)}{2}.$$

For instance, a smooth curve C given by a quadratic equation (a conic, such as an ellipse or a hyperbola) has genus 0, because

$$\text{genus}(C) = \frac{(2-1)(2-2)}{2} = 0.$$

Thus, conics correspond to compact orientable surfaces with no holes, such as a sphere.

An elliptic curve E (as in Section 1.4) is given by an equation $y^2 = x^3 + Ax + B$, with $4A^3 + 27B^2 \neq 0$ to ensure smoothness, so

$$\text{genus}(E) = \frac{(3-1)(3-2)}{2} = 1.$$

Thus, every elliptic curve is a curve of genus 1; i.e., it corresponds to a compact orientable surface with one hole (a torus).

A curve of genus 0 or 1 may have infinitely many rational points (see Examples 1.3.1 and 1.4.4). In contrast, in 1922, Louis Mordell conjectured that any curve with genus > 1 can only have finitely many rational points. This was proved by Gerd Faltings in 1983 (see Figure 1.9).

Figure 1.9. Gerd Faltings (born 1954) is a German mathematician known for his work in arithmetic geometry. Image source: Archives of the Mathematisches Forschungsinstitut Oberwolfach.

Theorem 1.5.3 (Faltings's theorem)**.** *Let C be a smooth curve defined over $\mathbb{Q}$ of genus $g > 1$. Then, C has only finitely many rational points.*

There is some progress on methods to find the rational points on curves of genus 2, but very little is known about how to find the rational points on a curve of genus ≥ 3.

1.6. Diophantine Equations

> *... his boyhood lasted $\frac{1}{6}$th of his life; he married after $\frac{1}{7}$th more; his beard grew after $\frac{1}{12}$th more, and his son was born 5 years later; the son lived to half his father's age, and the father died 4 years after the son.*
>
> Metrodorus ($\sim$ 600 AD), from the *Greek Anthology*, in reference to Diophantus of Alexandria's life

In previous sections, we have discussed examples of finding integral and rational points on polynomials, and curves. More generally, we may ask ourselves how to find integral and rational points on a surface, or on a higher-dimensional algebro-geometric object V (called a *variety*), which, in general will be given by a set of equations

$$V : \begin{cases} f_1(x_1, x_2, \ldots, x_n) = 0, \\ f_2(x_1, x_2, \ldots, x_n) = 0, \\ \vdots \\ f_r(x_1, x_2, \ldots, x_n) = 0, \end{cases}$$

where, for each $1 \leq i \leq r$, the polynomial f_i has n variables $x_1, \ldots, x_n$, and integer coefficients. In this case, we are interested in the integral and rational points of V, namely,

$$V(\mathbb{Z}) = \{(a_1, \ldots, a_n) \in \mathbb{Z}^n : f_1(a_1, \ldots, a_n) = \cdots = f_r(a_1, \ldots, a_n) = 0\} \text{ and}$$
$$V(\mathbb{Q}) = \{(t_1, \ldots, t_n) \in \mathbb{Q}^n : f_1(t_1, \ldots, t_n) = \cdots = f_r(t_1, \ldots, t_n) = 0\}.$$

Each of the equations above is usually called a diophantine equation.

Definition 1.6.1. A polynomial equation of the form $C : f(x_1, \ldots, x_n) = 0$, where f is a polynomial in n variables with integer coefficients, is called a *diophantine equation.* A *rational (resp. integral, resp. natural) solution* of C is an n-tuple $(a_1, \ldots, a_n)$ of rational numbers $a_i \in \mathbb{Q}$ (resp. integers $a_i \in \mathbb{Z}$, resp. natural numbers $a_i \in \mathbb{N}$) such that

$$f(a_1, \ldots, a_n) = 0.$$

The term "diophantine" was coined in honor of Diophantus of Alexandria, whose treaties are the first records of a systematic approach to the study of the rational solutions of algebraic equations. We will write more about Diophantus in Section 1.6.1 below.

Example 1.6.2. Problem 28 in Book II of Diophantus's *Arithmetica* reads as follows:

> *To find two square numbers such that their product added to either gives a square.*

If we write x^2 and y^2 for the squares, then the problem is equivalent to finding rational solutions of the system of equations

$$\begin{cases} x^2y^2 + x^2 = u^2, \\ x^2y^2 + y^2 = v^2. \end{cases}$$

Diophantus finds one rational solution, namely $(x, y) = (3/4, 7/24)$. Let us find all the integral solutions first. From the first equation we see that $x^2(y^2 + 1) = u^2$. Therefore, either $x = u = 0$ or y^2+1 itself is a square. Since the only two consecutive squares are 0 and 1 (see Exercise 1.8.14), it follows that $(x, y, u, v) = (n, 0, \pm n, 0)$ and $(0, m, 0, \pm m)$, for some integers m, n, are the only integral solutions of the problem.

Now, let us find the rational solutions. As before, $x = u = 0$ or $y^2 + 1$ is a square. Thus, there is $t \in \mathbb{Q}$, with $t \neq \pm 1$, such that $y = \frac{2t}{1-t^2}$ (this follows from parametrizing $y^2 + 1 = w^2$; see Exercise 1.8.14). Now the second equation says $y^2(x^2 + 1) = v^2$, so either $y = 0$ or $x^2 + 1$ is a square. We similarly conclude that $x = \frac{2s}{1-s^2}$ for some $s \in \mathbb{Q}$ with $s \neq \pm 1$. Hence, the rational solutions of the problem are given by

$$(x_s, y_t) = \left(\frac{2s}{1-s^2}, \frac{2t}{1-t^2} \right),$$

and there is one solution for each s and t in $\mathbb{Q}$, other than ± 1. Indeed,

$$\begin{aligned} x_s^2 y_t^2 + x_s^2 &= x_s^2(y_t^2 + 1) \\ &= \left(\frac{2s}{1-s^2} \right)^2 \cdot \left(\frac{1+t^2}{1-t^2} \right)^2 \\ &= \left(\frac{2s(1+t^2)}{(1-s^2)(1-t^2)} \right)^2, \end{aligned}$$

and, similarly, $x_s^2 y_t^2 + y_t^2 = (2t(1+s^2))^2/((1-s^2)(1-t^2))^2$, for any $s \in \mathbb{Q}$ and any $t \in \mathbb{Q}$ not equal to ± 1.

1.6.1. About Diophantus of Alexandria. Diophantus of Alexandria (born between AD 201 and 215 and died between 285 and 299 at, apparently, age 84) is sometimes called "the father of algebra". He was an Alexandrian Greek mathematician and the author of a series of books called *Arithmetica* (see Figure 1.10), a tract *On Polygonal Numbers*, and a collection of results under the title of *Porisms*. Of the original 13 books that formed *Arithmetica*, only six were thought to have survived and it was also thought that the others must have been lost quite soon after they were written. However, in 1968, F. Sezgin made a remarkable discovery of an Arabic manuscript in the library Astan-i Quds in Meshed (The Holy Shrine library of Iran). The book seems to be a translation by Qusta ibn Luqa, who died in 912, of Books IV to VII of the *Arithmetica* by Diophantus of Alexandria.

The *Arithmetica* is not only the major work of Diophantus, but also the most prominent work on algebra in Greek mathematics. The books form a collection of about 130 problems giving numerical solutions of algebraic equations. Here is the dedication at the beginning of *Arithmetica*:

> *Knowing, my most esteemed friend Dionysius, that you are anxious to learn how to investigate problems in numbers, I have tried, beginning from the foundations on which the science is built up, to set forth to you the nature and power subsisting in numbers.*
>
> *Perhaps the subject will appear rather difficult, inasmuch as it is not yet familiar (beginners are, as a rule, too ready to despair of success); but you, with the impulse of your enthusiasm and the benefit of my teaching, will find it easy to master; for eagerness to learn, when seconded by instruction, ensures rapid progress.*

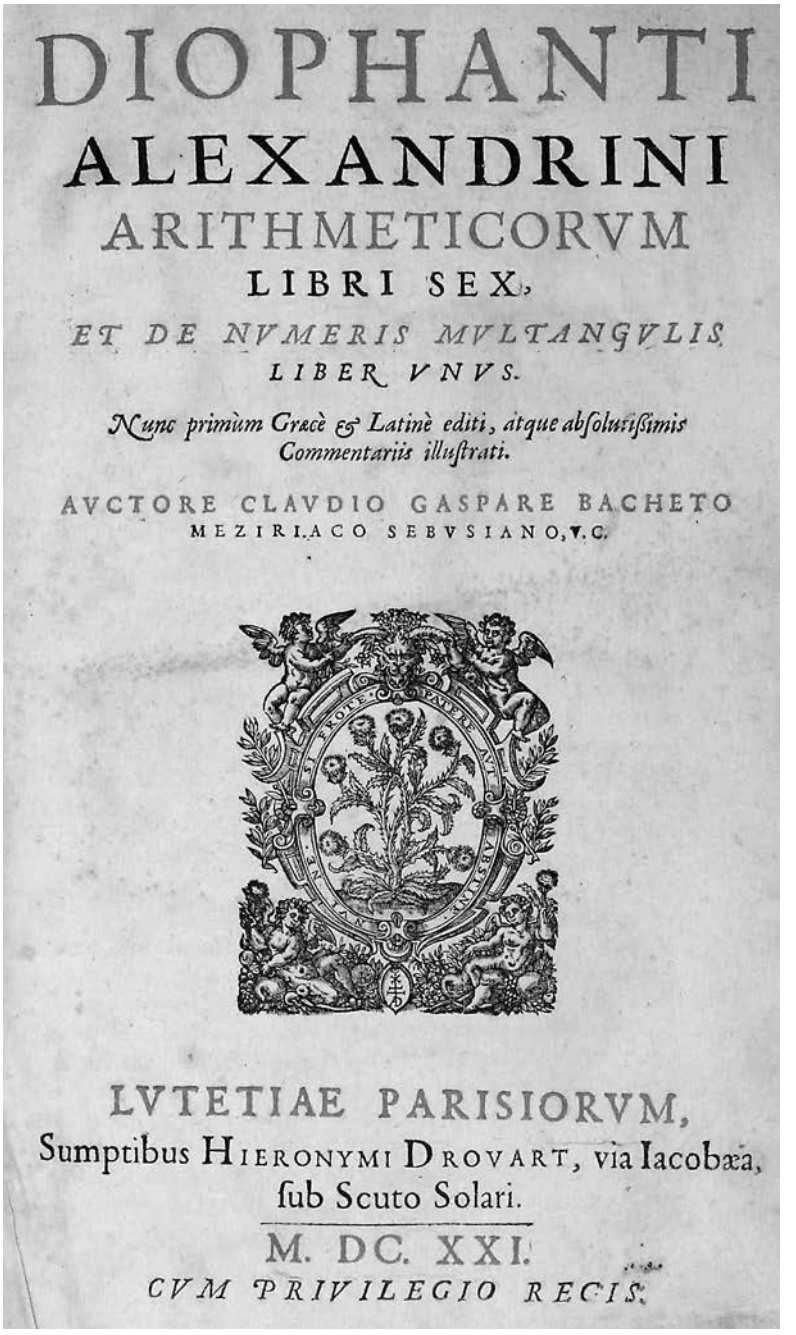
DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX,
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.
Nunc primùm Græcè & Latinè editi, atque absolutissimis Commentariis illustrati.
AVCTORE CLAVDIO GASPARE BACHETO
MEZIRIACO SEBVSIANO, V.C.
LVTETIAE PARISIORVM,
Sumptibus HIERONYMI DROVART, via Iacobæa,
sub Scuto Solari.
M. DC. XXI.
CVM PRIVILEGIO REGIS.

Figure 1.10. Title page of the 1621 edition of Diophantus's *Arithmetica*, translated from Greek into Latin by Claude Gaspard Bachet de Méziriac. Image source: Wikipedia Commons.

What follows is an example of the exposition in the *Arithmetica*, quoted from [**Hea10**].

Example 1.6.3 (Diophantus's *Arithmetica*, Book I, Problem 1)**.** *To divide a given number into two having a given difference.*

> *Given number* 100*, given difference* 40*.*
> *Lesser number required* x*. Therefore*
>
> $$2x + 40 = 100,$$
> $$x = 30.$$
>
> *The required numbers are* 70*,* 30*.*

In more modern terminology, the problem is as follows: given natural numbers N and n, find integers x and y, with $x < y$ such that $x + y = N$ and $y - x = n$. Diophantus solves the problem by subtracting both equations, to obtain $2x+n = N$, and therefore $2x = N - n$. If $N - n$ is even, then $x = (N - n)/2$ and $y = x + n = (N + n)/2$. For instance, if $N = 100$ and $n = 40$, then $x = (100 - 40)/2 = 30$ and $y = x + n = 70$.

Perhaps the most famous of all problems proposed by Diophantus in his *Arithmetica* is Problem 8 in Book II, which says

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

It is next to this proposition that, hundreds of years later, Fermat scribbled his famous note in which he enunciates what is known as "Fermat's last theorem". Pierre de Fermat (1601–1665) was a French lawyer at the Parlement of Toulouse and an amateur mathematician who is given credit for early developments that led to infinitesimal calculus and also for notable contributions to analytic geometry, probability, and optics. Nonetheless, he is particularly famous for his contributions to number theory.

During his lifetime Fermat proposed many challenges to other mathematicians, some of them quite difficult to solve. One by one, his challenges were resolved, except for one claim that took over 350 years to solve (it was proved by Andrew Wiles in 1995). Fermat's original claim was made in 1637, in an intriguing note in the margin of a copy of Diophantus's *Arithmetica*:

> *Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.*

Or, in English:

> *It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain.*

In more modern notation, Fermat's last theorem can be stated as follows.

Theorem 1.6.4 (Fermat's last theorem)**.** *The equation $x^n+y^n = z^n$ does not have any solutions $x, y, z \in \mathbb{Z}$ with $xyz \neq 0$, if $n \geq 3$.*

Figure 1.11. Pierre de Fermat (1601–1665). Image source: Wikimedia Commons.

The diophantine equation $x^n + y^n = z^n$ is, perhaps, the most studied in the theory of numbers, and it has generated thousands of pages of research articles. Notice that a non-trivial integral solution of $x^n + y^n = z^n$ corresponds to a non-trivial rational point $(\frac{x}{z}, \frac{y}{z})$ on the curve $F_n : X^n + Y^n = 1$ and, conversely, a rational point on F_n provides an integral solution of $x^n + y^n = z^n$. The curve F_n is known as the *nth Fermat curve*.

The curve $F_2 : X^2 + Y^2 = 1$ corresponds to the circle of radius 1 (which is a genus 0 curve), and it has infinitely many rational points. When $n = 3$, the Fermat curve $F_3 : X^3 + Y^3 = 1$ is an elliptic curve (a curve of genus 1), with no rational points other than $(1, 0)$ and $(0, 1)$ (and one point at "infinity"). For $n \geq 4$, the Fermat curve F_n has genus ≥ 2. Thus, by Faltings's theorem (Theorem 1.5.3) for each $n \geq 4$, the curve F_n can have at most finitely many rational points. The proof of the fact that F_n for all $n \geq 3$ has no non-trivial rational points had to wait until 1995, when Andrew Wiles announced the first complete proof of Fermat's last theorem and published it in **[Wil95]**. See **[Loz11]** for an introduction to the concepts that go into Wiles's proof.

1.7. Hilbert's Tenth Problem

Suppose $C : f(x_1, \ldots, x_n) = 0$ is a diophantine equation, as in Definition 1.6.1. The goal of the field of arithmetic geometry is to systematically study the integer and rational solutions of diophantine equations, so we ask ourselves three basic

questions:

(a) Can we determine if C has any integral solutions, or rational solutions?

(b) If so, can we find *any* of the integral or rational solutions of C?

(c) Finally, can we find *all* solutions and prove that we have found all of them?

The first question was formalized by David Hilbert (see Figure 1.12): *to devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.* This was Hilbert's tenth problem out of 23 fundamental questions that he proposed to the mathematical community during the Second International Congress of Mathematicians in Paris in the year 1900.

Figure 1.12. David Hilbert (1862–1943) was one of the most influential mathematicians of the 19th and early 20th centuries. Image source: Wikimedia Commons.

Julia Robinson's work in the late 1940s on Hilbert's tenth problem (using Pell's equation, a type of equation that we will discuss in Chapter 14) was central to the formulation of a mathematical-logic approach to the problem (see Figure 1.13). Further collaboration among Davis, Matiyasevich, Putnam, and Robinson led to the surprising discovery and proof that, in fact, *there is no such general algorithm* that decides whether a diophantine equation has integer solutions (see [**Mat93**]).

However, if we restrict our attention to solving diophantine equations of certain types, e.g., lines, conics, elliptic curves, then we can answer questions (a), (b), and (c) posed above, and this book is dedicated to describing the techniques that are known in these simpler (but fundamental) cases.

Figure 1.13. Julia Hall Bowman Robinson (1919–1985) was an American mathematician renowned for her contributions to computability theory and computational complexity theory. Her work on Hilbert's tenth problem played a crucial role in its ultimate resolution. Image author: George M. Bergman (Berkeley). Source: Archives of the Mathematisches Forschungsinstitut Oberwolfach.

1.8. Exercises

Exercise 1.8.1. Show that the decimal expansion $0.9999\ldots$, with infinitely many nines, is equal to the decimal expansion $1 = 1.0000\ldots$. In other words, show that the infinite series $\sum_{k=1}^{\infty} 9/10^k$ converges and the sum equals 1. (Hint: use the geometric series test.)

Exercise 1.8.2. Find all the natural, integral, and rational roots of the following polynomial equations:

(1) $x^5 - 9x^4 - 5x^3 + 45x^2 + 4x - 36 = 0$.

(2) $3x^4 + 5x^3 - 3x^2 - 5x = 0$.

(3) $x^4 + 5x^3 - 16x^2 - 17x - 21 = 0$.

(4) $x^4 + x^3 + 21 = 0$.

Exercise 1.8.3. Find $k \in \mathbb{Z}$ such that $x = 5$ is a root of $x^3 + kx^2 + 23x + 285 = 0$.

Exercise 1.8.4. Find integers m and n such that $x = -2$ and $x = 3$ are roots of the polynomial equation $x^3 + 10x^2 + mx + n = 0$.

Exercise 1.8.5. Let $p(x)$ and $q(x)$ be polynomials, and let $\alpha \in \mathbb{Q}$ be a root of $p(x) = 0$ and $\beta \in \mathbb{Q}$ a root of $q(x) = 0$. Show the following statements:

(1) The numbers α and β are roots of the polynomial equation $p(x) \cdot q(x) = 0$.

(2) If $\alpha = \beta$, then α is a root of the polynomial equation $p(x) + q(x) = 0$.

Exercise 1.8.6. The goal of this exercise is to show that every complex number $\alpha \in \mathbb{C}$ has a square root within the complex numbers; i.e., there is some $\sqrt{\alpha} \in \mathbb{C}$. Recall the definition of the complex numbers:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R} \text{ and } i^2 = -1\}.$$

(1) Find a square root of $\alpha = 1+i$, within $\mathbb{C}$; i.e., find $\beta \in \mathbb{C}$ such that $\beta^2 = 1+i$. (Hint: write $\beta = c + di$ and find a similar expression for its square β^2.)

(2) Find a square root of $\alpha = a + bi$, within $\mathbb{C}$; i.e., find $\beta = c + di$ such that $\beta^2 = \alpha$. In fact, show that

$$\beta = \pm \left(\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}} + \left(\frac{b}{|b|} \sqrt{\frac{\sqrt{a^2+b^2}-a}{2}} \right) \cdot i \right).$$

(3) For any real number θ, we define

$$e^{i\theta} = \cos(\theta) + \sin(\theta) \cdot i.$$

Show that any complex number α can be written uniquely as $\alpha = re^{i\theta}$, for some $r \geq 0$ and some $\theta \in [0, 2\pi)$. (Hint: find a geometric interpretation of r and θ in the complex plane.)

(4) Show that if $\alpha = re^{i\theta}$, then the square roots of α are $\beta = \sqrt{r}e^{i\theta/2}$ and $-\beta = \sqrt{r}e^{i(\theta/2+\pi)}$.

Exercise 1.8.7. Find all natural numbers n such that its cube minus its square plus itself equals 1.

Exercise 1.8.8. Let $P = (x_0, y_0)$ and $Q = (x_1, y_1)$ be two points in the plane, with $x_0 \neq x_1$, and let $m = (y_1 - y_0)/(x_1 - x_0)$. Show that the line L that passes through P and Q is given by

$$y - y_0 = m \cdot (x - x_0).$$

Exercise 1.8.9. Let $P = (1, 4)$ and $Q = (4, -2)$ be points on the plane.

(1) Find the equation $y = ax + b$ of a line L that passes through P and Q.

(2) Find a formula for all the rational points on L.

(3) Find a formula for all the integral points on L.

(4) How many points on L have natural coordinates; i.e., how many points $R = (x_0, y_0)$ on L are there with $x_0, y_0 \in \mathbb{N}$?

Exercise 1.8.10. Find all the rational points on the circle $x^2 + y^2 = 2$.

Exercise 1.8.11. Let C be the ellipse given by $x^2 + 3y^2 = 784$.

(1) Find all the integral points on C.

(2) Find a parametrization of all the rational points on C.

Exercise 1.8.12. Let C be the hyperbola given by the equation $x^2 - 7y^2 = 2$.

(1) Find all the rational points on the hyperbola $x^2 - 7y^2 = 2$.

(2) Find 3 distinct integral points with positive x-coordinate.

Exercise 1.8.13. Show that the hyperbola $C' : x^2 - 5y^2 = 3$ has no integral points.

Exercise 1.8.14. (1) Are there two perfect squares (i.e., integers of the form n^2, where n itself is an integer) that differ by 1? Write the problem in terms of a diophantine equation, find all integral solutions to the equation, and prove that you have found them all. (Hint: write one square as n^2 and the other square as $(n+m)^2$.)

(2) Find a parametrization of all the rational squares (i.e., rational numbers of the form t^2 for some $t \in \mathbb{Q}$) that differ by 1.

(3) Are there two consecutive integers such that their product is a perfect square? If so, find all such integers.

(4) Are there three consecutive integers such that their product is a perfect square? If so, find all such integers. (*This is hard! Here it suffices to find one diophantine equation in two variables that represents this problem.*)

(5) Are there three integers $u < v < w$ that differ by 5 (i.e., $u + 5 = v$ and $v + 5 = w$) and such that their product is a perfect square? If so, find all such integers. (*There are some .... Can you find any? Finding all solutions is hard! Again, here it suffices to find one diophantine equation in two variables that represents this problem.*)

Exercise 1.8.15. We say that a natural number $n \geq 1$ is a *congruent number* if there is a right triangle with rational sides and area equal to n. Is $n = 5$ a congruent number? If so, find a right triangle with rational sides and area equal to 5.

Exercise 1.8.16. A triple (a, b, c) of natural numbers $a, b, c \in \mathbb{N}$ is said to be *pythagorean* if they satisfy $a^2 + b^2 = c^2$.

(1) Show that $(a, b, c) = (n^2 - m^2, 2nm, n^2 + m^2)$ is a pythagorean triple for any two non-zero distinct integers $n > m > 0$.

(2) Show that if n and m satisfy (i) one of n and m is even and the other one is odd and (ii) n and m are relatively prime, then $(a, b, c) = (n^2 - m^2, 2nm, n^2 + m^2)$ is a *primitive* pythagorean triple; i.e., a, b, and c are pairwise relatively prime.

(3) Use (b) to find five distinct primitive pythagorean triples.

Exercise 1.8.17. An *Euler brick* is just a rectangular box in which all of the edges (length, depth, and height) have integer dimensions and in which the diagonals on all three sides are also integers.

(1) Find the dimensions of two distinct Euler bricks.

(2) A *perfect cuboid* is an Euler brick in which the space diagonal, that is, the distance from any corner to its opposite corner, is also an integer. Can you find a perfect cuboid? (*This is an* **open problem**. *Here, it suffices to find a system of diophantine equations that represents this problem.*)

Exercise 1.8.18 (Diophantus's *Arithmetica*, Book II, Problem 30)**.** Find two numbers such that their product plus or minus their sum gives a square; i.e., find a pair of rational numbers x and y such that there are $u, v \in \mathbb{Q}$ with

$$\begin{cases} xy + x + y = u^2, \\ xy - (x + y) = v^2. \end{cases}$$

Can you find *all* such rational numbers x and y?

Exercise 1.8.19. Find a copy of Book II of the *Arithmetica* by Diophantus of Alexandria and quote two problems (other than numbers 8, 29, or 30). Reproduce Diophantus's solution, and then rewrite it in a more modern language and notation.

Exercise 1.8.20. Let E be the elliptic curve given by $y^2 = x^3 + Ax + B$, for some $A, B \in \mathbb{Z}$.

(1) Use implicit differentiation to show that
$$y' = \frac{dy}{dx} = \frac{3x^2 + A}{2y}.$$
(2) Let E be $y^2 = x^3 - 2$ and $P = (3, 5)$. Find $\frac{dy}{dx}(P)$, i.e., the slope of the tangent line to E at the point P.
(3) Let E be $y^2 = x^3 - x$ and $P = (0, 0)$. What is the slope of the tangent line to E at the point P?

Exercise 1.8.21. Let $f(a, b, c, d) = ad - bc$ and let $C : ad - bc = 1$.

(1) Let $\mathrm{SL}(2, \mathbb{Z})$ be the set of 2×2 matrices with integer coefficients and determinant 1. Show that the set $C(\mathbb{Z})$ of integral points on the diophantine equation C is in bijection with $\mathrm{SL}(2, \mathbb{Z})$.
(2) Show that $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ belong to $\mathrm{SL}(2, \mathbb{Z})$.
(3) Show that if A and B are matrices in $\mathrm{SL}(2, \mathbb{Z})$, then $A \cdot B$ is also in $\mathrm{SL}(2, \mathbb{Z})$, where $\cdot$ here denotes matrix multiplication (see Example 5.2.5).
(4) Show that $Q_n = (S \cdot T^2)^n = (S \cdot T^2) \cdots (S \cdot T^2)$ is a matrix in $\mathrm{SL}(2, \mathbb{Z})$ for all $n \geq 1$. Describe the points on C that correspond to the matrices Q_n for $1 \leq n \leq 6$.
(5) Show that there are infinitely many integral points (a, b, c, d) in $C(\mathbb{Z})$ with all non-zero coordinates.

Note: this problem continues in Exercises 2.11.12 and 5.6.4.