2 Exponentiation Is Diophantine

The goal of this chapter is to show that the function b^c is a Diophantine function of two arguments. The proof is rather technical (and Sections 2.1–2.4 could be skipped at the first reading), but it opens a straightforward path to proving that many other interesting and important functions and relations are Diophantine. In particular, in Section 3.4 we show that the set of all prime numbers is Diophantine.

2.1 Special second-order recurrent sequences

As was just stated, the goal of this chapter is to show that the function b^c is Diophantine or, equivalently, to show that the set of triples

$$\{\langle a, b, c\rangle \mid a = b^c\} \tag{2.1.1}$$

is Diophantine. Clearly, this would imply that the set of pairs

$$\{\langle a, b\rangle \mid \exists n\,[a = b^n]\} \tag{2.1.2}$$

is Diophantine. The converse implication is also valid (see Exercise 2.2), but it is not easy even to establish that (2.1.2) is Diophantine.

The set of all powers of a given number b may be viewed as the set of all members of the first-order recurrent sequence

$$\beta_b(0) = 1, \qquad \beta_b(n+1) = b\beta_b(n). \tag{2.1.3}$$

In our proof an essential role will be played by the second-order recurrent sequence

$$\alpha_b(0) = 0, \qquad \alpha_b(1) = 1, \qquad \alpha_b(n+2) = b\alpha_b(n+1) - \alpha_b(n), \tag{2.1.4}$$

where $b \geq 2$. In this section we take the first step by showing that the set of pairs

$$\{\langle a, b\rangle \mid b \geq 2 \;\&\; \exists n\,[a = \alpha_b(n)]\} \tag{2.1.5}$$

is Diophantine. Strange as it may seem, it is much easier to prove this than it is to show that the set (2.1.2), which is so much more natural and commonplace, is Diophantine.

The second-order relation (2.1.4) can be rewritten as a first-order relation among the matrices

$$\mathrm{A}_b(n) = \begin{pmatrix} \alpha_b(n+1) & -\alpha_b(n) \\ \alpha_b(n) & -\alpha_b(n-1) \end{pmatrix}, \tag{2.1.6}$$

taking $\alpha_b(-1) = -1$. Namely,

$$\mathrm{A}_b(0) = \mathrm{E}, \qquad \mathrm{A}_b(n+1) = \mathrm{A}_b(n)\Xi_b, \tag{2.1.7}$$

where

$$\mathrm{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Xi_b = \begin{pmatrix} b & -1 \\ 1 & 0 \end{pmatrix}. \tag{2.1.8}$$

This implies that

$$\mathrm{A}_b(n) = \Xi_b^n, \tag{2.1.9}$$

and hence

$$\det(\mathrm{A}_b(n)) = 1, \tag{2.1.10}$$

i.e.,

$$\begin{aligned} \alpha_b^2(n) - \alpha_b(n+1)\alpha_b(n-1) &= \alpha_b^2(n+1) - b\alpha_b(n+1)\alpha_b(n) + \alpha_b^2(n) \\ &= \alpha_b^2(n-1) - b\alpha_b(n-1)\alpha_b(n) + \alpha_b^2(n) = 1. \end{aligned} \tag{2.1.11}$$

It turns out that equation (2.1.11) characterizes the sequence (2.1.4) in the following sense: *if*

$$x^2 - bxy + y^2 = 1, \tag{2.1.12}$$

then either

$$x = \alpha_b(m+1), \qquad y = \alpha_b(m) \tag{2.1.13}$$

or

$$x = \alpha_b(m), \qquad y = \alpha_b(m+1) \tag{2.1.14}$$

for some m. In order to distinguish which of the two cases, (2.1.13) or (2.1.14), holds, it is sufficient to note that (2.1.4) implies by induction that

$$0 = \alpha_b(0) < \alpha_b(1) < \cdots < \alpha_b(n) < \alpha_b(n+1) < \cdots \tag{2.1.15}$$

We now show that *equation* (2.1.12) *together with the inequality*

$$y < x \tag{2.1.16}$$

implies the existence of some m *for which* (2.1.13) *holds.* The proof will proceed by induction on y. If $y = 0$, then clearly $x = 1$; i.e., (2.1.13) holds with $m = 0$. If $y > 0$, then (2.1.12) and (2.1.16) imply that

$$x = by + \frac{1 - y^2}{x} \le by, \tag{2.1.17}$$

$$x = by + \frac{1}{x} - \frac{y^2}{x} > by - y. \tag{2.1.18}$$

Let $x_1 = y$ and $y_1 = by - x$. Then

$$\begin{aligned} x_1^2 - bx_1y_1 + y_1^2 &= y^2 - by(by - x) + (by - x)^2 \\ &= x^2 - bxy + y^2 \\ &= 1. \end{aligned} \tag{2.1.19}$$

By (2.1.18), $y_1 < x_1$, and by the induction hypothesis,

$$x_1 = \alpha_b(m_1 + 1), \qquad y_1 = \alpha_b(m_1) \tag{2.1.20}$$

for some m_1. Hence, for $m = m_1 + 1$,

$$x = bx_1 - y_1 = \alpha_b(m+1), \qquad y = x_1 = \alpha_b(m). \tag{2.1.21}$$

Thus, we have proved that the set (2.1.5) is defined by the formula

$$b \ge 2 \;\&\; \exists x\,[x^2 - abx + a^2 = 1]. \tag{2.1.22}$$

2.2 The special recurrent sequences are Diophantine (basic ideas)

Our first goal will be to show that the set of triples

$$\{\, \langle a, b, c\rangle \mid b \ge 4 \;\&\; a = \alpha_b(c) \,\} \tag{2.2.1}$$

is Diophantine. In this section we outline the underlying ideas, while the formal proof will be given in the next section.

It is convenient to consider the set (2.2.1) as the union of the terms of the sequences

$$\langle \alpha_b(0), b, 0\rangle, \ldots, \langle \alpha_b(n), b, n\rangle, \ldots \tag{2.2.2}$$

for $b = 4, 5, \ldots$. Using induction on definition (2.1.4), it is easy to derive that

$$\alpha_2(n) = n. \tag{2.2.3}$$

Hence, for $b = 2$ the sequence (2.2.2) is very simple:

$$\langle 0, 2, 0\rangle, \ldots, \langle n, 2, n\rangle, \ldots \tag{2.2.4}$$

However, we are concerned with the case $b \geq 4$, so $\alpha_b(n)$ cannot be defined by a simple equation like (2.2.3). Nevertheless, for $b > 2$ there is a weak analog of (2.2.3); namely, it follows by induction from (2.1.4) that

$$\alpha_{b_1}(n) \equiv \alpha_{b_2}(n) \pmod{q} \tag{2.2.5}$$

provided that

$$b_1 \equiv b_2 \pmod{q}. \tag{2.2.6}$$

Hence, in particular,

$$\alpha_b(n) \equiv \alpha_2(n) = n \pmod{b-2}, \tag{2.2.7}$$

so that the first $b-2$ members of the sequence (2.2.2) coincide with the first $b-2$ members of the sequence

$$\langle \alpha_b(0), b, \operatorname{rem}(\alpha_b(0), b-2)\rangle, \ldots, \langle \alpha_b(n), b, \operatorname{rem}(\alpha_b(n), b-2)\rangle, \ldots \tag{2.2.8}$$

The "advantage" of the sequence (2.2.8) (as compared with (2.2.2)) consists in the fact that here n enters only as an argument of α. Together with the facts that the set (2.1.5) and the function rem are Diophantine, this implies that the set of all triples from (2.2.8) is also Diophantine. The "disadvantage" consists in the fact that only finite initial segments of the sequences (2.2.8) and (2.2.2) are equal.

Using (2.2.5), we can construct another sequence that has the same "advantage" but avoids the "disadvantage." Namely, let

$$w \equiv b \pmod{v}, \tag{2.2.9}$$

$$w \equiv 2 \pmod{u}, \tag{2.2.10}$$

$$v > 2\alpha_b(k), \tag{2.2.11}$$

$$u > 2k. \tag{2.2.12}$$

Then, as we shall see, the first k members of the sequence (2.2.2) coincide with the

first k members of the sequence

$$\langle \operatorname{arem}(\alpha_w(0), v), b, \operatorname{arem}(\alpha_w(0), u)\rangle, \ldots, \langle \operatorname{arem}(\alpha_w(n), v), b, \operatorname{arem}(\alpha_w(n), u)\rangle, \ldots \tag{2.2.13}$$

(The function rem is replaced here by the function arem that was introduced in Section 1.6; the role of this substitution will become clear later.)

Now, the union of all the sequences of the form (2.2.13) with u, v, and w satisfying conditions (2.2.9) and (2.2.10) certainly contains all the triples from the sequence (2.2.2); however, this union may also contain some additional triples. To eliminate these additional triples, we begin by imposing on u and v, besides (2.2.9) and (2.2.10), some further conditions, and, moreover, we exclude from (2.2.13) those triples that do not satisfy the inequality

$$2\operatorname{arem}(\alpha_w(n), v) < u. \tag{2.2.14}$$

In order to understand the nature of these new conditions on v and u, we note that the recurrent relations (2.1.4) imply that for any positive v the sequence

$$\alpha_b(0), \ldots, \alpha_b(n), \ldots \tag{2.2.15}$$

is purely periodic modulo v. For our special choice of v we will be able to determine the length of the period and its structure. Namely, let

$$v = \alpha_b(m+1) - \alpha_b(m-1); \tag{2.2.16}$$

then

$$\alpha_b(m+1) \equiv \alpha_b(m-1) \pmod{v}. \tag{2.2.17}$$

The recurrent relation (2.1.4) can be rewritten as

$$\alpha_b(n-2) = b\alpha_b(n-1) - \alpha_b(n), \tag{2.2.18}$$

and hence

$$\begin{aligned} \alpha_b(m+2) &= b\alpha_b(m+1) - \alpha_b(m) \\ &\equiv b\alpha_b(m-1) - \alpha_b(m) \pmod{v} \\ &= \alpha_b(m-2), \end{aligned} \tag{2.2.19}$$

$$\begin{aligned}
&\alpha_b(m+3) \equiv \alpha_b(m-3) \pmod{v},\\
&\vdots\\
&\alpha_b(2m-1) \equiv \alpha_b(1) \pmod{v},\\
&\alpha_b(2m) \equiv \alpha_b(0) \pmod{v}.
\end{aligned} \tag{2.2.20}$$

Furthermore, we have

$$\alpha_b(2m) \equiv \alpha_b(0) = 0 = -\alpha_b(0) \pmod{v}, \tag{2.2.21}$$

$$\alpha_b(2m+1) = b\alpha_b(2m) - \alpha_b(2m-1) \equiv -\alpha_b(1) \pmod{v} \tag{2.2.22}$$

and hence

$$\alpha_b(2m+n) \equiv -\alpha_b(n) \pmod{v}. \tag{2.2.23}$$

Thus, for our choice of v, the sequence (2.2.15) modulo v has the following period of $4m$ terms:

$$\begin{aligned}
&0,\ 1,\ \ldots,\ \alpha_b(m-1),\ \alpha_b(m),\ \alpha_b(m-1),\ \ldots,\ 1,\\
&0,\ -1,\ \ldots,\ -\alpha_b(m-1),\ -\alpha_b(m),\ -\alpha_b(m-1),\ \ldots,\ -1.
\end{aligned} \tag{2.2.24}$$

According to (2.2.9), this is also the period modulo v of the sequence

$$\alpha_w(0), \ldots, \alpha_w(n), \ldots \tag{2.2.25}$$

Correspondingly, the sequence

$$\operatorname{arem}(\alpha_w(0), v), \ldots, \operatorname{arem}(\alpha_w(n), v), \ldots \tag{2.2.26}$$

has the period of $2m$ terms

$$0, 1, \ldots, \alpha_b(m-1), \alpha_b(m), \alpha_b(m-1), \ldots, 1, \tag{2.2.27}$$

because according to (2.1.15)

$$\begin{aligned}
v &= \alpha_b(m+1) - \alpha_b(m-1)\\
&= b\alpha_b(m) - 2\alpha_b(m-1)\\
&\geq 2\alpha_b(m)
\end{aligned}$$

for $b \geq 4$.

According to (2.2.10) and (2.2.7), the sequence (2.2.25) modulo u has the period of u terms

$$0, 1, \ldots, u-1. \tag{2.2.28}$$

Now we impose on u the very important condition

$$u \mid m. \tag{2.2.29}$$

This implies that the length of the period of the sequence (2.2.26) is a multiple of the length of the period of the sequence

$$\operatorname{arem}(\alpha_w(0), u), \ldots, \operatorname{arem}(\alpha_w(n), u), \ldots \tag{2.2.30}$$

and hence that the sequence (2.2.13) has an almost symmetrical period of length $2m$. Thus all the "extra" triples in (2.2.13) should appear among the first $m+1$ members of this sequence. For these initial triples, condition (2.2.14) can be rewritten as

$$2\alpha_b(n) < u, \tag{2.2.31}$$

and therefore

$$2n < u, \tag{2.2.32}$$

because according to (2.1.15)

$$n \leq \alpha_b(n). \tag{2.2.33}$$

Now, (2.2.32) implies that

$$\operatorname{arem}(\alpha_b(n), u) = \operatorname{arem}(n, u) = n; \tag{2.2.34}$$

thus, condition (2.2.14) indeed eliminates all the "extra" triples.

In trying to implement the plan described above, we encounter the following difficulty: how can we transform the pair of conditions (2.2.16) and (2.2.29) into Diophantine equations without first proving that α is a Diophantine function? To overcome this difficulty we shall employ the following property of α:

$$\alpha_b^2(k) \mid \alpha_b(m) \Longrightarrow \alpha_b(k) \mid m. \tag{2.2.35}$$

We shall put

$$u = \alpha_b(k) \tag{2.2.36}$$

and replace (2.2.29) by the stronger condition

$$u^2 \mid \alpha_b(m). \tag{2.2.37}$$

2.3 The special recurrent sequences are Diophantine (proof)

We begin by proving the implication (2.2.35). Let b, k, and m satisfy

$$\alpha_b^2(k) \mid \alpha_b(m). \tag{2.3.1}$$

Recall that $\alpha_b(k)$ and $\alpha_b(m)$ are elements of the matrices $\mathrm{A}_b(k)$ and $\mathrm{A}_b(m)$ defined by (2.1.6) and satisfying (2.1.9). Let

$$m = n + kl, \qquad 0 \le n < k. \tag{2.3.2}$$

We have

$$\begin{aligned}\begin{pmatrix}\alpha_b(m+1) & -\alpha_b(m)\\ \alpha_b(m) & -\alpha_b(m-1)\end{pmatrix} &= \mathrm{A}_b(m)\\ &= \Xi_b^m\\ &= \Xi_b^{n+kl}\\ &= \Xi_b^n(\Xi_b^k)^l\\ &= \mathrm{A}_b(n)\mathrm{A}_b^l(k)\\ &= \begin{pmatrix}\alpha_b(n+1) & -\alpha_b(n)\\ \alpha_b(n) & -\alpha_b(n-1)\end{pmatrix}\begin{pmatrix}\alpha_b(k+1) & -\alpha_b(k)\\ \alpha_b(k) & -\alpha_b(k-1)\end{pmatrix}^l.\end{aligned} \tag{2.3.3}$$

Passing to a congruence modulo $\alpha_b(k)$, we obtain

$$\begin{pmatrix}\alpha_b(m+1) & -\alpha_b(m)\\ \alpha_b(m) & -\alpha_b(m-1)\end{pmatrix} \equiv \begin{pmatrix}\alpha_b(n+1) & -\alpha_b(n)\\ \alpha_b(n) & -\alpha_b(n-1)\end{pmatrix}\begin{pmatrix}\alpha_b(k+1) & 0\\ 0 & -\alpha_b(k-1)\end{pmatrix}^l \pmod{\alpha_b(k)}, \tag{2.3.4}$$

and hence

$$\alpha_b(m) \equiv \alpha_b(n)\alpha_b^l(k+1) \pmod{\alpha_b(k)}. \tag{2.3.5}$$

By (2.1.11), $\alpha_b(k)$ and $\alpha_b(k+1)$ are coprime; thus, (2.3.1) and (2.3.5) imply that

$$\alpha_b(k) \mid \alpha_b(n). \tag{2.3.6}$$

Now it follows from (2.3.2) and (2.1.15) that $\alpha_b(n) < \alpha_b(k)$, so that (2.3.6) is possible only if $n = 0$, i.e., if $m = kl$. Furthermore, we have:

$$\begin{aligned}\mathrm{A}_b(m) &= \mathrm{A}_b^l(k)\\ &= [\alpha_b(k)\Xi_b - \alpha_b(k-1)\mathrm{E}]^l\\ &= \sum_{i=0}^{l}(-1)^{l-i}\binom{l}{i}\alpha_b^i(k)\alpha_b^{l-i}(k-1)\Xi_b^i.\end{aligned} \tag{2.3.7}$$

Passing from the equality to a congruence modulo $\alpha_b^2(k)$, we can omit all the summands except the first two:

$$\begin{aligned}\mathrm{A}_b(m) &= \begin{pmatrix}\alpha_b(m+1) & -\alpha_b(m)\\ \alpha_b(m) & -\alpha_b(m-1)\end{pmatrix}\\ &\equiv (-1)^l\alpha_b^l(k-1)\mathrm{E} + (-1)^{l-1}l\alpha_b(k)\alpha_b^{l-1}(k-1)\Xi_b \pmod{\alpha_b^2(k)},\end{aligned} \tag{2.3.8}$$

whence

$$\alpha_b(m) \equiv (-1)^{l-1}l\alpha_b(k)\alpha_b^{l-1}(k-1) \pmod{\alpha_b^2(k)}. \tag{2.3.9}$$

Together with (2.3.1) this implies that

$$\alpha_b(k) \mid l\alpha_b^{l-1}(k-1), \tag{2.3.10}$$

and because by (2.1.11) $\alpha_b(k)$ and $\alpha_b(k-1)$ are coprime,

$$\alpha_b(k) \mid l. \tag{2.3.11}$$

The implication (2.2.35) is proved.

Now we can exhibit a system of Diophantine conditions that is solvable if and only if the triple $\langle a, b, c\rangle$ belongs to the set (2.2.1):

$$b \ge 4, \tag{2.3.12}$$

$$u^2 - but + t^2 = 1, \tag{2.3.13}$$

$$s^2 - bsr + r^2 = 1, \tag{2.3.14}$$

$$r < s, \tag{2.3.15}$$

$$u^2 \mid s, \tag{2.3.16}$$

$$v = bs - 2r, \tag{2.3.17}$$

$$v \mid w - b, \tag{2.3.18}$$

$$u \mid w - 2, \tag{2.3.19}$$

$$w > 2, \tag{2.3.20}$$

$$x^2 - wxy + y^2 = 1, \tag{2.3.21}$$

$$2a < u, \tag{2.3.22}$$

$$a = \operatorname{arem}(x, v), \tag{2.3.23}$$

$$c = \operatorname{arem}(x, u). \tag{2.3.24}$$

We first prove that if the conditions (2.3.12)–(2.3.24) are satisfied, then

$$a = \alpha_b(c). \tag{2.3.25}$$

It was shown in Section 2.1 that (2.3.12) and (2.3.13) imply that for some k,

$$u = \alpha_b(k). \tag{2.3.26}$$

Likewise, (2.3.12), (2.3.14), and (2.3.15) imply that for some positive m,

$$s = \alpha_b(m), \qquad r = \alpha_b(m-1). \tag{2.3.27}$$

By (2.2.35), it follows from (2.3.16), (2.3.26), and (2.3.27) that

$$u \mid m. \tag{2.3.28}$$

By (2.1.4), it follows from (2.3.17) and (2.3.27) that

$$v = \alpha_b(m+1) - \alpha_b(m-1). \tag{2.3.29}$$

Furthermore, it follows from (2.3.20) and (2.3.21) that for some n,

$$x = \alpha_w(n). \tag{2.3.30}$$

From this, (2.3.18) and (2.3.19), and (2.2.5)–(2.2.7), we have that

$$x \equiv \alpha_b(n) \pmod{v}, \tag{2.3.31}$$

$$x \equiv n \pmod{u}. \tag{2.3.32}$$

Let

$$n = 2lm \pm j, \tag{2.3.33}$$

where

$$j \le m. \tag{2.3.34}$$

Using the matrix representation once again, we have:

$$\begin{aligned} \mathrm{A}_b(n) &= \Xi_b^n \\ &= \Xi_b^{2lm \pm j} \\ &= \left[[\Xi_b^m]^2\right]^l \Xi_b^{\pm j} \\ &= \left[[\mathrm{A}_b(m)]^2\right]^l [\mathrm{A}_b(j)]^{\pm 1}, \end{aligned} \tag{2.3.35}$$

$$\begin{aligned} \mathrm{A}_b(m) &= \begin{pmatrix} \alpha_b(m+1) & -\alpha_b(m) \\ \alpha_b(m) & -\alpha_b(m-1) \end{pmatrix} \\ &\equiv -\begin{pmatrix} -\alpha_b(m-1) & \alpha_b(m) \\ -\alpha_b(m) & \alpha_b(m+1) \end{pmatrix} \pmod{v} \\ &= -[\mathrm{A}_b(m)]^{-1}, \end{aligned} \tag{2.3.36}$$

$$[\mathrm{A}_b(m)]^2 \equiv -\mathrm{E} \pmod{v} \tag{2.3.37}$$

$$\mathrm{A}_b(n) \equiv \pm[\mathrm{A}_b(j)]^{\pm 1} \pmod{v}. \tag{2.3.38}$$

(In this last formula all four combinations of the signs "+" and "−" are possible.) Passing from the matrix congruence (2.3.38) to element-wise congruence, we have that

$$x \equiv \alpha_b(n) \equiv \pm\alpha_b(j) \pmod{v}. \tag{2.3.39}$$

By (2.1.15), it follows from (2.3.34) that

$$2\alpha_b(j) \le 2\alpha_b(m) \le (b-2)\alpha_b(m) < b\alpha_b(m) - 2\alpha_b(m-1) = v, \tag{2.3.40}$$

and hence

$$a = \operatorname{arem}(x, v) = \operatorname{arem}(\alpha_b(n), v) = \alpha_b(j). \tag{2.3.41}$$

From this and (2.3.22), using (2.2.33), we have that

$$2j \le 2\alpha_b(j) = 2a < u. \tag{2.3.42}$$

Finally, from (2.3.28), (2.3.31), (2.3.33), and (2.3.42) we obtain

$$c = \operatorname{arem}(x, u) = \operatorname{arem}(n, u) = j, \tag{2.3.43}$$

which together with (2.3.41) gives the desired equality (2.3.25).

Now we are going to prove the converse; i.e., we will show that if the numbers a, b, and c satisfy (2.3.12) and (2.3.25), then there are numbers s, r, u, t, v, w satisfying (2.3.13)–(2.3.24). The above considerations indicate how these numbers are to be chosen.

We begin by choosing u according to (2.3.26), selecting a and k so that the inequality (2.3.22) holds and u is odd. We are able to do this because by (2.1.15) the sequence $\alpha_b(0)$, $\alpha_b(1)$, ... increases monotonically and by (2.1.11) at least one of any two consecutive terms of the sequence is odd. Let

$$t = \alpha_b(k+1); \tag{2.3.44}$$

then using (2.1.11), equation (2.3.13) holds.

We choose r and s as in (2.3.27), with

$$m = uk; \tag{2.3.45}$$

then by (2.1.11) and (2.1.15), equation (2.3.14) and inequality (2.3.15) both hold. Using (2.3.9),

$$s = \alpha_b(uk) \equiv (-1)^{u-1} u \alpha_b(k) \alpha_b^{u-1}(k-1) \pmod{u^2}; \tag{2.3.46}$$

hence condition (2.3.16) is also valid.

We can find v satisfying (2.3.17) because using (2.1.15)

$$bs - 2r \geq 4\alpha_b(m) - 2\alpha_b(m-1) > 2\alpha_b(m). \tag{2.3.47}$$

We now verify that u and v are coprime. Suppose that $d|u$ and $d|v$; then by (2.3.16) $d \mid s$ and by (2.3.17) $d \mid 2r$. However, by our choice of u, d is odd; hence $d \mid r$ and by (2.3.14) $d \mid 1$. Thus by the Chinese Remainder Theorem (see the Appendix) we can find w satisfying (2.3.18), (2.3.19), and (2.3.20).

Finally, let

$$x = \alpha_w(c), \qquad y = \alpha_w(c+1); \tag{2.3.48}$$

then by (2.1.15), equation (2.3.21) holds.

Using (2.2.5) it follows from (2.3.25), (2.3.48), and (2.3.18) that

$$x = \alpha_w(c) \equiv \alpha_b(c) = a \pmod{v}. \tag{2.3.49}$$

From (2.3.17), (2.3.25), and (2.3.47) it follows that

$$v > 2a, \tag{2.3.50}$$

and hence (2.3.49) implies (2.3.23).

By (2.1.7) it follows from (2.3.48) that

$$x \equiv c \pmod{w-2}, \tag{2.3.51}$$

which together with (2.3.19) gives the congruence

$$x \equiv c \pmod{u}. \tag{2.3.52}$$

By (2.2.33) it follows from (2.3.25) and (2.3.22) that

$$2c \leq 2\alpha_b(c) = 2a < u, \tag{2.3.53}$$

which together with (2.3.52) implies (2.3.24).

All of the conditions (2.3.12)–(2.3.24) are Diophantine; thus, we have established that the set (2.2.1) is Diophantine.

2.4 Exponentiation is Diophantine

To begin with, we need to specify the value of 0^0. For a number of different reasons, it is convenient to make the definition $0^0 = 1$.

The recurrent relation (2.1.4) is close to (2.1.3) for large values of b, and $\alpha_b(n)$ grows approximately like $\beta_b(n) = b^n$. More precisely, it is easy to prove by induction that

$$(b-1)^n \leq \alpha_b(n+1) \leq b^n. \tag{2.4.1}$$

For a fixed n the relative error goes to 1 when $b \to \infty$, but we need a good approximation to b^c for a fixed b. That is why we introduce a new variable x with a large value. Later we'll verify that

$$b^c = \lim_{x \to \infty} \frac{\alpha_{bx+4}(c+1)}{\alpha_x(c+1)}. \tag{2.4.2}$$

Moreover, this relation holds for all values of b and c, including the case $b = 0$. Our language of Diophantine equations does not contain the operation lim, but in this case it can be replaced by the function div. Namely, (2.4.1) implies that

$$\frac{\alpha_{bx+4}(c+1)}{\alpha_x(c+1)} \geq \frac{(bx+3)^c}{x^c} \geq b^c \tag{2.4.3}$$

for large x and hence that

$$b^c = \alpha_{bx+4}(c+1) \operatorname{div} \alpha_x(c+1). \tag{2.4.4}$$

In order to determine when the value of x is sufficiently large, we estimate the left-hand side of (2.4.3) from above. We have to treat the cases $b = 0$ and $b > 0$ separately. For $b = c = 0$

$$\frac{\alpha_{bx+4}(c+1)}{\alpha_x(c+1)} = 1; \tag{2.4.5}$$

for $b = 0,\ c > 0,\ x > 4$,

$$\frac{\alpha_{bx+4}(c+1)}{\alpha_x(c+1)} < \frac{4^c}{(x-1)^c} \le 1; \tag{2.4.6}$$

for $b > 0,\ x > 16c$,

$$\begin{aligned}\frac{\alpha_{bx+4}(c+1)}{\alpha_x(c+1)} &\le \frac{(bx+4)^c}{(x-1)^c} \\ &\le \frac{\left(1+\frac{4}{x}\right)^c}{\left(1-\frac{1}{x}\right)^c} b^c \\ &\le \frac{b^c}{\left(1-\frac{1}{x}\right)^c \left(1-\frac{4}{x}\right)^c} \\ &\le \frac{b^c}{\left(1-\frac{4}{x}\right)^{2c}} \\ &\le \frac{b^c}{1-\frac{8c}{x}} \\ &\le b^c\left(1+\frac{16c}{x}\right).\end{aligned} \tag{2.4.7}$$

Thus (2.4.4) becomes valid as soon as

$$x > 16(c+1)(b+1)^c; \tag{2.4.8}$$

for example, we can take

$$x = 16(c+1)\alpha_{b+4}(c+1). \tag{2.4.9}$$

To obtain a Diophantine representation for $a = b^c$, we need to eliminate the function α from (2.4.4) and (2.4.9). For this purpose we can use three copies of the conditions (2.3.13)–(2.3.24) only, because condition (2.3.12) will be fulfilled automatically.

2.5 Exponential Diophantine equations

In what follows, an important role will be played by *exponential Diophantine equations*. These are equations of the form

$$E_1(x_1, \ldots, x_m) = E_2(x_1, \ldots, x_m), \tag{2.5.1}$$

where E_1 and E_2 are expressions constructed from variables and particular natural numbers using addition, multiplication, and exponentiation.

We do not allow the use of subtraction in exponential Diophantine equations in order to remain within the set of natural numbers and thus avoid such problems as specifying the value of

$$(x-y)^{2^{2^{x-y}}} \tag{2.5.2}$$

when $x = 1$, and $y = 3$. However, every now and then we shall use the sign "–" when writing down exponential Diophantine equations, but only in cases when this sign could be eliminated by transposing some terms to the other side or by similarly evident transformations.

Bearing this remark in mind, we see that any system of exponential Diophantine equations can be compressed into a single equation in a manner similar to passing from (1.2.1) to (1.2.2).

In analogy with Diophantine equations, one can consider parametric exponential Diophantine equations and introduce *exponential Diophantine representations of sets, properties, relations*, and *functions*. According to Section 1.5, having proved in Section 2.4 that exponentiation is Diophantine, we have a method for transforming any exponential Diophantine equation into an equivalent Diophantine equation with the same parameters, at the cost of an increase in the number of unknowns. Thus the class of sets (properties, relations, functions) having exponential Diophantine representations coincides with the class of Diophantine sets (properties, relations, functions, respectively). Following the convention in Section 1.5, we regard exponential Diophantine representations as being generalized Diophantine representations. However, exponential Diophantine representations may be of interest in themselves, because often they can be more compact than the corresponding genuine Diophantine representations. Also, exponential Diophantine representations may have certain additional properties that we still have been unable to obtain for Diophantine representations.

We can also consider a class of equations that is intermediate between exponential Diophantine equations and genuine Diophantine equations. Namely, a *unary*

exponential Diophantine equation is an exponential Diophantine equation in which only constants are raised to powers; i.e., instead of binary exponentials, only unary exponentials such as 2^c, 3^c, ... are used. If only one unary exponential, say 2^c, is used, perhaps several times, then we have a *unary exponential Diophantine equation to the base* 2.

As an example of an exponential Diophantine equation, we consider the famous Fermat equation

$$(p+1)^{s+3} + (q+1)^{s+3} = (r+1)^{s+3}. \tag{2.5.3}$$

It is written here in this form because then *Fermat's Last Theorem* is just the assertion that equation (2.5.3) has no solutions in the unknowns p, q, r, and s. Although (2.5.3) is a Diophantine equation in the unknowns p, q, and r for any fixed value of s, Fermat's Last Theorem, in its original form, is not an individual subproblem of Hilbert's Tenth Problem, because s occurs exponentially in (2.5.3). (Incidentally, Hilbert did not include Fermat's Last Theorem among his "Mathematical Problems.") However, at this point, we are able to construct a specific polynomial $\mathbf{F}$ with integer coefficients such that the equation

$$\mathbf{F}(p, q, r, s, x_1, \ldots, x_m) = 0 \tag{2.5.4}$$

is solvable in $x_1, \ldots, x_m$ if and only if p, q, r, and s satisfy (2.5.3), and hence Fermat's Last Theorem is equivalent to the assertion that (2.5.4) is an unsolvable Diophantine equation in $m+4$ unknowns. Thus, in spite of the fact that we still (1992) don't know whether Fermat's Last Theorem is true or false, we can find an individual subproblem of Hilbert's Tenth Problem to which it is equivalent.

Exercises

In Exercises 2.1–2.3, 2.8–2.10, and in Open Question 2.1, one is supposed to find constructions that are simpler than what would result from a straightforward application of the technique of Chapter 2.

1. Show that if the set (2.1.2) is Diophantine, then so is the set

$$\{\langle a_1, b_1, a_2, b_2\rangle \mid \exists n\,[a_1 = b_1^n \;\&\; a_2 = b_2^n]\}.$$

2. Show that if the set (2.1.2) is Diophantine, then so is the set (2.1.1).

3. For a fixed odd b, give a direct proof that the set

$$\{a \mid \exists n\,[a = b^n]\}$$

is Diophantine, given that it has an infinite Diophantine subset.

4. Show that for a fixed b, the set (2.1.5) is defined not only by formula (2.1.22) but also by the Pell equation

$$z^2 - \left(\left(\frac{b}{2}\right)^2 - 1\right)a^2 = 1.$$

5. To prove that exponentiation is Diophantine, we could have used the sequences defined by the relations

$$\gamma_b(0) = 0, \qquad \gamma_b(1) = 1, \qquad \gamma_b(n+2) = b\gamma_b(n+1) + \gamma_b(n),$$

where $b \geq 1$, instead of the sequences (2.1.4). Show that the set

$$\{\langle a, b\rangle \mid b \geq 1 \;\&\; \exists n\,[a = \gamma_b(n)]\},$$

which is an analog of the set (2.1.5), is Diophantine.

6. In (2.2.1) the inequality $b \geq 4$ is used (instead of $b \geq 2$) for the sake of a slight simplification (where?) of the proof. This limitation was not burdensome in (2.4.4) and (2.4.9) for achieving the main goal of the chapter. For the sake of generality, show that $\alpha_b(c)$ is a Diophantine function of two arguments defined for all $b \geq 2$ and all c.

7. In addition to (2.4.4), there is yet another, less evident, connection between the sequences α and exponentiation, namely

$$b^n \equiv \alpha_x(n+1) + (b-x)\alpha_x(n) \pmod{bx - b^2 - 1}.$$

Prove this and then use it to obtain another Diophantine representation of the set (2.1.1).

8. In (2.4.2) we used the fact (expressed by the inequalities (2.4.1)) that $\alpha_b(n)$ grows almost like b^n. It turns out that these inequalities can be replaced by much weaker ones. Let us say that a binary relation $\mathcal{J}$ *has exponential growth* when the following two conditions hold:

(a) for every u and v, $\mathcal{J}(u, v)$ implies that $v < u^u$;

(b) for every k there are u and v such that $\mathcal{J}(u,v)$ and $v > u^k$.

Find a generalized Diophantine representation of exponentiation, given that there exists a Diophantine relation of exponential growth.

9. The conditions of Exercise 2.8 can be weakened even further. Let us say that a binary relation $\mathcal{R}$ *has roughly exponential growth* if there is a number m such that

(a) for every u and v, $\mathcal{R}(u,v)$ implies that

$$v < u^{u^{\cdot^{\cdot^{u}}}},$$

where the tower of exponents is of height m.
(b) for every k there are u and v such that $\mathcal{R}(u,v)$ and $v > u^k$.

Show that exponentiation cannot be non-Diophantine if some relation of roughly exponential growth is Diophantine. (Note that here it is not required to find a corresponding generalized Diophantine representation.)

10. Show that if the equation

$$9(u^2+7v^2)^2 - 7(r^2+7s^2)^2 = 2$$

has only finitely many solutions, then there is a Diophantine relation of exponential growth.

Open questions

1. Is there a direct method for transforming a Diophantine relation of roughly exponential growth into a Diophantine relation of exponential growth?

2. Does the equation of Exercise 2.10 have an infinite number of solutions?

3. In a manner similar to (2.2.7), the kth-order recurrence relation

$$\delta(n+k) = b_{k-1}\delta(n+k-1) + \cdots + b_0\delta(n)$$

can be transformed into a first-order relation among matrices of order $k \times k$. If $b_0 = \pm 1$ then, as with (2.2.10), the determinants of the corresponding matrices are equal to $\pm c$ where the constant c is determined by $\delta(0), \ldots, \delta(k-1)$. As with (2.2.11), this condition can be stated in the form of a relation among the quantities $\delta(n), \ldots, \delta(n+k-1)$. When is it the case that this relation characterizes

the sequence, so that (as in the case of (2.2.12)) it furnishes a Diophantine equation all solutions of which are among the terms of the sequence defined by the given recurrence?

Commentary

As was stated in the Commentary to Chapter 1, the origin of systematic investigations of the class of Diophantine sets was connected with Tarski's conjecture that the set of all powers of 2 is not Diophantine. When Julia Robinson did not succeed in proving this, she began to incline to the conjecture that exponentiation is Diophantine. In an important paper [1952], she gave sufficient conditions for exponentiation to be Diophantine. In particular, she showed that it would be sufficient to find a Diophantine relation of exponential growth (see Exercise 2.8) or at least roughly exponential growth (see Exercise 2.9). Relations of exponential growth are also known as *Julia Robinson predicates* (see, for example, Davis [1963]).

Later, Robinson [1969a] found various conditions sufficient for the existence of Diophantine relations of exponential growth; namely, she showed that it would be sufficient to find an infinite Diophantine set consisting entirely of primes, or to show that the set of all powers of 2 is Diophantine.

Davis [1968] found another sufficient condition consisting in the uniqueness of the trivial solution $u = r = 1$, $v = s = 0$ of the equation from Exercise 2.10. However, Herrman [1971] established the existence of a non-trivial solution, and Shanks [1972], using a computer, also found a non-trivial solution:

$$u = 525692038369576, \quad r = 2484616164142152,$$
$$v = 1556327039191013, \quad s = 1381783865776981.$$

Nevertheless, as was mentioned in Davis, Matiyasevich, and Robinson [1976], this doesn't entirely spoil Davis's idea, because in fact it would suffice to show that the equation has only finitely many solutions (see Exercise 2.10).

These approaches have so far not led to success. Nevertheless, they remain of interest even after exponentiation was proved to be Diophantine in another way. This is because of the connection between these approaches and the so-called singlefold Diophantine representations (see Section 7.2 and the Commentary to Chapter 7). In addition to the conditions mentioned above, various other, more involved, conditions were proposed that also imply that exponentiation is Diophantine (see, for example, Davis [1962, 1966], Davis and Putnam [1958], Matiyasevich [1968b]).

The very first example of a Diophantine relation of exponential growth was published by Matiyasevich [1970]. It was the relation

$$v = \phi_{2u},$$

where $\phi_0, \phi_1, \ldots$ are the Fibonacci numbers defined by

$$\phi_0 = 0, \qquad \phi_1 = 1, \qquad \phi_{n+2} = \phi_n + \phi_{n-1}.$$

According to the above-mentioned criterion due to Julia Robinson, this implied that exponentiation was Diophantine. Chronologically, this example of a Diophantine relation of exponential growth turned out to be the last missing link in establishing the algorithmic unsolvability of Hilbert's Tenth Problem, because the algorithmic unsolvability of exponential Diophantine equations had previously been established (for more details see the Commentary to Chapter 5).

The Fibonacci numbers are a special case (namely $b = 1$) of the sequences γ_b from Exercise 2.5, which are closely related to our α_b. All of these sequences have similar properties that can be used to prove that they are Diophantine. In the case of the sequences γ_b, this was shown by Chudnovsky [1970, 1971, 1984], Davis [1971, 1973a], Kosovskiĭ [1971], and also by Simon Kochen (see Davis [1971]) and Kurt Schütte (see Robinson [1971] or Fenstad [1971]).

In Section 2.3 we did a bit more than simply finding a Diophantine relation of exponential growth, as was done in Matiyasevich [1970]. The system (2.3.12)–(2.3.24) defines a relation among three numbers (rather than two), and if this relation holds, then the numbers satisfy two-sided inequalities stronger than those required in the definition of a relation of exponential growth. This opens a somewhat shorter path (used in Section 2.4) to proving that exponentiation is Diophantine than would be obtained from a straightforward application of the Julia Robinson criterion. Yet another method, also originating from Robinson [1952], is outlined in Exercise 2.7. Diophantine representations of b^c are presented in full detail in particular by Davis [1971], Kosovskii [1971], Matiyasevich [1971a, 1971b], and Matiyasevich and Robinson [1975].

Diophantine equations of the type (2.5.4) were explicitly presented by Ruohonen [1972] and Baxa [1993]. It is highly unlikely that transforming Fermat's simple exponential Diophantine equation to an equivalent complicated Diophantine equation could be of any help in investigations on Fermat's Last Theorem. On the other hand, this reduction can be viewed as an informal "psychological" argument in favor of the unsolvability of Hilbert's Tenth Problem, because otherwise the *process*

required by the problem would permit one, in particular, to determine whether the Theorem is true or false.

Chudnovsky [1971] (cf. [1984]) states that Davis's question about the solvability of the equation from Exercise 2.10 "can be reduced to studying the arithmetical properties of the sequences (A_n, B_n) of solutions of the equation $9x^2 - 7y^2 = 2$" and by "studying these sequences one can obtain a Diophantine representation for $y = 2^x$ with $x > C$ where C is a constant. The result obtained answers Davis's question." Today we know that Davis's conjecture that the trivial solution is unique is not true, and it is not clear what Chudnovsky had in mind: did he mean that the number of solutions was finite or did he propose to use the sequences (A_n, B_n) in some other way? That is why Question 2.2 is stated as being open.

An answer to Open Question 2.3 will most likely be connected with an analysis of the multiplicative group of units of the field $\mathbb{Q}(\chi)$, where

$$\chi^k = b_{k-1}\chi^{k-1} + \cdots + b_0,$$

and most likely for the answer to be positive it is necessary that $b_0 = \pm 1$ and $k \leq 4$, and also in the case $k = 4$ that the equation have no real roots, while in the case $k = 3$ that it have only one real root (because, by Dirichlet's Theorem, it is only under these conditions that the field has a unique fundamental unit).