Chapter IX

Integral Points on Elliptic Curves

Many elliptic curves have infinitely many rational points, although the Mordell–Weil theorem assures us that the group of rational points is finitely generated. Another natural Diophantine question is that of determining how many of the rational points on a given (affine) Weierstrass equation have integral coordinates. In this chapter we prove a theorem of Siegel that says that there are only finitely many such integral points. Siegel gave two proofs of his theorem, which we present in (IX §3) and (IX §4). Both proofs make use of techniques from the theory of Diophantine approximation, and thus do not provide an effective procedure for actually finding all of the integral points. However, Siegel's second proof reduces the problem to that of solving the so-called unit equation, which in turn can be effectively resolved using methods from transcendence theory. We discuss effective solutions, without giving proofs, in (IX §5).

Unless otherwise specified, the notation and conventions for this chapter are the same as those for Chapter VIII. In addition, we set the following notation:

H, H_K height functions, see (VIII §5).

n_v $= [K_v : \mathbb{Q}_v]$, the local degree for $v \in M_K$, see (VIII §5).

S $\subset M_K$, generally a finite set of absolute values containing M_K^∞.

R_S the ring of S-integers of K,

$$R_S = \{x \in K : v(x) \ge 0 \text{ for all } v \in M_K \text{ with } v \notin S\}.$$

R_S^* the unit group of R_S.

J.H. Silverman, *The Arithmetic of Elliptic Curves, Second Edition*, Graduate Texts in Mathematics 106, DOI 10.1007/978-0-387-09494-6_IX,

IX.1 Diophantine Approximation

The fundamental problem in the subject of Diophantine approximation is the question of how closely an irrational number can be approximated by a rational number.

Example 1.1. For any rational number p/q, we know that the quantity $\bigl|p/q - \sqrt{2}\,\bigr|$ is strictly positive, and since $\mathbb{Q}$ is dense in $\mathbb{R}$, an appropriate choice of p/q makes it as small as desired. The problem is to make it small without taking p and q to be too large. The next two elementary results illustrate this idea.

Proposition 1.2. (Dirichlet) *Let $\alpha \in \mathbb{R}$ with $\alpha \notin \mathbb{Q}$. Then there are infinitely many rational numbers $p/q \in \mathbb{Q}$ such that*

$$\left|\frac{p}{q} - \alpha\right| \le \frac{1}{q^2}.$$

PROOF. Let Q be a (large) integer and look at the set of real numbers

$$\bigl\{q\alpha - [q\alpha] : q = 0, 1, \ldots, Q\bigr\},$$

where $[\ \cdot\ ]$ denotes greatest integer. Since α is irrational, this set contains $Q+1$ distinct numbers in the interval between 0 and 1. Dividing the interval $[0,1]$ into Q equal-sized pieces and applying the pigeonhole principle, we find that there are integers $0 \le q_1 < q_2 \le Q$ satisfying

$$\Bigl|\bigl(q_1\alpha - [q_1\alpha]\bigr) - \bigl(q_2\alpha - [q_2\alpha]\bigr)\Bigr| \le \frac{1}{Q}.$$

Hence

$$\left|\frac{[q_2\alpha] - [q_1\alpha]}{q_2 - q_1} - \alpha\right| \le \frac{1}{(q_2 - q_1)Q} \le \frac{1}{(q_2 - q_1)^2}.$$

This provides one rational approximation to α having the desired property.

Finally, having obtained a list of such approximations, let p/q be the one for which $|p/q - \alpha|$ is smallest. Then taking $Q > |p/q - \alpha|^{-1}$ ensures that we get a new approximation that is not already in our list. Hence there exist infinitely many rational numbers satisfying the conditions of the proposition. □

Remark 1.2.1. A result of Hurwitz says that the $1/q^2$ on the right-hand side of (IX.1.2) may be replaced by $1/(\sqrt{5}\,q^2)$, and that this result is best possible. See, e.g., [108, Theorem 194].

Proposition 1.3. (Liouville [151]) *Let $\alpha \in \bar{\mathbb{Q}}$ have degree $d \ge 2$ over $\mathbb{Q}$, i.e., $\bigl[\mathbb{Q}(\alpha) : \mathbb{Q}\bigr] = d$. There is a constant $C > 0$, depending on α, such that for all rational numbers p/q we have*

$$\left|\frac{p}{q} - \alpha\right| \ge \frac{C}{q^d}.$$

PROOF. We may assume that $\alpha \in \mathbb{R}$, since otherwise $C = \mathrm{Im}(\alpha)$ works. Let

$$f(T) = a_0T^d + a_1T^{d-1} + \cdots + a_d \in \mathbb{Z}[T]$$

be a minimal polynomial for α, and let

$$C_1 = \sup\bigl\{f'(t) : \alpha - 1 \le t \le \alpha + 1\bigr\}.$$

Then the mean value theorem tells us that

$$\left|f\left(\frac{p}{q}\right)\right| = \left|f\left(\frac{p}{q}\right) - f(\alpha)\right| \le C_1\left|\frac{p}{q} - \alpha\right|.$$

On the other hand, we know that $q^d f(p/q) \in \mathbb{Z}$, and further that $f(p/q) \ne 0$, since f has no rational roots. Hence

$$\left|q^d f\left(\frac{p}{q}\right)\right| \ge 1.$$

Setting $C = \min\{C_1^{-1}, 1\}$ and combining the last two inequalities yields

$$\left|\frac{p}{q} - \alpha\right| \ge \frac{C}{q^d} \qquad \text{for all } p/q \in \mathbb{Q}. \qquad \square$$

Remark 1.3.1. Liouville used his theorem to prove the existence of transcendental numbers; see Exercise 9.2. Note that in Liouville's theorem it is quite easy to find a value for the constant C explicitly in terms of α. This is in marked contrast to the results that we consider in the rest of this section.

Dirichlet's theorem (IX.1.2) says that every real number can be approximated by rational numbers to within $1/q^2$, while Liouville's result (IX.1.3) says that algebraic numbers of degree d can be approximated no closer than C/q^d. For quadratic irrationalities there is little more to say, but if $d \ge 3$, then it is natural to ask for the best exponent on q. There is no particular reason to restrict the approximating values to $\mathbb{Q}$, so we allow them to vary over any fixed number field K. Finally, in measuring the closeness of the approximation, we may use any absolute value on K.

Definition. Let $\tau(d)$ be a positive real-valued function on the natural numbers. A number field K is said to have *approximation exponent* τ if it has the following property:

> Let $\alpha \in \bar{K}$, let $d = \bigl[K(\alpha) : K\bigr]$, and let $v \in M_K$ be an absolute value on K that has been extended to $K(\alpha)$ in some fashion. Then for any constant C there exist only finitely many $x \in K$ satisfying the inequality
>
> $$|x - \alpha|_v < CH_K(x)^{-\tau(d)}.$$

Liouville's elementary estimate (IX.1.3) says that $\mathbb{Q}$ has approximation exponent $\tau(d) = d + \epsilon$ for any $\epsilon > 0$. This result has been successively improved by a number of mathematicians:

Liouville	1851	$\tau(d) = d + \epsilon$
Thue	1909	$\tau(d) = \frac{1}{2}d + 1 + \epsilon$
Siegel	1921	$\tau(d) = 2\sqrt{d} + \epsilon$
Gelfond, Dyson	1947	$\tau(d) = \sqrt{2d} + \epsilon$
Roth	1955	$\tau(d) = 2 + \epsilon$

In view of (IX.1.2), Roth's result is essentially best possible, although it has been conjectured that the ϵ can be replaced by some function $\epsilon(d)$ such that $\epsilon(d) \to 0$ as $d \to \infty$. We should also mention that Mahler showed how to handle several absolute values at once, and W. Schmidt [221, Chapter VI] dealt with the more difficult problem of simultaneously approximating several irrationals.

The main ideas that go into the proof of Roth's theorem are quite beautiful, and at least in theory, relatively elementary. Unfortunately, to develop these ideas fully would take us rather far afield. Hence rather than including a complete proof, we are content to state here the result that we will need. In (IX §8) we briefly sketch the proof of Roth's theorem without giving any of the myriad details.

Theorem 1.4. (Roth's Theorem) *For every $\epsilon > 0$, every number field K of degree d has approximation exponent*

$$\tau(d) = 2 + \epsilon.$$

PROOF. See (IX §8) for a brief sketch of the proof. A nice exposition for $K = \mathbb{Q}$ and the usual archimedean absolute value is given in [221, Chapter V]. For the general case, see [114, Part D] or [139, Chapter 7]. □

Example 1.5. How do theorems on Diophantine approximation lead to results about Diophantine equations? Consider the simple example of trying to solve the equation

$$x^3 - 2y^3 = a$$

in integers $x, y \in \mathbb{Z}$, where $a \in \mathbb{Z}$ is fixed. Suppose that (x, y) is a solution with $y \neq 0$. Let ζ be a primitive cube root of unity, and factor the equation as

$$\left(\frac{x}{y} - \sqrt[3]{2}\right)\left(\frac{x}{y} - \zeta\sqrt[3]{2}\right)\left(\frac{x}{y} - \zeta^2\sqrt[3]{2}\right) = \frac{a}{y^3}.$$

The second and third factors in the product are bounded away from 0, so we obtain an estimate of the form

$$\left|\frac{x}{y} - \sqrt[3]{2}\right| \le \frac{C}{y^3},$$

where the constant C is independent of x and y. Now (XI.1.4), or even Thue's original theorem with $\tau(d) = \frac{1}{2}d + 1 + \epsilon$, shows that there are only finitely many possibilities for x and y. Hence the equation

$$x^3 - 2y^3 = a$$

has only finitely many solutions in integers. This type of argument will reappear in the proof of (IX.4.1); see also Exercise 9.6.

Remark 1.6. The statement of (IX.1.4) says that *there exist* only finitely many elements of K having a certain property. This phrasing is felicitous because the proof of (IX.1.4) is not effective. In other words, the proof does not give an effective procedure that is guaranteed to produce all of the elements in the finite set. (See (IX.8.1) for a discussion of why this is so.) We note that as a consequence, all of the finiteness results that we prove in (IX §§2, 3) are ineffective, since they rely on (IX.1.4). Similarly, the proof in (IX.1.5) yields no explicit bound for $|x|$ and $|y|$ in terms of a. However, there are other methods, based on estimates for linear forms in logarithms, that are effective. We discuss such methods, without proof, in (IX §5).

IX.2 Distance Functions

A Diophantine inequality such as

$$|x-\alpha|_v < CH_K(x)^{-\tau(d)}$$

consists of two pieces. First, there is the height function $H_K(x)$, which measures the *arithmetic* size of x. We have already studied height functions and their transformation properties in some detail (VIII, §§5, 6). Second, there is the quantity $|x-\alpha|_v$, which is a *topological* or *metric* measure of the distance from x to α, i.e., it measures distance in the v-adic topology. In this section we define a notion of v-adic distance on curves, deduce some of its basic properties, and reinterpret the main Diophantine approximation result from (IX §1) in terms of this distance function.

Definition. Let C/K be a curve, let $v \in M_K$, and fix a point $Q \in C(K_v)$. Choose a function $t_Q \in K_v(C)$ that has a zero of order $e \ge 1$ at Q and no other zeros.[1] Then for $P \in C(K_v)$, we define the (v-*adic*) *distance from* P *to* Q by

$$d_v(P,Q) = \min\left\{ \bigl|t_Q(P)\bigr|_v^{1/e}, 1 \right\}.$$

(If t_Q has a pole at P, we formally set $\bigl|t_Q(P)\bigr| = \infty$, so $d_v(P,Q) = 1$.)

Remark 2.1. In practice, we fix the point Q and use the distance function $d_v(P,Q)$ to measure the distance from P to Q as P varies. It is clear that the distance function d_v has the right qualitative property, i.e., $d_v(P,Q)$ is small if P is v-adically close to Q. On the other hand, the value of $d_v(P,Q)$ certainly depends on the choice of the function t_Q, so possibly a better notation would be $d_v(P,t_Q)$. However, since we will use d_v only to measure the rate at which a varying point approaches a fixed point, the next result shows that the choice of t_Q is irrelevant for the statements of our theorems.

Proposition 2.2. *Let $Q \in C(K_v)$ and let $F \in K_v(C)$ be a function that vanishes at Q. Then the limit*

[1] To see that t_Q exists, we use the the Riemann–Roch theorem. Thus (II.5.5c) tells us that if C has genus g and if $e \ge g+1$, then $\ell\bigl(e(Q)\bigr) \ge 2$, so there is a nonconstant function $f \in \mathcal{L}\bigl(e(Q)\bigr)$. This function f has a pole at Q and no other poles, and we can take $t_Q = 1/f$.

$$\lim_{\substack{P\in C(K_v)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|F(P)\bigr|_v}{\log d_v(P,Q)} = \operatorname{ord}_Q(F)$$

exists and is independent of the choice of the function t_Q used to define $d_v(P,Q)$.

Here $\operatorname{ord}_Q(F)$ is the order of vanishing of F at Q as in (II §2)*, while the notation $P\underset{v}{\to}Q$ means that $P \in C(K_v)$ approaches Q in the v-adic topology, i.e., $d_v(P,Q)\to 0$.*

PROOF. Let t_Q be the function vanishing only at Q that we are using to define $d_v(\,\cdot\,,Q)$. Let $e=\operatorname{ord}_Q(t_Q)$ and $f=\operatorname{ord}_Q(F)$. Then the function $\phi = F^e/t_Q^f$ has neither a zero nor a pole at Q, so $\bigl|\phi(P)\bigr|_v$ is bounded away from 0 and ∞ as $P\underset{v}{\to}Q$. Hence

$$\begin{aligned}
\lim_{\substack{P\in C(K_v)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|F(P)\bigr|_v}{\log d_v(P,Q)} &= \lim_{\substack{P\in C(K_v)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|F(P)\bigr|_v}{\log\bigl|t_Q(P)\bigr|_v^{1/e}} \\
&= f + \lim_{\substack{P\in C(K_v)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|\phi(P)\bigr|_v}{\log\bigl|t_Q(P)\bigr|_v} \\
&= f. \qquad \square
\end{aligned}$$

Remark 2.2.1. The use of the function t_Q in the definition of distance is somewhat artificial and does not generalize well to higher-dimensional varieties. An alternative definition that does generalize uses a finite list of functions $t_1,\dots,t_r\in K(E)$ with the property that each t_i vanishes at Q and such that $t_1,\dots,t_r$ have no other common zeros. Then, if we let e_i denote the order of vanishing of t_i at Q, a distance function d_v may be defined by

$$d_v(P,Q) = \min\Bigl\{\max\bigl\{|t_1(P)|_v^{1/e_1},\dots,|t_r(P)|_v^{1/e_r}\bigr\},1\Bigr\}.$$

This function is an example of a local height function; see [139, Chapter 10], [114, §B.8], or [261] for further details.

Next we examine the effect of finite maps on the distance between points. The crucial observation is that this effect depends on the ramification of the map, not on its degree. To see the difference, compare (IX.2.3) with (VIII.5.6).

Proposition 2.3. *Let C_1/K and C_2/K be curves, and let $\phi: C_1\to C_2$ be a finite map defined over K. Let $Q\in C_1(K_v)$, and let $e_\phi(Q)$ be the ramification index of ϕ at Q* (II §2)*. Then*

$$\lim_{\substack{P\in C_1(K_v)\\ P\underset{v}{\to}Q}} \frac{\log d_v\bigl(\phi(P),\phi(Q)\bigr)}{\log d_v(P,Q)} = e_\phi(Q).$$

PROOF. Let $t_Q \in K_v(C_1)$ be a function that vanishes to order $e_1 \ge 1$ at Q and has no other zeros, and similarly let $t_{\phi(Q)} \in K_v(C_2)$ be a function that vanishes to order $e_2 \ge 1$ at $\phi(Q)$ and has no other zeros. It follows from the definition of ramification index that

$$\operatorname{ord}_Q t_{\phi(Q)} \circ \phi = e_\phi(Q) \operatorname{ord}_{\phi(Q)} t_{\phi(Q)} = e_\phi(Q) e_2,$$

so the functions $(t_{\phi(Q)} \circ \phi)^{e_1}$ and $t_Q^{e_\phi(P)e_2}$ vanish to the same order at Q. Hence the function

$$f = \frac{(t_{\phi(Q)} \circ \phi)^{e_1}}{t_Q^{e_\phi(Q)e_2}} \in K_v(C_1)$$

has neither a zero nor a pole at Q. It follows that $\bigl|f(P)\bigr|_v$ is bounded away from 0 and ∞ as $P \underset{v}{\to} Q$. Therefore

$$\begin{aligned}
\frac{\log d_v\bigl(\phi(P),\phi(Q)\bigr)}{\log d_v(P,Q)} &= \frac{\log\bigl|t_{\phi(Q)}\bigl(\phi(P)\bigr)\bigr|_v^{1/e_2}}{\log\bigl|t_Q(P)\bigr|_v^{1/e_1}} \\
&= \frac{e_\phi(Q)\log\bigl|t_Q(P)\bigr|_v^{1/e_1} + \log\bigl|f(P)\bigr|_v}{\log\bigl|t_Q(P)\bigr|_v^{1/e_1}} \\
&\longrightarrow e_\phi(Q) \qquad \text{as } P \underset{v}{\to} Q.
\end{aligned}$$

□

Finally, we reinterpret Roth's theorem (IX.1.4) in terms of distance functions.

Corollary 2.4. (of (IX.1.4)) *Fix an absolute value $v \in M_K$. Let C/K be a curve, let $f \in K(C)$ be a nonconstant function, and let $Q \in C(\bar{K})$. Then*

$$\liminf_{\substack{P\in C(K)\\ P\underset{v}{\to}Q}} \frac{\log d_v(P,Q)}{\log H_K\bigl(f(P)\bigr)} \ge -2.$$

(*If Q is not a v-adic accumulation point of $C(K)$, then we define the* $\liminf$ *to be* 0.)

PROOF. Replacing f by $1/f$ if necessary, we may assume that $f(Q) \ne \infty$. (Note that $H_K\bigl((1/f)(P)\bigr) = H_K\bigl(f(P)\bigr)$.) The function $f - f(Q)$ vanishes at Q, say to order e, so (IX.2.2) tells us that

$$\liminf_{\substack{P\in C(K)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|f(P)-f(Q)\bigr|_v}{\log d_v(P,Q)} = e.$$

Hence

$$\begin{aligned}
\liminf_{\substack{P\in C(K)\\ P\underset{v}{\to}Q}} \frac{\log d_v(P,Q)}{\log H_K\bigl(f(P)\bigr)} &= \liminf_{\substack{P\in C(K)\\ P\underset{v}{\to}Q}} \frac{\log\bigl|f(P)-f(Q)\bigr|_v}{e\log H_K\bigl(f(P)\bigr)} \\
&= \frac{1}{e}\liminf_{\substack{P\in C(K)\\ P\underset{v}{\to}Q}} \left(\frac{\log\bigl(H_K\bigl(f(P)\bigr)^\tau\bigl|f(P)-f(Q)\bigr|_v\bigr)}{\log H_K\bigl(f(P)\bigr)} - \tau\right).
\end{aligned}$$

We now set $\tau = 2 + \epsilon$. Then (IX.1.4) implies that

$$H_K\bigl(f(P)\bigr)^{\tau}\bigl|f(P) - f(Q)\bigr|_v \ge 1$$

for all but finitely many $P \in C(K)$. Therefore

$$\liminf_{\substack{P\in C(K)\\ P\xrightarrow[v]{}Q}} \frac{\log d_v(P,Q)}{\log H_K\bigl(f(P)\bigr)} \ge -\frac{\tau}{e} \ge -\frac{2+\epsilon}{e}.$$

Since $\epsilon > 0$ is arbitrary and $e \ge 1$, this is the desired result. □

IX.3 Siegel's Theorem

In this section we prove a result of Siegel that represents a significant improvement on the Diophantine approximation result (IX.2.4).

Theorem 3.1. (Siegel) *Let E/K be an elliptic curve with $\#E(K) = \infty$. Fix a point $Q \in E(\bar{K})$, a nonconstant even function $f \in K(E)$, and an absolute value $v \in M_{K(Q)}$. Then*

$$\lim_{\substack{P\in E(K)\\ h_f(P)\to\infty}} \frac{\log d_v(P,Q)}{h_f(P)} = 0.$$

Remark 3.1.1. Although we prove (IX.3.1) only for even functions, it is in fact true in general; see Exercise 9.14d.

Before proving (IX.3.1), we give some indication of its power.

Corollary 3.2.1. *Let E/K be an elliptic curve with Weierstrass coordinate functions x and y, let $S \subset M_K$ be a finite set of places containing M_K^∞, and let R_S be the ring of S-integers of K. Then*

$$\bigl\{P \in E(K) : x(P) \in R_S\bigr\}$$

is a finite set.

PROOF. We apply (IX.3.1) with the function $f = x$. Suppose that there is a sequence of distinct points $P_1, P_2, \ldots \in E(K)$ with every $x(P_i) \in R_S$. The definition of height then tells us that

$$h_x(P_i) = \frac{1}{[K:\mathbb{Q}]} \sum_{v\in S} \log\max\bigl\{1, \bigl|x(P_i)\bigr|_v^{n_v}\bigr\},$$

since the terms with $v \notin S$ have $\bigl|x(P_i)\bigr|_v \le 1$. Hence we can find a particular $v \in S$ and a subsequence of the P_i (which we relabel as $P_1, P_2, \ldots$) such that

$$h_x(P_i) \le \#S \cdot \log\bigl|x(P_i)\bigr|_v \qquad \text{for all } i = 1, 2, \ldots.$$

(Note that $n_v \le [K:\mathbb{Q}]$.) In particular, we see that $\bigl|x(P_i)\bigr|_v \to \infty$, and since O is the only pole of x, it follows that $d_v(P_i,O) \to 0$.

The function x has a pole of order 2 at O and no other poles, so we may take as our distance function

$$d_v(P_i,O) = \min\bigl\{\bigl|x(P_i)\bigr|_v^{-1/2}, 1\bigr\}.$$

Then, for all sufficiently large i, we have

$$\frac{-\log d_v(P_i,O)}{h_x(P_i)} \ge \frac{1}{2\#S}.$$

This contradicts (IX.3.1), which says that the left-hand side approaches 0 as $i \to \infty$. □

It is clear that the proof of (IX.3.2.1) works for any even function, not just x, since (IX.3.1) is given for all even functions. However, it is possible to reduce the case of arbitrary (not necessarily even) functions to the special case given in (IX.3.2.1). This reduction step, which we now give, is important in its own right, since it is used both in Siegel's second proof of finiteness (IX.4.3.1) and with the effective methods provided by linear forms in logarithms (IX.5.7).

Corollary 3.2.2. *Let C/K be a curve of genus one, let $f \in K(C)$ be a nonconstant function, and let S and R_S be as in* (IX.3.2.1)*. Then*

$$\bigl\{P \in C(K) : f(P) \in R_S\bigr\}$$

is a finite set. Further, (IX.3.2.2) *follows formally from* (IX.3.2.1)*.*

PROOF. We are clearly proving something stronger if we extend the field K and enlarge the set S. We may thus assume that $C(K)$ contains a pole Q of f, and taking Q to be the identity element, we view (C,Q) as an elliptic curve defined over K. Let x and y be coordinates on a Weierstrass equation for (C,Q), which we may take in the form

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

We have $f \in K(C) = K(x,y)$ and $\bigl[K(x,y) : K(x)\bigr] = 2$, so we can write

$$f(x,y) = \frac{\phi(x) + \psi(x)y}{\eta(x)}$$

with polynomials $\phi(x), \psi(x), \eta(x) \in K[x]$. Further, since

$$\operatorname{ord}_Q(x) = -2, \qquad \operatorname{ord}_Q(y) = -3, \qquad \text{and} \qquad \operatorname{ord}_Q(f) < 0,$$

it follows that

$$2\deg\eta < \max\{2\deg\phi, 2\deg\psi + 3\}.$$

(This is the condition for f to have a pole at Q.) Next we compute

$$\big(f\eta(x) - \phi(x)\big)^2 = \big(\psi(x)y\big)^2 = \psi(x)^2(x^3 + Ax + B).$$

Writing this out as a polynomial in x with coefficients in $K[f]$, we see that the highest power of x comes from one of the three terms $f^2\eta(x)^2$, $\phi(x)^2$, $\psi(x)^2x^3$. From above, the first of these has lower degree in x than the latter two, while the leading terms of $\phi(x)^2$ and $\psi(x)^2x^3$ cannot cancel, since they have different degrees. (One has even degree, the other odd degree.) It follows that x satisfies a *monic* polynomial with coefficients in $K[f]$, i.e., x is integral over $K[f]$. Multiplying this monic polynomial by an appropriate element of K to "clear denominators," we have shown that x satisfies a relation

$$a_0x^N + a_1(f)x^{N-1} + \cdots + a_{N-1}(f)x + a_N(f) = 0,$$

where $a_0 \in R_S$ is nonzero and $a_i(f) \in R_S[f]$ for $1 \le i \le N$. Enlarging the set S, we may assume that $a_0 \in R_S^*$, and then dividing the polynomial by a_0, we may assume that $a_0 = 1$.

Now suppose that $P \in C(K)$ satisfies $f(P) \in R_S$. Then P is not a pole of x, and the relation

$$x(P)^N + a_1\big(f(P)\big)x(P)^{N-1} + \cdots + a_{N-1}\big(f(P)\big)x(P) + a_N\big(f(P)\big) = 0$$

shows that $x(P)$ is integral over R_S. Since also $x(P) \in K$ and R_S is integrally closed, it follows that $x(P) \in R_S$. This proves that

$$\big\{P \in C(K) : f(P) \in R_S\big\} \subset \big\{P \in C(K) : x(P) \in R_S\big\},$$

and thus the finiteness assertion in (IX.3.2.1) implies the desired finiteness result described in (IX.3.2.2). □

Example 3.3. Consider the Diophantine equation

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

where $A, B \in \mathbb{Z}$ and $4A^3 + 27B^2 \ne 0$. The corollary (IX.3.2.1) says that this equation has only finitely many solutions $x, y \in \mathbb{Z}$. What does (IX.3.1) say in this situation, say if we take $Q = O$, $f = x$, and v the archimedean absolute value on $\mathbb{Q}$?

Label the nonzero rational points $P_1, P_2, \ldots \in E(\mathbb{Q})$ in order of nondecreasing height, and write

$$x(P_i) = \frac{a_i}{b_i} \in \mathbb{Q}$$

as a fraction in lowest terms. Then

$$\log d_v(P_i, O) = \frac{1}{2}\log\min\left\{\left|\frac{b_i}{a_i}\right|, 1\right\},$$
$$h_x(P_i) = \log\max\big\{|a_i|, |b_i|\big\}.$$

(Note that the $\frac{1}{2}$ appears because x^{-1} has a zero of order 2 at O.) We see from (IX.3.1) that

$$\lim_{i\to\infty} \frac{\min\bigl\{\log|b_i/a_i|,0\bigr\}}{\max\bigl\{\log|a_i|,\log|b_i|\bigr\}} = 0.$$

Next let Q_1 and Q_2 be the zeros of the function x, where we allow $Q_1 = Q_2$. Then it is not hard to check that

$$\log\min\bigl\{\bigl|x(P)\bigr|_v,1\bigr\} = \log d_v(P,Q_1)+\log d_v(P,Q_2)+O(1) \quad \text{for all } P\in E(K_v),$$

where the $O(1)$ depends on the choice of the distance functions $d_v(\,\cdot\,,Q_i)$, but is independent of P; see Exercise 9.16. Writing $v\in M_{\mathbb{Q}}^{\infty}$ for the usual archimedean absolute value on $\mathbb{Q}$, we use (IX.3.1) twice to obtain

$$\begin{aligned}
\lim_{i\to\infty} \frac{\min\bigl\{\log|a_i/b_i|,0\bigr\}}{\max\bigl\{\log|a_i|,\log|b_i|\bigr\}} &= \lim_{i\to\infty} \frac{\log\min\bigl\{\bigl|x(P_i)\bigr|,1\bigr\}}{h_x(P_i)} \\
&= \lim_{i\to\infty} \frac{\log d_v(P_i,Q_1)+\log d_v(P_i,Q_2)+O(1)}{h_x(P_i)} \\
&= 0.
\end{aligned}$$

Finally, combining the limit involving b_i/a_i with the limit involving a_i/b_i, it is easy to deduce that

$$\lim_{i\to\infty} \frac{\log|a_i|}{\log|b_i|} = 1.$$

In other words, when looking at the x-coordinates of the rational points on an elliptic curve, we will see that the numerators and the denominators tend to have about the same number of digits. This is a much stronger assertion than (IX.3.2.1), which merely says that there are only finitely many points whose denominator is 1.

Remark 3.4. Siegel's theorem (IX.3.2.1) is not effective, which means that the proof does not give an explicitly computable upper bound for the height of all integral points. However, Siegel's proof can be made quantitative in the following sense; see for example [81]:

> Given a nonsingular Weierstrass equation with coefficients in a number field K and given a finite set of absolute values S, there is a constant N, which can be explicitly calculated in terms of the field K, the set S, and the coefficients of the equation, such that the equation has no more than N integral solutions.

A subtler Diophantine problem, motivated by work of Dem'janenko and posed as a general conjecture by Serge Lang, is to give an intrinsic relationship between the number of integral points and the rank of the Mordell–Weil group.

Conjecture 3.5. (Lang [135, page 140]) *Let E/K be an elliptic curve, and choose a quasiminimal Weierstrass equation for E/K,*

$$E : y^2 = x^3 + Ax + B.$$

(*See Exercise 8.14*c.) *Let $S\subset M_K$ be a finite set of places containing M_K^{∞}, and let R_S be the ring of S-integers of K. There exists a constant C,* depending only on K, *such that*

$$\#\bigl\{P\in E(K) : x(P)\in R_S\bigr\} \le C^{\#S+\operatorname{rank}E(K)}.$$

This conjecture is known to be true if one restricts attention to elliptic curves having integral j-invariant. More generally, the following is known.

Theorem 3.6. *Let E/K, S, and R_S be as in* (IX.3.5).

(a) (Silverman [104, 262]) *There is a constant C, depending only on $[K:\mathbb{Q}]$ and on the number of places $v \in M_K^0$ with $\operatorname{ord}_v(j_E) < 0$, such that*

$$\#\bigl\{P \in E(K) : x(P) \in R_S\bigr\} \le C^{\#S+\operatorname{rank} E(K)}.$$

(b) (Hindry–Silverman [113]) *Assume that the ABC conjecture* (*with any exponent*) (VIII.11.4)*,* (VIII.11.6) *is true for the field K. Then there is a constant C, depending only on $[K:\mathbb{Q}]$ and on the constants appearing in the ABC conjecture, such that*

$$\#\bigl\{P \in E(K) : x(P) \in R_S\bigr\} \le C^{\#S+\operatorname{rank} E(K)}.$$

We turn now to the proof of (IX.3.1). In broad outline, the argument goes as follows. Our theorem on Diophantine approximation (IX.2.4) gives us a bound, in terms of the height of P, on how fast P can approach Q. Suppose now that we write $P = [m]P' + R$ and $Q = [m]Q' + R$. Then (IX.2.3) tells us that the distance from P' to Q' is about the same as the distance from P to Q, since the map $P \mapsto [m]P + R$ is unramified. On the other hand, the height of P' is much smaller than the height of P. Now applying (IX.2.4) to P' and Q' gives an improved estimate, and taking m sufficiently large gives the desired result.

PROOF OF (IX.3.1). Choose a sequence of distinct points $P_i \in E(K)$ satisfying

$$\lim_{i\to\infty} \frac{\log d_v(P_i,Q)}{h_f(P_i)} = L = \liminf_{\substack{P\in E(K)\\ h_f(P)\to\infty}} \frac{\log d_v(P,Q)}{h_f(P)}.$$

Since $d_v(P,Q) \le 1$ and $h_f(P) \ge 0$ for all points $P \in E(K)$, we have $L \le 0$. It thus suffices to prove that $L \ge 0$.

Let m be a large integer. From the weak Mordell–Weil theorem (VIII.1.1), the quotient group $E(K)/mE(K)$ is finite. Hence some coset contains infinitely many of the P_i. Replacing $\{P_i\}$ by a subsequence, we may assume that

$$P_i = [m]P_i' + R,$$

where $P_i', R \in E(K)$ and where R does not depend on i. We use standard properties of height functions to compute

$$\begin{aligned} m^2 h_f(P_i') &= h_f\bigl([m]P_i'\bigr) + O(1) && \text{using (VIII.6.4b),}\\ &= h_f(P_i - R) + O(1) &&\\ &\le 2h_f(P_i) + O(1) && \text{using (VIII.6.4a).} \end{aligned}$$

Note that the $O(1)$ is independent of i.

We next do an analogous computation with distance functions. If P_i is bounded away from Q in the v-adic topology, then $\log d_v(P_i,Q)$ is bounded, so clearly $L=0$. Otherwise we can replace P_i with a subsequence such that $P_i \underset{v}{\to} Q$. It follows that $[m]P_i' \underset{v}{\to} Q-R$, so the sequence P_i' accumulates to at least one of the m^2 possible m^{th} roots of $Q-R$. Again taking a subsequence, we can find a point $Q'\in E(\bar{K})$ satisfying

$$P_i' \underset{v}{\longrightarrow} Q' \qquad \text{and} \qquad Q=[m]Q'+R.$$

We next observe that the map $E\to E$ defined by $P\mapsto [m]P+R$ is everywhere unramified (III.4.10c), so (IX.2.3) tells us that

$$\lim_{i\to\infty} \frac{\log d_v(P_i,Q)}{\log d_v(P_i',Q')} = 1.$$

Combining this with the height inequality yields

$$L=\lim_{i\to\infty} \frac{\log d_v(P_i,Q)}{h_f(P_i)} \ge \lim_{i\to\infty} \frac{\log d_v(P_i',Q')}{\frac{1}{2}m^2 h_f(P_i')+O(1)}.$$

(Note that the $\log d_v$ expressions are negative, which reverses the inequality.)

We now apply the theorem on Diophantine approximation (IX.2.4) to the sequence $\{P_i'\}\subset E(K)$ as it converges v-adically to $Q'\in E(\bar{K})$. This yields

$$\liminf_{i\to\infty} \frac{\log d_v(P_i',Q')}{[K:\mathbb{Q}]h_f(P_i')} \ge -2.$$

(The factor of $[K:\mathbb{Q}]$, which in any case is not important, arises because h_f is the absolute height, while (IX.2.4) is stated using the relative height H_K.) Combining the last two inequalities yields

$$L\ge -\frac{4[K:\mathbb{Q}]}{m^2}.$$

The field K is fixed, while the value of m is arbitrary, which completes the proof that $L\ge 0$. □

IX.4 The S-Unit Equation

The finiteness of S-integral points on elliptic curves (IX.3.2.1) is a special case of Siegel's general result that an (affine) curve C/K of genus at least one has only finitely many S-integral points; see [114, Theorem D.9.1] or [139, Chapter 8, Theorem 2.4]. Of course, for curves C of genus two or greater, Siegel's result is superseded by Faltings' theorem [82, 84], which asserts that the full set of rational points $C(K)$ is finite.

Siegel gave a second proof of his theorem that applies to a restricted set of curves, but that does include all elliptic curves. This second method is important because, when combined with results from linear forms in logarithms (XI §5), it leads to

an effective procedure for finding all S-integral points. In this section we describe Siegel's alternative proof.

The idea is to reduce the problem of solving for S-integral points on a curve to the problem of solving several equations of the form

$$ax + by = 1$$

in S-units. We start with a quick sketch of how solving this S-unit equation can be reduced to a Diophantine approximation theorem such as (IX.1.4). This ineffective theorem can then be replaced by an effective estimate as described in (IX §5).

Theorem 4.1. *Let $S \subset M_K$ be a finite set of places, and let $a, b \in K^*$. Then the equation*

$$ax + by = 1$$

has only finitely many solutions in S-units $x, y \in R_S^*$.

INEFFECTIVE PROOF (SKETCH). Let m be a large integer. Dirichlet's S-unit theorem [142, V §1] implies that the quotient group $R_S^*/(R_S^*)^m$ is finite, so we can choose a finite set of coset representatives $c_1, \ldots, c_r \in R_S^*$. Then any solution (x, y) to the original equation can be written as

$$x = c_i X^m, \qquad y = c_j Y^m,$$

for some $X, Y \in R_S^*$ and some choice of c_i and c_j, and thus (X, Y) is a solution to the equation

$$ac_i X^m + bc_j Y^m = 1.$$

Since there are only finitely many choices for c_i and c_j, it suffices to prove that for any $\alpha, \beta \in K^*$, the equation

$$\alpha X^m + \beta Y^m = 1$$

has only finitely many solutions $X, Y \in R_S$.

Suppose that there are infinitely many such solutions. Then, since

$$H_K(Y) = \prod_{v \in S} \max\bigl\{1, |Y|_v^{n_v}\bigr\},$$

we can choose some $v \in S$ so that there are infinitely many solutions satisfying

$$|Y|_v \ge H_K(Y)^{1/([K:\mathbb{Q}]\#S)}.$$

(Note that $n_v \le [K : \mathbb{Q}]$.) Let $\gamma \in \bar{K}$ be a solution to

$$\gamma^m = -\beta/\alpha.$$

We will specify later which m^{th} root to take. The idea is that if m is large enough, then X/Y provides too close an approximation to γ.

We factor the left-hand side of the equation $\alpha X^m + \beta Y^m = 1$ to obtain

$$\prod_{\zeta\in\boldsymbol{\mu}_m} \left(\frac{X}{Y} - \zeta\gamma\right) = \frac{1}{\alpha Y^m}.$$

Since there are supposed to be infinitely many solutions, we may assume that $H_K(Y)$ is large, so also $|Y|_v$ is large. Then from the equality

$$\prod_{\zeta\in\boldsymbol{\mu}_m} \left|\frac{X}{Y} - \zeta\gamma\right|_v = \frac{1}{|\alpha Y^m|_v},$$

we see that X/Y must be close to one of the $\zeta\gamma$ values. Replacing γ by the appropriate $\zeta\gamma$, we may assume that $|X/Y - \gamma|_v$ is quite small. But then $|X/Y - \zeta\gamma|_v$ cannot be too small for $\zeta \neq 1$, since

$$\left|\frac{X}{Y} - \zeta\gamma\right|_v \geq \left|\gamma(1-\zeta)\right|_v - \left|\frac{X}{Y} - \gamma\right|_v.$$

Hence we can find a constant C_1, independent of X/Y, such that

$$\left|\frac{X}{Y} - \gamma\right|_v \leq \frac{C_1}{|Y|_v^m}.$$

(See Exercise 9.5.) Finally, from the expression

$$\alpha\left(\frac{X}{Y}\right)^m = \left(\frac{1}{Y}\right)^m - \beta,$$

one easily deduces that

$$H_K\left(\frac{X}{Y}\right) \leq C_2 H_K(Y),$$

where C_2 depends on only α, β, and m. Combining all of the above estimates yields

$$\left|\frac{X}{Y} - \gamma\right|_v \leq C H_K\left(\frac{X}{Y}\right)^{-m/([K:\mathbb{Q}]\#S)}.$$

But if we take $m > 2[K:\mathbb{Q}]\#S$, then Roth's theorem (IX.1.4) says that there are only finitely many possibilities for X/Y. Further, since

$$Y^m = \left(\alpha\left(\frac{X}{Y}\right)^m + \beta\right)^{-1} \qquad \text{and} \qquad X = \left(\frac{X}{Y}\right)Y,$$

each ratio X/Y corresponds to at most m possible pairs (X,Y). This contradicts our original assumption that there are infinitely many solutions, which completes the proof of (IX.4.1). □

Remark 4.2.1. There is a great similarity in the methods of proof for Siegel's theorem (IX.3.1) and the S-unit equation (IX.4.1). In both cases, we start with a point in a finitely generated group, namely $P \in E(K)$ for the former and $(x, y) \in R_S^* \times R_S^*$ for the latter. Next we pull back using the multiplication-by-m map in the group to produce a new solution whose height is much smaller than the original solution but that closely approximates another point defined over a finite extension of K. Finally, we invoke a theorem on Diophantine approximation, such as (IX.1.4), to complete the proof.

Remark 4.2.2. The proof that we have given for (IX.4.1) is ineffective because it makes use of Roth's theorem (IX.1.4). However, just as for Siegel's theorem, it is possible to make (IX.4.1) *quantitative*, i.e., to give an upper bound on the number of solutions. One might expect, a priori, that such a bound would depend on the field K and on the set of primes S, but Evertse proved the following uniform result for the S-unit equation that is an analogue of Lang's conjecture (IX.3.5) for elliptic curves. The proof, which we omit, is quite intricate.

Theorem 4.2.3. (Evertse [80]) *Let $S \subset M_K$ be a finite set of places containing M_K^∞, and let $a, b \in K^*$. Then the equation*

$$ax + by = 1$$

has at most $3 \times 7^{[K:\mathbb{Q}]+2\#S}$ solutions in S-units $x, y \in R_S^*$.

To see the analogy with (IX.3.5), note that R_S^* is a finitely generated group of rank $\#S - 1$. Thus the bound in (IX.3.5) has the form $C^{\operatorname{rank} R_S^* + \operatorname{rank} E(K)+1}$, while the bound in (IX.4.2.3) may be written as $C^{\operatorname{rank} R_S^*+1}$.

We next describe Siegel's reduction of S-integral points on hyperelliptic curves to solutions of the S-unit equation. Although we do not do so, the reader should note that every step in this reduction process can be made effective.

Theorem 4.3. (Siegel) *Let $f(x) \in K[x]$ be a polynomial of degree $d \ge 3$ with distinct roots in $\bar{K}$. Then the equation*

$$y^2 = f(x)$$

has only finitely many solutions in S-integers $x, y \in R_S$.

PROOF. We are clearly proving something stronger if we take a finite extension of K and enlarge the set S. Thus we may assume that f splits over K, say

$$f(x) = a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_d) \qquad \text{with } \alpha_1, \ldots, \alpha_d \in K.$$

Enlarging S, we may assume that the following statements are true:

(i) $a \in R_S^*$.

(ii) $\alpha_i - \alpha_j \in R_S^*$ for all $i \ne j$.

(iii) R_S is a principal ideal domain.

Now suppose that $x, y \in R_S$ satisfy $y^2 = f(x)$. Let $\mathfrak{p}$ be a prime ideal of R_S. Then $\mathfrak{p}$ divides at most one $x - \alpha_i$, since if it divides both $x - \alpha_i$ and $x - \alpha_j$, then it divides $\alpha_i - \alpha_j$, contradicting (ii). Further, we see from (i) that $\mathfrak{p}$ does not divide a. It follows from the equation

$$y^2 = a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_d)$$

that $\operatorname{ord}_{\mathfrak{p}}(x-\alpha_i)$ is even, and since this is true for all primes, the ideal $(x-\alpha_i)R_S$ is the square of an ideal in R_S. From (iii) we know that R_S is a principal ideal domain, so there are elements $z_i \in R_S$ and units $b_i \in R_S^*$ such that

$$x - \alpha_i = b_i z_i^2 \qquad \text{for} \quad i = 1, 2, \ldots, d.$$

Now let L/K be the extension of K obtained by adjoining to K the square root of every element of R_S^*. Note that L/K is a finite extension, since Dirichlet's S-unit theorem tells us that $R_S^*/(R_S^*)^2$ is finite. Let $T \subset M_L$ be the set of places of L lying over elements of S, and let R_T be the ring of T-integers in L. By construction, each b_i is a square in R_T, say $b_i = \beta_i^2$, so

$$x - \alpha_i = (\beta_i z_i)^2.$$

Taking the difference of any two of these equations yields

$$\alpha_j - \alpha_i = (\beta_i z_i - \beta_j z_j)(\beta_i z_i + \beta_j z_j).$$

Note that $\alpha_j - \alpha_i \in R_T^*$, while each of the two factors on the right is in R_T. It follows that each of these factors is a unit,

$$\beta_i z_i \pm \beta_j z_j \in R_T^* \qquad \text{for } i \neq j.$$

To complete the proof we use *Siegel's identity*:

$$\frac{\beta_1 z_1 \pm \beta_2 z_2}{\beta_1 z_1 - \beta_3 z_3} \mp \frac{\beta_2 z_2 \pm \beta_3 z_3}{\beta_1 z_1 - \beta_3 z_3} = 1.$$

This gives two elements of R_T^* that sum to 1, so (IX.4.1) says that there are only finitely many choices for

$$\frac{\beta_1 z_1 + \beta_2 z_2}{\beta_1 z_1 - \beta_3 z_3} \qquad \text{and} \qquad \frac{\beta_1 z_1 - \beta_2 z_2}{\beta_1 z_1 - \beta_3 z_3}.$$

Multiplying these two numbers, we find that there are only finitely many possibilities for

$$\frac{\alpha_2 - \alpha_1}{(\beta_1 z_1 - \beta_3 z_3)^2},$$

hence only finitely many for

$$\beta_1 z_1 - \beta_3 z_3,$$

and thus only finitely many for

$$\beta_1 z_1 = \frac{1}{2}\left((\beta_1 z_1 - \beta_3 z_3) + \frac{\alpha_3 - \alpha_1}{\beta_1 z_1 - \beta_3 z_3}\right).$$

Finally, since

$$x = \alpha_1 + (\beta_1 z_1)^2,$$

there are only finitely many possible values for x, and each x value gives at most two y values. □

Corollary 4.3.1. *Let C/K be a curve of genus one and let $f \in K(C)$ be a nonconstant function. Then there are only finitely many points $P \in C(K)$ such that $f(P) \in R_S$.*

PROOF. The reduction procedure described in (IX.3.2.2) says that it suffices to consider the case that f is the x-coordinate of a Weierstrass equation. The case $f = x$ is covered by (IX.4.3). □

IX.5 Effective Methods

In 1949, Gelfond and Schneider independently solved Hilbert's problem concerning the transcendence of $2^{\sqrt{2}}$. They actually proved the following strong transcendence criterion.

Theorem 5.1. (Gelfond, Schneider) *Let $\alpha, \beta \in \bar{\mathbb{Q}}$ with $\alpha \neq 0, 1$ and $\beta \notin \mathbb{Q}$. Then α^β is transcendental.*

Gelfond rephrased his result in terms of logarithms: If $\alpha_1, \alpha_2 \in \bar{\mathbb{Q}}^*$ and if $\log \alpha_1$ and $\log \alpha_2$ are linearly independent over $\mathbb{Q}$, then they are linearly independent over $\bar{\mathbb{Q}}$. He further showed that it is possible to give an explicit lower bound for

$$|\beta_1 \log \alpha_1 + \beta_2 \log \alpha_2|$$

whenever this quantity is nonzero, and he noted that many Diophantine problems could be solved effectively if one knew an analogous result for sums of arbitrarily many logarithms. Alan Baker proved such a theorem in 1966. The proof is quite involved, so we are content to quote the following version.

Theorem 5.2. (Baker) *Let $\alpha_1, \ldots, \alpha_n \in K^*$ and let $\beta_1, \ldots, \beta_n \in K$. For any constant κ, define*

$$\tau(\kappa) = \tau(\kappa; \alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) = h\big([1, \beta_1, \ldots, \beta_n]\big) h\big([1, \alpha_1, \ldots, \alpha_n]\big)^\kappa.$$

N.B. These are logarithmic height functions. Fix an embedding $K \subset \mathbb{C}$ and let $|\cdot|$ be the corresponding absolute value. Assume that

$$\beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n \neq 0.$$

Then there are effectively computable constants $C > 0$ and $\kappa > 0$, depending only on n and $[K : \mathbb{Q}]$, such that

$$|\beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n| > C^{-\tau(\kappa)}.$$

PROOF. See [11] or [135, VIII, Theorem 1.1]. □

Remark 5.2.1. We have restricted ourselves in (XI.5.2) to the case of the archimedean absolute value. There are analogous results in the nonarchimedean case, although minor technical difficulties arise due to the fact that the $\mathfrak{p}$-adic logarithm is defined only in a neighborhood of 1. See (IX.5.6) for a further discussion.

It is not immediately clear how Baker's theorem (IX.5.2) can be applied to give a bound for the solutions of the S-unit equation. We start with an elementary lemma; see also Exercise 9.8.

Lemma 5.3. *Let V be a finite-dimensional vector space over $\mathbb{R}$. Given any basis $\mathbf{e} = \{e_1, \dots, e_n\}$ for V, let $\| \cdot \|_{\mathbf{e}}$ be the sup norm with respect to $\mathbf{e}$, i.e.,*

$$\|x\|_{\mathbf{e}} = \Big\| \sum x_i e_i \Big\|_{\mathbf{e}} = \max\bigl\{|x_i|\bigr\}.$$

Let $\mathbf{f} = \{f_1, \dots, f_n\}$ be another basis for V. There are positive constants c_1 and c_2, depending on $\mathbf{e}$ and $\mathbf{f}$, such that for all $x \in V$,

$$c_1\|x\|_{\mathbf{e}} \le \|x\|_{\mathbf{f}} \le c_2\|x\|_{\mathbf{e}}.$$

PROOF. Let $A = (a_{ij})$ be the change of basis matrix from $\mathbf{e}$ to $\mathbf{f}$, so $e_i = \sum_j a_{ij} f_j$, and let $\|A\| = \max\bigl\{|a_{ij}|\bigr\}$. Then for any $x = \sum_i x_i e_i \in V$ we have $x = \sum_{i,j} x_i a_{ij} f_j$, so

$$\|x\|_{\mathbf{f}} = \max_j \left\{ \left| \sum_i x_i a_{ij} \right| \right\} \le n \max_{i,j}\bigl\{|a_{ij}|\bigr\} \max_i \bigl\{|x_i|\bigr\} = n\|A\|\,\|x\|_{\mathbf{e}}.$$

This gives one inequality, and the other follows by symmetry. □

We apply (IX.5.3) to the following situation. Let $S \subset M_K$ be a finite set of places containing M_K^∞, let $s = \#S$, and choose a basis $\alpha_1, \dots, \alpha_{s-1}$ for the free part of R_S^*. Then every $\alpha \in R_S^*$ can be written uniquely as

$$\alpha = \zeta \alpha_1^{m_1} \cdots \alpha_{s-1}^{m_{s-1}}$$

with integers $m_1, \dots, m_{s-1}$ and a root of unity ζ. Define the *size of α (relative to $\{\alpha_1, \dots, \alpha_{s-1}\}$)* by

$$m(\alpha) = \max\bigl\{|m_i|\bigr\}.$$

Lemma 5.4. *With notation as above, there are positive constants c_1 and c_2, depending only on K and S, such that every $\alpha \in R_S^*$ satisfies*

$$c_1 h(\alpha) \le m(\alpha) \le c_2 h(\alpha).$$

PROOF. Let $S = \{v_1, \dots, v_s\}$ and, to ease notation, let $n_i = n_{v_i}$ be the local degree corresponding to v_i. We consider the *S-regulator homomorphism*

$$\rho_S : R_S^* \longrightarrow \mathbb{R}^s, \qquad \alpha \longmapsto \bigl(n_1 v_1(\alpha), \dots, n_s v_s(\alpha)\bigr).$$

Note that the image of ρ_S lies in the hyperplane $H = \{x_1 + \cdots + x_s = 0\}$, and Dirichlet's S-unit theorem says that the image of ρ_S spans H. Let $\|\cdot\|_1$ be the sup norm on $\mathbb{R}^s$ relative to the standard basis, and let $\|\cdot\|_2$ be the sup norm relative to the basis

$$\big\{\rho_S(\alpha_1), \ldots, \rho_S(\alpha_{s-1}), (1, 1, \ldots, 1)\big\}.$$

Here $\rho_S(\alpha_1), \ldots, \rho_S(\alpha_{s-1})$ span H, and we have added one extra vector in order to span all of $\mathbb{R}^s$. From (IX.5.3) we find positive constants c_1 and c_2 such that

$$c_1\|x\|_1 \le \|x\|_2 \le c_2\|x\|_1 \qquad \text{for all } x \in \mathbb{R}^s.$$

Now let $\alpha \in R_S^*$ and write $\rho_S(\alpha) = \sum m_i \rho_S(\alpha_i)$. Then directly from the definitions we have

$$\begin{aligned}
\big\|\rho_S(\alpha)\big\|_2 &= \max\big\{|m_i|\big\} = m(\alpha),\\
\big\|\rho_S(\alpha)\big\|_1 &= \max\big\{n_i|v_i(\alpha)|\big\},\\
h_K(\alpha) &= \sum \max\big\{0, -n_i v_i(\alpha)\big\}.
\end{aligned}$$

(Note that the sum for $h_K(\alpha)$ needs to include only the absolute values in S, since by assumption $v(\alpha) = 0$ for all $v \notin S$.) It remains to compare $\big\|\rho_S(\alpha)\big\|_1$ and $h_K(\alpha)$.

In general, for any $x = (x_1, \ldots, x_s) \in H$, we can compare $\|x\|_1$ to the height $h(x) = \sum \max\{0, -x_i\}$. First, since $\max\{0, -x_i\} \le |x_i|$, we have the obvious estimate

$$h(x) \le s\|x\|_1.$$

On the other hand, if we sum the identity

$$x_i = \max\{0, x_i\} - \max\{0, -x_i\}$$

for $1 \le i \le s$ and use the fact that $x \in H$, i.e., $\sum x_i = 0$, we obtain

$$0 = h(-x) - h(x),$$

and hence $h(-x) = h(x)$. This allows us to compute

$$\begin{aligned}
2h(x) &= h(x) + h(-x)\\
&= \sum \big(\max\{0, -x_i\} + \max\{0, x_i\}\big)\\
&= \sum |x_i|\\
&\ge \max\big\{|x_i|\big\}\\
&= \|x\|_1.
\end{aligned}$$

Thus $\frac{1}{2}\|x\|_1 \le h(x) \le s\|x\|_1$, and combining this with the earlier estimates gives the desired result,

$$(c_1/s)h_K(\alpha) \le m(\alpha) \le 2c_2 h_K(\alpha). \qquad \square$$

We now have the tools needed to show how solving the S-unit equation can be reduced to the problem of giving bounds for linear forms in logarithms.

Theorem 5.5. *Fix $a, b \in K^*$. There exists an effectively computable constant $C = C(K, S, a, b)$ such that any solution $(\alpha, \beta) \in R_S^* \times R_S^*$ to the S-unit equation*

$$a\alpha + b\beta = 1$$

satisfies $H(\alpha) < C$.

PROOF. Let (α, β) be a solution and choose the absolute value v in S for which $|\alpha|_v$ is largest. Then, since $|\alpha|_w = 1$ for all $w \notin S$, we have

$$|\alpha|_v^{[K:\mathbb{Q}]s} \ge \prod_{w \in S} \max\bigl\{1, |\alpha|_w^{n_w}\bigr\} = H_K(\alpha),$$

and hence

$$|\alpha|_v \ge H(\alpha)^{1/s}.$$

(Here, as usual, $s = \#S$.)

To simplify our discussion, we will assume that v is archimedean, which is certainly true if, for example, $S = M_K^\infty$. (For arbitrary S, see the discussion in (IX.5.6).) The mean value theorem applied to the function $\log(x)$ yields

$$\left|\frac{\log x - \log y}{x - y}\right| \le \frac{1}{\min\bigl\{|x|, |y|\bigr\}}.$$

We apply this inequality with $x = a\alpha$ and $y = -b\beta$, so $x - y = 1$, and we obtain

$$\begin{aligned} |\log a\alpha - \log b\beta| &\le \min\bigl\{|a\alpha|, |a\alpha - 1|\bigr\}^{-1} \\ &\le 2\bigl(|a|H(\alpha)^{1/s}\bigr)^{-1}. \end{aligned}$$

(For the last line, we have assumed that $|\alpha| > 2/|a|$, since otherwise we have the excellent bound $H(\alpha) \le |\alpha|^s \le (2/|a|)^s$.)

Let $\alpha_1, \ldots, \alpha_{s-1}$ be a basis for R_S^*, and write

$$\alpha = \zeta \alpha_1^{m_1} \cdots \alpha_{s-1}^{m_{s-1}} \qquad \text{and} \qquad \beta = \zeta' \alpha_1^{m_1'} \cdots \alpha_{s-1}^{m_{s-1}'}.$$

Substituting this into the previous inequality yields

$$\left|\sum_{i=1}^{s-1} (m_i - m_i') \log \alpha_i + \log\left(\frac{a\zeta}{b\zeta'}\right)\right| \le \frac{c_1}{H(\alpha)^{1/s}},$$

where here and in what follows, the constants $c_1, c_2, \ldots$ are effectively computable and depend only on K, S, a, and b.

From the equality $a\alpha + b\beta = 1$, it is easy to obtain an estimate

$$\bigl|h(\alpha) - h(\beta)\bigr| \le c_2,$$

and applying (IX.5.4) yields

$$c_3 m(\alpha) \le m(\beta) \le c_4 m(\alpha).$$

(Clearly we may assume that $m(\alpha) \ge 1$ and $m(\beta) \ge 1$.) In particular,

$$|m_i - m_i'| \le m(\alpha) + m(\beta) \le c_5 h(\alpha).$$

Letting $q_i = m_i - m_i'$ and $\gamma = a\zeta/b\zeta'$ to ease notation, we have the inequality

$$|q_1 \log \alpha_1 + \cdots + q_{s-1} \log \alpha_{s-1} + \log \gamma| \le c_1 H(\alpha)^{-1/s}.$$

We now apply Baker's theorem (IX.5.2). This gives a lower bound of the form

$$|q_1 \log \alpha_1 + \cdots + q_{s-1} \log \alpha_{s-1} + \log \gamma| \ge c_6^{-\tau},$$

where

$$\tau = h\big([1, q_1, \ldots, q_{s-1}]\big) h\big([1, \alpha_1, \ldots, \alpha_{s-1}, \gamma]\big)^{\kappa}$$

and κ is a constant depending only on K and s. But from above,

$$h\big([1, q_1, \ldots, q_{s-1}]\big) = \log \max\big\{1, |q_1|, \ldots, |q_{s-1}|\big\} \le \log\big(c_5 h(\alpha)\big).$$

Combining the upper and lower bounds for the linear form in logarithms and using this estimate yields

$$c_7^{\log(c_5 h(\alpha))} \le c_1 H(\alpha)^{1/s}.$$

(Note that the basis $\alpha_1, \ldots, \alpha_{s-1}$ depends only on the field K and the set S, so we have absorbed the $h\big([1, \alpha_1, \ldots, \alpha_{s-1}, \gamma]\big)^{\kappa}$ into c_7.) Now a little bit of algebra gives

$$H(\alpha) \le c_8 h(\alpha)^{c_9},$$

and since $h(\alpha) = \log H(\alpha)$, this implies the desired bound for $H(\alpha)$. □

Remark 5.6. In order to apply the argument given in (IX.5.5) to a nonarchimedean absolute value, it is necessary to make some minor technical alterations. The main difficulty is that the logarithm function in the $\mathfrak{p}$-adic setting converges only in a neighborhood of 1. What one does is to take a subgroup of finite index in R_S^* that is generated by S-units that are $\mathfrak{p}$-adically close to 1, together with a uniformizer at $\mathfrak{p}$. Then, assuming that $|\alpha|_{\mathfrak{p}}$ is sufficiently large, one shows that $a\alpha/b\beta$ is $\mathfrak{p}$-adically close to 1. Now applying the above argument to some power of $a\alpha/b\beta$ gives a well-defined linear form in $\mathfrak{p}$-adic logarithms, and from then on the argument goes just the same. For the final step, of course, one must use a $\mathfrak{p}$-adic analogue of Baker's theorem. For further details on the reduction step, see for example [135, VI §1].

Remark 5.7. In order to obtain an effective bound for the points on an elliptic curve satisfying $f(P) \in R_S$, where f is an arbitrary nonconstant function, it is necessary to make the reduction step given in (IX.3.2.2) effective. This essentially involves giving an effective version of the Riemann–Roch theorem, which has been done by Coates [48]. As the reader might guess from the number of reduction steps involved, the effective bounds that come out of the proofs are quite large. To indicate the magnitudes involved, we quote two results; see also (IX.7.2), and (IX.7.4).

Theorem 5.8. (a) (Baker [11, page 45]) *Let $A, B, C, D \in \mathbb{Z}$ satisfy*

$$\max\bigl\{|A|, |B|, |C|, |D|\bigr\} \le H,$$

and assume that

$$E : Y^2 = AX^3 + BX^2 + CX + D$$

is an elliptic curve. Then any point $P = (x, y) \in E(\mathbb{Q})$ with $x, y \in \mathbb{Z}$ satisfies

$$\max\bigl\{|x|, |y|\bigr\} < \exp\left((10^6 H)^{10^6}\right).$$

(b) (Baker–Coates [12]) *Let $F(X, Y) \in \mathbb{Z}[X, Y]$ be an absolutely irreducible polynomial such that the curve $F(X, Y) = 0$ has genus one. Let n be the degree of F, and assume that the coefficients of F all have absolute value at most H. Then any solution to $F(x, y) = 0$ with $x, y \in \mathbb{Z}$ satisfies*

$$\max\bigl\{|x|, |y|\bigr\} < \exp\exp\exp\left((2H)^{10^{n^{10}}}\right).$$

Remark 5.8.1. There is an extensive literature on effective bounds for S-integral solutions to equations of the form $y^m = f(x)$; see for example [32, 96, 131, 268, 279, 301]. To quote one instance, we mention that [301] improves (IX.5.8a) to

$$\max\bigl\{|x|, |y|\bigr\} \le \exp\bigl(cH^{270}(\log H)^{54}\bigr)$$

for an absolute constant c.

Linear Forms in Elliptic Logarithms

Rather than reducing the problem of integral points on an elliptic curve to the question of solutions to the S-unit equation, and thence as above to bounds for linear forms in logarithms, one can instead work directly with the analytic parametrization of the elliptic curve. We briefly indicate how this is done in the simplest case.

Let $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation

$$E : y^2 = 4x^3 - g_2 x - g_3 \qquad \text{with} \quad g_2, g_3 \in \mathbb{Z}.$$

We are interested in bounding the height of points $P \in E(\mathbb{Q})$ that satisfy $x(P) \in \mathbb{Z}$. Let

$$\phi : \mathbb{C}/\Lambda \longrightarrow E(\mathbb{C})$$

be the analytic parametrization of $E(\mathbb{C})$ given by the Weierstrass $\wp$-function and its derivative (VI.5.1.1). We fix a basis $\{\omega_1, \omega_2\}$ for the lattice Λ. Let

$$\psi : E(\mathbb{C}) \longrightarrow \mathbb{C}$$

be the map that is inverse to ϕ and takes values in the fundamental parallelogram spanned by ω_1 and ω_2, shifted to be centered at 0. The map ϕ is the *elliptic exponential map*, and choosing a fundamental domain for the *elliptic logarithm map* ψ is analogous to choosing a principal value for the ordinary logarithm function $\log : \mathbb{C}^* \to \mathbb{C}$. (The analogy becomes even clearer if we identify $\mathbb{C}^*$ with $\mathbb{C}/\mathbb{Z}$.)

Fix a basis $P_1, \dots, P_r$ for the free part of $E(\mathbb{Q})$. Given any point $P \in E(\mathbb{Q})$, we can write

$$P = q_1 P_1 + \cdots + q_r P_r + T$$

with integers $q_1, \dots, q_r$ and a torsion point $T \in E_{\text{tors}}(\mathbb{Q})$. It follows that

$$\psi(P) = q_1\psi(P_1) + \cdots + q_r\psi(P_r) + \psi(T) \pmod{\Lambda},$$

so there are integers m_1 and m_2 such that

$$\psi(P) = q_1\psi(P_1) + \cdots + q_r\psi(P_r) + \psi(T) + m_1\omega_1 + m_2\omega_2.$$

Suppose now that P is a large integral point, i.e., $x(P) \in \mathbb{Z}$ and $\big|x(P)\big|$ is large. Then P is close to O in the complex topology on $E(\mathbb{C})$, so $\psi(P)$ is close to 0. More precisely, since $\wp(z) = x\big(\phi(z)\big)$ behaves like z^{-2} for z close to 0, we see that

$$\big|\psi(P)\big|^2 \le c_1\big|x(P)\big|^{-1} = c_1 H\big(x(P)\big)^{-1}.$$

We are using the fact that if $x \in \mathbb{Z}$ with $x \ne 0$, then $H(x) = |x|$. The constant c_1 depends on g_2 and g_3, but not on P.

On the other hand, since the canonical height is quadratic and positive definite from (VIII.9.3) and (VIII.9.6), we can estimate

$$\begin{aligned}\log H\big(x(P)\big) &= h_x(P) = 2\hat{h}(P) + O(1)\\ &= 2\hat{h}\Big(\sum q_i P_i + T\Big) + O(1)\\ &\ge c_2 \max\big\{|q_i|\big\}^2,\end{aligned}$$

where c_2 depends on E and the choice of the basis $P_1, \dots, P_r$. (See Exercise 9.8.) Substituting this above, we obtain an upper bound for our linear form in elliptic logarithms,

$$\big|q_1\psi(P_1) + \cdots + q_r\psi(P_r) + \psi(T) + m_1\omega_1 + m_2\omega_2\big| \le c_3^{-\max\{|q_i|\}^2}.$$

Further, since ω_1 and ω_2 are $\mathbb{R}$-linearly independent, it is easy to see that

$$\max\big\{|m_1|, |m_2|\big\} \le c_4 \max\big\{|q_i|\big\},$$

where c_4 depends on E, $\{P_i\}$, ω_1, and ω_2. Thus, if we let

$$q = \max\big\{|q_1|, \dots, |q_r|, |m_1|, |m_2|\big\},$$

then we obtain the estimate

$$\bigl|q_1\psi(P_1)+\cdots+q_r\psi(P_r)+\psi(T)+m_1\omega_1+m_2\omega_2\bigr|\le c_5^{-q^2}.$$

Now the desired finiteness result follows if we can find a lower bound for the left-hand side having the form $C^{-\tau(q)}$ with $\tau(q)/q^2\to 0$ as $q\to\infty$. The first effective estimate of this sort was proven by Masser [159] in the case that E has complex multiplication. The general case was proven by Wüstholz [313, 314], who had to overcome significant technical difficulties associated with the necessary zero and multiplicity estimates.

It remains to discuss the question of effectivity. The reduction to linear forms in ordinary logarithms via the S-unit equation is fully effective. It is possible to give an explicit upper bound for the height of any S-integral point of $E(K)$ in easily computed quantities associated to K, S, and E. One of these quantities, for example, is a bound for the heights of generators of the unit group R_S^*. In the analogous reduction to linear forms in elliptic logarithms, we similarly use a set of generators of the Mordell–Weil group $E(K)$, and the bound for the integral points depends on the heights of these generators. Unfortunately, as we have noted in (VIII.3.2) (see also Chapter X), the proof of the Mordell–Weil theorem is not effective. Thus although the approach to integral points on elliptic curves via elliptic logarithms is more natural than the roundabout route through the S-unit equation, it is likely to remain ineffective until an effective proof of the Mordell–Weil theorem is found. On the other hand, we should mention that if one is able to find a basis for the Mordell–Weil group, for example using the techniques in Chapter X, then the method of elliptic logarithms often provides the best known algorithm for finding the integral points on a given elliptic curve. See for example [58, 59, 96, 268, 279, 315].

IX.6 Shafarevich's Theorem

Recall that an elliptic curve E/K has good reduction at a finite place $v\in M_K$ if it has a Weierstrass equation whose coefficients are v-integral and whose discriminant is a v-adic unit (VII §5).

Theorem 6.1. (Shafarevich [242]) *Let $S\subset M_K$ be a finite set of places containing M_K^∞. Then up to isomorphism over K, there are only finitely many elliptic curves E/K having good reduction at all primes not in S.*

PROOF. Clearly we are proving something stronger if we enlarge S, so we may assume that S contains all primes of K lying over 2 and 3. Enlarging S further, we may also assume that the ring of S-integers R_S has class number one.

Under these assumptions, we see from (VIII.8.7) that any elliptic curve E/K has a Weierstrass equation of the form

$$E: y^2=x^3+Ax+B,\qquad A,B\in R_S,$$

with discriminant $\Delta=-16(4A^3+27B^2)$ satisfying

$$\Delta R_S=\mathcal{D}_{E/K}R_S.$$

Here $\mathcal{D}_{E/K}$ is the minimal discriminant of E/K; see (VIII §8). If we further assume that E has good reduction outside S, then $\operatorname{ord}_v(\mathcal{D}_{E/K}) = 0$ for all places $v \notin S$, so Δ is in R_S^*.

Assume now that we are given a list of elliptic curves $E_1/K, E_2/K, \ldots$, each of which has good reduction outside of S. We associate to each E_i a Weierstrass equation as above, say with coefficients $A_i, B_i \in R_S$ and discriminant $\Delta_i \in R_S^*$. Breaking the sequence of E_i into finitely many subsequences according to the residue class of Δ_i in the finite group $R_S^*/(R_S^*)^{12}$, we may replace the original sequence with an infinite subsequence satisfying $\Delta_i = CD_i^{12}$ for a fixed C and with $D_i \in R_S^*$.

The relation $\Delta = -16(4A^3 + 27B^2)$ implies that for each i, the point

$$\left(-\frac{12A_i}{D_i^4}, \frac{108B_i}{D_i^6}\right)$$

is an S-integral point on the elliptic curve

$$Y^2 = X^3 - 27C.$$

Siegel's theorem (IX.3.2.1) says that there are only finitely many such points, and thus only finitely many possibilities for A_i/D_i^4 and B_i/D_i^6. However, if

$$\frac{A_i}{D_i^4} = \frac{A_j}{D_j^4} \qquad \text{and} \qquad \frac{B_i}{D_i^6} = \frac{B_j}{D_j^6},$$

then the change of variables

$$x = (D_i/D_j)^2 x', \qquad y = (D_i/D_j)^3 y',$$

gives a K-isomorphism from E_i to E_j. Hence the sequence $E_1, E_2, \ldots$ contains only finitely many K-isomorphism classes of elliptic curves. □

Example 6.1.1. There are no elliptic curves $E/\mathbb{Q}$ having everywhere good reduction; see Exercise 8.15. There are 24 curves $E/\mathbb{Q}$ having good reduction outside of $\{2\}$ and 784 curves $E/\mathbb{Q}$ having good reduction outside of $\{2, 3\}$; for the complete list, see [19, Table 4]. Similar lists have been compiled for various quadratic fields; see for example [147] or [204].

Shafarevich's theorem (IX.6.1) has a number of important applications. We content ourselves with the following two corollaries.

Corollary 6.2. *Fix an elliptic curve E/K. Then there are only finitely many elliptic curves E'/K that are K-isogenous to E.*

PROOF. If E and E' are isogenous over K, then (VII.7.2) says that E and E' have the same set of primes of bad reduction. Now apply (IX.6.1). □

Corollary 6.3. (Serre) *Let E/K be an elliptic curve with no complex multiplication. Then for all but finitely many primes ℓ, the group of ℓ-torsion points $E[\ell]$ has no nontrivial $G_{\bar{K}/K}$-invariant subgroups.* (*In other words, the representation of $G_{\bar{K}/K}$ on $E[\ell]$ is irreducible.*)

PROOF. Suppose that $\Phi_\ell \subset E[\ell]$ is a nontrivial $G_{\bar{K}/K}$-invariant subgroup of $E[\ell]$. We know that $E[\ell] \cong (\mathbb{Z}/\ell\mathbb{Z})^2$, so Φ_ℓ is necessarily cyclic of order ℓ. We apply (III.4.12) to produce an elliptic curve E_ℓ/K and an isogeny $\phi_\ell : E \to E_\ell$ with $\ker(\phi_\ell) = \Phi$. The Galois invariance of Φ ensures that the curve E_ℓ and the isogeny ϕ_ℓ are defined over K.

Each E_ℓ is K-isogenous to E, so (IX.6.2) says that the E_ℓ fall into finitely many K-isomorphism classes. Suppose that $E_\ell \cong E_{\ell'}$ for two primes ℓ and ℓ'. Then the composition

$$E \xrightarrow{\phi_\ell} E_\ell \cong E_{\ell'} \xrightarrow{\hat{\phi}_{\ell'}} E$$

defines an endomorphism of E of degree

$$(\deg \phi_\ell)(\deg \hat{\phi}_{\ell'}) = \ell\ell'.$$

By assumption, $\operatorname{End}(E) = \mathbb{Z}$, so every endomorphism of E has degree n^2 for some $n \in \mathbb{Z}$. This shows that $\ell = \ell'$, and thus that $E_\ell \ncong E_{\ell'}$ for $\ell \neq \ell'$. Therefore there are only finitely many primes ℓ for which such a subgroup Φ_ℓ and curve E_ℓ can exist. □

Example 6.4. For $K = \mathbb{Q}$, results of Mazur [166] and Kenku [125] give a statement that is far more precise than (IX.6.2). They show that for a given elliptic curve $E/\mathbb{Q}$, there are at most eight $\mathbb{Q}$-isomorphism classes of elliptic curves $E'/\mathbb{Q}$ that are $\mathbb{Q}$-isogenous to E. Further, if $\phi : E \to E'$ is a $\mathbb{Q}$-isogeny whose kernel is a cyclic group, then either

$$1 \le \deg \phi \le 19 \qquad \text{or} \qquad \deg \phi \in \{21, 25, 27, 37, 43, 67, 163\}.$$

It is no coincidence that the possibilities for $\deg \phi$ are values of d for which $\mathbb{Q}(\sqrt{-d}\,)$ has class number one. The class number one condition means that the elliptic curve corresponding to the lattice

$$\mathbb{Z} + \mathbb{Z}\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{-d}\right)$$

via (VI.5.1.1) is isomorphic to an elliptic curve defined over $\mathbb{Q}$. (See (C.11.3.1) for details.) Now we need merely observe that multiplication by $\sqrt{-d}$ gives an isogeny from E to itself that is defined over $\mathbb{Q}$ and whose kernel Φ is invariant under the action of $G_{\bar{\mathbb{Q}}/\mathbb{Q}}$. Then $E \to E/\Phi$ is a cyclic isogeny of degree d between elliptic curves defined over $\mathbb{Q}$.

Remark 6.5. An examination of the proof of (IX.6.1) reveals an interesting possibility. If we had some other proof of (IX.6.1) that did not use either Siegel's theorem or Diophantine approximation techniques, then we could deduce that the equation

$$Y^2 = X^3 + D$$

has only finitely many solutions $X, Y \in R_S$. For given such a solution, the equation

$$y^2 = x^3 - Xx - Y$$

defines an elliptic curve with good reduction outside of the set

$$S \cup \{\text{primes dividing 2 and 3}\}.$$

Hence, assuming (IX.6.1), there can be only finitely many such curves, and we could argue back to the finiteness of the number of pairs (X, Y). Building on this idea, Parshin [203] showed how a generalization of (IX.6.1) to curves of higher genus (which had already been conjectured by Shafarevich [242]) could be used to prove Mordell's conjecture that curves of genus at least 2 have only finitely many *rational* points. The subsequent proof of Shafarevich's conjecture by Faltings [82, 84] completed this chain of reasoning. Faltings' proof, together with Parshin's idea, also gives a proof of Siegel's theorem (IX.3.2) that does not involve the use of Diophantine approximation. Subsequent to Faltings' proof of the Mordell conjecture, Vojta [299] gave a somewhat more natural proof based on Diophantine approximation methods. For an exposition of this latter proof, see for example [114, Part E].

IX.7 The Curve $Y^2 = X^3 + D$

Many of the general results known and conjectured about the arithmetic of elliptic curves were originally noticed and tested on various special sorts of equations, such as the one given in the title of this section. For example, long before the work of Mordell and Siegel led to general finiteness results such as (IX.3.2.1), many special cases had been proven by a variety of methods. (See, e.g., [185, Chapter 26].) The next result gives two examples in which the complete set of integral solutions can be obtained by relatively elementary means.

Proposition 7.1. (a) (V.A. Lebesgue) *The equation*

$$y^2 = x^3 + 7$$

has no solutions in integers $x, y \in \mathbb{Z}$.

(b) (Fermat) *The only integral solutions to the equation*

$$y^2 = x^3 - 2$$

are $(x, y) = (3, \pm 5)$.

PROOF. (a) Suppose that $x, y \in \mathbb{Z}$ satisfy $y^2 = x^3 + 7$. We first observe that x must be odd, since no integer of the form $8k + 7$ is a square. Next we rewrite the equation as

$$y^2 + 1 = x^3 + 8 = (x + 2)(x^2 - 2x + 4).$$

Since x is odd,

$$x^2 - 2x + 4 = (x - 1)^2 + 3 \equiv 3 \pmod 4,$$

so there exists at least one prime $p \equiv 3 \pmod 4$ that divides $x^2 - 2x + 4$. But then $y^2 + 1 \equiv 0 \pmod p$, which is not possible.

(b) Suppose that we have a solution $x, y \in \mathbb{Z}$ to $y^2 = x^3 - 2$. We factor the equation as

$$(y + \sqrt{-2}\,)(y - \sqrt{-2}\,) = x^3.$$

The ring $R = \mathbb{Z}[\sqrt{-2}\,]$ is a principal ideal domain, and the greatest common divisor of $y + \sqrt{-2}$ and $y - \sqrt{-2}$ in R divides $2\sqrt{-2}$, so we see that $y + \sqrt{-2}$ has one of the following forms:

$$y + \sqrt{-2} = \zeta^3 \quad \text{or} \quad \sqrt{-2}\zeta^3 \quad \text{or} \quad 2\zeta^3 \quad \text{for some } \zeta \in R.$$

Applying complex conjugation gives

$$y - \sqrt{-2} = \bar{\zeta}^3 \quad \text{or} \quad -\sqrt{-2}\bar{\zeta}^3 \quad \text{or} \quad 2\bar{\zeta}^3,$$

and taking the product yields

$$x^3 = y^2 + 2 = (\zeta\bar{\zeta})^3 \quad \text{or} \quad 2(\zeta\bar{\zeta})^3 \quad \text{or} \quad 4(\zeta\bar{\zeta}^3).$$

Since $x \in \mathbb{Z}$ and $\zeta\bar{\zeta} \in \mathbb{Z}$, only the first case is possible, so

$$y + \sqrt{-2} = \zeta^3 \qquad \text{and} \qquad y - \sqrt{-2} = \bar{\zeta}^3.$$

Subtracting these two equations gives

$$2\sqrt{-2} = \zeta^3 - \bar{\zeta}^3 = (\zeta - \bar{\zeta})(\zeta^2 + \zeta\bar{\zeta} + \bar{\zeta}^2).$$

We write $\zeta = a + b\sqrt{-2}$ with $a, b \in \mathbb{Z}$ and substitute to obtain

$$2\sqrt{-2} = 2\sqrt{-2}\,b(3a^2 - 2b^2).$$

Since a and b are in $\mathbb{Z}$, we must have

$$b = \pm 1 \qquad \text{and} \qquad 3a^2 - 2b^2 = \pm 1,$$

where the signs are the same. It follows that $(a, b) = (\pm 1, 1)$, and working back through the various substitutions yields the values $(x, y) = (3, \pm 5)$. □

Remark 7.1.1. It is worth remarking that the result in (IX.7.1b) is far more interesting than that in (IX.7.1a). The reason is that the Mordell–Weil group over $\mathbb{Q}$ of the elliptic curve $y^2 = x^3 + 7$ is trivial, so (IX.7.1a) reflects the fact that the equation has no rational solutions. On the other hand, the Mordell–Weil group of $y^2 = x^3 - 2$ is infinite cyclic (see Exercise 10.19), so (IX.7.1b) says that among the infinitely many rational points, only two have integer coordinates.

Baker applied his effective estimate for linear forms in logarithms to give an explicit upper bound, in terms of D, for the integral solutions to $y^2 = x^3 + D$. This bound was refined by Stark, who proved the following result.

Theorem 7.2. (Stark [273]) *For every $\epsilon > 0$ there is an effectively computable constant C_ϵ, depending only on ϵ, such that if $D \in \mathbb{Z}$ with $D \neq 0$ and if $x, y \in \mathbb{Z}$ are solutions to the equation*

$$y^2 = x^3 + D,$$

then

$$\log \max\bigl\{|x|, |y|\bigr\} \le C_\epsilon |D|^{1+\epsilon}.$$

Example 7.3. Stark's estimate (IX.7.2) gives a bound for x and y that is slightly worse than exponential in D. It is natural to ask whether this bound is of the correct order of magnitude. Various people have conducted computer searches for large solutions, see for example [75, 106, 134]. Among the interesting examples found, we mention:

$$\begin{aligned}
378{,}661^2 &= 5234^3 + 17,\\
911{,}054{,}064^2 &= 939{,}787^3 - 307,\\
149{,}651{,}610{,}621^2 &= 28{,}187{,}351^3 + 1090,\\
447{,}884{,}928{,}428{,}402{,}042{,}307{,}918^2 &= 5{,}853{,}886{,}516{,}781{,}223^3 - 1641843.
\end{aligned}$$

Although these examples show that x and y may be quite large in comparison to D, a close examination of the data led M. Hall to make the following conjecture, which was partly generalized by Lang.

Conjecture 7.4. (a) (Hall [106]) *For every $\epsilon > 0$ there is a constant C_ϵ, depending only on ϵ, such that for all $D \in \mathbb{Z}$ with $D \neq 0$ and for all $x, y \in \mathbb{Z}$ satisfying*

$$y^2 = x^3 + D,$$

we have

$$|x| \le C_\epsilon D^{2+\epsilon}.$$

(b) (Hall–Lang [138]) *There are absolute constants C and κ such that for every elliptic curve $E/\mathbb{Q}$ given by a Weierstrass equation*

$$y^2 = x^3 + Ax + B \qquad \text{with } A, B \in \mathbb{Z}$$

and for every integral point $P \in E(\mathbb{Q})$, i.e., satisfying $x(P) \in \mathbb{Z}$, we have

$$\bigl|x(P)\bigr| \le C \max\bigl\{|A|, |B|\bigr\}^{\kappa}.$$

The evidence for these conjectures is fragmentary. They are true for function fields, for which Davenport [57] proved (IX.7.4a) and Schmidt proved (IX.7.4b). Vojta [298, 4 §4] has shown that (IX.7.4a) over number fields is a consequence of his very general Nevanlinna-type conjectures for algebraic varieties. It is also easy to deduce (IX.7.4a) from the ABC conjecture; see Exercise 9.17. However, both Vojta's conjectures and the ABC conjecture are well beyond the reach of current techniques. (See also Exercise 9.10 for a proof that the exponent in (IX.7.4a) cannot

be improved.) Aside from these few facts, very little is known. It is worth pointing out that the effective techniques from (IX §5) seem intrinsically incapable of leading to estimates as strong as those described in (IX.7.4). We briefly explain the problem for the equation $y^2 = x^3 + D$.

When performing the reduction to the S-unit equation, we use a number field K whose discriminant looks like a power of D. The Brauer–Siegel theorem says that $\log(h_K \mathcal{R}_K) \sim \frac{1}{2} \log d_K$ as $[K : \mathbb{Q}]/\log d_K \to 0$, where h_K is the class number, $\mathcal{R}_K$ the regulator, and d_K the absolute discriminant of K. (See, e.g., [142, Chapter XVI].) In general there is no reason to expect the class number of K to be large, so the best that we can hope for is to find a bound for the regulator that is a power of $|D|$. Since the regulator is the determinant of the *logarithms* of a basis for the unit group R^*, the resulting bounds for the heights $H(\alpha_i)$ of generators $\alpha_i \in R^*$ will be exponential in $|D|$. This eventually leads to an exponential bound for x and y as in (IX.7.2).

There is a similar problem if we try to prove (IX.7.4) using linear forms in elliptic logarithms or by following Siegel's method of proof as in (IX.3.1), even assuming that we could prove strong effective versions of Roth's theorem and the Mordell–Weil theorem. The difficulty is that it is likely that the best possible upper bound for generators of the Mordell–Weil group of $y^2 = x^3 + D$ has the form $\hat{h}(P) \le C|D|^\kappa$, cf. (VIII.10.2). Here $\hat{h}$ is a logarithmic height, so this again leads to a bound for the x-coordinate of integral points that is exponential in D.

The problem in both cases can be explained most clearly by the analogy given in (IX.4.2.1). When solving the S-unit equation or when finding integral points on elliptic curves, one is initially given a finitely generated group ($R_S^* \times R_S^*$, respectively $E(K)$) and a certain exceptional subset (solutions to $ax + by = 1$, respectively points with $x(P) \in R_S$). The first step is to choose a basis for the finitely generated group and express the exceptional points in terms of the basis. The difficulty that arises in trying to prove (IX.7.4) or the analogous estimate for the S-unit equation is that in general, the best (conjectural) upper bound for the heights of the basis elements is exponentially larger than the desired upper bound for the exceptional points! The moral of this story, assuming the validity of various conjectures, is that a randomly chosen elliptic curve $E/\mathbb{Q}$ is unlikely to have any integral points at all.

IX.8 Roth's Theorem—An Overview

In this section we give a brief sketch of the principal steps that go into the proof of Roth's theorem (IX.1.4). None of the steps are tremendously deep, but the details required to make them rigorous are quite lengthy. For the full proof, see for example [114, Part D], [139, Chapter 7], or [221].

We assume that we are given an $\alpha \in \bar{K}$, an absolute value $v \in M_K$, and positive real numbers ϵ and C. We then want to prove that there are only finitely many $x \in K$ satisfying

$$|x-\alpha|_v \le CH_K(x)^{-2-\epsilon}.$$

Step I: An Auxiliary Polynomial

For any given integers $m, d_1, \dots, d_m$, one uses elementary estimates and the pigeonhole principle to construct a polynomial

$$P(X_1, \dots, X_m) \in R[X_1, \dots, X_m]$$

of degree d_i in X_i such that P vanishes to fairly high order (in terms of m and the d_i) at the point $(\alpha, \dots, \alpha)$. Further, one shows that P may be chosen with coefficients having fairly small heights, the bound for the heights being given explicitly in terms of α, m, and the d_i.

Step II: An Upper Bound for P

Suppose now that we are given elements $x_1, \dots, x_m \in K$ satisfying

$$|x_i-\alpha|_v \le CH_K(x_i)^{-2-\epsilon} \qquad \text{for } 1 \le i \le m.$$

Using the Taylor series expansion for $P(X_1, \dots, X_m)$ around $(\alpha, \dots, \alpha)$ and the fact that P vanishes to high order at $(\alpha, \dots, \alpha)$, one shows that $\big|P(x_1, \dots, x_m)\big|_v$ is fairly small.

Step III: A Nonvanishing Result (Roth's Lemma)

Suppose that the degrees $d_1, \dots, d_m$ are fairly rapidly decreasing, where the rate of decrease depends on m, and suppose that $x_1, \dots, x_m \in K$ have the property that their heights are fairly rapidly increasing, the rate of increase depending on m and $d_1, \dots, d_m$. Suppose further that $P(X_1, \dots, X_m) \in R[X_1, \dots, X_m]$ has degree d_i in X_i and coefficients whose heights are bounded in terms of d_1 and $h(x_1)$. Then one shows that P does not vanish to too high an order at $(x_1, \dots, x_m)$.

This is the hardest step in Roth's theorem. In Thue's original theorem, he used a polynomial of the form $P(X,Y) = f(X) + g(X)Y$ and obtained an approximation exponent $\tau(d) = \frac{1}{2}d + \epsilon$. The improvements of Siegel, Gelfond, and Dyson used a general polynomial in two variables. It was clear at that time that the way to obtain $\tau(d) = 2 + \epsilon$ was to use polynomials in more variables; the only stumbling block was the lack of a nonvanishing result such as the one that we have just described.

The proof of Roth's lemma is by induction on m, the number of variables in the polynomial P. If P factors as

$$P(X_1, \dots, X_m) = F(X_1)G(X_2, \dots, X_m),$$

then the induction proceeds fairly smoothly. Of course, such a factorization is unlikely to happen. What one does is to construct differential operators $\mathcal{D}_{ij}$ such that

the generalized Wronskian determinant $\det(\mathcal{D}_{ij}P)$ is a nonzero polynomial that does factor in the above fashion. It is then a delicate matter to estimate the degrees and heights of the coefficients of the resulting polynomial and to show that they have not grown too large to allow the inductive hypothesis to be applied.

Step IV: The Final Estimate

Suppose that the inequality

$$|x-\alpha|_v \le CH_K(x)^{-2-\epsilon}$$

has infinitely many solutions $x \in K$. We derive a contradiction as follows.

First choose a value for m, depending on ϵ, C, and $\big[K(\alpha):K\big]$. Second, choose $x_1,\dots,x_m \in K$ in succession satisfying

$$|x_i-\alpha|_v \le CH_K(x_i)^{-2-\epsilon},$$

such that $H_K(x_1)$ is large, depending on m, and such that $H_K(x_{i+1}) > H_K(x_i)^\kappa$ for some constant κ depending on m. Third, choose a large integer d_1, depending on m and the $H_K(x_i)$, and then choose $d_2,\dots,d_m$ in terms of d_1 and the $H_K(x_i)$. We are now ready to apply the initial three steps.

Using Step I, choose a polynomial $P(X_1,\dots,X_m)$ of degree d_i in X_i such that P vanishes to high order at $(\alpha,\dots,\alpha)$. The order of vanishing depends on m and $d_1,\dots,d_m$. From Step III, we know that P does not vanish to too high an order at $(x_1,\dots,x_m)$, so we can choose a low-order partial derivative that does not vanish,

$$z = \frac{\partial^{i_1+\cdots+i_m}}{\partial X_1^{i_1}\cdots\partial X_m^{i_m}}P(x_1,\dots,x_m) \ne 0.$$

From Step II, we know that $|z|_v$ is fairly small. On the other hand, since $z \ne 0$, we can use the product formula to show that $|z|_v$ cannot be too small. Specifically, we have $|z|_v \ge H_K(z)^{-1}$; see Exercise 9.9. Next, using elementary triangle inequality estimates, we find a lower bound for $H_K(z)^{-1}$. Combining this lower bound with the earlier upper bound, some algebra gives a contradiction. It follows that the inequality

$$|x-\alpha|_v \le CH_K(x)^{-2-\epsilon}$$

has only finitely many solutions.

Remark 8.1. In examining the proof sketch of Roth's theorem, especially the sequence of choices in Step IV, it is clear why we do not obtain an effective procedure for finding all $x \in K$ satisfying $|x-\alpha|_v \le CH_K(x)^{-2-\epsilon}$. What the proof shows is that we cannot find a long sequence of x_i whose heights grow sufficiently rapidly, where the terms "long sequence" and "sufficiently rapidly" can be made completely explicit in terms of K, α, ϵ, and C. The difficulty is that the required growth of the height of each x_i is given in terms of its predecessor. What this boils down to is that if we can find a large number of good approximations to α whose heights are

sufficiently large, then we can obtain a bound for all other good approximations to α in terms of the approximations that we already know. Unfortunately, the bounds that come out of Roth's theorem are so large that it is highly unlikely that there exists even a single good approximation to α having the requisite height.

Using an elaboration of the above argument, one can prove quantitative versions of Roth's theorem such as in the following result.

Theorem 8.2. ([103, 173]) *Let $K/\mathbb{Q}$ be a number field, let $\alpha \in \bar{K} \smallsetminus K$, and let $S \subset M_K$ be a finite set of absolute values, each of which is extended in some way to $K(\alpha)$. Let $\epsilon > 0$. There are constants C_1 and C_2, depending only on ϵ and $\bigl[K(\alpha) : K\bigr]$, such that the inequality*

$$\prod_{v\in S} \min\bigl\{|x-\alpha|_v^{n_v}, 1\bigr\} \le C H_K(x)^{-2-\epsilon}$$

has at most $4^{\#S}C_1$ solutions $x \in K$ satisfying $H_K(x) > \bigl(2H_K(\alpha)\bigr)^{C_2}$.

Of course, the constant C_2 in (IX.8.2) turns out to be sufficiently large that it is highly unlikely that there are any $x \in K$ satisfying the two conditions of the theorem. But the proof of Roth's theorem does not preclude the existence of large solutions, and it provides no tools with which to find them if they do exist!

Exercises

9.1. Let $\bigl(\phi(n)\bigr)_{n=1,2,\ldots}$ be a sequence of positive numbers. We say that a number $\alpha \in \mathbb{R}$ is *ϕ-approximable* (*over* $\mathbb{Q}$) if there are infinitely many $p/q \in \mathbb{Q}$ satisfying

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q\phi(q)}.$$

For example, Roth's theorem says that no element of $\bar{\mathbb{Q}}$ is $n^{1+\epsilon}$-approximable.

(a) Prove that for any $\epsilon > 0$, the set

$$\{\alpha \in \mathbb{R} : \alpha \text{ is } n^{1+\epsilon}\text{-approximable}\}$$

is a set of measure 0.

(b) More generally, prove that if the series $\sum_{n\ge1} 1/\phi(n)$ converges, then the set

$$\{\alpha \in \mathbb{R} : \alpha \text{ is } \phi\text{-approximable}\}$$

is a set of measure 0.

9.2. (a) Use Liouville's theorem (IX.1.3) to prove that the number $\sum_{n\ge1} 2^{-n!}$ is transcendental.

(b) More generally, let $\bigl(e(n)\bigr)_{n=1,2,\ldots}$ be a sequence of real numbers with the property that for every $d > 0$ there is a constant $C_d > 0$ such that

$$e(n) \ge C_d n^d \qquad \text{for all } n = 1, 2, \ldots.$$

(In complexity theory terminology, one says that the growth rate of the function $e(n)$ is faster than polynomial.) Let $b \ge 2$ be an integer. Prove that the number $\sum_{n\ge1} b^{-e(n)}$ is transcendental.

(c) Use (b) to prove that there are uncountably many transcendental numbers.

9.3. For each integer $m \neq 0$, let

$$N(m) = \#\big\{(x,y) \in \mathbb{Z} : y^2 = x^3 + m\big\}.$$

Note that (IX.3.2) tells us that $N(m)$ is finite.

(a) Prove that $N(m)$ can be arbitrarily large. (*Hint*. Choose an m_0 such that $y^2 = x^3 + m_0$ has infinitely many rational solutions. Then clear the denominators of a lot of them.)

(b) More precisely, prove that there is an absolute constant $c > 0$ such that

$$N(m) > c\big(\log|m|\big)^{1/3}$$

for infinitely many $m \in \mathbb{Z}$. (*Hint*. Use height functions to estimate the size of the denominators cleared in (a).)

(c) ** Prove or disprove that $N(m)$ is unbounded as m ranges over sixth-power-free integers, i.e., integers divisible by no nontrivial sixth power.

(d) Suppose that there is a value of m_0 such that the Mordell–Weil group $E_0(\mathbb{Q})$ of the elliptic curve $E_0 : y^2 = x^3 + m_0$ has rank r. Using an elaboration of the argument in (b), prove that there is an absolute constant $c > 0$ such that

$$N(m) > c\big(\log|m|\big)^{r/(r+2)}$$

for infinitely many $m \in \mathbb{Z}$.

(e) ** Let $\epsilon > 0$. Prove or disprove that

$$\lim_{|m|\to\infty} \frac{N(m)}{\big(\log|m|\big)^{1+\epsilon}} = 0.$$

9.4. Let $E/\mathbb{Q}$ be an elliptic curve and let $P \in E(\mathbb{Q})$ be a point of infinite order.

(a) For each prime $p \in \mathbb{Z}$ at which E has good reduction, let n_p be the order of the reduced point $\tilde{P}$ in the finite group $\tilde{E}(\mathbb{F}_p)$. Prove that the set

$$\{n_p : p \text{ prime}\}$$

contains all but finitely many positive integers. (*Hint*. You will need the strong form of Siegel's theorem; see (IX.3.3).)

(b) An alternative formulation for (a) is to write $x(nP) = a_n/d_n^2$ as a fraction in lowest terms. The sequence $(d_n)_{n\ge1}$ is an *elliptic divisibility sequence*.[2] A prime p is called a *primitive divisor of* d_n if $p \mid d_n$ and $p \nmid d_m$ for all $m < n$. Prove that all but finitely many terms in the sequence d_n have a primitive divisor. (This is an analogue for elliptic curves of a classical result for the multiplicative group that is due to Bang and Zsigmondy [317].)

9.5. (a) Let $f(T) = a_0T^n + \cdots + a_n \in \mathbb{Z}[T]$ be a polynomial with $a_0a_n \neq 0$ and with distinct roots $\xi_1, \ldots, \xi_n \in \mathbb{C}$. Let $A = \max\big\{|a_0|, \ldots, |a_n|\big\}$. Prove that for every rational number $t \in \mathbb{Q}$,

$$\big|f(t)\big| \ge (2n^2A)^{-n} \min\big\{|t-\xi_1|, \ldots, |t-\xi_n|\big\}.$$

[2] This definition differs from that given in exercises 3.34–3.36. In general, it may be necessary to take a subsequence $(d_{nk})_{n\ge1}$ in order to obtain a sequence satisfying the recurrence described in Exercise 3.34.

(b) Let $f(T) = a_0T^n + \cdots + a_n \in K[T]$ be a polynomial with $a_0a_n \neq 0$ and with distinct roots $\xi_1, \ldots, \xi_n \in \bar{K}$. Let $S \subset M_K$ be a finite set of places of K, each extended in some fashion to $\bar{K}$. Prove that there is a constant $C_f > 0$, depending only on f, such that for every $t \in K$,

$$\prod_{v\in S} \min\bigl\{1, \bigl|f(t)\bigr|_v^{n_v}\bigr\} \ge C_f \prod_{v\in S} \min_{1\le i\le n} \bigl\{1, |t-\xi_i|_v^{n_v}\bigr\}.$$

(c) Find an explicit expression for the constant C_f appearing in (b), where your value for C_f should depend only on n and $H_K\bigl([a_0, \ldots, a_n]\bigr)$.

9.6. (a) Let $F(X, Y) \in \mathbb{Z}[X, Y]$ be a homogeneous polynomial of degree $d \ge 3$ with nonzero discriminant. Prove that for every nonzero integer b, *Thue's equation*

$$F(X, Y) = b$$

has only finitely many solutions $(x, y) \in \mathbb{Z}^2$. (*Hint.* Let $f(T) = F(T, 1)$, and write $b = F(x, y) = y^d f(x/y)$. Now use Exercise 9.5a and (IX.1.4).)

(b) More generally, let $F(X, Y) \in K[X, Y]$ be a homogeneous polynomial of degree $d \ge 3$ with nonzero discriminant, and let $S \subset M_K$ be a finite set of places containing M_K^∞. Prove that for every $b \in K^*$, the equation

$$F(X, Y) = b$$

has only finitely many solutions $(x, y) \in R_S \times R_S$.

(c) Let $f(X) \in K[X]$ be a polynomial with at least two distinct roots in $\bar{K}$, let $S \subset M_K$ be as in (b), and let $n \ge 3$ be an integer. Prove that the equation

$$Y^n = f(X)$$

has only finitely many solutions $(x, y) \in R_S \times R_S$. (*Hint.* Mimic the proof of (IX.4.3) until you end up with a number of equations of the form $aW^n + bZ^n = c$, and then use (b).)

9.7. Let E/K be an elliptic curve without complex multiplication. Prove that for every prime ℓ, the representation of $G_{\bar{K}/K}$ on the $\mathbb{Q}_\ell$-vector space $T_\ell(E) \otimes \mathbb{Q}_\ell$ is irreducible.

9.8. (a) Let $\|\cdot\|$ be the usual Euclidean norm on $\mathbb{R}^n$, and let $\{v_1, \ldots, v_n\}$ be a basis for $\mathbb{R}^n$. Prove that there is a constant $c > 0$, depending only on n and $\{v_1, \ldots, v_n\}$, such that

$$\left\|\sum_{i=1}^n a_iv_i\right\| \ge c\max\bigl\{|a_i|\bigr\} \qquad \text{for all } a_1, \ldots, a_n \in \mathbb{R}.$$

(b) Let $\Lambda \subset \mathbb{R}^n$ be a lattice. Prove that there exist a basis $\{v_1, \ldots, v_n\}$ for Λ and a constant $c_n > 0$ *depending only on* n such that

$$\left\|\sum_{i=1}^n a_iv_i\right\| \ge c_n \sum_{i=1}^n \|a_iv_i\| \qquad \text{for all } a_1, \ldots, a_n \in \mathbb{R}.$$

(*Hint.* Ideally, one would like to choose an orthogonal basis for Λ. This is not generally possible, but mimic the Gram–Schmidt process to find a basis that is reasonably orthogonal.)

(c) Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on $\mathbb{R}^n$, i.e., they satisfy $\|v\| \ge 0$, $\|v\| = 0$ if and only if $v = 0$, $\|av\| \le |a|\,\|v\|$, and $\|v+w\| \le \|v\| + \|w\|$. Prove that there are constants $c_1, c_2 > 0$ such that

$$c_1\|v\|_1 \le \|v\|_2 \le c_2\|v\|_1 \qquad \text{for all } v \in \mathbb{R}^n.$$

(d) Let Q be a positive definite quadratic form on $\mathbb{R}^n$. Prove that there is a constant $c > 0$, depending on n and Q, such that for any integral lattice point $(a_1, \dots, a_n) \in \mathbb{Z}^n \subset \mathbb{R}^n$,

$$Q(a_1, \dots, a_n) \ge c \max\bigl\{|a_1|, \dots, |a_n|\bigr\}^2.$$

(e) Let E/K be an elliptic curve and let $P_1, \dots, P_r$ be a basis for the free part of $E(K)$. Prove that there is a constant $c > 0$, depending on E and $P_1, \dots, P_r$, such that for all integers $m_1, \dots, m_r$,

$$\hat{h}(m_1P_1 + \cdots + m_rP_r) \ge c \max\bigl\{|m_1|, \dots, |m_r|\bigr\}^2.$$

9.9. Let $z \in K$ with $z \ne 0$.

(a) Prove that for any $v \in M_K$,

$$|z|_v \ge H_K(z)^{-1}.$$

(b) More generally, prove that for any (not necessarily finite) set of absolute values $S \subset M_K$,

$$\prod_{v \in S} \min\bigl\{1, |z|_v^{n_v}\bigr\} \ge H_K(z)^{-1}.$$

(This lemma, as trivial as it appears, lies at the heart of all known proofs in Diophantine approximation and transcendence theory. In its simplest guise, namely for $K = \mathbb{Q}$, it asserts nothing more than the fact that there are no positive integers smaller than one!)

9.10. Prove that there is an (absolute) constant $C > 0$ such that the inequality

$$0 < |y^2 - x^3| < C\sqrt{|x|}$$

has infinitely many solutions $(x, y) \in \mathbb{Z}$. (*Hint*. Verify the identity

$$(t^2-5)^2\bigl((t+9)^2+4\bigr) - (t^2+6t-11)^3 = -1728(t-2).$$

Take solutions to $y^2 - 2v^2 = -1$ and set $t = 2u - 9$. Show that this leads to a value $C = 432\sqrt{2} + \epsilon$ for any $\epsilon > 0$.)

9.11. (a) Let $d \equiv 2 \pmod 4$ and let $D = d^3 - 1$. Prove that the equation

$$y^2 = x^3 + D$$

has no solution in integers $x, y \in \mathbb{Z}$.

(b) For each of the primes p in the set $\{11, 19, 43, 67, 163\}$, find all solutions $x, y \in \mathbb{Z}$ to the equation

$$y^2 = x^3 - p.$$

(*Hint*. Work in the ring $R = \mathbb{Z}\left[\frac{1}{2}(1 + \sqrt{-p}\,)\right]$. Note that R is a principal ideal domain and that 2 does not split in R.)

9.12. Let $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, \ldots, a_6 \in \mathbb{Z}$, and let $P \in E(\mathbb{Q})$ be a point of infinite order.

(a) Suppose that $x\big([m]P\big) \in \mathbb{Z}$ for some integer $m \ge 1$. Prove that $x(P) \in \mathbb{Z}$. (This result is often useful when one is searching for integral points on elliptic curves of rank 1. See Exercise 9.13 for an example.)

(b) More generally, for any $m \ge 1$, write $x(mP) = a_m/d_m^2 \in \mathbb{Q}$ as a fraction in lowest terms. Prove that

$$m \mid n \Longrightarrow d_m \mid d_n.$$

Thus the sequence $(d_m)_{m\ge1}$ is a *divisibility sequence*; see Exercise 3.36.

9.13. Let $E/\mathbb{Q}$ be the elliptic curve

$$E : y^2 + y = x^3 - x.$$

For this exercise you may assume that $E(\mathbb{Q})$ has rank 1. (For a proof that $\operatorname{rank} E(\mathbb{Q}) = 1$, see Exercise 10.9.)

(a) Prove that $E_{\text{tors}}(\mathbb{Q}) = \{O\}$, and hence that $E(\mathbb{Q}) \cong \mathbb{Z}$.

(b) Prove that $(0,0)$ is a generator for $E(\mathbb{Q})$. (*Hint.* Make a sketch of $E(\mathbb{R})$ and show that $(0,0)$ is not on the identity component. Use Exercise 9.12 to conclude that a generator for $E(\mathbb{Q})$ must be a point with *integer* coordinates on the nonidentity component, and find all such points.)

(c) Find all of the integer points in $E(\mathbb{Q})$. (*Hint.* Let $P = (0,0)$. Suppose that $[m]P$ is integral. Write $m = 2^a n$ with n odd and use Exercise 9.12 to show that$[n]P$ is integral. Use an argument as in (b) to find all possible values of n, and then do some computations to find the possible a values.)

(d) Solve the following classical number theory problem: Find all positive integers that are simultaneously the product of two consecutive integers and the product of three consecutive integers.

9.14. Let C/K be a curve and let $f, g \in K(C)$ be nonconstant functions.

(a) ***** Prove that

$$\lim_{\substack{P\in C(\bar K)\\ h_f(P)\to\infty}} \frac{h_f(P)}{h_g(P)} = \frac{\deg f}{\deg g}.$$

(b) Prove that for every $\epsilon > 0$ there exists a constant $c = c(f, g, \epsilon)$ such that

$$\left|\frac{1}{\deg f}h_f(P) - \frac{1}{\deg g}h_g(P)\right| < \epsilon h_f(P) + c \qquad \text{for all } P \in C(\bar K).$$

(c) Let C be an elliptic curve. Prove that there is a constant $c = c(f, m, \epsilon)$ such that

$$\big|h_f\big([m]P\big) - m^2h_f(P)\big| < \epsilon h_f(P) + c \qquad \text{for all } P \in C(\bar K).$$

(d) Prove that (IX.3.1) is true for all nonconstant functions $f \in K(E)$. Use this to prove the finiteness result (IX.3.2.2) directly, without first reducing to (IX.3.2.1).

9.15. For a given $Q \in C(K_v)$, let d_v be the distance function defined in (IX §2), and let d_v^{alt} denote the distance function given by the alternative definition in (IX.2.2.1). Prove that the ratio $d_v^{\text{alt}}(P, Q)/d_v(P, Q)$ is bounded for $P \in C(K_v)$.

9.16. Let C/K be a curve, let $f \in K(C)$ be a nonconstant function, and write the divisor of zeros of f as

$$\operatorname{div}_0(f) = \sum_{\substack{Q \in C(\bar{K}) \\ \operatorname{ord}_Q(f)>0}} \operatorname{ord}_Q(f)(Q) = n_1(Q_1) + n_2(Q_2) + \cdots + n_r(Q_r).$$

Replacing K by an extension field, we assume that $Q_1, \ldots, Q_r \in C(K)$. Let $v \in M_K$. Prove that

$$\log \min\bigl\{\bigl|f(P)\bigr|_v, 1\bigr\} = n_1 \log d_v(P, Q_1) + \cdots + n_r \log d_v(P, Q_r) + O(1) \quad \text{for all } P \in C(K_v),$$

where the $O(1)$ depends on f and the choice of distance functions, but is independent of P.

9.17. Let $\epsilon > 0$, and let m and n be positive integers satisfying $nm > n + m$. Assuming that the ABC conjecture (VIII.11.4) is true, prove the following assertions (see also Exercise 8.22):

(a) There is a constant $C = C(\epsilon, m, n)$ such that if

$$y^m = x^n + D \qquad \text{with } x, y, D \in \mathbb{Z} \text{ and } D \neq 0,$$

then

$$|x|^{nm-n-m} \le C|D|^{m+\epsilon} \qquad \text{and} \qquad |y|^{nm-n-m} \le C|D|^{n+\epsilon}.$$

(This is a generalized version of Hall's conjecture (IX.7.4).)

(b) Suppose now that $D \neq 0$ is fixed. If $\max\{m, n\}$ is sufficiently large, then the equation $y^m = x^n + D$ has no solutions $x, y \in \mathbb{Z}$ with $x, y \notin \{0, \pm 1\}$. (*Hint.* You'll need to keep track of how the constant in (a) depends on m and n.)

9.18. Let E be the elliptic curve $y^2 = x^3 + 2089$.

(a) Prove that the points

$$P_1 = (-12, 19), \quad P_2 = (-10, 33), \quad P_3 = (-4, 45), \quad P_4 = (3, 46),$$

are independent points in $E(\mathbb{Q})$.

(b) ***** Prove that $E(\mathbb{Q}) \cong \mathbb{Z}^4$ and that P_1, P_2, P_3, P_4 are a basis for $E(\mathbb{Q})$.

(c) Find 10 more points (x, y) in $E(\mathbb{Q})$ with $x, y \in \mathbb{Z}$ and $y > 0$. Express these integral points in terms of the basis listed in (a).