BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 23, Number 1, July 1990

OLD AND NEW CONJECTURED DIOPHANTINE INEQUALITIES

SERGE LANG *

The original meaning of diophantine problems is to find all solutions of equations in integers or rational numbers, and to give a bound for these solutions. One may expand the domain of coefficients and solutions to include algebraic integers, algebraic numbers, polynomials, rational functions, or algebraic functions. In the case of polynomial solutions, one tries to bound their degrees. Inequalities concerning the size of solutions of diophantine problems are called diophantine inequalities.

During the past few years, new insights have been gained in old problems combined with new ones, and great coherence has been achieved in understanding a number of diophantine inequalities. Some of these results, notably the first section, can be formulated in very simple terms, almost at the level of high school algebra.

Received by the editors July 24, 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 11D41, 11D75, 11G05, 11G30.

* Supported by NSF grant, but...: In 1981 the Yale administration turned down the NSF grant to me as principal investigator because I would not sign effort reports. (However the math dept. at Yale passed a resolution to return all effort reports unsigned to the administration, and the Yale administration did nothing about other people's grants.) I was without a grant for 4 years. Then I reapplied for some items (such as travel expenses) in 1985, via someone else's grant, but I have not reapplied for summer salary support since 1981, and have therefore been without such support since. The fact that I am listed by the NSF as being "supported" by the NSF has given rise to misunderstandings concerning the nature of the support I received this last decade, whence it is important to specify the nature of such support. In addition, I find that the shortage of funds affecting the NSF makes it impossible for them to fulfill properly their funding role by following past policies on the allocation of funds. Over the last few years I have found that on the whole, for each person getting a grant, there are several others equally qualified who have been dropped, thus causing chaos in the granting process, and serious misinterpretations as to its implications by university administrators. It is demoralizing for the younger generation of mathematicians when getting a grant becomes a crap-shoot, particularly when university administrators often interpret not having NSF support as making someone unsuitable for tenuring or promoting. As a result of all these factors (plus renewed bureaucratic pressures), I have decided not to apply for any further NSF support in any form whatsoever starting next year.

0273-0979/90 $1.00 + $.25 per page

I shall give a survey starting with these formulations, and ending with more sophisticated applications to elliptic curves. But I have made an attempt to have this article readable by a fairly broad audience by giving basic definitions, and limiting myself to the rational numbers whenever possible.

1. The abc conjecture. This conjecture evolved from the insights of Mason [Ma], Frey [Fr], Szpiro, and others. Mason started one recent trend of thoughts by discovering an entirely new relation among polynomials, in a very original work, as follows. Let $f(t)$ be a polynomial with coefficients in an algebraically closed field of characteristic 0. We define

$$n_0(f) = \text{number of distinct zeros of } f.$$

Thus $n_0(f)$ counts the zeros of f by giving each of them multiplicity one.

Mason's theorem. *Let* $a(t), b(t), c(t)$ *be relatively prime polynomials such that* $a + b = c$. *Then*

$$\max \deg\{a, b, c\} \leqq n_0(abc) - 1.$$

In the statement of Mason's theorem, observe that it does not matter whether we assume a, b, c relatively prime in pairs, or without common prime factor for a, b, c. These two possible assumptions are equivalent by the equation $a + b = c$. Also the statement is symmetric in a, b, c and we could have rewritten the equation in the form $a + b + c = 0$.

Mason's theorem is a theorem, not a conjecture, and can be proved as follows. Dividing by c, and letting $f = a/c$, $g = b/c$, we have

$$f + g = 1$$

where f, g are rational functions. Differentiating we get $f' + g' = 0$, which we rewrite as

$$\frac{f'}{f} f + \frac{g'}{g} g = 0,$$

so that

$$\frac{b}{a} = \frac{g}{f} = -\frac{f'/f}{g'/g}.$$

If R is a rational function, $R(t) = \prod (t - \rho_i)^{q_i}$ with $q_i \in \mathbf{Z}$, then

$$R'/R = \sum \frac{q_i}{t - \rho_i}$$

and the multiplicities disappear. Suppose

$$a(t) = \prod (t-\alpha_i)^{m_i}, \qquad b(t) = \prod (t-\beta_j)^{n_j}, \qquad c(t) = \prod (t-\gamma_k)^{r_k}.$$

Then

$$\frac{b}{a} = -\frac{f'/f}{g'/g} = -\frac{\sum \frac{m_i}{t-\alpha_i} - \sum \frac{r_k}{t-\gamma_k}}{\sum \frac{n_j}{t-\beta_j} - \sum \frac{r_k}{t-\gamma_k}}.$$

A common denominator for f'/f and g'/g is given by the product

$$N_0 = \prod (t-\alpha_i) \prod (t-\beta_j) \prod (t-\gamma_k),$$

whose degree is $n_0(abc)$. Observe that $N_0 f'/f$ and $N_0 g'/g$ are both polynomials of degrees at most $n_0(abc) - 1$. From the relation

$$\frac{b}{a} = -\frac{N_0 f'/f}{N_0 g'/g}$$

and the fact that a, b are assumed relatively prime we deduce the inequality in Mason's theorem.

As an application let us prove Fermat's theorem for polynomials. Thus let $x(t)$, $y(t)$, $z(t)$ be relatively prime polynomials such that one of them has degree $\geqq 1$, and such that

$$x(t)^n + y(t)^n = z(t)^n.$$

We want to prove that $n \leqq 2$. By Mason's theorem, we get

$$\deg x(t)^n \leqq \deg x(t) + \deg y(t) + \deg z(t) - 1$$

and similarly replacing x by y and z on the left-hand side. Adding, we find

$$n(\deg x + \deg y + \deg z) \leqq 3(\deg x + \deg y + \deg z) - 3.$$

This yields a contradiction if $n \geqq 3$.

Influenced by Mason's theorem, and considerations of Szpiro and Frey which we shall describe below, Masser and Oesterle formulated the *abc* conjecture for integers as follows. Let k be a non-zero integer. Define the **radical** of k to be

$$N_0(k) = \prod_{p|k} p$$

i.e. the product of the distinct primes dividing k. There is a classical analogy between polynomials and integers. Under that analogy, n_0 of a polynomial corresponds to $\log N_0$ of an integer. Thus for polynomials we had an inequality formulated additively,

whereas for integers we formulate the corresponding inequality multiplicatively. Note that if x, y are nonzero integers, then

$$N_0(xy) \leqq N_0(x)N_0(y),$$

and if x, y are relatively prime, then

$$N_0(xy) = N_0(x)N_0(y).$$

The abc conjecture. *Given $\varepsilon > 0$ there exists a number $C(\varepsilon)$ having the following property. For any nonzero relatively prime integers a, b, c such that $a + b = c$ we have*

$$\max(|a|, |b|, |c|) \leqq C(\varepsilon)N_0(abc)^{1+\varepsilon}.$$

Unlike the polynomial case, it is necessary to have the ε in the formulation of the conjecture, and the constant $C(\varepsilon)$ on the right-hand side. To simplify notation in dealing with the possible presence of such constants, if A, B are positive functions, we write

$$A \ll B$$

to mean that there exists a constant $C > 0$ such that $A \leqq CB$. Thus $A \ll B$ means that $A = O(B)$ in the big oh notation. We write

$$A \gg\ll B$$

to mean $A = O(B)$ and $B = O(A)$. In case the functions A, B depend on a parameter ε, the constant C also depends on ε.

The conjecture implies that many prime factors of abc occur to the first power, and that if some primes occur to high powers, then they have to be compensated by "large" primes, or many primes, occurring to the first power. Readers interested in seeing immediately the application to the Fermat problem and similar problems should skip the following remarks, having to do with the equation $2^n \pm 1 = k$. For n large, the abc conjecture would state that k is divisible by large primes to the first power or many primes to the first power. This phenomenon can be seen on the tables of [BLSTW]. The simplest examples showing the need for the constant $C(\varepsilon)$ in the abc conjecture were communicated to me by Wojtek Jastrzebowski and Dan Spielman as follows. We want to show that there is no constant $C > 0$ such that

$$\max(|a|, |b|, |c|) \leqq CN_0(abc).$$

Writing $3 = 1 + 2$, we find by induction that

$$2^n | (3^{2^n} - 1).$$

We consider the relations $a_n + b_n = c_n$ given by

$$3^{2^n} - 1 = c_n.$$

Then

$$N_0(a_n b_n c_n) = 3N_0(c_n) \leqq 3 \cdot 2 \frac{3^{2^n} - 1}{2^n}$$

so there is no constant C such that $c_n \leqq 3CN_0(c_n)$. Other examples can be constructed similarly since the role of 3 and 2 can be played by other integers: instead of 2 we use a prime p, and instead of 3 we use an integer $\equiv 1 \bmod p$.

In line with these examples, we now show that the *abc* conjecture implies a classical conjecture:

There are infinitely many primes p such that

$$2^{p-1} \not\equiv 1 \mod p^2.$$

We follow Silverman [Si]. First a remark. Let S be the set of primes such that

$$2^{p-1} \not\equiv 1 \mod p^2.$$

We claim that if n is a positive integer, and p is a prime such that $2^n \equiv 1 \bmod p$ but $2^n \not\equiv 1 \bmod p^2$, then p is in S. Indeed, let d be the period of 2 in the multiplicative group $(\mathbf{Z}/p\mathbf{Z})^*$ of units in $\mathbf{Z}/p\mathbf{Z}$, which has order $p-1$. Then d divides $p-1$ and also divides n. Furthermore $2^n \equiv 1 \bmod p$ but $2^n \not\equiv 1 \bmod p^2$ implies $2^d \not\equiv 1 \bmod p^2$. Hence $2^{p-1} \not\equiv 1 \bmod p^2$, as one sees by writing $p - 1 = dm$, with m prime to p, and $2^d = 1 + pk$ with k prime to p. Then

$$2^{p-1} \equiv 1 + pmk \not\equiv 1 \mod p^2,$$

so p is in S as claimed.

Now suppose that S is finite. Write

$$2^n - 1 = u_n v_n$$

where u_n is the product of primes in S, and all primes dividing v_n are not in S. Then u_n is bounded. If $p | v_n$ then by the claim, p^2 divides $2^n - 1$, so p^2 divides v_n. By the *abc* conjecture applied to the equation

$$(2^n - 1) + 1 = 2^n$$

we conclude that

$$u_n v_n \ll (u_n v_n^{1/2})^{1+\varepsilon} \ll v_n^{(1+\varepsilon)/2}$$

whence v_n is bounded, a contradiction.

Actually, in line with Lang-Trotter conjectures, the probability that $2^{p-1} \equiv 1 + pk (\mathrm{mod}\, p^2)$ with a fixed residue class $k \bmod p$ should be $O(1/p)$, so the number of primes $p \leqq x$ such that $2^{p-1} \equiv 1 \bmod p^2$ should be

$$O\left(\sum_{p \leqq x} \frac{1}{p}\right) = O(\log \log x).$$

Thus most primes should have the property that $2^{p-1} \not\equiv 1 \bmod p^2$.

We now pass to the application of the *abc* conjecture to various diophantine equations. In the case of polynomials, we got an explicit bound for the degree of Fermat's equation satisfied by polynomials which are not all constant. Since there is an unknown constant $C(\varepsilon)$ floating around in the *abc* conjecture, we shall get only an unknown bound for the classical case of Fermat's equation over the integers. Thus we define the **asymptotic Fermat problem** to state that there exists an integer n_1 such that for all $n \geqq n_1$ the equation

$$x^n + y^n = z^n$$

has only a trivial solution in integers, that is with one of x, y, z equal to 0. Of course, for the asymptotic Fermat problem as well as the ordinary one, we may and do assume that x, y, z are relatively prime.

The abc conjecture implies the asymptotic Fermat problem.

Indeed, suppose $x^n + y^n = z^n$ with x, y, z relatively prime. By the *abc* conjecture, we have

$$|x^n| \ll |xyz|^{1+\varepsilon}$$
$$|y^n| \ll |xyz|^{1+\varepsilon}$$
$$|z^n| \ll |xyz|^{1+\varepsilon}.$$

Taking the product yields

$$|xyz|^n \ll |xyz|^{3+\varepsilon},$$

whence for $|xyz| > 1$ we get n bounded. The extent to which the *abc* conjecture is proved with an explicit constant $C(\varepsilon)$ (or say

$C(1)$ to fix ideas) yields the corresponding explicit determination of the bound for n in the application.

We shall now see how the abc conjecture implies other conjectures by Hall, Szpiro, and Lang-Waldschmidt.

> **Hall's original conjecture** *is that if* u, v *are relatively prime*[1] *nonzero integers such that* $u^3 - v^2 \neq 0$ *then*
>
> $$|u^3 - v^2| \gg |u|^{1/2-\varepsilon}.$$

Note that Hall's conjecture describes how small $|u^3 - v^2|$ can be, and the answer is not too small, as described by the right-hand side. Furthermore, if $|u^3 - v^2|$ is small, then $|u^3| \gg\ll v^2$ so $|v| \gg\ll |u|^{3/2}$. The Hall conjecture can also be interpreted as giving a bound for integral relatively prime solutions of

$$v^2 = u^3 + b \quad \text{with integral } b.$$

Then we find

$$|u| \ll |b|^{2+\varepsilon}.$$

More generally, as in Lang-Waldschmidt [La3], p. 212, let us fix nonzero integers A, B and let u, v, k, m, n be variable, with u, v relatively prime and $mn > m + n$. Put

$$Au^m + Bv^n = k.$$

By the *abc* conjecture, we get

$$|u|^m \ll |uvN_0(k)|^{1+\varepsilon}$$
$$|v|^n \ll |uvN_0(k)|^{1+\varepsilon}.$$

If, say, $|Au^m| \leqq |Bv^n|$ then $|u| \ll |v|^{n/m}$. We substitute this estimate for u to get an inequality entirely in terms of v, namely

$$|v|^n \ll |v^{1+n/m}N_0(k)|^{1+\varepsilon} = |v|^{(1+n/m)(1+\varepsilon)}N_0(k)^{1+\varepsilon}.$$

We first bring all powers of v to the left-hand side. We have

$$n - 1 - \frac{n}{m} = \frac{mn - (m+n)}{m}.$$

[1] Actually the original conjecture does not make the assumption that u, v are relatively prime, only that $u^3 - v^2 \neq 0$. The original conjecture also follows from the *abc* conjecture by extracting a common factor, and using the same method as that indicated below. We make the relatively prime assumption to avoid secondary technical complications, and we leave the proof in general to the reader.

We let the reader take care of the extra ε, so we obtain

$$(1) \qquad |v| \ll N_0(k)^{m(1+\varepsilon)/(mn-(m+n))} \quad \text{and then also}$$
$$|u| \ll N_0(k)^{n(1+\varepsilon)/(mn-(m+n))}$$

because the situation is symmetric in u and v. Again by the *abc* conjecture, we have

$$|k| \ll |uvN_0(k)|^{1+\varepsilon},$$

so using the estimate for $|uv|$ coming from the product of the inequalities in (1) we find

$$(2) \qquad |k| \ll N_0(k)^{mn(1+\varepsilon)/(mn-(m+n))}.$$

The Hall conjecture concerning $u^3 - v^2 = k$ is a special case of (1), after we replace $N_0(k)$ with $|k|$, because $N_0(k) \leqq |k|$.

Again take $m = 3$, $n = 2$ and take $A = 4$, $B = -27$. In this case, we write D instead of k, and we find for

$$D = 4u^3 - 27v^2$$

that

$$(3) \qquad |u| \ll N_0(D)^{2+\varepsilon} \quad \text{and} \quad |v| \ll N_0(D)^{3+\varepsilon}.$$

These inequalities are supposed to hold at first for u, v relatively prime. Suppose we allow u, v to have some bounded common factor, say d. Write

$$u = u'd \quad \text{and} \quad v = v'd$$

with u', v' relatively prime. Then

$$D = 4d^3u'^3 - 27d^2v'^2.$$

Now we can apply inequalities (1) with $A = 4d^3$ and $B = -27d^2$, and we find the same inequalities (3), with the constant implicit in the sign $\ll$ depending also on d, or on some fixed bound for such a common factor. Under these circumstances, we call inequalities (3) the **generalized Szpiro conjecture.**

The original Szpiro conjecture was stated for what is called a "minimal discriminant" D. We shall discuss the notion of a minimal discriminant in §4, when we go further into the theory of elliptic curves. Szpiro's inequality was stated in the form

$$|D| \ll N(D)^{6+\varepsilon},$$

where $N(D)$ is a more subtle invariant which is commonly used in the literature, namely the conductor. But for our purposes, it is sufficient and much easier to use the radical $N_0(D)$, since the conductor is harder to define, and its subtleties are irrelevant for what we are doing.

Note that the generalized Szpiro conjecture actually bounds $|u|$, $|v|$ and not just $|D|$ itself in terms of the right power of $N_0(D)$.

The point of D is that it occurs as the discriminant of an elliptic curve. The recent trend of thoughts in the direction we are discussing was started by Frey [Fr], who associated with each solution of $a+b=c$ the elliptic curve

$$y^2 = x(x-a)(x+b),$$

which we call the **Frey curve**. The discriminant of the right-hand side is the product of the differences of the roots squared, and so

$$D = (abc)^2.$$

We make a translation

$$\xi = x + \frac{b-a}{3}$$

to get rid of the x^2-term, so that our equation can be rewritten

$$y^2 = \xi^3 - \gamma_2\xi - \gamma_3,$$

where γ_2, γ_3 are homogeneous in a, b of appropriate weight. The discriminant does not change because the roots of the polynomial in ξ are translations of the roots of the polynomial in x. Then

$$D = 4\gamma_2^3 - 27\gamma_3^2.$$

The translation with $(b-a)/3$ introduces a small denominator. One may avoid this denominator by using the curve

$$y^2 = x(x-3a)(x+3b),$$

so that γ_2, γ_3 then come out to be integers, and one can apply the generalized Szpiro conjecture to the discriminant, which then has an extra factor

$$D = 3^6(abc)^2 = 4\gamma_3^3 - 27\gamma_3^2.$$

The Szpiro conjecture implies asymptotic Fermat.

Indeed, suppose that

$$a = u^n, \quad b = v^n, \quad \text{and} \quad c = w^n$$

with relatively prime u, v, w. Then

$$4\gamma_2^3 - 27\gamma_3^2 = 3^6(uvw)^{2n},$$

and we get a bound on n from the Szpiro conjecture

$$|D| \ll N_0(D)^{6+\varepsilon}.$$

Of course any exponent would do, e.g. $|D| \ll N_0(D)^{100}$ for asymptotic Fermat.

We have already seen that the *abc* conjecture implies generalized Szpiro.

Conversely, generalized Szpiro implies abc.

Indeed, the correspondence

$$(a, b) \leftrightarrow (\gamma_2, \gamma_3)$$

is invertible, and has the "right" weight. A simple algebraic manipulation left to the reader shows that the generalized Szpiro estimates on γ_2, γ_3 imply the desired estimates on $|a|, |b|$. (I found this manipulation a good assignment for my undergraduate algebra class.)

From the equivalence between *abc* and generalized Szpiro, one can use the examples given at the beginning to show that the epsilon is needed in the Szpiro conjecture.

Hall made his conjecture in 1971, actually without the epsilon so it had to be adjusted later. The final setting of the proofs in the simple *abc* context which we gave above had to await Mason and the *abc* conjecture a decade later.

Let us return to the polynomial case and Mason's theorem. The proofs that the *abc* conjecture implies the other conjectures apply as well in this case, so the analog use of Hall, Szpiro and Lang-Waldschmidt are also proved in the polynomial case. Actually, it had already been conjectured in [BCHS] that if f, g are non-zero polynomials such that $f^3 - g^2 \neq 0$ then

$$\deg(f(t)^3 - g(t)^2) \geqq \tfrac{1}{2}\deg f(t) + 1.$$

This (and its analogue for higher degrees) was proved by Davenport [Dav] in 1965, but we now see it as a consequence of Mason's theorem. Both in the case of Hall's conjecture for integers and Davenport's theorem, the point is to determine what lower bound can occur in a difference between a cube and a square, in the simplest case. The result for polynomials is particularly clear

since, unlike the case of integers, there is no extraneous undetermined constant $C(\varepsilon)$ floating around, and there is even $+1$ on the right-hand side.

The polynomial case as in Davenport and the Hall conjecture for integers are of course not independent. Examples in the polynomial case parametrize cases with integers when we substitute integers for the variable. Examples are given in [BCHS], one of them due to Birch being:

$$f(t) = t^6 + 4t^4 + 10t^2 + 6 \quad \text{and} \quad g(t) = t^9 + 6t^7 + 21t^5 + 35t^3 + \tfrac{63}{2}t,$$

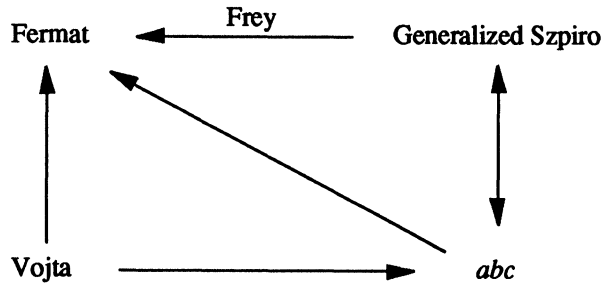
whence

$$\text{degree}(f(t)^3 - g(t)^2) = \tfrac{1}{2}\deg f + 1.$$

This example shows that Davenport's inequality is best possible, because the degree attains the lowest possible value permissible under his theorem. Substituting large integral values of $t \equiv 2 \bmod 4$ gives examples of similarly low values for $x^3 - y^2$. A fairly general construction is given by Danilov [Dan]. See also the discussion of similar questions relevant to the size of integral points on elliptic curves in [La2], for instance Conjecture 5.

For those who know the theory of function fields in one variable, it is clear that the *abc*-property can also be formulated in that case, and can also be proved in that case. In fact, historically it was done that way as in [Ma3]. However, we are interested in algebraic number fields, so we shall now go to the next level of exposition and discuss heights of points in number fields.

So far we have proved the implications and equivalences relating three corners of the following diagram.



Our purpose is to fill in the Vojta corner, and show how the conjectures follow from the Vojta conjecture. One of the crucial points

here is that we cannot stay within the realm of rational numbers, we must look at algebraic numbers. It will be seen how diophantine properties of a *family of curves* over the rational numbers depends on the diophantine properties of a *single curve* but uniformly for solutions in finite extensions of the rationals of bounded degree.

We want to estimate how big solutions of diophantine equations can be. What does "big" mean? Let $x = c/d$ be a rational number expressed in lowest form with relatively prime integers c, d. We define the **height** of x to be

$$h(x) = \log \max(|c|, |d|).$$

Similarly, if $P = (x_0, \ldots, x_M)$ is a point in projective space with integer coordinates x_j which are relatively prime, then we define the **height**

$$h(P) = \log \max |x_j|.$$

The abc conjecture thus bounds the height of the point (a, b, c), with relatively prime integers a, b, c which can be seen as representing a point in projective 2-space.

The next two sections which carry out the above program are independent of the final three sections which give applications of the abc conjecture to elliptic curves, via the Szpiro conjecture. The reader may therefore read these in any order.[2]

[2]**Remark for those who know or want to find out about modular curves.** I don't want to discuss modular curves here, I want to confine the discussion to diophantine inequalities. But for those interested, it may be useful to make the following comments, for which we assume that the reader will either know the required definitions or look them up elsewhere.

Frey started the recent train of thoughts concerning Fermat and elliptic curves by pointing out that the elliptic curve associated with a solution of Fermat would have remarkable properties which should contradict the Taniyama-Shimura conjecture that all elliptic curves over the rational are "modular" [Fr]. Frey's idea was to show that the elliptic curve then could not exist, whence Fermat follows. There were serious difficulties in realizing this idea. Serre [Se] pointed out that one needed apparently more than the Taniyama-Shimura conjecture, for instance another conjecture which he had made, concerning the modularity of Galois representations over the rationals. Then Ribet [Ri] proved enough of the Serre conjecture in the modular case to show that the Taniyama-Shimura conjecture suffices for the Fermat application. We do not go into this modular aspect here. On the contrary, in the present discussion, we are pointing in a different direction, namely the direction of diophantine analysis and diophantine inequalities. The Szpiro conjecture can be viewed as lying just in the middle, and is susceptible of being handled or proved in either direction. Thus the abc conjecture has a modular interpretation via the generalized Szpiro conjecture to which it is equivalent. I must also point out that the modular route to Fermat via Ribet-Serre-Taniyama-Shimura would prove Fermat unconditionally. The route via the diophantine inequalities depends

2. The height for algebraic points. In this section we discuss algebraic numbers, and we describe how to define the height for points with algebraic coordinates, which will be used in §3. After that we return to the rational numbers, but we shall use p-adic numbers also.

Let F be a number field, i.e. a finite extension of the rational numbers $\mathbf{Q}$. Then F has a family of absolute values, p-adic and archimedean (at infinity, as one says). We shall now describe these absolute values.

For each prime p there is the **p-adic absolute value** on $\mathbf{Q}$, defined on a rational number $a = p^r c/d$ with $(c, d) = 1$, $p \nmid cd$, by

$$|a|_p = 1/p^r .$$

This p-adic absolute value has a finite number of extensions to F, which are called **p-adic** also. Let v be one of these, extending v_p on $\mathbf{Q}$. Let F_v be the completion of F at v. Then F_v is a finite extension of the field of p-adic numbers $\mathbf{Q}_p$. The absolute value on $\mathbf{Q}_p$ extends uniquely to F_v. Each v is induced by an imbedding of F into the algebraic closure of $\mathbf{Q}_p$:

$$\sigma_v : F \to \mathbf{Q}_p^{\mathrm{a}}$$

and v is induced by the absolute value on $\mathbf{Q}_p^{\mathrm{a}}$. Conversely, given such an imbedding $\sigma : F \to \mathbf{Q}_p^a$ we let v_σ be the induced absolute value.

The rational numbers also have the ordinary absolute value, extending to the real numbers, and denoted by v_∞. An extension of v_∞ to F is said to be **at infinity**, and there is a finite number of such extension. Let v extend v_∞ to F. Then v is induced by an imbedding

$$\sigma_v : F \to \mathbf{R}^{\mathrm{a}} = \mathbf{C}$$

of F into the complex numbers. Thus the situations at infinity and at the ordinary primes are entirely similar. In each case, imbeddings of F into $\mathbf{Q}_v^{\mathrm{a}}$ which differ by an isomorphism of F_v over $\mathbf{Q}_v$ give rise to the same absolute value on F, and conversely. In the case of absolute values at infinity, a pair of complex conjugate imbeddings of F into $\mathbf{C}$ corresponds to an extension of the ordinary absolute value from $\mathbf{Q}$ to F.

---

on the constants in the estimates for these inequalities, and will be as effective as these constants can be made to be effective.

Now consider a point $P = (x_0, \ldots, x_n)$ in projective space $\mathbf{P}^n$, with coordinates in F so $x_j \in F$ and not all $x_j = 0$. We define the **height** of P by the formula

$$h(P) = \frac{1}{[F : \mathbf{Q}]} \sum_v [F_v : \mathbf{Q}_v] \log \max_j |x_j|_v$$

where $[F : \mathbf{Q}]$ denotes the degree of the extension, and where the sum is taken over all the above described absolute values v on F. The Artin-Whaples product formula (sum formula in our case) asserts that for $a \in F$, $a \neq 0$ we have

$$\sum_v [F_v : \mathbf{Q}_v] \log |a|_v = 0,$$

and so the height indeed depends only on the point in projective space. An elementary property of the absolute values implies that the height is independent of the field F in which the coordinates $x_0, \ldots, x_n$ lie: this is the reason for the normalizing factor $1/[F : \mathbf{Q}]$ in front of the formula defining the height.

Note that if $F = \mathbf{Q}$ and $x_0, \ldots, x_n$ are relatively prime integers, then

$$h(P) = \log \max |x_j|,$$

where the absolute value here is the ordinary one on the rational numbers.

We shall need another notion of algebraic number theory. The field F has a subring R, the ring of algebraic integers which are those elements of F satisfying an equation

$$T^n + a_{n-1} T^{n-1} + \cdots + a_0 = 0$$

whose coefficients $a_0, \ldots, a_{n-1}$ lie in the ordinary integers $\mathbf{Z}$. This ring R has a basis $\{w_1, \ldots, w_N\}$ over $\mathbf{Z}$, where $N = [F : \mathbf{Q}]$. If σ_j $(j = 1, \ldots, N)$ ranges over all imbeddings of F into $\mathbf{C}$, then the **logarithmic discriminant** $d(F)$ is defined by

$$\begin{aligned} d(F) &= \frac{1}{[F : \mathbf{Q}]} \log |\det \sigma_i w_j|^2 \\ &= \frac{1}{[F : \mathbf{Q}]} \log |\text{discriminant of } F \text{ over } \mathbf{Q}|. \end{aligned}$$

It is easy to prove that if F_1, F_2 are number fields, then

$$d(F_1 F_2) \leqq d(F_1) + d(F_2).$$

Also if $F_1 \subset F_2$ then $d(F_1) \leqq d(F_2)$.

Given a point P in projective space, we let

$$d(P) = d(F(P)).$$

Furthermore, the discriminant of the field $F(P)$ is insensitive to the higher power of primes dividing the coordinates of the point.

EXAMPLE. Suppose $x = a^{1/n}$ where a is a positive integer. Let

$$a = p_1^{\nu_1} \cdots p_r^{\nu_r}$$

be the factorization with $\nu_i \geqq 1$ and with distinct primes $p_1, \ldots, p_r$. Then

$$\log N_0(a) = \sum \log p_i.$$

Furthermore

$$\mathbf{Q}(a^{1/n}) \subset \mathbf{Q}(p_1^{1/n}, \ldots, p_r^{1/n}).$$

If α is a root of $\alpha^n = p$, then the discriminant is the norm of $n\alpha^{n-1}$, so the discriminant is bounded by $n^n p^{n-1}$. Hence

$$d(\mathbf{Q}(\alpha)) \leqq \log n + \frac{n-1}{n} \log p.$$

We apply this estimate to each prime dividing a to get

$$d(\mathbf{Q}(x)) \leqq r \log n + \frac{n-1}{n} \log N_0(a).$$

REMARK. For $N_0(a) \to \infty$ we have

$$r = o(\log N_0(a)).$$

Indeed, if r is bounded this is trivial. If r is unbounded, then $p_1 \cdots p_r \geqq r!$, so our assertion follows from Sterling's formula that $r! \geqq r^r e^{-r}$.

So far we have dealt with notions of algebraic number theory. We now pass to the height in the context of algebraic geometry, in the case of curves. We must assume that the reader is acquainted with the basic notion of an algebraic curve, imbeddable in projective space. Such a curve X will always be assumed irreducible. It is defined by a system of homogeneous polynomial equations when it is so imbedded, and is said to be **defined over** F if the coefficients of these equations lie in F. A **divisor** on X is a formal linear combination of points, with integer coefficients. The **degree** of a divisor is defined to be the sum of these coefficients. The curve X is covered by affine pieces, and a typical affine piece may be defined by one equation $f(x, y) = 0$, for instance. A point will usually be taken with coordinates x, y which lie in some number

field, i.e. we deal mostly with algebraic points. The set of points with coordinates in a field E is denote by $X(E)$. Then $X(\mathbf{C})$ is the set of complex points, and if the curve is nonsingular, then $X(\mathbf{C})$ is a compact Riemann surface. The genus g of X may be defined as the genus of this surface, but there are also algebraic definitions. For instance, if X is defined by one homogeneous irreducible equation

$$H(T_0, T_1, T_2) = 0$$

in the projective plane, if X is nonsingular, and H has degree d, then the genus of X is $(d-1)(d-2)/2$.

The genus is also equal to the dimension of the space of regular differential forms. Say over the complex numbers, let ydx be a meromorphic differential form on X where y, x are rational functions on X. Let P be a complex point. Let t be a local uniformizing parameter at P. We write

$$ydx = y(t)\frac{dx}{dt}\,dt,$$

where y, x are expressible as power series in t. Then we define

$$\operatorname{ord}_P ydx = \text{order of the power series } y(t)\frac{dx}{dt}.$$

We can associate a **divisor to the differential form** ydx by letting

$$(ydx) = \sum_P \operatorname{ord}_P(ydx)(P).$$

Then the degree of this form satisfies

$$\deg(ydx) = \sum_P \operatorname{ord}_P(ydx) = 2g - 2.$$

If f is a rational function on X, then one can also associate a **divisor** (f) to f, namely at the point P we define $\operatorname{ord}_P f =$ order of the power series $f(t)$, in terms of the local parameter t. Then $\deg(f) = 0$ (which corresponds to the product formula). A divisor is said to be **rationally equivalent to** 0 if it is the divisor of a rational function. Equivalence among divisors will always refer to rational equivalence in what follows. In light of the relation $\deg(f) = 0$ we see that the degree function is defined on equivalence classes. The class of divisors of rational differential forms is called the **canonical class**.

Let X be a projective nonsingular curve defined over a number field. To each divisor D on X one can associate a function, the

height,

$$h_D: X(\mathbf{Q}^a) \to \mathbf{R}$$

from the set of algebraic points into the reals, satisfying the following properties, and uniquely determined by them modulo bounded functions:

1. $D \mapsto h_D$ is well defined mod $O(1)$ on the rational equivalence class of D and is a homomorphism, i.e. is additive in D.
2. If D is a hyperplane section in a projective imbedding, then $h_D(P)$ is the height of a point P in that projective imbedding.

The uniqueness follows because given a divisor D there exists two projective imbeddings of X with hyperplane sections H_1 and H_2 respectively such that D is equivalent to $H_1 - H_2$ (this is a basic lemma of algebraic geometry).

3. The Vojta conjecture. The conjectures expressed by Vojta [Vo] are basic to the subject. I give here only one of the conjectures having to do with curves.

> **Vojta conjecture** Let X be a projective nonsingular curve defined over a number field. Let K be the canonical class of X. Then given $\varepsilon > 0$,
>
> $$h_K(P) \leqq (1+\varepsilon)d(P) + O_\varepsilon(1) \quad \text{for } P \in X(\mathbf{Q}^a).$$

In the first place, observe that for a curve of genus $\geqq 2$, the Vojta conjecture immediately implies that the set of rational points $X(F)$ is finite (Mordell conjecture—Faltings theorem). Indeed, in that case $d(P)$ is constant, so the height of such points is bounded, and it is easy to show that there is only a finite number of points of bounded degree and bounded height.

We shall now see how the Vojta conjecture implies asymptotic Fermat and also implies the *abc* conjecture. Note that we shall use the Vojta conjecture only for points P of bounded degree over F as in [Vo], *Appendix ABC* p. 84.

Suppose we want to prove Vojta implies asymptotic Fermat. Let

$$X_n: x^n + y^n = z^n$$

be the Fermat curve and consider X_4 which has genus 3. For a hypersurface of degree d in $\mathbf{P}^m$ the canonical class is that of $(d-(m+1)))H$, where H is a hyperplane. So for X_n in $\mathbf{P}^2$ the

canonical class is that of $(n-3)H$, and for X_4 the canonical class is

$$K = H.$$

Let $u^n + v^n = w^n$ in relatively prime integers. Associate the point

$$P = (x, y, z) = (u^{n/4} : v^{n/4} : w^{n/4}) \quad \text{on } X_4.$$

From the definition of the height, it is immediate that

$$h(P) = \frac{n}{4} \log \max(|u|, |v|, |w|).$$

By Vojta's conjecture we get

$$\frac{n}{4} \log \max|u|, |v|, |w| \ll \log N_0(uvw) + O(1),$$

which gives a bound for n.

Similarly to prove the abc conjecture from Vojta, suppose $a + b = c$. Fix n. Associate the point

$$P = (a^{1/n} : b^{1/n} : c^{1/n}) \quad \text{on } X_n.$$

Since $K_n = (n-3)H$ we get by Vojta's conjecture

$$h_K(P) = \frac{n-3}{n} \log \max(|a|, |b|, |c|) \leqq (1+\varepsilon_1) \log N_0(abc) + O_n(1).$$

This IS the log of the abc inequality. We let $n \to \infty$ with $\varepsilon = \varepsilon_1 + 4/n$ or whatever.

We shall conclude with sections on elliptic curves, showing how the abc conjecture implies diophantine properties of such curves over number fields.

4. Elliptic curves and minimal equations. An elliptic curve for our purposes will be a curve which can be represented by an equation in Weierstrass form

$$y^2 = x^3 - \gamma_2 x - \gamma_3,$$

with coefficients γ_2, γ_3 in some field F. Let A denote the curve. The set of points with coordinates $x, y \in F$ together with the point at infinity is denoted by $A(F)$, and is a group. As before, we shall take F to be mostly a subfield of the complex numbers, in which case $A(F)$ is a subgroup of $A(\mathbf{C})$. Over $\mathbf{C}$, the curve is parametrized by the Weierstrass functions

$$z \mapsto (\wp(z), \tfrac{1}{2}\wp'(z)),$$

giving an analytic isomorphism $\mathbf{C}/\Lambda \to A(\mathbf{C})$ where Λ is a lattice. The lattice points go to the point at infinity under this parametrization. The addition formula for the $\wp$-function defines the group law on $A(\mathbf{C})$, and this addition formula is given by rational functions on the coordinates (x, y), with coefficients in $\mathbf{Q}$, which is why $A(F)$ is a subgroup.

As before we have the **discriminants**

$$4\gamma_2^3 - 27\gamma_3^2 = D \quad \text{and} \quad \Delta = 16D.$$

For any number $c \neq 0$ we can change the elliptic curve by an isomorphism, which under the Weierstrass parametrization maps $z \mapsto cz$. Then the representation of this isomorphism on the equation has the effect:

$$x \mapsto x' = c^{-2}x, \qquad y \mapsto y' = c^{-3}y,$$
$$\gamma_2 \mapsto \gamma_2' = c^{-4}\gamma_2, \qquad \gamma_3 \mapsto \gamma_3' = c^{-6}\gamma_3,$$

so that the isomorphic curve satisfies the equation

$$y'^2 = x'^3 - \gamma_2' x' - \gamma_3'.$$

Suppose the elliptic curve is defined over $\mathbf{Q}$, that is $\gamma_2, \gamma_3 \in \mathbf{Q}$. By an appropriate such isomorphism, we can make $\gamma_2, \gamma_3 \in \mathbf{Z}$. Suppose p is a prime such that p^4 divides γ_2 and p^6 divides γ_3. Then we can change the elliptic curve by an isomorphism, letting $\gamma_2 \mapsto p^{-4}\gamma_2$ and $\gamma_3 \mapsto p^{-6}\gamma_3$. After having done so repeatedly until no further possible, we obtain what is a **minimal model** for the curve over $\mathbf{Z}$, and then Δ is called a **minimal discriminant**.

REMARK. The minimal discriminant is defined up to a factor of ± 1, and is an invariant of an isomorphism class of elliptic curves over $\mathbf{Q}$. We have slightly over simplified the situation in two ways. First by tying ourselves down to the Weierstrass model, we don't quite get the "right" notion for this minimal discriminant, because of the primes 2 and 3. One has to give a more general equation, first studied by Deuring, and carried out systematically in more recent times by Neron and Tate. Also to be absolutely correct, to take care of the primes 2 and 3, we really should consider a more general form of the Weierstrass model, namely:

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

One can obtain a minimal model, i.e. one with minimal discriminant, just as in the standard case of the Weierstrass equation. The

following statements refer to such a model, but the reader may think of $p \neq 2, 3$ and of the usual Weierstrass equation without harm.

The discussion of isomorphisms above contained the essential aspects of the minimal discriminant. Furthermore, by dealing over the rationals, we avoid all the problems which come from non-unique factorization in number fields, and the existence of a non-trivial group of units in the ring of integers of the number field. A reader acquainted with elementary algebraic number theory will then see that in case of a nontrivial ideal class group, there is no unique minimal model of the curve with minimal discriminant, but there is a finite number of models, with relatively minimal discriminants. These are secondary considerations for what we are doing here, where we want to see clearly the effect of the *abc* conjecture on the diophantine aspects of the curve. These essential aspects are all present for curves over the rationals.

We shall want to apply the Szpiro conjecture, but we must take into account that γ_2, γ_3 may not be relatively prime. We still want to see that for a minimal model, we have $|\Delta| \ll N_0(\Delta)^s$ for some s independent of A. Indeed, by the minimality assumption, if p^m is a common factor of γ_2, γ_3 then $m \leqq 5$. Therefore, if $d = (\gamma_2, \gamma_3)$ is the greatest common divisor, then

$$d \leqq N_0(d)^5.$$

By factoring out a common factor, canceling, and small algebraic manipulations, it is easy to see that for a minimal equation, by the *abc* conjecture we have

$$|\Delta| \ll N_0(\Delta)^{6+\varepsilon} d^4$$

and therefore

$$|\Delta| \ll N_0(\Delta)^s$$

for some low integer s which we may call the **Szpiro exponent**. Without loss of generality, we may then assume that

$$|\Delta| \leqq N_0(\Delta)^s$$

except for a finite number of Δ, and so except for a finite number of elliptic curves. We shall omit these exceptional curves, and *assume in the sequel that* $|\Delta| \leqq N_0(\Delta)^s$.

5. The Szpiro conjecture and torsion points. Let F be a number field, that is a finite extension of the rational numbers. The

Mordell-Weil theorem asserts that the group of rational points $A(F)$ is finitely generated. (Mordell originally proved the theorem over the rationals.) It is a major problem to describe the order of the torsion group and the rank of $A(F)$. We are here concerned with the torsion group, which is finite by the Mordell-Weil theorem. A standard conjecture states:

> *Given the number field F, there exists a positive number C such that for every elliptic curve A defined over F, the order of the torsion group $A(F)_{\text{tor}}$ is bounded by C.*

Over the rational numbers, this was proved in a very strong form by Mazur [Maz], who showed that the order of the torsion group is bounded by 16. For this purpose, Mazur developed a whole theory on modular curves. Here we are concerned with a statement which is weaker in the sense that no specific bound like 16, is given, but stronger in that we want the statement for any number field. Results for number fields have been obtained by Kubert [Ku]. In this section, we show that the *abc* conjecture for number fields implies the uniform boundedness of torsion as stated above by an argument due to Frey which he wrote me in 1986.

For simplicity, we shall work over the rational numbers. A reader acquainted with the basic properties of elliptic curves and number fields will immediately see that the arguments generalize.

As usual we define the isomorphism invariant

$$j = 3^3 4^6 \gamma_2^3 / \Delta.$$

REMARK. The primes which divide the denominator of j play a special role. These primes must also divide Δ, but the converse need not hold. We shall have to use a relatively deep fact:

> *If there exists a rational point in $A(\mathbf{Q})$ of prime order $n \geqq 5$, then γ_2 and γ_3 are relatively prime except for small powers of 2 and 3, and also possibly n itself.*

This fact is hard to prove, and we make some comments for those somewhat acquainted with elliptic curves, or those who may wish to pursue this matter further. Let p be a prime $\geqq 5$. One may reduce the minimal equation of an elliptic curve $\bmod p$. Then the reduced equation defines a possibly reducible curve, called the **fiber**. The group law on the original elliptic curve reduces to a group law on the set of nonsingular points on the fiber, giving

rise to the theory of the Néron model. Cf. [Ne], the last section on elliptic curves, and the last table, as well as Artin's exposition (somewhat sketchy) of Néron models [Ar], especially Proposition 1.15. Now suppose that j is p-integral. If p is not a common factor of γ_2 and γ_3 then we say that the elliptic curve is **semistable** at p. If p is a common factor, then the curve reduces mod p to $y^2 = x^3$. Let us look at the curve over the p-adic field $\mathbf{Q}_p$. By the theory of Néron models, it can be shown that the nonsingular part of the fiber is an algebraic group, which contains the additive group as a subgroup of index at most 4, and that the kernel of reduction is a p-adic Lie group, which does not contain points of finite order. From this structure, we conclude that the curve is semistable at p when there exists a rational point of prime order $n \geqq 5$, except possibly when $p = n$, because all points on the additive group in characteristic p have order p. For a generalization and examples, see Lenstra-Oort [L-O], especially §3 for examples of large p-torsion p-adically.

Admitting the above fact, we see that except possibly for 2, 3, and n, the primes dividing the denominator of j are precisely the primes dividing Δ.

To prove a theorem about torsion points over the rationals, we must have a model for them which will contain enough algebraic information, even though it may involve some analysis. The classical parametrization $\mathbf{C}/\Lambda \to A(\mathbf{C})$ by the Weierstrass functions is not good enough for this purpose. However, we may take the lattice to have a $\mathbf{Z}$-basis $\{\tau, 1\}$ where τ is in the upper half plane: $\mathrm{Im}(\tau) > 0$. Then we put

$$q_\tau = q = e^{2\pi i \tau}.$$

We can then also represent $A(\mathbf{C})$ as a quotient of the multiplicative group as follows. We have the Fourier series expansion, power series in q, in all standard texts, including [La1] and [La3]:

$$(2\pi i)^{-4}\gamma_2 = \frac{1}{48}\left[1 + 240\sum_{n=1}^{\infty}\frac{n^3 q^n}{1-q^n}\right]$$

$$(2\pi i)^{-6}\gamma_3 = \frac{1}{2^5 3^3}\left[-1 + 504\sum_{n=1}^{\infty}\frac{n^5 q^n}{1-q^n}\right]$$

$$(2\pi i)^{-12}\Delta = q\prod_{n=1}^{\infty}(1-q^n)^{24}$$

$$j = \frac{1}{q} + 744 + 196884q + \cdots$$

and for a variable $t \in \mathbf{C}^*$:

$$(2\pi i)^{-2}x(t) = \frac{1}{12}\sum_{m\in\mathbf{Z}}\frac{q^m t}{(1-q^m t)^2} - 2\sum_{n=1}^{\infty}\frac{nq^n}{1-q^n} = X(t)$$

$$(2\pi i)^{-3}y(t) = \frac{1}{2}\sum_{m\in\mathbf{Z}}\frac{q^m t(1+q^m t)}{(1-q^m t)^3} = Y(t).$$

We get an analytic isomorphism

$$\mathbf{C}^*/q^{\mathbf{Z}} \xrightarrow{\approx} A(\mathbf{C}) \quad \text{by} \quad t \mapsto (x(t), y(t)).$$

The power series expansions for $X(t)$ and $Y(t)$ have essentially integral coefficients, and as Tate remarked in the late fifties, they can be used to parametrize an elliptic curve in the p-adic domain, where they converge provided that

$|q| < 1$ or equivalently p divides the denominator of j.

This elliptic curve, called the **Tate curve**, is isomorphic to the given curve over a quadratic extension. For simplicity, we shall argue as if this isomorphism is over $\mathbf{Q}_p$ itself.

As with the complex numbers, the absolute value on $\mathbf{Q}_p$ extends uniquely to the algebraic closure $\mathbf{Q}_p^{\mathrm{a}}$ and to the completion of the algebraic closure, which we denote by $\mathbf{C}_p$, and which plays the role of $\mathbf{C}$.

In the p-adic field, using the notation $u \sim v$ to mean that u/v is a p-adic unit, we see that

$$j \sim \frac{1}{q} \quad \text{and} \quad \Delta \sim q.$$

Except for the primes 2 and 3, γ_2 and γ_3 are then p-adic units. Tate normalizes the equation and the power series further to get rid of denominators involving 2 and 3 completely, but we don't go into this here.

Thus $t \mapsto (X(t), Y(t))$ gives a homomorphism

$$\mathbf{C}_p^* \to A(\mathbf{C}_p) \quad \text{inducing} \quad \mathbf{Q}_p^* \to A(\mathbf{Q}_p)$$

from the multiplicative group of the field to the group of points of A in the field. The kernel is the infinite cyclic group $q^{\mathbf{Z}}$. Similarly,

if F is a finite extension of $\mathbf{Q}_p$ in $\mathbf{C}_p$ we have a parametrization $F^*/q^{\mathbf{Z}} \to A(F)$. We then obtain a model for the torsion points of A in $A(\mathbf{C}_p)$. The points of order n are parametrized by the group generated by

$$q^{1/n}, \zeta_n \mod q^{\mathbf{Z}}$$

where $q^{1/n}$ is any one of the nth roots of q, well defined modulo an nth root of unity; and ζ_n is a primitive nth root of unity.

We shall first prove that there is only a finite number of primes n such that an elliptic curve over $\mathbf{Q}$ has a rational point of order n, by showing that the minimal discriminants of such curves are divisible by only a finite number of primes.

Suppose that the elliptic curve A over the rationals has a point $P \in A(\mathbf{Q})$ of order precisely n with n prime. For each prime number p dividing the denominator of j, and thus dividing Δ, this point is represented by $q^{1/n}$ or ζ_n. If P corresponds to $q^{1/n}$, since $P \in A(\mathbf{Q}_p)$ it follows that $q^{1/n} \in \mathbf{Q}_p$, so $p^n | q = q_A$. Suppose on the other hand that P corresponds to ζ_n for some primitive nth root of unity. Let (P) be the cyclic group generated by P, and let

$$B = A/(P)$$

be the quotient elliptic curve, which is then also defined over $\mathbf{Q}$. Then we have the minimal discriminants Δ_A and Δ_B, and the two parameters q_A and q_B coming from the Tate parametrization p-adically. In the present case, we claim that $q_B = q_A^n$. Indeed, the system here works just as in the case of a lattice. The map raising to the nth power gives an isomorphism

$$\mathbf{C}_p^*/(q, \zeta_n) \xrightarrow{\approx} \mathbf{C}_p^{*n}/(q^n) = \mathbf{C}_p^*/(q^n).$$

Thus the "period" group of $\mathbf{C}_p^*/(q, \zeta_n)$ is generated by q^n, in case P corresponds to ζ_n. (See the Remark below.) Therefore $q_B = q_A^n$. It follows that in this case, $p^n | q_B$. Hence in both cases,

$$p^n \text{ divides } q_A q_B.$$

In terms of the minimal discriminants, this implies that

$$p^n \text{ divides } \Delta_A \Delta_B.$$

But it is known that Δ_A and Δ_B are divisible by the same primes. By the Szpiro conjecture applied to Δ_A and Δ_B separately, we get

$$|\Delta_A \Delta_B| \ll N_0(\Delta_A)^s N_0(\Delta_B)^s = N_0(\Delta)^{2s}.$$

If $p_1, \ldots, p_r$ are the primes dividing Δ and $\neq n, 2, 3$, then we get

$$(p_1 \cdots p_r)^n \ll (p_1 \cdots p_r)^{2s_n 2s},$$

which gives a bound on n as effective as the constant in Szpiro's conjecture, unless there are no primes $p_1, \ldots, p_r$. But in that case, Δ is divisible only by n (or 2 and 3). Then for $n > 5$, the curve has good reduction modulo 5, say, and reduction mod 5 induces an isomorphism on the points of order n, which is impossible for n sufficiently large since the cardinality of $A(\mathbf{F}_5)$ is bounded.

It follows that the discriminants of minimal models are divisible by only a finite number of primes. By what we saw at the end of §4, this implies that there is only a finite number of values Δ_A for minimal models A. It remains to be shown that for fixed Δ the equation

$$\Delta = 4\gamma_2^3 - 27\gamma_3^2$$

has only a finite number of solutions γ_2, γ_3 giving minimal models. This is a known theorem, but for our purposes we can deduce it from the Szpiro conjecture. Indeed, the powers of primes occurring as common factors of γ_2, γ_3 are bounded by the minimality condition. Hence the g.c.d. of γ_2, γ_3 is bounded for all elliptic curves with minimal equation over $\mathbf{Q}$. By the generalized Szpiro conjecture, it follows that $|\gamma_2|$ and $|\gamma_3|$ are also bounded, thus proving the desired implication.

REMARK. Concerning the assertion made that $q_B = q_A^n$ when $B = A/(P)$ and P corresponds to the nth roots of unity, the reader should think first in terms of the complex case. Suppose the lattice of the elliptic curve is $[\tau, 1]$ (i.e. the group generated by $\tau, 1$ over $\mathbf{Z}$), so $A(\mathbf{C}) \approx \mathbf{C}/[\tau, 1]$. Let P correspond to $1/n$. Then

$$A(\mathbf{C})/(P) \approx C/[\tau, 1/n] \approx \mathbf{C}/[n\tau, 1],$$

where the second isomorphism is induced by multiplication $n: \mathbf{C} \to \mathbf{C}$. Exponentiating, we see in the complex case that the "q" corresponding to $A/(P)$ is $e^{2\pi i n\tau}$. Now the reader must accept that essentially the same theory of parametrization holds in the p-adic domain, so the argument works the same way, as given in the above proof.

The same method should prove a more general conjecture as follows.

> *Let F be a number field. There exists a positive number C having the following property. If A is an elliptic curve over F, without complex multiplication, and if there exists a cyclic subgroup of order n, invariant under the Galois group over F, then $n \leqq C$.*

The cyclic subgroup is generated by one point P of order n, and the hypothesis means that the extension $F(P)$ is Galois, with group G such that for $\sigma \in G$ we have

$$\sigma P = \chi(\sigma)P,$$

with some element $\chi(\sigma) \in (\mathbf{Z}/n\mathbf{Z})^*$. Thus χ gives a representation of G as a subgroup of $(\mathbf{Z}/n\mathbf{Z})^*$. If n is a prime number, then in particular, the order of G is prime to n.

To follow the same pattern of proof, one needs to start with semistability. Assume for the moment that the curve is semistable, and let $F = \mathbf{Q}$ for simplicity as before. As before, we then get that a prime dividing Δ_A also divides the denominator of j.

In the first step of the proof, supposing that P corresponds to $q^{1/n}$ under the Tate parametrization, we note that either the polynomial $X^n - q$ is irreducible over $\mathbf{Q}_p$ or it has a root in $\mathbf{Q}_p$, by an elementary criterion of field theory. Therefore, if q is not an nth power in $\mathbf{Q}_p$ then

$$\mathbf{Q}_p(q^{1/n}) = \mathbf{Q}_p(P)$$

has degree n over $\mathbf{Q}_p$. Since the Galois group has order prime to n, this cannot happen and therefore q is an nth power in $\mathbf{Q}_p$. Thus we are in the same situation as before, and exactly the same arguments using the Szpiro conjecture show that n is bounded.

However, under the weaker hypothesis of the Galois invariant cyclic subgroup instead of a rational point of sufficiently high order, it is not clear how to reduce the general case to the semistable case, so at this time, the above arguments apply only to the family of semistable curves.

6. The height on elliptic curves. Let A be an elliptic curve over the rationals again, and let Δ_A be its minimal discriminant. We let $y^2 = x^3 - \gamma_2 x - \gamma_3$ be a minimal equation, with integer coefficients. A fundamental problem is to estimate the absolute values

of x, y as function of γ_2 and γ_3 for integral solutions in $\mathbf{Z}$, and to estimate the height $h(x(P))$ for rational solutions. Since $A(\mathbf{Q})$ is finitely generated, $A(\mathbf{Q})$ modulo its torsion group is a free abelian group, finitely generated. It is a problem to give an upper bound for the heights of free generators for this group. For conjectures see [La2]. The height $h(x(P))$ has a great deal of structure. On the group $A(\mathbf{Q})$ modulo torsion, according to a fundamental theorem of Néron-Tate, there exists a positive definite quadratic form

$$h_A: A(\mathbf{Q})/A(\mathbf{Q})_{\text{tor}} \to \mathbf{R}$$

such that

$$h_A(P) = \tfrac{1}{2}h(x(P)) + O(1).$$

In the next section we shall formulate a fundamental conjecture about this height. Here, we describe a number of basic properties which allow us to compute it, and which will be used to deal with that conjecture.

In the preceding section, we used certain explicit formulas parametrizing the functions $(\gamma_2, \gamma_3, \Delta, j, x, y)$. We now need similar formulas to express the height. It turns out that the height can be given as a sum

$$h_A(P) = \sum_v \lambda_v(P),$$

where λ_v is a function (the **Néron function**) given by an analytic expression on $A(\mathbf{Q}_v)$ for each absolute value v. We shall now describe these functions. As a matter of notation, for any absolute value v, we define

$$v(a) = -\log|a|_v.$$

For instance, if v is p-adic, then $v(p^m) = m\log p$, so $v(a)$ is the order m of a at p, times a normalizing factor $\log p$ which is used to get global formulas putting all the absolute values together. As a approaches 0 we want the p-adic order to approach ∞, and similarly for an absolute value at infinity.

The height h_A is a quadratic form, but the local functions λ_v cannot be quadratic: there has to be some extraneous term appearing in the quadratic relations locally, and only after taking the sum over all v does this term disappear. For elliptic curves, a

neat characterization for the Néron functions was given by Tate as follows.

Let F be a p-adic field, or the complex numbers. Let A be an elliptic curve defined over F. There exists a unique function $\lambda_v : A(F) - \{0\} \to \mathbf{R}$ satisfying the following conditions:

(i) *λ_v is continuous and bounded outside every neighborhood of 0.*

(ii) *Let z be a local uniformizing parameter at 0. Then there exists a bounded continuous function α on an open neighborhood of 0 such that for all P in that neighborhood, $P \neq 0$,* we have

$$\lambda_v(P) = v(z(P)) + \alpha(P).$$

(iii) For all $P, Q \in A(F)$ such that $P, Q, P \pm Q \neq 0$ we have

$$\lambda_v(P+Q) + \lambda_v(P-Q) = 2\lambda_v(P) + 2\lambda_v(Q) + v(x(P) - x(Q)) - \tfrac{1}{6}v(\Delta).$$

Without the last two terms, the third relation would be the relation defining a quadratic function. The final constant involving $v(\Delta)$ is a conveniently normalized term. When applying the relation of (iii) globally, the sum over all v of the last two terms will vanish by the product formula, when $x(P)$, $x(Q)$, are rational numbers, say; so summing over all v yields a quadratic function on the group of rational points. It is not difficult to show that this function is the quadratic height.

We shall now give explicit formulas due to Tate for the Néron functions λ_v in order to be able to estimate the height from below later. In each case, it is not difficult to verify that the formulas we give satisfy the desired quadratic relation. The other conditions (i) and (ii) simply express that the Néron function has a logarithmic singularity at the origin, and these conditions are also immediately verified in each case. We shall omit the verification.

$v = v_\infty$.

In dealing with torsion points on elliptic curves, we have already remarked that we have a complex analytic isomorphism

$$\mathbf{C}/[\tau, 1] \to A(\mathbf{C})$$

where τ lies in the upper half plane, and $[\tau, 1]$ is the lattice

generated by τ, 1 over $\mathbf{Z}$. We can change τ by any element of $\mathrm{SL}_2(\mathbf{Z})$, and in particular, we can take τ to be in the standard fundamental domain for $\mathrm{SL}_2(\mathbf{Z})$. If we let

$$q = q_\tau = e^{2\pi i\tau},$$

then

$$(*) \qquad \operatorname{Im}\tau \geqq \tfrac{1}{2}\sqrt{3} \quad \text{and so} \quad |q_\tau| \leqq e^{-\pi\sqrt{3}}.$$

We let $u = u_1\tau + u_2 \in \mathbf{C}$ $(u_1, u_2 \in \mathbf{R})$, and we let

$$t = q_u = e^{2\pi i u}$$

be a variable in $\mathbf{C}^*$. We let

$$g_0(t) = g_0(q, t) = (1-t)\prod_{n=1}^{\infty}(1-q^n t)(1-q^n/t).$$

Then we have a functional equation for g_0, namely

$$g_0(qt) = -t^{-1}g_0(t) = g_0(t^{-1}).$$

Let $\mathbf{B}_2$ be the second Bernoulli polynomial,

$$\mathbf{B}_2(T) = T^2 - T + \tfrac{1}{6}.$$

We define the **Néron function**

$$\lambda_v(u, \tau) = \tfrac{1}{2}\mathbf{B}_2(u_1)v(q) + v(g_0(q_u)) = \tfrac{1}{2}\mathbf{B}_2(u_1)v(q) + v(g_0(t)).$$

Immediately from the functional equation, we see that $\lambda_v(u, \tau)$ is even in u, that is

$$\lambda_v(-u, \tau) = \lambda_v(u, \tau).$$

The term involving $\mathbf{B}_2$ serves the purpose of making the function as we have defined it periodic, with periods 1, τ. This can be verified directly and simply from the functional equation for g_0. The function λ_v is real analytic, except for a logarithmic singularity at the lattice points, as one sees directly from its definition.

Proposition 6.1. *For $v = v_\infty$ there exists a constant $C_\infty > 0$ having the following property. Let $\operatorname{Im}\tau > \sqrt{3}/2$ and let $|u_1| \leqq 1/6$. Then*

$$\lambda_v(u, \tau) \geqq -C_\infty.$$

Proof. By the periodicity and the fact that $\lambda_v(u, \tau)$ is even in u, we can normalize a representative for a point by the condition

$$(**) \quad 0 \leqq u_1 \leqq \tfrac{1}{2} \text{ whence } |q^{1/2}| \leqq |q_u| \leqq 1 \quad \text{and} \quad |q^{1/2}/q_u| \leqq 1.$$

In that case, we conclude that the Néron function has the value

$$\lambda_v(u, \tau) = -\tfrac{1}{2}\mathbf{B}_2(u_1)\log|q| - \log|1 - q_u| - O(1),$$

and it is easy to compute an explicit value for $O(1)$, independent of u and τ. Namely, for $n \geqq 1$, we have estimates

$$|q^n q_u| \leqq e^{-n\pi\sqrt{3}/2}$$

and

$$|q^n/q_u| = |q^{n-1/2}q^{1/2}/q_u| \leqq |q^{n-1/2}| \leqq e^{-(n-1/2)\pi\sqrt{3}/2}.$$

These inequalities show that the term $-O(1)$ is independent of u and τ. In addition, the choice of $|u_1| \leqq 1/6$ was made so that $\mathbf{B}_2(u_1) \geqq 0$, and hence the term $-\mathbf{B}_2(u_1)\log|q|$ is actually $\geqq 0$ since $|q| \leqq 1$. Finally $|q_u| \leqq 1$, so

$$\log|1 - q_u| \leqq \log 2.$$

The uniform lower bound of Proposition 6.1 for $\lambda_v(u, \tau)$ follows at once from these estimates.

We have an analytic isomorphism

$$\mathbf{C}/[\tau, 1] \xrightarrow{\approx} A(\mathbf{C}) \text{ by } u \mapsto P_u.$$

We also write $u = u(P)$. The function $\lambda_v(u, \tau)$ being periodic in u, we use it to define the v-component of the height, namely for a rational point P corresponding to u we let

$$\lambda_v(P) = \lambda_v(u(P), \tau).$$

Proposition 6.1 can be restated in part by saying that if P is sufficiently close to the origin, then $\lambda_v(P)$ has a uniform lower bound as in Proposition 6.1.

We shall apply this to multiples of an arbitrary point P which lie close to the origin, thus giving us a lower bound for the component of the height at infinity.

Proposition 6.2. *Let C_∞ be the constant of Proposition 6.1. Let A be an elliptic curve over $\mathbf{Q}$, and let $P \in A(\mathbf{Q})$. Given an integer $M \geqq 1$, there exists an integer b satisfying*

$$1 \leqq b \leqq 6M$$

such that for $v = v_\infty$ we have

$$\lambda_v(mbP) \geqq -C_\infty \quad \text{for } 1 \leqq m \leqq M.$$

Proof. We get a homomorphism $A(\mathbf{Q}) \to \mathbf{R}/\mathbf{Z}$ by mapping

$$Q \mapsto u_1(Q) \mod \mathbf{Z}.$$

Let n be a suitably large integer. Partition $\mathbf{R}/\mathbf{Z}$ into n small intervals of length $1/n$. The multiples of the point P given by

$$0, P, 2P, \ldots, nP$$

map to $n+1$ elements of $\mathbf{R}/\mathbf{Z}$. Hence there exist integers $0 \leqq n_1 < n_2 \leqq n$ such that $n_1 P$ and $n_2 P$ lie in the same small interval. Let $b = n_2 - n_1$. Then $0 < b \leqq n$ and bP lies in the small interval containing the origin. Then

$$|u_1(bP)| \leqq \frac{1}{n}$$

and therefore

$$|u_1(mbP)| \leqq \frac{M}{n} \quad \text{for } 1 \leqq m \leqq M.$$

Thus if we let $n = 6M$ we get $|u_1(mbP)| \leqq 1/6$, and we can apply Proposition 6.1 to conclude the proof.

Next we deal with the absolute values associated with prime numbers, and we have to distinguish cases, depending on whether j is p-integral or not.

$v = v_p$ for p prime, and j is p-integral.

Given a prime p, one may reduce the minimal equation of an elliptic curve mod p. Then the reduced equation defines a possibly reducible curve, called the fiber. The group law on the original elliptic curve reduces to a group law on the set of nonsingular points on the fiber.

Proposition 6.3. *Let $v = v_p$. Then for all points P whose reduction mod p is nonsingular on the fiber, the Néron function is defined by*

$$\lambda_v(P) = \frac{1}{2}\max\{0, \log|x(P)|_v\} + \frac{1}{12}v(\Delta) \geqq \frac{1}{12}v(\Delta).$$

This proposition is proved essentially by brute force, cf. for instance Theorem 4.4 and Theorem 6.1 of [La3, Chapter III] for the standard Weierstrass equation.

The following result is crucial to know how far a point is from having singular reduction on the fiber.

Proposition 6.4. *Assume that we are dealing with a minimal model, and that j is p-integral. Then*

(i) *For every point $P \in A(F)$ the point $12P$ has nonsingular reduction on the fiber.*

(ii) *Furthermore, $\lambda_v(12mP) \geqq 1/12v(\Delta)$ for all positive integers m.*

Proof. The proof of the first statement is by computation, using the Néron-Kodaira classification of all possible cases of degeneracy which can occur. See Néron's last chapter on elliptic curves ([Ne], the final table), and for more recent insights, Tate's algorithm [Ta2]. It is a very tedious matter. To replace such computations by theoretical arguments takes very heavy machinery. The second statement follows by applying Theorem 6.3.

For our purposes the factor 12 is not important, all that matters is that there is some universal integer which can be used to multiply a point and land it into the nonsingular part of the fiber.

$v = v_p$ with p prime, and j is not p-integral.

Finally we must give a description of the Néron function for v_p when p divides the denominator of j, i.e. j is not p-integral. This case is a p-adic analogue of the case over the complex numbers, with essentially the same formulas. We deal here with the Tate curve as in §4.

Let $\mathbf{C}_p$ be the completion of the algebraic closure of $\mathbf{Q}_p$ and let t be a variable in $\mathbf{C}_p^*$. Define

$$g_0(t) = (1-t)\prod_{n=1}^{\infty}(1-q^n t)(1-q^n/t).$$

The functional equation was formal, and so we have again

$$g_0(qt) = g_0(t^{-1}) = -t^{-1}g_0(t),$$

valid for all $t \in \mathbf{C}_p^*$. Define

$$u(t) = \frac{v(t)}{v(q)} = \frac{\operatorname{ord}_p t}{\operatorname{ord}_p q}.$$

Proposition 6.5. *Let p divide the denominator of j. Let P_t be the image of t in $A(\mathbf{C}_p)$ under the Tate parametrization. Then*

$$\lambda_v(P_t) = v(g_0(t)) + \tfrac{1}{2}\mathbf{B}_2(u(t))v(q).$$

Thus we see the analogy with the complex case. From the functional equation of g_0 we conclude again that $\lambda_v(P_t)$ is periodic with period q, and so is defined on the elliptic curve.

Now let F be a finite extension of $\mathbf{Q}_p$ in $\mathbf{C}_p$, and suppose $q \in F^*$. For every $t \in F^*$ we can find a representative mod $q^{\mathbf{Z}}$ which we denote by t_P, and which is uniquely determined by the condition

$$0 \leqq u(t) < 1 \text{ or equivalently } |q|_v < |t|_v \leqq 1.$$

For such a representative we see from the formula for g_0 that $v(g_0(t_P)) \geqq 0$. It is therefore convenient to define the periodic function

$$\mathbf{B}\colon \mathbf{R}/\mathbf{Z} \to \mathbf{R} \quad \text{by} \quad \mathbf{B}(u) = \mathbf{B}_2(u) \text{ if } 0 \leqq u \leqq 1,$$

and $\mathbf{B}$ is extended by periodicity to all of $\mathbf{R}$. One can also write

$$\mathbf{B}(u) = \{u\}^2 - \{u\} + \tfrac{1}{6}$$

where $\{u\}$ is the fractional part of u, and we obtain:

Proposition 6.6. *Let p divide the denominator of j, and $v = v_p$. Then*

$$\lambda_v(P) \geqq \tfrac{1}{2}\mathbf{B}(u(t_P))v(q).$$

Just as Proposition 6.4 gave us a simple criterion to get a nice formula for the height of points whose reduction is nonsingular on the fiber, we can formulate a similar criterion which we can apply to Proposition 6.6. Indeed, if $u(t_P) = 0$ then we get the simple inequality

$$\lambda_v(P) \geqq \tfrac{1}{12}v(q) = \tfrac{1}{12}v(\Delta)$$

just as in Proposition 6.3. Suppose however that P is an arbitrary point in $A(\mathbf{Q}_p)$. Let

$$b = \operatorname{ord}_p(q) = \operatorname{ord}_p(\Delta).$$

Then from the definitions, $t_P^b \in q^{\mathbf{Z}}$, and therefore t_{bP} is a p-adic unit, and so $u(t_{bP}) = 0$. Combining the two cases of Proposition 6.4 and 6.6, we find:

Proposition 6.7. *Given a positive integer n_0 there exists an integer $b > 0$ having the following property. For all elliptic curves A over $\mathbf{Q}$, and non-torsion rational point $P \in A(\mathbf{Q})$, if $v = v_p$ is such*

that j is p-integral or $\operatorname{ord}_p(\Delta) \leqq n_0$ *then*

$$\lambda_v(bP) \geqq \tfrac{1}{12} v(\Delta).$$

7. The Szpiro conjecture implies the minimal height conjecture. Let A be an elliptic curve defined over $\mathbf{Q}$, and let h_A be the Néron-Tate height. In [La1] I conjectured that this height satisfies a minimum condition as follows. We let Δ_A be the minimal discriminant.

There exists constants $C_{\mathbf{Q}}$ and $C'_{\mathbf{Q}} > 0$ such that for all elliptic curves A over $\mathbf{Q}$ and a non-torsion point $P \in A(\mathbf{Q})$, we have

$$h_A(P) \geqq C_Q \log|\Delta_A| - C'_{\mathbf{Q}}.$$

Note that given an integer $\Delta \neq 0$ there is only a finite number of elliptic curves A over $\mathbf{Q}$ whose minimal discriminant is Δ. Consequently we really did not need to mention the constant $C'_{\mathbf{Q}}$ in the stated inequality, because for $\log|\Delta_A|$ sufficiently large, the right-hand side is $\geqq C''_{\mathbf{Q}} \log|\Delta_A|$ for some $C''_{\mathbf{Q}} > 0$, and by picking $\log|\Delta_A|$ sufficiently large, we are omitting only a finite number of values of $h_A(P)$. Hence by shrinking the constant $C_{\mathbf{Q}}$ suitably, we obtain the stated inequality without a $C'_{\mathbf{Q}}$.

On the other hand in [La2] I also conjectured an upper bound for the height of suitable free generators of $A(\mathbf{Q})/A(\mathbf{Q})_{\text{tor}}$, and I showed that the two conjectures are not independent: the lower bound conjecture is used to motivate the upper bound conjecture, and to carry out certain steps in trying to prove it. The essential point is as follows. Let $\langle\ ,\ \rangle$ be the symmetric bilinear form giving rise to the quadratic form h_A. Then from L-series considerations, conjecturally one gets a bound for the determinant $\det\langle P_i, P_j\rangle$ of a basis for $A(\mathbf{Q})$ mod torsion. It is possible to construct a basis $\{P_1, \ldots, P_r\}$ which is "almost" orthogonalized. "Almost" is because we are over $\mathbf{Z}$, not over $\mathbf{R}$. For a precise definition, see [La2] or [La3]. We order the points by ascending height, so

$$h_A(P_1) \leqq \cdots \leqq h_A(P_r).$$

Then we get a conjectural upper bound for the product

$$h_A(P_1) \cdots h_A(P_r).$$

In order to get an upper bound for $h_A(P_r)$ itself, we can then divide by the first $r-1$ terms, and this is where one needs a lower bound for $h_A(P_1)$.

It occurred to Marc Hindry and to me independently that the Szpiro conjecture should imply this minimum, and the implication was proved by Hindry-Silverman [H-S]. I shall describe their proof in this section.

In the first place, Silverman [S] had proved my conjecture in the case of integral j-invariant, and so the problem was how to expand his arguments so that they could apply to the non-integral case.

As before, we let s be the Szpiro exponent, so that we have

$$|\Delta| \leqq N_0(\Delta)^s$$

except for a finite number of elliptic curves, which we disregard. We always take the elliptic curve in minimal form, so Δ is the minimal discriminant.

For the subsequent proof, it will be convenient to decompose the minimal discriminant into a product. We let

$$\Delta = \Delta_1 \Delta_2$$

where:

Δ_1 is the product of the prime powers which occur with an exponent $< 2s$

Δ_2 is the product of those prime powers which occur with an exponent $\geqq 2s$.

Then

$$(*) \qquad \log|\Delta_1| \geqq \frac{1}{2s} \log|\Delta| .$$

This follows immediately from the inequality $N_0(\Delta_2) \leqq \Delta_2^{1/2s}$, the inequality

$$|\Delta| \leqq N_0(\Delta)^s \leqq |\Delta_1|^s N_0(\Delta_2)^s \leqq |\Delta_1|^s |\Delta_2|^{1/2},$$

and by substituting $\Delta_2 = \Delta/\Delta_1$ on the right-hand side.

The primes dividing Δ_2 are the ones which caused trouble in extending Silverman's proof. To cancel their negative contribution to the height, due to a term with the Bernoulli function, Hindry-Silverman use an averaging process, based on a simple inequality of analysis, which we extract here. As in the previous section, we let $\mathbf{B}$ be the periodic function obtain from the Bernoulli polynomial $\mathbf{B}_2$.

Lemma 7.1. *For all* $u \in \mathbf{R}/\mathbf{Z}$ *and all positive integers* M *we have*

$$\sum_{m=1}^{M} \left(1 - \frac{m}{M+1}\right) \mathbf{B}(mu) \geqq -\frac{1}{12}.$$

Proof. We have the Fourier expansion

$$\mathbf{B}(u) = \frac{1}{2\pi^2} \sum_{n \neq 0} \frac{1}{n^2} e^{2\pi i n u}.$$

But also

$$\frac{1}{M+1} \left(\sum_{n=1}^{M+1} z^n\right) \left(\sum_{k=1}^{M+1} z^{-k}\right) = \sum_{m=1}^{M+1} \left(1 - \frac{m}{M+1}\right) (z^m + z^{-m}) + 1.$$

Therefore

$$\begin{aligned}
&\sum_{m=1}^{M} \left(1 - \frac{m}{M+1}\right) \mathbf{B}(mu) \\
&\quad = \frac{1}{2\pi^2} \sum_{m=1}^{M} \sum_{n=1}^{\infty} \left(1 - \frac{m}{M+1}\right) \frac{1}{n^2} (e^{2\pi i n m u} + e^{-2\pi i n m u}) \\
&\quad = \frac{1}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \left(\frac{1}{M+1} \left| \sum_{m=1}^{M+1} e^{2\pi i m n u} \right|^2 - 1 \right) \\
&\quad \geqq -\frac{1}{12} \text{ because } \zeta(2) = \pi^2/6
\end{aligned}$$

as desired.

Hindry and Silverman place the lemma in the context of an inequality for Fourier transforms by Blanksby-Montgomery [B-M]. The lemma and its proof also fall under the pattern used by Elkies in estimating Green's functions on Riemannian manifolds, cf. [La4], Chapter VI, Theorem 6.1.

All that will be used of Lemma 7.1 is that a linear combination of $\mathbf{B}(mu)$ with suitable positive coefficients is uniformly bounded from below, and the sum of the coefficients tends to infinity. The specific form of the coefficients is not important for the applications we shall make.

We now give the proof of the minimal height conjecture assuming the Szpiro conjecture. First, by taking $n_0 = 2s$ (for instance), we can apply Theorem 6.7. Since for all points $P \in A(\mathbf{Q})$ we have

$$h_A(bP) = b^2 h_A(P),$$

it suffices to prove the lower bound for the height of points P which satisfy the inequality

$$\lambda_v(P) \geqq \frac{1}{12} v(\Delta_1)$$

for all $v = v_p$ such that $p|\Delta_1$. From now on, we assume that the point P has this property. We pick M sufficiently large as a function of s. For instance, $M = 4s$ suffices, as will become apparent from the following arguments. We let

$$c_m = 1 - \frac{m}{M+1} \text{ whence } \sum_{m=1}^{M} c_m = \frac{M}{2} \quad \text{and} \quad C_M = \sum_{m=1}^{M} c_m m^2 .$$

We select b as in Proposition 6.2. Then

$$\begin{aligned} C_M b^2 h_A(P) &= \sum_{m=1}^{M} c_m m^2 b^2 h_A(P) \\ &= \sum_{m=1}^{M} c_m h_A(mbP) = \sum_{v} \sum_{m=1}^{M} c_m \lambda_v(mbP) . \end{aligned}$$

We shall give a lower bound for the v-contribution $\sum c_m \lambda_v(mbP)$ for each v. We partition the absolute values v into four sets:

$v = v_\infty$;
$v = v_p$ with $p \nmid \Delta$;
$v = v_p$ with $p|\Delta_1$;
$v = v_p$ with $p|\Delta_2$.

We note that for every non-zero integer d we have trivially

$$\sum_{p|d} v_p(d) = \log|d| .$$

If $v = v_\infty$ then using Proposition 6.2 and the b of that proposition, we conclude that the v-contribution is $\geqq -MC_\infty$. If $v = v_p$ with $p \nmid \Delta$, then the v-contribution is $\geqq 0$ by Proposition 6.3. Therefore by Proposition 6.6 and Lemma 7.1 for $p|\Delta_2$, and Proposition 6.7 for $p|\Delta_1$, we get

$$\begin{aligned} C_M b^2 h_A(P) &\geqq -MC_\infty + \frac{1}{12}\frac{M}{2}\log|\Delta_1| - \frac{1}{24}\log|\Delta_2| \\ &\geqq -MC_\infty + \frac{1}{48s} M \log|\Delta| - \frac{1}{24}\log|\Delta| \end{aligned}$$

using $(*)$ and the trivial fact that $|\Delta_2| \leqq |\Delta|$. We may now pick for instance $M = 4s$. Then the right-hand side is

$$\geqq -4sC_\infty + \frac{1}{24}\log|\Delta| ,$$

which concludes the proof.

REFERENCES

[Ar] M. ARTIN, *Néron models*, Arithmetic Geometry (G. Cornell and J. Silverman, eds.), Springer-Verlag, Berlin and New York, 1986, pp. 213–230.

[BCHS] B. BIRCH, S. CHOWLA, M. HALL, AND A. SCHINZEL, *On the difference* $x^3 - y^2$, Norske Vid. Selsk. Forrh., **38** (1965), pp. 65–69.

[B-M] P. BLANKSBY AND H. MONTGOMERY, *Algebraic integers near the unit circle*, Acta Arith., **18** (1971), pp. 355–369.

[BLSTW] J. BRILLHART, D. H. LEHMER, J. L. SELFRIDGE, B. TUCKERMAN AND S. S. WAGSTAFF, JR., *Factorization of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11$ *up to high powers*, Contemporary Mathematics Vol. 22, AMS, 1983.

[Dan] L. V. DANILOV, *The diophantine equation* $x^3 - y^2 = k$ *and Hall's conjecture*, Mat. Zametki., **32** No. 3 (1982), pp. 273–275.

[Dav] H. DAVENPORT, *On* $f^3(t) - g^2(t)$, K. Norske Vid. Selsk. Forrh. (Trondheim), **38** (1965), pp. 86–87.

[Fr1] G. FREY, *Links between stable elliptic curves and certain diophantine equations*, Annales Universitatis Saraviensis, Series Mathematicae, **1** (1986), pp. 1–40.

[Fr2] ———, *Links between elliptic curves and solutions of* $A - B = C$, J. Indian Math. Soc., **51** (1987), pp. 117–145.

[Ha] M. HALL, *The diophantine equation* $x^3 - y^2 = k$, Computers in Number Theory (A. O. L. Atkin and B. J. Birch, eds.), Academic Press, London, 1971, pp. 173–198.

[H-S] M. HINDRY AND J. SILVERMAN, *The canonical height and integral points on elliptic curves*, Invent. Math., **93** (1988), pp. 419–450.

[Ku1] D. KUBERT, *Universal bounds on the torsion of elliptic curves*, Proc. London Math. Soc., vol. XXXIII (1976), pp. 193–237.

[Ku2] ———, *Universal bounds on the torsion of elliptic curves*, Compositio Math., **38** (1979), pp. 121–128.

[La1] S. LANG, *Elliptic functions*, Addison-Wesley, 1973, reprinted by Springer-Verlag, Berlin and New York, 1987.

[La2] ———, *Conjectured diophantine estimates on elliptic curves*, Arithmetic and Geometry, Volume dedicated to Shafarevich, Vol. I, edited by M. Artin and J. Tate, Birkhauser, 1983, pp. 155-171.

[La3] ———, *Elliptic Curves: Diophantine Analysis*, Springer-Verlag, Berlin and New York, 1978, pp. 212-213.

[La4] ———, *Introduction to Arakelov Theory*, Springer-Verlag, Berlin and New York, 1988.

[L-O] H. LENSTRA AND F. OORT, *Abelian varieties having purely additive reduction*, J. Pure and Applied Algebra, **36** (1985), pp. 281–198.

[Ma1] R. C. MASON, *Equations over function fields*, Springer Lecture Notes **1068** (1984, 149–157), in Number Theory, proceedings of the Noordwijkerhout, 1983.

[Ma2] ———, *Diophantine equations over function fields*, London Math. Soc. Lecture Note Series, vol. 96, Cambridge University Press, United Kingdom, 1984.

[Ma3] ———, *The hyperelliptic equation over function fields*, Math. Proc. Cambridge Philos. Soc., **93** (1983), pp. 219–230.

[Maz] B. MAZUR, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1978).

[Ne] A. NÉRON, *Modèles minimaux des variétés abéliènnes sur les corps locaux et globaux*, Inst. Hautes Études Sci. Publ. Math., **21** (1964), pp. 361–482.

[Ri] K. RIBET, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms* (to appear).

[Se] J. P. SERRE, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J., **54** (1987), pp. 179–230.

[Si] J. SILVERMAN, *Lower bound for the canonical height on elliptic curves*, Duke Math. J., **48** (1981), pp. 633–648.

[Ta1] J. TATE, *The arithmetic of elliptic curves*, Invent. Math., **23** (1974), pp. 179–206.

[Ta2] ———, *Algorithm for determining the type of a singular fiber in an elliptic pencil*, Modular Functions in One Variable IV, Lecture Notes in Math., vol. 476, Springer-Verlag, Berlin and New York, (Antwerp Conference).

[Ve] P. VOJTA, *Diophantine approximations and value distribution theory*, Lecture Notes in Math., vol. 1239, Springer-Verlag, Berlin and New York, 1987.

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CONNECTICUT 06520