Contemporary Mathematics

The Thue-Siegel method in diophantine geometry

Paul Vojta

Abstract. This mini-course described the Thue-Siegel method, as used in the proof of Faltings' theorem on the Mordell conjecture. The exposition followed Bombieri's variant of this proof, which avoids the machinery of Arakelov theory.

In the 1950s, K. F. Roth [**Rot55**] proved a much-anticipated theorem on diophantine approximation, building on work of Thue, Siegel, Gel'fond, Dyson, and others.

Theorem 0.1. *Given an algebraic number $\alpha \in \overline{\mathbb{Q}}$, a number $C \in \mathbb{R}$, and $\epsilon > 0$, there are only finitely many $p/q \in \mathbb{Q}$ (with $p, q \in \mathbb{Z}$ and $\gcd(p,q) = 1$) that satisfy*

$$\left| \frac{p}{q} - \alpha \right| \le \frac{C}{|q|^{2+\epsilon}} \,. \tag{0.1.1}$$

This theorem was proved using the *Thue-Siegel* method (which was described in M. Nakamaye's course).

The goal for this course is to briefly describe how the Thue-Siegel method was adapted by Vojta [**Voj91**], Faltings [**Fal91**], and Bombieri [**Bom90, Bom91**] to give a proof of the Mordell conjecture (which had already been proved by Faltings):

Theorem 0.2. [**Fal83**] *(Mordell's conjecture) Let k be a number field and let C be a smooth projective curve over k of genus > 1. Then $C(k)$ is finite.*

Faltings' original 1983 proof used results from the theory of moduli spaces of abelian varieties to prove a conjecture of Shafarevich on principally polarized abelian varieties of given dimension with good reduction outside of a fixed finite set of places of a number field. He then used "Parshin's trick" to obtain the Mordell conjecture. In short, this was very different from the Thue-Siegel method.

Analogies with Nevanlinna theory suggested that there should be a common proof of both Roth's theorem and the Mordell conjecture, and this led the author to try to prove the Mordell conjecture using the Thue-Siegel method. This led ultimately to the paper [**Voj91**], which gave another proof using the Thue-Siegel method, but which relied heavily on the Arakelov theory as developed by Gillet and Soulé. In particular, the use of Siegel's lemma was replaced by an argument using an adaptation of the Riemann-Roch-Hirzebruch-Grothendieck theorem to Arakelov theory, due to Gillet and Soulé.

Shortly thereafter, Faltings [**Fal91**] managed to eliminate the use of the Gillet-Soulé Riemann-Roch theorem, replacing it with arguments on the Jacobian of the curve C. This allowed him to make use of more than two rational points (as did

2010 *Mathematics Subject Classification.* Primary 11J68; Secondary 11J97, 11G30.

Roth), as well as extend his result to prove a special case of a conjecture of Serge Lang on rational points on closed subvarieties of abelian varieties not containing any nontrivial translated abelian subvarieties. (He later succeeded in removing the latter hypothesis [**Fal94**].)

Following that, Bombieri [**Bom90**] simplified the proof further, by removing Arakelov theory altogether, at the (minor) cost of some clever geometrical arguments. He was able to use the fact that Arakelov intersection numbers can be realized more classically as heights. Then, he was able to avoid the use of the Riemann-Roch theorem of Gillet and Soulé by applying a more classical theorem of Riemann-Roch type (due to Hirzebruch), together with some clever geometrical manipulations. In the end, he obtained a proof relying only on elementary facts from algebraic geometry and the theory of heights.

This course describes Bombieri's proof (but will omit some details, due to lack of time). Another description of his proof appears in [**HS00**, Part E].

For those interested in the earlier proof [**Voj91**], a good place to start would be an earlier paper [**Voj89**], which proved the function field variant of Mordell's conjecture (first proved by Manin [**Man63**]). This paper also used the Thue-Siegel method, but used classical intersection theory instead of Arakelov theory, and the Riemann-Roch-Grothendieck theorem in place of the Riemann-Roch theorem in Arakelov theory due to Gillet and Soulé.

This paper is organized as follows. Section 1 gives an overview of the Thue-Siegel method, as used to prove Roth's theorem and its predecessors. Section 2 defines the basic geometrical objects to be used in Bombieri's proof. Section 3 gives some basic information on curves and their Jacobians, as used by a result of Mumford [**Mum65**], and which forms the core of Bombieri's proof (and its predecessors by Vojta and Faltings). Section 4 then gives an upper bound for the height of a carefully chosen point (P_1, P_2) on $C \times C$, relative to a line bundle on $C \times C$ that is chosen based on properties of P_1 and P_2. Section 5 gives a lower bound for the same height, depending on the height of a global section of the line bundle. This global section plays a role comparable to that of the auxiliary polynomial in the classical Thue-Siegel method. Section 6 constructs such a global section with a bound on its height. In Section 7, the *index* of a local section of a line bundle is defined; this generalizes the index used in the classical Thue-Siegel method, and is a weighted order of vanishing. Finally, Section 8 concludes the proof by comparing the two bounds for the height derived in Sections 4 and 5, and deriving a contradiction to the assumption that C has infinitely many rational points.

In this paper, $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\mathbb{Z}_{>0} = \{1, 2, 3, \dots\}$. We will fix throughout a number field k, and work in the category of schemes (and morphisms) over k. A **variety** is an integral separated scheme of finite type over k, and a **curve** is a variety of dimension 1. When working with a product $X \times Y$, the associated projection morphisms are denoted $p_1 \colon X \times Y \to X$ and $p_2 \colon X \times Y \to Y$. On $\mathbb{P}^r \times \mathbb{P}^s$, $\mathscr{O}(n, m) = p_1^*\mathscr{O}(n) \otimes p_2^*\mathscr{O}(m)$. A **line sheaf** is an invertible sheaf. If D_1 and D_2 are divisors, then $D_1 \sim D_2$ denotes linear equivalence. If D is a divisor on a variety X, the notations $H^i(X, \mathscr{O}(D))$ and $h^i(X, \mathscr{O}(D))$ are shortened to $H^i(X, D)$ and $h^i(X, D)$, respectively. Finally, the function field of a variety X is denoted $K(X)$.

For the number field k, the set M_k is the set of all places of k; this is the disjoint union of the sets of real places, complex places, and non-archimedean places of k. The real and complex places are in canonical one-to-one correspondence with the set of injections $k \hookrightarrow \mathbb{R}$ and the set of unordered pairs $(\sigma, \bar\sigma)$ of injections $k \hookrightarrow \mathbb{C}$ with image not contained in $\mathbb{R}$, respectively. The set of non-archimedean places is in canonical one-to-one correspondence with the set of nonzero prime ideals in the ring $\mathscr{O}_k$ of integers of k ($\mathscr{O}_k$ is the integral closure of $\mathbb{Z}$ in k). For each place $v \in M_k$ we define a **norm** $\|\cdot\|_v$, as follows:

$$\|x\|_v = \begin{cases} |\sigma(x)| & \text{if } v \text{ is real and corresponds to } \sigma\colon k \hookrightarrow \mathbb{R}; \\ |\sigma(x)|^2 & \text{if } v \text{ is complex and corresponds to } (\sigma,\bar\sigma); \\ (\mathscr{O}_k : \mathfrak{p})^{-\operatorname{ord}_{\mathfrak{p}}(x)} & \text{if } v \text{ is non-archimedean and corresponds to } \mathfrak{p} \subseteq \mathscr{O}_k\,. \end{cases}$$

(In the non-archimedean case, the formula assumes $x \neq 0$; of course $\|0\|_v = 0$ for all v.) Note that these are called norms, not absolute values, because $\|\cdot\|_v$ does not obey the triangle inequality when v is complex.

We have the **product formula**

$$\prod_{v \in M_k} \|x\|_v = 1 \qquad \text{for all } x \in k^*\,. \tag{0.3}$$

Also, let

$$N_v = \begin{cases} 1 & \text{if } v \text{ is real,} \\ 2 & \text{if } v \text{ is complex, and} \\ 0 & \text{if } v \text{ is non-archimedean.} \end{cases}$$

Then

$$\sum_{v \in M_k} N_v = [k : \mathbb{Q}] \tag{0.4}$$

and

$$\|a_1 + \dots + a_n\|_v \le n^{N_v} \max\{\|a_1\|_v, \dots, \|a_n\|_v\} \tag{0.5}$$

for all $n \in \mathbb{Z}_{>0}$ and all $a_1, \dots, a_n \in k$. Finally, heights are always taken to be logarithmic and absolute. For example, the height of a point $P \in \mathbb{P}^n(k)$ with homogeneous coordinates $[x_0 : x_1 : \dots : x_n]$ is

$$h(P) = \frac{1}{[k : \mathbb{Q}]} \sum_{v \in M_k} \log \max\{\|x_0\|_v, \dots, \|x_n\|_v\}\,. \tag{0.6}$$

For more information on the basic properties of heights (which we shall assume), see [**HS00**, Part B] or [**Lan83**, Ch. 4].

1. The Thue-Siegel Method

This section briefly describes the Thue-Siegel method, which was originally used to prove Roth's theorem (and its predecessors), and later was extended to consider hyperplanes in $\mathbb{P}^n$ by W. M. Schmidt [**Sch72**, Lemma 7] (see also [**Sch80**, Ch. VI, Thm. 1F]), as well as to prove Mordell's conjecture as described in the Introduction to this paper.

The Thue-Siegel method is as described in Nakamaye's course, starting with Thue's seminal paper [**Thu09**]. Basically, it involved constructing an *auxiliary* *polynomial* in one variable [**Lio44**], two variables (Thue, Siegel [**Sie21**], Gel'fond [**GL48**] and [**Gel52**], Dyson [**Dys47**]), or many variables [**Rot55**].

In a nutshell, the Thue-Siegel method is as follows.

Step −1: Assume that there are infinitely many $p/q \in \mathbb{Q}$ for which (0.1.1) is false.

Step 0: Choose two (or n) exceptions p_1/q_1, p_2/q_2 to (0.1.1) with certain properties.

Step 1: Using information about those exceptions, construct a polynomial f in two (or n) variables with certain properties.

Steps 2, 3: Show that the above choices, considerations, etc. imply that

$$f\left(\frac{p_1}{q_1}, \frac{p_2}{q_2}\right) = 0 ,$$

and the same for certain partial derivatives of f.

Step 4: Derive a contradiction.

In trying to adapt this to give a proof of the Mordell conjecture, some obstacles arise:

(1) What should play the role of α?
(2) How do we work with polynomials, when we're trying to prove something about points on a (non-rational) *curve*?

Let's answer the second question first.

One can view a polynomial in x of degree d as a global section of a line sheaf $\mathscr{O}(d)$ on $\mathbb{P}^1$; for example, $ax^2 + bx + c$ corresponds to

$$ax_1^2 + bx_0x_1 + cx_0^2 \in \Gamma(\mathbb{P}^1, \mathscr{O}(2)) .$$

Likewise, $f \in \mathbb{Z}[x, y]$ of degree d_1 in x and d_2 in y corresponds to a global section of the line sheaf $\mathscr{O}(d_1, d_2) := p_1^*\mathscr{O}(d_1) \otimes p_2^*\mathscr{O}(d_2)$ on $\mathbb{P}^1 \times \mathbb{P}^1$, where for $i = 1, 2$, $p_i \colon \mathbb{P}^1 \times \mathbb{P}^1 \to \mathbb{P}^1$ is the projection to the i^{th} factor. The line sheaf $\mathscr{O}(d_1, d_2)$ corresponds to the divisor $d_1([\infty] \times \mathbb{P}^1) + d_2(\mathbb{P}^1 \times [\infty])$.

Having described that background, to answer the question let C be a smooth projective curve over k of genus $g > 1$, and fix $P \in C(k)$. (If $C(k) = \emptyset$, then there is nothing to prove.) The first thing to try is to look at sections of the line sheaf $\mathscr{O}(d_1([P] \times C) + d_2(C \times [P]))$. As it turns out, though, this does not work.

However, the product $C \times C$ has many other line sheaves. One possibility is to note that if $Q \in C(k)$ and $Q \ne P$, then $\mathscr{O}(Q) \not\cong \mathscr{O}(P)$ (otherwise $C \cong \mathbb{P}^1$; see [**Har77**, II Example 6.10.1]). Therefore, one can create other line sheaves by allowing P to vary. This does not work, either.

Instead, what actually does work is to look at the diagonal Δ on $C \times C$. We have $\deg\left(\mathscr{O}(\Delta)\big|_\Delta\right) = 2 - 2g$ (so $\Delta^2 = 2 - 2g$). Therefore $\mathscr{O}(\Delta)$ is not isomorphic to $p_1^*\mathscr{O}(D_1) \otimes p_2^*\mathscr{O}(D_2)$ for any divisors D_1 and D_2 on C. This is because, for a closed rational point $P \in C$, we have the intersection numbers

$$\begin{aligned} \Delta \,.\, (\{P\} \times C) = \Delta \,.\, (C \times \{P\}) = 1 \,, \\ \left(p_1^*D_1 + p_2^*D_2\right) . \,(\{P\} \times C) = \deg D_2 \,, \qquad \text{and} \\ \left(p_1^*D_1 + p_2^*D_2\right) . \,(C \times \{P\}) = \deg D_1 \,. \end{aligned}$$

This would imply $\deg D_1 = \deg D_2 = 1$, so the divisor Δ would need to be numerically equivalent to the divisor $([P] \times C) + (C \times [P])$, and that would give

$$2 - 2g = \Delta^2 = \big(([P] \times C) + (C \times [P])\big)^2 = 2 \;,$$

a contradiction since $g > 1$.

Here we have used *intersection theory on* $C \times C$: There is a function $(D_1 \,.\, D_2)$ from $\operatorname{Div}(C \times C) \times \operatorname{Div}(C \times C)$ to $\mathbb{Z}$ that satisfies:

(1) The pairing is *symmetric*: $(D_1 \,.\, D_2) = (D_2 \,.\, D_1)$ for all $D_1, D_2 \in \operatorname{Div}(C \times C)$;
(2) the pairing is *bilinear*: $((D_1 + D_2) \,.\, D_3) = (D_1 \,.\, D_3) + (D_2 \,.\, D_3)$ for all $D_1, D_2, D_3 \in \operatorname{Div}(C \times C)$;
(3) the kernels on the left and the right contain the subgroup of principal divisors (and hence this is actually a pairing $\operatorname{Pic}(C\times C)\times\operatorname{Pic}(C\times C) \to \mathbb{Z}$); and
(4) if D_1 and D_2 are prime divisors on $C \times C$ that cross transversally, then $(D_1 \,.\, D_2)$ is the number of points of intersection in $(C \times C)(\overline{\mathbb{Q}})$.

For more details on intersection theory, see [**Har77**, App. A].

So now let

$$\Delta' = \Delta - \{P\} \times C - C \times \{P\} \;.$$

This is a better divisor to work with than Δ, because (unlike Δ) it has degree 0 on the fibers of p_1 and p_2, so it is orthogonal (in the intersection pairing) to the "obvious" divisors on $C \times C$. We have

$$\begin{aligned}(\Delta')^2 &= \Delta^2 - 2\Delta(\{P\} \times C) - 2\Delta(C \times \{P\}) + 2(\{P\} \times C)(C \times \{P\}) \\ &= (2 - 2g) - 2 - 2 + 2 \\ &= -2g \;.\end{aligned}$$

For the actual proof of Faltings' Theorem (Mordell's conjecture), we will use a divisor of the form

$$dY = d(\Delta' + a_1(\{P\} \times C) + a_2(C \times \{P\})) \;, \tag{1.1}$$

where $a_1, a_2 \in \mathbb{Q}$, $d \in \mathbb{Z}$, $d > 0$, and d is sufficiently divisible so that $da_1, da_2 \in \mathbb{Z}$.

How will this divisor be used, and why is this form of Y useful?

Part of an answer to this question lies in *Siegel's lemma.*

Proposition 1.2. *(Siegel's lemma) Let M and N be positive integers with $N > M$, and let A be an $M \times N$ matrix with integer entries whose absolute values are bounded by $B \in \mathbb{Z}$. Then there is a nonzero element $\mathbf{x} \in \mathbb{Z}^N$ such that $A\mathbf{x} = \mathbf{0}$ and $|\mathbf{x}| \le (NB)^{M/(N-M)}$.*

Here $|\mathbf{x}| = \max\{|x_i| : 1 \le i \le N\}$.

Proof (sketch). For $c \in \mathbb{N}$, we have

$$\big|\{\mathbf{x} \in \mathbb{Z}^N : 0 \le x_i < c \;\forall\, i\}\big| = c^N \;.$$

For $\mathbf{x}$ in this set, we have $|\mathbf{x}| < c$ and $|A\mathbf{x}| < NBc$, so A maps a set of size c^N to a set of size $< (2NBc)^M$. If c is large enough so that $c^N > (2NBc)^M$, then there will be distinct $\mathbf{x}', \mathbf{x}'' \in \mathbb{Z}^N$ with $|\mathbf{x}'|, |\mathbf{x}''| < c$ and $A\mathbf{x}' = A\mathbf{x}''$. Let $\mathbf{x} = \mathbf{x}' - \mathbf{x}''$. Then $\mathbf{x} \ne \mathbf{0}$, $|\mathbf{x}| < c$, and $A\mathbf{x} = \mathbf{0}$. The inequality $c^N > (2NBc)^M$ is equivalent to $c > (2NB)^{M/(N-M)}$.

With a little more work, a similar result can be obtained with $c = (NB)^{M/(N-M)}$; see [**HS00**, Lemma D.4.1]. □

When using the Thue-Siegel method to prove Roth's theorem (or, since we'll be using only two variables, one of the earlier weaker versions of Roth's theorem), Siegel's lemma is used to construct a polynomial in two variables.

Given positive integers d_1 and d_2, Siegel's lemma is used to construct a polynomial $f \in \mathbb{Z}[x,y]$ of degree $\le d_1$ in x and $\le d_2$ in y. The proof requires some vanishing conditions at (α,α); these conditions lead to a system of linear equations in which the variables are the coefficients. The number of variables is $N = (d_1+1)(d_2+1)$.

As for the number M of equations, let's first look at a polynomial for which $d_2 = 1$. We will want this to be a polynomial of the form

$$f(x,y) = (x-\alpha)^n G(x) - (y-\alpha)H(x)\ .$$

A polynomial f will be of this form if and only if

$$f(\alpha,\alpha) = \frac{\partial}{\partial x} f(\alpha,\alpha) = \left(\frac{\partial}{\partial x}\right)^2 f(\alpha,\alpha) = \cdots = \left(\frac{\partial}{\partial x}\right)^{n-1} f(\alpha,\alpha) = 0\ .$$

For larger values of d_2, the linear conditions will come from requiring the vanishing of $\left(\frac{\partial}{\partial x}\right)^i \left(\frac{\partial}{\partial y}\right)^j f(\alpha,\alpha)$ for certain pairs $(i,j) \in \mathbb{N}^2$.

Returning to the context of a proof of Mordell's conjecture, we need to find out what N is in this case. This amounts to finding $\dim_k H^0(C\times C, Y)$.

A useful way to find this dimension (or at least a lower bound for it) is the following special case of the Riemann-Roch-Hirzebruch theorem.

Theorem 1.3. *Let $\mathscr{L}$ be a line sheaf on a nonsingular complete surface X over a field k, and let $d \in \mathbb{N}$. Then*

$$h^0(X,\mathscr{L}^{\otimes d}) - h^1(X,\mathscr{L}^{\otimes d}) + h^2(X,\mathscr{L}^{\otimes d}) = \frac{d^2}{2}(\mathscr{L}\,.\,\mathscr{L}) + O(d)\ . \tag{1.3.1}$$

(Here $h^i(X,\mathscr{L}^{\otimes d}) = \dim_k H^i(X,\mathscr{L}^{\otimes d})$.)

When $\mathscr{L} = \mathscr{O}(Y)$, the right-hand side of (1.3.1) is

$$\frac{d^2}{2}(-2g + 2a_1a_2 + o(1)) = d^2(-g + a_1a_2 + o(1))\ .$$

As long as this is positive, $H^0(C\times C, dY)$ will be nonzero. This is because $h^1(C\times C, dY) \ge 0$, and $h^2(C\times C, dY) = h^0(C\times C, K_{C\times C} - dY) = 0$ for $d \gg 0$ by duality and the fact that Y will have intersections $a_2 > 0$ and $a_1 > 0$ with the fibers of p_1 and p_2, respectively.

To give a little more information about the definition (1.1) of Y, a_1 and a_2 are positive rational numbers chosen such that $a_1a_2 > g$ (but is very close to g) and a_1/a_2 is close to $h(P_2)/h(P_1)$, where P_1 and P_2 will be two suitably chosen rational points on C and $h(P_1)$ and $h(P_2)$ are their heights (using a height function that will be described later).

In summary, the answer to the second question, about what object replaces the auxiliary polynomial in the proofs of Thue, Siegel, etc., is that the replacement is the line sheaf $\mathscr{O}(dY)$ (for large d).

Let us now return briefly to the first question: What replaces α? Mordell's conjecture does not contain any diophantine inequality, so this was perhaps the hardest question that came up when trying to find a proof.

The quick answer is that nothing replaces α, because $h^0(C \times C, dY)$ (which replaces $N-M$, a lower bound for the dimension of the solution space in Siegel's lemma) is already small.

But this is not all of the answer. In the proofs of Thue, Siegel, etc., as $N - M \to 0$ the bound on $|\mathbf{x}|$ becomes larger, so Steps 2 and 3 become harder. That phenomenon occurs in the proof of Mordell, too—see (6.1.1) as $\gamma \to 0$.

Note that the Riemann-Roch-Hirzebruch theorem does not say anything that corresponds to the bound on the coefficients in Siegel's lemma. This is a necessary aspect of the proof, yet it is handled differently in the three major variants of the proof:

- My own proof used Arakelov theory (as developed by Gillet and Soulé), in particular the "Riemann-Roch-Hirzebruch-Grothendieck-Gillet-Soulé" theorem, which extended the Riemann-Roch-Hirzebruch theorem by encompassing families of varieties (Riemann-Roch-Hirzebruch-Grothendieck) and further incorporated Arakelov theory.
- Faltings' proof replaced the Riemann-Roch-. . . -Gillet-Soulé Theorem with a fancier Siegel's lemma (and also extended the theorem to higher dimensions).
- Bombieri's proof also put Siegel's lemma back in, but used the basic Siegel lemma and used it in a different way.

This paper will follow Bombieri's proof.

Getting back to the question about what replaces α, for simplicity we restrict to considering the proof, in the classical Thue-Siegel method, that the auxiliary polynomial f vanishes at $(p_1/q_1, p_2/q_2)$. Assume that it is nonzero there. The approximation conditions, together with vanishing conditions for f at (α, α) and the bounds on the coefficients of f, imply an upper bound on $\|f(p_1/q_1, p_2/q_2)\|_v$ when $v = \infty$. By the product formula (0.3), this gives a lower bound on the product

$$\prod_{v \neq \infty} \left\| f\left(\frac{p_1}{q_1}, \frac{p_2}{q_2}\right) \right\|_v . \tag{1.4}$$

However, since the coefficients of f are integers, the denominator of $f(p_1/q_1, p_2/q_2)$ is bounded by $|q_1^{d_1} q_2^{d_2}|$. The proof obtains a contradiction from these two facts to conclude that $f(p_1/q_1, p_2/q_2) = 0$.

In the case of Mordell's conjecture, there is no α, so we can consider the equivalent of the product (1.4) over all places of the number field k. The geometry of the curve C (or, more precisely, of its Jacobian), implies a lower bound on this product. This is equivalent to an upper bound on minus the logarithm of such a product, which is closely related to the *height* of a point (P_1, P_2) on $C \times C$ relative to a certain line sheaf. The proof then proceeds to a contradiction by finding a contradictory lower bound for this height.

Thus, it is the geometry of the Jacobian that replaces α. This is very natural, from the point of view that the approximation conditions in the Thue-Siegel proof can be viewed as taking place on the log Jacobian of $\mathbb{P}^1$ minus the Galois conjugates of α.

2. Basic Constructions in Bombieri's Proof

Bombieri avoids Arakelov theory in his proof by taking suitable embeddings into projective space and suitable linear projections, and by expressing the divisor dY as a difference of two suitably chosen divisors on $C \times C$.

Here is the basic geometric construction used in Bombieri's proof.

Recall that C is a smooth projective curve of genus $g > 1$ over a number field k. We are assuming (by way of contradiction) that $C(k)$ is infinite.

Step 0. Fix a divisor A of degree 1 on C, chosen so that $(2g-2)A \sim K_C$, where K_C is the canonical divisor on C. (The reason for this condition will be explained in Section 3.) (This may require extending the base field k.) By abuse of notation, we write p_1^*A and p_2^*A as $A \times C$ and $C \times A$, respectively, instead, where $p_1, p_2 \colon C \times C \to C$ are the two projection morphisms.

Let

$$\Delta' = \Delta - A \times C - C \times A \,. \tag{2.1}$$

Step 1. Fix $s \in \mathbb{Z}$ sufficiently large such that the divisor

$$B := s(A \times C) + s(C \times A) - \Delta' \tag{2.2}$$

is very ample. Choose a projective embedding

$$\phi_B \colon C \times C \to \mathbb{P}^m$$

associated to the complete linear system $|B|$, and use ϕ_B to define a height h_B on $C \times C$:

$$h_B(P_1, P_2) = h(\phi_B(P_1, P_2)) \,, \tag{2.3}$$

where the second height $h(\cdot)$ is the standard Weil height (0.6) on $\mathbb{P}^m$.

Step 2. Choose $d \in \mathbb{Z}$ sufficiently large so that the map

$$H^0(\mathbb{P}^m, \mathscr{O}(d)) \longrightarrow H^0(C \times C, dB) \tag{2.4}$$

is surjective. (This can be done by choosing d such that $H^1(\mathbb{P}^m, \mathscr{I}_{C\times C} \otimes \mathscr{O}(d)) = 0$, where $\mathscr{I}_{C\times C}$ is the ideal sheaf on $\mathbb{P}^m$ corresponding to the image of ϕ_B, with reduced induced subscheme structure. This condition on d is true for sufficiently large d by [**Har77**, III 5.2].)

Step 3. Fix $N \in \mathbb{Z}$ such that NA is very ample. Let

$$\phi_{NA} \colon C \hookrightarrow \mathbb{P}^n \tag{2.5}$$

be some associated projective embedding, chosen suitably generically to satisfy conditions to be described later (see Section 6). As with h_B, let h_{NA} be the height on C defined by ϕ_{NA}:

$$h_{NA}(P) = h(\phi_{NA}(P)) \,, \tag{2.6}$$

and also let $h_A(P) = \frac{1}{N} h_{NA}(P)$, for all $P \in C(\bar{k})$.

Step 4. As before, the natural map

$$H^0(\mathbb{P}^n \times \mathbb{P}^n, \mathscr{O}(\delta_1, \delta_2)) \longrightarrow H^0(C \times C, \delta_1(NA \times C) + \delta_2(C \times NA)) \tag{2.7}$$

is surjective for all sufficiently large $\delta_1, \delta_2 > 0$. (In fact, it is true for all δ_1, δ_2 with $\min\{\delta_1, \delta_2\}$ sufficiently large, but for our purposes it suffices to know that for each positive $r \in \mathbb{Q}$ it is true for all sufficiently large δ_1, δ_2 with $\delta_1/\delta_2 = r$, with the bound depending on r. This special case follows directly from [**Har77**, III 5.2] as before.)

Step 5. We will be working with points $P_1, P_2 \in C(k)$, chosen such that $h_{NA}(P_1)$ and $h_{NA}(P_2)/h_{NA}(P_1)$ are large. Given such P_1 and P_2, choose rational a_1 and a_2 such that

$$a_1^2 \approx (g+\gamma)\frac{h_{NA}(P_2)}{h_{NA}(P_1)} \qquad \text{and} \qquad a_2^2 \approx (g+\gamma)\frac{h_{NA}(P_1)}{h_{NA}(P_2)}\ , \tag{2.8}$$

so that

$$\frac{a_1}{a_2} \approx \frac{h_{NA}(P_2)}{h_{NA}(P_1)}\ .$$

Here $\gamma > 0$ is small (to be determined later). Let

$$\delta_1 = (a_1 + s)\frac{d}{N} \qquad \text{and} \qquad \delta_2 = (a_2 + s)\frac{d}{N}\ , \tag{2.9}$$

and let

$$\begin{aligned} Y &= \delta_1 NA \times C + C \times \delta_2 NA - dB \\ &= d(a_1 A \times C + a_2 C \times A + \Delta')\ , \end{aligned} \tag{2.10}$$

where d is an integer chosen sufficiently divisible so that $\delta_1, \delta_2 \in \mathbb{Z}$. (This differs by a factor of d from the Y of (1.1).) Define

$$h_Y(P_1, P_2) = \delta_1 h_{NA}(P_1) + \delta_2 h_{NA}(P_2) - d h_B(P_1, P_2)\ . \tag{2.11}$$

By definition, the height h_Y behaves well as a_1, a_2, and d vary.

Bombieri's proof avoids Arakelov theory by using clever manipulations in the heights h_B, h_{NA}, and h_Y instead.

3. Divisors, Heights, and Jacobians

This section describes the connections with the Jacobian of C, which are central to all three variants of the proof discussed here.

Let J be the Jacobian of C, and define

$$j \colon C \hookrightarrow J$$

to be the map $j(x) = (x) - A \in J$. (More precisely, it is defined by taking a point $x \in C \times_k \bar{k}$ to the divisor $(x) - A$ in $J \times_k \bar{k}$, but this map comes from a morphism $C \to J$ over k; see [**Mil86**, Thm. 1.1].) Also let Θ be the associated theta divisor; this is the sum $j(C) + \cdots + j(C)$ with $g-1$ terms (under the group operation).

The following is a well-known fact about the theta divisor.

Lemma 3.1. [**HS00**, Lemma E.2.1] *With notation as above, we have*

- *(a). Θ is a symmetric divisor; i.e., $[-1]^*\Theta \sim \Theta$, where $[-1]$ denotes taking the inverse under the group operation;*
- *(b). $j^*\Theta \sim gA$; and*
- *(c). if $s \colon J \times J \to J$ is the group operation, then*

$$(j \times j)^*(s^*\Theta - p_1^*\Theta - p_2^*\Theta) \sim -\Delta'\ . \tag{3.1.1}$$

In addition, Θ is ample; see [**Mil86**, Thm. 6.6 and the preceding paragraph].

Since Θ is ample and symmetric, it gives rise to a (Néron-Tate) **canonical height** $\hat{h}_\Theta$ on J:

Theorem 3.2. [**HS00**, Thm. B.5.1] *Let J be an abelian variety over a number field k, and let D be a symmetric divisor on J. Then there exists a unique canonical height $\hat h_D\colon J(\bar k)\to\mathbb{R}$ such that*

$$\hat h_D([m]P)=m^2\hat h_D(P)\qquad \text{for all } P\in J(\bar k) \text{ and all } m\in\mathbb{Z}$$

and for all (conventional) Weil heights h_D on J relative to D,

$$\hat h_D(P)=h_D(P)+O(1)\qquad \text{for all } P\in J(\bar k)\ ,$$

with the implicit constant depending on h_D. Moreover:

(a). (Parallelogram Law)

$$\hat h_D(P+Q)+\hat h_D(P-Q)=2\hat h_D(P)+2\hat h_D(Q)\qquad \text{for all } P,Q\in J(\bar k)\ .$$

(b). The canonical height $\hat h_D$ is a quadratic form. The associated pairing

$$\langle\cdot,\cdot\rangle_D\colon J(\bar k)\times J(\bar k)\to\mathbb{R}$$

defined by

$$\langle P,Q\rangle_D=\frac{\hat h_D(P+Q)-\hat h_D(P)-\hat h_D(Q)}{2} \tag{3.2.1}$$

is bilinear and satisfies $\langle P,P\rangle_D=\hat h_D(P)$.

Since Δ is an effective divisor on $C\times C$, we have $h_\Delta(P_1,P_2)\ge O(1)$ for all $P_1,P_2\in C(\bar k)$ with $P_1\ne P_2$. This fact can be combined with (3.1.1) to give information on the canonical heights of rational points on C, as follows.

Since the bilinear form $\langle\cdot,\cdot\rangle_\Theta$ is positive definite on $J(\bar k)\otimes_{\mathbb{Z}}\mathbb{R}$, the latter is an (infinite dimensional) Euclidean space, and in particular

$$\langle x,y\rangle_\Theta=(\cos\theta)\sqrt{|x|^2|y|^2} \tag{3.3}$$

for all $x,y\in J(\bar k)\otimes_{\mathbb{Z}}\mathbb{R}$, where

$$|x|^2=\langle x,x\rangle_\Theta=\hat h_\Theta(x)$$

and θ is the angle between x and y. In particular, let $x=j(P_1)$ and $y=j(P_2)$. Then, by (3.3), (3.2.1), (3.1.1), functoriality of heights, and Lemma 3.1b,

$$\begin{aligned}(\cos\theta)\sqrt{|x|^2|y|^2}&=\langle x,y\rangle_\Theta\\ &=\frac{\hat h_\Theta(x+y)-\hat h_\Theta(x)-\hat h_\Theta(y)}{2}\\ &=-\frac12 h_{\Delta'}(P_1,P_2)+O(1)\\ &=-\frac12 h_\Delta(P_1,P_2)+\frac12 h_A(P_1)+\frac12 h_A(P_2)+O(1)\\ &\le\frac{|x|^2}{2g}+\frac{|y|^2}{2g}+O(1)\ .\end{aligned} \tag{3.4}$$

Since $\sqrt{|x|^2|y|^2}$ can be as large as $(|x|^2+|y|^2)/2$ (when $|x|=|y|$) and $g>1$, this means that $\cos\theta$ is bounded away from 1 (so θ is bounded away from 0) when $|x|$ and $|y|$ are close to each other.

When P_1 and P_2 are rational over k, we have $x,y\in J(k)$. By the Mordell-Weil theorem [**HS00**, Thm. C.0.1], $J(k)$ is finitely generated, so $J(k)\otimes_{\mathbb{Z}}\mathbb{R}$ is finite dimensional. Since the angle θ is bounded from below when $|x|$ and $|y|$ are close, this implies that not too many points in $C(k)$ can have heights sufficiently large and in a sufficiently small range. This is the analogue for Mordell's conjecture of the "gap principle" in the context of diophantine approximation [**HS00**, Exercises D.12 and D.13], and was originally discovered by Mumford [**Mum65**]; see also [**HS00**, Prop. B.6.6].

This idea will be used in the next section, using a different effective divisor in place of Δ.

4. An Upper Bound for the Height

The purpose of this section is to find an upper bound for the height $h_Y(P_1, P_2)$, which will be compared with the lower bound (5.5.1) to ultimately derive a contradiction in the proof of the theorem.

This will be done using the results of Section 3 on canonical heights on the Jacobian of C.

To begin, let a_1 and a_2 be positive rational numbers, to be determined later, and let $d_i = a_i d$ for $i = 1, 2$.

LEMMA 4.1. *There is a constant c_1, depending only on C, A, ϕ_{NA}, and ϕ_B (but not on P_1 or P_2), such that*

$$h_Y(P_1, P_2) \le \frac{d_1}{g}|j(P_1)|^2 + \frac{d_2}{g}|j(P_2)|^2 - 2d\langle j(P_1), j(P_2)\rangle_\Theta + c_1(d_1 + d_2 + d) \tag{4.1.1}$$

for all $P_1, P_2 \in C(\bar{k})$.

PROOF. Recall from (2.11) that h_Y is defined using h_{NA} and h_B. We also use (2.2) to define $h_{\Delta'}$:

$$h_B(P_1, P_2) = \frac{s}{N} h_{NA}(P_1) + \frac{s}{N} h_{NA}(P_2) - h_{\Delta'}(P_1, P_2) \, .$$

With this definition, (2.9) and (2.11) give

$$h_Y(P_1, P_2) = d\left(h_{\Delta'}(P_1, P_2) + \frac{a_1}{N} h_{NA}(P_1) + \frac{a_2}{N} h_{NA}(P_2)\right) \, . \tag{4.1.2}$$

Note that this holds *exactly* (i.e., not up to $O(1)$), by the various definitions. By Theorem 3.2 and Lemma 3.1b,

$$|j(P)|^2 = \hat{h}_\Theta(j(P)) = \frac{g}{N} h_{NA}(P) + O(1)$$

for all $P \in C(\bar{k})$. In addition, by (3.1.1) and (3.2.1),

$$h_{\Delta'}(P_1, P_2) = -2\langle j(P_1), j(P_2)\rangle_\Theta + O(1)$$

for all $P_1, P_2 \in C(\bar{k})$.

Applying these two equations to (4.1.2) gives (4.1.1). (Actually, (4.1.1) is an equality up to $O(d_1 + d_2 + d)$.) □

The next step is to use (3.4) to get the right-hand side of (4.1.1) to be negative (for sufficiently large $d_1 + d_2 + d$).

As noted at the end of Section 3, $J(k) \otimes_{\mathbb{Z}} \mathbb{R}$ is a *finite*-dimensional vector space. Therefore, since we have assumed that $C(k)$ is infinite, there is an infinite subset Σ' of $C(k)$ such that, for all $P_1, P_2 \in \Sigma'$, the angle between $j(P_1)$ and $j(P_2)$ in $J(k) \otimes \mathbb{R}$ is at most $\cos^{-1}(3/4)$; therefore

$$\langle j(P_1), j(P_2)\rangle_\Theta \ge \frac{3}{4}\sqrt{\hat{h}_\Theta(j(P_1))\hat{h}_\Theta(j(P_2))} \, .$$

Incorporating this into (4.1.1) gives

$$h_Y(P_1,P_2) \le \frac{d_1}{g}|j(P_1)|^2 + \frac{d_2}{g}|j(P_2)|^2 - \frac{3}{2}d|j(P_1)||j(P_2)| + c_1(d_1+d_2+d) \tag{4.2}$$

for all $P_1, P_2 \in \Sigma'$.

5. A Lower Bound

In Section 3, since Δ is an effective divisor and $P_1 \ne P_2$ (hence $(P_1,P_2) \notin \Delta$), the height $h_\Delta(P_1,P_2)$ is trivially bounded from below. In the present case, though, Δ will be replaced by an effective divisor associated to a global section s of $\mathscr{O}(Y)$. The section s will be constructed using Siegel's lemma, and it will be hard to guarantee that s will not vanish at (P_1,P_2). In fact, in general s will vanish there, but we will derive a contradiction based on its (weighted) multiplicity at (P_1,P_2).

Since Y depends on P_1 and P_2, it will be necessary to control the constant $O(1)$ in (3.4), and this is a major part of the proof.

We start by defining a notion of the height of a global section s of $\mathscr{O}(Y)$, which plays the role of the height of the auxiliary polynomial in the classical Thue-Siegel method.

Lemma 5.1. *Let Y be a divisor on $C \times C$ as in (2.10). Let $y_0,\dots,y_m$ be the global sections of $\mathscr{O}(B)$ corresponding to the coordinates of ϕ_B. Also let $x_0,\dots,x_n$ and $x'_0,\dots,x'_n$ be bases for the global sections on $C\times C$ of $p_1^*\mathscr{O}(NA)$ and $p_2^*\mathscr{O}(NA)$, respectively. Then, for each global section s of $\mathscr{O}(Y)$ there are polynomials*

$$F_0(\mathbf{x},\mathbf{x}'),\dots,F_m(\mathbf{x},\mathbf{x}')$$

such that

(i). each of the F_i is homogeneous of degree δ_1 in the variables $\mathbf{x} = (x_0,\dots,x_n)$ and homogeneous of degree δ_2 in $\mathbf{x}' = (x'_0,\dots,x'_n)$,

(ii). for all $i \in \{0,\dots,m\}$,

$$F_i(\mathbf{x},\mathbf{x}') = y_i^d s\ ,$$

and

(iii). for all $i,j \in \{0,\dots,m\}$,

$$\frac{F_i(\mathbf{x},\mathbf{x}')}{y_i^d} = \frac{F_j(\mathbf{x},\mathbf{x}')}{y_j^d} \tag{5.1.1}$$

everywhere on $C\times C \setminus \{y_iy_j = 0\}$.

Conversely, each set of polynomials $F_0,\dots,F_m$ satisfying (i) and (iii) uniquely determines a global section s of $\mathscr{O}(Y)$ that satisfies (ii). Thus, there is a one-to-many correspondence between global sections s and systems $(F_0,\dots,F_m)$ of polynomials satisfying (i) and (iii).

Proof. Let s be a global section of $\mathscr{O}(Y)$ on $C \times C$. For each i, the global section

$$y_i^d s \in \Gamma(C\times C,\ \delta_1 NA \times C + C \times \delta_2 NA)\ ,$$

lifts by surjectivity of (2.7) to

$$F_i(\mathbf{x},\mathbf{x}') \in \Gamma(\mathbb{P}^n \times \mathbb{P}^n, \mathscr{O}(\delta_1,\delta_2))\ .$$

These F_i are polynomials in $\mathbf{x}$ and $\mathbf{x}'$ that satisfy (i) and (ii) by construction. They also satisfy (iii), because the two sides of (5.1.1) are equal to s everywhere.

Conversely, let $F_0,\dots,F_m$ be polynomials satisfying conditions (i) and (iii). For each i, $F_i(\mathbf{x},\mathbf{x}')\big|_{C\times C}$ is a global section of $\mathscr{O}(\delta_1 NA\times C+C\times\delta_2 NA)$ on $C\times C$, and $y_i\big|_{C\times C}$ is a global section of $\mathscr{O}(B)$ on $C\times C$. Therefore

$$\frac{F_i}{y_i^d}\in\Gamma((C\times C)\setminus\{y_i=0\},\ Y)\ . \tag{5.1.2}$$

Since $y_0,\dots,y_m$ are never simultaneously zero, the sections (5.1.2) glue together to give a global section s of $\mathscr{O}(Y)$ on $C\times C$, by (5.1.1). □

Lemma 5.2. *Let s be a global section of $\mathscr{O}(Y)$ that does not vanish at (P_1,P_2), let $\mathscr{F}=(F_0,\dots,F_m)$ be as in Lemma 5.1 (as determined by s), and define*

$$h(\mathscr{F})=\frac{1}{[k:\mathbb{Q}]}\sum_{v\in M_k}\max_i\max_{c\in\{\textit{coefficients of }F_i\}}\|c\|_v\ .$$

Then

$$h_Y(P_1,P_2)\ge -h(\mathscr{F})-n\log((\delta_1+n)(\delta_2+n))\ . \tag{5.2.1}$$

(Recall that $n=h^0(C,NA)-1$.)

Proof. Let $\tilde{\mathbf{x}}=\phi_{NA}(P_1)$ and let $\tilde{x}_0,\dots,\tilde{x}_n\in k$ be homogeneous coordinates for $\tilde{\mathbf{x}}$. Define $\tilde{\mathbf{x}}'=\phi_{NA}(P_2)$ and $\tilde{x}'_0,\dots,\tilde{x}'_n\in k$ similarly, as well as $\tilde{\mathbf{y}}=\phi_B(P_1,P_2)$ and $\tilde{y}_0,\dots,\tilde{y}_m\in k$.

Then, by definition,

$$\begin{aligned} h_Y(P_1,P_2)&=\frac{\delta_1}{[k:\mathbb{Q}]}\sum_v\max_j\log\|\tilde{x}_j\|_v+\frac{\delta_2}{[k:\mathbb{Q}]}\sum_v\max_j\log\|\tilde{x}'_j\|_v\\ &\qquad-\frac{d}{[k:\mathbb{Q}]}\sum_v\max_i\log\|\tilde{y}_i\|_v\\ &=\frac{1}{[k:\mathbb{Q}]}\sum_v\min_i\max_{j,j'}\log\left\|\frac{\tilde{x}_j^{\delta_1}(\tilde{x}'_{j'})^{\delta_2}}{\tilde{y}_i^d}\right\|_v\ , \end{aligned}\tag{5.2.2}$$

where all sums are over all $v\in M_k$.

As before, we continue to let $x_0,\dots,x_n$ and $y_0,\dots,y_m$ be the global sections of $\mathscr{O}(NA)$ and $\mathscr{O}(B)$, respectively, corresponding to the coordinates of ϕ_{NA} and ϕ_B, respectively. We assume that x_0 does not vanish at P_1 or P_2, and that y_0 does not vanish at (P_1,P_2). Then $\tilde{x}_0\ne 0$, $\tilde{x}'_0\ne 0$, and $\tilde{y}_0\ne 0$. Assume that $\tilde{x}_0=\tilde{x}'_0=\tilde{y}_0=1$. For each j, x_j/x_0 is a rational function on C that is regular at P_1 and at P_2, and we have

$$\left(\frac{x_j}{x_0}\right)(P_1)=\frac{\tilde{x}_j}{\tilde{x}_0}=\tilde{x}_j\qquad\text{and}\qquad\left(\frac{x_j}{x_0}\right)(P_2)=\frac{\tilde{x}'_j}{\tilde{x}'_0}=\tilde{x}'_j\ . \tag{5.2.3}$$

Similarly, for each i, y_i/y_0 is a rational function on $C\times C$ that is regular at (P_1,P_2), and

$$\left(\frac{y_i}{y_0}\right)(P_1,P_2)=\tilde{y}_i\ . \tag{5.2.4}$$

Since s and $p_1^*(x_0^{\delta_1})p_2^*(x_0^{\delta_2})/y_0^d$ are both regular sections of $\mathscr{O}(Y)$ in a neighborhood of (P_1,P_2) and the latter section does not vanish at (P_1,P_2), the ratio

$$s(P_1,P_2)\ \Big/\ \frac{x_0(P_1)^{\delta_1}x_0(P_2)^{\delta_2}}{y_0(P_1,P_2)^d}$$

is a well-defined element of k, and it is nonzero at (P_1, P_2) by the assumption that s also does not vanish at (P_1, P_2). Therefore, by the product formula (0.3),

$$\frac{1}{[k:\mathbb{Q}]}\sum_v \log\left\| s(P_1,P_2) \Big/ \frac{x_0(P_1)^{\delta_1}x_0(P_2)^{\delta_2}}{y_0(P_1,P_2)^d}\right\|_v = 0 .$$

Subtracting this from (5.2.2) and applying (5.2.3) and (5.2.4) then gives

$$\begin{aligned} h_Y(P_1,P_2) &= \frac{1}{[k:\mathbb{Q}]}\sum_v \min_i \max_{j,j'} \log\left\|\frac{x_j(P_1)^{\delta_1}x_{j'}(P_2)^{\delta_2}}{s(P_1,P_2)y_i(P_1,P_2)^d}\right\|_v \\ &= \frac{1}{[k:\mathbb{Q}]}\sum_v \min_i \max_{j,j'} \log\left\|\frac{x_j(P_1)^{\delta_1}x_{j'}(P_2)^{\delta_2}}{F_i(\mathbf{x}(P_1),\mathbf{x}(P_2))}\right\|_v \\ &= -\frac{1}{[k:\mathbb{Q}]}\sum_v \max_i \min_{j,j'} \log\left\|F_i\left(\frac{\mathbf{x}(P_1)}{x_j(P_1)},\frac{\mathbf{x}(P_2)}{x_{j'}(P_2)}\right)\right\|_v , \end{aligned} \tag{5.2.5}$$

where $\mathbf{x} = (x_0, \dots, x_n)$.

For each v and i let $j_{v,i}$ and $j'_{v,i}$ be the values of j and j', respectively, where the minimum occurs; i.e.,

$$\min_{j,j'}\left\|F_i\left(\frac{\mathbf{x}(P_1)}{x_j(P_1)},\frac{\mathbf{x}(P_2)}{x_{j'}(P_2)}\right)\right\|_v = \left\|F_i\left(\frac{\mathbf{x}(P_1)}{x_{j_{v,i}}(P_1)},\frac{\mathbf{x}(P_2)}{x_{j'_{v,i}}(P_2)}\right)\right\|_v .$$

Then, by bihomogeneity of F_i, $\|x_j/x_{j_{v,i}}\|_v \le 1$ and $\|x'_{j'}/x'_{j'_{v,i}}\|_v \le 1$ for all j and j' (respectively). Since the number of nonzero terms in F_i is at most $(\delta_1+n)^n(\delta_2+n)^n$,

$$\min_{j,j'}\left\|F_i\left(\frac{\mathbf{x}(P_1)}{x_j(P_1)},\frac{\mathbf{x}(P_2)}{x_{j'}(P_2)}\right)\right\|_v \le \left((\delta_1+n)^n(\delta_2+n)^n\right)^{N_v} \max_{c\in\{\text{coefficients of } F_i\}} \|c\|_v \tag{5.2.6}$$

for all $v \in M_k$ and all i, by (0.5).

The inequality (5.2.1) then follows from (5.2.5), (5.2.6), (0.4), and the definition of $h_{\mathscr{F}}$. □

This will be used to do the equivalent of controlling the constant $O(1)$ in (3.4), because Y here will play the role of Δ later on.

However, things are a bit more complicated than that, because it is possible that s may vanish at (P_1, P_2). In general, what will be needed is a similar argument using "partial derivatives" of s. Since s is a section of a line bundle, as opposed to a function, the meaning of partial derivative is not so clear. However, as with s itself, under certain conditions it is possible to obtain a well-defined notion of whether certain partial derivatives of s are zero.

DEFINITION 5.3. Let s be a rational section of some line sheaf $\mathscr{L}$ on $C \times C$, and let $P_1, P_2 \in C(k)$ be points such that s is regular at (P_1, P_2). Fix a rational section s_0 of $\mathscr{L}$ that generates $\mathscr{L}$ in a neighborhood of (P_1, P_2), and let ζ_1 and ζ_2 be local coordinates on C at P_1 and P_2, respectively. Then we say that a pair $(i_1^*, i_2^*) \in \mathbb{N}^2$ is **admissible** if

$$\left(\frac{\partial}{\partial\zeta_1}\right)^{i_1}\left(\frac{\partial}{\partial\zeta_2}\right)^{i_2}\left(\frac{s}{s_0}\right)(P_1,P_2) = 0 \tag{5.3.1}$$

for all pairs $(i_1, i_2) \in \mathbb{N}^2$ such that $i_1 \le i_1^*$, $i_2 \le i_2^*$, and $(i_1, i_2) \ne (i_1^*, i_2^*)$.

We note that, by elementary computations and induction on (i_1, i_2), the condition (5.3.1) is independent of the choices of s_0, ζ_1, and ζ_2 when (i_1, i_2) is admissible. Therefore admissibility itself is independent of these choices.

The usefulness of admissibility of pairs stems from the following lemma.

LEMMA 5.4. *Let s, $\mathscr{L}$, P_1, P_2, s_0, ζ_1, and ζ_2 be as in Definition 5.3, and let (i_1^*, i_2^*) be an admissible pair. Then the quantity*

$$\left(\frac{\partial}{\partial\zeta_1}\right)^{i_1^*}\left(\frac{\partial}{\partial\zeta_2}\right)^{i_2^*}\left(\frac{s}{s_0}\right)(P_1,P_2)\cdot\left(s_0\otimes(d\zeta_1)^{\otimes i_1^*}\otimes(d\zeta_2)^{\otimes i_2^*}\right)\Big|_{(P_1,P_2)} \in \mathscr{L}\otimes p_1^*\Omega_C^{\otimes i_1^*}\otimes p_2^*\Omega_C^{\otimes i_2^*} \tag{5.4.1}$$

is independent of the choices of s_0, ζ_1 and ζ_2.

PROOF. Independence of s_0 is easy to see from the Leibniz rule and admissibility of (i_1^*, i_2^*), and independence of ζ_1 and ζ_2 follows from formalisms of differential geometry. □

For admissible pairs (i_1^*, i_2^*), we can then define

$$\left(\frac{\partial}{\partial\zeta_1}\right)^{i_1^*}\left(\frac{\partial}{\partial\zeta_2}\right)^{i_2^*} s(P_1,P_2)$$

to be the section of (5.4.1).

With more work, one can then generalize Lemma 5.2 as follows.

LEMMA 5.5. *Let s be a global section of $\mathscr{O}(Y)$, let $\mathscr{F} = (F_0, \dots, F_m)$ be as in Lemma 5.1 (as determined by s), and let (i_1^*, i_2^*) be an admissible pair at (P_1, P_2) for s. Assume that $(\partial/\partial\zeta_1)^{i_1^*}(\partial/\partial\zeta_2)^{i_2^*}s(P_1,P_2) \ne 0$. Then there are constants c_2 and c_3, depending only on C, ϕ_{NA}, and ϕ_B, such that*

(5.5.1)

$$h_Y(P_1,P_2) \ge -h(\mathscr{F}) - c_2(i_1^*h_A(P_1) + i_2^*h_A(P_2)) - c_3(i_1^* + i_2^*) - (1+o(1))(\delta_1+\delta_2)\ .$$

The proof of this is quite technical, so it will not be included here. The interested reader is referred to [**Bom90**, Lemma 6] for a detailed proof.

6. Construction of a Global Section

This section gives Bombieri's construction of a global section of $\mathscr{O}(Y)$ with bounds on its height $h(\mathscr{F})$. This is the core of Bombieri's additions to earlier proofs, as it replaces the use of advanced Arakelov theory with arguments using more classical algebraic geometry.

LEMMA 6.1. *Let $\gamma > 0$ and let a_1 and a_2 be positive rational numbers with*

$$(1-\gamma)a_1a_2 > g\ .$$

Let Y be the divisor defined by (2.10) for a suitable $d > 0$. Then there exist a nonzero global section s of $\mathscr{O}(Y)$; a constant c_4 depending only on C, ϕ_{NA}, and ϕ_B; and a representation $\mathscr{F} = (F_i)_{0\le i\le m}$ of s, such that

$$h(\mathscr{F}) \le c_4(d_1+d_2)/\gamma + o(d_1+d_2)\ . \tag{6.1.1}$$

PROOF. The section s will be constructed by finding $\mathscr{F}$ that satisfies the conditions of Lemma 5.1.

By the Riemann-Roch theorem for projective algebraic surfaces,

$$\begin{aligned} h^0(C\times C, Y) &= d^2(a_1a_2-g)+O(d_1+d_2+d) \\ &\ge d^2(a_1a_2\gamma)-O(\delta_1+\delta_2)\ . \end{aligned} \tag{6.1.2}$$

By Riemann-Roch for curves,

$$h^0(C\times C, \delta_1 NA\times C + C\times\delta_2 NA) = (N\delta_1+1-g)(N\delta_2+1-g)\ ,$$

and therefore the space of all possible $\mathscr{F}$ has dimension

$$(m+1)(N^2\delta_1\delta_2 - N(g-1)(\delta_1+\delta_2)+(g-1)^2)\ . \tag{6.1.3}$$

(Here (6.1.2) and (6.1.3) are close to the values of $N-M$ and N, respectively, in the application of Siegel's lemma; see the details later in this proof.)

We will work using (local) affine coordinates $\xi_j := (x_j/x_0)\circ p_1$, $\xi'_j := (x_j/x_0)\circ p_2$, and $\eta_i := (y_i/y_0)\big|_{C\times C}$ on $C\times C$.

For suitably generic choice of ϕ_{NA} (which we assume has been chosen), the rational map $\pi_1\colon \mathbb{P}^n \dashrightarrow \mathbb{P}^1$ given by $[x_0:x_1:\dots:x_n]\mapsto[x_0:x_1]$ is a morphism on $\phi_{NA}(C)$. Likewise, we assume that ϕ_{NA} has been chosen such that the projection $\pi_{1,2}\colon [x_0:x_1:\dots:x_n]\mapsto[x_0:x_1:x_2]$ maps $\phi_{NA}(C)$ birationally to its image in $\mathbb{P}^2$ (which is therefore a curve of degree N), and ξ_2 is integral over the ring $k[\xi_1]$.

Since the morphism $\phi_{NA}\times\phi_{NA}\colon C\times C\to\mathbb{P}^n\times\mathbb{P}^n$ is a closed immersion, composing with $\pi_{1,2}\times\pi_{1,2}$ gives a birational map from $C\times C$ to its image in $\mathbb{P}^2\times\mathbb{P}^2$, so we have

$$K(C\times C) = k(\xi_1,\xi_2,\xi'_1,\xi'_2)$$

and therefore, for all i,

$$\eta_i = \frac{P_i(\xi_1,\xi_2,\xi'_1,\xi'_2)}{Q_i(\xi_1,\xi_2,\xi'_1,\xi'_2)}$$

for polynomials P_i and Q_i with coefficients in k. The conditions $F_i/y_i^d = F_j/y_j^d$ then become

$$((P_iQ_j)^dF_j)\big|_{C\times C} = ((P_jQ_i)^dF_i)\big|_{C\times C} \qquad \text{for all } i,j\ .$$

Note that the height of $(P_iQ_j)^d$ is $O(d)$ for all i,j.

In order to apply Siegel's lemma, it is necessary to find a basis for the set of all possible $\mathscr{F}$ such that the linear conditions relative to this basis do not become much larger than the heights of $(P_iQ_j)^d$. The easiest way to do this is to use a linear subspace of slightly smaller dimension. This will change N and $N-M$ in Siegel's lemma by $O(\delta_1+\delta_2)$, which is small enough not to change the outcome appreciably.

The details of this are as follows. The image of the morphism $\pi_{1,2}\circ\phi_{NA}\colon C\to\mathbb{P}^2$ is a curve of degree N, and is birational to C. Moreover, the field extension $K(C)/k(\xi_1)$ is also of degree N, generated by ξ_2. We then restrict our choices of the F_i to the space generated by the monomials

$$x_0^{\delta_1-u_1-w_1}x_1^{u_1}x_2^{w_1}(x'_0)^{\delta_2-u_2-w_2}(x'_1)^{u_2}(x'_2)^{w_2}\ , \qquad \begin{aligned} &0\le w_1\le N\ , \quad 0\le u_1\le\delta_1-w_1\ , \\ &0\le w_2\le N\ , \quad 0\le u_2\le\delta_2-w_2\ . \end{aligned}$$

The number of ordered pairs (w_1, u_1) is

$$\begin{aligned}(\delta_1+1)+(\delta_1+1-1)+\cdots+(\delta_1+1-(N-1)) &= N\delta_1+N-\frac{N(N-1)}{2}\\ &= N\left(\delta_1-\frac{N-3}{2}\right)\ ,\end{aligned}$$

and similarly for the number of pairs (w_2, u_2). Therefore the dimension of this subspace is

$$(m+1)N^2\left(\delta_1-\frac{N-3}{2}\right)\left(\delta_2-\frac{N-3}{2}\right)\ ,$$

which differs from (6.1.3) by $O(\delta_1+\delta_2)$.

With the setup described above, the height of the linear forms remains $O(d)$, and therefore Siegel's lemma constructs $\mathscr{F}$ satisfying the conditions of Lemma 5.1 with

$$\begin{aligned}h(\mathscr{F}) &\lessapprox O\left(d\cdot\frac{(m+1)N^2\delta_1\delta_2}{\gamma d_1 d_2}\right)\\ &= O\left(\frac{d_1+d_2}{\gamma}\right)\ .\end{aligned}$$

(Recall that $N\delta_i = d(a_i+s)$ and $d_i = da_i$ $(i=1,2)$. Then $a_1\to\infty$ and $a_2\to 0$ as $h(P_2)/h(P_1)\to\infty$; therefore $N\delta_1/d_1\to$ constant and $N\delta_2/d_2$ grows like $h(P_2)/h(P_1)$ as $h(P_2)/h(P_1)\to\infty$.) □

The above proof omits many of the details that determine the bounds on $h(\mathscr{F})$. For a complete proof see [**Bom90**, § 8].

7. The Index

Since the inequality (5.5.1) will be applied when $h(P_2)/h(P_1)$ and $h(P_1)$ are large the effect of an increase in i_2^* will be much greater than the effect of the same increase in i_1^*. Therefore it will be useful to work with a measure of vanishing of s at (P_1, P_2) that assigns more weight to ζ_2 than to ζ_1.

This is consonant with the fact that the ratio δ_1/δ_2 is large, which stems from the fact that a_1 and a_2 are chosen such that their ratio a_1/a_2 is close to $h_A(P_2)/h_A(P_1)$ (see (2.8)). This ensures that the two components of the term $c_2(i_1^* h_A(P_1)+i_2^* h_A(P_2))$ in (5.5.1) have approximately the same size.

With this in mind, the Thue-Siegel method relies heavily on the "index" of a polynomial or a global section, which may be regarded as a weighted order of vanishing at a point.

Definition 7.1. Let k be a field, let d_1 and d_2 be positive integers, let $f\in k[x_1,x_2]$ be a nonzero polynomial, and let $\alpha_1,\alpha_2\in k$. Write

$$f(x_1,x_2)=\sum_{i_1,i_2\in\mathbb{N}} a_{i_1,i_2}(x_1-\alpha_1)^{i_1}(x_2-\alpha_2)^{i_2}\ .$$

Then the **index** of f at (α_1,α_2) relative to (d_1,d_2) is the (rational) number

$$\begin{aligned}\operatorname{ind}_{d_1,d_2}(f,(\alpha_1,\alpha_2)) &= \min\left\{\frac{i_1}{d_1}+\frac{i_2}{d_2} : i_1,i_2\in\mathbb{N},\ a_{i_1,i_2}\neq 0\right\}\\ &= \min\left\{\frac{i_1}{d_1}+\frac{i_2}{d_2} : i_1,i_2\in\mathbb{N},\ \left(\frac{\partial}{\partial x_1}\right)^{i_1}\left(\frac{\partial}{\partial x_2}\right)^{i_2} f(\alpha_1,\alpha_2)\neq 0\right\}\ ,\end{aligned}$$

where the second expression is valid only if $\operatorname{char} k = 0$.

This definition is used in the proofs of Thue and Siegel. In those proofs, d_1/d_2 is taken close to $(h(p_1/q_1)/h(p_2/q_2))^{-1}$ (with notation as in Section 1), so that $d_i h(p_i/q_i)$ is approximately independent of i. In addition, $|q_1|^{d_1} \approx |q_2|^{d_2}$, so since f has degree $\le d_1$ in x_1 and degree $\le d_2$ in x_2, the contributions to the denominators from each variable in the expression $f(p_1/q_1, p_2/q_2)$ are approximately the same.

In Bombieri's (and my) proof of the Mordell conjecture, $d_1 = a_1 d$ and $d_2 = a d_2$, which are approximately the intersection numbers of the divisor Y with fibers $A \times C$ and $C \times A$, respectively.

In the above definition, f may be a power series, and may be further extended to the index of a global section of $\mathscr{O}(Y)$ at (P_1, P_2) (by dividing by a local generator of $\mathscr{O}(Y)$ at (P_1, P_2) and expressing the quotient as a power series in local coordinates on the factors). This is easily seen to be well defined, for the same reason that admissibility is well defined.

8. The End of the Proof

So far we have constructed a global section s of $\mathscr{O}(Y)$ whose height $h(\mathscr{F})$ is bounded by (6.1.1), and obtained upper and lower bounds for the height $h_Y(P_1, P_2)$.

In more detail, combining (5.5.1) with (6.1.1) gives the lower bound

$$h_Y(P_1,P_2) \ge -c_4\left(\frac{d_1+d_2}{\gamma}\right) - c_2(i_1^* h_A(P_1) + i_2^* h_A(P_2)) - c_3(i_1^*+i_2^*) - (1+o(1))(\delta_1+\delta_2)\ .$$

Recall that the upper bound (4.2) is

$$h_Y(P_1,P_2) \le \frac{d_1}{g}|j(P_1)|^2 + \frac{d_2}{g}|j(P_2)|^2 - \frac{3}{2}d|j(P_1)||j(P_2)| + c_1(d_1+d_2+d)\ .$$

Combining these two bounds then gives

$$\begin{aligned} &c_2 \max\{d_1 h_A(P_1), d_2 h_A(P_2)\}\left(\frac{i_1^*}{d_1} + \frac{i_2^*}{d_2}\right) \\ &\qquad \ge -\left(\frac{d_1}{g}|j(P_1)|^2 + \frac{d_2}{g}|j(P_2)|^2 - \frac{3}{2}d|j(P_1)||j(P_2)|\right) \\ &\qquad\qquad - c_3(i_1^*+i_2^*) - c_4\left(\frac{d_1+d_2}{\gamma}\right) - O(d_1+d_2)\ . \end{aligned} \tag{8.1}$$

The first thing to notice about this inequality is that the quantity $i_1^*/d_1 + i_2^*/d_2$ on the left-hand side is the index $\operatorname{ind}_{d_1,d_2}(s,(P_1,P_2))$ of s at (P_1,P_2) relative to d_1 and d_2 (for suitable choices of i_1^* and i_2^*, which we now assume).

Another important fact about (8.1) is that all of the terms on the right-hand side grow linearly with d, as does the coefficient $c_2 \max\{d_1 h_A(P_1), d_2 h_A(P_2)\}$ on the left. Therefore, (8.1) implies a lower bound on the index that does not depend on d.

Of course, such a lower bound is not useful if it is negative. However, since $|j(P)|^2 = \hat{h}_\Theta(j(P)) = g\, h_A(P) + O(1)$ for all $P \in C(\bar{k})$,

(8.2)

$$-\frac{d_1}{g}|j(P_1)|^2 - \frac{d_2}{g}|j(P_2)|^2 + \frac{3}{2}d|j(P_1)||j(P_2)| \approx dg\sqrt{h_A(P_1)h_A(P_2)}\left(-2\frac{\sqrt{g+\gamma}}{g} + \frac{3}{2}\right)\ .$$

Since $g \ge 2$ (this is where the assumption on the genus is used), one can take $\gamma > 0$ sufficiently small so that $\sqrt{g+\gamma}/g < 3/4$, so this term is positive. Moreover, it grows at the same rate as $\max\{d_1 h_A(P_1), d_2 h_A(P_2)\}$.

Thus, if $h_A(P_1)$ and $h_A(P_2)$ are sufficiently large, then the last three terms of (8.1) are insignificant, and we have a lower bound

$$\operatorname{ind}_{d_1,d_2}(s,(P_1,P_2)) \ge \epsilon \;, \tag{8.3}$$

where $\epsilon > 0$ depends only on C, N, A, ϕ_{NA}, ϕ_B, etc., but not on P_1, P_2, or d.

The final contradiction will be obtained using *Roth's lemma:*

THEOREM 8.4. *Let f be a nonzero polynomial in m variables $x_1,\dots,x_m$, of degree at most $r_i \in \mathbb{Z}_{>0}$ in x_i for each i, and with coefficients in $\overline{\mathbb{Q}}$. Let $(b_1,\dots,b_m)$ be an algebraic point. Finally, let $\epsilon > 0$ be such that*

$$\frac{r_{i+1}}{r_i} \le \epsilon^{2^{m-1}}\;, \qquad \text{for all } i = 1,\dots,m-1 \tag{8.4.1}$$

and

$$r_i h(b_i) \ge \epsilon^{-2^{m-1}}(h(P) + 2mr_1)\;, \qquad \text{for all } i = 1,\dots,m\;. \tag{8.4.2}$$

Then the index of f at $(b_1,\dots,b_m)$ satisfies

$$\operatorname{ind}_{r_1,\dots,r_m}(P,(b_1,\dots,b_m)) \le 2m\epsilon\;.$$

This lemma appears (with proof) in [**HS00**, Prop. D.6.2] (and in many other places), and will not be proved here.

Of course we will use $m = 2$ in applying Theorem 8.4, and will also need to adjust ϵ in (8.3). But first, it will be necessary to address the fact that f is a polynomial in two variables (which can be regarded as a global section of $\mathscr{O}(r_1,r_2)$ on $\mathbb{P}^1 \times \mathbb{P}^1$), whereas s is a global section of a line sheaf on $C \times C$.

This difference can be handled as follows. For suitable a and b, fix finite maps $C \to \mathbb{P}^1$ given by the rational functions x_a/x_0 and x_b/x_0, and let $g = F_0/\eta_0^d$, which has poles only along $x_0 = 0$ and $x_0' = 0$. Now, recalling that $\xi_a = (x_a/x_0)\circ p_1$ and $\xi_b' = (x_b/x_0)\circ p_2$, take the norm

$$f = N^{K(C\times C)}_{k(\xi_a,\xi_b')} g$$

to get a section of $\mathscr{O}(N^2\delta_1, N^2\delta_2)$ on $\mathbb{P}^1 \times \mathbb{P}^1$. This is the polynomial f, and one can bound its height by

$$h(P) \le O\left(\frac{d_1+d_2}{\gamma}\right) + O(d_1+d_2)\;;$$

see [**Bom90**, § 9].

One will have $b_1 = \xi_a(P_1)$ and $b_2 = \xi_b(P_2)$, so $h(b_1) = h_{NA}(P_1) + O(1)$ and $h(b_2) = h_{NA}(P_2) + O(1)$ by elementary properties of heights [**HS00**, Thm. B.3.2(b) or Cor. B.2.6]. Therefore, one can choose $P_1 \in \Sigma'$ with $h_A(P_1)$ sufficiently large so that (8.3) and (8.4.2) hold, and subsequently choose $P_2 \in \Sigma'$ with $h_A(P_2)/h_A(P_1)$ sufficiently large so that (8.4.1) is true.

This then gives a contradiction, so the assumption that $C(k)$ is infinite must be false.

Note that much has been left out of the proof, most notably the derivations of the bounds of the heights in the application of Siegel's lemma, as well as the height inequalities in this section. The interested reader is referred to [**Bom90**] for more explicit details.

References

[Bom90] Enrico Bombieri, *The Mordell conjecture revisited*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **17** (1990), no. 4, 615–640. MR 1093712 (92a:11072)

[Bom91] ———, *Errata-corrige: "The Mordell conjecture revisited" [Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* **17** *(1990), no. 4, 615–640; MR1093712 (92a:11072)]*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **18** (1991), no. 3, 473. MR 1145319 (92k:11064)

[Dys47] F. J. Dyson, *The approximation to algebraic numbers by rationals*, Acta Math. **79** (1947), 225–240. MR 0023854 (9,412h)

[Fal83] Gerd Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366. MR 718935 (85g:11026a)

[Fal91] ———, *Diophantine approximation on abelian varieties*, Ann. of Math. (2) **133** (1991), no. 3, 549–576. MR 1109353 (93d:11066)

[Fal94] ———, *The general case of S. Lang's conjecture*, Barsotti Symposium in Algebraic Geometry (Abano Terme, 1991), Perspect. Math., vol. 15, Academic Press, San Diego, CA, 1994, pp. 175–182. MR 1307396 (95m:11061)

[Gel52] A. O. Gel′fond, *Transcendentnye i algebraičeskie čisla*, Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, 1952, [English translation: *Transcendental and algebraic numbers,* Dover, New York, 1960]. MR 0057921 (15,292e)

[GL48] A. O. Gel′fond and Yu. V. Linnik, *On Thue's method in the problem of effectiveness in quadratic fields*, Doklady Akad. Nauk SSSR (N.S.) **61** (1948), 773–776. MR 0027783 (10,354f)

[Har77] Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York-Heidelberg, 1977, Graduate Texts in Mathematics, No. 52. MR 0463157 (57 #3116)

[HS00] Marc Hindry and Joseph H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics, vol. 201, Springer-Verlag, New York, 2000, An introduction. MR 1745599 (2001e:11058)

[Lan83] Serge Lang, *Fundamentals of Diophantine geometry*, Springer-Verlag, New York, 1983. MR 715605 (85j:11005)

[Lio44] J. Liouville, *Sur des classes très étendues de quantités dont la valeur n'est ni algébrique ni même réductible à des irrationnelles algébriques*, C. R. Acad. Sci. Paris **18** (1844), 883–885 and 910–911.

[Man63] Ju. I. Manin, *Rational points on algebraic curves over function fields*, Izv. Akad. Nauk SSSR Ser. Mat. **27** (1963), 1395–1440. MR 0157971 (28 #1199)

[Mil86] J. S. Milne, *Jacobian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 167–212. MR 861976

[Mum65] David Mumford, *A remark on Mordell's conjecture*, Amer. J. Math. **87** (1965), 1007–1016. MR 0186624 (32 #4083)

[Rot55] K. F. Roth, *Rational approximations to algebraic numbers*, Mathematika **2** (1955), 1–20; corrigendum, 168. MR 0072182 (17,242d)

[Sch72] Wolfgang M. Schmidt, *Norm form equations*, Ann. of Math. (2) **96** (1972), 526–551. MR 0314761 (47 #3313)

[Sch80] ———, *Diophantine approximation*, Lecture Notes in Mathematics, vol. 785, Springer, Berlin, 1980. MR 568710 (81j:10038)

[Sie21] C. L. Siegel, *Approximation algebraischer Zahlen*, Math. Zeit. **10** (1921), 173–213.

[Thu09] A. Thue, *Über Annäherungswerte algebraischer Zahlen*, J. reine angew. Math. **135** (1909), 284–305.

[Voj89] Paul Vojta, *Mordell's conjecture over function fields*, Invent. Math. **98** (1989), no. 1, 115–138. MR 1010158 (90k:11076)

[Voj91] ———, *Siegel's theorem in the compact case*, Ann. of Math. (2) **133** (1991), no. 3, 509–548. MR 1109352 (93d:11065)

Department of Mathematics, University of California, 970 Evans Hall #3840, Berkeley, CA 94720-3840

Email address: vojta@math.berkeley.edu