The average elliptic curve has few integral points.

Levent Hasan Ali Alpöge.

Acknowledgements. First and foremost, I would like to thank my family. Of course I cannot put into words how much their support has meant to me. It was my sister who first taught me arithmetic during long car rides when I was very young, and, as is evidenced by this thesis, I still am just as excited about it all these years later.

Next I would like to thank my advisor, Jacob Tsimerman, for teaching me a tremendous amount of mathematics and always having a healthy dose of skepticism when listening to my arguments. I am greatly indebted to him and Arul Shankar for coming up with such a fun problem. I don't know how they did it, but I am lucky they are so skilled.

I would like to thank all my teachers, in mathematics, physics, computer science, from Harvard, high school, LISG, and otherwise, for teaching me patiently for all these years. In particular I would like to thank Prof.s Gaitsgory, Elkies, Gross, Miller, and Gallian for advising me during my time at Harvard, Williams, and Duluth. If it weren't for Prof. Gallian's constant encouragement I would have never even considered asking for a problem. Actually I really must thank all my professors at Harvard as well, but nobody wants to read a listing of the entire Harvard faculty in a senior thesis.

Finally, I would like to thank my friends for being awesome. Obviously you guys know who you are.

Also, more formally, I would like to thank Manjul Bhargava, Peter Bruin, Noam Elkies, Henry Cohn, John Cremona, Roger Heath-Brown, Harald Helfgott, Emmanuel Kowalski, Barry Mazur, Arul Shankar, Joseph Silverman, Katherine Stange, Jacob Tsimerman, and Yukihiro Uchida for helpful conversations related to work in this thesis.

I would of course also like to thank Anirudha Balasubramanian for helpful conversations related to the Irwin-Hall distribution.[2]

CHAPTER 1 INTRODUCTION

Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.
— Leopold Kronecker.

1. History.

In August 1659, in his correspondence with Carcavi, Pierre de Fermat proposed the following four problems. **[15]**

(1) *Il n'y a aucun cube divisible en deux cubes.*[1]
(2) *Il n'y a qu'un seul quarré en entiers qui, augmenté du binaire, fasse un cube. Le dit quarré est* 25.[2]
(3) *Il n'y a que deux quarrés en entiers, lesquels, augmentés de* 4*, fessent un cube. Les dits quarrés sont* 4 *et* 121.[3]
(4) *Toutes les puissances quarrées de* 2*, augmentées de l'unité, sont nombres premiers.*[4]

Statement (4) is false. Statement (1) is the first case of a problem that remained open for more than three-hundred and fifty years. In this thesis we will focus on statements (2) and (3).

These are statements about integer solutions to certain equations. Statement (2) asks about the number of integer solutions to $y^2 = x^3 - 2$. Statement (3) asks about the number of integer solutions to $y^2 = x^3 - 4$. The similarity of the forms of these equations is no coincidence. Fermat had developed his method of *descent*, in which one takes a putative minimal counterexample and produces an even smaller one to derive a contradiction, and applied it to each of these cases.

Earlier on in his letter to Carcavi, Fermat mentioned that

[1] *There is no cube divisible into two cubes.* **[1]**

[2] *There is only one square of whole numbers which, when two is added to it, makes a cube. Said square is* 25. **[1]**

[3] *There are only two squares of whole numbers which, when* 4 *is added, make a cube. Said squares are* 4 *and* 121. **[1]**

[4] *All the square powers of two, with one added, are prime numbers.* **[1]** By this, Fermat surely meant iterated powers of two. But both are false. $2^{3^2} + 1 = 513$ is certainly divisible by 3. More subtly, $2^{2^5} + 1 = 4294967297$ is divisible by 641.

> J'appelai cette manière de démontrer la *descente infinie* ou *indèfinie*, etc.; je ne m'en servis au commencement que pour démontrer les propositions négatives, comme, par exemple ...

(5) *Qu'il n'y a aucun triangle rectangle en nombres dont l'aire soit un nombre quarré.*[5]

Here he is asking about integer solutions to the simultaneous equations

$$a^2 + b^2 = c^2 \text{ (the Pythagorean theorem),} \tag{1.1}$$

$$ab = 2n^2 \text{ (the area of a triangle).} \tag{1.2}$$

Fermat knew that any solution of (1.1) and (1.2) in the integers must be of the form

$$\begin{aligned} a &= p^2 - q^2, \\ b &= 2pq, \\ c &= p^2 + q^2, \end{aligned}$$

for some integers p and q. Now let $x := \frac{p}{q}, y := \frac{n}{q^2}$. Then observe that $y^2 = x^3 - x$, by (1.2). In reverse, if $y^2 = x^3 - x$ with x and $y \neq 0$ rational, then let

$$\begin{aligned} A &:= \frac{x^2 - 1}{y}, \\ B &:= \frac{2x}{y}, \\ C &:= \frac{x^2 + 1}{y}, \end{aligned}$$

all rational numbers. Write them with a common denominator n, and let a, b, c be their respective numerators. Then observe that $a^2 + b^2 = c^2$, and $ab = 2n^2$. This establishes a bijection between rational solutions to $y^2 = x^3 - x$ with $y \neq 0$ and solutions of Fermat's problem (5). Therefore Fermat's achievement lay in showing that there were no such rational solutions, again by his method of descent.

2. This thesis.

It is again no coincidence that the equation $y^2 = x^3 - x$ shares the same form as the previously considered equations $y^2 = x^3 - 2$ and $y^2 = x^3 - 4$. These are the equations of *elliptic curves*, of the form

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

with A and B integral. In this thesis we consider the question of how many integral solutions such an equation can have. Specifically, we show that, on average, it is at most an effective absolute constant.[6] It is an old folklore conjecture that the true average is zero. This is the first time the average has been proved to be finite. This work is original to this thesis.

[5] *I have called this manner of demonstration* infinite descent, *or* indefinite descent, *etc.; at first I only used it to demonstrate negative propositions, such as ... that there is no right triangle in whole numbers whose area is a square number.* [**1**]

[6] By "on average," we mean the average taken when ordering by the *height* of the equation, $H(A, B) := \max(4|A|^3, 27|B|^2)$. Further, we always mean "lim sup of the average" when we say "average." See Chapter 5 for precise details.

To consider an average, one must first know that the numbers being averaged are finite. Therefore we present a proof of a famous theorem of Siegel stating that, indeed, these equations admit only finitely many integer solutions. This requires the development of tools from the theory of Diophantine approximation, chief among them the equally famous theorem of Roth determining how well an algebraic number can be approximated by rationals. Each of these rests on an understanding of the theory of heights, and so we explain the necessary results throughout, beginning with a Chapter dedicated entirely to the Weil height. The presentation is meant for a reader wholly unfamiliar with the tools of the field, and so we provide explanations with the goal of *exposition*, rather than taking the shortest path to presenting the titular result.

3. Notation.

Let us fix notation at the outset. By $A \ll_\theta B$ we mean that there exists a constant $C_\theta > 0$, potentially depending only on θ, such that $|A| \leq C_\theta |B|$. By $A \leq O_\theta(B)$ we mean $A \ll_\theta B$. By $A \leq o_\theta(B)$, with an implicit limit $n \to \infty$ understood, we mean that, for every $\epsilon > 0$, there is an N depending only on θ and ϵ for which, for every $n \geq N$, one has $A \leq \epsilon B$ as functions of n. In reverse, $A \gg_\theta B$ means $B \ll_\theta A$, and $A \geq \Omega_\theta(B)$ means $A \gg_\theta B$. Finally, $A \asymp_\theta B$ means $A \ll_\theta B$ and $A \gg_\theta B$. We will also list notations introduced later at the end of the thesis, for the reader's convenience.

Having set our notation, we will begin with the theory of heights.

CHAPTER

2 THE WEIL HEIGHT

1. Motivation.

We have mentioned Fermat's method of descent, which proceeds by contradiction, beginning with a minimal counterexample and producing one yet smaller. One delicate point is making precise what one means by *minimal*. For this one needs a measure of complexity of solutions to equations. In the case of integral solutions one such measure is furnished for free: given a solution $P = (x, y)$ to, e.g., $y^2 = x^3 + Ax + B$, take $H_x(P) := |x|$ as the "complexity" of the point P.

This is all well and good, but we have seen that rational solutions arise naturally as well. We might proceed naïvely and put, given $P = (x, y)$, $H_x(P) = |x|$ again as a measure of complexity. But this would have two problems. First, we would no longer be able to deduce a contradiction from Fermat's method of descent: there are strictly decreasing sequences of positive rational numbers uniformly bounded below by a positive constant. Second, this would not even agree with our intuitive notion of "complexity"!

Specifically, consider the curve **[12]**

$$y^2 = x^3 + 3917x.$$

This has a rational point ("smallest", in some sense) with x-coordinate equal to

$$\frac{1.319\ldots \times 10^{70}}{5.488\ldots \times 10^{68}}.$$

The reader will agree that this is a point of enormous complexity. However, its absolute value is $24.01\ldots$. So, by our measure, it is less complicated than an integral point with x-coordinate 25. Can this truly be considered an adequate measure of complexity?

2. The multiplicative and logarithmic Weil heights.

For this reason we propose the following variation. Given a rational number $\alpha = \frac{a}{b}$ with a and b coprime, define $H(\alpha) := \max(|a|, |b|)$. This is the so-called (multiplicative) *naïve*, or *Weil*, *height* of α. Note that this has been constructed both so that the point we mentioned above has enormous height, and also so that it extends our definition on nonzero integers. But now we are led to complain that this is a nonlocal definition — if we know the prime factorizations of a and b, then it seems unlikely we will be able to read off the multiplicative height of $\frac{a}{b}$.

Amazingly, our concerns are unfounded.

CLAIM 2.1. *Let $\alpha \in \mathbb{Q}$ be nonzero. Then*

$$H(\alpha) = \prod_v \max(|\alpha|_v, 1),$$

the product taken over the places[1] *of $\mathbb{Q}$.*

Therefore, to determine the height of α, we simply need to know the "sizes" of α in all completions of $\mathbb{Q}$: $\mathbb{R}$, and $\mathbb{Q}_p$ for each p. Motivated by this claim and the ubiquity of the function $\log^+$ in complex analysis, we take logarithms of both sides and define:

DEFINITION 2.2. *Let $\alpha \in \mathbb{Q}$ be nonzero. Then the (logarithmic)* Weil height *of α is*

$$h(\alpha) := \sum_v \log^+ |x|_v,$$

where $\log^+(x) := \max(0, \log x)$.

Thus the content of the claim is that $h(\alpha) = \max(\log|a|, \log|b|)$ if $\alpha = \frac{a}{b}$ with $(a, b) = 1$.

PROOF OF CLAIM 2.1. Write $\alpha = \frac{a}{b}$ with $(a, b) = 1$. Factor $a = \pm \prod_p p^{v_p(a)}$ and $b = \prod_p p^{v_p(b)}$. Then

$$|\alpha|_\infty = \frac{|a|}{b},$$

and, for each prime p,

$$|\alpha|_p = p^{v_p(b) - v_p(a)}.$$

Moreover, since a and b are coprime, $v_p(b) - v_p(a) = v_p(b)$ or $-v_p(a)$ for all p. Therefore $\max(|\alpha|_p, 1) = p^{v_p(b)}$ if $p|b$, and 1 otherwise. In particular

$$\prod_p \max(|\alpha|_p, 1) = b.$$

Hence the right-hand side is

$$\max(|\alpha|_\infty, 1) \cdot b = \max(|a|, |b|),$$

as desired. □

Thus we have extended our notion of complexity to the rationals. We are not yet done, since we will need to measure general algebraic numbers against rationals in due course, but we are close. Because Claim 2.1 reduces the issue to a local problem in the case of the rationals, we are led to consider piecing together local information in the case of a general algebraic number. After all, we understand the local behaviour of finite-degree extensions of the rationals quite well. The only

[1] That is, pairwise inequivalent nontrivial absolute values of $\mathbb{Q}$. These are given by the usual ("Archimedean", in the sense that the integers form an unbounded set) absolute value, and the p-adic absolute values for all primes p.

question is how to normalize so that our notion of height does not change under embeddings $K \subseteq L$ of subfields. This is solved by the following definition.

DEFINITION 2.3. *Let K be a number field. Let $\alpha \in K$ be a nonzero element. The* absolute Weil height *of α is*

$$h(\alpha) := \sum_v \frac{[K_v : \mathbb{Q}_{v|_\mathbb{Q}}]}{[K : \mathbb{Q}]} \log^+ |\alpha|_v,$$

the sum taken over the places v of K.

That is to say, the sum is taken over nonconjugate embeddings $K \hookrightarrow \mathbb{R}$ or $K \hookrightarrow \mathbb{C}$ (that is, pairwise inequivalent Archimedean absolute values on K extending that of $\mathbb{Q}$) and prime ideals $\mathfrak{p} \subseteq \mathfrak{o}_K$ of the ring of integers of K. The former induce absolute values simply by restriction, and the latter induce absolute values via $|x|_\mathfrak{p} := \mathcal{N}\mathfrak{p}^{-v_\mathfrak{p}(x)}$, where $v_\mathfrak{p}(x)$ is the power of $\mathfrak{p}$ appearing in the unique factorization of the fractional ideal $(x) \subseteq K$ generated by x, and $\mathcal{N}\mathfrak{p} := \#|\mathfrak{o}_K/\mathfrak{p}|$. Such a thing of course determines a place of $\mathbb{Q}$, by restriction — this is what we mean by the notation $v|_\mathbb{Q}$ above.

The purpose of the weights $\frac{[K_v:\mathbb{Q}_{v|_\mathbb{Q}}]}{[K:\mathbb{Q}]}$ is to make the height invariant under change of field — that is, independent of the field inside which we regard α. Indeed they do work for this purpose, per the following.

CLAIM 2.4. *Let $K \subseteq L$ be an extension of number fields, and $\alpha \in K$. Then*

$$\sum_v \frac{[K_v : \mathbb{Q}_{v|_\mathbb{Q}}]}{[K : \mathbb{Q}]} \log^+ |\alpha|_v = \sum_w \frac{[L_w : \mathbb{Q}_{w|_\mathbb{Q}}]}{[L : \mathbb{Q}]} \log^+ |\alpha|_w,$$

the sums taken over places v of K and w of L, respectively.

PROOF OF CLAIM 2.4. Let w be a place of L. Let v be its restriction to K, and u its restriction to $\mathbb{Q}$. Recall that we obtain all places of K in this way — given a place v of K, there is at least one place w of L extending v (simply embed L in an algebraic closure of K_v and take its closure — writing L_w/K_v for the field extension thus generated, $|x|_w := |\mathrm{Nm}_{L_w/K_v}(x)|_v^{\frac{1}{[L_w:K_v]}}$ works).

Therefore we may reorganize the sum over w into a sum over v, and then a sum over w extending v. That is, the right-hand side is just

$$\sum_v \frac{[K_v : \mathbb{Q}_u]}{[K : \mathbb{Q}]} \sum_{w|v} \frac{[L_w : K_v]}{[L : K]} \log^+ |\alpha|_w.$$

By definition $|\alpha|_w = |\alpha|_v$ since w extends v. Thus we may simplify this to

$$\sum_v \frac{[K_v : \mathbb{Q}_u]}{[K : \mathbb{Q}]} \log^+ |\alpha|_v \sum_{w|v} \frac{[L_w : K_v]}{[L : K]}.$$

But the fact that $\sum_{w|v}[L_w : K_v] = [L : K]$ is classical (after all, $L \otimes_K K_v \simeq \prod_{w|v} L_w$). □

Thus we have a height function on $\bar{\mathbb{Q}}$. For our purposes, in fact this is enough. But let us note two amusing properties of the absolute Weil height before we move to putting it into use. For the reader just becoming acquainted with heights, rest assured that they, as with any tool in mathematics, are much more easily understood when actually used in a calculation or sufficiently complicated proof.

3. Properties of the Weil height.

The first amusing (but structurally important) property is the so-called *Northcott property*.

PROPOSITION 3.1 (Northcott property of the Weil height.). *Let $C > 0$. The number of nonzero algebraic numbers $\alpha \in \bar{\mathbb{Q}}$ with $h(\alpha) < C$ and $\deg(\alpha) < C$ is finite.*

Therefore, for instance, the Weil height is an appropriate measure of complexity for Fermat's method of descent.

PROOF OF PROPOSITION 3.1. Note that it is certainly possible to have zero height — e.g., $h(1) = 0$. Thus we must first show that there are only finitely many algebraic numbers of *zero* height and bounded degree. But these are just the roots of unity of bounded degree — indeed, if $h(\zeta) = 0$, then $|\zeta| \leq 1$ under all absolute values of $\mathbb{Q}(\zeta)$, whence, by a classical theorem of Kronecker, ζ is a root of unity. (Proof: Note that ζ is an algebraic integer. Let $K/\mathbb{Q}$ be the Galois closure of $\mathbb{Q}(\zeta)$. This is of degree at most $C!$. Observe that all absolute values of the first $\deg(\zeta)$ symmetric polynomials in the $\sigma\zeta, \sigma \in \mathrm{Gal}(K/\mathbb{Q})$, are uniformly bounded in terms of C, by our hypothesis. These are all integers, whence there are finitely many of them. Since the hypothesis holds for all $\zeta^N, N \in \mathbb{Z}^+$, by the pigeonhole principle we find $a > b > 0$ and σ such that $\zeta^a = \sigma(\zeta^b)$. Thus $\zeta^{a^{[K:\mathbb{Q}]} - b^{[K:\mathbb{Q}]}} = 1$. Thus the claim. Note also that by the product formula $|\zeta| = 1$ at all places, but we didn't need this here.)

Now if $h(\alpha) < C$, let $d := \deg(\alpha)$, and write $\alpha = \alpha_1, \ldots, \alpha_d$ (respectively, $a_0 \in \mathbb{Z}^+$) for the roots (respectively, leading coefficient) of the minimal polynomial of α over $\mathbb{Q}$. The discriminant of $\mathbb{Q}(\alpha)$ divides $\Delta := a_0^{2d-2} \prod_{i<j} (\alpha_i - \alpha_j)^2 \in \mathbb{Z}$. The height of Δ is certainly at most $2C \log |a_0| + 2 \sum_{i<j} h(\alpha_i - \alpha_j)$. But, via the Archimedean and non-Archimedean triangle inequalities at the infinite and finite places, respectively, in general one has $h(x + y) \leq \log 2 + h(x) + h(y)$, the $\log 2$ coming from the infinite places (specifically, $|x + y| \leq 2 \max(|x|, |y|)$, whereas the finite places have no factor of 2). Thus

$$h(\Delta) \ll_C \log |a_0|.$$

Since the greatest common divisor of the coefficients of the minimal polynomial of α is 1, we must have $\log |a_0| \leq h(\sigma_k(\alpha_1, \ldots, \alpha_d)) \ll_C 1$ for some symmetric polynomial $\sigma_k, 1 \leq k \leq d$. We therefore find that $|\Delta|$ is bounded uniformly in terms of C.

Therefore $\mathbb{Q}(\alpha)$ has discriminant bounded in terms of C, and degree at most C. There are finitely many number fields of bounded degree and bounded discriminant, whence all such α lie in a single number field $K/\mathbb{Q}$.

But now α is forced to be integral at all primes with norm at least $e^{C[K:\mathbb{Q}]}$. Thus α is an S-integer of K for S a sufficiently large finite set. But if $\mathfrak{p} \in S$, $|\alpha|_{\mathfrak{p}} \leq e^{C[K:\mathbb{Q}]}$ as well. Thus by scaling by n a product of finitely many prime powers (bounded in terms of C and K) we may assume α is integral. But α then has norm bounded by $ne^{C[K:\mathbb{Q}]} \ll_C 1$ in all embeddings of K into $\mathbb{R}$ or $\mathbb{C}$. Thus all such α are roots of a finite list of polynomials of bounded degree with integral coefficients.[2] □

[2] Note to the reader: there is a slick proof of this theorem via Jensen's formula and the theory of the Mahler measure presented in Bombieri-Gubler's book [**11**] as well.

The second amusing property will not be proved here, simply because it is somewhat orthogonal to the focus of this thesis. However, the proof is fun, and the interested reader should consult [**11**] for the full details.

THEOREM 3.2 (Bilu's equidistribution theorem.). *Let $\alpha_i \in \bar{\mathbb{Q}}$ be distinct, nonzero, and such that $h(\alpha_i) \to 0$. Then the Galois orbits of the α_i equidistribute on the unit circle.*

That is, the measures determined by placing point masses at all conjugates of a fixed α_i limit to the Haar measure on the unit circle. For example, $h(2^{\frac{1}{n}}) = \frac{\log 2}{n} \to 0$ as $n \to \infty$. The Galois conjugates of $2^{\frac{1}{n}}$ are $\zeta_n^k 2^{\frac{1}{n}}, 0 \leq k < n$, ζ_n a primitive n-th root of unity. Since $2^{\frac{1}{n}} \to 1$ in $\mathbb{C}$, and since the roots of unity are dense on the unit circle, the result is certainly plausible.

Though we will not prove Bilu's theorem here, let us give an idea of what goes into the proof. By Banach-Alaoglu, it suffices to show that any convergent subsequence of the Galois orbit measures converges to Haar measure on the unit circle. By Jensen it converges to some measure supported only on the unit circle. It is seen to be absolutely continuous upon considering the growth rates of the discriminants of the α_i, all integers. By considering Fourier coefficients it also follows that its Fourier coefficients agree with those of Haar measure. Thus the theorem.

Finally, we leave the reader with an open problem.

QUESTION 3.3 (Lehmer's conjecture.). *Let $\alpha \in \bar{\mathbb{Q}}$ be nonzero and not a root of unity. Is $h(\alpha)\deg(\alpha)$ uniformly bounded below by a positive constant?*

The quantity $h(\alpha)\deg(\alpha)$ is the logarithm of the so-called *Mahler measure* of α, a complex-analytic measure of its minimal polynomial.[3] The infimum is conjectured to occur at the roots of $x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1$, a polynomial with Mahler measure $1.1762\ldots$.

Thus ends the introduction to Weil heights. We will now use them for something, namely proving that algebraic numbers cannot be approximated too well by rationals, steadily improving on what we mean by "too well" until we arrive at Roth's Fields Medal-winning theorem. We will prove the weaker results in full as well, if only to space out the introduction of the many new ideas in Roth's proof.

[3] Specifically, the *Mahler measure* of the polynomial $f(x_1, \dots, x_n)$ is

$$M(f) := \exp\left(\oint_{\zeta_1 \in S^1} \cdots \oint_{\zeta_n \in S^1} d\zeta_1 \cdots d\zeta_n \log|f(\zeta_1, \dots, \zeta_n)|\right).$$

By Jensen's formula this agrees with the definition given above.

CHAPTER

3

ROTH'S THEOREM

1. Theorems we will prove.

In this Chapter we will prove the following theorem.[1]

THEOREM 1.1 (Roth's theorem.). *Let $\alpha \in \bar{\mathbb{Q}}$. Let $\kappa > 2$. Then there are only finitely many rationals $\beta \in \mathbb{Q}$ for which*

$$|\alpha - \beta| \leq H(\beta)^{-\kappa}.$$

A few remarks about this theorem. First, by a construction of Dirichlet (the continued fraction expansion), for $\kappa = 2$ this theorem is, up to a constant, false. Thus in a sense this is the optimal result.

Second, for $\alpha \in \mathbb{Q}$, write $\alpha = \frac{a}{b}$ with $(a, b) = 1$ and write $\beta = \frac{c}{d}$ with $(c, d) = 1$. Then, if $\alpha \neq \beta$, $\left|\frac{a}{b} - \frac{c}{d}\right| \geq (bd)^{-1} \gg_\alpha H(\beta)^{-1}$, simply because the numerator is a nonzero integer, hence of absolute value at least 1. Thus the result is immediate in the case of rational α.

Finally, the proof of this theorem, were it to be presented in one and only one go, would be impossible to follow. Therefore we prove two weaker variants of the theorem, the first of which is rather basic but introduces the idea of auxiliary polynomials, and the second of which further introduces the Wronskian method and the central vanishing-nonvanishing tension (which arises via pigeonhole and tremendous asymmetry, respectively) exploited in the proof. These theorems are Liouville's and Thue's theorems, stated below. For the latter (and for Roth as well) we will need Siegel's Lemma, which implements the pigeonhole principle in producing multivariate polynomials with vanishing properties (though we will state it for underdetermined linear equations in the integers).

We state the aforementioned theorems below.

[1] As a reference the reader is encouraged to look in Bombieri-Gubler [**11**], whose bounds are often stronger and more uniform.

THEOREM 1.2 (Liouville's theorem.). *Let $\alpha \in \bar{\mathbb{Q}}$. Let $\kappa > \deg(\alpha)$. Then there are only finitely many rationals $\beta \in \mathbb{Q}$ for which*

$$|\alpha - \beta| \leq H(\beta)^{-\kappa}.$$

THEOREM 1.3 (Thue's theorem.). *Let $\alpha \in \bar{\mathbb{Q}}$. Let $\kappa > \frac{\deg(\alpha)}{2} + 1$. Then there are only finitely many rationals $\beta \in \mathbb{Q}$ for which*

$$|\alpha - \beta| \leq H(\beta)^{-\kappa}.$$

LEMMA 1.4 (Siegel's Lemma.). *Let $N > M$ be positive integers. Let $A = (A_{ij})_{i=1,j=1}^{M,N}$ ($A_{ij} \in \mathbb{Z}$) be an $M \times N$ matrix in the integers. Let $||A||_\infty := \max_{i,j} |A_{ij}|$. Then there is a nonzero integer vector $x = (x_j)_{j=1}^{N}$ ($x_i \in \mathbb{Z}$) for which $Ax = 0$ and*

$$||x||_\infty \leq (3||A||_\infty N)^{\frac{M}{N-M}},$$

where $||x||_\infty := \max_i |x_i|$.

We begin with Liouville.

2. Liouville's theorem.

Once one sees the degree of α factor in, one is immediately led to consider the minimal polynomial of α, which is the only invariant containing the degree of α that could possibly play a role. But then the point is clear: all the roots of the minimal polynomial are irrational if α is irrational (we have already seen the rational case, anyway), and so it cannot vanish on *any* rational number β. But if β is so close to α, it must evaluate to something quite small on β as well! Since we know the denominator of this nonzero number, we will derive a contradiction by measuring precisely what is meant by "quite small." Note that the whole proof proceeds through properties of the minimal polynomial of α, whence our terminology "auxiliary polynomial."

PROOF OF LIOUVILLE'S THEOREM. Let $p(x) \in \mathbb{Z}[x]$ be the minimal polynomial of α, of degree $d := \deg(\alpha)$. Since, without loss of generality, α is irrational, $p(\beta) \neq 0$. Therefore $|p(\beta)| \geq H(\beta)^{-d}$, since it is a nonzero rational with denominator at most $H(\beta)^d$. But

$$p(\beta) = p(\beta) - p(\alpha) = (\beta - \alpha) \cdot \frac{p(\beta) - p(\alpha)}{\beta - \alpha}.$$

The second factor is bounded above by a constant only depending on α simply by expanding out the relevant monomials (note that $|\beta| \leq |\alpha| + 1$ by hypothesis). Inserting the upper bound of the hypothesis into the first term, we see that

$$H(\beta)^{-d} \leq |p(\beta)| \ll_\alpha H(\beta)^{-\kappa}.$$

Once $H(\beta) \gg_\alpha 1$, this is a contradiction. Since there are finitely many rationals of bounded height, the result follows. □

Next we turn to Thue's theorem.

3. Thue's theorem.

Here the idea is to choose our auxiliary polynomial more intelligently. If we restrict ourselves to one variable, unfortunately we cannot do any better than Liouville — after all, if $p(x) \in \mathbb{Z}[x]$ vanishes on α, then automatically one has that the minimal polynomial of α divides p. Thus the degree of p is larger than that of the minimal polynomial, and so one obtains a worse bound for κ (namely, $\kappa \leq \deg p [\geq \deg(\alpha)]$).

But, as with many problems in mathematics (see Maynard's multidimensional Selberg sieve for bounded gaps between primes for a recent example), adding dimensions helps quite a bit. In the argument giving bounded gaps, one takes the dimension to be quite large in order to use the geometry of a high-dimensional simplex. We will see the exact same idea in Roth's theorem — except our simplex will be quite a bit larger (in volume) than the one used to prove the existence of bounded gaps between primes. But even before that, we will see that already in two dimensions one gets a significant improvement on Liouville's theorem.[2]

The idea of the proof will be the following. Rather than taking one extremely good approximation to α, let us take *two*. Specifically, take one approximation, say β_1, with *very large* height. Then take a second, say β_2, with height much, much larger than that of the first. This asymmetry in heights of these points will then be exploited by using a highly asymmetric auxiliary polynomial in two variables (asymmetric in the sense that its partial degrees will be orders of magnitude different). We will first produce such an auxiliary polynomial vanishing to high order at the point (α, α) via the pigeonhole principle. Then we will show that, due to its asymmetry, some small-order derivative of this polynomial will not vanish at (β_1, β_2). By replacing the original polynomial with this derivative and arguing as in Liouville (the upper bound from the high-order vanishing at (α, α), the lower bound from nonvanishing at (β_1, β_2) and rationality), we will obtain a contradiction.

Therefore it will follow that there cannot exist two such approximations with large, and quite different, heights. Thus finiteness will follow, but only in a qualitative sense, since this argument cannot rule out a single rational approximation of tremendous height. (However, it *can* control the number of approximations, which will be important to us in the proof of the titular result of this thesis.[3])

[2] Indeed, it is certain that the need for many dimensions was realized already in Thue's time, but the technical difficulty in producing an auxiliary polynomial not vanishing on a rational point constructed from approximations required some fifty years' wait until Roth proved his Lemma. The optimal result achieved by two-dimensional methods is the bound $\kappa \leq \sqrt{2d}$ due to Dyson and Gelfond, after Siegel proved the bound $\kappa \leq 2\sqrt{d}$. These results have historical significance because a bound of shape $\kappa \leq o(d)$ allowed Siegel to prove that there are only finitely many integral points on affine patches of curves of positive genus, and, more generally, those with at least three points at ∞, like $\mathbb{P}^1 - \{0, 1, \infty\}$. We will see the case of genus one curves shortly.

[3] In fact the control on the number of approximations is quite good. This is because one has a so-called *strong gap principle* forcing the approximations to be doubly exponentially far apart! Specifically, if $q > q'$ and $\frac{p}{q}, \frac{p'}{q'}$ are both approximations to α with exponent $\kappa = 2 + \epsilon > 2$, then

$$q^{-1} q'^{-1} \leq \left| \frac{p}{q} - \frac{p'}{q'} \right| \ll_\alpha q^{-2-\epsilon},$$

whence $q' \geq q^{1+\epsilon}$. That is to say, the denominators must grow like $q, q^{(1+\epsilon)}, q^{(1+\epsilon)^2}, \ldots, q^{(1+\epsilon)^k}, \ldots$ — doubly exponentially!

In order to start, however, we will need to make precise the use of the pigeonhole principle above. This is Siegel's Lemma.

3.1. Siegel's Lemma.

The statement of Lemma 1.4 suggests its proof, which is to simply use that the number of integral points in a large box in N dimensions grows much faster than the number of points in a similar box in M dimensions if $N > M$.

PROOF OF SIEGEL'S LEMMA. Let

$$B(X) := \{(x_i)_{i=1}^N \in \mathbb{Z}^N : 0 \leq x_i \leq X\}.$$

First, $\#|B(X)| = (X+1)^N$. Second,

$$A \cdot B(X) \subseteq \{(y_j)_{j=1}^M \in \mathbb{Z}^M : -||A||_\infty NX \leq y_j \leq ||A||_\infty NX\}.$$

The latter set has size $((2||A||_\infty + 1)NX)^M$. Once $X > ((2||A||_\infty + 1)N)^{\frac{M}{N-M}}$, $B(X)$ has more elements than $A \cdot B(X)$. Therefore there are $x \neq y \in B(X)$ with $Ax = Ay$. The vector $x - y$ is of the desired form. □

Thus we have made good on our repeated invocation of "the pigeonhole principle." Now let us apply it.

3.2. Proof of Thue's theorem.

PROOF OF THUE'S THEOREM. Suppose otherwise. By scaling (and then decreasing κ slightly) we had might as well take α to be an algebraic integer. Let β_1 be a solution to $|\alpha - \beta_1| \leq H(\beta_1)^{-\kappa}$ with $h(\beta_1) > A$ (we will choose A in terms of α in due course). Let β_2 be a solution to the same inequality with $h(\beta_2) > Ah(\beta_1)$.

Let $d_1 \in \mathbb{Z}^+$ be the nearest positive integer to $\frac{\deg(\alpha)}{2} \cdot \frac{h(\beta_2)}{h(\beta_1)}$. The asymmetry we will exploit is that d_1 is much larger than 1, and our auxiliary polynomial will be of partial degree bounded by $(d_1, 1)$.

Now to the construction of such a polynomial. First, let us find a polynomial that vanishes to high order on (α, α).

CLAIM 3.1. *Let $\epsilon > 0$. Let $M \leq \frac{2d_1}{\deg(\alpha)}(1-\epsilon)$ be a positive integer. Then there is a nonzero polynomial $P(x,y) \in \mathbb{Z}[x,y]$ with $\overrightarrow{\deg}(P) \leq (d_1, 1)$, coefficients bounded in absolute value by*

$$||P||_\infty \ll e^{O_{\alpha,\epsilon}(d_1)}$$

and such that

$$P(\alpha,\alpha), (\partial_x P)(\alpha,\alpha), \ldots, (\partial_x^M P)(\alpha,\alpha) = 0,$$

where $\partial_x := \frac{\partial}{\partial x}$ and $\overrightarrow{\deg}(P) := (\deg_x(P), \deg_y(P))$.

PROOF OF CLAIM 3.1. Consider the following linear equations in $a_{ij}, 0 \leq i \leq d_1, 0 \leq j \leq 1$:

$$\sum_{i=0}^{d_1} \sum_{j=0}^{1} a_{ij} \binom{i}{a} \alpha^{i+j-a} = 0.$$

Write, for $k \geq \deg(\alpha)$, α^k in terms of $1, \alpha, \ldots, \alpha^{\deg(\alpha)-1}$. Then set the coefficients of each of the α^i ($0 \leq i < \deg(\alpha)$) equal to zero. This gives $M \deg(\alpha)$ *linear* equations for the a_{ij}. The size of the coefficients of each of the a_{ij} is at most $O(1)^{d_1} \cdot O_\alpha(1)^{d_1} = e^{O_\alpha(d_1)}$, the first factor coming from the binomial coefficients (which are each at most 2^{d_1}), and the second from the repeated use of the linear relation $\alpha^{\deg(\alpha)} \in \operatorname{span}_{\mathbb{Z}}(1, \ldots, \alpha^{\deg(\alpha)-1})$.

The number of a_{ij} is $2(d_1+1) > M\deg(\alpha)$ by hypothesis. Therefore, by Siegel's Lemma, there are $a_{ij} \in \mathbb{Z}$ satisfying all of the considered equations, and all bounded by

$$||(a_{ij})_{i=0,j=0}^{d_1,1}||_\infty \leq \exp(O_{\alpha,\epsilon}(d_1)).$$

Let now $P(x,y) := \sum_{i=0}^{d_1}\sum_{j=0}^{1} a_{ij}x^iy^j$, and observe that therefore

$$(\partial_x^a P)(\alpha,\alpha) = a!\sum_{i=0}^{d_1}\sum_{j=0}^{1}\binom{i}{a}a_{ij}\alpha^{i+j-a} = 0.$$

This completes the proof. □

A few comments are in order. The bound on the coefficients of the polynomial P will be more than enough for our purposes, but it would have been far too weak had we not factored out the $a!$ term in expressing the vanishing of $(\partial_x^a P)(\alpha,\alpha)$. Second, the point of the $(1-\epsilon)$ term in the bound on M is for precisely the same reason: the bound on the coefficients Siegel provides has exponent $\frac{\text{dim}}{\text{codim}} = \frac{M\deg(\alpha)}{2(d_1+1)-M\deg(\alpha)}$. Since we would like to cancel the contribution of the numerator, we are forced to take M to be on the same scale, but slightly smaller, than $\frac{2d_1}{\deg(\alpha)}$. This will change nothing in the final bound.

Unfortunately for us our work does not end here. This is because we have no control on whether or not P vanishes on (β_1,β_2).[4] However, for the purposes of the final bound it will be enough to show that a derivative of small order — say, ϵM — does not vanish (whence we are left with basically all of the vanishing at (α,α) and nonvanishing at (β_1,β_2) if we work with this small-order derivative instead). This is the result that was the bottleneck for fifty years before Roth proved his Lemma. The method we use below will be of the same flavour.

CLAIM 3.2. *Let $\epsilon > 0$. Let $M := \lfloor\frac{2d_1}{\deg(\alpha)}(1-\epsilon)\rfloor$. Let $P(x,y)$ be as guaranteed by Claim 3.1. There is some $a \leq 1 + \frac{O_{\alpha,\epsilon}(d_1)}{A}$ such that*

$$(\partial_x^a P)(\beta_1,\beta_2) \neq 0.$$

PROOF OF CLAIM 3.2. Write $P(x,y) =: F(x) + yG(x)$. If F is proportional to G (or vice versa), then the result is immediate since α is (without loss of generality) irrational and thus $F(\beta_1) \neq 0$ and $\beta_2 \neq -1$ (since $h(\beta_2)$ is large). Otherwise, let a be minimal such that $(\partial_x^a P)(\beta_1,\beta_2) \neq 0$. Consider the following polynomial in x:

$$W(x) := \det\begin{pmatrix} F & \partial_x F \\ G & \partial_x G \end{pmatrix}.$$

Note that, for $0 \leq m \leq a-1$, $(\partial_x^m W)(\beta_1) = 0$ via bilinearity of the determinant, since there is a linear relation among the rows of any matrix that appears after differentiation:

$$\big((\partial_x^m F)(\beta_1), (\partial_x^{n+1}F)(\beta_1)\big) + \beta_2\cdot\big((\partial_x^m G)(\beta_1), (\partial_x^{n+1}G)(\beta_1)\big) = 0$$

[4] While an extensive commentary on where the naïve attempts to control this fail would likely be out of the scope of a senior thesis, suffice it to say that an attempt to both guarantee vanishing of all the derivatives and nonvanishing at a single point will preclude the use of the pigeonhole principle — indeed, were one able to do this in sufficient generality one would obtain a result contradicting Dirichlet's continued fraction construction! We leave it to the reader to do his own exploration.

for $0 \leq m, n \leq a-1$. Therefore, being a univariate polynomial vanishing on β_1 to order $a-1$, it must be divisible by the minimal polynomial of β_1, also known as $q_1 x - p_1$, where $\beta_1 =: \frac{p_1}{q_1}$, to the same order. That is to say,

$$(q_1 x - p_1)^{a-1} | W(x).$$

Since $W(x)$ is not identically zero (consider the numerator of $\left(\frac{F}{G}\right)'$ or $\left(\frac{G}{F}\right)'$), this implies that the largest coefficient of $W(x)$ is at least $\max(|p_1|, |q_1|)^{a-1} = H(\beta_1)^{a-1}$.

But we know a bound for the largest coefficient of P, and therefore we may bound the largest coefficient appearing in W. By definition it is at most

$$||W||_\infty \leq \exp(O_{\alpha,\epsilon}(d_1)).$$

Therefore $a \leq 1 + \frac{O_{\alpha,\epsilon}(d_1)}{h(\beta_1)}$. This completes the proof. □

Let now a be as guaranteed in Claim 3.2 for a polynomial P guaranteed by Claim 3.1. (Here, as before, we take $\epsilon > 0$ and $M := \lfloor \frac{2d_1}{\deg(\alpha)}(1-\epsilon) \rfloor$.) Let $Q(x,y) := (\frac{\partial_x^a}{a!} P)(x,y)$. Since certainly $a \ll_{\alpha,\epsilon} d_1$ (indeed, by quite a margin), we have that the bound on the coefficients of P persists, again because we have factored out the factorial term:

$$||Q||_\infty \leq \exp(O_{\alpha,\epsilon}(d_1)).$$

Moreover, Q still vanishes to quite high order at (α, α):

$$Q(\alpha,\alpha), (\partial_x Q)(\alpha,\alpha), \ldots, (\partial_x^{M-a} Q)(\alpha,\alpha) = 0.$$

Finally, we now have the crucial fact that

$$Q(\beta_1, \beta_2) \neq 0.$$

Now let us play these facts off one another to complete the proof. In the first place, $Q(\beta_1,\beta_2)$ is nonzero, so it is, in absolute value, at least

$$|Q(\beta_1,\beta_2)| \geq \exp(-d_1 h(\beta_1) - h(\beta_2)).$$

In the second place, $Q(\beta_1,\beta_2)$ is quite close to $Q(\alpha,\alpha)$, so it is quite small. Specifically, via Taylor expanding around (α,α), we have that

$$\begin{aligned} |Q(\beta_1,\beta_2)| &\ll e^{O_{\alpha,\epsilon}(d_1)} \left(|\alpha - \beta_1|^{M-a} + |\alpha - \beta_2| \right) \\ &\ll \exp\left(O_{\alpha,\epsilon}(d_1) - \min((M-a)\kappa h(\beta_1), \kappa h(\beta_2)) \right). \end{aligned}$$

Therefore we derive the inequality

$$\kappa \cdot \min\left((M-a) h(\beta_1), h(\beta_2) \right) \leq d_1 h(\beta_1) + h(\beta_2) + O_{\alpha,\epsilon}(d_1).$$

That is to say, since d_1 is within 1 of $\frac{\deg(\alpha)}{2} \frac{h(\beta_2)}{h(\beta_1)} > A$,

$$\kappa \leq \frac{d_1}{M-a} + 1 + \frac{O_{\alpha,\epsilon}(1)}{A} \leq \frac{\deg(\alpha)}{2}(1+\epsilon) + \frac{O_{\alpha,\epsilon}(1)}{A}.$$

Taking first ϵ very small (depending on α and κ) and then A very large (depending on α, κ, and ϵ) derives a contradiction, completing the proof. □

Reiterating, our method of proof was the following. We first supposed there were infinitely many such rational approximations in order to produce a sequence β_1, β_2 of two of them starting at tremendous height and spaced out with tremendous spaces between heights. Then we constructed a polynomial depending on these approximations that was both *very* zero (i.e., vanished to enormous order) at (α,α), and nonzero at (β_1,β_2). (To do this we first constructed a polynomial vanishing to high order at (α, α) by pigeonhole, and then showed that a small derivative of it would not vanish at (β_1, β_2) because we could reduce to the univariate case and force the coefficients of our polynomial to actually be quite large, a contradiction.) Using this polynomial and proceeding as in Liouville, we obtained a contradiction, thus affirming the original claim.

Now we may begin the proof of Roth's theorem.

4. Roth's theorem.

There will be many parameters in the proof. We leave it to the reader to trust us to choose them carefully (and in the correct order) at the end.

PROOF OF ROTH'S THEOREM. Again, without loss of generality α is an algebraic integer. Let $\beta_1, \ldots, \beta_m$ be m solutions to the inequality in question with $h(\beta_1) > A$ and $h(\beta_i) > Ah(\beta_{i-1})$ for $i > 1$.

Let $\frac{d_2}{d_1}, \ldots, \frac{d_m}{d_1}$ be continued fraction approximants to $\frac{h(\beta_1)}{h(\beta_2)}, \ldots, \frac{h(\beta_1)}{h(\beta_m)}$ with $d_1 > h(\beta_m)[= \max_i h(\beta_i)]$, so that

$$|d_i h(\beta_i) - d_1 h(\beta_1)| \ll 1.$$

Therefore

$$|d_i h(\beta_i) - d_j h(\beta_j)| \ll 1$$

as well.

Now let us construct from this data an auxiliary polynomial which vanishes to high order at $(\alpha, \alpha, \ldots, \alpha)$. But how should we measure "high order" with so many variables? We will measure it by the strength of the final upper bound implied by the vanishing. As in the two-variable case, the resulting upper bound will be of shape $\exp(-\kappa[\mu_1 h(\beta_1) + \cdots + \mu_m h(\beta_m)])$, where $\sum \mu_i$ is minimal so that $(\partial_1^{\mu_1} \cdots \partial_m^{\mu_m} P)(\alpha, \ldots, \alpha) \neq 0$, for P our auxiliary polynomial (there may be multiple such $\vec{\mu}$s, but let us forget about this for the moment). This to optimize this bound we would like $\mu_1 h(\beta_1) + \cdots + \mu_m h(\beta_m)$ to be as large as possible. But $d_1 h(\beta_1) \approx \cdots \approx d_m h(\beta_m)$, so that this is equivalent to asking that $\frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m}$ be as large as possible. This weighted degree of the operator $\partial_1^{\mu_1} \cdots \partial_m^{\mu_m}$ is what we will use to measure the vanishing of our auxiliary polynomial.

Specifically, let, for a polynomial $P \in \mathbb{Z}[x_1, \ldots, x_m]$ and $(\xi_1, \ldots, \xi_m) \in \bar{\mathbb{Q}}^m$,

$$\operatorname{ind}(P, (\xi_1, \ldots, \xi_m)) := \max \left\{ \frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} : (\partial_1^{\mu_1} \cdots \partial_m^{\mu_m} P)(\xi_1, \ldots, \xi_m) = 0 \right\}.$$

We will also use the notation $\vec{\xi} := (\xi_1, \ldots, \xi_m)$ and

$$\partial_{\vec{\mu}} := \frac{\partial_1^{\mu_1}}{\mu_1!} \cdots \frac{\partial_m^{\mu_m}}{\mu_m!},$$

where we divide out by the factorials for the same reason as in the proof of Thue's theorem — namely, dividing by these factorials preserves integrality and makes the coefficients quite a bit smaller. We will again employ the notation $\overrightarrow{\deg}(P) := (\deg_{x_1}(P), \ldots, \deg_{x_m}(P))$ for the partial degrees of P.

CLAIM 4.1. *Let $\epsilon > 0$. Let $\frac{1}{2} - \epsilon \leq t \leq \frac{1}{2} - \frac{1}{A} - O_\epsilon\left(\sqrt{\frac{\deg(\alpha)}{m}}\right)$. Then there is a nonzero polynomial $P(x_1, \ldots, x_m) \in \mathbb{Z}[x_1, \ldots, x_m]$ of partial degrees bounded by*

$\overrightarrow{\deg}(P) \le (d_1, \ldots, d_m)$, *coefficients bounded by*

$$||P||_\infty \le \exp(O_{\epsilon,\alpha}(d_1)),$$

and such that $\operatorname{ind}(P, (\alpha, \ldots, \alpha)) \ge mt$.

The point is that we have $\prod_{i=1}^m (d_i + 1) \approx d_1 \cdots d_m$ many variables to satisfy

$$\approx \deg(\alpha) \cdot \operatorname{vol}\left(\{(x_1, \ldots, x_m) \in \mathbb{R}^m : 0 \le x_1 \le d_1, \ldots, 0 \le x_m \le d_m, \frac{x_1}{d_1} + \cdots + \frac{x_m}{d_m} < mt\}\right)$$

many equations. Thus we need

$$\begin{aligned}&\deg(\alpha) \cdot \operatorname{vol}\left(\{(x_1, \ldots, x_m) \in \mathbb{R}^m : 0 \le x_1, \ldots, x_m \le 1, \sum x_i < mt\}\right)\\ &= \deg(\alpha) \cdot \Pr\left(X_1 + \cdots + X_m < mt\right)\end{aligned}$$

to be appreciably smaller than 1, where X_i are independent uniform random variables on $[0, 1]$. But the law of large numbers tells us that $\frac{X_1+\cdots+X_m}{m}$ is basically a point mass on $\frac{m}{2}$, so that the probability term will decay extremely rapidly in $\frac{1}{2} - t$![5] In fact the tail probability will decay as the tail probability of a Gaussian, by e.g. the Chernoff/Hoeffding bound (and one cannot do better because of the central limit theorem).

PROOF OF CLAIM 4.1. Consider the linear equations

$$\sum_{0 \le i_1 \le d_1, \ldots, 0 \le i_m \le d_m} a_{(i_1, \ldots, i_m)} \binom{i_1}{\mu_1} \cdots \binom{i_m}{\mu_m} \alpha^{i_1 + \cdots + i_m - \mu_1 - \cdots - \mu_m} = 0,$$

one for each $\vec{\mu}$ for which $0 \le \mu_1 \le d_1, \ldots, 0 \le \mu_m \le d_m$ and $\frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} \le mt$, in the variables $a_{(i_1, \ldots, i_m)} \in \mathbb{Z}$ $(0 \le i_1 \le d_1, \ldots, 0 \le i_m \le d_m)$.

Using the relation $\alpha^{\deg(\alpha)} \in \operatorname{span}_{\mathbb{Z}}(1, \ldots, \alpha^{\deg(\alpha)-1})$, write this as $\deg(\alpha)$ times as many equations with integral coefficients by setting the resulting coefficients of each α^i, $0 \le i < \deg(\alpha)$, equal to zero.

There are thus

$$\deg(\alpha) \cdot \# \left| \left\{ (\mu_1, \ldots, \mu_m) \in \mathbb{Z}^m : 0 \le \mu_1 \le d_1, \ldots, 0 \le \mu_m \le d_m, \frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} \le mt \right\} \right|$$

many *linear* equations, each with coefficients bounded by $\exp(O_\alpha(d_1 + \cdots + d_m))$. There are $\prod_{i=1}^m (d_i + 1) \ge \prod_{i=1}^m d_i$ many variables.[6] So long as

$$\begin{aligned}&\# \left| \left\{ (\mu_1, \ldots, \mu_m) \in \mathbb{Z}^m : 0 \le \mu_1 \le d_1, \ldots, 0 \le \mu_m \le d_m, \frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} \le mt \right\} \right|\\ &< \frac{d_1 \cdots d_m}{\deg(\alpha)}(1 - \epsilon),\end{aligned}$$

Siegel's Lemma will apply (again, we have added the $(1 - \epsilon)$ to keep the exponent in Siegel's Lemma from growing).

Now the count of lattice points in the convex region $\{(x_1, \ldots, x_m) \in \mathbb{R}^m : 0 \le x_1 \le d_1, \ldots, 0 \le x_m \le d_m, \frac{x_1}{d_1} + \cdots + \frac{x_m}{d_m} < mt\}$ is very nearly the volume of the region. Indeed, this count is certainly the same as the volume of the region formed

[5] In fact the distribution is exactly a so-called Irwin-Hall distribution, which incidentally arises when trying to prove that certain partition sequences are unimodal.

[6] The inequality we have used is not at all wasteful, since the d_i grow exponentially with very large factor.

by the union of closed hypercubes of length 1 centered at each of the lattice points. But this region is contained in the region

$$\left\{(x_1,\ldots,x_m)\in\mathbb{R}^m : 0\leq x_1\leq d_1,\ldots,0\leq x_m\leq d_m, \frac{x_1}{d_1}+\cdots+\frac{x_m}{d_m} < mt+\frac{1}{2}\left(\frac{1}{d_1}+\cdots+\frac{1}{d_m}\right)\right\}.$$

Since the d_i grow exponentially (and $A>2$ — indeed, by some margin) this is further contained in the region

$$\left\{(x_1,\ldots,x_m)\in\mathbb{R}^m : 0\leq x_1\leq d_1,\ldots,0\leq x_m\leq d_m, \frac{x_1}{d_1}+\cdots+\frac{x_m}{d_m} < mt+\frac{1}{A}\right\},$$

which has volume

$$(d_1\cdots d_m)\cdot\Pr\left(\frac{X_1+\cdots+X_m}{m} < t+\frac{1}{A}\right),$$

where the X_i are independent uniform random variables on $[0,1]$. The Hoeffding inequality tells us that this is at most

$$\leq (d_1\cdots d_m)\cdot e^{-2m(\frac{1}{2}-t-\frac{1}{A})^2}.$$[7]

Thus for t at most $\frac{1}{2}-\frac{1}{A}-O_\epsilon\left(\sqrt{\frac{\deg(\alpha)}{m}}\right)$ we have the desired upper bound for the number of equations.

Therefore Siegel's Lemma applies and we may find a solution

$$P(x_1,\ldots,x_m) =: \sum_{0\leq i_1\leq d_1,\ldots,0\leq i_m\leq d_m} a_{(i_1,\ldots,i_m)}x_1^{i_1}\cdots x_m^{i_m}\in\mathbb{Z}[x_1,\ldots,x_m]$$

such that

$$||P||_\infty \leq e^{O_{\epsilon,\alpha}(d_1+\cdots+d_m)}\leq e^{O_{\epsilon,\alpha}(d_1)}$$

(since $A>2$ and $d_1>Ad_2>A^2d_3>\cdots>A^{m-1}d_m$) and

$$\operatorname{ind}(P,(\alpha,\ldots,\alpha))\geq mt.$$

This completes the proof. □

So let $\epsilon>0$ be another parameter, and let P be as guaranteed in Claim 4.1 (once $m,A\gg_{\alpha,\epsilon}1$, so that there is some t for which the hypothesis holds). We would like to take some small derivative of P produce a polynomial Q that does not vanish on $(\beta_1,\ldots,\beta_m)$. To do this we will need the major achievement of Roth's proof, Roth's Lemma.

CLAIM 4.2 (Roth's Lemma.). *Let $P\in\mathbb{Z}[x_1,\ldots,x_m]$ be a polynomial with partial degrees bounded above by $\overrightarrow{\deg}P\leq(d_1,\ldots,d_m)$ with $d_1>Ad_2>\cdots>A^{m-1}d_m$ and coefficients bounded above by*

$$||P||_\infty\leq\exp(Cd_1),$$

with $A>C>28$. Let $\beta_i\in\mathbb{Q}$ be such that $|d_ih(\beta_i)-d_jh(\beta_j)|\ll 1$ and $h(\beta_1)>A$. Then

$$\operatorname{ind}(P,(\beta_1,\ldots,\beta_m))\leq 2m\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-1}}}.$$

[7] For an elementary argument giving this bound, see Bombieri-Gubler Lemma 6.3.5. I have simply cited Hoeffding here because the proof of such a bound, albeit quite simple, is not at all the focus of our efforts!

The proof will proceed by induction on the number of variables m, just as in the two-variable case (where we reduced to the univariate case via the Wronskian). We will write $||\vec{x}||_1 := \sum |x_i|$ for the ℓ^1-norm.

PROOF OF ROTH'S LEMMA. Let us first do the case $m = 1$. Note that, writing $\beta_1 =: \frac{p_1}{q_1}$,

$$(q_1 x - p_1)^{d_1 \operatorname{ind}(P,\beta_1)} | P(x_1).$$

Since P is nonzero this implies that

$$H(\beta_1)^{d_1 \operatorname{ind}(P,\beta_1)} \leq ||P||_\infty \leq \exp(Cd_1).$$

Therefore

$$\operatorname{ind}(P, \beta_1) \leq \frac{C}{h(\beta_1)} \leq \frac{C}{A},$$

which is even stronger than the desired form.

Next, in the case $m > 1$, write

$$P(x_1, \ldots, x_m) =: \sum_{i=0}^{s} f_i(x_1, \ldots, x_{m-1}) g_i(x_m)$$

for some $s \leq d_m$ and f_i and g_i all linearly independent. (An expression of this form exists by writing $P(x_1, \ldots, x_m) = \sum_{i=0}^{d_m} P_i(x_1, \ldots, x_{m-1}) x_m^i$ and then combining linearly independent terms (decreasing the number of summands at each step) until termination.)

The claim is that linear independence forces there to be some Wronskian $W_{\vec{\mu}^{(0)},\ldots,\vec{\mu}^{(s)}}(x_1, \ldots, x_{m-1})$ (with $||\vec{\mu}^{(i)}||_1 =: \sum_{j=1}^{m-1} \mu_j^{(i)} \leq i$) in the f_i that does not vanish identically, where

$$W_{\vec{\mu}^{(0)},\ldots,\vec{\mu}^{(s)}}(x_1, \ldots, x_{m-1}) := \det\left(\partial_{\vec{\mu}^{(i)}} f_j\right)_{i,j=0}^{s}.$$

For the same reason there is also a Wronskian $W_{\mu_m^{(0)},\ldots,\mu_m^{(s)}}(x_m)$ (with $\mu_m^{(i)} \leq i$) in the g_i that does not vanish identically, where

$$W_{\mu_m^{(0)},\ldots,\mu_m^{(s)}}(x_m) := \det\left(\partial_{\mu_m^{(i)}} g_j\right)_{i,j=0}^{s}.$$

(Here we have written $\partial_{\mu_m^{(i)}} := \frac{\partial_m^{\mu_m^{(i)}}}{\mu_m^{(i)}!}$.)

To see this claim, note that otherwise all such Wronskians would vanish as functions of t upon letting $x_1 = t, x_2 = t^N, \ldots, x_m = t^{N^m}$, where $N \gg_{\deg(f_i),\deg(g_i)} 1$ is sufficiently large. Since N is so large, all the f_i and g_i are linearly independent as functions of t. Thus we need only prove the nonvanishing of the Wronskian $W(t) := \det\left((\partial_t^i p_j)(t)\right)_{i,j=0}^{s}$ for univariate and linearly independent polynomials $p_0(t), \ldots, p_s(t) \in \mathbb{Z}[t]$, since then by the chain rule one of our original Wronskians would not vanish (since the univariate Wronskian is a sum of terms proportional to Wronskians formed with derivatives in the x_i, by multilinearity of the determinant and the product rule of differentiation).

But this is classical: the vanishing of the determinant implies that there are some $a_i(t) \in \mathbb{Z}[t]$ not all zero such that $\sum_{i=0}^{s} a_i(t)(\partial_t^i p_j)(t) = 0$ for each j. Thus we have found $s + 1$ linearly independent solutions to the (at most) s-th order differential equation

$$a_0(t)f(t) + a_1(t)f'(t) + \cdots + a_s(t)f^{(s)}(t) = 0,$$

which is impossible, by uniqueness of solutions to ordinary differential equations (there is only an (at most) s-dimensional space of initial conditions).

Let then $U(x_1, \dots, x_m) := W_{\vec{\mu}^{(0)},\dots,\vec{\mu}^{(s)}}(x_1, \dots, x_{m-1}), V(x_m) := W_{\mu_m^{(0)},\dots,\mu_m^{(s)}}(x_m)$ be the guaranteed nonvanishing Wronskians. Then

$$\begin{aligned} W_P(x_1, \dots, x_m) &:= U(x_1, \dots, x_{m-1}) \cdot V(x_m) \\ &= \det \left(\partial_{\left(\mu_1^{(i)},\dots,\mu_{m-1}^{(i)},\mu_m^{(j)}\right)} P \right)_{i,j=0}^{s} \end{aligned}$$

is a nonvanishing Wronskian of P, with each $||(\mu_1^{(i)}, \dots, \mu_{m-1}^{(i)}, \mu_m^{(j)})||_1 \le 2s \le 2d_m$.

Therefore, by expanding the determinant as a sum of products indexed by permutations,

$$||W_P(x_1, \dots, x_m)||_\infty \le s! \cdot 2^{2s} \cdot ||P||_\infty^s \le \exp(s(\log d_m + Cd_1 + 2)) \le \exp((C+4)[(s+1)d_1]).$$

Since U and V have disjoint sets of variables, the product UV has coefficients the pairwise products of coefficients of U and V, without any summation. Thus also

$$||W_P(x_1, \dots, x_m)||_\infty = ||U(x_1, \dots, x_{m-1})||_\infty \cdot ||V(x_m)||_\infty.$$

Hence the same upper bound applies to both U and V.

But now note that $\overrightarrow{\deg}(U) \le ((s+1)d_1, \dots, (s+1)d_{m-1})$, which are *still* separated by a factor of A! We also of course have $\deg(V) \le (s+1)d_m$.

Thus by the univariate case from the beginning of the argument we have that

$$\operatorname{ind}(V, \beta_m) \le \frac{C+4}{A}(s+1).$$

(The extra $s+1$ term comes from the fact that the univariate case we use has ind with weight $(s+1)d_m$, rather than d_m.) Moreover, by the induction hypothesis we have that

$$\begin{aligned} \operatorname{ind}(U, (\beta_1, \dots, \beta_{m-1})) &\le 2(s+1)(m-1) \left(\frac{(C+4)(m-1)}{A} \right)^{\frac{1}{2^{m-2}}} \\ &\le 2(s+1)(m-1) \left(\frac{(C+4)m}{A} \right)^{\frac{1}{2^{m-2}}}. \end{aligned}$$

(Again, the factor of $s+1$ comes from the difference in weights.) Therefore

$$\operatorname{ind}(W_P, (\beta_1, \dots, \beta_m)) \le 2(m-1)(s+1) \left(\frac{(C+4)m}{A} \right)^{\frac{1}{2^{m-2}}} + (s+1)\frac{C+4}{A}.$$

But if P vanishes to extremely high order at $(\beta_1, \dots, \beta_m)$, then, even with all these derivatives on P, this Wronskian determinant would still have many *entire rows* which vanish to high order at $(\beta_1, \dots, \beta_m)$. So certainly the determinant would vanish!

Let us make this precise. Consider the expansion of the determinant as a sum of permutations. Since the sum certainly vanishes if all its constituent terms vanish, we have that

$$
\begin{aligned}
\operatorname{ind}(W_P, (\beta_1, \ldots, \beta_m)) &\geq \min_{\sigma \in S_s} \sum_{i=0}^{s} \operatorname{ind}\left(\partial_{(\mu_1^{(i)}, \ldots, \mu_{m-1}^{(i)}, \mu_m^{(\sigma(i))})} P, (\beta_1, \ldots, \beta_m)\right) \\
&\geq \min_{\sigma \in S_s} \sum_{i=0}^{s} \max\left(0, \operatorname{ind}(P, (\beta_1, \ldots, \beta_m)) - \frac{\mu_1^{(i)}}{d_1} - \cdots - \frac{\mu_{m-1}^{(i)}}{d_{m-1}} - \frac{\mu_m^{(\sigma(i))}}{d_m}\right) \\
&\geq \sum_{i=0}^{s} \left[\max\left(0, \operatorname{ind}(P, (\beta_1, \ldots, \beta_m)) - \frac{i}{d_m}\right) - \frac{s}{d_{m-1}}\right] \\
&\geq \sum_{i=0}^{s} \left[\max\left(0, \operatorname{ind}(P, (\beta_1, \ldots, \beta_m)) - \frac{i}{s}\right) - \frac{1}{A}\right] \\
&= \sum_{i=0}^{s} \max\left(0, \operatorname{ind}(P, (\beta_1, \ldots, \beta_m)) - \frac{i}{s}\right) - \frac{s+1}{A},
\end{aligned}
$$

where we have used that

$$
\operatorname{ind}(\partial_{\vec{\mu}} P, (\xi_1, \ldots, x_m)) \geq \operatorname{ind}(P, (x_1, \ldots, x_m)) - \left\| \left(\frac{\mu_i}{d_i}\right)_{i=1}^{m} \right\|_1,
$$

and that ind is always nonnegative.

Now in general

$$
\sum_{i=0}^{s} \max\left(0, x - \frac{i}{s}\right) \geq \frac{s+1}{2} \min(x, x^2).
$$

Indeed, if $x \geq 1$, then $\max\left(0, x - \frac{i}{s}\right) = x - \frac{i}{s}$, so that the sum is equal to

$$
(s+1)\left(x - \frac{1}{2}\right) \geq \frac{s+1}{2} x.
$$

Else, the sum is equal to

$$
\begin{aligned}
\sum_{0 \leq i \leq xs} x - \frac{i}{s} &= (\lfloor xs \rfloor + 1)x - \frac{\lfloor xs \rfloor (\lfloor xs \rfloor + 1)}{2s} \\
&\geq \frac{x^2 s}{2}\left(1 + \frac{1 - \{xs\}}{xs}\right)\left(1 + \frac{\{xs\}}{xs}\right) \\
&\geq \frac{s+1}{2} x^2
\end{aligned}
$$

(here $\{a\}$ is the fractional part of $a \in \mathbb{R}$ and $\lfloor a \rfloor = a - \{a\}$ is the largest integer not exceeding a). Here we have applied the inequality

$$
\left(1 + \frac{1 - \{xs\}}{xs}\right)\left(1 + \frac{\{xs\}}{xs}\right) \geq 1 + \frac{1}{xs}\left[\geq 1 + \frac{1}{s}\right]
$$

— after all, for $a, b > 1$ one always has $ab \geq a + b - 1$, since $ab - a - b + 1 = (a-1)(b-1)$.

Therefore we see that

$$
\operatorname{ind}(W_P, (\beta_1, \ldots, \beta_m)) \geq (s+1)\left(\min\left(\frac{1}{2}\operatorname{ind}(P, (\beta_1, \ldots, \beta_m)), \frac{1}{2}\operatorname{ind}(P, (\beta_1, \ldots, \beta_m))^2\right) - \frac{1}{A}\right).
$$

But we have already proved that

$$\operatorname{ind}(W_P,(\beta_1,\dots,\beta_m)) \le 2(s+1)(m-1)\left(\frac{(C+4)m}{A}\right)^{\frac{1}{2^{m-2}}} + (s+1)\frac{(C+4)}{A}$$

Therefore, since of course $\operatorname{ind}(P,(\beta_1,\dots,\beta_m)) \le m$, multiplying through by $m \ge 1$ if necessary gives

$$\begin{aligned}\operatorname{ind}(P,(\beta_1,\dots,\beta_m))^2 &\le 4m(m-1)\left(\frac{(C+4)m}{A}\right)^{\frac{1}{2^{m-2}}} + 3m\frac{(C+4)}{A}\\ &\le 4m^2\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-2}}}.^{8}\end{aligned}$$

This completes the proof of Roth's Lemma. □

Therefore let $\epsilon > 0$, and $A, m \gg_{\alpha,\epsilon} 1$ so that Claim 4.1 applies for some t with $\frac{1}{2} - t \le \epsilon - \frac{1}{mA}$. Let P be the polynomial guaranteed by Claim 4.1, and observe that, by Roth's Lemma (with $C \le O_{\alpha,\epsilon}(1)$),

$$\operatorname{ind}(P,(\beta_1,\dots,\beta_m)) \le 2m\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-1}}}.$$

Therefore there is some derivative (dividing out by factorials as usual) $Q(x_1,\dots,x_m)$ of P with

$$\begin{aligned}\operatorname{ind}(Q,(\alpha,\dots,\alpha)) &\ge \operatorname{ind}(P,(\alpha,\dots,\alpha)) - 2m\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-1}}}\\ &\ge mt - 2m\left(\frac{O_{\alpha,\epsilon}(m)}{A}\right)^{\frac{1}{2^{m-1}}}\\ &\ge \left(\frac{1}{2}-\epsilon\right)m\end{aligned}$$

once $A \gg_{\alpha,\epsilon,m} 1$. Moreover, this Q has the crucial property that $Q(\beta_1,\dots,\beta_m) \ne 0$. Finally, the coefficients of Q are bounded by

$$||Q||_\infty \le \exp\left(\left(2m\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-1}}} + O_{d,\alpha}(1)\right)d_1\right).$$

Again because $A \gg_{\alpha,\epsilon,m} 1$, this implies that

$$||Q||_\infty \le \exp(O_{\alpha,\epsilon}(d_1)),$$

so the bound on the coefficients of P persists.

[8] Here we have used the inequality

$$(4m-1)\left(\frac{(C+4)m}{A}\right)^{\frac{1}{2^{m-2}}} \le 4m\left(\frac{Cm}{A}\right)^{\frac{1}{2^{m-2}}},$$

which holds once

$$1 + \frac{4}{C} \le \left(1 + \frac{1}{4m-1}\right)^{2^{m-2}},$$

e.g. once $C > 28$.

Now we are done. First, $Q(\beta_1, \ldots, \beta_m) \neq 0$ implies

$$\begin{aligned} |Q(\beta_1, \ldots, \beta_m)| &\geq \exp(-d_1 h(\beta_1) - \cdots - d_m h(\beta_m)) \\ &\geq \exp(-m d_1 h(\beta_1) + m). \end{aligned}$$

Second, Taylor expanding Q about $(\alpha, \ldots, \alpha)$ implies that

$$|Q(\beta_1, \ldots, \beta_m)| \ll \exp\left(O(m \log d_1) + O_{\alpha,\epsilon}(d_1) - \kappa\left(\frac{1}{2} - \epsilon\right) m d_1 h(\beta_1)\right),$$

where the $\exp(O(m \log d_1))$ term comes from the $O(m)$ derivatives on Q as well as the $d_1^{O(m)}$ many $\vec{\mu} \geq \vec{0}$ with $\sum_{i=1}^m \frac{\mu_i}{d_i} > \left(\frac{1}{2} - \epsilon\right) m$ and $\sum_{i=1}^m \mu_i \leq \left(\frac{1}{2} - \epsilon\right) m d_1$. (The count is at most the volume of the latter region with ϵ replaced by 0, for instance.)

This implies the inequality

$$\kappa \leq \frac{1}{\frac{1}{2} - \epsilon} + O_{\alpha,\epsilon}\left(\frac{m}{A}\right).$$

Thus taking ϵ so small so that $\frac{1}{\frac{1}{2}-\epsilon} < \kappa$, and then taking $m \gg_{\alpha,\epsilon} 1$ sufficiently large so that the required inequalities from above hold, *and then* taking $A \gg_{\alpha,\epsilon,m} 1$ sufficiently large so that the required inequalities above hold *and* so that $\frac{1}{\frac{1}{2}-\epsilon} + O_{\alpha,\epsilon}\left(\frac{m}{A}\right) < \kappa$, we arrive at a contradiction, completing the proof of Roth's theorem. □

After this incredibly difficult proof, we would be remiss if we did not present at least *some* application of our hard work. We will in fact do one better: we will prove the utterly spectacular fact that an elliptic curve can only have finitely many integral points.

CHAPTER

4

SIEGEL'S THEOREM

1. Siegel's theorem and a sketch of proof.

In this Chapter we will prove Siegel's theorem: that affine patches of elliptic curves can only have finitely many integral points. Specifically, we will prove the following theorem.[1]

THEOREM 1.1 (Siegel's theorem.). *Let $A, B \in \mathbb{Z}$ be such that $4A^3 + 27B^2 \neq 0$. Then $y^2 = x^3 + Ax + B$ has finitely many solutions $x, y \in \mathbb{Z}$.*

In particular this gives a qualitative response to Fermat's challenges concerning $y^2 = x^3 - n$ for some $n \in \mathbb{Z}$: for each fixed n, there are only finitely many such solutions. Finding all of them is a different matter, of course!

The idea will be as follows. Write $E : y^2 = x^3 + Ax + B$ for the curve in $\mathbb{P}^2$ determined by this equation. An infinite sequence of integral points $P_n = (x_n, y_n)$ of increasing height $h(P_n) := h(x_n) = \log |x_n|$ forms a sequence "converging" to the point at infinity, which is the origin in the group law. But this convergence is rather slow when measured in height. To speed it up, we use a tensor-power trick of sorts, in that we use the ability to scale on an elliptic curve to get much faster convergence. Specifically, by Mordell-Weil (which we won't bother proving here), there is a rational point R for which $P_n \equiv R \bmod 3$ for infinitely many n (hence without loss of generality all n). Write $P_n = 3Q_n + R$, with Q_n a rational point.

Then by just writing down explicit formulas for tripling and adding on the curve one sees that $h(3Q_n + R) = 9h(Q_n) + O_{E,R}(\sqrt{h(Q_n)})$. (Alternatively, one can use the theory of the canonical height.) Moreover, since the P_n were converging to ∞, the Q_n must be converging to some solution of $3\tilde{R} = -R$ (without loss of generality they all become near one $\tilde{R}$, since there are nine $\tilde{R}$'s and infinitely many P_n), a point over a degree 18 extension of $\mathbb{Q}$. In particular, its x-coordinate

[1] As a reference the reader is encouraged to look in Silverman's book [**31**], where much stronger, but less explicit, results are derived.

is algebraic. But the distance from P_n to ∞ (e.g., inside $\mathbb{C}$, via the universal covering map) is roughly $|x(P_n)|^{-\frac{1}{2}}$. Since $P_n = 3Q_n + R$, by writing down explicit formulas one sees that this is essentially $\prod_{3\tilde{R}=-R} \left|x(Q_n) - x(\tilde{R})\right|$. By Roth, except for finitely many exceptions, this is always at least

$$\gg_R \exp\left(-(2+\epsilon)h(Q_n)\right) \geq \exp\left(-\left(\frac{2}{9} + O_R(\epsilon)\right) h(P_n)\right)$$

in absolute value. On the other hand, it was roughly equal to $|x(P_n)|^{-\frac{1}{2}} = \exp\left(-\frac{1}{2}h(P_n)\right)$ by integrality. Therefore $\frac{1}{2} \leq \frac{2}{9} + O_R(\epsilon)$! Taking ϵ sufficiently small (with respect to R) results in a contradiction.

Let us now fill in the details to make this a rigorous argument.

2. Details.

PROOF OF SIEGEL'S THEOREM. Suppose otherwise. Let $P_n \in E(\mathbb{Z})$ be an infinite sequence of integral points with increasing height. By Mordell-Weil the group $E(\mathbb{Q})/3E(\mathbb{Q})$ is finite. By passing to a subsequence we may assume that all P_n are congruent modulo 3. Write e.g. $R := P_1$ (which is, without loss of generality, not a 6-torsion point), so that all $P_n \equiv R \bmod 3$. Thus we may write $P_n =: 3Q_n + R$, with each $Q_n \in E(\mathbb{Q})$.

By passing to another subsequence we may assume that, in the complex topology (e.g. via the uniformization $\wp : \mathbb{C} \twoheadrightarrow E(\mathbb{C})$), among the nine solutions to $3\tilde{R} = -R$ in $E(\bar{\mathbb{Q}})$, all the Q_n are closest to a fixed $\tilde{R} \in E(\bar{\mathbb{Q}})$. Note that $x(\tilde{R})$ is algebraic, since multiplication by 3 is algebraic on E.

REMARK 2.1. *In fact we have an explicit description of multiplication by n on an elliptic curve: there are universal polynomials ψ_m in A, B, x, y, called the* division polynomials *of E, for which*

$$x(nP) = x(P) - \frac{\psi_{n-1}(P)\psi_{n+1}(P)}{\psi_n^2(P)}.$$

These division polynomials are defined by

$$\begin{aligned}
\psi_0 &= 0,\\
\psi_1 &= 1,\\
\psi_2 &= 2y,\\
\psi_3 &= 3x^4 + 6Ax^2 + 12Bx - A^2,\\
\psi_4 &= 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3), \text{ and, inductively,}\\
\psi_{2m+1} &= \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3,\\
\psi_{2m} &= \left(\frac{\psi_m}{2y}\right) \cdot (\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2).
\end{aligned}$$

By our description of multiplication by n and an examination of the recurrence, we may obtain a precise description of ψ_n as a polynomial as follows. Its roots are precisely the nonzero n-torsion points. Its leading coefficient is n. It is degree $\frac{n^2-1}{2}$ when x is given weight 1 and y given weight $\frac{3}{2}$. In fact under this grading it is homogeneous of this degree if we also give A weight 2 and B weight 3.

By e.g. the explicit description in the remark (or on general principles, since both are functions vanishing to single order when $3Q_n = -R$), we find that

$$\log|x(P_n)|^{-\frac{1}{2}} = \sum_{3\tilde{R}=-R} \log\left|x(Q_n) - x(\tilde{R})\right| + O_{E,R}(1).$$

Indeed,

$$\begin{aligned}\prod_{3\tilde{R}=-R}\left(x(Q_n) - x(\tilde{R})\right) &= \psi_3(Q_n)^2(x(Q_n) - x(R)) - \psi_2(Q_n)\psi_4(Q_n)\\ &= \psi_3(Q_n)^2 \cdot (x(3Q_n) - x(R))\\ &= \psi_3(Q_n)^2 \cdot (x(P_n - R) - x(R))\\ &= \psi_3(Q_n)^2 \cdot \left(-x(P_n) - 2x(R) + \frac{(y(P_n) + y(R))^2}{(x(P_n) - x(R))^2}\right)\\ &= \psi_3(Q_n)^2 \cdot \frac{2y(P_n)y(R) + 3x(P_n)x(R)^2 + Ax(P_n) + y(R)^2 - 2x(R)^3 + B}{(x(P_n) - x(R))^2},\end{aligned}$$

where we have used the explicit formulas for multiplication by 3 and point addition on the curve. But now as $n \to \infty$ the dominant term in the numerator of the second factor is $y(P_n) \sim x(P_n)^{\frac{3}{2}}$. The denominator grows like $x(P_n)^2$, whence the claimed overall growth of this expression as $x(P_n)^{-\frac{1}{2}}$ times a constant independent of n (since $Q_n \to \tilde{R}$ for some $\tilde{R}$ such that $3\tilde{R} = -R$, which is a nonzero distance away from a 3-torsion point — hence it is bounded away, by a bound depending on E and R, from any zero or pole of ψ_3).

Finally, it remains to show that this passage from P_n to Q_n actually speeds up convergence. Note first that (again)

$$\begin{aligned}x(P_n - R) &= \frac{(y(P_n) + y(R))^2 - (x(P_n) - x(R))^2(x(P_n) + x(R))}{(x(P_n) - x(R))^2}\\ &= \frac{x(P_n)^2x(R) + 2y(P_n)y(R) + x(P_n)x(R)^2 + Ax(P_n) + B + y(R)^2 - x(R)^3}{(x(P_n) - x(R))^2}.\end{aligned}$$

Since these are integral points, we therefore have that the numerator and denominator of this last expression are both integral. Note that the numerator grows like $x(P_n)^2$ times a constant independent of n and the denominator grows like $x(P_n)^2$. Since they can only decrease if they share a factor, this tells us that

$$h(3Q_n) = h(P - R) \leq 2h(P_n) + O_{E,R}(1).^2$$

Now consider $h(3Q_n)$. Via our explicit formula from before, we have that

$$x(3Q_n) = x(Q_n) - \frac{\psi_2(Q_n)\psi_4(Q_n)}{\psi_3(Q_n)}, \tag{2.1}$$

[2] We have not introduced the canonical height

$$\hat{h}(P) := \lim_{k\to\infty} \frac{h(2^kP)}{4^k}$$

in order to remain self-contained. But the bound we have given is actually suboptimal, and we could improve it through the theory of the canonical height. Specifically, the canonical height induces a positive-definite inner product on $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}}$, and therefore satisfies the Cauchy-Schwarz inequality. Moreover it differs from the Weil height by a function bounded by a constant only depending on the curve E. Thus we would derive $h(P - R) \leq h(P) + O_{E,R}(\sqrt{h(P)})$, but anyway the suboptimal bound is enough for our purposes.

where

$$\psi_2(Q_n) = 2y(Q_n),$$
$$\psi_3(Q_n) = 3x(Q_n)^4 + 6Ax(Q_n)^2 + 12Bx(Q_n) - A^2, \text{ and}$$
$$\psi_4(Q_n) = 4y(Q_n)\left(x(Q_n)^6 + 5Ax(Q_n)^4 + 20Bx(Q_n)^3 - 5A^2x(Q_n)^2 - 4ABx(Q_n) - 8B^2 - A^3\right).$$

Note that the numerator of $x(3Q_n)$, $\psi_3(Q_n)^2x(Q_n) - \psi_2(Q_n)\psi_4(Q_n)$, is, upon using the relation $y^2 = x^3 + Ax + B$, monic of degree 9. The denominator is of degree 8. Therefore, on writing $Q_n =: \left(\frac{x}{D^2}, \frac{y}{D^3}\right)$ and clearing denominators, we see that the resulting numerator, an integer, is coprime to D. Thus any shared factor with the resulting denominator, also an integer, must be coprime to D.

The resultant of the resulting numerator and denominator (over $\mathbb{Z}$, as polynomials in x) is invariant under the change of variables $x \mapsto xD$.[3] Thus we may calculate this resultant in the case $D = 1$. In this case one may calculate explicitly that it is Δ_E^{12}, where $\Delta_E := -16(4A^3 + 27B^2)$ is the discriminant of the cubic equation. Alternatively without setting $D = 1$ one would get this multiplied by a power of D.[4] In any case, the greatest common divisor is therefore uniformly bounded in terms of E.

Now note that, since Q_n is bounded away from the zeroes of ψ_3, if $x(3Q_n) = x(P_n - R)$ is large it must be because $x(Q_n)$ is large. But $x(P_n - R)$ grows like $x(P_n)$ times a constant independent of n, simply from our explicit expression. Therefore $|x(Q_n)| = \frac{|x|}{|D|^2} \to \infty$. Therefore $h(Q_n) = \log|x|$ for n sufficiently large.

Hence upon clearing out factors of D in Equation (2.1), we see that the x^9 term dominates in the numerator, and the x^8D^2 term dominates in the denominator. Since the resulting numerator and denominator are essentially coprime (since their greatest common divisor is bounded independently of n), when written in least common terms $x(3Q_n)$ has height

$$h(3Q_n) \geq \max(9\log|x|, 8\log|x| + 2\log|D|) - O_E(1) = 9h(Q_n) - O_E(1),$$

where the $O_E(1)$ term is from the potential decrease in numerator and denominator due to the presence of a common factor.

Thus, in sum we see that

$$9h(Q_n) \leq 2h(P_n) + O_E(1).$$[5]

Let us now put all these estimates together. First, let us express the fact that the P_n are "converging to ∞." This is expressed by

$$\frac{1}{2} = \frac{\log|x(P_n)|^{\frac{1}{2}}}{h(P_n)}.$$

Second, let us express the fact that P_n converges to ∞ if and only if Q_n, defined by $3Q_n + R = P_n$, converges to a solution $\tilde{R}$ to $3\tilde{R} = -R$. By our estimate above, this

[3] More precisely, one immediately sees that the greatest common divisor over $\mathbb{Q}$ is 1. Clearing denominators one gets that a linear combination of the two polynomials is an integer, whence invariant under any change of variables.

[4] On general principles it is immediate that it should be some power of the discriminant. In any case, one can simply ask e.g. Mathematica to do the computation.

[5] Moreover, we could have removed the 2 with the use of the canonical height.

is expressed by

$$\log|x(P_n)|^{\frac{1}{2}} = \log\left(\prod_{3\tilde{R}=-R}\left|x(Q_n)-x(\tilde{R})\right|^{-1}\right) + O_{E,R}(1).$$

Third, let us express the fact that the convergence of Q_n to an $\tilde{R}$ is "faster" than that of P_n to ∞. This is expressed by

$$\frac{9}{2}h(Q_n) - O_E(1) \leq h(P_n),$$

or

$$\frac{1}{h(P_n)} \leq \frac{2}{9}\cdot\frac{1}{h(Q_n)} + o(1).$$

Putting these together, we derive the following chain of inequalities once n is large enough so that $\prod_{3\tilde{R}=-R}\left|x(Q_n)-x(\tilde{R})\right| \leq 1$.

$$\begin{aligned}\frac{1}{2} &= \frac{\log|x(P_n)|^{\frac{1}{2}}}{h(P_n)}\\ &= \frac{\log\left(\sum_{3\tilde{R}=-R}\left|x(Q_n)-x(\tilde{R})\right|^{-1}\right)}{h(P_n)} + o(1)\\ &\leq \left(\frac{2}{9}+o(1)\right)\frac{\log\left(\sum_{3\tilde{R}=-R}\left|x(Q_n)-x(\tilde{R})\right|^{-1}\right)}{h(Q_n)}.\end{aligned}$$

But, since there are nine *distinct* solutions to $3\tilde{R} = -R$, Q_n can only get very close to one of them to multiplicity *one* in this expression. Moreover, the closest it can get, except for finitely many exceptions, is

$$\left|x(Q_n)-x(\tilde{R})\right|^{-1} \leq H(Q_n)^{2+\epsilon},$$

for any $\epsilon > 0$. Therefore, except for finitely many exceptions,

$$\frac{\log\left(\sum_{3\tilde{R}=-R}\left|x(Q_n)-x(\tilde{R})\right|^{-1}\right)}{h(Q_n)} \leq 2+\epsilon+\frac{O_R(1)}{h(Q_n)} = 2+\epsilon+o(1).$$

Therefore we have derived the inequality

$$\frac{1}{2} \leq \frac{4+2\epsilon}{9} + o(1).$$

This is false. Thus the contradiction, and the theorem is proved. □

3. Remarks on the method of proof.

Much of the work in the preceding section was to control the Weil height under operations on the curve. This can be done in a much more succinct way using the canonical height, but actually the explicit formulas are of utmost importance in the philosophy behind the argument in the next Chapter. Specifically, it is an explicit computation of the x-coordinate of the difference of two integral points, as we have done twice now, that expresses the fact that *integral points* are quasiorthogonal in the Mordell-Weil lattice $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}}$ under the pairing given by the canonical height. This quasiorthogonality, analogous to the quasiorthogonality enjoyed by rational points on higher genus curves (as observed by Mumford), is the "reason" there are few integral points on an elliptic curve. We will see this fleshed out in what follows.

Moreover, as a side remark, almost every exposition of Siegel's theorem that the author has found proceeds in precisely the same way. These turn out to be difficult arguments to make explicit, simply because they pass from the Weil height to the canonical height and back multiple times, and this becomes problematic when looking for bounds uniform in the curve E (especially when dealing with points of small height). The presentation here errs on the side of being too explicit in order to be elementary and concrete. If nothing else, this gives a different angle on the argument.

CHAPTER

5

THE AVERAGE NUMBER OF INTEGRAL POINTS ON AN ELLIPTIC CURVE

1. Remarks on integral points.

We have now seen that the equation $E_{A,B} : y^2 = x^3 + Ax + B$, once the right-hand side has no double roots, has finitely many solutions in the integers. So then, how many?

First, there cannot be a uniform bound on the number of integral solutions.[1] This is because there are elliptic curves with infinitely many rational points.[2] Take such a curve $y^2 = x^3 + Ax + B$ and take an infinite sequence of rational points $\left(\frac{x_n}{D_n^2}, \frac{y_n}{D_n^3}\right)$ on it. Now observe that the curve

$$y^2 = x^3 + \left[A \cdot \prod_{n=1}^{N} D_n^4\right] x + \left[B \cdot \prod_{n=1}^{N} D_n^6\right]$$

has at least N integral solutions:

$$\left(x_n \cdot \left[\prod_{1 \le m \le N, m \ne n} D_m^2\right], y_n \cdot \left[\prod_{1 \le m \le N, m \ne n} D_m^3\right]\right) \text{ for } 1 \le n \le N.$$

Second, this number is not an invariant of the curve, but rather of the affine equation. So, while integral points may be extremely classical objects of study, one

[1] However, it may well be that if $(A, B) = 1$ a uniform bound might exist. We do not know what to expect one way or another.

[2] For instance, the curve $E : -139y^2 = x^3 + 10x^2 - 20x + 8$ is a famous example, but anyway these curves arise with positive probability (conjecturally $\frac{1}{2}$) if one chooses A and B at random, a very recent result of Bhargava and Skinner. [**9**]

has to work a bit to pick out integral points from rational points, since one cannot just e.g. evaluate the given functor of points on $\operatorname{Spec}\mathbb{Z}$. In general one can pick these points out via the height associated to an ample line bundle, like the one provided by the double cover $x : E \to \mathbb{P}^1$ (our "Weil height"), but we will just work completely explicitly. For the purposes of counting each elliptic curve once and only once, however, we will impose the further restriction that if $p^4|A$, then $p^6 \nmid B$. This associates exactly one such $E_{A,B}$ to each elliptic curve E over $\mathbb{Q}$.

Third, there is actually an entirely explicit bound on the number of integral points a curve can have. In fact, there is a bound on the height of any integral point! This is provided by Baker's method of lower bounds on linear forms in logarithms (and, more generally, lower bounds on linear forms in elliptic logarithms). The resulting bound is as follows.

THEOREM 1.1 (Baker, [**3**]). *Suppose $|A| \leq T^2$ and $|B| \leq T^3$. Let $(x, y) \in \mathbb{Z}^2$ be an integral solution of $y^2 = x^3 + Ax + B$. Then*

$$|x| \leq e^{(10^6 T)^{10^6}}.$$

For the purposes of counting points, this bound appears to be horrible. But this is only the case if one applies it naïvely. In fact, Helfgott-Venkatesh [**22**], in deriving their bounds, used Baker's result as a "stopping point" for their sphere packing argument. Since the bound of Helfgott-Venkatesh will be a crucial input for our arguments, in fact we will indirectly use Baker's result as well. The bound has since been improved by Stark and many others (see [**31**] for a discussion), but still remains superexponential in T. Its interest lies in the fact that it is effective, while Siegel's theorem was not. This was because Roth's theorem (or even Thue's theorem!) could not rule out a single approximation of tremendous height — indeed, recall that we had to work with multiple approximations to our algebraic number to get a contradiction — so that it could give no such bound on the heights of integral points.

Now let us discuss the pointwise bounds on the number of integral points of an elliptic curve that have been derived since Baker's theorem.

2. Pointwise bounds.

Silverman and Hindry-Silverman were the first to make Siegel's theorem explicit. In doing so, they proved the following theorems.

THEOREM 2.1 (Silverman, [**29**]). *There is an effective absolute constant C for which*

$$\#|E_{A,B}(\mathbb{Z})| \ll C^{\operatorname{rank}(E_{A,B})+\omega(\Delta)},$$

where $\omega(n)$ is the number of prime factors of n, and $\Delta_{A,B} = -16(4A^3 + 27B^2)$ is the discriminant of $E_{A,B}$.

THEOREM 2.2 (Hindry-Silverman, [**23**]). *There is an effective absolute constant C for which*

$$\#|E_{A,B}(\mathbb{Z})| \ll C^{\operatorname{rank}(E_{A,B})+\sigma_{E_{A,B}}},$$

where

$$\sigma_{E_{A,B}} := \frac{\log|\Delta_{A,B}|}{\log|N_{A,B}|}$$

is the Szpiro ratio *of $E_{A,B}$ (here $N_{A,B}$ is the conductor of $E_{A,B}$).*[3]

Conjecturally, the Szpiro ratio is always at most $6+\epsilon$. This is equivalent to the ABC conjecture. In any case, this absolute constant C is, in both cases, at least on the order of 10^{10}. In fact, even if one uses recent improvements to inputs to the arguments in Hindry-Silverman (due to Petsche [**28**], who obtained a better lower bound on the canonical height of a nontorsion point) one cannot make the constant smaller than this order of magnitude. On the other hand it is quite easy to show that most curves have Szpiro ratio at most, say, 100, so one might think that this makes the second bound amenable to averaging.

But finiteness of the average of $\left(10^{10}\right)^{\operatorname{rank}(E_{A,B})}$ is far out of the reach of current techniques.[4] Recent spectacular results of Bhargava-Shankar (which will feature centrally in this argument) have proven that the average of $5^{\operatorname{rank}(E_{A,B})}$ is finite — indeed, it is at most 6. This is the extent of current techniques. Let us restate this theorem for future reference below.

THEOREM 2.4 (Bhargava-Shankar, [**6, 5, 8, 7**]). *Let $n=2,3,4,$ or 5. Then the average size of the n-Selmer groups of the elliptic curves $E_{A,B}$, when ordered by height, is $\sigma(n)$, the sum of divisors of n.*

Here by *height* we mean

$$H(E_{A,B}) := \max(4|A|^3, 27|B|^2).$$

When combined with the inequality $n^{\operatorname{rank}(E_{A,B})} \leq \#|\mathrm{Sel}_n(E_{A,B})|$ provided by Galois cohomology, this implies that

$$\limsup_{T\to\infty} \frac{\sum_{H(E_{A,B})\leq T^6} n^{\operatorname{rank}(E_{A,B})}}{\sum_{H(E_{A,B})\leq T^6} 1} \leq \sigma(n).$$

This is what we mean when we say that the average of $5^{\operatorname{rank}(E_{A,B})}$ is at most 6 — in particular, an implicit $\limsup$ is to be understood.

Next, there is another bound on the number of integral points due to Helfgott and Venkatesh, who prove the following.

THEOREM 2.5 (Helfgott-Venkatesh, [**22**]). *Let S be the set of primes dividing the discriminant $\Delta_{A,B}$, along with ∞. Let $s := \#|S| = \omega(\Delta)+1$. Then*

$$\#|E_{A,B}(\mathbb{Z})| \ll O(1)^s \cdot (\log|\Delta|)^2 \cdot 1.33^{\operatorname{rank}(E_{A,B})}.$$

To achieve this, Helfgott and Venkatesh observe that, much as in the case of rational points on higher genus curves, integral points repel in the Mordell-Weil lattice $E_{A,B}(\mathbb{Q})/E_{A,B}(\mathbb{Q})_{\text{tors}}$. Specifically, one sees explicitly that points of close

[3] That is, $N_{A,B} = \prod_{p|\Delta} p^{e_p}$, where $e_p = 1$ if $E_{A,B}$ has multiplicative reduction at p, and $e_p \geq 2$ if $E_{A,B}$ has additive reduction at p, with equality if $p \neq 2,3$. We will not bother with the description at 2 or 3 — we will only use that $e_2 \leq 8$ and $e_3 \leq 5$.

[4] Assuming the Generalized Riemann Hypothesis (GRH) as well as the Birch and Swinnerton-Dyer conjecture (BSD), however, we may average a quantity like $\left(10^{10}\right)^{\operatorname{rank}(E_{A,B})}$, by the following result of Heath-Brown.

THEOREM 2.3 (Heath-Brown, [**20**]). *Assume GRH and BSD. Then the proportion of curves with rank at least R, when ordered by height, is $\ll R^{-\Omega(R)}$.*

Therefore our main result follows on GRH and BSD from the work of Heath-Brown and Hindry–Silverman. (One needs to show that most curves have nonnegligible conductor, but this is easy.)

height have nontrivial angle between them under the pairing determined by the canonical height. They then use bounds of Kabatiansky-Levenshtein on sphere packing to bound the number of integral points in dyadic annuli of the Mordell-Weil lattice. These dyadic annuli range all the way up to Baker's bound, at which point they may stop by Baker's theorem.

This latter result is amenable to averaging. Indeed, on combining it with the aforementioned theorem of Bhargava-Shankar and the observation that an integer X has at most $\ll \frac{\log X}{\log\log X}$ prime divisors, an application of Hölder's inequality gives the following.

COROLLARY 2.6 (Helfgott-Venkatesh). *Let $\epsilon > 0$. Then*

$$\#|E_{A,B}(\mathbb{Z})| \ll_\epsilon T^\epsilon \cdot 1.33^{\mathrm{rank}(E_{A,B})}.$$

Averaging this gives the following.

COROLLARY 2.7 (Helfgott-Venkatesh). *Let $\epsilon > 0$. Then the average number of integral points on the elliptic curves $E_{A,B}$, when ordered by height, is $\ll_\epsilon T^\epsilon$.*

Again, by this we mean that

$$\limsup_{T\to\infty} \frac{\sum_{H(E_{A,B})\leq T^6} \#|E_{A,B}(\mathbb{Z})|}{\sum_{H(E_{A,B})\leq T^6} 1} \ll_\epsilon T^\epsilon.$$

There are also various theorems of Heath-Brown [**19**], Bombieri–Pila [**10**], and others that control rational points of small height, but this bound resulting from the work of Helfgott-Venkatesh is the best that may be derived from the literature.

Finally, let us also mention a recent result of Bhargava-Gross which deals with rational points on higher genus hyperelliptic curves, for which Faltings's theorem plays the role of our Siegel's theorem. They obtain the following bound via invariant-theoretic means.

THEOREM 2.8 (Bhargava-Gross, [**4**]). *The average number of rational points on odd hyperelliptic curves of genus $g > 2$, when ordered by height, is $\ll 1$. Indeed, it is less than* 20.

3. This thesis.

It is a longstanding folklore conjecture that the number of integral points on a randomly chosen elliptic curve should be zero. This expectation was mentioned in Silverman's 1986 edition of *The Arithmetic of Elliptic Curves*, for instance. From the above, we see that the best bounds we have so far are that this number, conjecturally zero, does not grow too quickly.

In what follows we will prove that (the $\limsup$ of) this average is finite. The constant will be effective, but we will not calculate it here. The method is quite general and needs only an input of Bhargava-Shankar type that allows one to average $4^{\mathrm{rank}(E)}$ over a given family.[5] Moreover, if one allows a small, but growing, number of quadratic twists of these curves, it seems that the method can prove

[5] For instance, a recent result of Kane [**25**] allows one to average all moments of 2-Selmers over a certain family of quadratic twists of a curve with full rational 2-torsion. This improves on results of Heath-Brown [**17, 18**] and Swinnerton-Dyer [**33**]. The corresponding result for curves with no rational 2-torsion has been studied by Klagsbrun-Mazur-Rubin [**26**], but it appears a result of Kane type has not yet been achieved for these curves.

that the average is indeed zero, but we will not present this here. The method will be a blend of Helfgott-Venkatesh-type sphere packing, a uniform version of Siegel's theorem, and various improvements on bounds due to Silverman on the difference between the Weil and canonical heights for the "average" elliptic curve.

The numerology behind the result is as follows. First, the Kabatiansky-Levenshtein bound on the number of points on an $(n-1)$-sphere with pairwise angles at least 60° is $\ll 1.33^n$. Second, Siegel's theorem, as we have presented it, proceeds by applying Roth's theorem to $3^{\mathrm{rank}(E_{A,B})}$ many points of very large height. We will have to apply Kabatiansky-Levenshtein (in $E_{A,B}(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$, a $\mathrm{rank}(E_{A,B})$-dimensional Euclidean space) a constant number of times for each of these points, whence we will obtain a bound of shape

$$\ll 3^{\mathrm{rank}(E_{A,B})} \cdot 1.33^{\mathrm{rank}(E_{A,B})} = 3.99^{\mathrm{rank}(E_{A,B})}$$

in this range. For points of small height we will proceed by a simple counting argument. For points of intermediate height we will use Kabatiansky-Levenshtein a constant number of times. The worst situation in this range will occur when the heights of our points are too large to be bounded by our small height argument, but very small so that the sphere packing obtains the worst bounds. This will occur for integral points with x-coordinate on the order of T^5. It will turn out that the resulting bound gotten by Kabatiansky-Levenshtein will be $2.2^{\mathrm{rank}(E_{A,B})}$, which is fine for our purposes.

Thus the bottleneck will be averaging $3.99^{\mathrm{rank}(E_{A,B})}$ over our curves. But of course we can do this, as we have noted above. Notice that the argument would have failed had we written our large integral points P_n as $P_n = 4Q_n + R$ in the previous Chapter, since averaging $4^{\mathrm{rank}(E_{A,B})} \cdot 1.33^{\mathrm{rank}(E_{A,B})} = 5.32^{\mathrm{rank}(E_{A,B})}$ is out of reach. Moreover, we would not have been able to prove Siegel's theorem by writing $P_n = 2Q_n + R$. So in some sense things work out quite nicely for us.

In any case, let us get to the details. We will prove the following.

THEOREM 3.1. *The average number of integral points on elliptic curves, when ordered by height, is finite. In other words,*

$$\limsup_{T\to\infty} \frac{\sum_{H(E_{A,B})\leq T^6} \#|E_{A,B}(\mathbb{Z})|}{\sum_{H(E_{A,B})\leq T^6} 1} \ll 1.$$

As always, the sum is taken over A, B for which $\Delta := -16(4A^3 + 27B^2) \neq 0$ and such that there is no prime p for which $p^4|A$ and $p^6|B$. Note also that the denominator is of order T^5. Thus we must prove that the numerator is bounded by $\ll T^5$.

As a final remark, though we have presented the previous Chapters for a reader unfamiliar with the field, we will now assume that the reader is familiar with the theory of elliptic curves. For instance, we will freely use that the canonical height is quadratic and determines a positive-definite pairing on the Mordell-Weil lattice, as well as other standard facts.

4. Proof of the Main Theorem.

PROOF OF THEOREM 3.1. We will break the sum into three parts. The first part will account for the "trivial solutions". If we knew the Hall-Lang conjecture we would (almost) be able to stop there. The second part will account for the integral points with intermediate height. For these we will use sphere packing arguments. The third will account for integral points of large height. For these we will apply the estimate of Davenport-Roth [**14**], in the form given by Bombieri-Gubler [**11**], on the *number* of large-height approximations that may be present in the setting of Roth's theorem. Before we do this we will have to "repel" our integral point P away from the smallest integral point in its coset modulo 3, which we have written R, via sphere packing arguments. Once P is of large enough height as compared with R, Roth's Lemma will kick in and tell us that there are only a few possibilities left for P.

But first let us implement a few reductions. Fix a very small $\delta > 0$ for the rest of the argument. First, let us reduce to working with curves $E_{A,B}$ with A, B nearly coprime. This will guarantee that we have little additive reduction to worry about. For this, apply the bound of Helfgott-Venkatesh. Specifically,

$$\sum_{(A,B)\geq T^\delta} \#|E_{A,B}(\mathbb{Z})| \ll_\epsilon \sum_{(A,B)\geq T^\delta} T^\epsilon \cdot 1.33^{\mathrm{rank}(E_{A,B})}.$$

By Cauchy-Schwarz this is

$$\ll_\epsilon T^\epsilon \left(\sum_{(A,B)\geq T^\delta} 1\right)^{\frac{1}{2}} \left(\sum_{H(E_{A,B})\leq T^6} 1.7689^{\mathrm{rank}(E_{A,B})}\right)^{\frac{1}{2}}.$$

By Bhargava-Shankar,

$$\sum_{H(E_{A,B})\leq T^6} 1.7689^{\mathrm{rank}(E_{A,B})} \ll T^5.$$

Moreover, the first factor is

$$\sum_{(A,B)\geq T^\delta} 1 \ll \sum_{g=T^\delta}^{O(T^2)} \left(\frac{T^2}{g}\right)\left(\frac{T^3}{g}\right) \ll T^{5-\delta}.$$

Combining these two bounds gives an overall bound of

$$\sum_{(A,B)\geq T^\delta} \#|E_{A,B}(\mathbb{Z})| \ll_\epsilon T^{5-\frac{\delta}{2}+\epsilon}.$$

This is negligible in the average once $\epsilon < \frac{\delta}{4}$, say.

So we have reduced to the case of $(A, B) \leq T^\delta$. Next let us reduce to the case of Δ having no large square factor. This will again proceed via Helfgott-Venkatesh. To make our lives easy, we will quote the following theorem of Helfgott-Venkatesh, which will allow us to reduce to the case of Δ "mostly" squarefree, after which point reducing to Δ almost entirely squarefree will be easy. The theorem is as follows.

THEOREM 4.1 (Helfgott-Venkatesh, [**22**]). *The number of elliptic curves over $\mathbb{Q}$ with conductor N is $\ll N^{0.22378}$.*

Because our curves now have essentially no additive reduction, the conductor is essentially the radical of $\Delta \ll T^6$. Specifically, the powers of 2 and 3 appearing in the conductor are uniformly bounded (by 2^8 and 3^5, respectively). Moreover, for a bad prime $p \geq 5$, p appears with multiplicity one unless $E_{A,B}$ has additive reduction at p, in which case it appears with multiplicity two. Finally, $E_{A,B}$ has additive reduction at p if and only if $p|A$ and $p|B$. Therefore

$$\operatorname{rad}(\Delta) \leq N_{A,B} \ll \operatorname{rad}(\Delta) \cdot T^{\delta}.$$

Let us next show that we may reduce to the case of curves having conductor at least $T^{4.08}$. The sum over all the remaining curves is, again by Helfgott-Venkatesh,

$$\begin{aligned}\sum_{N_{A,B} \leq T^{4.08}} \#|E_{A,B}(\mathbb{Z})| &\ll T^{0.001} \cdot \sum_{N=1}^{T^{4.08}} \#|\{(A,B) : N_{A,B} = N\}| \\ &\ll T^{0.001} \cdot \sum_{N=1}^{T^{4.08}} N^{0.22378} \\ &\ll T^{4.08 \cdot 1.22378 + 0.001} \\ &\ll T^{4.995}.\end{aligned}$$

This is again negligible in the average.

So we may restrict to A, B for which both $(A,B) \leq T^{\delta}$ and $N_{A,B} \geq T^{4.08}$, whence $\operatorname{rad}(\Delta) \geq T^{4.08-\delta}$. Note in particular that if $g^2|\Delta$ then $g|\frac{\Delta}{\operatorname{rad}(\Delta)}$, whence $g \ll T^{1.92+\delta}$.

Now the sum over curves with Δ divisible by some g^2 with $g \geq T^{\frac{\delta}{2}}$ is

$$\ll \sum_{g=T^{\frac{\delta}{2}}}^{O(T^{1.92+\delta})} \sum_{A,B: g^2|\Delta} \#|E_{A,B}(\mathbb{Z})|.$$

As always, via Helfgott-Venkatesh and Hölder, this reduces to bounding

$$\sum_{g=T^{\frac{\delta}{2}}}^{O(T^{1.92+\delta})} \#|\{A, B : -16(4A^3 + 27B^2) \equiv 0 \pmod{g^2}\}|$$

by something of shape $\ll_{\epsilon} T^{5-\epsilon}$. But e.g. fixing A and g, if

$$-432B^2 \equiv 64A^3 \pmod{g^2},$$

then, writing g_0 for the prime-to-6 part of g, this implies that B is congruent to one of at most two square roots of $-\frac{64A^2}{432}$ modulo p^2 for each $p|g_0$. Thus there are at most $2^{\omega(g_0)}$ many congruence classes into which B may fall modulo g_0^2.

Now in general, given a subset $S \subseteq \mathbb{Z}/n\mathbb{Z}$, the size of the preimage of S under the reduction map $\{1, \ldots, N\} \to \mathbb{Z}/n\mathbb{Z}$ is

$$\ll \#|S| \left(\frac{N}{n} + 1\right).^{6}$$

Therefore the number of $|A| \ll T^2$ and $|B| \ll T^3$ for which $g^2|\Delta$ is at most

$$\ll T^2 \cdot 2^{\omega(g_0)} \cdot \left(\frac{T^3}{g_0^2} + 1\right) \ll \frac{T^{5+\epsilon}}{g_0^2} + T^{2+\epsilon},$$

since $\omega(g_0) \leq \omega(g) \leq \frac{\log g}{\log \log g} \leq \epsilon \log T + O_{\epsilon}(1)$.

[6] After all, the largest fibre of the map is over $1 \in \mathbb{Z}/n\mathbb{Z}$, of size $\lfloor \frac{N-1}{n} \rfloor + 1$.

Thus, summing this over g, we get that the sum over curves with Δ having a square factor of size at least T^{δ} is at most

$$\ll_{\epsilon} \left(T^{5+\epsilon} \cdot \sum_{g=T^{\frac{\delta}{2}}}^{O(T^{1.92+\delta})} \frac{1}{g_0^2} \right) + \left(T^{3.92+\epsilon} \right).$$

But the function $g \mapsto g_0$ is multiplicative, and via the usual Dirichlet series calculation using Perron's formula, one finds that the former sum is $\ll T^{-\delta}$.[7] Therefore the sum over curves with discriminant having a nonnegligible square factor is at most

$$\ll_{\epsilon} T^{5+\epsilon-\delta} + T^{3.92+\epsilon}.$$

This is negligible in the average once ϵ is chosen sufficiently small with respect to δ.

Finally, by precisely the same method, we may reduce to the case of $E_{A,B}$ having no rational torsion. Specifically, by Mazur's theorem the torsion is uniformly bounded. Moreover, Harron and Snowden [**16**] have determined the precise proportions of curves with given rational torsion subgroup, and as a corollary one sees that the proportion of curves having nontrivial rational torsion is at most $\ll T^{-2}$. Since this is certainly smaller than $T^{-\epsilon}$ for some ϵ, the same argument combining Helfgott-Venkatesh and Hölder's inequality works in this case as well.

Therefore we have reduced to considering curves with no rational torsion, with $(A, B) \leq T^{\delta}$, and with Δ having no square factor larger than T^{δ}. In particular $N_{A,B} \geq \mathrm{rad}(\Delta) \geq \Delta \cdot T^{-\delta}$, so that the conductor is quite large. The reason we have done this is because the most troublesome part of the difference between the Néron local height and the Weil local height at a point P and prime p of multiplicative reduction is roughly of the form $\frac{a(v_p(\Delta)-a)}{v_p(\Delta)}$, where $0 \leq a < v_p(\Delta)$ is the component of the special fibre of the Néron local model of $E_{A,B}$ at p onto which P reduces. In particular, if $v_p(\Delta) = 1$, this expression is forced to be zero, since a is a nonnegative integer smaller than 1.

In other words, we have reduced to the following situation. For us, most bad primes are of multiplicative reduction. Moreover, most primes of multiplicative reduction give no contribution to the difference between the canonical and Weil heights. Therefore the difference between the canonical and Weil heights must be quite small! (We will make this precise in a bit.) Hence any repulsion result between integral points in the Weil height translates to a corresponding result in the canonical height. Since we have already seen a repulsion result between integral points in the Weil height — we will restate this in due course — we therefore have a repulsion result between integral points in the canonical height, which allows us to apply sphere packing bounds. While this may seem like a slight technical nuisance, the point is that to bound the count of all integral points we must in particular bound the count of integral points of intermediate height, and the known uniform bounds on differences between Weil and canonical heights (due to Silverman) are simply too weak to allow for this.

[7] One may e.g. calculate the sum of the first X coefficients of $\zeta(s+2)\cdot(1-2^{-s-2})\cdot(1-3^{-s-2}) = \sum_{g\geq 1} \frac{g_0^{-2}}{g_0^s}$ as $\sum_{g=1}^{X} g_0^{-2} = \frac{\pi^2}{9} + O(X^{-1})$ because of the pole at $s = -1$. Of course one can extract the second-order term as well, but we will not need this here.

First let us precisely state the claimed improvement on Silverman's bounds in our situation.

LEMMA 4.2. *Let A, B be integers such that $|A|^2, |B|^3 \ll T^6$, $(A, B) \leq T^\delta$, and $\Delta = \Delta_{A,B}$ has no square factor larger than T^δ. Let $P \in E_{A,B}(\mathbb{Q})$ be a rational point on $E_{A,B}$. Then*

$$-O(1) - \frac{1}{6}\log^+ |j(E_{A,B})| - \frac{1}{6}\log|\Delta| \leq h(P) - \hat{h}(P) \leq \frac{1}{4}\log^+ |j(E_{A,B})| + 4\delta \log T + O(1),$$

where $j(E_{A,B}) = -1728\frac{(4A)^3}{\Delta}$ is the j-invariant of $E_{A,B}$ and

$$\hat{h}(P) := \lim_{k\to\infty} \frac{h(2^k P)}{4^k}$$

is the canonical height of E.

The upper bound is the claimed improvement. Note that the lower bound — in other words, the upper bound on the canonical height — is often thought to be the "easier" bound, so that the constants can likely be improved on this side as well.[8]

PROOF OF LEMMA 4.2. The lower bound is copied verbatim from Theorem 1.1 in Silverman's **[30]**. (Note that Silverman's normalizations differ from ours by a factor of 2.) The upper bound, while discovered in another manner (using a result of Stange which we will state during the discussion of constants), also follows from Silverman's methods. Specifically, write the difference as a sum of local differences:

$$h(P) - \hat{h}(P) = \sum_v \log^+ |x(P)|_v - \lambda_v(P).$$

At good primes p the local contribution is zero. At primes p of multiplicative reduction dividing the discriminant to order 1, by the result of Tate in the same paper (Theorem 4.1, part (b) of **[30]**), the local contribution is bounded above by

$$\log^+ |x(P)|_p - \lambda_p(P) \leq -\frac{1}{6}\log^+ |j(E_{A,B})|_p.$$

At primes p of additive reduction or of multiplicative reduction dividing the discriminant to order at least 2, by the result of Tate (Theorem 4.1, now part (a) of **[30]**) the local contribution is bounded by

$$\log^+ |x(P)|_p - \lambda_p(P) \leq \frac{1}{12}\log^+ |j(E_{A,B})|_p.$$

The remaining prime at ∞ has local contribution bounded by

$$\log^+ |x(P)| - \lambda_\infty(P) \leq \frac{1}{6}\log|\Delta| + \frac{1}{4}\log^+ |j(E_{A,B})| + O(1),$$

by Theorem 5.5 in the same paper **[30]**.

[8] In fact we are able to derive an even better upper bound for *integral* points of large height via the aforementioned result of Stange. This result replaces the given upper bound with $\epsilon \log T + O_\epsilon(1)$ for any $\epsilon > 0$. The idea is to first scale the integral point and then observe that *both* the Weil and canonical heights essentially scale correctly, up to a factor that is the greatest common divisor of certain combinations of division polynomials evaluated on the integral point. Since these are simply certain recurrences *in the integers*, it stands to reason that their prime factorizations can be controlled. Then one applies Silverman and divides. We have not used this here simply because the result is still very new and hence unpublished.

It remains to observe two things. First, that

$$\sum_p \log^+ |j(E_{A,B})|_p = \log|\Delta| - \log|(\Delta, A^3)| - O(1) \geq \log|\Delta| - 3\delta \log T - O(1).$$

This is from the product formula and our bound on the greatest common divisor of A and B. (Indeed, if $g|(\Delta, A^3)$, then $g|432B^2$, so that $g|(A^3, 432B^3)$, which, in turn, divides $432(A,B)^3$.) Second, that

$$\sum_{v_p(\Delta)>1} \log^+ |j(E_{A,B})|_p \leq 2\delta \log T.$$

This is because the left-hand side is at most the logarithm of $\prod_{v_p(\Delta)>1} p^{v_p(\Delta)}$, which divides

$$\left(\prod_{v_p(\Delta)>1} p^{\lfloor \frac{v_p(\Delta)}{2} \rfloor}\right)^2 \cdot \left(\prod_{v_p(\Delta)>1} p\right)^2,$$

a product of two square divisors of Δ, whence both at most T^δ.

Now we may sum the local contributions. We get:

$$\begin{aligned}
h(P) - \hat{h}(P) &= \sum_v \log^+ |x(P)|_v - \lambda_v(P) \\
&= \left(\log^+ |x(P)| - \lambda_\infty(P)\right) + \left(\sum_{v_p(\Delta)=1} \log^+ |x(P)|_p - \lambda_p(P)\right) \\
&\quad + \left(\sum_{v_p(\Delta)>1} \log^+ |x(P)|_p - \lambda_p(P)\right) \\
&\leq \left(\frac{1}{6}\log|\Delta| + \frac{1}{4}\log^+ |j(E_{A,B})| + O(1)\right) + \left(-\frac{1}{6}\sum_{v_p(\Delta)=1} \log^+ |j(E_{A,B})|_p\right) \\
&\quad + \left(\frac{1}{12}\sum_{v_p(\Delta)>1} \log^+ |j(E_{A,B})|_p\right) \\
&= \left(\frac{1}{6}\log|\Delta| + \frac{1}{4}\log^+ |j(E_{A,B})| + O(1)\right) + \left(-\frac{1}{6}\sum_{p|\Delta} \log^+ |j(E_{A,B})|_p\right) \\
&\quad + \left(\frac{1}{4}\sum_{v_p(\Delta)>1} \log^+ |j(E_{A,B})|_p\right) \\
&\leq \left(\frac{1}{6}\log|\Delta| + \frac{1}{4}\log^+ |j(E_{A,B})| + O(1)\right) + \left(-\frac{1}{6}\log|\Delta| + 3\delta \log T + O(1)\right) \\
&\quad + \left(\frac{\delta}{2}\log T\right) \\
&= \frac{1}{4}\log^+ |j(E_{A,B})| + 4\delta \log T + O(1),
\end{aligned}$$

as desired. □

We will in fact only use a very crude corollary of this lemma. Recall that the conductor was seen to be at least $T^{4.08}$. Therefore the discriminant is of order at least $T^{4.08}$. Since $|A| \leq T^2$, we must therefore have $|j(E_{A,B})| \ll T^{1.92}$. Thus we derive the following.[9]

COROLLARY 4.3. *Let A, B be integers such that $|A|^2, |B|^3 \ll T^6$, $(A, B) \leq T^\delta$, and $\Delta = \Delta_{A,B}$ has no square factor larger than T^δ. Suppose the conductor of the curve $N_{A,B} \geq T^{4.08}$. Let $P \in E_{A,B}(\mathbb{Q})$ be a rational point on $E_{A,B}$. Then*

$$-O(1) - 1.32 \log T \leq h(P) - \hat{h}(P) \leq 0.49 \log T + O(1)$$

once δ is sufficiently small.

We have finished the preliminaries. Let us now move to the proof.

Let $\epsilon > 0$. Break the sum into three parts:

$$\sum_{H(E_{A,B}) \leq T^6}^{\star} \#|E_{A,B}(\mathbb{Z})| \leq \mathrm{I}_\epsilon + \sum_{H(E_{A,B}) \leq T^6}^{\star} \mathrm{II}_{A,B} + \sum_{H(E_{A,B}) \leq T^6}^{\star} \mathrm{III}_{A,B},$$

where

$$\begin{aligned} \mathrm{I}_\epsilon &:= \#|\{(x, y, A, B) : 4|A|^3, 27|B|^2 \leq T^6, |x| \leq T^{5-\epsilon}, y^2 = x^3 + Ax + B\}|, \\ \mathrm{II}_{A,B} &:= \#|\{(x, y) : T^{5-\epsilon} \leq |x| \leq T^{10^{10}}, y^2 = x^3 + Ax + B\}|, \\ \mathrm{III}_{A,B} &:= \#|\{(x, y) : T^{10^{10}} \leq |x|, y^2 = x^3 + Ax + B\}|, \end{aligned}$$

and we have written $\sum^\star$ to remind the reader that we are now summing only over A, B for which $E_{A,B}$ has no rational torsion, $(A, B) \leq T^\delta$, $\Delta = -16(4A^3 + 27B^2)$ has largest square factor at most T^δ, and the conductor $N_{A,B} \geq T^{4.08}$.

LEMMA 4.4. *We have the following bound on the number of integral points of small height:*

$$\mathrm{I}_\epsilon \ll T^{5-\epsilon}.$$

Therefore the small height points are negligible.

PROOF OF LEMMA 4.4. First, note that B is determined by x, y, and A. Fix first x, A such that $|x| \leq 10^{10}T$, $|A| \ll T^2$. Then the number of y with $|y^2 - x^3 - Ax| \ll T^3$ is at most $\ll T^{\frac{3}{2}}$. Indeed, the given inequality implies that $y^2 \ll T^3$. Thus the count of such (x, y, A, B) is at most

$$\ll T \cdot T^2 \cdot T^{\frac{3}{2}} = T^{5-\frac{1}{2}},$$

whence negligible.

Hence we may assume $|x| > 10^{10}T$. In this case the constraints imply that $|y| \asymp |x|^{\frac{3}{2}}$. Now fix x for which $|x| > 10^{10}T$. Suppose y and y' are both positive and such that there exist A, B and A', B' such that (x, y, A, B) and (x, y', A', B') satisfy the constraints. Then

$$y^2 - y'^2 = (A - A')x + (B - B').$$

[9] The numbers 1.32 and 0.49 arise because $1.32 = 1 + \frac{1.92}{6}$ and $0.49 > 0.48 + 4\delta = \frac{1.92}{4} + 4\delta$ for δ sufficiently small.

Taking absolute values, we find that

$$|y-y'| \ll \frac{T^2}{|x|^{\frac{1}{2}}} + \frac{T^3}{|x|^{\frac{3}{2}}} \ll \frac{T^2}{|x|^{\frac{1}{2}}}.$$

Similarly if y and y' were both negative.

Therefore the number of y for which there exist such A and B is

$$\ll \frac{T^2}{|x|^{\frac{1}{2}}} + 1,$$

since our bound cannot rule out the existence of e.g. one positive y — it can only rule out the existence of two positive y's that are far apart.

Now, for fixed x and y, the number of A for which $|A| \ll T^2$ and $|y^2 - x^3 - Ax| \ll T^3$ is at most

$$\ll \frac{T^3}{|x|} + 1.$$

Indeed, we play the same game. If A and A' are two such solutions to this inequality, then by considering differences we see that

$$|A - A'| = \frac{|(y^2 - x^3 - Ax) - (y^2 - x^3 - A'x)|}{|x|} \ll \frac{T^3}{|x|}.$$

Combining these bounds, we see that, for fixed $|x| \geq 10^{10}T$, there are at most

$$\ll \left(\frac{T^2}{|x|^{\frac{1}{2}}} + 1\right) \cdot \left(\frac{T^3}{|x|} + 1\right)$$

choices of y, A, B such that (x, y, A, B) satisfies the constraints.

Therefore the number of points in range I_ϵ is at most

$$\mathrm{I}_\epsilon \ll T^{5-\frac{1}{2}} + \sum_{x=10^{10}T}^{T^{5-\epsilon}} \left(\frac{T^2}{|x|^{\frac{1}{2}}} + 1\right) \cdot \left(\frac{T^3}{|x|} + 1\right) \ll T^{5-\epsilon},$$

as desired. □

Before we bound the intermediate and large height points, let us establish the repulsion between integral points that we have mentioned over and over again.

LEMMA 4.5 (Helfgott-Mumford gap principle [**21**].). *Let $P \neq Q \in E_{A,B}(\mathbb{Z})$ be integral points such that $h(P) \geq h(Q) \geq (1+\epsilon)\log T$. Then*

$$h(P+Q) \leq 2h(P) + h(Q) + O(1).$$

If we replace h by $\hat{h}$ and take $4\log T \leq h(P) \leq h(Q) \leq (1+\epsilon)h(P)$, then this says that $\cos\theta_{P,Q} = \frac{\hat{h}(P+Q)-\hat{h}(P)-\hat{h}(Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}} \leq \frac{\sqrt{1+\epsilon}}{2} + \frac{O(1)}{\log T}$. Thus the angle between P and Q is at least $\approx 60°$! Thankfully passing to $\hat{h}$ from h will not lose us too much, and this result will still approximately be true.

PROOF OF LEMMA 4.5. Write $P =: (x, y)$ and $Q =: (X, Y)$. Then

$$\begin{aligned} x(P+Q) &= \frac{(y+Y)^2 - (x-X)^2(x+X)}{(x-X)^3} \\ &= \frac{x^2X + xX^2 + 2yY + A(x+X) + 2B}{(x-X)^2}. \end{aligned}$$

Therefore

$$\begin{aligned} h(P+Q) &\leq \max\left(\log|x^2X + xX^2 + 2yY + A(x+X) + 2B|, 2\log|x-X|\right) \\ &\leq \max\left(2h(P)+h(Q), h(P)+2h(Q), \frac{3}{2}h(P)+\frac{3}{2}h(Q), 2\log T + h(P), 3\log T, 2h(P)\right) \\ &\quad + O(1) \\ &= 2h(P)+h(Q)+O(1), \end{aligned}$$

as desired. (The second bound simply replaced a sum of terms by a constant multiple of the maximum of their absolute values.) □

COROLLARY 4.6. *Let A, B be such that $E_{A,B}$ has no rational torsion, $(A,B) \leq T^{\delta}$, the largest square factor of Δ is at most T^{δ}, and the conductor of $E_{A,B}$ is at least $N_{A,B} \geq T^{4.08}$. Let $P \neq Q \in E_{A,B}(\mathbb{Z})$ be integral points such that $(5-\epsilon)\log T \leq h(P) \leq h(Q) \leq (1+\delta)h(P)$. Then, in $E_{A,B}(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$ under the inner product induced by the canonical height, P and Q have angle at least*

$$\theta_{P,Q} \geq 0.628$$

once δ and ϵ are sufficiently small.

PROOF OF COROLLARY 4.6. If $\cos\theta_{P,Q} = \frac{\hat{h}(P+Q)-\hat{h}(P)-\hat{h}(Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}}$ is negative, we are done, since then the angle is at least $\frac{\pi}{2}$. Otherwise, recall the inequalities relating the Weil and canonical heights that we derived in Corollary 4.3. They imply that $\hat{h}(P+Q) \leq h(P) + 1.32\log T$, $\hat{h}(P) \geq h(P) - 0.49\log T$, and $\hat{h}(Q) \geq h(Q) - 0.49\log T$. Therefore

$$\begin{aligned} \cos\theta_{P,Q} &= \frac{\hat{h}(P+Q)-\hat{h}(P)-\hat{h}(Q)}{2\sqrt{\hat{h}(P)\hat{h}(Q)}} \\ &\leq \frac{\hat{h}(P+Q)-\hat{h}(P)-\hat{h}(Q)}{2\sqrt{h(P)h(Q)}} \cdot \left(1-\frac{0.49\log T}{h(P)}\right)^{-1} \\ &\leq \left(\frac{h(P+Q)-h(P)-h(Q)}{2\sqrt{h(P)h(Q)}} + \frac{2.3\log T}{2h(P)}\right) \cdot \left(1-\frac{0.49\log T}{h(P)}\right)^{-1} \\ &\leq \left(\frac{\sqrt{1+\epsilon}}{2} + \frac{2.3\log T}{2h(P)}\right) \cdot \left(1-\frac{0.49\log T}{h(P)}\right)^{-1}. \end{aligned}$$

Thus, since $h(P) \geq (5-\epsilon)\log T$, for ϵ sufficiently small we obtain the inequality

$$\cos\theta_{P,Q} \leq 0.8093.$$

Thus $\theta_{P,Q} \geq 0.628$, as desired. □

Now we may use sphere packing arguments to control points of intermediate and large height. For this we will need the following amazing bound of Kabatiansky-Levenshtein.

THEOREM 4.7 (Kabatiansky-Levenshtein, [**24**]). *Let $X \subseteq S^{n-1}$ be a collection of unit vectors v_i such that $\theta_{i,j} = \arccos\left(\frac{\langle v_i, v_j\rangle}{2}\right) \geq \theta$ for every $i \neq j$. Then*

$$\#|X| \leq \exp\left(\left[\frac{1+\sin\theta}{2\sin\theta}\log\left(\frac{1+\sin\theta}{2\sin\theta}\right) - \frac{1-\sin\theta}{2\sin\theta}\log\left(\frac{1-\sin\theta}{2\sin\theta}\right) + o(1)\right] n\right),$$

where $o(1) \to 0$ as $n \to \infty$.

Let us now bound points of intermediate height.

LEMMA 4.8. *Let A, B be such that $E_{A,B}$ has no rational torsion, $(A, B) \leq T^{\delta}$, $\Delta = \Delta_{A,B}$ has largest square factor at most T^{δ}, and the conductor of $E_{A,B}$ is bounded below by $N_{A,B} \geq T^{4.08}$. Suppose δ is sufficiently small so that Corollary 4.6 holds for some ϵ. Then we have the following bound on the number of integral points of intermediate height:*

$$\mathrm{II}_{A,B} \ll_{\delta} 2.2^{\mathrm{rank}(E_{A,B})}.$$

Thus the intermediate height points can be averaged.

PROOF OF LEMMA 4.8. Let $\epsilon > 0$ be sufficiently small so that Corollary 4.6 holds for δ and ϵ. Break the interval $[5-\epsilon, 10^{10}]$ into

$$K \leq 1 + \frac{\log\left(\frac{10^{10}}{5-\epsilon}\right)}{\log(1+\epsilon)} \leq O_{\epsilon}(1)$$

"dyadic" ranges $[N_i, N_{i+1}]$, where $N_{i+1} = (1+\epsilon)N_i$ except for at the last interval. Write

$$\mathrm{II}_{A,B} = \sum_{i=1}^{K} \mathrm{II}_{A,B}^{(i)},$$

with

$$\mathrm{II}_{A,B}^{(i)} := \#|\{P \in E_{A,B}(\mathbb{Z}) : h(P) \in [N_i \log T, N_{i+1} \log T]\}|.$$

We will show that

$$\mathrm{II}_{A,B}^{(i)} \ll 2.2^{\mathrm{rank}(E_{A,B})},$$

from which the result will follow.

Note that if $P \neq Q$ are both counted by $\mathrm{II}_{A,B}^{(i)}$, then we have seen that $\theta_{P,Q} \geq 0.628 =: \theta_0$. Similarly, $\theta_{P,-Q} \geq 0.628$ as well. Thus in particular P and Q are not both parallel. Let S_i be the set of projections of the $P \in \mathrm{II}_{A,B}^{(i)}$ to the unit sphere in $E_{A,B}(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$ under the inner product induced by the canonical height. Note that the map

$$P \mapsto P \otimes \frac{1}{\sqrt{\hat{h}(P)}} \in S_i$$

is injective, whence we need only upper bound the size of S_i.[10] This upper bound is provided for us by Kabatiansky-Levenshtein. Indeed,

$$\exp\left(\frac{1+\sin\theta_0}{2\sin\theta_0}\log\left(\frac{1+\sin\theta_0}{2\sin\theta_0}\right) - \frac{1-\sin\theta_0}{2\sin\theta_0}\log\left(\frac{1-\sin\theta_0}{2\sin\theta_0}\right)\right) \leq 2.17.$$

Therefore, since $o(n) \leq cn + O_c(1)$ for c sufficiently small (for us, so that $\exp(\log 2.17 + c) \leq 2.2$), the result follows. □

Finally, let us turn to points of large height. For these we will need the following result of Davenport-Roth, as improved by Bombieri-Gubler, on the number of approximations whose finiteness Roth's theorem guarantees.

[10] Here we use that the curve has no rational torsion.

THEOREM 4.9 (Davenport-Roth, Bombieri-Gubler). *Let $\kappa \geq 2.1$. Let $\alpha \in \bar{\mathbb{Q}}$ be of degree $r \leq 9$. Then the number of $\beta \in \mathbb{Q}$ such that both $h(\beta) > 10^5 h(\alpha)$ and*

$$|\alpha - \beta| \leq H(\beta)^{-\kappa}$$

is $\ll 1$.

PROOF OF THEOREM 4.9. See Bombieri-Gubler, section 6.5.7 (page 173). The reason such a uniform bound exists is because of the strong gap principle for good approximations that we have described. □

We will also need the following result of Mignotte bounding from below the distance between two roots of a polynomial with integral coefficients in terms of the height of the polynomial (i.e., its largest coefficient).

THEOREM 4.10 (Mignotte, [**27**]). *Let $P(x) \in \mathbb{Z}[x]$ be a separable polynomial of degree d. Let $\alpha_1, \ldots, \alpha_d$ be its roots. Then, for every $i \neq j$,*

$$|\alpha_i - \alpha_j| \gg_d ||P||_\infty^{-d+1},$$

where $||P||_\infty$ is the largest coefficient appearing in $P(x)$.

Now we may handle points of large height.

LEMMA 4.11. *Let A, B be such that $E_{A,B}$ has no rational torsion, $(A, B) \leq T^\delta$, $\Delta = \Delta_{A,B}$ has largest square factor at most T^δ, and the conductor of $E_{A,B}$ is bounded below by $N_{A,B} \geq T^{4.08}$. Suppose δ is sufficiently small so that Corollary 4.6 holds for some ϵ. Then we have the following bound on the number of integral points of large height:*

$$\mathrm{III}_{A,B} \ll_\delta 3.99^{\mathrm{rank}(E_{A,B})}.$$

Thus points of large height may be averaged.

PROOF OF LEMMA 4.11. Let $\epsilon > 0$ be sufficiently small so that Corollary 4.6 holds for δ and ϵ.

Write

$$\mathrm{III}_{A,B} = \sum_{R \in E_{A,B}(\mathbb{Q})/3E_{A,B}(\mathbb{Q})} \mathrm{III}_{A,B}^{(R)},$$

where

$$\mathrm{III}_{A,B}^{(R)} := \#|\{P \in E_{A,B}(\mathbb{Z}) : h(P) > 10^{10} \log T, P \equiv R \bmod 3\}|,$$

the sum taken over R an integral point of minimal height larger than $10^{10} \log T$ in its coset (if one does not exist, then of course $\mathrm{III}_{A,B}^{(R)} = 0$). There are at most $3^{\mathrm{rank}(E_{A,B})}$ such summands. We will show each

$$\mathrm{III}_{A,B}^{(R)} \ll_\delta 1.33^{\mathrm{rank}(E_{A,B})}.$$

Next, again, break the interval $[1, 10^{10}]$ into

$$K \leq 1 + \frac{\log(10^{10})}{\log(1+\epsilon)} \leq O_\epsilon(1)$$

"dyadic" ranges $[N_i, N_{i+1}]$, where $N_{i+1} = (1+\epsilon)N_i$ for all but the last interval. Then write

$$\mathrm{III}_{A,B}^{(R)} = \sum_{i=1}^{K} \mathrm{III}_{A,B}^{(R,i)} + \mathrm{III}_{A,B}^{(R,\infty)},$$

where

$$\mathrm{III}_{A,B}^{(R,i)} := \#|\{P \in E_{A,B}(\mathbb{Z}) : h(P) \in [N_i \log T, N_{i+1} \log T], P \equiv R \bmod 3\}|$$

and

$$\mathrm{III}_{A,B}^{(R,\infty)} := \#|\{P \in E_{A,B}(\mathbb{Z}) : h(P) > 10^{10} \log T, P \equiv R \bmod 3\}|.$$

It will be enough to show that

$$\mathrm{III}_{A,B}^{(R,i)} \ll 1.33^{\mathrm{rank}(E_{A,B})}$$

and

$$\mathrm{III}_{A,B}^{(R,\infty)} \ll 1.$$

The bound on the $\mathrm{III}_{A,B}^{(R,i)}$ is exactly the same as in Lemma 4.8, except that in the bound on the angle one may now use $h(P), h(R) > 10^{10} \log T$ instead of $h(P), h(R) > (5-\epsilon) \log T$. This changes the angle lower bound θ_0 to e.g. 0.50001, which results in a Kabatiansky-Levenshtein bound of at most $1.33^{\mathrm{rank}(E_{A,B})}$. Thus

$$\mathrm{III}_{A,B}^{(R,i)} \ll 1.33^{\mathrm{rank}(E_{A,B})}.$$

Therefore it remains only to bound $\mathrm{III}_{A,B}^{(R,\infty)}$. Let $P \in E_{A,B}(\mathbb{Z})$ be such that $h(P) > 10^{10} h(R)$. Write $P =: 3Q + R$ with $Q \in E_{A,B}(\mathbb{Q})$. We proceed as in Siegel. Namely, we begin with the simple equality

$$\frac{1}{2} = \frac{\log |x(P)|^{\frac{1}{2}}}{h(P)}.$$

Now

$$\begin{aligned} h(P) &\geq \hat{h}(P) - 1.32 \log T - O(1) \\ &= \hat{h}(3Q+R) - 1.32 \log T - O(1) \\ &\geq 9\hat{h}(Q) + \hat{h}(R) - \sqrt{\hat{h}(P-R)\hat{h}(R)} - 1.32 \log T - O(1) \\ &\geq 9\hat{h}(Q) + h(R) - \sqrt{\hat{h}(P-R) \cdot (1+10^{-7}) h(R)} - 1.81 \log T - O(1), \quad (4.1) \end{aligned}$$

where we have used Cauchy-Schwarz to bound the cross-term. But

$$\begin{aligned} \hat{h}(P-R) &\geq \hat{h}(P) - 2\sqrt{\hat{h}(P)\hat{h}(R)} + \hat{h}(R) \\ &\geq h(P) - 0.98 \log T - 2\sqrt{(h(P) + 1.32 \log T) h(R)} + h(R). \end{aligned}$$

Since $h(P) > 10^{10} h(R) > 10^{20} \log T$, this implies that $\hat{h}(P-R) \geq (1 - 2 \cdot 10^{-5}) h(P)$.

Thus

$$h(P) \geq (9 - 3 \cdot 10^{-5}) h(Q),$$

by (4.1) and the fact that $h(P) > 10^{10} h(R) > 10^{20} \log T$, so that the other terms on the right-hand side are bounded in terms of $h(P)$ (whence one can subtract and then divide to get the result).

Similarly, via

$$\begin{aligned} h(P) &\leq \hat{h}(P) + 0.49 \log T + O(1) \\ &\leq 9\hat{h}(Q) + h(R) + \sqrt{\hat{h}(P-R)\hat{h}(R)} + 0.49 \log T + O(1), \end{aligned}$$

we find that $h(Q) \geq 10^8 h(R) > 10^{18} \log T$.

Therefore we find that

$$\begin{aligned}\frac{1}{2} &\leq \left(\frac{1}{9}+10^{-5}\right)\cdot\frac{\log|x(P)|^{\frac{1}{2}}}{h(Q)}\\ &= \left(\frac{1}{9}+10^{-5}\right)\cdot\frac{\log\left(\prod_{3\tilde{R}=-R}\left|x(Q)-x(\tilde{R})\right|^{-1}\right)}{h(Q)}+\frac{1}{18}\cdot\frac{\log\left(|x(3Q+R)|\cdot\prod_{3\tilde{R}=-R}\left|x(Q)-x(\tilde{R})\right|^{2}\right)}{h(Q)}.\end{aligned}$$

Let us first show that the second term in this expression is negligible. This is plausible simply because $x(3Q+R)$ has a pole of order two at each $\tilde{R}$, so that the expression appearing in the logarithm is the absolute value of a rational function with no zero or pole at any $\tilde{R}$, and Q is quite close to an $\tilde{R}$ since P is quite close to ∞.

To see that it is small, first recall that we have seen an expression for $\prod_{3\tilde{R}=-R}\left(x(Q)-x(\tilde{R})\right)$. Specifically, it is

$$\prod_{3\tilde{R}=-R}\left(x(Q)-x(\tilde{R})\right)=(x(Q)-x(R))\psi_3(Q)^2-\psi_2(Q)\psi_4(Q),$$

where

$$\begin{aligned}&\psi_2(Q)=2y(Q),\\ &\psi_3(Q)=3x(Q)^4+6Ax(Q)^2+12Bx(Q)-A^2, \text{ and}\\ &\psi_4(Q)=4y(Q)\left(x(Q)^6+5Ax(Q)^4+20Bx(Q)^3-5A^2x(Q)^2-4ABx(Q)-8B^2-A^3\right),\end{aligned}$$

and we use the relation $y(Q)^2=x(Q)^3+Ax(Q)+B$ to write the expression as a polynomial in $x(Q)$. In particular this is a polynomial whose largest coefficient is $x(R)\cdot A^4 \ll |x(R)|\cdot T^8$, so that $||P||_\infty \ll H(R)\cdot T^8$.

Note that it is also equal to

$$\psi_3(Q)^2\cdot(x(3Q)-x(R)).$$

Therefore we see that

$$\begin{aligned}&x(3Q+R)\cdot\prod_{3\tilde{R}=-R}\left(x(Q)-x(\tilde{R})\right)^2\\ &=\psi_3(Q)^2\cdot\left((y(3Q)-y(R))^2-(x(3Q)-x(R))^2(x(3Q)+x(R))\right).\end{aligned}$$

But note that

$$\begin{aligned}x(3Q)&=x(P-R)\\ &=\frac{x(P)^2x(R)+x(P)x(R)^2+2y(P)y(R)+A(x(P)+x(R))+2B}{(x(P)-x(R))^2}\\ &\ll |x(R)|,\end{aligned}$$

since the dominant term in the numerator is $2x(P)^2x(R)$ and the dominant term in the denominator is $x(P)^2$. (After all, $|x(P)|>|x(R)|^{10^{10}}$.)

Therefore, since $y(3Q)^2=x(3Q)^3+Ax(3Q)+B$, we also have that $y(3Q)\ll |x(R)|^{\frac{3}{2}}$.

Thus we have found that

$$\begin{aligned}
&x(3Q+R)\cdot\prod_{3\tilde{R}=-R}\Big(x(Q)-x(\tilde{R})\Big)^2\\
&\ll \psi_3(Q)^2\cdot\big((y(3Q)-y(R))^2-(x(3Q)-x(R))^2(x(3Q)+x(R))\big)\\
&\ll |\psi_3(Q)|^2\cdot|x(R)|^3.
\end{aligned}$$

But now

$$x(3Q)=\frac{\psi_3(Q)^2x(Q)-\psi_2(Q)\psi_4(Q)}{\psi_3(Q)^2},$$

a ratio of a monic degree 9 polynomial in $x(Q)$ to a degree 8 polynomial in $x(Q)$. If $|x(Q)|>|x(R)|^2$, then the dominant terms in the numerator and denominator would be the leading terms of these polynomials, since we have explicit descriptions for them and their coefficients are negligible in comparison to $|x(R)|>T^{10^{10}}$. But then $|x(3Q)|\gg|x(R)|^2$ as well, a contradiction. Therefore $|x(Q)|\leq|x(R)|^2$. Thus the polynomial $\psi_3(Q)$ in $x(Q)$ is bounded by

$$\psi_3(Q)\ll|x(R)|^4.$$

Hence this implies that

$$x(3Q+R)\cdot\prod_{3\tilde{R}=-R}\Big(x(Q)-x(\tilde{R})\Big)^2\ll|x(R)|^{11}.$$

Taking this upper bound into account, we get that

$$\begin{aligned}
\frac{1}{2}&\leq\left(\frac{1}{9}+10^{-5}\right)\cdot\frac{\log|x(P)|^{\frac{1}{2}}}{h(Q)}\\
&=\left(\frac{1}{9}+10^{-5}\right)\cdot\frac{\log\left(\prod_{3\tilde{R}=-R}\left|x(Q)-x(\tilde{R})\right|^{-1}\right)}{h(Q)}+\frac{11}{18}\cdot\frac{\log|x(R)|}{h(Q)}.
\end{aligned}$$

The second term is at most 10^{-6}, for instance. So we are left with the first term. If

$$|x(Q)-x(\tilde{R})|\geq\frac{\min_{\tilde{R}\neq\tilde{R}'}|x(\tilde{R})-x(\tilde{R}')|}{2}\Big[\gg||P||_\infty^{-8}\gg H(R)^{-8}\cdot T^{-64}\Big]$$

(by Mignotte's bound) for each of the nine $\tilde{R}$'s, then the first term is bounded above by

$$\begin{aligned}
\left(\frac{1}{9}+10^{-5}\right)\cdot\frac{\log\left(\prod_{3\tilde{R}=-R}\left|x(Q)-x(\tilde{R})\right|^{-1}\right)}{h(Q)}&\leq\left(\frac{1}{9}+10^{-5}\right)\cdot\frac{72h(R)+576\log T+O(1)}{h(Q)}\\
&\leq 10^{-7}
\end{aligned}$$

once T is sufficiently large. Thus in this case we get $\frac{1}{2}\leq 2\cdot10^{-6}$, which is a contradiction.

Otherwise, there is an $\tilde{R}$ for which

$$|x(Q)-x(\tilde{R})|<\frac{\min_{\tilde{R}\neq\tilde{R}'}|x(\tilde{R})-x(\tilde{R}')|}{2}.$$

In this case, for each $\tilde{R}' \neq \tilde{R}$, we then have that

$$|x(Q) - x(\tilde{R}')| \geq \frac{\min_{\tilde{R} \neq \tilde{R}'} |x(\tilde{R}) - x(\tilde{R}')|}{2}.$$

Therefore in this case the first term is at most

$$\begin{aligned}
&\left(\frac{1}{9} + 10^{-5}\right) \cdot \frac{\log\left(\prod_{3\tilde{R}=-R} \left|x(Q) - x(\tilde{R})\right|^{-1}\right)}{h(Q)} \\
&\leq \left(\frac{1}{9} + 10^{-5}\right) \cdot \frac{\log|x(Q) - x(\tilde{R})|^{-1}}{h(Q)} + \left(\frac{1}{9} + 10^{-5}\right) \cdot \frac{64h(R) + 512\log T + O(1)}{h(Q)} \\
&\leq \left(\frac{1}{9} + 10^{-5}\right) \cdot \frac{\log|x(Q) - x(\tilde{R})|^{-1}}{h(Q)} + 10^{-7}.
\end{aligned}$$

In particular this implies that

$$|x(Q) - x(\tilde{R})| \leq H(Q)^{-4.4}.$$

Note also that $h(Q) > \left(\frac{1}{9} - 3 \cdot 10^{-5}\right) h(P) > 10^8 h(R)$ and

$$\begin{aligned}
h(R) &\geq \hat{h}(R) - 1.32\log T - O(1) \\
&\geq \frac{1}{9}\hat{h}(\tilde{R}) - 1.32\log T - O(1) \\
&\geq \frac{1}{9}h(\tilde{R}) - 2\log T - O(1),
\end{aligned}$$

so that $h(\tilde{R}) < 10h(R)$. Therefore $h(Q) > 10^7 h(\tilde{R})$. Thus Theorem 4.9 applies, implying that there are at most $\ll 1$ choices for $x(Q)$. Since there are at most two solutions $y(Q)$ to $y(Q)^2 = x(Q)^3 + Ax(Q) + B$, we therefore have that there are at most $\ll 1$ choices for Q. Since $3Q + R = P$, we therefore have that there are at most $\ll 1$ choices for P. That is to say, we have proven that

$$\text{III}_{A,B}^{(R,\infty)} \ll 1,$$

as desired. □

Therefore, combining Lemmas 4.4, 4.8, and 4.11, as well as our previous reductions, we see that, once T is sufficiently large (and δ and ϵ are chosen sufficiently small to begin the argument), the full sum is at most

$$\ll_\epsilon T^{5-\epsilon} + \sum_{H(E_{A,B}) \leq T^6} 3.99^{\operatorname{rank}(E_{A,B})} \ll_\epsilon T^5,$$

thanks to Bhargava-Shankar. This completes the proof. □

Having proved that the $\limsup$ of the average is finite and effectively bounded, let us discuss the resulting upper bound.

5. The constant.

First let us show that our bound is quite good if the proportion of curves with rank at least 2 is zero and the proportion of rank 0 and 1 curves is $\frac{1}{2}$ each, the "minimalist conjecture."[11] This will require a result of Stange.

[11] It seems difficult to improve this result, since (heuristically) the smallest generator of a rank one curve will, on average, have *logarithmic* height polynomial in T, by the Gross-Zagier formula. We have

THEOREM 5.1. *Assume the minimalist conjecture. Then the average elliptic curve has at most one integral point. That is,*

$$\limsup_{T\to\infty} \frac{\sum_{H(E_{A,B})\leq T^6} \#|E_{A,B}(\mathbb{Z})|}{\sum_{H(E_{A,B})\leq T^6} 1} \leq 1.$$

PROOF OF THEOREM 5.1. Combine the above bounds with Hölder to see that the contribution of the rank at least 2 curves is zero. Rank zero curves have only torsion points, and we showed that the average does not change upon restricting to curves with no rational torsion. So it suffices to show that rank 1 curves have, on average, at most two integral points. So let $E_{A,B}$ be of rank 1.

First, if $P := \left(\frac{x}{D^2}, \frac{y}{D^3}\right)$ has $|D| > 1$, then observe that

$$x(nP) = \frac{\psi_n(P)^2 x(P) - \psi_{n-1}(P)\psi_{n+1}(P)}{\psi_n(P)^2}.$$

The numerator is a monic polynomial in $x(P)$ of degree n^2. The denominator is a polynomial in $x(P)$ of degree n^2-1. Therefore clearing denominators of D gives a fraction of the form

$$x(nP) = \frac{x^{n^2} + (\in D^2\mathbb{Z})}{(\in D^2\mathbb{Z})},$$

which cannot be integral.

Therefore if $E_{A,B}$ has an integral point, its generator must also be integral. So let P be this generator. Note that $-P$ is also an integral point. So it suffices to show that no nontrivial multiple of P can be an integral point in our regime.

But we have already shown that it suffices to work with A, B for which there is no integral point P of height smaller than $(5-\epsilon)\log T$. Now we invoke a result of Stange, which arises from the study of elliptic divisibility sequences (like $\psi_n(P)$ for an integral point P).

THEOREM 5.2 (Stange, [**32**]). *Let $P \in E_{A,B}(\mathbb{Z})$ be an integral point. Let D^2 be the denominator (in lowest terms) of $x(nP)$. Then*

$$\log|D|^2 \leq 2\log|\psi_n(P)| \leq O(1) + \log|D|^2 + n^2 \sum_{p|\Delta} (-\log|\Delta|_p) \frac{a_p(P)(v_p(\Delta) - a_p(P))}{v_p(\Delta)^2},$$

where $0 \leq a_p(P) < v_p(\Delta)$ is the component of the special fibre of the Néron local model of $E_{A,B}$ onto which P reduces.

The upper bound on the denominator is immediate from the explicit formula, since it provides $x(nP)$ as a ratio of two integers with denominator $\psi_n(P)^2$. But the lower bound is quite powerful.

Recall that $v_p(\Delta) = 1$ except for a small proportion of the primes dividing Δ. This forces the last expression in the lower bound for $\log|D|$ to be quite small. Indeed, recall that we had reduced to the case of Δ with no square factor larger than T^δ for δ a small parameter we were free to choose.

only been able to rule out points of logarithmic height *logarithmic* in T by counting. Thus we would need to have control many orders of magnitude further than we have at the moment.

Given this reduction, the term in the upper bound is at most (since $x(1-x) \leq \frac{1}{4}$ on $[0,1]$)

$$\begin{aligned}\sum_{p|\Delta}(-\log|\Delta|_p)\frac{a_p(P)(v_p(\Delta)-a_p(P))}{v_p(\Delta)^2} &\leq \frac{1}{4}\sum_{v_p(\Delta)>1} -\log|\Delta|_p \\ &= \frac{1}{4}\log\left|\prod_{v_p(\Delta)>1} p^{v_p(\Delta)}\right| \\ &\leq \frac{\delta}{2}\log T,\end{aligned}$$

since, as we saw, $\prod_{v_p(\Delta)>1} p^{v_p(\Delta)}$ divides a product of two square factors of Δ.

Therefore as a result we find that the denominator of $x(nP)$, which we have written D^2, is bounded below by

$$\log|D|^2 \geq 2\log|\psi_n(P)| - \frac{\delta n^2}{2}\log T.$$

Since $|x(P)| \geq T^{5-\epsilon}$ and $\psi_n(P)^2$ is of degree n^2-1 in $x(P)$ with negligible coefficients (recall that it was homogeneous when A was given weight 2, B weight 3, and x weight 1), we see that $\log|D|^2 \geq \left(n^2\left(1-\frac{\delta}{2}\right)-1\right)\log T$. Once $|n| \geq 2$ and δ is sufficiently small this is a positive multiple of $\log T$, whence in particular nonzero. Thus $x(nP)$ is not integral, so that nP is not integral.

This completes the proof. □

Having stated this result of Stange, let us demonstrate a better lower bound on the canonical height in the situation of the "average" curve which improves the constants we get in our bounds. (Cf. Lemmas 4.2 and 4.3.)

LEMMA 5.3. *Let A, B be integers such that $|A|^2, |B|^3 \ll T^6$, $(A,B) \leq T^\delta$, and $\Delta = \Delta_{A,B}$ has no square factor larger than T^δ. Let $P \in E_{A,B}(\mathbb{Z})$ be an* integral *point on $E_{A,B}$ with $h(P) \geq (5-\epsilon)\log T$. Then*

$$-O_\delta(1) - \delta\log T \leq h(P) - \hat{h}(P) \leq \delta\log T + O_\delta(1).$$

PROOF OF LEMMA 5.3. Consider $h(nP) - \hat{h}(nP) = h(nP) - n^2\hat{h}(P)$. As we have seen, $x(nP)$ is a ratio of two polynomials in $x(P)$, and since $x(P)$ is so large the numerator and denominator are dominated by the leading terms. Since the numerator is degree n^2 and the denominator is degree n^2-1, $h(nP) \geq n^2h(P) - O(1) - \log|\gcd|$, where "gcd" denotes the greatest common divisor of this numerator and denominator. But Stange has proven that the reduced denominator is not so far from $\psi_n(P)^2$. That is, by Theorem 5.2 the gcd is at most

$$\log|\gcd| \leq O(1) + n^2\sum_{p|\Delta}(-\log|\Delta|_p)\frac{a_p(P)(v_p(\Delta)-a_p(P))}{v_p(\Delta)^2}.$$

As we have seen, the right-hand side reduces to

$$\log|\gcd| \leq O(1) + \frac{\delta n^2}{2}\log T.$$

Thus $h(nP) \geq n^2h(P) - \frac{\delta n^2}{2}\log T - O(1)$. Of course we also have the upper bound $h(nP) \leq n^2\log T + O(1)$, since the reduced numerator and denominator divide our two polynomials $\psi_n(P)^2x(P) - \psi_{n-1}(P)\psi_{n+1}(P)$ and $\psi_n(P)^2$, respectively.

Moreover, for general rational points on any elliptic curve we have bounds of the form $-O(1) - c\log T \leq h(nP) - \hat{h}(nP) \leq C\log T + O(1)$, where c and C are absolute constants, due e.g. to Silverman or our previous work (in which case we obtained 1.32 and 0.49, respectively).

Therefore we see that

$$\begin{aligned} h(P) - \hat{h}(P) &= \frac{1}{n^2}\left(n^2h(P) - \hat{h}(nP)\right) \\ &\leq \frac{1}{n^2}\left(\left[\frac{\delta n^2}{2}\log T\right] + O(1) + \left[h(nP) - \hat{h}(nP)\right]\right) \\ &\leq O(1) + \frac{\delta}{2}\log T + \frac{C}{n^2}\log T \end{aligned}$$

Taking $n \gg_{C,\delta} 1$, the desired bound follows. The reverse direction is precisely the same, except there we need not invoke the result of Stange. □

Now let us discuss the constant in the general situation, without supposing the minimalist conjecture.

Note that Lemma 5.3 improves the bound on the cosine of the angle between two integral points P and R with $h(R) < h(P)$ to, roughly,

$$\cos\theta_{P,R} \leq \frac{1}{2}\sqrt{\frac{h(P)}{h(R)}} + \frac{1.32}{2h(R)} + O(\delta).$$

The second term arises from the upper bound $\hat{h}(Q) \leq h(Q) + 1.32\log T + O(1)$ that we found earlier, and surely can be improved in our situation (remember Q is a general rational point, so that the result we have just proven cannot be applied).

In any case, let us suppose for heuristic reasons that the bound can be improved to $\cos\theta_{P,R} \leq \frac{1}{2}$.[12] Then the relevant sphere packing problem is the well-studied *kissing number* problem. In particular there are quite strong upper bounds for the kissing numbers in dimensions up to 24.

Now, the range I_ϵ does not contribute to the average. The range II contributes to the average only via the Kabatiansky-Levenshtein bound, which we will replace by the best kissing number bounds in dimension up to 24. The range III contributes to the average similarly, except here we need to average $3^{\mathrm{rank}(E_{A,B})}$ times the best known kissing number constant in dimension $\mathrm{rank}(E_{A,B})$.[13] By asking Mathematica to solve the relevant linear programming problem with constraints given by the Bhargava-Shankar bounds[14], it seems that, even if we allow ourselves the kissing number constant despite taking $h(P) < h(R) < 2h(P)$, say — i.e., so that the cosine of the angle between P and R is now only bounded above by $\cos\theta_{P,R} \leq \frac{1}{\sqrt{2}} + \epsilon$, we still find a constant of 96.4063 (and this is not allowing ranks higher than 24 to have nontrivial proportions).

[12] The strongest we could possibly expect is $\frac{1}{2} + \epsilon$, but this is just a heuristic discussion.

[13] The Roth contribution is in fact at most 2, since we end up with exponent approximately $4.5 > \sqrt{2 \cdot 9}$, whence the improvement to Thue-Siegel by Dyson and Gelfond applies, which needs only two good approximations for a contradiction. (Alternatively, the Wronskian method we have presented gives exponent $\sqrt{2 \cdot \deg(\alpha)}$ in two variables as well.)

[14] (— along with their lower bounds on the proportion of rank 0 and 1 curves.)

If we break into dyadic ranges $h(P) < h(R) < (4-\epsilon)h(P)$ and still allow ourselves the kissing number constants, the bound drops below 50, but this is rather unscientific.[15] In the more reasonable regime of $h(P) < h(R) < 1.1h(R)$ with kissing number constants, the resulting bound is below 670.

The author has not yet made the Kabatiansky-Levenshtein bound effective, mainly because this seems quite difficult. Thus the bound gotten is in principle effective, but we do not yet have a constant to show for it.

Finally, it seems allowing a small, but growing, number of quadratic twists will be able to improve the bound to zero, but we will not present this here.

Email address: `alpoge@college.harvard.edu`.
QUINCY HOUSE, HARVARD COLLEGE, CAMBRIDGE, MA 02138.

[15] Interestingly, the maximum is always achieved with all nontrivial probability put into rank 13. Note that the Kabatiansky-Levenshtein bound of 1.33^{dim} grows much slower than the growth of the kissing number upper bounds in "low" dimensions. They are as follows (starting in 2 dimensions, from [**13**]):

$$6, 12, 24, 44, 78, 134, 240, 364, 554, 870, 1357, 2069, 3183, 4866,$$
$$7355, 11072, 16572, 24812, 36764, 54584, 82340, 124416, 196560.$$

For instance, the final bound, 196560, is optimal in 24 dimensions, achieved by the Leech lattice. In comparison, $1.33^{24} = 938.5\ldots.$

LIST OF NOTATION

$\mathbb{Z}$	$= \{\ldots, -1729, \ldots, -1, 0, 1, \ldots, 691, \ldots\}$, the integers.
$\mathbb{Z}^+$	$= \{1, 2, \ldots, 65537, \ldots\}$, the positive integers.
$\mathbb{Q}$	$= \{\frac{p}{q} : (p, q) = 1, p \in \mathbb{Z}, q \in \mathbb{Z}^+\}$, the rational numbers.
$\mathbb{R}$	The real numbers.
$\mathbb{C}$	The complex numbers.
$f \ll_\theta g, f \leq_\theta O(g)$	There is a constant C, depending only on θ, such that, for all x, $\lvert f(x)\rvert \leq C\lvert g(x)\rvert$.
$f \gg_\theta g, f \geq \Omega_\theta(g)$	$g \ll_\theta f$.
$f \asymp_\theta g$	$f \ll_\theta g$ and $g \ll_\theta f$.
$f \leq o(g)$	$\frac{f}{g} \to 0$ as $n \to \infty$, with an implicit parameter $n \to \infty$ understood.
(a, b)	The greatest common divisor of a and b.
$\lvert \cdot \rvert, \lvert \cdot \rvert_\infty$	The usual absolute value on $\mathbb{C}$.
$v_p, v_\mathfrak{p}$	The p- and $\mathfrak{p}$-adic valuations on $\mathbb{Q}$ and a number field K, respectively.
v	A place (prime ideal or real or complex embedding) of a number field. Thus, over $\mathbb{Q}$, $v = \infty$ or $v = p$ for some prime p, by Ostrowski.
$\lvert \cdot \rvert_p, \lvert \cdot \rvert_\mathfrak{p}$	The p- and $\mathfrak{p}$-adic absolute values on $\mathbb{Q}$ and a number field K, respectively. Thus $\lvert x\rvert_p = p^{-v_p(x)}$ and $\lvert x\rvert_\mathfrak{p} = \mathcal{N}\mathfrak{p}^{-v_\mathfrak{p}(x)}$.
∂_i	$= \frac{\partial}{\partial x_i}$.
$\partial_{\vec{\mu}}$	$= \frac{\partial_1^{\mu_1}}{\mu_1!} \cdots \frac{\partial_m^{\mu_m}}{\mu_m!}$.
$\mathrm{ind}(P, (\xi_1, \ldots, \xi_m))$	$= \max\{\frac{\mu_1}{d_1} + \cdots + \frac{\mu_m}{d_m} : (\partial_{\vec{\mu}} P)(\xi_1, \ldots, \xi_m) = 0\}$. In particular, weights d_i will always be understood.
$E_{A,B}$	The elliptic curve $y^2 = x^3 + Ax + B$.
$\Delta, \Delta_{A,B}$	$= -16(4A^3 + 27B^2)$, the discriminant of $E_{A,B}$.
ψ_n	The n-th division polynomial of an elliptic curve $E_{A,B} : y^2 = x^3 + Ax + B$.
$h(x)$	The absolute Weil height of $x \in \bar{\mathbb{Q}}$, normalized so that, on rational $x = \frac{p}{q}$ in lowest terms, $h(x) = \max(\log \lvert p\rvert, \log \lvert q\rvert)$.
$H(x)$	$= \exp(h(x))$.
$h(P)$	$= h(x(P))$.
$H(P)$	$= H(x(P))$.
$\hat{h}(P)$	$= \lim_{k\to\infty} \frac{h(2^k P)}{4^k}$, the canonical height of P.

REFERENCES

[1] Translation taken from http://science.larouchepac.com/fermat/16590800%20Fermat%20to%20Carcavi.pdf.

[2] Levent Alpoge. Proof of a conjecture of stanley-zanello. *Journal of Combinatorial Theory, Series A*, 2014.

[3] A. Baker. Linear forms in the logarithms of algebraic numbers. I, II, III. *Mathematika 13 (1966), 204-216; ibid. 14 (1967), 102-107; ibid.*, 14:220–228, 1967.

[4] Manjul Bhargava and Benedict H. Gross. The average size of the 2-selmer group of jacobians of hyperelliptic curves having a rational weierstrass point. See http://arxiv.org/abs/1208.1007, preprint (2013).

[5] Manjul Bhargava and Arul Shankar. The average number of elements in the 4-selmer groups of elliptic curves is 7. See http://arxiv.org/abs/1312.7333, preprint (2013).

[6] Manjul Bhargava and Arul Shankar. The average size of the 5-selmer group of elliptic curves is 6, and the average rank is less than 1. See http://arxiv.org/abs/1312.7859, preprint (2013).

[7] Manjul Bhargava and Arul Shankar. Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves. See http://arxiv.org/abs/1006.1002, preprint (2010).

[8] Manjul Bhargava and Arul Shankar. Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0. See http://arxiv.org/abs/1007.0052, preprint (2010).

[9] Manjul Bhargava and Christopher Skinner. A positive proportion of elliptic curves over q have rank one. See http://arxiv.org/abs/1401.0233, preprint (2014).

[10] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.

[11] Enrico Bombieri and Walter Gubler. *Heights in Diophantine geometry*, volume 4 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2006.

[12] Andrew Bremner and Duncan A. Buell. Three points of great height on elliptic curves. *Math. Comp.*, 61(203):111–115, 1993.

[13] J. H. Conway and N. J. A. Sloane. *Sphere packings, lattices and groups*, volume 290 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, third edition, 1999. With additional contributions by E. Bannai, R. E. Borcherds, J. Leech, S. P. Norton, A. M. Odlyzko, R. A. Parker, L. Queen and B. B. Venkov.

[14] H. Davenport and K. F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2:160–167, 1955.

[15] Pierre Fermat. *Œuvres de Pierre Fermat. I*. Collection Sciences dans l'Histoire. [Science in History Collection]. Librairie Scientifique et Technique Albert Blanchard, Paris, 1999. La théorie des nombres. [Number theory], Translated by Paul Tannery, With an introduction and commentary by R. Rashed, Ch. Houzel and G. Christol.

[16] Robert Harron and Andrew Snowden. Counting elliptic curves with prescribed torsion. See http://arxiv.org/abs/1311.4920, preprint (2013).

[17] D. R. Heath-Brown. The size of Selmer groups for the congruent number problem. *Invent. Math.*, 111(1):171–195, 1993.

[18] D. R. Heath-Brown. The size of Selmer groups for the congruent number problem. II. *Invent. Math.*, 118(2):331–370, 1994. With an appendix by P. Monsky.

[19] D. R. Heath-Brown. The density of rational points on curves and surfaces. *Ann. of Math. (2)*, 155(2):553–595, 2002.

[20] D. R. Heath-Brown. The average analytic rank of elliptic curves. *Duke Math. J.*, 122(3):591–623, 2004.

[21] H. A. Helfgott. On the square-free sieve. *Acta Arith.*, 115(4):349–402, 2004.

[22] H. A. Helfgott and A. Venkatesh. Integral points on elliptic curves and 3-torsion in class groups. *J. Amer. Math. Soc.*, 19(3):527–550 (electronic), 2006.

[23] M. Hindry and J. H. Silverman. The canonical height and integral points on elliptic curves. *Invent. Math.*, 93(2):419–450, 1988.

[24] G. A. Kabatjanskiĭ and V. I. Levenšteĭn. Bounds for packings on the sphere and in space. *Problemy Peredači Informacii*, 14(1):3–25, 1978.

[25] Daniel Kane. On the ranks of the 2-Selmer groups of twists of a given elliptic curve. *Algebra Number Theory*, 7(5):1253–1279, 2013.

[26] Zev Klagsbrun, Barry Mazur, and Karl Rubin. A markov model for selmer ranks in families of twists. See http://arxiv.org/abs/1303.6507, preprint (2013).

[27] Maurice Mignotte. On the distance between the roots of a polynomial. *Appl. Algebra Engrg. Comm. Comput.*, 6(6):327–332, 1995.

[28] Clayton Petsche. Small rational points on elliptic curves over number fields. *New York J. Math.*, 12:257–268 (electronic), 2006.

[29] Joseph H. Silverman. A quantitative version of Siegel's theorem: integral points on elliptic curves and Catalan curves. *J. Reine Angew. Math.*, 378:60–100, 1987.

[30] Joseph H. Silverman. The difference between the Weil height and the canonical height on elliptic curves. *Math. Comp.*, 55(192):723–743, 1990.

[31] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[32] Katherine E. Stange. Integral points on elliptic curves and explicit valuations of division polynomials. See http://arxiv.org/abs/1108.3051, preprint (2013).

[33] Peter Swinnerton-Dyer. The effect of twisting on the 2-Selmer group. *Math. Proc. Cambridge Philos. Soc.*, 145(3):513–526, 2008.