Algebra & Number Theory

Volume 2
2008
No. 2

A finiteness property of torsion points

Matthew Baker, Su-ion Ih and Robert Rumely

mathematical sciences publishers

A finiteness property of torsion points

Matthew Baker, Su-ion Ih and Robert Rumely

Let k be a number field, and let G be either the multiplicative group $\mathbb{G}_m/k$ or an elliptic curve E/k. Let S be a finite set of places of k containing the archimedean places. We prove that if $\alpha \in G(\bar{k})$ is nontorsion, then there are only finitely many torsion points $\xi \in G(\bar{k})_{\mathrm{tors}}$ that are S-integral with respect to α. We also formulate conjectural generalizations for dynamical systems and for abelian varieties.

Introduction

Let k be a number field, with ring of integers $\mathcal{O}_k$ and algebraic closure $\bar{k}$. In this paper we prove finiteness theorems for torsion points that are integral with respect to a given nontorsion point, for the multiplicative group $\mathbb{G}_m/k$ and for elliptic curves E/k. We then attempt to place these results in a conceptual framework, and conjecture generalizations to dynamical systems and abelian varieties.

Let S be a finite set of places of k containing the archimedean places. Given $\alpha, \beta \in \mathbb{P}^1(\bar{k})$, let $\mathrm{cl}(\alpha), \mathrm{cl}(\beta)$ be their Zariski closures in $\mathbb{P}^1_{\mathcal{O}_k}$. By definition, β is S-integral relative to α if $\mathrm{cl}(\beta)$ does not meet $\mathrm{cl}(\alpha)$ outside S. Thus, β is S-integral relative to α if and only if for each place v of k not in S, and each pair of k-embeddings $\sigma : k(\beta) \hookrightarrow \bar{k}_v$, $\tau : k(\alpha) \hookrightarrow \bar{k}_v$, we have $\|\sigma(\beta), \tau(\alpha)\|_v = 1$ under the spherical metric on $\mathbb{P}^1(\bar{k}_v)$. Equivalently, for all σ, τ,

$$\begin{cases} |\sigma(\beta) - \tau(\alpha)|_v \geq 1 & \text{if } |\tau(\alpha)|_v \leq 1, \\ |\sigma(\beta)|_v \leq 1 & \text{if } |\tau(\alpha)|_v > 1. \end{cases}$$

Theorem 0.1. *Let k be a number field, and let S be a finite set of places of k containing all the archimedean places. Fix $\alpha \in \mathbb{P}^1(\bar{k})$ with Weil height $h(\alpha) > 0$; that is, identifying $\mathbb{P}^1(\bar{k})$ with $\bar{k} \cup \{\infty\}$, α is not 0 or ∞ or a root of unity. Then there are only finitely many roots of unity in $\bar{k}$ that are S-integral with respect to α.*

Similarly, let E/k be an elliptic curve, and let $\mathscr{E}/\mathrm{Spec}(\mathcal{O}_k)$ be a model of E.

MSC2000: primary 11G05; secondary 11J71, 11J86, 37F10, 11G50.
Keywords: elliptic curve, equidistribution, canonical height, torsion point, integral point.
Work supported in part by NSF grant DMS-0300784.

Theorem 0.2. *Let k be a number field, and let S be a finite set of places of k containing all the archimedean places. If $\alpha \in E(\bar{k})$ is nontorsion (has canonical height $\hat{h}(\alpha) > 0$), there are only finitely many torsion points $\xi \in E(\bar{k})_{\mathrm{tors}}$ which are S-integral with respect to α.*

By S-integrality we mean that the Zariski closures of ξ and α in $\mathscr{E}/\mathrm{Spec}(\mathcal{O}_k)$ do not meet outside fibres above S. Since any two models are isomorphic outside a finite set of places, it follows from the theorem that the finiteness property is independent of the choice of the set S and the model $\mathscr{E}$.

The main ingredients of the proofs of Theorems 0.1 and 0.2 are linear forms in logarithms (Baker's theorem for $\mathbb{G}_m$, and David/Hirata-Kohno's theorem for elliptic curves), properties of local height functions, and a strong form of equidistribution for torsion points at all places v. In outline, both theorems are proved as follows. By base change, one reduces to the case where α is rational over k. Given a place v of k, let $\bar{k}_v$ be the algebraic closure of the completion k_v, and let λ_v be the normalized canonical local height occurring in the decomposition of the global height. On the one hand, well known properties of local and global heights can be used to show that since α is nontorsion, for any torsion point ξ_n one has

$$0 < \hat{h}(\alpha) = \frac{1}{[k(\xi_n):k]} \sum_v \sum_{\sigma : k(\xi_n)/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)), \tag{1}$$

where $\sigma : k(\xi_n)/k \hookrightarrow \bar{k}_v$ means σ is an embedding of $k(\xi_n)$ in $\bar{k}_v$ fixing k. On the other hand, if $\{\xi_n\}$ is a sequence of distinct torsion points which are S-integral with respect to α, then for each v, by equidistribution and the normalization of λ_v,

$$\lim_{n\to\infty} \frac{1}{[k(\xi_n):k]} \sum_{\sigma : k(\xi_n)/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)) = 0. \tag{2}$$

By the integrality hypothesis, the outer sum in (1) can be restricted to $v \in S$, allowing the limit and the sum to be interchanged. This gives $\hat{h}(\alpha) = 0$, contradicting the assumption that α is nontorsion.

Examples show that the conclusion is false if α is a torsion point, and that it can fail if $\{\xi_n\}$ is merely a sequence of small points (that is, a sequence of points with $\hat{h}(\xi_n) \to 0$). In particular, our results cannot be strengthened to theorems of Bogomolov type.

The paper is divided into three sections. In Section 1, we prove Theorem 0.1 for $\mathbb{G}_m$; in Section 2, we prove Theorem 0.2 for elliptic curves. In Section 3, we attempt to provide perspective on these results, comparing them with other arithmetic finiteness theorems, and formulating conjectural generalizations.

Throughout the paper, we use the following notation. For each place v of k, let k_v be the completion of k at v and let $|x|_v$ be the normalized absolute value which

coincides with the modulus of additive Haar measure on k_v. If v is archimedean and $k_v \cong \mathbb{R}$, then $|x|_v = |x|$, while if $k_v \cong \mathbb{C}$ then $|x|_v = |x|^2$. If v is nonarchimedean and lies over the rational prime p, then $|p|_v = p^{-[k_v:\mathbb{Q}_p]}$. For $0 \neq \alpha \in k$, the product formula reads

$$\prod_v |\alpha|_v = 1.$$

If $\bar{k}_v$ is an algebraic closure of k_v, there is a unique extension of $|x|_v$ to $\bar{k}_v$, also denoted $|x|_v$. Given a finite extension L/k, for each place w of L we have the normalized absolute value $|x|_w$ on L_w. If we embed L_w in $\bar{k}_v$, then $|x|_w = |x|_v^{[L_w:k_v]}$ for each $x \in L_w$. Write $\log x$ for the natural logarithm of x. Given $\beta \in L$ and a place v of k, as σ ranges over all embeddings of L into $\bar{k}_v$ fixing k we have

$$\sum_{\sigma: L/k \hookrightarrow \bar{k}_v} \log |\sigma(\beta)|_v = \sum_{w \mid v} \log |\beta|_w . \tag{3}$$

The absolute Weil height of $\alpha \in k$ (also called the naive height) is defined to be

$$h(\alpha) = \frac{1}{[k:\mathbb{Q}]} \sum_v \max(0, \log |\alpha|_v),$$

with the convention that $\log 0 = -\infty$. It is well known that for $\alpha \in \overline{\mathbb{Q}}$, $h(\alpha)$ is independent of the field k containing $\mathbb{Q}(\alpha)$ used to compute it, so h extends to a function on $\overline{\mathbb{Q}}$. Furthermore $h(\alpha) \geq 0$, with $h(\alpha) = 0$ if and only if $\alpha = 0$ or α is a root of unity.

1. The finiteness theorem for $\mathbb{G}_m$

Limitations. Before giving the proof of Theorem 0.1, we note some examples that limit possible strengthenings of the theorem.

(A) The hypothesis $h(\alpha) > 0$ is necessary. To see this, take $k = \mathbb{Q}$. If $\alpha = 0$ or $\alpha = \infty$, then each root of unity ζ_n is integral with respect to α at all finite places. If $\alpha = 1$, then each root of unity whose order is divisible by at least two distinct primes is integral with respect to α at all finite places. If $\alpha = \zeta_N$ is a primitive N-th root of unity with $N > 1$, let ζ_m be a primitive m-th root of unity with $(m, N) = 1$ and $m > 1$. Then $\zeta_N^{-1}\zeta_m$ is a primitive mN-th root of unity whose order is divisible by at least two distinct primes, so $1 - \zeta_N^{-1}\zeta_m$ is a unit in $\overline{\mathbb{Z}}$, the ring of all algebraic integers, and $\zeta_N - \zeta_m$ is also a unit. This holds for all conjugates of ζ_N and ζ_m. Hence ζ_m is integral with respect to α at all finite places.

(B) When $h(\alpha) > 0$, one can ask if the theorem could be strengthened to a result of Bogomolov type: is there a number $B = B(\alpha) > 0$ such that there are only finitely

many points $\beta \in \bar{k}$ with $h(\beta) < B$ which are S-integral with respect to α? That is, could finiteness for roots of unity be strengthened to finiteness for small points?

The following example[1] shows this is not possible (see [Autissier 2006] for similar examples). Take $k = \mathbb{Q}$, $\alpha = 2$, and $S = \{\infty\}$. For each n, let β_n be a root of the polynomial

$$f_n(x) \;=\; x^{2^n-1}(x-2) - 1\,.$$

Here $f_n(x+1)$ is Eisenstein with respect to the prime $p = 2$, so $f_n(x)$ is irreducible over $\mathbb{Q}$. Note that each β_n is a unit. By Rouché's theorem, β_n has one conjugate very near 2 and the rest of its conjugates very close to the unit circle; this can be used to show that $\lim_{n\to\infty} h(\beta_n) = 0$. Finally, $\beta_n - 2$ is also a unit, so β_n is integral with respect to 2 at all finite places.

Proof of Theorem 0.1. By replacing k with $k(\alpha)$, and S with the set of places $S_{k(\alpha)}$ lying over S, we are reduced to proving the theorem when $\alpha \in k$. Indeed, if ζ is a root of unity which is S-integral with respect to α over k, then each k-conjugate of ζ is $S_{k(\alpha)}$-integral with respect to α over $k(\alpha)$.

Suppose $\alpha \in k$, and that there are infinitely many distinct roots of unity ζ_n which are S-integral with respect to α. For each n, we will evaluate the sum

$$A_n \;=\; \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{v \text{ of } k} \sum_{\sigma : k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v) \tag{4}$$

in two different ways. On the one hand, we will see that each $A_n = 0$. On the other hand, by applying the integrality hypothesis, A. Baker's theorem on linear forms in logarithms, and a strong form of equidistribution for roots of unity, we will show that $\lim_{n\to\infty} A_n = h(\alpha) > 0$. This contradiction will give the desired result. The details are as follows.

First, using (3), formula (4) can be rewritten as

$$A_n \;=\; \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{w \text{ of } k(\zeta_n)} \log|\zeta_n - \alpha|_w\,.$$

Since α is not a root of unity, we have $\zeta_n - \alpha \neq 0$; hence the product formula gives $A_n = 0$.

Next, take $v \notin S$. If $|\alpha|_v > 1$, we have $|\sigma(\zeta_n) - \alpha|_v = |\alpha|_v$ for each $\sigma : k(\zeta_n)/k \hookrightarrow \bar{k}_v$, by the ultrametric inequality. On the other hand, if $|\alpha|_v \le 1$, the integrality hypothesis gives $|\sigma(\zeta_n) - \alpha|_v = 1$. It follows that for each $v \notin S$

$$\frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{\sigma : k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v) \;=\; \frac{1}{[k:\mathbb{Q}]} \max(0, \log|\alpha|_v), \tag{5}$$

[1] The authors thank Pascal Autissier for correcting an error in an earlier version of this example.

so

$$A_n = \sum_{v\in S} \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{\sigma: k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v) + \frac{1}{[k:\mathbb{Q}]} \sum_{v\notin S} \max(0, \log|\alpha|_v).$$

Now let $n \to \infty$. Since S is finite, we can interchange the limit and the sum over $v \in S$, obtaining

$$0 = \sum_{v\in S} \left(\lim_{n\to\infty} \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{\sigma: k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v) \right) + \frac{1}{[k:\mathbb{Q}]} \sum_{v\notin S} \max(0, \log|\alpha|_v).$$

We will now show that for each $v \in S$,

$$\lim_{n\to\infty} \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{\sigma: k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v) = \frac{1}{[k:\mathbb{Q}]} \max(0, \log|\alpha|_v). \quad (6)$$

Inserting this in the previous equation gives $h(\alpha) = 0$, a contradiction.

For each nonarchimedean $v \in S$, (6) is trivial if $|\alpha|_v > 1$ or $|\alpha|_v < 1$. In the first case $|\sigma(\zeta_n) - \alpha|_v = |\alpha|_v$ for all n and all σ, and in the second case $|\sigma(\zeta_n) - \alpha|_v = 1$ for all n and all σ. Hence we can assume that $|\alpha|_v = 1$. We can then apply the following result, part (i) of which is a special case of the Tate–Voloch conjecture for semiabelian varieties proved by Scanlon [1999].

Lemma 1.1. *Let v be nonarchimedean, and suppose $|\alpha|_v = 1$. Then*

(i) *there is a bound $M(\alpha) > 0$ such that $|\zeta - \alpha|_v \geq M(\alpha)$ for all roots of unity $\zeta \in \bar{k}_v$ and*

(ii) *for each $0 < r < 1$, there are only finitely many roots of unity $\zeta \in \bar{k}_v$ with $|\zeta - \alpha|_v < r$.*

Proof. Since α is not a root of unity, (i) follows immediately from (ii). For (ii), note that if ζ and ζ' are roots of unity with $|\zeta - \alpha|_v < r$ and $|\zeta' - \alpha|_v < r$, then $|\zeta - \zeta'|_v < r$ and so $\zeta'' = \zeta^{-1}\zeta'$ is a root of unity with $|1 - \zeta''|_v < r$. There are only finitely many such ζ''. Indeed, if p is the rational prime under v, the only roots of unity $\xi \in \bar{k}_v$ with $|1 - \xi|_v < 1$ are those with order p^n for some n. If ξ is a primitive p^n-th root of unity, then $|1 - \xi|_v = p^{-[k_v:\mathbb{Q}_p]/p^{n-1}(p-1)}$ so $1 > r > |1 - \xi|_v$ for only finitely many n. □

Assuming v is nonarchimedean and $|\alpha|_v = 1$, let $M(\alpha)$ be as in Lemma 1.1. Fix $0 < r < 1$, and let $N(r)$ be the number of roots of unity in $\bar{k}_v$ with $|\zeta - \alpha|_v < r$.

For each ζ_n and each $\sigma : k(\zeta_n)/k \to \bar{k}_v$, we have $|\sigma(\zeta_n) - \alpha|_v \leq 1$, so

$$0 \geq \lim_{n\to\infty} \frac{1}{[k(\zeta_n):\mathbb{Q}]} \sum_{\sigma : k(\zeta_n)/k \hookrightarrow \bar{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v)$$
$$\geq \lim_{n\to\infty} \frac{1}{[k(\zeta_n):\mathbb{Q}]} \big(([k(\zeta_n):k] - N(r)) \log r + N(r) \cdot \log M(\alpha)\big) = \frac{1}{[k:\mathbb{Q}]} \log r.$$

Since $r < 1$ is arbitrary, the limit in (6) is 0, verifying (6) in this case.

Now suppose v is archimedean. To simplify notation, view k as a subfield of $\mathbb{C}$ and identify $\bar{k}_v$ with $\mathbb{C}$. (Thus, the way k is embedded depends on the choice of v.)

By Jensen's formula [Conway 1973, p. 280] applied to $f(z) = z - \alpha$,

$$\frac{1}{2\pi} \int_0^{2\pi} \log|e^{i\theta} - \alpha|\, d\theta \;=\; \max(0, \log|\alpha|). \tag{7}$$

Here $|x|$ can be replaced by $|x|_v$, since $|x|_v$ is either $|x|$ or $|x|^2$.

The $\mathrm{Gal}(\bar{k}/k)$-conjugates of roots of unity equidistribute in the unit circle. We will give a direct proof of this below, but we note that it also follows from Bilu's theorem [1997] and restriction of scalars, or from the equidistribution theorem for polynomial dynamical systems given in [Baker and Hsia 2005]. Those theorems show that if μ_n is the discrete measure

$$\mu_n \;=\; \frac{1}{[k(\zeta_n):k]} \sum_{\sigma : k(\zeta_n)/k \hookrightarrow \mathbb{C}} \delta_{\sigma(\zeta_n)}(x),$$

where $\delta_P(x)$ is the Dirac measure with mass 1 at P, then the μ_n converge weakly to the Haar measure $\mu = (1/2\pi)d\theta$ on the unit circle.

If $|\alpha|_v > 1$ or $|\alpha|_v < 1$ then $\log|z-\alpha|_v$ is continuous on the unit circle. In these cases, (6) follows from (7) and weak convergence. If $|\alpha|_v = 1$ then $\log|z-\alpha|_v$ is not continuous on $|z| = 1$ and weak convergence is not enough to give $\int_{|z|=1} \log|z-\alpha|_v\, d\mu_n(z) \to 0$: there could be a problem if some conjugate of ζ_n were extremely close to α,or if too many conjugates of ζ_n clustered near α.

The first problem is solved by A. Baker's theorem on lower bounds for linear forms in logarithms [Baker 1975, Theorem 3.1, p. 22]. We are assuming that $|\alpha|_v = 1$, and α is not a root of unity. Fix a branch of log with $\log z = \log|z| + i\theta$, for $-\pi < \theta \leq \pi$, and write $\log\alpha = i\theta_0$. For another branch, $\log 1 = 2\pi i$. The following is a special case of Baker's theorem. (In his statement of the theorem, Baker uses an exponential height having bounded ratio with $H(\beta) = e^{h(\beta)}$.)

Proposition 1.2 (A. Baker). *There is a constant $C = C(\alpha) > 0$ such that for each $\beta = a/N \in \mathbb{Q}$, with $a, N \in \mathbb{Z}$ coprime,*

$$|\, i\theta_0 - \beta \cdot 2\pi i\,| \;\geq\; e^{-C\cdot\max(1,h(\beta))},$$

where $h(\beta) = \log\max(|a|, |N|)$ is the Weil height of β.

The second problem is settled by a strong form of equidistribution for roots of unity, proved starting on page 226. It says that for any $0 < \gamma < 1$, the conjugates of the ζ_n are asymptotically equidistributed in arcs of length $[k(\zeta_n) : k]^{-\gamma}$. Note that weak convergence is equivalent to equidistribution in arcs of fixed length.

Proposition 1.3 (Strong equidistribution). *Let $k \subset \mathbb{C}$ be a number field. Then the $\mathrm{Gal}(\bar{k}/k)$-conjugates of the roots of unity in $\bar{k}$ (viewed as embedded in $\mathbb{C}$) are strongly equidistributed in the unit circle, in the following sense.*

Given an arc I in the unit circle, write $\mu(I) = \frac{1}{2\pi}\mathrm{length}(I)$ for its normalized Haar measure. If $\zeta \in \bar{k}$ is a root of unity, put

$$N(\zeta, I) \;=\; \#\big\{\sigma(\zeta) \in I : \sigma \in \mathrm{Gal}(\bar{k}/k)\big\}.$$

Fix $0 < \gamma < 1$. Then for all roots of unity ζ and all I,

$$\frac{N(\zeta, I)}{[k(\zeta) : k]} \;=\; \mu(I) + O_{k,\gamma}([k(\zeta) : k]^{-\gamma}). \tag{8}$$

Assuming Proposition 1.3, we will now complete the proof of Theorem 0.1 by showing that (6) holds for archimedean v such that $|\alpha|_v = 1$.

Let $\mu = (1/2\pi)\, d\theta$ be the normalized Haar measure on the unit circle, and for each n, put

$$\mu_n \;=\; \frac{1}{[k(\zeta_n) : k]} \sum_{\sigma : k(\zeta_n)/k \to \mathbb{C}} \delta_{\sigma(\zeta_n)}(x).$$

Then the μ_n are supported on the unit circle and converge weakly to μ as $n \to \infty$. We must show that

$$\int_{|z|=1} \log|z - \alpha|\, d\mu_n(z) \;=\; \frac{1}{[k(\zeta_n) : k]} \sum_{\sigma} \log(|\sigma(\zeta_n) - \alpha|) \;\to\; 0.$$

The idea is to split the integrand $\log|z - \alpha|$ into two parts: a continuous "background" function that can be handled by weak convergence, and a function with a logarithmic pole at α supported in a small neighborhood of α. The terms nearest α can then be dealt with using Baker's theorem, while the other terms can be treated by strong equidistribution. Define

$$\mathrm{larg}_{\alpha,\varepsilon}(z) \;= \min\left(0, \log(|\theta - \theta_0|/\varepsilon)\right),$$

taking $\mathrm{larg}_{\alpha,\varepsilon}(\theta_0) = -\infty$. Then there is a continuous function $g_{\alpha,\varepsilon}(z)$ on $|z| = 1$ for which $\log|z - \alpha| = \mathrm{larg}_{\alpha,\varepsilon}(z) + g_{\alpha,\varepsilon}(z)$.

Fix $0 < \epsilon < 1$. We will show that for all sufficiently large n,

$$\left|\int_{|z|=1} \log|z - \alpha|\, d\mu_n(z)\right| \;<\; 6\epsilon. \tag{9}$$

Note that $\int_0^\varepsilon \log(t/\varepsilon)\,dt = -\varepsilon$. For the remainder of the proof, we restrict to $|z|=1$; write $\alpha = e^{i\theta_0}$ where $-\pi < \theta_0 \le \pi$, and write $z = e^{i\theta}$ where $\theta_0 - \pi < \theta \le \theta_0 + \pi$. Recalling that $\int_{|z|=1} \log|z-\alpha|\,d\mu(z) = 0$, we have

$$\int_{|z|=1} g_{\alpha,\varepsilon}(z)\,d\mu(z) \;=\; -\int_{|z|=1} \operatorname{larg}_{\alpha,\varepsilon}(z)\,d\mu(z) \;=\; -2\int_0^\varepsilon \log(\theta/\varepsilon)\,\frac{d\theta}{2\pi} \;=\; \frac{\varepsilon}{\pi}.$$

By weak convergence, it follows that for all sufficiently large n,

$$\left|\int_{|z|=1} g_{\alpha,\varepsilon}(z)\,d\mu_n(z)\right| \;<\; \varepsilon. \tag{10}$$

To obtain (9), it suffices to show that for all sufficiently large n,

$$\left|\int_{|z|=1} \operatorname{larg}_{\alpha,\varepsilon}(z)\,d\mu_n(z)\right| \;<\; 5\varepsilon.$$

For each interval $[c,d]$ let $I_\alpha([c,d])$ be the arc $\{\alpha e^{2\pi i t} : t \in [c,d]\}$. Noting that $\operatorname{larg}_{\alpha,\varepsilon}(z)$ is supported on $I_\alpha([-\varepsilon,\varepsilon])$, put $D = D_n = \lceil [k(\zeta_n):k]^{1/2} \rceil$ and divide $I_\alpha([-\varepsilon,\varepsilon])$ into $2D$ equal subarcs. Taking $\gamma = 2/3$ in Proposition 1.3, it follows that if n is sufficiently large, each such subarc contains at most $2\varepsilon[k(\zeta_n):k]^{1/2}$ conjugates of ζ_n.

First consider the union of the two central subarcs, $I_\alpha([-\varepsilon/D, \varepsilon/D])$. Let N be the order of ζ_n. Let $\sigma_0(\zeta_n) = e^{2\pi i a/N}$ be the conjugate of ζ_n closest to $\alpha = e^{i\theta_0}$. We can assume that $|a/N| \le 1$, which implies that $h(a/N) = \max(\log|a|, \log N) = \log N$. By Baker's theorem,

$$|2\pi(a/N) - \theta_0| \;>\; e^{-C\max(1,\log N)}.$$

Hence if n is sufficiently large,

$$\operatorname{larg}_{\alpha,\varepsilon}(\sigma_0(\zeta_n)) \;>\; -C\log N - \log\varepsilon \;\ge\; -C\log N.$$

Since there are at most $4\varepsilon[k(\zeta_n):k]^{1/2}$ conjugates of ζ_n in $I_\alpha([-\varepsilon/D, \varepsilon/D])$,

$$0 \;\ge\; \int_{I_\alpha([-\varepsilon/D,\varepsilon/D])} \operatorname{larg}_{\alpha,\varepsilon}(|z-\alpha|)\,d\mu_n(z) \;>\; -4\frac{C\log N}{[k(\zeta_n):k]^{1/2}}\varepsilon.$$

Note that $[k(\zeta_n):k] \ge [\mathbb{Q}(\zeta_n):\mathbb{Q}]/[k:\mathbb{Q}] = \varphi(N)/[k:\mathbb{Q}]$. For all large N, $\varphi(N) \ge N^{1/2}$, so there is a constant B such that $[k(\zeta_n):k]^{1/2} \ge BN^{1/4}$. Thus for all sufficiently large n,

$$\left|\int_{I_\alpha([-\varepsilon/D,\varepsilon/D])} \log|z-\alpha|\,d\mu_n(z)\right| \;<\; \varepsilon. \tag{11}$$

Finally, consider the remaining subarcs. For $\ell = 1, \ldots, D-1$, if

$$z \in I_\alpha\big([\ell\varepsilon/D, (\ell+1)\varepsilon/D]\big) \quad \text{or} \quad z \in I_\alpha\big([-(\ell+1)\varepsilon/D, -\ell\varepsilon/D]\big)$$

then $0 \geq \operatorname{larg}_{\alpha,\varepsilon}(z) \geq \log(\ell/D)$. As before, by Proposition 1.3, for sufficiently large n, each subarc contains at most $2\,[k(\zeta_n):k]\,(\varepsilon/D)$ conjugates of ζ_n. It follows that

$$\begin{aligned} 0 \;&\geq \int_{I_\alpha([-\varepsilon,\varepsilon])\setminus I_\alpha([-\varepsilon/D,\varepsilon/D])} \operatorname{larg}_{\alpha,\varepsilon}(z)\,d\mu_n(z) \\ &\geq 2\cdot\sum_{\ell=1}^{D-1}\log\big((\frac{\ell\varepsilon}{D})/\varepsilon\big)\cdot\frac{2\varepsilon}{D} \;>\; 4\int_0^{\varepsilon}\log(t/\varepsilon)\,dt \;=\; -4\epsilon\,. \end{aligned} \tag{12}$$

Combining (10), (11), and (12) gives (9), which completes the proof of Theorem 0.1. □

In the course of writing this paper, the authors learned of several results related to Theorem 0.1, some of which imply it in special cases.

A. Bang's theorem [1886] says that if $\alpha \neq \pm 1$ is a nonzero rational number, then for all sufficiently large integers n there is a prime p such that the order of α modulo p is exactly n. This can be rephrased as saying that for all sufficiently large n, there exists a primitive n-th root of unity ζ_n and a nonzero prime ideal $\mathfrak{p}$ of $\mathbb{Z}[\zeta_n]$ such that $\alpha \equiv \zeta_n \pmod{\mathfrak{p}}$. Since all primitive n-th roots are conjugate over $\mathbb{Q}$, this implies Theorem 0.1 in the case $\alpha \in \mathbb{Q}$. A. Schinzel [1974] gave an effective generalization of Bang's theorem to arbitrary number fields; Schinzel's theorem implies Theorem 0.1 for number fields k which are linearly disjoint from the maximal cyclotomic field $\mathbb{Q}^{ab}$, and $\alpha \in k$.

J. Silverman [1995] has shown that if $\alpha \in \overline{\mathbb{Q}}$ is an algebraic unit which is not a root of unity, there are only finitely many m for which $\Phi_m(\alpha)$ is a unit, where $\Phi_m(x)$ is the m-th cyclotomic polynomial. In fact, if $d = [\mathbb{Q}(\alpha):\mathbb{Q}]$ he shows there is an absolute, effectively computable constant C such that the number of such m's is at most

$$C\cdot d^{1+0.7/\log\log d}\,.$$

In the case when α is a unit, this yields Theorem 0.1 in the same situations as Schinzel's theorem.

G. Everest and T. Ward [1999, Lemma 1.10] show that if $F(x) \in \mathbb{Z}[x]$ is monic and irreducible, with roots $\alpha_1, \ldots, \alpha_d$, and if $F(x)$ is not a constant multiple of x or a cyclotomic polynomial $\Phi_m(x)$, then the quantity $\Delta_n(F) = \prod_{i=1}^{d}(\alpha_i^n - 1)$ satisfies

$$\lim_{n\to\infty}\frac{1}{n}\log\Delta_n(F) \;=\; m(F) \;>\; 0\,, \tag{13}$$

where $m(F) = \deg(F)\cdot h(\alpha_i)$ is the logarithm of the Mahler measure of $F(x)$. When $k = \mathbb{Q}$, and $\alpha = \alpha_1$ is an algebraic integer, the product formula tells us that $\prod_{v \text{ of } \mathbb{Q}} |\Delta_n(F)|_v = 1$, so for all large n there must be some nonarchimedean v and some α_i such that $|\alpha_i^n - 1|_v < 1$, and this in turn means there is some n-th root of unity ζ with $|\alpha_i - \zeta|_v < 1$. This implies there are infinitely many roots of unity

which are not integral with respect to some α_i, as also follows from Theorem 0.1. However, the Everest–Ward theorem does not yield Theorem 0.1.

Strong equidistribution for roots of unity. We will now prove Proposition 1.3, the strong equidistribution theorem for roots of unity. At least when $k = \mathbb{Q}$, the result is well known to analytic number theorists, but we are not aware of a reference in the literature.

The proof rests on the following lemma, for which we thank Carl Pomerance. Let $\varphi(N)$ denote Euler's function and let $d(N) = \sum_{m\,|\,N, m\geq 1} 1$ be the divisor function. We write $\lambda(m)$ for the number of distinct primes dividing m, and use $\theta(x)$ to denote a quantity satisfying $-|x| \leq \theta(x) \leq |x|$.

Lemma 1.4 (Pomerance). *Fix an integer $Q > 1$ and an integer b coprime to Q. Then for each integer $N \geq 1$ divisible by Q and each interval $(C, D] \subset \mathbb{R}$,*

$$\#\big\{a \in (C, D] \cap \mathbb{Z} : (a, N) = 1,\ a \equiv b \pmod{Q}\big\} = \frac{\varphi(N)}{N\varphi(Q)}(D - C) + \theta(d(N)).$$

In particular, the error depends only on N, and not on Q or $(C, D]$.

Proof. Let $p_1, \ldots, p_r$ be the distinct primes dividing N but not Q. (If there are no such primes, take $p_1 \cdots p_r = 1$ below.) Take $b_0 \in \mathbb{Z}$ with $b_0 \equiv b \pmod{Q}$, $b_0 \equiv 0 \pmod{p_1 \ldots p_r}$. Then

$$\big\{a \in (C, D] \cap \mathbb{Z} : a \equiv b \pmod{Q}, (a, N) = 1\big\}$$
$$= \big\{a \in (C, D] \cap \mathbb{Z} :\ Q \,|\, a - b_0,\ \ p_1,\ \ldots,\ \ p_r \nmid a - b_0\big\}.$$

If m is a positive integer dividing $p_1 \cdots p_r$, put

$$r_{m,b,Q}(C, D) = \#\{a \in (C, D] \cap \mathbb{Z} :\ Qm \,|\, a - b_0\}.$$

Then

$$r_{m,b,Q}(C, D) = \left\lfloor \frac{d - b_0}{Qm} \right\rfloor - \left\lfloor \frac{c - b_0}{Qm} \right\rfloor = \frac{1}{Qm}(D - C) + \theta(1).$$

Carrying out inclusion/exclusion relative to the primes $p_1, \ldots, p_r$, we have

$$\#\big\{a \in (C, D] \cap \mathbb{Z} : a \equiv b \pmod{Q},\ (a, N) = 1\big\}$$
$$= \sum_{m\,|\,p_1 \cdots p_r} (-1)^{\lambda(m)} r_{m,b,Q}(C, D) = \frac{1}{Q}\prod_{i=1}^{r}\Big(1 - \frac{1}{p_i}\Big)(D - C) + \theta(d(p_1 \cdots p_r))$$
$$= \frac{\varphi(N)}{N\varphi(Q)}(D - C) + \theta(d(N)). \qquad \square$$

Proof of Proposition 1.3. Let ζ_N denote a primitive N-th root of unity. There are only finitely many subfields of k, so there are only finitely subfields of the form $k_N = k \cap \mathbb{Q}(\zeta_N)$ for some N. For each N there is a minimal Q for which $k_N = k_Q$,

and then $\mathbb{Q}(\zeta_Q) \subset \mathbb{Q}(\zeta_N)$ so $Q \mid N$. We will call $Q = Q_N$ the cyclotomic conductor of ζ_N relative to k, and write $T_N = [\mathbb{Q}(\zeta_{Q_N}) : k_N]$.

As $\mathbb{Q}(\zeta_N)$ is galois over $\mathbb{Q}$, it is linearly disjoint from k over k_N, and

$$\mathrm{Gal}(k(\zeta_N)/k) \cong \mathrm{Gal}(\mathbb{Q}(\zeta_N)/k_N).$$

Since $k_N \subset \mathbb{Q}(\zeta_{Q_N}) \subset \mathbb{Q}(\zeta_N)$, the conjugates of ζ_N over k are a union of T_N sets of the form

$$\{e^{2\pi i a/N} : a \equiv b_i (\mathrm{mod}\ Q_N),\ (a, N) = 1\},$$

for certain numbers b_i coprime to Q_N.

Let I be an arc of the unit circle corresponding to an angular interval $(\theta_1, \theta_2]$. Put $(C, D] = (N/2\pi)(\theta_1, \theta_2]$. Then $e^{2\pi i a/N} \in I$ if and only if $a \in (C, D]$. By Lemma 1.4,

$$N(\zeta_N, I) = T_N \cdot \frac{\varphi(N)}{N\varphi(Q_N)} \cdot \frac{N}{2\pi}(\theta_2 - \theta_1) + \theta(T_N \cdot d(N)). \tag{14}$$

Recall that for any $\delta > 0$, if N is sufficiently large then $d(N) \le N^\delta$ and $\varphi(N) \ge N^{1-\delta}$ [Hardy and Wright 1954, Theorem 315, p. 260, and Theorem 327, p. 267]. Take δ such that $0 < 2\delta < 1 - \gamma$. Noting that $[k(\zeta_N) : k] = T_N\, \varphi(N)/\varphi(Q_N)$, and that $\varphi(Q_N)$ is bounded independent of N, (14) gives

$$\frac{N(\zeta_N, I)}{[k(\zeta_N) : k]} = \mu(I) + O_\gamma(N^{-\gamma}). \tag{15}$$

Since $[k(\zeta_N) : k] \le N$, the error bound in (15) holds with N replaced by $[k(\zeta_N) : k]$. Since $[k(\zeta_N) : k]/N^\gamma \to \infty$ as $N \to \infty$, adjoining or removing endpoints of I will not affect the form of the estimate, so (8) applies to all intervals. □

2. The finiteness theorem for elliptic curves

Preliminaries. Let k be a number field, and let E/k be an elliptic curve. We can assume E is defined by a Weierstrass equation

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6 \tag{16}$$

with coefficients in $\mathbb{O}_k$. More precisely, E is the hypersurface in $\mathbb{P}^2/\mathrm{Spec}(k)$ defined by the homogenization of (16). Let Δ be its discriminant.

Given a nonarchimedean place v of k and points $\alpha, \beta \in E(\bar{k})$, we will say that β is integral with respect to α at v if the Zariski closures $\mathrm{cl}(\beta)$ and $\mathrm{cl}(\alpha)$ do not meet in the model $\mathscr{E}_v/\mathrm{Spec}(\mathbb{O}_v)$ defined by the homogenization of (16). Equivalently, if $\|z, w\|_v$ is the restriction of the spherical metric on $\mathbb{P}^2(\bar{k}_v)$ to $E(\bar{k}_v)$ [Rumely 1989, § 1.1], then for each pair of embeddings $\sigma, \tau : \bar{k}/k \hookrightarrow \bar{k}_v$,

$$\|\sigma(\beta), \tau(\alpha)\|_v = 1.$$

If S is a set of places of k containing all the archimedean places, we say β is S-integral with respect to α if β is integral with respect to α at each $v \notin S$.

Write $\hat{h}(\alpha)$ for the canonical height on $E(\bar{k})$, defined by

$$\hat{h}(\alpha) = \frac{1}{2} \lim_{n\to\infty} \frac{1}{4^n} h_{\mathbb{P}^1}(x([2^n]\alpha)) = \frac{1}{3} \lim_{n\to\infty} \frac{1}{4^n} h_{\mathbb{P}^2}([2^n]\alpha),$$

where $h_{\mathbb{P}^1}$ (respectively, $h_{\mathbb{P}^2}$) is the naive height on $\mathbb{P}^1(\bar{k})$ (respectively, $\mathbb{P}^2(\bar{k})$), x is the coordinate function on the Weierstrass model (16), and $[m]$ is multiplication by m on $E(\bar{k})$. (For a discussion of $\hat{h}(\alpha)$ and its properties, see [Silverman 1986, pp. 227–231 and 365–366; or 1994, § VI].) Recall that $\hat{h}(\alpha) \geq 0$, that $\hat{h}([m]\alpha) = m^2 \hat{h}(\alpha)$ for all m, and that $\hat{h}(\alpha) = 0$ if and only if $\alpha \in E(\bar{k})_{\mathrm{tors}}$. From these facts it follows (as is well known) that if $\xi \in E(\bar{k})_{\mathrm{tors}}$, then

$$\hat{h}(\alpha) = \hat{h}(\alpha - \xi). \tag{17}$$

There is also a decomposition of $\hat{h}(\alpha)$ as a sum of local terms. For each place v of k, let $\lambda_v(P)$ be the local Néron–Tate height function on $E(\bar{k}_v)$. For compatibility with our absolute values we normalize $\lambda_v(P)$ so that $\lambda_v(P) = [k_v : \mathbb{Q}_p] \cdot \lambda_{v,\mathrm{Sil}}(P)$, where $\lambda_{v,\mathrm{Sil}}(P)$ is the local Néron–Tate height defined in Silverman [1986, p. 365]. For each $0 \neq \alpha \in E(k)$ we have

$$\hat{h}(\alpha) = \frac{1}{[k:\mathbb{Q}]} \sum_{v \text{ of } k} \lambda_v(\alpha); \tag{18}$$

see [Silverman 1986, Theorem 18.2, p. 365]. Only finitely many terms in the sum are nonzero.

If L/k is a finite extension, for each place w of L there is a normalized local Néron–Tate height $\lambda_w(P)$ on $E(\bar{L}_w)$. If we fix a k_v-isomorphism $\bar{L}_w \cong \bar{k}_v$, then for all $P \in E(\bar{k}_v)$,

$$\lambda_w(P) = [L_w : k_v]\, \lambda_v(P). \tag{19}$$

It follows that if $\beta \in E(L)$, then for each place v of k, as σ runs over all embeddings of L into $\bar{k}_v$ fixing k,

$$\sum_{\sigma : L/k \hookrightarrow \bar{k}_v} \lambda_v(\sigma(\beta)) = \sum_{w \mid v} \lambda_w(\beta). \tag{20}$$

We will use the following explicit formulas.

Proposition 2.1. *Let k be a number field, and let E/k be an elliptic curve. Let v be a place of k.*

(i) *If v is archimedean, fix an isomorphism $E(\bar{k}_v) \cong \mathbb{C}/\Lambda$ for an appropriate lattice $\Lambda \subset \mathbb{C}$. Let $\sigma(z, \Lambda)$ be the Weierstrass σ-function, let $\Delta(\Lambda) = g_2(\Lambda)^3 - 27 g_3(\Lambda)^2$ be the discriminant of Λ, and let $\eta : \mathbb{C} \to \mathbb{R}$ be the $\mathbb{R}$-linearized*

period map associated to the Weierstrass ζ-function $\zeta(z,\Lambda)$. If $P \in E(\bar{k}_v)$ corresponds to $z \in \mathbb{C}$, then

$$\lambda_v(P) \;=\; -\log\bigl(|\Delta(\Lambda)^{1/12} e^{-z\eta(z)/2}\sigma(z,\Lambda)|_v\bigr).$$

If $\mu_v(z)$ is the additive Haar measure on $E(\bar{k}_v)$ that gives $E(\bar{k}_v) \cong \mathbb{C}/\Lambda$ total mass 1, then

$$\int_{E(\bar{k}_v)} \lambda_v(z)\, d\mu_v(z) \;=\; 0.$$

(ii) *If v is nonarchimedean and E has split multiplicative reduction at v (so E is k_v-isomorphic to a Tate curve), fix a Tate isomorphism $E(\bar{k}_v) \cong \bar{k}_v^\times/q^{\mathbb{Z}}$ where $q \in \bar{k}_v^\times$ satisfies $|q|_v = |1/j(E)|_v < 1$. Let $B_2(x) = x^2 - x + \frac{1}{6}$ be the second Bernoulli polynomial, and put*

$$\tilde{\lambda}_v(x) = \frac{1}{2} B_2\Bigl(\frac{x}{\mathrm{ord}_v(q)}\Bigr)(-\log|q|_v).$$

If $P \in E(\bar{k}_v)$ corresponds to $z \in \bar{k}_v^\times$, with z chosen so that $|q|_v < |z|_v \le 1$, then

$$\lambda_v(P) \;=\; -\log|1-z|_v + \tilde{\lambda}_v(\mathrm{ord}_v(z)).$$

If μ_v is the Haar measure $dx/\mathrm{ord}_v(q)$, which gives the loop $\mathbb{R}/(\mathbb{Z}\cdot\mathrm{ord}_v(q))$ total mass 1, then

$$\int_0^{\mathrm{ord}_v(q)} \tilde{\lambda}_v(x)\, d\mu_v(x) \;=\; 0.$$

(iii) *If v is nonarchimedean and E has good reduction at v, let $\|z,w\|_v$ be the spherical metric on $E(\bar{k}_v)$ induced by a projective embedding $E \hookrightarrow \mathbb{P}^2$ corresponding to a minimal Weierstrass model for E at v. Then for each $P \in E_v(\bar{k}_v)$*

$$\lambda_v(P) \;=\; -\log\|P, O\|_v.$$

Proof. This is a summary of results in [Silverman 1994, § VI]; see in particular Theorems 1.1 (p. 455), 3.2 (p. 466), 3.3 (p. 468) and 4.1 (p. 470). □

The finiteness theorem. For the convenience of the reader, we recall Theorem 0.2 from the Introduction:

Theorem 0.2. *Let k be a number field, and let S be a finite set of places of k containing all the archimedean places. If $\alpha \in E(\bar{k})$ is nontorsion (has canonical height $\hat{h}(\alpha) > 0$), there are only finitely many torsion points $\xi \in E(\bar{k})_{\mathrm{tors}}$ which are S-integral with respect to α.*

Again there are limitations to possible strengthenings of the theorem:

(A) As noted by Silverman, it is necessary that α be nontorsion. If $\alpha = O$ and S is the set of archimedean places, then by Cassels' generalization of the Lutz–Nagell theorem (Proposition 2.4 below), each torsion point whose order is divisible by at least two distinct primes is S-integral with respect to α.

Similarly, if α is a torsion point of order $N > 1$, let S contain all places of bad reduction for E. Then for each q coprime to N, all q-torsion points are S-integral with respect to α.

(B) When $\hat{h}(\alpha) > 0$, Zhang has pointed out that Theorem 0.2 cannot in general be strengthened to a result of Bogomolov type. A result of E. Ullmo [1995, Theorem 2.4] shows that if E has good reduction at all finite places, then for each $\varepsilon > 0$, there are infinitely many distinct points $\beta \in E(\bar{k})$ with $\hat{h}(\beta) < \varepsilon$ which are S_∞-integral with respect to α, where S_∞ is the set of archimedean places of k.

Proof of Theorem 0.2. The argument is similar to the proof of Theorem 0.1, but requires more machinery. It should be possible to axiomatize some of the arguments and combine both proofs, but for overall clarity of exposition we have chosen not to.

We begin with some reductions.

First, after replacing k by $k(\alpha)$, and S by the set $S_{k(\alpha)}$ of places lying over S, we can assume that $\alpha \in k$.

Second, after replacing k by a finite extension K/k, and replacing S with the set S_K of places of K lying above places in S, we can assume that E has semistable reduction. Thus we can assume without loss of generality that for nonarchimedean v, either E has good reduction, or E is k_v-isomorphic to a Tate curve.

Third, after enlarging S if necessary, we can assume that S contains all v for which $|\Delta|_v \neq 1$. In particular, we can assume that the model of E defined by (16) has good reduction for all $v \notin S$.

We claim that if $\xi_n \in E(\bar{k})_{\mathrm{tors}}$ is any torsion point, then

$$\hat{h}(\alpha) = \frac{1}{[k(\xi_n):\mathbb{Q}]} \sum_v \sum_{\sigma : k(\xi_n)/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)). \tag{21}$$

To see this, let L be the galois closure of $k(\xi_n)$ in $\bar{k}$ over k. By (17) and (18), for each conjugate $\sigma(\xi_n)$,

$$\hat{h}(\alpha) = \hat{h}(\alpha - \sigma(\xi_n)) = \frac{1}{[L:\mathbb{Q}]} \sum_{w \text{ of } L} \lambda_w(\alpha - \sigma(\xi_n)).$$

Averaging over all k-embeddings $\sigma : L \hookrightarrow \bar{k}$, fixing a k-embedding $\bar{k} \hookrightarrow \bar{k}_v$ for each place v of K, using (19), and noting that there are only finitely many nonzero

terms in each sum, we have

$$\begin{aligned}\hat{h}(\alpha) &= \frac{1}{[L:k]} \sum_{\sigma: L/k \hookrightarrow \bar{k}} \frac{1}{[L:\mathbb{Q}]} \sum_{w \text{ of } L} \lambda_w(\alpha - \sigma(\xi_n)) \\ &= \frac{1}{[L:\mathbb{Q}]} \sum_{v \text{ of } k} \sum_{\sigma: L/k \hookrightarrow \bar{k}_v} \frac{1}{[L:k]} \sum_{w \mid v} [L_w : k_v] \cdot \lambda_v(\alpha - \sigma(\xi_n)) \\ &= \frac{1}{[L:\mathbb{Q}]} \sum_{v \text{ of } k} \sum_{\sigma: L/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)).\end{aligned}$$

Since each conjugate $\sigma(\xi_n)$ occurs $[L : k(\xi_n)]$ times in the final inner sum, this is equivalent to (21).

Suppose there were an infinite sequence of distinct torsion points $\{\xi_n\}$ which were S-integral with respect to α.

If $v \notin S$, our initial reductions assure that E has good reduction at v. By Proposition 2.1(iii) and the integrality hypothesis, $\lambda_v(\alpha - \sigma(\xi_n)) = 0$ for each n and σ. It follows that

$$\hat{h}(\alpha) = \sum_{v \in S} \frac{1}{[k(\xi_n) : k]} \sum_{\sigma: k(\xi_n)/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)). \tag{22}$$

From now through page 237, we will show in a series of cases that for each $v \in S$,

$$\lim_{n \to \infty} \left(\frac{1}{[k(\xi_n) : \mathbb{Q}]} \sum_{\sigma: k(\xi_n)/k \hookrightarrow \bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)) \right) = 0. \tag{23}$$

This will complete the proof of Theorem 0.2, for then, combining (22) and (23) and letting $n \to \infty$ in (22), we would have $\hat{h}(\alpha) = 0$, contradicting the assumption that α is nontorsion.

The archimedean case. Let v be an archimedean place of k. To simplify notation we view k as embedded in $\mathbb{C}$ and fix an isomorphism of $\bar{k}_v$ with $\mathbb{C}$. Thus, the way k is embedded depends on the choice of v.

To prove (23) we will need a theorem of David and Hirata-Kohno on linear forms in elliptic logarithms and a strong form of equidistribution for torsion points.

Proposition 2.2 (a special case of [David and Hirata-Kohno 2002, Theorem 1]). *Let E/k be an elliptic curve defined over a number field $k \subset \mathbb{C}$. Fix an isomorphism $\theta : \mathbb{C}/\Lambda \cong E(\mathbb{C})$ for an appropriate lattice $\Lambda \subset \mathbb{C}$. Let ω_1, ω_2 be generators for Λ. Fix a nontorsion point $\alpha \in E(k)$ and let $a \in \mathbb{C}$ be such that $\theta(a \bmod \Lambda) = \alpha$. There is a constant $C = C(E, \alpha) > 0$ such that for all rational numbers ℓ_1/N, ℓ_2/N with $\ell_1, \ell_2, N \in \mathbb{Z}$,*

$$\left| a - \left(\frac{\ell_1}{N} \omega_1 + \frac{\ell_2}{N} \omega_2 \right) \right| \geq e^{-C \max(1, \log N)}.$$

By the Szpiro–Ullmo–Zhang theorem [1997], the galois conjugates of the ξ_n are equidistributed in $E(\mathbb{C})$. As we will see, they are in fact strongly equidistributed, in a sense analogous to that in Proposition 1.3.

If $\xi \in E(\bar{k})_{\mathrm{tors}}$, write $\mathrm{Gal}(\bar{k}/k)\cdot\xi$ for the orbit $\{\sigma(\xi) : \sigma \in \mathrm{Gal}(\bar{k}/k)\}$. For each set $U \subset E(\mathbb{C})$, write

$$N(\xi, U) = \#\big((\mathrm{Gal}(\bar{k}/k)\cdot\xi) \cap U\big).$$

Let $\mathcal{S} \subset \mathbb{C}$ be a bounded, convex, centrally symmetric set with 0 in its interior. For each $a \in \mathbb{C}$ and $0 \le r \in \mathbb{R}$, write $\mathcal{S}(a, r) = \{a + rz : z \in \mathcal{S}\}$. For example, if $\mathcal{S} = B(0, 1)$ then $\mathcal{S}(a, r) = B(a, r)$.

Let $\Lambda \subset \mathbb{C}$ be a lattice such that $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$. Let $r_0 = r_0(\mathcal{S}, \Lambda) > 0$ be the largest number such that $\mathcal{S}(a, r)$ injects into $\mathbb{C}/\Lambda \cong E(\mathbb{C})$ under the natural projection for all $a \in \mathbb{C}$ and all $0 \le r < r_0$. Write $\mathcal{S}_E(a, r)$ for the image of $\mathcal{S}(a, r)$ in $E(\mathbb{C})$.

Proposition 2.3 (Strong equidistribution). *Let $k \subset \mathbb{C}$ be a number field, and let E/k be an elliptic curve. Then the $\mathrm{Gal}(\bar{k}/k)$-conjugates of the torsion points in $E(\bar{k})$ are strongly equidistributed in $E(\mathbb{C})$ in the following sense*:

Let μ be the additive Haar measure on $E(\mathbb{C})$ with total mass 1. Fix γ with $0 < \gamma < 1/2$, and fix a bounded, convex, centrally symmetric set $\mathcal{S}$ with 0 in its interior. Then for each r such that $\mathcal{S}(a, r)$ injects into $E(\mathbb{C})$, and for all $\xi \in E(\bar{k})_{\mathrm{tors}}$,

$$\frac{N\big(\xi, \mathcal{S}_E(a, r)\big)}{[k(\xi) : k]} = \mu\big(\mathcal{S}_E(a, r)\big) + O([k(\xi) : k]^{-\gamma})$$

where the implied constant depends only on k, $\mathcal{S}$, E, and γ.

The proof will be given starting on page 237.

We can now complete the proof of (23) in the archimedean case. The argument is similar to the one in the proof of Theorem 0.1. By the Szpiro–Ullmo–Zhang theorem [1997], or by Proposition 2.3 when $\mathcal{S}$ has the shape of a period parallelogram (so E can be tiled with sets $\mathcal{S}_E(a, r)$), one knows that as $n \to \infty$ the discrete measures

$$\mu_n = \frac{1}{[k(\xi_n) : k]} \sum_{\sigma : k(\xi_n)/k \hookrightarrow \mathbb{C}} \delta_{\sigma(\xi_n)}(x)$$

converge weakly to the Haar measure μ on $E(\mathbb{C})$ having total mass 1. Proving (23) is equivalent to showing that

$$\lim_{n\to\infty} \int_{E(\mathbb{C})} \lambda_v(\alpha - z)\, d\mu_n(z) = 0.$$

Choose a lattice $\Lambda \subset \mathbb{C}$ such that $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$, and let F be the area of a fundamental domain for Λ. After scaling Λ, if necessary, we can assume that

$F = 1$. After this normalization, μ coincides with Lebesgue measure. Let $\theta : \mathbb{C}/\Lambda \cong E(\mathbb{C})$ be an isomorphism as in the David/Hirata-Kohno theorem, and let $a \in \mathbb{C}$ be a point with $\theta(a \bmod \Lambda) = \alpha$.

Fix $\varepsilon > 0$ small enough that $B(a, \varepsilon)$ injects into $\mathbb{C}/\Lambda$, and identify $B(a, \varepsilon)$ with its image $B_E(a, \varepsilon) = \theta(B(a, \varepsilon)) \subset E(\mathbb{C})$. (In particular, identify a with α). Without loss, we can assume that $\varepsilon < 1/\pi$, so $\pi\varepsilon^2 < \varepsilon$. We will show that for all large n,

$$\left| \int_{E(\mathbb{C})} \lambda_v(\alpha - z)\, d\mu_n(z) \right| < 6\varepsilon. \tag{24}$$

Put

$$\operatorname{labs}_{\alpha,\varepsilon}(z) = \begin{cases} \infty & \text{if } z = a, \\ -[k_v : \mathbb{R}] \log \dfrac{|z-a|}{\varepsilon} & \text{if } z \in B(a, r)\backslash\{a\}, \\ 0 & \text{if } z \in E(\mathbb{C})\backslash B(a, r), \end{cases}$$

and note that

$$\begin{aligned} 0 < \int_{E(\mathbb{C})} \operatorname{labs}_{\alpha,\varepsilon}(z)\, d\mu(z) &= \int_{B(a,\varepsilon)} -[k_v : \mathbb{R}] \log(|z-a|/\varepsilon)\, d\mu(z) \\ &= [k_v : \mathbb{R}] \int_0^{\varepsilon} -2\pi t \log \frac{t}{\varepsilon}\, dt = [k_v : \mathbb{R}] \frac{\pi\varepsilon^2}{2} < \varepsilon. \end{aligned}$$

By Proposition 2.1(i) there is a continuous function $g_{\alpha,\varepsilon}(z)$ on $E(\mathbb{C})$ such that

$$\lambda_v(\alpha - z) = \operatorname{labs}_{\alpha,\varepsilon}(z) + g_{\alpha,\varepsilon}(z).$$

Since $\int_{E(\mathbb{C})} \lambda_v(\alpha - z)\, d\mu(z) = 0$ (also by Proposition 2.1(i)), we get

$$\left| \int_{E(\mathbb{C})} g_{\alpha,\varepsilon}(z)\, d\mu(z) \right| = \left| \int_{B(a,\varepsilon)} -\operatorname{labs}_{\alpha,\varepsilon}(z)\, d\mu(z) \right| < \varepsilon.$$

By weak convergence, it follows that for all sufficiently large n,

$$\left| \int_{E(\mathbb{C})} g_{\alpha,\varepsilon}(z)\, d\mu_n(z) \right| < 2\varepsilon. \tag{25}$$

To complete the proof of (24), it suffices to show that for all sufficiently large n,

$$\left| \int_{B(a,r)} \log(|z-a|/\varepsilon)\, d\mu_n(z) \right| < 2\varepsilon. \tag{26}$$

For this, put $D = D_n = \lceil [k(\xi_n) : k]^{1/8} \rceil$, and subdivide $B(a, \varepsilon)$ into a disc $A_0(n) = B(a, \varepsilon/D)$ and annuli $A_\ell(n) = B(a, (\ell+1)\varepsilon/D)\backslash B(a, \ell\varepsilon/D)$ for $\ell = 1, \ldots, D-1$.

For the central disc, we have $\mu(A_0(n)) = \pi\varepsilon^2/D^2 \le \pi\varepsilon^2/[k(\xi_n) : k]^{1/4}$. Applying Proposition 2.3 when $\mathcal{S}$ is a disc, taking $\gamma = 3/8$, gives

$$N(\xi_n, A_0(n))/[k(\xi_n) : k] \le 2\mu(A_0(n))$$

for all sufficiently large n. If ξ_n has order N_n, the David/Hirata-Kohno theorem tells us that for each conjugate $\sigma(\xi_n) \in A_0(n)$ (where as before we are identifying $B(a,\varepsilon)$ with its image $\theta(B(a,\varepsilon)) \subset E(\mathbb{C})$)

$$\big|\log|\sigma(\xi_n) - a|\big| \,\le\, C \log N_n\,.$$

Using (41) and (42) below, one sees that $[k(\xi_n):k] \ge N_n^{1/2}$ for all sufficiently large n. Thus $0 \le \big|\log|\sigma(\xi_n)-\alpha|\big| \,\le\, 2C\log[k(\xi_n):k]$ and

$$0 \,\le\, \Big|\int_{A_0(n)} \log|z-\alpha|\,d\mu_n(z)\Big| \,\le\, 4\pi\varepsilon^2 C\,\frac{\log[k(\xi_n):k]}{[k(\xi_n):k]^{1/4}} \,<\, \varepsilon \tag{27}$$

for all sufficiently large n.

For each annulus $A_\ell(n)$, $\ell = 1,\dots,D-1$, one has

$$\mu(A_\ell(n)) \,=\, \pi(2\ell+1)\,\varepsilon^2/D^2 \,\cong\, \pi(2\ell+1)\,\varepsilon^2/[k(\xi):k]^{1/4}.$$

Since $A_\ell(n)$ is the difference of two sets to which Proposition 2.3 applies, we find as above that for sufficiently large n,

$$N(\xi_n, A_\ell(n))/[k(\xi_n):k] \,\le\, 2\mu(A_\ell(n))\,.$$

Note that on $A_\ell(n)$, $\big|\log(|z-\alpha|/\varepsilon)\big| \,\le\, -\log(\ell/D)$. Summing over these annuli, and bounding the resulting Riemann sum by an integral, we find that

$$\begin{aligned}\Big|\int_{B(a,\varepsilon)\setminus A_0(n)} \log\frac{|z-a|}{\varepsilon}\,d\mu_n(z)\Big| \,&\le\, \sum_{\ell=1}^{D-1} -\log\Big(\frac{\ell\varepsilon/D}{\varepsilon}\Big)\cdot 2\mu(A_\ell(n)) \\ &<\, 2\int_{B(a,\varepsilon)} -2\pi t\log(t/\varepsilon)\,dt \,=\, \pi\varepsilon^2 \,<\, \varepsilon.\end{aligned}$$

Combining this with (27) gives (26), which completes the proof of (23) in the archimedean case (assuming Proposition 2.3).

The nonarchimedean case. In the nonarchimedean case, the proof of (23) depends on a well known result of Cassels on the denominators of torsion points [Silverman 1986, Theorem 3.4, p. 177]. Write $\overline{\mathbb{O}}_v$ for the ring of integers of $\bar{k}_v$.

Proposition 2.4 (Cassels). *Let k_v be a local field of characteristic 0 and residue characteristic $p > 0$, and let E/k_v be an elliptic curve defined by a Weierstrass equation*

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

whose coefficients belong to $\mathbb{O}_v$ (note that the Weierstrass equation need not be minimal). Let $P \in E(\bar{k}_v)_{\mathrm{tors}}$ be a point of exact order $m \ge 2$.

(i) *If m is not a power of p, then $x(P), y(P) \in \overline{\mathbb{O}}_v$.*

(ii) *If $m = p^n$, then $x(P) = a/D^2$, $y(P) = b/D^3$ where $a, b, D \in \overline{\mathbb{O}}_v$ and*

$$\operatorname{ord}_v(D) \;\le\; \frac{\operatorname{ord}_v(p)}{p^n - p^{n-1}}.$$

Proof. Silverman [1986, Theorem 3.4] states the theorem for torsion points belonging to $E(k_v)$, with $a, b, D \in k_v$ in part (ii) and D satisfying

$$\operatorname{ord}_v(D) \;=\; \left\lfloor \frac{\operatorname{ord}_v(p)}{p^n - p^{n-1}} \right\rfloor, \tag{28}$$

where $\lfloor x \rfloor$ denotes the floor of x. Since the Weierstrass equation for E need not be minimal, we can replace k_v by an arbitrary finite extension L_w/k_v, and if $e_{w/v}$ is the ramification index of L_w/k_v, then for $P \in E(L_w)_{\text{tors}}$ and $a, b, D \in L_w$, (28) becomes

$$\operatorname{ord}_v(D) \;=\; \frac{1}{e_{w/v}} \cdot \left\lfloor \frac{e_{w/v} \operatorname{ord}_v(p)}{p^n - p^{n-1}} \right\rfloor. \tag{29}$$

This yields the result for all $P \in E(\bar{k}_v)_{\text{tors}}$. □

As a consequence, we obtain the following result, part (i) of which is a special case of the Tate–Voloch conjecture proved in [Scanlon 1999].

Corollary 2.5. *Let E/k_v be an elliptic curve defined over a nonarchimedean local field. Then for each nontorsion point $\alpha \in E(\bar{k}_v)$:*

(i) *There is a number $M = M(\alpha)$ such that for all $\xi \in E(\bar{k}_v)_{\text{tors}}$,*

$$\lambda_v(\alpha - \xi) \;\le\; M.$$

(ii) *If E has good reduction, then for each $\varepsilon > 0$, there are only finitely many $\xi \in E(\bar{k}_v)_{\text{tors}}$ with $\lambda_v(\alpha - \xi) > \varepsilon$. If E is a Tate curve, then for each $\varepsilon > 0$, there are only finitely many $\xi \in E(\bar{k}_v)_{\text{tors}}$ with $\lambda_v(\alpha - \xi) > \varepsilon + \frac{1}{12}(-\log|\Delta(E)|_v)$.*

Proof. After a finite base extension, we can assume that E either has good reduction or is a Tate curve. Since (ii) implies (i), it suffices to prove (ii). Fix $\varepsilon > 0$.

First suppose E has good reduction. Then $\lambda_v(x - y) = -\log \|x, y\|_v$, where $\|x, y\|_v$ is the spherical distance on the minimal Weierstrass model for E/k_v. If $\xi_1, \xi_2 \in E(\bar{k}_v)_{\text{tors}}$ satisfy $\lambda_v(\alpha - \xi_i) > \varepsilon$, then $\|\xi_1, \alpha\|_v$, $\|\xi_2, \alpha\|_v < (Nv)^{-\varepsilon}$, where Nv is the order of the residue field of $\mathbb{O}_v$. By the ultrametric inequality for the spherical distance [Rumely 1989, § 1.1], $\|\xi_1, \xi_2\|_v < (Nv)^{-\varepsilon}$. By translation invariance, $\|\xi_1 - \xi_2, 0\|_v < (Nv)^{-\varepsilon}$. Put $\xi := \xi_1 - \xi_2$. By the definition of the spherical distance, if x, y are the coordinate functions in the minimal Weierstrass model,

$$-\log \|\xi, 0\|_v \;=\; \min\big(\operatorname{ord}_v(x(\xi)), \operatorname{ord}_v(y(\xi))\big) \cdot \log(Nv).$$

By Cassels' theorem, there are only finitely many torsion points for which

$$\min\big(\operatorname{ord}_v(x(\xi)), \operatorname{ord}_v(y(\xi))\big) \;>\; \varepsilon/\log(Nv).$$

Next suppose E is a Tate curve. Fix a Tate isomorphism $E(\bar{k}_v) \cong \bar{k}_v^\times / q^{\mathbb{Z}}$ where $|q|_v = |\Delta(E)|_v < 1$, and let $y^2 + xy = x^3 + a_4(q)x + a_6(q)$ be the corresponding Weierstrass equation. Let $a, u_1, u_2 \in \bar{k}_v^\times$ correspond to α, ξ_1, ξ_2 respectively; we can assume that $|q|_v < |a|_v, |u_1|_v, |u_2|_v \le 1$. By the formula for $\lambda_v(x-y)$ in Proposition 2.1(ii), if $\lambda_v(\alpha - \xi_i) > \varepsilon + \frac{1}{12}(-\log|\Delta(E)|_v)$, then $|a|_v = |u_1|_v = |u_2|_v$ and

$$-\log|1 - a^{-1}u_i|_v \;=\; \mathrm{ord}_v(1 - a^{-1}u_i)\cdot \log(Nv) \;>\; \varepsilon\,.$$

Put $\xi = \xi_1 - \xi_2$ and $u = u_2^{-1}u_1$. Then ξ corresponds to u under the Tate isomorphism, and $\mathrm{ord}_v(1-u) > \varepsilon/\log(Nv)$. By the formulas for $x(\xi)$, $y(\xi)$ in [Silverman 1994, p. 425],

$$\mathrm{ord}_v(x(\xi)) = 2\,\mathrm{ord}_v(1-u) \quad \text{and} \quad \mathrm{ord}_v(y(\xi)) = 3\,\mathrm{ord}_v(1-u).$$

Again by Cassels' theorem, only finitely many torsion points ξ can satisfy

$$\min\big(\mathrm{ord}_v(x(\xi)), \mathrm{ord}_v(y(\xi))\big) > \varepsilon/\log(Nv). \qquad \square$$

We can now prove (23) when E has good reduction at v.

Fix $\varepsilon > 0$. Let M be the upper bound in Corollary 2.5(i), and let N be the number of points $\xi \in E(\bar{k}_v)_{\mathrm{tors}}$ with $\lambda_v(\alpha - \xi) > \varepsilon$ given by Corollary 2.5(ii). For all sufficiently large n, $MN/[k(\xi_n):k] < \varepsilon$, giving

$$0 \le \frac{1}{[k(\xi_n):k]} \sum_{\sigma:\bar{k}/k\hookrightarrow\bar{k}_v} \lambda_v(\alpha - \sigma(\xi_n)) \le \frac{([k(\xi_n):k]-N)}{[k(\xi_n):k]}\,\varepsilon + \frac{N}{[k(\xi_n):k]}\,M \;<\; 2\varepsilon\,.$$

Thus

$$\lim_{n\to\infty} \frac{1}{[k(\xi_n):k]} \sum_{\sigma:\bar{k}/k\hookrightarrow\bar{k}_v} \lambda_v(\sigma(\xi_n) - \alpha) \;=\; 0\,.$$

To prove (23) when E is a Tate curve at v, we will need the following equidistribution theorem of Chambert-Loir [2006, corollaire 5.5].

Fix a Tate isomorphism $E(\bar{k}_v) \cong \bar{k}_v/q^{\mathbb{Z}}$, put $L = \mathbb{Z}\cdot \mathrm{ord}_v(q) \subset \mathbb{R}$, and define a "reduction map" $r : E(\bar{k}) \to \mathbb{R}/L$ by setting $r(P) = \mathrm{ord}_v(a) \pmod L$ if $P \in E(\bar{k}_v)$ corresponds to $a \in \bar{k}_v^\times$.

For each global point $P \in E(\bar{k})$, define a measure $\mu_{P,v}$ on $\mathbb{R}/L$ by

$$\mu_{P,v}(z) \;=\; \frac{1}{[k(P):k]} \sum_{\sigma:\bar{k}/k\hookrightarrow\bar{k}_v} \delta_{r(\sigma(P))}(z)$$

and let μ_v be the Haar measure on $\mathbb{R}/L$ with total mass 1.

Proposition 2.6 (Chambert-Loir). *For each sequence of distinct points $\{P_n\}$ in $E(\bar{k})$ with $\hat{h}(P_n) \to 0$, the sequence of measures $\{\mu_{P_n,v}\}$ converges weakly to μ_v.*

We can now prove (23) when E is a Tate curve. Recall that $\{\xi_n\}$ is a sequence of distinct torsion points which are S-integral with respect to α.

Fix $\varepsilon > 0$. Let M be the upper bound in Corollary 2.5(i). Put $a = r(\alpha)$ and let $\delta > 0$ be such that $\mu((a-\delta, a+\delta)) < \varepsilon/M$, where by abuse of notation we identify a sufficiently short interval in $\mathbb{R}$ with its image in $\mathbb{R}/L$. By Chambert-Loir's theorem, $\mu_{\xi_n,v}((a-\delta, a+\delta)) < 2\varepsilon/M$ for all sufficiently large n.

By the formulas in Proposition 2.1(ii), $\int_{\mathbb{R}/L} \tilde{\lambda}_v(z)\, d\mu_v(z) = 0$ and

$$\Big|\frac{1}{[k(\xi_n):k]} \sum_{\sigma:\bar{k}/k\hookrightarrow\bar{k}_v} \lambda_v(\sigma(\xi_n)-\alpha)\Big| \le \left|\int_{\mathbb{R}/L} \tilde{\lambda}_v(z-a)\, d\mu_{\xi_n,v}(z)\right| + M\,\mu_{\xi_n,v}((a-\delta, a+\delta)).$$

For all sufficiently large n the right side is at most 3ε. Hence

$$\lim_{n\to\infty} \frac{1}{[k(\xi_n):k]} \sum_{\sigma:\bar{k}/k\hookrightarrow\bar{k}_v} \lambda_v(\sigma(\xi_n)-\alpha) = 0.$$

This completes the proof of Theorem 0.2. □

Several results in the literature use methods related to ours.

J. Cheon and S. Hahn [1999] proved an elliptic curve analogue of Schinzel's theorem [1974]. Likewise, Everest and B. Ní Flathúin [1996] evaluate "elliptic Mahler measures" in terms of limits involving division polynomials, obtaining results similar to (13). They use David/Hirata-Kohno's theorem on elliptic logarithms in place of Baker's theorem, much as we do.

More recently, L. Szpiro and T. Tucker [2005] proved that local canonical heights for a dynamical system can be evaluated by taking limits over "division polynomials" for the dynamical system. (These polynomials have periodic points as their roots.) Their work uses Roth's theorem rather than Baker's or David/Hirata-Kohno's theorem. It would be interesting to see if this could be brought to bear on Conjecture 3.1 below.

Strong equidistribution for torsion points on elliptic curves. We will now prove Proposition 2.3, the strong equidistribution theorem for galois orbits of torsion points on elliptic curves, which was used in the proof of Theorem 0.2.

Proof of Proposition 2.3. The proof breaks into two cases, depending on whether or not E has complex multiplication. Both cases are similar, and are modeled on Proposition 1.3. We find an extension field over which there is a two-dimensional geometric interpretation of the galois orbits, and by carrying out inclusion/exclusion, we are able to count the number of conjugates over that field lying in a convex,

centrally symmetric set, with a good error bound. The conjugates over the original field can then be counted by breaking into cosets.

Case 1. Suppose E does not have complex multiplication. The action of $\mathrm{Gal}(\bar{k}/k)$ on $E(\bar{k})_{\mathrm{tors}}$ induces an injective homomorphism

$$\eta : \mathrm{Gal}(\bar{k}/k) \to \varprojlim \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \cong \prod_p \mathrm{GL}_2(\mathbb{Z}_p).$$

By Serre's theorem [1972, théorème 3], the image of $\mathrm{Gal}(\bar{k}/k)$ in $\prod_p \mathrm{GL}_2(\mathbb{Z}_p)$ is open. Hence there is a number Q such that $\mathrm{Im}(\eta)$ contains the subgroup

$$\prod_{p \mid Q} (1 + QM_2(\mathbb{Z}_p)) \times \prod_{p \nmid Q} \mathrm{GL}_2(\mathbb{Z}_p).$$

Let $G_Q \subset \mathrm{Gal}(\bar{k}/k)$ be the preimage of this subgroup.

Step 1: Determining the size of a galois orbit under G_Q. Let $\xi \in E(\bar{k})_{\mathrm{tors}}$ have order N, and put $Q_N = \gcd(Q, N)$. For suitable right coset representatives $\sigma_1, \ldots, \sigma_T$ of G_Q in $\mathrm{Gal}(\bar{k}/k)$, the galois orbit $\mathrm{Gal}(\bar{k}/k) \cdot \xi$ decomposes as a disjoint union of G_Q-orbits:

$$\mathrm{Gal}(\bar{k}/k) \cdot \xi = \bigcup_{i=1}^{T} G_Q \cdot \sigma_i(\xi).$$

Since G_Q is normal in $\mathrm{Gal}(\bar{k}/k)$, the orbits $G_Q \cdot \sigma_i(\xi) = \sigma_i(G_Q \cdot \xi)$ all have the same size. Thus $[k(\xi) : k] = T \cdot \#(G_Q \cdot \xi)$. By considering the action of G_Q on the p-parts of ξ, one sees that

$$\#(G_Q \cdot \xi) = \prod_{p \mid Q_N} p^{2(\mathrm{ord}_p(N) - \mathrm{ord}_p(Q_N))} \prod_{\substack{p \mid N \\ p \nmid Q_N}} p^{2\,\mathrm{ord}_p(N)} \left(1 - \frac{1}{p^2}\right) = \frac{N^2}{Q_N^2} \cdot \prod_{\substack{p \mid N \\ p \nmid Q}} \left(1 - \frac{1}{p^2}\right). \tag{30}$$

Indeed, let ξ_p be the p-component of ξ in $E[N] \cong \prod_{p \mid N} (\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})^2$. Identify ξ_p with an element of $(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})^2$: then ξ_p generates that group. If p divides Q_N, the image of G_Q in $\mathrm{GL}_2(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})$ is $I + p^{\mathrm{ord}_p(Q_N)} M_2(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})$, and

$$G_Q \cdot \xi_p = \xi_p + p^{\mathrm{ord}_p(Q_N)} \cdot (\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})^2.$$

On the other hand, if $p \nmid Q_N$, the image of G_Q in $\mathrm{GL}_2(\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})$ is the full group, so

$$G_Q \cdot \xi_p = (\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})^2 \backslash p \cdot (\mathbb{Z}/p^{\mathrm{ord}_p(N)}\mathbb{Z})^2.$$

Step 2: Counting translated lattice points in convex domains. Let $\mathcal{F}$ be a fundamental domain for Λ; we can assume $\mathcal{F}$ is bounded and contains 0. Let C be such that $\mathcal{F} \subset \mathcal{S}(0, C)$. Note that since $\mathcal{S}$ is convex, if $z_1 \in \mathcal{S}(a_1, r_1)$ and $z_2 \in \mathcal{S}(a_2, r_2)$, then $z_1 + z_2 \in \mathcal{S}(a_1 + a_2, r_1 + r_2)$. Put $F = \mathrm{area}\ \mathcal{F}$ and $S = \mathrm{area}\ \mathcal{S}$.

For each $0 < t \in \mathbb{R}$, we have $\text{area}(t\mathscr{F}) = t^2 F$ and $\text{area}\,\mathscr{S}(a,r) = r^2 S$. Each lattice $t\Lambda_N$ is homothetic to Λ_N, and hence has fundamental domain $t\mathscr{F} \subset \mathscr{S}(0, tC)$. Fix $x_0 \in \mathbb{C}$. As y runs over $x_0 + t\Lambda$, the sets $y + t\mathscr{F}$ are pairwise disjoint and cover $\mathbb{C}$. If $y \in \mathscr{S}(a,r)$, then $y + t\mathscr{F} \subset \mathscr{S}(a, r + tC)$. Hence

$$\#\big((x_0 + t\Lambda) \cap \mathscr{S}(a,r)\big) \le \frac{\text{area}\big(\mathscr{S}(a, r+tC)\big)}{\text{area}\,(t\mathscr{F})} = \frac{r^2 S}{F}\cdot\frac{1}{t^2} + \frac{2CSr}{F}\cdot\frac{1}{t} + \frac{C^2 S}{F}. \quad (31)$$

Similarly, if $r > tC$, take $z \in \mathscr{S}(a, r - tC)$, and let $y \in x_0 + t\Lambda$ be such that $z \in y + t\mathscr{F}$. Then $z - y \in t\mathscr{F}$, so $z - y \in \mathscr{S}(0, tC)$, and since $\mathscr{S}$ is centrally symmetric $y - z \in \mathscr{S}(0, tC)$. Thus $y = z + (y - z) \in \mathscr{S}(a,r)$. It follows that $\mathscr{S}(a, r - tC) \subset \bigcup_{y \in (x_0 + t\Lambda) \cap \mathscr{S}(a,r)} (y + t\mathscr{F})$, so

$$\#\big((x_0 + t\Lambda) \cap \mathscr{S}(a,r)\big) \ge \frac{\text{area}\big(\mathscr{S}(a, r-tC)\big)}{\text{area}\,(t\mathscr{F})} > \frac{r^2 S}{F}\cdot\frac{1}{t^2} - \frac{2CSr}{F}\cdot\frac{1}{t} - \frac{C^2 S}{F}. \quad (32)$$

If $r \le tC$, the right side of (32) is negative, so the inequality between the first and last quantities holds trivially.

Now let D be a positive divisor of N/Q_N. Taking $t = Q_N D/N$, and combining (31), (32), we obtain

$$\left| \#\left(\left(x_0 + \frac{Q_N D}{N}\Lambda_N\right) \cap \mathscr{S}(a,r)\right) - \frac{\text{area}\big(\mathscr{S}(a,r)\big)}{\text{area}\,(\mathscr{F})} \cdot \frac{N^2}{Q_N^2 D^2} \right| \le \frac{2CSr}{F}\cdot\frac{N}{Q_N D} + \frac{C^2 S}{F}. \quad (33)$$

Step 3: Inclusion/exclusion. Write $\Lambda_N = \frac{1}{N}\Lambda$, fix σ_i, and let $x \in \Lambda_N$ correspond to $\sigma_i(\xi)$. Since $E[N] \cong \Lambda_N/\Lambda$, the considerations above show there is a one-to-one correspondence between elements of $G_Q \cdot \sigma_i(\xi)$, and cosets $y + \Lambda$ for $y \in \Lambda_N$ such that $y - x \in Q_N\Lambda_N$ and $y + \Lambda$ has exact order N in Λ_N/Λ. Equivalently, $y - x \in Q_N\Lambda_N$ and $y \notin p\Lambda_N$ for each prime p dividing N but not Q.

Let $p_1, \ldots, p_R$ be the distinct primes dividing N but not Q; if there are no such primes, take $p_1 \cdots p_R = 1$. Since Q_N and $p_1, \cdots, p_R$ are pairwise coprime, there is an $x_0 \in \Lambda_N$ such that $x_0 \equiv x \pmod{Q_N\Lambda_N}$ and $x_0 \equiv 0 \pmod{p_1 \cdots p_R \Lambda_N}$. Then $y - x_0 \in Q_N\Lambda_N$ if and only if $y \in x_0 + Q_N\Lambda_N$, and $y \in p_i\Lambda_N$ if and only if $y \in x_0 + p_i\Lambda_N$. Note that if $D | p_1 \cdots p_R$ then $Q_N\Lambda_N \cap D\Lambda_N = Q_N D\Lambda_N$. Recalling that r_0 is the supremum over positive numbers r for which $\mathscr{S}(a,r)$ injects into $\mathbb{C}/\Lambda$, take $a \in \mathbb{C}$ and take $0 < r \le r_0$. Applying inclusion/exclusion, we obtain

$$\#\big(G_Q \cdot \sigma_i(\xi) \cap \mathscr{S}_E(a,r)\big) = \sum_{D \,|\, p_1 \cdots p_R} (-1)^{\lambda(D)} \cdot \#\big((x_0 + Q_N D\Lambda_N) \cap \mathscr{S}(a,r)\big), \quad (34)$$

where $\lambda(D)$ is the number of distinct primes dividing D.

Inserting (33) in (34) and summing over all $\sigma_i(\xi)$, $i = 1, \dots, T$, we find

$$N\big(\xi, \mathcal{S}_E(a,r)\big) = \frac{\text{area } \mathcal{S}(a,r)}{\mathcal{F}} \cdot \frac{TN^2}{Q_N^2} \prod_{\substack{p \mid N \\ p \nmid Q}} \Big(1 - \frac{1}{p^2}\Big) + \theta\Big(\frac{2CSr}{F} \cdot \frac{TN}{Q_N} \prod_{\substack{p \mid N \\ p \nmid Q}} \Big(1 + \frac{1}{p}\Big)\Big) + \theta\Big(\frac{C^2 S}{F} \cdot T 2^R\Big),$$

where, as before, $\theta(x)$ denotes a quantity with $-x \le \theta(x) \le x$. By (30),

$$[k(\xi) : k] = T \cdot \#(G_Q \cdot \xi) = \frac{TN^2}{Q_N^2} \prod_{\substack{p \mid N \\ p \nmid Q}} \Big(1 - \frac{1}{p^2}\Big). \tag{35}$$

Since $r \le r_0$, it follows that

$$\frac{N\big(\xi, \mathcal{S}_E(a,r)\big)}{[k(\xi) : k]} = \frac{\text{area } \mathcal{S}(a,r)}{\text{area } \mathcal{F}} + \theta\Big(\frac{2CSr_0}{F} \cdot \frac{Q_N}{N \prod_{\substack{p \mid N \\ p \nmid Q}} \big(1 - \frac{1}{p}\big)}\Big) + \theta\Big(\frac{C^2 S}{F} \cdot \frac{2^R Q_N^2}{N^2 \prod_{\substack{p \mid N \\ p \nmid Q}} \big(1 - \frac{1}{p^2}\big)}\Big).$$

Here area $\mathcal{S}(a,r)$/area $\mathcal{F} = \mu(\mathcal{S}_E(a,r))$. Note that T is bounded by the order of $\mathrm{GL}_2(\mathbb{Z}/Q\mathbb{Z})$), Q_N is bounded by Q, and

$$N \prod_{p \mid N} \Big(1 - \frac{1}{p}\Big) \ge N^{1-\varepsilon}$$

for each $\varepsilon > 0$ and each sufficiently large N. Using (35) and the fact that

$$1 \ge \prod_{p \mid N, p \nmid Q} \Big(1 - \frac{1}{p^2}\Big) \ge 1/\zeta(2)$$

one sees that the first error term is $O_\gamma([k(\xi) : k]^{-\gamma})$ for each $\gamma < 1/2$. Similarly, $2^R \le d(N) \le N^\varepsilon$ for each $\varepsilon > 0$ and each sufficiently large N. Thus the second error term is negligible in comparison with the first. This completes the proof when E does not have complex multiplication.

Case 2. Suppose E has complex multiplication. Let K be the CM field of E, and let $\mathbb{O} \subset \mathbb{O}_K$ be the order corresponding to E. After enlarging k if necessary, we can assume that $K \subset k$. Let $\Lambda \subset \mathbb{C}$ be a lattice such that $E \cong \mathbb{C}/\Lambda$. Without loss of generality, we can assume that $\Lambda \subset K$. Fix an analytic isomorphism $\vartheta : \mathbb{C}/\Lambda \cong E(\mathbb{C})$.

By the theory of complex multiplication (see [Shimura 1971], [Lang 1973], or [Silverman 1994, Chapter II]), $E(\bar{k})_{\text{tors}}$ is rational over k^{ab}, the maximal abelian extension of k. Let $k_{\mathbb{A}}^{\times}$ be the idèle ring of k, and for $s \in k_{\mathbb{A}}^{\times}$ let $[s,k]$ be the Artin map acting on k^{ab}. Given $\sigma \in \text{Gal}(\bar{k}/k)$, take $s \in k_{\mathbb{A}}^{\times}$ with $\sigma|_{k^{ab}} = [s,k]$, and put $w = N_{k/K}(s) \in K_{\mathbb{A}}^{\times}$. There is an action of $K_{\mathbb{A}}^{\times}$ on lattices, defined semilocally, which associates to w and Λ a new lattice $w^{-1}\Lambda$. This action extends to a map $w^{-1} : K/\Lambda \to K/w^{-1}\Lambda$. There is also a homomorphism $\psi : k_{\mathbb{A}}^{\times} \to K^{\times}$, the grössencharacter of E, which has the property that $\psi(s)N_{k/K}(s)^{-1}\Lambda = \Lambda$. Put $\kappa = \psi(s) \in K^{\times}$.

With this notation, there is a commutative diagram

$$\begin{array}{ccccc}
K/\Lambda & \hookrightarrow & \mathbb{C}/\Lambda & \xrightarrow{\vartheta} & E(\bar{k})_{\text{tors}} \\
\downarrow{\scriptstyle w^{-1}} & & & & \downarrow{\scriptstyle \sigma} \\
K/w^{-1}\Lambda & \hookrightarrow & \mathbb{C}/w^{-1}\Lambda & \longrightarrow & E(\bar{k})_{\text{tors}} \\
\downarrow{\scriptstyle \kappa} & & & & \downarrow{\scriptstyle id} \\
K/\Lambda & \hookrightarrow & \mathbb{C}/\Lambda & \xrightarrow{\vartheta} & E(\bar{k})_{\text{tors}}
\end{array}$$

in which the vertical arrows on the left are multiplication by w^{-1} and κ respectively, and those on the right are the galois action (see [Shimura 1971, Proposition 7.40, p. 211], or [Lang 1973, Theorem 8, p. 137]). Note that the same analytic isomorphism ϑ appears in the top and bottom rows. Thus, if $\xi \in E(\bar{k})_{\text{tors}}$ corresponds to $x \in K/\Lambda$, and $\sigma|_{k^{ab}} = [s,k]$, then

$$\sigma(\xi) \;=\; \vartheta(\psi(s)N_{k/K}(s)^{-1}x)\,.$$

This gives an explicit description of the galois action on torsion points in terms of adelic "multiplication".

The action of $K_{\mathbb{A}}^{\times}$ in the diagram is as follows. Let $L \subset K$ be a lattice. For each rational prime p of $\mathbb{Q}$, write $L_p = L \otimes_{\mathbb{Z}} \mathbb{Z}_p$ and $K_p = K \otimes_{\mathbb{Q}} \mathbb{Q}_p$; if $w \in K_{\mathbb{A}}^{\times}$, let w_p be its p-component. Then $w_p^{-1}L_p$ is a $\mathbb{Z}_p$-lattice in K_p. There is a unique lattice $M \subset K$ such that $M_p = w_p^{-1}L_p$ for each p [Lang 1973, Theorem 8, p. 97], and $w^{-1}L$ is defined to be M. Likewise, if $x \in K/L$, lift it to an element of $K \subset K_{\mathbb{A}}$ and write $x_p \in K_p$ for its p-component; there is a $y \in K$ such that $w_p^{-1}x_p \pmod{w^{-1}L_p} = y \pmod{M_p}$ for each p, and $w^{-1}(x \pmod{L})$ is defined to be $y \pmod{M}$.

The order $\mathbb{O}$ has the form $\mathbb{O} = \mathbb{Z} + c\mathbb{O}_K$ for some integer $c \geq 1$, and c is called the conductor of $\mathbb{O}$. The lattice Λ is a proper $\mathbb{O}$-lattice, meaning that $\mathbb{O} = \{x \in K : x\Lambda \subset \Lambda\}$. For any order $\mathbb{O}$, there are only finitely many homothety classes of proper $\mathbb{O}$-lattices [Lang 1973, Theorem 7, p. 95]. Write $\mathbb{O}_p = \mathbb{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ and $\mathbb{O}_{K,p} = \mathbb{O}_K \otimes_{\mathbb{Z}} \mathbb{Z}_p$.

If $p \nmid c$, then $\mathcal{O}_p = \mathcal{O}_{K,p} \cong \prod_{\mathfrak{p}\mid p} \mathcal{O}_{K,\mathfrak{p}}$, where $\mathfrak{p}$ runs over the primes of K lying over p, and $\mathcal{O}_{K,\mathfrak{p}}$ is the completion of $\mathcal{O}_K$ at $\mathfrak{p}$.

Let U be the kernel of the grössencharacter $\psi : k_{\mathbb{A}}^{\times} \to K^{\times}$, and take $W = N_{k/K}(U) \subset K_{\mathbb{A}}^{\times}$. Since ψ is continuous, there is an integer $Q \geq 1$ such that, for each $p \mid Q$, the subgroup $1 + Q\mathcal{O}_{K,p} \subset \mathcal{O}_{K,p}^{\times}$ is contained in W_p and for each $p \nmid Q$, $\mathcal{O}_{K,p}^{\times} \subset W_p$. If $w \in W$, then $w^{-1}\Lambda = \Lambda$, so $w_p \in \mathcal{O}_p^{\times}$. Hence $c \mid Q$.

Noting that $\mathcal{O}_p = \mathcal{O}_{K,p}$ if $p \nmid Q$, let $W_Q \subset K_{\mathbb{A}}^{\times}$ be the subgroup

$$\mathbb{C}^{\times} \times \prod_{p\mid Q} (1 + Q\mathcal{O}_p) \times \prod_{p \nmid Q} \mathcal{O}_p^{\times} \subset W,$$

and let U_Q be its preimage in $k_{\mathbb{A}}^{\times}$ under the norm map. Put

$$G_Q = \{\sigma \in \mathrm{Gal}(\bar{k}/k) : \sigma|_{k^{ab}} = [s, k] \text{ for some } s \in U_Q\}.$$

Then G_Q is open and normal in $\mathrm{Gal}(\bar{k}/k)$.

Step 1: Determining the size of a galois orbit under G_Q. Fix $\xi \in E(\bar{k})_{\mathrm{tors}}$. Suppose ξ has order N; put $Q_N = \gcd(Q, N)$. For suitable right coset representatives $\sigma_1, \ldots, \sigma_T$ of G_Q in $\mathrm{Gal}(\bar{k}/k)$, the orbit $\mathrm{Gal}(\bar{k}/k) \cdot \xi$ decomposes as a disjoint union of G_Q-orbits:

$$\mathrm{Gal}(\bar{k}/k) \cdot \xi = \bigcup_{i=1}^{T} G_Q \cdot \sigma_i(\xi).$$

As before, the orbits $G_Q \cdot \sigma_i(\xi) = \sigma_i(G_Q \cdot \xi)$ all have the same size, and $[k(\xi) : k] = T \cdot \#(G_Q \cdot \xi)$.

Let ξ correspond to $x + \Lambda \in K/\Lambda$. Write $\Lambda(x)$ for the $\mathcal{O}$-lattice $\mathcal{O}x + \Lambda$; since ξ has order N, $[\Lambda(x) : \Lambda] \geq N$. More generally, for any integer m, put $\Lambda(mx) = \mathcal{O} \cdot mx + \Lambda = m\mathcal{O}x + \Lambda$. Note that

$$\Lambda(mx)/\Lambda \cong \prod_{p\mid N} \Lambda(mx)_p/\Lambda_p = \prod_{p\mid N} (m\mathcal{O}_p x + \Lambda_p)/\Lambda_p.$$

If $p \mid Q$, then G_Q acts on ξ_p through the subgroup $1 + p^{\mathrm{ord}_p(Q)}\mathcal{O}_p \subset \mathcal{O}_p^{\times}$. Noting that $\mathrm{ord}_p(Q_N) = \min(\mathrm{ord}_p(Q), \mathrm{ord}_p(N))$ and that $p^{\mathrm{ord}_p(Q)}x \in \Lambda_p$ if $\mathrm{ord}_p(Q) \geq \mathrm{ord}_p(N)$, we have

$$G_Q \cdot \xi_p \cong (x + p^{\mathrm{ord}_p(Q)}\mathcal{O}_p x + \Lambda_p)/\Lambda_p = (x + \Lambda(p^{\mathrm{ord}_p(Q_N)}x)_p)/\Lambda_p.$$

Thus $\#(G_Q \cdot \xi_p) = [\Lambda(p^{\mathrm{ord}_p(Q_N)}x)_p : \Lambda_p]$.

If $p \nmid Q$, then $\mathcal{O}_p = \mathcal{O}_{K,p}$ and G_Q acts on ξ_p through $\mathcal{O}_p^{\times} \cong \prod_{\mathfrak{p}\mid p} \mathcal{O}_{K,\mathfrak{p}}^{\times}$. For each $\mathfrak{p} \mid p$, and each $\mathcal{O}$-lattice L, we have $L_p \cong (\mathcal{O}_K L)_p$ where $\mathcal{O}_K L$ is an $\mathcal{O}_K$-fractional ideal. Thus $\mathrm{ord}_{\mathfrak{p}}(L) := \mathrm{ord}_{\mathfrak{p}}(\mathcal{O}_K L)$ is well defined. Write $\mathrm{ord}_{\mathfrak{p}}(\xi) =$

$\mathrm{ord}_{\mathfrak{p}}(\Lambda) - \mathrm{ord}_{\mathfrak{p}}(\Lambda(x))$. Then $\Lambda(x)_p/\Lambda_p \cong \prod_{\mathfrak{p}\mid p} \mathbb{O}_K/\mathfrak{p}^{\mathrm{ord}_{\mathfrak{p}}(\xi)}$ and

$$\#(G_Q \cdot \xi_p) = [\Lambda(x)_p : \Lambda_p] \cdot \prod_{\substack{\mathfrak{p}\mid p \\ \mathrm{ord}_{\mathfrak{p}}(\xi)>0}} \Big(1 - \frac{1}{N\mathfrak{p}}\Big),$$

where $N\mathfrak{p} = \#(\mathbb{O}_K/\mathfrak{p})$ is the norm of $\mathfrak{p}$.

Combining these formulas, and using that

$$\prod_{p\mid N} [\Lambda(p^{\mathrm{ord}_p(Q_N)}x)_p : \Lambda_p] = [\Lambda(Q_N x) : \Lambda],$$

we obtain

$$\#(G_Q \cdot \xi) = [\Lambda(Q_N x) : \Lambda] \cdot \prod_{\substack{\mathfrak{p}\mid N, \mathfrak{p}\nmid Q \\ \mathrm{ord}_{\mathfrak{p}}(\xi)>0}} \Big(1 - \frac{1}{N\mathfrak{p}}\Big). \tag{36}$$

Step 2: Counting translated lattice points in convex domains. If L is any $\mathbb{O}$-lattice, and $F(L)$ is the area of a fundamental domain for $\mathbb{C}/L$, then by Minkowski's theorem there is a point $0 \neq \ell \in L$ with $|\ell| \leq (4/\pi)^{1/2} F(L)^{1/2}$. Here L is a proper $\mathbb{O}'$-lattice for some order $\mathbb{O}'$ with conductor $c' \mid c$. There are only finitely many such orders $\mathbb{O}'$, and for each $\mathbb{O}'$ there are only finitely many homothety classes of proper $\mathbb{O}'$-lattices, so there are only finitely many homothety classes of $\mathbb{O}$-lattices. Hence there is a constant C_1, independent of L, such that L has a fundamental domain $\mathcal{F}(L)$ contained in the ball $B(0, C_1 \cdot F(L)^{1/2})$. In turn, there is a constant C, independent of L, such that $\mathcal{F}(L) \subset \mathcal{S}(0, C \cdot F(L)^{1/2})$. This fact is the crux of the argument in the CM case.

Again, if L is an $\mathbb{O}$-lattice, then for each ideal ϖ of $\mathbb{O}_K$ coprime to c, there is a unique lattice ϖL defined by the property that $(\varpi L)_q = (\varpi \mathbb{O}_K L)_q$ for all primes $q \mid N\varpi$, and $(\varpi L)_q = L_q$ for all primes $q \nmid N\varpi$. This lattice has index $[L : \varpi L] = N\varpi$.

We will apply this taking $L = \Lambda(Q_N x) = Q_N \mathbb{O} x + \Lambda$. Note that the fundamental domain $\mathcal{F}(\varpi \Lambda(Q_N x))$ has area $F \cdot N\varpi / [\Lambda(Q_N x) : \Lambda]$, where F is the area of a fundamental domain $\mathcal{F}$ for Λ. By the same argument leading to (33) we find that for each $x_0 \in \mathbb{C}$

$$\Big| \#\big((x_0 + \varpi \Lambda(Q_N x)) \cap \mathcal{S}(a, r)\big) - \frac{\text{area } \mathcal{S}(a,r)}{\text{area } \mathcal{F}} \cdot \frac{[\Lambda(Q_N x) : \Lambda]}{N\varpi} \Big|$$
$$\leq \frac{2CSr}{F} \cdot \Big(\frac{[\Lambda(Q_N x) : \Lambda]}{N\varpi}\Big)^{1/2} + \frac{C^2 S}{F}. \tag{37}$$

Step 3: Inclusion/exclusion. Now consider a set $\mathcal{S}(a, r)$, where $a \in \mathbb{C}$ and $r \leq r_0$. For each $\sigma_i(\xi)$, we will compute $\#\big((G_Q \cdot \sigma_i(\xi)) \cap \mathcal{S}_E(a, r)\big)$. Fix σ_i, and replace ξ by $\sigma_i(\xi)$ in the discussion above. Let $x \in K/\Lambda$ correspond to $\sigma_i(\xi)$, and let $\mathfrak{p}_1, \ldots, \mathfrak{p}_R$ be the distinct primes of $\mathbb{O}_K$ dividing N but not Q, for which

$\mathrm{ord}_{\mathfrak{p}}(\Lambda(x)) \neq \mathrm{ord}_{\mathfrak{p}}(\Lambda)$. If there are no such primes, take $\mathfrak{p}_1 \cdots \mathfrak{p}_R = 1$ in the argument below. (Note that the $\mathfrak{p}_j$ are independent of σ_i, since $K \subset k$ and for $p \nmid Q$, σ_i acts on ξ through $\mathcal{O}_p^\times$.) Thus there is a one-to-one correspondence between elements of $G_Q \cdot \sigma_i(\xi)$, and cosets $y + \Lambda$ for $y \in K$ such that $y \in x + \Lambda(Q_N x)$ and $y \notin \mathfrak{p}_j \Lambda(x)$ for $j = 1, \ldots, R$. Since $\Lambda(Q_N x) \subset \Lambda(x)$, such y necessarily belong to $\Lambda(x)$. The index $[\Lambda(Q_N x) : \Lambda]$ in (37) is independent of σ_i by (36), since $\#(G_Q \cdot \sigma_i(\xi))$ and the $\mathfrak{p}_j$ are independent of σ_i.

The lattices $\Lambda(Q_N x)$ and $\mathfrak{p}_1 \cdots \mathfrak{p}_R \Lambda(x)$ have coprime indices in $\Lambda(x)$, so there is an $x_0 \in \Lambda(x)$ such that $x_0 \equiv x \pmod{\Lambda(Q_N x)}$ and $x_0 \equiv 0 \pmod{\mathfrak{p}_1 \cdots \mathfrak{p}_R \Lambda(x)}$. Further, for any $\mathcal{O}_K$-ideal ϖ dividing $\mathfrak{p}_1 \cdots \mathfrak{p}_R$,

$$\Lambda(Q_N x) \cap \Big(\bigcap_{\mathfrak{p}_j \mid \varpi} \mathfrak{p}_j \Lambda(x) \Big) = \varpi \Lambda(Q_N x).$$

Clearly $y \in x + \Lambda(Q_N x)$ if and only if $y \in x_0 + \Lambda(Q_N x)$, and $y \in \mathfrak{p}_j \Lambda(x)$ if and only if $y \in x_0 + \mathfrak{p}_j \Lambda(x)$. Since $\mathcal{S}(a, r)$ injects into $\mathbb{C}/\Lambda$, by inclusion/exclusion

$$\begin{aligned} &\#\Big((G_Q \cdot \sigma_i(\xi)) \cap \mathcal{S}_E(a, r) \Big) \\ &\qquad = \sum_{\varpi \mid \mathfrak{p}_1 \cdots \mathfrak{p}_R} (-1)^{\lambda_K(\varpi)} \cdot \#\Big((x_0 + \varpi \Lambda(Q_N x)) \cap \mathcal{S}(a, r) \Big), \end{aligned} \tag{38}$$

where $\lambda_K(\varpi)$ is the number of distinct prime ideals of $\mathcal{O}_K$ dividing ϖ.

Inserting (37) in the inclusion/exclusion formula (38) and summing over all $\sigma_i(\xi)$, we get

$$\begin{aligned} &N\big(\xi, \mathcal{S}_E(a, r)\big) \\ &\quad = \frac{\text{area } \mathcal{S}(a, r)}{\text{area } \mathcal{F}} \cdot T[\Lambda(Q_N x) : \Lambda] \prod_{j=1}^{R} \Big(1 - \frac{1}{N\mathfrak{p}_j}\Big) \\ &\qquad + \theta\Big(\frac{2CSr}{F} \cdot T[\Lambda(Q_N x) : \Lambda]^{1/2} \prod_{j=1}^{R} \Big(1 + \frac{1}{N\mathfrak{p}_j^{1/2}}\Big) \Big) + \theta\Big(\frac{C^2 S}{F} \cdot T 2^R \Big). \end{aligned}$$

By (36), $[k(\xi) : k] = T[\Lambda(Q_N x) : \Lambda] \prod_{j=1}^{R} \big(1 - \frac{1}{N\mathfrak{p}_j}\big)$. Since $r \leq r_0$ and

$$\prod_{j=1}^{R} \Big(1 + \frac{1}{N\mathfrak{p}_j^{1/2}}\Big) \leq 2^R,$$

we have

$$\begin{aligned} \frac{N\big(\xi, \mathcal{S}_E(a, r)\big)}{[k(\xi) : k]} &= \frac{\text{area } \mathcal{S}(a, r)}{\text{area } \mathcal{F}} + \theta\Big(\frac{C^2 S}{F} \cdot \frac{T 2^R}{[k(\xi) : k]}\Big) \\ &\quad + \theta\Big(\frac{2CSr_0}{F} \cdot \frac{T^{1/2} 2^R}{\big(\prod_{j=1}^{R} (1 - 1/N\mathfrak{p}_j)\big)^{1/2}} \cdot \frac{1}{[k(\xi) : k]^{1/2}} \Big). \end{aligned} \tag{39}$$

As before, area $\mathcal{S}(a,r)$/area $\mathcal{F} = \mu(\mathcal{S}_E(a,r))$. Here $T \le [\mathrm{Gal}(\bar{k}/k) : G_Q]$ is fixed. For each $\varepsilon > 0$ and each sufficiently large N, $2^R \le 2^{\Lambda_K(N)} \le 2^{2\lambda(N)} \le d(N)^2 \le N^\varepsilon$. Likewise, $\prod_{j=1}^{R}(1-1/N\mathfrak{p}) \ge \prod_{p\,|\,N}(1-1/p)^2 \ge C/(\log\log N)^2$ for some constant $C > 0$, where the last inequality follows from [Hardy and Wright 1954, Theorem 328, p. 267]. Finally, since ξ has order N and $Q_N \le Q$ is bounded, $[\Lambda(Q_N x) : \Lambda] \ge N/Q$, and so

$$[k(\xi) : k] \;\ge\; T \cdot N/Q \cdot C/(\log\log N)^2 \;\ge\; TC/Q \cdot N^{1-\varepsilon} \tag{40}$$

for all large N. Combining these shows that for each $0 < \gamma < 1/2$, the first error term is $\mathbb{O}_\gamma([k(\xi) : k]^{-\gamma})$. The same estimates show the second error term is negligible in comparison to the first. This completes the proof when E has complex multiplication. □

Before leaving this section, we note that the arguments above provide lower bounds for the degree $[k(\xi) : k]$ in terms of the order N of ξ, as required by (27). When E does not have complex multiplication, then since T is fixed, $Q_N \le Q$, and $\prod_p(1-1/p^2)$ converges to a nonzero limit, (35) shows there is a constant C_1 depending only on E such that

$$[k(\xi) : k] \;\ge\; C_1 N^2. \tag{41}$$

When E has complex multiplication, then since T and Q are fixed, (40) shows that there is a constant C_2 depending only on E such that

$$[k(\xi) : k] \;\ge\; C_2 N/(\log\log N)^2. \tag{42}$$

3. Context

Theorems 0.1 and 0.2 are the first known cases of general conjectures by the second author (which were refined through conversations with J. Silverman and S. Zhang) concerning dynamical systems and abelian varieties.

As before, let k be a number field, and let S be a finite set of places of k containing the archimedean places. Let $\mathbb{O}_{k,S}$ be the ring of S-integers of k.

Conjecture 3.1 (Su-Ion Ih). *Let $R(x) \in k(x)$ be a rational function of degree at least 2, and consider the dynamical system associated to the map $R_* : \mathbb{P}^1 \to \mathbb{P}^1$. Let $\alpha \in \mathbb{P}^1(\bar{k})$ be nonpreperiodic for R_*. Then there are only finitely many preperiodic points $\xi \in \mathbb{P}^1(\bar{k})$ that are S-integral with respect to α, that is, whose Zariski closures in $\mathbb{P}^1/\mathrm{Spec}(\mathbb{O}_{k,S})$ do not meet the Zariski closure of α.*

Conjecture 3.2 (Su-Ion Ih). *Let A/k be an abelian variety, and let $\mathcal{A}_S/\mathrm{Spec}(\mathbb{O}_{k,S})$ be a model of A. Let D be a nonzero effective divisor on A, defined over $\bar{k}$, at least one of whose irreducible components is not the translate of an abelian subvariety by a torsion point, and let $\mathrm{cl}(D)$ be its Zariski closure in $\mathcal{A}_S$. Then the set*

Type of variety	Type of rationality	k	$\bar{k}$
Compact	$k,\bar{k}$-rationality	Mordell–Lang Conjecture	Manin–Mumford Conjecture
Noncompact	$\mathbb{O}_k,\overline{\mathbb{Z}}$-rationality	Lang's Conjecture	Ih's Conjecture 3.2

$A_{D,S}(\overline{\mathbb{Z}})_{\mathrm{tors}}$, *consisting of all torsion points of* $A(\bar{k})$ *whose closure in* $\mathcal{A}_S$ *is disjoint from* $\mathrm{cl}(D)$, *is not Zariski dense in* A.

Theorem 0.1 establishes Conjecture 3.1 for the maps $R(x) = x^d$ with $|d| \geq 2$, whose preperiodic points are 0, ∞ and the roots of unity. It is possible to prove the conjecture for Chebyshev maps by similar methods, though we do not do so here.

Theorem 0.2, in addition to being the one-dimensional case of Conjecture 3.2, is equivalent to Conjecture 3.1 for Lattès maps. That is, if E/k is an elliptic curve, let $R \in k(x)$ be the degree 4 map on the x-coordinate corresponding to the doubling map on E, so that the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{[2]} & E \\ {\scriptstyle x}\downarrow & & \downarrow{\scriptstyle x} \\ \mathbb{P}^1 & \xrightarrow{R_*} & \mathbb{P}^1 \end{array}$$

Then $\beta \in E(\bar{k})$ is a torsion point if and only $x(\beta)$ is preperiodic for R_*.

Part of the motivation for Conjecture 3.2 is the following analogy between diophantine theorems over k and $\bar{k}$, and over $\mathbb{O}_k$ and $\overline{\mathbb{Z}}$ (the ring of all algebraic integers). Let A/k be an abelian variety, and let X be a nontorsion subvariety of A (that is, X is not the translate of an abelian subvariety by a torsion point). Recall that the Mordell–Lang Conjecture (proved by Faltings) says that $A(k) \cap X$ is not Zariski dense in X; while the Manin–Mumford Conjecture (first proved by Raynaud) says that $A(\bar{k})_{\mathrm{tors}} \cap X$ is not Zariski dense in X. Likewise, Lang's conjecture (also proved by Faltings) says that if D is an effective ample divisor on A, then the set $A_D(\mathbb{O}_k)$ of $\mathbb{O}_k$-integral points of A not meeting $\mathrm{supp}(D)$ is finite. Note that A is compact, whereas $A_D = A\backslash\mathrm{supp}(D)$ is noncompact.

Conjecture 3.1 is motivated by Conjecture 3.2 and the familiar analogy between torsion points of abelian varieties and preperiodic points of rational maps.

J. Silverman [1993] proved the following result, which is somewhat related to Conjecture 3.1: If the backward orbit of $\alpha \in \mathbb{P}^1(\bar{k})$ under a rational function R of degree ≥ 2 is infinite, then for every $\beta \in \mathbb{P}^1(\bar{k})$, there are only finitely many points in the forward orbit of β under R that are S-integral with respect to α.

More recently, C. Petsche [2007] has proved Conjecture 3.1 under the additional hypothesis that α is "totally Fatou", meaning that for every place v of k and every embedding σ of $\bar{k}$ into $\bar{k}_v$, $\sigma(\alpha)$ is in the v-adic Fatou set of R.

In closing, we note that an important ingredient of the proofs of Theorems 0.1 and 0.2 was a quantitative equidistribution theorem for torsion points. A quantitative equidistribution theorem for points of small height with respect to an arbitrary dynamical system on $\mathbb{P}^1$ has recently been proved by C. Favre and J. Rivera-Letelier [2006, théorème 6].

Acknowledgments

The authors would like to thank the referee for some useful pointers to the literature. The second author would also like to thank J. Silverman and S. Zhang for helping to refine the conjectures in Section 3.

References

[Autissier 2006] P. Autissier, "Sur une question d'équirépartition de nombres algébriques", *C. R. Math. Acad. Sci. Paris* **342**:9 (2006), 639–641. MR 2007b:11163 Zbl pre05045962

[Baker 1975] A. Baker, *Transcendental number theory*, Cambridge University Press, London, 1975. MR 54 #10163 Zbl 0297.10013

[Baker and Hsia 2005] M. H. Baker and L.-C. Hsia, "Canonical heights, transfinite diameters, and polynomial dynamics", *J. Reine Angew. Math.* **585** (2005), 61–92. MR 2006i:11071 Zbl 1071.11040

[Bang 1886] A. S. Bang, "Taltheoretiske undersøgelser", *Zeuthen Tidskr.* **4** (1886), 70–80, 130–137. JFM 19.0168.02

[Bilu 1997] Y. Bilu, "Limit distribution of small points on algebraic tori", *Duke Math. J.* **89**:3 (1997), 465–476. MR 98m:11067 Zbl 0918.11035

[Chambert-Loir 2006] A. Chambert-Loir, "Mesures et équidistribution sur les espaces de Berkovich", *J. Reine Angew. Math.* **595** (2006), 215–235. MR 2008b:14040 Zbl 05039459

[Cheon and Hahn 1999] J. Cheon and S. Hahn, "The orders of the reductions of a point in the Mordell–Weil group of an elliptic curve", *Acta Arith.* **88**:3 (1999), 219–222. MR 2000i:11084 Zbl 0933.11029

[Conway 1973] J. B. Conway, *Functions of one complex variable*, Graduate Texts in Mathematics **11**, Springer, New York, 1973. MR 56 #5843 Zbl 0277.30001

[David and Hirata-Kohno 2002] S. David and N. Hirata-Kohno, "Recent progress on linear forms in elliptic logarithms", pp. 26–37 in *A panorama of number theory or the view from Baker's garden* (Zürich, 1999), edited by G. Wüstholz, Cambridge Univ. Press, Cambridge, 2002. MR 2004e:11076 Zbl 1041.11053

[Everest and Fhlathúin 1996] G. R. Everest and B. N. Fhlathúin, "The elliptic Mahler measure", *Math. Proc. Cambridge Philos. Soc.* **120**:1 (1996), 13–25. MR 97e:11064 Zbl 0865.11068

[Everest and Ward 1999] G. Everest and T. Ward, *Heights of polynomials and entropy in algebraic dynamics*, Universitext, Springer, London, 1999. MR 2000e:11087 Zbl 0919.11064

[Favre and Rivera-Letelier 2006] C. Favre and J. Rivera-Letelier, "Équidistribution quantitative des points de petite hauteur sur la droite projective", *Math. Ann.* **335**:2 (2006), 311–361. MR 2007g:11074 Zbl pre05035986

[Hardy and Wright 1954] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 3rd ed., Clarendon Press, Oxford, 1954. MR 16,673c Zbl 0058.03301

[Lang 1973] S. Lang, *Elliptic functions*, Addison-Wesley, Reading, MA, 1973. MR 53 #13117 Zbl 0316.14001

[Petsche 2007] C. Petsche, "S-integral preperiodic points for dynamical systems over number fields", preprint, 2007. arXiv 0709.3879v2

[Rumely 1989] R. S. Rumely, *Capacity theory on algebraic curves*, Lecture Notes in Mathematics **1378**, Springer, Berlin, 1989. MR 91b:14018 Zbl 0679.14012

[Scanlon 1999] T. Scanlon, "The conjecture of Tate and Voloch on p-adic proximity to torsion", *Internat. Math. Res. Notices* 17 (1999), 909–914. MR 2000i:11100 Zbl 0986.11038

[Schinzel 1974] A. Schinzel, "Primitive divisors of the expression $A^n - B^n$ in algebraic number fields", *J. Reine Angew. Math.* **268/269** (1974), 27–33. MR 49 #8961 Zbl 0287.12014

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR 52 #8126 Zbl 0235.14012

[Shimura 1971] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan **11**, Iwanami Shoten, Tokyo, 1971. MR 47 #3318 Zbl 0221.10029

[Silverman 1986] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1986. MR 87g:11070 Zbl 0585.14026

[Silverman 1993] J. H. Silverman, "Integer points, Diophantine approximation, and iteration of rational maps", *Duke Math. J.* **71**:3 (1993), 793–829. MR 95e:11070 Zbl 0811.11052

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Silverman 1995] J. H. Silverman, "Exceptional units and numbers of small Mahler measure", *Experiment. Math.* **4**:1 (1995), 69–83. MR 96j:11150 Zbl 0851.11064

[Szpiro and Tucker 2005] L. Szpiro and T. Tucker, "Equidistribution and generalized Mahler measures", preprint, 2005. arXiv math/0510404v3

[Szpiro, Ullmo and Zhang 1997] L. Szpiro, E. Ullmo, and S. Zhang, "Équirépartition des petits points", *Invent. Math.* **127**:2 (1997), 337–347. MR 98i:14027 Zbl 0991.11035

[Ullmo 1995] E. Ullmo, "Points entiers, points de torsion et amplitude arithmétique", *Amer. J. Math.* **117**:4 (1995), 1039–1055. MR 96j:14016 Zbl 0863.14016

Communicated by Karl Rubin
Received 2007-10-29 Revised 2008-01-11 Accepted 2008-01-11

mbaker@math.gatech.edu *School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332-0160, United States*

ih@math.colorado.edu *Department of Mathematics, University of Colorado at Boulder, Campus Box 395, Boulder, CO 80309-0395, United States*

rr@math.uga.edu *Department of Mathematics, University of Georgia, Athens, Georgia 30602-0002, United States*