A FINITENESS PROPERTY OF TORSION POINTS

MATTHEW BAKER, SU-ION IH, AND ROBERT RUMELY

Abstract. 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 $\alpha \in G(\overline{k})$ is nontorsion, then there are only finitely many torsion points $\xi \in G(\overline{k})_{\text{tors}}$ which are $S$-integral with respect to $\alpha$. We also formulate conjectural generalizations for dynamical systems and for abelian varieties.

0. Introduction.

Let $k$ be a number field, with ring of integers $\mathcal{O}_k$ and algebraic closure $\overline{k}$. In this paper we prove finiteness theorems for torsion points which 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 $\alpha, \beta \in \mathbb{P}^1(\overline{k})$, let $\text{cl}(\alpha), \text{cl}(\beta)$ be their Zariski closures in $\mathbb{P}^1/\text{Spec}(\mathcal{O}_k)$. By definition, $\beta$ is $S$-integral relative to $\alpha$ if $\text{cl}(\beta)$ does not meet $\text{cl}(\alpha)$ outside $S$. Thus, $\beta$ is $S$-integral relative to $\alpha$ if and only if for each place $v$ of $k$ not in $S$, and each pair of $k$-embeddings $\sigma : k(\beta) \hookrightarrow \overline{k}_v$, $\tau : k(\alpha) \hookrightarrow \overline{k}_v$, we have $||\sigma(\beta), \tau(\alpha)||_v = 1$ under the spherical metric on $\mathbb{P}^1(\overline{k}_v)$. Equivalently, for all $\sigma, \tau$,

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

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 $\alpha \in \mathbb{P}^1(\overline{k})$ with Weil height $h(\alpha) > 0$; that is, identifying $\mathbb{P}^1(\overline{k})$ with $\overline{k} \cup \{\infty\}$, $\alpha$ is not 0 or $\infty$ or a root of unity. Then there are only finitely many roots of unity in $\overline{k}$ which are $S$-integral with respect to $\alpha$.

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

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 $\alpha \in E(\overline{k})$ is nontorsion (has canonical height $\hat{h}(\alpha) > 0$), there are only finitely many torsion points $\xi \in E(\overline{k})_{\text{tors}}$ which are $S$-integral with respect to $\alpha$.

By $S$-integrality we mean that the Zariski closures of $\xi$ and $\alpha$ in the model $E/\text{Spec}(\mathcal{O}_k)$ 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 $E$.

Date: October 22, 2007.

2000 Mathematics Subject Classification. Primary 11G05, 37F10, Secondary 11J86, 11J71, 11G50.

Key words and phrases. elliptic curve, equidistribution.

Work supported in part by NSF grant DMS-0300784.
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 \( \alpha \) is rational over \( k \). Given a place \( v \) of \( k \), let \( \overline{k}_v \) be the algebraic closure of the completion \( k_v \), and let \( \lambda_v \) be the normalized canonical local height occurring in the decomposition of the global height. On the one hand, elementary properties of heights show that since \( \alpha \) is nontorsion, for each torsion point \( \xi_n \) one has

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

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

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

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 \( \alpha \) is nontorsion.

Examples show that the conclusion is false if \( \alpha \) 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 the first, we prove Theorem 0.1 for \( \mathbb{G}_m \); in the second, we prove Theorem 0.2 for elliptic curves. In the third, 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 \( \overline{k}_v \) is an algebraic closure of \( k_v \), there is a unique extension of \( |x|_v \) to \( \overline{k}_v \), 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 \( \overline{k}_v \), 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 \( \sigma \) ranges over all embeddings of \( L \) into \( \overline{k}_v \) fixing \( k \) we have

\[
\sum_{\sigma : L / k \hookrightarrow \overline{k}_v} \log(|\sigma(\beta)|_v) = \sum_{w|v} \log(|\beta|_w).
\]
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 $\alpha \in \mathbb{Q}$, $h(\alpha)$ 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 $\alpha$ is a root of unity.

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

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

A) The hypothesis $h(\alpha) > 0$ is necessary:

Take $k = \mathbb{Q}$. If $\alpha = 0$ or $\alpha = \infty$, then each root of unity $\zeta_n$ is integral with respect to $\alpha$ 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 $\alpha$ at all finite places. If $\alpha = \zeta_N$ is a primitive $N$th root of unity with $N > 1$, let $\zeta_m$ 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 $\mathbb{Z}$, the ring of all algebraic integers, and $\zeta_N - \zeta_m$ is also a unit. This holds for all conjugates of $\zeta_N$ and $\zeta_m$. Hence $\zeta_m$ is integral with respect to $\alpha$ 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 $\beta \in \overline{\mathbb{Q}}$ with $h(\beta) < B$ which are $S$-integral with respect to $\alpha$? That is, could finiteness for roots of unity be strengthened to finiteness for small points?

The following example\textsuperscript{1} shows this is not possible. Take $k = \mathbb{Q}$, $\alpha = 2$, and $S = \{\infty\}$. For each $n$, let $\beta_n$ be a root of the polynomial

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

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 $\beta_n$ is a unit. By Rouché’s theorem, $\beta_n$ 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 $\beta_n$ is integral with respect to 2 at all finite places.

1.2. The proof of Theorem 0.1.

Proof. 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 $\zeta$ is a root of unity which is $S$-integral with respect to $\alpha$ over $k$, then each $k$-conjugate of $\zeta$ is $S_{k(\alpha)}$-integral with respect to $\alpha$ over $k(\alpha)$.

\textsuperscript{1}The authors thank Pascal Autissier for correcting an error in an earlier version of this example.
Suppose $\alpha \in k$, and that there are infinitely many distinct roots of unity $\zeta_n$ which are $S$-integral with respect to $\alpha$. For each $n$, we will evaluate the sum

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

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

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

Since $\alpha$ 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 $|\alpha|_v > 1$ then by the ultrametric inequality, for each $\sigma : k(\zeta_n)/k \hookrightarrow K_v$ we have $|\sigma(\zeta_n) - \alpha|_v = |\alpha|_v$. On the other hand, if $|\alpha|_v \leq 1$, the integrality hypothesis gives $|\sigma(\zeta_n) - \alpha|_v = 1$. It follows that for each $v \notin S$

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

so that

$$A_n = \sum_{v \in S} \frac{1}{[k(\zeta_n) : Q]} \sum_{\sigma : k(\zeta_n)/k \hookrightarrow K_v} \log(|\sigma(\zeta_n) - \alpha|_v)$$

$$+ \frac{1}{[k : Q]} \sum_{v \notin S} \max(0, \log(|\alpha|_v)).$$

Now let $n \to \infty$ in (6). Since $S$ is finite, we can interchange the limit and the sum over $v \in S$, obtaining

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

$$+ \frac{1}{[k : Q]} \sum_{v \notin S} \max(0, \log(|\alpha|_v)).$$

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

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

Inserting this in (7) gives $h(\alpha) = 0$, a contradiction.

For each nonarchimedean $v \in S$, (8) 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 $\sigma$, and in the second case $|\sigma(\zeta_n) - \alpha|_v = 1$ for all $n$ and all $\sigma$. Hence we can assume that $|\alpha|_v = 1$. 


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

(A) There is a bound $M(\alpha) > 0$ such that $|\zeta - \alpha|_v \geq M(\alpha)$ for all roots of unity $\zeta \in \overline{k}_v$.

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

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

Assuming $v$ is nonarchimedean and $|\alpha|_v = 1$, let $M(\alpha)$ be as in the Lemma. Fix $0 < r < 1$, and let $N(r)$ be the number of roots of unity in $\overline{k}_v$ with $|\zeta - \alpha|_v < r$. For each $\zeta_n$ and each $\sigma: k(\zeta_n)/k \to \overline{k}_v$, we have $|\sigma(\zeta_n) - \alpha|_v \leq 1$, so

$$0 \geq \lim_{n \to \infty} \frac{1}{[k(\zeta_n) : Q]} \sum_{\sigma: k(\zeta_n)/k \to \overline{k}_v} \log(|\sigma(\zeta_n) - \alpha|_v)$$

$$\geq \lim_{n \to \infty} \frac{1}{[k(\zeta_n) : Q]} \left( ([k(\zeta_n) : k] - N(r)) \cdot \log(r) + N(r) \cdot \log(M(\alpha)) \right)$$

$$= \frac{1}{|k : Q|} \log(r).$$

Since $r < 1$ is arbitrary, the limit in (8) is 0, verifying (8) in this case.

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

By Jensen’s formula (see [Co86], p.280) applied to $f(z) = z - \alpha$,

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

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

The $\text{Gal}(\overline{k}/k)$-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 ([Bi97]) and restriction of scalars, or from the equidistribution theorem for polynomial dynamical systems given in Baker-Hsia ([BH05]). Those theorems show that if $\mu_n$ is the discrete measure

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

where $\delta_P(x)$ is the Dirac measure with mass 1 at $P$, then the $\mu_n$ 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, (8) follows from (9) 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 $\zeta_n$ were extremely close to $\alpha$, or if too many conjugates of $\zeta_n$ clustered near $\alpha$.

The first problem is solved by A. Baker’s theorem on lower bounds for linear forms in logarithms (see Baker [Ba75], Theorem 3.1, p.22). We are assuming that $|\alpha|_v = 1$, and $\alpha$
is not a root of unity. Fix a branch of log with \( \log(z) = \log(|z|) + i\theta, -\pi < \theta \leq \pi \), 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 \( \beta \).

The second problem is settled by a strong form of equidistribution for roots of unity, proved in §1.3 below. It says that for any \( 0 < \gamma < 1 \), the conjugates of the \( \zeta_n \) 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 \( \text{Gal}(\overline{k}/k) \)-conjugates of the roots of unity in \( \overline{k} \) (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 \( \mu(I) = \frac{1}{2\pi} \cdot \text{length}(I) \) for its normalized Haar measure. If \( \zeta \in \overline{k} \) is a root of unity, put

\[
N(\zeta, I) = \# \{ \sigma(\zeta) \in I : \sigma \in \text{Gal}(\overline{k}/k) \}.
\]

Fix \( 0 < \gamma < 1 \). Then for all roots of unity \( \zeta \) and all \( I \),

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

Assuming Proposition 1.3, we will now complete the proof of Theorem 0.1 by showing that (8) holds for archimedean \( v \) where \( |\alpha|_v = 1 \).

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

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

Then the \( \mu_n \) are supported on the unit circle and converge weakly to \( \mu \) 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 \( \log(|z - \alpha|) \) two parts: a continuous “background” function which can be handled by weak convergence, and a function with a logarithmic pole at \( \alpha \) supported in a small neighborhood of \( \alpha \). The terms nearest \( \alpha \) can then be dealt with using Baker’s theorem, while the other terms can be treated by strong equidistribution. Define

\[
larg_{\alpha, \epsilon}(z) = \min(0, \log(|\theta - \theta_0|/\epsilon)) ,
\]
taking \( \larg_{\alpha, \epsilon}(\theta_0) = -\infty \). Then there is a continuous function \( g_{\alpha, \epsilon}(z) \) on \( |z| = 1 \) for which \( \log(|z - \alpha|) = \larg_{\alpha, \epsilon}(z) + g_{\alpha, \epsilon}(z) \).
Fix $0 < \varepsilon < 1$. We will show that for all sufficiently large $n$,

\begin{equation}
| \int_{|z|=1} \log(|z - \alpha|) \, d\mu_n(z) | < 6\varepsilon .
\end{equation}

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

\[ \int_{|z|=1} g_{\alpha,\varepsilon}(z) \, d\mu(z) = - \int_{|z|=1} \log_{\alpha,\varepsilon}(z) \, d\mu(z) = -2 \int_0^\varepsilon \log(\theta/\varepsilon) \frac{d\theta}{2\pi} = \frac{\varepsilon}{\pi} . \]

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

\begin{equation}
| \int_{|z|=1} g_{\alpha,\varepsilon}(z) \, d\mu_n(z) | < \varepsilon .
\end{equation}

To obtain (11), it will suffice to show that for all sufficiently large $n$,

\[ | \int_{|z|=1} \log_{\alpha,\varepsilon}(z) \, d\mu_n(z) | < 5\varepsilon . \]

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 $\log_{\alpha,\varepsilon}(z)$ is supported on $I_\alpha([-\varepsilon, \varepsilon])$, put $D = D_n = \lfloor [k(\zeta_n) : k]^{1/2} \rfloor$ 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 $\zeta_n$.

First consider the union of the two central subarcs, $I_\alpha([\varepsilon/D, \varepsilon/D])$. Let $N$ be the order of $\zeta_n$. Let $\sigma_0(\zeta_n) = e^{2\pi i a/N}$ be the conjugate of $\zeta_n$ closest to $\alpha = e^{i\theta_0}$. We can assume that $|a/N| \leq 1$, 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,

\[ \log_{\alpha,\varepsilon}(\sigma_0(\zeta_n)) > -C \log(N) - \log(\varepsilon) \geq -C \log(N) . \]

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

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

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

\begin{equation}
| \int_{I_\alpha([-\varepsilon/D, \varepsilon/D])} \log(|z - \alpha|) \, d\mu_n(z) | < \varepsilon .
\end{equation}

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

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

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

\[
0 \geq \int_{I_n([-\varepsilon,\varepsilon])\setminus I_n([-\varepsilon/D,\varepsilon/D])} \text{larg}_{\alpha,\varepsilon}(z) \, d\mu_n(z) \\
\geq 2 \sum_{l=1}^{D-1} \log\left(\frac{\ell}{D}\right) \cdot \frac{2\varepsilon}{D} > 4 \int_{0}^{\varepsilon} \log(t/\varepsilon) \, dt = -4\varepsilon.
\]

Combining (12), (13), and (14) gives (11), which completes the proof of Theorem 0.1. \( \square \)

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 [B1886] (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 \( \alpha \) 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 \( \zeta_n \) and a nonzero prime ideal \( p \) of \( \mathbb{Z}[\zeta_n] \) such that \( \alpha \equiv \zeta_n \pmod{p} \). 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 [Sc74] 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 [Si95] 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 \( \alpha \) is a unit, this yields Theorem 0.1 in the same situations as Schinzel’s theorem.

G. Everest and T. Ward ([EW99], 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

\[
\lim_{n \to \infty} \frac{1}{n} \log(\Delta_n(F)) = m(F) > 0,
\]

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 \( \alpha_i \) such that \( |\alpha_i^n - 1|_v < 1 \), and this in turn means there is some \( n \)-th root of unity \( \zeta \) with \( |\alpha_i - \zeta|_v < 1 \). This implies there are infinitely many roots of unity which are not integral with respect to some \( \alpha_i \), as also follows from Theorem 0.1. However, the Everest-Ward theorem does not yield Theorem 0.1.

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

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 \( b_0 \equiv b \pmod{Q} \), \( b_0 \equiv 0 \pmod{p_1 \cdots p_r} \). Then
\[
\{ a \in (C, D) \cap \mathbb{Z} : a \equiv b \pmod{Q}, (a, N) = 1 \} = \{ a \in (C, D) \cap \mathbb{Z} : Q|a - b_0, p_1, \ldots, p_r|a - b_0 \}
\]
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
\[
r_{m,b,Q}(C, D) = [(d - b_0)/Qm] - [(c - b_0)/Qm] = \frac{1}{Qm}(D - C) + \theta(1) .
\]
Carrying out inclusion/exclusion relative to the primes \( p_1, \ldots, p_r \) we have
\[
\#\{ a \in (C, D) \cap \mathbb{Z} : a \equiv b \pmod{Q}, (a, N) = 1 \} = \sum_{\text{m|} p_1 \cdots p_r} (-1)^{\lambda(m)} r_{m,b,Q}(C, D) = \frac{1}{Q} \prod_{i=1}^r (1 - \frac{1}{p_i}) \cdot (D - C) + \theta(d(p_1 \cdots p_r))
\]
\[
= \frac{\varphi(N)}{N \varphi(Q)} (D - C) + \theta(d(N)) .
\]

\[\square\]

**Proof.** (of Proposition 1.3.) Let \( \zeta_N \) 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 \( Q|N \). We will call \( Q = Q_N \) the cyclotomic conductor of \( \zeta_N \) relative to \( k \), and write \( T_N = [\mathbb{Q}(\zeta_Q) : k_N] \).
As \( \mathbb{Q}(\zeta_N) \) is galois over \( \mathbb{Q} \), it is linearly disjoint from \( k \) over \( k_N \), and \( \text{Gal}(k(\zeta_N)/k) \cong \text{Gal}(\mathbb{Q}(\zeta_Q)/k_N) \). Since \( k_N \subset \mathbb{Q}(\zeta_Q) \subset \mathbb{Q}(\zeta_N) \), the conjugates of \( \zeta_N \) over \( k \) are a union of \( T_N \) sets of the form
\[
\{ e^{2\pi i a/N} : a \equiv b_i \pmod{Q_N}, (a, N) = 1 \} ,
\]
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 \( (C, D) = \frac{N}{2\pi} (\theta_1, \theta_2) \). Then \( e^{2\pi i a/N} \in I \) if and only if \( a \in (C, D) \). By Lemma 1.4,
\[
N(\zeta_N, I) = T_N \cdot \frac{\varphi(N)}{N \varphi(Q_N)} \cdot \frac{N}{2\pi}(\theta_2 - \theta_1) + \theta(T_N \cdot d(N)) .
\]

Recall that for any \( \delta > 0 \), if \( N \) is sufficiently large then \( d(N) \leq N^\delta \) and \( \varphi(N) \geq N^{1-\delta} \) (see Hardy and Wright [HW71], Theorem 315, p.260, and Theorem 327, p.267). Take \( \delta \) such
that $0 < 2\delta < 1 - \gamma$. Noting that $[k(\zeta_N) : k] = T_N \phi(N)/\phi(Q_N)$, and that $\phi(Q_N)$ is bounded independent of $N$, (16) gives

$$N(\zeta_N, I) \quad \mu(I) + O_\gamma(N^{-\gamma}).$$

Since $[k(\zeta_N) : k] \leq N$, the error bound in (17) 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 (10) applies to all intervals. □

2. The finiteness theorem for elliptic curves

2.1. 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

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

with coefficients in $O_k$. More precisely, $E$ is the hypersurface in $\mathbb{P}^2/\text{Spec}(k)$ defined by the homogenization of (18). Let $\Delta$ be its discriminant.

Given a nonarchimedean place $v$ of $k$ and points $\alpha, \beta \in E(\overline{k})$, we will say that $\beta$ is integral with respect to $\alpha$ at $v$ if the Zariski closures $\text{cl}(\beta)$ and $\text{cl}(\alpha)$ do not meet in the model $\mathcal{E}_v/\text{Spec}(O_v)$ defined by the homogenization of (18). Equivalently, if $\|z, w\|_v$ is the restriction of the spherical metric on $\mathbb{P}^2(\overline{k})$ to $E(\overline{k})$ (see [Ru89], §1.1), then for each pair of embeddings $\sigma, \tau : \overline{k}/k \hookrightarrow \overline{k}$,

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

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

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

$$\widehat{h}(\alpha) = \frac{1}{2} \lim_{n \to \infty} \frac{1}{4^n} \mathcal{h}_p(x([2^n]\alpha)) = \frac{1}{3} \lim_{n \to \infty} \frac{1}{4^n} \mathcal{h}_p([2^n]\alpha),$$

where $\mathcal{h}_p$ (resp. $\mathcal{h}_p^{\ast}$) is the naive height on $\mathbb{P}^1(\overline{k})$ (resp. $\mathbb{P}^2(\overline{k})$), $x$ is the coordinate function on the Weierstrass model (18), and $[m]$ is multiplication by $m$ on $E(\overline{k})$. (For a discussion of $\widehat{h}(\alpha)$ and its properties, see [Si86], pp.227-231 and 365-366; or [Si99], §VI.) Recall that $\widehat{h}(\alpha) \geq 0$, that $\widehat{h}([m]\alpha) = m^2\widehat{h}(\alpha)$ for all $m$, and that $\widehat{h}(\alpha) = 0$ if and only if $\alpha \in E(\overline{k})_{\text{tors}}$. From these facts it follows (as is well known) that if $\xi \in E(\overline{k})_{\text{tors}}$, then

$$\widehat{h}(\alpha) = \widehat{h}(\alpha - \xi).$$

There is also a decomposition of $\widehat{h}(\alpha)$ as a sum of local terms. For each place $v$ of $k$, let $\lambda_v(P)$ be the local Néron-Tate height function on $E(\overline{k})$. For compatibility with our absolute values we normalize $\lambda_v(P)$ so that $\lambda_v(P) = [k_v : \mathbb{Q}_p] \cdot \lambda_v,_{\text{Sil}}(P)$, where $\lambda_v,_{\text{Sil}}(P)$ is the local Néron-Tate height defined in Silverman ([Si86], p.365). For each $0 \neq \alpha \in E(k)$

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

(see [Si86], Theorem 18.2, p.365). Note that 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 $E(\overline{k}_w)$. If we fix a $k_v$-isomorphism $\overline{T}_w \cong \overline{k}_v$, then for all $P \in E(\overline{k}_v)$,

$$\lambda_w(P) = \lfloor L_w : k_v \rfloor \lambda_v(P) .$$

(21)

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

$$\sum_{\sigma : L/k \to \overline{k}_v} \lambda_v(\sigma(\beta)) = \sum_{w|v} \lambda_w(\beta) .$$

(22)

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 $k$.

A) If $v$ is archimedean, fix an an isomorphism $E(\overline{k}_v) \cong \mathbb{C}/\Lambda$ for an appropriate lattice $\Lambda \subset \mathbb{C}$. Let $\sigma(z, \Lambda)$ be the Weierstrass $\sigma$-function, let $\Delta(\Lambda) = g_2(\Lambda)^3 - 27g_3(\Lambda)^2$ be the discriminant of $\Lambda$, and let $\eta : \mathbb{C} \to \mathbb{R}$ be the $\mathbb{R}$-linearized period map associated to the Weierstrass $\zeta$-function $\zeta(z, \Lambda)$. If $P \in E(\overline{k}_v)$ corresponds to $z \in \mathbb{C}$, then

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

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

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

B) 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 $E(\overline{k}_v) \cong \overline{k}_v^\times/q^2$ where $q \in \overline{k}_v^\times$ 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 $\tilde{\lambda}_v(x) = \frac{1}{2} B_2(\frac{x}{\text{ord}_v(q)}) \cdot (-\log(|q|_v))$. If $P \in E(\overline{k}_v)$ corresponds to $z \in \overline{k}_v^\times$, with $z$ chosen so that $|q|_v < |z|_v \leq 1$, then

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

Furthermore, if $\mu_v$ is the Haar measure $dx/\text{ord}_v(q)$ giving the loop $\mathbb{R}/(\mathbb{Z} \cdot \text{ord}_v(q))$ total mass 1, then

$$\int_0^{\text{ord}_v(q)} \lambda_v(x) d\mu_v(x) = 0 .$$

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

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

Proof. This is a summary of results in ([Si99], §VI); see in particular Theorem 1.1, p. 455; Theorem 3.2, p.466; Theorem 3.3, p.468; and Theorem 4.1, p.470. □
2.2. 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 $\alpha \in E(\overline{k})$ is nontorsion (has canonical height $\hat{h}(\alpha) > 0$), there are only finitely many torsion points $\xi \in E(\overline{k})_{\text{tors}}$ which are $S$-integral with respect to $\alpha$.

Again there are limitations to possible strengthenings of the theorem:

A) As noted by Silverman, it is necessary that $\alpha$ 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 $\alpha$.

Similarly, if $\alpha$ 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 $\alpha$.

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 ([U95], 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 $\beta \in E(k)$ with $\hat{h}(\beta) < \varepsilon$ which are $S_\infty$-integral with respect to $\alpha$, where $S_\infty$ 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 semi-stable 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 (18) has good reduction for all $v \notin S$.

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

\[
\hat{h}(\alpha) = \frac{1}{[k(\xi_n) : \mathbb{Q}]} \sum_v \sum_{\sigma : k(\xi_n)/k \to \overline{k}, \sigma v = v} \lambda_v(\alpha - \sigma(\xi_n)).
\]

To see this, let $L$ be the galois closure of $k(\xi_n)$ in $\overline{k}$ over $k$. By (19) and (20), for each conjugate $\sigma(\xi_n)$,

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

Averaging over all $k$-embeddings $\sigma : L \hookrightarrow \overline{k}$, fixing a $k$-embedding $\overline{k} \hookrightarrow \overline{k}_v$ for each place $v$ of $K$, using (21), and noting that there are only finitely many nonzero terms in each sum,
we have
\[ \hat{h}(\alpha) = \frac{1}{[L : k]} \sum_{\sigma : L/k \hookrightarrow \mathbb{Q}} \frac{1}{[L : \mathbb{Q}]} \sum_{w \text{ of } L} \lambda_w(\alpha - \sigma(\xi_n)) \]
\[ = \frac{1}{[L : \mathbb{Q}]} \sum_{v \text{ of } k} \sum_{\sigma : L/k \hookrightarrow \mathbb{Q}_v} \frac{1}{[L : k]} \sum_{w|v} [L_w : k_v] \cdot \lambda_v(\alpha - \sigma(\xi_n)) \]
\[ = \frac{1}{[L : \mathbb{Q}]} \sum_{v \text{ of } k} \sum_{\sigma : L/k \hookrightarrow \mathbb{Q}_v} \lambda_v(\alpha - \sigma(\xi_n)) \cdot \]

Since each conjugate \( \sigma(\xi_n) \) occurs \( [L : k(\xi_n)] \) times in the final inner sum, this is equivalent to (23).

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

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

\[ \hat{h}(\alpha) = \sum_{v \in S} \frac{1}{[k(\xi_n) : k]} \sum_{\sigma : k(\xi_n)/k \hookrightarrow \mathbb{Q}_v} \lambda_v(\alpha - \sigma(\xi_n)) \cdot \]

In the following two subsections, we will show that for each \( v \in S \),

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

This will complete the proof of Theorem 0.2 for then, combining (24) and (25) and letting \( n \to \infty \) in (24), we would have \( \hat{h}(\alpha) = 0 \), contradicting the assumption that \( \alpha \) is nontorsion.

2.2.1. **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 \( k \) with \( \mathbb{C} \). Thus, the way \( k \) is embedded depends on the choice of \( v \).

To prove (25) we will need two facts: David/Hirata-Kohno’s theorem on linear forms in elliptic logarithms, and a strong form of equidistribution for torsion points.

The following is a special case of ([DHK02], Theorem 1, p.31):

**Proposition 2.2.** (David/Hirata-Kohno)

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 \( \omega_1, \omega_2 \) be generators for \( \Lambda \). Fix a nontorsion point \( \alpha \in E(k) \) and let \( a \in \mathbb{C} \) be such that \( \theta(a \mod \Lambda) = \alpha \).

Then there is a constant \( C = C(E, \alpha) > 0 \) such that for all rational numbers \( \ell_1/N, \ell_2/N \) with \( \ell_1, \ell_2, N \in \mathbb{Z} \),

\[ |a - (\frac{\ell_1}{N} \omega_1 + \frac{\ell_2}{N} \omega_2)| \geq e^{-C \max(1, \log(N))} \cdot \]

By the Szpiro-Ullmo-Zhang theorem ([SUZ97]), the galois conjugates of the \( \xi_n \) 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 \( \xi \in E(\overline{k})_{\text{tors}} \), write \( \text{Gal}(\overline{k}/k) : \xi \) for the orbit \( \{ \sigma(\xi) : \sigma \in \text{Gal}(\overline{k}/k) \} \). For each set \( U \subset E(\mathbb{C}) \), write

\[
N(\xi, U) = \#((\text{Gal}(\overline{k}/k) : \xi) \cap U).
\]

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

Let \( \Lambda \subset \mathbb{C} \) be a lattice such that \( E(\mathbb{C}) \cong \mathbb{C}/\Lambda \). Let \( r_0 = r_0(\Lambda) > 0 \) be the largest number such that \( S(a, r) \) injects into \( \mathbb{C}/\Lambda \cong E(\mathbb{C}) \) under the natural projection for all \( a \in \mathbb{C} \) and all \( 0 \leq r < r_0 \). Write \( S_E(a, r) \) for the image of \( S(a, r) \) 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 \( \text{Gal}(\overline{k}/k) \)-conjugates of the torsion points in \( E(\overline{k}) \) are strongly equidistributed in \( E(\mathbb{C}) \) in the following sense:

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

\[
\frac{N(\xi, S_E(a, r))}{[k(\xi) : k]} = \mu(S_E(a, r)) + O([k(\xi) : k]^{-\gamma})
\]

where the implied constant depends only on \( k, S, E, \) and \( \gamma \).

The proof will be given in §2.3 below.

We can now complete the proof of (25) in the archimedean case. The argument is similar to the one in the proof of Theorem 0.1. By the Szpiro-Ullmo-Zhang theorem ([SUZ97]), or by Proposition 2.3 when \( S \) has the shape of a period parallelogram (so \( E \) can be tiled with sets \( S_E(a, r) \)), one knows that as \( n \to \infty \) the discrete measures

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

converge weakly to the Haar measure \( \mu \) on \( E(\mathbb{C}) \) having total mass 1. Proving (25) is equivalent to showing that

\[
\lim_{n \to \infty} \int_{E(\mathbb{C})} \lambda_\nu(\alpha - z) d\mu_n(z) = 0.
\]

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 \( \Lambda \). After scaling \( \Lambda \), if necessary, we can assume that \( F = 1 \). After this normalization, \( \mu \) 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 \( \theta(a \mod \Lambda) = \alpha \).

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 \( \alpha \).) Without loss, we can assume that \( \varepsilon < 1/\pi \), so \( \pi \varepsilon^2 < \varepsilon \). We will show that for all large \( n \),

\[
| \int_{E(\mathbb{C})} \lambda_\nu(\alpha - z) d\mu_n(z) | < 6\varepsilon.
\]
Put
\[
\text{labs}_{\alpha, \varepsilon}(z) = \begin{cases} 
\infty & \text{if } z = a, \\
-\lfloor k_v : \mathbb{R} \rfloor \log(|z - a|/\varepsilon) & \text{if } z \in B(a, r) \backslash \{a\}, \\
0 & \text{if } z \in E(\mathbb{C}) \backslash B(a, r).
\end{cases}
\]
and note that
\[
0 < \int_{E(\mathbb{C})} \text{labs}_{\alpha, \varepsilon}(z) \, d\mu(z) = \int_{B(a, \varepsilon)} -\lfloor k_v : \mathbb{R} \rfloor \log(|z - a|/\varepsilon) \, d\mu(z)
\]
\[
= \lfloor k_v : \mathbb{R} \rfloor \int_0^\varepsilon -2\pi t \log(t/\varepsilon) \, dt = \lfloor k_v : \mathbb{R} \rfloor \frac{\pi\varepsilon^2}{2} < \varepsilon.
\]
By Proposition 2.1.A there is a continuous function \(g_{\alpha, \varepsilon}(z)\) on \(E(\mathbb{C})\) such that
\[
\lambda_v(\alpha - z) = \text{labs}_{\alpha, \varepsilon}(z) + g_{\alpha, \varepsilon}(z).
\]
Since \(\int_{E(\mathbb{C})} \lambda_v(\alpha - z) \, d\mu(z) = 0\) (also by Proposition 2.1.A), we get
\[
\left| \int_{E(\mathbb{C})} g_{\alpha, \varepsilon}(z) \, d\mu(z) \right| = \left| \int_{B(a, \varepsilon)} -\text{labs}_{\alpha, \varepsilon}(z) \, d\mu(z) \right| < \varepsilon.
\]
By weak convergence, it follows that for all sufficiently large \(n\),
\[
\left| \int_{E(\mathbb{C})} g_{\alpha, \varepsilon}(z) \, d\mu_n(z) \right| < 2\varepsilon.
\]
To complete the proof of (26), it will suffice to show that for all sufficiently large \(n\),
\[
\left| \int_{B(a, r)} \log(|z - a|/\varepsilon) \, d\mu_n(z) \right| < 2\varepsilon.
\]
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 \(\mu(A_0(n)) = \frac{\pi\varepsilon^2}{D^2} \leq \pi\varepsilon^2/[k(\xi_n) : k]^{1/4}\). Applying Proposition 2.3 when \(S\) is a disc, taking \(\gamma = 3/8\), gives
\[
N(\xi_n, A_1(n))/[k(\xi_n) : k] \leq 2\mu(A_0(n))
\]
for all sufficiently large \(n\). If \(\xi_n\) 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 \(\theta(B(a, \varepsilon)) \subset E(\mathbb{C})\))
\[
|\log(|\sigma(\xi_n) - a|)| \leq C \log(N_n).
\]
Using (44) and (45) below, one sees that \([k(\xi_n) : k] \geq N_n^{1/2}\) for all sufficiently large \(n\). Thus
\[
0 \leq \left| \int_{A_0(n)} \log(|z - a|) \, d\mu_n(z) \right| \leq 4\pi\varepsilon^2 C \cdot \log([k(\xi_n) : k]) [k(\xi_n) : k]^{1/4} < \varepsilon
\]
for all sufficiently large \(n\).

For each annulus \(A_\ell(n), \ell = 1, \ldots, D - 1\), one has
\[
\mu(A_\ell(n)) = \pi(2\ell + 1)\varepsilon^2/D^2 \simeq \pi(2\ell + 1)\varepsilon^2/[k(\xi_n) : k]^{1/4}.
\]
Since \(A_\ell(n)\) is the difference of two sets to which Proposition 2.3 applies, we find as above that for sufficiently large \(n\),
\[
N(\xi_n, A_\ell(n))/[k(\xi_n) : k] \leq 2\mu(A_\ell(n)).
\]
Note that on $A_{\ell}(n)$, $|\log(|z - a|/\varepsilon)| \leq -\log(\ell/D)$. Summing over these annuli, and bounding the resulting Riemann sum by an integral, we find that

$$\left| \int_{B(a,\varepsilon) \setminus A_0(n)} \log(|z - a|/\varepsilon) \, d\mu_n(z) \right| \leq \sum_{\ell=1}^{D-1} -\log((\ell\varepsilon)/D) \cdot 2\mu(A_\ell(n))$$

$$< 2 \cdot \int_{B(a,\varepsilon)} -2\pi t \log(t/\varepsilon) \, dt$$

$$= \pi\varepsilon^2 < \varepsilon.$$  

Combining (29) and (30) gives (28), which completes the proof of (25) in the archimedean case (assuming Proposition 2.3).

2.2.2. The Nonarchimedean Case: In the nonarchimedean case, the proof of (25) depends on a well-known result of Cassels on the denominators of torsion points (see [Si86], Theorem 3.4, p.177). Write $\mathcal{O}_v$ for the ring of integers of $\mathbb{K}_v$.

**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 $\mathcal{O}_v$ (N. B. the Weierstrass equation need not be minimal.) Let $P \in E(\mathbb{K}_v)_{\text{tors}}$ be a point of exact order $m \geq 2$.

(A) If $m$ is not a power of $p$, then $x(P), y(P) \in \mathcal{O}_v$.

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

$$\text{ord}_v(D) \leq \frac{\text{ord}_v(p)}{p^n - p^{n-1}}.$$  

**Proof.** Silverman ([Si86], Theorem 3.4) states the theorem for torsion points belonging to $E(k_v)$, with $a, b, D \in k_v$ in part B) and $D$ satisfying

$$\text{ord}_v(D) = \left\lfloor \frac{\text{ord}_v(p)}{p^n - p^{n-1}} \right\rfloor.$$  

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 $a, b, D \in L_w$, (31) becomes

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

This yields the result for all $P \in E(\mathbb{K}_v)_{\text{tors}}$.  \hfill \Box

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

(A) There is a number $M$ such that for all $\xi \in E(\mathbb{K}_v)_{\text{tors}}$,

$$\lambda_v(\alpha - \xi) \leq M.$$  

(B) If $E$ has good reduction, then for each $\varepsilon > 0$, there are only finitely many $\xi \in E(\mathbb{K}_v)_{\text{tors}}$ with $\lambda_v(\alpha - \xi) > \varepsilon$. If $E$ is a Tate curve, then for each $\varepsilon > 0$, there are only finitely many $\xi \in E(\mathbb{K}_v)_{\text{tors}}$ 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 (B) implies (A), it suffices to prove (B). 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 $E/k$. If $\xi_1, \xi_2 \in E(k_v)_{\text{tors}}$ 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 $O_v$. By the ultrametric inequality for the spherical distance ([Ru89], §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,

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

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

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

Next suppose $E$ is a Tate curve. Fix a Tate isomorphism $E(k_v) \cong \mathbb{F}_q^\times / q^2$ 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 $a, u_1, u_2 \in \mathbb{F}_q^\times$ correspond to $\alpha, \xi_1, \xi_2$ respectively; we can assume that $|q|_v < |a|_v, |u_1|_v, |u_2|_v \leq 1$. By the formula for $\lambda_v(x - y)$ in Proposition 2.1.B, if $\lambda_v(\alpha - \xi_i) > \varepsilon + \frac{1}{12}(-\log(|\Delta(E)|_v))$, then $|a|_v = |u_1|_v = |u_2|_v$ and

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

Put $\xi = \xi_1 - \xi_2$ and $u = u_2^{-1}u_1$. Then $\xi$ corresponds to $u$ under the Tate isomorphism, and $\text{ord}_v(1 - u) > \varepsilon / \log(Nv)$. By the formulas for $x(\xi), y(\xi)$ in ([Si99], p. 425), $\text{ord}_v(x(\xi)) = 2\text{ord}_v(1 - u)$ and $\text{ord}_v(y(\xi)) = 3\text{ord}_v(1 - u)$. Again by Cassels’ theorem, only finitely many torsion points $\xi$ can satisfy $\min(\text{ord}_v(x(\xi)), \text{ord}_v(y(\xi))) > \varepsilon / \log(Nv)$. 

We can now prove (25) when $E$ has good reduction at $v$.

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

$$0 \leq \frac{1}{[k(\xi_n) : k]} \sum_{\sigma \in \mathbb{F}/k \cong \mathbb{F}_v} \lambda_v(\alpha - \sigma(\xi_n)) \leq \frac{([k(\xi_n) : k] - N)}{[k(\xi_n) : k]} \cdot \varepsilon + \frac{N}{[k(\xi_n) : k]} \cdot M < 2\varepsilon.$$

Thus

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

To prove (25) when $E$ is a Tate curve at $v$, we will need the following equidistribution theorem of Chambert-Loir ([CL07], Corollaire 5.5).

Fix a Tate isomorphism $E(k_v) \cong \mathbb{F}_q^\times / q^2$, put $L = \mathbb{Z} \cdot \text{ord}_v(q) \subset \mathbb{R}$, and define a “reduction map” $r : E(k) \to \mathbb{R}/L$ by setting $r(P) = \text{ord}_v(a) \pmod{L}$ if $P \in E(k_v)$ corresponds to $a \in \mathbb{F}_v$. 


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

$$\mu_{P,v}(z) = \frac{1}{[k(P) : k]} \sum_{\sigma : k/k \rightarrow k_v} \delta_{\sigma(P)}(z)$$

and let $\mu_v$ 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 $E(\overline{k})$ with $\hat{h}(P_n) \to 0$, the sequence of measures $\{\mu_{P_n,v}\}$ converges weakly to $\mu_v$.

We can now prove (25) when $E$ is a Tate curve. Recall that $\{\xi_n\}$ is a sequence of distinct torsion points which are $S$-integral with respect to $\alpha$.

Fix $\varepsilon > 0$. Let $M$ be the upper bound in Corollary 2.5.A. 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.B, $\int_{\mathbb{R}/L} \hat{\lambda}_v(z) d\mu_v(z) = 0$ and

$$\left| \frac{1}{[k(\xi_n) : k]} \sum_{\sigma : k/k \rightarrow k_v} \lambda_v(\sigma(\xi_n) - \alpha) \right| \leq \left| \int_{\mathbb{R}/L} \hat{\lambda}_v(z - a) d\mu_{\xi_n,v}(z) \right| + M \cdot \mu_{\xi_n,v}((a-\delta, a+\delta)) .$$

For all sufficiently large $n$ the right side is at most $3\varepsilon$. Hence

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

This completes the proof of Theorem 0.2. □

Several results in the literature use methods related to ours:

J. Cheon and S. Hahn [CH99] proved an elliptic curve analogue of Schinzel’s theorem [Sc74]. Likewise, Everest and B. Ní Flathúin [EF96] evaluate ‘elliptic Mahler measures’ in terms of limits involving division polynomials, obtaining results similar to (15). 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 [STpp] 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.

2.3. 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 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 \( \text{Gal}(\overline{k}/k) \) on \( E(\overline{k})_{\text{tors}} \) induces an injective homomorphism

\[
\eta : \text{Gal}(\overline{k}/k) \rightarrow \lim_{\longrightarrow} \text{GL}_2(\mathbb{Z}/N\mathbb{Z}) \cong \prod_p \text{GL}_2(\mathbb{Z}_p).
\]

By Serre’s theorem ([Se72], Théorème 3), the image of \( \text{Gal}(\overline{k}/k) \) in \( \prod_p \text{GL}_2(\mathbb{Z}_p) \) is open. Hence there is a number \( Q \) such that \( \text{Im}(\eta) \) contains the subgroup

\[
\prod_{p \mid Q} (1 + QM_2(\mathbb{Z}_p)) \times \prod_{p \not\mid Q} \text{GL}_2(\mathbb{Z}_p).
\]

Let \( G_Q \subset \text{Gal}(\overline{k}/k) \) be the preimage of this subgroup.

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

\[
\text{Gal}(\overline{k}/k) \cdot \xi = \bigcup_{i=1}^T G_Q \cdot \sigma_i(\xi).
\]

Since \( G_Q \) is normal in \( \text{Gal}(\overline{k}/k) \), 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 \( \xi \), one sees that

\[
\#(G_Q \cdot \xi) = \prod_{p|Q_N} p^{2(\text{ord}_p(N) - \text{ord}_p(Q_N))} \prod_{p \nmid Q_N, p^2 | Q} p^{2\text{ord}_p(N)}(1 - \frac{1}{p^2})
\]

\[= \frac{N^2}{Q_N^2} \prod_{p|Q_N} (1 - \frac{1}{p^2}).\]

Indeed, let \( \xi_p \) be the \( p \)-component of \( \xi \) in \( E[N] \cong \prod_{p|N}(\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z})^2 \). Identify \( \xi_p \) with an element of \( (\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z})^2 \), and note that it is a generator for that group. If \( p|Q_N \), the image of \( G_Q \) in \( \text{GL}_2(\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z}) \) is \( I + p^{\text{ord}_p(Q_N)}M_2(\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z}) \), and

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

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

\[
G_Q \cdot \xi_p = (\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z})^2p \cdot (\mathbb{Z}/p^{\text{ord}_p(N)}\mathbb{Z})^2.
\]

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

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

\[
\#((x_0 + t\Lambda) \cap \mathcal{S}(a, r)) \leq \frac{\text{area}(\mathcal{S}(a, r + tC))}{\text{area}(t\mathcal{F})} = \frac{r^2 S}{F} \cdot \frac{1}{t^2} + \frac{2CSr}{F} \cdot \frac{1}{t} + \frac{C^2 S}{F}.
\]

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

\[
\#((x_0 + t\Lambda) \cap \mathcal{S}(a, r)) \geq \frac{\text{area}(\mathcal{S}(a, r - tC))}{\text{area}(t\mathcal{F})} > \frac{r^2 S}{F} \cdot \frac{1}{t^2} - \frac{2CSr}{F} \cdot \frac{1}{t} - \frac{C^2 S}{F}.
\]

If \(r \leq tC\) then the right side of (35) is negative, so the inequality between the first and last quantities holds trivially.

Now let \(D\) be a positive divisor of \(N/QN\). Taking \(t = QND/N\), and combining (34), (35), we obtain

\[
\left| \#((x_0 + \frac{QND}{N}\Lambda_N) \cap \mathcal{S}(a, r)) - \frac{\text{area}(\mathcal{S}(a, r))}{\text{area}(t\mathcal{F})} \cdot \frac{N^2}{QND^2} \right| \leq \frac{2CSr}{F} \cdot \frac{N}{QND} + \frac{C^2 S}{F}.
\]

Step 3: Inclusion/Exclusion. Write \(\Lambda_N = \frac{1}{N}\Lambda\), fix \(\sigma_i\), 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 \(y - x \in QN\Lambda_N\) and \(y + \Lambda\) has exact order \(N\) in \(\Lambda_N/\Lambda\). Equivalently, \(y - x \in QN\Lambda_N\) and \(y \notin p\Lambda\) 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 \(QN\) and \(p_1, \ldots, p_R\) are pairwise coprime, there is an \(x_0 \in \Lambda_N\) such that \(x_0 \equiv x \pmod{QN\Lambda_N}\) and \(x_0 \equiv 0 \pmod{p_1 \cdots p_R\Lambda_N}\). Then \(y - x_0 \in QN\Lambda_N\) if and only if \(y \equiv x_0 + QN\Lambda_N\), and \(y \in p_i\Lambda_N\) if and only if \(y \equiv x_0 + p_i\Lambda_N\). Note that if \(D|p_1 \cdots p_R\) then \(QN\Lambda_N \cap D\Lambda_N = QND\Lambda_N\). Recalling that \(r_0\) is the supremum over positive numbers \(r\) for which \(\mathcal{S}(a, r)\) injects into \(\mathbb{C}/\Lambda\), take \(a \in \mathbb{C}\) and take \(0 < r \leq r_0\). Applying inclusion-exclusion, we obtain

\[
\#(G_Q \cdot \sigma_i(\xi) \cap \mathcal{S}_E(a, r)) = \sum_{D|p_1 \cdots p_R} (-1)^{\lambda(D)} \cdot \#((x_0 + QND\Lambda_N) \cap \mathcal{S}(a, r)) ,
\]

where \(\lambda(D)\) is the number of distinct primes dividing \(D\).
Inserting (36) in (37) and summing over all \( \sigma_i(\xi) \), \( i = 1, \ldots, T \), we find
\[
N(\xi, \mathcal{S}_E(a, r)) = \frac{\text{area}(\mathcal{S}(a, r))}{\text{area}(\mathcal{F})} \cdot TN^2 \prod_{p \mid N, p \mid Q} \left(1 - \frac{1}{p^2}\right)
+ \theta\left(\frac{2CSr}{F} \cdot \frac{TN}{Q_N} \prod_{p \mid N, p \mid Q} (1 + \frac{1}{p})\right) + \theta\left(\frac{C^2 S}{F} \cdot T2^R\right)
\]
where as before, \( \theta(x) \) denotes a quantity with \(-x \leq \theta(x) \leq x\). By (33),
\[
[k(\xi) : k] = T \cdot \#(G_Q \cdot \xi) = \frac{T N^2}{Q_N^2} \prod_{p \mid N, p \mid Q} \left(1 - 1/p^2\right).
\]
Since \( r \leq r_0 \), it follows that
\[
\frac{N(\xi, \mathcal{S}_E(a, r))}{[k(\xi) : k]} = \frac{\text{area}(\mathcal{S}(a, r))}{\text{area}(\mathcal{F})} + \theta\left(\frac{2CSr_0}{F} \cdot \frac{N}{\prod_{p \mid N, p \mid Q} (1 - 1/p^2)}\right)
+ \theta\left(\frac{C^2 S}{F} \cdot \frac{2^R Q_N^2}{N^2 \prod_{p \mid N, p \mid Q} (1 - 1/p^2)}\right).
\]

Here \( \text{area}(\mathcal{S}(a, r))/\text{area}(\mathcal{F}) = \mu(\mathcal{S}_E(a, r)) \). Note that \( T \leq \#(\text{GL}_2(\mathbb{Z}/(Q\mathbb{Z})) \) is bounded, \( Q_N \leq Q \) is bounded, and \( N \prod_{p \mid N} (1 - 1/\gamma) \geq N^{1-\varepsilon} \) for each \( \varepsilon > 0 \) and each sufficiently large \( N \). Using (38) and the fact that \( 1 \geq \prod_{p \mid N, p \mid Q} (1 - 1/p^2) \geq 1/\zeta(2) \) one sees that the first error term is \( O_s([k(\xi) : k]^{-\gamma}) \) for each \( \gamma < 1/2 \). Similarly, \( 2^R \leq d(N) \leq N^a \) 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 \( \mathcal{O} \subset \mathcal{O}_K \) 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 [Sh71], [L73], or [Si99], Chapter II), \( E(\overline{\mathbb{F}})_{\text{tors}} \) is rational over \( k^{ab} \), the maximal abelian extension of \( k \). Let \( k^{x}_\Lambda \) be the id`le ring of \( k \), and for \( s \in k^{x}_\Lambda \) let \([s, k]\) be the Artin map acting on \( k^{ab} \). Given \( \sigma \in \text{Gal}(\overline{\mathbb{F}}/k) \), take \( s \in k^{x}_\Lambda \) with \( \sigma|_{k^{ab}} = [s, k] \), and put \( w = N_k/K(s) \in k^{x}_\Lambda \). There is an action of \( k^{x}_\Lambda \) on lattices, defined semi-locally, which associates to \( w \) and \( \Lambda \) a new lattice \( w^{-1}\Lambda \). This action extends to a map \( w^{-1} : K/\Lambda \rightarrow K/w^{-1}\Lambda \). There is also a homomorphism \( \psi : k^{x}_\Lambda \rightarrow k^{x} \), the ‘grössencharacter’ of \( E \), which has the property that \( \psi(s)N_k/K(s)^{-1}\Lambda = \Lambda \). Put \( \kappa = \psi(s) \in k^{x} \).

With this notation, there is a commutative diagram:
\[
\begin{array}{ccc}
w^{-1} & K/\Lambda & \rightarrow & \mathbb{C}/\Lambda & \rightarrow & E(\overline{\mathbb{F}})_{\text{tors}} \\
\downarrow & \downarrow & \vartheta & \downarrow & \sigma & \downarrow \\
K/w^{-1}\Lambda & \rightarrow & \mathbb{C}/w^{-1}\Lambda & \rightarrow & E(\overline{\mathbb{F}})_{\text{tors}} & \rightarrow \\
\kappa & \downarrow & id & \downarrow & id & \\
K/\Lambda & \rightarrow & \mathbb{C}/\Lambda & \rightarrow & E(\overline{\mathbb{F}})_{\text{tors}} & \rightarrow \\
\end{array}
\]
in which the vertical arrows on the left are multiplication by \( w^{-1} \) and \( \kappa \) respectively, and those on the right are the galois action (see [Sh71], Proposition 7.40, p.211, or [L73], Theorem...
Note that the same analytic isomorphism $\vartheta$ appears in the top and bottom rows. Thus, if $\xi \in E(\overline{k})_{\text{tors}}$ corresponds to $x \in K/\Lambda$, and $\sigma|_{x^{ab}} = [s, k]$, 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 $K^\times$ in the diagram is as follows. Let $L \subset K$ be a lattice. For each rational prime $p$ of $\mathbb{Q}$, write $L_p = L \otimes \mathbb{Z}_p$ and $K_p = K \otimes \mathbb{Q}_p$; if $w \in K^\times$, 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 $p$ ([L73], 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 $K \subset K_\Lambda$ 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 $\mathcal{O}$ has the form $\mathcal{O} = \mathbb{Z} + c\mathcal{O}_K$ for some integer $c \geq 1$, and $c$ is called the conductor of $\mathcal{O}$. The lattice $\Lambda$ is a proper $\mathcal{O}$-lattice, meaning that $\mathcal{O} = \{x \in K : x\Lambda \subset \Lambda \}$. For any order $\mathcal{O}$, there are only finitely many homothety classes of proper $\mathcal{O}$-lattices ([L73], Theorem 7, p.95). Write $\mathcal{O}_p = \mathcal{O} \otimes \mathbb{Z}_p$ and $\mathcal{O}_{K,p} = \mathcal{O}_K \otimes \mathbb{Z}_p$. If $p \mid \mathfrak{c}$, then $\mathcal{O}_p = \mathcal{O}_{K,p} \cong \prod_{p \mid \mathfrak{c}} \mathcal{O}_{K,p}$, where $p$ runs over the primes of $K$ lying over $p$, and $\mathcal{O}_{K,p}$ is the completion of $\mathcal{O}_K$ at $p$.

The kernel of the grössencharacter $\psi : K^\times \rightarrow K^\times$ is open in $k_\Lambda^\times$, so its image $W = N_{k/K}(U) \subset K_\Lambda^\times$ is open. Thus there is an integer $Q \geq 1$ such that for each $p \mid Q$, the subgroup $1 + Q \mathcal{O}_{K,p} \subset \mathcal{O}_{K,p}$ is contained in $W_p$, and for each $p \nmid Q$, $\mathcal{O}_{K,p} \subset W_p$. 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 $W_Q \subset K^\times_\Lambda$ be the subgroup

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

and let $U_Q$ be its preimage in $K^\times_\Lambda$ under the norm map. Put

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

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

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

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

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 $\xi$ correspond to $x + \Lambda \in K/\Lambda$. Write $\Lambda(x)$ for the $\mathcal{O}$-lattice $\mathcal{O}x + \Lambda$; since $\xi$ 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

$$\Lambda(mx)/\Lambda \cong \prod_{p|N} \Lambda(mx)_p/\Lambda_p = \prod_{p|N} (m\mathcal{O}_p x + \Lambda_p)/\Lambda_p.$$
If \( p \mid Q \), then \( G_Q \) acts on \( \xi_p \) through the subgroup \( 1 + p^{\text{ord}_p(Q)}O_p \subset \mathcal{O}_p^\times \). Noting that \( \text{ord}_p(Q_N) = \min(\text{ord}_p(Q), \text{ord}_p(N)) \) and that \( p^{\text{ord}_p(Q)}x \in \Lambda_p \) if \( \text{ord}_p(Q) \geq \text{ord}_p(N) \), we have

\[
G_Q \times \xi_p \cong (x + p^{\text{ord}_p(Q)}x + \Lambda_p)/\Lambda_p \quad (x + \Lambda(p^{\text{ord}_p(Q)}x)/\Lambda_p).
\]

Thus \( \#(G_Q \times \xi_p) = [\Lambda(p^{\text{ord}_p(Q)}x): \Lambda_p] \).

If \( p \mid Q \), then \( O_p = \mathcal{O}_{K,p} \) and \( G_Q \) acts on \( \xi_p \) through \( \mathcal{O}_p^\times \cong \prod_{p \mid p} \mathcal{O}_{K,p}^\times \). For each \( p \mid p \), and each \( \mathcal{O} \)-lattice \( L \), we have \( L_p = (\mathcal{O}_{K,L})_p \) where \( \mathcal{O}_{K,L} \) is an \( \mathcal{O}_K \)-fractional ideal. Thus \( \text{ord}_p(L) := \text{ord}_p(\mathcal{O}_{K,L}) \) is well defined. Write \( \text{ord}_p(\xi) = \text{ord}_p(\Lambda) - \text{ord}_p(\Lambda(x)) \). Then \( \Lambda(x)/\Lambda_p \cong \prod_{p \mid p} \mathcal{O}_K/p^{\text{ord}_p(\xi)} \) and

\[
\#(G_Q \times \xi_p) = [\Lambda(x): \Lambda_p] \cdot \prod_{p \mid p \mid \text{ord}_p(\xi) > 0} (1 - \frac{1}{N_p}),
\]

where \( N_p = \#(\mathcal{O}_K/p) \) is the norm of \( p \).

Combining these formulas, and using that \( \prod_{p \mid [\mathcal{O}_K : \mathcal{O}_p]} [\Lambda(p^{\text{ord}_p(Q)}x) : \Lambda_p] = [\Lambda(Q_N x) : \Lambda] \), we obtain

\[
(39) \quad \#(G_Q \times \xi) = [\Lambda(Q_N x) : \Lambda] \cdot \prod_{p \mid [\mathcal{O}_K : \mathcal{O}_p], \text{ord}_p(\xi) > 0} (1 - \frac{1}{N_p}).
\]

Step 2: Counting translated lattice points in convex domains. If \( L \) is any \( \mathcal{O} \)-lattice, and \( F(L) \) is the area of a fundamental domain for \( \mathcal{O}/L \), 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 \( \mathcal{O} \)-lattice for some order \( \mathcal{O}' \) with conductor \( c'|c \). There are only finitely many such orders \( \mathcal{O}' \), and for each \( \mathcal{O}' \) there are only finitely many homothety classes of proper \( \mathcal{O}' \)-lattices, so there are only finitely many homothety classes of \( \mathcal{O} \)-lattices. Hence there is a constant \( C_1 \), independent of \( L \), such that \( L \) has a fundamental domain \( F(L) \) contained in the ball \( B(0, C_1 \cdot F(L)^{1/2}) \). In turn, there is a constant \( C_2 \), independent of \( L \), such that \( F(L) \subset S(0, C \cdot F(L)^{1/2}) \). This fact is the crux of the argument in the CM case.

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

We will apply this taking \( L = \Lambda(Q_N x) = Q_N \mathcal{O} x + \Lambda \). Note that the fundamental domain \( F(\varpi \Lambda(Q_N x)) \) has area \( F \cdot N \varpi / [\Lambda(Q_N x) : \Lambda] \), where \( F \) is the area of a fundamental domain \( F \) for \( \Lambda \). By the same argument leading to (36) we find that for each \( x_0 \in \mathbb{C} \)

\[
\#((x_0 + \varpi \Lambda(Q_N x)) \cap S(a, r)) - \frac{\text{area}(S(a, r))}{\text{area}(F)} \cdot \frac{[\Lambda(Q_N x) : \Lambda]}{N \varpi} 
\]
\[
\leq \frac{2CSr}{F} \cdot \left( \frac{[\Lambda(Q_N x) : \Lambda]}{N \varpi} \right)^{1/2} + C^2 S \frac{F}{F}.
\]

Step 3: Inclusion/Exclusion. Now consider a set \( S(a, r) \), where \( a \in \mathbb{C} \) and \( r \leq r_0 \). For each \( \sigma_i(\xi) \), we will compute \( \#((G_Q \cdot \sigma_i(\xi)) \cap S_E(a, r)) \). Fix \( \sigma_i \), and replace \( \xi \) by \( \sigma_i(\xi) \) in the discussion above. Let \( x \in K/\Lambda \) correspond to \( \sigma_i(\xi) \), and let \( p_1, \ldots, p_R \) be the distinct primes of \( \mathcal{O}_K \) dividing \( N \) but not \( Q \), for which \( \text{ord}_p(\Lambda(x)) \neq \text{ord}_p(\Lambda) \). If there are no such primes, take \( p_1, \ldots, p_R = 1 \) in the argument below. (Note that the \( p_j \) are independent of \( \sigma_i \), since \( K \subseteq k \) and for \( p \mid Q \), \( \sigma_i \) acts on \( \xi \) 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 $y \in x + \Lambda(Q_Nx)$ and $y \notin p_j\Lambda(x)$ for $j = 1, \ldots, R$. Since $\Lambda(Q_Nx) \subset \Lambda(x)$, such $y$ necessarily belong to $\Lambda(x)$. The index $[\Lambda(Q_Nx) : \Lambda]$ in (40) is independent of $\sigma_i$ by (39), since $\#(G_Q \cdot \sigma_i(\xi))$ and the $p_j$ are independent of $\sigma_i$.

The lattices $\Lambda(Q_Nx)$ and $p_1 \cdots p_R\Lambda(x)$ have coprime indices in $\Lambda(x)$, so there is an $x_0 \in \Lambda(x)$ such that $x_0 \equiv x \pmod{\Lambda(Q_Nx)}$ and $x_0 \equiv 0 \pmod{p_1 \cdots p_R\Lambda(x)}$. Further, for any $\mathcal{O}_K$-ideal $\varpi$ dividing $p_1 \cdots p_R$,

$$\Lambda(Q_Nx) \cap \left( \bigcap_{p_j \mid \varpi} \Lambda(x) \right) = \varpi\Lambda(Q_Nx).$$

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

$$\#((G_Q \cdot \sigma_i(\xi)) \cap \mathcal{S}_E(a,r)) = \sum_{\varpi|p_1 \cdots p_R} (-1)^{\lambda_K(\varpi)} \cdot \#((x_0 + \varpi\Lambda(Q_Nx)) \cap \mathcal{S}(a,r))$$

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

Inserting (40) in the inclusion-exclusion formula (41) and summing over all $\sigma_i(\xi)$,

$$N(\xi, \mathcal{S}_E(a,r)) \cdot \frac{\text{area}(\mathcal{S}(a,r))}{\text{area}(\mathcal{F})} \cdot T[\Lambda(Q_Nx) : \Lambda] \prod_{j=1}^R \left( 1 - \frac{1}{Np_j} \right)$$

$$+ \theta\left( \frac{2CSr}{F} \cdot T[\Lambda(Q_Nx) : \Lambda]^{1/2} \prod_{j=1}^R \left( 1 + \frac{1}{Np_j^{1/2}} \right) \right)$$

$$+ \theta\left( \frac{C^2S}{F} \cdot T^{2R} \right).$$

By (39), $[k(\xi) : k] = T[\Lambda(Q_Nx) : \Lambda] \prod_{j=1}^R (1 - \frac{1}{Np_j})$. Since $r \leq r_0$ and $\prod_{j=1}^R (1 + \frac{1}{Np_j^{1/2}}) \leq 2^R$,

$$\frac{N(\xi, \mathcal{S}_E(a,r))}{[k(\xi) : k]} = \frac{\text{area}(\mathcal{S}(a,r))}{\text{area}(\mathcal{F})}

+ \theta\left( \frac{2CSr_0}{F} \cdot \frac{T^{1/2}2^R}{\prod_{j=1}^R (1 - \frac{1}{Np_j})^{1/2}} \cdot \frac{1}{[k(\xi) : k]^{1/2}} \right)

+ \theta\left( \frac{C^2S}{F} \cdot \frac{T^{2R}}{[k(\xi) : k]} \right).$$

As before $\text{area}(\mathcal{S}(a,r))/\text{area}(\mathcal{F}) = \mu(\mathcal{S}_E(a,r))$. Here $T \leq |\text{Gal}({\overline{k}}/k) : G_Q|$ is fixed. For each $\varepsilon > 0$ and each sufficiently large $N$, $2^R \leq 2^{\lambda_K(N)} \leq 2^{2\lambda(N)} \leq d(N)^2 \leq N^\varepsilon$. Likewise, $\prod_{j=1}^R (1 - \frac{1}{Np_j}) \geq \prod_{p \mid N} (1 - 1/p)^2 \geq C/(\log \log(N))^2$ for some constant $C > 0$, where the last inequality follows from ([HW71], Theorem 328, p.267). Finally, since $\xi$ has order $N$ and $Q_N \leq Q$ is bounded, $[\Lambda(Q_Nx) : \Lambda] \geq N/Q$, and so

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

for all large $N$. Combining these shows that for each $0 < \gamma < 1/2$, the first error term is $O_\gamma([k(\xi) : k]^{-\gamma})$. The same estimates show the second error term is negligible in comparison to the first. This completes the proof when $E$ has complex multiplication. \qed
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 \(\xi\), as required by (29). When \(E\) does not have complex multiplication, then since \(T\) is fixed, \(Q_N \leq Q\), and \(\prod_p (1 - 1/p^2)\) converges to a nonzero limit, (38) shows there is a constant \(C_1\) depending only on \(E\) such that

\[(44) \quad [k(\xi) : k] \geq C_1 N^2.\]

When \(E\) has complex multiplication, then since \(T\) and \(Q\) are fixed, (43) shows that there is a constant \(C_2\) depending only on \(E\) such that

\[(45) \quad [k(\xi) : k] \geq C_2 N/(\log \log(N))^2.\]

3. Context.

Theorems 0.1 and 0.2 are the first known cases of general conjectures by the second author (refined by 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 \(O_{k,S}\) be the ring of \(S\)-integers of \(k\).

Conjecture 3.1. (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 \(R_\ast : \mathbb{P}^1 \to \mathbb{P}^1\). Let \(\alpha \in \mathbb{P}^1(\overline{k})\) be nonpreperiodic for \(R_\ast\). Then there are only finitely many preperiodic points \(\xi \in \mathbb{P}^1(\overline{k})\) which are \(S\)-integral with respect to \(\alpha\), i.e. whose Zariski closures in \(\mathbb{P}^1/\text{Spec}(O_{k,S})\) do not meet the Zariski closure of \(\alpha\).

Conjecture 3.2. (Ih) Let \(A/k\) be an abelian variety, and let \(A_S/\text{Spec}(O_{k,S})\) be a model of \(A\). Let \(D\) be a nonzero effective divisor on \(A\), defined over \(k\), at least one of whose irreducible components is not the translate of an abelian subvariety by a torsion point, and let \(\text{cl}(D)\) be its Zariski closure in \(A_S\). Then the set \(A_{D,S}(\overline{\mathbb{Z}})_{\text{tors}}\), consisting of all torsion points of \(A(\overline{k})\) whose closure in \(A_S\) is disjoint from \(\text{cl}(D)\), 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, \(\infty\), and the roots of unity. It is possible to prove Conjecture 3.1 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:

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

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

Part of the motivation for Conjecture 3.2 is the following analogy between diophantine theorems over \(k\) and \(\overline{k}\), and over \(O_k\) and \(\mathbb{Z}\) (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 \(A(\overline{k})_{\text{tors}} \cap X\) 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 $A_D(\mathcal{O}_k)$ of $\mathcal{O}_k$-integral points of $A$ not meeting $\text{supp}(D)$ is finite. Note that $A$ is compact, whereas $A_D = A \setminus \text{supp}(D)$ is noncompact.

<table>
<thead>
<tr>
<th>Type of variety</th>
<th>Type of rationality</th>
<th>$k$</th>
<th>$\mathcal{k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compact</td>
<td>$k, k$-rationality</td>
<td>Mordell-Lang Conjecture</td>
<td>Manin-Mumford Conjecture</td>
</tr>
<tr>
<td>Noncompact</td>
<td>$\mathcal{O}_k, \mathbb{Z}$-rationality</td>
<td>Lang’s Conjecture</td>
<td>Ih’s Conjecture 3.2</td>
</tr>
</tbody>
</table>

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.

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 ([FRL06], Théorème 6).

**References**


MATTHEW BAKER, SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GEORGIA 30332-0160, USA
E-mail address: mbaker@math.gatech.edu

SU-ION IH, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO AT BOULDER, CAMPUS BOX 395, BOULDER, CO 80309-0395, USA
E-mail address: ih@math.colorado.edu

ROBERT RUMELY, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GEORGIA, ATHENS, GEORGIA 30602-0002, USA
E-mail address: rr@math.uga.edu