Algebra & Number Theory

Volume 18
2024
No. 11

Galois orbits of torsion points near atoral sets

Vesselin Dimitrov and Philipp Habegger

msp

ALGEBRA AND NUMBER THEORY 18:11 (2024)
https://doi.org/10.2140/ant.2024.18.1945

msp

Galois orbits of torsion points near atoral sets

Vesselin Dimitrov and Philipp Habegger

We prove that the Galois equidistribution of torsion points of the algebraic torus $\mathbb{G}_m^d$ extends to the singular test functions of the form $\log|P|$, where P is a Laurent polynomial having algebraic coefficients that vanishes on the unit real d-torus in a set whose Zariski closure in $\mathbb{G}_m^d$ has codimension at least 2. Our result includes a power-saving quantitative estimate of the decay rate of the equidistribution. It refines an ergodic theorem of Lind, Schmidt, and Verbitskiy, of which it also supplies a purely Diophantine proof. As an application, we confirm Ih's integrality finiteness conjecture on torsion points for a class of atoral divisors of $\mathbb{G}_m^d$.

1. Introduction

1A. ***Main results.*** Let $d \geq 1$ be an integer and let $\mathbb{G}_m^d$ denote the d-dimensional algebraic torus. We will identify $\mathbb{G}_m^d$ with $(\mathbb{C}\backslash\{0\})^d$, the group of its $\mathbb{C}$-points.

Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ be a *torsion point*, i.e., a point of a finite order. We define

$$\delta(\boldsymbol{\zeta}) = \inf\{|a| : a \in \mathbb{Z}^d\backslash\{0\} \text{ with } \boldsymbol{\zeta}^a = 1\} \tag{1-1}$$

where, here and throughout the article, $|\cdot|$ denotes the maximum-norm; we refer to Section 2 for the notation $\boldsymbol{\zeta}^a$.

MSC2010: primary 11J83; secondary 11R06, 14G40, 37A45, 37P30.
Keywords: number theory, Diophantine approximation, Mahler measure.

It is well-known that the Galois orbit

$$\{\boldsymbol{\zeta}^\sigma : \sigma \in \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/\mathbb{Q})\}$$

becomes equidistributed in $\mathbb{G}_m^d$ with respect to the Haar measure as $\delta(\boldsymbol{\zeta}) \to \infty$. More precisely, if $f : \mathbb{G}_m^d \to \mathbb{R}$ is a continuous function with compact support, then

$$\frac{1}{[\mathbb{Q}(\boldsymbol{\zeta}) : \mathbb{Q}]} \sum_{\sigma \in \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/\mathbb{Q})} f(\boldsymbol{\zeta}^\sigma) \to \int_{[0,1)^d} f(\boldsymbol{e}(x))\, dx \tag{1-2}$$

as $\delta(\boldsymbol{\zeta}) \to \infty$ where

$$\boldsymbol{e}(x) = (e^{2\pi\sqrt{-1}x_1}, \dots, e^{2\pi\sqrt{-1}x_d}) \tag{1-3}$$

for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Our aim is to investigate the equidistribution result for test functions $f = \log|P|$ where P is a nonzero Laurent polynomial in d unknowns and with algebraic coefficients. Such P may vanish on $(S^1)^d$, where $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is the unit circle, and so f is not defined everywhere. But for $\delta(\boldsymbol{\zeta})$ large in terms of P, Laurent's theorem [1984] also known as the Manin–Mumford conjecture for $\mathbb{G}_m^d$, implies that P does not vanish at any conjugate of $\boldsymbol{\zeta}$; see also [Sarnak and Adams 1994] for another proof. Moreover, the integral of f over $(S^1)^d$ exists as the singularity is merely logarithmic. It is known as the *Mahler measure*

$$m(P) = \int_{[0,1)^d} \log|P(\boldsymbol{e}(x))|\, dx, \tag{1-4}$$

see for instance Section 3.4 in [Schinzel 2000] for the convergence of this integral for arbitrary $P \in \mathbb{C}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$.

A *torsion coset* of $\mathbb{G}_m^d$ is the translate of a connected algebraic subgroup of $\mathbb{G}_m^d$ by a point of finite order. We call a torsion coset *proper* if it does not equal $\mathbb{G}_m^d$.

We call $P \in \mathbb{C}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$ *essentially atoral* if the Zariski closure of

$$\{(z_1, \dots, z_d) \in (S^1)^d : P(z_1, \dots, z_d) = 0\}$$

in $\mathbb{G}_m^d$ is a finite union of irreducible algebraic sets of codimension at least 2 and proper torsion cosets.

For example, if $d = 1$ then P is essentially atoral if and only if it does not vanish at any point of infinite multiplicative order in S^1.

Lind, Schmidt and Verbitskiy [Lind et al. 2013, Definition 2.1] define the notion of an *atoral* Laurent polynomial $P \in \mathbb{Z}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$. An atoral Laurent polynomial is essentially atoral in our sense. Moreover, if P is irreducible then it is atoral if and only if the intersection of its zero locus with $(S^1)^d$ has dimension at most $d-2$ as a semialgebraic set, see Proposition 2.2 [Lind et al. 2013]. A related, but not quite equivalent, definition of atoral Laurent polynomials with complex coefficients was introduced earlier by Agler, McCarthy and Stankus [Agler et al. 2006].

Let $\overline{\mathbb{Q}}$ denote the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. We are ready to state our first result.

Theorem 1.1. *For each essentially atoral $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$ there exists $\kappa > 0$ with the following property. Suppose $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has finite order with $\delta(\boldsymbol{\zeta})$ sufficiently large. Then $P(\boldsymbol{\zeta}^\sigma) \neq 0$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/\mathbb{Q})$ and*

$$\frac{1}{[\mathbb{Q}(\boldsymbol{\zeta}) : \mathbb{Q}]} \sum_{\sigma \in \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/\mathbb{Q})} \log|P(\boldsymbol{\zeta}^\sigma)| = m(P) + O(\delta(\boldsymbol{\zeta})^{-\kappa})$$

as $\delta(\boldsymbol{\zeta}) \to \infty$, where the implicit constant depends only on d and P.

Theorem 8.8 below is a more precise version of this result. In particular, we allow σ to range over subgroups of $\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/\mathbb{Q})$ whose index and conductor grow sufficiently slow, the conductor is defined in Section 3. Moreover, κ depends only on d and the number of nonzero terms appearing in P. Our method of proof allows one to determine an explicit value for κ.

Torsion points in $\mathbb{G}_m^d$ are characterized as the algebraic points of height zero; see Section 2 for the definition of the height $h : \mathbb{G}_m^n(\overline{\mathbb{Q}}) \to [0, \infty)$. Bilu [1997] proved that Galois orbits of algebraic points $\boldsymbol{\alpha} \in \mathbb{G}_m^d$ of small height satisfy an analogous equidistribution statement as (1-2), asymptotically as $h(\boldsymbol{\alpha}) \to 0$ and $\delta(\boldsymbol{\alpha}) \to \infty$; the definition (1-1) extends naturally to nontorsion points and may take infinity as a value. It is natural to ask whether Theorem 1.1 admits a suitable generalization to points of small height. Autissier's example [2006] rules out the verbatim generalization already for $\mathbb{G}_m$. He constructed a sequence $(\alpha_n)_{n\in\mathbb{N}}$ of pairwise distinct algebraic numbers whose height tends to 0 but such that $(1/[\mathbb{Q}(\alpha_n) : \mathbb{Q}]) \sum_\sigma \log|\sigma(\alpha_n) - 2|$ tends to 0 for $n \to \infty$. But the integral of the corresponding test function against the unit circle is $\log 2$. An interesting problem still arises if the test function has at worst a logarithmic singularity of real codimension at least 2 on $(S^1)^d$. Suppose that $|f(z)|$ is $O\big(\big|\log(|P(z)|^2 + |Q(z)|^2)\big|\big)$ on an open neighborhood of $(S^1)^d$ in $\mathbb{G}_m^d$, where P and Q are nonconstant and coprime Laurent polynomials with algebraic coefficients, and that f vanishes on the complement of a compact set in $\mathbb{G}_m^d$. One may then ask about comparing the average of f over the Galois orbit of $\boldsymbol{\alpha} \in \mathbb{G}_m^d(\overline{\mathbb{Q}})$ with the average of f over $(S^1)^d$: is their difference bounded by $\ll_f (h(\boldsymbol{\alpha}) + \delta(\boldsymbol{\alpha})^{-1})^\kappa$, for some $\kappa > 0$ depending only on P and Q? We also mention Chambert-Loir and Thuillier's Théorème 1.2 [2009] which is a general equidistribution result for points of small height, allowing $\log|P|$ as a test function if the zero locus of P in $\mathbb{G}_m^d$ is a finite union of torsion cosets. In this paper we allow $\log|P|$ as a test function if P is essentially atoral but we average over points of finite order.

Our Theorem 1.1 recovers a variant of the result of Lind, Schmidt and Verbitskiy [2013]. In their work, the sum is not over the Galois orbit of a single point of finite order but rather over a finite subgroup G of $\mathbb{G}_m^d$. For this purpose we define

$$\delta(G) = \inf\{|a| : a \in \mathbb{Z}^d\backslash\{0\} \text{ such that } \boldsymbol{\zeta}^a = 1 \text{ for all } \boldsymbol{\zeta} \in G\}. \tag{1-5}$$

Each finite subgroup of $\mathbb{G}_m^d$ is a disjoint union of Galois orbits. This observation allows us to recover the theorem of Lind, Schmidt, and Verbitskiy with an estimate on the decay rate.

Theorem 1.2. *Let $P \in \mathbb{Q}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$ be essentially atoral. There exists $\kappa > 0$ such that for any finite subgroup $G \subset \mathbb{G}_m^d$ we have*

$$\frac{1}{\#G} \sum_{\substack{\zeta \in G \\ P(\zeta) \neq 0}} \log|P(\zeta)| = m(P) + O(\delta(G)^{-\kappa}) \tag{1-6}$$

where the implicit constant depends only on d and P.

To relate (1-6) to the expression in Lind, Schmidt, and Verbitskiy's Theorem 1.3 [Lind et al. 2013] we refer to [Lind et al. 2010, Lemma 2.1] as well as the comments on pages 1063 and 1064 of [Lind et al. 2013]. Note that G is Ω_Γ and $\#G$ is $|\mathbb{Z}^d/\Gamma|$ in the notation of [Lind et al. 2013].

Lind, Schmidt, and Verbitskiy's approach is based on an in-depth study [Schmidt and Verbitskiy 2009; Lind et al. 2010; 2013] of an associated dynamical system: the algebraic $\mathbb{Z}^d$-action on a closed, shift-invariant subgroup of $(S^1)^{\mathbb{Z}^d}$ whose dual is $\mathbb{Z}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]/(P)$. The atoral condition, in the sense of [Lind et al. 2013], turns out to be equivalent to the existence of a nontrivial summable homoclinic point.

Theorem 1.2 may be read as a strong quantitative estimate on the growth of periodic points for such dynamical systems. The refinement to Galois orbits, Theorem 1.1, does not seem to be directly possible by the homoclinic method, nor does it seem to follow formally from the case (1-6) of finite subgroups, which is where the dynamical method applies.

Our method of proof draws its origins in work of Duke [2007]. It differs from the method of Lind, Schmidt, and Verbitskiy. However, it is striking that the notion of atoral appears crucially in both approaches.

The first-named author [Dimitrov 2016] was able to prove Theorem 1.2 for a general Laurent polynomial when G equals the group of N-torsion elements in $\mathbb{G}_m^d$.

Let us return to Galois orbits. We believe that the hypothesis on P being essentially atoral is also unnecessary in Theorem 1.1 on Galois orbits. The next conjecture sums up our expectations. It is related to Schmidt's conjecture [1995, Remark 21.16(2)].

Conjecture 1.3. *For each $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$ there exists $\kappa > 0$ with the following property. Suppose $\zeta \in \mathbb{G}_m^d$ has finite order with $\delta(\zeta)$ sufficiently large. Then $P(\zeta^\sigma) \neq 0$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$ and*

$$\frac{1}{[\mathbb{Q}(\zeta) : \mathbb{Q}]} \sum_{\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})} \log|P(\zeta^\sigma)| = m(P) + O(\delta(\zeta)^{-\kappa})$$

as $\delta(\zeta) \to \infty$, where the implicit constant depends only on d and P.

For $d = 1$ this conjecture follows from work of M. Baker, Ih, and Rumely [Baker et al. 2008], see their statement around (6). They use a version of A. Baker's deep estimates on linear forms in logarithms. Our Theorem 1.1 in the case $d = 1$ does not cover polynomials that vanish at a point of infinite order on the unit circle and therefore avoids the use of linear forms in logarithms. The conjecture is open already for $d = 2$ and

$$P = X_1 + X_1^{-1} + X_2 + X_2^{-1} - 3.$$

1B. ***Ih's conjecture on integral torsion points.*** As another application of our results we derive a special case of Ih's conjecture [Baker et al. 2008] in the multiplicative setting. Let $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$. Ih's conjecture predicts that the set of torsion points $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ such that $P(\boldsymbol{\zeta})$ is an algebraic unit is not Zariski dense in $\mathbb{G}_m^d$, unless the zero set of P in $\mathbb{G}_m^d$ is itself a finite union of proper torsion cosets. M. Baker, Ih, and Rumely [2008] cover the case $d = 1$ for arbitrary polynomials using their work on Conjecture 1.3. Here we mimic the approach of M. Baker, Ih, and Rumely and solve a case of Ih's conjecture for essentially atoral polynomials with integral coefficients and any d.

Corollary 1.4. *Let $K \subset \mathbb{C}$ be a number field with ring of integers $\mathbb{Z}_K$ and let $P \in \mathbb{Z}_K[X_1^{\pm 1}, \dots, X_d^{\pm 1}] \backslash \{0\}$. Suppose that the zero set of P in $\mathbb{G}_m^d$ is not a finite union of torsion cosets. Suppose in addition that $\tau(P)$ is essentially atoral for all field embeddings $\tau : K \to \mathbb{C}$. Then there exists $B \geq 1$ such that if $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has finite order and $P(\boldsymbol{\zeta})$ is an algebraic unit, then $\delta(\boldsymbol{\zeta}) \leq B$.*

Ih's conjecture expects the existence of B without assuming that each $\tau(P)$ is essentially atoral. Observe that the result of M. Baker, Ih, and Rumely is not a direct consequence of this corollary, as we do not allow univariate polynomials that vanish at a point of infinite multiplicative order on the unit circle. Our approach does not depend on the theory of linear forms in logarithms.

A special class of atoral polynomials, to which our results apply *a fortiori*, are the irreducible integer Laurent polynomials $P \in \mathbb{Z}[X_1^{\pm 1}, \dots, X_n^{\pm 1}] \setminus \{0\}$ that are not fixed up to a monomial factor and up to a sign by the involution sending each X_i to $1/X_i$. We call these P asymmetric. They are atoral in the sense of Lind, Schmidt and Verbitskiy, see the proof of [Lind et al. 2013, Proposition 2.2]. Hence an asymmetric Laurent polynomial is essentially atoral. The converse is false as the Laurent polynomial

$$X_1 + X_1^{-1} + X_2 + X_2^{-1} - 4.$$

is essentially atoral; indeed, its zero locus on $(S^1)^2$ consists of the single point $(1, 1)$.

If $K = \mathbb{Q}$, Corollary 1.4 in the case of an asymmetric, and thus necessarily irreducible Laurent polynomial P, can be deduced as follows from the Manin–Mumford conjecture for $\mathbb{G}_m^d$. Indeed, if γ is a unit in the ring of algebraic integers of a cyclotomic field, then $\eta = \bar{\gamma}/\gamma$ is an algebraic integer whose Galois conjugates lie on S^1. So η is a root of unity by Kronecker's theorem, see [Bombieri and Gubler 2006, Theorem 1.5.9]. We consider the zero $(\eta, \boldsymbol{\zeta})$ of $P(X_1^{-1}, \dots, X_d^{-1}) - X_0 P(X_1, \dots, X_d)$, which is irreducible and defines an algebraic subset of $\mathbb{G}_m^{d+1}$ none of whose geometric irreducible components is a torsion coset. A similar argument applies if K is a totally real number field.

1C. ***Overview of the proof.*** We close the introduction by describing the method of proof of Theorem 1.1, which builds upon work of the second-named author [Habegger 2018] and is related to the approach of Duke [2007]. The basic idea is to reduce the multivariate statement in Theorem 1.1 to the univariate case. Whereas we worked with torsion points of prime order in [Habegger 2018], a new technical difficulty in this paper is that we allow torsion points of arbitrary order.

Any torsion point $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ of order N takes on the form $(\zeta^{a_1}, \dots, \zeta^{a_d})$ where $\zeta = \boldsymbol{e}(1/N)$ is a root of unity of order N and $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$. The precise manner how the nonunique a is chosen is

delicate and will be discussed below. The notation $\boldsymbol{\zeta} = \zeta^a$ will be quite useful. A nonboldface ζ denotes a root of unity and boldface $\boldsymbol{\zeta}$ suggests a torsion point of $\mathbb{G}_m^d$.

If P is as in Theorem 1.1, but for simplicity with coefficients in $K = \mathbb{Q}$, we define the univariate polynomial

$$Q(X) = P(X^a) = P(X^{a_1}, \dots, X^{a_d}) \in \mathbb{Q}[X^{\pm 1}]. \tag{1-7}$$

Multiplying Q by a power of X turns out to be harmless, so one can assume that Q is a polynomial. The values $|P(\boldsymbol{\zeta}^\sigma)|$ equal the values of $|Q(\zeta^\sigma)|$ as σ ranges over $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$.

The univariate case and root separation (Section 4)**.** Let us suppose for the moment that $\boldsymbol{\zeta} = \zeta$ is a root of unity. It is classical that the Galois conjugates of ζ are equidistributed around the unit circle; we recall of these facts in Section 3. So (1-2) holds for $f(z) = \log|Q(z)|$ provided Q has no zero on the unit circle. In Proposition 4.5 we make convergence quantitative for such Q. Roughly speaking, for all $\epsilon > 0$ we have

$$\frac{1}{[\mathbb{Q}(\zeta) : \mathbb{Q}]} \sum_\sigma \log|Q(\zeta^\sigma)| = m(Q) + O_{P,\epsilon}\left(\frac{|a|^{1+\epsilon}}{N^{1-\epsilon}}\right) \tag{1-8}$$

where σ runs over $\mathrm{Gal}(\mathbb{Q}(\zeta)/\mathbb{Q})$. Actually, the hypothesis on Q is slightly weaker as we allow it to vanish at roots of unity, if all $Q(\zeta^\sigma) \neq 0$. This hypothesis is ultimately a reflection of the hypothesis that the multivariate P is essentially atoral in Theorem 1.1. Indeed, in the univariate case, being essentially atoral boils down to not vanishing at any point of infinite multiplicative order in S^1. The hypothesis on Q is crucial for our method to work. The main difficulty we encounter in the average (1-8) are exceptionally small values of Q at some ζ^σ. The burden is to show, in a uniform sense, that no complex root z of Q can be too close to ζ^σ in a suitable sense.

If z is itself a root of unity, doing this is straightforward as $|z - 1| \gg 1/\mathrm{ord}(z)$.

The difficulty lies in the case when z has infinite multiplicative order. Here it is tempting to apply a version of Baker's theorem on linear forms in logarithms, as did M. Baker, Ih, and Rumely [2008]. However, and as already discussed by Duke [2007, Section 3] this seems unhelpful for the problem at hand. Indeed, estimates on linear forms in two logarithms such as [Laurent et al. 1995] lead to a factor $[\mathbb{Q}(z) : \mathbb{Q}]^2 = O(|a|^2)$ in a bound for any member of the sum in (1-8). This is not good enough for our application as $|a|^2/[\mathbb{Q}(\zeta) : \mathbb{Q}]$ may spoil the average in (1-8).

Our solution is to use the banal inequality $|z - \zeta| \geq \big||z| - 1\big|$ which lies at the heart of the method here and in [Habegger 2018]. As z is no root of unity, and as Q does not vanish at points of infinite multiplicative order on S^1, we have $|z| \neq 1$ and so the banal inequality provides a nontrivial lower bound. We now explain how it leads to a useful estimate on $|z - \zeta|$ via lower bounding $\big||z| - 1\big|$.

If z is close to the unit circle, then $\big||z| - 1\big|$ is approximately $|z - 1/\bar{z}|$. In [Habegger 2018] a result of Mahler [1964] on the separation of roots of an integer polynomial led to a suitable lower bound for $|z - 1/\bar{z}|$. In that paper, Habegger used his counting result on approximations to a set definable in an o-minimal structure. This allowed to make Mahler's estimate uniform over the various zeros z of Q.

The main tool of the present paper is a uniform generalization of Mahler's inequality for the separation of several pairs of roots of Q. Such a generalization was obtained by Mignotte [1995]. In Section 4 we give a variant of Mignotte's theorem that is tailored to our application and is self-contained. We thus bypass the o-minimal theory used in [Habegger 2018]. We still require Bombieri, Masser, and Zannier's theorem [Bombieri et al. 2007] to be mentioned below. Moreover, our Theorem 1.1 is effective in nature.

A possible approach towards Conjecture 1.3 lies in extending (1-8) to Q that are allowed to vanish at any point of S^1. As observed, we lack a suitable lower bound for $|z-\zeta|$ if z is an algebraic number of infinite multiplicative order on the unit circle. As suggested in the similar setting of [Habegger 2018, Lemma 4.2], it turns out that the z of interest have small height $h(z)$. We therefore propose the following conjecture.

Conjecture 1.5. *For all $B \geq 1$ and $\epsilon > 0$ there exists a constant $c = c(B,\epsilon) > 0$ with the following property. Let $z \in \mathbb{C}$ be an algebraic number with $|z| = 1$ and $h(z) \leq B/D$ where $D = [\mathbb{Q}(z) : \mathbb{Q}]$. If $\zeta \in \mathbb{C}\backslash\{z\}$ is a root of unity of order N, then $\log|\zeta - z| \geq -cD^{1+\epsilon}N^{\epsilon}$.*

The crux of this conjecture is its dependency on the degree D. In comparison, the state-of-the-art results in the theory of linear forms in two logarithms of algebraic numbers in the D-aspect, such as Laurent, Mignotte, and Nesterenko's Théorème 3 [Laurent et al. 1995], have only a quadratic dependency on D.

Equidistribution of torsion points (Section 3)**.** We return to the case $\boldsymbol{\zeta} = \zeta^a$ of a general torsion point in $\mathbb{G}_m^d$ of order N. The exponent vector a used to define Q as in (1-7) depends on $\boldsymbol{\zeta}$. For this reason it is important that the error term in Proposition 4.5 is explicit in terms of Q. Moreover, it is important to choose a with $|a|$ as small as possible. For fixed ζ the exponent a is well-defined up to addition of an element in $N\mathbb{Z}^d$. So clearly we may assume $|a| \leq N$, although this is not good enough in view of (1-8). Fortunately, there is a second degree of freedom, namely we can replace $\boldsymbol{\zeta}$ by any Galois conjugate of itself.

This leads us to classical questions of equidistribution of the Galois orbit of $\boldsymbol{\zeta}$; we compile the necessary statements in Section 3. Using the Erdös–Turán theorem and the theory of Gauss sums, Lemma 3.7 produces a with $|a| = O(N\delta(\boldsymbol{\zeta})^{-1/(3d)})$ such that ζ^a is a Galois conjugate of $\boldsymbol{\zeta}$.

Let us return to the error term in (1-8). One factor N cancels out and the error term becomes $N^{2\epsilon}\delta(\boldsymbol{\zeta})^{-(1+\epsilon)/(3d)}$. The innocuous ϵ in (1-8) is ultimately responsible for the factor $N^{2\epsilon}$. Although $\delta(\boldsymbol{\zeta}) \leq N$, there is no nontrivial bound in the reverse direction and $N^{2\epsilon}\delta(\boldsymbol{\zeta})^{-(1+\epsilon)/(3d)}$ could explode.

Factoring $\boldsymbol{\zeta}$ (Section 5)**.** The solution to this problem is described in Section 5. In Proposition 5.1 we factor $\boldsymbol{\zeta}$ into a product $\boldsymbol{\eta}\boldsymbol{\xi}$ where $\boldsymbol{\xi}$ has finite order M such that $\boldsymbol{\xi} = \boldsymbol{e}(a/M)$ where $|a| = O(M^{1-\kappa})$. Moreover, the order of $\boldsymbol{\eta}$ is bounded from above by a small power of N. The power saving obtained in the exponent of N is small even when compared to the saving obtained for $|a|$. The methods employed come from the geometry of numbers and slopes of lattices in $\mathbb{R}^d$.

We will replace $\boldsymbol{\zeta}$ by $\boldsymbol{\xi}$ and the univariate polynomial $Q(X) = P(X^a)$ by $P(\boldsymbol{\eta}^a X^a)$. This last transformation does not change the height or the monomial structure of Q. But it can change the field generated by its coefficients as the order of $\boldsymbol{\eta}$ and hence its field of definition vary as $\boldsymbol{\zeta}$ varies. For this reason, we must keep track of the base field of Q throughout the whole argument.

Putting everything together (Sections 6, 7, 8)**.** In Sections 6 and 8 we put all ingredients together to prove the final result. Here we apply a result of Bombieri, Masser, and Zannier [2007] on the intersections of a subvariety in $\mathbb{G}_m^d$ of codimension at least 2 with all 1-dimensional algebraic subgroups of $\mathbb{G}_m^d$. Roughly speaking, this result shows that if P is essentially atoral, then for "most" choices of a the univariate polynomial Q as in (1-7) does not vanish at any point of infinite multiplicative order on S^1. Recall that this property of Q was crucial to deduce (1-8). Bombieri, Masser, and Zannier's result is related to the study of unlikely intersections, for an overview we refer to Zannier's book [2012]. Another tool that makes an appearance is Lawton's theorem [1983].

The intermediate Section 7 contains a weak version of a result of Hlawka [1971] on the numerical integration of a continuous, multivariate function. The results obtained there are useful in connection with the function attaching the Mahler measure to a nonzero polynomial.

Appendices. In Appendix A we give a quantitative version of Lawton's theorem [1983] regarding the convergence of a sequence of Mahler measures. Unfortunately, we are not able to use the very closely related theorem in [Habegger 2018] as we require additional uniformity. The arguments in this appendix follow closely Lawton's strategy. After this paper was submitted, Brunault, Guilloux, Mehrabdollahi, and Pengo proved a higher dimensional generalization of our version of Lawton's theorem with explicit constants [Brunault et al. 2022].

Finally, in Appendix B we show how to deduce Theorem 1.2, the theorem of Lind, Schmidt and Verbitskiy, from our Theorem 1.1.

1D. ***Final remarks.*** The results mentioned above, in particular the theorem of Bombieri, Masser, and Zannier, also play an important role in Le's approach [2014]. The question on how small a sum of roots of unity can be was raised by Myerson [1986] in connection with a combinatorial question [Myerson 1979; 1980] which was later studied by Duke [2007]. Dubickas [2018] has more recent work in this direction for sums of 2 and 3 roots of unity of prime order.

2. Notation and preliminaries

Apart from the notation already introduced we use $\mathbb{N}$ to denote the natural numbers $\{1, 2, 3, \dots\}$. If $x = (x_1, \dots, x_m)$ with all x_i elements in an abelian group G and if $A = (a_{i,j})_{i,j} \in \mathrm{Mat}_{m,n}(\mathbb{Z})$ we write $x^A = (x_1^{a_{1,1}} \cdots x_m^{a_{m,1}}, \dots, x_1^{a_{1,n}} \cdots x_m^{a_{m,n}}) \in G^n$. So if $B \in \mathrm{Mat}_{n,p}(\mathbb{Z})$, then $(x^A)^B = x^{AB}$. For a commutative ring R with 1 we let $R^\times$ denote its group of units. Euler's function φ maps $N \in \mathbb{N}$ to the cardinality of $(\mathbb{Z}/N\mathbb{Z})^\times$. The group of all roots of unity in $\mathbb{C}^\times$ is μ_∞. We often identify $\mathbb{G}_m^d$ with the set of its complex points $(\mathbb{C}^\times)^d$ and let 1 denote the unit element $(1, \dots, 1) \in \mathbb{G}_m^d$. If $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ is a torsion point, we write

$\operatorname{ord}(\boldsymbol{\zeta})$ for its order. We write $\langle \cdot, \cdot \rangle$ for the Euclidean inner product on $\mathbb{R}^d$, $|\cdot|_2$ for the Euclidean norm on $\mathbb{R}^d$, and $|\cdot|$ for the maximum-norm on $\mathbb{R}^d$ and $\mathrm{Mat}_{m,n}(\mathbb{R})$. We define $\log^+ x = \log \max\{1, x\}$ for all $x \geq 0$.

The constants implicit in Vinogradov's notation $\ll_{x,y,z,\ldots}$, $\gg_{x,y,z,\ldots}$, and in $O_{x,y,z,\ldots}(\cdots)$ depend only on the values $x, y, z, \ldots$ appearing in the subscript.

Let $P \in \mathbb{C}[X_1^{\pm 1}, \ldots, X_d^{\pm 1}]\backslash\{0\}$, then $|P|$ denotes the maximum-norm of the coefficient vector of P and we set $|0| = 0$. Recall that $m(P)$ is the Mahler measure of P. It follows from Corollaries 4 and 6 in Chapter 3.4 of [Schinzel 2000] that $\exp(m(P))$ is at most the Hermitian norm of the coefficient vector of P. Suppose P has at most $k \geq 1$ nonzero terms, we find

$$m(P) \leq \log|P| + \tfrac{1}{2}\log k. \tag{2-1}$$

The following result of Dobrowolski and Smyth [2017, Corollary 2] provides a reverse inequality of the same quality.

Theorem 2.1 (Dobrowolski and Smyth). *Suppose $P \in \mathbb{C}[X_1^{\pm 1}, \ldots, X_d^{\pm 1}]\backslash\{0\}$ has at most $k \geq 2$ nonzero terms with k an integer. Then $m(P) \geq \log|P| - (k-2)\log 2$.*

Therefore,

$$|m(P) - \log|P|| \ll k \tag{2-2}$$

with absolute implied constant. Observe that if P is a polynomial, then

$$m(P) \geq \log|P| - \log(2) \sum_{i=1}^{d} \deg_{X_i} P$$

by the classical Lemma 1.6.10 of [Bombieri and Gubler 2006]. So (2-2) is stronger when the number of terms in P is known to be bounded, which is often the case in our work.

Let x be an element of a number field K. The absolute logarithmic Weil height, or just height, of x is

$$h(x) = \frac{1}{[K:\mathbb{Q}]} \sum_v [K_v : \mathbb{Q}_v] \log \max\{1, |x|_v\}; \tag{2-3}$$

here v runs over all places of K normalized such that $|2|_v = 2$ for an infinite place v and $|p|_v = 1/p$ if v lies above the rational prime p, the completion of K with respect to v is K_v and the completion of $\mathbb{Q}$ with respect to the restriction of v is $\mathbb{Q}_v$. Let P be a nonzero Laurent polynomial with coefficients $x_0, \ldots, x_n \in K$. The absolute logarithmic Weil height, or just height, of P is

$$h(P) = \frac{1}{[K:\mathbb{Q}]} \sum_v [K_v : \mathbb{Q}_v] \log \max\{|x_0|_v, \ldots, |x_n|_v\}. \tag{2-4}$$

See [Bombieri and Gubler 2006, Chapter 1] for more details on heights. For example, $h(x)$ and $h(P)$ are well-defined for $x \in \overline{\mathbb{Q}}$ and $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \ldots, X_d^{\pm 1}]$, i.e., the values do not depend on the number field K containing x and the coefficients of P, respectively. Moreover $h(P) = h(\lambda P)$ for all $\lambda \in \overline{\mathbb{Q}}^\times$.

3. Quantitative Galois equidistribution for torsion points

We need a strong enough quantitative version of the Galois equidistribution of torsion points $\boldsymbol{\zeta}$ of $\mathbb{G}_m^d$, with a power saving discrepancy in $\delta(\boldsymbol{\zeta})$ defined in (1-1).

Different approaches are possible and we opt to use the Erdös–Turán–Koksma bound. This reduces the problem to the estimation of certain exponential sums, which happen to be Gauss sums that can be explicitly evaluated.

Let $N \in \mathbb{N}$. For a divisor $f \in \mathbb{N}$ of N we work with the canonical surjective, homomorphism $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/f\mathbb{Z})^\times$ induced by reducing modulo f.

The *conductor* f_G of a subgroup G of $(\mathbb{Z}/N\mathbb{Z})^\times$ is the least positive integer $f \mid N$ such that G contains $\ker((\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/f\mathbb{Z})^\times)$.

Certainly, f_G is well-defined as $\ker((\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N\mathbb{Z})^\times)$ is the trivial subgroup. Moreover, $f_{(\mathbb{Z}/N\mathbb{Z})^\times} = 1$. But one should take care as the conductor of $G = \{1\}$ is $N/2$ for $N \equiv 2 \pmod 4$. Observe that $[(\mathbb{Z}/N\mathbb{Z})^\times : G] \le \varphi(f_G)$.

The group $(\mathbb{Z}/N\mathbb{Z})^\times$ is naturally isomorphic to the Galois group of $\mathbb{Q}(\zeta)/\mathbb{Q}$, where ζ is a root of unity of order N. Let $L \subset \mathbb{Q}(\zeta)$ be the fixed field of G. Then L lies in the fixed field of $\ker((\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/f_G\mathbb{Z})^\times)$ which equals $\mathbb{Q}(\zeta_{f_G})$ where ζ_{f_G} is a root of unity of order f_G.

Let $f \ge 1$ be an integer and ζ_f of order f. We claim $L \subset \mathbb{Q}(\zeta_f)$ if and only if $f_G \mid f$. Indeed, if the inclusion holds, then $L \subset \mathbb{Q}(\zeta_f) \cap \mathbb{Q}(\zeta_{f_G})$. It is well-known that the intersection is generated by a root of unity of order $\gcd(f, f_G)$. By minimality of f_G we find $f_G \mid f$. The converse direction follows as $\mathbb{Q}(\zeta_{f_G}) \subset \mathbb{Q}(\zeta_f)$ if $f_G \mid f$.

So f_G is the greatest common divisor of all f, for which $L \subset \mathbb{Q}(\zeta_f)$. Equivalently f_G is the greatest common divisor of all $f \mid N$, for which $\ker((\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/f\mathbb{Z})^\times) \subset G$.

By class field theory, f_G is the finite part of the conductor of the abelian extension $L/\mathbb{Q}$.

The next lemma collects some classical facts on Gauss sums. We write $f_\chi = f_{\ker \chi}$ for a character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$. We recall that $\boldsymbol{e}(\,\cdot\,)$ was defined in (1-3).

Lemma 3.1. *Let $N \in \mathbb{N}$ and say $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ is a character. For $k \in \mathbb{Z}$ we define $\tau = \sum_{\sigma \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(\sigma)\boldsymbol{e}(k\sigma/N)$, then the following hold true*:

(i) *If* $\gcd(k, N) = 1$ *then* $|\tau| \le f_\chi^{1/2}$.

(ii) *For unrestricted k we set $N' = N/\gcd(k, N)$. Then*

$$|\tau| \le \frac{\varphi(N)}{\varphi(N')} f_\chi^{1/2}.$$

Proof. If $k = 1$, part (i) follows directly from [Iwaniec and Kowalski 2004, Lemma 3.1, Section 3.4]. The more general case $\gcd(k, N) = 1$ follows as $\sum_{\sigma \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(\sigma)\boldsymbol{e}(k\sigma/N) = \sum_{\sigma \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(k'\sigma)\boldsymbol{e}(\sigma/N)$ where $kk' \equiv 1 \pmod N$ and since χ is completely multiplicative.

To prove (ii) set $N' = N/\gcd(k, N)$ and $k' = k/\gcd(k, N)$. Then τ is

$$\sum_{\sigma\in(\mathbb{Z}/N\mathbb{Z})^\times} \chi(\sigma)\boldsymbol{e}\left(\frac{k'}{N'}\sigma\right) = \sum_{\sigma'\in(\mathbb{Z}/N'\mathbb{Z})^\times} \Bigg(\sum_{\substack{\sigma\in(\mathbb{Z}/N\mathbb{Z})^\times \\ \sigma\equiv\sigma' \pmod{N'}}} \chi(\sigma)\Bigg)\boldsymbol{e}\left(\frac{k'}{N'}\sigma\right).$$

The inner sum on the right runs over a coset of the kernel of $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N'\mathbb{Z})^\times$. Since χ is a character, the inner sum equals 0 if the said kernel does not lie in the kernel of χ. In this case, $\tau = 0$ and we are done.

Otherwise, $\ker((\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N'\mathbb{Z})^\times) \subset \ker\chi$, and then $f_\chi \mid N'$. We find moreover that χ factors through a character $\chi' : (\mathbb{Z}/N'\mathbb{Z})^\times \to \mathbb{C}^\times$ and $f_{\chi'} \mid f_\chi$. As the kernel of $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N'\mathbb{Z})^\times$ has order $\varphi(N)/\varphi(N')$ we have

$$\tau = \frac{\varphi(N)}{\varphi(N')} \sum_{\sigma'\in(\mathbb{Z}/N'\mathbb{Z})^\times} \chi'(\sigma')\boldsymbol{e}\left(\frac{k'}{N'}\sigma\right).$$

Part (ii) now follows from (i) since $\gcd(k', N') = 1$. □

Lemma 3.2. *Let $N \in \mathbb{N}$, let G be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$, and let $k \in \mathbb{Z}$. We define $N' = N/\gcd(k, N)$, then*

$$\frac{1}{\#G}\left|\sum_{\sigma\in G} \boldsymbol{e}(k\sigma/N)\right| \le \frac{[(\mathbb{Z}/N\mathbb{Z})^\times : G]}{\varphi(N')} f_G^{1/2}.$$

Proof. Let $\chi_1', \dots, \chi_m' : (\mathbb{Z}/N\mathbb{Z})^\times/G \to \mathbb{C}^\times$ be all characters and $m = [(\mathbb{Z}/N\mathbb{Z})^\times : G]$. Then $\sum_{i=1}^m \chi_i'(\sigma) = 0$ for all $\sigma \in (\mathbb{Z}/N\mathbb{Z})^\times/G$ expect for the neutral element, where this sum equals m. Write χ_i for $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N\mathbb{Z})^\times/G$ composed with χ_i'. Then $\sum_{i=1}^m \chi_i(\sigma) = 0$ if and only if $\sigma \in (\mathbb{Z}/N\mathbb{Z})^\times \backslash G$, otherwise this sum is m. Therefore,

$$\sum_{\sigma\in G} \boldsymbol{e}(k\sigma/N) = \frac{1}{m}\sum_{i=1}^m \sum_{\sigma\in(\mathbb{Z}/N\mathbb{Z})^\times} \chi_i(\sigma)\boldsymbol{e}(k\sigma/N) \tag{3-1}$$

and Lemma 3.1(ii) implies

$$\left|\sum_{\sigma\in(\mathbb{Z}/N\mathbb{Z})^\times} \chi_i(\sigma)\boldsymbol{e}(k\sigma/N)\right| \le \frac{\varphi(N)}{\varphi(N')} f_{\chi_i}^{1/2}.$$

Note that $G \subset \ker\chi_i$ because χ_i factors through $(\mathbb{Z}/N\mathbb{Z})^\times \to (\mathbb{Z}/N\mathbb{Z})^\times/G$. So $f_{\chi_i} \le f_G$, by the minimality of f_{χ_i}. The current lemma now follows from (3-1). □

Let $d, n \in \mathbb{N}$ and $x_1, \dots, x_n \in [0, 1)^d$. The *discrepancy* of $(x_1, \dots, x_n)$ is

$$\mathcal{D}(x_1, \dots, x_n) = \sup_B \left|\frac{\#\{i : x_i \in B\}}{n} - \mathrm{vol}(B)\right| \tag{3-2}$$

where B ranges over all products $\prod_{i=1}^d [\alpha_i, \beta_i)$ with $0 \le \alpha_i < \beta_i \le 1$. Note that the discrepancy lies in $[0, 1]$. In some references such as [Harman 1998], the discrepancy is not normalized by dividing by n and can be greater than 1.

In the next proposition we bound from above the discrepancy of the Galois orbit of a point of finite order in $\mathbb{G}_m^d$ using the Gauss sum estimates above. Below, $d_0(N)$ denotes the number of divisors of a natural number N.

Proposition 3.3. *Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have order N and let G be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ such that $\{\boldsymbol{\zeta}^\sigma : \sigma \in G\} = \{\boldsymbol{e}(x_i) : 1 \le i \le \#G\}$ with all x_i in $[0,1)^d$:*

(i) *We have*

$$\mathcal{D}(x_1, \dots, x_{\#G}) \ll_d [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{(\log 2\delta(\boldsymbol{\zeta}))^{d-1} \log\log 3\delta(\boldsymbol{\zeta})}{\delta(\boldsymbol{\zeta})^{1/2}}.$$

(ii) *If $d = 1$, then*

$$\mathcal{D}(x_1, \dots, x_{\#G}) \ll [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{\log(2N) d_0(N)}{\varphi(N)}.$$

Proof. We abbreviate $n = \#G$. We fix $a \in \mathbb{Z}^d$ with $\boldsymbol{\zeta} = \boldsymbol{e}(a/N)$. Then N and the entries of a are coprime. Let $H \ge 4$ be an integer. We use the Erdös–Turán–Koksma inequality [Harman 1998, Theorem 5.21], to bound the discrepancy $\mathcal{D} = \mathcal{D}(x_1, \dots, x_n)$ as follows

$$\mathcal{D} \ll_d \frac{1}{H} + \sum_{\substack{b \in \mathbb{Z}^d \setminus \{0\} \\ |b| \le H}} \frac{1}{r(b)} \left| \frac{1}{n} \sum_{\sigma \in G} \boldsymbol{e}\left(\frac{\langle a, b\rangle}{N} \sigma \right) \right| \tag{3-3}$$

here $r(b_1, \dots, b_d) = \max\{1, |b_1|\} \cdots \max\{1, |b_d|\}$.

By Lemma 3.2, the expression inside the modulus is at most $C/\varphi(N/\gcd(\langle a, b\rangle, N))$ with $C = [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2}$. We have $\varphi(M) \gg M/\log\log(3+M)$ for all integers $M \ge 1$ with an absolute and effective implicit constant, see for example [Rosser and Schoenfeld 1962, Theorem 15]. Therefore,

$$\mathcal{D} \ll_d \frac{1}{H} + C \sum_{\substack{b \in \mathbb{Z}^d \setminus \{0\} \\ |b| \le H}} \frac{1}{r(b)} \frac{\gcd(\langle a, b\rangle, N)}{N} \log\log(3 + N/\gcd(\langle a, b\rangle, N)).$$

If $b \in \mathbb{Z}^d \setminus \{0\}$, then

$$\left\langle a, \frac{N}{\gcd(\langle a, b\rangle, N)} b \right\rangle = N \frac{\langle a, b\rangle}{\gcd(\langle a, b\rangle, N)} \in N\mathbb{Z}$$

which implies $\boldsymbol{\zeta}^{bN/\gcd(\langle a,b\rangle,N)} = 1$. So $N/\gcd(\langle a, b\rangle, N) \ge \delta/|b| > 0$ where $\delta = \delta(\boldsymbol{\zeta})$. As $t \mapsto (\log\log(3+t))/t$ is decreasing on $t > 0$ we find

$$\mathcal{D} \ll_d \frac{1}{H} + C \frac{1}{\delta} \sum_{\substack{b \in \mathbb{Z}^d \setminus \{0\} \\ |b| \le H}} \frac{|b|}{r(b)} \log\log(3 + \delta).$$

The sum of $|b|/r(b)$ over all $b \in \mathbb{Z}^d$ with $1 \le |b| \le H$ is $\ll_d H(\log H)^{d-1}$, so we find

$$\mathcal{D} \ll_d \frac{1}{H} + C \frac{\log\log(3\delta)}{\delta} H (\log H)^{d-1}.$$

Part (i) follows by fixing H to be the least integer with $H \ge \delta^{1/2}$ and $H \ge 4$.

In part (ii) we have $d = 1$. We may assume $N \geq 4$ as the discrepancy is at most 1. Here a is coprime to N and so $\gcd(ab, N) = \gcd(b, N)$. In (3-3) we take $H = N$ and use again Lemma 3.2 with C as before to find

$$\mathcal{D} \ll \frac{1}{N} + \sum_{b=1}^{N} \frac{C}{b\varphi(N/\gcd(b,N))} \leq \frac{1}{N} + \sum_{g|N} \frac{C}{g\varphi(N/g)} \sum_{e=1}^{N/g} \frac{1}{e}.$$

In the sum over g we have $g\varphi(N/g) \geq \varphi(N)$ and the harmonic sum is $\ll \log N$. So $\mathcal{D} \ll 1/N + C(\log N)d_0(N)/\varphi(N)$, which implies (ii). □

A variant of the case $d = 1$ already appears in [Baker et al. 2008, Lemma 1.3] which is attributed to Pomerance.

The discrepancy bound in (i) depends on $\delta(\boldsymbol{\zeta})$. But $\delta(\boldsymbol{\zeta})$ is always bounded above by N. So estimates that decay in N are stronger than estimates that decay in $\delta(\boldsymbol{\zeta})$. However, there can be no upper bound for the discrepancy in terms of the order N.

Let us assume for the moment $d = 1$. Then we have $\delta(\boldsymbol{\zeta}) = N$. If $[(\mathbb{Z}/N\mathbb{Z})^\times : G]$ and f_G are fixed, the decay of the discrepancy is $1/N$ up to terms of subpolynomial growth by part (ii) of the preceding proposition and standard estimates for Euler's function φ.

The *total variation* on $[a, b]$ of a real valued function F whose domain contains the interval $[a, b]$ with $a \leq b$ is

$$\mathrm{Var}_a^b(F) = \sup_{a \leq x_0 \leq \cdots \leq x_m \leq b} \sum_{i=1}^{m} |F(x_i) - F(x_{i-1})|.$$

For $a = 0$ and $b = 1$ we abbreviate $\mathrm{Var}(F) = \mathrm{Var}_a^b(F)$.

The next lemma requires Koksma's inequality.

Lemma 3.4. *Let $F : [0, 1] \to \mathbb{R}$ be a function with $\mathrm{Var}(F) < \infty$. If $N \geq 1$ is an integer and G is a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ such that $\{\zeta^\sigma : \sigma \in G\} = \{\boldsymbol{e}(x_i) : 1 \leq i \leq \#G\}$ with all x_i in $[0, 1)$, then*

$$\left| \frac{1}{\#G} \sum_{i=1}^{\#G} F(x_i) - \int_0^1 F(x)\,dx \right| \ll [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{\log(2N)d_0(N)}{\varphi(N)} \mathrm{Var}(F).$$

Proof. The claim follows from Theorems 1.3 and 5.1 in Chapter 2 of [Kuipers and Niederreiter 1974] together with Proposition 3.3(ii). □

3A. *A univariate average.*

Lemma 3.5. *Let $\alpha \in \mathbb{C}$ and $r > 0$. For $x \in [0, 1]$ we define*

$$F_{\alpha,r}(x) = \log \max(r, |\boldsymbol{e}(x) - \alpha|).$$

Then $F_{\alpha,r} : [0, 1] \to \mathbb{R}$ satisfies $\mathrm{Var}(F_{\alpha,r}) \leq 3 \log(1 + 2/r)$.

Proof. We abbreviate $F = F_{\alpha,r}$. By elementary geometry we can find $m \leq 3$ and $0 = x_0 < x_1 < \cdots < x_m = 1$ such that F is monotone on all $[x_{i-1}, x_i]$. Then $\mathrm{Var}_{x_{i-1}}^{x_i}(F) = |F(x_i) - F(x_{i-1})|$ and

$\mathrm{Var}(F) = \sum_{i=1}^{m} \mathrm{Var}_{x_{i-1}}^{x_i}(F)$. We have $\log \max\{r, |\alpha|-1\} \le F(x) \le \log\max\{r, |\alpha|+1\}$ for all $x \in [0,1]$. Hence $\mathrm{Var}_{x_{i-1}}^{x_i}(F) \le \log(\max\{r,|\alpha|+1\}/\max\{r,|\alpha|-1\})$ which we see is at most $\log(1+2/r)$ by considering the cases $|\alpha| \ge 1+r$ and $|\alpha| < 1+r$. Thus $\mathrm{Var}(F) \le 3\log(1+2/r)$. □

The value r serves as a truncation parameter. We now apply Koksma's inequality to $F_{\alpha,r}$.

Lemma 3.6. *Let $\zeta \in \mu_\infty$ have order N and let G be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$. Let $\alpha \in \mathbb{C}$ and $r \in (0,1]$, then*

$$\frac{1}{\#G} \sum_{\substack{\sigma \in G \\ |\zeta^\sigma - \alpha| > r}} \log|\zeta^\sigma - \alpha| = \log^+|\alpha| + O\left(\left([(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{\log(2N) d_0(N)}{\varphi(N)} + r\right)\left|\log\frac{r}{2}\right|\right). \quad (3\text{-}4)$$

Proof. We let I denote the left-hand side of (3-4). Then $I = I_1 + I_2$ with

$$I_1 = \frac{1}{\#G}\sum_{i=1}^{\#G} F_{\alpha,r}(x_i) \quad \text{and} \quad I_2 = \frac{1}{\#G} \sum_{\substack{i \\ |\boldsymbol{e}(x_i)-\alpha| \le r}} -\log r$$

with the $x_i \in [0,1)$ as in Lemma 3.4 and $F_{\alpha,r}$ as in Lemma 3.5. The integrals below are understood to be over subsets of $[0,1]$. Applying Lemmas 3.4 and 3.5 to $F_{\alpha,r}$ yields

$$I_1 = \int_0^1 F_{\alpha,r}(x)\,dx + O\left([(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{\log(2N)d_0(N)}{\varphi(N)}\left|\log\frac{r}{2}\right|\right). \quad (3\text{-}5)$$

The set of $x \in [0,1]$ with $|\boldsymbol{e}(x) - \alpha| \le r$ is of the form $\varnothing$, $[a,b]$, or $[0,a]\cup[b,1]$. So its characteristic function has total variation at most 2. Lemma 3.4 applied to this characteristic function yields

$$I_2 = -\int_{|\boldsymbol{e}(x)-\alpha|\le r} \log r\,dx + O\left([(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \frac{\log(2N)d_0(N)}{\varphi(N)}\right). \quad (3\text{-}6)$$

The sum of the integrals in (3-5) and (3-6) equals

$$\int_0^1 \log|\boldsymbol{e}(x) - \alpha|\,dx - \int_{|\boldsymbol{e}(x)-\alpha|\le r} \log|\boldsymbol{e}(x)-\alpha|\,dx.$$

Jensen's formula [Bombieri and Gubler 2006, Proposition 1.6.5] implies that the first integral equals $\log^+|\alpha|$. To complete the proof it suffices to show that the second integral is $O(r|\log r/2|)$.

The integral is nonpositive as $r \le 1$ and we may assume that it is nonzero. First assume, $|\alpha| \le \frac{1}{2}$. In this case $|\boldsymbol{e}(x)-\alpha| \ge \frac{1}{2}$ and the integral is $O(r)$. Second, say $|\alpha| > \frac{1}{2}$. [Rahman and Schmeisser 2002, Lemma 11.6.1] implies $|\boldsymbol{e}(x)-\alpha| \ge |\alpha|^{1/2}|\boldsymbol{e}(x)-\boldsymbol{e}(y)| \ge 2^{-1/2}|\boldsymbol{e}(x-y)-1|$ where $\alpha = |\alpha|\boldsymbol{e}(y)$ and $|x-y| \le \frac{1}{2}$. There is an absolute and effectively computable constant $C > 0$ with $|\boldsymbol{e}(x-y)-1| \ge C|x-y|$ and thus $|\boldsymbol{e}(x)-\alpha| \ge 2^{-1/2}C|x-y|$. In the integral we have $r \ge |\boldsymbol{e}(x)-\alpha|$ and so the desired bound follows from elementary analysis. □

3B. ***A Galois conjugate near*** **1.** We will also need an estimate on the minimal distance of a Galois conjugate of a torsion point to the unit element.

Lemma 3.7. *Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have order N and let G be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$. There exist $\sigma \in G$ and $a \in \mathbb{Z}^d$ with $\boldsymbol{\zeta} = \boldsymbol{e}(a\sigma/N)$, $|a| < N$, and*

$$\frac{|a|}{N} \ll_d \frac{[(\mathbb{Z}/N\mathbb{Z})^\times : G]^{1/d} f_G^{1/(2d)}}{\delta(\boldsymbol{\zeta})^{1/(3d)}}. \tag{3-7}$$

Proof. Let $\boldsymbol{\zeta} = \boldsymbol{e}(b/N)$ with $b \in \mathbb{Z}^d$, the entries of b and N have no common prime divisor. Suppose $x_1, \dots, x_n$ are as in Proposition 3.3 coming from the $\boldsymbol{\zeta}^\sigma$ as σ ranges over G where $n = \#G$. There exists $c(d) > 0$ depending only on d with $\mathcal{D}(x_1, \dots, x_n) \le c(d)[(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \delta(\boldsymbol{\zeta})^{-1/3}$; note that $\delta(\boldsymbol{\zeta}) = N$ if $d = 1$. We set $\kappa = 2c(d)^{1/d}[(\mathbb{Z}/N\mathbb{Z})^\times : G]^{1/d} f_G^{1/(2d)} \delta(\boldsymbol{\zeta})^{-1/(3d)}$. If $\kappa \ge 1$ we take σ the identity and fix $a \in \mathbb{Z}^d$ with $|a| < N$ and $\boldsymbol{\zeta} = \boldsymbol{e}(a/N)$. Otherwise, by the definition of the discrepancy the hypercube $[0, \kappa)^d$ contains some $x_i = a/N$. Hence a satisfies, $|a| < N$, (3-7), and $\boldsymbol{e}(a/N) = \boldsymbol{\zeta}^{\sigma^{-1}}$ for some $\sigma \in G$. □

4. Theorem of Mahler and Mignotte

In this section, we establish the separation of pairs of roots of an integer polynomial. Theorem 4.1 below was shown by Mahler [1964] for the case $k = 1$ of a single pair of roots. Mignotte [1995] generalized Mahler's inequality to products over several disjoint pairs of roots (see his Theorem 1). We reproduce here a lightened version of Mignotte's theorem that is suitable for our needs. The proof is an adaptation of Mahler's original argument about a single pair, guided by the principle that Liouville's Inequality bounds an algebraic number at an arbitrary set of places in terms of the height. Let us also mention Güting's proof [1961] of a less precise earlier result involving the length of a polynomial instead of the Mahler measure.

Let $Q \in \mathbb{C}[X]$ be a nonzero univariate polynomial. By Jensen's formula its Mahler measure equals

$$m(Q) = \log|a_0| + \sum_{i=1}^{D} \log^+|z_i| \tag{4-1}$$

if $Q = a_0(X - z_1) \cdots (X - z_D)$ and where the z_i are complex. If Q is nonconstant, we let $\operatorname{disc}(Q)$ denote its discriminant as a degree $\deg Q$ polynomial.

Theorem 4.1. *Let $Q \in \mathbb{C}[X] \setminus \mathbb{C}$ be of degree D and with no repeated roots. If $z_1, \dots, z_k, z_1', \dots, z_k'$ are pairwise distinct complex roots of Q, then*

$$\sum_{j=1}^{k} -\log|z_j - z_j'| \le \tfrac{1}{2}(D + 2k)\log D - \tfrac{k}{2}\log 3 + (D-1)m(Q) - \tfrac{1}{2}\log|\operatorname{disc}(Q)| \tag{4-2}$$

with strict inequality for $k \ge 1$.

Proof. We modify Mahler's and Mignotte's argument as follows.

Both sides of (4-2) are invariant under multiplication Q by a nonzero scalar. So we may assume that Q is monic. After possibly swapping z_j with z'_j we may assume $|z_j| \geq |z'_j|$ for all j.

We augment $z_1, \dots, z_k$ to all complex roots $z_1, \dots, z_D$ of Q. Then we consider the Vandermonde determinant

$$V = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_D \\ \vdots & \vdots & & \vdots \\ z_1^{D-1} & z_2^{D-1} & \cdots & z_D^{D-1} \end{pmatrix},$$

which is nonzero as $z_1, \dots, z_D$ are pairwise distinct. For $j \in \{1, \dots, k\}$, let $i_j > k$ be the index with $z'_j = z_{i_j}$. For these j, we subtract the i_j-th column from the j-th column and factoring each difference $z_j - z_{i_j}$ out of the determinant with the identities $z_j^m - z_{i_j}^m = (z_j - z_{i_j})(z_j^{m-1} + z_j^{m-2} z_{i_j} + \cdots + z_{i_j}^{m-1})$, $1 \leq m \leq D-1$. We obtain an expression

$$V = W \prod_{j=1}^{k} (z_j - z_{i_j}) = W \prod_{j=1}^{k} (z_j - z'_j), \tag{4-3}$$

where $W \neq 0$ is the determinant of the matrix having

$$\begin{pmatrix} 0 \\ 1 \\ z_j + z'_j \\ \vdots \\ z_j^{D-2} + z_j^{D-3} z'_j + \cdots + {z'_j}^{D-2} \end{pmatrix}$$

for its j-th column, $j \in \{1, \dots, k\}$, and the same entries as in the Vandermonde matrix in the remaining columns. By Hadamard's inequality, $|W|$ is bounded from above by the product of the Hermitian norms of all these columns. The j-th column, for some $j \in \{1, \dots, k\}$, has Hermitian norm

$$\sqrt{\sum_{m=0}^{D-2} |z_j^m + z_j^{m-1} z'_j + \cdots + {z'_j}^m|^2} \leq \sqrt{\sum_{m=0}^{D-2} (m+1)^2} \max\{1, |z_j|, |z'_j|\}^{D-2} < \sqrt{D^3/3} \cdot \max\{1, |z_j|\}^{D-1}$$

where we used $|z'_j| \leq |z_j|$. The Hermitian norm of the j-th column with $j \in \{k+1, \dots, D\}$ is at most $\sqrt{D} \max\{1, |z_j|\}^{D-1}$.

Applying Hadamard's inequality, using these two bounds, and taking the logarithm yields

$$\begin{aligned} \log|W| &\leq \frac{k}{2} \log\left(\frac{D^3}{3}\right) + \frac{D-k}{2} \log D + (D-1) \sum_{j=1}^{D} \log^+ |z_j| \\ &= \frac{D+2k}{2} \log D - \frac{k}{2} \log 3 + (D-1) m(Q) \end{aligned}$$

as Q is monic; the inequality is strict for $k \geq 1$. If $k = 0$ no column operations are necessary and $V = W$.

The monic polynomial Q has discriminant $\operatorname{disc}(Q) = V^2$. Consequently $|V| = |\operatorname{disc}(Q)|^{1/2}$, and in view of (4-3) we have

$$\sum_{j=1}^{k} -\log|z_j - z_j'| = \log|W| - \log|V| \le \tfrac{1}{2}(D+2k)\log D - \tfrac{k}{2}\log 3 + (D-1)m(Q) - \tfrac{1}{2}\log|\operatorname{disc}(Q)|$$

with a strict inequality for $k \ge 1$. This concludes the proof. □

While Theorem 4.1 suffices for our needs here, we remark that it is possible to relax the hypothesis to having $z_1, \dots, z_k$ pairwise distinct and $\{z_1, \dots, z_k\} \cap \{z_1', \dots, z_k'\} = \varnothing$, at the cost of a slightly worse upper bound (4-2).

The following corollary holds for integral polynomials that are not necessarily squarefree.

Corollary 4.2. *Let $Q \in \mathbb{Z}[X] \setminus \mathbb{Z}$ be of degree D. If $z_1, \dots, z_k, z_1', \dots, z_k'$ are pairwise distinct complex roots of Q, then*

$$\sum_{j=1}^{k} -\log|z_j - z_j'| \le \frac{D+2k}{2}\log D - \frac{k}{2}\log 3 + (D-1)m(Q) \tag{4-4}$$

with strict inequality for $k \ge 1$.

Proof. We may assume $k \ge 1$. We begin by splitting off the squarefree part of Q. More precisely, we factor $Q = \widetilde{Q}R$ where $\widetilde{Q}$, $R \in \mathbb{Z}[X]$ and $\widetilde{Q}$ is squarefree and vanishes at all complex roots of Q. The discriminant $\operatorname{disc}(\widetilde{Q})$ is a nonzero integer, and so $|\operatorname{disc}(\widetilde{Q})| \ge 1$. Moreover, $m(\widetilde{Q}) \ge 0$. Theorem 4.1 applied to $\widetilde{Q}$ and $1 \le \deg \widetilde{Q} \le D$ implies that the sum on the left of (4-4) is at most $\frac{1}{2}(D+2k)\log D - \frac{k}{2}\log 3 + (D-1)m(\widetilde{Q})$. The corollary follows from $m(\widetilde{Q}) = m(Q) - m(R) \le m(Q)$. □

4A. ***A repulsion property of the unit circle.*** A key point in [Habegger 2018] is that while Mahler's theorem does not give a strong enough bound for the distance of a complex root of $Q \in \mathbb{Z}[X] \setminus \{0\}$ to an N-th root of unity (the product $(X^N - 1)Q(X)$ has an exceedingly large degree), it can be used to bound the distance from the unit circle to the locus of roots of P lying off the unit circle. With Corollary 4.2, this repulsion property of the unit circle can be strengthened as follows.

Lemma 4.3. *Let $Q \in \mathbb{Z}[X] \setminus \mathbb{Z}$ and $Q = a_0(X - z_1)\cdots(X - z_D)$ where $z_1, \dots, z_D \in \mathbb{C}$. Then*

$$\sum_{\substack{j=1 \\ |z_j| \ne 1}}^{D} \log^+ \frac{1}{\big||z_j| - 1\big|} \le D\log\left(\frac{3+\sqrt{5}}{2}\right) + 2D\log(2D) + 4Dm(Q) \le 4D(\log(2D) + m(Q)). \tag{4-5}$$

Before we come to the proof let us remark that $||z| - 1|$ is the distance $\operatorname{dist}(z, S^1)$ of $z \in \mathbb{C}$ to the unit circle S^1. Thus inequality (4-5) can be restated as providing

$$\frac{1}{D}\sum_{\substack{j=1 \\ |z_j| \ne 1}}^{D} \log^+ \frac{1}{\operatorname{dist}(z_j, S^1)} \le \log\left(\frac{3+\sqrt{5}}{2}\right) + 2\log(2D) + 4m(Q).$$

Our result suggests that the unit circle repels roots of Q that lie off the unit circle. Related estimates are implicit in work of Dubickas [1997], see his Theorem 2.

Proof. The second bound in (4-5) is elementary, so it suffices to prove the first one.

Say $Q = a_0 Q_1 \cdots Q_n$ where each $Q_i \in \mathbb{Z}[X]$ is irreducible of positive degree with $a_0 \in \mathbb{Z}$. Observe that $0 \le m(Q_i) \le m(Q)$ and $\sum_i \deg Q_i = \deg Q$. So it suffices to prove (4-5) for Q irreducible in $\mathbb{Z}[X]$. We may also assume $Q(0) \neq 0$.

We will apply Corollary 4.2 to the polynomial $\widetilde{Q} \in \mathbb{Z}[X]$ constructed from Q in the following manner. If $Q(1/X)X^D \neq \pm Q$ we take $\widetilde{Q} = Q(X)Q(1/X)X^D$ and $\widetilde{Q} = Q$ otherwise. So $\widetilde{D} = \deg \widetilde{Q} = \delta D$ and $m(\widetilde{Q}) = \delta m(Q)$ with $\delta = 2$ in the first case and $\delta = 1$ in the second case. Indeed, a polynomial and its reciprocal have the same Mahler measure. For any root z of $\widetilde{Q}$ we also have $\widetilde{Q}(1/\bar{z}) = 0$.

The following basic observation for a complex number z will prove useful. We have $|z - 1/\bar{z}| \le 1$ if and only if $\phi^{-1} \le |z| \le \phi$ with $\phi = (1+\sqrt{5})/2$ the golden ratio.

Let $w_1, \dots, w_k$ be the roots of $\widetilde{Q}$ without repetition such that $\phi^{-1} \le |w_j| < 1$. Then $w_j' = 1/\overline{w_j}$ is a root of $\widetilde{Q}$ for each $j \in \{1, \dots, k\}$ with $|w_j'| > 1$. Corollary 4.2 yields

$$\sum_{j=1}^{k} \log^+ \frac{1}{|w_j - 1/\overline{w_j}|} \le \delta D \log(\delta D) + \delta^2 D m(Q) \tag{4-6}$$

because $k \le \widetilde{D}/2 = \delta D/2$ and $m(Q) \ge 0$.

Suppose z_j is a root of Q with $|z_j| \neq 1$ and $\phi^{-1} \le |z_j| \le \phi$. Then $z_j \in \{w_l, 1/\overline{w}_l\}$ for some unique l. The mapping $j \mapsto l$ is at worst 2-to-1 and injective if $\delta = 2$ as Q is irreducible.[1] This leads to the factor $2/\delta$ in

$$\sum_{\substack{|z_j| \neq 1 \\ 1/\phi \le |z_j| \le \phi}} \log^+ \frac{1}{|z_j - 1/\overline{z_j}|} \le \frac{2}{\delta} \sum_{l=1}^{k} \log^+ \frac{1}{|w_l - 1/\overline{w}_l|} \tag{4-7}$$

For a complex number z with $|z| \ge \phi^{-1}$ we have $|z - 1/\bar{z}| = \frac{|z|+1}{|z|}\big||z| - 1\big| \le (1+\phi)\big||z| - 1\big|$. This allows us to get

$$\sum_{\substack{|z_j| \neq 1 \\ 1/\phi \le |z_j| \le \phi}} \log^+ \frac{1}{\big||z_j| - 1\big|} \le s \log(1+\phi) + \sum_{\substack{|z_j| \neq 1 \\ 1/\phi \le |z_j| \le \phi}} \log^+ \frac{1}{|z_j - 1/\bar{z}_j|}$$

where s is the number of terms in the first sum. There are $D - s$ remaining roots of Q and if $|z_j| < \phi^{-1}$ or $|z_j| > \phi$ we get $\log^+ 1/\big||z| - 1\big| \le \log(1+\phi)$. Together with (4-6) and (4-7) we find

$$\sum_{|z_j| \neq 1} \log^+ \frac{1}{\big||z_j| - 1\big|} \le D \log(1+\phi) + 2D \log(\delta D) + 2\delta D m(Q).$$

We have established (4-5) for Q as $\delta \le 2$. □

Next we generalize our bound to a polynomial with coefficients in a number field. Recall that $h(Q)$ is the absolute logarithmic projective Weil height of a nonzero polynomial Q with algebraic coefficients.

[1] Indeed, if $z_j, z_k \in \{w_l, 1/\overline{w}_l\}$ with $z_j \neq z_k$, then $z_j = 1/\bar{z}_k$. So $\widetilde{Q} = Q$ and hence $\delta = 1$ in this case.

Corollary 4.4. *Let $F \subset \mathbb{C}$ be a number field and let $Q \in F[X] \setminus F$ and $Q = a_0(X - z_1)\cdots(X - z_D)$ where $z_1, \dots, z_D \in \mathbb{C}$. Then*

$$\sum_{\substack{j=1 \\ |z_j|\neq 1}}^{D} \log^+ \frac{1}{\big||z_j| - 1\big|} \le 10[F : \mathbb{Q}]^2 D(\log(2D) + h(Q)).$$

Proof. Let $\widetilde{Q}$ be the product of the $\mathbb{Q}$-Galois conjugates of Q. Then $\widetilde{Q}$ has rational coefficients and degree $\widetilde{D} \le D[F : \mathbb{Q}]$. Let $\lambda \in \mathbb{N}$ such that $\lambda\widetilde{Q}$ is integral with content 1. For the projective height we find $h(\widetilde{Q}) = \log|\lambda\widetilde{Q}|$. Together with [Bombieri and Gubler 2006, Lemma 1.6.7] we get $m(\lambda\widetilde{Q}) \le \frac{1}{2}\log(1+\widetilde{D}) + h(\widetilde{Q})$. As all $\mathbb{Q}$-Galois conjugates of Q have the same projective height we use elementary estimates at local places, see [Bombieri and Gubler 2006, Remark 1.6.14], to find

$$h(\widetilde{Q}) \le [F : \mathbb{Q}]\log(1 + D) + [F : \mathbb{Q}]h(Q).$$

By Lemma 4.3 applied to $\lambda\widetilde{Q}$, the sum $\sum_{j=1:|z_j|\neq 1}^{D} \log^+ 1/\big||z_j| - 1\big|$ is at most

$$4\widetilde{D}\big(\log(2\widetilde{D}) + \tfrac{1}{2}\log(1 + \widetilde{D}) + [F : \mathbb{Q}]\log(1 + D) + [F : \mathbb{Q}]h(Q)\big).$$

We use $1 + \widetilde{D} \le 2\widetilde{D} \le 2D[F : \mathbb{Q}] \le (2D)^{[F:\mathbb{Q}]}$ to complete the proof. □

4B. ***Averages over roots of unity.*** In this subsection we apply the repulsion property of the unit circle, Corollary 4.4, to estimate the norm of cyclotomic integers of the form $Q(\zeta)$, where ζ is a varying root of unity and Q is a moderately controlled univariate polynomial with algebraic coefficients and without zeros in $S^1 \setminus \mu_\infty$. This gives a fairly uniform solution of the one dimensional essentially atoral case and forms the basis for the higher dimensional case to be taken up in the next sections.

Proposition 4.5. *Let $F \subset \mathbb{C}$ be a number field and let $Q \in F[X]\setminus\{0\}$ be of degree at most $D \ge 1$ with no roots in $S^1\setminus\mu_\infty$. Let $\zeta \in \mu_\infty$ be of order N and G a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ such that $Q(\zeta^\sigma) \neq 0$ for all $\sigma \in G$. Then*

$$\frac{1}{\#G}\sum_{\sigma\in G} \log|Q(\zeta^\sigma)| = m(Q) + O\left([F : \mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} D(\log(2D) + h(Q))\frac{(\log 2N)^3 d_0(N)}{N}\right). \quad (4\text{-}8)$$

Proof. We may assume that Q is nonconstant and $D = \deg Q$. Let $Q = a_0(X - z_1)\cdots(X - z_D)$. The idea is that each given root z_j may get within distance of $\le 1/N^2$ to at most a single conjugate of ζ.

We call z_j exceptional if $|\zeta^{\sigma_j} - z_j| \le 1/N^2$ for some $\sigma_j \in G$. As $|\xi - \xi'| \ge 4/N$ for distinct roots of unity ξ, ξ' of order N we see that σ_j is uniquely determined by z_j. Note that $\zeta^{\sigma_j} \neq z_j$ because $Q(\zeta^{\sigma_j}) \neq 0 = Q(z_j)$.

We apply Lemma 3.6 with $\alpha = z_j$ and $r = 1/N^2$. Thus

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|\zeta^\sigma - z_j| = \log^+|z_j| + \frac{1}{\#G}\log|\zeta^{\sigma_j} - z_j| + O\bigg([(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2}\frac{(\log 2N)^2 d_0(N)}{\varphi(N)} + \frac{\log 2N}{N^2}\bigg), \quad (4\text{-}9)$$

if z_j is exceptional, otherwise the same bound without the term $(\#G)^{-1}\log|\zeta^{\sigma_j} - z_j|$ holds true. As $1/N^2 \le 1/\varphi(N)$ we merge $(\log 2N)/N^2$ into the first term of the error term. Summing (4-9) over all $j \in \{1, \dots, D\}$ and adding $\log|a_0|$ gives

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|Q(\zeta^\sigma)| = m(Q) - \frac{1}{\#G}\sum_{j=1}^{D}{}' \log\frac{1}{|\zeta^{\sigma_j} - z_j|} + O\bigg([(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} D\frac{(\log 2N)^2 d_0(N)}{\varphi(N)}\bigg), \quad (4\text{-}10)$$

the dash signifies that we only sum over those j for which z_j is exceptional.

To bound the dashed sum we require Corollary 4.4. If z_j is exceptional, then $|\zeta^{\sigma_j} - z_j| \le 1$. Therefore, the dashed sum is nonnegative.

First, we consider the subsum over all exceptional $z_j \notin \mu_\infty$. Then $|z_j| \ne 1$ and $|\zeta^{\sigma_j} - z_j| \ge ||z_j| - 1|$ by the reverse triangle inequality. By Corollary 4.4 we find

$$0 \le \sum_{\substack{j=1\\ z_j\notin\mu_\infty}}^{D}{}' \log\frac{1}{|\zeta^{\sigma_j} - z_j|} \le \sum_{\substack{j=1\\ z_j\notin\mu_\infty}}^{D}{}' \log^+\frac{1}{||z_j| - 1|} = O([F:\mathbb{Q}]^2 D(\log(2D) + h(Q))). \quad (4\text{-}11)$$

Second, we consider the subsum over all exception $z_j \in \mu_\infty$, which is harmless. Recall that $\zeta^{\sigma_j} \ne z_j$. Since the order of z_j is $\ll [\mathbb{Q}(z_j) : \mathbb{Q}]^2 \le (D[F : \mathbb{Q}])^2$ and the order of ζ^{σ_j} is N we find $|\zeta^{\sigma_j} - z_j| \gg N^{-1}(D[F : \mathbb{Q}])^{-2}$. On the other hand, $|\zeta^{\sigma_j} - z_j| \le N^{-2}$ and hence $N \ll (D[F : \mathbb{Q}])^2$. We obtain the crude estimate $|\zeta^{\sigma_j} - z_j| \gg (D[F : \mathbb{Q}])^{-4} \gg (2D)^{-4[F:\mathbb{Q}]}$ and finally bound the at most D terms below separately to get

$$0 \le \sum_{\substack{j=1\\ z_j\in\mu_\infty}}^{D}{}' \log\frac{1}{|\zeta^{\sigma_j} - z_j|} = O([F:\mathbb{Q}]D\log(2D)). \quad (4\text{-}12)$$

We divide the sum of (4-11) and (4-12) by $\#G$ to find

$$0 \le \frac{1}{\#G}\sum_{j=1}^{D}{}' \log\frac{1}{|\zeta^{\sigma_j} - z_j|} = O\bigg([F:\mathbb{Q}]^2 D(\log(2D) + h(Q))\frac{[(\mathbb{Z}/N\mathbb{Z})^\times : G]}{\varphi(N)}\bigg).$$

The proposition follows from (4-10) and $\varphi(N) \gg N/\log\log(3N)$, a consequence of [Rosser and Schoenfeld 1962, Theorem 15]. □

Proposition 4.5 and ultimately Theorem 4.1 may be viewed as our input from transcendence theory. If this or a comparable bound held without the restrictive condition that Q has no roots in $S^1 \setminus \mu_\infty$ then it could be used to attack Conjecture 1.3. We were unable to prove or disprove that a suitable version of Proposition 4.5 extends to general polynomials. Progress on Conjecture 1.5 could indicate a path towards this goal.

5. Geometry of numbers

Let $d \geq 1$ and suppose $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has order N. It would be useful if $\boldsymbol{\zeta}$ had a Galois conjugate close to the unit element 1. If the distance were at most a small power of N^{-1}, this conjugate could be used to help reduce the multivariate Theorem 1.1 to the univariate Proposition 4.5, see [Habegger 2018].

Unfortunately, such a conjugate need not exist. Take for example $\boldsymbol{\zeta} = \boldsymbol{e}(1/p, 1/p^n)$ where p is a prime and $n \in \mathbb{N}$, here $N = p^n$. Any conjugate of $\boldsymbol{\zeta}$ has distance $\gg 1/p$ to 1 regardless of the value of n. The problem is that $\boldsymbol{\zeta}$ is up to a point of order p contained in the algebraic subgroup $\{1\} \times \mathbb{G}_m$.

We overcome this difficult by constructing a factorization $\boldsymbol{\zeta} = \boldsymbol{\eta}\boldsymbol{\xi}$ into torsion points $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ that satisfy the following properties for prescribed $\epsilon > 0$. First, the order of $\boldsymbol{\eta}$ is small relative to N, more precisely it is $O_{d,\epsilon}(N^\epsilon)$. Second, some Galois conjugate of $\boldsymbol{\xi}$ is at distance at most $O_{d,\epsilon}(N^{-\kappa(\epsilon)})$ to 1. Here $\kappa(\epsilon)$ is expected to be small for small ϵ. But we will see that $\kappa(\epsilon)/\epsilon$ is large. This is of central importance for our application.

We use the geometry of numbers to construct this factorization. An important tool is the slope of a lattice.

A lattice Λ in $\mathbb{R}^d$ is a finitely generated and discrete subgroup of $\mathbb{R}^d$. The rank of Λ is denoted by $\mathrm{rk}(\Lambda)$ and its determinant by $\det(\Lambda)$. We consider the set

$$A = \{(r, \log\det(\Omega)) : r \in \mathbb{Z} \text{ and } \Omega \text{ is a subgroup of } \Lambda \text{ with } \mathrm{rk}(\Omega) = r\}$$

and use the convention $\det(\{0\}) = 1$. In contrast to the convention in Arakelov theory, we have no sign in front of $\log\det(\Lambda)$. Observe that the second coordinate is bounded from below on A. Stuhler [1976, Proposition 1] proved that for each $j \in \{0, \dots, \mathrm{rk}(\Lambda)\}$ there exists a sublattice $\Lambda_j \subset \Lambda$ of rank j, possibly nonunique, with $\log\det(\Lambda_j)$ minimal among all sublattices of rank j. The lower boundary of the convex hull of A is the graph of a piecewise linear, continuous, convex function $P : [0, \mathrm{rk}(\Lambda)] \to \mathbb{R}$. As $\Lambda_0 = \{0\}$ and $\Lambda_{\mathrm{rk}(\Lambda)} = \Lambda$ we find $P(0) = 0$ and $P(\mathrm{rk}(\Lambda)) = \log\det(\Lambda)$.

For each $j \in \{1, \dots, \mathrm{rk}(\Lambda)\}$, the slope of P on $[j-1, j]$ is

$$\mu_j(\Lambda) = P(j) - P(j-1).$$

By convexity we have

$$\mu_1(\Lambda) \leq \mu_2(\Lambda) \leq \cdots \leq \mu_{\mathrm{rk}(\Lambda)}(\Lambda).$$

Moreover, $\mu_1(\Lambda) + \cdots + \mu_j(\Lambda) = P(j) - P(0) = P(j)$ for all j as $P(0) = \log\det(\Lambda_0) = 0$.

Assume $\Lambda \neq \{0\}$ and let $\nu \in \left(0, \frac{1}{2}\right]$ be a parameter. Suppose that

$$\mu_j(\Lambda) < \nu^{\mathrm{rk}(\Lambda)-j+1} \log \det(\Lambda)$$

for all $j \in \{1, \dots, \mathrm{rk}(\Lambda)\}$. Taking the sum yields

$$\log \det(\Lambda) < (\nu + \nu^2 + \cdots + \nu^{\mathrm{rk}(\Lambda)}) \log \det(\Lambda).$$

As $\nu \in \left(0, \frac{1}{2}\right]$ we must have $\det(\Lambda) < 1$.

Let us now assume $\det(\Lambda) \geq 1$, then there exists a unique $j_0 \in \{0, \dots, \mathrm{rk}(\Lambda) - 1\}$ such that

$$\mu_k(\Lambda) < \nu^{\mathrm{rk}(\Lambda)-k+1} \log \det(\Lambda) \quad \text{for all } 1 \leq k \leq j_0 \text{ and} \quad \mu_{j_0+1}(\Lambda) \geq \nu^{\mathrm{rk}(\Lambda)-j_0} \log \det(\Lambda). \tag{5-1}$$

We write $\Lambda(\nu)$ for the rank j_0 lattice Λ_{j_0}, indicating its dependency on ν. It satisfies $\mathrm{rk}(\Lambda/\Lambda(\nu)) \geq 1$.

Note that $\mu_{j_0}(\Lambda(\nu)) < \mu_{j_0+1}(\Lambda(\nu))$ if $j_0 \geq 1$. Therefore, $\Lambda(\nu)$ appears in the Harder–Narasimhan filtration of Λ as considered by Stuhler [1976] and Grayson [1984], if we include $\{0\}$ as a member of the filtration. In particular, $\Lambda(\nu)$ is the unique lattice in Λ of rank $\mathrm{rk}(\Lambda(\nu))$ and minimal determinant.

Here are two simple properties:

First, for the Euclidean norm $|\cdot|_2$ we claim

$$\log |v|_2 \geq \nu^{\mathrm{rk}(\Lambda/\Lambda(\nu))} \log \det(\Lambda) \quad \text{for all } v \in \Lambda \backslash \Lambda(\nu). \tag{5-2}$$

Indeed, the lattice Λ' generated by $\Lambda(\nu)$ and v contains $\Lambda(\nu)$ strictly. We must have $\mathrm{rk}(\Lambda') > \mathrm{rk}(\Lambda(\nu))$, as $\det(\Lambda')$ would otherwise be strictly less than $\det \Lambda(\nu)$. (This shows in particular that $\Lambda/\Lambda(\nu)$ is torsion free; a well-known property of the Harder–Narasimhan filtration.) So $\mathrm{rk}(\Lambda') = \mathrm{rk}(\Lambda) + 1$ and by convexity of P we find $\log \det(\Lambda') \geq \log \det(\Lambda(\nu)) + \mu_{j_0+1}(\Lambda)$. On the other hand, $\det(\Lambda') \leq \det(\Lambda') \det(\Lambda(\nu) \cap v\mathbb{Z}) \leq \det(\Lambda(\nu)) \det(v\mathbb{Z})$ is well-known, for a proof see [Stuhler 1976, Proposition 2]. We conclude $\log \det(v\mathbb{Z}) \geq \mu_{j_0+1}(\Lambda)$. Now $\det(v\mathbb{Z}) = |v|_2$, so (5-2) follows from (5-1).

Second, (5-1) and $\nu \in \left[0, \frac{1}{2}\right)$ imply

$$\log \det(\Lambda(\nu)) \leq \mu_1(\Lambda) + \cdots + \mu_{j_0}(\Lambda) \leq 2\nu^{1+\mathrm{rk}(\Lambda/\Lambda(\nu))} \log \det(\Lambda). \tag{5-3}$$

We now make things more concrete. Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have order N and set

$$\Lambda_{\boldsymbol{\zeta}} = \{u \in \mathbb{Z}^d : \boldsymbol{\zeta}^u = 1\}. \tag{5-4}$$

We consider the homomorphism $\mathbb{Z}^d \to \mathbb{G}_m$ defined by $u \mapsto \boldsymbol{\zeta}^u$ and see that $\mathbb{Z}^d/\Lambda_{\boldsymbol{\zeta}}$ is isomorphic to the finite subgroup of $\mathbb{G}_m$ generated by the coordinates of $\boldsymbol{\zeta}$. So $\mathbb{Z}^d/\Lambda_{\boldsymbol{\zeta}}$ is cyclic of order N. In particular, $\Lambda_{\boldsymbol{\zeta}}$ is a lattice in $\mathbb{R}^d$ of rank d with $\det(\Lambda_{\boldsymbol{\zeta}}) = [\mathbb{Z}^d : \Lambda_{\boldsymbol{\zeta}}] = N \geq 1$. The saturation

$$\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu) = \{u \in \mathbb{Z}^d : \text{there is } n \in \mathbb{Z}\backslash\{0\} \text{ such that } nu \in \Lambda_{\boldsymbol{\zeta}}(\nu)\} \tag{5-5}$$

of $\Lambda_{\boldsymbol{\zeta}}$ in $\mathbb{Z}^d$ will also be useful for us. It is a lattice of the same rank as $\Lambda_{\boldsymbol{\zeta}}(\nu)$.

For any lattice $\Lambda \subset \mathbb{R}^d$ of positive rank, we set

$$\lambda_1(\Lambda) = \min\{|u| : u \in \Lambda\backslash\{0\}\} \tag{5-6}$$

where as usual $|\cdot|$ denotes the maximum-norm. It is convenient to define $\lambda_1(\{0\}) = \infty$.

Proposition 5.1. *Let $\nu \in \left(0, \frac{1}{4}\right]$ and let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ be of order N. There exists $V \in \mathrm{GL}_d(\mathbb{Z})$ and a decomposition $\boldsymbol{\zeta} = \boldsymbol{\eta}\boldsymbol{\xi}$ with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ in $\mathbb{G}_m^d$ of finite order E and M, respectively, such that the following holds. We abbreviate $r = \mathrm{rk}(\Lambda_{\boldsymbol{\zeta}}/\Lambda_{\boldsymbol{\zeta}}(\nu)) \in \{1, \dots, d\}$:*

(i) *We have $E \mid N$, $M \mid N$, and $E \le N^{2\nu^{1+r}}$. In particular, $\mathbb{Q}(\boldsymbol{\eta}, \boldsymbol{\xi}) = \mathbb{Q}(\boldsymbol{\zeta})$.*

(ii) *We have $|V| \ll_d N^{2\nu^{1+r}}$ with $\boldsymbol{\xi}^V = (1, \dots, 1, \boldsymbol{\xi}')$ and $\boldsymbol{\xi}' \in \mathbb{G}_m^r$.*

(iii) *If G is a subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$ there exist $a \in \mathbb{Z}^r$ and $\sigma \in G$ such that $\boldsymbol{\xi}' = \boldsymbol{e}(a\sigma/M)$,*

$$|a| < M \quad \text{and} \quad \frac{|a|}{M} \ll_d \frac{[(\mathbb{Z}/M\mathbb{Z})^\times : G] f_G^{1/2}}{N^{\nu^r/(6d)}}. \tag{5-7}$$

(iv) *With the definition* (1-1) *we have $\delta(\boldsymbol{\xi}) \ge d^{-1/2} \min\{\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu)), N^{\nu^d/2}\}$.*

Moreover, if $r = d$, or equivalently $\Lambda(\nu) = \{0\}$, then V is the identity matrix.

Proof. We abbreviate $\Lambda = \Lambda_{\boldsymbol{\zeta}}$ as well as $\Lambda(\nu) = \Lambda_{\boldsymbol{\zeta}}(\nu)$ and $\widetilde{\Lambda}(\nu) = \widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu)$. Note $\det(\Lambda) = N$.

We can find a collection of $d - r = \mathrm{rk}(\widetilde{\Lambda}(\nu))$ linearly independent vectors in $\widetilde{\Lambda}(\nu)$ whose norms are at most $\ll_d \det(\widetilde{\Lambda}(\nu))$ by applying Minkowski's second theorem, see Theorem V in Chapter VIII of [Cassels 1959], and using $\lambda_1(\Lambda) \ge 1$. By appending suitable standard basis vectors of $\mathbb{Z}^d$ we find d linearly independent vectors in $\mathbb{Z}^d$. By Corollary 2, Chapter I.2 of [loc. cit.] applied to $\mathbb{Z}^d$ and these vectors we get a basis of $\mathbb{Z}^d$ whose entries have norm $\ll_d \det(\widetilde{\Lambda}(\nu))$. By the said corollary, the original linearly independent vectors can be expressed via an triangular matrix in terms of the new basis vectors. So the first $\mathrm{rk}(\widetilde{\Lambda}(\nu))$ entries of this basis are a basis of the saturated group $\widetilde{\Lambda}(\nu)$. Thus there exists $V \in \mathrm{GL}_d(\mathbb{Z})$ whose first $\mathrm{rk}(\widetilde{\Lambda}(\nu))$ columns constitute a basis of $\widetilde{\Lambda}(\nu)$ and

$$|V| \ll_d \det(\widetilde{\Lambda}(\nu)). \tag{5-8}$$

As $\det(\widetilde{\Lambda}(\nu)) \le \det(\Lambda(\nu))$, the bound for $|V|$ in (ii) follows from (5-3).

We write $\boldsymbol{\zeta}^V = (\boldsymbol{\eta}', \boldsymbol{\xi}')$ where $\boldsymbol{\eta}' \in \mathbb{G}_m^{d-r}$ and $\boldsymbol{\xi}' \in \mathbb{G}_m^r$ both have finite order dividing N. We take $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ from the assertion to equal $(\boldsymbol{\eta}', 1, \dots, 1)^{V^{-1}}$ and $(1, \dots, 1, \boldsymbol{\xi}')^{V^{-1}}$, respectively. So $\boldsymbol{\zeta} = \boldsymbol{\eta}\boldsymbol{\xi}$.

Observe that $[\widetilde{\Lambda}(\nu) : \Lambda(\nu)]\widetilde{\Lambda}(\nu) \subset \Lambda(\nu) \subset \Lambda$. So the first $\mathrm{rk}(\widetilde{\Lambda}(\nu))$ entries of $\boldsymbol{\zeta}^{[\widetilde{\Lambda}(\nu):\Lambda(\nu)]V}$ are $\boldsymbol{\eta}'^{[\widetilde{\Lambda}(\nu):\Lambda(\nu)]} = 1$. This implies that $E = \mathrm{ord}(\boldsymbol{\eta})$ from the assertion satisfies $E \mid [\widetilde{\Lambda}(\nu) : \Lambda(\nu)]$ and thus $E \le \det(\Lambda(\nu)) \le N^{2\nu^{1+r}}$ by (5-3).

To verify (iii) let us fix $v \in \mathbb{Z}^r \setminus \{0\}$ such that $\boldsymbol{\xi}'^v = 1$ and $|v| = \delta(\boldsymbol{\xi}')$. Then $\boldsymbol{\xi}^{V'v} = 1$ where $V' \in \mathrm{Mat}_{dr}(\mathbb{Z})$ consists of the final r columns of V. Raising to the E-th power to kill $\boldsymbol{\eta}$ yields $\boldsymbol{\zeta}^{EV'v} = 1$. Therefore, $EV'v \in \Lambda$. Note that $EV'v \notin \Lambda(\nu)$, indeed otherwise $V'v$ would lie in the saturation $\widetilde{\Lambda}(\nu)$. This is impossible as no nontrivial linear combination of columns of V' lies in $\widetilde{\Lambda}(\nu)$ which is generated by the first $\mathrm{rk}(\widetilde{\Lambda}(\nu))$ columns of V. Thus (5-2) implies $|EV'v|_2 \ge N^{\nu^r}$. By (5-8) we have

$$|EV'v| \ll_d E|V'||v| \ll_d E|V||v| \ll_d [\widetilde{\Lambda}(\nu) : \Lambda(\nu)] \det(\widetilde{\Lambda}(\nu))|v| = \det(\Lambda(\nu))|v|$$

we conclude $N^{\nu^r} \ll_d \det(\Lambda(\nu))|v|$. The determinant bound in (5-3) gives

$$\delta(\boldsymbol{\xi}') = |v| \gg_d N^{\nu^r - 2\nu^{1+r}} \gg_d N^{\nu^r/2}$$

and the last inequality used $\nu \le \frac{1}{4}$.

To complete the proof of (iii) let G be a subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$ where $M = \mathrm{ord}(\boldsymbol{\xi}) = \mathrm{ord}(\boldsymbol{\xi}')$. By Lemma 3.7 applied to $\boldsymbol{\xi}'$ there are $\sigma \in G$ and $a \in \mathbb{Z}^r$ with $\boldsymbol{\xi}' = \boldsymbol{e}(a\sigma/M)$, $|a| < M$, and

$$\frac{|a|}{M} \ll_d \frac{[(\mathbb{Z}/M\mathbb{Z})^\times : G]^{1/r} f_G^{1/(2r)}}{\delta(\boldsymbol{\xi}')^{1/(3r)}} \ll_d \frac{[(\mathbb{Z}/M\mathbb{Z})^\times : G] f_G^{1/2}}{N^{\nu^r/(6d)}}.$$

It remains to check (iv). Say $v \in \mathbb{Z}^d \backslash \{0\}$ with $\boldsymbol{\xi}^v = 1$ and $|v| = \delta(\boldsymbol{\xi})$. Then $\boldsymbol{\zeta}^v = \boldsymbol{\eta}^v \boldsymbol{\xi}^v = \boldsymbol{\eta}^v$. Thus $Ev \in \Lambda$ and there are two cases to consider. If $v \in \widetilde{\Lambda}(\nu)$, then $\sqrt{d}|v| \geq |v|_2 \geq \lambda_1(\widetilde{\Lambda}(\nu))$ by definition. Otherwise, $v \notin \widetilde{\Lambda}(\nu)$ in which case $Ev \notin \Lambda(\nu)$ by saturation. Here we can use (5-2) and the bound for E from (i) to conclude $|v|_2 \geq E^{-1} N^{\nu^r} \geq N^{\nu^r - 2\nu^{1+r}} \geq N^{\nu^r/2}$. So $|v| \geq |v|_2/\sqrt{d} \geq N^{\nu^r/2}/\sqrt{d} \geq N^{\nu^d/2}/\sqrt{d}$, as claimed in (iv). □

The situation simplifies in the following two cases. If $r = d$, then $\boldsymbol{\xi} = \boldsymbol{\zeta}$, $\boldsymbol{\eta} = 1$, $M = N$, $E = 1$, and V is the identity matrix. If N is a prime, then $E = 1$ as $E \mid N$ and $E \leq N^{2\nu^{1+r}} < N$ by part (i) above. Thus again $\boldsymbol{\xi} = \boldsymbol{\zeta}$ and $\boldsymbol{\eta} = 1$.

6. A preliminary result

Let $d \geq 1$ be an integer.

Definition 6.1. We use the convention $\inf \varnothing = \infty$. For $u \in \mathbb{Z}^d$ we define

$$\rho(u) = \inf\{|v| : v \in \mathbb{Z}^d \backslash \{0\} \text{ and } \langle u, v\rangle = 0\}, \tag{6-1}$$

as usual $|\cdot|$ is the maximum-norm on $\mathbb{R}^d$. Let $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$ be a Laurent polynomial. If the equation $P(\boldsymbol{\eta} z^u) = 0$ has no solution in $\boldsymbol{\eta} \in (\mu_\infty)^d$, $z \in S^1 \backslash \mu_\infty$, and $u \in \mathbb{Z}^d$, we set $\mathcal{B}(P) = 1$. Else wise we set

$$\mathcal{B}(P) = \inf\{B \in \mathbb{N} : \text{ if } \boldsymbol{\eta} \in (\mu_\infty)^d, z \in S^1 \backslash \mu_\infty \text{ is algebraic, and } u \in \mathbb{Z}^d \text{ with } P(\boldsymbol{\eta} z^u) = 0 \text{ then } \rho(u) \leq B\}.$$

Let us spell this out for $d = 1$. Then $\rho(u) = 1$ for $u = 0$ and $\rho(u) = \infty$ otherwise. If P vanishes at a point S^1 of infinite order, then $\mathcal{B}(P) = \infty$. Conversely, if P does not vanish at any point of $S^1 \backslash \mu_\infty$ then we have $\mathcal{B}(P) = 1$. In particular, if $d = 1$ and P is essentially atoral, then $\mathcal{B}(P) = 1$.

Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have order N and say $\nu \in \left(0, \frac{1}{4}\right]$. Below we make use of the canonically determined lattice $\Lambda_{\boldsymbol{\zeta}}(\nu)$ attached to $(\boldsymbol{\zeta}, \nu)$ as in Section 5. Recall that $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu))$ is the least positive Euclidean norm of a vector in the saturation of $\Lambda_{\boldsymbol{\zeta}}(\nu)$ in $\mathbb{Z}^d$. For technical reasons we work with

$$\tilde{\lambda}(\boldsymbol{\zeta}; \nu) = \min\{\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu)), N^{\nu^d/2}\}. \tag{6-2}$$

For example, if $\Lambda_{\boldsymbol{\zeta}}(\nu)$ is $\{0\}$, then the minimum equals $N^{\nu^d/2}$.

An important goal is to generalize Proposition 4.5 to multivariate polynomials. Proposition 6.2 below is a step in this direction.

Proposition 6.2. *Let $K \subset \mathbb{C}$ be a number field, $0 < \nu \le 1/(128d^2)$, and suppose $P \in K[X_1, \dots, X_d]\backslash\{0\}$ has at most k nonzero terms for an integer $k \ge 2$ and satisfies $\mathcal{B}(P) < \infty$. Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have order N and suppose G is a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ with $P(\boldsymbol{\zeta}^\sigma) \ne 0$ for all $\sigma \in G$. Then the following properties hold true with $r = d - \mathrm{rk}(\Lambda_{\boldsymbol{\zeta}}(\nu)) \ge 1$:*

(i) *If $d = 1$, then*

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^\sigma)| = m(P) + O_{d,k}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^2(1+h(P))}{N^{\nu^r/(20d)}}\right).$$

(ii) *If $d \ge 2$ and $\tilde{\lambda}(\boldsymbol{\zeta};\nu) > d^{1/2}\max\{\mathcal{B}(P), \deg P\}$, then*

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^\sigma)|$$
$$= m(P) + O_{d,k,\nu}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^2(1+h(P))}{N^{\nu^r/(20d)}} + \frac{\deg(P)^{16d^2}}{\tilde{\lambda}(\boldsymbol{\zeta};\nu)^{1/(16(k-1))}}\right). \quad (6\text{-}3)$$

Proof. We may assume that P is nonconstant. Part (i) follows with ample margin from Proposition 4.5 with $Q = P$ and $F = K$. Indeed, we require the standard estimate $d_0(N) \ll_\epsilon N^\epsilon$ which holds for all $\epsilon > 0$. We refrain from stating better bounds in (i) for the purpose of better comparability with the bounds in part (ii).

We split the proof of part (ii) up into 5 steps:

Step 1: Reduction to the univariate case. We write L for the fixed field of G in $\mathbb{Q}(\boldsymbol{\zeta})$. Note that G is the Galois group of $\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/L) = \mathrm{Gal}(L(\boldsymbol{\zeta})/L)$.

By Proposition 5.1 applied to $\boldsymbol{\zeta}$ we obtain $V \in \mathrm{GL}_d(\mathbb{Z})$ and a decomposition $\boldsymbol{\zeta} = \boldsymbol{\eta}\boldsymbol{\xi}$. Let $E = \mathrm{ord}(\boldsymbol{\eta})$ and $M = \mathrm{ord}(\boldsymbol{\xi})$. By (i) of Proposition 5.1 we find

$$E \le N^{2\nu^{1+r}} \quad \text{and thus} \quad M \ge N/E \ge N^{1-2\nu^{1+r}}. \quad (6\text{-}4)$$

The group used in Proposition 5.1(iii) is obtained as follows; we denote it with H to avoid a clash of notation with G from above. Let H be the subgroup of $(\mathbb{Z}/M\mathbb{Z})^\times$ corresponding to $\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\xi})/\mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta}))$. By Galois theory, see for example [Lang 2002, Theorem VI.1.12], the restriction homomorphism $\mathrm{Gal}(L(\boldsymbol{\xi})/L(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta})) \to \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\xi})/\mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta}))$ is an isomorphism. Using this isomorphism we will identify H with $\mathrm{Gal}(L(\boldsymbol{\xi})/L(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta}))$.

For future reference we estimate the conductor of $H \subset (\mathbb{Z}/M\mathbb{Z})^\times$. The fixed field of H in $\mathbb{Q}(\boldsymbol{\xi})$ is $\mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta})$. By the characterization of f_G, the field L is contained in $\mathbb{Q}(\boldsymbol{e}(1/f_G))$. So $\mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta}) \subset \mathbb{Q}(\boldsymbol{e}(1/M)) \cap \mathbb{Q}(\boldsymbol{e}(1/f_G), \boldsymbol{e}(1/E))$ since $\boldsymbol{\xi}$ has order M and $\boldsymbol{\eta}$ has order E. This final intersection is generated by a root of unity of order $\gcd(M, \mathrm{lcm}(f_G, E))$. We conclude

$$f_H \le \mathrm{lcm}(f_G, E) \le f_G E \le f_G N^{2\nu^{1+r}} \quad (6\text{-}5)$$

having used (6-4).

We use basic Galois theory to compute

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^{\sigma})| = \frac{1}{[L(\boldsymbol{\eta}):L]}\sum_{\tau\in\mathrm{Gal}(L(\boldsymbol{\eta})/L)}\ \sum_{\substack{\sigma\in\mathrm{Gal}(L(\boldsymbol{\xi})/L)\\ \tau|_{L(\boldsymbol{\eta})\cap L(\boldsymbol{\xi})}=\sigma|_{L(\boldsymbol{\eta})\cap L(\boldsymbol{\xi})}}} \frac{1}{\#H}\log|P(\boldsymbol{\eta}^{\tau}\boldsymbol{\xi}^{\sigma})|.$$

Observe that the inner sum is over a coset of $\tilde{\tau}H$ of H inside $(\mathbb{Z}/M\mathbb{Z})^{\times}$; here $\tilde{\tau}\in(\mathbb{Z}/M\mathbb{Z})^{\times}$ restricts to the restriction $\tau|_{L(\boldsymbol{\eta})\cap L(\boldsymbol{\xi})}$. Below, τ is as in the outer sum. The inner sum equals

$$S_{\tau} = \frac{1}{\#H}\sum_{\sigma\in\tilde{\tau}H}\log|P(\boldsymbol{\eta}^{\tau}\boldsymbol{\xi}^{\sigma})| = \frac{1}{\#H}\sum_{\sigma\in H}\log|P(\boldsymbol{\eta}^{\tau}\boldsymbol{\xi}^{\tilde{\tau}\sigma})|, \tag{6-6}$$

and the complete sum is

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^{\sigma})| = \frac{1}{[L(\boldsymbol{\eta}):L]}\sum_{\tau\in\mathrm{Gal}(L(\boldsymbol{\eta})/L)} S_{\tau}. \tag{6-7}$$

By Proposition 5.1(iii) applied to H we get $a\in\mathbb{Z}^r$ satisfying (5-7) and $\sigma_0\in H$ with

$$\boldsymbol{\xi}^{V} = (1,\dots,1,\boldsymbol{e}(a\sigma_0/M)).$$

We extend a to the left by $d-r$ zeros and obtain a row vector $(0,a)\in\mathbb{Z}^d$. We set $u=(0,a)V^{-1}\in\mathbb{Z}^d$ and use Proposition 5.1(ii) to get $\boldsymbol{\xi}=\boldsymbol{e}(u\sigma_0/M)$. Let us set

$$Q = P(\boldsymbol{\eta}^{\tau}X^{u})X^{l} \tag{6-8}$$

in the unknown X; it depends on τ and the exponent l is chosen to make sure that Q is a polynomial. So $0\neq|P(\boldsymbol{\eta}^{\tau}\boldsymbol{\xi}^{\tilde{\tau}\sigma})| = |Q(\boldsymbol{e}(\tilde{\tau}\sigma\sigma_0/M))|$ and in particular $Q\neq 0$. We may assume that $Q(0)\neq 0$. The coefficients of Q lie in $F=K(\boldsymbol{\eta})$ and Q has at most k nonzero terms as P has at most this many nonzero terms. All this allows us to rewrite (6-6) using a univariate polynomial, σ_0 above is absorbed by the sum

$$S_{\tau} = \frac{1}{\#H}\sum_{\sigma\in H}\log|Q(\boldsymbol{e}(\tilde{\tau}\sigma/M))|. \tag{6-9}$$

Step 2: Nonvanishing of Q on $S^1\backslash\mu_\infty$. Suppose $w\in\mathbb{Z}^d\backslash\{0\}$ satisfies $\langle u,w\rangle=0$ and $|w|=\rho(u)$. Recall that $\boldsymbol{\xi}=\boldsymbol{e}(u\sigma_0/M)$, so $\boldsymbol{\xi}^{w}=1$. Thus $|w|\geq\delta(\boldsymbol{\xi})$ and Proposition 5.1(iv) together with (6-2) yield

$$\rho(u) = |w| \geq d^{-1/2}\tilde{\lambda}(\boldsymbol{\zeta};\nu). \tag{6-10}$$

Let $z\in S^1\backslash\mu_\infty$ be algebraic. If $Q(z)=0$ then $P(\boldsymbol{\eta}^{\tau}z^{u})=0$ by (6-8). By Definition 6.1 we have $\rho(u)\leq\mathcal{B}(P)$. This and (6-10) contradict the lower bound $\tilde{\lambda}(\boldsymbol{\zeta};\nu)>d^{1/2}\mathcal{B}(P)$ in the hypothesis. Hence $Q(z)\neq 0$.

Thus Q, having algebraic coefficients, does not vanish at any point of $S^1\backslash\mu_\infty$. As $\rho(u)>1$ we also have $u\neq 0$.

Step 3: Bounding quantities in preparation for Proposition 4.5. This step is mainly bookkeeping. We aim to apply Proposition 4.5 to Q, the root of unity $\boldsymbol{e}(\tilde{\tau}/M)$, and the subgroup $H \subset (\mathbb{Z}/M\mathbb{Z})^\times$ to determine the asymptotic behavior of S_τ. To proceed we bound the various quantities below separately:

$$\begin{aligned}
[(\mathbb{Z}/M\mathbb{Z})^\times : H] &\le [(\mathbb{Z}/N\mathbb{Z})^\times : G] N^{2\nu^{1+r}},\\
f_H &\le f_G N^{2\nu^{1+r}},\\
\deg(Q) &\ll_d \deg(P) \min\{[(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} N^{1-\nu^r/(10d)}, N^2\},\\
h(Q) &= h(P),\\
[K(\boldsymbol{\eta}) : \mathbb{Q}] = [F : \mathbb{Q}] &\le [K : \mathbb{Q}] N^{2\nu^{1+r}}.
\end{aligned} \tag{6-11}$$

Note that $\#H = [\mathbb{Q}(\boldsymbol{\xi}) : \mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta})] = [\mathbb{Q}(\boldsymbol{\xi}) : \mathbb{Q}]/[\mathbb{Q}(\boldsymbol{\xi}) \cap L(\boldsymbol{\eta}) : \mathbb{Q}] \ge [\mathbb{Q}(\boldsymbol{\xi}) : \mathbb{Q}]/[L(\boldsymbol{\eta}) : \mathbb{Q}]$ and since $[\mathbb{Q}(\boldsymbol{\xi}) : \mathbb{Q}] = \#(\mathbb{Z}/M\mathbb{Z})^\times$ we find $[(\mathbb{Z}/M\mathbb{Z})^\times : H] \le [L(\boldsymbol{\eta}) : \mathbb{Q}] \le [L : \mathbb{Q}]E$. The first bound follows from (6-4) and as $[L : \mathbb{Q}] = [(\mathbb{Z}/N\mathbb{Z})^\times : G]$.

We already proved the bound for f_H in (6-5).

Next comes $\deg(Q)$. Observe that

$$\begin{aligned}
\deg(Q) &\ll_d |a||V^{-1}| \deg(P) \ll_d |a||V|^{d-1} \deg(P)\\
&\ll_d [(\mathbb{Z}/M\mathbb{Z})^\times : H] f_H^{1/2} \deg(P) N^{1+2(d-1)\nu^{1+r}-\nu^r/(6d)}\\
&\ll_d [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \deg(P) N^{1+2\nu^{1+r}+\nu^{1+r}+2(d-1)\nu^{1+r}-\nu^r/(6r)}
\end{aligned}$$

having used the bounds in Proposition 5.1, $M \le N$, and the first two bounds in (6-11). As $\nu \le 1/(128d^2)$ the exponent of N is at most $1 + (2d+1)\nu^{1+r} - \nu^r/(6d) \le 1 - \nu^r/(10d)$ and thus we obtain

$$\deg(Q) \ll_d [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \deg(P) N^{1-\nu^r/(10d)}$$

which is part of the third inequality in (6-11). The bound $\deg(Q) \ll_d \deg(P)N^2$ is proved similarly, but requires only the trivial estimate $|a| < M \le N$ from (5-7) and $|V^{-1}| \ll_d N^{2d\nu^{1+r}}$.

We claim that the coefficients of $P(\boldsymbol{\eta}^\tau X^u)$ are equal to the coefficients of P up to multiplication by a root of unity. In view of the definition of the height (2-4) this will imply the fourth claim in (6-11). Indeed, it suffices to rule out that two distinct monomials in P lead to the same power of X after the substitution. Hence it suffices to verify $\rho(u) > \deg P$. But this follows from (6-10) and as $\tilde{\lambda}(\boldsymbol{\zeta}; \nu) > d^{1/2} \deg P$ by hypothesis.

The degree of the number field F containing the coefficients of Q satisfies

$$[F : \mathbb{Q}] = [K(\boldsymbol{\eta}) : \mathbb{Q}] \le [K : \mathbb{Q}][\mathbb{Q}(\boldsymbol{\eta}) : \mathbb{Q}] \le [K : \mathbb{Q}]E \le [K : \mathbb{Q}] N^{2\nu^{1+r}}$$

where we used (6-4). This implies the fifth claim in (6-11).

Step 4: Applying Proposition 4.5 in the univariate case. Our aim is to determine the asymptotics of (6-9). We use the bounds from the last step to control the error term in (4-8) arise in Proposition 4.5

applied to F, Q, $\boldsymbol{e}(\tilde{\tau}/M)$ of order M, and to H. By (6-11) the error is

$$\ll [F:\mathbb{Q}]^2[(\mathbb{Z}/M\mathbb{Z})^\times : H] f_H^{1/2} \deg(Q)(\log(2\deg Q)+h(Q))\frac{(\log 2M)^3 d_0(M)}{M}$$

$$\ll_d [K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)(\log(2N^2\deg P)+h(P))N^{9\nu^{1+r}+1-\nu^{r/(10d)}}\frac{(\log 2M)^3 d_0(M)}{N}$$

where we use $\deg Q \ll N^2 \deg P$ to bound $\log(2\deg Q)$ from above and the lower bound for M in (6-4).

The exponent of N is $9\nu^{1+r}-\nu^r/(10d) \le -\nu^r/(19d)$ as $\nu \le 1/(128d^2) \le 1/(256d)$. As $M \mid N$ we find $d_0(M) \le d_0(N)$. As in the proof of (i) we use $d_0(N) \ll_\epsilon N^\epsilon$ for all ϵ. We also anticipate $\log(2N^2)$ coming from $\log(2N^2\deg P)$ to find

$$\log(2N^2)N^{9\nu^{1+r}-\nu^{r/(10d)}}(\log 2M)^3 d_0(M) \ll_{d,\nu} N^{-\nu^r/(20d)}.$$

Using the crude inequality $\log\deg P \le \deg P$ the error term is thus

$$\ll_{d,\nu} [K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^2(1+h(P))N^{-\nu^r/(20d)}.$$

Applying Proposition 4.5 and recalling $m(Q)=m(P(\boldsymbol{\eta}^\tau X^u))$ we find

$$S_\tau = m(P(\boldsymbol{\eta}^\tau X^u)) + O_{d,\nu}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^2(1+h(P))}{N^{\nu^r/(20d)}}\right). \tag{6-12}$$

Step 5: Applying a quantitative version of Lawton's theorem. To determine the asymptotics of the Mahler measure we apply our quantitative variant of Lawton's theorem, Theorem A.1 to $P(\boldsymbol{\eta}^\tau(X_1,\dots,X_d)) \neq 0$ and u. This polynomial has the same degree and number of terms as P. The exponent vector satisfies $\rho(u) \ge d^{-1/2}\tilde{\lambda}(\boldsymbol{\zeta};\nu)$ by (6-10). Our hypothesis implies $\rho(u) > \deg P$, as required by Theorem A.1. We find

$$m(P(\boldsymbol{\eta}^\tau X^u)) = m(P(\boldsymbol{\eta}^\tau(X_1,\dots,X_d))) + O_{d,k}\left(\frac{\deg(P)^{16d^2}}{\tilde{\lambda}(\boldsymbol{\zeta};\nu)^{1/(16(k-1))}}\right). \tag{6-13}$$

The Mahler measure of P and $P(\boldsymbol{\eta}^\tau(X_1,\dots,X_d))$ are equal as translating by $\boldsymbol{\eta}^\tau \in (S^1)^d$ does not affect the value of the integral (1-4).

By combining (6-12) and (6-13) we conclude

$$S_\tau = m(P) + O_{d,k,\nu}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^2(1+h(P))}{N^{\nu^r/(20d)}} + \frac{\deg(P)^{16d^2}}{\tilde{\lambda}(\boldsymbol{\zeta};\nu)^{1/(16(k-1))}}\right).$$

Part (ii) of the proposition follows from (6-7). □

We now explain why the situation simplifies when the order N of $\boldsymbol{\zeta}$ is a prime number. In this case, after the proof of Proposition 5.1 we observed that $\boldsymbol{\eta}=1$ and $\boldsymbol{\zeta}=\boldsymbol{\xi}$. In the proof above, inequality (6-10) can be replaced by $\rho(u) \ge \delta(\boldsymbol{\zeta})$. So the hypothesis on $\boldsymbol{\zeta}$ in (ii) of the proposition can be replaced by $\delta(\boldsymbol{\zeta}) > \max\{\mathcal{B}(P), \deg P\}$; see also the argument near (6-13). This is certainly satisfied for $\delta(\boldsymbol{\zeta}) \to \infty$. Moreover, $\tilde{\lambda}(\boldsymbol{\zeta};\nu)$ can be replaced by $\delta(\boldsymbol{\zeta})$ in (6-3). From this point it is not difficult to deduce Theorem 1.1 when N is a prime.

The remaining argument is required to treat general N. We need to keep track of extra information such as $[K:\mathbb{Q}]$, $[(\mathbb{Z}/N\mathbb{Z})^\times : G]$, f_G, and the dependency on P to anticipate a monomial change of coordinates.

7. Equidistribution

Proposition 6.2 closes in on Theorem 1.1. Indeed, suppose that for some choice of ν the value $\tilde{\lambda}(\boldsymbol{\zeta}; \nu)$ grows polynomially in $\delta(\boldsymbol{\zeta})$. Then the error term of (6-3) tends to 0 as $\delta(\boldsymbol{\zeta}) \to \infty$ and we are done.

However, consider the following example, already found in the beginning of Section 5. Suppose $n \geq 2$ and $\boldsymbol{\zeta}_p$ and $\boldsymbol{\zeta}_{p^n}$ are roots of unity of order p and p^n, respectively. Say $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_p, \boldsymbol{\zeta}_{p^n})$, it has order p^n. The lattice $\Lambda_{\boldsymbol{\zeta}}$ contains $(p, 0)^t$ and this vector has minimal positive Euclidean norm in $\Lambda_{\boldsymbol{\zeta}}$. For n large enough in terms of ν we have $\Lambda(\nu) = (p, 0)^t\mathbb{Z}$ and $\widetilde{\Lambda}(\nu) = (1, 0)^t\mathbb{Z}$. Thus $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu)) = 1$ and this yields $\tilde{\lambda}(\boldsymbol{\zeta}; \nu) = 1$.

This example suggests a monomial change of coordinates which we will do in the next section. In the current section we lay the groundwork for this change of coordinates.

7A. ***Numerical integration.*** We require a higher dimensional replacement of the Koksma bound [Harman 1998, Theorem 5.4]. The classical analog is called the Koksma–Hlawka Inequality and applies to functions of bounded variation in the sense of Hardy and Krause. Let $\theta : U \to \mathbb{R}$ be a function whose domain U is a nonempty subset of $\mathbb{R}^d$. In this subsection we use the *modulus of continuity* of θ defined by

$$\omega(\theta; t) = \sup_{\substack{x, y \in U \\ |x-y| \leq t}} |\theta(x) - \theta(y)| \tag{7-1}$$

for all $t \geq 0$; as usual $|\cdot|$ denotes the maximum-norm on $\mathbb{R}^d$. We define $\omega(\theta; t) = 0$ if $U = \varnothing$. We will use it to estimate a mean in terms of the corresponding integral in Proposition 7.1. Hlawka [1971] has a related and more precise result. For the reader's convenience we give a self-contained treatment that suffices for our purposes.

Proposition 7.1. *Let $\theta : [0, 1]^d \to \mathbb{R}$ be a continuous function and let $x_1, \ldots, x_n \in [0, 1)^d$ with discrepancy $\mathcal{D} = \mathcal{D}(x_1, \ldots, x_n)$. Then*

$$\left| \frac{1}{n} \sum_{i=1}^{n} \theta(x_i) - \int_{[0,1)^d} \theta(x)\, dx \right| \leq (1 + 2^{d+1})\omega(\theta, \mathcal{D}^{1/(d+1)}). \tag{7-2}$$

Proof. Both sides of (7-2) are invariant under adding a constant function to θ. So we may assume $\theta(0) = 0$.

Let $T \geq 1$ be an integral parameter to be determined below. We write $[0, 1)^d$ as a disjoint union of T^d half-open hypercubes Q_j with side length $1/T$. Let $\overline{Q}_j$ denote the closure of Q_j in $[0, 1]^d$. The mean value theorem tells us that for each j there exists $y_j \in \overline{Q}_j$ such that $\int_{Q_j} \theta(x)\, dx = \mathrm{vol}(Q_j)\theta(y_j) = T^{-d}\theta(y_j)$.

For each j we write $n_j = \#\{i \in \{1, \ldots, n\} : x_i \in Q_j\}$. So

$$\frac{1}{n}\left| \sum_{i=1}^{n} \theta(x_i) - \sum_j n_j \theta(y_j) \right| \leq \frac{1}{n} \sum_j \sum_{\substack{i=1 \\ x_i \in Q_j}}^{n} |\theta(x_i) - \theta(y_j)| \leq \frac{1}{n} \sum_j \omega(\theta; 1/T) n_j = \omega(\theta; 1/T). \tag{7-3}$$

On the other hand, $\frac{1}{n}\sum_j n_j\theta(y_j)$ equals

$$\sum_j \frac{n_j}{n} T^d \int_{Q_j} \theta(x)\,dx = \sum_j (1+\delta_j T^d) \int_{Q_j} \theta(x)\,dx = \int_{[0,1)^d} \theta(x)\,dx + T^d \sum_j \delta_j \int_{Q_j} \theta(x)\,dx$$

where $\delta_j = n_j/n - T^{-d}$. The definition of discrepancy implies $|\delta_j| \le \mathcal{D}$. Hence

$$\left| \frac{1}{n}\sum_j n_j\theta(y_j) - \int_{[0,1)^d} \theta(x)\,dx \right| \le T^d \mathcal{D} \int_{[0,1)^d} |\theta(x)|\,dx \le T^{d+1}\mathcal{D}\omega(\theta; 1/T) \tag{7-4}$$

where we used $|\theta(x)| \le T\omega(\theta; 1/T)$ for all $x \in [0,1]^d$; recall that $\theta(0) = 0$.

We apply the triangle inequality to (7-3) and (7-4) and conclude that the left-hand side of (7-2) is at most $(1 + T^{d+1}\mathcal{D})\omega(\theta; 1/T)$. To complete the proof observe that $0 < \mathcal{D} \le 1$ and fix $T = \lceil \mathcal{D}^{-1/(d+1)} \rceil$ which satisfies $\mathcal{D}^{-1/(d+1)} \le T \le \mathcal{D}^{-1/(d+1)} + 1$. □

7B. ***Averaging the Mahler measure.*** This subsection is purely in the complex setting. Let $P \in \mathbb{C}[X_1, \dots, X_d]\backslash\{0\}$ have at most $k \ge 2$ nonzero terms, where k is an integer.

Let $l \in \{1, \dots, d-1\}$. For $x \in \mathbb{R}^l$ we define $P_{\boldsymbol{e}(x)} = P(\boldsymbol{e}(x), X_1, \dots, X_{d-l}) \in \mathbb{C}[X_1, \dots, X_{d-l}]$. Next we construct an auxiliary Laurent polynomial $\widehat{P}$ in l variables whose value at $\boldsymbol{e}(x)$ is comparable to $|P_{\boldsymbol{e}(x)}|$. For $i \in \mathbb{Z}^{d-l}$ we denote $p_i \in \mathbb{C}[X_1, \dots, X_l]$ the coefficients of P, taken as a Laurent polynomial in $X_{l+1}, \dots, X_d$, and define

$$\widehat{P} = \sum_i p_i(X_1, \dots, X_l)\,\overline{p_i}(X_1^{-1}, \dots, X_l^{-1}) \in \mathbb{C}[X_1^{\pm 1}, \dots, X_l^{\pm 1}] \tag{7-5}$$

where the bar denotes complex conjugation.

Lemma 7.2. *In the notation above the following properties hold true*:

(i) *The Laurent polynomial $\widehat{P}$ has at most k^2 nonzero terms.*

(ii) *The product $(X_1 \cdots X_l)^{\deg P}\,\widehat{P}$ is a polynomial of degree at most $(l+1)\deg P$.*

Proof. If each p_i consists of k_i nonzero terms, then $\widehat{P}$ consists of at most $\sum_i k_i^2$ terms. Since $\sum_i k_i \le k$ we find that $\widehat{P}$ has at most k^2 nonzero terms. This implies part (i).

Part (ii) follows from (7-5). □

Observe that $\widehat{P}(\boldsymbol{e}(x)) = \sum_i |p_i(\boldsymbol{e}(x))|^2 \ge 0$. As $|P_{\boldsymbol{e}(x)}|$ is the maximum of $|p_i(\boldsymbol{e}(x))|$ as i varies, we find

$$\frac{1}{k^{1/2}}\widehat{P}(\boldsymbol{e}(x))^{1/2} \le |P_{\boldsymbol{e}(x)}| \le \widehat{P}(\boldsymbol{e}(x))^{1/2}. \tag{7-6}$$

So $P_{\boldsymbol{e}(x)} = 0$ if and only if $\widehat{P}(\boldsymbol{e}(x)) = 0$.

The main result of this subsection is:

Proposition 7.3. *Assume $P \in \mathbb{C}[X_1, \dots, X_d]\backslash\mathbb{C}$ has at most k nonzero terms for an integer $k \geq 2$. Let $l \in \{1, \dots, d-1\}$ and let $\widehat{P}$ be as above. Suppose $x_1, \dots, x_n \in [0,1)^l$ with discrepancy $\mathcal{D} = \mathcal{D}(x_1, \dots, x_n)$. If $P_{\boldsymbol{e}(x_i)} \neq 0$ for all $i \in \{1, \dots, n\}$, then*

$$\frac{1}{n}\sum_{i=1}^{n} m(P_{\boldsymbol{e}(x_i)}) = m(P) + O_{d,k}\left(\deg(P)\mathcal{D}^{1/(16(d+1)k^2)} + \left|m(\widehat{P}) - \frac{1}{n}\sum_{i=1}^{n}\log\widehat{P}(\boldsymbol{e}(x_i))\right|\right). \tag{7-7}$$

By a theorem of Boyd [1998], the Mahler measure is a continuous function in the coefficients of a nonzero polynomial of fixed degree (below in Lemma A.5 we prove that it is even Hölder continuous). Therefore, if the $P_{\boldsymbol{e}(x_i)}$ in the proposition above are uniformly bounded away from 0, then the average on the left in (7-7) converges to the integral $\int_{[0,1)^l} m(P_{\boldsymbol{e}(x)})\,dx$ as the discrepancy tends to 0. But even when $|P| = 1$ it is conceivable that $|P_{\boldsymbol{e}(x_i)}|$ is small for some x_i, then $P_{\boldsymbol{e}(x_i)}$ is near the Mahler measure's logarithmic singularity. This happens if and only if $\widehat{P}(\boldsymbol{e}(x_i))$ is small by (7-6). The proposition states that we can handle the mean for arbitrary x_i if we can control the logarithmic mean of $\widehat{P}$ over the $\boldsymbol{e}(x_i)$.

The proof follows a series of lemmas. We first note a useful property of the modulus of continuity as defined in (7-1). Let $\theta : [0,1]^d \to \mathbb{R}\cup\{-\infty\}$ be a function and $c \in \mathbb{R}$, such that $\theta_c(x) = \max\{c, \theta(x)\}$ defines a continuous function $[0,1]^d \to \mathbb{R}$. We claim that

$$\omega(\theta_c; t) \leq \omega(\theta|_{\theta^{-1}((c,\infty))}; t) \quad \text{for all } t \geq 0. \tag{7-8}$$

This inequality follows by definition if $\theta^{-1}((c,\infty))$ is empty. Say $x, y \in [0,1]^d$ with $|x-y| \leq t$. To bound $|\theta_c(x) - \theta_c(y)|$ from above by the right-hand side of (7-8) we may assume $\theta_c(x) > c = \theta_c(y)$. By continuity of θ_c there is for all small enough $\epsilon > 0$ a $z \in [0,1]^d$ on the line segment connecting x and y with $c+\epsilon = \theta_c(z) = \theta(z)$. Then $|\theta_c(x)-\theta_c(y)| = |\theta(x)-c| \leq |\theta(x)-\theta(z)| + |\theta(z)-c| \leq \omega(\theta|_{\theta^{-1}((c,\infty))}; t) + \epsilon$. Our claim (7-8) follows as ϵ can be made arbitrarily small.

Let P and k be as in Proposition 7.3 and assume in addition that $|P| = 1$.

Lemma 7.4. *let $x_1, \dots, x_n \in [0,1)^d$ have discrepancy $\mathcal{D} = \mathcal{D}(x_1, \dots, x_n)$. If $r \in (0,1]$, then*

$$\frac{1}{n}\#\{i \in \{1, \dots, n\} : |P(\boldsymbol{e}(x_i))| \leq r\} \ll_{d,k} r^{1/(2k)} + \deg(P)\mathcal{D}^{1/(d+1)}/r. \tag{7-9}$$

Proof. For $x \in [0,1]^d$ we set

$$\chi(x) = \max\{0, 2 - |P(\boldsymbol{e}(x))|/r\}$$

and this defines a continuous function on $[0,1]^d$ with values in $[0,2]$.

We note that $\chi(x) \geq 1$ if $|P(\boldsymbol{e}(x))| \leq r$. As χ is nonnegative the average $\frac{1}{n}\sum_{i=1}^{n}\chi(x_i)$ is at least the proportion of the i among $\{1, \dots, n\}$ such that $|P(\boldsymbol{e}(x_i))| \leq r$. On the other hand, Lemma A.3(i) implies

$$\int_{[0,1)^d} \chi(x)\,dx \leq 2\mathrm{vol}(\{x \in [0,1)^d : |P(\boldsymbol{e}(x))| < 2r\}) \ll_{d,k} r^{1/(2(k-1))} \ll_{d,k} r^{1/(2k)}. \tag{7-10}$$

We will apply Proposition 7.1 to bound the proportion on the left in (7-9). Say $t > 0$, let us verify

$$\omega(\chi; t) \ll_{d,k} \deg(P)t/r. \tag{7-11}$$

We apply (7-8) to $\theta(x) = 2 - |P(\boldsymbol{e}(x))|/r$ and $c = 0$. Say $x, y \in \theta^{-1}((0,\infty))$ with $|x-y| \le t$, so in particular $|P(\boldsymbol{e}(x))| < 2r$ and $|P(\boldsymbol{e}(y))| < 2r$. Then $|\theta(x)-\theta(y)| = |P(\boldsymbol{e}(x)) - P(\boldsymbol{e}(y))|/r \ll_{d,k} \deg(P)t/r$, where we used $|x-y| \le t$ and $|P| = 1$. We obtain (7-11).

Let us set $t = \mathcal{D}^{1/(d+1)}$. We apply numerical integration, Proposition 7.1, and use (7-10) to conclude the proof. □

In the next lemma we truncate the singularity of $x \mapsto \log|P(\boldsymbol{e}(x))|$ using a parameter r and bound the modulus of continuity of the resulting function.

Lemma 7.5. *Let $r \in (0,1]$, for $x \in [0,1]^d$ we define $\psi(x) = \max\{\log r, \log|P(\boldsymbol{e}(x))|\}$ as above* (7-8). *Then $\psi : [0,1]^d \to \mathbb{R}$ is continuous and for all $t > 0$ we have*

$$\omega(\psi; t) \ll_{d,k} \frac{\deg(P)t}{r}.$$

Proof. Clearly, ψ is continuous on $[0,1]^d$. We apply (7-8) to $\theta(x) = \log|P(\boldsymbol{e}(x))|$ and $c = \log r$. Say $x, y \in [0,1]^d$ with $|P(\boldsymbol{e}(x))| \ge |P(\boldsymbol{e}(y))| \ge r$ and $|x-y| \le t$. Then as in the proof of Lemma 7.4 we find $\big||P(\boldsymbol{e}(x))/P(\boldsymbol{e}(y))| - 1\big| \ll_{d,k} \deg(P)t/|P(\boldsymbol{e}(y))| \ll_{d,k} \deg(P)t/r$. Applying the logarithm and using $0 \le \log s \le s-1$ for all $s \ge 1$ yields

$$\big|\log|P(\boldsymbol{e}(x))| - \log|P(\boldsymbol{e}(y))|\big| \ll_{d,k} \frac{\deg(P)t}{r},$$

as desired. □

Lemma 7.6. *We keep the notation of Lemma 7.5. Then*

$$\left| m(P) - \int_{[0,1)^d} \psi(x)\, dx \right| \ll_{d,k} r^{1/(4k)}.$$

Proof. The absolute value in question is

$$\mathcal{E} = \left| \int_\Sigma \log|P(\boldsymbol{e}(x))|\, dx - \mathrm{vol}(\Sigma)\log r \right|$$

where $\Sigma = S(P,r) = \{x \in [0,1)^d : |P(\boldsymbol{e}(x))| < r\}$ in the notation of (A-2). Hence $\mathrm{vol}(\Sigma) \ll_{d,k} r^{1/(2(k-1))}$ by Lemma A.3(i). So

$$\mathcal{E} \ll_{d,k} \int_\Sigma \big|\log|P(\boldsymbol{e}(x))|\big|\, dx + r^{1/(2k)}$$

as $r \le 1$. To bound the final integral we use Lemma A.4 which implies $\mathcal{E} \ll_{d,k} r^{1/(4(k-1))} + r^{1/(2k)} \ll_{d,k} r^{1/(4k)}$. □

Lemma 7.7. *Let $x_1, \dots, x_n \in [0,1)^d$ with $P(\boldsymbol{e}(x_i)) \ne 0$ for all i and discrepancy $\mathcal{D} = \mathcal{D}(x_1, \dots, x_n)$. We set*

$$\epsilon = \left| m(P) - \frac{1}{n}\sum_{i=1}^{n} \log|P(\boldsymbol{e}(x_i))| \right|.$$

If $r \in (0, 1]$, *then*

$$\frac{1}{n} \sum_{|P(\boldsymbol{e}(x_i))|<r} \big|\log|P(\boldsymbol{e}(x_i))|\big| \ll_{d,k} \deg(P)\mathcal{D}^{1/(d+1)}r^{-2} + r^{1/(4k)} + \epsilon.$$

Proof. By the triangle inequality and with ψ as in Lemma 7.5 we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\psi(x_i) - \log|P(\boldsymbol{e}(x_i))|\right| \le \left|\frac{1}{n}\sum_{i=1}^{n}\psi(x_i) - \int_{[0,1)^d}\psi(x)\,dx\right| + \left|\int_{[0,1)^d}\psi(x)\,dx - m(P)\right| + \epsilon.$$

We use Proposition 7.1 and Lemma 7.5 with $t = \mathcal{D}^{1/(d+1)}$ to bound the first term on the right by $\ll_{d,k} \deg(P)\mathcal{D}^{1/(d+1)}/r$. The second term is $\ll_{d,k} r^{1/(4k)}$ by Lemma 7.6.

The term on the left equals $\frac{1}{n}\sum_{|P(\boldsymbol{e}(x_i))|<r}(\log r - \log|P(\boldsymbol{e}(x_i))|)$. Observe that $-\log|P(\boldsymbol{e}(x_i))| = \big|\log|P(\boldsymbol{e}(x_i))|\big|$ in this sum as $r \le 1$. We rearrange and find

$$\frac{1}{n} \sum_{|P(\boldsymbol{e}(x_i))|<r} \big|\log|P(\boldsymbol{e}(x_i))|\big| \ll_{d,k} \deg(P)\mathcal{D}^{1/(d+1)}r^{-1} + r^{1/(4k)} + \frac{|\log r|}{n}\left(\sum_{|P(\boldsymbol{e}(x_i))|<r} 1\right) + \epsilon.$$

By Lemma 7.4, the term corresponding to the sum over i on the right is

$$\ll_{d,k} r^{1/(2k)}|\log r| + \deg(P)\mathcal{D}^{1/(d+1)}r^{-1}|\log r|.$$

Combining our bounds and absorbing $|\log r|$ in an appropriate power of r^{-1} we find

$$\frac{1}{n} \sum_{|P(\boldsymbol{e}(x_i))|<r} \big|\log|P(\boldsymbol{e}(x_i))|\big| \ll_{d,k} \deg(P)\mathcal{D}^{1/(d+1)}r^{-2} + r^{1/(4k)} + \epsilon,$$

as desired. □

After this warming-up we prove variants of Lemmas 7.5 and 7.6 where $\log|\cdot|$ is replaced by the Mahler measure. We also truncate at the parameter r.

Lemma 7.8. *Let* $r \in (0, 1]$, *for* $x \in [0, 1]^l$ *we define* $\mu(x) = \max\{\log r, m(P_{\boldsymbol{e}(x)})\}$ *as above* (7-8) *where we interpret the Mahler measure of* 0 *as* $-\infty$. *Then* $\mu : [0, 1]^l \to \mathbb{R}$ *is continuous and for all* $t > 0$ *we have*

$$\omega(\mu; t) \ll_{d,k} \left(\frac{\deg(P)t}{r}\right)^{1/(8k)} (1 + |\log r|).$$

Proof. By Boyd's theorem [1998] the Mahler measure is continuous on the space of nonzero polynomials of bounded degree. Thus μ is continuous on $[0, 1]^l$. Observe that $\omega(\mu; t) \ll_k 1 + |\log r|$ as $m(P_{\boldsymbol{e}(x)}) \ll_k 1$ by (2-1) and $|P| = 1$. So we may assume that $\deg(P)t/r$ is sufficiently small in terms of d and k.

We again use (7-8), this time with $\theta(x) = m(P_{\boldsymbol{e}(x)})$ and $c = \log r$. Let $x, y \in [0, 1]^d$ with $m(P_{\boldsymbol{e}(x)}) \ge \log r$ and $m(P_{\boldsymbol{e}(y)}) \ge \log r$ and $|x - y| \le t$. Then $|P_{\boldsymbol{e}(x)}| \gg_k r$ and $|P_{\boldsymbol{e}(y)}| \gg_k r$ by (2-1). As in the proof of Lemma 7.4 we find $|P_{\boldsymbol{e}(x)} - P_{\boldsymbol{e}(y)}| \ll_{d,k} \deg(P)t$. Since $\deg(P)t/r$ is smaller than some prescribed constant depending only on d and k we may assume $|P_{\boldsymbol{e}(x)} - P_{\boldsymbol{e}(y)}|/\min\{|P_{\boldsymbol{e}(x)}|, |P_{\boldsymbol{e}(y)}|\} \le \frac{1}{2}$. Lemma A.5

implies

$$|m(P_{\boldsymbol{e}(x)}) - m(P_{\boldsymbol{e}(y)})| \ll_{d,k} \left(\frac{|P_{\boldsymbol{e}(x)} - P_{\boldsymbol{e}(y)}|}{\min\{|P_{\boldsymbol{e}(x)}|, |P_{\boldsymbol{e}(y)}|\}} \right)^{1/(8(k-1))} \ll_{d,k} \left(\frac{\deg(P)t}{r} \right)^{1/(8(k-1))},$$

as desired. □

Before continuing we recall the p_i and the auxiliary Laurent polynomial $\widehat{P}$ determined by P and l in (7-5). By Lemma 7.2(i) $\widehat{P}$ has at most k^2 nonzero terms, so $\sup_{x\in[0,1]^l} |\widehat{P}(\boldsymbol{e}(x))| \le k^2|\widehat{P}|$. There exists i with $|p_i| = |P| = 1$. The definition of the Mahler measure implies

$$m(p_i) \le \sup_{x\in[0,1]^l} \log|p_i(\boldsymbol{e}(x))| \le \tfrac{1}{2} \sup_{x\in[0,1]^l} \log|\widehat{P}(\boldsymbol{e}(x))| \le \tfrac{1}{2}(2\log k + \log|\widehat{P}|).$$

Using $|p_i| = 1$ and the theorem of Dobrowolski and Smyth, Theorem 2.1, we conclude $m(p_i) \ge -(k-2)\log 2$. Thus $|\widehat{P}| \gg_k 1$. Bounding $|\widehat{P}|$ from above is more straight-forward. Indeed, $|\widehat{P}| \ll_k 1$ by (7-5) and since $|P| = 1$. Therefore,

$$1 \ll_k |\widehat{P}| \ll_k 1. \tag{7-12}$$

We let $\widetilde{P}$ denote the polynomial from Lemma 7.2(ii) divided by $|\widehat{P}|$, so $|\widetilde{P}| = 1$.

Lemma 7.9. *We keep the notation of Lemma 7.8. Then*

$$\left| m(P) - \int_{[0,1)^l} \mu(x)\, dx \right| \ll_{d,k} r^{1/(2k^2)}.$$

Proof. We recall that $|\widehat{P}|\widetilde{P}$ equals $\widehat{P}$ up to a monomial factor. By (7-6), (7-12), and Theorem 2.1 there exists $c > 0$ depending only on k such that $|\widetilde{P}(\boldsymbol{e}(x))| \ge cr^2$ implies $m(P_{\boldsymbol{e}(x)}) \ge \log r$. By Fubini's theorem we have $\int_{[0,1)^l} m(P_{\boldsymbol{e}(x)})\, dx = m(P)$, so the absolute value in question is

$$\mathcal{E} = \left| \int_\Sigma m(P_{\boldsymbol{e}(x)})\, dx - \mathrm{vol}(\Sigma)\log r \right|$$

where $\Sigma = S(\widetilde{P}, cr^2)$; indeed $m(P_{\boldsymbol{e}(x)}) = \mu(x)$ for all $x \in [0,1]^l \setminus \Sigma$.

Note that $\mathrm{vol}(\Sigma) \ll_{d,k} r^{1/(k^2-1)}$ by Lemma A.3(i) applied to $\widetilde{P}$. So

$$\mathcal{E} \ll_{d,k} r^{1/(k^2-1)}|\log r| + \left| \int_\Sigma m(P_{\boldsymbol{e}(x)})\, dx \right| \ll_{d,k} r^{1/k^2} + \int_\Sigma |m(P_{\boldsymbol{e}(x)})|\, dx. \tag{7-13}$$

To bound the integral in (7-13) from above we will replace $m(P_{\boldsymbol{e}(x)})$ by $\log|P_{\boldsymbol{e}(x)}|$. Say $x \in \Sigma$ and $P_{\boldsymbol{e}(x)} \ne 0$, then $\big|m(P_{\boldsymbol{e}(x)}) - \log|P_{\boldsymbol{e}(x)}|\big| \ll_k 1$ by (2-2) and thus $|m(P_{\boldsymbol{e}(x)})| \ll_k 1 + \big|\log|P_{\boldsymbol{e}(x)}|\big|$. The function $x \mapsto \big|\log|P_{\boldsymbol{e}(x)}|\big|$ is integrable over $[0,1)^l$ in the sense of Lebesgue and so is $x \mapsto |m(P_{\boldsymbol{e}(x)})|$; both take the value $+\infty$ on a measure zero subset of $[0,1)^l$. We find

$$\mathcal{E} \ll_{d,k} r^{1/k^2} + \int_\Sigma \big(1 + \big|\log|P_{\boldsymbol{e}(x)}|\big|\big)\, dx.$$

From (7-6) and (7-12) we deduce $\big|\log|P_{\boldsymbol{e}(x)}|\big| \ll_k \big|\log|\widetilde{P}(\boldsymbol{e}(x))|\big| + 1$ if $\widehat{P}(\boldsymbol{e}(x)) \neq 0$. So $\mathcal{E} \ll_{d,k} r^{1/k^2} + \int_\Sigma \big(1 + \big|\log|\widetilde{P}(\boldsymbol{e}(x))|\big|\big)\, dx$. By Lemma A.4 applied to $\widetilde{P}$ and the volume estimate for Σ, the integral on the right is $\ll_{d,k} r^{1/(k^2-1)} + r^{1/(2(k^2-1))} \ll_{d,k} r^{1/(2k^2)}$, as desired. □

Proof of Proposition 7.3. If we scale P by a factor λ, then $\widehat{P}$, defined in (7-5), is scaled by $|\lambda|^2$. So the proposition is invariant under nonzero scaling and we may assume $|P| = 1$. Later on we will choose the parameter r in terms of $\deg(P)$ and $\mathcal{D}$. In the meantime we assume that $r \in \big(0, \frac{1}{2}\big]$.

We want to bound $\mathcal{E} = |m(P) - n^{-1}\sum_{i=1}^n m(P_{\boldsymbol{e}(x_i)})|$ from above. We replace the Mahler measure with $\mu(\cdot)$ coming from Lemma 7.8. Indeed, the triangle inequality implies

$$\mathcal{E} \le \left| m(P) - \int_{[0,1)^l} \mu(x)\, dx \right| + \left| \int_{[0,1)^l} \mu(x)\, dx - \frac{1}{n}\sum_{i=1}^n \mu(x_i) \right| + \left| \frac{1}{n}\sum_{i=1}^n \mu(x_i) - m(P_{\boldsymbol{e}(x_i)}) \right|.$$

The first term on the right is $\ll_{d,k} r^{1/(2k^2)}$ by Lemma 7.9 applied to P. By Proposition 7.1 applied to μ and $t = \mathcal{D}^{1/(d+1)}$ and Lemma 7.8 the second term is $\ll_{d,k} (\deg(P)\mathcal{D}^{1/(d+1)} r^{-1})^{1/(8k)} |\log r|$. So

$$\mathcal{E} \ll_{d,k} r^{1/(2k^2)} + (\deg(P)\mathcal{D}^{1/(d+1)} r^{-2})^{1/(8k)} + \mathcal{E}' \tag{7-14}$$

after absorbing $|\log r|$ in a multiple $r^{-1/(8k)}$ and where $\mathcal{E}'$ is the third term above. Only terms with $m(P_{\boldsymbol{e}(x_i)}) \le \log r$ contribute to the average, so $\mathcal{E}'$ equals

$$\left| \frac{1}{n} \sum_{m(P_{\boldsymbol{e}(x_i)}) \le \log r} \log r - m(P_{\boldsymbol{e}(x_i)}) \right| \le \frac{|\log r|}{n} \#\{i : m(P_{\boldsymbol{e}(x_i)}) \le \log r\} + \frac{1}{n} \sum_{m(P_{\boldsymbol{e}(x_i)}) \le \log r} |m(P_{\boldsymbol{e}(x_i)})|.$$

By Theorem 2.1 we may replace $m(P_{\boldsymbol{e}(x_i)})$ by $\log|P_{\boldsymbol{e}(x_i)}|$ at the cost of introducing a constant $c_1 > 0$ depending only on k, i.e.,

$$\mathcal{E}' \ll_{d,k} \frac{|\log r|}{n} \#\{i : |P_{\boldsymbol{e}(x_i)}| \le c_1 r\} + \frac{1}{n} \sum_{|P_{\boldsymbol{e}(x_i)}| \le c_1 r} \big(1 + \big|\log|P_{\boldsymbol{e}(x_i)}|\big|\big).$$

Recall that $\widetilde{P}$ was defined after the proof of Lemma 7.8. If $|P_{\boldsymbol{e}(x_i)}| \le c_1 r$, then $|\widetilde{P}(\boldsymbol{e}(x_i))| = |\widehat{P}(\boldsymbol{e}(x_i))|/|\widehat{P}| \le c_2 r^2$ for some c_2 depending only on k by (7-6) and (7-12). The same inequalities imply $\big|\log|P_{\boldsymbol{e}(x_i)}|\big| \ll_k \big|\log|\widetilde{P}(\boldsymbol{e}(x_i))|\big| + 1$, the "+1" is absorbed in the first term in

$$\mathcal{E}' \ll_{d,k} \frac{|\log r|}{n} \#\{i : |\widetilde{P}(\boldsymbol{e}(x_i))| \le c_2 r^2\} + \frac{1}{n} \sum_{|\widetilde{P}(\boldsymbol{e}(x_i))| \le c_2 r^2} \big|\log|\widetilde{P}(\boldsymbol{e}(x_i))|\big|.$$

Recall that $\deg \widetilde{P} \ll_d \deg P$ and that $\widetilde{P}$ has at most k^2 terms and norm 1. Lemma 7.4 applied to $\widetilde{P}$ and $c_2 r^2$ implies

$$\mathcal{E}' \ll_{d,k} r^{1/k^2} |\log r| + \deg(P)\mathcal{D}^{1/(l+1)} r^{-2} |\log r| + \frac{1}{n} \sum_{|\widetilde{P}(\boldsymbol{e}(x_i))| \le c_2 r^2} \big|\log|\widetilde{P}(\boldsymbol{e}(x_i))|\big|.$$

We use Lemma 7.7, applied to $\widetilde{P}$ and $c_2 r^2$, to bound the final sum and thus obtain

$$\mathcal{E}' \ll_{d,k} r^{1/k^2}|\log r| + \deg(P)\mathcal{D}^{1/(l+1)}r^{-2}|\log r| + \deg(P)\mathcal{D}^{1/(l+1)}r^{-4} + r^{1/(2k^2)} + \epsilon,$$

here $\epsilon = \left|m(\widehat{P}) - \frac{1}{n}\sum_{i=1}^{n}\log|\widehat{P}(\boldsymbol{e}(x_i))|\right|$; note that multiplying $\widehat{P}$ with a nonzero scalar and a monomial leaves ϵ invariant.

We return to the total error term $\mathcal{E}$. By (7-14) together with $l \le d$, $r \le 1$, and $\mathcal{D} \le 1$ we get

$$\mathcal{E} \ll_{d,k} r^{1/(2k^2)} + (\deg(P)\mathcal{D}^{1/(d+1)}r^{-4})^{1/(8k)} + \deg(P)\mathcal{D}^{1/(d+1)}r^{-4} + \epsilon.$$

We choose $r = \frac{1}{2}\mathcal{D}^{1/(8(d+1))}$, then the proposition follows as $\mathcal{D} \le 1$. □

8. Endgame

In this section we prove a stronger version of Theorem 1.1 from the introduction.

8A. ***Preliminaries.*** Suppose $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$. For $V \in \mathrm{GL}_d(\mathbb{Z})$ we set $Q \in \overline{\mathbb{Q}}[X_1, \dots, X_d]$ to be $P(X^{V^{-1}})$ multiplied by a suitable monomial in $X_1, \dots, X_d$ such that Q is coprime to $X_1 \cdots X_d$. Let $l \in \{0, \dots, d-1\}$. For $z = (z_1, \dots, z_l) \in \mathbb{C}^l$ we set

$$P_{V,z} = Q(z_1, \dots, z_l, X_1, \dots, X_{d-l}) \tag{8-1}$$

this is a polynomial in $d - l$ variables. Note that $P_{V,\boldsymbol{e}(x)} = Q_{\boldsymbol{e}(x)} \in \mathbb{C}[X_1, \dots, X_{d-l}]$ in the notation introduced near the beginning of Section 7B. It is useful to allow $l = 0$ in which case $P_{V,z} = Q$. In our typical application $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has finite order. We write $\boldsymbol{\zeta}^V = (\boldsymbol{\eta}, \boldsymbol{\xi})$ where $\boldsymbol{\eta} \in \mathbb{G}_m^l$, $\boldsymbol{\xi} \in \mathbb{G}_m^{d-l}$ and see $|P_{V,\boldsymbol{\eta}}(\boldsymbol{\xi})| = |P(\boldsymbol{\zeta})|$.

The following lemma requires a result of Bombieri, Masser, and Zannier [2007] and relies crucially on the hypothesis that P is essentially atoral.

Lemma 8.1. *Suppose $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm}]\backslash\{0\}$ is essentially atoral. There exists $c \ge 1$ depending only on P and d such that for all $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ of finite order with $\delta(\boldsymbol{\zeta}) \ge c$, for all $V \in \mathrm{GL}_d(\mathbb{Z})$, and all $l \in \{0, \dots, d-1\}$, we have $P_{V,\boldsymbol{\eta}} \ne 0$ and $\mathcal{B}(P_{V,\boldsymbol{\eta}}) \le c|V^{-1}|$ where $\boldsymbol{\zeta}^V = \{\boldsymbol{\eta}\} \times \mathbb{G}_m^{d-l}$.*

Proof. The Zariski closure W in $\mathbb{G}_m^d$ of all algebraic zeros of P in $(S^1)^d$ is defined over $\overline{\mathbb{Q}}$.

By hypothesis, P is essentially atoral. So each irreducible component of the Zariski closure of all complex roots of P on $(S^1)^d$ is of codimension at least 2 in $\mathbb{G}_m^d$ or a proper torsion coset of $\mathbb{G}_m^d$. Therefore, each irreducible component of W is also of this type.

Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ be of finite order with $\delta(\boldsymbol{\zeta}) \ge c$, where c is to be determined, and $\boldsymbol{\zeta}^V = \{\boldsymbol{\eta}\} \times \mathbb{G}_m^{d-l}$ with V and l as in the hypothesis.

Let $\boldsymbol{\eta}' \in \mathbb{G}_m^{d-l}$ be of finite order, $z \in S^1\backslash\mu_\infty$ be algebraic, and $u \in \mathbb{Z}^{d-l}$ with $P_{V,\boldsymbol{\eta}}(\boldsymbol{\eta}' z^u) = 0$. We must find $v'' \in \mathbb{Z}^{d-l}\backslash\{0\}$ with $|v''| \le c|V^{-1}|$ such that $\langle u, v''\rangle = 0$. The existence of such a v'' establishes in particular $P_{V,\boldsymbol{\eta}} \ne 0$ (as c depends only on P and d).

Now $P(x) = 0$ for the algebraic point $x = (\boldsymbol{\eta}, \boldsymbol{\eta}' z^u)^{V^{-1}} \in (S^1)^d$. So x is contained in an irreducible component W' of W and in a 1-dimensional algebraic subgroup of $\mathbb{G}_m^d$.

If $\dim W' \le d-2$, we apply Bombieri, Masser, and Zannier's Theorem 1.5 [2007] to $\mathcal{X} = W'$. We get a proper torsion coset of $\mathbb{G}_m^d$ containing x and coming from a finite set depending only on W', and thus only on P. We find $v \in \mathbb{Z}^d \backslash \{0\}$ with $|v| \ll_{d,P} 1$ and $x^v = 1$.

If W' is a proper torsion coset of $\mathbb{G}_m^d$ there exists $v \in \mathbb{Z}^d \backslash \{0\}$, depending only on W' such that $y^v = 1$ holds for all $y \in W'$. Again we find $|v| \ll_{d,P} 1$ and $x^v = 1$.

In either case we have

$$1 = x^v = (\boldsymbol{\eta}, \boldsymbol{\eta}' z^u)^{V^{-1}v} = \boldsymbol{\eta}^{v'} (\boldsymbol{\eta}' z^u)^{v''} \quad \text{where } V^{-1}v = \begin{pmatrix} v' \\ v'' \end{pmatrix} \in \mathbb{Z}^l \times \mathbb{Z}^{d-l}. \tag{8-2}$$

In particular, $\langle u, v'' \rangle = 0$ as z has infinite order.

If $v'' \neq 0$, then we are done. Indeed, $|v''| \le |V^{-1}v| \le d|V^{-1}||v|$ and $|v|$ is bounded from above solely in terms of P and d.

Let us assume $v'' = 0$ and derive a contradiction for c large in terms of P and d. Note $l \ge 1$ as v cannot be 0. Then $v' \neq 0$ and by equality (8-2) we find $\boldsymbol{\eta}^{v'} = 1$. Recall that $\boldsymbol{\eta}$ consists of the first l coordinates of $\boldsymbol{\zeta}^V$. Thus $\boldsymbol{\zeta}^v = 1$ and hence $\delta(\boldsymbol{\zeta}) \le |v|$ where $|v| \ll_{d,P} 1$. But $\delta(\boldsymbol{\zeta}) \ge c$, a contradiction for large enough c. □

Definition 8.2. Let $c \ge 1$ be a real number. Suppose $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}] \backslash \{0\}$ and $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ is of finite order. The pair $(P, \boldsymbol{\zeta})$ is called c-admissible if for all $V \in \mathrm{GL}_d(\mathbb{Z})$ and all $l \in \{0, \dots, d-1\}$, we have $P_{V,\boldsymbol{\eta}} \neq 0$ and $\mathcal{B}(P_{V,\boldsymbol{\eta}}) \le c|V^{-1}|$ where $\boldsymbol{\zeta}^V \in \{\boldsymbol{\eta}\} \times \mathbb{G}_m^{d-l}$.

The case $l = 0$ yields in particular $\mathcal{B}(P) \le c$ if there exists $\boldsymbol{\zeta}$ such that $(P, \boldsymbol{\zeta})$ is c-admissible; indeed take V as the identity matrix.

Let P be an essentially atoral Laurent polynomial with algebraic coefficient. By Lemma 8.1 there exists $c \ge 1$ such that $(P, \boldsymbol{\zeta})$ is c-admissible for all $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ of finite order with $\delta(\boldsymbol{\zeta}) \ge c$.

In the definition of admissibility, it will be useful to keep track of $\boldsymbol{\zeta}$ when passing it down in an induction step. The next lemma makes this precise.

Lemma 8.3. *Let $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}] \backslash \{0\}$ and let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ be of finite order such that $(P, \boldsymbol{\zeta})$ is c-admissible with $c \ge 1$. Say $l \in \{0, \dots, d-1\}$, $V \in \mathrm{GL}_d(\mathbb{Z})$, and $\boldsymbol{\zeta}^V = (\boldsymbol{\eta}, \boldsymbol{\xi})$ with $\boldsymbol{\eta} \in \mathbb{G}_m^l$ and $\boldsymbol{\xi} \in \mathbb{G}_m^{d-l}$. Then $(P_{V,\boldsymbol{\eta}}, \boldsymbol{\xi})$ is $(cd|V^{-1}|)$-admissible.*

Proof. Throughout the proof we use that $|\cdot|$ is the maximum-norm on matrices.

We abbreviate $R = P((\boldsymbol{\eta}, X_1, \dots, X_{d-l})^{V^{-1}})$ which equals $P_{V,\boldsymbol{\eta}}$ up to a monomial factor. It suffices to show that $(R, \boldsymbol{\xi})$ is $(cd|V|^{-1})$-admissible.

To this end say $k \in \{0, \dots, d-l-1\}$, $W \in \mathrm{GL}_{d-l}(\mathbb{Z})$, and $\boldsymbol{\xi}^W = \{\boldsymbol{\eta}'\} \times \mathbb{G}_m^{d-l-k}$ with $\boldsymbol{\eta}' \in \mathbb{G}_m^k$. We must bound $\mathcal{B}(R_{W,\boldsymbol{\eta}'})$. So say $z \in S^1 \backslash \mu_\infty$, $u \in \mathbb{Z}^{d-l-k}$, and $\boldsymbol{\eta}'' \in \mathbb{G}_m^{d-l-k}$ is of finite order with $R_{W,\boldsymbol{\eta}'}(\boldsymbol{\eta}'' z^u) = 0$. Thus $R((\boldsymbol{\eta}', \boldsymbol{\eta}'' z^u)^{W^{-1}}) = 0$ and hence $P((\boldsymbol{\eta}, (\boldsymbol{\eta}', \boldsymbol{\eta}'' z^u)^{W^{-1}})^{V^{-1}}) = 0$. We abbreviate $W' = \left(\begin{smallmatrix} E_l & 0 \\ 0 & W \end{smallmatrix}\right)$ with E_l the $l \times l$ identity matrix. So $P((\boldsymbol{\eta}, \boldsymbol{\eta}', \boldsymbol{\eta}'' z^u)^{(VW')^{-1}}) = 0$ which means $P_{VW',(\boldsymbol{\eta},\boldsymbol{\eta}')}(\boldsymbol{\eta}'' z^u) = 0$.

Observe that $\boldsymbol{\zeta}^{VW'} = (\boldsymbol{\eta}, \boldsymbol{\xi})^{W'} = (\boldsymbol{\eta}, \boldsymbol{\xi}^W) = (\boldsymbol{\eta}, \boldsymbol{\eta}', *)$. By hypothesis $(P, \boldsymbol{\zeta})$ is c-admissible. Therefore, $\mathcal{B}(P_{VW',(\boldsymbol{\eta},\boldsymbol{\eta}')}) \le c|(VW')^{-1}| = c|W'^{-1}V^{-1}| \le cd|V^{-1}||W'^{-1}| = cd|V^{-1}||W^{-1}|$. In other words, there

exists $v \in \mathbb{Z}^{d-l-k}\backslash\{0\}$ with $|v| \le cd|V^{-1}||W^{-1}|$ and $\langle u, v\rangle = 0$. Thus $\mathcal{B}(R_{W,\boldsymbol{\eta}'}) \le cd|V^{-1}||W^{-1}|$, as desired. Moreover, $R_{W,\boldsymbol{\eta}'} \neq 0$. □

Lemma 8.4. *Let $P \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]\backslash\{0\}$ and let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ be of finite order such that $(P, \boldsymbol{\zeta})$ is c-admissible with $c \ge 1$. Say $l \in \{1, \dots, d-1\}$ and let $\widehat{P} \in \overline{\mathbb{Q}}[X_1^{\pm 1}, \dots, X_l^{\pm 1}]$ be as in* (7-5) *and $\boldsymbol{\zeta} \in \{\boldsymbol{\eta}\} \times \mathbb{G}_m^{d-l}$. Then $(\widehat{P}, \boldsymbol{\eta})$ is c-admissible.*

Proof. Suppose $V \in \mathrm{GL}_l(\mathbb{Z})$ such that $\boldsymbol{\eta}^V = (\boldsymbol{\eta}', *)$ where $\boldsymbol{\eta} \in \mathbb{G}_m^{l'}$ and $l' \in \{0, \dots, l-1\}$. Following the definition of admissibility and recalling (8-1) we are in the following situation. There is $\boldsymbol{\eta}'' \in \mathbb{G}_m^{l-l'}$, $z \in S^1\backslash\mu_\infty$ algebraic, and $u' \in \mathbb{Z}^{l-l'}$ such that

$$\widehat{P}((\boldsymbol{\eta}', \boldsymbol{\eta}'' z^{u'})^{V^{-1}}) = 0.$$

It follows from the definition of $\widehat{P}$ that $P((\boldsymbol{\eta}', \boldsymbol{\eta}'' z^{u'})^{V^{-1}}, X_{l+1}, \dots, X_d) = 0$ as a polynomial in $X_{l+1}, \dots, X_d$. We extend $\widetilde{V} = \left(\begin{smallmatrix} V & 0 \\ 0 & E_{d-l} \end{smallmatrix}\right)$ where E_{d-l} is the $(d-l) \times (d-l)$ identity matrix. Then $P((\boldsymbol{\eta}', \boldsymbol{\eta}'' z^{u'}, z^{u''})^{\widetilde{V}^{-1}}) = 0$ for all $u'' \in \mathbb{Z}^{d-l}$.

By hypothesis, $(P, \boldsymbol{\zeta})$ is c-admissible and $\boldsymbol{\zeta}^{\widetilde{V}} = (\boldsymbol{\eta}^V, *) = (\boldsymbol{\eta}', *, *)$. Now $P_{\widetilde{V},\boldsymbol{\eta}'}(\boldsymbol{\eta}'' z^{u'}, z^{u''}) = 0$, so by definition there exist $v' \in \mathbb{Z}^{l-l'}$, $v'' \in \mathbb{Z}^{d-l}$, not both zero, such that $\langle u', v'\rangle + \langle u'', v''\rangle = 0$ and $|(v', v'')| \le c|\widetilde{V}^{-1}| = c|V^{-1}|$ for the maximum-norm.

As we are free to vary u'' we see that $\{u'\} \times \mathbb{Q}^{d-l}$ is contained in a finite union of proper vector subspaces of $\mathbb{Q}^d$, each defined as the kernel of $\langle \cdot, (v', v'')\rangle$ with v', v'' as above. So $\{u'\} \times \mathbb{Q}^{d-l} \subset V$ for one of these vector spaces V defined by some (v', v''). We must have $v'' = 0$ and hence $\langle u', v'\rangle = 0$. Then $v' \neq 0$ and as $|v'| \le c|V^{-1}|$ we conclude that $\widehat{P}$ is c-admissible. □

Here are some basic estimates involving $P_{V,\boldsymbol{\eta}}$.

Lemma 8.5. *Let $P \in \overline{\mathbb{Q}}[X_1, \dots, X_d]\backslash\{0\}$, $l \in \{0, \dots, d-1\}$, and $V \in \mathrm{GL}_d(\mathbb{Z})$. Say $\boldsymbol{\eta} \in \mathbb{G}_m^l$ has finite order and $P_{V,\boldsymbol{\eta}} \neq 0$: The following hold true.*

(i) *We have* $\deg P_{V,\boldsymbol{\eta}} \ll_d |V|^{d-1} \deg P$.

(ii) *We have $h(P_{V,\boldsymbol{\eta}}) \le \log(k) + h(P)$ where $k \ge 2$ is an upper bound for the number of nonzero terms of P.*

Proof. Both parts follow are elementary consequences of the degree and the height of a polynomial. For (i) we require $|V^{-1}| \ll_d |V|^{d-1}$. For (ii) we note that Q from the beginning of this subsection has the same coefficients and thus the same height as P. We decompose $h(P_{V,\boldsymbol{\eta}})$ in local heights as in (2-4). The triangle inequality at the archimedean places leads to $\log k$. □

We continue with further basic estimates involving $\widehat{P}$ as in (7-5).

Lemma 8.6. *Let $K \subset \mathbb{C}$ be a number field and suppose $P \in K[X_1, \dots, X_d]\backslash\{0\}$ has at most $k \ge 2$ terms, where k is an integer. Say $l \in \{1, \dots, d-1\}$ with $\widehat{P} \in \mathbb{C}[X_1^{\pm 1}, \dots, X_l^{\pm 1}]$ as in* (7-5). *Then the following properties hold true*:

(i) *We have* $\widehat{P} \in K'[X_1^{\pm 1}, \dots, X_l^{\pm 1}]$ *where* K' *is a number field such that* $K \subset K' \subset \mathbb{C}$ *and* $[K' : \mathbb{Q}] \leq [K : \mathbb{Q}]^2$.

(ii) *We have* $h(\widehat{P}) \ll_k 1 + h(P)$.

Proof. The coefficients of $\widehat{P}$ are contained in the subfield K' of $\mathbb{C}$ generated by a primitive element of $K/\mathbb{Q}$ and its complex conjugate. So $[K' : \mathbb{Q}] \leq [K : \mathbb{Q}]^2$ and (i) follows. For (ii) we remark that each p_i as in (7-5) has at most k terms and that there are at most k nonzero p_i. Using the local decomposition of the height together with the ultrametric and archimedean triangle inequality yields the claim. □

8B. ***Completion of proof.*** The next lemma will setup a monomial change of coordinates. We recall that $\Lambda_{\boldsymbol{\xi}}$ was defined in (5-4), $\widetilde{\Lambda}_{\boldsymbol{\xi}}(\nu)$ was defined in (5-5) and $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}}(\nu))$ is as in (5-6).

Lemma 8.7. *Suppose* $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ *has order* N *and let* $\delta \geq 1$, $\epsilon \in \left(0, \frac{1}{2}\right]$, $\nu_1, \dots, \nu_{d-1} \in \left(0, \frac{1}{2}\right]$ *with* $\nu_1 + \cdots + \nu_{d-1} \leq \frac{1}{2}$. *Then there exist* $l \in \{0, \dots, d-1\}$ *and* $V \in \mathrm{GL}_d(\mathbb{Z})$ *such that the following hold*:

(i) *We have* $|V| \ll_d \delta^{2\epsilon^{d-l}}$ *and* V *is the identity matrix if* $l = 0$.

(ii) *We have* $\boldsymbol{\zeta}^V = (\boldsymbol{\eta}, \boldsymbol{\xi})$ *where* $\boldsymbol{\eta} \in \mathbb{G}_m^l$, $\boldsymbol{\xi} \in \mathbb{G}_m^{d-l}$, $\mathrm{ord}(\boldsymbol{\eta}) \leq N^{\nu_1 + \cdots + \nu_l}$, $\boldsymbol{\xi}$ *has finite order at least* $N^{1/2}$. *Finally, if* $l \leq d-2$ *then* $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}}(\nu_{l+1})) > \delta^{\epsilon^{d-l-1}}$.

Proof. Set $\boldsymbol{\xi}_1 = \boldsymbol{\zeta}$ and let V_0 be the identity matrix in $\mathrm{GL}_d(\mathbb{Z})$. For all $l \in \{1, \dots, d-1\}$ with $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l)) \leq \delta^{\epsilon^{d-l}}$ we will construct inductively $V_l \in \mathrm{GL}_d(\mathbb{Z})$, $\boldsymbol{\xi}_{l+1} \in \mathbb{G}_m^{d-l}$ of order at most N, and $\eta_l \in \mathbb{G}_m$ of order at most N^{ν_l} such that $\boldsymbol{\zeta}^{V_l} = (\eta_1, \dots, \eta_l, \boldsymbol{\xi}_{l+1})$ and

$$|V_l| \ll_d \delta^{\epsilon^{d-1} + \cdots + \epsilon^{d-l}}. \tag{8-3}$$

Suppose $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l)) \leq \delta^{\epsilon^{d-l}}$, there exists $v \in \widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l) \backslash \{0\}$ such that $|v| \leq \delta^{\epsilon^{d-l}}$ and v is primitive. Note that $[\widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l) : \Lambda_{\boldsymbol{\xi}_l}(\nu_l)]v$ lies in $\Lambda_{\boldsymbol{\xi}_l}$, so $\mathrm{ord}(\boldsymbol{\xi}_l^v) \leq [\widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l) : \Lambda_{\boldsymbol{\xi}_l}(\nu_l)] \leq \det(\Lambda_{\boldsymbol{\xi}_l}(\nu_l)) \leq N^{2\nu_l^2} \leq N^{\nu_l}$ by (5-3) and since $\det(\Lambda_{\boldsymbol{\xi}_l}) = \mathrm{ord}(\boldsymbol{\xi}_l) \leq N$. We can realize v as the first column of a matrix $V_l' \in \mathrm{GL}_{d-l+1}(\mathbb{Z})$ with $|V_l'| \ll_d |v| \ll_d \delta^{\epsilon^{d-l}}$, see the proof of Proposition 5.1. Let E_{l-1} denote the $(l-1) \times (l-1)$ identity matrix and set

$$V_l = V_{l-1} \begin{pmatrix} E_{l-1} & 0 \\ 0 & V_l' \end{pmatrix} \in \mathrm{GL}_d(\mathbb{Z}).$$

By step $l-1$ we have $\boldsymbol{\zeta}^{V_{l-1}} = (\eta_1, \dots, \eta_{l-1}, \boldsymbol{\xi}_l)$. We define η_l and $\boldsymbol{\xi}_{l+1}$ via $\boldsymbol{\zeta}^{V_l} = (\eta_1, \dots, \eta_l, \boldsymbol{\xi}_{l+1})$. Note $\eta_l = \boldsymbol{\xi}_l^v$, so $\mathrm{ord}(\eta_l) \leq N^{\nu_l}$ by the bound above. Finally, $|V_l| \ll_d |V_{l-1}||V_l'| \ll_d \delta^{\epsilon^{d-1} + \cdots + \epsilon^{d-l}}$, and $\boldsymbol{\xi}_l^N = 1$. This completes our construction.

Otherwise, fix the largest $l \in \{1, \dots, d-1\}$ for which $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}_l}(\nu_l)) \leq \delta^{\epsilon^{d-l}}$; if no l satisfies the inequality we take $l = 0$. Then define $V = V_l$. Thus V is the identity matrix if $l = 0$ and claims (i) and (ii) are immediate as $\boldsymbol{\xi} = \boldsymbol{\zeta}$. So say $l \geq 1$. Then (i) holds by (8-3) as $\epsilon \leq \frac{1}{2}$. To verify (ii) observe that $\boldsymbol{\zeta}^V = (\eta_1, \dots, \eta_l, \boldsymbol{\xi})$ with $\boldsymbol{\xi} = \boldsymbol{\xi}_{l+1} \in \mathbb{G}_m^{d-l}$ and $(\eta_1, \dots, \eta_l)$ has order at most $N^{\nu_1 + \cdots + \nu_l} \leq N^{1/2}$. Thus $\boldsymbol{\xi}$ has finite order at least $N^{1/2}$ since $\boldsymbol{\zeta}^V$ has order N. If $l \leq d-2$, then $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}}(\nu_{l+1})) > \delta^{\epsilon^{d-l-1}}$, because the construction does not continue. □

We are ready to prove a theorem that will quickly imply our Theorem 1.1 and its refinements.

Theorem 8.8. *Let $c \geq 1$, let $K \subset \mathbb{C}$ be a number field, and suppose $P \in K[X_1, \dots, X_d]\backslash\{0\}$ has at most k terms for an integer $k \geq 2$. There are constants $C = C(d,k) \geq 1$ and $\kappa = \kappa(d,k) > 0$ depending only on d and k with the following property. Let $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ have finite order N and suppose G is a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ with $P(\boldsymbol{\zeta}^\sigma) \neq 0$ for all $\sigma \in G$. If $(P, \boldsymbol{\zeta}^\sigma)$ is c-admissible for all $\sigma \in G$ and if*

$$\delta(\boldsymbol{\zeta}) \geq C \max\{c, \deg P\}^C \tag{8-4}$$

then

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^\sigma)| = m(P) + O_{d,k}\left(\frac{[K:\mathbb{Q}]^{2^d}[(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^\kappa}\right).$$

Proof. The case $d = 1$ follows from Proposition 6.2(i) as $\delta(\boldsymbol{\zeta}) = N$ in this case and as $\mathcal{B}(P) < \infty$. So we may assume $d \geq 2$. We may also assume that P is nonconstant.

We work with the parameters $\nu_1, \dots, \nu_{d-1} \in (0, 1/(128d^2)]$, $\epsilon \in \left(0, \frac{1}{2}\right]$ in this proof. They are assumed to be small in terms of d and k but independent of P and $\boldsymbol{\zeta}$. We may assume that ϵ is small in terms of the ν_l, e.g., $\epsilon \leq \nu_l^d/4$ for all l. We determine them during the argument.

We apply Lemma 8.7 to $\boldsymbol{\zeta}$, $\delta = \delta(\boldsymbol{\zeta})$, ϵ, and the ν_l. Say l, V, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$ are given by this lemma, in particular $\boldsymbol{\zeta}^V = (\boldsymbol{\eta}, \boldsymbol{\xi})$ and $|V| \ll_d \delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}}$. We have

$$\operatorname{ord}(\boldsymbol{\eta}) \leq N^{\nu_1+\cdots+\nu_l} \quad \text{and} \quad \operatorname{ord}(\boldsymbol{\xi}) \geq N^{1/2}. \tag{8-5}$$

The case $l = 0$ is straightforward. Here V is the identity matrix, $\boldsymbol{\xi} = \boldsymbol{\zeta}$, and $\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\zeta}}(\nu_1)) > \delta(\boldsymbol{\zeta})^{\epsilon^{d-1}}$ as we are in case $d - l = d \geq 2$ of Lemma 8.7(ii). So $\tilde{\lambda}(\boldsymbol{\zeta}; \nu_1) \geq \delta(\boldsymbol{\zeta})^{\min\{\epsilon^{d-1}, \nu_1^d/2\}}$ using (6-2) and $\delta(\boldsymbol{\zeta}) \leq N$. As $(P, \boldsymbol{\zeta})$ is c-admissible we have $\mathcal{B}(P) \leq c$. We will apply Proposition 6.2(ii) to P and $\nu = \nu_1$, so we must verify $\tilde{\lambda}(\boldsymbol{\zeta}; \nu_1) > d^{1/2}\max\{c, \deg P\}$. This inequality is satisfied if $\delta(\boldsymbol{\zeta})$ is as in (8-4) with C large in terms of ϵ, ν_1, d, and k. So if $l = 0$, the theorem follows from (6-3).

Step 1: A monomial change of coordinates. From now on we assume $l \geq 1$, i.e., $l \in \{1, \dots, d-1\}$. We fix $\sigma \in G$ throughout this set. We have $\boldsymbol{\zeta}^{\sigma V} = (\boldsymbol{\eta}^\sigma, \boldsymbol{\xi}^\sigma) \in \mathbb{G}_m^l \times \mathbb{G}_m^{d-l}$ and $|P(\boldsymbol{\zeta}^\sigma)| = |P_{V,\boldsymbol{\eta}^\sigma}(\boldsymbol{\xi}^\sigma)|$. This time we apply Proposition 6.2 to $P_{V,\boldsymbol{\eta}^\sigma} \in K(\boldsymbol{\eta})[X_1, \dots, X_{d-l}]$, $\boldsymbol{\xi}^\sigma$, and ν_{l+1}. We often use that $\delta(\cdot)$ is Galois invariant, i.e., $\delta(\boldsymbol{\zeta}^\sigma) = \delta(\boldsymbol{\zeta})$.

If $d - l = 1$ we will apply Proposition 6.2(i) and there is nothing further to check.

But for $d - l \geq 2$ we must verify the hypothesis in the second part of this proposition. This step is similar as in the case $l = 0$.

Note that $P_{V,\boldsymbol{\eta}^\sigma} \neq 0$ as $(P, \boldsymbol{\zeta}^\sigma)$ is c-admissible; this polynomial has at most k nonzero terms. By Lemma 8.3 the pair $(P_{V,\boldsymbol{\eta}^\sigma}, \boldsymbol{\xi}^\sigma)$ is $(cd|V^{-1}|)$-admissible. Observe $|V^{-1}| \ll_d |V|^{d-1} \ll_d \delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}(d-1)}$. So the said pair is $c_1 c\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}$-admissible; here and below $c_1, c_2, \dots$ denote positive constants that depend only on d. In particular, $\mathcal{B}(P_{V,\boldsymbol{\eta}^\sigma}) \leq c_1 c\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}$. A similar argument and Lemma 8.5(i) yield

$$\max\{\mathcal{B}(P_{V,\boldsymbol{\eta}^\sigma}), \deg P_{V,\boldsymbol{\eta}^\sigma}\} \leq c_2\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}\max\{c, \deg P\}.$$

In conclusion, to apply Proposition 6.2(ii) we must verify the inequality in

$$\tilde{\lambda}(\boldsymbol{\xi}^{\sigma};\nu_{l+1}) = \max\{\lambda_1(\widetilde{\Lambda}_{\boldsymbol{\xi}^{\sigma}}(\nu_{l+1})), \operatorname{ord}(\boldsymbol{\xi}^{\sigma})^{\nu_{l+1}^{d-l}/2}\} > c_2\sqrt{d}\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}\max\{c,\deg P\}.$$

But $\Lambda_{\boldsymbol{\xi}^{\sigma}} = \Lambda_{\boldsymbol{\xi}}$ and the order are Galois invariant, see (5-4) and (5-5). So it suffices to prove the lower bound for $\tilde{\lambda}(\boldsymbol{\xi};\nu_{l+1})$. By Lemma 8.7(ii) we have

$$\tilde{\lambda}(\boldsymbol{\xi};\nu_{l+1}) \geq \min\{\delta(\boldsymbol{\zeta})^{\epsilon^{d-l-1}}, N^{\nu_{l+1}^{d-l}/4}\} \geq \delta(\boldsymbol{\zeta})^{\min\{\epsilon^{d-l-1},\nu_{l+1}^{d}/4\}} = \delta(\boldsymbol{\zeta})^{\epsilon^{d-l-1}} \tag{8-6}$$

as $\epsilon \leq \nu_{l+1}^{d}/4$. We may assume $\epsilon^{d-l-1} - 2\epsilon^{d-l}d \geq \epsilon^{d-l-1}/2$; this is equivalent to $\epsilon \leq 1/(4d)$. By (8-4) and (8-6) the desired inequality is satisfied when C is large in terms of ϵ, d, and k. We may thus apply Proposition 6.2.

We collect the following bounds from Lemmas 8.5 and 8.7:

$$\begin{aligned} \deg P_{V,\boldsymbol{\eta}^{\sigma}} &\ll_{d,k} \delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}\deg P,\\ h(P_{V,\boldsymbol{\eta}^{\sigma}}) &\ll_{k} 1 + h(P),\\ [K(\boldsymbol{\eta}):\mathbb{Q}] \leq \operatorname{ord}(\boldsymbol{\eta})[K:\mathbb{Q}] &\leq N^{\nu_1+\cdots+\nu_l}[K:\mathbb{Q}], \text{ and}\\ \operatorname{ord}(\boldsymbol{\xi}) &\geq N^{1/2} \end{aligned} \tag{8-7}$$

Recall that $\boldsymbol{\zeta}^{V} = (\boldsymbol{\eta},\boldsymbol{\xi})$. So $\boldsymbol{\xi}^{u} = 1$ for some $u \in \mathbb{Z}^{d-l}$ with $|u| = \delta(\boldsymbol{\xi})$, hence $\boldsymbol{\zeta}^{V\binom{0}{u}} = 1$. We conclude $\delta(\boldsymbol{\zeta}) \leq \left|V\binom{0}{u}\right| \leq d|V|\delta(\boldsymbol{\xi})$. We find

$$\delta(\boldsymbol{\xi}^{\sigma}) = \delta(\boldsymbol{\xi}) \gg_d \delta(\boldsymbol{\zeta})^{1-2\epsilon^{d-l}} \gg_d \delta(\boldsymbol{\zeta})^{1/2} \tag{8-8}$$

as we may assume $\epsilon^{d-l} \leq \frac{1}{4}$.

We must specify a subgroup $G \subset (\mathbb{Z}/M\mathbb{Z})^{\times}$ in Proposition 6.2 where $M = \operatorname{ord}(\boldsymbol{\xi})$; we will denote it by H here. Let L denote the fixed field of G in $\mathbb{Q}(\boldsymbol{\zeta})$. Let H be the subgroup of $(\mathbb{Z}/M\mathbb{Z})^{\times} = \operatorname{Gal}(\mathbb{Q}(\boldsymbol{\xi})/\mathbb{Q})$ corresponding to $\operatorname{Gal}(\mathbb{Q}(\boldsymbol{\xi})/\mathbb{Q}(\boldsymbol{\xi})\cap L(\boldsymbol{\eta}))$. We identify H with $\operatorname{Gal}(L(\boldsymbol{\zeta})/L(\boldsymbol{\eta}))$ under the isomorphism $\operatorname{Gal}(L(\boldsymbol{\zeta})/L(\boldsymbol{\eta})) \to \operatorname{Gal}(\mathbb{Q}(\boldsymbol{\xi})/\mathbb{Q}(\boldsymbol{\xi})\cap L(\boldsymbol{\eta}))$ induced by restriction. The fixed field of H in $\mathbb{Q}(\boldsymbol{\xi})$ is contained in $L(\boldsymbol{\eta})$, so

$$[(\mathbb{Z}/M\mathbb{Z})^{\times}:H] \leq [L(\boldsymbol{\eta}):\mathbb{Q}] \leq \operatorname{ord}(\boldsymbol{\eta})[L:\mathbb{Q}] = \operatorname{ord}(\boldsymbol{\eta})[(\mathbb{Z}/N\mathbb{Z})^{\times}:G] \leq N^{\nu_1+\cdots+\nu_l}[(\mathbb{Z}/N\mathbb{Z})^{\times}:G]$$

having used the bound for the order of $\boldsymbol{\eta}$ from (8-5). Moreover, the conductor of H satisfies

$$f_H \leq \operatorname{lcm}(f_G, \operatorname{ord}(\boldsymbol{\eta})) \leq f_G \operatorname{ord}(\boldsymbol{\eta}) \leq f_G N^{\nu_1+\cdots+\nu_l}.$$

If $\tau \in H$, then $\boldsymbol{\eta}^{\tau} = \boldsymbol{\eta}$. Therefore, $|P(\boldsymbol{\zeta}^{\tau\sigma})| = |P_{V,\boldsymbol{\eta}^{\tau\sigma}}(\boldsymbol{\xi}^{\tau\sigma})| = |P_{V,\boldsymbol{\eta}^{\sigma}}(\boldsymbol{\xi}^{\tau\sigma})| \neq 0$ for all $\tau \in H$. To cover the case $l = d-1$ it is useful to set $\nu_d = 1/(128d^2)$. By applying Proposition 6.2 to $P_{V,\boldsymbol{\eta}^{\sigma}}$, $\boldsymbol{\xi}^{\sigma}$, ν_{l+1},

and H while using the various estimates above, include (8-6) and (8-7), we find

$$\begin{aligned}
&\frac{1}{\#H}\sum_{\tau\in H}\log|P_{V,\boldsymbol{\eta}^\sigma}(\boldsymbol{\xi}^{\tau\sigma})| \\
&\quad=\frac{1}{\#H}\sum_{\tau\in H}\log|P(\boldsymbol{\zeta}^{\tau\sigma})| \\
&\quad=m(P_{V,\boldsymbol{\eta}^\sigma})+O_{d,k}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times:G]^2 f_G\deg(P)^2(1+h(P))N^{(2+2+1)(\nu_1+\cdots+\nu_l)}\delta(\boldsymbol{\zeta})^{4\epsilon^{d-l}d}}{N^{\nu_{l+1}^d/(40d)}}\right) \\
&\qquad\qquad+O_{d,k}\left(\frac{\deg(P)^{16d^2}}{\delta(\boldsymbol{\zeta})^{\epsilon^{d-l-1}/(16k)-32\epsilon^{d-l}d^3}}\right)
\end{aligned}$$

here we used $r\le d$ and $M\ge N^{1/2}$; the second error term, which appears in the their line of the expression, can be omitted if $l=d-1$ as then we apply Proposition 6.2(i).

At this point we reap the benefit of having split the error term in Proposition 6.2 into a part depending on N and a part depending on $\delta(\boldsymbol{\zeta})$. Indeed, the order of $\boldsymbol{\eta}$, which we bound in terms of N, does not affect the second error term above. Recall that $\delta(\boldsymbol{\zeta})\le N$, but there can be no meaningful lower bound for $\delta(\boldsymbol{\zeta})$ in terms of N. Introducing a dependency on N in the second error term $\delta(\boldsymbol{\zeta})$ would spoil the result.

We use the crude bound $\delta(\boldsymbol{\zeta})\le N$ and we may assume the parameters satisfy

$$5(\nu_1+\cdots+\nu_l)+4\epsilon^{d-l}d\le\frac{\nu_{l+1}^d}{80d},$$

for all $l\in\{1,\dots,d-1\}$, and

$$32\epsilon^{d-l}d^3\le\frac{\epsilon^{d-l-1}}{32k}.$$

Such a choice is possible. Indeed, we may fix $\nu_d,\nu_{d-1},\dots,\nu_1$ to decay quickly enough and ϵ is allowed to be small in terms of $\nu_1,\dots,\nu_{d-1}$ and d,k.

We now combine both contributions to the error term and get

$$\frac{1}{\#H}\sum_{\tau\in H}\log|P(\boldsymbol{\zeta}^{\tau\sigma})|=m(P_{V,\boldsymbol{\eta}^\sigma})+O_{d,k}\left(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^\times:G]^2 f_G\deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^\kappa}\right)\quad(8\text{-}9)$$

if

$$\kappa\le\min\left\{\frac{\nu_{l+1}^d}{80},\frac{\epsilon^{d-l-1}}{32k}\right\}.$$

Later we may shrink κ.

Step 2: Induction on d. Let $\boldsymbol{\zeta}$ be as in the hypothesis. Recall that $\boldsymbol{\zeta}^V=(\boldsymbol{\eta},\boldsymbol{\xi})$. We still assume $l\ge1$ and we find that

$$\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^\sigma)|=\frac{1}{[L(\boldsymbol{\eta}):L]}\sum_{\tau\in\mathrm{Gal}(L(\boldsymbol{\eta})/L)}\frac{1}{\#H}\sum_{\sigma\in H}\log|P(\boldsymbol{\zeta}^{\tilde\tau\sigma})|$$

with $\tilde{\tau}$ a fixed lift of τ to $\mathrm{Gal}(L(\boldsymbol{\zeta})/L) = \mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/L) = G$, recall that $H = \mathrm{Gal}(L(\boldsymbol{\zeta})/L(\boldsymbol{\eta}))$. Thus (8-9) with $\tilde{\tau}$ for σ implies

$$\begin{aligned}&\frac{1}{\#G}\sum_{\sigma\in G}\log|P(\boldsymbol{\zeta}^{\sigma})|\\&=\frac{1}{[L(\boldsymbol{\eta}):L]}\sum_{\tau\in\mathrm{Gal}(L(\boldsymbol{\eta})/L)} m(P_{V,\boldsymbol{\eta}^{\tau}})+O_{d,k}\bigg(\frac{[K:\mathbb{Q}]^2[(\mathbb{Z}/N\mathbb{Z})^{\times}:G]^2 f_G \deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^{\kappa}}\bigg).\end{aligned}\tag{8-10}$$

We set Q to equal $P(X^{V^{-1}})$ times a monomial such that Q is a polynomial coprime to $X_1\cdots X_d$. We apply the construction (7-5) to Q and l and obtain $\widehat{Q}$. Recall Lemma 7.2 and write $\widetilde{Q}$ for $\widehat{Q}$ times the monomial from part (ii) of this lemma. Then $\widetilde{Q}$ has at most k^2 nonzero terms and using also Lemma 8.6 we find

$$\begin{aligned}&\widetilde{Q}\in K'[X_1^{\pm1},\dots,X_l^{\pm1}]\quad\text{where } [K':\mathbb{Q}]\le[K:\mathbb{Q}]^2,\\&\deg\widetilde{Q}\ll_d \deg Q\ll_d |V^{-1}|\deg P\ll_d |V|^{d-1}\deg P\ll_d \delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}\deg P,\ \text{and}\\&h(\widetilde{Q})\ll_k 1+h(Q)\ll_k 1+h(P).\end{aligned}\tag{8-11}$$

By Lemma 8.3, with $l=0$, the pair $(Q,(\boldsymbol{\eta}^{\sigma},\boldsymbol{\xi}^{\sigma}))$ is $c_3c\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}$-admissible for all $\sigma\in G$. Now $(\widetilde{Q},\boldsymbol{\eta}^{\sigma})$ is also $c_3c\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d}$-admissible by Lemma 8.4 for all $\sigma\in G$ (multiplying a polynomial by a monomial has no effect on admissibility).

We want to apply the current theorem to $\widetilde{Q}$ and $\boldsymbol{\eta}\in\mathbb{G}_m^l$ by induction on the number of variables, recall $l\le d-1$. For this we must verify

$$\delta(\boldsymbol{\eta})\ge c_4C(l,k^2)\delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}dC(l,k^2)}\max\{c,\deg P\}^{C(l,k^2)}$$

having used the bound for $\deg\widetilde{Q}$ in (8-11). As above and in (8-8), the bound $\delta(\boldsymbol{\eta})\ge\delta(\boldsymbol{\zeta})/|V|\gg_d\delta(\boldsymbol{\zeta})^{1/2}$. So it suffices to check

$$\delta(\boldsymbol{\zeta})^{1-4\epsilon^{d-l}dC(l,k^2)}\ge c_5C(l,k^2)^2\max\{c,\deg P\}^{2C(l,k^2)}.\tag{8-12}$$

We may assume that $1-4\epsilon^{d-l}dC(l,k^2)\ge\frac{1}{2}$ as we may choose ϵ small in terms of d and $C(l,k^2)$. So (8-12) follows from (8-4) if $C=C(d,k)$ is large enough in terms of d and k.

To apply this theorem by induction we must specify a subgroup of $H'\subset(\mathbb{Z}/E\mathbb{Z})^{\times}=\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\eta})/\mathbb{Q})$ with $E=\mathrm{ord}(\boldsymbol{\eta})$. We take H' as identified with $\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\eta})/\mathbb{Q}(\boldsymbol{\eta})\cap L)\cong\mathrm{Gal}(L(\boldsymbol{\eta})/L)$ under the isomorphism induced by restriction. For all $\tau\in\mathrm{Gal}(L(\boldsymbol{\eta})/L)$ we have $P_{V,\boldsymbol{\eta}^{\tau}}\ne0$ and so $\widetilde{Q}(\boldsymbol{\eta}^{\tau})\ne0$ by (7-6). By induction and (8-11) we have

$$\frac{1}{\#H'}\sum_{\tau\in H'}\log|\widetilde{Q}(\boldsymbol{\eta}^{\tau})|=m(\widetilde{Q})+O_{d,k}\bigg(\frac{[K:\mathbb{Q}]^{2^d}[(\mathbb{Z}/E\mathbb{Z})^{\times}:H']^2 f_{H'}\deg(P)^{16d^2}\delta(\boldsymbol{\zeta})^{32\epsilon^{d-l}d^3}(1+h(P))}{\delta(\boldsymbol{\eta})^{\kappa(l,k^2)}}\bigg).$$

Note that $[(\mathbb{Z}/E\mathbb{Z})^\times : H'] = [\mathbb{Q}(\boldsymbol{\eta}) \cap L : \mathbb{Q}] \le [L : \mathbb{Q}] = [(\mathbb{Z}/N\mathbb{Z})^\times : G]$ and $f_{H'} \le f_G$. Using again $\delta(\boldsymbol{\eta}) \gg_d \delta(\boldsymbol{\zeta})^{1/2}$ we get

$$\frac{1}{\#H'} \sum_{\tau \in H'} \log|\widehat{Q}(\boldsymbol{\eta}^\tau)| = m(\widehat{Q}) + O_{d,k}\left(\frac{[K:\mathbb{Q}]^{2^d} [(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^{16d^2} (1+h(P))}{\delta(\boldsymbol{\zeta})^{\kappa(l,k^2)/2 - 32\epsilon^{d-l}d^3}} \right) \quad (8\text{-}13)$$

as passing from $\widetilde{Q}$ to $\widehat{Q}$ is harmless. We may assume that $\kappa(l,k^2)/4 \ge 32\epsilon^{d-l}d^3$.

Recall that Q equals $P(X^{V^{-1}})$ up to a monomial factor. We will soon apply Proposition 7.3 to Q. Consider $(x_1, \dots, x_{\#H'})$, with each $x_i \in [0,1)^l$, a tuple of discrepancy $\mathcal{D}$ as in (3-2), where the $\boldsymbol{e}(x_i)$ are the $\boldsymbol{\eta}^\tau$. So $P_{V,\boldsymbol{\eta}^\tau} = P_{V,\boldsymbol{e}(x_i)} = Q_{\boldsymbol{e}(x_i)}$. Proposition 7.3 together with (8-13) imply

$$\begin{aligned}\frac{1}{\#H'} \sum_{\tau \in H'} & m(P_{V,\boldsymbol{\eta}^\tau}) \\ &= m(Q) + O_{d,k}\left(\deg(Q)\mathcal{D}^{1/(16(d+1)k^2)} + \frac{[K:\mathbb{Q}]^{2^d} [(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^{\kappa(l,k^2)/4}} \right).\end{aligned}$$

By Proposition 3.3(i) for $\boldsymbol{\eta}$ and H' and estimates used above we find

$$\mathcal{D} \ll_d [(\mathbb{Z}/E\mathbb{Z})^\times : H'] f_{H'}^{1/2} \delta(\boldsymbol{\eta})^{-1/3} \ll_d [(\mathbb{Z}/N\mathbb{Z})^\times : G] f_G^{1/2} \delta(\boldsymbol{\zeta})^{-1/6}.$$

From above we find $\deg Q \ll_d |V^{-1}| \deg P \ll_d \delta(\boldsymbol{\zeta})^{2\epsilon^{d-l}d} \deg P$. The Mahler measure is invariant under a monomial change of coordinates by [Schinzel 2000, Corollary 8, Chapter 3.4], thus $m(P) = m(Q)$. As H' is identified with $\mathrm{Gal}(L(\boldsymbol{\eta})/L)$ we get

$$\frac{1}{[L(\boldsymbol{\eta}):L]} \sum_{\tau \in \mathrm{Gal}(L(\boldsymbol{\eta})/L)} m(P_{V,\boldsymbol{\eta}^\tau}) = m(P) + O_{d,k}\left(\frac{[K:\mathbb{Q}]^{2^d} [(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^{\min\{1/(96(d+1)k^2) - 2\epsilon^{d-l}d, \kappa(l,k^2)/4\}}} \right).$$

We shrink ϵ a final time to achieve $1/(96(d+1)k^2) - 2\epsilon^{d-l}d > 1/(100(d+1)k^2)$. The theorem follows on combining this asymptotic estimate with (8-10), when $\kappa = \kappa(d,k)$ is small in terms of $\kappa(l,k^2)$, d, and k. □

To prove Theorem 1.1 we can multiply P by any monomial, so we may assume that it is a polynomial. Thus the theorem is a direct consequence of the following more precise corollary one taking $G = \Gamma_N$ which has conductor 1.

Corollary 8.9. *Let $K \subset \mathbb{C}$ be a number field and suppose $P \in K[X_1, \dots, X_d]\backslash\{0\}$ is essentially atoral and has at most k nonzero terms for an integer $k \ge 2$. There exists $\kappa = \kappa(d,k) > 0$ with the following property. Suppose $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has finite order N and suppose G is a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ and $\delta(\boldsymbol{\zeta})$ is large in terms of d, P, $[K:\mathbb{Q}]$, f_G, and $[(\mathbb{Z}/N\mathbb{Z})^\times : G]$. Then $P(\boldsymbol{\zeta}^\sigma) \ne 0$ for all $\sigma \in G$ and*

$$\frac{1}{\#G} \sum_{\sigma \in G} \log|P(\boldsymbol{\zeta}^\sigma)| = m(P) + O_{d,k}\left(\frac{[K:\mathbb{Q}]^{2^d} [(\mathbb{Z}/N\mathbb{Z})^\times : G]^2 f_G \deg(P)^{16d^2}(1+h(P))}{\delta(\boldsymbol{\zeta})^{\kappa}} \right).$$

Proof. By Lemma 8.1 there is $c \ge 1$, depending only on P, such that $(P, \boldsymbol{\zeta})$ is c-admissible for all $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ of finite order with $\delta(\boldsymbol{\zeta}) \ge c$.

Suppose $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has finite order and $P(\boldsymbol{\zeta}) = 0$. By the Manin–Mumford conjecture, $\delta(\boldsymbol{\zeta})$ is bounded in terms of d and P only. Hence for $\delta(\boldsymbol{\zeta})$ sufficiently large in terms of these quantities we have $P(\boldsymbol{\zeta}) \neq 0$. The same also holds with $\boldsymbol{\zeta}$ replaced by a Galois conjugate as $\delta(\cdot)$ is Galois invariant. Our corollary now follows from Theorem 8.8. □

Proof of Corollary 1.4. We may assume that $K/\mathbb{Q}$ is Galois and, after multiplying with a suitable monomial, that P is a polynomial. Our hypothesis implies that P is not a monomial. The product P' for $\tau(P)$ as τ ranges over $\mathrm{Gal}(K/\mathbb{Q})$ has rational coefficients. The coefficients are even integers as the coefficients of P lie in $\mathbb{Z}_K$.

The Mahler measure of any nonzero, integral polynomial is nonnegative. By a theorem attributed to Boyd [1981], Lawton [1977], and Smyth [1981], the fact that the zero set of P in $\mathbb{G}_m^d$ has an irreducible component not equal to the translate of an algebraic subgroup by a point of finite order implies $m(P') > 0$.

Suppose $\boldsymbol{\zeta} \in \mathbb{G}_m^d$ has order N. Take for G the subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$ associated to $\mathrm{Gal}(\mathbb{Q}(\boldsymbol{\zeta})/K \cap \mathbb{Q}(\boldsymbol{\zeta}))$. Then $[(\mathbb{Z}/N\mathbb{Z})^\times : G] \leq [K : \mathbb{Q}]$. As $\boldsymbol{\zeta}$ varies, there are only finitely many possibilities for the number field $K \cap \mathbb{Q}(\boldsymbol{\zeta})$, being a subfield of the field K. So there is a fixed $n \in \mathbb{N}$, independent of $\boldsymbol{\zeta}$, such that $K \cap \mathbb{Q}(\boldsymbol{\zeta})$ is contained in the number field generated by a root of unity of order n. So f_G is bounded from above solely in terms of K. For any $\tau \in \mathrm{Gal}(K/\mathbb{Q})$ choose an extension $\tilde{\tau} \in \mathrm{Gal}(K(\boldsymbol{\zeta})/\mathbb{Q})$. We apply Corollary 8.9 to the polynomial $\tau(P)$ which is essentially atoral by hypothesis. If $\delta(\boldsymbol{\zeta}^{\tilde{\tau}}) = \delta(\boldsymbol{\zeta})$ is large enough in terms of the fixed data, then

$$\frac{1}{\#G} \sum_{\sigma \in G} \log|\tau(P)(\boldsymbol{\zeta}^{\tilde{\tau}\sigma})| = m(\tau(P)) + o(1)$$

as $\delta(\boldsymbol{\zeta}) \to \infty$, here and below the implied constant is independent of $\boldsymbol{\zeta}$.

The average of the left-hand side over $\tau \in \mathrm{Gal}(K/\mathbb{Q})$ equals the left-hand side in

$$\frac{1}{[K(\boldsymbol{\zeta}) : \mathbb{Q}]} \sum_{\sigma : K(\zeta) \to \mathbb{C}} \log|\sigma(P(\boldsymbol{\zeta}))| = \frac{1}{[K : \mathbb{Q}]} \sum_{\tau \in \mathrm{Gal}(K/\mathbb{Q})} m(\tau(P)) + o(1).$$

As the Mahler measure is additive, the average on the right-hand side is $m(P')/[K : \mathbb{Q}] > 0$. But the left-hand side vanishes if $P(\boldsymbol{\zeta})$ is an algebraic unit. In this case, we see that $\delta(\boldsymbol{\zeta})$ is bounded from above. □

Appendix A. A theorem of Lawton re-revisited

The following theorem makes explicit a result of Lawton [1983]. It is a more precise version of a result of Habegger [2018] which is unfortunately insufficient for our purposes. We closely follow the proof presented in [Habegger 2018] which itself is based on Lawton's approach [1983]. We also show how to correct an inaccuracy in the proof of [Habegger 2018, Lemma A.4(i)].

Recall the definition of $\rho(\cdot)$ in (6-1) where $d \geq 1$ is an integer.

Theorem A.1. *Suppose $P \in \mathbb{C}[X_1, \dots, X_d]\backslash\{0\}$ has at most k nonzero terms for an integer $k \geq 2$. For $a = (a_1, \dots, a_d) \in \mathbb{Z}^d\backslash\{0\}$ with $\rho(a) > \deg P$ we have*

$$m(P(X^{a_1}, \dots, X^{a_d})) = m(P) + O_{d,k}\left(\frac{\deg(P)^{16d^2}}{\rho(a)^{1/(16(k-1))}}\right) \tag{A-1}$$

where the implicit constant depends only on d and k.

In the univariate case $d = 1$ we have $\rho(a) = \infty$ for all $a \in \mathbb{Z}\backslash\{0\}$ by definition. Then we should interpret (A-1) as stating $m(P(X^a)) = m(P)$. This identity is an easy consequence of (4-1). So throughout this subsection we assume $d \geq 2$.

We did not strive to obtain the best-possible exponent in $\rho(a)^{1/(16(k-1))}$ that our method can produce.

We must assume $\rho(a) > \deg P$ to avoid interaction of coefficients in $P(X^{a_1}, \dots, X^{a_d})$. Indeed, take for example $P = X_1(X_2 - 1 + \epsilon)$ with $\epsilon \in (0, 1)$ small and $a = (1, 0)$. Then $P(X, 1) = X\epsilon$ whose Mahler measure is $\log \epsilon$. On the other hand $m(P) = m(X_2 - 1 + \epsilon) = \log \max\{1, |1 - \epsilon|\} = 0$ by Jensen's formula. The difference

$$m(P(X, 1)) - m(P) = \log \epsilon$$

is unbounded as $\epsilon \to 0$. This does not contradict our theorem as $\rho(a) = 1$.

The Lebesgue measure on $\mathbb{R}^d$ is denoted by $\mathrm{vol}(\cdot)$. For $P \in \mathbb{C}[X_1^{\pm 1}, \dots, X_d^{\pm 1}]$ and $r > 0$ we define

$$S(P, r) = \{x \in [0, 1)^d : |P(\boldsymbol{e}(x))| < r\} \tag{A-2}$$

where $\boldsymbol{e}$ is as in (1-3).

Dobrowolski extended Lawton's Theorem 1 [1983] to polynomials that are not necessarily monic.

Theorem A.2 [Dobrowolski 2017, Theorem 1.1]. *Suppose $P \in \mathbb{C}[X]\backslash\{0\}$ has at most k nonzero terms for an integer $k \geq 2$. Then $\mathrm{vol}(S(P, r)) \ll_k \min\{1, r/|P|\}^{1/(k-1)}$ for all $r > 0$.*

Dobrowolski requires that P as at least 2 nonzero terms. But it is convenient to allow P to have a single term, as above. It is also convenient to apply the estimate in the case $P = 0$, we then interpret the minimum to be 1.

Until the end of this appendix and if not stated otherwise we assume that $P \in \mathbb{C}[X_1, \dots, X_d]\backslash\mathbb{C}$ has at most k nonzero terms for an integer $k \geq 2$ and $|P| = 1$.

Lemma A.3. (i) *If $r > 0$ then $\mathrm{vol}(S(P, r)) \ll_{d,k} r^{1/(2(k-1))}$.*

(ii) *We have $\int_{[0,1)^d} \left|\log|P(\boldsymbol{e}(x))|\right|^2 dx \ll_{d,k} 1$.*

Proof. To ease notation we drop d, k in the subscript $\ll_{d,k}$.

Because of the trivial bound $\mathrm{vol}(S(P, r)) \leq 1$ we may assume $r \leq 1$.

The case $d = 1$ follows from Theorem A.2. So let us now assume $d \geq 2$. We consider P as a polynomial in the unknown X_d and coefficients among $\mathbb{C}[X_1, \dots, X_{d-1}]$. We pick a coefficient P_i with maximal norm, i.e., P has a term $P_i X_d^i$ such that $P_i \in \mathbb{C}[X_1, \dots, X_{d-1}]$ and $|P_i| = |P| = 1$.

For $x' \in \mathbb{R}^{d-1}$ we let $P_{\boldsymbol{e}(x')}$ denote $P(\boldsymbol{e}(x'), X) \in \mathbb{C}[X]$. Recall that

$$S(P, r) = \{(x', t) \in [0, 1)^{d-1} \times [0, 1) : |P_{\boldsymbol{e}(x')}(\boldsymbol{e}(t))| < r\}.$$

We splice the hypercube and apply Fubini's theorem to find

$$\operatorname{vol}(S(P, r)) = \int_{[0,1)^{d-1}} \operatorname{vol}(S(P_{\boldsymbol{e}(x')}, r))\, dx'.$$

The measure zero set of $x' \in [0, 1)^{d-1}$ with $P_{\boldsymbol{e}(x')} = 0$ is harmless. By Theorem A.2 we find

$$\operatorname{vol}(S(P, r)) \ll \int_{[0,1)^{d-1}} \min\left\{1, \frac{r}{|P_{\boldsymbol{e}(x')}|}\right\}^{1/(k-1)} dx'.$$

The coefficient of X^i in $P_{\boldsymbol{e}(x')}$ is $P_i(\boldsymbol{e}(x'))$. So $|P_{\boldsymbol{e}(x')}| \geq |P_i(\boldsymbol{e}(x'))|$ and

$$\operatorname{vol}(S(P, r)) \ll \int_{[0,1)^{d-1}} \min\left\{1, \frac{r}{|P_i(\boldsymbol{e}(x'))|}\right\}^{1/(k-1)} dx' = I_1 + r^{1/(k-1)} I_2 \tag{A-3}$$

where

$$I_1 = \int_{|P_i(\boldsymbol{e}(x'))|<r} dx' \quad \text{and} \quad I_2 = \int_{|P_i(\boldsymbol{e}(x'))|\geq r} \frac{dx'}{|P_i(\boldsymbol{e}(x'))|^{1/(k-1)}};$$

both integrals are over subsets of $[0, 1)^{d-1}$. We will bound I_1 and I_2 from above.

We have $I_1 = \operatorname{vol}(S(P_i, r))$. This lemma applied by induction to P_i, a polynomial in $d-1$ variables with at most k nonzero terms and $|P_i| = 1$, yields

$$I_1 \ll r^{1/(2(k-1))}. \tag{A-4}$$

To bound I_2 we consider real numbers $r = r_0 < r_1 < \cdots < r_{N+1} = k+1$, with $r_{n+1} \leq r_n + \delta$ where $\delta \in (0, 1]$ is a small parameter. We split the domain of integration up into measurable parts

$$\Sigma_n = \{x' \in [0, 1)^{d-1} : r_n \leq |P_i(\boldsymbol{e}(x'))| < r_{n+1}\} \quad \text{for } n \in \{0, \ldots, N\}.$$

Observe that $|P_i(\boldsymbol{e}(x'))| \leq k < r_{N+1}$ for all x'. Thus

$$I_2 = \sum_{n=0}^{N} \int_{\Sigma_n} \frac{dx'}{|P_i(\boldsymbol{e}(x'))|^{1/(k-1)}} \leq \sum_{n=0}^{N} \frac{\operatorname{vol}(\Sigma_n)}{r_n^{1/(k-1)}} = \sum_{n=0}^{N} a_n b_n \tag{A-5}$$

where $a_n = r_n^{-1/(k-1)}$ and $b_n = \operatorname{vol}(\Sigma_n)$.

As the Σ_n are pairwise disjoint, the partial sums satisfy

$$B_n = \sum_{l=0}^{n} b_l = \operatorname{vol}\left(\bigcup_{l=0}^{n} \Sigma_l\right) \leq \operatorname{vol}(\{x' \in [0, 1)^{d-1} : |P_i(\boldsymbol{e}(x'))| < r_{n+1}\}) = \operatorname{vol}(S(P_i, r_{n+1})).$$

In particular, we have the trivial bound $B_n \leq 1$. As in the bound for I_1 we apply this lemma by induction to P_i and find

$$0 \leq B_n \leq \operatorname{vol}(S(P_i, r_{n+1})) \ll r_{n+1}^{1/(2(k-1))}. \tag{A-6}$$

Summation by parts implies

$$I_2 \le \sum_{n=0}^{N} a_n b_n = a_N B_N - \sum_{n=0}^{N-1} B_n(a_{n+1} - a_n) \le 1 + \sum_{n=0}^{N-1} B_n(a_n - a_{n+1});$$

we used $a_N = r_N^{-1/(k-1)} \le 1$ as $r_N \ge r_{N+1} - \delta \ge 1$ and $B_N \le 1$. By (A-6) and the definition of a_n we find

$$I_2 \ll 1 + \sum_{n=0}^{N-1} r_{n+1}^{1/(2(k-1))}(r_n^{-1/(k-1)} - r_{n+1}^{-1/(k-1)}).$$

We use the mean value theorem to bound

$$r_n^{-1/(k-1)} - r_{n+1}^{-1/(k-1)} \ll r_n^{-1/(k-1)-1}(r_{n+1} - r_n) \ll r_{n+1}^{-1/(k-1)-1}(r_{n+1} - r_n);$$

for the second bound we assume, as we may, that $\delta \le r$ and so $r_{n+1} \le r_n + \delta \le 2r_n$. Thus $I_2 \ll 1 + \int_r^{k+1} t^{-1/(2(k-1))-1} dt \ll r^{-1/(2(k-1))}$.

This bound together with (A-4) implies $I_1 + r^{1/(k-1)} I_2 \ll r^{1/(2(k-1))}$. Therefore, $\mathrm{vol}(S(P, r)) \ll r^{1/(2(k-1))}$ by (A-3), completing the induction step and the proof of (i).

We define $p_n(x) = \min\{n, |\log|P(\boldsymbol{e}(x))||^2\} \ge 0$ where $n \ge 0$ is an integer. We must find an upper bound for the nondecreasing sequence $I_n = \int_{[0,1)^d} p_n(x)\, dx$. Observe that $|P(\boldsymbol{e}(x))| \le k|P| = k$, so if $n \ge (\log k)^2$, then $|P(\boldsymbol{e}(x))| \le e^{\sqrt{n}}$. We fix m to be the least integer with $m \ge 1 + (\log k)^2$, so $m \ge 2$. Say $n \ge m$. Then p_n equals n on $S(P, e^{-\sqrt{n}})$ and it equals p_{n+1} outside this set. Thus

$$I_{n+1} - I_n = \int_{S(P, e^{-\sqrt{n}})} (p_{n+1}(x) - p_n(x))\, dx \le \mathrm{vol}(S(P, e^{-\sqrt{n}})) \ll e^{-\lambda\sqrt{n}}$$

from part (i), here $\lambda = 1/(2(k-1))$. A telescoping sum trick shows

$$I_n - I_m \ll \sum_{l \ge m} e^{-\lambda\sqrt{l}} \ll \int_{m-1}^{\infty} e^{-\lambda\sqrt{l}} dl \ll 1.$$

The initial term satisfies $I_m \le m \ll 1$ as m depends only on k, this completes the proof. □

A more careful analysis should lead to $\mathrm{vol}(S(P, r)) \ll_{d,k} (1 + |\log r|)^{d-1} r^{1/(k-1)}$ for all $r > 0$ in part (i) of Lemma A.3. But this improvement has little effect on the main results of the current work.

Brunault, Guilloux, Mehrabdollahi, and Pengo pointed out that the argument for second-named author's [Habegger 2018, Lemma A.4(i)] leads (for $k \ge 2$) to an estimate $O(y^{f(n)/(2(k-1))})$ where $f(n)$ depends on the number of variables n, as opposed to the claimed bound $O(y^{1/(2(k-1))})$. However, the claimed bound holds true by Lemma A.3(i). Alternatively and in the proof of Lemma A.3(i) one can replace Dobrowolski's Theorem 1.1 [2017] by Lawton's Theorem 1 [1983] which is sufficient for the applications in [Habegger 2018].

Lemma A.4. *If $r > 0$ then*

$$\int_{S(P,r)} \big|\log|P(\boldsymbol{e}(x))|\big|\, dx \ll_{d,k} r^{1/(4(k-1))}.$$

Proof. As $|P(\boldsymbol{e}(x))| \le |P|k \le k$ for all $x \in [0,1)^d$ we may assume $r \le 1$ by the Cauchy–Schwarz inequality and Lemma A.3(ii).

With $\Sigma = S(P, r)$ we find

$$0 \le -\int_\Sigma \log|P(\boldsymbol{e}(x))|\,dx = -\sum_{n=0}^{\infty} \int_{r/2^{n+1} \le |P(\boldsymbol{e}(x))| < r/2^n} \log|P(\boldsymbol{e}(x))|\,dx \le \sum_{n=0}^{\infty} \log\left(\frac{2^{n+1}}{r}\right) \mathrm{vol}(S(P, r/2^n)).$$

Let $\lambda = 1/(2(k-1)) \le \frac{1}{2}$. We use Lemma A.3(i) to bound $\mathrm{vol}(S(P, r/2^n)) \ll_{d,k} (r/2^n)^\lambda$. Note that $\log(2t) \ll_k t^{\lambda/2}$ on $t \in [1, \infty)$. We take $t = 2^n/r \ge 1$ and conclude

$$-\int_\Sigma \log|P(\boldsymbol{e}(x))|\,dx \ll_{d,k} \sum_{n=0}^{\infty} \left(\frac{r}{2^n}\right)^{\lambda/2} \ll_{d,k} r^{\lambda/2}. \qquad \square$$

Boyd [1998] proved that the Mahler measure is continuous on the nonzero polynomials of fixed degree. Here we show that the Mahler measure is Hölder continuous away from 0. For the next lemma we momentarily drop our usual assumptions on P.

Lemma A.5. *Suppose $P, Q \in \mathbb{C}[X_1, \dots, X_d]\backslash\{0\}$ such that P and Q both have at most k nonzero terms for an integer $k \ge 2$. If $\delta = |P - Q|/|Q| \le \frac{1}{2}$, then*

$$m(P) \le m(Q) + C(d, k)\delta^{1/(8(k-1))}$$

where $C(d, k) > 0$ is effective and depends only on d and k.

Proof. It suffices to prove the lemma when $|Q| = 1$; indeed, just replace P and Q by $P/|Q|$ and $Q/|Q|$, respectively, to reduce to this case.

Suppose for the moment that $x \in \mathbb{R}^d$ with $P(\boldsymbol{e}(x))Q(\boldsymbol{e}(x)) \ne 0$. Then $|P(\boldsymbol{e}(x)) - Q(\boldsymbol{e}(x))| \le 2k|P - Q| = 2k\delta$ and so

$$\log\left|\frac{P(\boldsymbol{e}(x))}{Q(\boldsymbol{e}(x))}\right| \le \left|\frac{P(\boldsymbol{e}(x))}{Q(\boldsymbol{e}(x))}\right| - 1 \le 2k\frac{\delta}{|Q(\boldsymbol{e}(x))|} \tag{A-7}$$

where the first inequality used $\log t \le t - 1$ for all $t > 0$.

The difference of Mahler measures $m(P) - m(Q)$ can be written as

$$\int_{[0,1)^d\backslash\Sigma} (\log|P(\boldsymbol{e}(x))| - \log|Q(\boldsymbol{e}(x))|)\,dx + \int_\Sigma (\log|P(\boldsymbol{e}(x))| - \log|Q(\boldsymbol{e}(x))|)\,dx$$

with $\Sigma = S(Q, \delta^{1/2})$.

The first integral is at most $2k\delta^{1/2}$ by (A-7). We proceed by bounding the second integral I from above. First, we note that $|P(\boldsymbol{e}(x))| \le k|P| \le 3k/2$ as $|P - Q| \le \delta \le \frac{1}{2}$ and thus $|P| \le \frac{3}{2}$. So

$$I \le \log(3k/2)\mathrm{vol}(\Sigma) - \int_\Sigma \log|Q(\boldsymbol{e}(x))|\,dx \le \log(3k/2)\mathrm{vol}(\Sigma) + c\delta^{1/(8(k-1))}$$

where we applied Lemma A.4 to Q and $\delta^{1/2}$, the case Q constant being trivial; here $c = c(d,k) > 0$ depends only on d and k. Finally, Lemma A.3(i) yields $\mathrm{vol}(\Sigma) = \mathrm{vol}(S(Q,\delta^{1/2})) \ll_{d,k} \delta^{1/(4(k-1))}$ and the lemma follows as $\delta \le 1$. □

Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $b \in \mathbb{N}_0$ let $C^b(\mathbb{R}^d)$ denote the set of real valued functions on $\mathbb{R}^d$ whose derivatives exist and are continuous up to and including order b. For a multiindex $i = (i_1, \dots, i_d) \in \mathbb{N}_0^d$ we set $\ell(i) = i_1 + \cdots + i_d$. If $g \in C^b(\mathbb{R}^d)$ and $\ell(i) \le b$, we set $\partial^i g = (\partial/\partial x_1)^{i_1} \cdots (\partial/\partial x_d)^{i_d} g \in C^0(\mathbb{R}^d)$ and

$$|g|_{C^b} = \max_{\substack{i \in \mathbb{N}_0^d \\ \ell(i) \le b}} \sup_{x \in \mathbb{R}^d} |\partial^i g(x)| \in \mathbb{R} \cup \{\infty\}.$$

We recall here the construction of $f_r \in C^b(\mathbb{R}^d)$ as in [Habegger 2018] where $r \in \left(0, \frac{1}{2}\right]$ is a parameter. This function equals $\log|P(\cdot))|$ away from the singularity, i.e., the locus where $P(\boldsymbol{e}(\cdot))$ vanishes.

We fix the antiderivative ϕ of $x^b(1-x)^b$ on $[0,1]$ with $\phi(0) = 0$ and multiply it with a positive number to ensure $\phi(1) = 1$. Then we extend it by 0 on $x < 0$ and by 1 for $x > 1$ to obtain a nondecreasing step function $\phi \in C^b(\mathbb{R})$ whose derivative ϕ' has support $[0,1]$. Finally, we rescale and define $\phi_r(x) = \phi(((2/r)^2 x - 1)/3)$. So ϕ_r is a nondecreasing function which vanishes on $(-\infty, (r/2)^2]$, equals 1 on $[r^2, \infty)$, and satisfies

$$\left|\frac{d^i \phi_r}{dx^i}\right|_{C^0} \ll_b r^{-2i} \quad \text{for all } 0 \le i \le b, \text{ hence} \quad |\phi_r|_{C^b} \ll_b r^{-2b}.$$

The function ϕ_r takes values in $[0,1]$. Moreover, we define

$$\psi_r(x) = \begin{cases} \frac{1}{2}\phi_r(x)\log x, & x > 0, \\ 0, & x \le 0. \end{cases}.$$

Then ψ_r vanishes on $(-\infty, (r/2)^2]$, coincides with $\frac{1}{2}\log x$ on $[r^2, \infty)$, and satisfies

$$|\psi_r|_{C^b} \ll_b r^{-2b}|\log r|. \tag{A-8}$$

We consider $g : x \mapsto |P(\boldsymbol{e}(x))|^2$, then

$$|g|_{C^b} \ll_{k,b} (\deg P)^b; \tag{A-9}$$

recall that $|P| = 1$. Next we compose $f_r = \psi_r \circ g \in C^b(\mathbb{R}^d)$, so for $x \in \mathbb{R}^d$ we have

$$f_r(x) = \begin{cases} 0, & \text{if } |P(\boldsymbol{e}(x))| \le r/2, \\ \log|P(\boldsymbol{e}(x))|, & \text{if } |P(\boldsymbol{e}(x))| \ge r. \end{cases}.$$

By [Habegger 2018, Lemma A.5], which follows from the chain rule, together with (A-8) and (A-9) we find

$$|f_r|_{C^b} \ll_{k,b} r^{-2b}|\log r|(\deg P)^{b^2}. \tag{A-10}$$

For the following lemmas we suppose $b \ge d+1$. As above we have $r \in \left(0, \frac{1}{2}\right]$.

Lemma A.6. *Suppose $a \in \mathbb{Z}^d \backslash \{0\}$, then*

$$\int_0^1 f_r(as)\,ds = \int_{[0,1)^d} f_r(x)\,dx + O_{d,k,b}\bigg(\frac{|\log r|}{r^{2b}}\frac{(\deg P)^{b^2}}{\rho(a)^{b-d}}\bigg).$$

We follow and adapt the proof of [Habegger 2018, Lemma A.6].

Proof. For $m \in \mathbb{Z}^d$ let $\widehat{f_r}(m)$ denote the Fourier coefficient of f_r. By [Grafakos 2014, Theorem 3.2.9(a)] with derivative up to order b and using $|\widehat{\partial^i f_r}(m)| \le |\partial^i f_r|_{C^0} \le |f_r|_{C^b}$ where $\ell(i) = b$ we conclude $|\widehat{f_r}(m)| \ll_{d,k,b} |f_r|_{C^b}|m|^{-b}$ if $m \neq 0$. So $|\widehat{f_r}(m)| \ll_{d,k,b} r^{-2b}|\log r|(\deg P)^{b^2}|m|^{-b}$ for all $m \in \mathbb{Z}^d\backslash\{0\}$ by (A-10). Then

$$\sum_{|m|\ge\rho(a)} |\widehat{f_r}(m)| \ll_{d,k,b} \frac{|\log r|}{r^{2b}}(\deg P)^{b^2} \sum_{|m|\ge\rho(a)} \frac{1}{|m|^b} \ll_{d,k,b} \frac{|\log r|}{r^{2b}}\frac{(\deg P)^{b^2}}{\rho(a)^{b-d}} \tag{A-11}$$

as $b \ge d+1$. In particular, the Fourier coefficients of f_r are absolutely summable and the Fourier series converges absolutely and uniformly to f_r, see [Grafakos 2014, Proposition 3.1.14]. Hence

$$\int_0^1 f_r(as)\,ds = \sum_{m\in\mathbb{Z}^d} \int_0^1 \widehat{f_r}(m)e^{2\pi\sqrt{-1}\langle a,m\rangle s}\,ds = \int_{[0,1)^d} f_r(x)\,dx + \sum_{\substack{m\in\mathbb{Z}^d\backslash\{0\}\\ \langle a,m\rangle=0}} \widehat{f_r}(m).$$

The lemma follows from (A-11) as only those m with $|m| \ge \rho(a)$ contribute to the final sum. □

Lemma A.7. *Suppose $a \in \mathbb{Z}^d\backslash\{0\}$ such that $\rho(a) > \deg P$. For all $s \in [0,1)$, up to finitely many exceptions, we have $|P(\boldsymbol{e}(as))| \neq 0$ and*

$$\int_0^1 \log|P(\boldsymbol{e}(as))|\,ds = \int_0^1 f_r(as)\,ds + O_k(r^{1/(k-1)}|\log r|).$$

We follow and adapt the proof of [Habegger 2018, Lemma A.7].

Proof. Say $a = (a_1, \dots, a_d)$ with $\rho(a) > \deg P$. Then the coefficients of the univariate Laurent polynomial $Q = P(X^{a_1}, \dots, X^{a_d})$ are precisely the coefficients of P. Hence $|Q| = |P| = 1$ and Q has at most k nonzero terms.

The first claim follows as $P(\boldsymbol{e}(as)) = Q(\boldsymbol{e}(s))$ for all $s \in \mathbb{R}$ and since $Q \neq 0$.

To prove the second claim we note that the difference of the two integrals equals

$$\int_{S(Q,r)} (\log|Q(\boldsymbol{e}(s))| - f_r(as))\,ds$$

with $S(Q,r)$ as in (A-2). Note that $\int_{S(Q,r)} \log|Q(\boldsymbol{e}(s))|\,ds \le 0$ as $r \le 1$. Recall Theorem A.2 which yields $\mathrm{vol}(S(Q,r)) \ll_k r^{1/(k-1)}$. As in the proof of [Lawton 1983, Lemma 4], see also [Schinzel 2000, Theorem 7, Appendix G], we find

$$\int_{S(Q,r)} \log|Q(\boldsymbol{e}(s))|\,ds \ge -Cr^{1/(k-1)}|\log r|,$$

where $C > 0$ depends only on k. Finally, by the definition of f_r we find $\log(r/2) \le f_r(as) \le 0$ if $|Q(\boldsymbol{e}(s))| < r$. Thus $\int_{S(Q,r)} f_r(as)\,ds$ is also $O_k(r^{1/(k-1)}|\log r|)$. □

Lemma A.8. *We have*

$$\left|\int_{[0,1)^d} (f_r(x) - \log|P(\boldsymbol{e}(x))|)\,dx\right| \ll_{d,k} r^{1/(4(k-1))}.$$

We follow and adapt the proof of [Habegger 2018, Lemma A.8].

Proof. We have

$$\begin{aligned}\left|\int_{[0,1)^d} (f_r(x)-\log|P(\boldsymbol{e}(x))|)\,dx\right| &= \left|\int_{[0,1)^d} (\phi_r(|P(\boldsymbol{e}(x))|^2)-1)\log|P(\boldsymbol{e}(x))|\,dx\right| \\ &\le \int_{[0,1)^d} \big|\phi_r(|P(\boldsymbol{e}(x))|^2)-1\big|\big|\log|P(\boldsymbol{e}(x))|\big|\,dx \\ &\le \left(\int_{[0,1)^d} \big|\phi_r(|P(\boldsymbol{e}(x))|^2)-1\big|^2\,dx\right)^{1/2}\left(\int_{[0,1)^d} \big|\log|P(\boldsymbol{e}(x))|\big|^2\,dx\right)^{1/2}\end{aligned}$$

by the definition of f_r and where we used the Cauchy–Schwarz inequality in the last step. The second integral on the final line is $\ll_{d,k} 1$ by Lemma A.3(ii). The first integral is

$$\int_{S(P,r)} \big|\phi_r(|P(\boldsymbol{e}(x))|^2) - 1\big|^2\,dx \le \mathrm{vol}(S(P,r)) \ll_{d,k} r^{1/(2(k-1))}$$

by Lemma A.3(i). We take the square root to complete the proof. □

Proof of Theorem A.1. As stated below Theorem A.1 we may assume $d \ge 2$. We may also assume that P is nonconstant. As we have seen in the proof of Lemma A.7, the condition $\rho(a) > \deg P$ guarantees $P(X^{a_1}, \dots, X^{a_d}) \ne 0$. Moreover, replacing P by $P/|P|$ leaves $m(P(X^{a_1}, \dots, X^{a_d})) - m(P)$ invariant. So it suffices to prove the theorem if $|P| = 1$.

We fix the parameters $b = 4d \ge d+1$ and $r = \rho(a)^{-1/4}/2 \le \frac{1}{2}$.

We write $|m(P(X^{a_1}, \dots, X^{a_d})) - m(P)|$ as $\int_0^1 \log|P(\boldsymbol{e}(as))|\,ds - \int_{[0,1)^d} \log|P(\boldsymbol{e}(x))|\,dx$ and find that it is at most

$$\left|\int_0^1 f_r(as)\,ds - \int_{[0,1)^d} f_r(x)\,dx\right| + \left|\int_0^1 (\log|P(\boldsymbol{e}(as))| - f_r(as))\,ds\right| + \left|\int_{[0,1)^d} (f_r(x)-\log|P(\boldsymbol{e}(x))|)\,dx\right|.$$

Then by Lemmas A.6, A.7, and A.8 this sum is

$$\ll_{d,k} \frac{|\log r|}{r^{2b}} \frac{(\deg P)^{b^2}}{\rho(a)^{b-d}} + r^{1/(k-1)}|\log r| + r^{1/(4(k-1))}.$$

By our choice of r and $\rho(a) \ge 2$, the sum is

$$\ll_{d,k} \frac{\log \rho(a)}{\rho(a)^{b-d-b/2}} (\deg P)^{b^2} + \frac{\log \rho(a)}{\rho(a)^{1/(4(k-1))}} + \frac{1}{\rho(a)^{1/(16(k-1))}}.$$

Finally, as $b = 4d$ the sum is

$$\ll_{d,k} (\deg P)^{16d^2} \frac{\log \rho(a)}{\rho(a)^d} + \frac{\log \rho(a)}{\rho(a)^{1/(4(k-1))}} + \frac{1}{\rho(a)^{1/(16(k-1))}}. \qquad \square$$

Appendix B. Recovering the theorem of Lind, Schmidt, and Verbitskiy

In this appendix we recover from our work a variant of Lind, Schmidt, and Verbitskiy's Theorem 1.1 [2013]. This variant is stated in the introduction as Theorem 1.2. Let $d \in \mathbb{N}$. For a finite subgroup $G \subset \mathbb{G}_m^d$, recall that we defined $\delta(G)$ in (1-5).

Lemma B.1. *Let G be a finite subgroup of $\mathbb{G}_m^d$. If $a \in \mathbb{Z}^d \backslash \{0\}$, then*

$$\frac{1}{\#G} \#\{\boldsymbol{\zeta} \in G : \boldsymbol{\zeta}^a = 1\} \le \frac{|a|}{\delta(G)}.$$

Proof. We will detect $\boldsymbol{\zeta}^a = 1$ using the character $\chi(\boldsymbol{\zeta}) = \boldsymbol{\zeta}^a$ of G. The image $\chi(G)$ is a cyclic subgroup of $\mathbb{C}^\times$ of order N, say. For $\boldsymbol{\zeta} \in G$, the sum $\sum_{k=0}^{N-1} \chi(\boldsymbol{\zeta})^k = 0$ equals N if $\boldsymbol{\zeta}^a = 1$ and vanishes otherwise. The number of solutions $\boldsymbol{\zeta} \in G$ of $\boldsymbol{\zeta}^a = 1$ is thus

$$\sum_{\boldsymbol{\zeta} \in G} \frac{1}{N} \sum_{k=0}^{N-1} \chi(\boldsymbol{\zeta}^k) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{\boldsymbol{\zeta} \in G} \chi(\boldsymbol{\zeta})^k = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\#G}{N} \sum_{\xi \in \chi(G)} \xi^k = \frac{\#G}{N}.$$

We conclude the proof as $\boldsymbol{\zeta}^{aN} = \chi(\boldsymbol{\zeta})^N = 1$ for all $\boldsymbol{\zeta} \in G$ and hence $N \ge \delta(G)/|a|$. $\square$

Lemma B.2. *Let G be a finite subgroup of $\mathbb{G}_m^d$:*

(i) *If $T \ge 1$, then*

$$\frac{1}{\#G} \#\{\boldsymbol{\zeta} \in G : \delta(\boldsymbol{\zeta}) \le T\} \le \frac{3^d T^{d+1}}{\delta(G)}.$$

(ii) *If $\kappa > 0$, then*

$$\frac{1}{\#G} \sum_{\boldsymbol{\zeta} \in G} \delta(\boldsymbol{\zeta})^{-\kappa} \le \frac{4^d}{\delta(G)^{\kappa/(d+1+\kappa)}}.$$

Proof. Any $\boldsymbol{\zeta} \in G$ with $\delta(\boldsymbol{\zeta}) \le T$ satisfies $\boldsymbol{\zeta}^a = 1$ for some $a \in \mathbb{Z}^d \backslash \{0\}$ and $|a| \le T$. The number of such a is at most $(2T+1)^d \le 3^d T^d$ and each a leads to at most $|a|\#G/\delta(G) \le T\#G/\delta(G)$ different $\boldsymbol{\zeta}$ by Lemma B.1. This implies (i).

For the second assertion we split up the elements in G into those with $\delta(\boldsymbol{\zeta}) \le T$ and those with $\delta(\boldsymbol{\zeta}) > T$; here $T \ge 1$ is a parameter to be chosen.

For the lower range, we use the trivial lower bound $\delta(\boldsymbol{\zeta}) \ge 1$ and part (i) to obtain

$$\frac{1}{\#G} \sum_{\substack{\boldsymbol{\zeta} \in G \\ \delta(\boldsymbol{\zeta}) \le T}} \delta(\boldsymbol{\zeta})^{-\kappa} \le \frac{3^d T^{d+1}}{\delta(G)}.$$

For the higher range, we have

$$\frac{1}{\#G} \sum_{\substack{\boldsymbol{\zeta} \in G \\ \delta(\boldsymbol{\zeta}) > T}} \delta(\boldsymbol{\zeta})^{-\kappa} \leq \frac{1}{T^{\kappa}}.$$

The lemma follows by taking the sum of these two bounds with $T = \delta(G)^{1/(d+1+\kappa)}$. □

Proof of Theorem 1.2. Without loss of generality we can assume that P is a polynomial.

Any finite subgroup of $\mathbb{G}_m^d$ is defined over $\mathbb{Q}$, i.e., it is mapped to itself under the action of the absolute Galois group of $\mathbb{Q}$, see [Bombieri and Gubler 2006, Corollary 3.2.15]. We decompose G into a disjoint union $G_1 \cup \cdots \cup G_m$ of Galois orbits. It is useful to fix a representative $\boldsymbol{\zeta}_i \in G_i$ for each $i \in \{1, \ldots, m\}$ and define $N_i = \operatorname{ord}(\boldsymbol{\zeta}_i)$. All elements in G_i have the same order and the Galois action is the natural action of $(\mathbb{Z}/N_i\mathbb{Z})^{\times}$ on G_i. Moreover, $\#G_i = \varphi(N_i)$. Note that δ is constant on each G_i as $\delta(\boldsymbol{\zeta}^{\sigma}) = \delta(\boldsymbol{\zeta})$ for all field automorphisms σ. Moreover, $P(\boldsymbol{\zeta}_i) \neq 0$ if and only if P does not vanish at any point of G_i; indeed P has rational coefficients by hypothesis.

Let $T \geq 1$ be a parameter depending on $\delta(G)$ and large in terms of P, d which we will fix in due time. We split our average (1-6) up into those $\boldsymbol{\zeta}$ with $\delta(\boldsymbol{\zeta}) \leq T$ and those with $\delta(\boldsymbol{\zeta}) > T$.

First, we will show that the sum

$$\frac{1}{\#G} \sum_{\substack{\boldsymbol{\zeta} \in G \\ \delta(\boldsymbol{\zeta}) \leq T, P(\boldsymbol{\zeta}) \neq 0}} \log|P(\boldsymbol{\zeta})| = \frac{1}{\#G} \sum_{\substack{i=1 \\ \delta(\boldsymbol{\zeta}_i) \leq T, P(\boldsymbol{\zeta}_i) \neq 0}}^{m} \sum_{\sigma \in (\mathbb{Z}/N_i\mathbb{Z})^{\times}} \log|P(\boldsymbol{\zeta}_i^{\sigma})| \tag{B-1}$$

is negligible. Say $P(\boldsymbol{\zeta}_i) \neq 0$. Then $P(\boldsymbol{\zeta}_i)$ lies in a number field of degree $\varphi(N_i)$ over $\mathbb{Q}$. So

$$\left| \sum_{\sigma \in (\mathbb{Z}/N_i\mathbb{Z})^{\times}} \log|P(\boldsymbol{\zeta}_i^{\sigma})| \right| \leq \sum_{\sigma \in (\mathbb{Z}/N_i\mathbb{Z})^{\times}} \left| \log|P(\boldsymbol{\zeta}_i^{\sigma})| \right| \leq 2\varphi(N_i) h(P(\boldsymbol{\zeta}_i)) \ll_P \varphi(N_i)$$

where we used basic properties of the height (2-3) and $P(\boldsymbol{\zeta}_i^{\sigma}) = P(\boldsymbol{\zeta}_i)^{\sigma}$ for all $\sigma \in \operatorname{Gal}(\mathbb{Q}(\boldsymbol{\zeta}_i)/\mathbb{Q})$. So the absolute value of (B-1) is at most

$$\ll_P \frac{1}{\#G} \sum_{\substack{i=1 \\ \delta(\boldsymbol{\zeta}_i) \leq T}}^{m} \varphi(N_i) \ll_P \frac{1}{\#G} \sum_{\substack{\boldsymbol{\zeta} \in G \\ \delta(\boldsymbol{\zeta}) \leq T}} 1 \ll_{d,P} \frac{T^{d+1}}{\delta(G)}. \tag{B-2}$$

by Lemma B.2(i).

The remaining sum is

$$\frac{1}{\#G} \sum_{\substack{i=1 \\ \delta(\boldsymbol{\zeta}_i) > T}}^{m} \sum_{\sigma \in (\mathbb{Z}/N_i\mathbb{Z})^{\times}} \log|P(\boldsymbol{\zeta}_i^{\sigma})|;$$

note that $P(\boldsymbol{\zeta}_i^\sigma) \neq 0$ for T large enough by the Manin–Mumford conjecture for $\mathbb{G}_m^d$. Using Theorem 1.1 we have

$$\frac{1}{\#G}\sum_{\substack{i=1\\ \delta(\boldsymbol{\zeta}_i)>T}}^{m} \varphi(N_i)(m(P)+O_{d,P}(\delta(\boldsymbol{\zeta}_i)^{-\kappa})) = \frac{1}{\#G}\Bigg(\sum_{\boldsymbol{\zeta}\in G:\delta(\boldsymbol{\zeta})>T} 1\Bigg)m(P) + O_{d,P}\Bigg(\frac{1}{\#G}\sum_{\boldsymbol{\zeta}\in G:\delta(\boldsymbol{\zeta})>T}\delta(\boldsymbol{\zeta})^{-\kappa}\Bigg)$$
$$= \Bigg(1-\frac{1}{\#G}\sum_{\boldsymbol{\zeta}\in G,\delta(\boldsymbol{\zeta})\leq T} 1\Bigg)m(P) + O_{d,P}\Big(\delta(G)^{-\kappa/(d+1+\kappa)}\Big)$$

where we used Lemma B.2(ii). The remaining average in the last line is $O_d(T^{d+1}/\delta(G))$ by Lemma B.2(i).

We combine this estimate with the first bound (B-2) to conclude that the average (1-6) equals

$$m(P) + O_{d,P}(T^{d+1}\delta(G)^{-1} + \delta(G)^{-\kappa/(d+1+\kappa)})$$

The theorem follows with the choice $T = c\delta(G)^{1/(2(d+1))}$ where $c \geq 1$ is sufficiently large in terms of d and P. The exponent κ in (1-6) is $\min\{\frac{1}{2}, \kappa/(d+1+\kappa)\}$ in the notation here. □

We leave to the interested reader the task of generalizing the previous theorem to polynomials defined over an arbitrary number field.

Acknowledgments

The authors thank Pierre Le Boudec for references regarding Gauss sums, Peter Sarnak for the reference to Le's work [2014], and Shouwu Zhang for pointing out the work of Chambert-Loir and Thuillier [2009]. We also thank the referee for carefully reading this text and for providing many valuable comments that led to improvements of the text and some simplifications. The authors thank François Brunault, Antonin Guilloux, Mahya Mehrabdollahi, and Riccardo Pengo for pointing out a mistake in an earlier attempt to prove Lemma A.3(i) and for making us aware of the work of Dobrowolski [2017]. Vesselin Dimitrov gratefully acknowledges support from the European Research Council via ERC grant GeTeMo 617129. Philipp Habegger has received funding from the Swiss National Science Foundation project n° 200020_184623.

References

[Agler et al. 2006] J. Agler, J. E. McCarthy, and M. Stankus, "Toral algebraic sets and function theory on polydisks", *J. Geom. Anal.* **16**:4 (2006), 551–562. MR Zbl

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

[Baker et al. 2008] M. Baker, S.-i. Ih, and R. Rumely, "A finiteness property of torsion points", *Algebra Number Theory* **2**:2 (2008), 217–248. MR Zbl

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

[Bombieri and Gubler 2006] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, New Math. Monogr. **4**, Cambridge Univ. Press, 2006. MR Zbl

[Bombieri et al. 2007] E. Bombieri, D. Masser, and U. Zannier, "Anomalous subvarieties: structure theorems and applications", *Int. Math. Res. Not.* **2007**:19 (2007), art. id. rnm057. MR Zbl

[Boyd 1981] D. W. Boyd, "Kronecker's theorem and Lehmer's problem for polynomials in several variables", *J. Number Theory* **13**:1 (1981), 116–121. MR Zbl

[Boyd 1998] D. W. Boyd, "Uniform approximation to Mahler's measure in several variables", *Canad. Math. Bull.* **41**:1 (1998), 125–128. MR Zbl

[Brunault et al. 2022] F. Brunault, A. Guilloux, M. Mehrabdollahei, and R. Pengo, "Limits of Mahler measures in multiple variables", preprint, 2022. arXiv 2203.12259

[Cassels 1959] J. W. S. Cassels, *An introduction to the geometry of numbers*, Grundl. Math. Wissen. **99**, Springer, 1959. MR Zbl

[Chambert-Loir and Thuillier 2009] A. Chambert-Loir and A. Thuillier, "Mesures de Mahler et équidistribution logarithmique", *Ann. Inst. Fourier* (*Grenoble*) **59**:3 (2009), 977–1014. MR Zbl

[Dimitrov 2016] V. Dimitrov, "Convergence to the Mahler measure and the distribution of periodic points for algebraic Noetherian $\mathbb{Z}^d$-actions", preprint, 2016. arXiv 1611.04664

[Dobrowolski 2017] E. Dobrowolski, "A note on Lawton's theorem", *Canad. Math. Bull.* **60**:3 (2017), 484–489. MR Zbl

[Dobrowolski and Smyth 2017] E. Dobrowolski and C. Smyth, "Mahler measures of polynomials that are sums of a bounded number of monomials", *Int. J. Number Theory* **13**:6 (2017), 1603–1610. MR Zbl

[Dubickas 1997] A. Dubickas, "Algebraic conjugates outside the unit circle", pp. 11–21 in *New trends in probability and statistics*, *IV* (Palanga, Lithuania, 1996), edited by A. Laurinčikas et al., VSP, Utrecht, Netherlands, 1997. MR Zbl

[Dubickas 2018] A. Dubickas, "On sums of two and three roots of unity", *J. Number Theory* **192** (2018), 65–79. MR Zbl

[Duke 2007] W. Duke, "A combinatorial problem related to Mahler's measure", *Bull. Lond. Math. Soc.* **39**:5 (2007), 741–748. MR Zbl

[Grafakos 2014] L. Grafakos, *Classical Fourier analysis*, 3rd ed., Grad. Texts in Math. **249**, Springer, 2014. MR Zbl

[Grayson 1984] D. R. Grayson, "Reduction theory using semistability", *Comment. Math. Helv.* **59**:4 (1984), 600–634. MR Zbl

[Güting 1961] R. Güting, "Approximation of algebraic numbers by algebraic numbers", *Michigan Math. J.* **8** (1961), 149–159. MR Zbl

[Habegger 2018] P. Habegger, "The norm of Gaussian periods", *Q. J. Math.* **69**:1 (2018), 153–182. MR Zbl

[Harman 1998] G. Harman, *Metric number theory*, Lond. Math. Soc. Monogr. (N.S.) **18**, Oxford Univ. Press, 1998. MR Zbl

[Hlawka 1971] E. Hlawka, "Discrepancy and Riemann integration", pp. 121–129 in *Studies in pure mathematics*, edited by L. Mirsky, Academic Press, London, 1971. MR Zbl

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc. Colloq. Publ. **53**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Kuipers and Niederreiter 1974] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Wiley, New York, 1974. MR Zbl

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Grad. Texts in Math. **211**, Springer, 2002. MR Zbl

[Laurent 1984] M. Laurent, "Équations diophantiennes exponentielles", *Invent. Math.* **78**:2 (1984), 299–327. MR Zbl

[Laurent et al. 1995] M. Laurent, M. Mignotte, and Y. Nesterenko, "Formes linéaires en deux logarithmes et déterminants d'interpolation", *J. Number Theory* **55**:2 (1995), 285–321. MR Zbl

[Lawton 1977] W. M. Lawton, "A generalization of a theorem of Kronecker", *J. Sci. Fat. Chiangmai Unit.* **4** (1977), 15–23. Zbl

[Lawton 1983] W. M. Lawton, "A problem of Boyd concerning geometric means of polynomials", *J. Number Theory* **16**:3 (1983), 356–362. MR Zbl

[Le 2014] T. Le, "Homology torsion growth and Mahler measure", *Comment. Math. Helv.* **89**:3 (2014), 719–757. MR Zbl

[Lind et al. 2010] D. Lind, K. Schmidt, and E. Verbitskiy, "Entropy and growth rate of periodic points of algebraic $\mathbb{Z}^d$-actions", pp. 195–211 in *Dynamical numbers*: *interplay between dynamical systems and number theory* (Bonn, Germany, 2009), edited by S. Kolyada et al., Contemp. Math. **532**, Amer. Math. Soc., Providence, RI, 2010. MR Zbl

[Lind et al. 2013] D. Lind, K. Schmidt, and E. Verbitskiy, "Homoclinic points, atoral polynomials, and periodic points of algebraic $\mathbb{Z}^d$-actions", *Ergodic Theory Dynam. Systems* **33**:4 (2013), 1060–1081. MR Zbl

[Mahler 1964] K. Mahler, "An inequality for the discriminant of a polynomial", *Michigan Math. J.* **11** (1964), 257–262. MR Zbl

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

[Myerson 1979] G. Myerson, "A combinatorial problem in finite fields, I", *Pacific J. Math.* **82**:1 (1979), 179–187. MR Zbl

[Myerson 1980] G. Myerson, "A combinatorial problem in finite fields, II", *Q. J. Math.* **31**:122 (1980), 219–231. MR Zbl

[Myerson 1986] G. Myerson, "Unsolved problems: how small can a sum of roots of unity be?", *Amer. Math. Monthly* **93**:6 (1986), 457–459. MR Zbl

[Rahman and Schmeisser 2002] Q. I. Rahman and G. Schmeisser, *Analytic theory of polynomials*, Lond. Math. Soc. Monogr. (N.S.) **26**, Oxford Univ. Press, 2002. MR Zbl

[Rosser and Schoenfeld 1962] J. B. Rosser and L. Schoenfeld, "Approximate formulas for some functions of prime numbers", *Illinois J. Math.* **6** (1962), 64–94. MR Zbl

[Sarnak and Adams 1994] P. Sarnak and S. Adams, "Betti numbers of congruence groups", *Israel J. Math.* **88**:1-3 (1994), 31–72. MR Zbl

[Schinzel 2000] A. Schinzel, *Polynomials with special regard to reducibility*, Encycl. Math. Appl. **77**, Cambridge Univ. Press, 2000. MR Zbl

[Schmidt 1995] K. Schmidt, *Dynamical systems of algebraic origin*, Progr. Math. **128**, Birkhäuser, Basel, 1995. MR Zbl

[Schmidt and Verbitskiy 2009] K. Schmidt and E. Verbitskiy, "Abelian sandpiles and the harmonic model", *Comm. Math. Phys.* **292**:3 (2009), 721–759. MR Zbl

[Smyth 1981] C. J. Smyth, "On measures of polynomials in several variables", *Bull. Austral. Math. Soc.* **23**:1 (1981), 49–63. MR Zbl

[Stuhler 1976] U. Stuhler, "Eine Bemerkung zur Reduktionstheorie quadratischer Formen", *Arch. Math.* (*Basel*) **27**:6 (1976), 604–610. MR Zbl

[Zannier 2012] U. Zannier, *Some problems of unlikely intersections in arithmetic and geometry*, Ann. of Math. Stud. **181**, Princeton Univ. Press, 2012. MR Zbl

Communicated by Antoine Chambert-Loir
Received 2019-10-01 Revised 2023-05-25 Accepted 2023-11-27

vesselin.dimitrov@gmail.com *School of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States*

philipp.habegger@unibas.ch *Departement Mathematik und Informatik, Universität Basel, Basel, Switzerland*

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2024 is US $525/year for the electronic version, and $770/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebra & Number Theory

Volume 18 No. 11 2024