Journal of Number Theory 292 (2027) 35–67

Contents lists available at ScienceDirect

Journal of Number Theory

journal homepage: www.elsevier.com/locate/jnt

General Section

Galois orbits of torsion points over polytopes near atoral sets

Chenying Lin

University of Regensburg, Universitätsstraße 31, Regensburg, 93053, Germany

ARTICLE INFO

Article history:
Received 26 January 2025
Received in revised form 16 April 2026
Accepted 16 May 2026
Available online 17 July 2026
Communicated by L. Smajlovic

Keywords:
Galois equidistribution
Torsion points
Heights
Discrepancy
Koksma's inequality
Polytopes

ABSTRACT

Given an essentially atoral Laurent polynomial P, we show an equidistribution theorem for the function $\log |P|$ on specific subsets of Galois orbits of torsion points of the d-dimensional algebraic torus $\mathbb{G}_m^d(\overline{\mathbb{Q}})$. The specific subsets under consideration are the preimages of d-dimensional polytopes within the hypercube $[0,1]^d$ under the cotropicalization map. This generalizes an equidistribution theorem of V. Dimitrov and P. Habegger, who considered only all Galois orbits that correspond to the entire hypercube $[0,1]^d$. In addition, we provide an estimate for the convergence speed of this equidistribution, expressed as a negative power of the strictness degree. Our approach is based on an alternative version of Koksma's inequality over polytopes.

As an application, we provide the convergence speed of heights on a sequence of projective points for a specific two-dimensional example, answering a question posed by R. Gualdi and M. Sombra.

Contents

E-mail address: Chenying.Lin@mathematik.uni-regensburg.de.

https://doi.org/10.1016/j.jnt.2026.05.014

1. Introduction

Let $\overline{\mathbb{Q}}$ be a fixed algebraic closure of $\mathbb{Q}$. Let $d \geq 1$ be an integer and let $\mathbb{G}_m^d$ denote the d-dimensional multiplicative group variety over $\overline{\mathbb{Q}}$, where $\mathbb{G}_m$ is the group variety $\operatorname{Spec}\overline{\mathbb{Q}}[T^{\pm 1}]$. Let $\iota : \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ be a fixed embedding; let also denote by ι the induced map $(\overline{\mathbb{Q}}^*)^d \hookrightarrow (\mathbb{C}^*)^d$. We identify

- $\mathbb{G}_m^d$ with its $\overline{\mathbb{Q}}$-points $\mathbb{G}_m^d(\overline{\mathbb{Q}}) = (\overline{\mathbb{Q}}^*)^d$,
- $\iota(x)$ with x for every $x \in \overline{\mathbb{Q}}$ and $\iota(\omega)$ with ω for every $\omega \in \mathbb{G}_m^d$.

Let $\omega \in \mathbb{G}_m^d$ be a torsion point. We define the *strictness degree* of ω to be

$$\delta(\omega) := \inf\left\{|a| : a \in \mathbb{Z}^d\backslash\{0\},\, \omega^a = 1\right\},$$

where $\omega^a = \omega_1^{a_1} \cdot \dots \cdot \omega_d^{a_d}$ for $\omega = (\omega_1, \dots, \omega_d)$ and $a = (a_1, \dots, a_d)$, and $|\cdot|$ is the maximum norm. By [BG06, Chapter 3], an algebraic subgroup of $\mathbb{G}_m^d$ of dimension $d-1$ takes the form $H_a = \{\omega \in \mathbb{G}_m^d : \omega^a = 1\}$ with $a \in \mathbb{Z}^d\backslash\{0\}$. Moreover, every proper algebraic subgroup of $\mathbb{G}_m^d$ is included in H_a for some a.

The well-known Bilu's equidistribution theorem for the Galois orbits of points of small heights [Bil97] implies that when $\delta(\omega)$ becomes arbitrarily large, the average of the Dirac measure at ω^σ over $\sigma \in \operatorname{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})$ weakly converges to the normalized Haar measure on $(S^1)^d := \{(z_1, \dots, z_d) \in (\mathbb{C}^*)^d : |z_1| = \dots = |z_d| = 1\}$. In other words, for a continuous function $f : (\mathbb{C}^*)^d \to \mathbb{R}$ with compact support, we have

$$\frac{1}{[\mathbb{Q}(\omega_\ell) : \mathbb{Q}]} \sum_{\sigma \in \operatorname{Gal}(\mathbb{Q}(\omega_\ell)/\mathbb{Q})} f(\omega_\ell^\sigma) \longrightarrow \int_{[0,1)^d} f(e^{\mathrm{i}2\pi x_1}, \dots, e^{\mathrm{i}2\pi x_d})\mathrm{d}x_1 \cdots \mathrm{d}x_d \tag{1.1}$$

for any sequence $(\omega_\ell)_{\ell \geq 1}$ with $\delta(\omega_\ell) \to \infty$.

In [DH24], V. Dimitrov and P. Habegger proved a quantitative version of the equidistribution result for functions $f = \log|P|$ with $P \in \overline{\mathbb{Q}}[T_1^{\pm 1}, \dots, T_d^{\pm 1}]$ an *essentially atoral* Laurant polynomial. For the definition of essentially atoral polynomial, we refer to Section 2. Note that the function $f = \log|P|$ has singularities at the vanishing set of P. Therefore, the result of Dimitrov and Habegger is an extension of (1.1). In this case, the integral in the right hand side of (1.1) becomes the *(logarithmic) Mahler measure*

$$m(P) \coloneqq \int\limits_{[0,1)^d} \log\left|P(e^{\mathrm{i}2\pi x_1},\dots,e^{\mathrm{i}2\pi x_d})\right|\mathrm{d}x_1\cdots\mathrm{d}x_d$$

of the polynomial P. The study of Mahler measure plays an important role in height theory, which studies the arithmetic complexity of points and cycles in varieties.

Note that given any torsion point $\omega = (\omega_1,\dots,\omega_d) \in \mathbb{G}_m^d$, for every j there exists a real number $x_j \in [0,1)$ such that $\omega_j = e^{\mathrm{i}2\pi x_j}$. The main result of V. Dimitrov and P. Habegger, as presented in [DH24, Theorem 1.1], considers the sum over all Galois conjugates of ω, i.e. the whole higher dimensional cube $[0,1)^d$. However, in certain applications, it becomes necessary to focus on the sum restricted to specific subsets of the full Galois orbits, i.e. a subset of $[0,1)^d$. Here we consider subsets obtained by partitioning the unit hypercube into polyhedral subsets. The following result is the main theorem of this article, generalizing [DH24, Theorem 1.1].

Denote $e(x) = (e^{\mathrm{i}2\pi x_1},\dots,e^{\mathrm{i}2\pi x_d})$ for $x = (x_1,\dots,x_d) \in \mathbb{R}^d$. We define the *cotropicalization map* by

$$\mathrm{ct} : (\mathbb{C}^*)^d \longrightarrow [0,1)^d, \quad (z_1,\dots,z_d) \longmapsto (x_1,\dots,x_d),$$

where $z_i = r_i e^{\mathrm{i}2\pi x_i}$ for $r_i > 0$ and $x_i \in [0,1)$.

Theorem 1.1. *Let d,k be integers and let $P \in \overline{\mathbb{Q}}[T_1^{\pm1},\dots,T_d^{\pm1}]\backslash\{0\}$ be essentially atoral with at most k nonzero terms. Let $\Delta \subset [0,1)^d$ be a polytope of dimension d. Then there exists a constant $\kappa = \kappa(d,k) > 0$ such that the following holds.*

Given any torsion point $\omega \in \mathbb{G}_m^d$ suppose that the strictness degree $\delta(\omega)$ is large enough. Then $P(\omega^\sigma) \neq 0$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})$, and as $\delta(\omega) \to \infty$ we have

$$\left| \frac{1}{[\mathbb{Q}(\omega):\mathbb{Q}]} \sum_{\substack{\sigma\in\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}) \\ \omega^\sigma\in\mathrm{ct}^{-1}(\Delta)}} \log\left|P(\omega^\sigma)\right| - \int\limits_{\Delta} \log\left|P(e(x))\right|\mathrm{d}x \right| \ll_{\Delta,P} \delta(\omega)^{-\kappa}.$$

Moreover, an explicit value of $\kappa = \kappa(d,k)$ can be computed by applying Remark 4.1 and an algorithm in the extended arXiv version [Lin24, Appendix A].

Other types of subsets of the full Galois orbits were considered in [DH24]. In [DH24, Theorem 8.8], the case in which the subset is made by the orbit of ω under the action of a subgroup of the Galois group, was already studied. When the subset is restricted to a finite subgroup of $\mathbb{G}_m^d$, this was discussed in [DH24, Theorem 1.2], which recovers [LSV13, Theorem 1.3] by a different approach. It is worth to point out that Theorem 1.1 cannot be covered by the previous two cases.

This theorem was inspired by the work of R. Gualdi and M. Sombra [GS23]: they studied the distribution of the height of the intersection between the projective line defined by the linear polynomial $x + y + z$ and its translate by a torsion point. This serves as an example of how the height of the solution set is concluded from the arithmetic complexity of a system. They computed the limit of such heights of a sequence of torsion points by Bilu's equidistribution theorem [Bil97] and proposed a question regarding a quantitative version for their particular case [GS23, Question 6.2]. It is shown in [GS23, Section 5] that the formula of this height is locally of the form $\frac{1}{[\mathbb{Q}(\omega)/\mathbb{Q}]}\sum_\sigma \log|P(\omega^\sigma)|$, where σ runs over the elements of the Galois group such that ω^σ corresponds to points in a triangle contained in unit square $[0,1)^2$. As an application of Theorem 1.1, we solve [GS23, Question 6.2] by deriving the following proposition.

Proposition 1.2. *Let $(\omega_\ell)_{\ell\geq 1}$ be a strict sequence of non-trivial torsion points in $\mathbb{G}_m^2$, i.e. any proper algebraic subgroup of $\mathbb{G}_m^2$ contains ω_ℓ for only finitely many values of ℓ. For every $\omega_\ell = (\omega_{\ell,1}, \omega_{\ell,2})$, let $\mathcal{P}(\omega_\ell) \in \mathbb{P}^2(\overline{\mathbb{Q}})$ be the solution of the system of linear equations*

$$x + y + z = x + \omega_{\ell,1}^{-1} y + \omega_{\ell,2}^{-1} z = 0$$

in the 2-dimensional projective space. Then as $\ell \to \infty$ we have

$$\left| \mathrm{h}(\mathcal{P}(\omega_\ell)) - \frac{2\zeta(3)}{3\zeta(2)} \right| \ll \delta(\omega_\ell)^{-1/(2^{61}\times 5^5)},$$

where ζ is the Riemann zeta function.

The following provides an overview of the proof of Theorem 1.1 and outlines the structure of this article.

All notations are stated in Section 2. Also, we have written down some formal definitions and preliminaries in this section.

A standard way from measure theory to estimate an equidistribution result is Koksma's inequality, see for example [Har98, Theorem 5.4]. It says that if a function $f : [0,1) \to \mathbb{R}$ has bounded *variation*, i.e. $V(f) := \sup_{0\leq a_0\leq\cdots\leq a_m<1} \sum_{i=1}^m |f(a_i) - f(a_{i-1})|$ is bounded, then for every finite sequence $(x_1, x_2, \ldots, x_n)$ in $[0,1)$ we have

$$\left| \frac{1}{n}\sum_{i=1}^n f(x_i) - \int_{[0,1]} f(x)\mathrm{d}x \right| \leq D(x_1,\ldots,x_n)V(f), \tag{1.2}$$

where

$$D(x_1,\ldots,x_n) := \sup_{0\leq a<b\leq 1} \left| \frac{\sharp\{i : a \leq x_i < b\}}{n} - (b-a) \right|$$

is called the *discrepancy* of $(x_1,\ldots,x_n)$.

We say a sequence $(x_1, x_2, \ldots, x_n, \ldots)$ in $[0,1)$ is *equidistributed* if $D(x_1,\ldots,x_n) \to 0$ as $n \to \infty$. Hence, if the quantity $D(x_1,\ldots,x_n)$ is smaller, then the finite sequence $(x_1,\ldots,x_n)$ is more "equidistributed" in $[0,1)$.

However, in our case, we need to deal with functions having logarithmic singularities, which don't have bounded variation. It leads us to estimate the log function using the "bounded log function" $\log\max\{r, x\}$ for $r > 0$.

The bounded log function also appears in [DH24], where some useful property, that we recall and enrich in Section 4, is shown. By the triangle inequality, we have

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{\substack{\sigma\in\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}) \\ \omega^\sigma\in\mathrm{ct}^{-1}(\Delta)}} \log|P(\omega^\sigma)| - \int_\Delta \log|P(e(x))|\mathrm{d}x \right| \\
&\qquad \le \left| \frac{1}{n} \sum_{\substack{\sigma\in\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}) \\ \omega^\sigma\in\mathrm{ct}^{-1}(\Delta)}} \log\max\{r, |P(\omega^\sigma)|\} - \int_\Delta \log\max\{r, |P(e(x))|\}\mathrm{d}x \right| \\
&\qquad\qquad + \frac{1}{n} \left| \sum_{\substack{\sigma\in\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q}) \\ \omega^\sigma\in\mathrm{ct}^{-1}(\Delta)}} \big(\log\max\{r, |P(\omega^\sigma)|\} - \log|P(\omega^\sigma)|\big) \right| \\
&\qquad\qquad\qquad + \left| \int_\Delta \big(\log\max\{r, |P(e(x))|\} - \log|P(e(x))|\big)\mathrm{d}x \right|. \qquad (1.3)
\end{aligned}$$

For the first difference in the right hand side of (1.3), we need to develop a version of Koksma's inequality for polytopes to estimate it, because the standard generalization of the Koksma's inequality (1.2) only works for the whole d-dimensional interval $[0, 1)^d$. There are some generalization of Koksma's inequality into more general sets like [BCGT13] and [Har10], but both of them have some shortcomings for our problems: [BCGT13] contains the partial derivatives of functions that are difficult to compute in our case, and [Har10] somehow hides the geometric meaning of discrepancy. Therefore, based on the method in [DH24, Proposition 7.1], in Section 3, a version of Koksma's inequality over polytope is developed and applied to our bounded log function. As [DH24, Proposition 7.1] only works for continuous function, we will construct "continuous characteristic functions" over polytopes for our purpose.

For the second difference in the right hand side of (1.3), we use the triangle inequality again to divide it into two sums. The sum regarding $\log\max\{r, x\}$ is easy to estimate because the function is bounded. For the sum involves log functions, the main theorem in [DH24] will be applied. The estimation for the third difference is similar with the second one.

In Section 4, we put everything together to give a complete proof of Theorem 1.1.

Section 5 contains a proof of Proposition 1.2, which gives an application of Theorem 1.1 and answers the question in the paper of R. Gualdi and M. Sombra we mentioned above [GS23, Question 6.2].

2. Preliminaries

2.1. Conventions and notations

Throughout the article we use the following notations.

- $|\cdot|$: maximum norm in $\mathbb{R}^d$ and $\mathrm{GL}_d(\mathbb{Z})$. If $A = (a_{ij}) \in \mathrm{GL}_d(\mathbb{Z})$, then $|A| = \max_{i,j} |a_{i,j}|$.
- $|\cdot|_2$: Euclidean norm in $\mathbb{R}^d$.
- $\langle \cdot, \cdot \rangle$: Euclidean inner product on $\mathbb{R}^d$.
- A *cubic ball* in $\mathbb{R}^d$ is a subset of the form $\{y \in \mathbb{R}^d : |x - y| \le r\}$ for $x \in \mathbb{R}^d$ and $r \in \mathbb{R}$, where x is called the center of the ball and r is called the radius of the ball.
- $\mathrm{diam}(A)$: the diameter of a subset $A \subset \mathbb{R}^d$ with respect to the maximum norm, which means $\mathrm{diam}(A) = \max\{|x - y| : x, y \in A\}$.
- μ: Lebesgue measure on $\mathbb{R}^d$.
- χ_A: the characteristic function of a subset $A \subset \mathbb{R}^d$, defined as the real-valued function given by $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$.
- $e(x) := (e^{\mathrm{i}2\pi x_1}, \dots, e^{\mathrm{i}2\pi x_d})$ for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.
- $m(P)$: the (logarithmic) Mahler measure for a Laurent polynomial P in d variables over $\mathbb{C}$ defined as

$$m(P) := \int\limits_{[0,1)^d} \log \left|P(e(x))\right| \mathrm{d}\mu(x).$$

- $|P|$: the maximal norm of the coefficient vector of a Laurent polynomial P.
- $\overline{\mathbb{Q}}$: the fixed algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$.
- For $\omega = (\omega_1, \dots, \omega_d) \in (\overline{\mathbb{Q}}^*)^d$, $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$ and $A = (a_{ij}) \in \mathrm{GL}_d(\mathbb{Z})$, we set

$$\omega^a := \omega_1^{a_1} \cdots \omega_d^{a_d} \quad \text{and} \quad \omega^A := (\omega_1^{a_{11}} \cdots \omega_d^{a_{d1}},\ \dots,\ \omega_1^{a_{1d}} \cdots \omega_d^{a_{dd}}).$$

- $\delta(\omega)$: the strictness degree of $\omega \in (\overline{\mathbb{Q}}^*)^d$ defined as

$$\inf \left\{ |a| : a \in \mathbb{Z}^d \backslash \{0\},\, \omega^a = 1 \right\}.$$

- $S^1 := \{z \in \mathbb{C} : |z| = 1\}$.
- For $z \in S^1$ and $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$, we set $z^a = (z^{a_1}, \dots, z^{a_d})$.

- $\ll_{\bullet}$: the Vinogradov's notation where the constants implicit in it only depend on $\bullet$. For example, for a polynomial P, $\ll_P$ denotes the Vinogradov's notation, in which the implicit constants only depend on P.
- ct: the cotropicalization map defined by

$$(\mathbb{C}^*)^d \longrightarrow [0,1)^d, \quad (z_1, \dots, z_d) \longmapsto (x_1, \dots, x_d),$$

where $z_i = r_i e^{\mathrm{i}2\pi x_i}$ for $r_i > 0$ and $x_i \in [0,1)^d$.

2.2. *Convex geometry*

Let $d \geq 1$ be an integer. In the statement of Theorem 1.1, we consider subsets of Galois orbits of torsion points that lie in the preimages of polytopes within $[0,1)^d$ under the cotropicalization map. Therefore, we need to introduce some basic definitions and properties from convex geometry.

Let $N \cong \mathbb{Z}^d$ be a lattice of rank d and $M := N^\vee$ be its dual. Then $N_\mathbb{R} := N \otimes \mathbb{R} \cong \mathbb{R}^d$ is a real vector space of dimension d and $M_\mathbb{R} := M \otimes \mathbb{R} = N_\mathbb{R}^\vee$ is its dual space.

A subset $C \subset N_\mathbb{R}$ is *convex* if it contains all the segments with vertices in C. In other words, for any $u_1, u_2 \in C$, we have the line segment

$$\overline{u_1 u_2} := \{tu_1 + (1-t)u_2 : t \in [0,1]\} \subset C.$$

An *affine space* is of the form $\mathbb{R}v_1 + \cdots + \mathbb{R}v_n + p$, where $\mathbb{R}v_1 + \cdots + \mathbb{R}v_n$ is a linear space with $v_i \in \mathbb{R}^d$ and $p \in \mathbb{R}^d$. Let $C \subset \mathbb{R}^d$ be a convex subset. The *affine hull* of C, denoted $\mathrm{aff}(C)$, is the minimal affine space which contains it. The *dimension* of C is defined as the dimension of its affine hull. The *relative interior* of C, denoted $\mathrm{ri}(C)$, is defined as the interior of C relative to its affine hull.

A convex subset $F \subset C$ is called a *face* of C if, for every closed line segment $\overline{u_1 u_2} \subset C$ such that $\mathrm{ri}(\overline{u_1 u_2}) \cap F \neq \varnothing$, the inclusion $\overline{u_1 u_2} \subset F$ holds. A face of C of codimension 1 is called a *facet*. A face of C of dimension 0 is called a *vertex*. A non-empty subset $F \subset C$ is called an *exposed face* of C if there exists $x \in M_\mathbb{R}$ such that

$$F = \{u \in C : \langle x, u \rangle \leq \langle x, v \rangle, \forall v \in C\}.$$

Any exposed face of C is a face.

Pick $x \in M_\mathbb{R}$ and $c \in \mathbb{R}$. Then the set

$$H_{x,c} := \{u \in N_\mathbb{R} : \langle x, u \rangle \geq c\}$$

is called a *closed half-space*. A *polyhedron* in $N_\mathbb{R}$ is the intersection of finitely many closed half-spaces in $N_\mathbb{R}$. A *polytope* in $N_\mathbb{R}$ is a bounded polyhedron in $N_\mathbb{R}$. A polytope can be equivalently defined as the *convex hull* of finitely many points. Indeed, a subset $P \subset N_\mathbb{R}$ is a polytope if and only if there exists $p_1, \dots, p_s \in N_\mathbb{R}$ such that

$$P = \operatorname{conv}(p_1, \ldots, p_s) := \left\{ \sum_{i=1}^{s} \lambda_i p_i : \lambda_i \geq 0,\ \sum_{i=1}^{s} \lambda_i = 1 \right\}.$$

We refer to [Brø83, Theorem 1.9.2] for the equivalence. The structure of the faces of P is given in [Brø83, Theorem 1.7.5]. Specifically, every face of P is of the form

$$\{u \in P : \langle x, u \rangle = c\}$$

for some $x \in M_{\mathbb{R}}$ and $c \in \mathbb{R}$.

Definition 2.1. Let $P \subset N_{\mathbb{R}}$ be a polytope of the whole dimension d and F be a facet of P. Let v_F be a vector orthogonal to F with $|v_F|_2 = 1$. We set

$$P_F := \{u + \lambda v_F : u \in F,\ 0 \leq \lambda \leq 1\}.$$

Then $P_F \subset N_{\mathbb{R}}$ is a polytope of dimension d. We call $\operatorname{vol}(F) := \mu(P_F)$ the *volume* of F and the sum of the volumes of all facets of P the *surface area* of P.

To prove our version of Koksma's inequality over polytopes, we will construct a continuous version of characteristic function of each polytope in Section 3. The idea is to shrink the polytope by an arbitrary small scalar and connect the new smaller polytope and the original polytope by some linear functions. Therefore, we also introduce the notion of piecewise affine function.

Recall that an *affine function* $f : A \to B$ from an affine space A to an affine space B is a function which satisfies

$$f(\lambda a + (1 - \lambda)c) = \lambda f(a) + (1 - \lambda) f(c)$$

for $a, c \in A$ and $\lambda \in \mathbb{R}$. In the more specific case where $A = L + x$ for a linear subspace L of $\mathbb{R}^d$ and a point $x \in \mathbb{R}^d$, and $B = \mathbb{R}^d$, it follows from [Brø83, Section 1.1] that there exists a linear function $g : L \to \mathbb{R}^d$ and a point $y \in \mathbb{R}^d$ such that $f(z + x) = g(z) + y$ for all $z \in L$.

Let $P \subset N_{\mathbb{R}}$ be a polytope. A function $f : P \to \mathbb{R}$ is *piecewise affine* if there is a finite cover of P by closed subsets such that the restriction of f to each of these closed subsets is the restriction of an affine function. Notice that every piecewise affine function is continuous. Hence, if we construct a piecewise affine characteristic function, then it will be continuous.

To show that our continuous version of characteristic function is piecewise affine, we first need to understand the set difference between the original polytope and the shrunk polytope. It will be described using a *complete fan* consisting of *polyhedral cones*.

A *polyhedral cone* is a polyhedron σ such that $\lambda\sigma = \sigma$ for all $\lambda \geq 0$. A *fan* Σ in $N_{\mathbb{R}}$ is a finite collection of polyhedral cones in $N_{\mathbb{R}}$ that such that

- For all $\sigma \in \Sigma$, every face of σ is also in Σ and σ does not contain any line.
- For all $\sigma, \sigma' \in \Sigma$, we have $\sigma \cap \sigma'$ is a face of both σ and σ'.

If $\bigcup_{\sigma\in\Sigma} \sigma = N_{\mathbb{R}}$, then we say Σ is *complete*.

2.3. Discrepancy

Let $n \geq 1$ be an integer and let $x_1, \ldots, x_n$ be points in the hypercube $[0,1)^d$. We define the *discrepancy* of the finite sequence $(x_1, \ldots, x_n)$ to be

$$D(x_1, \ldots, x_n) \coloneqq \sup_B \left| \frac{\sum_{i=1}^n \chi_B(x_i)}{n} - \mu(B) \right|,$$

where B ranges over all products $\prod_{i=1}^d [a_i, b_i)$ with $0 \leq a_i < b_i \leq 1$. By definition, the discrepancy lies in $[0,1]$. In some references, however, for example in [Har98], the discrepancy is defined without normalization by n, thus allowing values greater than 1.

In this paper, we consider the discrepancy of the arguments of the Galois orbit of a torsion point $\omega \in \mathbb{G}_m^d$. It has been estimated in [DH24, Proposition 3.3]:

Proposition 2.2. *Let $\omega \in \mathbb{G}_m^d$ be a torsion point of order n, and let $\{\omega^\sigma : \sigma \in \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})\} = \{e(x_i) : 1 \leq i \leq n\}$, where all $x_i \in [0,1)^d$. Then we have*

$$D(x_1, \ldots, x_n) \ll_d \frac{(\log 2\delta(\omega))^{d-1} \log\log 3\delta(\omega)}{\delta(\omega)^{1/2}}$$

as $\delta(\omega) \to \infty$.

As we are concerned about polytopes, the following notion of isotropic discrepancy will also be useful. The reference for it is [KN74, Chapter 2].

Definition 2.3. Let $d, n \in \mathbb{N}$ and $x_1, \ldots, x_n \in [0,1)^d$. The *isotropic discrepancy* of the finite sequence $(x_1, \ldots, x_n)$ is

$$J(x_1, \ldots, x_n) \coloneqq \sup_C \left| \frac{\sum_{i=1}^n \chi_C(x_i)}{n} - \mu(C) \right|$$

where C ranges over all convex subsets of $[0,1)^d$.

Proposition 2.4. *Let $d, n \in \mathbb{N}$ and $x_1, \ldots, x_n \in [0,1)^d$. Then we have*

$$D(x_1, \ldots, x_d) \leq J(x_1, \ldots, x_n) \leq (4d\sqrt{d} + 1) D(x_1, \ldots, x_n)^{1/d}.$$

Proof. The first inequality $D(x_1, \ldots, x_d) \leq J(x_1, \ldots, x_n)$ follows immediately from the definitions. For the second inequality, we refer the reader to [KN74, Theorem 2.1.6]. □

2.4. Essentially atoral polynomials

The main theorem Theorem 1.1 in this article is a quantitative version of equidistribution of function $\log|P(e(x))|$ where P is an *essentially atoral Laurent polynomial.* We give the definition of essentially atoral Laurent polynomial here.

Definition 2.5. A *torsion coset* of $\mathbb{G}_m^d$ is a coset ωH, where $\omega \in \mathbb{G}_m^d$ is a torsion point and H is a connected algebraic subgroup of $\mathbb{G}_m^d$. A torsion coset is *proper* if it is not equal to $\mathbb{G}_m^d$. A non-zero Laurent polynomial $P \in \mathbb{C}[T_1^{\pm 1}, \dots, T_d^{\pm 1}]$ is called *essentially atoral* if the Zariski closure of

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

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

Example 2.6. If $d = 1$, then a non-zero Laurent polynomial $P \in \mathbb{C}[T^{\pm 1}]$ is essentially atoral if and only if it does not vanish at any non-torsion point in S^1.

Example 2.7. Let $d \geq 2$ and $P(T_1, \dots, T_d) = T_1^{m_1} \cdots T_d^{m_d} - T_1^{n_1} \cdots T_d^{n_d}$ be a nonzero Laurent binomial over $\mathbb{C}$, where all m_i, n_i are integers. Then P is essentially atoral. Indeed, let

$$a \coloneqq (m_1 - n_1, \dots, m_d - n_d)$$

and

$$H_a \coloneqq \{(z_1, \dots, z_d) \in (\mathbb{C}^*)^d : z_1^{a_1} \cdots z_d^{a_d} = 1\}.$$

By [BG06, Proposition 3.2.10], there exists a finite subgroup $S \subset S^1$ such that $H_a \cong S \times \mathbb{G}_m^{d-1}$. Therefore, the Zariski closure of

$$\big\{(z_1, \dots, z_d) \in \mathbb{C}^d : P(z_1, \dots, z_d) = 0\big\} \cap (S^1)^d = H_a \cap (S^1)^d \cong S \times (S^1)^{d-1}$$

is isomorphic to a finite union of $\mathbb{G}_m^{d-1}$ and each $\mathbb{G}_m^{d-1}$ is a proper torsion coset of $\mathbb{G}_m^d$.

2.5. Height

Height functions assign a notion of arithmetic size or complexity to mathematical objects, such as points and cycles in varieties. In this paper, we are interested in Weil height of points in projective space.

Let K be a number field. The notion of absolute values of K will be used in the definition of height functions. A standard reference for absolute values is [Neu99, Chapter

2]. Here, we normalize the absolute values of K as follows. Let v be a non-archimedean place of K. Then there exists a prime number p such that v extends the p-adic place. Recall that $|p|_p = \frac{1}{p}$. For any $x \in K$, we define

$$|x|_v := \left|N_{K_v/\mathbb{Q}_p}(x)\right|_p^{1/[K:\mathbb{Q}]}$$

where K_v (resp. $\mathbb{Q}_p$) is the completion of K with respect to v (resp. $\mathbb{Q}$ with respect to the p-adic place) and $N_{K_v/\mathbb{Q}_p}$ is the norm of the finite extension $K_v/\mathbb{Q}_p$.

Let M_K be the set of all places on K. For example, if $K = \mathbb{Q}$, then $M_{\mathbb{Q}}$ consists of an archimedean place ∞ and non-archimedean places v_p corresponding to prime numbers p.

Definition 2.8. For a point $x = [x_0 : \cdots : x_d] \in \mathbb{P}^d(\overline{\mathbb{Q}})$ in d-dimensional projective space over $\overline{\mathbb{Q}}$, let K be a number field containing all coordinates of x. Then the *height* of x is defined as

$$\mathrm{h}(x) := \sum_{w \in M_K} \max_j \log |x_j|_w.$$

For each place $w \in M_K$, the *w-adic height* of the homogeneous coordinates vector $(x_0, \ldots, x_d)$ of x is defined as

$$\mathrm{h}_w(x_0, \ldots, x_d) := \max_j \log |x_j|_w.$$

For each place $v \in M_{\mathbb{Q}}$, the *v-adic height* of the homogeneous coordinates vector $(x_0, \ldots, x_d)$ of x is defined as

$$\mathrm{h}_v(x_0, \ldots, x_d) = \sum_{w \text{ extends } v} \mathrm{h}_w(x_0, \ldots, x_d).$$

The height is well-defined in this way, see [BG06, Chapter 1] for details.

3. Koksma's inequality over polytopes

Throughout this section, we identify points and vectors in the real vector space $\mathbb{R}^d$.

Definition 3.1. Let $f : B \to \mathbb{R}$ be a function, where B is a non-empty subset of $\mathbb{R}^d$. Then the *modulus of continuity* of f is defined by

$$\rho(f, t) = \sup\left\{|f(x) - f(y)| : x, y \in B,\ |x - y| \leq t\right\}, \quad \forall t \geq 0.$$

Let $f : [0, 1]^d \to \mathbb{R}$ be a continuous function, and let $(x_1, \ldots, x_n)$ be a finite sequence in $[0, 1)^d$ with discrepancy $D = D(x_1, \ldots, x_n)$. Then [DH24, Proposition 7.1] gives the following proposition, which shares the same idea with simplest Koksma's inequality in dimension one (1.2), i.e. the upper bound of

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \int\limits_{[0,1)^d} f \mathrm{d}\mu\right|$$

is related to the discrepancy D and the variation of f.

Proposition 3.2. *Let $f : [0,1]^d \to \mathbb{R}$ be a continuous function and let $(x_1, \ldots, x_n)$ be a finite sequence in $[0,1)^d$ with discrepancy $D = D(x_1, \ldots, x_n)$. Then*

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \int\limits_{[0,1)^d} f \mathrm{d}\mu\right| \le (1 + 2^{d+1})\rho\left(f, D^{1/(d+1)}\right).$$

Now let $\Delta \subset [0,1]^d$ be a polytope of full dimension, i.e. $\dim(\Delta) = d$. In this section, we would like to find an upper bound for the difference

$$\frac{1}{n}\sum_{\substack{i=1 \\ x_i \in \Delta}}^{n} f(x_i) - \int\limits_{\Delta} f \mathrm{d}\mu = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\chi_\Delta(x_i) - \int\limits_{[0,1)^d} f\chi_\Delta \mathrm{d}\mu,$$

where χ_Δ is the characteristic function of Δ such that $\chi_\Delta(x) = 1$ if $x \in \Delta$ and $\chi_\Delta(x) = 0$ otherwise.

In order to apply Proposition 3.2 to our case, we need to construct a continuous version of characteristic function and compute its modulus of continuity. The proof below works for the case $d \ge 2$. However, the result we proved also works for $d = 1$ and the proof is essentially the same. The only difference is that if $d = 1$ then we don't need to intersect the polytope with 2-dimensional planes to reduce some problems to the 2-dimensional case.

3.1. Construction of "continuous characteristic function"

Consider $0 < \epsilon < 1$. Let $\Delta \subset \mathbb{R}^d$ be a polytope. Let us construct a continuous function $\chi^c_{\Delta,\epsilon}$ to estimate the characteristic function χ_Δ. The main idea is to shrink the polytope Δ to Δ_ϵ in a linear way. The function $\chi^c_{\Delta,\epsilon}$ will be 0 outside Δ and 1 inside Δ_ϵ. We will then connect the facets of $(\Delta, 0)$ and $(\Delta_\epsilon, 1)$ by linear functions. In order to define the function, we need to pick a special point in the interior of Δ. Recall that a cubic ball in $\mathbb{R}^d$ is of the form $\{y \in \mathbb{R}^d : |x - y| \le r\}$ for $x \in \mathbb{R}^d$ and $r \in \mathbb{R}_{>0}$ using the maximum norm $|\cdot|$.

Definition 3.3. An *inscribed cubic ball* of a polytope is a ball of maximal radius that is contained within the polytope.

Proposition 3.4. *Every polytope has an inscribed cubic ball.*

Proof. Let $P \subset \mathbb{R}^d$ be a polytope. If P is a point, then the only point in the inscribed cubic ball is P. We suppose that P is not a point. First, we claim that for every point $x \in \mathrm{ri}(P)$ there exists a ball centered at x of maximal radius that is contained within P. Indeed, for every $x \in \mathrm{ri}(P)$, there exists $r_x \geq 0$ such that

$$r_x = \min \{|x - y| : y \in P \setminus \mathrm{ri}(P)\},$$

since

$$P \setminus \mathrm{ri}(P) \to \mathbb{R}, \quad y \mapsto |x - y|$$

is a continuous function on the compact set $P \setminus \mathrm{ri}(P)$. Then the ball centered at x with radius r_x is contained within P, and it has maximal radius among all balls centered at x that are contained within P. The claim follows.

Consider the function

$$r : P \to \mathbb{R}, \quad x \mapsto \min \{|x - y| : y \in P \setminus \mathrm{ri}(P)\}.$$

Note that $r(x) = r_x$ if $x \in \mathrm{ri}(P)$ and $r(x) = 0$ if $x \in P \setminus \mathrm{ri}(P)$. Let us show that r is continuous. Take $x_0, x \in P$. For any $y \in P \setminus \mathrm{ri}(P)$, we have

$$|x - y| \leq |x - x_0| + |x_0 - y|$$

and thus

$$\min \{|x - y| : y \in P \setminus \mathrm{ri}(P)\} \leq \min \{|x_0 - y| : y \in P \setminus \mathrm{ri}(P)\} + |x - x_0|.$$

Hence $r(x) - r(x_0) \leq |x - x_0|$. Symmetrically, we have $r(x_0) - r(x) \leq |x - x_0|$ and therefore r is continuous. The proposition follows from the fact that P is compact. $\square$

Definition 3.5. Let $P \subset \mathbb{R}^d$ be a polytope. Then the *inner radius* of P, denoted $\mathrm{inrad}(P)$, is the radius of an inscribed cubic ball of P.

Let x_c be the center of an inscribed cubic ball of Δ. For each $0 < \epsilon < 1$, we define

$$\begin{aligned} \varphi_\epsilon &: \mathbb{R} \to \mathbb{R}^d \\ & x \mapsto x_c + (1 - \epsilon)(x - x_c) = (1 - \epsilon)x + \epsilon x_c, \end{aligned}$$

which is the shrinking of $\mathbb{R}^d$ by a scalar $1-\epsilon$ with a specified center at x_c. Set $\Delta_\epsilon := \varphi_\epsilon(\Delta)$.

Lemma 3.6. *For each $0 < \epsilon < 1$, the set Δ_ϵ is a polytope contained in Δ.*

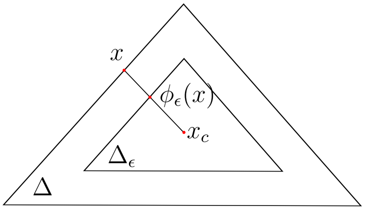

Fig. 1. An Example of $\Delta_\epsilon \subset \Delta$ in dimension 2. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Proof. It is stated in [Brø83, Exercise 2.8.1] that the image of a polytope under an affine function remains a polytope. Since φ_ϵ is an affine function, the set $\Delta_\epsilon = \varphi_\epsilon(\Delta)$ is a polytope. By the convexity of Δ and the fact that $x_c \in \Delta$, we have $\Delta_\epsilon \subset \Delta$. See Fig. 1. $\square$

Remark 3.7. One can prove that $\Delta_\epsilon \to \Delta$ as $\epsilon \to 0$ in terms of *Hausdorff distance*. See [RW98, Chapter 4] for more details on Hausdorff distance. Indeed, the Hausdorff distance between Δ and Δ_ϵ is

$$d_H(\Delta_\epsilon, \Delta) := \max\left\{\sup_{x\in\Delta} d(x, \Delta_\epsilon),\ \sup_{y\in\Delta_\epsilon} d(\Delta, y)\right\},$$

where

$$\begin{aligned}
&d(\Delta, y) := \inf_{x\in\Delta} |x-y| = 0, \quad \forall y \in \Delta_\epsilon,\\
&d(x, \Delta_\epsilon) := \inf_{y\in\Delta_\epsilon} |x-y| \le |x-\varphi_\epsilon(x)| = \epsilon|x-x_c|, \quad \forall x \in \Delta.
\end{aligned}$$

Therefore, we have

$$d_H(\Delta_\epsilon, \Delta) \le \epsilon \cdot \sup_{x\in\Delta} |x-x_c| \le \epsilon \cdot \operatorname{diam}(\Delta).$$

It follows that $d_H(\Delta, \Delta_\epsilon) \to 0$ as $\epsilon \to 0$.

Before constructing the continuous version of characteristic function, let us first study the set difference $\Delta \setminus \operatorname{ri}(\Delta_\epsilon)$. It will help us to show that the constructed function is piecewise affine and thus continuous.

Definition 3.8 *(truncated hyperpyramid).* Let $B \subset \mathbb{R}^d$ be a $(d-1)$-dimensional polytope that is the convex hull of the points $p_1, \ldots, p_s \in \mathbb{R}^d$, and let $x \in \mathbb{R}^d \setminus B$. The polytope

$$P := \operatorname{conv}(p_1, \ldots, p_s, x)$$

is called a *hyperpyramid* with *base* B and *peak* x. Let $H \subset \mathbb{R}^d$ be a hyperplane such that H is parallel to B and $H \cap P \notin \{\varnothing, \{x\}, B\}$. Then there exist $q_1, \ldots, q_s \in \mathbb{R}^d$ such that $H \cap P = \operatorname{conv}(q_1, \ldots, q_s)$. The polytope

$$\operatorname{conv}(p_1, \dots, p_s, q_1, \dots, q_s)$$

is called a *truncated hyperpyramid* with *top* $H \cap P$, *base* B and *peak* x.

Lemma 3.9. *Let F be a facet of Δ. Then $\varphi_\epsilon(F)$ and F are parallel for every $0 < \epsilon < 1$.*

Proof. Let $v \in \mathbb{R}^d$ such that v is orthogonal to F. Then for any $x, y \in F$, we have $\langle v, x - y\rangle = 0$. Thus $\langle v, \varphi_\epsilon(x) - \varphi_\epsilon(y)\rangle = (1 - \epsilon)\langle v, x - y\rangle = 0$, which means that v is orthogonal to $\varphi_\epsilon(F)$. It follows that F and $\varphi_\epsilon(F)$ are parallel. $\square$

Lemma 3.10. *Let $F_1, \dots, F_e$ be all the facets of Δ. For each $0 < \epsilon < 1$, the set $\Delta \setminus \operatorname{ri}(\Delta_\epsilon)$ can be written as a union of truncated hyperpyramids $P_1, \dots, P_e$, where each P_i has top $\varphi_\epsilon(F_i)$, base F_i and peak x_c.*

Proof. For each $1 \le i \le e$, let us consider the polyhedral cone

$$\angle(F_i) := \{t(x - x_c) : x \in F_i,\ t \ge 0\}.$$

Let Σ be the collection of $\angle(F_1)$, $\dots$, $\angle(F_e)$ and all their faces. Then Σ is a complete fan. By the completeness of Σ, we have

$$\Delta \setminus \operatorname{ri}(\Delta_\epsilon) = \big(\angle(F_1) \cap (\Delta \setminus \operatorname{ri}(\Delta_\epsilon))\big) \cup \dots \cup \big(\angle(F_e) \cap (\Delta \setminus \operatorname{ri}(\Delta_\epsilon))\big).$$

It follows from Lemma 3.9 that each $\angle(F_i) \cap (\Delta \setminus \operatorname{ri}(\Delta_\epsilon))$ is a truncated hyperpyramid with top $\phi_\epsilon(F_i)$, base F_i and peak x_c. $\square$

The following lemma will be used when we prove our version of Koksma's inequality over polytopes.

Lemma 3.11. *Let $0 < \epsilon < 1$. Let $S(\Delta)$ be the surface area of $\Delta \subset \mathbb{R}^d$. Then*

$$\mu\left(\Delta \setminus \Delta_\epsilon\right) \le \frac{\epsilon S(\Delta)\operatorname{diam}(\Delta)}{\sqrt{d}}.$$

Proof. By Lemma 3.10, $\Delta \setminus \operatorname{ri}(\Delta_\epsilon)$ can be written as the union of truncated hyperpyramid P_i, where each P_i has top $\varphi_\epsilon(F_i)$, base F_i and peak x_c. For each i, we can pick a point $x \in \operatorname{aff}(F_i)$ such that the vector $x_c - x$ is orthogonal to F_i.

The volume of a hyperpyramid in $\mathbb{R}^d$ with base B and peak p can be computed using [Mat99, Theorem 1.2.10], which says that such a hyperpyramid has volume $\operatorname{vol}(B) \cdot h/d$, where h is the Euclidean distance from p to B. Applying this theorem, we get

$$\begin{aligned}\mu(P_i) &\le \frac{\operatorname{vol}(\phi_\epsilon(F_i))|\varphi_\epsilon(x) - x|_2}{d} \\ &= \frac{\operatorname{vol}(\phi_\epsilon(F_i))\epsilon|x_c - x|_2}{d}\end{aligned}$$

$$\leq \frac{\mathrm{vol}(F_i)\epsilon \cdot \sup\{|a-b|_2 : a \in \Delta, b \in \Delta\}}{d}.$$

Recall that $\mathrm{diam}(\Delta)$ is the diameter of Δ with respect to the maximum norm. For every $a, b \in \Delta$, we have $|a-b|_2 \leq \sqrt{d}|a-b|$. It implies that $\sup\{|a-b|_2 : a \in \Delta, b \in \Delta\} \leq \sqrt{d}\,\mathrm{diam}(\Delta)$ and thus

$$\mu(P_i) \leq \frac{\mathrm{vol}(F_i)\epsilon\,\mathrm{diam}(\Delta)}{\sqrt{d}},$$

which proves the result. □

We now construct the continuous version of characteristic function of Δ with respect to ϵ. For every $y \in \Delta \backslash \Delta_\epsilon$, consider the ray

$$\{x_c + \lambda(y - x_c) : \lambda \in \mathbb{R}_{\geq 0}\}$$

with starting point x_c and passing through y. Take

$$\lambda_y := \sup\left\{\lambda \in \mathbb{R}_{\geq 0} : x_c + \lambda(y - x_c) \in \Delta\right\}.$$

and set $x_y := x_c + \lambda_y(y - x_c) \in \Delta$. Then we have

$$y = x_c + (1 - \delta_y)(x_y - x_c) = \varphi_\delta(x_y)$$

for $\delta_y := 1 - 1/\lambda_y$. Hence $y \in \varphi_\delta(\Delta) = \Delta_\delta$. As $y \in \Delta \backslash \Delta_\epsilon$, we have $0 \leq \delta_y < \epsilon$. With this expression, we can define a function $\chi^c_{\Delta,\epsilon} : \mathbb{R}^d \to \mathbb{R}$ such that

$$\chi^c_{\Delta,\epsilon}(y) := \begin{cases} 0 & \text{for } y \notin \Delta, \\ \delta_y/\epsilon & \text{for } y \in \Delta\backslash\Delta_\epsilon, \\ 1 & \text{for } y \in \Delta_\epsilon. \end{cases} \tag{3.1}$$

Proposition 3.12. *For each $0 < \epsilon < 1$, the function $\chi^c_{\Delta,\epsilon} : \mathbb{R}^d \to \mathbb{R}$ is piecewise affine.*

Proof. By definition, $\chi^c_{\Delta,\epsilon} = 0$ on $\mathbb{R}^d \backslash \Delta$ and $\chi^c_{\Delta,\epsilon} = 1$ on Δ_ϵ. By Lemma 3.10, we have

$$\Delta = \Delta_\epsilon \cup (P_1 \cup \cdots \cup P_e),$$

where every P_i is a truncated hyperpyramid. Let us consider $\chi^c_{\Delta,\epsilon}$ on each closed subset P_i.

Suppose $y_1, y_2 \in P_i$ with $\chi^c_{\Delta,\epsilon}(y_1) = \delta_1/\epsilon$ and $\chi^c_{\Delta,\epsilon}(y_2) = \delta_2/\epsilon$. Let $y = ty_1 + (1-t)y_2 \in P_i$ for $t \in \mathbb{R}$. Then there exists $0 \leq \delta < \epsilon$ such that $\chi^c_{\Delta,\epsilon}(y) = \delta/\epsilon$. It is easy to check that $\delta = t\delta_1 + (1-t)\delta_2$ using elementary Euclidean geometry. Therefore,

$$\begin{aligned}\chi^c_{\Delta,\epsilon}(y) &= \frac{\delta}{\epsilon}\\ &= t\frac{\delta_1}{\epsilon} + (1-t)\frac{\delta_2}{\epsilon}\\ &= t\chi^c_{\Delta,\epsilon}(y_1) + (1-t)\chi^c_{\Delta,\epsilon}(y_2)\end{aligned}$$

and so $\chi^c_{\Delta,\epsilon}$ is affine on P_i. The conclusion is proved. $\square$

Corollary 3.13. *For each $0 < \epsilon < 1$, the function $\chi^c_{\Delta,\epsilon}$ is continuous.*

Remark 3.14. One can prove that $\chi^c_{\Delta,\epsilon}$ converges pointwise to the characteristic function χ_Δ in the interior of Δ as $\epsilon \to 0$. Indeed, take $y \in \mathrm{ri}(\Delta)$. Then there exists $\epsilon > 0$ and $x \in \Delta \setminus \mathrm{ri}(\Delta)$ such that

$$y = x_c + (1-\epsilon)(x - x_c),$$

where x is in the intersection of $\Delta \setminus \mathrm{ri}(\Delta)$ and the ray with starting point x_c and passing through y. Then $y \in \Delta_\epsilon$. Therefore, for any $0 < \delta < \epsilon$, we have $y \in \Delta_\epsilon \subset \Delta_\delta$ and therefore $\chi^c_{\Delta,\delta}(y) = 1 = \chi_\Delta(y)$, which proves the desired result.

3.2. Computation of modulus of continuity

To apply Proposition 3.2, for $\epsilon > 0$ and $t \geq 0$, we need to compute the modulus of continuity $\rho(f \cdot \chi^c_{\Delta,\epsilon}, t)$. First, let us compute $\rho(\chi^c_{\Delta,\epsilon}, t)$. As we will apply our version of Koksma's inequality over polytopes to a sequence in $[0,1)^d$ with small enough discrepancy, it is sufficient to compute $\rho(\chi^c_{\Delta,\epsilon}, t)$ when t is small enough.

Proposition 3.15. *Let $\Delta \subset \mathbb{R}^d$ be a polytope, and let $\chi^c_{\Delta,\epsilon}$ be the continuous version of the characteristic function of Δ that we constructed in (3.1) for $0 < \epsilon < 1$. Then for sufficiently small real numbers $\epsilon > 0$ and $t \geq 0$, we have*

$$\rho(\chi^c_{\Delta,\epsilon}, t) \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)}.$$

Let us now prove Proposition 3.15. During the proof, we intersect the polytopes with affine two-dimensional planes multiple times to make our arguments visual.

By definition,

$$\rho(\chi^c_{\Delta,\epsilon}, t) = \sup\left\{\left|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)\right| : x, y \in \mathbb{R}^d,\ |x - y| \leq t\right\}.$$

Let $x, y \in \mathbb{R}^d$ with $|x-y| \leq t$. Our goal is to find an upper bound for $|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)|$. If $x, y \notin \Delta$, then $|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)| = |0 - 0| = 0$. If $x, y \in \Delta_\epsilon$, then $|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)| = |1 - 1| = 0$.

If $x \notin \Delta$ and $y \in \Delta \backslash \operatorname{ri}(\Delta_\epsilon)$, then by considering the line passing through x and y, we can find an $x' \in \Delta \backslash \operatorname{ri}(\Delta)$ such that $|x'-y| \le |x-y| \le t$ and $\chi^c_{\Delta,\epsilon}(x) = \chi^c_{\Delta,\epsilon}(x') = 0$. The case $y \notin \Delta$ and $x \in \Delta \backslash \operatorname{ri}(\Delta_\epsilon)$ is in the same way.

Similarly, if $x \in \Delta \backslash \operatorname{ri}(\Delta_\epsilon)$ and $y \in \Delta_\epsilon$, then by considering the line passing through x and y, we can find an $y' \in \Delta_\epsilon \backslash \operatorname{ri}(\Delta_\epsilon)$ such that $|x-y'| \le |x-y| \le t$ and $\chi^c_{\Delta,\epsilon}(y) = \chi^c_{\Delta,\epsilon}(y') = 1$. The case $y \in \Delta \backslash \operatorname{ri}(\Delta_\epsilon)$ and $x \in \Delta_\epsilon$ is in the same way.

Hence, it remains to consider the case $x, y \in \Delta \backslash \operatorname{ri}(\Delta_\epsilon)$.

Recall that x_c is the center of an inscribed cubic ball of Δ.

Reduction Step. Let $F_1, \dots, F_e$ be the facets of Δ. By Lemma 3.10, the set $\Delta \backslash \operatorname{ri}(\Delta_\epsilon)$ is a union of truncated hyperpyramids $P_1, \dots, P_e$, where each P_i has top $\varphi_\epsilon(F_i)$, base F_i and peak x_c. We may assume that $x, y \in P_i$ for one i.

Indeed, otherwise x, y are not in the same P_i for one i. When t is small enough, x, y, x_c are not on the same line. Let H be the affine two-dimensional plane containing x, y, x_c. By definition $\chi^c_{\Delta,\epsilon}(x) = \delta_x/\epsilon$ and $\chi^c_{\Delta,\epsilon}(y) = \delta_y/\epsilon$ for $0 \le \delta_x, \delta_y < \epsilon$. If $\delta_x = \delta_y$, then $|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)| = 0$. Without loss of generality, we suppose that $\delta_x < \delta_y$.

We claim that $\varphi_{\delta_x}(\Delta) \cap H$ is a 2-dimensional polytope with facets $\varphi_{\delta_x}(F_{k_1}) \cap H, \dots, \varphi_{\delta_x}(F_{k_l}) \cap H$, where $F_{k_1}, \dots, F_{k_l}$ are the facets of Δ such that $\varphi_{\delta_x}(F_{k_i}) \cap H$ is a line segment. Indeed, each nonempty $\varphi_{\delta_x}(F_j) \cap H$ is a face of the polytope $\varphi_{\delta_x}(\Delta) \cap H$. Note that $\overline{xy} \subset \varphi_{\delta_x}(\Delta) \cap H$ and $\overline{xx_c} \subset \varphi_{\delta_x}(\Delta) \cap H$ are two line segments which do not lie on the same line. Therefore, we have $\dim(\varphi_{\delta_x}(\Delta) \cap H) = 2$. Since $\dim(\varphi_{\delta_x}(F_{k_i}) \cap H) = 1$, then $\varphi_{\delta_x}(F_{k_i}) \cap H$ is a facet of $\varphi_\delta(\Delta) \cap H$.

Similarly, $\varphi_{\delta_y}(\Delta) \cap H$ is a 2-dimensional polytope with facets

$$\varphi_{\delta_y}(F_{k_1}) \cap H, \dots, \varphi_{\delta_y}(F_{k_l}) \cap H.$$

Let

$$d_i := \inf \left\{ |a-b| : a \in \operatorname{aff}(\varphi_{\delta_x}(F_{k_i}) \cap H),\ b \in \operatorname{aff}(\varphi_{\delta_y}(F_{k_i}) \cap H) \right\}.$$

Without loss of generality, we may assume that $d_1 \le d_2 \le \cdots \le d_l$. Then when ϵ is small enough, we have

$$\inf \left\{ |a-b| : a \in (\varphi_{\delta_x}(\Delta) \cap H) \backslash \operatorname{ri}(\varphi_{\delta_x}(\Delta) \cap H),\ b \in \varphi_{\delta_y}(\Delta) \cap H \right\} = d_1$$

and we can take $x' \in \varphi_{\delta_x}(F_{k_1}) \cap H$ and $y' \in \varphi_{\delta_y}(F_{k_1}) \cap H$ such that $|x'-y'| = d_1$. Then $|x'-y'| \le |x-y| \le t$, $\chi^c_{\Delta,\epsilon}(x') = \chi^c_{\Delta,\epsilon}(x) = \delta_x/\epsilon$ and $\chi^c_{\Delta,\epsilon}(y') = \chi^c_{\Delta,\epsilon}(y) = \delta_y/\epsilon$. Therefore, we can assume $x, y \in P_i$ for one i.

By the definition of inner radius Definition 3.5, we have $|z - x_c| \ge \operatorname{inrad}(\Delta)$ for any z in the facets of Δ.

Case 1. Suppose x, y, x_c are on the same line, and $\chi^c_{\Delta,\epsilon}(x) = \delta_x/\epsilon$ and $\chi^c_{\Delta,\epsilon}(y) = \delta_y/\epsilon$. Then there exists $z \in \Delta \backslash \operatorname{ri}(\Delta)$ such that

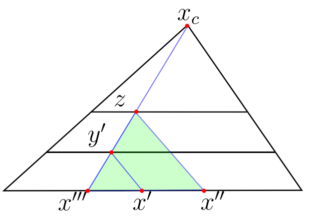

Fig. 2. An example of the truncated hyperpyramid $P \cap H$.

$$x = x_c + (1 - \delta_x)(z - x_c),$$
$$y = x_c + (1 - \delta_y)(z - x_c).$$

Thus $|x - y| = |z - x_c| \cdot |\delta_x - \delta_y|$ and so

$$\left|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)\right| = \frac{|\delta_x - \delta_y|}{\epsilon} = \frac{|x-y|}{\epsilon|z - x_c|} \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)}.$$

Case 2. Suppose x, y, x_c are not on the same line. The Reduction Step allows us to assume that $x, y \in P$ for the truncated hyperpyramid P corresponding to a facet F of Δ. First we suppose that $x \in F$ and $\chi^c_{\Delta,\epsilon}(x) = 0$. Let H be the affine 2-dimensional plane containing x, y, x_c. Suppose $\chi^c_{\Delta,\epsilon}(y) = \delta/\epsilon$ for $0 < \delta < \epsilon$. Then $y \in \varphi_\delta(F)$.

Notice that $F \cap H$ is a line segment, otherwise $F \cap H = \{x\}$ and $\varphi_\delta(F) \cap H = \{y\}$ and thus x, y, x_c are on the same line, a contradiction. It follows from the definition of φ_ϵ and φ_δ that both $\varphi_\epsilon(F) \cap H$ and $\varphi_\delta(F) \cap H$ are line segments. By Lemma 3.9, the intersections $\varphi_\epsilon(F) \cap H$, $\varphi_\delta(F) \cap H$ and $F \cap H$ are parallel. As in Fig. 2, when ϵ is small enough, we can take $x', x'' \in F \cap H$, $y' \in \varphi_\delta(F) \cap H$ and $z \in \varphi_\epsilon(F) \cap H$ such that

- y', z, x_c are on the same line,
- $|x' - y'| = \inf\{|a - b| : a \in \mathrm{aff}(F \cap H),\ b \in \mathrm{aff}(\varphi_\delta(F) \cap H)\}$,
- $|z - x''| = \inf\{|a - b| : a \in \mathrm{aff}(F \cap H),\ b \in \mathrm{aff}(\varphi_\epsilon(F) \cap H)\}$,
- the line $\ell_{x'y'}$ passing through x', y' and the line $\ell_{x''z}$ passing through x'', z are parallel.

With these points we have

- $|x' - y'| \leq |x - y| \leq t$,
- $\chi^c_{\Delta,\epsilon}(x') = \chi^c_{\Delta,\epsilon}(x) = 0$, $\chi^c_{\Delta,\epsilon}(y') = \chi^c_{\Delta,\epsilon}(y) = \delta/\epsilon$.

Set

$$d_F := \inf\left\{|a - b| : a \in \mathrm{aff}(F \cap H),\ b \in \mathrm{aff}(\varphi_\epsilon(F) \cap H)\right\},$$
$$D_F := \inf\left\{|a - x_c| : a \in \mathrm{aff}(F \cap H)\right\}.$$

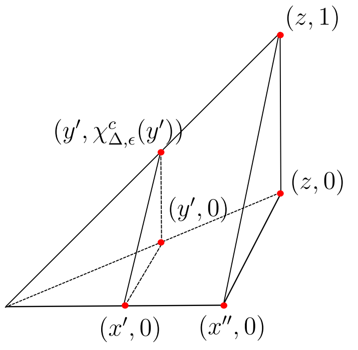

Fig. 3. Similar triangles for computation.

Then $|z - x''| = d_F$. Let x''' be the intersection of $F \cap H$ and the ray passing through y' with start point x_c, as in Fig. 2. Then by similar triangles

$$\frac{d_F}{D_F} = \frac{|x''' - z|_2}{|x''' - x_c|_2} = \epsilon$$

and thus $|x'' - z| = d_F = \epsilon D_F \geq \epsilon \cdot \mathrm{inrad}(\Delta)$. By Proposition 3.12 and its proof, we have $\chi^c_{\Delta,\epsilon}$ is affine on P. Using similar triangles as in Fig. 3, we get

$$\left|\chi^c_{\Delta,\epsilon}(x') - \chi^c_{\Delta,\epsilon}(y')\right| = \frac{|x' - y'|_2}{|x'' - z|_2} = \frac{|x' - y'|}{|x'' - z|} \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)},$$

where $\frac{|x'-y'|_2}{|x''-z|_2} = \frac{|x'-y'|}{|x''-z|}$ holds since x''', y', z are in the same line and $\ell_{x'y'}$ and $\ell_{x''z}$ are parallel.

It follows that

$$\left|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)\right| = \left|\chi^c_{\Delta,\epsilon}(x') - \chi^c_{\Delta,\epsilon}(y')\right| \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)}.$$

For the general case $x, y \in \Delta \setminus \mathrm{ri}(\Delta_\epsilon)$, the argument is essentially the same and we still get

$$\left|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)\right| \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)}.$$

Therefore,

$$\rho(\chi^c_{\Delta,\epsilon}, t) \leq \frac{t}{\epsilon \cdot \mathrm{inrad}(\Delta)}. \quad \square$$

3.3. Proof of Koksma's inequality over polytopes

Now we can prove one of the main tools in this paper, a version of Koksma's inequality over polytopes.

Theorem 3.16. *Let $f : [0,1]^d \to \mathbb{R}^d$ be a continuous function and let $(x_1, \ldots, x_n)$ be a finite sequence in $[0,1)^d$ with discrepancy $D = D(x_1, \ldots, x_n)$. Let $\Delta \subset [0,1]^d$ be a polytope of dimension d. Then for D small enough we have*

$$\left|\frac{1}{n}\sum_{\substack{i=1\\x_i\in\Delta}}^{n} f(x_i) - \int_{\Delta} f \mathrm{d}\mu\right| \le (1+2^{d+1})\rho(f, D^{1/(d+1)}) + M\Big(\frac{(1+2^{d+1})D^{1/(2d+2)}}{\operatorname{inrad}(\Delta)} + (4d\sqrt{d}+1)D^{1/d}e(\Delta) + \frac{2\operatorname{diam}(\Delta)S(\Delta)D^{1/(2d+2)}}{\sqrt{d}}\Big),$$

where

- *$\rho(f, D^{1/(d+1)})$ is the modulus of continuity of f for $D^{1/(d+1)}$,*
- *M is an upper bound of $|f|$, i.e. $|f(x)| \le M$ for all $x \in [0,1)^d$,*
- *$e(\Delta)$ is the number of facets of Δ,*
- *$S(\Delta)$ is the surface area of Δ, i.e. it is the sum of the volumes of all facets of Δ.*

Proof. We want to apply Proposition 3.2 to prove this theorem. To do this, we approximate $f \cdot \chi_\Delta$ by a continuous function. For $\epsilon > 0$, let $\chi^c_{\Delta,\epsilon}$ be the continuous characteristic function we constructed in Section 3.1. Then we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i)\chi_\Delta(x_i) - \int_{[0,1]^d} f\chi_\Delta \mathrm{d}\mu\right| &\le \left|\frac{1}{n}\sum_{i=1}^{n} f(x_i)\chi^c_{\Delta,\epsilon}(x_i) - \int_{[0,1]^d} f\chi^c_{\Delta,\epsilon}\mathrm{d}\mu\right| \\ &+ \frac{1}{n}\left|\sum_{i=1}^{n}\big(f(x_i)\chi_\Delta(x_i) - f(x_i)\chi^c_{\Delta,\epsilon}(x_i)\big)\right| \\ &+ \left|\int_{[0,1]^d}\big(f\chi_\Delta - f\chi^c_{\Delta,\epsilon}\big)\,\mathrm{d}\mu\right|.\end{aligned}$$

Let us first estimate the first difference. In order to apply Proposition 3.2 to it, we only need to understand the modulus of continuity of $f \cdot \chi^c_{\Delta,\epsilon}$. For any two points $x, y \in [0,1]^d$, by the triangle inequality,

$$\begin{aligned}&\big|f(x)\chi^c_{\Delta,\epsilon}(x) - f(y)\chi^c_{\Delta,\epsilon}(y)\big| \\ &\quad\le \big|f(x)\chi^c_{\Delta,\epsilon}(x) - f(y)\chi^c_{\Delta,\epsilon}(x) + f(y)\chi^c_{\Delta,\epsilon}(x) - f(y)\chi^c_{\Delta,\epsilon}(y)\big|\end{aligned}$$

$$\leq |f(x) - f(y)| + |f(y)| \left|\chi^c_{\Delta,\epsilon}(x) - \chi^c_{\Delta,\epsilon}(y)\right|.$$

Since f is continuous on $[0,1]^d$, there exists a positive real number M such that $|f(x)| \leq M$ for all x. By Proposition 3.15, for sufficiently small ϵ and t, we have

$$\rho\left(f \cdot \chi^c_{\Delta,\epsilon}, t\right) \leq \rho(f,t) + M \cdot \rho(\chi^c_{\Delta,\epsilon}, t) \leq \rho(f,t) + \frac{Mt}{\epsilon \cdot \mathrm{inrad}(\Delta)}.$$

Therefore, we get

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i)\chi^c_{\Delta,\epsilon}(x_i) - \int\limits_{[0,1]^d} f\chi^c_{\Delta,\epsilon}\mathrm{d}\mu\right| \leq (1+2^{d+1})\left(\rho(f, D^{1/(d+1)}) + \frac{MD^{1/(d+1)}}{\epsilon \cdot \mathrm{inrad}(\Delta)}\right)$$

for small enough ϵ.

For the second difference, we know that χ_Δ and $\chi^c_{\Delta,\epsilon}$ are only different on $\Delta \setminus \mathrm{ri}(\Delta_\epsilon)$. Therefore,

$$\begin{aligned}
&\frac{1}{n}\left|\sum_{i=1}^{n}\left(f(x_i)\chi_\Delta(x_i) - f(x_i)\chi^c_{\Delta,\epsilon}(x_i)\right)\right| \\
&\qquad \leq \frac{1}{n}\sum_{\substack{i=1 \\ x_i \in \Delta\setminus \mathrm{ri}(\Delta_\epsilon)}}^{n} |f(x_i)| \left|\chi_\Delta(x_i) - \chi^c_{\Delta,\epsilon}(x_i)\right| \\
&\qquad \leq M\frac{\sum_{i=1}^{n}\chi_{\Delta\setminus \mathrm{ri}(\Delta_\epsilon)}(x_i)}{n}.
\end{aligned}$$

By Lemma 3.10, $\Delta \setminus \mathrm{ri}(\Delta_\epsilon)$ is a union of polytopes $P_1, \ldots, P_e$ such that $\mathrm{ri}(P_i) \cap \mathrm{ri}(P_j) = \varnothing$ for $i \neq j$, where $e \coloneqq e(\Delta)$ is the number of facets of Δ. Therefore, we can find convex subsets $P'_1, \ldots, P'_e$ such that

- $\mathrm{ri}(P'_i) = \mathrm{ri}(P_i)$,
- $P'_i \cap P'_j = \varnothing$ for all $i \neq j$,
- $\Delta \setminus \mathrm{ri}(\Delta_\epsilon) = P'_1 \cup \cdots \cup P'_e$.

Let $J(x_1, \ldots, x_n)$ be the isotropic discrepancy of $(x_1, \ldots, x_n)$. It follows from Proposition 2.4 that

$$\begin{aligned}
\left|\frac{\sum_{i=1}^{n}\chi_{\Delta\setminus \mathrm{ri}(\Delta_\epsilon)}(x_i)}{n} - \mu(\Delta\setminus\Delta_\epsilon)\right| &\leq \sum_{j=1}^{e}\left|\frac{\sum_{i=1}^{n}\chi_{P'_j}(x_i)}{n} - \mu(P'_j)\right| \\
&\leq eJ(x_1, \ldots, x_n) \\
&\leq e(4d\sqrt{d}+1)D^{1/d}.
\end{aligned}$$

Thus

$$\frac{\sum_{i=1}^{n} \chi_{\Delta\setminus \mathrm{ri}(\Delta_\epsilon)}(x_i)}{n} \leq e(4d\sqrt{d}+1)D^{1/d} + \mu(\Delta\backslash\Delta_\epsilon)$$

and so

$$\begin{aligned}\frac{1}{n}\left|\sum_{i=1}^{n}\big(f(x_i)\chi_\Delta(x_i) - f(x_i)\chi^c_{\Delta,\epsilon}(x_i)\big)\right| \\ &\leq M\frac{\sum_{i=1}^{n} \chi_{\Delta\setminus \mathrm{ri}(\Delta_\epsilon)}(x_i)}{n} \\ &\leq M(e(4d\sqrt{d}+1)D^{1/d} + \mu(\Delta\backslash\Delta_\epsilon)).\end{aligned}$$

Similarly, for the third difference, we have

$$\begin{aligned}\left|\int\limits_{[0,1]^d}\big(f\chi_\Delta - f\chi^c_{\Delta,\epsilon}\big)\,\mathrm{d}\mu\right| &\leq \int\limits_{\Delta\backslash\Delta_\epsilon} |f|\cdot\left|\chi_\Delta - \chi^c_{\Delta,\epsilon}\right|\mathrm{d}\mu \\ &\leq M\mu(\Delta\backslash\Delta_\epsilon).\end{aligned}$$

By Lemma 3.11 we have $\mu(\Delta\backslash\Delta_\epsilon) \leq \epsilon\cdot\mathrm{diam}(\Delta)S(\Delta)/\sqrt{d}$. When D is small enough, we can take $\epsilon = D^{1/(2d+2)}$ such that ϵ is also small enough. Combined with all these inequalities, we get the result. □

Remark 3.17. Note that Theorem 3.16 does not imply Proposition 3.2. If $\Delta = [0,1]^d$, then we can apply Proposition 3.2 directly to get an upper bound of the difference without the term

$$M\Big(\frac{(1+2^{d+1})D^{1/(2d+2)}}{\mathrm{inrad}(\Delta)} + (4d\sqrt{d}+1)D^{1/d}e(\Delta) + \frac{2\,\mathrm{diam}(\Delta)S(\Delta)D^{1/(2d+2)}}{\sqrt{d}}\Big)$$

in Theorem 3.16.

Corollary 3.18. *Let $f : [0,1]^d \to \mathbb{R}^d$ be a continuous function and let $(x_1,\ldots,x_n)$ be a finite sequence in $[0,1)^d$ with discrepancy $D = D(x_1,\ldots,x_n)$. Let $\Delta \subset [0,1]^d$ be a polytope of dimension d. Then for D small enough we have*

$$\left|\frac{1}{n}\sum_{\substack{i=1\\ x_i\in\Delta}}^{n} f(x_i) - \int\limits_{\Delta} f\mathrm{d}\mu\right| \ll_\Delta \rho(f, D^{1/(d+1)}) + MD^{1/(2d+2)},$$

where $M = \max\left\{|f(x)| : x \in [0,1]^d\right\}$.

4. Proof of the main theorem

In this section Theorem 1.1 is proved.

As in Theorem 1.1, we let d, k be integers, let $P \in \overline{\mathbb{Q}}[T_1^{\pm 1}, \dots, T_d^{\pm 1}] \backslash \{0\}$ be an essentially atoral Laurent polynomial with at most k nonzero terms, and let $\Delta \subset [0,1)^d$ be a d-dimensional polytope.

For a torsion point $\omega \in \mathbb{G}_m^d$, according to Laurent's Theorem [Lau84], if $\delta(\omega)$ is large in terms of d and P, then $P(\omega^\sigma) \neq 0$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})$. We fix a torsion point $\omega \in \mathbb{G}_m^d$ with $\delta(\omega)$ sufficiently large such that $P(\omega^\sigma) \neq 0$ for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})$. Denote $n := [\mathbb{Q}(\omega) : \mathbb{Q}]$. For the Galois orbit of ω, we get a finite set $\{x_1, \dots, x_n\} \subset [0,1)^d$ such that

$$\{e(x_1), \dots, e(x_n)\} = \{\omega^\sigma : \sigma \in \mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})\}.$$

Then $P(e(x_i)) \neq 0$ for $1 \leq i \leq n$. Let $D = D(x_1, \dots, x_n)$ (resp. $J = J(x_1, \dots, x_n)$) be the discrepancy (resp. isotropic discrepancy) of the sequence $(x_1, \dots, x_n)$. Our goal is to estimate the difference

$$\left| \frac{1}{n} \sum_{\substack{i=1 \\ x_i \in \Delta}}^{n} \log |P(e(x_i))| - \int_{\Delta} \log |P(e(x))| \mathrm{d}\mu(x) \right|.$$

We first examine the easiest case that P is a constant. In this case,

$$\left| \frac{1}{n} \sum_{\substack{i=1 \\ x_i \in \Delta}}^{n} \log |P(e(x_i))| - \int_{\Delta} \log |P(e(x))| \mathrm{d}\mu(x) \right| \ll_P \left| \frac{\sum_{i=1}^n \chi_\Delta(x_i)}{n} - \mu(\Delta) \right| \leq J$$

by the definition of J. By Proposition 2.4 and Proposition 2.2, we have

$$J \ll_d D^{1/d} \ll_d \left(\frac{(\log \delta(\omega))^d}{\delta(\omega)^{1/2}} \right)^{1/d} \ll_d \delta(\omega)^{-1/(4d)}$$

as $\delta(\omega) \to \infty$. The desired result follows.

Now suppose $P \in \overline{\mathbb{Q}}[T_1^{\pm 1}, \dots, T_d^{\pm 1}] \backslash \overline{\mathbb{Q}}$. We may assume that P is a polynomial, as multiplying by a monomial does not change the difference between the discrete sum and the integral stated above. Let $c > 0$ be the largest maximum norm among the coefficients of P. We may assume that the maximum norm of the coefficient vector of P is 1; otherwise, we replace P by a suitable scalar multiple, which does not affect the estimate since the implicit constant depends on P.

Suppose $k \geq 2$. By Proposition 2.2, $\delta(\omega)$ is large implies that D is small. At the end of proof, we will take r in terms of D such that the small D implies that r is also small.

By the triangle inequality,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{\substack{i=1\\x_i\in\Delta}}^{n}\log|P(e(x_i))| - \int_{\Delta}\log|P(e(x))|\mathrm{d}\mu(x)\right| \\
&\qquad \le \left|\frac{1}{n}\sum_{\substack{i=1\\x_i\in\Delta}}^{n}\log\max\{r,|P(e(x_i))|\} - \int_{\Delta}\log\max\{r,|P(e(x))|\}\mathrm{d}\mu(x)\right| \\
&\qquad\quad + \frac{1}{n}\left|\sum_{i=1}^{n}\big(\log|P(e(x_i))| - \log\max\{r,|P(e(x_i))|\}\big)\right| \\
&\qquad\quad + \int_{\Delta}\big|\log|P(e(x))| - \log\max\{r,|P(e(x))|\}\big|\mathrm{d}\mu(x). \qquad (4.1)
\end{aligned}$$

For the first difference, as D is small, we can apply Corollary 3.18 to it. In order to apply Corollary 3.18, we need to estimate the maximum value of $\log\max\{r,|P(e(x))|\}$ on $[0,1]^d$. As $|P(e(x))|$ is continuous on the compact set $[0,1]^d$, we have that $|P(e(x))|$ is bounded. Taking r sufficiently small, we have $\big|\log\max\{r,|P(e(x))|\}\big| \le |\log r|$ for all $x \in [0,1]^d$. By Corollary 3.18,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{\substack{i=1\\x_i\in\Delta}}^{n}\log\max\{r,|P(e(x_i))|\} - \int_{\Delta}\log\max\{r,|P(e(x))|\}\mathrm{d}\mu(x)\right| \\
&\qquad \ll_{\Delta} \rho(\log\max\{r,|P(e(x))|\}; D^{1/(d+1)}) + (-\log r)D^{1/(2d+2)}
\end{aligned}$$

for D small enough. The modulus of continuity of the function $\log\max\{r,|P(e(x))|\}$ can be estimated by [DH24, Lemma 7.5], which states that

$$\rho(\log\max\{r,|P(e(x))|\}; t) \ll_{d,k} \frac{\deg(P)t}{r}$$

for any $r \in (0,1]$ and $t > 0$. In our setting, we conclude that

$$\rho(\log\max\{r,|P(e(x))|\}; D^{1/(d+1)}) \ll_P \frac{D^{1/(d+1)}}{r}$$

and thus

$$\left|\frac{1}{n}\sum_{\substack{i=1\\x_i\in\Delta}}^{n}\log\max\{r,|P(e(x_i))|\} - \int_{\Delta}\log\max\{r,|P(e(x))|\}\mathrm{d}\mu(x)\right|$$

$$\ll_{\Delta,P} \frac{D^{1/(d+1)}}{r} + (-\log r)D^{1/(2d+2)}. \quad (4.2)$$

For the second difference, the triangle inequality gives

$$\frac{1}{n}\left|\sum_{i=1}^{n}\big(\log|P(e(x_i))| - \log\max\{r,|P(e(x_i))|\}\big)\right|$$
$$\leq \frac{1}{n}\sum_{\substack{i=1\\|P(e(x_i))|<r}}^{n}\big|\log\max\{r,|P(e(x_i))|\}\big| + \frac{1}{n}\sum_{\substack{i=1\\|P(e(x_i))|<r}}^{n}\big|\log|P(e(x_i))|\big|.$$

By definition, $\log\max\{r,|P(e(x))|\} = \log r$ if $|P(e(x))| < r$. Hence,

$$\frac{1}{n}\sum_{\substack{i=1\\|P(e(x_i))|<r}}^{n}\big|\log\max\{r,|P(e(x_i))|\}\big| = \frac{\sharp\{i : |P(e(x_i))| < r\}\ |\log r|}{n}.$$

This can be further estimated using [DH24, Lemma 7.4], which indicates that

$$\frac{\sharp\,\{i : |P(e(x_i))| \leq r\}}{n} \ll_{d,k} r^{1/(2k)} + \deg(P)\frac{D^{1/(d+1)}}{r}.$$

Thus,

$$\frac{1}{n}\sum_{\substack{i=1\\|P(e(x_i))|<r}}^{n}\big|\log\max\{r,|P(e(x_i))|\}\big| \ll_P (-\log r)\left(r^{1/(2k)} + \frac{D^{1/(d+1)}}{r}\right).$$

Note that [DH24, Lemma 7.7] shows

$$\frac{1}{n}\sum_{\substack{i=1\\|P(e(x_i))|<r}}^{n}\big|\log|P(e(x_i))|\big|$$
$$\ll_{d,k} \frac{(\deg P)D^{1/(d+1)}}{r^2} + r^{1/(4k)} + \left|m(P) - \frac{1}{n}\sum_{i=1}^{n}\log|P(e(x_i))|\right|.$$

We conclude from the main theorem of [DH24] that

$$\left|m(P) - \frac{1}{n}\sum_{i=1}^{n}\log|P(e(x_i))|\right| \ll_P \delta(\omega)^{-\gamma}$$

for a constant $\gamma = \gamma(d,k) > 0$. Putting the above inequalities together, for r small enough, we get

$$\frac{1}{n}\left|\sum_{i=1}^{n}\log|P(e(x_i))| - \log\max\{r, |P(e(x_i))|\}\right| \ll_Q \delta(\omega)^{-\gamma} + r^{1/(4k)} + \frac{D^{1/(d+1)}}{r^2}. \quad (4.3)$$

Let $\mathcal{S}_P(r) = \{x \in [0,1)^d : |P(e(x))| < r\}$. Similarly, for the third difference, we have

$$\int_{\Delta} \big|\log|P(e(x))| - \log\max\{r, |P(e(x))|\}\big| \mathrm{d}\mu(x)$$
$$\leq \int_{\mathcal{S}_P(r)} \big|\log\max\{r, |P(e(x))|\}\big| \mathrm{d}\mu(x) + \int_{\mathcal{S}_P(r)} \big|\log|P(e(x))|\big| \mathrm{d}\mu(x).$$

By [DH24, Lemma A.3(i)], the Lebesgue volume of $\mathcal{S}_P(r)$ is $\ll_{d,k} r^{1/(2k-2)}$. Applying this and using $\log\max\{r, |P(e(x))|\} = \log r$ if $|P(e(x))| < r$, we get

$$\int_{\mathcal{S}_P(r)} \big|\log\max\{r, |P(e(x))|\}\big| \mathrm{d}x \leq (-\log r)\cdot \mu(\mathcal{S}_P(r)) \ll_{d,k} (-\log r) r^{1/(2k-2)}.$$

Moreover, it is shown in [DH24, Lemma A.4] that

$$\int_{\mathcal{S}_P(r)} \big|\log|P(e(x))|\big| \mathrm{d}\mu(x) \ll_{d,k} r^{1/(4k-4)},$$

and so for r small enough we have

$$\int_{\Delta} \big|\log|P(e(x))| - \log\max\{r, |P(e(x))|\}\big| \mathrm{d}\mu(x) \ll_{d,k} r^{1/(4k-4)}. \quad (4.4)$$

Choose $r = D^{1/(4d+4)}$ now. Then gathering the inequalities (4.1), (4.2), (4.3) and (4.4) together we have

$$\left|\frac{1}{n}\sum_{\substack{i=1 \\ x_i \in \Delta}}^{n} \log|P(e(x_i))| - \int_{\Delta} \log|P(e(x))| \mathrm{d}\mu(x)\right| \ll_{\Delta,P} \delta(\omega)^{-\gamma} + D^{1/(16k(d+1))}$$

if $\delta(\omega)$ is sufficiently large. The conclusion follows from Proposition 2.2. □

Remark 4.1. An explicit value of κ in Theorem 1.1 can be deduced from this proof. Indeed, by Proposition 2.2,

$$D^{1/(16k(d+1))} \ll_d \left(\frac{(\log\delta(\omega))^d}{\delta(\omega)^{1/2}}\right)^{1/(16k(d+1))}$$

$$\begin{aligned}&\ll_d \left(\frac{\delta(\omega)^{1/4}}{\delta(\omega)^{1/2}}\right)^{1/(16k(d+1))}\\&= \delta(\omega)^{-1/(64k(d+1))}.\end{aligned}$$

Therefore, one can choose

$$\kappa = \min\left\{\gamma, \frac{1}{64k(d+1)}\right\}$$

and the proof of [DH24, Theorem 8.8] contains a way to determine $\gamma = \gamma(d,k)$. This approach is summarized in the extended arXiv version [Lin24, Appendix A] with an algorithm.

5. Application

In this section, we answer [GS23, Question 6.2], posed by Gualdi and Sombra, by applying Theorem 1.1. Let us introduce the background of this question. The paper [GS23] studies the height of the intersection of

$$L = Z(x_0 + x_1 + x_2) \subset \mathbb{P}^2(\overline{\mathbb{Q}})$$

with its translate ωL by a torsion point $\omega = (\omega_1, \omega_2) \in \mathbb{G}_m^2(\overline{\mathbb{Q}}) \cong (\overline{\mathbb{Q}}^\times)^2$, where

$$\omega L = Z(x_0 + \omega_1^{-1}x_1 + \omega_2^{-1}x_2).$$

When ω is nontrivial, i.e. $\omega \neq (1,1)$, the intersection $L \cap \omega L$ consists of the point

$$\mathcal{P}(\omega) = [\omega_2^{-1} - \omega_1^{-1} : 1 - \omega_2^{-1} : \omega_1^{-1} - 1] \in \mathbb{P}^2(\overline{\mathbb{Q}}).$$

Let us also denote by $\mathcal{P}(\omega)$ the vector of homogeneous coordinates $(\omega_2^{-1} - \omega_1^{-1}, 1 - \omega_2^{-1}, \omega_1^{-1} - 1)$. We fix an embedding $\iota : \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$. Let $d = \mathrm{ord}(\omega)$. Then [GS23, Lemma 2.3] gives an explicit formula for the archimedean height of $\mathcal{P}(\omega)$:

$$\mathrm{h}_\infty(\mathcal{P}(\omega)) = \frac{1}{\phi(d)} \sum_{k \in (\mathbb{Z}/d\mathbb{Z})^\times} \log\max(|\iota(\omega_2^k) - \iota(\omega_1^k)|, |\iota(\omega_2^k) - 1|, |\iota(\omega_1^k) - 1|),$$

where ∞ is the unique archimedean place of $\mathbb{Q}$ and $|\cdot|$ denotes the usual absolute value on $\mathbb{C}$.

We will later consider a sequence in $\mathbb{G}_m^2(\overline{\mathbb{Q}})$ that avoids all proper algebraic subgroups of $\mathbb{G}_m^2(\overline{\mathbb{Q}})$. According to [BG06, Chapter 3], over any field F of characteristic 0, a one-dimensional algebraic subgroup of $\mathbb{G}_m^2(F)$ is of the form

$$H_a := \{\omega \in \mathbb{G}_m^2(F) : \omega^a = 1\} \tag{5.1}$$

for an $a \in \mathbb{Z}^2 \backslash \{0\}$. We say a sequence $(\omega_\ell)_{\ell \geq 1}$ in $\mathbb{G}_m^2(\overline{\mathbb{Q}})$ is *strict* if for all $a = (a_1, a_2) \in \mathbb{Z}^2 \backslash \{0\}$, there exists an integer $\ell_0 > 0$ such that ω_ℓ does not lie in H_a for $\ell > \ell_0$, which means

$$\omega_{\ell,1}^{a_1} \omega_{\ell,2}^{a_2} \neq 1, \quad \forall \ell > \ell_0,$$

where $\omega_\ell = (\omega_{\ell,1}, \omega_{\ell,2})$. Note that for a sequence $(\omega_\ell)_{\ell \geq 1}$ in $\mathbb{G}_m^2(\overline{\mathbb{Q}})$, the strictness degree $\delta(\omega_\ell) \to \infty$ as $\ell \to \infty$ if and only if $(\omega_\ell)_{\ell \geq 1}$ is strict.

Let $(\omega_\ell)_{\ell \geq 1}$ be a strict sequence of non-trivial torsion points in $\mathbb{G}_m^2(\overline{\mathbb{Q}})$ now. It is shown in [GS23, Proposition 4.1] that the limit of archimedean heights $\mathrm{h}_\infty(\mathcal{P}(\omega_\ell))$ is an integral by equidistribution theorems. Indeed, fix an integer ℓ and write $n_\ell = \phi(\operatorname{ord}(\omega_\ell))$ for simplicity, set

$$\{e(\alpha_{\ell,1}), \ldots, e(\alpha_{\ell,n_\ell})\} \coloneqq \left\{\iota(\omega_\ell^k) : k \in (\mathbb{Z}/\operatorname{ord}(\omega_\ell)\mathbb{Z})^\times\right\},$$

where $e(\alpha_{\ell,k}) = (e^{\mathrm{i}2\pi x_{\ell,k}},\ e^{\mathrm{i}2\pi y_{\ell,k}})$ for each $\alpha_{\ell,k} = (x_{\ell,k}, y_{\ell,k}) \in [0,1)^2$. Let

$$\begin{aligned} f &: [0,1)^2 \to \mathbb{R} \cup \{-\infty\}, \\ &(x,y) \mapsto \log \max \left\{ \left|e^{\mathrm{i}2\pi(x-y)} - 1\right|,\ \left|e^{\mathrm{i}2\pi x} - 1\right|,\ \left|e^{\mathrm{i}2\pi y} - 1\right| \right\}. \end{aligned}$$

Then we have

$$\lim_{\ell \to \infty} \mathrm{h}_\infty\left(\mathcal{P}(\omega_\ell)\right) = \lim_{\ell \to \infty} \frac{1}{n_\ell} \sum_{k=1}^{n_\ell} f(\alpha_{\ell,k}) = \int\limits_{[0,1)^2} f \mathrm{d}\mu.$$

In [GS23, Question 6.2], a quantitative version of this equidistribution result is requested. The following is our solution.

Let $\omega \in \mathbb{G}_m^2(\overline{\mathbb{Q}})$ be a non-trivial torsion point. For the Galois orbit of ω, we get a finite set $\{\alpha_1, \ldots, \alpha_n\} \subset [0,1)^2$ such that

$$\{e(\alpha_1), \ldots, e(\alpha_n)\} \coloneqq \left\{\iota(\omega^k) : k \in (\mathbb{Z}/\operatorname{ord}(\omega)\mathbb{Z})^\times\right\},$$

where $n = \phi(\operatorname{ord}(\omega))$ and $\alpha_k = (x_k, y_k) \in [0,1)^2$. Let us estimate

$$\left| h(\mathcal{P}(\omega)) - \int\limits_{[0,1)^2} f \mathrm{d}\mu \right|.$$

As in [GS23, Section 5], the square $[0,1)^2$ has a partition consisting of 12 closed triangles as in Fig. 4. Among these 12 triangles, in 4 of them, $f(x,y) = \log\left|e^{i2\pi x} - 1\right|$; in 4 of them, $f(x,y) = \log\left|e^{i2\pi y} - 1\right|$; in the remaining 4 triangles, $f(x,y) = \log\left|e^{i2\pi(x-y)} - 1\right|$. The first (resp. second / third) group of triangles are marked by green (resp. blue / red) in the figure. We denote these 12 subsets as Ω_{ij} for $1 \leq i \leq 3$ and $1 \leq j \leq 4$, where

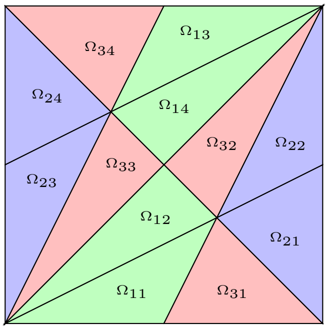

Fig. 4. The partition of $[0, 1]^2$ into triangles.

$\Omega_{11}, \ldots, \Omega_{14}$ correspond to green triangles, $\Omega_{21}, \ldots, \Omega_{24}$ correspond to blue triangles and $\Omega_{31}, \ldots, \Omega_{34}$ correspond to red triangles.

Furthermore, for any point in the boundaries of these triangles, its image under the complex exponential map

$$[0,1)^2 \to \mathbb{G}_m^2(\mathbb{C}), \quad (x,y) \mapsto (e^{\mathrm{i}2\pi x}, e^{\mathrm{i}2\pi y})$$

lies in one of the algebraic subgroups

$$H_{(1,0)}, H_{(0,1)}, H_{(1,1)}, H_{(1,-1)}, H_{(2,-1)}, H_{(1,-2)},$$

as defined by the notation H_a in (5.1). Therefore, if $\delta(\iota(\omega^k)) \geq 3$, then none of the points $\alpha_1, \ldots, \alpha_n$ lie on the boundaries of these 12 triangles. It follows that

$$f(\alpha_k) = \sum_{\substack{1 \leq i \leq 3 \\ 1 \leq j \leq 4}} f \cdot \chi_{\Omega_{ij}}(\alpha_k), \quad \forall 1 \leq k \leq n$$

if $\delta(\omega) \geq 3$, and thus

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(\alpha_k) - \int\limits_{[0,1)^2} f \mathrm{d}\mu \right| \leq \sum_{\substack{1 \leq i \leq 3 \\ 1 \leq j \leq 4}} \left| \frac{1}{n} \sum_{k=1}^{n} f \cdot \chi_{\Omega_{ij}}(\alpha_k) - \int\limits_{[0,1)^2} f \cdot \chi_{\Omega_{ij}} \mathrm{d}\mu \right| \tag{5.2}$$

if $\delta(\omega) \geq 3$.

Set $P_1(T_1, T_2) = T_1 - 1$, $P_2(T_1, T_2) = T_2 - 1$ and $P_3(T_1, T_2) = T_1 - T_2$. In $\Omega_{11}, \ldots, \Omega_{14}$ we have $f(x, y) = \log |P_1(e(x, y))|$; in $\Omega_{21}, \ldots, \Omega_{24}$ we have $f(x, y) = \log |P_2(e(x, y))|$; in $\Omega_{31}, \ldots, \Omega_{34}$ we have $f(x, y) = \log |P_3(e(x, y))|$. The three polynomials P_1, P_2 and P_3 are all binomials and, therefore, are essentially atoral by Example 2.7. Hence, by Theorem 1.1, if $\delta(\omega)$ is sufficiently large, then there exists a constant $\kappa > 0$ such that

$$\left|\frac{1}{n}\sum_{k=1}^{n} f\cdot\chi_{\Omega_{ij}}(\alpha_k) - \int\limits_{[0,1)^2} f\cdot\chi_{\Omega_{ij}}\mathrm{d}\mu\right| \ll \delta(\omega)^{-\kappa} \tag{5.3}$$

for all $1 \leq i \leq 3$ and $1 \leq j \leq 4$, where the constant is specified by P_1, P_2, P_3 and the choice of Ω_{ij}. Moreover, by Remark 4.1, we can choose

$$\kappa = \min\left\{\gamma(2,2), \frac{1}{64\times 2\times 3}\right\},$$

where $\gamma(2,2)$ can be calculated to be $1/(2^{61}\times 5^5)$ by [Lin24, Algorithm 1] with the input $\gamma(1,2) = \gamma(1,4) = 1/2$. Thus, $\kappa = 1/(2^{61}\times 5^5)$. It follows from (5.2) and (5.3) that

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\alpha_k) - \int\limits_{[0,1)^2} f\mathrm{d}\mu\right| \ll \delta(\omega)^{-1/(2^{61}\times 5^5)}.$$

This provides the convergence speed of the archimedean part of the height of the point $\mathcal{P}(\omega)$ as $\delta(\omega)\to\infty$ and answers [GS23, Question 6.2]. Moreover, it is shown in [GS23, Proposition 6.1] that $\int_{[0,1)^2} f\mathrm{d}\mu = \frac{2\zeta(3)}{3\zeta(2)}$, where ζ is the Riemann zeta function. The non-archimedean part of the height of $\mathcal{P}(\omega)$ was given in [GS23, Corollary 3.4], which states that

$$\sum_{v_p\in M_{\mathbb{Q}}\setminus\{\infty\}} \mathrm{h}_{v_p}(\mathcal{P}(\omega)) = -\frac{\Lambda(\mathrm{ord}(\omega))}{\phi(\mathrm{ord}(\omega))},$$

where Λ is the Von Mangoldt function. More precisely, if $\mathrm{ord}(\omega)$ is not a prime power, then the non-archimedean part of the height of $\mathcal{P}(\omega)$ is 0; if $\mathrm{ord}(\omega) = p^e$ for a prime number p and a positive integer e, then

$$\sum_{v_p\in M_{\mathbb{Q}}\setminus\{\infty\}} \mathrm{h}_{v_p}(\mathcal{P}(\omega)) = -\frac{\log p}{p^{e-1}(p-1)}.$$

In the case where $d \coloneqq \mathrm{ord}(\omega) = p^e$, we have

$$\left|\sum_{v_p\in M_{\mathbb{Q}}\setminus\{\infty\}} \mathrm{h}_{v_p}(\mathcal{P}(\omega))\right| = \frac{\log p}{p^{e-1}(p-1)} = \frac{p}{e(p-1)}\frac{\log d}{d} \leq \frac{2\log d}{d}.$$

Since $\delta(\omega) \leq d$ and the function $x \mapsto (\log x)/x$ decreases as $x\to\infty$, we get

$$\left|\sum_{v_p\in M_{\mathbb{Q}}\setminus\{\infty\}} \mathrm{h}_{v_p}(\mathcal{P}(\omega))\right| \ll \frac{2\log\delta(\omega)}{\delta(\omega)} \ll \delta(\omega)^{-1/2}$$

and so $\sum_{v_p \in M_{\mathbb{Q}} \setminus \{\infty\}} \mathrm{h}_{v_p}(\mathcal{P}(\omega)) \to 0$ as $\delta(\omega) \to \infty$. In summary, if $\delta(\omega)$ is sufficiently large, then we have

$$\left| \mathrm{h}(\mathcal{P}(\omega)) - \frac{2\zeta(3)}{3\zeta(2)} \right| \ll \delta(\omega)^{-1/(2^{61} \times 5^5)}$$

as $\delta(\omega) \to \infty$. Therefore, if $(\omega_\ell)_{\ell \geq 1}$ is a strict sequence of non-trivial torsion points in $\mathbb{G}_m^d(\overline{\mathbb{Q}})$, then

$$\left| \mathrm{h}(\mathcal{P}(\omega_\ell)) - \frac{2\zeta(3)}{3\zeta(2)} \right| \ll \delta(\omega_\ell)^{-1/(2^{61} \times 5^5)}$$

as $\ell \to \infty$, which proves Proposition 1.2.

Acknowledgment

This is part of the author's PhD project. The author gratefully acknowledges many helpful suggestions and carefully reading provided by the supervisors, Roberto Gualdi and Walter Gubler. The author also wishes to thank Riccardo Pengo for the reference suggestions regarding limits of Mahler measures.

The author is supported by the collaborative research center SFB 1085 *Higher Invariants* (Project-ID 224262486), funded by the DFG and Project PID2023-147642NB-I00 of the Universitat de Barcelona.

Data availability

Data will be made available on request.

References

[BCGT13] L. Brandolini, L. Colzani, G. Gigante, G. Travaglini, On the Koksma-Hlawka inequality, J. Complex. 29 (2) (2013) 158–172.

[BG06] E. Bombieri, W. Gubler, Heights in Diophantine Geometry, New Math. Monogr., vol. 4, Cambridge University Press, Cambridge, 2006.

[Bil97] Y. Bilu, Limit distribution of small points on algebraic tori, Duke Math. J. 89 (3) (1997) 465–476.

[Brø83] A. Brøndsted, An Introduction to Convex Polytopes, Grad. Texts Math., vol. 90, Springer, Cham, 1983.

[DH24] V. Dimitrov, P. Habegger, Galois orbits of torsion points near atoral sets, Algebra Number Theory 18 (11) (2024) 1945–2001.

[GS23] R. Gualdi, M. Sombra, Limit heights and special values of the Riemann zeta function, preprint, https://arxiv.org/abs/2304.01966, 2023.

[Har98] G. Harman, Metric Number Theory, Lond. Math. Soc. Monogr., New Ser., vol. 18, Clarendon Press, Oxford, 1998.

[Har10] G. Harman, Variations on the Koksma-Hlawka inequality, Unif. Distrib. Theory 5 (1) (2010) 65–78.

[KN74] L. Kuipers, H. Niederreiter, Uniform Distribution of Sequences, Pure and Applied Mathematics, John Wiley & Sons, a Wiley-Interscience Publication, New York etc., 1974, xiv, 390 pp.

[Lau84] M. Laurent, Exponential Diophantine equations, Invent. Math. 78 (1984) 299–327.
[Lin24] Chenying Lin, Galois orbits of torsion points over polytopes near atoral sets, arXiv preprint, 2024.
[LSV13] D. Lind, K. Schmidt, E. Verbitskiy, Homoclinic points, atoral polynomials, and periodic points of algebraic $\mathbb{Z}^d$-actions, Ergod. Theory Dyn. Syst. 33 (4) (2013) 1060–1081.
[Mat99] A.M. Mathai, An Introduction to Geometrical Probability. Distributional Aspects with Applications, Gordon and Breach Science Publishers, Amsterdam, 1999.
[Neu99] J. Neukirch, Algebraic Number Theory. Transl. from the German by Norbert Schappacher, Grundlehren Math. Wiss., vol. 322, Springer, Berlin, 1999.
[RW98] R.T. Rockafellar, R.J.B. Wets, Variational Analysis, Grundlehren Math. Wiss., vol. 317, Springer, Berlin, 1998.