Invent. math.
https://doi.org/10.1007/s00222-022-01101-4

Check for updates

Billiards, heights, and the arithmetic of non-arithmetic groups

Curtis T. McMullen[1]

Received: 8 June 2020 / Accepted: 13 January 2022

Abstract. In this paper we introduce a new height on $\mathbb{P}^1(K)$ associated to an Abelian variety with real multiplication by K, and use it to study non-arithmetic triangle groups, Teichmüller curves, and billiards in lattice polygons. Complementary results on matrix coefficients and measures are obtained using modular symbols. In particular, we show the matrix entries m of the classical Hecke group $\Delta(2, 5, \infty)$ are constrained by the condition that $-\gamma^{-2}(m'/m)$ lies in a countable, closed semigroup $S \subset [-1, 1]$ homeomorphic to $\omega^\omega + 1$.

Contents

Research supported in part by the NSF Typeset 2022-01-18 09:05.

✉ Curtis T. McMullen

[1] Mathematics Department, Harvard University, Cambridge, MA 02138-2901, USA

Journal: **222** Article No.: **1101** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2022/4/1** Pages: **43** Layout: **Small-X**

1 Introduction

In this paper we use methods from Hodge theory to describe the arithmetic of certain non-arithmetic lattices $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$. A key role is played by a new height on $\mathbb{P}^1(K)$ determined by an Abelian variety with real multiplication by K.

As a complement, we also use the modular symbols of [Mc6] to give a qualitative description of the matrix coefficients of Γ. The circle of ideas we will discuss has its origins in the theory of polygonal billiards.

Curves on Hilbert modular surfaces. We begin by briefly stating one of our two main results, discussed in detail in Sect. 5.

Theorem 1.1 *Let $V = \mathbb{H}/\Gamma \looparrowright X_K$ be a complex geodesic curve on a Hilbert modular surface. Then either V is a Shimura curve, or the cusps of Γ coincide with $\mathbb{P}^1(K)$ and satisfy quadratic height bounds.*

Here K is a real quadratic field, and V is isometrically immersed for the Kobayashi metric on X_K.

To illustrate this result, we describe in detail several motivating applications and complements. References are collected at the end of this section.

Triangle groups. Consider the triangle group

$$\Delta(p, q, \infty) = \langle S, T\rangle \subset \mathrm{SL}_2(\mathbb{R}),$$

generated by

$$S = \begin{pmatrix} \cos(\pi/p) & \sin(\pi/p) \\ -\sin(\pi/p) & \cos(\pi/p) \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix},$$

where τ is chosen so that $\mathrm{Tr}(ST) = -2\cos(\pi/q)$. Its invariant trace field is given by:

$$\begin{aligned} K_{pq} &= \mathbb{Q}(\mathrm{Tr}(g^2) \ : \ g \in \Delta(p, q, \infty)) \\ &= \mathbb{Q}(\cos(2\pi/p), \cos(2\pi/q), \cos(\pi/p)\cos(\pi/q)); \end{aligned}$$

and $\Delta(p, q, \infty)$ is arithmetic if and only if $K_{pq} = \mathbb{Q}$ (cf. [Tak, Prop. 5], [MR, Ex. 4.9]).

One can readily survey the global properties of an arithmetic group such as $\Delta(2, 3, \infty) = \mathrm{SL}_2(\mathbb{Z})$: every integer occurs as a matrix entry, every pair of relatively prime integers (a, c) occurs as a matrix column, and the cusps of $\mathrm{SL}_2(\mathbb{Z})$ coincide with $\mathbb{Q} \cup \{\infty\}$.

The non-arithmetic triangle groups are more mysterious. While it is easy to describe their generators, it is difficult to characterize the matrices they

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

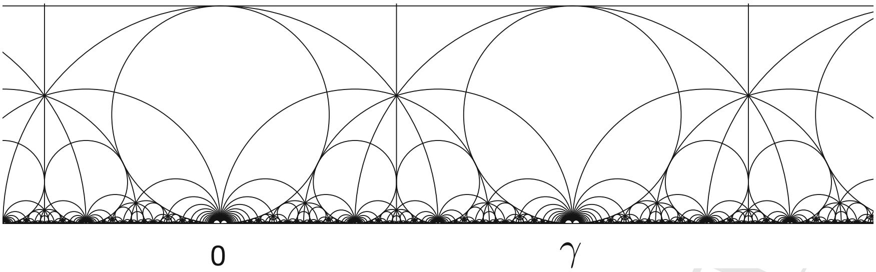

Fig. 1 Cusps for the triangle group $\Delta(2, 5, \infty)$

contain. On the other hand, it can be shown that every non-arithmetic triangle group arises (up to commensurability) from a geodesic curve on a Hilbert modular variety of dimension $g = \deg(K_{pq}/\mathbb{Q}) > 1$. Thus Theorem 1.1 yields (Sect. 6):

Corollary 1.2 *The cross-ratios of the cusps of the triangle group* $\Delta(p, q, \infty)$ *coincide with* $\mathbb{P}^1(K_{pq}) - \{0, 1, \infty\}$ *whenever* $\deg(K_{pq}/\mathbb{Q}) = 2$.[1]

Many instances of this result were first proved with a case-by-case analysis. The proof we present in Sect. 5 is a conceptual and effective descent argument.

Golden fractions. To explain the height bounds in Theorem 1.1, we will discuss the non-arithmetic group $\Gamma = \Delta(2, 5, \infty)$ in more detail.

Let $\gamma = (1 + \sqrt{5})/2$. Then $\mathcal{O} = \mathbb{Z}[\gamma]$ is the maximal order in the field $K = \mathbb{Q}(\gamma)$, γ is a fundamental unit, and the $(2, 5, \infty)$ triangle group is given by

$$\Gamma = \left\langle \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \right\rangle \subset \mathrm{SL}_2(\mathbb{R}).$$

Although it is a lattice in $\mathrm{SL}_2(\mathbb{R})$, it is equally natural to regard Γ as a *thin subgroup* of the arithmetic group $\mathrm{SL}_2(\mathcal{O}) \subset \mathrm{SL}_2(\mathbb{R})^2$. Via Theorem 1.1, the latter perspective yields the following three equivalent assertions.

1. The cusps of $\Delta(2, 5, \infty)$ coincide with $\mathbb{Q}(\gamma) \cup \{\infty\}$ (see Fig. 1).
2. Every $x \in \mathbb{Q}(\gamma)$ can be expanded as a finite golden continued fraction,

$$x = [a_1, a_2, a_3, \ldots, a_N] = a_1\gamma + \cfrac{1}{a_2\gamma + \cfrac{1}{a_3\gamma + \cdots \cfrac{1}{a_N\gamma}}} \tag{1.1}$$

[1] Quadratic trace fields occur for signature $(2, q, \infty)$ with $q = 5, 8, 10, 12$, $(3, q, \infty)$ with $q = 4, 5, 6$, and $(4, q, \infty)$ with $q = 6, 12$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

with $a_i \in \mathbb{Z}$.

3. Every $x \in \mathbb{Q}(\gamma)$ can be expressed as a *golden fraction* $x = a/c$, characterized by the property that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ for some b, d. This expression is unique up to a sign change, $x = (-a)/(-c)$.

Let us elaborate the last point. Since K has class number one, we can certainly write $x = A/B$ as a ratio of relatively prime integers $A, B \in \mathbb{Z}[\gamma]$. In fact, since γ is a unit, there are many such expressions: we also have $x = (\gamma^k A)/(\gamma^k B)$ for any $k \in \mathbb{Z}$.

The golden fraction expression $x = a/c$ uses the thin group Γ to pick out a particular value of k. The complexity of this expression is controlled by the height bounds in Theorem 1.1; in this case, they yield (Sect. 6):

Corollary 1.3 *The height of any nonzero golden fraction $x = a/c$ satisfies*

$$h(a) + h(c) = O(1 + h(x)^2). \tag{1.2}$$

Moreover the continued fraction of x satisfies $N + \max \log |a_i| = O(1 + h(x))$.

Here $h(x)$ is the absolute logarithmic height on $K = \mathbb{Q} \oplus \mathbb{Q}\gamma$; for $x = (p/q) + (r/s)\gamma$, it satisfies

$$1 + h(x)^2 \asymp 1 + (\log \max\{|p|, |q|, |r|, |s|\})^2.$$

Material on heights, abelian varieties with real multiplication, and Hilbert modular varieties is developed in Sects. 2, 3 and 4, in preparation for the proof of Theorem 1.1 in Sect. 5. The exponent 2 in Eq. (1.2) is sharp; see Sect. 7.

Matrix coefficients. We now turn to complementary results, based on modular symbols.

Let $M \subset \mathbb{Z}[\gamma]$ denote the set of all matrix entries that occur in Γ. The discussion of golden fractions above shows that

$$\mathbb{Z}[\gamma] = \bigcup_{k \in \mathbb{Z}} \gamma^k M.$$

As noted by Leutbecher in the 1970s [Le], there is no known characterization of the elements of M. The next result gives a qualitative description of M and also reveals its hidden multiplicative structure.

Derived sets. In preparation for the statement, recall that any compact, countable metric space E is homeomorphic to a countable ordinal. The *derived set* DE is E with its isolated points removed; $D^{n+1}(E) = D^n(DE)$; and $D^\infty E = \bigcap D^n(E)$. The derived set $D^\infty E$ is a single point if and only if E is homeomorphic to the ordinal $\omega^\omega + 1$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

The ratio set. Now let $x \mapsto x'$ denote the generator of $\mathrm{Gal}(K/\mathbb{Q}) = \mathbb{Z}/2$, define

$$\delta : K^{\times} \to K^{\times}$$

by $\delta m = m'/m$, and let

$$\delta M = \{m'/m \ : \ 0 \neq m \in M\}. \tag{1.3}$$

Define the (signed) *ratio set* for Γ by

$$S = -\gamma^{-2} \cdot \delta M. \tag{1.4}$$

In Sect. 7 we will show:

Theorem 1.4 *The closure of the signed ratio set* $\overline{S}$ *is a compact, countable subset of* $[-1, 1]$ *satisfying* $\overline{S} \cdot \overline{S} \subset \overline{S}$. *In fact we have*

$$\overline{S} = \langle S \rangle \cup \{0\} \cong \omega^{\omega} + 1,$$

and $D^{\infty}\overline{S} = \{0\}$.

Here $\langle S \rangle$ denotes the multiplicative semigroup generated by S.

Unlike Corollary 1.3, this result is not limited to quadratic fields. Similar results on matrix entries hold for all triangle groups with $\deg(K_{pq}/\mathbb{Q}) > 1$, and the general statement, Theorem 1.7 below, applies to any group with a contracting twist.

For a hint of the complexity of the set δM, see Fig. 2. Note that if Γ were replaced by the arithmetic group $\mathrm{SL}_2 \mathbb{Z}[\gamma]$, the set of ratios m'/m would become dense in $\mathbb{R}$; in fact every element in K with norm 1 would arise, by Hilbert's Theorem 90.

Billiards. Many more examples of non-arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$ arise from the theory of polygonal billiards, leading to dynamical applications of Theorem 1.1.

To illustrate some of these, consider a finite polygon $P \subset \mathbb{C}$ with internal angles $\pi(a_1, \ldots, a_n)/q$, where $\gcd(a_1, \ldots, a_n, q) = 1$. A standard unfolding construction associates to P a holomorphic 1-form (X, ω) on a compact Riemann surface, together with an action of the dihedral group D_{2q} such that $(X, |\omega|)/D_{2q}$ is isometric to $(P, |dz|)$. The affine symmetries of (X, ω) give rise to a discrete group $\mathrm{SL}(X, \omega) \subset \mathrm{SL}_2(\mathbb{R})$.

Lattice polygons. We say P is a *lattice polygon*, with *trace field* K, if $V = \mathbb{H}/\mathrm{SL}(X, \omega)$ has finite volume and K is the trace field of $\mathrm{SL}(X, \omega)$. It is known that K is a number field, and that $\mathrm{SL}(X, \omega)$ is arithmetic if and only if $K = \mathbb{Q}$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

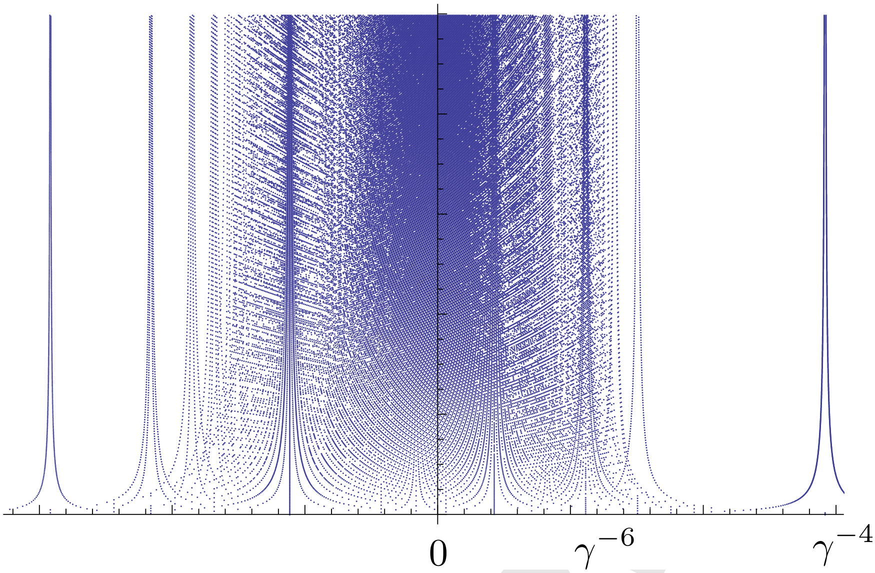

Fig. 2 Matrix coefficients of $\Delta(2, 5, \infty)$. The scatterplot shows part of the image of M under $m = x + y\gamma \mapsto (m'/m, \max(|x|, |y|))$

By a well-known result of Veech [V, Prop. 2.11], lattice polygons enjoy optimal dynamics: either all geodesics with slope s are periodic, or all are uniformly distributed. In the periodic case, we denote the length of the longest periodic trajectory by $L(s)$.

Let us say P is *normalized* if (i) it has a vertical or horizontal edge, and (ii) when $q = 2$, $s = 1$ is the slope of a periodic trajectory. The first condition can be achieved rotating P. The second condition, which arises only when all sides of P are vertical or horizontal, can be achieved by an affine stretch (which respects billiard paths). Theorem 1.1 then entails (Sect. 6):

Corollary 1.5 *Let P be a normalized lattice polygon with quadratic trace field K. Let $\alpha = \tan(\pi/q)$ if $q \geq 3$, and $\alpha = 1$ if $q = 2$. Then its periodic slopes are given by*

$$S(P) = \alpha K \cup \{\infty\}.$$

Moreover, for any $s \in K$ we have

$$\log L(\alpha s) = O(1 + h(s)^2). \tag{1.5}$$

A similar result (Theorem 6.1) holds for Teichmüller curves.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

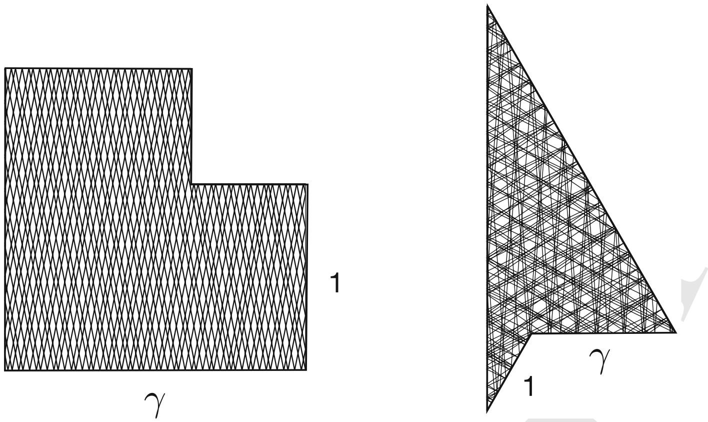

Fig. 3 Long periodic billiard paths, each with over 200 segments, with initial slopes 5 and $8\sqrt{3}$ respectively

Example 1. The golden L**.** One of the simplest examples of a lattice polygon with quadratic trace field is the symmetric L-shaped region P built by attaching two golden rectangles to the unit square, as shown at the left in Fig. 3.

Computer experiments quickly reveal that even small, rational slopes lead to very long trajectories in P; for example, $L(5) \approx 479$, while $L(6765) \approx 1.734 \times 10^{25}$. This rapid growth, which is consistent with Eq. (1.5), is explained in Sect. 7.

The 1-form associated to P satisfies $\mathrm{SL}(X, \omega) = \Delta(2, 5, \infty)$. Using this connection, we will give a simple dynamical proof that

$$a + b\gamma \in M \implies ab \geq 0$$

and hence

$$-\gamma^{-2} \leq m'/m \leq 1 \tag{1.6}$$

for all matrix entries $m \neq 0$ in $\Delta(2, 5, \infty)$. Equality occurs when $m = 1$ and $m = \gamma$; see Corollary 7.4.

Example 2. The golden arrow. A second lattice polygon, also based on the golden ratio, is shown at the right in Fig. 3; its internal angles are $\pi(1, 1, 2, 8)/6$, and its periodic slopes are given by $S(P) = \sqrt{3} \cdot \mathbb{Q}(\gamma) \cup \{\infty\}$.

Both examples belong to infinite families, discussed in [Mc1, §9] and [EMMW, §8] respectively, and their side lengths can be varied to produce infinitely many different quadratic trace fields.

Cubic and rational trace fields. We remark that quadratic trace fields are the main case of interest for lattice billiards. Indeed, when $K = \mathbb{Q}$ the billiards in P is closely related to billiards in a square, and a closed trajectory at slope

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

$s = p/q$ has length $O(|p| + |q|)$. On the other hand, lattice polygons with trace fields of cubic and higher degree are rigid; that is, they are determined by their internal angles up to finitely many choices [EFW, Cor. 1.6]. These rigid examples include the regular polygons, discovered by Veech, provided we exclude the n-gons with $n = 3, 4, 5, 6, 8, 10$ and 12.

Distribution of long billiard trajectories. We now turn to an application which was our original motivation for proving Theorem 1.4.

Let P be a lattice polygon with trace field $K \neq \mathbb{Q}$. Even though billiards in P has optimal dynamics, long periodic trajectories in P need not be evenly distributed [DL]. To describe their behavior, let M_s denote the set of probability measures on P that arise as limits of closed trajectories with slopes $s_n \to s$ and lengths tending to infinity. When s is not a periodic slope, equidistribution holds and M_s is a single point; but for periodic slopes, equidistribution fails and we have $M_s \cong \omega^\omega + 1$ [Mc6].

Now suppose P is the golden L and $s = 0$. In this case, M_0 can be described directly in terms of the set δM defined by Eq. (1.3). For a precise statement, first observe that P can be regarded as two stacked golden rectangles, A_1 and A_2, the first of width 1 and the second of width γ. Let $\alpha_i = \chi_{A_i}|dz|^2$, $i = 1, 2$ denote their respective area measures, and let $\widehat{\nu}(r)$ denote the unique probability measure on P proportional to

$$\nu(r) = (1 - r)\alpha_1 + (1 + \gamma^{-2} r)\alpha_2. \tag{1.7}$$

We then have (Sect. 7):

Theorem 1.6 *The limit measures for billiards in the golden L at slope $s = 0$ are given by*

$$M_0 = \{\widehat{\nu}(r) \ : \ r \in \delta M\}.$$

By (1.6), the two most unevenly distributed measures in M_0 are those proportional to

$$\nu_R = \alpha_2 \quad \text{and} \quad \nu_L = \gamma\alpha_1 + \alpha_2.$$

These arise, as $n \to \infty$, as limits of trajectories of slope $1/(n\gamma)$ starting near the right and left of the bottom edge of P, respectively. Since $\Delta(2, 5, \infty)$ has only one cusp, the measures M_s for other periodic slopes are essentially the same as those for M_0.

Compression of cusps. A dynamical proof of Theorems 1.4 and 1.6 will be given in Sect. 7. We conclude by stating our second main theorem, which provides another route to Theorem 1.4 (and many similar results) in the spirit of Theorem 1.1.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Let $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ be a lattice with at least one cusp. Let

$$\rho : \Gamma \to \Gamma' \subset \mathrm{SL}_2(\mathbb{R})$$

be an isomorphism of abstract groups, such that

$$g \textit{ is parabolic} \iff \rho(g) \textit{ is parabolic}. \tag{1.8}$$

We do not require that Γ' is discrete. Let $F : \mathbb{H} \to \mathbb{H}$ be a holomorphic map, such that

$$F(g \cdot t) = \rho(g) \cdot F(t) \tag{1.9}$$

for all $g \in \Gamma$ and $t \in \mathbb{H}$. Then the pair (ρ, F) defines a *twist* of Γ. If F is an isometry, then ρ is simply conjugation by F. Otherwise, the twist is *contracting*.

Given a pair of distinct cusps x, y of Γ, fixed by parabolics $g, h \in \Gamma$, let

$$r(x, y) = \frac{\mathrm{Tr}(\rho(gh)) - 2}{\mathrm{Tr}(gh) - 2}. \tag{1.10}$$

This quantity is independent of the choice of g and h.

We will see that $r(x, y) \in [0, 1]$. In fact, when g and h generate the parabolic stabilizers of x and y, the quantity

$$D(x, y) = \log|\mathrm{Tr}(gh) - 2|$$

can be interpreted as the *renormalized distance* between these two cusps, and $\log r(x, y)$ measures the amount this distance is reduced by F.

Define the (absolute) *ratio set* by

$$R = \{r(x, y) \ : \ x, y \textit{ are distinct cusps of } \Gamma\} \subset [0, 1]. \tag{1.11}$$

In Sect. 8 we will show:

Theorem 1.7 *The ratio set of a contractive twist of* Γ *as above satisfies* $\overline{R} = \langle R \rangle \cup \{0\} \cong \omega^\omega + 1$ *and* $D^\infty(\overline{R}) = \{0\}$.

This result implies Theorem 1.4 and related results for other triangle groups $\Delta(p, q, \infty)$, as well as for the lattices $\mathrm{SL}(X, \omega)$ associated to Teichmüller curves.

For these groups, a finite cover of $V = \mathbb{H}/\Gamma$ can be realized as a complex geodesic on a Hilbert modular variety X_K, $d = \deg(K/\mathbb{Q})$, and one obtains $d - 1$ contracting twists by projecting to the factors of its universal cover $\mathbb{H}^d$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

See [CW, §2] and corrections in [Ri, §2] for the case of triangle groups, and [Mo1] for the case of Teichmüller curves.

Questions. We conclude with some open problems related to fields of cubic and higher degree.

1. Is there an $n > 0$ such that $\log L(s) = O(1 + h(s)^n)$ for all periodic slopes in the regular heptagon?
2. Let $V \subset X_K$ be a Kobayashi geodesic curve on a Hilbert modular threefold. If V has no cusps, is V a Shimura curve?
3. Suppose $\mathrm{SL}(X, \omega)$ is a lattice with trace field K, and the cross-ratios of its cusps coincide with $\mathbb{P}^1(K) - \{0, 1, \infty\}$. Does this imply that $\deg(K/\mathbb{Q}) \le 2$?

Notes and references. This paper is a sequel to [Mc5] and [Mc6].

The theory of triangle groups has a long history, including works by Schwarz, Fricke, Klein, Hecke and many others (see e.g. [Sch, He, Mag], [SG, Ch. 14]). Many cases of Corollary 1.2 were proved first in [Le, Be] and [Se]; the general case follows from [Mc2] and [BM] (see Sect. 6), and was also addressed recently in [Pa]. See [Bh, BR, Wo, AS] for work on Question (3) in the case of triangle groups, resolving the case $\Delta(2, q, \infty)$; some additional cases are covered by [CSc].

The geodesic curves on Hilbert modular varieties coming from triangle groups are discussed in [CW]. For more on connections between Hilbert modular surfaces, Teichmüller curves and Kobayashi geodesics, see [Mc1, Mc4, Mo1, MV], and [We].

An encoding of the periodic trajectories in the regular pentagon and the golden L is studied in detail in [DL]; Theorem 1.6 and [Mc6] address [DL, §4, Conj. 4.6]. See [Bo, Theorem 7.9] for related results on periodic points for interval exchange transformations.

Notation. The expressions $A = O(B)$ and $A \asymp B$ mean $A \le CB$ and $A/C \le B \le CA$ for some unspecified constant $C > 0$.

2 Heights

In this section we review the theory of heights for number fields, projective spaces, and groups such as $\mathrm{SL}_2(K)$, with an emphasis on formulations using integers and infinite places.

Useful references include [BG, La] and [HS].

Absolute values. Let K be a number field of degree d over $\mathbb{Q}$, and let $\mathcal{O}$ be the ring of integers in K. Each place v of K determines a normalized absolute value $|x|_v$ on K; taken together, these satisfy the product formula $\prod_v |x|_v = 1$. For $K = \mathbb{Q}$ the absolute values are normalized so that $|p|_p = 1/p$ and $|x|_\infty = |x|$;

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

in general, the absolute values are normalized so that

$$|x|_v = |N^{K_v}_{\mathbb{Q}_p}(x)|_p^{1/d}$$

whenever $v|p$. For example, if v is a real place of K, and $\rho_v : K \to K_v \cong \mathbb{R}$ is the associated completion, then $|x|_v = |\rho_v(x)|^{1/d}$.

Heights on projective space. The absolute multiplicative height on $\mathbb{P}^n(K)$ is given by

$$H(x) = H(x_0 : x_1 : \cdots : x_n) = \prod_v \max_i |x_i|_v.$$

It is well-defined by the product formula, which also implies that $H(x) \geq 1$.

The normalizations above are chosen so that $H(x)$ remains constant under finite extensions of K. In addition, an automorphism f of $\mathbb{P}^n(K)$ changes the height by at most a bounded factor; we have

$$H(f(x)) \asymp H(x) \tag{2.1}$$

for all $x \in \mathbb{P}^n(K)$ (see e.g. [HS, Theorem B.3.1].)

Logarithmic height. The logarithmic height on $\mathbb{P}^1(K)$ is defined by

$$h(x) = \log H(x) \geq 0.$$

Throughout this paper we adopt the usual convention that multiplicative heights (such as H, H_A and H_τ) are written in upper case, and their logarithms (such as h, h_A, h_τ) in lower case.

Integer coordinates. A closely related height on $\mathbb{P}^n(K)$ can be defined by

$$\widetilde{H}(x) = \inf_a \prod_{v|\infty} \max_i |a_i|_v, \tag{2.2}$$

where the infimum is taken over vectors of *integers* $a \in \mathcal{O}^{n+1}$ such that $[a_0 : \cdots : a_n] = x$. This height is comparable to the standard one; using finiteness of the class number, one can show that

$$H(x) \asymp \widetilde{H}(x),$$

and equality holds when $\mathcal{O}$ is a UFD (cf. [La, §3.1]). The implicit constants depend on (K, n).

One useful feature of formula (2.2) is that it involves only the *integers* $\mathcal{O}$ and the *infinite places* of K. This motivates our definition of the height $H_A(x)$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

on $\mathbb{P}^1_A(K)$ in Sect. 3, where $\mathcal{O}^2$ will be replaced by the integral homology $H_1(A, \mathbb{Z})$ of an Abelian variety with real multiplication by K.

Heights on affine space. The restriction of H to the affine part of $\mathbb{P}^n(K)$ gives a natural height on K^n. The height on K itself, given explicitly by

$$H(x) = \prod_v \max(1, |x|_v), \tag{2.3}$$

satisfies $H(x^n) = H(x)^n$, $H(xy) \leq H(x)H(y)$ and $H(x+y) \leq 2H(x)H(y)$ (an extreme case is $x = y = 1$).

Lemma 2.1 *For any embedding $\rho_v : K \to \mathbb{R}$, we have*

$$H(x)^d \geq \rho_v(x) \geq H(x)^{-d}$$

for all $x \neq 0$ in K.

Proof. Since each term in the product (2.3) is at least 1, we have $H(x) \geq |x|_v = |\rho_v(x)|^{1/d}$, giving the upper bound. For the lower bound, use the fact that $H(1/x) = H(x)$. $\square$

Totally real fields and $\mathrm{SL}_2(\mathbb{R})$. We conclude with some observations of use in the sequel.

Let $\rho = |dt|/\operatorname{Im}(t)$ be the hyperbolic metric on $\mathbb{H} = \{t : \operatorname{Im}(t) > 0\}$, and let $d(p, q)$ denote the hyperbolic distance between a pair of points.

The group $\mathrm{SL}_2(\mathbb{R})$ acts linearly on $\mathbb{R}^2$ and isometrically on $\mathbb{H}$. The operator norm of $g \in \mathrm{SL}_2(\mathbb{R})$ and its translation distance on $\mathbb{H}$ are related by

$$\log \|g\|_2^2 = d(i, g(i)) \geq 0. \tag{2.4}$$

To see this, use the polar decomposition to reduce to the case where $g = \left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)$, $a \geq 1$, and both sides become $\log(a^2)$. We also note that, since any two norms on $M_2(\mathbb{R})$ are equivalent, we have

$$1 \leq \left\| \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \right\|_2 \asymp \max(|a|, |b|, |c|, |d|). \tag{2.5}$$

Now let K be a totally real field of degree d over $\mathbb{Q}$. The infinite places of K determine an embedding $K \subset \mathbb{R}^d$ sending k to (k_i); similarly, we have an embedding

$$\mathrm{SL}_2(K) \subset \mathrm{SL}_2(\mathbb{R})^d$$

sending g to (g_i). We define the height of $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(K)$ by

$$H(g) = H(a, b, c, d) \quad \text{and} \quad h(g) = \log H(g), \tag{2.6}$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

using the inclusion $\mathrm{SL}_2(K) \subset K^4 \subset \mathbb{P}^4(K)$. Note that $H(g) = H(-g) = H(g^{-1})$, and

$$\|g_i\|_2 = O(H(g)^d) \tag{2.7}$$

for all i, by Lemma 2.1.

Lemma 2.2 *For all $g \in \mathrm{SL}_2(\mathcal{O})$, we have $H(g) \asymp \prod_1^d \|g_i\|_2^{1/d}$.*

Proof. In the product formula for the $H(g)$, the finite places each contribute 1 since the entries of g are integers; and at the infinite places, which are indexed by i, we can apply Eq. (2.5). □

Hyperbolic elements. For hyperbolic elements $g \in \mathrm{SL}_2(\mathbb{R})$, we recall for later use that the translation length

$$T(g) = \inf_{t \in \mathbb{H}} d(t, g(t))$$

and trace are related by

$$|\operatorname{Tr}(g)| = 2\cosh(T(g)/2), \tag{2.8}$$

as can be verified by reducing to the case where g is diagonal.

3 Abelian varieties with real multiplication

In this section we use Hodge theory to introduce a natural height $H_A(x)$ on the space of slopes $\mathbb{P}^1_A(K)$ attached to an Abelian variety A with real multiplication by K. Our terminology is justified by

Proposition 3.1 *For any projective linear isomorphism $\iota : \mathbb{P}^1_A(K) \cong \mathbb{P}^1(K)$, we have*

$$H_A(x) \asymp H(\iota(x)).$$

This height, of interest in its own right, underlies the descent argument used to prove Theorem 1.1 in Sect. 5.

In the next section we study the behavior of $H_A(x)$ as A varies in a Hilbert modular variety. For background on these topics, see [BL] and [vG].

Abelian varieties. Let A be a polarized Abelian variety of dimension d. We can naturally identify A with the quotient space

$$A = \Omega(A)^*/H_1(A, \mathbb{Z}),$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

where $\Omega(A) \cong \mathbb{C}^d$ is the space of holomorphic 1-forms on A, and its paring with $H_1(A, \mathbb{Z}) \cong \mathbb{Z}^{2d}$ is given by $\langle C, \omega\rangle = \int_C \omega$.

The polarization of A is recorded by a positive-definite Hermitian inner product on $\Omega(A)^*$, with the property that the symplectic form

$$[C, D] = -\operatorname{Im}\langle C, D\rangle \tag{3.1}$$

takes integral values on $H_1(A, \mathbb{Z})$. We denote the associated *Hodge norm* on $H_1(A, \mathbb{R})$ by $\|C\|_A = \langle C, C\rangle^{1/2}$. The polarization also determines a norm and inner product on $\Omega(A)$, via duality.

Example: The Jacobian. When $A = \Omega(X)^*/H_1(X, \mathbb{Z})$ is the Jacobian of a compact Riemann surface X of genus g, we have a natural (principal) polarization given by the dual of the Hermitian form

$$\langle \omega_1, \omega_2\rangle = \frac{i}{2}\int_X \omega_1 \wedge \overline{\omega}_2$$

on $\Omega(X)$. The associated symplectic form $[C, D]$ agrees with the usual intersection form on $H_1(X, \mathbb{Z})$, and the Hodge norm on homology is given by

$$\|C\|_X = \sup\left\{\left|\int_C \omega\right| \,:\, \langle\omega, \omega\rangle = 1\right\}.$$

Real multiplication. Now let K be a totally real field of degree d over $\mathbb{Q}$. We say A has *real multiplication by* K if it is equipped with an inclusion

$$K \subset \operatorname{End}(A) \otimes \mathbb{Q},$$

sending each $k \in K$ to a self-adjoint operator T_k. (This means $\langle T_k C, D\rangle = \langle C, T_k D\rangle$ for all $C, D \in H_1(A, \mathbb{R})$.)

The action of T_k on homology makes $H_1(A, \mathbb{Q})$ into a two-dimensional vector space over K; hence we can form the associated projective line

$$\mathbb{P}^1_A(K) = (H_1(A, \mathbb{Q}) - \{0\})\,/K^\times \cong \mathbb{P}^1(K). \tag{3.2}$$

We can also decompose $H_1(A, \mathbb{R}) \cong \mathbb{R}^{2d}$ into orthogonal eigenspaces $S_v \cong \mathbb{R}^2$, indexed by the infinite places of K.

Let $\pi_v : H_1(A, \mathbb{R}) \to S_v$ be orthogonal projection to the v-eigenspace, and define $N_A : H_1(A, \mathbb{R}) \to \mathbb{R}$ by

$$N_A(C) = \prod_{v|\infty} \|\pi_v(C)\|_A.$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Note that

$$N_A(T_k(C)) = |N_{\mathbb{Q}}^K(k)| \cdot N_A(C) \tag{3.3}$$

for all $k \in K$; in particular, the value of $N_A(C)$ is invariant under the action of the units in K. When $C \in H^1(A, \mathbb{Z})$ is an integral cycle, we have

$$N_A(C) = \prod_{v|\infty} \left| \int_C \omega_v \right|, \tag{3.4}$$

where (ω_v) is an orthonormal eigenbasis in $\Omega(A)$.

Hodge norms and height. Using real multiplication and the Hodge norm, we now define a height on $\mathbb{P}^1_A(K)$ by

$$H_A(x) = \inf\{N_A(C)^{1/d} \ : \ C \in H_1(A, \mathbb{Z}) \quad \text{and} \quad [C] = x\}.$$

Here $[C]$ denotes the point in $\mathbb{P}^1_A(K)$ represented by C. This height measures the minimum size of an integral homology class C in a given orbit of $K^\times$ acting on $H_1(A, \mathbb{Q})$.

Proof of Proposition 3.1. First recall that the height $\widetilde{H}(x)$ on $\mathbb{P}^1(K)$ defined by (2.2) is comparable to $H(x)$. Moreover, $\widetilde{H}(x)$ changes by at most a bounded factor if we change its definition by (i) replacing $\mathcal{O}^2$ by a commensurable lattice $L \subset K^2$; or (ii) replacing $\max\{|a_1|_v, |a_2|_v\}$ by another norm on $K_v^2 = \mathbb{R}^2$.

Lift ι to a K-linear map $I : H_1(A, \mathbb{Q}) \cong K^2$. Since (i) I sends $H_1(A, \mathbb{Z})$ to a lattice commensurable to $\mathcal{O}^2$, and (ii) I_v sends the Hodge norm on S_v to a norm on K_v^2, we have $H_A(x) \asymp \widetilde{H}(\iota(x)) \asymp H(x)$. $\square$

Remark One can similarly use the Hodge norm on any Abelian variety A to define a height

$$H_A(x) = \inf\{\|C\|_A \ : \ C \in H_1(A, \mathbb{Z}) \text{ and } [C] = x\}$$

on $\mathbb{P}H^1(A, \mathbb{Q}) \cong \mathbb{P}^{2d-1}(\mathbb{Q})$. Other rings of endomorphisms besides $K \subset \mathrm{End}(A) \otimes \mathbb{Q}$ can also be considered.

4 Hilbert modular varieties

In this section we introduce a height $H_\tau(x)$ on $\mathbb{P}^1(K)$ for each point $\tau \in \mathbb{H}^d$, the universal cover of the Hilbert modular variety X_K. This height is given by

$$H_\tau(x) = H_{A_\tau}(x),$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

where $A_\tau = \mathbb{C}^d/\mathcal{O} \oplus \mathcal{O}^\vee \tau$.

We will establish two results that control the behavior of $H_\tau(x)$ as τ varies. The first gives some insight into the geometric meaning of the height. Let

$$\delta(\tau) = \inf_x H_\tau(x) > 0$$

Proposition 4.1 *The function $\delta(\tau)$ is comparable to the length of the shortest closed geodesic on A_τ.*

Corollary 4.2 *For each $r > 0$, the locus $X_K(r) = \{[\tau] \in X_K : \delta(\tau) \geq r\}$ is compact, and $X_K = \bigcup_{r>0} X_K(r)$.*

The second result is conveniently phrased in terms of the logarithmic heights

$$h(x) = \log H(x) \quad \text{and} \quad h_\tau(x) = \log H_\tau(x). \tag{4.1}$$

Proposition 4.3 *For any compact set $D \subset \mathbb{H}^d$, there is an $M > 0$ such that*

$$|h_\tau(x) - h(x)| \leq M$$

for all $x \in \mathbb{P}^1(K)$ and $\tau \in D$.

Construction of A_τ. We begin with some definitions. Let $\mathcal{O}$ be the ring of integers in a totally real field K of degree d, and let

$$\mathcal{O}^\vee = \{a \in K : \mathrm{Tr}_{\mathbb{Q}}^K(ab) \in \mathbb{Z} \ \ \forall b \in \mathcal{O}\}$$

be its dual (the inverse different). There is a natural unimodular symplectic form on $\mathcal{O} \oplus \mathcal{O}^\vee$, defined by

$$[(a, b), (c, d)] = \mathrm{Tr}_{\mathbb{Q}}^K \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{4.2}$$

Let $(v_1, \ldots, v_d)$ denote the infinite places of K, and let $a \mapsto (a_i)$ be the corresponding embedding $K \to \mathbb{R}^d$. Then for each $\tau \in \mathbb{H}^d$ we have a natural map $K^2 \to \mathbb{C}^d$, sending (a, b) to $(a_i + b_i\tau_i)$. The image of $\mathcal{O} \oplus \mathcal{O}^\vee$ is a lattice, and

$$A_\tau = \mathbb{C}^d/(\mathcal{O} \oplus \mathcal{O}^\vee \tau)$$

is a principally polarized Abelian variety with real multiplication by K. In fact, we have an inclusion $\mathcal{O} \subset \mathrm{End}(A)$ sending a to the linear map $T_a(z) = (a_i z_i)$ on $\mathbb{C}^d$. The polarization of A_τ is uniquely determined by the symplectic form

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

(4.2). The corresponding inner product on $\mathbb{C}^d$, given by

$$\langle z, w\rangle_\tau = \sum_1^d \frac{z_i \overline{w}_i}{\operatorname{Im}\tau_i},$$

is readily verified to satisfy (3.1); and T_a is self-adjoint because it is given by a real diagonal matrix. The inner product above makes A_τ into a flat torus of volume one, since the symplectic form in (4.2) is unimodular.

The moduli space X_K. The isomorphism class of A_τ depends only on the location of τ in the Hilbert modular variety

$$X_K = \mathbb{H}^d / \operatorname{SL}(\mathcal{O} \oplus \mathcal{O}^\vee).$$

Here $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \operatorname{SL}_2(K)$ acts on $\mathbb{H}^d$ by

$$g \cdot \tau = ((a_i\tau_i + b)/(c_i\tau_i + d_i)),$$

and $\operatorname{SL}(\mathcal{O} \oplus \mathcal{O}^\vee)$ is the subgroup stabilizing the lattice $\mathcal{O} \oplus \mathcal{O}^\vee \subset K^2$. The action on $\mathbb{P}^1(K)$, used below, is given similarly by

$$g \cdot x = (ax + b)/(cx + d).$$

Markings and periods. We will work in the universal cover $\mathbb{H}^d$ of X_K, so that for $A = A_\tau$ we have a natural isomorphism

$$\mathcal{O} \oplus \mathcal{O}^\vee \cong H_1(A, \mathbb{Z}). \tag{4.3}$$

An orthonormal eigenbasis for $\Omega(A)$ is given by

$$\omega_i = dz_i / \sqrt{\operatorname{Im}\tau_i}, \tag{4.4}$$

$i = 1, \ldots, g$, and for $C = (a, b) \in \mathcal{O} \oplus \mathcal{O}^\vee \cong H_1(A, \mathbb{Z})$ we have

$$\int_C \omega_i = a_i + b\tau_i.$$

Norms and heights. We also have an isomorphism

$$\mathbb{P}^1_A(K) \cong \mathbb{P}^1(K) = K \cup \{\infty\},$$

compatible with (4.3), which we normalize so that $x \in K \subset \mathbb{P}^1(K)$ corresponds to the line in K^2 defined by

$$a + bx = 0. \tag{4.5}$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

(The point $x = \infty$ corresponds to the line $b = 0$.) With these identifications, we can regard N_A and H_A as functions N_τ and H_τ on $\mathcal{O} \oplus \mathcal{O}^\vee$ and $\mathbb{P}^1(K)$ respectively, given explicitly by

$$N_\tau(a, b) = \prod_1^d \frac{|a_i + b_i \tau_i|}{\sqrt{\operatorname{Im} \tau_i}} \quad \text{and}$$

$$H_\tau(x) = \inf_{a+bx=0} N_\tau(a, b)^{1/d},$$

as can be seen using Eq. (3.4).

Naturality. Since the height $H_A(x)$ is functorial, we have

$$H_\tau(x) = H_{g\cdot\tau}(g \cdot x) \tag{4.6}$$

for all $g \in \mathrm{SL}(\mathcal{O} \oplus \mathcal{O}^\vee)$. It follows that $\delta(g \cdot \tau) = \delta(\tau)$. One can also check that $\delta(\tau)$ is continuous, using the formula above.

The case of elliptic curves. It is easy to describe the height and its minimum $\delta(\tau)$ geometrically when $d = \dim(A) = 1$. In this case $A_\tau = \mathbb{Z} \oplus \mathbb{Z}\tau$ carries a unique flat metric $|dz|/\operatorname{Im}(\tau)^{1/2}$ of area one, $H_\tau(x)$ is the length of a shortest closed geodesic on A_τ with homological slope x, and the shortest closed geodesic on A_τ has length $\delta(\tau) > 0$.

Units: Proof of Proposition 4.1. These statements can be generalized to $d > 1$ as follows.

Note that for any unit $\epsilon \in \mathcal{O}$ we have $\prod |\epsilon_i| = 1$ and hence

$$N_\tau(\epsilon a, \epsilon b) = N_\tau(a, b).$$

By Dirichlet's unit theorem, the map $\epsilon \to \log |\epsilon_i|$ sends $\mathcal{O}^\times$ to a lattice in $\mathbb{R}^{d-1}$. Thus we can always adjust (a, b) by a unit so that the terms in the product formula for $N_\tau(a, b)$ have approximately the same size. Using the inequality between the arithmetic and geometric means, it follows that:

$$H_\tau(x) \asymp \inf_{a+bx=0} \|a + b\tau\|_\tau,$$

using the norm associated to the inner product (4.4). Proposition 4.1 follows immediately: we have

$$\delta(\tau) \asymp \inf\{\|\lambda\|_\tau \ : \ 0 \neq \lambda \in \mathcal{O} \oplus \mathcal{O}^\vee \tau\}. \tag{4.7}$$

$\square$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Proof of Corollary 4.2. By Mahler's criterion, the space of lattices in $\mathbb{C}^d$ with volume one and shortest vector of length $\geq r > 0$ is compact. Since $\mathrm{vol}(A_\tau) = 1$, the condition $\delta(\tau) \geq r > 0$ also defines a compact subset of X_K by Eq. (4.7). $\square$

Control of H_τ. To conclude we estimate the variation of $h_\tau(x)$ with respect to τ.

Proposition 4.4 *Given $(a, b) \neq (0, 0) \in \mathbb{R}^2$ and $t \in \mathbb{H}$, let*

$$\eta(t) = \log \frac{|a + bt|^2}{\operatorname{Im} t}.$$

Then the level sets of η are horocycles resting on $x = -a/b \in \partial\mathbb{H}$, and $\|d\eta\| = 1$ in the hyperbolic metric.

Proof. Using the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$ and $\mathbb{R}^2$, we can reduce to the case where $(a, b) = (1, 0)$. Then $\eta(\tau) = -\log \operatorname{Im} \tau$, $-a/b = \infty$ and the result is immediate. $\square$

Corollary 4.5 *The trajectories of the vector field $-\nabla\eta$ are unit speed geodesics converging to $x = -a/b$.*

The relation $x = -a/b$ above explains our convention (4.5).

Now observe that we can write

$$h_\tau(x) = \frac{1}{2d} \inf_{a+bx=0} \sum_1^d \log \frac{|a_i + b_i \tau_i|^2}{\operatorname{Im} \tau_i}. \tag{4.8}$$

Applying the estimate above to each term, we obtain

Corollary 4.6 *For any $\sigma, \tau \in \mathbb{H}^d$ we have*

$$\sup_x |h_\sigma(x) - h_\tau(x)| \leq (1/2) \max_i d(\sigma_i, \tau_i).$$

(The maximum on the right gives the Kobayashi distance $d_K(\sigma, \tau)$; see Sect. 5.)

Proof of Proposition 4.3. The function $M(\tau) = \sup_x |h_\tau(x) - h(x)|$ is finite by Proposition 3.1, and continuous by the estimate above, so $\sup_D M(\tau)$ is finite by compactness of D. $\square$

5 Geodesic curves on Hilbert modular surfaces

In this section we prove Theorem 1.1, which we state in detail as follows.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Theorem 5.1 *Let*

$$V = \mathbb{H}/\Gamma \looparrowright X_K = \mathbb{H}^2/\operatorname{SL}(\mathcal{O} \oplus \mathcal{O}^\vee)$$

be a complex geodesic curve on the Hilbert modular surface associated to a real quadratic field K. Then either V is a Shimura curve, or for every $x \in \mathbb{P}^1(K)$ there exists a parabolic element $g \in \Gamma$ satisfying $g \cdot x = x$ and

$$h(g) = O(1 + h(x)^2). \tag{5.1}$$

Here $h(g)$ is the logarithmic height defined in (2.6).

Remarks. 1. Theorem 5.1 implies the same result with $\operatorname{SL}(\mathcal{O}\oplus\mathcal{O}^\vee)$ replaced by any commensurable subgroup of $\operatorname{SL}_2(K)$, e.g. $\operatorname{SL}_2(\mathcal{O})$ or $\operatorname{SL}_2(\mathfrak{A} \oplus \mathfrak{B})$.
2. When V is a Shimura curve, Γ is arithmetic. Then either V is compact (and Γ has no cusps), or Γ is commensurable to $\operatorname{SL}_2(\mathbb{Z})$, and its cusps are a copy of $\mathbb{P}^1(\mathbb{Q})$.
3. In general, given a number field K and a subgroup Γ of $\operatorname{SL}_2(K)$, we say the cusps of Γ satisfy *quadratic height bounds* if each cusp $x \in \mathbb{P}^1(K)$ is fixed by a parabolic $g \in \Gamma$ satisfying (5.1). This property is inherited under commensurability.
4. We can take g in (5.1) to be the generator of the stabilizer of x in Γ (up to $\pm I$). Note that we also have

$$\log \|g\|_2 = O(1 + h(x)^2),$$

by Eq. (2.7).

Conventions. In this section, $K = \mathbb{Q}(\sqrt{D})$ will denote a real quadratic field with a *distinguished* real embedding. Thus we will regard K as a subfield of $\mathbb{R}$, and $\operatorname{SL}_2(K)$ as a subgroup of $\operatorname{SL}_2(\mathbb{R})$, with Galois involutions $x \mapsto x'$ and $g \mapsto g'$ respectively.

The action of $\operatorname{SL}(\mathcal{O} \oplus \mathcal{O}^\vee)$ on $\mathbb{H}^2$ is given by $g \cdot \tau = (g \cdot \tau_1, g' \cdot \tau_2)$. Note that the map $\iota(\tau) = (\tau_2, \tau_1)$ satisfies $\iota(g \cdot \tau) = g' \cdot \iota(\tau)$, and descends to an involution on X_K.

Geodesic curves. We say an algebraic curve $V = \mathbb{H}/\Gamma \looparrowright X_K$ is a *complex geodesic* if it is isometrically immersed for the Kobayashi metrics on its domain and range.

On $\mathbb{H}^d$ the Kobayashi distance d_K is given in terms of the hyperbolic metric d by

$$d_K(z, w) = \max_i d(z_i, w_i),$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

and its infinitesimal form is the Finsler metric give by $\max |dz_i|/\operatorname{Im}(z_i)$. The Kobayashi metrics on V and X_K are inherited from their universal covers.

By the Schwarz lemma, every holomorphic map $I : \mathbb{H} \to \mathbb{H}^d$ is either a contraction or an isometry for the Kobayashi metric; and in the isometric case, the composition of I with projection to one of the factors of $\mathbb{H}^d$ is an isomorphism. It follows that, up to the action of ι, any geodesic curve $V \looparrowright X_K$ is presented by the following data:

- A lattice $\Gamma \subset \mathrm{SL}(\mathcal{O} \oplus \mathcal{O}^\vee) \subset \mathrm{SL}_2(\mathbb{R})$; and
- A holomorphic map $F : \mathbb{H} \to \mathbb{H}$; such that
- We have

$$F(g \cdot t) = g' \cdot F(t) \tag{5.2}$$

  for all $t \in \mathbb{H}$ and $g \in \Gamma$.

The immersion of $V = \mathbb{H}/\Gamma$ into X_K is then given by $I(t) = (t, F(t))$. Equivalently, V is covered by the graph of F in $\mathbb{H} \times \mathbb{H}$.

Galois contraction. Since F itself is either an isometry or a contraction, Eqs. (5.2), (2.4) and (2.8) imply:

Proposition 5.2 *For any $g \in \Gamma$, we have*

$$\|g'\|_2 = O(\|g\|_2) \quad \textit{and} \quad |\operatorname{Tr}(g')| \le \max(2, |\operatorname{Tr}(g)|).$$

In particular, the traces of hyperbolic elements in Γ lie in the discrete subset of $\mathcal{O} \subset \mathbb{R}$ where $|x'| \le |x|$. The number of such x with $|x| \le R$ grows like R^2.

Shimura curves. We say V is a *Shimura curve* if F is an isometry. In this case V is also isometrically immersed for the symmetric Riemann metric on X_K, and Γ has trace field $\mathbb{Q}$. For more on these much-studied curves, see e.g. [vG, §V] and references therein.

For the remainder of this section, we will assume V is *not* a Shimura curve, and thus F is a contraction: the norm of its derivative in the hyperbolic metric satisfies $\|Df(t)\| < 1$ for all $t \in \mathbb{H}$. Our goals is to prove Theorem 5.1.

Thick-thin decomposition. Since Γ is a lattice, V is a finite volume hyperbolic surface with a finite number of cusps m. Let $V_c \subset V$ be the union of m disjoint closed horoballs B_i, one for each cusp, chosen so the length of ∂B_i is a small, universal constant $\epsilon_0 > 0$. Let $V_{\text{thick}} = \overline{V - V_c}$. Then V_{thick}, the *thick part* of V, is a compact submanifold bounded by m closed horocycles.

Let $\pi : \mathbb{H} \to V = \mathbb{H}/\Gamma$ be the quotient map. By the functional equation (5.2), $\|Df(t)\|$ is Γ-invariant; so by compactness, there exists an $\alpha > 0$ such that

$$\pi(t) \in V_{\text{thick}} \implies \|DF(t)\| < 1 - \alpha. \tag{5.3}$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Choose a compact set $D \subset \mathbb{H}$ such that $\pi(D) = V_{\mathrm{thick}}$. Shrinking V_c if necessary, we can arrange that $i = \sqrt{-1}$ is in the interior of D; it will serve as a convenient basepoint.

Height bounds. Let $\tau(t) = (t, F(t))$ denote the isometric map $\mathbb{H} \to \mathbb{H}^2$ covering the immersion $V \to X_K$. Note that for $g \in \Gamma$ we have

$$\tau(g \cdot t) = g \cdot \tau(t). \tag{5.4}$$

By Propositions 4.1 and 4.3, there exists an $M > 0$ such that

$$\pi(t) \in V_{\mathrm{thick}} \implies -M < \inf_x h_{\tau(t)}(x), \tag{5.5}$$

and

$$\sup_x |h_{\tau(t)}(x) - h(x)| < M \tag{5.6}$$

for all $t \in D$; in particular, for $t = i$.

Descending geodesics. Now fix a point $x \in \mathbb{P}^1(K)$. To show x is a cusp of Γ, we consider the unit-speed geodesic ray $\gamma : [0, \infty) \to \mathbb{H}$ with $\gamma(0) = i$ and $\gamma(s) \to x$ as $s \to \infty$.

We aim to analyze the excursions of $\pi \circ \gamma : [0, \infty) \to V$ into the cusps, and its sojourns in the thick part. These alternating behaviors are described by a sequence of consecutive closed segments,

$$[0, \infty) = T_1 \cup C_1 \cup T_2 \cup C_3 \cup \cdots \tag{5.7}$$

meeting end to end, such that $\bigcup T_i$ and $\bigcup C_i$ form the preimages under $\pi \circ \gamma$ of V_{thick} and V_c respectively. Since ∂V_c is convex, every sojourn in the thick part lasts a definite amount of time: we have $|T_i| > c(\epsilon_0) > 0$ for all i.

Let $T(s) = \sum |T_i \cap [0, s]|$. The following are equivalent:

1. The point x is a cusp of Γ.
2. The total amount of time spent in the thick part, $\sup_s T(s)$, is finite.
3. The decomposition (5.7) terminates with an interval of the form $C_n = [s_n, \infty)$.

In this case we refer to s_n as the *exit time* for the geodesic descending to x.

Evolution of height. To begin the analysis, we will show the amount of time spent in the thick part is controlled by $h(x)$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Proposition 5.3 *For any $s > 0$, we have $h_{\tau(\gamma(s))}(x) \le h(x) + M - \alpha T(s)/4$.*

Proof. Recall that for any $(a, b) \in \mathcal{O} \oplus \mathcal{O}^\vee$ with $a + bx = 0$ we have, by Eq. (4.8),

$$h_{\tau(\gamma(s))}(x) \le \frac{1}{4}\left(\log \frac{|a_1 + b_1\gamma(s)|^2}{\operatorname{Im}\gamma(s)} + \log \frac{|a_2 + b_2 F(\gamma(s))|^2}{\operatorname{Im} F(\gamma(s))}\right)$$
$$= \frac{\eta_1(s) + \eta_2(s)}{4}.$$

Moreover, we can choose (a, b) such that equality nearly holds for $s = 0$, and thus, by the bound (5.6), we have

$$\frac{\eta_1(0) + \eta_2(0)}{4} \le h(x) + M. \tag{5.8}$$

Fixing these values of a and b, we wish to study the behavior of $\eta(s) = \eta_1(s) + \eta_2(s)$. By following the geodesic $\gamma(s)$, we are driving $\eta_1(s)$ to zero as fast as possible; at the same time, the growth of $\eta_2(s)$ is damped by the contraction of F. More precisely, Corollary 4.5 yields:

$$\eta_1'(s) = -1 \quad \text{and} \quad \eta_2'(s) \le \|DF(\gamma(s))\| < 1.$$

These derivative estimates, combined with the bound (5.3) on the size of $\|DF\|$ over the thick part of V, yield

$$\eta(s) \le \eta(0) - \alpha T(s).$$

Combining this inequality with Eq. (5.8), we obtain the stated result. □

Corollary 5.4 *The total amount of time the geodesic $\pi \circ \gamma(s)$ spends in the thick part of V is $O(1 + h(x))$.*

Proof. The lower bound (5.5) on the height, combined with the upper bound in Proposition 5.3, gives $\sup_s T(s) \le (4/\alpha)(h(x) + 2M) = O(h(x) + 1)$. □

Corollary 5.5 *The point x is a cusp of Γ.*

Cuspidal excursions. Next we show the height also controls the cuspidal excursions of $\pi \circ \gamma$, organized by Eq. (5.7).

Proposition 5.6 *We have $|C_i| = O(1 + h(x))$, provided $|C_i|$ is finite.*

Proof. Suppose V has m cusps, represented by points $x_1, \ldots, x_m \in \mathbb{P}^1(K)$. Let $B_j \subset \mathbb{H}$ be a closed horoball resting on x_j, chosen so that $V_c = \pi(\bigcup B_j)$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Recall the bound (5.6) holds for all $t \in D$. Redefine D if necessary so that $\pi(D \cap \bigcup B_j)$ contains ∂V_c.

Let $C_i = [s_i, t_i]$. Since $\pi \circ \gamma(C_i)$ is contained in V_c, there exists a horoball B_j and a $g \in \Gamma$ such that $g \cdot \gamma(s) \in D \cap B_j$. Then $g \cdot \gamma[s_i, \infty)$ is a geodesic ray that begins in the compact set $D \cap \partial B_j$, and converges to $g \cdot x \neq x_j$.

Suppose for simplicity that $x_j = 0$. By the height comparison (5.6), the functional equation (5.4), naturality (4.6), and Proposition 5.3, we have

$$\begin{aligned} h(g \cdot x) - M &\leq h_{\tau(g \cdot \gamma(s_i))}(g \cdot x) \\ &= h_{g \cdot \tau(\gamma(s_i))}(g \cdot x) \\ &= h_{\tau(\gamma(s_i))}(x) \leq h(x) + M. \end{aligned}$$

Lemma 2.1 then controls how close $g \cdot x$ can be to $x_j = 0$: we have

$$\log |g \cdot x| \geq -2h(g \cdot x) \geq -2h(x) - 4M.$$

Now a geodesic segment that descends from height y_1 to y_2 in $\mathbb{H}$ has length $O(1+\log(y_1/y_2))$. For the segment $g \cdot \gamma[s_i, t_i]$ we have $y_1 \asymp 1$ and $y_2 \asymp |g \cdot x|^2$, which gives

$$|C_i| = |t_i - s_i| = O(1 + h(x))$$

as desired. To handle the general case, choose an automorphism f_j of $\mathbb{P}^1(K)$ for each j, such that $f_j(x_j) = 0$, and observe that $h(f_j(x)) = h(x) + O(1)$ by Eq. (2.1). $\square$

Corollary 5.7 *The exit time for the geodesic $\pi \circ \gamma(s)$ is $O(1 + h(x)^2)$, and the number of cuspidal excursions is $n = O(1 + h(x))$.*

Proof. Suppose $C_n = [s_n, \infty)$ is the final cuspidal excursion. Then the exit time is given by $s_n = |T_1| + |C_1| + |T_2| + |C_2| + \cdots + |T_n|$. By Corollary 5.4, we have $\sum_1^n |T_i| = O(1 + h(x))$; and since $|T_i|$ is bounded below, we have $n = O(1 + h(x))$. The preceding result then gives

$$\sum_1^{n-1} |C_i| = O(n(1 + h(x)) = O(1 + h(x)^2),$$

completing the proof. $\square$

Proof of Theorem 5.1. Given $x \in \mathbb{P}^1(K)$, let $\gamma : [0, \infty) \to \mathbb{H}$ be the unique geodesic ray with $\gamma(0) = i$ and $\gamma(s) \to x$, as above. We have seen that the

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

exit time for $\pi \circ \gamma$ is $O(1 + h(x)^2)$. Thus there exists a closed horoball B, resting on x and forming a component of $\pi^{-1}(V_c)$, such that

$$d(i, B) = O(1 + h(x)^2). \tag{5.9}$$

Let $g \in \Gamma$ generate the stabilizer of B in Γ. We wish to control its height, which satisfies

$$2h(g) = \log \|g\|_2 + \log \|g'\|_2 + O(1), \tag{5.10}$$

by Lemma 2.2 and the fact that Γ is conjugate to a subgroup of $\mathrm{SL}_2(\mathcal{O})$. By elementary hyperbolic geometry, the parabolic translation distance $D(t) = d(t, g(t))$ grows linearly as $t \in \mathbb{H}$ moves away from B: that is, we have $D(t) = O(1 + d(t, B))$. Thus Eqs. (2.4) and (5.9) give

$$\log \|g\|_2^2 = d(i, g(i)) = O(d(i, B)) = O(1 + h(x)^2),$$

and a bound of the same form holds for $\log \|g'\|_2$, by Proposition 5.2, Thus (5.10) yields $h(g) = O(1 + h(x)^2)$, as desired. □

6 Triangle groups and Teichmüller curves

In this section we bring Teichmüller curves into play, to connect triangle groups, billiards and curves on Hilbert modular varieties. We then prove Corollaries 1.2, 1.3, and 1.5.

We will begin by showing:

Theorem 6.1 *Let (X, ω) be a holomorphic 1-form of genus g such that $\mathrm{SL}(X, \omega)$ is a lattice with real quadratic trace field K. Assume 0, 1 and ∞ are periodic slopes for (X, ω). Then:*

1. *The cusps of $\mathrm{SL}(X, \omega) \subset \mathrm{SL}_2(K)$ coincide with $\mathbb{P}^1(K)$ and satisfy quadratic height bounds; and*
2. *The periodic slopes s for (X, ω) coincide with $\mathbb{P}^1(K)$, and satisfy $\log L(s) = O(1 + h(x)^2)$.*

Here $L(s)$ is the length of the longest closed geodesic with slope s, and the quadratic height bounds are of the form $h(g) = O(1 + h(x)^2)$ as in Eq. (5.1).

Teichmüller curves. For background material, see e.g. the surveys [Mas, Mo2, Z] and [Mc7], as well as [Mc1] and [Mc6].

Let $\Omega\mathcal{M}_g \to \mathcal{M}_g$ denote the moduli space of nonzero holomorphic 1-forms (X, ω) of genus g. The stabilizer of a given form under the natural action of $\mathrm{SL}_2(\mathbb{R})$ will be denoted by $\mathrm{SL}(X, \omega)$. It is related to the group of real affine

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

automorphism of (X, ω) via the exact sequence

$$1 \to \mathrm{Aut}(X, \omega) \to \mathrm{Aff}^+(X, \omega) \xrightarrow{D} \mathrm{SL}(X, \omega) \to 1.$$

Suppose $\mathrm{SL}(X, \omega)$ is a lattice. Then the orbit of (X, ω) covers an isometrically immersed *Teichmüller curve*

$$V = \mathbb{H}/\,\mathrm{SL}(X, \omega) \looparrowright \mathcal{M}_g.$$

The curve V has at least one cusp, and the trace field $K \subset \mathbb{R}$ of $\mathrm{SL}(X, \omega)$ is totally real (cf. [HL]).

The geodesic flow on the singular flat surface $(X, |\omega|)$ satisfies the Veech dichotomy: every geodesic is either periodic or uniformly distributed. The map $s \mapsto x = -1/s$ gives a bijection between the periodic slopes and the cusps of $\mathrm{SL}(X, \omega)$; in particular, periodic slopes are dense. In the periodic case, the smooth closed geodesics with slope s sweep out a collection of open cylinders

$$C(s) = \{A_1, \ldots, A_n\}, \tag{6.1}$$

whose closures cover X.

Replacing the original form with $h \cdot (X, \omega)$ for some $h \in \mathrm{SL}_2(\mathbb{R})$, one can always normalize so that 0, 1 and ∞ are periodic slopes. We then have $\mathrm{SL}(X, \omega) \subset \mathrm{SL}_2(K)$ (cf. [CSm]).

Modular embeddings and triangle groups. We will use the following two important results about Teichmüller curves.

I. [Mo1, Cor. 2.11] After passing to a finite cover, V can be immersed as a totally geodesic curve in the Hilbert modular variety X_K. Its image is a Shimura curve only when $K = \mathbb{Q}$ (in which case $X_K = \mathbb{H}/\,\mathrm{SL}_2(\mathbb{Z})$).

II. [BM] Every triangle group $\Delta(p, q, \infty)$ is commensurable to the fundamental group $\mathrm{SL}(X, \omega)$ of a Teichmüller curve V.

Remarks on the proof of (I). One begins by showing that $\mathrm{Jac}(X)$ admits a canonical factor A with real multiplication by K, and ω is the pullback of an eigenform on A. The variation of A over V gives a map from V to a suitable moduli space of abelian varieties with real multiplication $Y = \mathbb{H}^d/\Delta$. Although A need not be principally polarized, $\Delta \subset \mathrm{SL}_2(K)$ stabilizes a lattice M commensurable to $\mathcal{O} \oplus \mathcal{O}^\vee$, so after passing to a finite cover we obtain a map $V \to X_K$. This map is a geodesic immersion because the periods of the eigenform ω vary as fast as possible; the other eigenforms vary more slowly, so it is not a Shimura curve.

For $d \geq 3$ the proof that A admits real multiplication rests on rigidity theorems of Deligne and Schmid. The case $d = 2$, which is sufficient for

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Theorem 6.1, can be handled by more elementary means; see [Mc1], and [Mc3].

Proof of Theorem 6.1. Part (1) follows immediately from Theorem 1.1 and (I) above. For part (2), let s be a slope in $\mathbb{P}^1(K)$. Then $x = -1/s$ is a cusp of $\mathrm{SL}(X, \omega)$, whose stabilizer is generated by a parabolic element $g = D\phi$ for some $\phi \in \mathrm{Aff}^+(X, \omega)$. The geodesics at slope s decompose X into cylinders $(A_1, \dots, A_n)$ as in Eq. (6.1). Replacing ϕ with ϕ^n for some $n = O(1)$, we can assume $\phi(A_i) = A_i$ for all i, and ϕ fixes ∂A_i pointwise.

Let (h_i, c_i) be the height and circumference of A_i. Note that $L(s) = \max c_i$. The map $\phi|A_i$ is a power of a Dehn twist, and hence it sends a geodesic segment of length h_i to one of length greater than c_i. It follows that

$$\|g\|_2 \asymp \|g^n\|_2 \geq c_i/h_i = c_i^2/(h_i c_i) \geq c_i^2/\,\mathrm{area}(X, \omega) \asymp c_i^2.$$

By the quadratic height bounds from part (1), we also have

$$\log \|g\|_2 = O(h(g)) = O(1 + h(x)^2) = O(1 + h(s)^2),$$

and hence $\log L(s) = \log \max c_i = O(1 + h(s)^2)$ as desired. □

Proof of Corollary 1.2. Combining Theorem 1.1 with (II) above shows that every triangle group of the form $\Delta(p, q, \infty)$ is conjugate to a group with cusps $\mathbb{P}^1(K)$. □

Let us explicitly state the following complement:

Corollary 6.2 *The cusps of $\Delta(p, q, \infty)$ satisfy quadratic height bounds whenever K_{pq} is a real quadratic field.*

Proof of Corollary 1.3. Let $u = \left(\begin{smallmatrix} 1 & \gamma \\ 0 & 1 \end{smallmatrix}\right) \in \Delta(2, 5, \infty)$, and let $x = a/c$ be the golden fraction for a cusp x of $\Delta(2, 5, \infty)$. Then $h = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ for some b, d, and hence x is fixed by the parabolic element

$$g = huh^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \gamma \begin{pmatrix} -ac & a^2 \\ -c^2 & ac \end{pmatrix} \in \Gamma.$$

We then have

$$h(a) + h(c) = O(1 + h(g)) = O(1 + h(x)^2),$$

by the quadratic height bounds for $\Delta(2, 5, \infty)$.

The golden continued fraction (1.1) for x describes the behavior of the geodesic ray γ descending to x: each term a_i corresponds to a cuspidal excursion of length $O(1 + \log |a_i|)$. Thus N are $\max \log |a_i|$ are both $O(1 + h(x))$ by Proposition 5.6 and Corollary 5.7. □

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Proof of Corollary 1.5. Let P be a lattice polygon with quadratic trace field K, and internal angles $(\pi a_i/q)$ as in Sect. 1. The periodic slopes $S(P)$ for P coincide with those for the 1-form (X, ω) obtained by unfolding P.

By Theorem 6.1, the set of all cross-ratios of 4-tuples in $S(P)$ coincides with $K - \{0, 1\}$. Equivalently, there exists a vector space $A \subset \mathbb{C}$, dense in $\mathbb{C}$ and 2-dimensional over K, such that $S(P)$ coincides with the slopes of vectors in A. Here A is uniquely determined by $S(P)$, up to multiplication by a real scalar.

Now assume P is normalized so one of its sides is horizontal. Since $S(P)$ is invariant under the dihedral group D_{2q} generated by reflections through the sides of P, the same is true for A. Thus A is invariant under $z \mapsto \overline{z}$ and $z \mapsto \zeta_q z$, where $\zeta_q = \exp(2\pi i/q)$. Consequently A contains a real number $a = z + \overline{z} \neq 0$; rescaling by $1/a$, we can assume that $1 \in A$. We then have

$$A = K \oplus K\zeta_q,$$

provided $q > 2$. In particular $S(P)$ contains 0, ∞ and $\alpha = \tan(\pi/q)$, these being the slopes of 1, $\zeta_q - \overline{\zeta}_q$ and $1 + \zeta_q$. Since all cross-ratios of periodic slopes lie in K, it follows that

$$S(P) = \alpha K \cup \{\infty\},$$

as desired. When $q = 2$, the normalization of P implies that $0, 1, \infty \in S(P)$, and hence we can take $\alpha = 1$ in the equation above.

The height bound on $\log L(\alpha s)$ is a consequence of Theorem 6.1. $\square$

We remark that Corollary 1.5 immediately implies:

Corollary 6.3 *Let P be a lattice polygon with quadratic trace field. Then every edge of P has periodic slope.*

Notes and references. Theorem 6.1, without the quadratic height bounds, first appears in [Mc2, Theorem A.1]. Its proof there, which uses interval exchange transformations and the Veech dichotomy, proceeds by contradiction. In [Mc5] we give an effective proof of the Veech dichotomy, using Hodge theory instead of interval exchange maps, and the argument above continues this development. One can also deduce Corollary 1.2 from [Mc2] and [BM].

7 Study of triangle groups via dynamics

In this section we use the connection between Teichmüller curves and triangle groups, made explicit in the case of $\Delta(2, 5, \infty)$, to establish the results on matrix coefficients, ratio sets and limit measures stated as Theorems 1.4 and

1.6. We also present a plot of matrix coefficients and some examples of golden fractions, showing our height bounds are sharp.

Algebraic preliminaries. Let $\gamma = (1+\sqrt{5})/2$, so $\mathcal{O} = \mathbb{Z}[\gamma]$ is the maximal order, of discriminant $D = 5$, in $K = \mathbb{Q}(\gamma) \subset \mathbb{R}$. Define a $\mathbb{Q}$-linear map $\psi : K \to \mathbb{Q}$ by

$$\psi(x + y\gamma) = y,$$

where $x, y \in \mathbb{Q}$. Using the fact that $\mathcal{O}^\vee = D^{-1/2}\mathcal{O}$, we find that

$$[(a,b),(c,d)] = \mathrm{Tr}_{\mathbb{Q}}^{K}\, D^{-1/2} \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \psi(ad - bc) \tag{7.1}$$

gives a unimodular symplectic form on $\mathcal{O}^2$. (This modification of Eq. (4.2) is made to account for the conventional normalization of triangle groups.)

We define a ring homomorphism $I : K = \mathbb{Q} \oplus \mathbb{Q}\gamma \to \mathrm{M}_2(\mathbb{R})$ by

$$I(x + y\gamma) = x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + y \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}. \tag{7.2}$$

The matrix $I(m)$ represents multiplication by m on K with respect to the basis $(1, \gamma)$. Up to scale, $I(m)$ depends only on $\delta(m) = m'/m$; thus we have a continuous homomorphism of multiplicative semigroups J making the diagram

$$\begin{tikzcd} K^\times \arrow[r, "I"] \arrow[d, "\delta"'] & \mathrm{M}_2(\mathbb{R}) \arrow[d] \\ \mathbb{R} \arrow[r, "J"] & M_2(\mathbb{R})/\mathbb{R}^\times \end{tikzcd} \tag{7.3}$$

commute.

Remarks. The target of J, with $[0]$ removed, is homeomorphic to $\mathbb{P}^3$; it consists of matrices of rank ≥ 1, up to scale. The map J is a homeomorphism to its image, which is an affine line $\mathbb{R} \subset \mathbb{P}^1 \subset \mathbb{P}^3$. Every matrix in $J(\mathbb{R})$ is invertible, except for

$$J(0) = \left[\begin{pmatrix} 1 & \gamma \\ \gamma & \gamma^2 \end{pmatrix}\right]. \tag{7.4}$$

With respect to a basis where $I(\gamma) = \begin{pmatrix} \gamma & 0 \\ 0 & \gamma' \end{pmatrix}$, we have $J(x) = [\begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}]$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

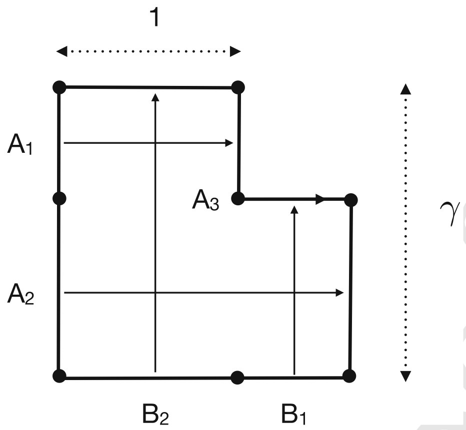

Fig. 4 Core curves of cylinders on the surface $(X, \omega) = (P, dz)/\sim$

The golden surface. Let $P \subset \mathbb{C}$ be the golden L from Sect. 1. With the vertices shown in Fig. 4, P is a combinatorial octagon. By identifying parallel sides, we obtain a holomorphic 1-form $(X, \omega) = (P, dz)/\sim$ of genus two, with a single zero of order two coming from the marked points.

The surface X decomposes into a pair of horizontal cylinders $A = C(0) = \{A_1, A_2\}$ and vertical cylinders $B = C(\infty) = \{B_1, B_2\}$. We associate to each of these cylinders an integral homology class on X, using the orientations shown. Since the intersection matrix $[A_i, B_j] = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ is unimodular, these cycles generate the integral homology of X. It is then readily verified that the period map

$$\pi : H_1(X, \mathbb{Z}) \to \mathcal{O}^2,$$

characterized by $\pi(C) = (a, b) \iff \int_C \omega = a + ib$, is a symplectic isomorphism. For example, we have $\pi(A_1) = (1, 0)$, $\pi(B_2) = (0, \gamma)$, and

$$[A_1, B_2] = 1 = \mathrm{Tr}_{\mathbb{Q}}^K(\gamma/\sqrt{5}).$$

Affine automorphisms. Next we observe that the affine group of (X, ω) can be naturally identified with the $(2, 5, \infty)$ triangle group; that is, we have an isomorphism

$$D : \mathrm{Aff}^+(X, \omega) \cong \mathrm{SL}(X, \omega) = \Delta(2, 5, \infty) \subset \mathrm{SL}_2(\mathcal{O}).$$

The action of first group on $H_1(X, \mathbb{Z})$, and of the last group on $\mathcal{O}^2$, are intertwined by the period map π.

Since $\Delta(2, 5, \infty)$ is a maximal Fuchsian group, to show that D is an isomorphism it suffices to observe that $\mathrm{Aut}(X, \omega)$ is trivial, and that there exist

$\sigma, \tau \in \mathrm{Aff}^+(X, \omega)$ such that

$$D\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad D\tau_A = \begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix}.$$

Indeed, we can take σ to be a 90° rotation coming from the Euclidean symmetry of $P/\sim$, and τ_A to be a product of Dehn twists on A_1 and A_2.

Since $\Delta(2, 5, \infty)$ is a lattice, (X, ω) generates a Teichmüller curve $V = \mathbb{H}/\operatorname{SL}(X, \omega) \to \mathcal{M}_2$. (Indeed, in the terminology of [Mc1], V is the Weierstrass curve W_D, $D = 5$.)

Matrix entries and columns. Let $e_1 = (1, 0)$ and $e_2 = (0, 1)$ in $\mathcal{O}^2$. We remark that the first and second columns of matrices in $\Gamma = \Delta(2, 5, \infty)$ represent the same vectors in $\mathcal{O}^2$; that is, $\Gamma \cdot e_1 = \Gamma \cdot e_2$. Similarly, the set of matrix entries in a given position, $e_i \cdot \Gamma \cdot e_j$, is independent of i and j. These assertions follow from the fact that Γ contains $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and $-I$.

As in Sect. 1, we denote the set of all matrix entries in Γ by $M \subset \mathcal{O} \subset \mathbb{R}$, and let

$$\delta M = \{m'/m \ : \ 0 \neq m \in M\}.$$

As a hint of the multiplicative structure hidden in δM, we observe that

$$\overline{\delta M} \supset -\gamma^{-2} \cdot \delta M. \tag{7.5}$$

To see this, note that

$$\begin{pmatrix} 1 & n\gamma \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + n\gamma c & * \\ * & * \end{pmatrix};$$

thus if $c \in M$, we also have $c_n = a + n\gamma c \in M$ for all $n \in \mathbb{Z}$, and $c_n'/c_n \to -\gamma^{-2}c'/c$. (Here $*$ denotes a matrix element whose value is irrelevant.)

Classification of closed geodesics. Since V has only one cusp, for each periodic slope s there exists a $\phi \in \mathrm{Aff}^+(X, \omega)$ such that

$$C(s) = (\phi(A_1), \phi(A_2)) = (C_1, C_2).$$

When represented in this way, the cylinders C_i inherit orientations from A_i and thus represent homology classes. The geodesics at slope s are of two types: *short* geodesics lying in C_1, and *long* geodesics in C_2. Their lengths differ by a factor of γ. Since

$$\pi(C_1) = D\phi \cdot \pi(A_1) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix},$$

we have:

Proposition 7.1 *A cycle $C \in H_1(X, \mathbb{Z})$ is represented by a short geodesic if and only if its period $\pi(C)$ occurs as the column of a matrix in $\Delta(2, 5, \infty)$. Any long geodesic parallel to C has period $\gamma \cdot \pi(C)$.*

In the terminology of Sect. 1, one can compute the period of a short geodesic with slope s by computing the golden fraction expression of its slope, $s = c/a$.

Corollary 7.2 *There exists a closed geodesic C with period $\pi(C) = (*, c)$ if and only if $c \in M \cup \gamma M$.*

Corollary 7.3 *The entries $m = x + y\gamma$ of matrices in $\Delta(2, 5, \infty)$ satisfy $xy \geq 0$.*

Proof. By the preceding Proposition, there exists an $a \in \mathcal{O}$ and a geodesic C such that $\pi(C) = (a, m)$. Since the closed geodesics A_1 and A_3 shown in Fig. 4 have periods $(1, 0)$ and $(\gamma - 1, 0)$ respectively, Eq. (7.1) gives

$$\langle A_1, C\rangle = \psi(x + y\gamma) = y \quad \text{and} \quad \langle A_3, C\rangle = \psi((\gamma - 1)x + y) = x.$$

But A_1 and A_3 have the same direction, so their intersection numbers with C have the same sign. □

Using the fact that $\gamma' = -1/\gamma$, the previous Corollary easily implies:

Corollary 7.4 *We have $\delta M \subset [-\gamma^{-2}, 1]$.*

The endpoints are achieved when $m = \gamma$ and $m = 1$.

Proof of Theorem 1.6. The set M_0 of probability measures arising from billiard trajectories in P, with slopes tending to zero, can be identified with the set of measures on (X, ω) arising from closed geodesics with slopes tending to zero. By [Mc6, Theorem 1.3], a dense subset of M_0 is given by the probability measures $\widehat{\mu}(C)$ proportional to

$$\mu(C) = \sum_{i=1}^{2} \frac{\langle A_i, C\rangle}{c(A_i)} \chi_{A_i} |\omega|^2,$$

where C ranges over all closed geodesics and $c(A_i)$ denotes the circumference of A_i (given by 1 and γ for $i = 1$ and 2). Thus if $\pi(C) = (*, c)$ and $c = x + y\gamma$, we have

$$\mu(C) = y\alpha_1 + \gamma^{-1}(x + y)\alpha_2$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

by Eq. (7.1). A simple computation also shows that, if we set $r = c'/c$, then the ratio of the coefficients of α_1 and α_2 above satisfies

$$\frac{\gamma^{-1}(x+y)}{y} = \frac{1+\gamma^{-2}r}{1-r}.$$

Thus $\mu(C)$ is proportional to the measure $\nu(r)$ defined by (1.7), and hence $\widehat{\mu}(C) = \widehat{\nu}(r)$.

By Proposition 7.1, there exists a C such that $\pi(C) = (*, c)$ if and only if $c \in M \cup \gamma M$. For $c \in M$, the corresponding values of r range in δM, while for $c \in \gamma M$, they range in $-\gamma^{-2} \cdot \delta M$. Thus $\widehat{\nu}$ provides a homeomorphism

$$\widehat{\nu} : \overline{\delta M \,\cup\, -\gamma^{-2} \cdot \delta M} \to M_0.$$

But the domain of this map is simply the closure of δM, by Eq. (7.5); and thus the image δM is dense in M_0, as desired. □

The last step in the proof says it suffices to consider the measures coming from 'short' closed geodesics. Geometrically, this reflects the fact that the homology class of the short geodesic $\tau_A^n(B_1)$, $n \gg 0$, is close to a multiple of the long geodesic A_2.

Intersection matrices. Let $\phi \in \mathrm{Aff}(X, \omega)$ be an affine automorphism. The intersections between the cylinder systems A and $\phi(A)$, given by the matrix

$$\langle A, \phi(A)\rangle = \begin{pmatrix} \langle A_1, \phi(A_1)\rangle & \langle A_1, \phi(A_2)\rangle \\ \langle A_2, \phi(A_1)\rangle & \langle A_2, \phi(A_2)\rangle \end{pmatrix},$$

are conveniently described using the function (7.2).

Proposition 7.5 *If* $D\phi = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, *then* $\langle A, \phi(A)\rangle = I(c\gamma)$.

Proof. Note that $\pi(\phi(A_1)) = (a, c)$ and $\pi(\phi(A_2)) = \gamma \cdot (a, c)$. Thus if $c = x + y\gamma$ we have, by Eq. (7.1),

$$\langle A, \phi(A)\rangle = \psi \begin{pmatrix} c & c\gamma \\ c\gamma & c\gamma^2 \end{pmatrix} = x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + y \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

□

Proof of Theorem 1.4. Let

$$T = \{[\langle A, \phi(A)\rangle] \; : \; \phi \in \mathrm{Aff}^+(X, \omega)\} \subset M_2(\mathbb{R})/\mathbb{R}^\times.$$

The set T coincides with the set of all possible intersection matrices between pairs of cylinder systems on (X, ω), up to scale. We also note that, by

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

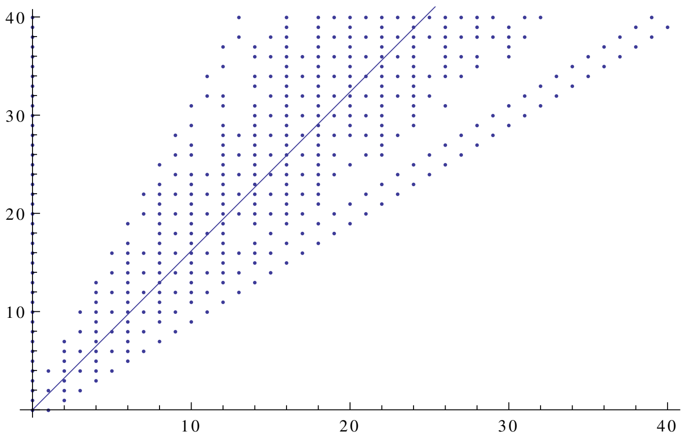

Fig. 5 Matrix entries $(x, y) = x + y\gamma$. The line $y = \gamma x$ is shown

Eq. (7.4), the heights and circumferences of the cylinders (A_1, A_2) satisfy $J(0) = [h(A_i)c(A_j)]$. Thus, by Theorems 1.7 and 1.8 of [Mc6], we have

$$\overline{T} = \langle T\rangle \cup \{J(0)\} \cong \omega^\omega + 1. \tag{7.6}$$

Now let $S = \delta(\gamma M) = -\gamma^{-2} \cdot \delta M$, as in Eq. (1.4). By Proposition 7.5 and the commutative diagram (7.3), we have

$$T = [I(\gamma M)] = J(\delta(\gamma M)) = J(S).$$

Since the map $J : [-1, 1] \to J[-1, 1]$ is an isomorphism of topological semigroups, Eq. (7.6) implies that

$$\overline{S} = \langle S\rangle \cup \{0\} \cong \omega^\omega + 1.$$

$\square$

Survey of matrix entries. Next we present a picture of M. Recall that every matrix entry $m = x + y\gamma$ of the triangle group $\Delta(2, 5, \infty)$ determines a point $(x, y) \in \mathbb{Z}^2$ in the first or third quadrant, by Corollary 7.3.

The matrix entries with $0 \le x, y \le 40$ are plotted in Fig. 5. By Theorem 1.4, if this figure were extended to the whole first quadrant, the slopes of all marked points would form a set P satisfying $\overline{P} \cong \omega^\omega + 1$ and $D^\infty \overline{P} = \{\gamma\}$. The line with slope γ is shown.

Since $\overline{P}$ is a closed set of measure zero, the marked points in Fig. 5 can be covered by finitely many cones with small total angle, leading to the estimate:

$$|\{x + y\gamma \in M \ :\ |x| + |y| \le N\}| = o(N^2).$$

Question. Is δM already closed in $\mathbb{R}^\times$? Equivalently, is $\gamma^{-2}\delta M$ itself a semi-group?

Golden fractions. We conclude the study of $\Delta(2, 5, \infty)$ with some examples of golden fractions.

The golden fraction a/c for an algebraic integer $x \in \mathbb{Z}[\gamma]$ is conveniently described by the unique integer $N = N(x)$ such that $a/c = (\gamma^N x)/\gamma^N$. Note that

$$N(x + n\gamma) = N(x) \tag{7.7}$$

for any $n \in \mathbb{Z}$, since adding a multiple of γ only changes the first term in the golden continued fraction for x.

1. For $k \geq 0$, we have $N(\gamma^{2k}) = k^2 - k + 1$. This shows the exponent 2 in Eq. (1.2) is sharp.
   The proof is by induction on k, using the fact that $n = \gamma^{2k-1} - \gamma^{1-2k} \in \mathbb{Z}$, and hence
   $$\gamma^{2k} = n\gamma + 1/\gamma^{2k-2}.$$
   By similar reasoning we have $N(\gamma^{2k+1}) = k^2$.
2. The Fibonacci sequence $(f_1, f_2, f_3, \ldots) = (1, 1, 2, 3, 5, 8 \ldots)$ satisfies $N(f_{2k}) = k^2$ and $N(f_{2k-1}) = k^2 - k + 1$. In fact $\gamma^n = f_{n-1} + f_n\gamma$, so $N(f_{n-1}) = N(\gamma^n)$ by Eq. (7.7).
   Applying this observation to $f_{20} = 6765$ gives the trajectory length $L(6765) = 2\gamma^{101}\sqrt{f_{20}^2 + 1} \approx 10^{25}$ for billiards in the golden L, as stated in the Introduction.
3. On the other hand, $N(n\gamma) = 0$ for all n. This shows the lower bound $h(x) \leq h(a) + h(c)$ on the complexity of a golden fraction is best possible.

8 Compression of cusps

In this section we prove Theorem 1.7 which, like Theorem 1.1, does not make direct reference to Teichmüller curves. As an application, we give a second proof of Theorem 1.4.

We begin by defining the renormalized distance between a pair of cusps on a hyperbolic surface.

Parabolics. Recall that a matrix $g \in \mathrm{SL}_2(\mathbb{R})$ is parabolic (or unipotent) if and only if $n = g - I$ is nilpotent of rank one. These nilpotent matrices are naturally identified with the open light-cone in the Lie algebra $\mathrm{sl}_2(\mathbb{R})$. A parabolic is *positive* if it is conjugate in $\mathrm{SL}_2(\mathbb{R})$ to $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. The positive parabolics correspond to one component of the light cone.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Using the trace form, we define a pairing on parabolics by

$$\langle g_1, g_2 \rangle = \mathrm{Tr}(g_1 g_2) - 2 = \mathrm{Tr}(n_1 n_2),$$

where $n_i = g_i - I$. Note that

$$\langle g_1^i, g_2^j \rangle = ij \langle g_1, g_2 \rangle \tag{8.1}$$

for all $i, j \in \mathbb{Z}$.

Cusps. Let Γ be a lattice in $\mathrm{SL}_2(\mathbb{R})$. For convenience, assume that $(\pm I) \subset \Gamma$. We say $x \in \partial\mathbb{H}$ is a *cusp of* Γ if it is fixed by a parabolic element $g \in \Gamma$. The set of all cusps will be denoted $\kappa(\Gamma)$.

For each $x \in \kappa(\Gamma)$ there exists a unique positive parabolic $g = \mathrm{gen}(x) \in \Gamma$ such that every other parabolic fixing x is a power of g. Given $r \in \mathbb{R}$ let $B_r(x) \subset \mathbb{H}$ denote the unique horoball resting on x such that the length of its boundary, modulo the action of g, satisfies

$$|\partial B_r(x)/\langle g \rangle| = \exp(r). \tag{8.2}$$

Note that $B_r(x)$ and $B_s(x)$ are bounded by parallel horocycles, distance $|r - s|$ apart. One can think of $B_r(x)$ as the ball of *renormalized* radius r centered at x; we emphasize that r can be negative.

Let

$$\pi : \mathbb{H} \to V = \mathbb{H}/\Gamma$$

denote the quotient map. By regarding $B_r(x) \cup \{x\}$ as a neighborhood of x, we obtain a natural topology on the extended upper halfplane such that the quotient

$$\mathbb{H} \cup \kappa(\Gamma)/\Gamma = \overline{V}$$

is a compact Riemann surface. The added points $\kappa(V) = \overline{V} - V$ are the *cusps of* V.

Renormalized distance. We define the *renormalized distance* between a pair of distinct cusps x, y by

$$D(x, y) = \lim_{r \to -\infty} d(B_r(x), B_r(y)) + 2r. \tag{8.3}$$

Let $\gamma \subset \mathbb{H}$ denote the hyperbolic geodesic joining x to y. Geometrically, $D(x, y)$ measures the length of the geodesic $\pi(\gamma) \subset V$, truncated at standard horoball neighborhoods of the cusps of V.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Proposition 8.1 *We have* $\exp(D(x_1, x_2)) = -\langle g_1, g_2\rangle$, *where* $g_i = \mathrm{gen}(x_i)$.

Proof. Normalizing so $(x_1, x_2) = (\infty, 0)$, we can write

$$g_1 = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad g_2 = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix}$$

with $b, c > 0$. Then $B_0(x_1)$ is the region $\mathrm{Im}(t) \geq b$, $B_0(x_2)$ is a Euclidean ball of radius $1/c$ resting on $t = 0$, and—provided these balls are disjoint—we have

$$\exp(D(x, y)) = \exp \int_{1/c}^{b} du/u = bc = 2 - \mathrm{Tr}(g_1 g_2) = -\langle g_1, g_2\rangle.$$

To treat the general case, use $B_r(x_i)$ with r small enough to insure disjointness. $\square$

Indiscrete groups. The discussion above also goes through for an arbitrary group $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$, subject only to the condition that *its parabolic subgroups are cyclic*. In particular, Γ need not be discrete. Although the quotients spaces V and $\overline{V}$ cannot be defined in this setting, Eq. (8.3) still gives a geometric meaning to $D(x, y)$ in terms of the universal cover $\mathbb{H}$.

Contracting twists. We now return to the setting of Theorem 1.7. Let $\rho : \Gamma \to \Gamma' \subset \mathrm{SL}_2(\mathbb{R})$ and $F : \mathbb{H} \to \mathbb{H}$ define a contractive twist of Γ. For brevity, we will write $g' = \rho(g)$; thus $F(g \cdot t) = g' \cdot F(t)$. Since the parabolic subgroups of Γ are cyclic, the same is true for Γ', by condition (1.8).

Let x be a cusp of Γ and let $g = \mathrm{gen}(x)$. Then there is a unique cusp x' of Γ' fixed by g'. Below, functions of primed variables such as $\mathrm{gen}(x')$ and $B_r(x')$ depend implicitly on Γ'; those of unprimed variables depend on Γ.

Proposition 8.2 *The parabolic* $g' = \mathrm{gen}(x)'$ *is positive, and* $F(B_r(x)) \subset B_r(x')$ *for all* $r \in \mathbb{R}$.

Proof. The map F descends to a holomorphic map between punctured disks,

$$f : \Delta^* \cong \mathbb{H}/\langle g\rangle \to \mathbb{H}/\langle g'\rangle \cong \Delta^*$$

inducing the identity on π_1. Completing f across the origin and applying the Schwarz lemma yields the result above. $\square$

Corollary 8.3 *We have* $\mathrm{gen}(x') = \mathrm{gen}(x)'$.

Compression. Consistent with Eq. (1.10), given $x, y \in \kappa(\Gamma)$ with $(g, h) = (\mathrm{gen}(x), \mathrm{gen}(y))$, we now define

$$r(x, y) = \frac{\langle g', h'\rangle}{\langle g, h\rangle}.$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

By Eq. (8.1), this quantity does not change if we replace g and h with a different pair of parabolics in Γ fixing x and y. By Proposition 8.1 and Corollary 8.3, we have

$$r(x, y) = \exp(D(x', y') - D(x, y)). \tag{8.4}$$

This expression makes it clear that $r(x, y)$ measures the rate at which F pulls the cusps x and y closer together. A quantitative estimate for this compression is given by:

Proposition 8.4 *Let $\gamma : \mathbb{R} \to \mathbb{H}$ be a unit speed hyperbolic geodesic running from x to y. Then*

$$D(x, y) - D(x', y') \geq \int_{-\infty}^{\infty} 1 - \|F'(\gamma(s))\| \, ds.$$

Here $\|F'\| < 1$ is the norm of the derivative of F in the hyperbolic metric.

Proof. Choose $r > 0$ small enough that equality holds in Eq. (8.3) for $D(x, y)$ and $D(x', y')$. Let $s_1 < s_2$ be the moments when $\gamma(s)$ leaves and enters $B_r(x)$ and $B_r(y)$ respectively. Then by Proposition 8.2, $F \circ \gamma : [s_1, s_2] \to \mathbb{H}$ provides a path connecting $B_r(x')$ to $B_r(y')$. It follows that $D(x, y) + 2r = s_2 - s_1$ and

$$D(x', y') + 2r = d(B_r(x'), B_r(y')) \leq \int_{s^1}^{s_2} \|F'(\gamma(s)\| \, ds.$$

Taking the difference and letting $r \to 0$ yields the bound above. □

Let V_{thick} denote the thick part of V, as in Sect. 5. Since $\|F'(t)\|$ depends only on $\pi(t)$, we have

$$\delta = \inf_{\pi(t) \in V_{\text{thick}}} 1 - \|F'(t)\| > 0.$$

Corollary 8.5 *Let T be the total amount of time the geodesic $\pi \circ \gamma$ spends in the thick part of V. Then $r(x, y) \leq \exp(-\delta T)$.*

Proof. Immediate from Proposition 8.4 and Eq. (8.4). □

Modular symbols. In preparation for the proof of Theorem 1.7, we briefly review some material from [Mc6, §2].

A *modular symbol of degree d* is a formal product

$$\sigma = \gamma_1 * \cdots * \gamma_n$$

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

of oriented geodesics on V, joining consecutive cusps $(c_0, \ldots, c_n)$ to form a path from c_0 to c_n on $\overline{V}$. We let $\mathcal{S}^d(V)$ denote the modular symbols of degree d, and $\mathcal{S}(V) = \bigcup_0^\infty \mathcal{S}^d(V)$. The modular symbols form the morphisms in a graded category whose objects are the cusps of V.

The *geometric topology* on $\mathcal{S}(V)$ is defined so that $\sigma_n \to \tau$ if the corresponding paths converge uniformly, when suitably parameterized. Assuming V has at least one cusp, we have $\mathcal{S}(V) \cong \omega^\omega$; $\mathcal{S}^1(V)$ is dense among modular symbols of degree $d \geq 1$; and $\sigma_n \to \infty$ in $\mathcal{S}(V)$ if and only if

$$L(\sigma_n \cap V_{\text{thick}}) \to \infty.$$

We can also identify $\mathcal{S}^1(V)$ with the set of ordered pairs of distinct cusps of Γ, subject to the equivalence relation $[x, y] \sim [gx, gy]$ for all $g \in \Gamma$. The corresponding geodesic γ on V has a lift to $\mathbb{H}$ which runs from x to y. In the geometric topology, if x, y, z are distinct cusps and $g = \text{gen}(y)$, then

$$[x, g^n(z)] \to [x, y] * [y, z] \tag{8.5}$$

as $n \to \infty$.

Continuity of compression. By invariance of the trace, we have $r(gx, gy) = r(x, y)$ for all $g \in \Gamma$. Thus we can regard r as a function on $\mathcal{S}^1(V)$; $r(\gamma)$ measures the compression of γ under F. Extending the definition to all modular symbols by

$$r(\gamma_1 * \gamma_2 * \cdots * \gamma_n) = r(\gamma_1) r(\gamma_2) \cdots r(\gamma_n),$$

we obtain a *functor*

$$r : \mathcal{S}(V) \to [0, 1];$$

this means simply that $r(\sigma * \tau) = r(\sigma) r(\tau)$.

Proposition 8.6 *The functor r is continuous, and $r(\sigma_n) \to 0$ if $\sigma_n \to \infty$.*

Proof. We begin by proving continuity in the representative case described by Eq. (8.5). Let x, y, z be distinct cusps of Γ, and let $g_x = \text{gen}(x)$, $g_y = \text{gen}(y)$ and $g_z = \text{gen}(z)$. Normalize coordinates so that $x = x' = \infty$, $y = y' = 0$,

$$g_x = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \quad g_y = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}, \quad g_z = \begin{pmatrix} * & c \\ * & * \end{pmatrix},$$

and the same equations hold with primes on all variables. Then $\langle g_x, g_y \rangle = ab$ and $\langle g_y, g_z \rangle = bc$, and hence $r(x, y) = (a'b')/(ab)$ and $r(y, z) = (b'c')/(bc)$.

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Since $\mathrm{gen}(g_y^n z) = g_y^n g_z g_y^{-n}$, and

$$\langle g_x, g_y^n g_z g_y^{-n} \rangle = -n^2(ab)(bc) + O(n),$$

we similarly find that $r(x, g_y^n z) \to (a'b')(b'c')/((ab)(bc)) = r(x, y)r(y, z)$, as required for continuity under Eq. (8.5). The general case is similar. (Cf. [Mc6, Thm 4.1].)

Finally observe that if $\sigma_n \to \infty$ in $\mathcal{S}(V)$ then $L(\sigma_n \cap V_{\mathrm{thick}}) \to \infty$ and hence $r(\sigma_n) \to 0$ by Corollary 8.5. □

Recall from Eq. (1.11) that the absolute ratio set is given by $R = r(\mathcal{S}^1(V))$.

Proposition 8.7 *The ratio set R contains an accumulation point $p \neq 0$.*

Proof. Let x and y be two distinct cusps of Γ, normalized as in the preceding proof. Let $y_n = g_x^n(y)$ and $z = g_y(x)$. It is readily verified that

$$\langle g_x^n g_y g_x^{-n}, g_y g_x g_y^{-1} \rangle = ab(abn - 1)^2,$$

and hence $r(y_n, z) \to p = (a'b')^3/(ab)^3 \neq 0$. Using the fact that $r(x, y) = (a'b')/(ab) < 1$, we also find that $r(y_n, z)$ takes on infinitely many values as n varies, and hence p is an accumulation point of R (cf. [Mc6, Cor. 6.2]). □

Proof of Theorem 1.7. Let $S = \mathcal{S}(V) \cup \{\infty\}$ denote the 1-point compactification of $\mathcal{S}(V)$. By Proposition 8.6, if we set $r(\infty) = 0$ we obtain a continuous map $r : S \to [0, 1]$. Since

$$\langle \mathcal{S}^1(V) \rangle \cup \{\infty\} = \overline{\mathcal{S}^1(V)}$$

in S, and r is a functor, the set $R = r(\mathcal{S}^1(V))$ similarly satisfies

$$\langle R \rangle \cup \{0\} = \overline{R}$$

in $[0, 1]$; and since $D^\infty S = \{\infty\}$, we have $D^\infty R \subset \{0\}$. To see equality holds, observe that there exists a point $0 < p < 1$ in $D\overline{R}$ by Proposition 8.7; then $p^n \in D^n\overline{R}$, and hence $0 = \lim p^n \in D^\infty\overline{R}$. Thus $D^\infty\overline{R} = \{0\}$ is a single point, and hence the compact set $\overline{R}$ itself is homeomorphic to $\omega^\omega + 1$. □

Remarks. The analysis above can be extended to the case where $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ is any finitely generated discrete group, and $\rho : \Gamma \to \mathrm{SL}_2(\mathbb{R})$ is any homomorphism compatible with a contracting map F. The essential point is not that V has finite volume, but rather that the thick part of the *convex core* of V is compact. In the more general case, $r(x, y)$ can assume the value zero (e.g. when ρ maps a parabolic element to an elliptic element). Examples where ρ is not injective arise naturally for square-tiled surfaces; see e.g. [Mc6, §9].

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

Second proof of Theorem 1.4. To conclude, we return to the triangle group $\Gamma = \Delta(2, 5, \infty)$. The group Γ admits a contractive twist (ρ, F) with $\rho(g) = g'$ given by Galois conjugation. (See [Mc1, §10]; with the current normalizations, F is antiholomorphic, but this causes no difficulties.) The map ρ preserves parabolics, since they are characterized by the rational condition $\mathrm{Tr}(g) = \pm 2$.

Let $p = \left(\begin{smallmatrix} 1 & \gamma \\ 0 & 1 \end{smallmatrix}\right) = \mathrm{gen}(\infty)$, and let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. As we have seen in Sect. 7, every matrix entry for Γ occurs in every possible position, so it suffices to consider the values of c which occur for $g \in \Gamma$.

We readily compute that

$$r(\infty, g(\infty)) = r(\infty, a/c) = \frac{\langle p', (gpg^{-1})' \rangle}{\langle p, gpg^{-1} \rangle} = \left(\frac{\gamma' c'}{\gamma c}\right)^2.$$

Since Γ has only one cusp, every element of its absolute ratio set R arises as above. Referring to Eq. (1.4), we see that the absolute and signed ratio sets for Γ are related $R = S^2$. Since $\overline{R} \cong \omega^\omega + 1$, it follows easily that $\overline{S} \cong \omega^\omega + 1$ as well.

The statement $\overline{S} = \langle S \rangle \cup \{0\}$ is more subtle. It can be handled by defining a new functor $s : \mathcal{S}(V) \to [-1, 1]$ satisfying

$$s(\infty, g(\infty)) = (\gamma' c')/(\gamma c).$$

This distinguished square root of r is still continuous, and the proof of Theorem 1.4 for S can then be completed along the same lines as the proof of Theorem 1.7 for R. □

References

[AS] Arnoux, P., Schmidt, T.A.: Veech surfaces with non-periodic directions in the trace field. J. Mod. Dyn. **3**, 611–629 (2009)

[Be] Berg, F.: Dreiecksgruppen mit Spitzen in quadratischen Zahlkörpern. Abh. Math. Sem. Univ. Hamburg **55**, 191–200 (1985)

[BL] Birkenhake, C., Lange, H.: Complex Abelian Varieties. Springer, Berlin (1992)

[BG] Bombieri, E., Gubler, W.: Heights in Diophantine Geometry. Cambridge University Press, Cambridge (2006)

[Bh] Borho, W.: Kettenbrüche im Galoisfeld. Abh. Math. Sem. Univ. Hamburg **39**, 76–82 (1973)

[BR] Borho, W., Rosenberger, G.: Eine Bemerkung zur Hecke-Gruppe $G(\lambda$. Abh. Math. Sem. Univ. Hamburg **39**, 83–87 (1973)

[Bo] Boshernitzan, M.D.: Rank two interval exchange transformations. Ergod. Theory Dyn. Syst. **8**, 379–394 (1988)

[BM] Bouw, I.I., Möller, M.: Teichmüller curves, triangle groups and Lyapunov exponents. Ann. Math. **172**, 139–185 (2010)

[CSc] Calta, K., Schmidt, T.A.: Infinitely many lattice surfaces with special pseudo-Anosov maps. J. Mod. Dyn. **7**, 1–16 (2013)

[CSm] Calta, K., Smillie, J.: Algebraically periodic translation surfaces. J. Mod. Dyn. **2**, 209–248 (2008)

Journal: **222** Article No.: **1101** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2022/4/1** Pages: **43** Layout: **Small-X**

[CW] Cohen, P., Wolfart, J.: Modular embeddings for some non-arithmetic Fuchsian groups. Acta Arith. **56**, 93–110 (1990)
[DL] Davis, D., Lelièvre, S.: Periodic paths on the pentagon, double pentagon and golden L. Preprint (2018)
[EFW] Eskin, A., Filip, S., Wright, A.: The algebraic hull of the Kontsevich–Zorich cocycle. Ann. Math. **188**, 281–313 (2018)
[EMMW] Eskin, A., McMullen, C., Mukamel, R., Wright, A.: Billiards, quadrilaterals and moduli spaces. J. Am. Math. Soc. **33**, 1039–1086 (2020)
[vG] van der Geer, G.: Hilbert Modular Surfaces. Springer, Berlin (1987)
[He] Hecke, E.: Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichungen. Math. Ann. **112**, 664–699 (1936)
[HS] Hindry, M., Silverman, J.: Diophantine Geometry: An Introduction. Springer, Berlin (2000)
[HL] Hubert, P., Lanneau, E.: Veech groups without parabolic elements. Duke Math. J. **133**, 335–346 (2006)
[La] Lang, S.: Fundamentals of Diophantine Geometry. Springer, Berlin (1983)
[Le] Leutbecher, A.: Über die Heckeschen Gruppen $G(\lambda)$. II. Math. Ann. **211**, 63–86 (1974)
[MR] Maclachlan, C., Reid, A.W.: The Arithmetic of Hyperbolic 3-Manifolds. Springer, Berlin (2003)
[Mag] Magnus, W.: Noneuclidean Tesselations and Their Groups. Elsevier, Amsterdam (1974)
[Mas] Masur, H.: Ergodic theory of translation surfaces. In: Handbook of Dynamical Systems, vol. 18, pp. 527–547. Elsevier, Amstredam (2006)
[Mc1] McMullen, C.: Billiards and Teichmüller curves on Hilbert modular surfaces. J. Am. Math. Soc. **16**, 857–885 (2003)
[Mc2] McMullen, C.: Teichmüller geodesics of infinite complexity. Acta Math. **191**, 191–223 (2003)
[Mc3] McMullen, C.: Prym varieties and Teichmüller curves. Duke Math. J. **133**, 569–590 (2006)
[Mc4] McMullen, C.: Foliations of Hilbert modular surfaces. Am. J. Math. **129**, 183–215 (2007)
[Mc5] McMullen, C.: Teichmüller dynamics and unique ergodicity via currents and Hodge theory. J. Reine Angew. Math. **768**, 39–54 (2020)
[Mc6] McMullen, C.: Modular symbols for Teichmüller curves. J. Reine Angew. Math **777**, 89–125 (2021)
[Mc7] McMullen, C.: Billiards and Teichmüller curves. Preprint (2021)
[Mo1] Möller, M.: Variations of Hodge structures of a Teichmüller curve. J. Am. Math. Soc. **19**, 327–344 (2006)
[Mo2] Möller, M.: Affine groups of flat surfaces. In: Papadopoulos, A. (ed.) Handbook of Teichmüller Theory, vol. II, pp. 369–387. European Mathematical Society, Zurich (2009)
[MV] Möller, M., Viehweg, E.: Kobayashi geodesics in $\mathcal{A}_g$. J. Differ. Geom. **86**, 355–379 (2010)
[Pa] Panti, Giovanni: Decreasing height along continued fractions. Ergod. Theory Dyn. Syst. **40**, 763–788 (2020)
[Ri] Ricker, S.: Symmetric Fuchsian quadrilateral groups and modular embeddings. Q. J. Math. **53**, 75–86 (2002)
[SG] Sansone, G., Gerretsen, J.: Lectures on the Theory of Functions of a Complex Variable. P. Noordhoff, Groningen (1960)
[Sch] Schwarz, H.A.: Über diejenigen Fälle, in welchen die Gaußische hypergeomtrische Reihe eine algebraische Funktion ihres vierten Elements darstellt. J. Reine Angew. Math. **75**, 292–335 (1873)
[Se] Seibold, F.: Zahlentheoretische Eigenschaften der Heckeschen Gruppen $G(\lambda)$ und verwandter Transformationsgruppen. Dissertation, Technischen Universität München (1985)
[Tak] Takeuchi, K.: Arithmetic triangle groups. J. Math. Soc. Jpn. **29**, 91–106 (1977)

Journal: 222 Article No.: 1101 TYPESET DISK LE CP Disp.:2022/4/1 Pages: 43 Layout: Small-X

[V] Veech, W.: Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. Invent. Math. **97**, 553–583 (1989)
[We] Weiss, C.: Twisted Teichmüller Curves. Lecture Notes in Mathematics, vol. 204. Springer, Berlin (2014)
[Wo] Wolfart, J.: Eine Bemerkung über Heckes Modulgruppen. Arch. Math. (Basel) **29**, 72–77 (1977)
[Z] Zorich, A.: Flat surfaces. In: Frontiers in Number Theory, Physics, and Geometry. I, pp. 437–583. Springer, Berlin (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal: **222** Article No.: **1101** TYPESET DISK LE CP Disp.:**2022/4/1** Pages: **43** Layout: **Small-X**