Ergod. Th. & Dynam. Sys. (2020), **40**, 763–788 doi:10.1017/etds.2018.55
© *Cambridge University Press, 2018*

Decreasing height along continued fractions

GIOVANNI PANTI

Department of Mathematics, University of Udine, via delle Scienze 206, 33100 Udine, Italy
(e-mail: giovanni.panti@uniud.it)

(*Received* 7 *March* 2018 *and accepted in revised form* 20 *June* 2018)

Abstract. The fact that the euclidean algorithm eventually terminates is pervasive in mathematics. In the language of continued fractions, it can be stated by saying that the orbits of rational points under the Gauss map $x \mapsto x^{-1} - \lfloor x^{-1} \rfloor$ eventually reach zero. Analogues of this fact for Gauss maps defined over quadratic number fields have relevance in the theory of flows on translation surfaces, and have been established via powerful machinery, ultimately relying on the Veech dichotomy. In this paper, for each commensurability class of non-cocompact triangle groups of quadratic invariant trace field, we construct a Gauss map whose defining matrices generate a group in the class; we then provide a direct and self-contained proof of termination. As a byproduct, we provide a new proof of the fact that non-cocompact triangle groups of quadratic invariant trace field have the projective line over that field as the set of cross-ratios of cusps. Our proof is based on an analysis of the action of non-negative matrices with quadratic integer entries on the Weil height of points. As a consequence of the analysis, we show that long symbolic sequences in the alphabet of our maps can be effectively split into blocks of predetermined shape having the property that the height of points which obey the sequence and belong to the base field decreases strictly at each block end. Since the height cannot decrease infinitely, the termination property follows.

1. *Introduction*

The ordinary continued fractions algorithm and its variants (ceiling fractions, centered, odd, even, Farey, etc; see [**3**]) can be seen as the action of a Gauss map on an interval in $\mathrm{P}^1\,\mathbb{R}$. Here, by a *Gauss map* we mean a piecewise-projective map T, all of whose pieces have the form $T(x) = (ax + b)/(cx + d)$ for some matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ in the extended modular group $\mathrm{PSL}_2^{\pm}\,\mathbb{Z}$, *extended* referring to the fact that matrices of determinant -1, as well as $+1$, are allowed. As remarked in [**1**, p. 36], this setting has two natural generalizations, obtained by replacing $\mathrm{PSL}_2^{\pm}\,\mathbb{Z}$ either with $\mathrm{PSL}_n^{\pm}\,\mathbb{Z}$ (*multidimensional continued fractions*; [**5**, **22**]), or with the *extended Hilbert modular group* $\mathrm{PSL}_2^{\pm}\,\mathbb{O}$ for some totally real ring of algebraic integers $\mathbb{O}$; in this paper we are concerned with the latter.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

Given a continued fractions algorithm, which we identify with the corresponding map T, we define $\Gamma_T^{\pm}$ as the subgroup of $\mathrm{PSL}_2^{\pm}\,\mathbb{R}$ generated by the matrices defining T. Algorithms generating the same group—or, more generally, commensurable groups—should be considered close cousins. As maps conjugate by a projective transformation can be identified, commensurability is to be intended in the wide sense: two subgroups Γ, Γ' of $\mathrm{PSL}_2^{\pm}\,\mathbb{R}$ are *commensurable* if there exists $C \in \mathrm{PSL}_2^{\pm}\,\mathbb{R}$ such that $C\Gamma C^{-1} \cap \Gamma'$ has finite index both in $C\Gamma C^{-1}$ and in Γ'. Since a certain degree of rigidity is required, we have restrictions on the allowed $\Gamma_T^{\pm}$. One such restriction is the requirement that any unimodular interval (see §2 for definitions) be mapped to any other in at least one and at most two ways, one of them being order-preserving and the other order-reversing. Other restrictions arise—although we do not actually use flows in the present work—from the desire to preserve the classical connection between continued fractions and flows on the modular surface [**24**], [**9**, Ch. 9]. We then make the standing assumption that $\Gamma_T^{\pm}$ acts on the hyperbolic plane $\mathcal{H}$ in a properly discontinuous way, and that the resulting fuchsian group $\Gamma_T = \Gamma_T^{\pm} \cap \mathrm{PSL}_2\,\mathbb{R}$ be of the first kind. We will see in Theorem 5.1 that this assumption yields, as a side effect, that points are uniquely defined by their symbolic orbits.

When Γ_T has invariant trace field $\mathbb{Q}$ (see §2 for definitions), it is commensurable with $\mathrm{PSL}_2\,\mathbb{Z}$; this is the arithmetic case, which is well understood. In this paper we take advantage of the recent classification [**19**] of hyperbolic triangle groups in terms of their arithmetic dimension (namely, the number of split real places of the associated quaternion algebra) to introduce continued fractions with good properties for every commensurability class of non-cocompact triangle groups of arithmetic dimension 2. In §2, after a brief review of basic concepts and notation, we apply Margulis's rigidity theorem to refine the classification in [**19**] to a classification into commensurability classes. It turns out that the set of 16 non-cocompact triangle groups of arithmetic dimension 2 is split into four subsets according to the invariant trace field $F_2 \in \{\mathbb{Q}(\sqrt{2}), \mathbb{Q}(\sqrt{3}), \mathbb{Q}(\sqrt{5}), \mathbb{Q}(\sqrt{6})\}$, and further split into nine commensurability classes. In §3 we introduce a continued fractions algorithm for each of these classes. Our algorithms extend previous work; the algorithm for the class containing the Hecke group $\Delta(2, 8, \infty)$ coincides with the octagon Farey map in [**26**, **27**], which in turn is a 'folded version' of the map in [**1**]. In [**8**] continued fractions—different from ours—generating the groups $\Delta(3, m, \infty)$ (which are related to the Veech surfaces in [**31**], [**7**, §6.1]) are introduced; these triangle groups are 2-arithmetic precisely for $m = 4, 5, 6$.

Let T be one of our algorithms. The following is an informal version of the good property referred to above; we will formulate a precise version in Theorem 5.3.

(H) All entries of all matrices defining T are in the integer ring of the invariant trace field F_2 of the commensurability class of Γ_T. The T-symbolic orbit of every $\alpha \in F_2$ can be split into finite blocks of predetermined shape having the property that the height of the points in the T-orbit of α decreases strictly at each block end. This block-splitting process stops only when the symbolic orbit of a parabolic fixed point is reached.

Since by the Northcott theorem [**6**, Theorem 1.6.8] the height of points cannot decrease infinitely, property (H) implies that the T-orbit of any point of F_2 must end up in a parabolic fixed point. This fact is far from granted; in [**2**, Theorem 1] it is proved that

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

it cannot hold for the Hecke groups $\Delta(2, m, \infty)$ unless the degree of the invariant trace field is 1 or 2 (see also the beginning of §5 of the present paper), and it is conjectured that the same degree restriction holds for every Veech group. Note that a finitely generated non-elementary fuchsian group can be realized as a Veech group only if its trace and invariant trace fields coincide [**12**, Lemma 10.1].

Property (H) can be seen from various points of view: geometrically it implies that all points of F_2 are cusps of Γ_T (see Theorem 5.4), while dynamically it means that all directions of vanishing Sah–Arnoux–Fathi (SAF) invariant are parabolic (see [**2**] for definitions and references). These facts are known in degree 2, but the proofs are involved. The geometric fact was proved in [**15**, Satz 2] for the specific case of the Hecke groups $\Delta(2, m, \infty)$ ($m = 5, 8, 10, 12$), and the dynamic fact in [**18**, Appendix], making crucial use of the Veech dichotomy [**30**]; this is indeed the approach used in [**27**, Theorem 2.3.3]. Our proof is direct, algorithmic and self-contained, making no appeal to the Veech dichotomy.

In §4 we briefly review the basics of the Weil height. Then, in Theorem 4.3, we characterize the matrices $A \in \mathrm{PSL}_2^{\pm}\,\mathbb{O}$ ($\mathbb{O}$ being the integer ring of a real quadratic field K) which increase strictly the height of all points of K contained in a given interval. In §5 we introduce the notion of decreasing block, and use it to formulate property (H) in a precise way. In the final Theorem 5.3 we prove that all the algorithms introduced in §3 indeed satisfy property (H).

2. *Triangle groups and commensurability*

A matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{PSL}_2^{\pm}\,\mathbb{R}$ acts on $z \in \mathrm{P}^1\,\mathbb{C}$ in the standard way: $A * z$ equals $(az + b)/(cz + d)$ if $\det(A) = 1$, and equals $(a\bar{z} + b)/(c\bar{z} + d)$ if $\det(A) = -1$. We use square brackets for matrices to emphasize that we are actually considering their classes in the projective group. The above action preserves the upper half plane $\mathcal{H}$ and its boundary $\mathrm{P}^1\,\mathbb{R}$.

The *trace field* of a fuchsian group Γ is the field $F_1 = \mathbb{Q}(\mathrm{tr}[\Gamma])$ generated over the rationals by the traces of the elements of Γ. An element $A \in \Gamma$ is *parabolic* if its trace has absolute value 2; every $\alpha \in \mathrm{P}^1\,\mathbb{R}$ fixed by a parabolic element of Γ is a *cusp* of Γ. Two fuchsian groups intersecting in a subgroup of finite index in both have the same set of cusps (of course, the two groups may partition this common set of cusps in group orbits in different ways). Fuchsian groups of finite covolume are automatically of the first kind, geometrically finite and finitely generated; see [**13**, Ch. 4] for definitions and proofs.

Notation 2.1. We let K be a totally real number field of degree $d \geq 1$, with ring of algebraic integers $\mathbb{O}$. We denote by $\Gamma^{\pm}$ any subgroup of the extended Hilbert modular group $\mathrm{PSL}_2^{\pm}\,\mathbb{O}$ such that:

(i) $\Gamma^{\pm}$ acts on $\mathcal{H}$ in a properly discontinuous way (i.e., for each point α and each compact C the set of elements of $\Gamma^{\pm}$ mapping α to C is finite);
(ii) the fuchsian group $\Gamma = \Gamma^{\pm} \cap \mathrm{PSL}_2\,\mathbb{R}$ is of finite covolume and has trace field K;
(iii) $0, 1, \infty$ are cusps of Γ.

A $\Gamma^{\pm}$-*unimodular interval* is an interval in $\mathrm{P}^1\,\mathbb{R}$ which is the image $A * I_0$ of the base unimodular interval $I_0 = [0, \infty]$ under some element A of $\Gamma^{\pm}$. Intervals are ordered

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

'counterclockwise' with respect to the natural cyclic order of $\mathrm{P}^1\,\mathbb{R}$ (the one obtained by looking at $\mathrm{P}^1\,\mathbb{R}$ as the boundary of the Poincaré disk). A $\Gamma^{\pm}$-*unimodular partition* of I_0 is a finite family $I_1, \ldots, I_r$ (with $r \geq 2$) of unimodular intervals such that $I_0 = \bigcup_{1\leq a\leq r} I_a$, distinct I_a intersect at most in a common endpoint, and the set of endpoints other than ∞ generates K as a vector space over $\mathbb{Q}$. Unimodular partitions are always numbered in accordance with the above order, so that $0 \in I_1$, $\infty \in I_r$, and I_a shares a common endpoint with I_{a+1}, for $0 \leq a < r$. Whenever $\Gamma^{\pm}$ is clear from the context, we say just unimodular interval and unimodular partition.

LEMMA 2.2.

(1) *The only element of* $\Gamma^{\pm}$ *that fixes* I_0 *setwise and fixes* $\{0, \infty\}$ *pointwise is the identity.* $\Gamma^{\pm}$ *may or may not contain an element that fixes* I_0 *setwise and exchanges* 0 *with* ∞*; if it does, then the element is unique and has the form* $\left[\begin{smallmatrix} & \alpha \\ \alpha^{-1} & \end{smallmatrix}\right]$ *(blank entries standing for* 0*), which acts on* $\mathcal{H}$ *as a reflection (i.e., Möbius inversion) through the geodesic connecting* $-\alpha$ *with* α.

(2) *If* $\Gamma^{\pm}$ *contains a reflection as above then, for each pair* U, V *of unimodular intervals,* $\Gamma^{\pm}$ *contains precisely two elements mapping* U *to* V*, one of them being order-preserving and the other order-reversing. If* $\Gamma^{\pm}$ *does not contain such a reflection, then each unimodular* U *can be mapped to each unimodular* V *by precisely one element of* $\Gamma^{\pm}$.

Proof. Let $A \in \mathrm{PSL}_2^{\pm}\,\mathbb{R}$ be a matrix that fixes I_0 setwise. If it fixes 0 and ∞, then it is of the form $\left[\begin{smallmatrix} \alpha & \\ & \alpha^{-1} \end{smallmatrix}\right]$ for some $\alpha > 0$. Therefore, if A belongs to $\Gamma^{\pm}$, then it belongs to Γ; since ∞ is a cusp of Γ, we have $\alpha = 1$. On the other hand, if A exchanges 0 and ∞ then it must fix some $0 < \alpha < \infty$, and it readily follows that $A = \left[\begin{smallmatrix} & \alpha \\ \alpha^{-1} & \end{smallmatrix}\right]$. If $\Gamma^{\pm}$ contains $\left[\begin{smallmatrix} & \beta \\ \beta^{-1} & \end{smallmatrix}\right]$ as well, then Γ contains $\left[\begin{smallmatrix} \alpha\beta^{-1} & \\ & \alpha^{-1}\beta \end{smallmatrix}\right]$ and, by the above, $\alpha = \beta$. This proves (1), and (2) is an easy consequence. □

We can now make precise our definition of continued fraction algorithm, which we identify with a Gauss map.

Definition 2.3. Let K, d, $\mathbb{O}$, $\Gamma^{\pm}$ be as in Notation 2.1, and let $A_1, \ldots, A_r \in \Gamma^{\pm}$ be such that:

(iv) the family of intervals $I_a = A_a * I_0$, for $a = 1, \ldots, r$, is a unimodular partition of I_0;

(v) $\det A_r = +1$.

The *slow continued fraction algorithm* determined by $A_1, \ldots, A_r$ is the Gauss map $T : I_0 \to I_0$ which is induced on I_a by A_a^{-1} (by convention, the endpoint in $I_a \cap I_{a+1}$ is mapped using A_a^{-1}). The maps A_a^{-1} are the *pieces* of T. If $d = 2$ (respectively, $d = 3$), then we say that T is *quadratic* (*cubic*). We denote by $\Gamma_T^{\pm}$ the subgroup of $\Gamma^{\pm}$ generated by $A_1, \ldots, A_r$, and we let $\Gamma_T = \Gamma_T^{\pm} \cap \mathrm{PSL}_2\,\mathbb{R}$.

The two restrictions in Definition 2.3—about the determinant of A_r and the definition of T at endpoints—simplify some of the proofs and are essentially irrelevant. Indeed, ambiguity at endpoints may occur at most once along the T-orbit of a point, and is analogous to the existence of two finite ordinary continued fractions expansions for rational numbers. On the other hand, requiring $\det A_r = +1$ amounts to requiring that

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

∞ be a fixed point for T. This can always be achieved, either by replacing T with T^2 if $\det A_r = -1 = \det A_1$, or by conjugating T with $x \mapsto x^{-1}$ if $\det A_r = -1 = -\det A_1$.

As assumed in Notation 2.1 (and already used in the proof of Lemma 2.2), the point ∞ is a cusp of Γ, and thus an indifferent fixed point of T. This forces any T-invariant measure absolutely continuous with respect to Lebesgue measure to be infinite, and gives poor approximation properties to the convergents; hence the adjective 'slow' in Definition 2.3. One remedies this by accelerating the algorithm (i.e., considering the first-return map to some subinterval of I_0); note that the acceleration process leaves $\Gamma_T^{\pm}$ unaltered [**21**, §2]. Since we are concerned only with the eventual orbits of points, we do not need any acceleration.

LEMMA 2.4. *The group $\Gamma_T^{\pm}$ has finite index in $\Gamma^{\pm}$, and hence so does Γ_T in Γ.*

Proof. By our assumptions $\Gamma^{\pm}$ has a fundamental domain of finite hyperbolic area; it is then sufficient to show that the same holds for $\Gamma_T^{\pm}$. Let $\mathcal{H}^+ = \{z \in \mathcal{H} : \operatorname{Re} z \geq 0\}$. For every $a \in \{1, \ldots, r\}$, we have

$$A_a * \mathcal{H}^+ \subset \mathcal{H}^+ \quad \text{and} \quad \mathcal{H} = \bigcup \{A_a^k * \mathcal{H}^+ : k \in \mathbb{Z}\}.$$

Therefore $\mathcal{H}^+ \backslash (A_a * \mathcal{H}^+)$ is a fundamental domain for the infinite cyclic group generated by A_a, and this implies that $\mathcal{R} = \mathcal{H}^+ \backslash \bigcup \{A_a * \mathcal{H}^+ : 1 \leq a \leq r\}$ contains a fundamental domain for the action of $\Gamma_T^{\pm}$. Let $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_r = \infty$ be the points determining the unimodular partition $I_1, \ldots, I_r$. Then $\mathcal{R}$ is bounded by the full geodesics $g_0, g_1, \ldots, g_r$, where g_i is the geodesic having endpoints α_{i-1} and $\alpha_i \pmod{r+1}$ (see the thick-bordered region in Figure 1 (left)). By elementary considerations $\mathcal{R}$ has finite hyperbolic area, and our claim follows. □

We note as an aside that in the case $\Gamma^{\pm} = \mathrm{PSL}_2^{\pm} \mathbb{Z}$, Lemma 2.4 has a sharper formulation: for every T, the index of $\Gamma_T^{\pm}$ in $\mathrm{PSL}_2^{\pm} \mathbb{Z}$ is bounded by 8 [**21**, Theorem 4.1].

We now specialize to the case in which $\Gamma^{\pm}$ is an extended triangle group. Let $2 \leq l \leq m \leq n \leq \infty$ be integers such that $1/l + 1/m + 1/n < 1$; then there exists a triangle in $\mathcal{H}$, unique up to isometry, having angles $\pi/l, \pi/m, \pi/n$. The group generated by reflections in the sides of this triangle is the *extended triangle group* $\Delta^{\pm}(l, m, n) < \mathrm{PSL}_2^{\pm} \mathbb{R}$, and its index-2 subgroup of orientation-preserving elements is the *triangle group* $\Delta(l, m, n)$, which is a fuchsian group of cofinite volume. Clearly the original triangle is a fundamental domain for $\Delta^{\pm}(l, m, n)$; we will call it the *fundamental triangle*. Let T be a Gauss map obtained from $\Delta^{\pm}(l, m, n)$ via the construction in Definition 2.3. As remarked above, our assumptions imply that Γ_T contains parabolic elements, and thus at least one of the vertices of the fundamental triangle must lie at infinity. Therefore, n equals ∞ and $\Delta(l, m, \infty)$ is not cocompact.

Example 2.5. Consider the group $\Delta^{\pm}(4, \infty, \infty)$ and fix as fundamental triangle the one with vertices 0, ∞, and $(1 + i \tan(\pi/8))/2$. The matrices expressing reflections in the triangle sides are (see Figure 1 (left))

$$J = \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}, \quad P = \begin{bmatrix} -1 & 1 \\ & 1 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & \\ 2 + \sqrt{2} & -1 \end{bmatrix}. \tag{1}$$

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

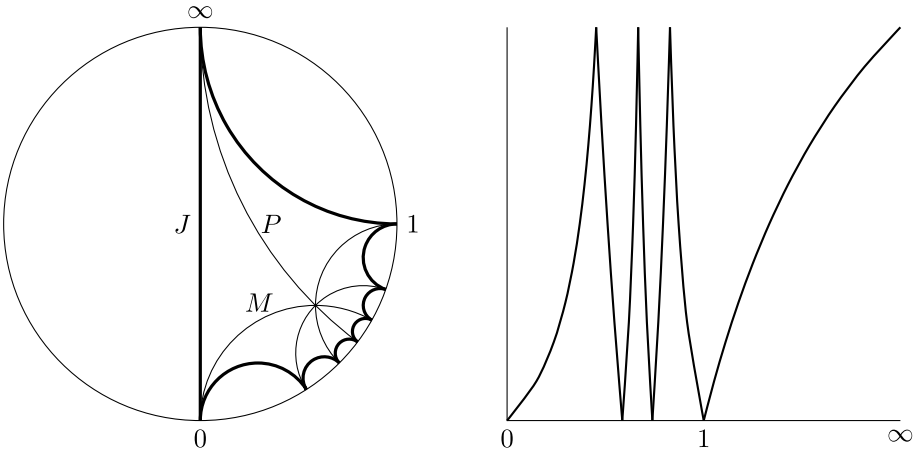


FIGURE 1. The $(4, \infty, \infty)$ case. Fundamental domain and some images; associated interval map.

Postcomposing J first with M, and then with powers of the counterclockwise rotation MP (of order 4), we obtain matrices $A_1, \dots, A_7$ as follows:

$$A_1 = MJ = \begin{bmatrix} 1 & 0 \\ 2+\sqrt{2} & 1 \end{bmatrix}, \quad A_2 = (MP)J = \begin{bmatrix} 1 & 1 \\ 2+\sqrt{2} & 1+\sqrt{2} \end{bmatrix},$$
$$A_3 = (MP)MJ, \quad A_4 = (MP)^2 J, \quad A_5 = (MP)^2 MJ, \quad A_6 = (MP)^3 J,$$
$$A_7 = (MP)^3 MJ = (MP)^{-1} MJ = PJ = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}.$$

The corresponding unimodular partition is

$$0 < 1 - \tfrac{1}{2}\sqrt{2} < -1 + \sqrt{2} < \tfrac{1}{2} < 2 - \sqrt{2} < \tfrac{1}{2}\sqrt{2} < 1 < \infty.$$

The matrices A_1, A_2, A_7 clearly generate $\Delta^{\pm}(4, \infty, \infty)$, which then equals $\Gamma_T^{\pm}$; of course $K = \mathbb{Q}(\sqrt{2})$, $d = 2$, $\mathbb{O} = \mathbb{Z}[\sqrt{2}]$. We draw in Figure 1 (left) the fundamental triangle, its M-image (the union of the two is a fundamental domain for $\Delta(4, \infty, \infty)$), and the images of both under successive powers of MP. The image in the unit disk of the region $\mathcal{H}^+ \setminus \bigcup\{A_j * \mathcal{H}^+ : 1 \le j \le 7\}$ of Lemma 2.4 is bordered by thick geodesics. We also draw in Figure 1 (right) the resulting map T which equals, up to a conjugation, the octagon Farey map in [**26**, **27**]; indeed, setting $E = \left[\begin{smallmatrix} 1-\sqrt{2} & 1 \\ & 2 \end{smallmatrix}\right]$ and $i = 1, \dots, 7$, the matrix γv_i of [**26**, p. 34] is $E^{-1} A_i^{-1} E$. Note that, since we want ∞ to appear as an ordinary point, here and elsewhere we draw the hyperbolic plane in the disk model via the Cayley map $2^{-1/2}\left[\begin{smallmatrix} 1 & -i \\ -i & 1 \end{smallmatrix}\right]$. For the same reason, the graph in Figure 1 (right) is actually the graph of the conjugate of T under the projective transformation $\left[\begin{smallmatrix} 1 & \\ 1 & 1 \end{smallmatrix}\right] : [0, \infty] \to [0, 1]$.

We remark that $\Delta^{\pm}(4, \infty, \infty)$ does not contain elements of the form $\left[\begin{smallmatrix} \alpha^{-1} & \alpha \end{smallmatrix}\right]$ (otherwise the fundamental triangle would not be a fundamental domain) and therefore, by Lemma 2.2, the pieces of T are uniquely determined by the intervals. In other words, the only matrix in (this specific embedding of) $\Delta^{\pm}(4, \infty, \infty)$ that maps I_a to I_0 is A_a^{-1}. Thus, we cannot avoid pieces with negative slopes; if we insist on having all pieces in $\Delta(4, \infty, \infty)$ we must necessarily unfold the map, as in the original construction in [**1**].

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

We need more vocabulary. Let Γ be a fuchsian group of finite covolume, and let Γ^2 be the subgroup generated by the squares of the elements of Γ; since Γ is finitely generated, Γ^2 is a normal subgroup of finite index. The *invariant trace field* of Γ is the trace field F_2 of Γ^2; of course it is a subfield of the trace field F_1 of Γ. Both F_2 and the *invariant quaternion algebra* $F_2[\Gamma^2]$ are invariants (although not complete) of the commensurability class of Γ [**16**, Theorem 3.3.4 and Corollary 3.3.5].

If $\Gamma = \Delta(l, m, n)$ then F_1, F_2, and $F_2[\Gamma^2]$ can be easily computed. Let $\lambda_l = 2\cos(\pi/l)$, and analogously for m and n ($\lambda_\infty = 2$); then

$$F_1 = \mathbb{Q}(\lambda_l, \lambda_m, \lambda_n), \quad F_2 = \mathbb{Q}(\lambda_l^2, \lambda_m^2, \lambda_n^2, \lambda_l\lambda_m\lambda_n).$$

In the non-cocompact case $n = \infty$—which is the one concerning us—the algebra $F_2[\Gamma^2]$ always equals $\mathrm{Mat}_2\, F_2$ (because Γ^2 contains a parabolic element, whose difference with the identity is nilpotent). In the cocompact case $F_2[\Gamma^2]$ is given by the Hilbert symbol [**16**, Lemma 3.5.7]

$$\left(\frac{\lambda_m^2(\lambda_m^2 - 4),\ \lambda_m^2\lambda_n^2(\lambda_l^2 + \lambda_m^2 + \lambda_n^2 + \lambda_l\lambda_m\lambda_n - 4)}{F_2}\right).$$

Each archimedean place $\sigma : F_2 \to \mathbb{R}$ yields a real quaternion algebra $F_2[\Gamma^2] \otimes_\sigma \mathbb{R}$, which necessarily equals either $\mathrm{Mat}_2\, \mathbb{R}$ (in which case we say that σ is *split*) or the Hamilton quaternions. The number of archimedean split places is the *arithmetic dimension* of Γ.

In our non-cocompact case, the arithmetic dimension of Γ is just the degree of F_2 over $\mathbb{Q}$, and the list of all triangle groups of arithmetic dimension 2 could be generated by exhaustive search. Indeed, if $\Delta(l, m, \infty)$ is such a group, then its invariant trace field must contain $\cos(2\pi/l)$, which has degree $\varphi(l)/2$ over $\mathbb{Q}$. Thus we have the bound $\varphi(l) \leq 4$, yielding $l \in \{2, 3, 4, 5, 6, 8, 10, 12\}$, and similarly for m if different from ∞. We spare such search by appealing to [**19**], which proves that the set of triangle groups, including the cocompact ones, of bounded arithmetic dimension is finite, and provides efficient algorithms for its enumeration. In particular, [**19**, §6] gives explicit lists of all triangle groups of low dimension: there are 76 cocompact and 9 non-cocompact groups of dimension 1 (this is the arithmetic case, already settled in [**28**]) and, for dimension from 2 to 5, the corresponding pairs of cardinalities are (148, 16), (111, 13), (286, 31), (94, 6).

In [**29**, Proposition 1] Takeuchi proves that in the arithmetic case the pair $(F_2, F_2[\Gamma^2])$ is a complete invariant of the commensurability class of Γ, and uses this fact to refine his own work in [**28**] into a classification of all 85 triangle groups of arithmetic dimension 1 up to commensurability. The only non-cocompact commensurability class is that of the modular group $\mathrm{PSL}_2\, \mathbb{Z} = \Delta(2, 3, \infty)$, and it turns out that there are 18 cocompact classes. When the arithmetic dimension is greater than 1, the pair $(F_2, F_2[\Gamma^2])$ is not complete anymore, so we apply a different strategy.

THEOREM 2.6. *Two non-arithmetic triangle groups Γ, Δ are commensurable if and only if there exist $C \in \mathrm{PSL}_2\, \mathbb{R}$ and a triangle group Ξ containing both Γ and $C\Delta C^{-1}$ with finite index.*

By the results in [**11**, **25**], the relation of inclusion among triangle groups is easily settled, so we can split the classes in [**19**, §6] into commensurability subclasses.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

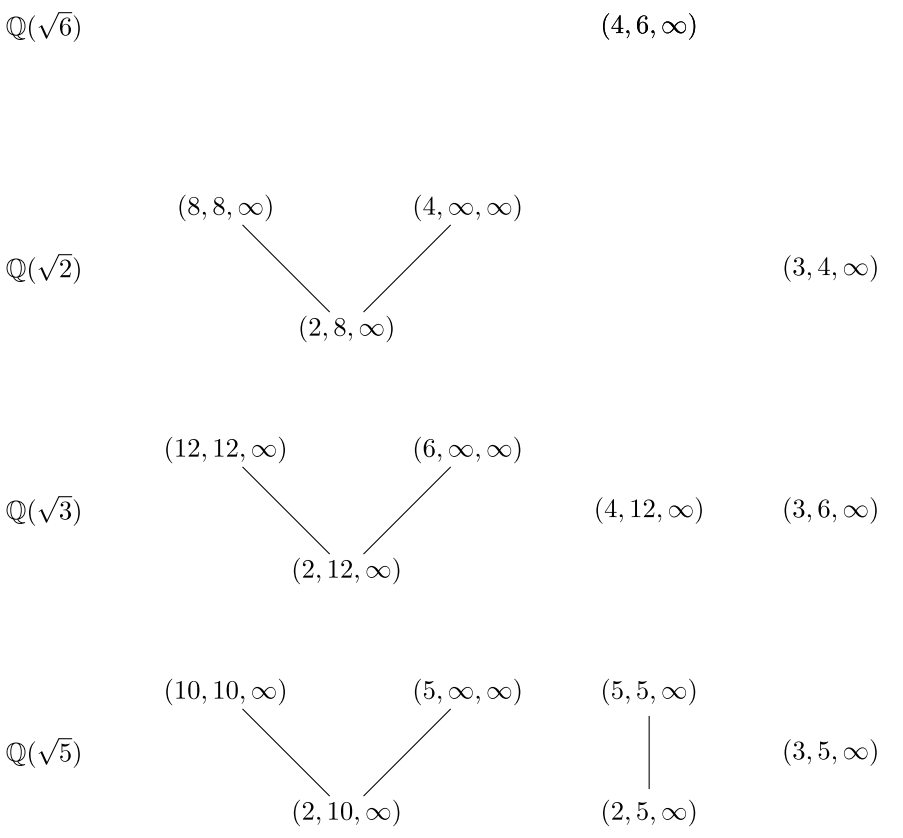


FIGURE 2. Non-cocompact commensurability classes in dimension 2.

In particular, the 16 non-cocompact triangle groups of arithmetic dimension 2 are split as shown in Figure 2. There, the invariant trace field is displayed to the left, and a group lies over another if and only if the first one is a subgroup of the second (always of index 2). Here and later on, for short and when no confusion may arise, we tacitly suppress the initial Δ (or $\Delta^{\pm}$) from our notation for triangle groups.

Proof of Theorem 2.6. A proof is sketched in [**10**, p. 759]; we supply a more detailed version for the reader's convenience. For the non-trivial direction, assume that Γ and Δ are commensurable. Then, after replacing Δ with an appropriate $C\Delta C^{-1}$, we assume that Γ and Δ intersect in a subgroup of finite index in both. Therefore Γ and Δ have the same *commensurator* Ξ, the latter being the group of all $D \in \mathrm{PSL}_2\,\mathbb{R}$ such that Γ and $D\Gamma D^{-1}$ (respectively, Δ and $D\Delta D^{-1}$) intersect in a subgroup of finite index in both. Since we are in the non-arithmetic case, the Margulis dichotomy [**17**] implies that both Γ and Δ have finite index in Ξ; hence Ξ is fuchsian as well. But it is known that the only fuchsian groups that can contain a triangle group with finite index are themselves triangle groups (because the Teichmüller space of the larger group embeds in the space of the subgroup, and triangle groups are precisely those non-cyclic fuchsian groups having trivial spaces; see [**25**, §6]). □

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

3. *Continued fraction algorithms*

In this section we introduce, for each commensurability class in Figure 2, a Gauss map T as in Definition 2.3 such that:

(1) T is continuous (equivalently, $\det A_a = (-1)^{r-a}$ for every $a = 1, \ldots, r$);
(2) K is the invariant trace field of the class;
(3) Γ_T is a group in the class, with the exception of the cases $(4, 6, \infty)$ and $(4, 12, \infty)$, in which Γ_T has index 2 in the unique group of the class;
(4) property (H) holds.

Note that the exception in (3) is unavoidable, since the invariant trace field of $\Delta(4, 6, \infty)$ and of $\Delta(4, 12, \infty)$ is strictly contained in the respective trace fields.

The integer rings of our invariant trace fields $F_2 = K = \mathbb{Q}(\sqrt{d})$, for $d = 2, 3, 5, 6$, are $\mathbb{Z}[\sqrt{2}]$, $\mathbb{Z}[\sqrt{3}]$, $\mathbb{Z}[\tau]$, $\mathbb{Z}[\sqrt{6}]$, respectively ($\tau = (1 + \sqrt{5})/2$ is the golden ratio). All of them have class number 1; since the class group is in natural correspondence with the $\mathrm{PSL}_2\,\mathbb{O}$-orbits in $\mathrm{P}^1\,K$, this is surely necessary for all points in K to be cusps of Γ_T and hence, as remarked in §1, for property (H) to hold. Note that each commensurability class in Figure 2 contains a group whose cusps form a single group orbit.

3.1. *Case* $(l, m, \infty) \in \{(2, 5, \infty), (3, 4, \infty), (3, 5, \infty), (3, 6, \infty)\}$. This is the simplest and most flexible case: we take as fundamental triangle the one having vertices $-\exp(-i\pi/l)$, $\exp(i\pi/m)$, and ∞.

Let

$$F = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, \quad Q_l = \begin{bmatrix} \lambda_l & 1 \\ -1 & \end{bmatrix}, \quad R_m = \begin{bmatrix} & 1 \\ -1 & \lambda_m \end{bmatrix}.$$

Then Q_l is a counterclockwise rotation of order l around $-\exp(-i\pi/l)$, and R_m is a counterclockwise rotation of order m around $\exp(i\pi/m)$. Also, $\Delta(l, m, \infty)$ is a free product $\langle Q_l \rangle * \langle R_m \rangle = Z_l * Z_m$ of cyclic groups, while $\Delta^{\pm}(l, m, \infty)$ is an amalgamated product $\langle Q_l, F \rangle *_{\langle F \rangle} \langle R_m, F \rangle = D_l *_{Z_2} D_m$ of dihedral groups. Let T' be the Gauss map determined by the $(l-1)(m-1)$ matrices $A'_{ij} = R_m^i Q_l^j$, for $1 \le i < m$ and $1 \le j < l$. Figure 3 (left), which is drawn for $(l, m, \infty) = (3, 4, \infty)$, clarifies the definition; appropriate powers of Q_l map I_0 to subintervals of $[\infty, 0]$, and powers of R_m map back these subintervals inside I_0 (for short, in Figure 3 we dropped the subscripts in Q_3, R_4). All matrices A'_{ij} have positive determinant, so T' is not continuous and has all pieces of positive slope. Replacing any A'_{ij} with $A'_{ij}F$, we have full flexibility in converting each slope from positive to negative. In order to obtain a continuous map we perform this matrix replacement precisely at those intervals $I_1, \ldots, I_r$ ($r = (l-1)(m-1)$) having odd index; this completely defines our final map T. It is easy to check that the resulting matrices $A_1, \ldots, A_r$ generate $\Delta^{\pm}(l, m, \infty)$, and therefore $\Gamma_T = \Delta(l, m, \infty)$. Since here λ_l equals 0 or 1, for this group the trace field agrees with the invariant trace field. The required properties (1), (2), (3) are thus satisfied; we will prove (4) in §5.

3.2. *Cases* (l, ∞, ∞), *for* $l = 4, 5, 6$. The $(4, \infty, \infty)$ case has already been dealt with in Example 2.5. In order to treat all cases in a uniform way, we first prove a lemma.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

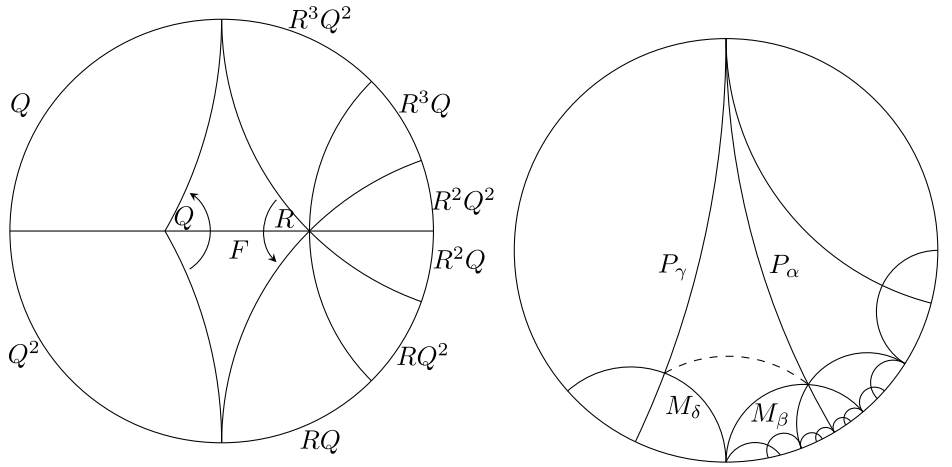


FIGURE 3. Fundamental domain and some images in the $(3, 4, \infty)$ and $(4, 6, \infty)$ cases.

LEMMA 3.1. *Let $\alpha, \beta \in \mathbb{R}^*$, and let*

$$P_\alpha = \begin{bmatrix} -1 & \alpha \\ & 1 \end{bmatrix}, \quad M_\beta = \begin{bmatrix} 1 & \\ \beta & -1 \end{bmatrix}.$$

Then these matrices express the reflections of $\mathcal{H}$ in the geodesic g_α connecting ∞ with $\alpha/2$, and the geodesic g_β connecting 0 with $2/\beta$, respectively. Let g_0 be the geodesic connecting 0 with ∞, which is the axis of reflection of the matrix J in Example 2.5. The three geodesics g_α, g_β, and g_0 form a hyperbolic triangle with two vertices at infinity and another—the one at which g_α and g_β meet—inside $\mathcal{H}$. Let us call the angle at this third vertex the inner angle, of measure $2\pi/p$ (we still let $\lambda_p = 2\cos(\pi/p)$, even though here p might not be an integer). We then have $2\pi/p \le \pi/2$ if and only if $2 \le \alpha\beta < 4$; if this happens, then $\alpha\beta = \lambda_p^2 = 2 + \lambda_{p/2}$, and the bisectrix of the inner angle is the geodesic connecting $-\sqrt{\alpha/\beta}$ with $\sqrt{\alpha/\beta}$.

Proof. The first statement is easily checked, and implies that g_α, g_β intersect if and only if (either $0 < \alpha/2 < 2/\beta$ or $2/\beta < \alpha/2 < 0$) if and only if $0 < \alpha\beta < 4$. Assume that they intersect with an inner angle of $2\pi/p$, for some $2 < p$. Then, by the duplication formula for cosines, $|4\cos^2(\pi/p) - 2| = |2\cos(2\pi/p)| = |\mathrm{tr}(M_\beta P_\alpha)| = |\alpha\beta - 2|$, so that

- either $\lambda_p^2 - 2 = \alpha\beta - 2$, i.e., $2/\beta = 2\alpha/\lambda_p^2$,
- or $\lambda_p^2 - 2 = -\alpha\beta + 2$, i.e., $2/\beta = 2\alpha/(4 - \lambda_p^2)$.

These two cases correspond to the sign ambiguity in $\cos(2\pi/p)$, namely, to the inner angle being acute or obtuse. Since we want it to be acute, we must choose $2/\beta$ to be smaller in absolute value. As $2\pi/p \le \pi/2$ implies $\lambda_p^2 \ge 2$, the second case is ruled out and $\alpha\beta = \lambda_p^2$. The last claim follows by solving for x the equation $P_\alpha = \left[\begin{smallmatrix} x^{-1} & x \end{smallmatrix}\right] M_\beta \left[\begin{smallmatrix} x^{-1} & x \end{smallmatrix}\right]^{-1}$. □

Generalizing Example 2.5, we treat the case (l, ∞, ∞) by using a fundamental triangle bounded by g_α, g_β, g_0, having an inner angle of π/l, and such that J, P_α, $M_\beta \in \mathrm{PSL}_2^\pm \mathbb{O}$, where $\mathbb{O}$ is the appropriate integer ring. According to Lemma 3.1, the ways of

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

accomplishing this are in one-to-one correspondence with the factorization of $2+\lambda_l$ in $\mathbb{O}$; since $\mathbb{O}$ is a unique factorization domain, there is just one way, up to conjugation by units.

For $l=4$, the number $2+\lambda_4 = 2+\sqrt{2}$ is prime in $\mathbb{Z}[\sqrt{2}]$, and the latter has $1+\sqrt{2}$ as fundamental unit. Hence the admissible pairs α, β are exactly, up to exchanging α with β,

$$\alpha = (1+\sqrt{2})^k, \quad \beta = (1+\sqrt{2})^{-k}(2+\sqrt{2}),$$

for $k \in \mathbb{Z}$. In Example 2.5 we just set $k=0$. The choice $k=1$ would have been slightly more symmetric, as α/β is then nearer to 1, but would have made the construction in §5 more involved; hence we maintain $k=0$ for our official T.

For $l=5$, $2+\lambda_5 = 2+\tau$ is prime in $\mathbb{Z}[\tau]$, while for $l=6$, $2+\lambda_6 = 2+\sqrt{3}$ is the fundamental unit in $\mathbb{Z}[\sqrt{3}]$. We retain $\alpha=1$ in both cases, and set β equal to either $2+\tau$ or $2+\sqrt{3}$. Writing P for P_1, the rotation $M_\beta P$ then has order l, and T has $2l-1$ pieces, with $A_1, \ldots, A_{2l-1}$ given by

$$A_{1+2q} = (M_\beta P)^q M_\beta J, \quad A_{2+2q} = (M_\beta P)^{q+1} J, \quad A_{2l-1} = (M_\beta P)^{l-1} M_\beta J = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix},$$

for $0 \le q \le l-2$. We still have $\Gamma_T^{\pm} = \Delta^{\pm}(l,\infty,\infty)$ and (1), (2), (3) are satisfied. Here the situation is much more rigid than in §3.1. Indeed, $\Delta^{\pm}(l,\infty,\infty)$ does not contain any reflection of the form given in Lemma 2.2, and no flexibility is allowed in the orientation of the pieces of T.

3.3. *Cases* $(4, m, \infty)$*, for* $m = 6, 12$. In both cases the trace field is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$, which contains strictly the invariant trace fields $\mathbb{Q}(\sqrt{6})$ and $\mathbb{Q}(\sqrt{3})$ (for $m=6$ and $m=12$, respectively); therefore our $\Gamma_T^{\pm}$ has to be a proper subgroup of $\Delta^{\pm}(4,m,\infty)$. We will use a reflection group, with fundamental domain the union of two $(4,m,\infty)$-triangles, images of each other under a reflection in a common edge (the dashed geodesic arc in Figure 3 (right)).

Let us first treat the case $m=6$. According to Lemma 3.1, we must find $\alpha, \beta, \gamma, \delta \in \mathbb{Z}[\sqrt{6}]$ such that

$$\begin{aligned} \alpha\beta = \lambda_6^2 = 3, &\quad \alpha, \beta > 0, \\ \gamma\delta = \lambda_4^2 = 2, &\quad \gamma, \delta < 0, \\ \alpha/\beta = \gamma/\delta. & \end{aligned} \tag{2}$$

In $\mathbb{Z}[\sqrt{6}]$ we have the fundamental unit $\varepsilon = 5+2\sqrt{6}$, and the unique factorizations $3 = \varepsilon(3-\sqrt{6})^2$, $2 = \varepsilon^{-1}(-2-\sqrt{6})^2$. Equations (2) imply $6 = (\alpha\delta)^2$, so that $\alpha\delta = -\sqrt{6} = (3-\sqrt{6})(-2-\sqrt{6})$. Therefore the general solution of (2) is

$$\begin{aligned} \alpha &= \varepsilon^k(3-\sqrt{6}), & \beta &= \varepsilon^{-k+1}(3-\sqrt{6}), \\ \gamma &= \varepsilon^{k-1}(-2-\sqrt{6}), & \delta &= \varepsilon^{-k}(-2-\sqrt{6}). \end{aligned}$$

The value of k that minimizes the distance from $\alpha/\beta = \varepsilon^{2k-1}$ to 1 is 0, so we set

$$P_\alpha = \begin{bmatrix} -1 & 3-\sqrt{6} \\ & 1 \end{bmatrix}, \quad M_\beta = \begin{bmatrix} 1 & \\ 3+\sqrt{6} & -1 \end{bmatrix},$$
$$P_\gamma = \begin{bmatrix} -1 & 2-\sqrt{6} \\ & 1 \end{bmatrix}, \quad M_\delta = \begin{bmatrix} 1 & \\ -2-\sqrt{6} & -1 \end{bmatrix}.$$

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

Now consider the matrices

$$\begin{aligned} &U_1 = P_\gamma, && U_2 = P_\gamma M_\delta, && U_3 = P_\gamma M_\delta P_\gamma, \\ &V_1 = M_\beta, && V_2 = M_\beta P_\alpha, && V_3 = (M_\beta P_\alpha) M_\beta, \\ &V_4 = (M_\beta P_\alpha)^2, && V_5 = (M_\beta P_\alpha)^2 M_\beta. \end{aligned}$$

As in §3.1, the U_i map I_0 to subintervals of $[\infty, 0]$, and the V_j map these subintervals back inside I_0. A little reflection shows that the matrices

$$\begin{aligned} &A_1 = V_1 U_3, && A_2 = V_1 U_2, && A_3 = V_1 U_1, \\ &A_4 = V_2 U_1, && A_5 = V_2 U_2, && A_6 = V_2 U_3, \\ &A_7 = V_3 U_3, && A_8 = V_3 U_2, && A_9 = V_3 U_1, \\ &A_{10} = V_4 U_1, && A_{11} = V_4 U_2, && A_{12} = V_4 U_3, \\ &A_{13} = V_5 U_3, && A_{14} = V_5 U_2, && A_{15} = V_5 U_1 \end{aligned}$$

(which alternately have determinant ± 1) determine a unimodular partition of I_0, whose 15 intervals are displayed along the right border of the disk model in Figure 3 (right), starting with $I_1 = A_1 * I_0 = [0, 5 - 2\sqrt{6}]$, $I_2 = A_2 * I_0 = [5 - 2\sqrt{6}, (8 - 3\sqrt{6})/5]$, and ending with $I_{15} = A_{15} * I_0 = [1, \infty]$. As a consequence, the matrices $A_1, \ldots, A_{15}$ determine a Gauss map T, and one easily checks that $\Gamma_T^\pm$ equals $\langle P_\alpha, M_\beta, P_\gamma, M_\delta \rangle$, which is an index-2 subgroup of $\Delta^\pm(4, 6, \infty)$. Conditions (1), (2), (3) are thus satisfied. The fact that $\Gamma_T^\pm$ does not contain a reflection as in Lemma 2.2 corresponds to the fact that $\alpha/\beta = 5 - 2\sqrt{6}$ is not a square in $\mathbb{Z}[\sqrt{6}]$, but becomes a square—namely, $(\sqrt{3} - \sqrt{2})^2$—in the integer ring of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

The case $(4, 12, \infty)$ is analogous; we solve (2) (with λ_6^2 replaced by $\lambda_{12}^2 = 2 + \sqrt{3}$) for $\alpha, \beta, \gamma, \delta \in \mathbb{Z}[\sqrt{3}]$. Now $2 + \sqrt{3}$ is the fundamental unit ε of $\mathbb{Z}[\sqrt{3}]$, while $2 = \varepsilon(1 - \sqrt{3})^2$. Arguing as above, the general solution to (2) is $\alpha = \varepsilon^k$, $\beta = \varepsilon^{-k+1}$, $\gamma = \varepsilon^k(1 - \sqrt{3})$, $\delta = \varepsilon^{-k+1}(1 - \sqrt{3})$. We take $k = 0$ and obtain the four matrices

$$P_\alpha = \begin{bmatrix} -1 & 1 \\ & 1 \end{bmatrix}, \qquad M_\beta = \begin{bmatrix} 1 & \\ 2 + \sqrt{3} & -1 \end{bmatrix},$$

$$P_\gamma = \begin{bmatrix} -1 & 1 - \sqrt{3} \\ & 1 \end{bmatrix}, \quad M_\delta = \begin{bmatrix} 1 & \\ -1 - \sqrt{3} & -1 \end{bmatrix}.$$

The matrices U_1, U_2, U_3 are defined precisely as in the previous case, while the list of V_j is extended by $V_6 = (M_\beta P_\alpha)^3$, $V_7 = (M_\beta P_\alpha)^3 M_\beta, \ldots, V_{11} = (M_\beta P_\alpha)^5 M_\beta$. The products $V_j U_i$, in full analogy with the above, yield matrices $A_1, \ldots, A_{33}$ and a corresponding Gauss map, for which all of the above considerations hold verbatim.

4. *The height on the projective line*

We recall the definition and the basic properties of the height. Let K be an extension of $\mathbb{Q}$ of degree d. For each place P of K we denote by $|\ |_P$ the valuation in P normalized by $|\alpha|_P = |\sigma(\alpha)|^{e/d}$ (if P is archimedean associated to the embedding $\sigma : K \to \mathbb{C}$, with $e = 1$ or 2 according whether σ is real or complex), and by $|p|_P = p^{-d(P|p)/d}$ (if P is non-archimedean lying over the rational prime p, with local degree $d(P|p) = \deg[K_P : \mathbb{Q}_p]$). Given $\alpha = [\alpha_1, \alpha_2] \in \mathrm{P}^1 K$, its *absolute height* is

$$H(\alpha) = \prod_P \max(|\alpha_1|_P, |\alpha_2|_P).$$

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

The number $H(\alpha)$ is in $\mathbb{R}_{\geq 1}$ and does not depend on K, nor on the representative chosen for α [**6**, Lemmas 1.5.2 and 1.5.3]. We let $\mathbb{O}$ be the ring of algebraic integers in K.

LEMMA 4.1. *Let* $(\alpha_1, \alpha_2) \in K^2 \backslash \{(0,0)\}$, *let* $A \in \mathrm{PSL}_2^{\pm}\,\mathbb{O}$, *and let* P *be a non-archimedean prime. Let* $\binom{\beta_1}{\beta_2} = A\binom{\alpha_1}{\alpha_2}$. *Then* $\max(|\alpha_1|_P, |\alpha_2|_P) = \max(|\beta_1|_P, |\beta_2|_P)$.

Proof. The fractional ideals $\mathbb{O}\alpha_1 + \mathbb{O}\alpha_2$ and $\mathbb{O}\beta_1 + \mathbb{O}\beta_2$ coincide. Since P is non-archimedean, both $\max(|\alpha_1|_P, |\alpha_2|_P)$ and $\max(|\beta_1|_P, |\beta_2|_P)$ are equal to the maximum of $|\gamma|_P$, for γ varying in that fractional ideal. □

We are interested in the behavior of the height along T-orbits of points in K, where T, K are as in Definition 2.3. Let $A \in \mathrm{PSL}_2^{\pm}\,\mathbb{O}$. If, up to multiplication by -1, all entries of A are greater than or equal to 0, with at most one entry equal to 0, then we call A *positive*. If all entries are greater than 0 we call A *strictly positive*. Since $r \geq 2$, the matrices A_1, A_r are positive (but not strictly), while $A_2, \ldots, A_{r-1}$ are strictly positive. When writing positive matrices, we will always tacitly assume that the displayed entries are greater than or equal to 0.

The case $K = \mathbb{Q}$ is trivial but instructive. The height of $\beta = u/v \in \mathrm{P}^1\,\mathbb{Q}$ is $|u| \vee |v|$ ($\vee$ denoting the max function), provided that u, v are relatively prime integers. Assume that $0 < \beta < \infty$, and let A be a positive matrix in $\mathrm{PSL}_2^{\pm}\,\mathbb{Z}$; setting $\alpha = A * \beta$, it is then clear that $H(\alpha) > H(\beta)$. Since the pieces of T are given by the inverse of positive matrices, this implies that, for every $\alpha \in \mathrm{P}^1\,\mathbb{Q} \cap [0, \infty]$, the height of $T(\alpha)$ is strictly less than the height of α, unless $T(\alpha) \in \{0, \infty\}$. As the height is a positive integer, it cannot decrease indefinitely; hence the T-orbit of every rational point must end up in $\{0, \infty\}$.

By Northcott's theorem [**6**, Theorem 1.6.8], the number of points of bounded height in any $\mathrm{P}^1\,K$ is finite; the argument above would then work in any number field, if only the action of positive matrices in $\mathrm{PSL}_2^{\pm}\,\mathbb{O}$ were height-increasing. However, this is not the case.

Example 4.2. Let $K = \mathbb{Q}(\sqrt{6})$, and consider the strictly positive matrices

$$A_9 = \begin{bmatrix} 2 & -1+\sqrt{6} \\ 3+\sqrt{6} & 2+\sqrt{6} \end{bmatrix}, \quad A_{13} = \begin{bmatrix} 1+\sqrt{6} & 3-\sqrt{6} \\ 2+\sqrt{6} & 1 \end{bmatrix},$$

introduced in §3.3 when treating case $(4, 6, \infty)$. The matrix A_9 is height-increasing: for every $\beta \in K \cap [0, \infty]$, we have $H(A_9 * \beta) > H(\beta)$; this will follow from Theorem 4.3.

On the other hand, let

$$\beta = \frac{703 - 240\sqrt{6}}{380}, \quad A_{13} * \beta = \frac{403 + 83\sqrt{6}}{346 + 223\sqrt{6}}.$$

We can check easily that numerator and denominator in the above expression for β are relatively prime in $\mathbb{Z}[\sqrt{6}]$, and thus we compute

$$H(\beta) = \left[380(703 + 240\sqrt{6})\right]^{1/2} = 700.3809\ldots > H(A_{13} * \beta) = 422.6795\ldots.$$

The situation is actually worse: as we will see in Theorem 4.3, the set of points in K on which A_{13} is height-decreasing is dense in $[0, \infty]$.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

We now introduce our main tool. Let K be any real quadratic field with integer ring $\mathbb{O}$, denote by $'$ the non-trivial automorphism of K, and let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{PSL}_2^{\pm}\,\mathbb{O}$ be positive. We define

$$
\begin{aligned}
f(x) &= \frac{(ax+b)\vee(cx+d)}{x\vee 1},\\
t &= (|a'|+|c'|)\vee(|b'|+|d'|) = |a'+c'|\vee|a'-c'|\vee|b'+d'|\vee|b'-d'|,\\
E^\sharp &= \{x\in[0,\infty] : f(x) > t\},\\
E^\natural &= \{x\in[0,\infty] : f(x) = t\},\\
E^\flat &= \{x\in[0,\infty] : f(x) < t\}.
\end{aligned}
$$

When dealing with more than one matrix, we will use notation such as $t(A)$, $E^\sharp(A)$, f_A for clarity.

THEOREM 4.3. *Assuming the above notation, the following statements hold.*

(1) *f is non-decreasing on* $[0, 1]$ *and non-increasing on* $[1, \infty]$.
(2) *$E^\sharp$ is either empty or an open interval containing* 1.
(3) *$E^\natural$ either contains at most two points or equals one of the intervals* $[0, 1]$, $[1, \infty]$.
(4) *$H(A * \beta) > H(\beta)$ for every $\beta \in E^\sharp \cap K$; the same holds if A is strictly positive and β is a boundary point of $E^\sharp$ such that $A * \beta \neq 1$.*
(5) *If $E^\natural$ is an interval, then $\{\beta \in E^\natural \cap K : H(A * \beta) = H(\beta)\}$ is dense in $E^\natural$.*
(6) *$\{\beta \in E^\flat \cap K : H(A * \beta) < H(\beta)\}$ is dense in $E^\flat$.*

Assuming Theorem 4.3, let us justify our claims in Example 4.2. We have

$$t(A_9) = -1 + \sqrt{6} < \min f_{A_9} = f_{A_9}(0) \wedge f_{A_9}(\infty) = 2 + \sqrt{6};$$

hence $E^\sharp(A_9) = [0, \infty]$ and statement (4) applies. On the other hand,

$$t(A_{13}) = 4 + \sqrt{6} > \max f_{A_{13}} = f_{A_{13}}(1) = 3 + \sqrt{6};$$

hence $E^\sharp(A_{13}) = E^\natural(A_{13}) = \emptyset$ and $E^\flat(A_{13}) = [0, \infty]$, so (6) applies; see Figure 4 (right).

Proof of Theorem 4.3. Since $a, b, c, d \geq 0$, statements (1) and (2) are clear. The function f is best visualized by precomposing it with the order-preserving homeomorphism $\varphi : [0, 2] \to [0, \infty]$ which is the identity on $[0, 1]$ and equals $(2 - x)^{-1}$ on $[1, 2]$. Then $f \circ \varphi$ is piecewise-linear, namely,

$$(f\circ\varphi)(x) = \begin{cases} (ax+b)\vee(cx+d) & \text{on } [0,1],\\ (-bx+a+2b)\vee(-dx+c+2d) & \text{on } [1,2]. \end{cases}$$

We have $f(0) = b \vee d$, $f(\infty) = a \vee c$, and $f(1) = (a + b) \vee (c + d) = \max f$. Figure 4 (left) shows $f \circ \varphi$ and t (dashed) for the matrix

$$A_2 A_{10}^9 = \begin{bmatrix} 1 & 10+9\sqrt{3} \\ \sqrt{3} & 28+10\sqrt{3} \end{bmatrix},$$

where A_2, A_{10} are the matrices introduced in §3.1 when treating case $(3, 6, \infty)$. The dashed vertical segments mark the intervals $I_1, \ldots, I_{10}$ of that case. Figure 4 (right) is the analogous one for the matrix A_{13} of Example 4.2.

We formulate two remarks.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

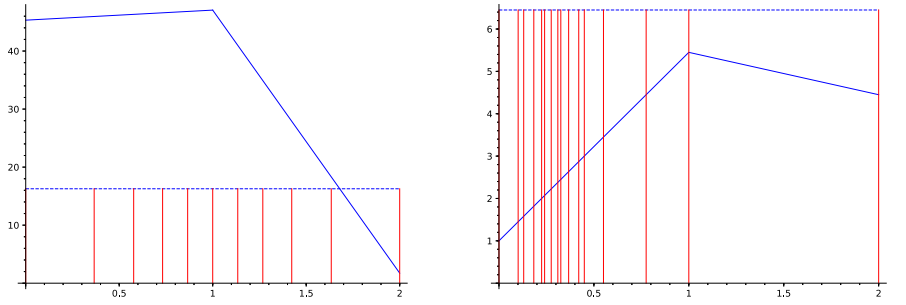


FIGURE 4. Superimposed graphs of $f_A \circ \varphi$ and $t(A)$: (left) $A = A_2 A_{10}^9$ of §3.1; (right) $A = A_{13}$ of Example 4.2.

- Exchanging the rows of A does not modify f and t, and corresponds to postcomposing A with the involution $x \leftrightarrow 1/x$, which leaves the height invariant. On the other hand, exchanging the columns does not modify t, and transforms $f(x)$ to $f(1/x)$ (from the point of view of $f \circ \varphi$, it is a reflection in the $x = 1$ axis). This corresponds to precomposing A with $x \leftrightarrow 1/x$, which simply exchanges $E^\sharp$, $E^\natural$, $E^\flat$ with their images under the involution, leaving our claims unaffected.
- The identities $|a'| = |c'|$ and $|b'| = |d'|$ cannot hold simultaneously. This is clear if one of the entries of A is 0. Otherwise, assume by contradiction that A is strictly positive and both identities hold. Then we have

$$\begin{aligned} 1 &= a'd' - b'c' \\ &= \operatorname{sgn}(a'd')|a'||d'| - \operatorname{sgn}(b'c')|b'||c'| \\ &= (\operatorname{sgn}(a'd') - \operatorname{sgn}(b'c'))|a'||b'|, \end{aligned}$$

  which is impossible, since $\operatorname{sgn}(a'd') - \operatorname{sgn}(b'c') \in \{-2, 0, 2\}$ and $a'b' \neq \pm 1/2$ (because a', b' are algebraic integers).

We therefore assume, without loss of generality, that $|a'| < |c'|$ and that a is the only entry of A which can possibly be 0.

We prove (3). If $a \neq 0$, then f is strictly increasing on $[0, 1]$ and strictly decreasing on $[1, \infty]$; hence $E^\natural$ contains at most two points. Otherwise, A has the form $\left[\begin{smallmatrix} b \\ b^{-1} & d \end{smallmatrix}\right]$, for some unit $b \in \mathbb{O}$. Since $E^\natural$ contains three or more points, it contains an interval, which necessarily takes the form $[0, r]$, for some $0 < r \leq 1$; indeed, such an interval is the only region where f has derivative 0. Looking for a contradiction, we assume $r < 1$; then

$$d < b^{-1}r + d = f(r) = b = t = |(b^{-1})'| \vee (|b'| + |d'|) < f(1) = b \vee (b^{-1} + d).$$

As $|b'| = b^{-1}$, we obtain

$$d < b = b \vee (b^{-1} + |d'|) < b \vee (b^{-1} + d),$$

from which we derive

$$\begin{aligned} d < b < b^{-1} + d, \\ b^{-1} < b^{-1} + |d'| \leq b, \\ b - b^{-1} < d < b, \end{aligned}$$

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

$$-b + b^{-1} \le d' \le b - b^{-1},$$
$$-b^{-1} < d - b < 0,$$
$$-b + b^{-1} - b' \le d' - b' \le b - b^{-1} - b'.$$

Now, b^{-1} equals either b' or $-b'$; in both cases the last line in the above chain yields $|d' - b'| \le b$. Therefore, writing N for the norm function, $|N(d-b)| = |d-b| \cdot |d' - b'| < b^{-1}b = 1$; as $d - b \in \mathbb{O}\backslash\{0\}$ and the norm takes integer values, this is impossible and concludes the proof of (3).

Now let β range in $K \cap [0, \infty)$. We have $A\binom{\beta}{1} = \binom{a\beta+b}{c\beta+d}$; by Lemma 4.1,

$$H(A * \beta) \gtreqless H(\beta)$$

if and only if

$$(|a\beta + b| \vee |c\beta + d|)(|a'\beta' + b'| \vee |c'\beta' + d'|) \gtreqless (|\beta| \vee 1)(|\beta'| \vee 1)$$

if and only if (since $\beta, a, b, c, d \ge 0$)

$$f(\beta)(|a'\beta' + b'| \vee |c'\beta' + d'|) \gtreqless |\beta'| \vee 1. \tag{$*$}$$

We note that the corresponding condition for $\beta = \infty$ is

$$f(\infty)(|a'| \vee |c'|) \gtreqless 1. \tag{$**$}$$

For a given $x \in [0, \infty]$, we denote by $g_x(y)$ the function of the real variable y defined by

$$g_x(y) = f(x)(|a'y + b'| \vee |c'y + d'|).$$

Since $|a'| < |c'|$, the graph of g_x has a truncated V shape, with two outer half lines whose slope has absolute value $f(x)|c'|$, and an inner segment whose slope has absolute value $f(x)|a'|$. These affine linear parts are connected by two distinct *cornerpoints*, which we compute by solving the systems

$$\begin{cases} f(x)a'y + f(x)b' = z \\ f(x)c'y + f(x)d' = z \end{cases} \qquad \begin{cases} -f(x)a'y - f(x)b' = z \\ f(x)c'y + f(x)d' = z \end{cases}$$

in the variables (y, z)—say, via the Cramer rule—and then taking the absolute value of the resulting z. After the computation, the cornerpoints turn out to be

$$q_1 = \left(\frac{b' - d'}{-a' + c'}, \frac{f(x)}{|-a' + c'|}\right), \quad q_2 = \left(\frac{-b' - d'}{a' + c'}, \frac{f(x)}{|a' + c'|}\right).$$

CLAIM. *Let β belong to the closure of $E^\sharp$, and let q_1, q_2 be the cornerpoints of g_β; then the following hold.*

(i) *The absolute value $f(\beta)|c'|$ of the slopes of the outer half lines of g_β is greater than* 1.

(ii) *q_1 and q_2 belong to the epigraph of $|y| \vee 1$, namely the convex region $\{(y, z) : y + z \ge 0, z \ge 1, -y + z \ge 0\}$. If $\beta \in E^\sharp$, then they belong to the strict epigraph (defined analogously, with $>$ replacing $\ge$) and hence, by (i), the entire graph of g_β lies in the strict epigraph of $|y| \vee 1$.*

(iii) *If $a \ne 0$ then q_1, q_2 do not lie in the same affine linear part of the graph of $|y| \vee 1$.*

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

Proof of Claim. (i) From the triangle inequality we get

$$\begin{aligned} 1 = a'd' - b'c' &\le |a'd'| + |b'c'| \\ &< |c'|(|d'| + |b'|) \\ &\le |c'|t \\ &\le |c'|f(\beta). \end{aligned}$$

(ii) The cornerpoint q_1 belongs to the epigraph of $|y| \vee 1$ if and only if $\pm(b' - d') + f(\beta) \ge 0$ and $f(\beta) \ge |-a' + c'|$, and this condition amounts to $f(\beta) \ge |a' - c'| \vee |b' - d'|$. Analogously, q_2 belongs to the epigraph if and only if $f(\beta) \ge |a' + c'| \vee |b' + d'|$. By our assumptions on β we have $f(\beta) \ge t$, and the above inequalities hold by definition of t. If $\beta \in E^\sharp$ then $f(\beta) < t$ and the inequalities are strict.
(iii) Suppose that q_1, q_2 lie on $\{-y + z = 0\}$. Then $|b' - d'| = f(\beta) = |b' + d'|$, which is impossible since b, d are both different from 0. The same argument works for the linear parts along $\{z = 1\}$ and $\{y + z = 0\}$. □

We now prove (4). For $\beta = \infty$, the greater-than inequality in $(**)$ is just (i) in the claim; so let $\beta \ne \infty$. If $\beta \in E^\sharp$ then, by (ii) in the claim, $g_\beta(y) > |y| \vee 1$ for every $y \in \mathbb{R}$. In particular, $g_\beta(\beta') > |\beta'| \vee 1$, which amounts to the greater-than inequality in $(*)$. Assume that A is strictly positive and that β is a boundary point of $E^\sharp$ with $A * \beta \ne 1$; then $f(\beta) = t$. By (i) and (iii) of the claim, the graph of g_β is contained in the strict epigraph of $|y| \vee 1$, with the exception of at least one and at most two of q_1, q_2. The first coordinates of these cornerpoints are $(A^{-1} * 1)'$ and $(A^{-1} * (-1))'$. Neither of these can be β', since otherwise we would have $A * \beta = 1$ (which is excluded by hypothesis) or $A * \beta = -1$ (which is impossible). Therefore $g_\beta(\beta') > |\beta'| \vee 1$ holds as well and (4) is established.

We prove (5). Let $J \subseteq E^\natural = [0, 1]$ be an interval of positive length. By the proof of (3), f has value b on J. Thus, for every $x \in J$, the cornerpoints of g_x are

$$q_1 = \left(\frac{b' - d'}{N(b)b}, 1\right), \quad q_2 = \left(\frac{-b' - d'}{N(b)b}, 1\right).$$

Indeed, under our hypotheses, a equals 0 and b is a unit, so that $c' = (b^{-1})' = (b^{-1})'bb^{-1} = N(b^{-1})b = N(b)b$. Since $|b' - d'| \vee |-b' - d'| \le t = b$, the interval J' of endpoints $(b' - d')/(N(b)b)$ and $(-b' - d')/(N(b)b)$ is contained in $[-1, 1]$ and has positive length. By the Artin–Whaples approximation theorem [**14**, Ch. XII, Theorem 1.2], there exists $\beta \in K$ such that $\beta \in J$ and $\beta' \in J'$. Therefore $g_\beta(\beta') = 1 = |\beta'| \vee 1$, which means equality in $(*)$ and establishes (5).

Finally, let $J \subseteq E^\flat$ be an interval of positive length; by possibly shrinking it, we find $m < t$ such that $f \le m$ on J. By definition of t, either $m < |a' + c'| \vee |a' - c'|$ or $m < |b' + d'| \vee |b' - d'|$ (or both). In any case, at least one of the cornerpoints of the graph of $m(|a'y + b'| \vee |c'y + d'|)$ is strictly below the graph of $|y| \vee 1$; in particular, there exists an interval of positive length J' such that $g_x(y) < |y| \vee 1$ for every pair $(x, y) \in J \times J'$. Again by the Artin–Whaples theorem, there exists $\beta \in K$ with $(\beta, \beta') \in J \times J'$; hence $g_\beta(\beta') < |\beta'| \vee 1$ and $H(A * \beta) < H(\beta)$. This establishes (6) and concludes the proof of Theorem 4.3. □

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

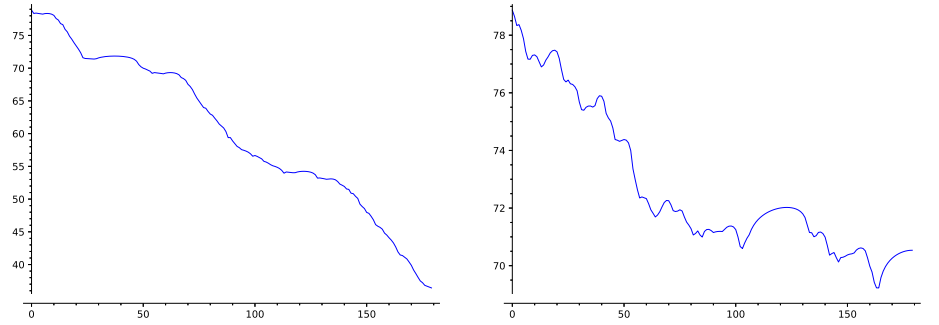


FIGURE 5. Logarithmic heights along an orbit: a quadratic (left) and a cubic case (right).

5. *Decreasing blocks*

Let T and K be a Gauss map and a number field as in Definition 2.3. For a 'generic' input $\alpha \in K$, the height along the T-orbit of α is 'generically decreasing'. This is of course a vague statement, but is exemplified in Figure 5 (left), which shows the logarithmic height (i.e., the logarithm of the height) along the first 180 points in the T-orbit of α, where

- T is the map introduced in §3.1 for the $(3, 4, \infty)$ case;
- α is the rational number

$$\alpha = \frac{1787256914210011674751898944150441 1}{5689015449433817447221222422292025},$$

whose only—as far as we know—claim to glory is being the 66th convergent of the ordinary continued fractions expansion of π.

In this final section we will give a precise formulation to property (H), and prove it for all the algorithms in §§3.1, 3.2 and 3.3. This will remove, in a quite effective way, all vagueness from the above assertion. Note that, as remarked in §1, the validity of (H) implies that any point in K must end its orbit in a parabolic fixed point; explicit computation shows that α lands at ∞ at the 656th iterate of T.

It is interesting to compare the behavior of the previous quadratic T with that of a cubic map. Let $\lambda_7 = 2\cos(\pi/7)$, whose minimal polynomial is $x^3 - x^2 - 2x + 1$; again the maximal order $\mathbb{Z}[\lambda_7]$ in $\mathbb{Q}(\lambda_7)$ has class number 1. Precisely as in §3.1 for the case $(2, 5, \infty)$, we define a Gauss map T' via the matrices

$$R_7Q_2F, \quad R_7^2Q_2, \quad R_7^3Q_2F, \quad R_7^4Q^2, \quad R_7^5Q_2F, \quad R_7^6Q_2,$$

which generate a copy of the extended Hecke group $\Delta^{\pm}(2, 7, \infty)$. The graphs of T and T' have the same structure; they are both saw-like, with six pieces and ∞ as the only parabolic fixed point (similar to Figure 1 (right), but with the leftmost piece deleted). Figure 5 (right) shows the behavior of the logarithmic height for the first 180 points of the T'-orbit of the same input α. The decreasing drift is clearly much more erratic, and indeed α takes 4583 steps to reach ∞. Note that, as far as we know, there is no a priori guarantee that α should land at ∞, nor that it is a cusp of $\Delta(2, 7, \infty)$. Indeed, property (H) definitely fails: the point $\lambda_7^2 - \lambda_7 - 1$ (which has minimal height greater than 1 in $K = \mathbb{Q}(\lambda_7)$) is a fixed point for T', because it is in I_1 and is fixed by $A_1 = R_7Q_2F = \left[\begin{smallmatrix} 1 & 1 \\ \lambda_7 & \end{smallmatrix}\right]$. As it is fixed by the hyperbolic A_1^2, it is not a cusp of $\Delta(2, 7, \infty)$. This is an instance

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

of a fixed point for a so-called special hyperbolic matrix, which corresponds to an affine pseudo-Anosov diffeomorphism with direction of vanishing SAF invariant; see [**2**, §5] and references therein.

Given a Gauss map T as in Definition 2.3, we regard $\mathfrak{A} = \{1, \dots, r\}$ as a finite alphabet; as usual, $\mathfrak{A}^*$ denotes the set of all finite words over $\mathfrak{A}$, while $\mathfrak{A}^\omega$ is the space of all one-sided infinite sequences. For every $\alpha \in [0, \infty]$, a *T-symbolic orbit* for α is a sequence $\mathbf{a} = a_0 a_1 \cdots \in \mathfrak{A}^\omega$ such that $T^k(\alpha) \in I_{a_k}$, for $k \geq 0$. Every point has at least one and at most two symbolic orbits; it has two if and only if it eventually lands at an endpoint of one of $I_2, \dots, I_{r-1}$ (so that, at the next step, it gets mapped to either 0 or ∞). Note that this ambiguity is in the nature of things; if the map $\mathbf{a} \mapsto \alpha$ were injective then, since it it surjective and obviously continuous, it would constitute a homeomorphism from the Cantor space $\mathfrak{A}^\omega$ to $[0, \infty]$. The symbolic orbit $\overline{r}$ of ∞ is unique, however, as is that of 0, namely $\overline{1}$ or $1\overline{r}$ according whether $\det A_1$ equals $+1$ or -1.

The following result generalizes [**27**, Lemma 2.2.15], as well as [**20**, Observation 3], to arbitrary extended fuchsian groups; note that the proof uses a different technique.

THEOREM 5.1. *For every $\mathbf{a} \in \mathfrak{A}^\omega$, the nested strictly decreasing sequence of $\Gamma^\pm$-unimodular intervals*

$$I_0 \supset A_0 * I_0 \supset A_0 A_1 * I_0 \supset A_0 A_1 A_2 * I_0 \supset \cdots$$

shrinks to a singleton $\{\alpha\}$. The point α is the unique point having symbolic orbit **a**.

Proof. The second statement is an obvious consequence of the first. The intersection of the above sequence of intervals is surely a closed interval $[\alpha, \beta]$, with $0 \leq \alpha \leq \beta \leq \infty$; we show that $\alpha = \beta$. Suppose not, and let $B_k = A_0 \cdots A_k$. Then there exists a strictly increasing sequence $k_0, k_1, \dots$ such that $B_{k_l} * 0$ converges, as $l \to \infty$, to one of α, β, say to α, and $B_{k_l} * \infty$ converges to β. Since the projective space $\mathrm{P}\,\mathrm{Mat}_2\,\mathbb{R} = \mathrm{P}^3\,\mathbb{R}$ is compact, we can find a subsequence $(k_{l_j})_{j \in \omega}$ such that $\lim_{j \to \infty} [B_{k_{l_j}}] = [C]$, for some $C \in (\mathrm{Mat}_2\,\mathbb{R}) \backslash 0$, square brackets denoting projective equivalence. As $C * 0 = \alpha \neq \beta = C * \infty$, the matrix C is non-singular and determines an isometry of $\mathcal{H}$, possibly orientation-reversing. Now, the point $C * i$ belongs to $\mathcal{H}$, and is an accumulation point of the sequence $(B_{k_{l_j}} * i)_{j \in \omega}$. Since the B_k belong to $\Gamma^\pm$ and are all distinct, this contradicts the fact that $\Gamma^\pm$ acts on $\mathcal{H}$ in a properly discontinuous way. □

Definition 5.2. Let T be a continuous quadratic Gauss map as in Definition 2.3. A *decreasing block* is a word $b_0 \cdots b_u \in \mathfrak{A}^*$, of length $u + 1 \geq 2$, such that:

(i) $b_0, b_u \neq r$;

(ii) I_{b_u} is contained in the closure of $E^\sharp(A)$, where $A = A_{b_0} \cdots A_{b_{u-1}} = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$;

(iii) if A contains a 0 entry (this happens if and only if $\det A_1 = +1$ and A is a power of A_1, or $\det A_1 = -1$ and $A = A_1$) and $\beta \in \partial E^\sharp(A) \cap \partial I_{b_u}$, then

$$t(A)(|a'\beta' + b'| \vee |c'\beta' + d'|) > |\beta'| \vee 1.$$

A set $\mathcal{B}$ of decreasing blocks is *complete* for T if:

(iv) $\det A_1 = -1$ and every $\mathbf{a} \in \mathfrak{A}^\omega$ that does not start with r, and does not end with $\overline{r}$, begins with a block in $\mathcal{B}$;

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

(v) $\det A_1 = +1$ and every $\mathbf{a} \in \mathfrak{A}^\omega$ that does not start with r, and does not end either with $\overline{r}$ or with $\overline{1}$, begins with a block in $\mathcal{B}$.

We can at last formulate property (H) in a precise way.

THEOREM 5.3. *Each of the algorithms introduced in §§3.1, 3.2 and 3.3 admits a complete set of decreasing blocks.*

The content of Theorem 5.3 is as follows. Choose any symbolic sequence $\mathbf{a} \in \mathfrak{A}^\omega$ that does not end in $\overline{r}$ (nor in $\overline{1}$, if $\det A_1 = +1$). Then, by Theorem 5.3, there exists an infinite sequence of times $0 \le t_0 < t_1 < t_2 < \cdots$ such that each block $a_{t_k} \cdots a_{t_{k+1}}$ is decreasing. Choose any initial very long finite subword w of $\mathbf{a}$, and let α be any element of K having a T-symbolic orbit that starts with w; let $h = \max\{l : t_l < \text{length}(w)\}$. Then the following fact is true.

CLAIM. *For every $0 \le k < h$, the strict inequality*

$$H(T^{t_k}(\alpha)) > H(T^{t_{k+1}}(\alpha))$$

holds, with the only possible exception of the last index $k = h - 1$, where we may have equality; however, this exception occurs if and only if $T^{t_{h-1}}(\alpha) = 1$, and forces $T^{t_h}(\alpha) = 0$.

Proof of Claim. Let us fix a decreasing block $b_0 \cdots b_u$ and $\gamma \in K$ having a symbolic orbit that starts with $b_0 \cdots b_u$. Let $A = A_{b_0} \cdots A_{b_{u-1}}$ and $\beta = T^u(\gamma)$; then $\beta \in I_{b_u}$ and $\gamma = A * \beta$. By the definition of decreasing block, β is in the closure of $E^\sharp(A)$. If $\beta \in E^\sharp(A)$, or ($\beta \in \partial E^\sharp(A)$, A is strictly positive, and $A * \beta \ne 1$), then $H(A * \beta) > H(\beta)$ by Theorem 4.3(4). If $\beta \in \partial E^\sharp(A)$ then $f_A(\beta) = t(A)$; also, by Definition 5.2(ii), $\beta \in \partial I_{b_u}$. If A contains a 0 entry then, by Definition 5.2(iii),

$$f_A(\beta)(|a'\beta' + b'| \vee |c'\beta' + d'|) > |\beta'| \vee 1,$$

which is (∗) in the proof of Theorem 4.3; thus, $H(A * \beta) > H(\beta)$ again. In the exceptional case in which $\beta \in \partial E^\sharp(A)$, A is strictly positive, and $A * \beta = 1$, we have $H(\gamma) \ge H(\beta)$ anyway, so $H(A * \beta) = H(\beta) = 1$ and $\beta \in \{0, 1, \infty\}$. Since $b_u \ne r$, the case $\beta = \infty$ is excluded. If β were equal to 1, then 1 would be fixed by a strictly positive matrix in Γ, namely A or A^2. It is easy to show that this contradicts our assumption that 1 is a cusp of Γ; we thus conclude $\beta = 0$. Note that the exceptional case may occur only at the final decreasing block in w, because as soon as $T^t(\alpha)$ reaches value 0, it can no longer assume value 1. Our claim is thus established. □

We now state formally the geometric consequence of property (H) sketched in §1. Let X be a subset of $\mathrm{P}^1\,\mathbb{R}$, of cardinality $\sharp X$ at least three. We let $\mathrm{CR}(X)$ be the set of all cross-ratios $(\alpha_1, \alpha_2; \alpha_3, \alpha_4)$, for $\alpha_i \in X$ and $\sharp\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} \ge 3$; we have $\mathrm{CR}(X) \subseteq \mathrm{P}^1\,\mathbb{R}$, and $(\alpha_1, \alpha_2; \alpha_3, \alpha_4) \in \{0, 1, \infty\}$ if and only if $\sharp\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} = 3$.

THEOREM 5.4. *Let Γ be a non-cocompact triangle group with quadratic invariant trace field F_2. Then* $\mathrm{CR}(\textit{cusps of } \Gamma) = \mathrm{P}^1\,F_2$.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

Proof. The set CR(cusps of Γ) is a commensurability invariant of Γ; indeed, it does not change either by conjugating Γ (by the invariance of cross-ratios under projective transformations), or by passing to subgroups of finite index (by the remark at the beginning of §2). We can then assume that Γ equals Γ_T, for precisely one of our Gauss maps T. By construction, all matrices in Γ_T have entries in the integer ring of F_2, and this easily implies that every cusp of Γ_T is in $\mathrm{P}^1\, F_2$. For the reverse inclusion, let $\alpha \in \mathrm{P}^1\, F_2$. By applying a power of A_r (the parabolic matrix in Definition 2.3(v)) to α, we may assume $\alpha \in [0, \infty]$. By Theorem 5.3 and the Northcott property, there exists a product A of $A_1^{-1}, \ldots, A_r^{-1} \in \Gamma_T^{\pm}$ such that $A * \alpha$ equals 0 or ∞. Postcomposing A, if necessary, with an element of $\Gamma_T^{\pm}$ that fixes $A * \alpha$ and has negative determinant, we may assume $A \in \Gamma_T$. As 0 and ∞ are cusps of Γ_T, so is α. □

Remark 5.5. As noted in §1, Theorem 5.4 was proved by Leutbecher for the four quadratic Hecke groups, and reestablished by McMullen [**18**, Theorem A.2], using the Veech theory, for three more groups, namely $\Delta(3, 4, \infty)$, $\Delta(3, 5, \infty)$, $\Delta(3, 6, \infty)$; these three groups had been previously examined by the Leutbecher school [**4**, **23**]. As pointed out by the referee, [**18**, Theorem A.1] and the further realizations of Veech groups in [**7**] yield the full result.

Theorem 5.3 is proved by direct verification. One computes $E^\sharp(A_a)$, for each $a = 1, \ldots, r-1$; whenever its closure contains some I_b, with $b \neq r$, the block ab is decreasing (for $a = 1$, boundary points must be checked according to Definition 5.2(iii)). If I_b is not contained in the closure of $E^\sharp(A_a)$, then one computes $E^\sharp(A_a A_b)$ and repeats the process. After some practice a good candidate for $\mathcal{B}$ is readily reached, and checking that indeed it constitutes a complete set of decreasing blocks is just a matter of patience. The simplest cases are those in which the intervals are evenly spread around 1, namely those in §3.1. The cases in §3.2 are slightly more involved, due to the presence of a second parabolic fixed point at 0 (note that ∞ and 0 do not play a symmetric role). The case $(4, 6, \infty)$ is the most difficult one, since all intervals $I_1, \ldots, I_{14}$ cumulate in $[0, 1]$, and several of the $E^\sharp$ are small. Despite the larger number of intervals, case $(4, 12, \infty)$ is actually simpler, because intervals are more evenly spread, and the central ones behave in the same way. We start with the simplest case.

5.1. *Case* $(2, 5, \infty)$. The matrices are

$$A_1 = \begin{bmatrix} & 1 \\ 1 & \tau \end{bmatrix}, \quad A_2 = \begin{bmatrix} \tau & 1 \\ \tau & \tau \end{bmatrix}, \quad A_3 = \begin{bmatrix} \tau & \tau \\ \tau & 1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & \tau \\ & 1 \end{bmatrix},$$

and the four intervals have endpoints $0 < \tau^{-1} < 1 < \tau < \infty$. Let $\mathcal{B}$ be the family of all blocks of the form $a4^k b$, where $a, b \neq 4$ and $k \geq 0$. For convenience, we introduce some shorthand for block families; notation such as $\{a, b\}^c\, c\{d, e\} f^*\{f, g\}^c$ will denote the set of all blocks of the form $xcyf^k z \in \mathfrak{A}^*$ such that $x \in \{a, b\}^c = \mathfrak{A}\backslash\{a, b\}$, $y \in \{d, e\}$, $z \in \{f, g\}^c$, and $k \geq 0$. The completeness of the above $\mathcal{B}$ (which, in the previous notation, is $\{4\}^c 4^* \{4\}^c$) is clear: every $\mathbf{a} \in \{1, 2, 3, 4\}^\omega$ that does not start with 4 and does not end

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

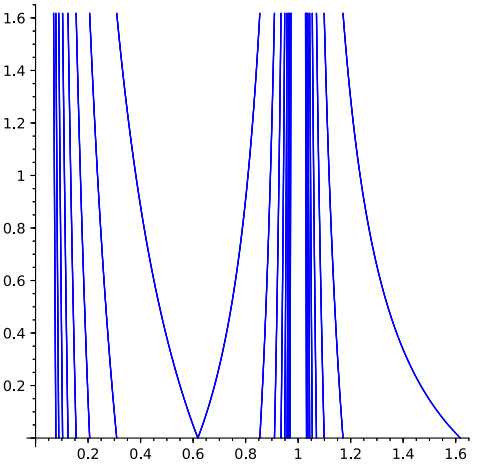


FIGURE 6. The first-return map in the $(2, 5, \infty)$ case.

with $\overline{4}$ (det $A_1 = -1$ here) must begin with a block in $\mathcal{B}$. For $k \geq 0$ we have

$$A_1 A_4^k = \begin{bmatrix} & 1 \\ 1 & (k+1)\tau \end{bmatrix}, \quad A_2 A_4^k = \begin{bmatrix} \tau & (k+1)+k\tau \\ \tau & k+(k+1)\tau \end{bmatrix}.$$

We compute $t(A_1 A_4^k) = 1 + (k+1)\tau^{-1}$, $E^\sharp(A_1) = (0, \tau^2)$, and $E^\sharp(A_1 A_4^k) = [0, \tau^2)$ for $k \geq 1$. We must make sure that the boundary point $0 \in \partial E^\sharp(A_1) \cap \partial I_1$ satisfies Definition 5.2(iii), which is indeed the case since $t(A_1)(|a'0 + b'| \vee |c'0 + d'|) = (1 + \tau^{-1})(1 \vee |\tau'|) = 1 + \tau^{-1} > 1$. Since $\tau < \tau^2$, we have $I_1 \cup I_2 \cup I_3 = [0, \tau] \subset [0, \tau^2] =$ (closure of $E^\sharp(A_1 A_4^k)$), and all blocks of the form $14^*\{4\}^c$ are decreasing.

The number $t(A_2 A_4^k)$ has value τ for $k = 0, 1$, and value $(2k+1)(2-\tau)$ for $k \geq 2$. We have $E^\sharp(A_2) = (0, \xi(0))$ and $E^\sharp(A_1 A_4^k) = [0, \xi(k))$ for some number $\xi(k)$ which we do not need to compute since it is surely greater than τ; indeed, $f_{A_2 A_4^k}(\tau) = \tau + k/\tau + (k+1)$, which is strictly greater than $t(A_2 A_4^k)$ for every k. Again, $0 \in \partial E^\sharp(A_2) \cap \partial I_1$, so we must check the inequality in Definition 5.2(iii), that is, that $t(A_2)(1 \vee |\tau'|) > 1$, which is the case. We conclude that all blocks $24^*(4)^c$ are decreasing. Since $A_3 A_4^k$ is obtained from $A_2 A_4^k$ by exchanging rows and this operation, as remarked in the proof of Theorem 4.3, leaves t and f unaltered, all blocks $34^*\{4\}^c$ are decreasing as well. This concludes the proof of Theorem 5.3 for the map of case $(2, 5, \infty)$.

The particular form of the blocks $\{4\}^c 4^* \{4\}^c$ implies that, starting with any input in $[0, \tau] \cap \mathbb{Q}(\tau)$ and waiting until the T-orbit returns to $[0, \tau]$ guarantees that the height decreases strictly. Thus, the *first-return map* induced by T on $[0, \tau]$ (see Figure 6) is strictly height-decreasing. In particular, the first-return orbit of every point in $K = \mathbb{Q}(\tau)$ is eventually undefined: we have to wait forever.

Remark 5.6. In general, conjugating a map changes radically the structure of decreasing blocks. For example, conjugating the map T above by $x \mapsto \tau^{-1}x$, we obtain a Gauss map T' for which the block 11 is not decreasing anymore.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

5.2. *Cases* $(3, 4, \infty)$ *and* $(3, 5, \infty)$. These cases are identical to case $(2, 5, \infty)$. The set of all blocks of the form $\{5\}^c 5^* \{5\}^c$ (respectively, $\{7\}^c 7^* \{7\}^c$) is a complete family of decreasing blocks, and the first-return map on the first $r - 1$ intervals is strictly height-decreasing.

5.3. *Case* $(3, 6, \infty)$. This case is more involved. There are 10 intervals, the rightmost two being $I_9 = [\sqrt{3}, 1 + \sqrt{3}]$ and $I_{10} = [1 + \sqrt{3}, \infty]$. The blocks $1(10)^*\{10\}^c$ are height-decreasing (we appeal to the reader's patience for the extra parentheses, which are needed to avoid digit ambiguities). Now, computing $E^\sharp$ for the matrix

$$A_2 A_{10}^k = \begin{bmatrix} 1 & (k+1) + k\sqrt{3} \\ \sqrt{3} & (3k+1) + (k+1)\sqrt{3} \end{bmatrix}$$

yields an interval $[0, \xi(k))$. The number $\xi(k)$ is in I_{10} for $k = 0, \ldots, 16$ (see Figure 4 (left)), but enters I_9 from $k = 17$ on, converging monotonically from the right to $(3 + \sqrt{3})/2$, which is in the topological interior of I_9. It follows that the blocks $2(10)^k\{10\}^c$ are indeed decreasing for $k \le 16$, but no more so for higher k. We therefore include in $\mathcal{B}$ all blocks of the form $2(10)^k\{10\}^c$ for $k \le 16$ and of the form $2(10)^k 9(10)^*\{10\}^c$ for $k \ge 17$; computing the relative $E^\sharp$s shows that the latter blocks are decreasing. Identical considerations apply to blocks starting with the digits $a = 3, \ldots, 9$. Note that if we are satisfied with a coarser $\mathcal{B}$ (i.e., a set of decreasing blocks which, although complete, is non-optimal, as it contains blocks longer than necessary), then we may avoid computing the minimal k corresponding to each a, and just put in $\mathcal{B}$, in addition to $1(10)^*\{10\}^c$, all blocks of the form $\{1, 10\}^c(10)^*\{9, 10\}^c$ and $\{1, 10\}^c(10)^*9(10)^*\{10\}^c$.

In case $(3, 6, \infty)$, the first-return map to $I_1 \cup \cdots \cup I_9 = [0, 1 + \sqrt{3}]$ is *not* height-decreasing. The point

$$\alpha = \frac{-106298232 + 90029733\sqrt{3}}{-87526426 + 100742393\sqrt{3}}$$

is in I_2 and has logarithmic height $18.8328\ldots$. It gets mapped to I_{10} by T, and stays there for 16 parabolic steps, returning to I_9 as

$$\beta = A_2^{-1} * \alpha - 17(1 + \sqrt{3}) = \frac{-284852703 + 186193328\sqrt{3}}{-18821875 + 18821875\sqrt{3}},$$

which has logarithmic height $18.8341\ldots$.

We have thus proved Theorem 5.3 for the cases in §3.1. In the remaining cases both extremal intervals I_1 and I_r correspond to a parabolic matrix; note again that in the construction of $\mathcal{B}$ the role of the digits 1 and r is not symmetric.

5.4. *Cases* $(4, \infty, \infty)$, $(5, \infty, \infty)$, *and* $(6, \infty, \infty)$. These are similar, although not completely overlapping. The alphabets are $\mathfrak{A} = \{1, \ldots, r\}$, with r equal to 7, 9, and 11, respectively. In each case we provide a list of block families whose union gives a complete set of decreasing blocks. The completeness of the sets is apparent, and the fact that the blocks are decreasing reduces to a tedious but straightforward computation of the relative $E^\sharp$s.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

- $(4, \infty, \infty)$

$$11^*\{1, 7\}^c,$$
$$11^*77^*\{7\}^c,$$
$$\{1, 6, 7\}^c\ 7^*\{7\}^c,$$
$$6\{1, 7\}^c,$$
$$611^*\{1, 2, 7\}^c,$$
$$611^*\{2, 7\}7^*\{7\}^c,$$
$$677^*\{7\}^c.$$

- $(5, \infty, \infty)$

$$11^*\{1, 2, 9\}^c,$$
$$11^*\{2, 9\}9^*\{9\}^c,$$
$$\{1, 8, 9\}^c\ 9^*\{9\}^c,$$
$$8\{1, 9\}^c,$$
$$811^*\{1\}^c\ 9^*\{9\}^c,$$
$$899^*\{9\}^c.$$

- $(6, \infty, \infty)$

$$11^*\{1, 11\}^c,$$
$$11^*(11)(11)^*\{11\}^c,$$
$$\{1, 10, 11\}^c\ (11)^*\{11\}^c,$$
$$(10)\{1, 11\}^c,$$
$$(10)11^*\{1, 11\}^c,$$
$$(10)11^*(11)(11)^*\{11\}^c,$$
$$(10)(11)(11)^*\{11\}^c.$$

5.5. *Case* $(4, 12, \infty)$. This case involves the largest alphabet $\mathfrak{A} = \{1, \ldots, 33\}$ but it is not difficult, since all digits from 2 to 29 behave in the same way.

$$11^*\{1, 33\}^c,$$
$$11^*(33)(33)^*\{33\}^c,$$
$$\{1, 30, 31, 32, 33\}^c\ (33)^*\{33\}^c,$$
$$(30)\{1, 33\}^c,$$
$$(30)11^*(33)^*\{33\}^c,$$
$$(30)(33)(33)^*\{33\}^c,$$
$$\{31, 32\}\{1, 2, 3, 33\}^c,$$
$$\{31, 32\}11^*\{1, 33\}^c,$$
$$\{31, 32\}11^*(33)(33)^*\{33\}^c,$$
$$\{31, 32\}\{2, 3, 33\}(33)^*\{33\}^c.$$

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

5.6. *Case* $(4, 6, \infty)$. The alphabet is $\mathfrak{A} = \{1, \ldots, 15\}$. This is the most difficult case, and the only one in which a matrix, namely A_{13} of Example 4.2, has its $E^\flat$ equal to all of $[0, \infty]$; see Figure 4 (right). The closure of the set of points in $K = \mathbb{Q}(\sqrt{6})$ on which T is strictly increasing is also large, and includes all of I_{13}; however, further applications of T make up for these temporary increases. The following is a list of block families whose union provides a complete set of decreasing blocks.

$$11^*\{1, 15\}^c,$$
$$11^*(15)(15)^*\{15\}^c,$$
$$\{2, 3, 6, 7, 11\}1^*\{1\}^c(15)^*\{15\}^c,$$
$$\{4, 5, 8, 9\}(15)^*\{15\}^c,$$
$$(10)\{1, 2, 3, 4, 5, 6, 7, 15\}^c,$$
$$(10)11^*(15)(15)^*\{15\}^c,$$
$$(10)\{2, 3, 4, 5, 6, 7, 15\}(15)^*\{15\}^c,$$
$$(12)\{13, 14\},$$
$$(12)11^*\{1\}^c(15)^*\{15\}^c,$$
$$(12)\{2, 3, 4, 5, 8, 9, 10, 11, 12\}(15)^*\{15\}^c,$$
$$(12)\{6, 7\}1^*\{1, 15\}^c,$$
$$(12)\{6, 7\}1^*(15)(15)^*\{15\}^c,$$
$$(13)11^*\{1\}^c(15)^*\{15\}^c,$$
$$(13)\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15\}(15)^*\{15\}^c,$$
$$(13)\{12, 13\}1^*\{1, 15\}^c,$$
$$(13)\{12, 13\}1^*(15)(15)^*\{15\}^c,$$
$$(14)\{1, 2, 3, 4, 15\}^c,$$
$$(14)\{1, 2, 3\}1^*\{1, 15\}^c,$$
$$(14)\{1, 2, 3\}1^*(15)(15)^*\{15\}^c,$$
$$(14)\{4, 15\}(15)^*\{15\}^c.$$

This completes the proof of Theorem 5.3.

Acknowledgement. The main idea behind Theorem 4.3 was suggested to me by Umberto Zannier, whom I very sincerely thank. I also thank John Voight for a clarifying discussion on the results in [**19**], and the referee for the detailed reading and valuable suggestions. Some of the pictures and a few computations were done using SageMath.

References

[**1**] P. Arnoux and P. Hubert. Fractions continues sur les surfaces de Veech. *J. Anal. Math.* **81** (2000), 35–64.
[**2**] P. Arnoux and T. A. Schmidt. Veech surfaces with nonperiodic directions in the trace field. *J. Mod. Dyn.* **3**(4) (2009), 611–629.
[**3**] V. Baladi and B. Vallée. Euclidean algorithms are Gaussian. *J. Number Theory* **110**(2) (2005), 331–386.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press

[**4**] F. Berg. Dreiecksgruppen mit Spitzen in quadratischen Zahlkörpern. *Abh. Math. Sem. Univ. Hamburg* **55** (1985), 191–200.
[**5**] V. Berthé. Multidimensional Euclidean algorithms, numeration and substitutions. *Integers* **11B** (2011), Paper no. A2, 34pp.
[**6**] E. Bombieri and W. Gubler. *Heights in Diophantine Geometry*. Cambridge University Press, Cambridge, 2006.
[**7**] I. I. Bouw and M. Möller. Teichmüller curves, triangle groups, and Lyapunov exponents. *Ann. of Math. (2)* **172**(1) (2010), 139–185.
[**8**] K. Calta and T. A. Schmidt. Continued fractions for a class of triangle groups. *J. Aust. Math. Soc.* **93**(1–2) (2012), 21–42.
[**9**] M. Einsiedler and T. Ward. *Ergodic Theory with a View towards Number Theory*. Springer, Berlin, 2011.
[**10**] E. Girondo, D. Torres-Teigell and J. Wolfart. Shimura curves with many uniform dessins. *Math. Z.* **271**(3–4) (2012), 757–779.
[**11**] L. Greenberg. Maximal Fuchsian groups. *Bull. Amer. Math. Soc.* **69** (1963), 569–573.
[**12**] W. P. Hooper. Grid graphs and lattice surfaces. *Int. Math. Res. Not. IMRN* **2013**(12) (2013), 2657–2698.
[**13**] S. Katok. *Fuchsian Groups*. University of Chicago Press, Chicago, 1992.
[**14**] S. Lang. *Algebra*, 3rd edn. Addison-Wesley, Reading, MA, 1993.
[**15**] A. Leutbecher. Über die Heckeschen Gruppen $G(\lambda)$. II. *Math. Ann.* **211** (1974), 63–86.
[**16**] C. Maclachlan and A. W. Reid. *The Arithmetic of Hyperbolic 3-Manifolds*. Springer, New York, 2003.
[**17**] G. A. Margulis. Discrete groups of motions of manifolds of nonpositive curvature. *Proc. Int. Congress of Mathematicians (Vancouver, BC, 1974). Amer. Math. Soc. Transl. Series 2* **109** (1977), 33–45.
[**18**] C. T. McMullen. Teichmüller geodesics of infinite complexity. *Acta Math.* **191**(2) (2003), 191–223.
[**19**] S. Nugent and J. Voight. On the arithmetic dimension of triangle groups. *Math. Comp.* **86**(306) (2017), 1979–2004.
[**20**] G. Panti. A general Lagrange theorem. *Amer. Math. Monthly* **116**(1) (2009), 70–74.
[**21**] G. Panti. Slow continued fractions, transducers, and the Serret theorem. *J. Number Theory* **185** (2018), 121–143.
[**22**] F. Schweiger. *Multidimensional Continued Fractions*. Oxford University Press, Oxford, 2000.
[**23**] F. Seibold. Zahlentheoretische Eigenschaften der Heckeschen Gruppen $G(\lambda)$ und verwandter Transformationsgruppen. *Inauguraldissertation*, Techn. Univ. München, 1985.
[**24**] C. Series. The modular surface and continued fractions. *J. Lond. Math. Soc. (2)* **31**(1) (1985), 69–80.
[**25**] D. Singerman. Finitely maximal Fuchsian groups. *J. Lond. Math. Soc. (2)* **6** (1972), 29–38.
[**26**] J. Smillie and C. Ulcigrai. Geodesic flow on the Teichmüller disk of the regular octagon, cutting sequences and octagon continued fractions maps. *Dynamical Numbers—Interplay between Dynamical Systems and Number Theory*. American Mathematical Society, Providence, RI, 2010, pp. 29–65.
[**27**] J. Smillie and C. Ulcigrai. Beyond Sturmian sequences: coding linear trajectories in the regular octagon. *Proc. Lond. Math. Soc. (3)* **102**(2) (2011), 291–340.
[**28**] K. Takeuchi. Arithmetic triangle groups. *J. Math. Soc. Japan* **29**(1) (1977), 91–106.
[**29**] K. Takeuchi. Commensurability classes of arithmetic triangle groups. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **24**(1) (1977), 201–212.
[**30**] W. A. Veech. Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* **97**(3) (1989), 553–583.
[**31**] C. C. Ward. Calculation of Fuchsian groups associated to billiards in a rational triangle. *Ergod. Th. & Dynam. Sys.* **18**(4) (1998), 1019–1042.

https://doi.org/10.1017/etds.2018.55 Published online by Cambridge University Press