DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS
Volume **40**, Number **7**, July **2020**

doi:10.3934/dcds.2020183

pp. **4341–4378**

BILLIARDS ON PYTHAGOREAN TRIPLES AND THEIR MINKOWSKI FUNCTIONS

Giovanni Panti

Department of Mathematics, Computer Science and Physics
University of Udine
via delle Scienze 206
33100 Udine, Italy

(Communicated by Jairo Bochi)

Abstract. It has long been known that the set of primitive pythagorean triples can be enumerated by descending certain ternary trees. We unify these treatments by considering hyperbolic billiard tables in the Poincaré disk model. Our tables have $m \geq 3$ ideal vertices, and are subject to the restriction that reflections in the table walls are induced by matrices in the triangle group $\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$. The resulting billiard map $\widetilde{B}$ acts on the de Sitter space $x_1^2+x_2^2-x_3^2=1$, and has a natural factor B on the unit circle, the pythagorean triples appearing as the B-preimages of fixed points. We compute the invariant densities of these maps, and prove the Lagrange and Galois theorems: A complex number of unit modulus has a preperiodic (purely periodic) B-orbit precisely when it is quadratic (and isolated from its conjugate by a billiard wall) over $\mathbb{Q}(i)$.

Each B as above is a $(m-1)$-to-1 orientation-reversing covering map of the circle, a property shared by the group character $T(z)=z^{-(m-1)}$. We prove that there exists a homeomorphism Φ, unique up to postcomposition with elements in a dihedral group, that conjugates B with T; in particular Φ —whose prototype is the classical Minkowski question mark function— establishes a bijection between the set of points of degree ≤ 2 over $\mathbb{Q}(i)$ and the torsion subgroup of the circle. We provide an explicit formula for Φ, and prove that Φ is singular and Hölder continuous with exponent $\log(m-1)$ divided by the maximal periodic mean free path in the associated billiard table.

1. **Introduction.** Rational points in the real projective line $\mathrm{P}^1\,\mathbb{R}$ involve two integers, a numerator and a denominator; we can enumerate them by reversing the euclidean algorithm or —equivalently— taking inverse branches of continued fraction maps. Rational points in the unit circle S^1 involve three integers, the two legs and the hypotenuse of a pythagorean triangle. As the line and the circle can be mutually parametrized with preservation of rational points, the complexity of the enumeration is the same, and there is a line of work (starting from [6], and running through [4], [11], [3], [33], [15] and references therein) describing how pythagorean triples can be generated by descending trees.

2010 *Mathematics Subject Classification.* Primary: 37D40, 11J70.

Key words and phrases. Pythagorean triples, Romik map, billiards, Minkowski function, joint spectral radius.

The author is partially supported by the research project SiDiA of the University of Udine.

Ascending the same trees amounts to iterating continued fraction maps, and in [42] Romik analyzes one such map, relating it to the geodesic flow on the three-punctured sphere. It turns out that Romik's map can also be seen as the Gauss map of even continued fractions; see [2, §4], [15, §5], [7, §2] for various developments.

Although there is a birational bijection with rational coefficients between the line and the circle, continued fraction maps on the two spaces are not exactly the same thing. Indeed, the rational symmetry group of the projective line is the extended modular group $\mathrm{PSL}_2^{\pm}\mathbb{Z}$, while that of the circle is $\mathrm{SO}_{2,1}\mathbb{Z}$, the stabilizer of the Lorentz form inside $\mathrm{SL}_3\mathbb{Z}$. When embedded in a larger ambient group —say $\mathrm{PSL}_2^{\pm}\mathbb{R}$— they appear as the $(2,3,\infty)$ and the $(2,4,\infty)$ extended triangle groups, and neither is a subgroup of the other (of course, they are commensurable).

In this paper we develop continued fraction maps (of the "slow" type, that is with parabolic fixed points) directly on the circle, as factors of billiard maps determined by ideal polygons in the hyperbolic plane. We summarize our main results as follows:

- Let D be a polygon in the Poincaré disk having $m \geq 3$ vertices, all at the boundary at infinity S^1. Let $B : S^1 \to S^1$ be the map that sends the interval between two vertices to the union of the remaining intervals via reflection in the corresponding polygon side. Let T be the group character $z \mapsto z^{-(m-1)}$. Then B and T are conjugate by an essentially unique homeomorphism Φ, which provides a bijection between the set of points of degree at most 2 over $\mathbb{Q}(i)$ and the torsion subgroup of S^1. The homeomorphism Φ is singular and Hölder continuous, of exponent $\log(m-1)$ divided by the maximal mean free path (see Definition 10.3) of periodic trajectories in the hyperbolic billiard determined by D.

The route leading to the above statement is somehow long; we offer two justifications.

1. The end result is a flexible and applicable tool. Indeed, the maximal mean free path referred to above equals twice the logarithm of the joint spectral radius of the set Σ of matrices expressing reflections in the billiard walls. When the vertices of D determine a unimodular partition of S^1 (an arithmetical condition explained in §5), this joint spectral radius can often be explicitly computed; see Example 10.6.
2. Along that route we encounter fair landscapes.

We describe our route: in §2 we determine finite sets of reflections generating the orthogonal group $\mathrm{O}_{2,1}\mathbb{Z}$ and its subgroups $\mathrm{SO}_{2,1}\mathbb{Z}$ and $\mathrm{O}^{\uparrow}_{2,1}\mathbb{Z}$, the latter being the stabilizer of the upper sheet of the hyperboloid $x_1^2 + x_2^2 - x_3^2 = -1$. Then, as a warmup, in §3 we review the construction of the Romik map using our formalism. In §4 we provide explicit $\mathrm{PSL}_2^{\pm}\mathbb{R}$-equivariant bijections between the homogeneous space $\mathrm{PSL}_2\mathbb{R}/\{\text{diagonal matrices}\}$, the de Sitter space $x_1^2 + x_2^2 - x_3^2 = 1$, the space of oriented geodesics in the hyperbolic plane, and that of quadratic forms of discriminant 1. These correspondences are known, but since they appear scattered in the literature and some care is required to extend the acting group from the usual $\mathrm{PSL}_2\mathbb{R}$ to the full $\mathrm{PSL}_2^{\pm}\mathbb{R}$, our brief self-contained treatment in Theorem 4.1 may have some value. In §5 we treat unimodular partitions of the circle; a reader not interested in arithmetical issues may safely skip Theorems 5.3 and 5.5.

The preliminaries being over, we introduce in §6 our continued fraction maps B as factors of billiard maps $\widetilde{B}$ associated to ideal polygons whose vertices form a unimodular partition of the circle. Reflections in the table walls are expressed by

elements of $\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$ —which we naturally take as matrices— in the Poincaré model, and by matrices in $\mathrm{O}^{\uparrow}_{2,1}\,\mathbb{Z}$ in the Klein model. Here the de Sitter space plays a twofold rôle, as the phase space of $\widetilde{B}$ as well as the space of shrinking intervals, this double nature being reflected in a double action of $\mathrm{PSL}^{\pm}_2\,\mathbb{R}$; see Remark 5.2. In §7 we use the bijections in §4 to characterize the natural extension and the absolutely continuous invariant measure of B. In §8 we show that the map B and the extended fuchsian group generated by Σ are orbit-equivalent, and prove the following statement, which combines the classical Lagrange and Galois theorems. A complex number of unit modulus is quadratic over $\mathbb{Q}(i)$ if and only if its B-orbit is eventually periodic; moreover, if this is the case, then the conjugate point has the reverse period, and the two points are purely periodic precisely when they are separated by a billiard wall.

In §9 we introduce the conjugacy alluded to above. It is a natural conjugacy; indeed, B is an $(m-1)$-to-1 orientation-reversing covering map of the circle, a topological property shared by precisely one group character, namely $T(z) = z^{-(m-1)}$. We thus have a "linearized" version of a continued fraction map, precisely as the tent map on $[0,1]$ is a linearized version of the Farey map. It turns out (Lemma 9.3) that the natural symbolic coding of points via B, as well as the analogous coding via T, characterizes the ternary betweenness relation on the circle. Since the latter relation determines the circle topology, we obtain in Theorem 9.2 that B and T are conjugate by a homeomorphism Φ, unique up to postcomposition with elements of the dihedral group with $2m$ elements. This homeomorphism is the analogue of the classical Minkowski question mark function [19], [43], [27], which conjugates the Farey map with the tent map. We provide in Theorem 9.4 an explicit expression for Φ analogous to the Denjoy-Salem formula [43, p. 436] for the question mark function, and show in Examples 8.4 and 9.5 how the arithmetic properties of B and T are intertwined by Φ. In Theorem 10.1 we provide an ergodic-theoretic proof of the fact that Φ has zero derivative at Lebesgue-all points.

In the final Theorem 10.5 we complete the proof of the connection sketched above between the joint spectral radius of Σ and the Hölder exponent of Φ. In all instances we examined the Lagarias-Wang finiteness conjecture ([31], see §10) turned out to be true for Σ, and a maximizing periodic billiard trajectory was easily guessed and verified. It is plausible that the conjecture holds for all billiard tables determined by unimodular partitions of the circle, and we leave this as an interesting open problem.

2. **Notation and preliminaries.** Since we treat various spaces of matrices, we will distinguish them notationally, by using boldface for 3×3 matrices and lightface for 2×2 ones. Points in $\mathbb{R}^3$ are written in boldface and are always column vectors, although we may write $\boldsymbol{x} = (x_1, x_2, x_3)$ for typographical reasons. We will use square or round brackets for vectors and matrices, according whether we are in a projective setting (that is, up to multiplication by nonzero scalars) or in a linear-algebra one. Zero entries in matrices are replaced by blank spaces.

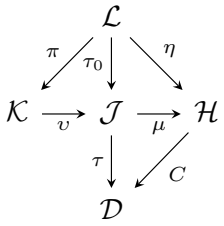
Let

$$\boldsymbol{L} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}$$

be the matrix of the three-variable Lorentz quadratic form, and let $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\top}\boldsymbol{L}\boldsymbol{y}$ be the corresponding symmetric bilinear map. The upper sheet $\mathcal{L} = \{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{x}\rangle =$

$-1,\, x_3 > 0\}$ of the 2-sheeted hyperboloid $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = -1$ is one of the standard models of the hyperbolic plane, other models being the upper halfplane $\mathcal{H} = \{z \in \mathbb{C} : \operatorname{im} z > 0\}$, the Klein disk $\mathcal{K} = \{[x_1, x_2, x_3] \in \mathrm{P}^2\,\mathbb{R} : x_1^2 + x_2^2 < x_3^2\}$, and the Poincaré disk $\mathcal{D} = \{z \in \mathbb{C} : |z| < 1\}$; we refer the reader to [10] for an enjoyable introduction to hyperbolic geometry. We need explicit bijections between these models, so we introduce a fifth auxiliary model, namely the upper hemisphere $\mathcal{J} = \{\boldsymbol{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1,\, x_3 > 0\}$, and state a lemma.

Lemma 2.1. *The spaces $\mathcal{L}$, $\mathcal{H}$, $\mathcal{K}$, $\mathcal{D}$, $\mathcal{J}$ are in bijective correspondence via the commuting diagram*



where

- $\pi : \mathbb{R}^3 \setminus \{0\} \to \mathrm{P}^2\,\mathbb{R}$ *is the natural quotient map,*
- τ_0 *is the stereographic projection through* $(0, 0, -1)$*,*
- $\eta(\boldsymbol{x}) = (x_1 + i)/(x_3 - x_2)$*,*
- $\upsilon([x_1, x_2, x_3]) = \big(x_1/x_3, x_2/x_3, (x_3^2 - x_1^2 - x_2^2)^{1/2}/x_3\big)$ *is the "vertical" projection,*
- μ *is the stereographic projection through* $(0, 1, 0)$ *to the halfplane* $\{x_2 = 0, x_3 > 0\}$*, followed by the obvious identification of the latter with* $\mathcal{H}$*,*
- τ *is the stereographic projection through* $(0, 0, -1)$ *to the disk* $\{x_1^2 + x_2^2 < 1, x_3 = 0\}$*, followed by the obvious identification of the latter with* $\mathcal{D}$*,*
- C *is the Möbius transformation* $z \mapsto C * z = (z - i)/(-iz + 1)$ *induced by the Cayley matrix* $C = 2^{-1/2}\left[\begin{smallmatrix} 1 & -i \\ -i & 1 \end{smallmatrix}\right] \in \mathrm{PSL}_2\,\mathbb{C}$ *(as customary, we blur the distinction between matrices and the maps they induce).*

These correspondences extend to the respective ideal boundaries.

Proof. The proof reduces to a commentary on the figure on page 70 of [10]. The upper-left triangle commutes because $\upsilon\circ\pi$ sends $\boldsymbol{x} = (x_1, x_2, x_3) \in \mathcal{L}$ to $\big(x_1, x_2, (x_3^2 - x_1^2 - x_2^2)^{1/2}\big)/x_3 = (x_1, x_2, 1)/x_3 = (1/x_3)(x_1, x_2, x_3) + (1 - 1/x_3)(0, 0, -1)$. The upper-right triangle commutes because μ sends $(x_1, x_2, 1)/x_3 \in \mathcal{J}$ to $x_1/(x_3 - x_2) + i/(x_3 - x_2) = \eta(\boldsymbol{x})$. The lower-right triangle commutes because, given $\boldsymbol{y} \in \mathcal{J}$,

$$C * \big(\mu(\boldsymbol{y})\big) = \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} * \frac{y_1 + y_3 i}{1 - y_2} = \frac{y_1 + y_2 i}{1 + y_3} = \tau(\boldsymbol{y}).$$

The fact that these correspondences extend to the ideal boundaries is obvious as soon as the boundary $\partial\mathcal{L}$ of $\mathcal{L}$ and the maps π, τ_0, η on it are properly defined. We see $\partial\mathcal{L}$ as the intersection of the projective closure of $\mathcal{L} \cup (-\mathcal{L})$ (i.e., the variety $x_1^2 + x_2^2 - x_3^2 + x_4^2 = 0$ in $\mathrm{P}^3\,\mathbb{R}$) with the plane at infinity $x_4 = 0$, and we set

$$\begin{aligned} \pi([x_1, x_2, x_3, 0]) &= [x_1, x_2, x_3], \\ \tau_0([x_1, x_2, x_3, 0]) &= (x_1/x_3, x_2/x_3, 0), \\ \eta([x_1, x_2, x_3, 0]) &= x_1/(x_3 - x_2). \end{aligned}$$

We can then view $[x_1, x_2, x_3, 0] \in \partial\mathcal{L}$ as the limit (in the euclidean metric of an appropriate local chart) of $\boldsymbol{x}(t) = t\big(x_1, x_2, (x_1^2 + x_2^2 + 1/t^2)^{1/2}\big) \in \mathcal{L}$, for $t \to +\infty$. An easy computation shows that the π-, τ_0-, η-images of $[x_1, x_2, x_3, 0] \in \partial\mathcal{L}$, as defined above, agree with the limits (in the euclidean metric) of $\pi(\boldsymbol{x}(t))$, $\tau_0(\boldsymbol{x}(t))$, $\eta(\boldsymbol{x}(t))$, for $t \to +\infty$. This guarantees the required commutativity. □

It is well known that the orthogonal group $\mathrm{O}_{2,1}\,\mathbb{R}$ of the Lorentz form has four connected components, namely the component of the identity (which is a normal subgroup) and its cosets with respect to the diagonal matrices having diagonal entries $(-1,1,1)$, $(1,1,-1)$, $(-1,1,-1)$. The union of the component of the identity with its $(-1,1,-1)$-coset is the special orthogonal group $\mathrm{SO}_{2,1}\,\mathbb{R}$, while its union with the $(-1,1,1)$-coset is the group $\mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R}$ of all matrices that preserve $\mathcal{L}$; equivalently, $\mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R} = \{\boldsymbol{A} \in \mathrm{O}_{2,1}\,\mathbb{R} : \text{the } (3,3)\text{-entry of } \boldsymbol{A} \text{ is } > 0\}$. We will write $\mathrm{SO}^{\uparrow}_{2,1}\,\mathbb{R} = \mathrm{SO}_{2,1}\,\mathbb{R} \cap \mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R}$ for the component of the identity.

The group of isometries (including the orientation-reversing ones) of $\mathcal{H}$ is $\mathrm{PSL}_2^{\pm}\,\mathbb{R}$ $= \{A \in \mathrm{GL}_2\,\mathbb{R} : |\det A| = 1\}/\{\pm I\}$, which acts on $\mathcal{H}$ as follows: given $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, then $A * z$ equals $(az+b)/(cz+d)$ if $\det A = 1$, and equals $(a\bar{z}+b)/(c\bar{z}+d)$ if $\det A = -1$. Conjugating $\mathrm{PSL}_2^{\pm}\,\mathbb{R}$ with the Cayley matrix we obtain the group

$$\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{C} = \left\{ \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \in \mathrm{GL}_2\,\mathbb{C} : \big||\alpha|^2 - |\beta|^2\big| = 1 \right\} \Big/ \pm I,$$

which acts on $\mathcal{D}$ via

$$\begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix} * z = \begin{cases} (\alpha z + \beta)/(\bar{\beta} z + \bar{\alpha}), & \text{if } |\alpha|^2 - |\beta|^2 = 1; \\ (\beta \bar{z} + \alpha)/(\bar{\alpha}\bar{z} + \bar{\beta}), & \text{if } |\alpha|^2 - |\beta|^2 = -1. \end{cases}$$

We construct an isomorphic representation $\mathrm{PSL}_2^{\pm}\,\mathbb{R} \to \mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R}$ by identifying the vector $\boldsymbol{w} = (w_1, w_2, w_3) \in \mathbb{R}^3$ with the matrix

$$W = \begin{pmatrix} -w_2 + w_3 & -w_1 \\ -w_1 & w_2 + w_3 \end{pmatrix}, \tag{1}$$

on which $A \in \mathrm{PSL}_2^{\pm}\,\mathbb{R}$ acts on the left by $W \mapsto (A^{-1})^{\top} W A^{-1}$. This is a well defined action, independent from the lift of A to $\mathrm{SL}_2^{\pm}\,\mathbb{R}$, linear, and preserving the form $\langle \boldsymbol{w}, \boldsymbol{w} \rangle = -\det W$. Computing the images of the 1-parameter subgroups in the Iwasawa decomposition of $\mathrm{PSL}_2\,\mathbb{R}$ provides a geometric picture of the representation, namely

$$\begin{aligned} \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} &\mapsto \begin{pmatrix} \cos(-2t) & -\sin(-2t) & \\ \sin(-2t) & \cos(-2t) & \\ & & 1 \end{pmatrix}, \\ \begin{bmatrix} \exp(t/2) & \\ & \exp(-t/2) \end{bmatrix} &\mapsto \begin{pmatrix} 1 & & \\ & \cosh(t) & \sinh(t) \\ & \sinh(t) & \cosh(t) \end{pmatrix}, \\ \begin{bmatrix} 1 & t \\ & 1 \end{bmatrix} &\mapsto \begin{pmatrix} 1 & -t & t \\ t & 1 - t^2/2 & t^2/2 \\ t & -t^2/2 & 1 + t^2/2 \end{pmatrix}. \end{aligned} \tag{2}$$

Convention 2.2. In order to simplify notation we adopt the convention that, whenever a matrix in $\mathrm{PSL}_2^{\pm}\,\mathbb{R}$ is denoted by a certain capital letter, then its image under

the above representation, and its C-conjugate, are denoted by the same capital letter in bold and in calligraphic fonts, respectively. With this understanding, we give names to a few matrices that will recur throughout this paper.

$$\begin{aligned}
J &= \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}, & \mathscr{J} &= \begin{bmatrix} & -i \\ i & \end{bmatrix}, & \boldsymbol{J} &= \begin{pmatrix} -1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \\
F &= \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, & \mathscr{F} &= \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, & \boldsymbol{F} &= \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix}, \\
P &= \begin{bmatrix} -1 & 2 \\ & 1 \end{bmatrix}, & \mathscr{P} &= \begin{bmatrix} i & 1-i \\ 1+i & -i \end{bmatrix}, & \boldsymbol{P} &= \begin{pmatrix} -1 & -2 & 2 \\ -2 & -1 & 2 \\ -2 & -2 & 3 \end{pmatrix}, \\
G &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, & \mathscr{G} &= \frac{1}{\sqrt{2}} \begin{bmatrix} & 1+i \\ 1-i & \end{bmatrix}, & \boldsymbol{G} &= \begin{pmatrix} & 1 & \\ 1 & & \\ & & 1 \end{pmatrix}.
\end{aligned} \tag{3}$$

Explicit computation —which we omit— shows that $\eta \circ \boldsymbol{A} = A \circ \eta$ on $\mathcal{L}$, for every A in the above 1-parameter subgroups, and also for $A = J$; therefore the identity $\eta \circ \boldsymbol{A} = A \circ \eta$ holds for every $A \in \mathrm{PSL}_2^{\pm}\mathbb{R}$. The action of $\mathrm{O}_{2,1}^{\uparrow}\mathbb{R}$ on $\mathbb{R}^3$ descends to a projective action on $\mathrm{P}^2\mathbb{R}$ that fixes the Klein model $\mathcal{K}$ and its boundary $\partial\mathcal{K}$. These observations, together with Lemma 2.1, imply that for every $A \in \mathrm{PSL}_2^{\pm}\mathbb{R}$ the diagram

$$\begin{array}{ccccc}
\mathcal{K} & \xrightarrow{\tau\circ\upsilon} & \mathcal{D} & \xleftarrow{C} & \mathcal{H} \\
{\scriptstyle \boldsymbol{A}}\downarrow & & {\scriptstyle \mathscr{A}}\downarrow & & \downarrow{\scriptstyle A} \\
\mathcal{K} & \xrightarrow[\tau\circ\upsilon]{} & \mathcal{D} & \xleftarrow[C]{} & \mathcal{H}
\end{array} \tag{4}$$

commutes. The analogous diagram involving the ideal boundaries of $\mathcal{K}, \mathcal{D}, \mathcal{H}$ commutes as well, and actually simplifies. Indeed, the nontrivial bijection $\tau \circ \upsilon$ reduces on $\partial\mathcal{K}$ to the obvious identification $[x_1, x_2, x_3] \mapsto (x_1 + x_2 i)/x_3$, while $C^{-1} \circ \tau \circ \upsilon$ reduces to the stereographic projection through $[0, 1, 1]$, namely $[x_1, x_2, x_3] \mapsto x_1/(x_3 - x_2)$. We will thus switch freely between $\partial\mathcal{K}$ and $\partial\mathcal{D}$, using S^1 as a neutral name for both.

Let D be a polygon in $\mathcal{H}$, bounded by $m \geq 3$ geodesics $l_0, \dots, l_{m-1}$, and having angles at vertices $\pi/e_0, \dots, \pi/e_{m-1}$, with $e_0, \dots, e_{m-1}$ integers ≥ 2 or ∞ (if the corresponding vertex lies in $\partial\mathcal{H}$); the Gauss-Bonnet formula forces $m - 2 > \sum_a e_a^{-1}$. The *extended Coxeter group* associated to D is the subgroup $\Gamma^{\pm}$ of $\mathrm{PSL}_2^{\pm}\mathbb{R}$ generated by the reflections in the sides of D. It has the presentation

$$\langle x_0, \dots, x_{m-1} | x_0^2 = \dots = x_{m-1}^2 = (x_0 x_1)^{e_0} = \dots = (x_{m-1} x_0)^{e_{m-1}} = 1 \rangle$$

(with the understanding that relators $(x_a x_{a+1})^{\infty}$ do not appear), and D is a fundamental domain for it. Its index-2 subgroup of orientation-preserving elements $\Gamma = \Gamma^{\pm} \cap \mathrm{PSL}_2\mathbb{R}$ is a fuchsian group of finite covolume; see [28], [32]. When D is a triangle we write $\Delta(e_0, e_1, e_2)$ and $\Delta^{\pm}(e_0, e_1, e_2)$ for Γ and $\Gamma^{\pm}$, referring to them as a *triangle group* and an *extended triangle group*, respectively (the adjective *extended* stresses the fact that orientation-reversing isometries are allowed; in both cases, the action on $\mathcal{H}$ is properly discontinuous). Note that the numbers e_0, e_1, e_2 determine the triangle up to isometry, and hence the groups up to conjugation. We

will freely use all of the above terminology when working in other models of the hyperbolic plane.

Let us return to the Lorentz form $\langle -,-\rangle$. We recall that, given a nonisotropic vector $\boldsymbol{w}$, the *reflection* $\boldsymbol{R_w}$ is the unique linear involution of $\mathbb{R}^3$ that fixes pointwise the polar hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{w}, \boldsymbol{x}\rangle = 0\}$ and exchanges $\boldsymbol{w}$ with $-\boldsymbol{w}$. An easy computation (of course, all of this is well known) shows that:

(i)
$$\boldsymbol{R_w}(\boldsymbol{x}) = \boldsymbol{x} - \frac{2\langle \boldsymbol{w}, \boldsymbol{x}\rangle}{\langle \boldsymbol{w}, \boldsymbol{w}\rangle}\boldsymbol{w}, \tag{5}$$

(ii) $\boldsymbol{R_w}$ preserves $\langle -,-\rangle$,

(iii) in terms of matrices,
$$\boldsymbol{R_w} = \boldsymbol{I} - \frac{2}{\langle \boldsymbol{w}, \boldsymbol{w}\rangle}\boldsymbol{w}\,\boldsymbol{w}^{\top}\boldsymbol{L}, \tag{6}$$

(iv) $\boldsymbol{R_w} \in \mathrm{O}^{\uparrow}_{2,1}\mathbb{R}$ if and only if $\langle \boldsymbol{w}, \boldsymbol{w}\rangle > 0$.

Notation 2.3.
- $\mathrm{O}_{2,1}\mathbb{Z}$ (respectively, $\mathrm{SO}_{2,1}\mathbb{Z}$, $\mathrm{O}^{\uparrow}_{2,1}\mathbb{Z}$, $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z}$) is the intersection of $\mathrm{O}_{2,1}\mathbb{R}$ (respectively, $\mathrm{SO}_{2,1}\mathbb{R}$, $\mathrm{O}^{\uparrow}_{2,1}\mathbb{R}$, $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{R}$) with $\mathrm{GL}_3\mathbb{Z}$.
- $\mathrm{PSL}^{\pm}_2\mathbb{Z} = \{A \in \mathrm{PSL}^{\pm}_2\mathbb{R} : A \text{ has entries in } \mathbb{Z}\}$.
- $\mathrm{PSU}^{\pm}_{1,1}\mathbb{Z}[i] = \{\mathscr{A} \in \mathrm{PSU}^{\pm}_{1,1}\mathbb{C} : \mathscr{A} \text{ has entries in } \mathbb{Z}[i]\}$.
- $\langle F, P, G\rangle^{+}$ (and analogously for other groups generated by involutions) is the group of all products of an even number of elements in $\{F, P, G\}$.

The four matrices $\boldsymbol{J}, \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{G}$ in (3) are in $\mathrm{O}^{\uparrow}_{2,1}\mathbb{Z}$; in particular they are of the form $\boldsymbol{R_w}$, for $\boldsymbol{w}$ equal to $(1,0,0)$, $(0,1,0)$, $(1,1,1)$, $(-1,1,0)$, respectively. In [35] it is proved that the five reflections $\boldsymbol{J}$, $\boldsymbol{F}$, $\boldsymbol{R}_{(0,0,1)} = \mathrm{diag}(1,1,-1)$, $\boldsymbol{R}_{(1,1,0)} = \boldsymbol{JGJ}$, $\boldsymbol{P}$ generate $\mathrm{O}_{2,1}\mathbb{Z}$ (see [17] for an elementary proof which avoids the theory of Kac-Moody Lie algebras); we give an independent and expanded version in the following theorem.

Theorem 2.4. *We have* $\mathrm{O}^{\uparrow}_{2,1}\mathbb{Z} = \langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{G}\rangle$*, which is isomorphic to the extended triangle group* $\Delta^{\pm}(2,4,\infty)$*; adding* $\boldsymbol{R}_{(0,0,1)}$ *as a further generator we obtain the full group* $\mathrm{O}_{2,1}\mathbb{Z}$*. The group* $\langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{J}\rangle$ *is an index-2 subgroup of* $\mathrm{O}^{\uparrow}_{2,1}\mathbb{Z}$*, and equals* $\Delta^{\pm}(2,\infty,\infty)$*; its image* $\langle \mathscr{F}, \mathscr{P}, \mathscr{J}\rangle$ *inside* $\mathrm{PSU}^{\pm}_{1,1}\mathbb{C}$ *is* $\mathrm{PSU}^{\pm}_{1,1}\mathbb{Z}[i]$*.*

Proof. We work in $\mathcal{H}$. Let $\Gamma = \{A \in \mathrm{PSL}_2\mathbb{R} : \boldsymbol{A} \in \mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z}\}$; then, by definition, Γ is an arithmetic fuchsian group. We observe that $\langle F, P, G\rangle^{+}$ is the triangle group $\Delta(2,4,\infty)$. Indeed F, P, G are the reflections in the three geodesics

- l_0, whose endpoints are 1 and -1;
- l_1, whose endpoints are ∞ and 1;
- l_2, whose endpoints are $1-\sqrt{2}$ and $1+\sqrt{2}$.

These geodesics determine a triangle D in $\mathcal{H}$ with vertices at $1+i\sqrt{2}$ with angle $\pi/2$, at i with angle $\pi/4$, and at the ideal point 1 with angle 0.

Clearly $\langle F, P, G\rangle^{+}$ is a subgroup of Γ, and it is well-known that a fuchsian group containing a triangle group must itself be a triangle group [44, §6]. The partially ordered set of all nine non-cocompact arithmetic triangle groups has been determined by Takeuchi in [46], and $\Delta(2,4,\infty)$ is maximal in it; therefore $\Gamma = \langle F, P, G\rangle^{+}$. Adding F as a further generator to $\langle F, P, G\rangle^{+}$ we obtain $\langle F, P, G\rangle = \{A \in \mathrm{PSL}^{\pm}_2\mathbb{R} : \boldsymbol{A} \in \mathrm{O}^{\uparrow}_{2,1}\mathbb{Z}\}$, as claimed.

For the second statement, observe that replacing the generator G with J means replacing l_2 with the geodesic l_2' whose endpoints are 0 and ∞. The polygon determined by l_0, l_1, l_2' is the triangle $D' = D \cup G[D]$, with angles $\pi/2$ at i, and 0 at 1 and at ∞; hence $\langle F, P, J\rangle$ is the extended triangle group $\Delta^{\pm}(2,\infty,\infty)$. Clearly $\langle \mathscr{F}, \mathscr{P}, \mathscr{J}\rangle \leq \mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$, and by computing

$$C^{-1}\begin{bmatrix} a+bi & c+di \\ c-di & a-bi \end{bmatrix} C = \begin{bmatrix} a+d & b+c \\ -b+c & a-d \end{bmatrix},$$

we see that $C^{-1}\big(\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]\big)C$ is a subgroup of $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$. Taking into account the respective fundamental domains, it is easy to check that $\langle F, P, J\rangle$ has index 3 in $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$; therefore $C^{-1}\big(\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]\big)C$ equals either $\langle F, P, J\rangle$ or the full $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$. However, this second possibility is ruled out by the fact that $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$ (which is the extended $(2,3,\infty)$ triangle group) contains elements of order 3, and hence of trace 1 (up to sign), while clearly no element of $\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$ may have trace 1. □

3. **Pythagorean triples and the Romik map.** A *[primitive] pythagorean triple* is a point $\boldsymbol{t} = (t_1, t_2, t_3) \in \mathbb{Z}^3$ such that $t_3 > 0$, $\gcd(t_1, t_2, t_3) = 1$, and $t_1^2 + t_2^2 = t_3^2$. Pythagorean triples correspond bijectively to rational points in the unit circle, which in turn correspond, via stereographic projection, to points in $P^1\mathbb{Q}$. These correspondences provide various techniques for enumerating triples, among which the one known to Euclid: given any reduced fraction a/b, the triple $(a^2-b^2, 2ab, a^2+b^2)/\gcd(a^2-b^2, 2ab, a^2+b^2)$ is pythagorean, and every pythagorean triple is uniquely obtainable in this way (the gcd in the denominator is 1 if $2 \mid ab$, and 2 otherwise). As noted in the introduction, many techniques are cast in the form of the descent of a binary or ternary tree.

A remarkable connection with the theory of continued fractions is offered in [42]; as a warmup, we sketch it using our notation. We partition S^1 in four quarters I_0, I_1, I_2, I_3, with $I_a = \{\exp(2\pi t i) : a/4 \leq t \leq (a+1)/4\}$. Let $\boldsymbol{A} = \boldsymbol{R}_{(1,-1,1)} = \boldsymbol{FPF}$. Then $\boldsymbol{A}$ acts on S^1 (viewed as $\partial\mathcal{K}$, see the diagram (4) and the resulting identifications) by exchanging $\boldsymbol{x}$ with the other point of intersection of S^1 with the line through $\boldsymbol{x}$ and $[1,-1,1]$; the interval I_3 is thus bijectively mapped to the union of the other three intervals. We fold back $I_0 \cup I_1 \cup I_2$ to I_3 via the reflection $\boldsymbol{F}$ acting on I_0, the rotation $\boldsymbol{JF}$ on I_1, and the reflection $\boldsymbol{J}$ on I_2; see Figure 1. Conjugating this process via the stereographic projection through $[0,1,1]$ we obtain the Romik map in Figure 2. By construction, it is a continuous piecewise-projective selfmap of the real unit interval $[0,1]$. It is composed of three *pieces*, each one mapping bijectively a subinterval of $[0,1]$ to the whole interval. The computation of these pieces is built-in in our formalism: indeed, since stereographic projection from $[0,1,1]$ is $C^{-1} \circ \tau \circ \upsilon$ on $\partial\mathcal{K}$, computation amounts to switching from boldface to lightface. Thus, the first piece is induced by $J(FPF) = \left[\begin{smallmatrix} 1 & \\ -2 & 1 \end{smallmatrix}\right]$ acting on $FPFJ * [0,1] = [0,1/3]$, the second one by $(JF)(FPF) = JPF = \left[\begin{smallmatrix} -2 & 1 \\ 1 & \end{smallmatrix}\right]$ acting on $FPJ * [0,1] = [1/3,1/2]$, and the third by $F(FPF) = PF = \left[\begin{smallmatrix} 2 & -1 \\ 1 & \end{smallmatrix}\right]$ on $FP * [0,1] = [1/2,1]$.

We adopt another notational shorthand, by consistently writing $\boldsymbol{t}, \theta$ (or $\boldsymbol{s}, \sigma, \ldots$) for pairs $\boldsymbol{t} = [t_1, t_2, t_3] \in \partial\mathcal{K}$, $\theta = (t_1 + t_2 i)/t_3 \in \partial\mathcal{D}$, identified as in the discussion following the diagram (4). We recall that the *residue field* of the point $\boldsymbol{t} = [t_1, t_2, t_3]$ in the projective variety $\{x_1^2 + x_2^2 - x_3^2 = 0\} = \partial\mathcal{K}$ is $\mathbb{Q}(\boldsymbol{t}) = \mathbb{Q}(t_1/t_3, t_2/t_3)$. If $\mathbb{Q}(\boldsymbol{t}) = \mathbb{Q}$ we say that $\boldsymbol{t}$ is a *rational point*; in this case $\boldsymbol{t}$ has a canonical presentation as a pythagorean triple. The corresponding $\theta \in \mathbb{Q}(i)$ has a canonical presentation

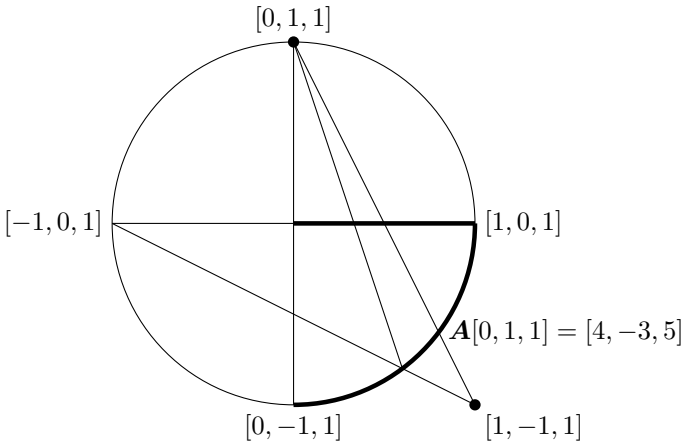


FIGURE 1. A hint of the construction of the Romik map; the interval I_3 and its stereographic projection to $[0, 1]$ as thick lines

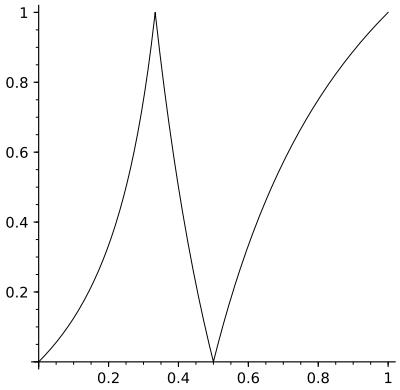


FIGURE 2. The Romik map.

as well, but a subtler one. For each prime integer $p \equiv 1 \pmod 4$, write uniquely $p = a^2 + b^2$, for integers $a > b > 0$, and let $\theta_p = (a + bi)/(a - bi)$ (corresponding, as in Euclid's setting, to $\boldsymbol{t}_p = [a^2 - b^2, 2ab, a^2 + b^2]$). It is well known —and easy to prove [20]— that every $\theta \in S^1 \cap \mathbb{Q}(i)$ factors uniquely in $\mathbb{Q}(i)$ as a product of a unit in $\mathbb{Z}[i]$ and finitely many numbers θ_p and their inverses. This implies that the set of primitive pythagorean triples forms a multiplicative group, isomorphic to the direct sum of the cyclic group of order 4 with countably many copies of the infinite cyclic group. We thus obtain our second canonical presentation: every $\theta \in S^1 \cap \mathbb{Q}(i)$ can be uniquely expressed as $\theta = \kappa\mu/\bar{\mu}$, with $\kappa \in \{1, i, -1, -i\}$ and $\mu \in \mathbb{Z}[i]$ having prime decomposition of the form

$$\mu = (a_1 + b_1 i)^{e_1} \cdots (a_q + b_q i)^{e_q},$$

with $a_j > |b_j| > 0$, $e_j > 0$ for every j, and the pairs $(a_1, |b_1|), \ldots, (a_q, |b_q|)$ all distinct.

4. **The de Sitter space.** The *de Sitter space* is the one-sheeted hyperboloid $\mathcal{S} = \{\boldsymbol{x} \in \mathbb{R}^3 : \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1\}$; it is a lorentzian manifold of constant positive curvature [37], [36]. The de Sitter space is in natural bijection with various spaces of interest to us: these bijections are well known, albeit a bit scattered in the literature. We collect the relevant facts in Theorem 4.1, whose nonstandard feature is the rôle of $\mathrm{PSL}_2^{\pm}\mathbb{R}$ as the acting group, instead of the usual $\mathrm{PSL}_2\mathbb{R}$.

We recall from §2 that $A \mapsto \boldsymbol{A}$ is a group isomorphism from $\mathrm{PSL}_2^{\pm}\mathbb{R}$ to $\mathrm{O}_{2,1}^{\uparrow}\mathbb{R}$. We define now another isomorphism $\Lambda : \mathrm{PSL}_2^{\pm}\mathbb{R} \to \mathrm{SO}_{2,1}\mathbb{R}$ by $\Lambda(A) = (\det A)\boldsymbol{A}$. In the following theorem we let $e : \{1, -1\} \to \{0, 1\}$ have value 0 on 1, and 1 on -1; also, we denote any group action by a star.

Theorem 4.1. *The spaces in the following list, together with the specified base points and transitive left actions of* $\mathrm{PSL}_2^{\pm}\mathbb{R}$*, are in bijective correspondence. These correspondences preserve the base points and are equivariant with respect to the actions.*

(S1) *The de Sitter space* $\mathcal{S}$*, with base point* $(1, 0, 0)$ *and action* $A * \boldsymbol{x} = \Lambda(A)\boldsymbol{x}$*.*
(S2) *The coset space* $\mathrm{PSL}_2\mathbb{R}/\mathfrak{A}$*, for* $\mathfrak{A}$ *the subgroup of diagonal matrices, with base point* $\mathfrak{A}$ *and action* $A * E\mathfrak{A} = AEJ^{e(\det A)}\mathfrak{A}$*.*
(S3) $(\mathrm{P}^1\mathbb{R} \times \mathrm{P}^1\mathbb{R}) \setminus (\text{diagonal})$*, with base point* $(\infty, 0)$ *and action* $A * (\omega, \alpha) = (A * \omega, A * \alpha)$*.*
(S4) $(S^1 \times S^1) \setminus (\text{diagonal})$*, with base point* $(i, -i)$ *and action* $A * (\sigma, \rho) = (\mathscr{A} * \sigma, \mathscr{A} * \rho)$*.*
(S5) *The space of oriented geodesics in* $\mathcal{D}$*, with base point the geodesic from* $-i$ *to* i *and action* $A * g = \mathscr{A}[g]$*.*
(S6) *The space of quadratic forms* $q\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = q_1x^2 - q_2xy + q_3y^2$ *of discriminant* 1*, with base point* $-xy$ *and action* $(A * q)\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = (\det A)q\left(A^{-1}\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right)$*.*

Each space carries a $\mathrm{PSL}_2^{\pm}\mathbb{R}$*-invariant infinite measure, which is the quotient Haar measure in (S2), and is induced by the form* $(\omega - \alpha)^{-2}\,\mathrm{d}\omega\,\mathrm{d}\alpha$ *in (S3). In (S1), the measure of a Borel subset* B *of* $\mathcal{S}$ *is the euclidean volume of the cone* $\{t\boldsymbol{x} : t \in [0, 1], \boldsymbol{x} \in B\}$*, and analogously for (S6).*

Proof. The natural bijections among the spaces in (S3), (S4), (S5) are the obvious ones resulting from the diagram (4). Here we will first describe the bijections among (S2), (S3), (S6), and then the one between (S1) and (S6).

Let q be a form as in (S6), associated to the symmetric matrix

$$Q = \begin{pmatrix} q_1 & -q_2/2 \\ -q_2/2 & q_3 \end{pmatrix}, \tag{7}$$

of determinant $-1/4$. We obtain a pair (ω, α) as in (S3) by labeling the two roots of $q(x, 1)$ as follows:

(a) if $q_1 = 0$ and $q_2 = 1$, then $\omega = \infty$ and $\alpha = q_3$;
(b) if $q_1 = 0$ and $q_2 = -1$, then $\omega = -q_3$ and $\alpha = \infty$;
(c) if $q_1 \neq 0$, then

$$\omega = \frac{q_2 + 1}{2q_1}, \qquad \alpha = \frac{q_2 - 1}{2q_1}.$$

Given a pair (ω, α) as in (S3), we set

$$E = \begin{cases} \begin{bmatrix} 1 & \alpha \\ & 1 \end{bmatrix}, & \text{if } \omega = \infty; \\ \begin{bmatrix} \omega & -1 \\ 1 & \end{bmatrix}, & \text{if } \alpha = \infty; \\ |\omega - \alpha|^{-1/2} \begin{bmatrix} \omega & \alpha \\ 1 & 1 \end{bmatrix} J^{e(\operatorname{sgn}(\omega-\alpha))}, & \text{otherwise;} \end{cases}$$

thus defining a coset $E\mathfrak{A}$ as in (S2).

Finally, any $E\mathfrak{A}$ in (S2) determines a symmetric matrix Q' of determinant $-1/4$ via

$$Q' = -\frac{1}{2}(E^{-1})^\top \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} E^{-1};$$

note that Q' is well defined, i.e., independent from the choice of a representative in $E\mathfrak{A}$ and from the lift of this representative to $\mathrm{SL}_2\,\mathbb{R}$.

It is clear that each of these constructions preserves the base points and is equivariant with respect to the listed actions. Therefore, the claimed correspondence between (S2), (S3), (S6) follows as soon as we prove that the final Q' equals the starting Q. We check case (c), leaving the simpler cases (a) and (b) to the reader. By definition,

$$E = \frac{1}{2|q_1|^{1/2}} \begin{bmatrix} q_2+1 & q_2-1 \\ 2q_1 & 2q_1 \end{bmatrix} J^{e(\operatorname{sgn} q_1)},$$

so that

$$E^{-1} = \frac{1}{2|q_1|^{1/2}} J^{e(\operatorname{sgn} q_1)} \begin{bmatrix} 2q_1 & -q_2+1 \\ -2q_1 & q_2+1 \end{bmatrix}.$$

Hence

$$\begin{aligned} Q' &= -\frac{1}{8|q_1|} \begin{pmatrix} 2q_1 & -2q_1 \\ -q_2+1 & q_2+1 \end{pmatrix} J^{e(\operatorname{sgn} q_1)} \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} J^{e(\operatorname{sgn} q_1)} \begin{pmatrix} 2q_1 & -q_2+1 \\ -2q_1 & q_2+1 \end{pmatrix} \\ &= -(\operatorname{sgn} q_1)\frac{1}{8|q_1|} \begin{pmatrix} 2q_1 & -2q_1 \\ -q_2+1 & q_2+1 \end{pmatrix} \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \begin{pmatrix} 2q_1 & -q_2+1 \\ -2q_1 & q_2+1 \end{pmatrix} \\ &= -\frac{1}{8q_1} \begin{pmatrix} -8q_1^2 & 4q_1q_2 \\ 4q_1q_2 & -2q_2^2+2 \end{pmatrix}, \end{aligned} \tag{8}$$

which is the initial Q; note the use of the identity $J^{\pm1}\left(\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix}\right)J^{\pm1} = \pm 1\left(\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix}\right)$ in the computation.

The bijection between (S1) and (S6) is a simple change of variables, namely

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} & 1 & \\ -1 & & 1 \\ 1 & & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}. \tag{9}$$

This change of variables transforms the matrix Q in (7) to $W/2$, where W is the matrix in (1). This implies that the bijection is equivariant with respect to the actions listed in (S1) and (S6); see also Remark 5.2.

The statement about invariant measures is well known; see, e.g., [22, §8]. □

For future reference we list here the form q and the point $\boldsymbol{w} \in \mathcal{S}$ as a function of (ω, α):

$$\begin{aligned} &q = -xy + \alpha y^2, && \boldsymbol{w} = (1, \alpha, \alpha), && \text{if } \omega = \infty; \\ &q = xy - \omega y^2, && \boldsymbol{w} = -(1, \omega, \omega), && \text{if } \alpha = \infty; \\ &q = \frac{x^2 - (\omega+\alpha)xy + \omega\alpha y^2}{\omega - \alpha}, && \boldsymbol{w} = \frac{(\omega+\alpha, \omega\alpha - 1, \omega\alpha + 1)}{\omega - \alpha}, && \text{otherwise.} \end{aligned} \tag{10}$$

5. **Circle intervals.** The unit circle S^1 is cyclically ordered by the ternary betweenness relation $\boldsymbol{t} \prec \boldsymbol{x} \prec \boldsymbol{t}'$, which reads "$\boldsymbol{t}, \boldsymbol{t}', \boldsymbol{x}$ are pairwise distinct, and traveling from $\boldsymbol{t}$ to $\boldsymbol{t}'$ counterclockwise we meet $\boldsymbol{x}$". Every pair of distinct points $\boldsymbol{t}, \boldsymbol{t}'$ determines two closed intervals, namely $[\boldsymbol{t}, \boldsymbol{t}'] = \{\boldsymbol{t}, \boldsymbol{t}'\} \cup \{\boldsymbol{x} : \boldsymbol{t} \prec \boldsymbol{x} \prec \boldsymbol{t}'\}$ and $[\boldsymbol{t}', \boldsymbol{t}]$. Given $\boldsymbol{w}$ in the de Sitter space, the set $I_{\boldsymbol{w}} = \{\boldsymbol{x} \in S^1 : x_3 \langle \boldsymbol{w}, \boldsymbol{x} \rangle \geq 0\}$ is an interval as well (the factor x_3, i.e., the third coordinate of $\boldsymbol{x}$, makes the definition independent from the choice of a representative for $\boldsymbol{x}$). Let us denote the ordinary cross product of two vectors in $\mathbb{R}^3$ by $\boldsymbol{x} \times \boldsymbol{y}$.

Lemma 5.1. *Let $\boldsymbol{t}, \boldsymbol{t}' \in S^1$ be distinct, and let*

$$\boldsymbol{w} = \frac{\boldsymbol{L}\boldsymbol{t}' \times \boldsymbol{L}\boldsymbol{t}}{\langle \boldsymbol{t}', \boldsymbol{t} \rangle}, \tag{11}$$

the right-hand side being independent from the chosen lifts of $\boldsymbol{t}, \boldsymbol{t}'$ to $\mathbb{R}^3 \setminus \{0\}$. Then the following statements hold.

(i) $\boldsymbol{w} \in \mathcal{S}$, *and* $I_{\boldsymbol{w}} = [\boldsymbol{t}, \boldsymbol{t}']$.

(ii) Let $(\omega, \alpha) \in (\mathrm{P}^1\mathbb{R} \times \mathrm{P}^1\mathbb{R}) \setminus (\text{diagonal})$ *be the pair corresponding to* $\boldsymbol{w}$ *according to Theorem 4.1. Then we have*

$$(\omega, \alpha) = \big((\mu \circ \upsilon)(\boldsymbol{t}'), (\mu \circ \upsilon)(\boldsymbol{t})\big).$$

(iii) For every $\boldsymbol{A} \in \mathrm{O}^{\uparrow}_{2,1}\mathbb{R}$, *we have* $\boldsymbol{A}[I_{\boldsymbol{w}}] = I_{\boldsymbol{A}\boldsymbol{w}}$, *which equals* $[\boldsymbol{A}\boldsymbol{t}, \boldsymbol{A}\boldsymbol{t}']$ *if* $\det \boldsymbol{A} = 1$, *and* $[\boldsymbol{A}\boldsymbol{t}', \boldsymbol{A}\boldsymbol{t}]$ *otherwise.*

(iv) $\boldsymbol{w} \in \mathbb{Q}^3$ *if and only if both* $\boldsymbol{t}$ *and* $\boldsymbol{t}'$ *are rational points.*

(v) The arclength of $[\boldsymbol{t}, \boldsymbol{t}']$ *and the third coordinate* w_3 *of* $\boldsymbol{w}$ *are related by* $\operatorname{arclength}([\boldsymbol{t}, \boldsymbol{t}']) = 2 \operatorname{arccot}(w_3)$.

(vi) If $\boldsymbol{t}$ *and* $\boldsymbol{t}'$ *do not lie on the same diameter (i.e., by (v), if* $w_3 \neq 0$*), then the unique circle in* $\mathbb{R}^2$ *perpendicular to* S^1 *and passing through* $\boldsymbol{t}, \boldsymbol{t}'$ *has center* $(w_1/w_3, w_2/w_3)$ *and curvature* $|w_3|$.

(vii) Assume that

$$I_{\boldsymbol{w}_0} \supseteq I_{\boldsymbol{w}_1} \supseteq I_{\boldsymbol{w}_2} \supseteq \cdots,$$

with arclength tending to 0 (i.e., $\lim_{t\to\infty} w_{t,3} = \infty$*). Then* $\lim_{t\to\infty} \operatorname{arclength}(I_{\boldsymbol{w}_t}) / (2/w_{t,3}) = 1$.

Proof. (i) Every rotation

$$\boldsymbol{S} = \begin{pmatrix} \cos s & -\sin s & \\ \sin s & \cos s & \\ & & 1 \end{pmatrix}$$

leaves invariant the arclength of $[\boldsymbol{t}, \boldsymbol{t}']$ and the third coordinate of $\boldsymbol{w}$ (because $\boldsymbol{S}$ belongs to $\mathrm{SO}_3\mathbb{R}$ as well as to $\mathrm{SO}_{2,1}\mathbb{R}$, and hence $(\boldsymbol{L}\boldsymbol{S}\boldsymbol{t}' \times \boldsymbol{L}\boldsymbol{S}\boldsymbol{t}) / \langle \boldsymbol{S}\boldsymbol{t}', \boldsymbol{S}\boldsymbol{t} \rangle = \boldsymbol{S}\boldsymbol{w}$). Therefore we assume without loss of generality $\boldsymbol{t} = [1, 0, 1]$ and $\boldsymbol{t}' = [\cos r, \sin r, 1]$, for some $0 < r < 2\pi$. Then, by explicit computation, $\boldsymbol{w} = \big((\sin r)/(1 - \cos r), 1, (\sin r)/(1 - \cos r)\big)$, which is indeed in $\mathcal{S}$. Let $\boldsymbol{x}(u) = [\cos u, \sin u, 1]$, and let $f(u) =$

$\langle \boldsymbol{w}, \boldsymbol{x}(u)\rangle : [0, 2\pi) \to \mathbb{R}$. Then, by elementary projective geometry, f takes value 0 in precisely two points, namely in $u = 0$ and in the unique solution to $\boldsymbol{x}(u) = \boldsymbol{t}'$. Again by explicit computation, f has derivative $f'(u) = \cos u - (\sin r)(\sin u)/(1 - \cos r)$, which is positive at 0. This, and extending f to be periodic, then implies that $\langle \boldsymbol{w}, \boldsymbol{x}\rangle \geq 0$ if and only if $\boldsymbol{x} \in [\boldsymbol{t}, \boldsymbol{t}']$, as claimed.
(ii) We have $(\mu \circ \upsilon)^{-1}(\omega) = (2\omega, \omega^2 - 1, \omega^2 + 1)$, and analogously for α. Our statement amounts then to the verification that the vector

$$\frac{\boldsymbol{L}(2\omega, \omega^2 - 1, \omega^2 + 1) \times \boldsymbol{L}(2\alpha, \alpha^2 - 1, \alpha^2 + 1)}{\langle (2\omega, \omega^2 - 1, \omega^2 + 1), (2\alpha, \alpha^2 - 1, \alpha^2 + 1)\rangle}$$

resulting from (11) equals the vector $\boldsymbol{w}$ given by (10). This is a straightforward computation.
(iii) Let $\boldsymbol{x}$ be a point in S^1, and choose a representative for it with positive third coordinate. Then, for every $\boldsymbol{A} \in \mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R}$, the third coordinate of $\boldsymbol{A}^{-1}\boldsymbol{x}$ is still positive; we thus have $\boldsymbol{x} \in \boldsymbol{A}[I_{\boldsymbol{w}}]$ iff $\boldsymbol{A}^{-1}\boldsymbol{x} \in I_{\boldsymbol{w}}$ iff $\langle \boldsymbol{w}, \boldsymbol{A}^{-1}\boldsymbol{x}\rangle \geq 0$ iff $\langle \boldsymbol{A}\boldsymbol{w}, \boldsymbol{x}\rangle \geq 0$ iff $\boldsymbol{x} \in I_{\boldsymbol{A}\boldsymbol{w}}$. The second statement follows from the first and the remark that $\boldsymbol{t} \prec \boldsymbol{A}^{-1}\boldsymbol{x} \prec \boldsymbol{t}'$ is equivalent to $\boldsymbol{A}\boldsymbol{t} \prec \boldsymbol{x} \prec \boldsymbol{A}\boldsymbol{t}'$ if $\det \boldsymbol{A} = 1$, and to $\boldsymbol{A}\boldsymbol{t}' \prec \boldsymbol{x} \prec \boldsymbol{A}\boldsymbol{t}$ if $\det \boldsymbol{A} = -1$.
(iv) The right-to-left implication follows from the definition of $\boldsymbol{w}$. Conversely, if $\boldsymbol{w} \in \mathbb{Q}^3$ then the proof of the equivalence between (S1) and (S6) in Theorem 4.1 yields that the form q corresponding to $\boldsymbol{w}$ has rational coefficients. Since q has discriminant 1, the roots of $q(x, 1)$ (given by (a), (b), (c) in the proof of the same Theorem 4.1) are rational numbers. By (ii), $\boldsymbol{t}$ and $\boldsymbol{t}'$ are the reverse stereographic projections through $[0, 1, 1]$ of these roots, and thus are rational points.
(v) As in (i), we assume $\boldsymbol{t} = [1, 0, 1]$ and $\boldsymbol{t}' = [\cos r, \sin r, 1]$. Then, as computed in (i), $w_3 = (\sin r)/(1 - \cos r) = \cot(r/2)$, and our statement follows.
(vi) Looking at $\boldsymbol{w}$ as a point in $\mathrm{P}^2\,\mathbb{R}$, the identities $\langle \boldsymbol{w}, \boldsymbol{t}\rangle = \langle \boldsymbol{w}, \boldsymbol{t}'\rangle = 0$ mean that $\boldsymbol{w}$ is the intersection point of the two lines tangent to S^1 at $\boldsymbol{t}$ and $\boldsymbol{t}'$; thus the described circle has center $(w_1/w_3, w_2/w_3)$. Upon applying the rotation in the proof of (i), the statement about the curvature follows by direct inspection.
(vii) This is clear. □

Remark 5.2. Since, as it is easily seen, the map $\boldsymbol{w} \mapsto I_{\boldsymbol{w}}$ is a bijection between $\mathcal{S}$ and the space of closed circle intervals, it is tempting to add a seventh item to the list in Theorem 4.1. However this would not be correct, since the action in Lemma 5.1(iii) does not agree with the one in Theorem 4.1(S1). In other words, $\mathrm{PSL}^{\pm}_2\,\mathbb{R}$ acts on the space of intervals via the "bold" isomorphism $A \mapsto \boldsymbol{A}$, while it acts on the de Sitter space via Λ. The following commuting diagram may clarify the situation

$$\begin{array}{ccccc} & & & & \mathrm{O}^{\uparrow}_{2,1}\,\mathbb{R} & \hookrightarrow & \mathrm{O}_{2,1}\,\mathbb{R} \\ & & & \overset{\text{bold}}{\nearrow} & \uparrow & & \uparrow \\ \mathrm{PSU}^{\pm}_{1,1}\,\mathbb{C} & \overset{C^{-1}-C}{\twoheadrightarrow} & \mathrm{PSL}^{\pm}_2\,\mathbb{R} & & & & \\ & & & \underset{\Lambda}{\searrow} & \downarrow & & \downarrow \\ & & & & \mathrm{SO}_{2,1}\,\mathbb{R} & \hookrightarrow & \mathrm{O}_{2,1}\,\mathbb{R} \end{array} \tag{12}$$

In (12), the rightmost vertical arrow is the involutive automorphism $\boldsymbol{A} \mapsto (\det \boldsymbol{A})(\operatorname{sgn} \boldsymbol{A}_{3,3})\boldsymbol{A}$ of $\mathrm{O}_{2,1}\,\mathbb{R}$, which restricts to the isomorphisms $\Lambda \circ \mathrm{bold}^{-1}$ and $\mathrm{bold} \circ \Lambda^{-1}$.

Since these isomorphisms obviously preserve the fact that a matrix has integer entries, Theorem 2.4 implies that $\mathrm{SO}_{2,1}\mathbb{Z} = \Lambda\big[\langle F,P,G\rangle\big] = \langle -\boldsymbol{F},-\boldsymbol{P},-\boldsymbol{G}\rangle \simeq \Delta^{\pm}(2,4,\infty)$ and $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z} = \Lambda\big[\langle F,P,G\rangle^{+}\big] = \langle \boldsymbol{F},\boldsymbol{P},\boldsymbol{G}\rangle^{+} \simeq \Delta(2,4,\infty)$.

When working with continued fractions algorithms one naturally deals with unimodular intervals in $\mathrm{P}^1\mathbb{R}$, namely intervals $[p/q, p'/q']$ with rational endpoints and such that $\det\left(\begin{smallmatrix} p & p' \\ q & q' \end{smallmatrix}\right) = -1$; for example, the intervals $[1/(a+1), 1/a]$ of continuity for the Gauss map $x \mapsto 1/x - \lfloor 1/x \rfloor$ are unimodular. It is a trivial —but key— fact that the modular group $\mathrm{PSL}_2\mathbb{Z}$ acts simply transitively on such intervals. The situation for intervals on the circle is more involved.

Theorem 5.3. *The set $\mathcal{S}\cap\mathbb{Z}^3$ is partitioned in two orbits, corresponding to the parity of w_3, by the action of $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z}$. On each orbit the action is simply transitive. Replacing $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z}$ with its index-2 subgroup $\Lambda\big[\langle F,P,J\rangle^{+}\big]$ each orbit is further split in two.*

Proof. It is easy to check that each of $-\boldsymbol{F}$, $-\boldsymbol{P}$, $-\boldsymbol{G}$ preserves the parity of w_3; hence there are at least two orbits.

Choose $\boldsymbol{w} \in \mathcal{S}\cap\mathbb{Z}^3$ and let $(\omega,\alpha) \in (\mathrm{P}^1\mathbb{Q} \times \mathrm{P}^1\mathbb{Q}) \setminus (\text{diagonal})$ be the corresponding ordered pair according to Theorem 4.1. An appropriate power $(FP)^k$ of the parabolic matrix FP (that fixes 1) sends (ω,α) to a new pair (ω',α') with $0 \le \omega' \le 1$. By [42, Theorem 2(i)], the orbit $\omega' = \omega'_0, \omega'_1, \omega'_2, \dots$ of ω' under the Romik map ends up after finitely many steps, say the nth step, in one of the two parabolic fixed points 0, 1. For each $0 \le t < n$, let

$$A_t = \begin{cases} JFPF, & \text{if } 0 < \omega'_t < 1/3; \\ JPF, & \text{if } 1/3 \le \omega'_t < 1/2; \\ PF, & \text{if } 1/2 \le \omega'_t < 1; \end{cases}$$

be the matrix acting at time t. Then $A = FJA_{n-1}A_{n-2}\cdots A_0(FP)^k \in \langle F,P,J\rangle$, and $A*(\omega,\alpha) = (\omega'',\alpha'')$ is such that $\omega'' \in \{\infty,-1\}$. Postcomposing A, if necessary, with J (if $\omega''=\infty$) or with F (if $\omega''=-1$), we have $A \in \langle F,P,J\rangle^{+}$.

Suppose $\omega''=\infty$. Then $\alpha'' \in \mathbb{Z}$ because the point $\boldsymbol{w}''$ corresponding to (∞,α'') equals $(1,\alpha'',\alpha'')$ by (10), and also equals $\Lambda(A)\boldsymbol{w}$, which is a point in $\mathbb{Z}^3$. This implies that an appropriate power of the parabolic matrix $PJ = \left[\begin{smallmatrix} 1 & 2 \\ 1 & \end{smallmatrix}\right]$ maps (∞,α'') either to $(\infty,0)$ or to $(\infty,1)$. If, on the other hand, $\omega'' = -1$, then the same argument with PJ replaced by $(JPJ)F = \left[\begin{smallmatrix} 2 & \\ -1 & 1 \end{smallmatrix}\right]$ (which is parabolic fixing -1) yields that a power of $JPJF$ maps $(-1,\alpha'')$ either to $(-1,1)$ or to $(-1,\infty)$.

Summing up, we have proved that the pair (ω,α) is in the $\langle F,P,J\rangle^{+}$-orbit of one of the pairs $(\infty,0),(\infty,1),(-1,1),(-1,\infty)$. Now, the rotation $GF \in \langle F,P,G\rangle^{+}$ maps the first pair to the third, and the second to the fourth. By Theorem 4.1 this means that the original point $\boldsymbol{w}$ is in the $\Lambda\big[\langle F,P,G\rangle^{+}\big]$-orbit of either $(1,0,0)$ or of $(1,1,1)$. Since $\Lambda\big[\langle F,P,G\rangle^{+}\big] = \mathrm{SO}^{\uparrow}_{2,1}\mathbb{Z}$ by Remark 5.2, our first claim is established.

Simple transitivity follows from the fact that both $(\infty,0)$ and $(\infty,1)$ have trivial stabilizer in $\langle F,P,G\rangle^{+}$ (because an element of a fuchsian group that fixes two distinct cusps must be the identity).

Finally, the pairs $(\infty,0),(\infty,1),(-1,1),(-1,\infty)$ remain distinct modulo $\langle F,P,J\rangle^{+}$. Indeed, the latter is the triangle group $\Delta(2,\infty,\infty)$, which has two distinct cusp orbits, and it is easy to check that any identification of the above four pairs would collapse these two orbits. □

We can now define unimodularity for circle intervals.

Definition 5.4. Let $\boldsymbol{t}, \boldsymbol{t}'$ be distinct rational points in S^1, and let $\boldsymbol{w} \in \mathcal{S} \cap \mathbb{Q}^3$ be the point corresponding to $[\boldsymbol{t}, \boldsymbol{t}']$ according to Lemma 5.1. If $\boldsymbol{w} \in \mathbb{Z}^3$ and w_3 is even (odd), then we say that $[\boldsymbol{t}, \boldsymbol{t}']$ is an *even* (*odd*) *unimodular interval*.

Theorem 5.5. *Let $\boldsymbol{t}, \boldsymbol{t}', \boldsymbol{w}$ be as in Definition 5.4; then the following conditions are equivalent.*

1. *$[\boldsymbol{t}, \boldsymbol{t}']$ is unimodular (either even or odd).*
2. *$\boldsymbol{R}_{\boldsymbol{w}}$ has integer entries.*
3. *$[\boldsymbol{t}, \boldsymbol{t}']$ is the image either of $\big[[0,-1,1],[0,1,1]\big]$ or of $\big[[1,0,1],[0,1,1]\big]$ under some (necessarily unique) element of $\mathrm{SO}^{\uparrow}_{2,1}\,\mathbb{Z}$.*
4. *$\langle \boldsymbol{t}, \boldsymbol{t}' \rangle \in \{-1, -2\}$ (here $\boldsymbol{t}, \boldsymbol{t}'$ are the canonical presentations of $\boldsymbol{t}, \boldsymbol{t}'$ as primitive pythagorean triples).*

If these conditions hold, then $[\boldsymbol{t}, \boldsymbol{t}']$ is odd iff it is the image of $\big[[1,0,1],[0,1,1]\big]$ iff $\langle \boldsymbol{t}, \boldsymbol{t}' \rangle = -1$. Moreover, $\boldsymbol{R}_{\boldsymbol{w}}$ belongs to $\langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{J} \rangle$, and the matrix $\mathscr{R}_{\boldsymbol{w}} \in \mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$ corresponding to it under Convention 2.2 is

$$\begin{bmatrix} \theta & \theta' \\ 1 & 1 \end{bmatrix} J \begin{bmatrix} \theta & \theta' \\ 1 & 1 \end{bmatrix}^{-1}, \tag{13}$$

where $\theta, \theta' \in S^1 \cap \mathbb{Q}(i)$ are identified with $\boldsymbol{t}, \boldsymbol{t}'$ as in §3.

Proof. (1) ⇒ (2) Since $\langle \boldsymbol{w}, \boldsymbol{w} \rangle = 1$, this is immediate from the explicit formula for $\boldsymbol{R}_{\boldsymbol{w}}$ in (6).
(2) ⇒ (3) Let

$$(\omega, \alpha) = \big((\mu \circ \upsilon)(\boldsymbol{t}'), (\mu \circ \upsilon)(\boldsymbol{t})\big) \in (\mathrm{P}^1\,\mathbb{Q} \times \mathrm{P}^1\,\mathbb{Q}) \setminus (\text{diagonal})$$

(see Lemma 5.1(ii)). Then, as in the proof of Theorem 5.3, we construct $A \in \langle F, P, J \rangle^+$ such that $A * (\omega, \alpha)$ equals either (∞, α'') or $(-1, \alpha'')$. Since $FG * (-1) = \infty$, there exists $B \in \langle F, P, G \rangle^+$ with $B * (\omega, \alpha) = (\infty, q)$, for some $q \in \mathbb{Q}$. Hence, $\Lambda(B)\boldsymbol{w} = (1, q, q) = \boldsymbol{v}$. We then have

$$\Lambda(B)\boldsymbol{R}_{\boldsymbol{w}}\Lambda(B)^{-1} = \boldsymbol{R}_{\Lambda(B)\boldsymbol{w}} = \boldsymbol{R}_{\boldsymbol{v}} = \boldsymbol{I} - \frac{2}{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \boldsymbol{v}\,\boldsymbol{v}^{\top}\boldsymbol{L},$$

and the leftmost entry in the display is a matrix with integer entries. Multiplying through by -1, subtracting the identity matrix $\boldsymbol{I}$, and multiplying by $\boldsymbol{L}$ on the right, we see that the matrix

$$\frac{2}{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \boldsymbol{v}\boldsymbol{v}^{\top} = 2 \begin{pmatrix} 1 & q & q \\ q & q^2 & q^2 \\ q & q^2 & q^2 \end{pmatrix}$$

must have integer entries. This implies that the denominator of the rational number q must divide 2, and so must do the denominator of q^2; therefore q is an integer. Thus, as in the proof of Theorem 5.3, an appropriate power $(\boldsymbol{PJ})^k$ will map $(1, q, q)$ either to $(1, 0, 0)$ or to $(1, 1, 1)$; therefore, $\Lambda\big((PJ)^k B\big)\boldsymbol{w} \in \{(1,0,0), (1,1,1)\}$. Now, $(PJ)^k B \in \langle F, P, G \rangle^+$, and Λ equals the "bold" isomorphism on $\langle F, P, G \rangle^+$, with range $\mathrm{SO}^{\uparrow}_{2,1}\,\mathbb{Z}$. Thus $\boldsymbol{w}$ is the image either of $(1, 0, 0)$ or of $(1, 1, 1)$ under some element of $\mathrm{SO}^{\uparrow}_{2,1}\,\mathbb{Z}$, a statement equivalent to (3) by Remark 5.2.
(3) ⇒ (4) This is clear, since $\langle (0,-1,1), (0,1,1) \rangle = -2$ and $\langle (1,0,1), (0,1,1) \rangle = -1$.
(4) ⇒ (1) If $\langle \boldsymbol{t}, \boldsymbol{t}' \rangle = -1$, then $\boldsymbol{w} \in \mathbb{Z}^3$ by the definition of $\boldsymbol{w}$ in Lemma 5.1; assume then $\langle \boldsymbol{t}, \boldsymbol{t}' \rangle = -2$. In every pythagorean triple one of the legs must be even, and the

other leg and the hypotenuse both odd. The condition $t_1t'_1+t_2t'_2-t_3t'_3=-2$ forces t_1, t'_1 to be both even and t_2, t'_2 both odd (or conversely). Since t_3, t'_3 are surely both odd, all the entries in $\boldsymbol{Lt}' \times \boldsymbol{Lt}$ must be even; thus $\boldsymbol{w} \in \mathbb{Z}^3$.

The stated characterization of $[\boldsymbol{t}, \boldsymbol{t}']$ being even/odd is clear from the previous proof.

By Theorem 5.3, $\boldsymbol{w}$ is in the $\langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{J}\rangle^+$-orbit of one of $(1,0,0)$, $(1,1,1)$, $(0,1,0)$, $(-1,1,1)$. Hence $\boldsymbol{R}_{\boldsymbol{w}}$ is a conjugate either of $\boldsymbol{R}_{(1,0,0)} = \boldsymbol{J}$, or of $\boldsymbol{R}_{(1,1,1)} = \boldsymbol{P}$, or of $\boldsymbol{R}_{(0,1,0)} = \boldsymbol{F}$, or of $\boldsymbol{R}_{(-1,1,1)} = \boldsymbol{JPJ}$ by a matrix in $\langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{J}\rangle^+$; in any case, it belongs to $\langle \boldsymbol{F}, \boldsymbol{P}, \boldsymbol{J}\rangle$.

Finally, let $\mathscr{S}$ be the matrix in (13). By direct computation

$$\mathscr{S} = (\theta-\theta')^{-1}\begin{bmatrix} -\theta-\theta' & 2\theta\theta' \\ -2 & \theta+\theta' \end{bmatrix},$$

which has the form $\left[\begin{smallmatrix} \alpha & \beta \\ \bar\beta & \bar\alpha \end{smallmatrix}\right]$, as can easily be checked; hence $\mathscr{S} \in \mathrm{PSU}^{\pm}_{1,1}\,\mathbb{C}$. If we can prove that $\mathscr{S}$ has entries in $\mathbb{Z}[i]$, then necessarily $\mathscr{S} = \mathscr{R}_{\boldsymbol{w}}$. Indeed, the matrix $\mathscr{S}^{-1}\mathscr{R}_{\boldsymbol{w}}$ would then belong to the fuchsian group $\mathrm{PSU}_{1,1}\,\mathbb{Z}[i]$, and would fix the two cusps θ, θ'; hence, it must be the identity matrix.

Write uniquely $\theta = \kappa\mu/\bar\mu$, $\theta' = \lambda\nu/\bar\nu$, as explained in §3. By Theorem 5.3, there exists $\mathscr{A} \in \langle \mathscr{F}, \mathscr{P}, \mathscr{J}\rangle^+ = \mathrm{PSU}_{1,1}\,\mathbb{Z}[i]$ such that

$$\mathscr{A}\begin{bmatrix} \kappa\mu & \lambda\nu \\ \bar\mu & \bar\nu \end{bmatrix} \in \left\{ \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & i \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} i & -1 \\ 1 & 1 \end{bmatrix} \right\}.$$

This implies that the determinant $\delta = \kappa\mu\bar\nu - \lambda\bar\mu\nu$ divides 2 in $\mathbb{Z}[i]$. Since

$$\begin{bmatrix} \theta & \theta' \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \kappa\mu & \lambda\nu \\ \bar\mu & \bar\nu \end{bmatrix}\begin{bmatrix} \bar\mu & \\ & \bar\nu \end{bmatrix}^{-1},$$

we have

$$\begin{aligned} \begin{bmatrix} \theta & \theta' \\ 1 & 1 \end{bmatrix} J \begin{bmatrix} \theta & \theta' \\ 1 & 1 \end{bmatrix}^{-1} &= \begin{bmatrix} \kappa\mu & \lambda\nu \\ \bar\mu & \bar\nu \end{bmatrix} J \begin{bmatrix} \kappa\mu & \lambda\nu \\ \bar\mu & \bar\nu \end{bmatrix}^{-1} \\ &= \delta^{-1}\begin{bmatrix} -\kappa\mu\bar\nu - \lambda\bar\mu\nu & 2\kappa\lambda\mu\nu \\ -2\bar\mu\bar\nu & \lambda\bar\mu\nu + \kappa\mu\bar\nu \end{bmatrix} \\ &= \delta^{-1}\begin{bmatrix} \delta - 2\kappa\mu\bar\nu & 2\kappa\lambda\mu\nu \\ -2\bar\mu\bar\nu & \delta + 2\lambda\bar\mu\nu \end{bmatrix}, \end{aligned}$$

which has entries in $\mathbb{Z}[i]$. □

6. **Billiard maps.** Having arranged our tools in working order, we proceed to our core objects.

Definition 6.1. A *unimodular partition* of the unit circle S^1 is a counterclockwise cyclically ordered m-uple $\boldsymbol{t}_0, \boldsymbol{t}_1, \dots, \boldsymbol{t}_{m-1}$ of pythagorean triples, of cardinality at least 3, such that each interval $[\boldsymbol{t}_a, \boldsymbol{t}_{a+1}]$ is unimodular (including $[\boldsymbol{t}_{m-1}, \boldsymbol{t}_0]$; here and in the following we are writing indices modulo m). We will write $\boldsymbol{w}_a = (\boldsymbol{Lt}_{a+1} \times \boldsymbol{Lt}_a)/\langle \boldsymbol{t}_{a+1}, \boldsymbol{t}_a\rangle \in \mathcal{S}$ for the points defined by Lemma 5.1.

According to our conventions, and without further notice, we will often switch to a complex-numbers setting, thus writing θ_a for $\boldsymbol{t}_a$.

For every a, let l_a be the geodesic in $\mathcal{D}$ of ideal endpoints θ_a and θ_{a+1}; of the two halfplanes determined by l_a, let D_a be the one containing all other l_b, for $b \neq a$. Then $D = \bigcap\{D_a : a = 0, \dots, m-1\}$ is a polygon with sides $l_0, \dots, l_{m-1}$ and ideal

vertices $\theta_0, \dots, \theta_{m-1}$, on which we can play billiards in the usual way. Namely, any unit velocity vector attached to an infinitesimal ball in the interior of D determines an oriented geodesic g starting from an ideal point ρ and ending at σ. The ball travels along g at unit speed, until it hits the side l_a determined by the half-open interval $[\theta_a, \theta_{a+1})$ to which σ belongs (unless σ is one of the vertices, in which case the ball is lost at infinity). When hitting l_a, the ball rebounces with angle of reflection equal to the angle of incidence, and continues its trajectory along the geodesic g' which is the image of g with respect to the reflection with mirror l_a. This reflection is induced by the matrix $\mathscr{R}_{\boldsymbol{w}_a}$ in (13) (with $\theta = \theta_a$ and $\theta' = \theta_{a+1}$), and thus has ideal initial and terminal points $\mathscr{R}_{\boldsymbol{w}_a} * \rho$ and $\mathscr{R}_{\boldsymbol{w}_a} * \sigma$, respectively. All of this naturally suggests the following standard definition [18, Chapter 6], [16, §IV.1].

Definition 6.2. The *billiard map* determined by the unimodular partition $\theta_0, \dots, \theta_{m-1}$ is the map $\widetilde{B}$ from $(S^1 \times S^1) \setminus (\text{diagonal})$ to itself defined by $\widetilde{B}(\sigma, \rho) = (\mathscr{A}_a * \sigma, \mathscr{A}_a * \rho)$, where a is the index of the unique half-open interval $I_a = [\theta_a, \theta_{a+1})$ containing σ, and $\mathscr{A}_a = \mathscr{R}_{\boldsymbol{w}_a}$. The map $\widetilde{B}$ is continuous, and determines a topological dynamical system. We denote by (S^1, B) the factor system naturally induced by the projection $(\sigma, \rho) \mapsto \sigma$; in short, $B(\sigma) = \mathscr{A}_a * \sigma$ for $\sigma \in I_a$.

We will freely use Theorem 4.1 to conjugate $\widetilde{B}$ to a map acting on any of the spaces (S1)–(S6); we will still denote the conjugated map by $\widetilde{B}$, slightly abusing notation. For ease of visualization (and crucially in §9 and §10) we will also conjugate $\widetilde{B}$ and B to maps on $[0,1)^2 \setminus (\text{diagonal})$ and $[0,1)$, respectively; these last conjugations are realized through the normalized (i.e., the image is divided by 2π) argument function $\arg : \partial\mathcal{D} \to [0,1)$.

Example 6.3. The ordered 6-uple

$$\theta_0 = 1,\ \theta_1 = \frac{12+5i}{13},\ \theta_2 = \frac{4+3i}{5},\ \theta_3 = i,\ \theta_4 = -i, \theta_5 = \frac{4-3i}{5},$$

is a unimodular partition, whose corresponding billiard table is shown in Figure 3 (left). The matrices $\mathscr{A}_0, \dots, \mathscr{A}_5$ are

$$\begin{bmatrix} -5i & -1+5i \\ -1-5i & 5i \end{bmatrix}, \quad \begin{bmatrix} -8i & -4+7i \\ -4-7i & 8i \end{bmatrix}, \quad \begin{bmatrix} -2i & -2+i \\ -2-i & 2i \end{bmatrix},$$
$$\begin{bmatrix} -i & \\ & i \end{bmatrix} = \mathscr{J}, \quad \begin{bmatrix} -2i & 2+i \\ 2-i & 2i \end{bmatrix}, \quad \begin{bmatrix} -3i & 1+3i \\ 1-3i & 3i \end{bmatrix}.$$

The graph of the arg-conjugate of B is shown in Figure 3 (right); it requires caution in two respects. First, B is a *continuous* map on S^1 and, second, it is piecewise-defined via *six* pieces, whose endpoints are given by the six B-fixed points ($0 = 1$ included). We plot in Figure 4 (left) 5000 points of the $\widetilde{B}$-orbit of a "typical" point in the de Sitter space $\mathcal{S}$, and in Figure 4 (right) their arg-images. The cluster points apparent in this latter figure correspond to the six fixed points cited above. These are indifferent fixed points (i.e., the derivative of B has absolute value 1), and this forces the unique B-invariant measure absolutely continuous with respect to the Lebesgue measure to be infinite; see Theorem 7.2 and Figure 5. Note that $\widetilde{B}$ is not injective: the points $(\theta_0, \mathscr{A}_0 * \theta_2)$ and $(\mathscr{A}_2 * \theta_0, \theta_2)$ are different, but both get mapped to (θ_0, θ_2) (see however Theorem 7.1(i)).

We let $\Gamma_B^{\pm}$ be the group generated by $\mathscr{A}_0, \dots, \mathscr{A}_{m-1}$, and $\Gamma_B = \Gamma_B^{\pm} \cap \mathrm{PSU}_{1,1}\,\mathbb{Z}[i]$ the associated fuchsian group. By conjugating with an appropriate element of

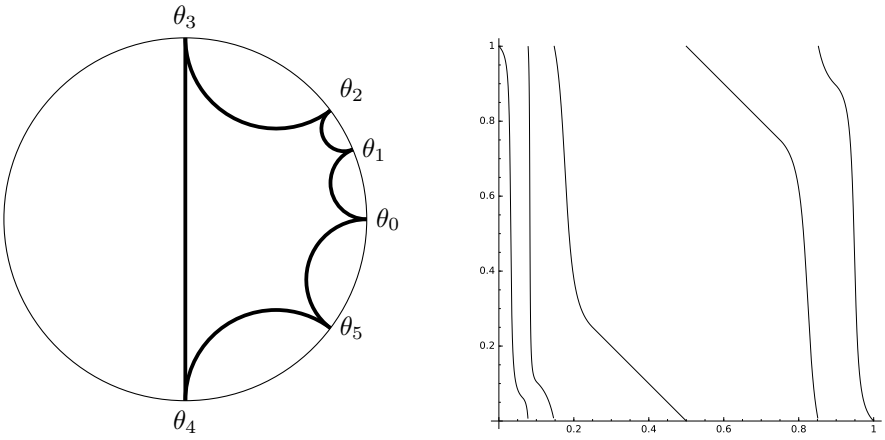


FIGURE 3. A unimodular billiard table and its associated factor map B.

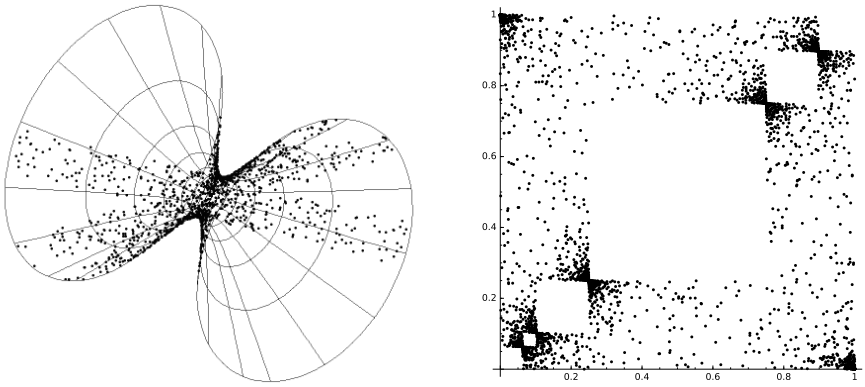


FIGURE 4. A typical $\widetilde{B}$-orbit on the de Sitter space and its arg-image

$\mathrm{PSU}_{1,1}\,\mathbb{Z}[i]$ we always assume, without loss of generality, that $\theta_0 = 1$. As noted in §2, $\Gamma_B^{\pm}$ admits the presentation $\langle x_0, \dots, x_{m-1} \mid x_0^2 = x_1^2 = \cdots = x_{m-1}^2 = 1\rangle$, and hence is isomorphic to the free product of m copies of the group of order two. Equivalently stated, each element of $\Gamma_B^{\pm}$ can be uniquely written as a word in the generators $\mathscr{A}_0, \dots, \mathscr{A}_{m-1}$, subject to the only condition that the same generator does not appear in two consecutive positions. Since D has finite hyperbolic area, Γ_B and $\Gamma_B^{\pm}$ have finite index in $\mathrm{PSU}_{1,1}^{\pm}\,\mathbb{Z}[i]$.

Definition 6.4. Let $B, I_0, \dots, I_{m-1}$ be as in Definition 6.2. For each $t = 0, 1, 2, \dots$, let a_t be determined by $B^t(\sigma) \in I_{a_t}$; the point $\varphi(\sigma) = a_0 a_1 a_2 \ldots = \mathbf{a}$ in the Cantor space $\{0, \dots, m-1\}^\omega$ is the *B-symbolic sequence* of σ.

Lemma 6.5. *The B-symbolic-sequence map $\varphi : S^1 \to \{0, \dots, m-1\}^\omega$ is injective. Its range is the set of all sequences $\mathbf{a}$ such that:*

(i) if $a_t = a_{t+1}$ for some t, then $a_t = a_{t+h}$ for every $h \geq 0$;

(ii) for any $a \in \{0, \dots, m-1\}$, the tail of $\mathbf{a}$ is neither of the form $\overline{a(a+1)}$, nor of the form $(a-1)\overline{a}$ (the bar denoting periodicity).

Remark 6.6. Since we are considering half-open intervals, each σ has precisely one B-symbolic sequence; thus φ is well defined. This differs slightly form other treatments of Gauss-like maps (see, e.g., [29, §2.1] or [45, §1.2.1]), in which rational points have two symbolic sequences. Note that φ is not continuous; indeed, if it were it would have compact image, which is not the case (e.g., all sequences of the form $(01)^n\overline{0}$ lie in the image, but the resulting sequence of sequences does not have a limit point in $\varphi[S^1]$).

Proof of Lemma 6.5. Each $\mathscr{A}_a$ is an involution, and exchanges $\overline{I}_a$ with $\bigcup_{b\neq a}\overline{I}_b$, the bar denoting topological closure. However, in this proof we carefully distinguish B (which maps bijectively $\overline{I}_a$ to $\bigcup_{b\neq b}\overline{I}_b$) from $\mathscr{A}_a$ (which is one of the branches of B^{-1}, the one that maps bijectively $\bigcup_{b\neq a}\overline{I}_b$ to $\overline{I}_a$). We do so in order to prepare the ground for the proof of Theorem 9.2, where the argument we are going to provide will be adapted to another $(m-1)$-to-1 covering map of S^1.

Let $\mathbf{a}=\varphi(\sigma)$. If $a_t=a_{t+1}=a$, then $B^t(\sigma)\in I_a\cap B^{-1}[I_a]=\{\theta_a\}$. Since θ_a is a B-fixed point, we have $a_{t+h}=a$ for every $h\geq 0$. Moreover, if $t\geq 1$ and $a_{t-1}\neq a$, then we have $\theta_a=B^t(\sigma)\in B[I_{a_{t-1}}]$, which implies $a_{t-1}\neq a-1$, because $\theta_a\notin B[I_{a-1}]$. Hence $\mathbf{a}$ cannot have tail $(a-1)\overline{a}$. The fact that $\mathbf{a}$ cannot have tail $\overline{a(a+1)}$ is proved in [12, Theorem 2.1]. We conclude that every B-symbolic sequence must satisfy (i) and (ii).

Conversely, we fix $\mathbf{a}$ satisfying (i) and (ii) and show that there exists a unique point having $\mathbf{a}$ as B-symbolic sequence. We need a preliminary remark: suppose we know that σ is the unique point having B-symbolic sequence $\mathbf{b}$. Then, by direct inspection, we have:

(a) if σ is in the interior of I_{b_0} and $b\neq b_0$, then $\mathscr{A}_b*\sigma$ is in the interior of I_b and is the unique point having B-symbolic sequence $b\mathbf{b}$;
(b) the same conclusion holds if $\sigma=\theta_{b_0}$, provided that $b\notin\{b_0,b_0-1\}$.

Case 1. The sequence $\mathbf{a}$ has tail $\overline{a}$, say from time t on. If $t=0$, then there exists a unique point having B-symbolic sequence $\overline{a}$, namely θ_a. If $t>0$, then the previous remark and induction show that $\mathscr{A}_{a_0}\cdots\mathscr{A}_{a_{t-1}}*\theta_a$ is the only point having B-symbolic sequence $\mathbf{a}$.

Case 2. The sequence $\mathbf{a}$ does not have tail $\overline{a}$, for any a. Since $a_t\neq a_{t+1}$ for every t, we have strict inclusions $\overline{I}_{a_t}\supset\mathscr{A}_{a_t}[\overline{I}_{a_{t+1}}]$ for every t, and hence a strictly decreasing sequence of nested intervals

$$\overline{I}_{a_0}\supset\mathscr{A}_{a_0}[\overline{I}_{a_1}]\supset\mathscr{A}_{a_0}\mathscr{A}_{a_1}[\overline{I}_{a_2}]\supset\cdots. \tag{14}$$

We claim that this sequence shrinks to a singleton. Indeed, each set in (14) is a unimodular interval, strictly containing the following one. By Lemma 5.1(v) the third coordinates of the corresponding points $\boldsymbol{w}_{a_0},\boldsymbol{A}_{a_0}\boldsymbol{w}_{a_1},\boldsymbol{A}_{a_0}\boldsymbol{A}_{a_1}\boldsymbol{w}_{a_2},\dots$ on the de Sitter space form a strictly increasing sequence. Since we are dealing with unimodular intervals, these third coordinates are integer numbers, and a strictly increasing sequence of integers must go to infinity. Therefore the arclengths of the intervals go to 0, and the intersection of the sequence in (14) contains at least one point —by compactness— but no more than one.

Let σ be the shrinking point of (14) and let $\varphi(\sigma)=\mathbf{b}$; we prove $\mathbf{a}=\mathbf{b}$ by induction (note that, clearly, no point other than σ may have B-symbolic sequence $\mathbf{a}$). We have $\sigma\in\overline{I}_{a_0}\cap I_{b_0}$; if a_0 were different from b_0, then necessarily $\sigma=\theta_{b_0}$ and $b_0=a_0+1$. Therefore, for every $t\geq 1$ we have $\sigma=B^t(\sigma)\in B^t[\mathscr{A}_{a_0}\cdots\mathscr{A}_{a_{t-1}}[\overline{I}_{a_t}]]=\overline{I}_{a_t}$, and thus σ belongs to $\overline{I}_{a_t}$. This implies $\mathbf{a}=\overline{a_0(a_0+1)}$, which contradicts (ii);

hence $a_0 = b_0$. For the inductive step, assume $a_r = b_r$ for $0 \le r < t$. Then $B^t(\sigma)$ has B-symbolic sequence $b_t b_{t+1} \dots$ and is the unique shrinking point of the chain

$$\overline{I}_{a_t} \supset \mathscr{A}_{a_t}[\overline{I}_{a_{t+1}}] \supset \mathscr{A}_{a_t}\mathscr{A}_{a_{t+1}}[\overline{I}_{a_{t+2}}] \supset \cdots .$$

Applying the base step above to $B^t(\sigma)$ we get $a_t = b_t$. □

7. **Natural extension and invariant measures.** If $\varphi(\sigma)$ has constant tail $\overline{a}$ for some $a \in \{0, \dots, m-1\}$, i.e., $B^h(\sigma) = \theta_a$ for some h, we say that σ is *B-terminating*. If $\varphi(\sigma)$ has periodic tail $\overline{a_h \cdots a_{h+p-1}}$ with minimal preperiod h and period $p \ge 2$, we say that σ is *B-periodic* or *B-preperiodic*, according whether h is 0 or greater than 0.

We will push the identification of the de Sitter space with $(S^1 \times S^1) \setminus (\text{diagonal})$ a bit further by using the symbol $\mathcal{S}$ for both; this is unambiguous since writing $\boldsymbol{w} \in \mathcal{S}$ or $(\sigma, \rho) \in \mathcal{S}$ clearly distinguishes the two uses. With this understanding, we denote by $\mathcal{S}_B$ the set of all pairs (σ, ρ) such that:

(i) both σ and ρ are B-nonterminating;
(ii) σ and ρ belong to different intervals.

For the map B of Example 6.3, the orbit in Figure 4 is dense in $\mathcal{S}_B$.

Theorem 7.1. *The following facts hold.*

(i) $\widetilde{B} \restriction \mathcal{S}_B$ *is a bijection on* $\mathcal{S}_B$.
(ii) If $(\sigma, \rho) \in \mathcal{S}$ *is such that both* σ *and* ρ *are* B*-nonterminating, then* $\widetilde{B}^t(\sigma, \rho) \in \mathcal{S}_B$ *for some* $t \ge 0$.
(iii) Let $\tilde{\mu}$ *be the* $\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{C}$*-invariant measure on* $(S^1 \times S^1) \setminus (\text{diagonal})$ *given by Theorem 4.1. Then* $(\mathcal{S}_B, \tilde{\mu}, \widetilde{B})$ *is a measure-preserving system, and so is its factor* (S^1, μ, B)*, where* $\mu = \pi_* \tilde{\mu}$ *is the pushforward measure induced by the projection* $\pi(\sigma, \rho) = \sigma$.
(iv) The invertible system $(\mathcal{S}_B, \tilde{\mu}, \widetilde{B})$ *is the natural extension of* (S^1, μ, B).

Proof. (i) The fact that $\widetilde{B}$ maps $\mathcal{S}_B$ into itself is clear. Writing f for the involution $(\sigma, \rho) \mapsto (\rho, \sigma)$ of $\mathcal{S}_B$, it is also clear that $f \circ \widetilde{B} \circ f = \widetilde{B}^{-1}$ on $\mathcal{S}_B$. In terms of symbolic sequences, all of this just amounts to $\widetilde{B} : (a_0 a_1 \dots, b_0 b_1 \dots) \mapsto (a_1 \dots, a_0 b_0 b_1 \dots)$ and $f \circ \widetilde{B} \circ f : (a_0 a_1 \dots, b_0 b_1 \dots) \mapsto (b_0 a_0 a_1 \dots, b_1 \dots)$.
(ii) Let $\sigma \ne \rho$ be both B-nonterminating. By Lemma 6.5 there exists $t \ge 0$ such that $B^t(\sigma)$ and $B^t(\rho)$ belong to different intervals. By the definitions of $\widetilde{B}$ and of $\mathcal{S}_B$, we have $\widetilde{B}^t(\sigma, \rho) \in \mathcal{S}_B$.
(iii) Any measurable $M \subseteq \mathcal{S}_B$ is the disjoint union $M = \dot{\bigcup}\{M_a : a \in \{0, \dots, m-1\}\}$, where $M_a = \{(\sigma, \rho) \in M : \rho \in I_a\}$. Thus $\widetilde{B}^{-1} M = \dot{\bigcup}_a \widetilde{B}^{-1} M_a = \dot{\bigcup}_a \mathscr{A}_a[M_a]$ and, as $\tilde{\mu}\big(\mathscr{A}_a[M_a]\big) = \tilde{\mu}(M_a)$, we have $\tilde{\mu}(\widetilde{B}^{-1} M) = \tilde{\mu}(M)$.
(iv) The set $\{\sigma \in S^1 : \sigma \text{ is } B\text{-terminating}\}$ is clearly B-invariant and has μ-measure 0; modulo this nullset and its π-counterimage, we have the commuting square

$$\begin{array}{ccc} (\mathcal{S}_B, \tilde{\mu}) & \xrightarrow{\widetilde{B}} & (\mathcal{S}_B, \tilde{\mu}) \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ (S^1, \mu) & \xrightarrow{B} & (S^1, \mu) \end{array}$$

By the very definition of the natural extension [41, p. 22], the metric system $(\mathcal{S}_B, \tilde{\mu}, \widetilde{B})$ is the natural extension of its factor (S^1, μ, B) if the supremum of the

family of measurable partitions

$$\{\widetilde{B}^t(\text{fibers of } \pi) : t \geq 0\}$$

is —modulo nullsets— the partition of $\mathcal{S}_B$ in singletons. This condition amounts to the request that if $(\sigma,\rho) \neq (\sigma',\rho')$, then there exists $t \geq 0$ such that $\pi\big(\widetilde{B}^{-t}(\sigma,\rho)\big) \neq \pi\big(\widetilde{B}^{-t}(\sigma',\rho')\big)$. This request is clearly satisfied: if $\sigma \neq \sigma'$ we take $t = 0$, while if $\sigma = \sigma'$ we take $t = h + 1$, there h is the least nonnegative integer such that $B^t(\rho)$ and $B^t(\rho')$ lie in different intervals. □

As usual in the context of Gauss-like maps, once a model of the natural extension has been determined the computation of the (unique) absolutely continuous B-invariant measure is easy; we state the result for the arg-conjugates of $\widetilde{B}$ and B.

Theorem 7.2. *Let $X = \{(\arg\sigma, \arg\rho) : (\sigma,\rho) \in \mathcal{S}_B\} \subset [0,1)^2$ and write —abusing language— $\widetilde{B}$ and B for $\arg\circ\widetilde{B}\circ\arg^{-1}$ and $\arg\circ B\circ\arg^{-1}$, respectively. For $a = 0,\dots,m-1$, let $x_a = \arg\theta_a$, and let $h_a : [0,1) \to \mathbb{R}_{\geq 0}$ be the function defined by*

$$h_a(x) = \frac{\pi}{\tan(\pi(x - x_a))} - \frac{\pi}{\tan(\pi(x - x_{a+1}))}$$

on (x_a, x_{a+1}), and having value 0 elsewhere. Then the following facts hold.

(i) The unique (up to constants) $\widetilde{B}$-invariant measure on X absolutely continuous with respect to the Lebesgue measure is $\mathrm{d}\tilde{\mu} = \pi^2\big(\sin(\pi(x-y))\big)^{-2}\,\mathrm{d}x\,\mathrm{d}y$.

(ii) The unique (up to constants) B-invariant measure on $[0,1)$ absolutely continuous with respect to the Lebesgue measure is $\mathrm{d}\mu = \big(\sum_a h_a\big)\,\mathrm{d}x$.

(iii) Both systems $(X,\tilde{\mu},\widetilde{B})$, $([0,1),\mu,B)$ are ergodic and conservative.

Proof. (i) This is just a change of variables, easily performed in two steps. Let $F_1, F_2 : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F_1(x,y) = \big(\pi(x-y), \pi(x+y)\big) = (x',y'),$$
$$F_2(x',y') = \left(\frac{\cos(x'+y')}{1-\sin(x'+y')}, \frac{\cos(-x'+y')}{1-\sin(-x'+y')}\right) = (\omega,\alpha).$$

Then $F_2 \circ F_1$ is a bijection from $[0,1)^2 \setminus \{\text{diagonal}\}$ to $(\mathrm{P}^1\,\mathbb{R} \times \mathrm{P}^1\,\mathbb{R}) \setminus \{\text{diagonal}\}$; indeed, it amounts to the componentwise application of $C^{-1}\circ\arg^{-1}$, with C the Cayley matrix. This implies that the pushforward of the infinite invariant measure $(\omega-\alpha)^{-2}\,\mathrm{d}\omega\,\mathrm{d}\alpha$ of Theorem 4.1 via $\arg\circ C$ is $(F_2\circ F_1)^*\big((\omega-\alpha)^{-2}\,\mathrm{d}\omega\,\mathrm{d}\alpha\big)$. One now computes

$$F_2^*\left(\frac{1}{(\omega-\alpha)^2}\,\mathrm{d}\omega\,\mathrm{d}\alpha\right) = \frac{1/2}{\sin^2(x')}\,\mathrm{d}x'\,\mathrm{d}y',$$
$$F_1^*\left(\frac{1/2}{\sin^2(x')}\,\mathrm{d}x'\,\mathrm{d}y'\right) = \frac{\pi^2}{\sin^2(\pi(x-y))}\,\mathrm{d}x\,\mathrm{d}y.$$

(ii) Let $x \in (x_a, x_{a+1})$. Then $h_a(x)$ is the integral

$$\int_0^{x_a} \frac{\pi^2\,\mathrm{d}y}{\sin^2(\pi(x-y))} + \int_{x_{a+1}}^1 \frac{\pi^2\,\mathrm{d}y}{\sin^2(\pi(x-y))}$$

of the invariant density in (i) along the fiber $\{x\} \times \big([0,x_a] \cup [x_{a+1},1]\big)$.

(iii) It is easy to check that B^2 satisfies Thaler's conditions [47, p. 69(1)–(4)]. This implies that B^2 is ergodic and conservative; therefore so is B and its natural extension $\widetilde{B}$ [1, Theorem 3.1.7]. □

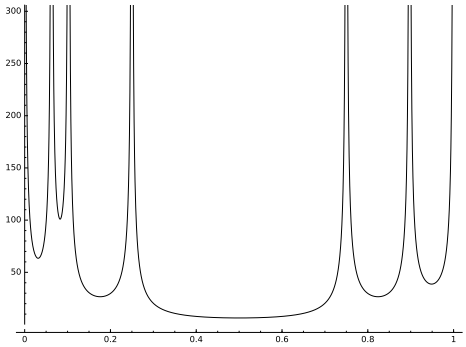


FIGURE 5. The invariant density for the map of Example 6.3

We draw in Figure 5 the invariant density $\sum_a h_a$ for the map B of Example 6.3. We note that, in case $m = 3$, a direct geometric proof of Theorem 7.2(ii) was given by Kołodziej and Misiurewicz, using Ptolemy's theorem on quadrilaterals inscribed in a circle [30], [34].

8. **The Lagrange theorem.** Our next result is a version of Serret's theorem (two real numbers have the same tail in their continued fraction expansion precisely when they are $\mathrm{PSL}_2^{\pm}\,\mathbb{Z}$-equivalent [24, §10.11], [39]) in modern language.

Theorem 8.1. *The map B and the group $\Gamma_B^{\pm}$ are orbit equivalent. More precisely, given $\sigma, \sigma' \in S^1$, there exists $\mathscr{A} \in \Gamma_B^{\pm}$ such that $\sigma' = \mathscr{A} * \sigma$ if and only if there exist $h, k \geq 0$ such that $B^h(\sigma) = B^k(\sigma')$. In particular, if σ belongs to $\mathbb{Q}(i)$ then it is B-terminating, its orbit landing in the unique vertex of D which is Γ_B-equivalent to σ.*

Proof. We begin proving the last assertion, for which the $\partial\mathcal{K}$ setting is expedient. Let then $\boldsymbol{s}$ be a rational point, and let $(\boldsymbol{w}_0)_3, \dots, (\boldsymbol{w}_{m-1})_3 \in \mathbb{Z}$ be the third coordinates of the points $\boldsymbol{w}_0, \dots \boldsymbol{w}_{m-1}$ of Definition 6.1. We need a preliminary step.

Claim. By conjugating B by an appropriate element of $\mathrm{SO}_{2,1}^{\uparrow}\,\mathbb{Z}$, we may assume that $(\boldsymbol{w}_0)_3, \dots, (\boldsymbol{w}_{m-1})_3$ are all greater than 0, with at most one exception that may equal 0.
Proof of Claim. By Lemma 5.1(v), the greater is the arclength of I_a, the smaller is $(\boldsymbol{w}_a)_3$, with $(\boldsymbol{w}_a)_3 = 0$ corresponding to arclength π. This implies that no more than one of the above third coordinates may be negative or 0. Say that $(\boldsymbol{w}_a)_3 < 0$. If I_a is even, then by Theorem 5.3 we may conjugate B by the matrix in $\mathrm{SO}_{2,1}^{\uparrow}\,\mathbb{Z}$ that sends $\boldsymbol{w}_a$ to $(0, 1, 0)$, and we are through. If I_a is odd, than we conjugate by the matrix that sends $\boldsymbol{w}_a$ to $(1, 1, 1)$; the image of I_a will then have arclength $\pi/2$. One of the new third coordinates may now have value 0, but none may have value -1 or less, since value -1 already corresponds to an arclength of $3\pi/2$, and the sum of the arclengths would exceed 2π.

Having proved our claim we perform, if needed, this preliminary conjugation, which does not affect the validity of our statement; renaming indices, we assume $(\boldsymbol{w}_0)_3 \geq 0$ and $(\boldsymbol{w}_1)_3, \dots, (\boldsymbol{w}_{m-1})_3 > 0$. If $\boldsymbol{s}$ is one of $\boldsymbol{t}_0, \dots, \boldsymbol{t}_{m-1}$, we are through. Otherwise, $\boldsymbol{s}$ is in the interior of precisely one interval, say I_a; let $\boldsymbol{s}' = B(\boldsymbol{s})$. Then,

lifting $\boldsymbol{s}$ and $\boldsymbol{s}'$ to their canonical representatives (i.e., to pythagorean triples), we have the identity in $\mathbb{Z}^3$

$$\boldsymbol{s}' = \boldsymbol{A}_a \boldsymbol{s} = \boldsymbol{s} - 2\frac{\langle \boldsymbol{w}_a, \boldsymbol{s}\rangle}{\langle \boldsymbol{w}_a, \boldsymbol{w}_a\rangle}\boldsymbol{w}_a. \tag{15}$$

Now, $\langle \boldsymbol{w}_a, \boldsymbol{w}_a\rangle = 1$ since $\boldsymbol{w}_a \in \mathcal{S}$, and $\langle \boldsymbol{w}_a, \boldsymbol{s}\rangle > 0$ since $\boldsymbol{s}$ is in the interior of I_a. This implies that the third coordinate of $\boldsymbol{s}'$ is strictly less than the third coordinate of $\boldsymbol{s}$, unless $a = 0$ and $(\boldsymbol{w}_0)_3 = 0$, in which case we have equality. But the third coordinates of $\boldsymbol{s}$ and $\boldsymbol{s}'$ are positive integers, and the exceptional case of equality is always preceded and followed by nonexceptional cases. Hence the process must stop, and this may happen only when the B-orbit of $\boldsymbol{s}$ lands in one of the interval endpoints $\boldsymbol{t}_0, \dots, \boldsymbol{t}_{m-1}$.

For the first assertion, the "if" implication is clear. Assume $\sigma' = \mathscr{A} * \sigma$. If one of σ, σ' is in $\mathbb{Q}(i)$ then so is the other, and by the first part of the proof both σ and σ' land in one of $\theta_0, \dots, \theta_{m-1}$. Since the vertices of D are $\Gamma_B^{\pm}$-inequivalent, they must land in the same θ_a. Let then $\sigma, \sigma' \notin \mathbb{Q}(i)$ and $\varphi(\sigma) = \boldsymbol{a}$. As noted in §6, $\mathscr{A}$ factors uniquely as $\mathscr{A} = \mathscr{A}_{b_0} \dots \mathscr{A}_{b_{r-1}}$, for certain $b_0, \dots, b_{r-1} \in \{0, \dots, m-1\}$. Let $0 \le h \le r$ be minimum such that $a_h \ne b_{r-1-h}$. Then

$$\begin{aligned}
\sigma' &= \mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{r-1}} * \sigma \\
&= \mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{r-1}} \mathscr{A}_{a_0} \cdots \mathscr{A}_{a_{r-1}} * B^r(\sigma) \\
&= \mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{r-1-h}} \mathscr{A}_{a_h} \cdots \mathscr{A}_{a_{r-1}} * B^r(\sigma) \\
&= \mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{r-1-h}} * B^h(\sigma).
\end{aligned}$$

By (a) in the proof of Lemma 6.5, $\varphi(\sigma') = b_0 \dots b_{r-1-h} a_h a_{h+1} \dots$, and $B^{r-h}(\sigma') = B^h(\sigma)$. □

The bijection between $\partial\mathcal{D} \cap \mathbb{Q}(i)$ and rational points in $\partial\mathcal{K}$ extends to higher degrees.

Lemma 8.2. *Let $\boldsymbol{s} = [s_1, s_2, s_3] \in \partial\mathcal{K}$ correspond to $\sigma = (s_1 + s_2 i)/s_3 \in \partial\mathcal{D}$ as usual, and let $\omega = C^{-1} * \sigma = (\mu \circ \upsilon)(\boldsymbol{s}) \in \mathrm{P}^1\mathbb{R}$. Then $\mathbb{Q}(\boldsymbol{s}) = \mathbb{Q}(\omega)$ and $[\mathbb{Q}(\omega) : \mathbb{Q}] = [\mathbb{Q}(i)(\sigma) : \mathbb{Q}(i)]$. If $\mathbb{Q}(\omega)/\mathbb{Q}$ is Galois totally real, then the Galois groups $\mathrm{Gal}(\mathbb{Q}(\omega)/\mathbb{Q})$ and $\mathrm{Gal}(\mathbb{Q}(i)(\sigma)/\mathbb{Q}(i))$ are naturally isomorphic. In particular, assume that σ is quadratic over $\mathbb{Q}(i)$ and let σ' be its Galois conjugate. Then $\sigma' \in \partial\mathcal{D}$ and $\omega' = C^{-1} * \sigma'$ is the Galois conjugate of ω with respect to the quadratic extension $\mathbb{Q}(\omega)/\mathbb{Q}$.*

Proof. Since the stereographic projection through $[0, 1, 1]$ is a rational map with rational coefficients, the identity $\mathbb{Q}(\boldsymbol{s}) = \mathbb{Q}(\omega)$ holds (with the convention that $\mathbb{Q}(\infty) = \mathbb{Q}$). All statements follow from elementary Galois theory, as soon as one realizes that $\mathbb{Q}(i, \sigma) = \mathbb{Q}(i, s_1/s_3, s_2/s_3)$. In this identity the left-to-right containment is obvious, and the other one follows from $s_1/s_3 = (\sigma + \sigma^{-1})/2$. □

The question of the validity of Lagrange's theorem (preperiodic points correspond to quadratic irrationals) for the Romik map is left open in [42, §5.1]. It can be settled in the affirmative by the result in [38]; see also [14] for this issue, and [13] for diophantine approximation aspects of the Romik map. Here we provide a different proof, valid not only for the Romik map but for all maps based on unimodular partitions. Note that our proof covers not only Lagrange's, but Galois's theorem [40, Chapter III]: periodic points correspond to reduced irrationals.

Theorem 8.3. *The point $\sigma \in S^1$ is B-preperiodic if and only if it is quadratic over $\mathbb{Q}(i)$. If this is the case and $a_0 \dots a_{h-1}\overline{a_h \dots a_{h+p-1}}$ is the B-symbolic sequence of σ (with p the minimal period and h the minimal preperiod, so that $a_{h-1} \neq a_{h+p-1}$), then the B-symbolic sequence of the Galois conjugate σ' is $a_0 \dots a_{h-1}\overline{a_{h+p-1} \dots a_h}$. In particular, the preperiodic σ is periodic iff so is σ' iff $(\sigma, \sigma') \in \mathcal{S}_B$.*

Proof. Let σ be B-preperiodic. Clearly, for every $\mathscr{A} \in \mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$, we have $\mathbb{Q}(i)(\mathscr{A} * \sigma) = \mathbb{Q}(i)(\sigma)$; we can then assume that σ is B-periodic, with B-symbolic sequence $\overline{a_0 a_1 \dots a_{p-1}}$. Let $\mathscr{B} = \mathscr{A}_{a_0}\mathscr{A}_{a_1} \cdots \mathscr{A}_{a_{p-1}}$. By looking at the decreasing sequence (14) in the proof of Lemma 6.5, we obtain

$$\bigcap_{n \geq 0} \mathscr{B}^n[\overline{I}_{a_0}] = \{\sigma\}.$$

Since $\mathscr{B} * \sigma$ is also in the above intersection, it equals σ, and this yields a quadratic polynomial with coefficients in $\mathbb{Q}(i)$ and having σ as root. This polynomial is not the zero polynomial, as $\mathscr{B}$ is not the identity matrix, and is irreducible over $\mathbb{Q}(i)$ because σ is B-nonterminating and Theorem 8.1 applies.

Conversely, let $\sigma \in S^1$ be quadratic over $\mathbb{Q}(i)$. By Lemma 8.2 the conjugate σ' is in S^1 as well. For $t \geq 0$, let $\widetilde{B}^t(\sigma, \sigma') = (\sigma_t, \sigma'_t)$, and let g_t be the oriented geodesic of origin σ'_t and endpoint σ_t. By Theorem 7.1 there exists $h \geq 0$ such that, for $0 \leq t < h$, the points σ_t and σ'_t belong to the same interval (so that g_t does not cut the billiard table D), while g_t cuts D for every $t \geq h$. In particular, the B-symbolic sequences of σ and σ' agree up to time $h-1$ included, and disagree at time h. Let $\omega = C^{-1} * \sigma_h$, $\omega' = C^{-1} * \sigma'_h$; since σ_t and σ'_t are still conjugate in $\mathbb{Q}(i)(\sigma)/\mathbb{Q}(i)$, by Lemma 8.2 ω and ω' are conjugate in $\mathbb{Q}(\omega)/\mathbb{Q}$. Let $O = \{\xi \in \mathbb{Q}(\omega) : \xi(\mathbb{Z}\omega + \mathbb{Z}) \subseteq \mathbb{Z}\omega + \mathbb{Z}\}$ be the coefficient ring of the module $\mathbb{Z}\omega + \mathbb{Z}$ [8, Chapter 2 §2.2]. Then O is an order in $\mathbb{Q}(\omega)$ with fundamental unit $\varepsilon > 1$, and thus the matrix

$$H = \begin{pmatrix} \omega & \omega' \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon & \\ & \varepsilon' \end{pmatrix} \begin{pmatrix} \omega & \omega' \\ 1 & 1 \end{pmatrix}^{-1} \tag{16}$$

(where ε' is the conjugate of ε) is in $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$.

Now, $\langle F, P, J \rangle = C^{-1}\big(\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]\big)C$ is an index-3 subgroup of $\mathrm{PSL}^{\pm}_2\,\mathbb{Z}$ (see the end of the proof of Theorem 2.4), and $\Gamma^{\pm}_B$ is a finite-index subgroup of $\mathrm{PSU}^{\pm}_{1,1}\,\mathbb{Z}[i]$ (see §6). Hence, replacing H with an appropriate power, we obtain a matrix $\mathscr{H}^l = CH^lC^{-1} \in \Gamma^{\pm}_B$ which induces on $\mathcal{D}$ either a hyperbolic translation of axis g_h (if $\det \mathscr{H}^l = 1$), or a glide reflection, again of axis g_h (if $\det \mathscr{H}^l \neq 1$). As noted in §6, $\mathscr{H}^l$ can be uniquely written as $\mathscr{H}^l = \mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{q-1}}$ for certain $b_0, \dots, b_{q-1} \in \{0, \dots, m-1\}$. We claim that $\overline{b_0 \cdots b_{q-1}}$ and $\overline{b_{q-1} \cdots b_0}$ are the B-symbolic sequences of σ_h and σ'_h, respectively (q might be a proper multiple of the minimal period p); this will conclude the proof of Theorem 8.3.

We must have $b_0 \neq b_{q-1}$. Indeed, if not, then $\mathscr{H}^l$ would factor as

$$\mathscr{H}^l = (\mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{t-1}})(\mathscr{A}_{b_t} \cdots \mathscr{A}_{b_{t+k-1}})(\mathscr{A}_{b_{t-1}} \cdots \mathscr{A}_{b_0}),$$

for some $k \geq 2$, with $t = (q-k)/2$ and $b_t \neq b_{t+k-1}$. Hence g_h would be the $(\mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{t-1}})$-image of the geodesic stabilized by $(\mathscr{A}_{b_t} \cdots \mathscr{A}_{b_{t+k-1}})$, which has endpoints in the two distinct intervals I_{b_t} and $I_{b_{t+k-1}}$. Since b_t and b_{t+k-1} are different from b_{t-1}, the endpoints of g_h would both lie in I_{b_0}, which is impossible since g_h cuts D; therefore $b_0 \neq b_{q-1}$.

The sequence $\overline{b_0 \cdots b_{q-1}}$ satisfies (i) in Lemma 6.5 (because $b_0 \neq b_{q-1}$), as well as (ii) (because otherwise $\mathscr{H}^l$ would be a power of some $\mathscr{A}_a\mathscr{A}_{a+1}$ and thus would be

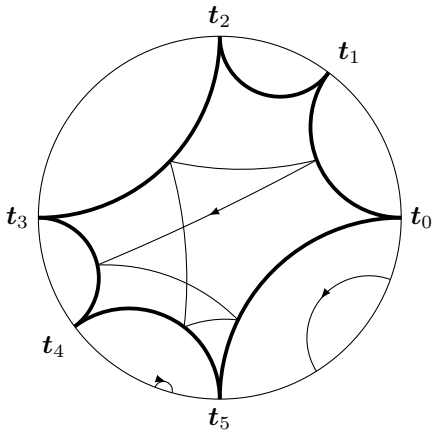


FIGURE 6. A periodic orbit in a billiard table

parabolic, which is not possible because any power of the matrix in (16) has trace of absolute value greater than 2). Therefore, $\overline{b_0 \cdots b_{q-1}}$ is the B-symbolic sequence of a unique point of S^1, and this point is necessarily σ_h, because σ_h is the ideal endpoint of g_h, and thus the shrinking point of

$$\bigcap_{n\geq 0} (\mathscr{A}_{b_0} \cdots \mathscr{A}_{b_{q-1}})^n [\overline{I}_{b_0}].$$

The same argument, applied to $\mathscr{H}^{-1} = \mathscr{A}_{b_{q-1}} \cdots \mathscr{A}_{b_0}$, shows that σ'_h has B-symbolic sequence $\overline{b_{q-1} \cdots b_0}$. □

Example 8.4. Consider the unimodular partition given by the pythagorean triples

$$\boldsymbol{t}_0 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \boldsymbol{t}_1 = \begin{bmatrix}3\\4\\5\end{bmatrix}, \boldsymbol{t}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix}, \boldsymbol{t}_3 = \begin{bmatrix}-1\\0\\1\end{bmatrix}, \boldsymbol{t}_4 = \begin{bmatrix}-4\\-3\\5\end{bmatrix}, \boldsymbol{t}_5 = \begin{bmatrix}0\\-1\\1\end{bmatrix};$$

in Figure 6 we draw the corresponding billiard table by thick geodesics.

Let $q(x, y) = 4091x^2 + 1302xy + 101y^2$, which has discriminant $D = 42440$. The roots of $q(x, 1)$ are

$$\omega_0 = \frac{-1302 + \sqrt{D}}{2 \cdot 4091} \simeq -0.13395, \quad \alpha_0 = \frac{-1302 - \sqrt{D}}{2 \cdot 4091} \simeq -0.18430.$$

We work directly on the de Sitter space; by (9), q corresponds to

$$\frac{1}{\sqrt{D}} \begin{pmatrix} & 1 & \\ -1 & & 1 \\ 1 & & 1 \end{pmatrix} \begin{pmatrix} 4091 \\ -1302 \\ 101 \end{pmatrix} \in \mathcal{S}.$$

Since we may safely multiply by a constant, and we prefer working with integer vectors, we multiply by $\sqrt{D}/2$ and define

$$\boldsymbol{v} = \frac{1}{2} \begin{pmatrix} & 1 & \\ -1 & & 1 \\ 1 & & 1 \end{pmatrix} \begin{pmatrix} 4091 \\ -1302 \\ 101 \end{pmatrix} = \begin{pmatrix} -651 \\ -1995 \\ 2096 \end{pmatrix} \in \frac{\sqrt{D}}{2} \mathcal{S} \cap \mathbb{Z}^3.$$

By the equivariance between (S1) and (S5) in Theorem 4.1, the billiard map $\widetilde{B}$ on [any dilated copy of] $\mathcal{S}$ is piecewise defined by the following matrices in $\mathrm{SO}_{2,1}\,\mathbb{Z}$:

$$\Lambda(A_0) = -\boldsymbol{A}_0 = \begin{pmatrix} 7 & 4 & -8 \\ 4 & 1 & -4 \\ 8 & 4 & -9 \end{pmatrix}, \qquad -\boldsymbol{A}_1 = \begin{pmatrix} 1 & 6 & -6 \\ 6 & 17 & -18 \\ 6 & 18 & -19 \end{pmatrix},$$
$$-\boldsymbol{A}_2 = \begin{pmatrix} 1 & -2 & 2 \\ -2 & 1 & -2 \\ -2 & 2 & -3 \end{pmatrix}, \qquad -\boldsymbol{A}_3 = \begin{pmatrix} 17 & 6 & 18 \\ 6 & 1 & 6 \\ -18 & -6 & -19 \end{pmatrix},$$
$$-\boldsymbol{A}_4 = \begin{pmatrix} 1 & 4 & 4 \\ 4 & 7 & 8 \\ -4 & -8 & -9 \end{pmatrix}, \qquad -\boldsymbol{A}_5 = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & 2 \\ 2 & -2 & -3 \end{pmatrix}.$$

In order to apply $\widetilde{B}$ we must determine the pair $(\boldsymbol{s}, \boldsymbol{r}) \in (S^1 \times S^1) \setminus (\text{diagonal})$ associated to $\boldsymbol{v}$, and the interval I_a to which $\boldsymbol{s}$ belongs. The intervals $I_0, \dots, I_5$ correspond as in Definition 6.1 to the points in $\mathcal{S}$

$$\boldsymbol{w}_0 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} -3 \\ -1 \\ 3 \end{pmatrix}, \boldsymbol{w}_4 = \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix}, \boldsymbol{w}_5 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

A straightforward computation along the lines of the proof of Theorem 4.1 shows that $\boldsymbol{s}, \boldsymbol{r}$ are given, as a function of $\boldsymbol{v} \in (\sqrt{D}/2)\mathcal{S}$, by

$$\boldsymbol{s} = \begin{bmatrix} -(v_1 + \sqrt{D}/2)(v_2 - v_3) \\ v_1(v_1 + \sqrt{D}/2) + v_3(v_2 - v_3) \\ v_1(v_1 + \sqrt{D}/2) + v_2(v_2 - v_3) \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} -(v_1 - \sqrt{D}/2)(v_2 - v_3) \\ v_1(v_1 - \sqrt{D}/2) + v_3(v_2 - v_3) \\ v_1(v_1 - \sqrt{D}/2) + v_2(v_2 - v_3) \end{bmatrix},$$

and that the 3rd coordinates s_3, r_3 displayed above are always strictly positive. This implies that all values $\langle \boldsymbol{w}_0, \boldsymbol{s} \rangle, \dots, \langle \boldsymbol{w}_5, \boldsymbol{s} \rangle$ are strictly negative, with precisely one strictly positive exception. The index a of that exception is the index of the interval I_a to which $\boldsymbol{s}$ belongs, and thus the index of the matrix $-\boldsymbol{A}_a$ to be applied.

In our case, $\langle \boldsymbol{w}_4, \boldsymbol{s} \rangle = 1.64125\ldots$ and $\langle \boldsymbol{w}_4, \boldsymbol{r} \rangle = 1.94758\ldots$; thus both $\boldsymbol{s} = \boldsymbol{s}_0$ and $\boldsymbol{r} = \boldsymbol{r}_0$ lie in I_4, and the $\widetilde{B}$-image of $\boldsymbol{v} = \boldsymbol{v}_0$ is $-\boldsymbol{A}_4 \boldsymbol{v}_0 = (-247, 199, -300) = \boldsymbol{v}_1$. Repeating the computation we see that both $\boldsymbol{s}_1$ and $\boldsymbol{r}_1$ are in I_5, so that $\boldsymbol{v}_2 = -\boldsymbol{A}_5 \boldsymbol{v}_1 = (-45, 93, 8)$. Now $\boldsymbol{s}_2$ and $\boldsymbol{r}_2$ belong to different intervals, namely the 3rd and the 0th; thus $\boldsymbol{v}_2$ belongs to $\mathcal{S}_B$ and the periodicity starts. Proceeding with the computation we obtain

$$\begin{pmatrix} -651 \\ -1995 \\ 2096 \end{pmatrix} \mapsto \begin{pmatrix} -247 \\ 199 \\ -300 \end{pmatrix} \mapsto \begin{pmatrix} -45 \\ 93 \\ 8 \end{pmatrix} \mapsto \begin{pmatrix} -63 \\ -129 \\ 100 \end{pmatrix} \mapsto$$
$$\begin{pmatrix} -5 \\ 197 \\ -168 \end{pmatrix} \mapsto \begin{pmatrix} 111 \\ 15 \\ -44 \end{pmatrix} \mapsto \begin{pmatrix} -7 \\ -119 \\ -60 \end{pmatrix} \mapsto \begin{pmatrix} -45 \\ 93 \\ 8 \end{pmatrix}.$$

The B-symbolic sequence of ω_0 is thus $45\overline{35420}$, and that of α_0 is $45\overline{02453}$. We draw in Figure 6 the resulting billiard trajectory, along with the two geodesics corresponding to the preperiodic points $\boldsymbol{v}_0$ and $\boldsymbol{v}_1$.

9. **Minkowski functions.** Let $B : S^1 \to S^1$ be the factor of some fixed billiard map as in Definition 6.2. Clearly B is an orientation-reversing $(m-1)$-to-1 covering map of S^1 onto itself. The same properties are shared by precisely one continuous

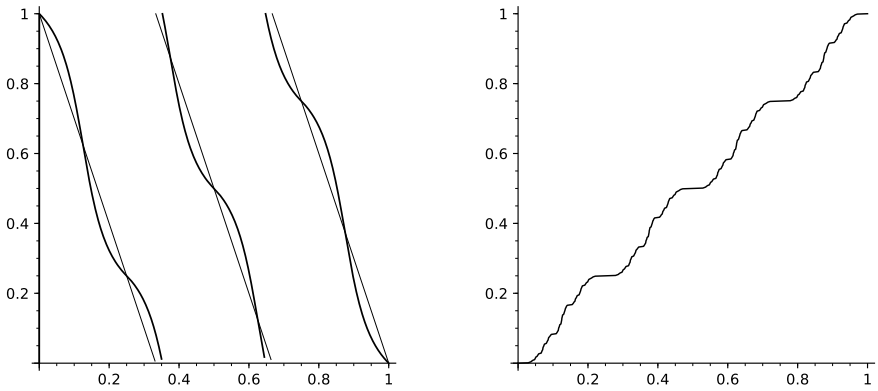


FIGURE 7. Superimposed graphs of B and T, and the resulting Minkowski function.

group homomorphism $T : S^1 \to S^1$, namely $T(z) = z^{-(m-1)}$. In this section we prove that there exists a self-homeomorphism Φ of S^1 that conjugates B with T. We provide an explicit expression for Φ, and prove that Φ is unique up to postcomposition with the elements of the dihedral group of order $2m$. In the final section we will show that Φ is purely singular with respect to the Lebesgue measure on S^1, and Hölder continuous with exponent equal to $\log(m-1)$ divided by the maximal periodic mean free path in the hyperbolic billiard associated to $\widetilde{B}$.

Example 9.1. The prototype of such homeomorphisms is the Minkowski *question mark* function, which conjugates the Farey map $x \mapsto \min(x/(1-x), (1-x)/x)$ on $[0,1]$ with the tent map $x \mapsto \min(2x, -2x+2)$, see [43], [27], [7] and references therein. For an example in our setting, let us consider the unimodular partition determined by $1, i, -1, -i$; we have then a "square billiard table". For ease of visualization we look at B and T as maps from $[0,1)$ to itself; in particular, $T(x) = -3x \pmod 1$. We show in Figure 7 (left) the superimposed graphs of B and T, and the resulting function Φ (right). As noted in Example 6.3, B is defined via 4 pieces, with endpoints the indifferent fixed points 0, 1/4, 1/2, 3/4, and has (apparent) discontinuities at 0, $\arg(\mathscr{A}_1 * 1) = \arccos(-3/5)/(2\pi) = 0.35241\ldots$, $\arg(\mathscr{A}_2 * 1) = 1 - \arg(\mathscr{A}_1 * 1)$. In this quite specific case T shares the set of fixed points (which of course are now expansive) with B; the graph of T has (apparent) discontinuities at 0, 1/3, 2/3. We will return to this example at the end of the paper.

In order to state the next result, we recall that the torsion subgroup S^1_{tor} of S^1 is the internal direct sum of the Prüfer groups $S^1_{p\text{-tor}} = \{\sigma \in S^1 : \text{ord}(\sigma) \text{ is a power of } p\}$, for p ranging over the primes. We let $\zeta = \exp(2\pi i/(m-1))$.

Theorem 9.2. *There exists a homeomorphism $\Phi : S^1 \to S^1$ such that $\Phi \circ B = T \circ \Phi$. This homeomorphism is unique up to postcomposition with elements of the dihedral group $z \mapsto \zeta^h z^e$, with $h \in \{0, \ldots, m-1\}$ and $e \in \{-1, 1\}$. The map Φ establishes a bijection between the set of points in S^1 of degree ≤ 2 over $\mathbb{Q}(i)$ and S^1_{tor}, the set $S^1 \cap \mathbb{Q}(i)$ corresponding to the direct sum of the subgroup $\langle\zeta\rangle$ generated by ζ and the finitely many $S^1_{p\text{-}tor}$, for $p \mid m-1$.*

Before proving Theorem 9.2 we need some preliminaries. We already encountered the ternary betweenness relation on S^1 in §5, and we now introduce the same relation

on the index set $\{0,\dots,m-1\}$, cyclically ordered in the natural way. The powers of ζ determine a partition of S^1 in the half-open intervals $J_a = \{\zeta^a\} \cup \{x : \zeta^a \prec x \prec \zeta^{a+1}\} = [\zeta^a, \zeta^{a+1})$. We define a binary relation $<_B$ on S^1 as follows: $\sigma <_B \sigma'$ if and only if σ and σ' lie in the same interval I_a, for some $a \in \{0,\dots,m-1\}$, and $\arg(\sigma) < \arg(\sigma')$. The relation $<_T$ is defined in the analogous way, using the intervals J_a. Precisely as in Definition 6.4, but using the intervals J_a, we introduce the *T-symbolic-sequence* map $\psi : S^1 \to \{0,\dots,m-1\}^\omega$.

Lemma 9.3. *All statements in Lemma 6.5 hold for ψ; in particular φ and ψ have identical range $X \subset \{0,\dots,m-1\}^\omega$, which is described by (i) and (ii) in that lemma. The betweenness and the $<_B$ relations on S^1 are characterized in terms of B-symbolic sequences and the betweenness relation on $\{0,\dots,m-1\}$ as follows: let $\varphi(\sigma) = \boldsymbol{a}$, $\varphi(\sigma') = \boldsymbol{a}'$, $\varphi(\sigma'') = \boldsymbol{a}''$. Then:*

- *(1) $\sigma <_B \sigma'$ if and only if there exists $t \geq 0$ such that:*
  - *(1.1) $a_h = a'_h$ for every $0 \leq h \leq t$,*
  - *(1.2) $a_{t+1} \neq a'_{t+1}$,*
  - *(1.3) one of the following mutually exclusive conditions holds:*
    - *(1.3.1) t is even and ($a_{t+1} = a_t$ or $a_{t+1} \prec a_t \prec a'_{t+1}$),*
    - *(1.3.2) t is odd and ($a'_{t+1} = a'_t$ or $a'_{t+1} \prec a_t \prec a_{t+1}$);*
- *(2) $\sigma \prec \sigma' \prec \sigma''$ if and only if one of the following mutually exclusive conditions holds:*
  - *(2.1) $a_0 \prec a'_0 \prec a''_0$,*
  - *(2.2) $a_0 = a'_0 \neq a''_0$ and $\sigma <_B \sigma'$,*
  - *(2.3) $a_0 \neq a'_0 = a''_0$ and $\sigma' <_B \sigma''$,*
  - *(2.4) $a_0 = a'_0 = a''_0$ and $\sigma <_B \sigma'$ and $\sigma' <_B \sigma''$.*

We have an analogous characterization of betweenness and $<_T$ in terms of T-symbolic sequences.

Proof. The proof of Lemma 6.4 easily extends to the case of the map T. Apart from the obvious modifications (use J_a for I_a, and ζ^a for θ_a), one has to replace the occurrences of B with occurrences of T, and those of $\mathscr{A}_a$ with T_a^{-1}, the latter being the ath inverse branch of T, i.e., the map that associates to $\sigma \in \bigcup_{b\neq a} \overline{J}_b$ its unique $-(m-1)$th root lying in $\overline{J}_a$. The fact that no T-symbolic sequence has tail $\overline{a(a+1)}$ is easy; indeed, any point having that symbolic sequence should jump forever from J_a to J_{a+1}. But at each jump its arclength distance from the fixed point ζ^{a+1} increases by a factor $m-1$, so the point will eventually escape from $J_a \cup J_{a+1}$. Finally, the analogue of the sequence (14) surely shrinks to a singleton, because at each step the arclengths shrink by a factor $m-1$. With these modifications, the proof carries through verbatim.

We prove statement (1). Suppose σ and σ' are different, but lie in the same interval I_{a_0}. Then there exists $t \geq 0$ such that for t steps the successive B-images of σ and σ' keep on lying in the same interval, while $B^{t+1}(\sigma)$ and $B^{t+1}(\sigma')$ lie in the different intervals $I_{a_{t+1}}$ and $I_{a'_{t+1}}$, respectively. Since B is orientation-reversing, $\sigma <_B \sigma'$ if and only if either t is even and $B^t(\sigma) <_B B^t(\sigma')$, or t is odd and $B^t(\sigma') <_B B^t(\sigma)$. We can then assume without loss of generality $t = 0$, and observe that $\sigma <_B \sigma'$ holds if and only if $\sigma = \theta_{a_0}$ (which is equivalent to $a_1 = a_0$), or $B(\sigma) \prec \theta_{a_0} \prec B(\sigma')$ (which is equivalent to $a_1 \prec a_0 \prec a'_1$, since now $B(\sigma)$ and $B(\sigma')$ lie in different intervals, both different from I_{a_0}).

Statement (2) is clear, as is the fact that all of the proof applies to the map T. □

Proof of Theorem 9.2. Let S be the shift on $X = \varphi[S^1] = \psi[S^1]$, and define $\Phi = \psi^{-1} \circ \varphi$. Then the inner squares in

$$\begin{array}{ccc} S^1 & \xrightarrow{B} & S^1 \\ \Big\downarrow{\scriptstyle \varphi} & & \Big\downarrow{\scriptstyle \varphi} \\ X & \xrightarrow{S} & X \\ \Big\uparrow{\scriptstyle \psi} & & \Big\uparrow{\scriptstyle \psi} \\ S^1 & \xrightarrow{T} & S^1 \end{array} \qquad \Phi \circlearrowright \Phi \tag{17}$$

commute, so the outer rectangle commutes as well. Let σ, σ', σ'' be distinct points of S^1. Then $\sigma \prec \sigma' \prec \sigma''$ holds if and only if the conditions of Lemma 9.3 apply to $\varphi(\sigma)$, $\varphi(\sigma')$, $\varphi(\sigma'')$. By construction, $\varphi(\sigma) = \psi\big(\Phi(\sigma)\big)$ and analogously for σ' and σ''; therefore σ' is between σ and σ'' if and only if $\Phi(\sigma')$ is between $\Phi(\sigma)$ and $\Phi(\sigma'')$. Since the topology of S^1 is definable in terms of betweenness, Φ is a homeomorphism.

Let Φ_1 be any homeomorphism that makes the outer rectangle in (17) commute. For every $h \in \{0, \dots, m-1\}$ and every $e \in \{1, -1\}$, the map $Q(z) = \zeta^h z^e$ commutes with T, so that $Q \circ \Phi_1$ too makes the outer rectangle commute. We therefore assume that Φ_1 is orientation-preserving and fixes 1, and prove $\Phi_1 = \Phi$. As Φ_1 and Φ are homeomorphisms and the set of B-terminating points is dense in S^1, it is enough to show that Φ_1 agrees with Φ on this set; in other words, that if σ has B-symbolic sequence $a_0 \dots a_{t-1}\overline{a_t}$ with $a_{t-1} \neq a_t$, then $\Phi_1(\sigma)$ has T-symbolic sequence $a_0 \dots a_{t-1}\overline{a_t}$.

We work by induction on t. If $t = 0$, then $\sigma = \theta_{a_0}$. Since Φ_1 is orientation-preserving, sends the set $\{\theta_0, \dots, \theta_{m-1}\}$ of B-fixed points to the set $\{\zeta^0, \dots, \zeta^{m-1}\}$ of T-fixed points, and fixes $1 = \theta_0 = \zeta^0$, we have $\Phi_1(\theta_a) = \zeta^a$ for every a. In particular, $\Phi_1(\sigma) = \zeta^{a_0}$, which has T-symbolic sequence $\overline{a_0}$. Let $t > 0$; then $a_0 \neq a_1$, which implies $\sigma \neq \theta_{a_0}$ and $\Phi_1(\sigma) \neq \zeta^{a_0}$. By the inductive hypothesis, the statement is true for all points that land in a B-fixed point in $t - 1$ steps. Since $B(\sigma)$ is one of these points, we have

$$\varphi\big(B(\sigma)\big) = \psi\big(\Phi_1(B(\sigma))\big) = \psi\big(T(\Phi_1(\sigma))\big) = a_1 \dots a_{t-1}\overline{a_t}.$$

Thus $\psi\big(\Phi_1(\sigma)\big) = ba_1 \dots a_{t-1}\overline{a_t}$ for some b, and we must show $b = a_0$. Suppose not; then we have $\zeta^{a_0} \prec \zeta^b \prec \Phi_1(\sigma)$, while $\zeta^{a_0} \prec \Phi(\sigma) \prec \zeta^b$. Applying the order-preserving homeomorphism Φ_1^{-1} to the former relation, and Φ^{-1} to the latter, we get $\theta_{a_0} \prec \theta_b \prec \sigma$ and $\theta_{a_0} \prec \sigma \prec \theta_b$, which is impossible; therefore $b = a_0$ and our first statement is proved.

By Theorems 8.1 and 8.3 the set of points in S^1 of degree 1 (respectively, 2) over $\mathbb{Q}(i)$ is the set of B-terminating (respectively, B-preperiodic) points. Their Φ-images are then the T-terminating (respectively, T-preperiodic) points. It is easily seen the every T-terminating or T-preperiodic point must have the form $\exp(2\pi i q)$ for some rational number q, i.e., must lie in S^1_{tor}. We have the decomposition $S^1_{\text{tor}} = H_1 \cdot H_2$, where H_1 (respectively, H_2) is the inner sum of all Prüfer groups $S^1_{p\text{-tor}}$ with $p \nmid m - 1$ (respectively, $p \mid m - 1$). Now, given $\sigma \in S^1_{\text{tor}}$, repeated applications of T kill the H_2 part, and as soon as this happens the periodicity starts. More precisely, let $h \geq 0$ be minimum such that $T^h(\sigma) \in H_1$. Then $T^h(\sigma)$ is T-periodic, because raising to the $-(m-1)$th power is an automorphism of H_1 of finite order. In particular, σ is T-terminating if and only if $T^h(\sigma)$ is a fixed point, i.e., a power of ζ. Thus, σ is T-terminating precisely when it belongs to $\langle\zeta\rangle \cdot H_2$. □

We note as an aside that the pushforward probability measure $\Phi_*^{-1}\lambda$, where λ is the Lebesgue measure on the circle, is B-invariant, and is the measure of maximal entropy for B.

For the rest of this paper we consider B, T, Φ as selfmaps of $[0,1)$, as in Figure 7. This improves visualization, and makes $\Phi = \psi^{-1} \circ \varphi$ the unique homeomorphism of $[0,1)$ (with the topology inherited from $\mathbb{R}$, not from S^1) that conjugates B with T. Accordingly, $<$ will now denote the standard non-circular orders on $[0,1)$ and on $\{0,\dots,m-1\}$. We will abuse language by writing I_a and J_a for the arg-images in $[0,1)$ of the intervals I_a and J_a of S^1.

In the next Theorem 9.4 we provide an explicit formula for $\Phi(x)$, analogous to the Denjoy-Salem formula for the classical case [19], [43, pp. 435-436], and to the formula in [7, Theorem 1] for the Minkowski function induced by the Romik map. We define a function $d : \{0,\dots,m-1\}^2 \setminus \{\text{diagonal}\} \to \{0,\dots,m-1\}$ by

$$d(a,b) = \begin{cases} a+1, & \text{if } a < b; \\ a, & \text{otherwise.} \end{cases}$$

Theorem 9.4. *Let $x \in [0,1)$ have B-symbolic sequence $\boldsymbol{a}$. Then*

$$\Phi(x) = \frac{1}{m-1} \sum_{t=0}^{\infty} d(a_t, a_{t+1}) \left(-\frac{1}{m-1}\right)^t. \tag{18}$$

Proof. The statement amounts to saying that $\psi^{-1}(\boldsymbol{a})$ equals the value of the absolutely convergent series on the right-hand side of (18). By construction,

$$\psi^{-1}(\boldsymbol{a}) = \lim_{n\to\infty} T_{a_0}^{-1} T_{a_1}^{-1} \cdots T_{a_{n-1}}^{-1}(0),$$

where $T_{a_t}^{-1}$ is the a_tth inverse branch of T discussed in the proof of Lemma 9.3 (instead of 0, any point in $[0,1)$ would do). We recall that, by definition, T_a^{-1} is that inverse branch of T that sends $\bigcup_{b\neq a} \overline{J}_b$ onto $\overline{J}_a$. Here a picture may help: rotate the graph of T in Figure 7 (left) along the diagonal, and look at its $m = 4$ inverse branches, the first two being

$$T_0^{-1}(x) = -x/3 + 1/3, \quad \text{on } [1/4, 1];$$

$$T_1^{-1}(x) = \begin{cases} -x/3 + 1/3, & \text{on } [0, 1/4]; \\ -x/3 + 2/3, & \text{on } [1/2, 1]. \end{cases}$$

A brief pondering over such a picture shows that $T_a^{-1}(x)$ equals $-x/(m-1) + (a+1)/(m-1)$ on $\bigcup_{b>a} \overline{J}_b$, and equals $-x/(m-1) + a/(m-1)$ on $\bigcup_{b<a} \overline{J}_b$; in short,

$$T_{a_t}^{-1}(x) = -\frac{x}{m-1} + \frac{d(a_t, a_{t+1})}{m-1}.$$

Applying induction to the above formula one easily proves that

$$T_{a_0}^{-1} T_{a_1}^{-1} \cdots T_{a_{n-1}}^{-1}(0) = \frac{1}{m-1} \sum_{t=0}^{n-1} d(a_t, a_{t+1}) \left(-\frac{1}{m-1}\right)^t,$$

(where we set $a_n = 0$), and the statement follows by letting n tend to infinity. □

If x is B-preperiodic, (18) yields a finite expression for $\Phi(x)$. Indeed, writing for short $d_t = d(a_t, a_{t+1})$ and $\boldsymbol{d} = d_0 d_1 \dots$, we have that the map $\boldsymbol{a} \mapsto \boldsymbol{d}$ is shift-invariant; in particular, it sends preperiodic sequences to preperiodic ones. Hence,

for $\boldsymbol{a} = \varphi(x)$ and $\boldsymbol{d} = \boldsymbol{d}(\boldsymbol{a}) = d_0 \dots d_{h-1}\overline{d_h \dots d_{h+p-1}}$ we set

$$y = \sum_{t=0}^{h-1} d_t \left(-\frac{1}{m-1}\right)^t, \qquad z = \sum_{t=0}^{p-1} d_{h+t} \left(-\frac{1}{m-1}\right)^t,$$

and obtain by a straightforward computation

$$\Phi(x) = \psi^{-1}(\boldsymbol{a}) = \frac{1}{m-1}\left(y + \frac{(-1)^h (m-1)^{-h} z}{1 + (-1)^{p+1}(m-1)^{-p}}\right). \tag{19}$$

Example 9.5. The point ω_0 of Example 8.4 has B-symbolic sequence $\boldsymbol{a} = 45\overline{35420}$, and $m = 6$. Thus $\boldsymbol{d} = 55\overline{45421}$ and, applying (19),

$$\psi^{-1}(\boldsymbol{a}) = \frac{32243}{39075} = \frac{1}{3} + \frac{11}{25} + \frac{27}{521}.$$

Multiplying successively by $-(m-1) = -5$, and working in $\mathbb{Q}/\mathbb{Z} \simeq S^1_{\mathrm{tor}}$, the summand $1/3$ is fixed (because $-5 \equiv 1$ modulo 3), and $11/25$ gets killed in two steps. So it only remains the summand $27/521$, which yields a periodic orbit of length 5 (because -5 has order 5 modulo 521), as expected.

The Galois conjugate α_0 of ω_0 has B-symbolic sequence $\boldsymbol{a}' = 45\overline{02453}$ and

$$\psi^{-1}(\boldsymbol{a}') = \frac{62873}{78150} = \frac{1}{2} + \frac{2}{3} + \frac{23}{25} + \frac{374}{521} + \text{integer part},$$

with identical dynamical behaviour. The appearance of the same primes at the denominators is not surprising. Indeed, given a periodic orbit of length p, a simple computation shows that the only primes whose powers may appear as denominators of summands are those dividing $(m-1)^p + (-1)^{p+1}$, in our case 2, 3, 521.

10. **Singularity and Hölder exponent.** We maintain the setting described before Theorem 9.4. Since Φ is a monotonically increasing homeomorphism of $[0,1)$, it is differentiable λ-a.e. (λ referring to the Lebesgue measure) with finite derivative.

Theorem 10.1. *The function Φ is purely singular (i.e., $\Phi' = 0$ λ-a.e.).*

We need a preliminary lemma, for which we refer to the notation introduced in Definition 6.1.

Lemma 10.2. *For every a, we have $\boldsymbol{w}_{a-1} + \boldsymbol{w}_a = q_a \boldsymbol{t}_a$ for some $q_a \in \mathbb{Z}_{>0}$. Moreover, the identities*

$$\begin{aligned} \boldsymbol{A}_{a-1}\boldsymbol{w}_a &= \boldsymbol{w}_{a-1} + q_a \boldsymbol{t}_a, \\ \boldsymbol{A}_a \boldsymbol{w}_{a-1} &= \boldsymbol{w}_a + q_a \boldsymbol{t}_a, \end{aligned} \tag{20}$$

hold.

Proof. It is easy to show that $\langle \boldsymbol{w}_{a-1}, \boldsymbol{w}_a \rangle = -1$; for example, applying an appropriate element of $\mathrm{SO}^{\uparrow}_{2,1}\mathbb{R}$ we may assume $\boldsymbol{t}_{a-1} = [0,-1,1]$, $\boldsymbol{t}_a = [1,0,1]$, $\boldsymbol{t}_{a+1} = [0,1,1]$, and compute directly. As a consequence, $\langle \boldsymbol{w}_{a-1} + \boldsymbol{w}_a, \boldsymbol{w}_{a-1} + \boldsymbol{w}_a \rangle = 1 - 2 + 1 = 0$, and $\boldsymbol{w}_{a-1} + \boldsymbol{w}_a$ lies on the isotropic cone of the Lorentz form. By the formula (11), the plane tangent to this cone at $\boldsymbol{t}_a$ contains both $\boldsymbol{w}_{a-1}$ and $\boldsymbol{w}_a$; hence $\boldsymbol{w}_{a-1} + \boldsymbol{w}_a$ must be an integer multiple of $\boldsymbol{t}_a$. We thus have $\boldsymbol{w}_{a-1} + \boldsymbol{w}_a = q_a \boldsymbol{t}_a$ for some $q_a \in \mathbb{Z}$, and must prove $q_a > 0$. Now, we can surely construct a parabolic transformation $\boldsymbol{P} \in \mathrm{SO}^{\uparrow}_{2,1}\mathbb{R}$ that fixes $\boldsymbol{t}_a$ and is such that $I_{\boldsymbol{P}\boldsymbol{w}_{a-1}}$ and $I_{\boldsymbol{P}\boldsymbol{w}_a}$ have both arclength strictly less than π. By Lemma 5.1(v), $\boldsymbol{P}\boldsymbol{w}_{a-1}$ and $\boldsymbol{P}\boldsymbol{w}_a$ have both strictly positive third coordinate. Since $\boldsymbol{P}\boldsymbol{w}_{a-1} + \boldsymbol{P}\boldsymbol{w}_a = q_a \boldsymbol{t}_a$ and $\boldsymbol{t}_a$ has positive third coordinate too, q_a must be strictly positive.

For the second statement we observe that $\boldsymbol{t}_a$ is a fixed point of $\boldsymbol{A}_{a-1} = \boldsymbol{R}_{\boldsymbol{w}_{a-1}}$, as well as of $\boldsymbol{A}_a = \boldsymbol{R}_{\boldsymbol{w}_a}$. We thus compute $\boldsymbol{A}_{a-1}\boldsymbol{w}_a = \boldsymbol{A}_{a-1}(-\boldsymbol{w}_{a-1} + q_a\boldsymbol{t}_a) = \boldsymbol{w}_{a-1} + q_a\boldsymbol{t}_a$, and analogously for the other identity in (20). □

Let $x \in [0,1)$ have B-symbolic sequence $\boldsymbol{a}$. If, for some $t \geq 0$, we have $a_t = a_{t+2}$ while $a_{t+1} \in \{a_t - 1, a_t + 1\}$, then we say that x *moves parabolically* at time t.

Proof of Theorem 10.1. Let μ be the infinite measure induced by the density $\sum_a h_a$ of Theorem 7.2(ii). Since $([0,1), \mu, B)$ is ergodic and conservative, by the Halmos version of the Poincaré recurrence theorem the set P of points that move parabolically at infinitely many times has full μ-measure. As $\sum_a h_a$ is bounded from below by some positive constant, $\mu(P^c) = 0$ implies $\lambda(P^c) = 0$. In particular, the set P' of points x that move parabolically at infinitely many times, and are such that $\Phi'(x)$ exists finite, has full Lebesgue measure. We claim that $\Phi'(x) = 0$ for every $x \in P'$.

Fix such an x, and let $\boldsymbol{a}$ be its B-symbolic sequence. Then, for each $t \geq 0$, x belongs to the cylinder $B_{a_0}^{-1} \cdots B_{a_{t-1}}^{-1}[I_{a_t}]$, whose closure is the arg-image of $\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{t-1}}[I_{\boldsymbol{w}_{a_t}}]$. To be fully precise we clarify that, according to Definition 6.2, I_a is the half-open interval $[\boldsymbol{t}_a, \boldsymbol{t}_{a+1})$ (or, here, its arg-image), while $I_{\boldsymbol{w}_a}$ is, as defined in §5, the closed interval $[\boldsymbol{t}_a, \boldsymbol{t}_{a+1}]$. However, our fixed x is surely not B-terminating, so interval endpoints are of no concern here.

It is easy to show that

$$\Phi'(x) = \lim_{t\to\infty} \frac{m^{-1}(m-1)^{-(t+1)}}{\lambda(B_{a_0}^{-1} \cdots B_{a_t}^{-1}[I_{a_{t+1}}])}.$$

Suppose by contradiction that the above limit is different from 0. Then, taking the quotient of two consecutive terms and multiplying by $m-1$, we obtain

$$\lim_{t\to\infty} \frac{\lambda(B_{a_0}^{-1} \cdots B_{a_t}^{-1}[I_{a_{t+1}}])}{\lambda(B_{a_0}^{-1} \cdots B_{a_{t+1}}^{-1}[I_{a_{t+2}}])} = m - 1.$$

Up to a factor of 2π, the length of $B_{a_0}^{-1} \cdots B_{a_t}^{-1}[I_{a_{t+1}}]$ equals the arclength of $\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_t}[I_{\boldsymbol{w}_{a_{t+1}}}]$ which, by Lemma 5.1(vii), is asymptotic to the inverse of $(\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_t}\boldsymbol{w}_{a_{t+1}})_3$, the index 3 referring to the 3rd coordinate. Therefore, writing $\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{t-1}} = \boldsymbol{C}_{t-1}$ for short, we have

$$\lim_{t\to\infty} \frac{(\boldsymbol{C}_{t-1}\boldsymbol{A}_{a_t}\boldsymbol{A}_{a_{t+1}}\boldsymbol{w}_{a_{t+2}})_3}{(\boldsymbol{C}_{t-1}\boldsymbol{A}_{a_t}\boldsymbol{w}_{a_{t+1}})_3} = m - 1. \tag{21}$$

Assume now that t is a parabolic time and write $a_t = a_{t+2} = a$; without loss of generality $a_{t+1} = a - 1$. Using Lemma 10.2 and observing that $\boldsymbol{A}_a\boldsymbol{t}_a = \boldsymbol{t}_a$, we compute

$$\begin{aligned}
\frac{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{A}_{a-1}\boldsymbol{w}_a)_3}{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{w}_{a-1})_3} &= \frac{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a(\boldsymbol{w}_{a-1} + q_a\boldsymbol{t}_a))_3}{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{w}_{a-1})_3} \\
&= 1 + \frac{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a q_a\boldsymbol{t}_a)_3}{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{w}_{a-1})_3} \\
&= 1 + \frac{(\boldsymbol{C}_{t-1}q_a\boldsymbol{t}_a)_3}{(\boldsymbol{C}_{t-1}(\boldsymbol{w}_a + q_a\boldsymbol{t}_a))_3} \\
&= 1 + \frac{(\boldsymbol{C}_{t-1}q_a\boldsymbol{t}_a)_3}{(\boldsymbol{C}_{t-1}\boldsymbol{w}_a)_3 + (\boldsymbol{C}_{t-1}q_a\boldsymbol{t}_a)_3}.
\end{aligned} \tag{22}$$

Since $(\boldsymbol{C}_{t-1}\boldsymbol{w}_a)_3$ is eventually positive (actually, it goes to infinity for $t \to \infty$), the last term in the above chain of equalities is less than 2 for all sufficiently large parabolic times. If $m \geq 4$ this contradicts (21) and establishes Theorem 10.1.

If $m = 3$ we need one more parabolic iteration. Namely, we redefine a parabolic time as a time t at which the B-symbolic sequence of x has the form either $a(a-1)a(a-1)a$ or $a(a+1)a(a+1)a$. Then the chain of equalities in (22) starts with

$$\frac{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{A}_{a-1}\boldsymbol{A}_a\boldsymbol{A}_{a-1}\boldsymbol{w}_a)_3}{(\boldsymbol{C}_{t-1}\boldsymbol{A}_a\boldsymbol{A}_{a-1}\boldsymbol{A}_a\boldsymbol{w}_{a-1})_3},$$

and ends up with

$$1 + \frac{(\boldsymbol{C}_{t-1}q_a\boldsymbol{t}_a)_3}{(\boldsymbol{C}_{t-1}\boldsymbol{w}_a)_3 + (\boldsymbol{C}_{t-1}3q_a\boldsymbol{t}_a)_3},$$

which is eventually less than $4/3$, again contradicting (21). □

In §6 we set $\Gamma_B^{\pm} = \langle \mathscr{A}_0, \dots, \mathscr{A}_{m-1} \rangle < \mathrm{PSU}_{1,1}^{\pm}\mathbb{Z}[i]$; let us now define $\Gamma_B^{\pm} = C^{-1}\Gamma_B^{\pm}C = \langle A_0, \dots, A_{m-1} \rangle < \mathrm{PSL}_2^{\pm}\mathbb{Z}$ and $\boldsymbol{\Gamma}_B^{\pm} = \langle \boldsymbol{A}_0, \dots, \boldsymbol{A}_{m-1} \rangle < \mathrm{O}_{2,1}^{\uparrow}\mathbb{Z}$; see the diagram (12). Let $A \in \Gamma_B^{\pm}$; then A^2 has positive determinant and is conjugate to a matrix either of the form $\left[\begin{smallmatrix} \exp(t/2) & \\ & \exp(-t/2) \end{smallmatrix}\right]$ or of the form $\left[\begin{smallmatrix} 1 & t \\ & 1 \end{smallmatrix}\right]$ (Γ_B does not contain elliptic elements). The formulas in (2) show immediately that the spectral radius $\rho(\boldsymbol{A}^2)$ of $\boldsymbol{A}^2$ is the square of the spectral radius of A^2; taking square roots we obtain $\rho(\boldsymbol{A}) = \rho(A)^2$.

We fix a lifting —whose choice is irrelevant— of $A_0, \dots, A_{m-1}$ to $\mathrm{SL}_2^{\pm}\mathbb{Z}$, and we denote by Σ^k (respectively, $\boldsymbol{\Sigma}^k$) the set of all products of k elements of $\Sigma = \Sigma^1 = \{A_0, \dots, A_{m-1}\}$ (respectively, $\{\boldsymbol{A}_0, \dots, \boldsymbol{A}_{m-1}\}$), repetitions allowed. We recall that the *joint spectral radius* of Σ is the number

$$\rho(\Sigma) = \lim_{k\to\infty} \big(\max\{\|A\|^{1/k} : A \in \Sigma^k\}\big),$$

where $\|\ \|$ is the operator norm induced by some vector norm, whose choice is irrelevant; see [5], [21], [23] for a detailed treatment. By the Berger-Wang theorem

$$\rho(\Sigma) = \limsup_{k\to\infty} \big(\max\{\rho(A)^{1/k} : A \in \Sigma^k\}\big),$$

and the previous remarks imply that $\rho(\boldsymbol{\Sigma}) = \rho(\Sigma)^2$.

The *finiteness conjecture* [31, p. 19] states the following:

- For every finite set of matrices Π there exists $k \geq 1$ and $A \in \Pi^k$ such that $\rho(\Pi) = \rho(A)^{1/k}$.

Although the conjecture has been refuted in [9], counterexamples are difficult to construct, and are widely believed to be rare; see [26] for a detailed discussion and references to the literature. We do not know if the sets $\Sigma = \{A_0, \dots, A_{m-1}\}$ defining our billiard maps always satisfy the conjecture. However, for any specific example we examined it was easy to guess an appropriate k and $A \in \Sigma^k$, and the guess was proved correct by explicitly constructing an appropriate matrix norm; see Example 10.6.

Definition 10.3. Let $(\sigma, \rho) \in \mathcal{S}_B$, and let $\gamma : \mathbb{R} \to \mathcal{D}$ be the geodesic path of ideal endpoints $\gamma(-\infty) = \rho$ and $\gamma(+\infty) = \sigma$, parametrized by arclength, and entering the table D at $t = 0$. Then γ descends to a billiard trajectory $\bar{\gamma} : \mathbb{R} \to D = \Gamma_B^{\pm}\backslash\mathcal{D}$, and we define the *mean free path* of $\bar{\gamma}$ to be

$$\mathrm{mfp}(\bar{\gamma}) = \lim_{t\to\infty} \frac{t}{\text{number of bounces between time 0 and time } t},$$

provided that the limit exists (it surely does if $\bar\gamma$ is periodic).

Theorem 10.4. *For $\tilde\mu$-every (σ,ρ), the mean free path of $\bar\gamma$ equals 0. The supremum of the family of mean free paths of periodic trajectories equals $2\log(\rho(\Sigma))$, and this supremum is a maximum if and only if the finiteness conjecture holds for Σ.*

Proof. Let $f:\mathcal{S}_B\to\mathbb{R}_{>0}$ be defined by $f(\sigma,\rho)=\sup\{t>0:\gamma(t)\in D\}$, where γ depends on (σ,ρ) as in Definition 10.3. Then the integral of f with respect to $\tilde\mu$ is finite, since it equals one half of the volume of the unit tangent bundle of $\Gamma_B\backslash\mathcal{D}$. Since the measure-preserving system $(\mathcal{S}_B,\tilde\mu,\widetilde{B})$ is conservative, a basic result of infinite ergodic theory [25, §4] yields that for $\tilde\mu$-every (σ,ρ) we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f\big(\widetilde{B}^k(\sigma,\rho)\big)=0.$$

As the limit above is precisely the free mean path of $\bar\gamma$, our first statement follows.

Let $M=\sup\{\mathrm{mfp}(\bar\gamma):\bar\gamma \text{ is a periodic billiard trajectory}\}$. Given $k\ge 3$, let A have maximum spectral radius in Σ^k. Surely A^2 cannot be parabolic and, by the unique factorization of A as a product of elements in Σ, we see that there exists $B=A_{b_0}\cdots A_{b_{h-1}}\in\Sigma^h$ such that $2\le h\le k$, $b_0\ne b_{h-1}$, and A is conjugate to B. Define $\gamma:\mathbb{R}\to\mathcal{D}$ by $\gamma(t)=CB*\exp(ti)$, where C is the Cayley matrix. Then γ descends to a h-bounces periodic billiard trajectory $\bar\gamma$ on D, which we claim to have length $2\log(\rho(B))$. Indeed, if h is even then B is hyperbolic; thus, by the proof of [12, Proposition 1], $\bar\gamma$ has length $2\arccosh(|\operatorname{tr}B|/2)$, which is indeed $2\log(\rho(B))$. If h is odd, then we replace B with B^2 and obtain that $\bar\gamma$ has length $\log(\rho(B^2))$, which again equals $2\log(\rho(B))$. As $\bar\gamma$ involves h bounces, we have $\mathrm{mfp}(\bar\gamma)=2\log(\rho(B)^{1/h})$; we conclude that $2\log(\rho(A)^{1/k})\le 2\log(\rho(B)^{1/h})=\mathrm{mfp}(\bar\gamma)$, and thus $2\log(\rho(\Sigma))\le M$.

Conversely, any periodic trajectory $\bar\gamma$ involving k bounces can be lifted (nonuniquely) to a unit speed geodesic path $\gamma:\mathbb{R}\to\mathcal{D}$. The B-symbolic sequence $\boldsymbol{a}$ of $\gamma(+\infty)=\sigma\in S^1$ is periodic of period k and the argument above, applied to $A=A_{a_0}\cdots A_{a_{k-1}}$, shows that $\bar\gamma$ has mean free path $2\log(\rho(A)^{1/k})$; therefore $M\le 2\log(\rho(\Sigma))$. □

Theorem 10.5. *The function Φ is Hölder continuous of exponent*

$$\alpha=\frac{\log(m-1)}{2\log(\rho(\Sigma))}.$$

If the finiteness conjecture holds for Σ, then α is the best Hölder exponent (i.e., Φ is not Hölder continuous of exponent β, for any $\beta>\alpha$).

Proof. Let $\|\boldsymbol{x}\|=\max\{|x_1|,|x_2|,|x_3|\}$ denote the ∞-norm in $\mathbb{R}^3$; note that $\|\boldsymbol{x}\|=|x_3|$ on $\mathcal{S}\cap\mathbb{Z}^3$, exception being made for the four points $(\pm1,0,0)$, $(0,\pm1,0)$ only. As noted in the proof of Theorem 10.1, the closure of the cylinder $B_{a_0}^{-1}\cdots B_{a_{k-1}}^{-1}[I_{a_k}]$ is the arg-image of $\boldsymbol{A}_{a_0}\cdots\boldsymbol{A}_{a_{k-1}}[I_{\boldsymbol{w}_{a_k}}]$. Taking into account Lemma 5.1(iii) and (vii), the length of the former is asymptotic, as k increases, to $\pi^{-1}\|\boldsymbol{A}_{a_0}\cdots\boldsymbol{A}_{a_{k-1}}\boldsymbol{w}_{a_t}\|^{-1}$. Once fixed a constant $C>\pi\max\{\|\boldsymbol{w}_{a_0}\|,\dots,\|\boldsymbol{w}_{a_{m-1}}\|\}>1$, this implies that there exists a level k_0 such that, for every $k\ge k_0$ and every cylinder $B_{a_0}^{-1}\cdots B_{a_{k-1}}^{-1}[I_{a_k}]$ of level k, we have

$$C^{-1}\|\boldsymbol{A}_{a_0}\cdots\boldsymbol{A}_{a_{k-1}}\|^{-1}<\lambda(B_{a_0}^{-1}\cdots B_{a_{k-1}}^{-1}[I_{a_k}])<1/2,$$

where the matrix norm is the one induced by the vector norm.

Fix now $\varepsilon > 0$. Then there exists $k_1 \geq k_0$ such that, for every $k \geq k_1$ and every matrix $\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{k-1}} \in \boldsymbol{\Sigma}^k$, we have $\rho(\boldsymbol{\Sigma}) + \varepsilon > \|\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{k-1}}\|^{1/k}$. Let $0 \leq x < x' < 1$ be such that

$$x' - x \leq l_1 = \min\{l : l \text{ is the length of a cylinder of level } k_1\}.$$

Let $k \geq k_1$ be minimum such that the interval $[x, x']$ contains a cylinder $B_{a_0}^{-1} \cdots B_{a_{k-1}}^{-1}[I_{a_k}]$ of level k; then we have

$$x' - x > C^{-1}\|\boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{k-1}}\|^{-1} > C^{-1}(\rho(\boldsymbol{\Sigma}) + \varepsilon)^{-k},$$

which implies

$$k > -\frac{\log(C)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)} - \frac{\log(x' - x)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)}. \tag{23}$$

On the other hand, the interval $[x, x']$ may contain at most $1+(m-2)+(m-2) = 2m-3$ endpoints of cylinders of level k; therefore

$$\Phi x' - \Phi x < (2m-2)m^{-1}(m-1)^{-k},$$

which implies

$$k < \frac{\log(2m^{-1}(m-1))}{\log(m-1)} - \frac{\log(\Phi x' - \Phi x)}{\log(m-1)}. \tag{24}$$

Eliminating k from (23) and (24) and rearranging terms, we obtain

$$\frac{\log(\Phi x' - \Phi x)}{\log(m-1)} < \frac{\log(C)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)} + \frac{\log(2m^{-1}(m-1))}{\log(m-1)} + \frac{\log(x' - x)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)},$$

whence

$$\begin{aligned}\log(\Phi x' - \Phi x) &< \frac{\log(m-1)\log(C)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)} + \log(2m^{-1}(m-1)) + \frac{\log(m-1)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)}\log(x' - x) \\ &< \log(E) + \frac{\log(m-1)}{\log(\rho(\boldsymbol{\Sigma}) + \varepsilon)}\log(x' - x),\end{aligned}$$

where

$$E = \exp\left(\frac{\log(m-1)\log(C)}{\log(\rho(\boldsymbol{\Sigma}))} + \log(2m^{-1}(m-1))\right).$$

We thus obtained

$$\Phi x' - \Phi x < E(x' - x)^{\log(m-1)/\log(\rho(\boldsymbol{\Sigma})+\varepsilon)}.$$

Since E does not depend on ε, we let ε tend to 0 and obtain the Hölder condition $\Phi x' - \Phi x \leq E(x' - x)^{\alpha}$, valid for $x' - x \leq l_1$ (remember that $\rho(\boldsymbol{\Sigma}) = \rho(\Sigma)^2$). Replacing E with $\max\{E, l_1^{-\alpha}\}$, the condition holds for every pair $x < x'$.

Assume now that the finiteness conjecture holds for $\boldsymbol{\Sigma}$, and let $\boldsymbol{A} = \boldsymbol{A}_{a_0} \cdots \boldsymbol{A}_{a_{k-1}} \in \boldsymbol{\Sigma}^k$ be a maximizing matrix (i.e., $\rho(\boldsymbol{\Sigma}) = \rho(\boldsymbol{A})^{1/k}$). We must have $a_0 \neq a_{k-1}$, since otherwise $\boldsymbol{A}$ would be conjugate to a matrix $\boldsymbol{B}$ in $\boldsymbol{\Sigma}^{k-2}$ and we would have $\rho(\boldsymbol{B})^{1/(k-2)} > \rho(\boldsymbol{A})^{1/k} = \rho(\boldsymbol{\Sigma})$, which is impossible. The eigenvalues of $\boldsymbol{A}$ are $(-1)^k$, $\rho(\boldsymbol{A})$, and $\rho(\boldsymbol{A})^{-1}$; let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ be the corresponding eigenvectors. The vector $\boldsymbol{w}_{a_0}$ cannot lie in the subspace spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$, because $\|\boldsymbol{A}^n \boldsymbol{w}_{a_0}\| \to \infty$ for $n \to \infty$. This easily implies that the length of the cylinder $(B_{a_0}^{-1} \cdots B_{a_{k-1}}^{-1})^n[I_{a_0}]$, of level kn and endpoints $x_n < x_n'$, is asymptotic to $C\rho(\boldsymbol{A})^{-n}$ as $n \to \infty$, for some constant C. But then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{\Phi x_n' - \Phi x_n}{(x_n' - x_n)^{\alpha+\varepsilon}} = \lim_{n\to\infty} \frac{m^{-1}(m-1)^{-kn}}{C^{\alpha+\varepsilon}\rho(\boldsymbol{A})^{-(\alpha+\varepsilon)n}} = \infty,$$

because $\rho(\boldsymbol{\Sigma})^{\alpha+\varepsilon} > m-1$ implies $\rho(\boldsymbol{A})^{\alpha+\varepsilon} > (m-1)^k$, and thus $(m-1)^{-k}/\rho(\boldsymbol{A})^{-(\alpha+\varepsilon)} > 1$. □

Example 10.6. Consider the square billiard table of Example 9.1. By the symmetries of the table, the graph of the induced Minkowski function Φ in Figure 7 (right) results from the gluing of four identical pieces, the fourth piece corresponding to the interval $[-i,1]$ in S^1. Since the foldings $\boldsymbol{F}, \boldsymbol{JF}, \boldsymbol{J}$ involved in the construction of the Romik map in §3 are isometries, it is not difficult to realize that this fourth piece is conjugate via stereographic projection from $[0,1,1]$ to the Minkowski function Q_E introduced in [7] for the Romik map. As the above stereographic projection is a Lipschitz bijection with Lipschitz inverse between $[-i,1]$ and $[0,1]$, the Hölder exponents of Φ and of Q_E must agree.

The set Σ contains the four matrices

$$A_0 = \begin{pmatrix} -1 & 2 \\ & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 2 \\ & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & \\ -2 & -1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -1 & \\ -2 & 1 \end{pmatrix}.$$

By looking at our square billiard table, we obviously conjecture that the maximum periodic mean free path should be realized by bouncing between two opposite walls; in other words, that the finiteness conjecture should hold for Σ, with witnessing matrix $A_3A_1 \in \Sigma^2$ (or its conjugate A_2A_0).

Denote by $\|\ \|_2$ the spectral norm on 2×2 real matrices induced by the euclidean norm on $\mathbb{R}^2$. Then, as it is well known, $\|A\|_2 = \rho(A^\top A)^{1/2}$, and one checks immediately that $\|A_a\|_2 = \sqrt{3+\sqrt{8}}$ for every $a \in \{0,1,2,3\}$. Since $\rho(A_3A_1)^{1/2} \le \rho(\Sigma) \le \max\{\|A\|_2 : A \in \Sigma^1\}$, and $\rho(A_3A_1)^{1/2}$ equals $\sqrt{3+\sqrt{8}} = 1+\sqrt{2}$ as well, our conjecture is confirmed. Theorem 10.4 now yields that Φ, and thus Q_E, has Hölder best exponent $\log(3)/(2\log(1+\sqrt{2}))$, in agreement with [7, Theorem 2].

REFERENCES

[1] J. Aaronson, *An Introduction to Infinite Ergodic Theory*, Vol. 50, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1997.

[2] J. Aaronson and M. Denker, The Poincaré series of $\mathbb{C}\setminus\mathbb{Z}$, *Ergodic Theory Dynam. Systems*, **19** (1999), 1–20.

[3] R. C. Alperin, The modular tree of Pythagoras, *Amer. Math. Monthly*, **112** (2005), 807–816.

[4] F. J. M. Barning, On Pythagorean and quasi-Pythagorean triangles and a generation process with the help of unimodular matrices, *Math. Centrum Amsterdam Afd. Zuivere Wisk.*, **1963** (1963), 37 pp.

[5] M. A. Berger and Y. Wang, Bounded semigroups of matrices, *Linear Algebra Appl.*, **166** (1992), 21–27.

[6] B. Berggren, Pytagoreiska trianglar, *Tidskrift för elementär matematik, fysik och kemi*, **17** (1934), 129–139.

[7] F. P. Boca and C. Linden, On Minkowski type question mark functions associated with even or odd continued fractions, *Monatsh. Math.*, **187** (2018), 35–57.

[8] A. I. Borevich and I. R. Shafarevich, *Number Theory*, Vol. 20, Pure and Applied Mathematics, Academic Press, New York-London, 1966.

[9] T. Bousch and J. Mairesse, Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture, *J. Amer. Math. Soc.*, **15** (2002), 77–111.

[10] J. W. Cannon, W. J. Floyd, R. Kenyon and W. R. Parry, Hyperbolic geometry, in *Flavors of Geometry*, Vol. 31, Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, Cambridge, 1997.

[11] D. Cass and P. J. Arpaia, Matrix generation of Pythagorean n-tuples, *Proc. Amer. Math. Soc.*, **109** (1990), 1–7.

[12] S. Castle, N. Peyerimhoff and K. F. Siburg, Billiards in ideal hyperbolic polygons, *Discrete Contin. Dyn. Syst.*, **29** (2011), 893–908.

[13] B. Cha and D. H. Kim, Lagrange spectrum of Romik's dynamical system, preprint, arXiv:1903.02882.

[14] B. Cha and D. H. Kim, Number theoretical properties of Romik's dynamical system, *Bull. Korean Math. Soc.*, **57** (2020), 251–274.
[15] B. Cha, E. Nguyen, and B. Tauber, Quadratic forms and their Berggren trees, *J. Number Theory*, **185** (2018), 218–256.
[16] N. Chernov and R. Markarian, *Introduction to the Ergodic Theory of Chaotic Billiards*, 2^{nd} edition, IMPA Mathematical Publications, Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2003.
[17] K. T. Conrad, Pythagorean descent, *Semantic Scholar*, 2007.
[18] I. P. Cornfeld, S. V. Fomin and Y. G. Sinaĭ, *Ergodic Theory*, Vol. 245, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, New York, 1982.
[19] A. Denjoy, Sur une fonction réelle de Minkowski, *J. Math. Pures Appl.*, **17** (1938), 105–155.
[20] E. J. Eckert, The group of primitive Pythagorean triangles, *Math. Mag.*, **57** (1984), 22–27.
[21] L. Elsner, The generalized spectral-radius theorem: An analytic-geometric proof, *Linear Algebra Appl.*, **220** (1995), 151–159.
[22] D. Fried, Symbolic dynamics for triangle groups, *Invent. Math.*, **125** (1996), 487–521.
[23] N. Guglielmi and M. Zennaro, Stability of linear problems: Joint spectral radius of sets of matrices, in *Current Challenges in Stability Issues for Numerical Differential Equations*, Vol. 2082, Lecture Notes in Math., Springer, Cham, 2014.
[24] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5^{th} edition, The Clarendon Press, Oxford University Press, New York, 1979.
[25] S. Isola, From infinite ergodic theory to number theory (and possibly back), *Chaos Solitons Fractals*, **44** (2011), 467–479.
[26] O. Jenkinson and M. Pollicott, Joint spectral radius, Sturmian measures and the finiteness conjecture, *Ergodic Theory Dynam. Systems*, **38** (2018), 3062–3100.
[27] T. Jordan and T. Sahlsten, Fourier transforms of Gibbs measures for the Gauss map, *Math. Ann.*, **364** (2016), 983–1023.
[28] S. Katok, Fuchsian groups, geodesic flows on surfaces of constant negative curvature and symbolic coding of geodesics, in *Homogeneous Flows, Moduli Spaces and Arithmetic*, Vol. 10, Clay Math. Proc., Amer. Math. Soc., 2010.
[29] M. Kesseböhmer and B. O. Stratmann, A multifractal analysis for Stern-Brocot intervals, continued fractions and Diophantine growth rates, *J. Reine Angew. Math.*, **605** (2007), 133–163.
[30] R. Kołodziej, An infinite smooth invariant measure for some transformation of a circle, *Bull. Acad. Polon. Sci. Sér. Sci. Math.*, **29** (1981), 549–551.
[31] J. C. Lagarias and Y. Wang, The finiteness conjecture for the generalized spectral radius of a set of matrices, *Linear Algebra Appl.*, **214** (1995), 17–42.
[32] C. Maclachlan and A. W. Reid, *The Arithmetic of Hyperbolic 3-manifolds*, Vol. 219, Graduate Texts in Mathematics, Springer-Verlag, New York, 2003.
[33] A. Miller, Trees of integral triangles with given rectangular defect, *Discrete Math.*, **313** (2013), 50–66.
[34] M. Misiurewicz, The result of Rafał Kołodziej, in *Ergodic Theory (Sem., Les Plans-sur-Bex, 1980)*, Vol. 29, Monograph. Enseign. Math., Univ. Genève, Geneva, 1981.
[35] J. Morita, A transformation group of the Pythagorean numbers, *Tsukuba J. Math.*, **10** (1986), 151–153.
[36] U. Moschella, The de Sitter and anti-de Sitter sightseeing tour, in *Einstein, 1905–2005*, Vol. 47, Prog. Math. Phys., Birkhäuser, Basel, 2006.
[37] K. Nomizu, The Lorentz-Poincaré metric on the upper half-space and its extension, *Hokkaido Math. J.*, **11** (1982), 253–261.
[38] G. Panti, A general Lagrange theorem, *Amer. Math. Monthly*, **116** (2009), 70–74.
[39] G. Panti, Slow continued fractions, transducers, and the Serret theorem, *J. Number Theory*, **185** (2018), 121–143.
[40] A. M. Rockett and P. Szüsz, *Continued Fractions*, World Scientific Publishing Co., Inc., River Edge, NJ, 1992.
[41] V. A. Rohlin, Exact endomorphisms of a Lebesgue space, *Izv. Akad. Nauk SSSR Ser. Mat.*, **25** (1961), 499–530.
[42] D. Romik, The dynamics of Pythagorean triples, *Trans. Amer. Math. Soc.*, **360** (2008), 6045–6064.
[43] R. Salem, On some singular monotonic functions which are strictly increasing, *Trans. Amer. Math. Soc.*, **53** (1943), 427–439.

[44] D. Singerman, Finitely maximal Fuchsian groups, *J. London Math. Soc.*, **6** (1972), 29–38.
[45] J. Smillie and C. Ulcigrai, Geodesic flow on the Teichmüller disk of the regular octagon, cutting sequences and octagon continued fractions maps, in *Dynamical Numbers–Interplay Between Dynamical Systems and Number Theory*, Vol. 532, Contempt. Math, Amer. Math. Soc., Providence, RI, 2010.
[46] K. Takeuchi, Arithmetic triangle groups, *J. Math. Soc. Japan*, **29** (1977), 91–106.
[47] M. Thaler, Transformations on [0, 1] with infinite invariant measures, *Israel J. Math.*, **46** (1983), 67–96.

Received July 2019; revised January 2020.

E-mail address: giovanni.panti@uniud.it