

THE DYNAMICS OF SUPER-APOLLONIAN CONTINUED FRACTIONS

SNEHA CHAUBEY, ELENA FUCHS, ROBERT HINES, AND KATHERINE E. STANGE

ABSTRACT. We examine a pair of dynamical systems on the plane induced by a pair of spanning trees in the Cayley graph of the Super-Apollonian group of Graham, Lagarias, Mallows, Wilks and Yan. The dynamical systems compute Gaussian rational approximations to complex numbers and are “reflective” versions of the complex continued fractions of A. L. Schmidt. They also describe a reduction algorithm for Lorentz quadruples, in analogy to work of Romik on Pythagorean triples. For these dynamical systems, we produce an invertible extension and an invariant measure, which we conjecture is ergodic. We consider some statistics of the related continued fraction expansions, and we also examine the restriction of these systems to the real line, which gives a reflective version of the usual continued fraction algorithm. Finally, we briefly consider an alternate setup corresponding to a tree of Lorentz quadruples ordered by arithmetic complexity.

1. INTRODUCTION

This paper unites three distinct lines of research that have appeared over the past 40 years. In what was chronologically the first, Asmus Schmidt developed a theory of Gaussian complex continued fractions based on *Farey circles and triangles* analogous to the Farey subdivision of the real line [Sc1, Sc2, Sc3, Sc4]. In the second, Graham, Lagarias, Mallows, Wilks and Yan laid much of the algebraic groundwork for the study of Apollonian circle packings and super-packings [GLMWY1, GLMWY2, GLMWY3]. In the third, Romik studied a dynamical system on the tree of Pythagorean triples which is conjugate to a Euclidean algorithm related to the Gauss map for real continued fractions [R].

In this work, we define a dynamical system on the complex plane which can be viewed in at least three distinct ways: as a “reflective” version of Gaussian complex continued fractions; as a system of reduction on the Descartes quadruples of all Apollonian circle packings; or as a dynamical system on the tree of Lorentz quadruples, i.e. solutions to $x^2 + y^2 + z^2 = t^2$. Restricted to the real line, it produces a reflective continued fraction algorithm and Euclidean algorithm.

Our work fits into a long line of work on the number theoretical aspects of Apollonian circle packings. One of the first papers on this subject was [GLMWY1], paving the way for many subsequent works on the arithmetic of such packings. Indeed, many of the results in these subsequent works started as conjectures in [GLMWY1] (see [Fu] and [Kon] for a summary of these results). By and large, the arithmetic questions inspired by the work of Graham et. al. have been of the following nature. First, one observes that certain Apollonian circle packings are *integral*, meaning their curvatures (inverse radii) are all integral. Consider the set of integer curvatures (counted with or without multiplicity) appearing in such a fixed primitive integral Apollonian packing. What can one say about this set of integers: can one describe the integers using local conditions alone, and

Date: January 13, 2021.

2010 Mathematics Subject Classification. Primary: 11A55, 11J70, 11E20, 37A45, 37F30, 52C26.

Key words and phrases. Bianchi group, dynamical system, ergodic theory, geodesic flow, invertible extension, invariant measure, Pythagorean triple, Lorentz quadruple, Descartes quadruple, orthogonal group, continued fractions, projective linear group, Möbius transformation, hyperbolic isometry, quadratic form, Apollonian circle packing.

The work of the fourth author was sponsored by the National Security Agency under Grant Number H98230-14-1-0106 and NSF DMS-1643552. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation herein. The work of the second author was supported by NSF DMS-1501970 and a Sloan Research Fellowship.

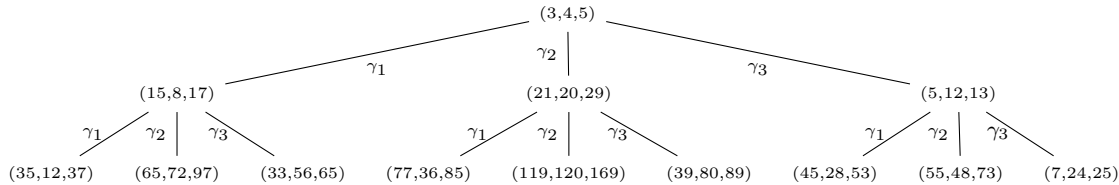


FIGURE 1. Tree of Pythagorean Triples, shown to the third level

if so, what should these local conditions be; are there infinitely many primes in any one packing, and if so, how many primes of bounded size are there?

While the question of rational approximation of complex numbers is quite different in flavor than the number theoretic questions just described, it turns out both questions hang on studying the same symmetry group. In the study of integer curvatures this is the Apollonian group, which lives naturally inside the Super-Apollonian group of [GLMWY2, GLMWY3], which, in turn, describes the collection of all integral Apollonian circle packings simultaneously. However, the Super-Apollonian group can be realized as an index 48 subgroup of the extended Bianchi group $\mathrm{PGL}_2(\mathbb{Z}[i]) \rtimes \langle \mathfrak{c} \rangle$, where $\mathfrak{c}$ is complex conjugation. This Bianchi group, analogous to $\mathrm{PGL}_2(\mathbb{Z})$, is the natural setting in which to study Gaussian complex continued fraction expansions. As an indication of this phenomenon, the Farey circles and triangles of Schmidt (see Figure 17), when iterated, actually produce an Apollonian super-packing as illustrated in [GLMWY3] or Figure 7. See the work of A. Hurwitz for a different approach to complex continued fractions [H1].

The Super-Apollonian continued fraction expansion we describe in this paper was, however, initially inspired by a process described by Romik in [R]. In that paper, Romik used the group generated by

$$\gamma_1 = \begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & 2 \\ -2 & 2 & 3 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & 2 \\ 2 & -2 & 3 \end{pmatrix}$$

to create a dynamical system on the positive quadrant of the unit circle. This is a subgroup of the orthogonal group $\mathrm{O}_F(\mathbb{Z})$ where $F(x, y, z) = x^2 + y^2 - z^2$, and any point (x, y) on the unit circle corresponds to the solution $(x, y, 1)$ of the equation $F(x, y, z) = 0$, or the Pythagorean equation. Indeed, the dynamical system Romik constructed acts naturally to walk to the root of a tree of primitive integer Pythagorean triples as depicted in Figure 1.

The fact that Pythagorean triples form a tree has been observed many times; see [Be, Ba, H, Ka, P]. Given any primitive Pythagorean triple (a, b, c) , which corresponds to the rational point $(a/c, b/c)$ on the unit circle, Romik's dynamical system allows one to quickly determine the finite word in $\gamma_1, \gamma_2, \gamma_3$ which describes the path in the tree from $(3, 4, 5)$ to (a, b, c) . It also associates an *infinite* word or expansion in $\gamma_1, \gamma_2, \gamma_3$ to any irrational point on the first quadrant of the unit circle, and this gives a type of continued fraction expansion. We refer the reader to [R] for further details on this beautiful work.

The present paper can be thought of as a version of Romik's work in four dimensions. As we discuss below, Apollonian packings are naturally connected to primitive *Lorentz quadruples*, or coprime quadruples of integers satisfying $x^2 + y^2 + z^2 = t^2$. In this higher dimensional setting, a number of new difficulties arise: for example, the analog of the ternary Pythagorean tree above is no longer a tree, and there is not a unique path from any given vertex to the root (see Figure 4 for a picture of the graph).

We will now briefly describe the idea behind Super-Apollonian continued fractions. An Apollonian packing is obtained by starting with four pairwise tangent circles, and repeatedly inscribing

into every region bounded by three existing pairwise tangent circles the unique circle which is tangent to all three. The number theoretical interest in these packings comes from the fact that if the starting four circles all have integer curvature, then all of the circles in the packing have integer curvature. This process is depicted in Figure 2.

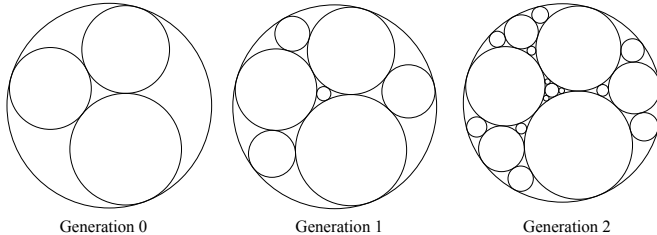


FIGURE 2. Constructing an Apollonian packing

These packings have a symmetry which can be described both algebraically and geometrically. The algebraic interpretation stems from Descartes' theorem from 1643, that if four pairwise circles have curvatures a, b, c, d , then $Q(a, b, c, d) = 2(a^2 + b^2 + c^2 + d^2) - (a + b + c + d)^2 = 0$; the symmetries of the packing are symmetries of this quadratic form. On the other hand, geometrically, one can embed the Apollonian circle packing in the complex plane and describe the symmetries as Möbius transformations.

While the curvatures of circles in any one primitive integral Apollonian packing give only a sparse subset of all primitive integer solutions to $Q(a, b, c, d) = 0$, one can use the geometric interpretation of the algebraic symmetry to choose a natural extension to the Super-Apollonian group of [GLMWY2]. The Super-Apollonian group acts on Descartes quadruples. The orbit of one Descartes quadruple under this larger group will now include all primitive integral Apollonian circle packings, nested one inside another densely in the complex plane; see Figure 7.

To approximate a complex number $z \in \mathbb{C}$, we approach z with a sequence of Descartes quadruples of circles, in the sense that the tangency points of the quadruples (all of which are rational) approach z . The sequence of quadruples is generated by repeated applications of generators of the Super-Apollonian group, so that the continued fraction expansion is a word in the generators. Figure 3 depicts this process when approximating the number $\pi + ei$. In contrast to the real or Pythagorean case, since the Super-Apollonian group is not free, the Cayley graph is not a tree, and there are multiple approximation sequences for a given $z \in \mathbb{C}$; a natural choice is given by the normal form of Graham, Lagarias, Mallows, Wilks and Yan [GLMWY2].

The Apollonian super-packing of Figure 7 is the same fractal subdivision of $\mathbb{C}$ used in the Gaussian continued fraction algorithm of Schmidt. However, Schmidt's setup involves a different choice of group and generators to navigate the subdivision. We develop a dynamical system, based on the Super-Apollonian group, which acts on the plane to generate the continued fraction algorithm; define an invertible extension to an action on hyperbolic geodesics; and find an invariant measure. This process closely follows the development of Schmidt's system by Schmidt and Nakada, as detailed in an appendix.

Having described these dynamical systems, we go on to analyse some statistics of the continued fraction expansion. We also include an analysis of the restriction of the Super-Apollonian dynamics to the real line, where we recover "reflective" real continued fractions. Finally, in part to emphasize the variability in possible systems, we briefly consider another dynamical system with the goal of organizing Lorentz quadruples by arithmetic complexity.

In future work, we plan to provide a proof of ergodicity for our continued fraction algorithm, and to develop parallel theories for other imaginary quadratic fields such as $\mathbb{Q}(\sqrt{-2})$.

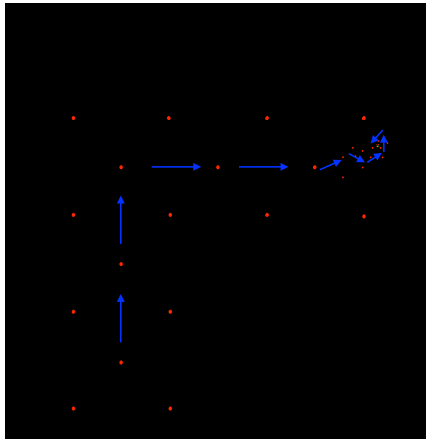


FIGURE 3. Estimating $\pi + ei$ using tangency points in an Apollonian super-packing

The paper proceeds as follows. Section 2 contains background information on classical continued fractions, Lorentz and Descartes quadruples, the Apollonian and Super-Apollonian groups and their geometric interpretation as Möbius transformations of the complex plane. This section includes a discussion of Graham, Lagarias, Mallows, Wilks and Yan’s swap normal form and the associated spanning tree of the Cayley graph of the Super-Apollonian group. Section 3 defines a dynamical system associated to this choice of spanning tree. First a pair of dynamical systems on the complex plane are described, which compute a Gaussian continued fraction expansion of a complex number. We describe an invertible extension and invariant measure. In Section 4, the closely related dynamical systems of Lorentz and Descartes quadruples are described. These systems travel to the root along the the swap normal and invert normal form spanning trees. We relate these to the Reduction Algorithm for Descartes quadruples of Graham, Lagarias, Mallows, Wilks and Yan. In Section 5, we give some consequences of the conjectural ergodicity of the systems, and compare to experimental data. We hope to demonstrate ergodicity in a follow-up paper. In Section 6, we restrict this system to the real line, and recover a “reflective” variation on the classical continued fraction algorithm. In Section 7, we briefly consider an alternate dynamical system which respects the arithmetic complexity of Lorentz quadruples. Finally, in an appendix, we review the Gaussian continued fraction algorithm of Schmidt and Nakada, and its explicit connection to our work.

A note on figures. Figures and experimental data were created with Sage Mathematics Software [Sa] and Mathematica [W].

Acknowledgements. We would like to thank Dan Romik for writing the paper that inspired this work, and for several helpful conversations. We also thank Jayadev Athreya and the anonymous referee for helpful comments and suggestions.

2. QUADRUPLES AND THE SUPER-APOLLONIAN GROUP

2.1. Simple continued fractions. We begin with a brief overview of simple continued fractions on the real line, described from the perspective and in the language we plan to use for complex continued fractions in the remainder of this paper. The purpose is to provide an explicit analogy for much of what follows.

Over the integers, the Euclidean algorithm is the iteration of the division algorithm, which, for $a, b \in \mathbb{Z}$, returns $q, r \in \mathbb{Z}$ so that $a = bq + r$, $0 \leq r < |b|$. Replacing the pair (a, b) with (b, r) , we repeat. We can illustrate this as $(a, b) \xrightarrow{q} (b, r)$, e.g.

$$(355, 113) \xrightarrow{3} (113, 16) \xrightarrow{7} (16, 1) \xrightarrow{16} (1, 0).$$

The transformation taking (a, b) to (b, r) can be written in terms of matrices as

$$\begin{pmatrix} b \\ r \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

If a, b are coprime, the Euclidean algorithm terminates at $(1, 0)$. We can work backwards from $(1, 0)$ to recover (a, b) , e.g.

$$\begin{pmatrix} 355 \\ 113 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 7 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 16 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

The Euclidean algorithm gives rise to a map on rationals

$$\frac{b}{a} \xrightarrow{q} \frac{r}{b} = \frac{a}{b} - q,$$

(say with a, b coprime), and the algorithm produces expressions such as

$$\frac{355}{113} = 3 + \frac{1}{7 + \frac{1}{16}}.$$

We can extend this map to the interval $(0, 1)$, i.e. the Gauss map $T(x) = \{1/x\}$ mapping $x \in (0, 1)$ to the fractional part of $1/x$, which is defined piecewise by Möbius transformations

$$T(x) = \frac{-qx + 1}{x}, \quad x \in \left[\frac{1}{q+1}, \frac{1}{q} \right).$$

The effect of this is an infinite continued fraction

$$x = [a_0; a_1, \dots, a_n, \dots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}}, \quad a_0 = \lfloor x \rfloor, \quad a_n = \left\lfloor \frac{1}{T^{n-1}(x - a_0)} \right\rfloor,$$

and a sequence of convergents p_n/q_n to x :

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}, \quad \lim_{n \rightarrow \infty} \frac{p_n}{q_n} = x.$$

For irrational $x \in (0, 1)$, T is the left shift map on $\mathbb{N}^{\mathbb{N}}$, $T([0; a_1, a_2, \dots]) = [0; a_2, a_3, \dots]$.

The Gauss map is ∞ -to-1, but we can find an invertible extension $\tilde{T}$ defined on pairs $(y, x) \in (-\infty, -1) \times (0, 1)$,

$$\tilde{T}(y, x) = (1/y - \lfloor x \rfloor, T(x)).$$

The pairs (y, x) can be identified with oriented geodesics in the hyperbolic plane, realized as the upper half plane above the real line. Specifically, the points y and x are the past and future points on the ideal boundary in the upper half-plane model. The extension $\tilde{T}$ acts piecewise by isometries, $\text{Isom}(\mathbb{H}^2) \cong \text{PGL}_2(\mathbb{R})$. (The space of geodesics $(-\infty, -1) \times (0, 1)$ corresponds to Gauss' reduced indefinite binary quadratic forms.) The space of geodesics carries an isometry invariant measure $\frac{dx dy}{(x-y)^2}$ coming from hyperbolic area or Haar measure on $\text{SL}_2(\mathbb{R})$, which restricts to a $\tilde{T}$ -invariant measure on $(-\infty, -1) \times (0, 1)$ since $\tilde{T}$ is defined piecewise by isometries. Pushing this measure forward to the second coordinate gives the T -invariant Gauss measure $d\mu(x) = \frac{dx}{1+x}$ on $(0, 1)$, which is ergodic:

$$d\mu(x) = dx \int_{-\infty}^{-1} \frac{dy}{(x-y)^2}.$$

For more on continued fractions, see [Kh] (basic properties), [wS] (Diophantine approximation), [Bi] (ergodicity), [EW] (connection to the geodesic flow and hyperbolic dynamics). For an excellent exposition of this view of the Gauss measure, see [Ke].

2.2. Lorentz and Descartes quadruples. We consider the Lorentz $(1, 3)$ -form

$$Q_L(\mathbf{x}) = x_0^2 - x_1^2 - x_2^2 - x_3^2,$$

and the solutions to $Q_L(\mathbf{x}) = 0$ are called *Lorentz quadruples*. Consider the following matrices:

$$\begin{aligned} L_1 &:= \begin{pmatrix} 2 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{pmatrix}, L_2 := \begin{pmatrix} 2 & -1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & -1 \\ -1 & 1 & -1 & 0 \end{pmatrix}, \\ L_3 &:= \begin{pmatrix} 2 & 1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 0 \end{pmatrix}, L_4 := \begin{pmatrix} 2 & 1 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \\ L_1^\perp &:= \begin{pmatrix} 2 & 1 & 1 & 1 \\ -1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{pmatrix}, L_2^\perp := \begin{pmatrix} 2 & 1 & -1 & -1 \\ -1 & 0 & 1 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & 1 & -1 & 0 \end{pmatrix}, \\ L_3^\perp &:= \begin{pmatrix} 2 & -1 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 0 \end{pmatrix}, L_4^\perp := \begin{pmatrix} 2 & -1 & -1 & 1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}. \end{aligned}$$

We create these matrices by conjugating L_1 by all eight possible diagonal matrices with diagonals $(1, \pm 1, \pm 1, \pm 1)$. Each of the eight resulting matrices is an involution, and preserves Q_L in the sense that

$$L_i^T G_L L_i = G_L, \quad (L_i^\perp)^T G_L L_i^\perp = G_L$$

for the Gram matrix

$$G_L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

of Q_L .

We create a Cayley graph, $\mathcal{C}_L$, whose vertices are the elements of the group generated by the $L_i, L_i^\perp$ and where two vertices M_1 and M_2 are joined by an edge exactly when $M_1 M_2^{-1}$ is one of the $L_i, L_i^\perp$. If one takes a quotient of this Cayley graph by considering the action of the $L_i, L_i^\perp$ on Lorentz quadruples, one obtains Figure 4.

The form Q_L is equivalent to the Descartes form

$$Q_D(\mathbf{y}) := (y_0 + y_1 + y_2 + y_3)^2 - 2(y_0^2 + y_1^2 + y_2^2 + y_3^2)$$

by the relationship $2Q_D(J\mathbf{x}) = Q_L(\mathbf{x})$ where

$$(2.1) \quad J = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

$$\begin{aligned}
S_3 &:= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 2 & -1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, S_4 := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 2 & 2 & -1 \end{pmatrix}, \\
S_1^\perp &:= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}, S_2^\perp := \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix}, \\
S_3^\perp &:= \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix}, S_4^\perp := \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 \end{pmatrix}.
\end{aligned}$$

The group $\mathcal{A}^S$ preserves Q_D in the sense that $M^t G_D M = G_D$ for all $M \in \mathcal{A}^S$. The corresponding Cayley graph $\mathcal{C}_D$ is isomorphic to $\mathcal{C}_L$.

One presentation of the Super-Apollonian group is [GLMWY2, Section 6]

$$\mathcal{A}^S = \left\langle S_1, S_2, S_3, S_4, S_1^\perp, S_2^\perp, S_3^\perp, S_4^\perp : S_j^2 = (S_j^\perp)^2 = 1, S_j S_k^\perp = S_k^\perp S_j \ (j \neq k) \right\rangle.$$

It is a right-angled hyperbolic reflection group generated by involutions $S_i, S_i^\perp$.

The group $\mathcal{A}_S$ is of index 48 in the full orthogonal group of Q_D [GLMWY3, Theorem 7.1].

2.3. Descartes quadruples, normal form and a spanning tree. The motivation for studying $\mathcal{A}^S$ came from the study of Apollonian circle packings; in this section we describe this geometry. We consider circles in $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ identified with $\mathbb{P}^1(\mathbb{C})$, having projective equation

$$aX\bar{X} - bY\bar{X} - \bar{b}YX + cY\bar{Y} = 0, \quad [X : Y] \in \mathbb{P}^1(\mathbb{C}), \quad ac - b\bar{b} = -1.$$

Such a circle is said to have *curvature* a (inverse of the radius in $[X : 1]$), *co-curvature* c (curvature of the circle in $[1 : Y]$), and *curvature-center* b (center times curvature). We denote it by a four-dimensional real vector (c, a, b_1, b_2) , where $b = b_1 + b_2 i$. These are known as ACC coordinates (*augmented curvature-center coordinates*) [GLMWY2]. When $a = 0$, the zero set in $[X : 1]$ is a line and b becomes a unit normal vector.

Four circles C_i as row vectors of a 4×4 matrix C in ACC coordinates are said to be in *Descartes configuration* if

$$C^t G_D C = \begin{pmatrix} 0 & -4 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

It is a theorem of Graham, Lagarias, Mallows, Wilks and Yan that circles are in *Descartes configuration* according to this algebraic condition if and only if, as circles, they are all mutually tangent with disjoint interiors, where sign of the curvature indicates orientation (hence interior) [LMW, Theorem 3.3] and [GLMWY2, Theorem 3.2]. This is nicely explained by interpreting the Descartes form as a bilinear pairing on circles which measure angle of intersection or hyperbolic distance; see for example [Koc].

Since it preserves Q_D , the Super-Apollonian group $\mathcal{A}^S$ acts on Descartes quadruples, in the form of such 4×4 matrices, from the left. There is a nice interpretation of this action geometrically [GLMWY2]. For such a configuration of circles, there is a dual Descartes quadruple consisting of circles passing orthogonally through the first quadruple, sharing the same set of tangency points (Figure 5). Then S_i acts as what we call a “swap” replacing C_i with its inversion in the dual circle orthogonal to the other three, and $S_i^\perp$ fixes C_i while replacing the other three circles with their inversions in C_i (we refer to the action of $S_i^\perp$ as simply an “inversion”). See Figure 5.

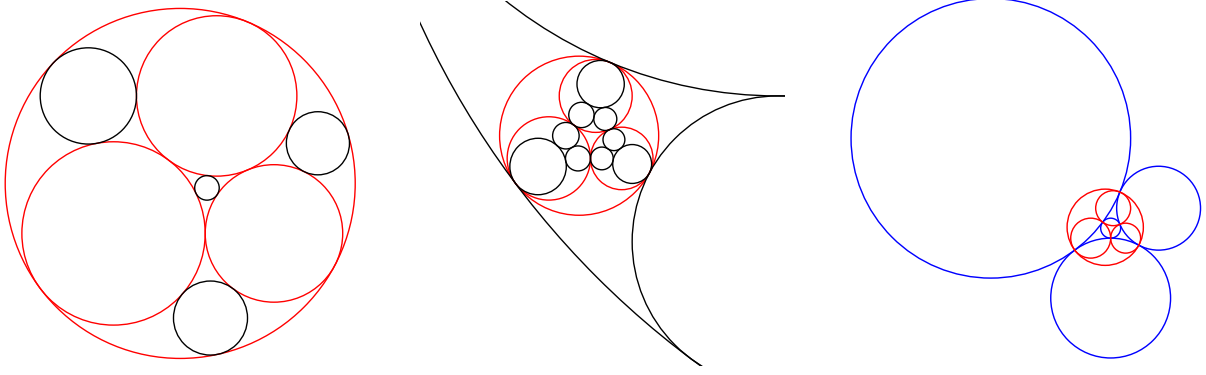


FIGURE 5. In all images, in red, is a Descartes quadruple. At left, the new circles produced by swaps are shown in black. At center, the new circles produced by inversions are shown in black. At right, the dual quadruple is shown in blue.

There are two “natural” ways of uniquely writing elements of $\mathcal{A}^S$ given the commutation relations in the group (i.e. two natural spanning trees for the Cayley graph of $\mathcal{A}^S$ with respect to the given generators), and we will be working with both of them. These were first defined in [GLMWY2].

Definition 2.1. A word $W = M_n M_{n-1} \cdots M_1$ in the Super-Apollonian generators is in swap normal form if $M_i \neq M_{i+1}$ and if whenever $M_i = S_j$ and $M_{i+1} = S_k^\perp$, then $j = k$; i.e. the “swaps” are pushed as far left as possible (equivalently the “inversions” are as far right as possible).

A word $W = M_n M_{n-1} \cdots M_1$ in the Super-Apollonian generators is in invert normal form if $M_i \neq M_{i+1}$ and if whenever $M_i = S_j^\perp$ and $M_{i+1} = S_k$, then $j = k$; i.e. the “inversions” are pushed as far left as possible (equivalently the “swaps” are as far right as possible).

The swap (invert) normal form of an element of the Super-Apollonian group is unique [GLMWY2]. In the Cayley graph $\mathcal{C}_Q$, travelling a path of length n to the origin, one can read off labels as $M_n, \dots, M_1$ where M_n is the distal edge and M_1 the proximal (to the origin). This gives a word $M_n \cdots M_1$ associated to the path.

Definition 2.2. Define the subgraph $\mathcal{T}_S$ of $\mathcal{C}_Q$ to be the union of all paths to the origin labelled by swap normal form words. This is called the swap down tree.

The reason for the terminology “swap down tree” is that, as one travels toward the origin in $\mathcal{C}_Q$ along the swap down tree, if one has a choice of $S_j^\perp$ followed by S_i or S_i followed by $S_j^\perp$ (both two-move sequences having the same endpoint closer to the origin), one must choose the latter, which is to say, one must swap before inverting. The following proposition asserts that besides being a spanning tree, it is minimal in a certain way.

Proposition 2.3. The swap down tree $\mathcal{T}_S$ is a spanning tree of $\mathcal{C}_Q$, and, from any vertex, the path to the origin along the tree is of minimal length among paths to the origin in $\mathcal{C}_Q$.

Proof. Every word can be put into a unique normal form, without increasing its length, by cancelling any double letters and moving each $S_i^\perp$ as far to the right as possible using the commutativity relations [GLMWY2, Proof of Theorem 6.1]. Therefore there is a unique path to the origin in $\mathcal{T}_S$ from any vertex of $\mathcal{C}_Q$. We may conclude that $\mathcal{T}_S$ is connected, is a tree, and spans $\mathcal{C}_Q$. Minimality follows from the observation that changing to swap normal form never increases length. $\square$

There are exactly analogous statements for the corresponding *invert down tree*.

2.4. Geometric realization of the Super-Apollonian group. The group $\mathrm{PGL}_2(\mathbb{Z}[i])$ acts on the extended complex plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ by the Möbius action

$$\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \cdot z = \frac{\alpha z + \gamma}{\beta z + \delta}.$$

This action can be extended to include complex conjugation,

$$\mathbf{c} \cdot z = \bar{z},$$

giving rise to the group $B[-1] = \mathrm{PGL}_2(\mathbb{Z}[i]) \rtimes \langle \mathbf{c} \rangle$ of Möbius transformations, the *extended Bianchi group*, a maximal discrete subgroup of $\mathrm{PSL}_2(\mathbb{C}) \rtimes \langle \mathbf{c} \rangle \cong \mathrm{Isom}(H^3)$. The group $B[-1]$ acts on the collection of circles of $\widehat{\mathbb{C}}$ (recall that lines are circles through ∞). In what follows, we will identify $\widehat{\mathbb{C}}$ with $\mathbb{P}^1(\mathbb{C})$ when convenient.

The orbit of the circle $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ under $\mathrm{PSL}_2(\mathbb{Z}[i])$ is a dense collection of nested circles called a *Schmidt arrangement*, denoted $\mathcal{S}_{\mathbb{Q}(i)}$. An image is shown in Figure 7. For more on Schmidt arrangements, see [St1, St2].

To describe our dynamical system, we choose a particular Descartes quadruple R_B , and its dual R_A , whose circles are the rows of

$$R_B = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 2 & 2 & 2 & 1 \end{pmatrix}, \quad R_A = R_B^\perp = \begin{pmatrix} 2 & 2 & 1 & 2 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

in ACC coordinates (see Figure 8), respectively. We call these the dual base quadruples. The terminology refers to the fact that R_A consists of the unique Descartes quadruple orthogonal to R_B and having the same intersection points (and vice versa). In particular, the “swaps” of R_A are the “inversions” of R_B and vice versa. Other choices of base quadruple would of course be possible, but the choice here coincides with a natural subset of the Schmidt arrangement and has particularly simple tangency points.

We embed the Super-Apollonian group into $\mathrm{PSL}_2(\mathbb{C}) \rtimes \langle \mathbf{c} \rangle$ using a form of the *exceptional isomorphism* $\mathrm{PGL}_2(\mathbb{C}) \rtimes \langle \mathbf{c} \rangle \cong O_{3,1}^+(\mathbb{R})$ (where the superscript $+$ denotes the orthochronous Lorentz transformations). To do so, we map each element of $\mathcal{A}^S$ to the Möbius transformation which acts the same way on R_B . The Möbius transformations corresponding to the Super-Apollonian generators are

$$\mathfrak{s}_1 = \frac{(1+2i)\bar{z}-2}{2\bar{z}-1+2i}, \quad \mathfrak{s}_2 = \frac{\bar{z}}{2\bar{z}-1}, \quad \mathfrak{s}_3 = -\bar{z}+2, \quad \mathfrak{s}_4 = -\bar{z},$$

$$\mathfrak{s}_1^\perp = \bar{z}, \quad \mathfrak{s}_2^\perp = \bar{z}+2i, \quad \mathfrak{s}_3^\perp = \frac{\bar{z}}{-2i\bar{z}+1}, \quad \mathfrak{s}_4^\perp = \frac{(1-2i)\bar{z}+2i}{-2i\bar{z}+1+2i}$$

(the $\mathfrak{s}_i$ are inversions in the circles of R_A and the $\mathfrak{s}_i^\perp$ are inversions in the circles of R_B). Let Γ denote the Möbius group generated by these generators; it is isomorphic to $\mathcal{A}^S$. Considering the Poincaré extension of the Möbius action to the upper-half-space model of hyperbolic space, one sees that Γ is the finite covolume Kleinian group generated by reflections in the sides of a right-angled ideal octahedron whose faces lie on the geodesic planes defined by the circles of R_B and R_A .

The orbit of the Super-Apollonian group on a particular Descartes quadruple is known as an Apollonian super-packing [GLMWY2]. For this choice of base quadruple, as a collection of circles, the corresponding Apollonian super-packing coincides with the Schmidt arrangement of $\mathcal{S}_{\mathbb{Q}(i)}$ [St2]. This orbit gives a sequence of partitions $\mathcal{P}_n$ of the plane into triangles and circles, each refining the last (see Figure 7). The regions of the partition $\mathcal{P}_n$ are indexed by the swap normal form words of length n in the Super-Apollonian generators; see Figure 6. A word $W' \in \mathcal{P}_{n+1}$ refines $W \in \mathcal{P}_n$ if and only if W is an initial segment of W' (right initial, speaking about $\mathcal{A}^S$). This will allow us to

coordinatize the plane by infinite words in swap normal form. The coordinates of a point will be produced by a dynamical system described below.

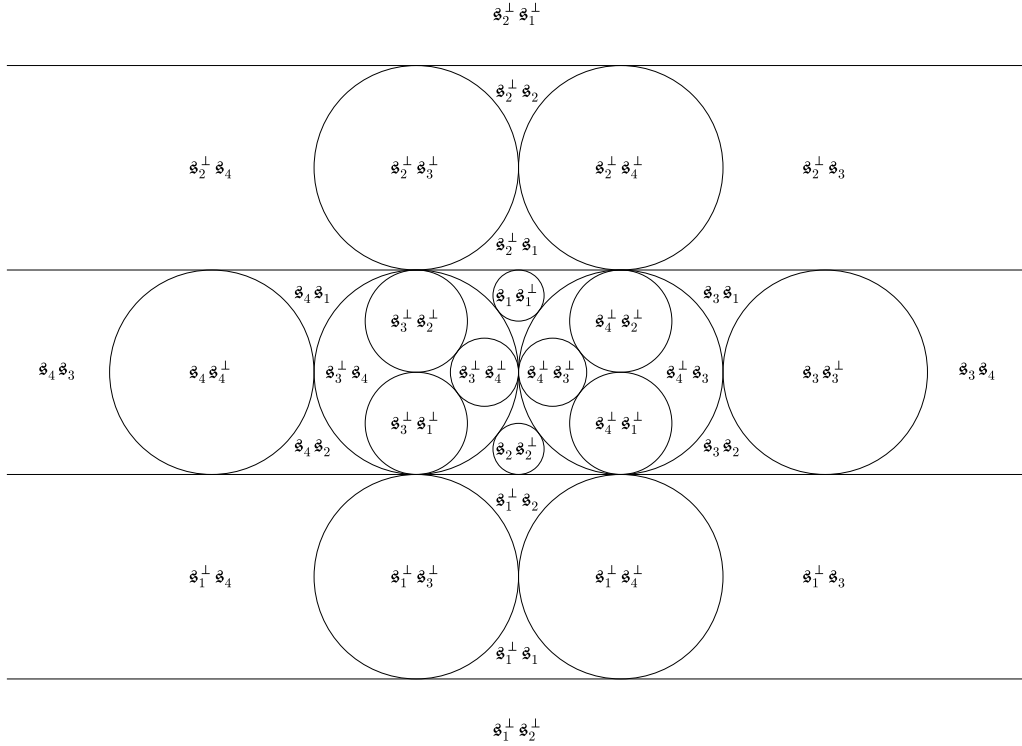


FIGURE 6. Partition of the complex plane induced by words of length two in swap normal form in the Möbius generators.

If $W = M_n M_{n-1} \cdots M_1$, $M_i \in \{S_j, S_j^\perp\}$ is a word in the Super-Apollonian generators and $Q = WR_B$, then in terms of the Möbius transformations the circles of Q are $\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_n c_i$ where c_i are the circles of R_B . The order reversal is a result of the fact that the M_i act on the left, while the $\mathbf{m}_i$ act on the right. The swap normal form for $\mathcal{A}^S$ passes to words in the Möbius generators, **reversing order as just noted**. Compare the following definition to Definition 2.1, noting the order reversal.

Definition 2.4. A word $\mathbf{w} = \mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_n$ in the Super-Apollonian Möbius generators is in swap normal form if $\mathbf{m}_i \neq \mathbf{m}_{i+1}$ and if whenever $\mathbf{m}_i = \mathbf{s}_j$ and $\mathbf{m}_{i+1} = \mathbf{s}_k^\perp$, then $j = k$; i.e. the “swaps” are pushed as far right as possible (equivalently the “inversions” as far left as possible).

A word $\mathbf{w} = \mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_n$ is in invert normal form if $\mathbf{m}_i \neq \mathbf{m}_{i+1}$ and if whenever $\mathbf{m}_i = \mathbf{s}_j^\perp$ and $\mathbf{m}_{i+1} = \mathbf{s}_k$, then $j = k$; i.e. “inversions” are as far right as possible.

Finally, we note that $\mathbf{s}_i^\perp = \mathfrak{d} \mathbf{s}_i \mathfrak{d}$, where

$$(2.2) \quad \mathfrak{d} = \frac{\bar{z} - 1 + i}{(1 - i)\bar{z} + i} = \mathfrak{d}^{-1}$$

is the isometry of the octahedron switching opposite faces, the Möbius version of the “duality operator”

$$D = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$$

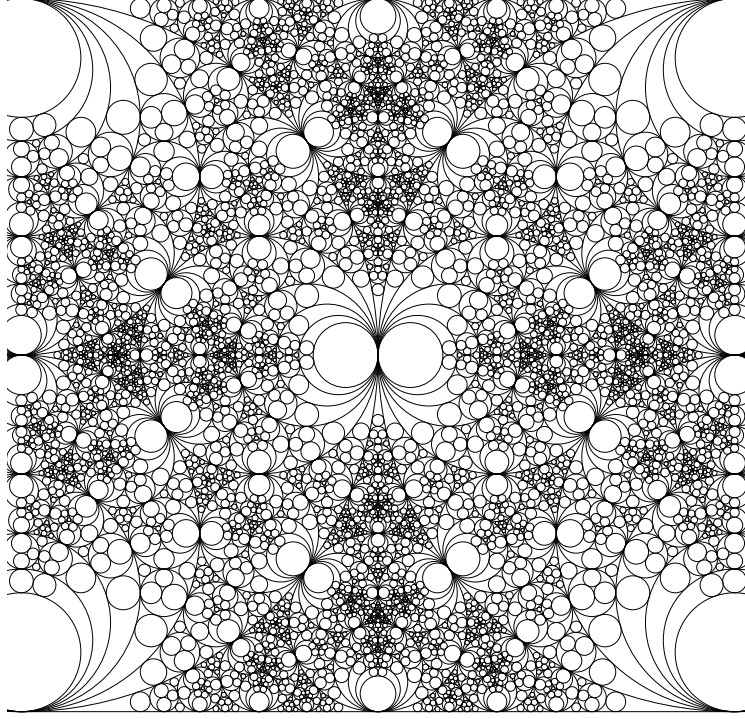


FIGURE 7. Circles in $[0, 1] \times [0, 1]$ generated by $\mathcal{A}^S$ swap normal form words of length ≤ 5 acting on R_B . The resulting image consists of a subcollection of the circles of the Schmidt arrangement $\mathcal{S}_{\mathbb{Q}(i)}$, which, as word length increases, will eventually include any circle of $\mathcal{S}_{\mathbb{Q}(i)}$.

from [GLMWY2]. This defines an involution on Super-Apollonian words in swap normal form, namely taking the transpose (or reversing order and conjugating by D)

$$M = M_n M_{n-1} \dots M_1, \quad M^\perp = M_1^\perp M_2^\perp \dots M_n^\perp.$$

On the level of Möbius transformations, $\mathfrak{m}^\perp = \mathfrak{d}\mathfrak{m}^{-1}\mathfrak{d}$. See [GLMWY2], [GLMWY3] for more information.

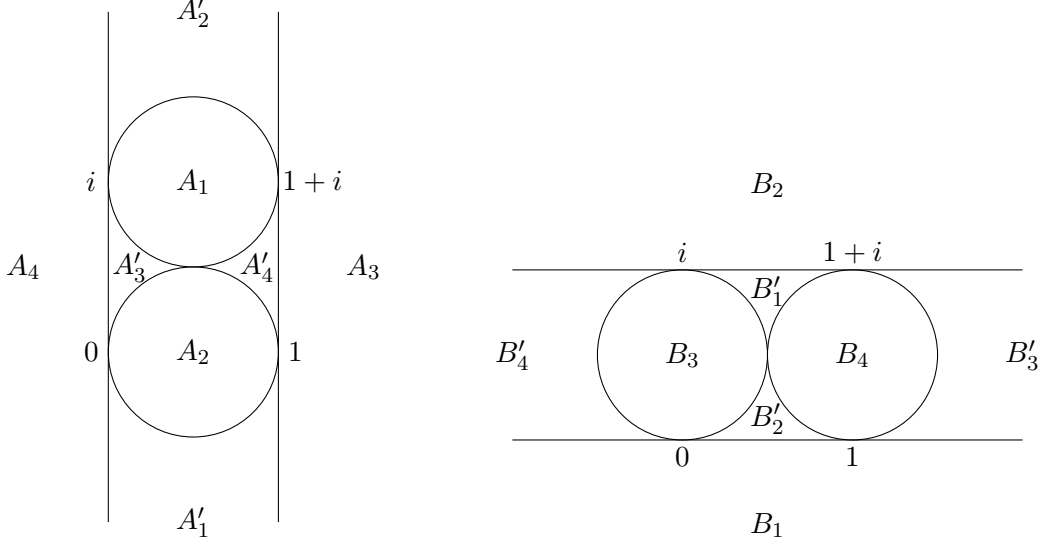
3. A PAIR OF DYNAMICAL SYSTEMS

3.1. Dynamics on $\mathbb{P}^1(\mathbb{C})$. We now define a pair of dynamical systems on $\mathbb{P}^1(\mathbb{C})$ associated to the base quadruples R_B and R_A . Let B_i, B'_i , be the open circular and closed triangular regions of the plane coming from the base quadruple R_B (A_i and A'_i defined similarly, see Figure 8), and define

$$T_B(z) = \begin{cases} \mathfrak{s}_i z & z \in B'_i, \\ \mathfrak{s}_i^\perp z & z \in B_i, \end{cases}, \quad T_A(w) = \begin{cases} \mathfrak{s}_i w & w \in A_i, \\ \mathfrak{s}_i^\perp w & w \in A'_i. \end{cases}$$

In words, if the point z is in one of the four closed triangular regions, we swap and if z is in one of the open circular regions, we invert in that circle. Each of T_A and T_B has six fixed points, the points of tangency $\{0, 1, \infty, i, i+1, 1/(1-i)\}$.

Theorem 3.1. *Under the dynamical systems T_A and T_B , every Gaussian rational $z \in \mathbb{Q}(i)$ reaches one of the fixed points in finite time. The fixed point reached is determined by the “parity” of the numerator and denominator of $z = p/q$, i.e. one of the six equivalence classes under the equivalence*

FIGURE 8. The regions A_i , A'_i , B_i , B'_i .

relation

$$\frac{p}{q} \sim \frac{r}{s} \iff ps \equiv qr \pmod{2}.$$

Proof. Recall that $B[-1] = \mathrm{PGL}_2(\mathbb{Z}[i]) \rtimes \langle \mathbf{c} \rangle$. The group Γ is the kernel of the surjective map

$$\mathrm{PGL}_2(\mathbb{Z}[i]) \rtimes \langle \mathbf{c} \rangle \rightarrow \mathrm{PGL}_2(\mathbb{Z}[i]/(2))$$

since Γ is in the kernel and both are of index 48 in $B[-1]$ (we have $[B[-1] : \Gamma] = 48$ comparing fundamental domains and $|\mathrm{PGL}_2(\mathbb{Z}[i]/(2))| = 48$ by direct computation). Hence Γ preserves parity, e.g.

$$\mathfrak{s}_1\left(\frac{p}{q}\right) = \frac{(1+2i)\bar{p} - 2\bar{q}}{2\bar{p} + \bar{q}(-1+2i)} \equiv \frac{p}{q} \pmod{2}.$$

Termination in finite time follows from the following version(s) of the Euclidean algorithm in $\mathbb{Z}[i]$. “Homogenizing” T_A and T_B gives dynamical systems on pairs $(0,0) \neq (p,q) \in \mathbb{Z}[i] \times \mathbb{Z}[i]$ that terminate when $p/q \in \{0, 1, \infty, i, 1+i, \frac{1}{1-i}\}$, i.e. act on pairs via complex conjugation and the matrices implied in the definitions of the $\mathfrak{s}_i$, $\mathfrak{s}_i^\perp$. For instance,

$$T_B(p, q) := (\bar{p}, 2\bar{p} - \bar{q}) \text{ for } \frac{p}{q} \in B'_2, \text{ where } T_B(p/q) = \mathfrak{s}_2(p/q) = \frac{\overline{p/q}}{2\overline{p/q} - 1}.$$

We’ll consider the case of T_B , noting that the proof for T_A is nearly identical:

$$(p, q) \mapsto \begin{cases} s_1(p, q) = ((1+2i)\bar{p} - 2\bar{q}, 2\bar{p} + (2i-1)\bar{q}) & p/q \in B'_1 \setminus \{i, 1+i, \frac{1}{1-i}\} \\ s_2(p, q) = (\bar{p}, 2\bar{p} - \bar{q}) & p/q \in B'_2 \setminus \{0, 1, \frac{1}{1-i}\} \\ s_3(p, q) = (2\bar{q} - \bar{p}, \bar{q}) & p/q \in B'_3 \setminus \{1, 1+i, \infty\} \\ s_4(p, q) = (-\bar{p}, \bar{q}) & p/q \in B'_4 \setminus \{0, i, \infty\} \\ s_1^\perp(p, q) = (\bar{p}, \bar{q}) & p/q \in B_1 \setminus \{0, 1, \infty\} \\ s_2^\perp(p, q) = (\bar{p} + 2i\bar{q}, \bar{q}) & p/q \in B_2 \setminus \{i, i+1, \infty\} \\ s_3^\perp(p, q) = (\bar{p}, \bar{q} - 2i\bar{p}) & p/q \in B_3 \setminus \{0, i, \frac{1}{1-i}\} \\ s_4^\perp(p, q) = ((1-2i)\bar{p} + 2i\bar{q}, (2i+1)\bar{q} - 2i\bar{p}) & p/q \in B_4 \setminus \{1, 1+i, \frac{1}{1-i}\}. \end{cases}$$

The inequalities defining the regions B_i , B'_i show that $|q|$ is reduced whenever s_1 , s_2 , $s_3^\perp$, or $s_4^\perp$ is applied. For example, when applying s_1 , the fact that p/q is in the triangle $B'_1 \setminus \{i, 1+i, \frac{1}{1-i}\}$ (or

the circle A_1) shows that $|q|$ is reduced when applying s_1 , as follows:

$$\frac{p}{q} \in B'_1 \setminus \left\{ i, 1+i, \frac{1}{1-i} \right\} \implies \left| \frac{p}{q} - \frac{1+2i}{2} \right|^2 < \frac{1}{4} \implies |2\bar{p} + (1-2i)\bar{q}| < |q|.$$

Note that the inequalities above define A_1 , but since B'_1 is contained in A_1 they hold true for B'_1 as well. Similarly, application of s_3 and $s_2^\perp$ both reduce $|p|$. Applying s_4 maps B'_4 onto $B'_1 \cup B'_2 \cup B'_4 \cup B'_3$ (from which one of $|p|$, $|q|$ will be reduced as just discussed). Finally, $s_1^\perp$ maps B_1 onto one of the other seven regions. Hence the algorithm terminates. $\square$

Iteration of the map T_B or T_A with input z produces a word $\mathfrak{z} = \mathfrak{m}_1 \cdots \mathfrak{m}_n \cdots$ in the Möbius generators $\mathfrak{s}_1, \mathfrak{s}_2, \mathfrak{s}_3, \mathfrak{s}_4, \mathfrak{s}_1^\perp, \mathfrak{s}_2^\perp, \mathfrak{s}_3^\perp, \mathfrak{s}_4^\perp$, where $\mathfrak{m}_n$ is defined by $T_B^n z = \mathfrak{m}_n(T_B^{n-1} z)$. We take the word to be finite for $z \in \mathbb{Q}(i)$, ending when a fixed point is reached. An example of this process is shown in Figure 9.

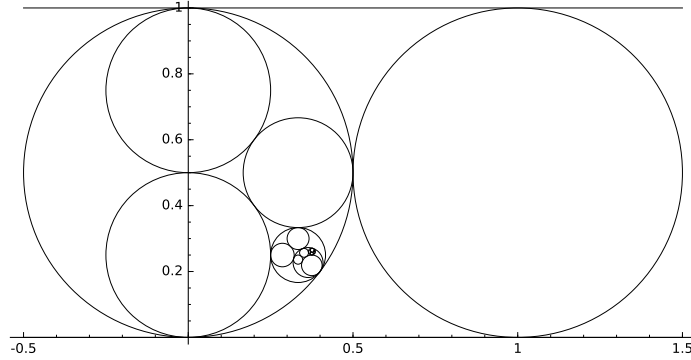


FIGURE 9. Approximating the irrational point $0.3828008104 \dots + i0.2638108161 \dots$ using T_B gives the normal form word $\mathfrak{s}_3^\perp \mathfrak{s}_2 \mathfrak{s}_2^\perp \mathfrak{s}_3^\perp \mathfrak{s}_1 \mathfrak{s}_1^\perp \mathfrak{s}_4 \mathfrak{s}_2 \mathfrak{s}_4 \mathfrak{s}_1 \mathfrak{s}_1^\perp \mathfrak{s}_3 \mathfrak{s}_2 \mathfrak{s}_3 \mathfrak{s}_3^\perp \mathfrak{s}_4 \mathfrak{s}_4^\perp \mathfrak{s}_2 \mathfrak{s}_1 \mathfrak{s}_2 \cdots$.

The collection $\mathcal{A}^S(n)$ of length n Super-Apollonian words in swap normal form partitions the plane into a collection of $9 \cdot 5^{n-1} - 1$ triangles and circles, which we call Farey circles and triangles following Schmidt [Sc1]. Specifically, we associate to each word in $\mathcal{A}^S(n)$ an open circular or closed triangular region, the notation being

$$F_B(\mathfrak{m}) = \mathfrak{m}_1 \cdots \mathfrak{m}_{n-1} B_i \text{ (circular), } \mathfrak{m}_1 \cdots \mathfrak{m}_{n-1} B_i' \text{ (triangular),}$$

for $\mathfrak{m} = \mathfrak{m}_1 \cdots \mathfrak{m}_n$ with $\mathfrak{m}_n = \mathfrak{s}_i^\perp$ (circular) or $\mathfrak{s}_i$ (triangular). This definition is set up so that the words of length one correspond to the eight regions of the base quadruple.

Theorem 3.2. *A word $\mathfrak{z} = \mathfrak{m}_1 \mathfrak{m}_2 \cdots$ produced by iteration by T_B (respectively T_A) on $z \in \mathbb{P}^1(\mathbb{C})$ is in swap (respectively invert) normal form. Furthermore,*

- (1) *If z is rational, then*

$$z = \mathfrak{z}b$$

for $b \in \{0, 1, \infty, i, 1+i, \frac{1}{1-i}\}$ matching z in parity as described in Theorem 3.1.

- (2) *If z is not rational, then $\mathfrak{z}$ is an infinite word with the property that*

$$\{z\} = \bigcap_{n \geq 1} F(\mathfrak{m}_1 \cdots \mathfrak{m}_n).$$

This gives a bijection $z \leftrightarrow \mathfrak{z}$, under which T_B (respectively T_A) can be considered to act on words, and this action is via the left-shift, $T_B(\mathfrak{m}_1 \mathfrak{m}_2 \cdots) = \mathfrak{m}_2 \mathfrak{m}_3 \cdots$.

Proof. That $\mathfrak{z}$ is in swap normal form is clear; the only circular region in $\mathfrak{s}_i(B'_i)$ is B_i .

Two Farey sets are either disjoint or one is contained in the other: if $\mathfrak{m}$ is a (left) initial segment of $\mathfrak{n}$, which we denote $\mathfrak{m} \leq \mathfrak{n}$, then $F_B(\mathfrak{m}) \supseteq F_B(\mathfrak{n})$. For $z \in \mathbb{C} \setminus \mathbb{Q}(i)$, the infinite swap normal form word $\mathfrak{z} = \mathfrak{m}_1 \mathfrak{m}_2 \dots$ produced by iterating T_B determines z since $z = \cap_n F(\mathfrak{m}_1 \dots \mathfrak{m}_n)$. For rational points, to determine z , we need both the finite word and the parity of the rational: then $z = \mathfrak{z}b$, where b is the element of $\{0, 1, \infty, i, 1+i, \frac{1}{1-i}\}$ of the specified parity. $\square$

From now on, we consistently use the variables $z, \mathfrak{z}$ for the B coordinate system, and $w, \mathfrak{w}$ for the A coordinate system, since we will be using both codings simultaneously.

3.2. Covering the boundary of the ideal octahedron and first approximation constant for $\mathbb{Q}(i)$. The purpose of this section and the next is to relate the Super-Apollonian continued fraction algorithm to classical statements of Diophantine approximation. In this section, we give the first value of the Lagrange spectrum for complex approximation by Gaussian rationals. With this as a point for comparison, in the second section we describe the goodness of the approximations obtained by the algorithm.

The “good” rational approximations to an irrational $z \in \mathbb{C}$

$$|z - p/q| \leq C/|q|^2$$

are determined by the collection of horoballs

$$\begin{aligned} B_C(p/q) &= \{(z, t) \in H^3 : |z - p/q|^2 + (t - C/|q|^2)^2 \leq C^2/|q|^4\}, \\ B_C(\infty) &= \{(z, t) \in H^3 : t \geq 1/2C\}, \end{aligned}$$

through which the geodesic $\overrightarrow{\infty z}$ passes (or through which any geodesic $\overrightarrow{wz}$ eventually passes). Over $\mathbb{Q}(i)$, the smallest value of C with the property that every irrational z has infinitely many rational approximations satisfying the above inequality was determined by Ford in [F]. Here we give a short proof of this fact using the geometry of the ideal octahedron.

Proposition 3.3. *Every $z \in \mathbb{C} \setminus \mathbb{Q}(i)$ has infinitely many rational approximations $p/q \in \mathbb{Q}(i)$ such that*

$$\left| z - \frac{p}{q} \right| \leq \frac{C}{|q|^2}, \quad C = \frac{1}{\sqrt{3}} \simeq 0.577350269 \dots$$

The constant $1/\sqrt{3}$ is the smallest possible, as witnessed by $z = \frac{1+\sqrt{-3}}{2}$.

Proof. The value of C for which the horoballs with parameter C based at the ideal vertices $\{0, 1, \infty, i, 1+i, 1/(1-i)\}$ cover the boundary of the fundamental octahedron is easily found to be $1/\sqrt{3}$ (one need only cover the face with vertices $\{0, 1, \infty\}$, see Figure 10). Hence as we follow the geodesic $\overrightarrow{\infty z}$ through the tessellation by octahedra, at least one of the six vertices of each octahedron satisfies the above inequality with $C = 1/\sqrt{3}$, one or two as it enters and one or two as it exits (some of which may coincide). This gives the smallest value of C for which the inequality above has infinitely many solutions for all irrational z , noting the the geodesic from $e^{-\pi i/3}$ to $e^{\pi i/3}$ passes orthogonally through the “centers” of the opposite faces of the octahedra through which it passes. $\square$

The sequence of octahedra we consider in our continued fraction algorithm are not necessarily along the geodesic path, but we do capture all rationals with $|z - p/q| < C/|q|^2$ with $C = 1/(1 + 1/\sqrt{2})$ as detailed in the next section.

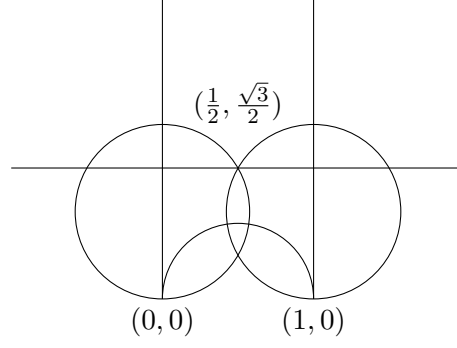


FIGURE 10. Horoball covering of the ideal triangular face with vertices $\{0, 1, \infty\}$ with coordinates (z, t) .

3.3. Quality of rational approximation. To any complex number we associate six sequences of Gaussian rational approximations by following the inverse orbit of the six points of tangency of our base quadruples R_A, R_B . Namely, if $z = \prod_{i=1}^{\infty} \mathfrak{z}_i = \prod_{i=1}^{\infty} \mathfrak{w}_i = w$ in the two codings, then the convergents $p_{n,\alpha}^A/q_{n,\alpha}^A, p_{n,\alpha}^B/q_{n,\alpha}^B$ are given by

$$\frac{p_{n,\alpha}^A}{q_{n,\alpha}^A} = \left(\prod_{i=1}^n \mathfrak{w}_i \right) (\alpha), \quad \frac{p_{n,\alpha}^B}{q_{n,\alpha}^B} = \left(\prod_{i=1}^n \mathfrak{z}_i \right) (\alpha), \quad \alpha \in \{0, 1, \infty, i, i+1, 1/(1-i)\},$$

with the property that

$$\lim_{n \rightarrow \infty} \frac{p_{n,\alpha}^A}{q_{n,\alpha}^A} = w, \quad \lim_{n \rightarrow \infty} \frac{p_{n,\alpha}^B}{q_{n,\alpha}^B} = z$$

for all α and $w \in \mathbb{C} \setminus \mathbb{Q}(i)$, $z \in \mathbb{C} \setminus \mathbb{Q}(i)$.

The following theorem is equivalent to a statement about approximation by Schmidt's continued fractions, given as Theorem 2.5 in [Sc1]. In particular, the approximations given by Schmidt's algorithm and the Super-Apollonian algorithm coincide. In [Sc1], it is stated without proof; here we provide a proof.

Theorem 3.4. *If p/q is such that*

$$|z_0 - p/q| < \frac{C}{|q|^2}, \quad C = \frac{\sqrt{2}}{1 + \sqrt{2}} \simeq 0.585786437 \dots,$$

then p/q is a convergent to z_0 (with respect to both T_A and T_B). Moreover, the constant C is the largest possible.

Proof. Note that the Apollonian super-packings associated to the root quadruples R_A, R_B , are invariant under the action of $\text{PSL}_2(\mathbb{Z}[i]) \rtimes \langle \mathfrak{c} \rangle$. Consider the quadruple where p/q first appears as a convergent to z_0 , and let $\gamma(z) = \frac{-Qz+P}{-qz+p} \in \text{PSL}_2(\mathbb{Z}[i])$ take this quadruple to the base quadruple (say R_B) with p/q mapping to infinity and infinity mapping to Q/q . For any value of C , the disk of radius $C/|q|^2$ centered at p/q gets mapped by γ to the exterior of the disk of radius $1/C$ centered at Q/q :

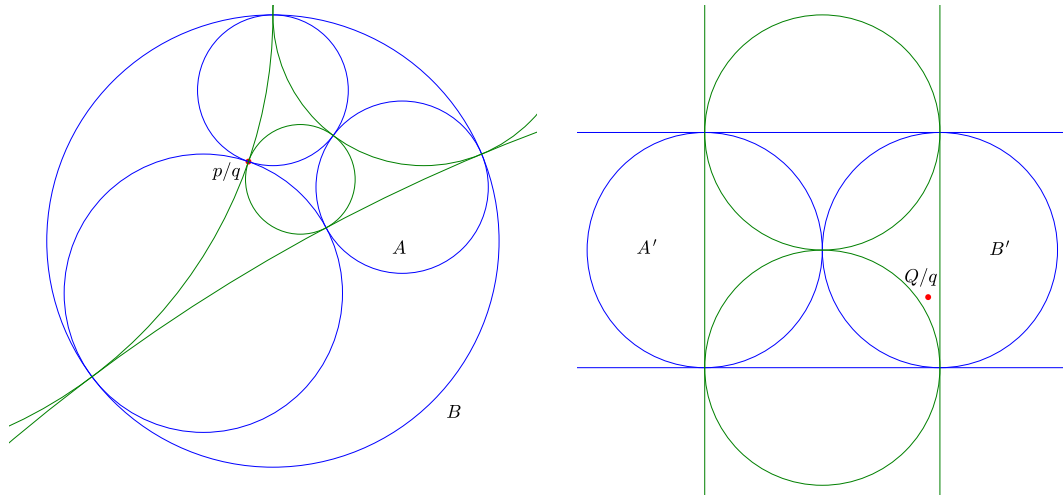
$$w = \frac{-Qz+P}{-qz+p} \Rightarrow |z - p/q| = \frac{1}{|w - Q/q||q|^2},$$

$$\frac{C}{|q|^2} \geq |z - p/q| = \frac{1}{|w - Q/q||q|^2} \Rightarrow |w - Q/q| \geq 1/C.$$

Consider the ways in which p/q can first appear as a convergent to z_0 in the sequence of partitions of the plane.

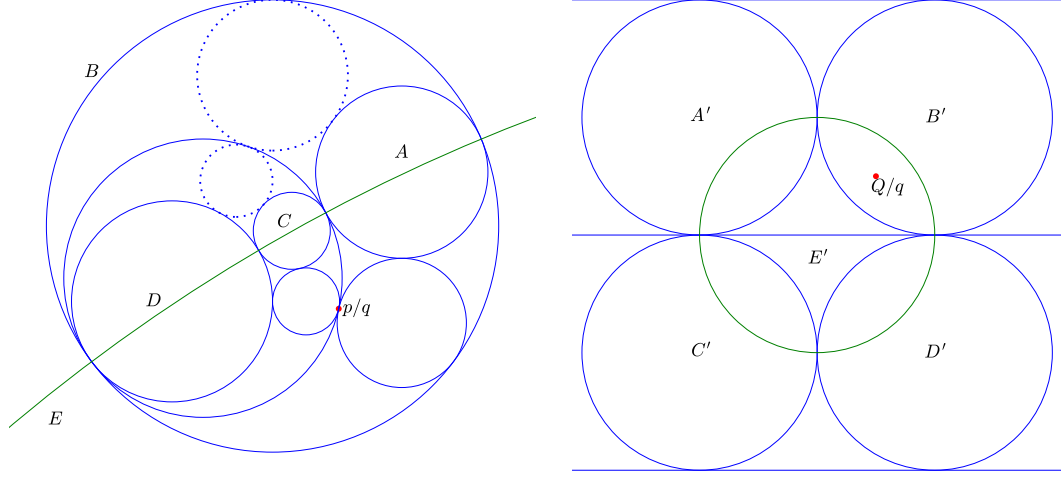
- We might invert into a circle containing z_0 . In particular, then, p/q is in the interior of the circle of inversion (since it is its first appearance as a convergent). In this case, by the discussion in Section 3.1, all z inside the circle also include this inversion in their expansion. Therefore, all z inside the circle will have p/q as a convergent. Our goal is to show that the circle of radius $1/|q|^2$ around p/q is contained in the circle of inversion.

Under γ above (perhaps after applying some binary tetrahedral symmetry of the base quadruple), the circles A, B , get mapped to A', B' as in the figures below, with $Q/q = \gamma(\infty)$ lying in the triangle inside B' as shown. The exterior of a disk of radius one centered at Q/q does not meet the interior of B' , hence, applying γ^{-1} , the disk of radius $1/|q|^2$ centered at p/q does not meet B . Therefore p/q is a convergent to any z with $|z - p/q| < 1/|q|^2$.



- We might swap into a triangle containing z_0 producing p/q as a point of tangency on an edge of this triangle. In the image, z_0 is in the quadrangle with sides formed by A, B, C, D ; this is the union of two triangles. The dotted circles in the figures indicate the two possible swaps associated to the initial creation of p/q as a convergent. In this case, any z in the indicated quadrangle will have p/q as a convergent. We aim to show that a circle of radius $C/|q|^2$ around p/q is contained in this region.

Under γ above (perhaps after applying some binary tetrahedral symmetry of the base quadruple), the circles A, B, C, D , and E are mapped to A', B', C', D' and E' , the three circles tangent to p/q are mapped to the lines in the second picture, with $Q/q = \gamma(\infty)$ lying in the intersection of the disks defined by B' and E' . The exterior of any circle of radius $1/C = 1 + 1/\sqrt{2}$ centered inside E' avoids the interiors of A', B', C' , and D' . Applying γ^{-1} shows that the disk of radius $C/|q|^2$ around p/q does not meet A, B, C , or D , so that p/q is a convergent to any z with $|z - p/q| < C/|q|^2$.



□

3.4. Invertible extension and invariant measures. In this section, we derive an invertible extension T of T_B , with the property that T^{-1} extends T_A , along with an invariant measure for T . This is done with a goal of eventually producing an ergodic measure preserving system, as is done both in [R] and [Sc2]. For this purpose, let H^3 denote hyperbolic 3-space, having boundary $\widehat{\mathbb{C}} \sim \mathbb{P}^1(\mathbb{C})$ (e.g. the Poincaré upper half space model).

The space of oriented geodesics in H^3 , identified with pairs in $\mathbb{P}^1(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C}) \setminus \Delta$, where Δ denotes the diagonal, carries an isometry invariant measure

$$|z - w|^{-4} du dv dx dy, \quad w = u + iv, \quad z = x + iy.$$

We restrict this measure to geodesics in the set

$$\mathcal{G} = \left(\bigcup_i A_i \times B_i \right) \cup \left(\bigcup_{i \neq j} A'_i \times B_j \right) \cup \left(\bigcup_{i \neq j} A_i \times B'_j \right) \cup \left(\bigcup_{i,j} A'_i \times B'_j \right),$$

consisting of geodesics between disjoint A and B regions of the base quadruples (see Figure 8). In what follows we use A coordinates for w in the first coordinate and B coordinates for z in the second coordinate. Define $T : \mathcal{G} \rightarrow \mathcal{G}$ by

$$T(w, z) = \begin{cases} (\mathfrak{s}_i w, \mathfrak{s}_i z) & z \in B_i, \mathfrak{z} = \mathfrak{s}_i \dots \\ (\mathfrak{s}_i^\perp w, \mathfrak{s}_i^\perp z) & z \in B'_i, \mathfrak{z} = \mathfrak{s}_i^\perp \dots \end{cases}$$

where $\mathfrak{z}$ is the Möbius transformation corresponding to z as described in Theorem 3.2. Equivalently,

$$T^{-1}(w, z) = \begin{cases} (\mathfrak{s}_i w, \mathfrak{s}_i z) & w \in A_i, \mathfrak{w} = \mathfrak{s}_i \dots \\ (\mathfrak{s}_i^\perp w, \mathfrak{s}_i^\perp z) & w \in A'_i, \mathfrak{w} = \mathfrak{s}_i^\perp \dots \end{cases}.$$

In other words, T is applying T_B diagonally depending on the second coordinate. In terms of the shifts on pairs $(\mathfrak{w}, \mathfrak{z}) = (\prod_i \mathfrak{w}_i, \prod_i \mathfrak{z}_i)$ corresponding to (w, z) , we have

$$T(w, z) = (\mathfrak{z}_1 \mathfrak{w}, \prod_{i=2}^{\infty} \mathfrak{z}_i) = (\mathfrak{z}_1 w, T_B(z)), \quad T^{-1}(w, z) = (\prod_{i=2}^{\infty} \mathfrak{w}_i, \mathfrak{w}_1 \mathfrak{z}) = (T_A(w), \mathfrak{w}_1 z).$$

See Figure 11 for a visualization of the invertible extension.

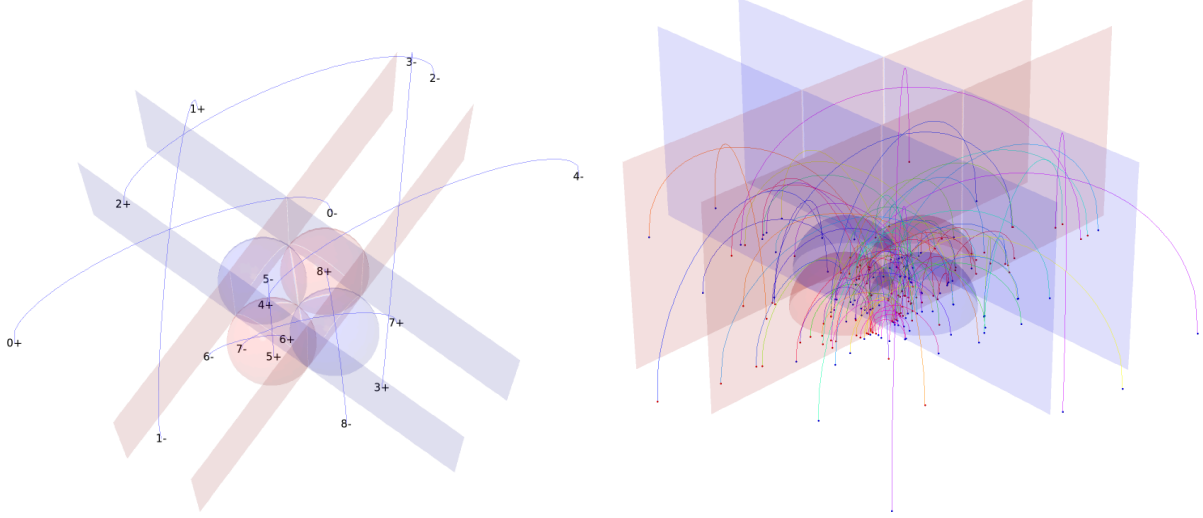


FIGURE 11. At left, eight iterations of the map T , labelled in sequence (with $+$ and $-$ indicating the orientation). At right, 100 iterations of T , with rainbow coloration indicating time. In both pictures, the geodesic planes of the base quadruple and its dual are shown.

Define the following regions (see Figure 12):

$$\begin{aligned}\mathcal{A}_i &= B_i \cup (\cup_{j \neq i} B'_j) \\ \mathcal{A}'_i &= (\cup_j B'_j) \cup (\cup_{j \neq i} B_j) \\ \mathcal{B}_i &= A_i \cup (\cup_{j \neq i} A'_j) \\ \mathcal{B}'_i &= (\cup_j A'_j) \cup (\cup_{j \neq i} A_j)\end{aligned}$$

Here $\mathcal{A}_i$ consists of the B regions not intersecting A_i , and so on.

Theorem 3.5. *The function $T : \mathcal{G} \rightarrow \mathcal{G}$ is a measure-preserving bijection. Consequently, the push-forward of this measure onto the first or second coordinate gives invariant measures μ_A and μ_B for T_A and T_B . Specifically, we have*

$$d\mu_A(w) = f_A(w) du dv = \begin{cases} f_{A_i}(u, v) du dv = du dv \int_{\mathcal{A}_i} |z - w|^{-4} dx dy, & w \in A_i \\ f_{A'_i}(u, v) du dv = du dv \int_{\mathcal{A}'_i} |z - w|^{-4} dx dy, & w \in A'_i \end{cases}$$

and

$$d\mu_B(z) = f_B(z) dx dy = \begin{cases} f_{B_i}(x, y) dx dy = dx dy \int_{\mathcal{B}_i} |z - w|^{-4} du dv, & z \in B_i \\ f_{B'_i}(x, y) dx dy = dx dy \int_{\mathcal{B}'_i} |z - w|^{-4} du dv, & z \in B'_i \end{cases}.$$

See Figure 13 for a graph of the invariant density f_B .

Proof. Note that $\mathcal{G} = \cup_i (A_i \times \mathcal{A}_i \cup A'_i \times \mathcal{A}'_i) = \cup_i (\mathcal{B}_i \times B_i \cup \mathcal{B}'_i \times B'_i)$, which is readily seen in Figure 12. Since

$$\begin{aligned}T(\mathcal{B}_i \times B_i) &= A'_i \times \mathcal{A}'_i, \\ T(\mathcal{B}'_i \times B'_i) &= A_i \times \mathcal{A}_i,\end{aligned}$$

we immediately get that T is a bijection.

As T is a bijection defined piecewise by isometries, it preserves the measure described above. $\square$

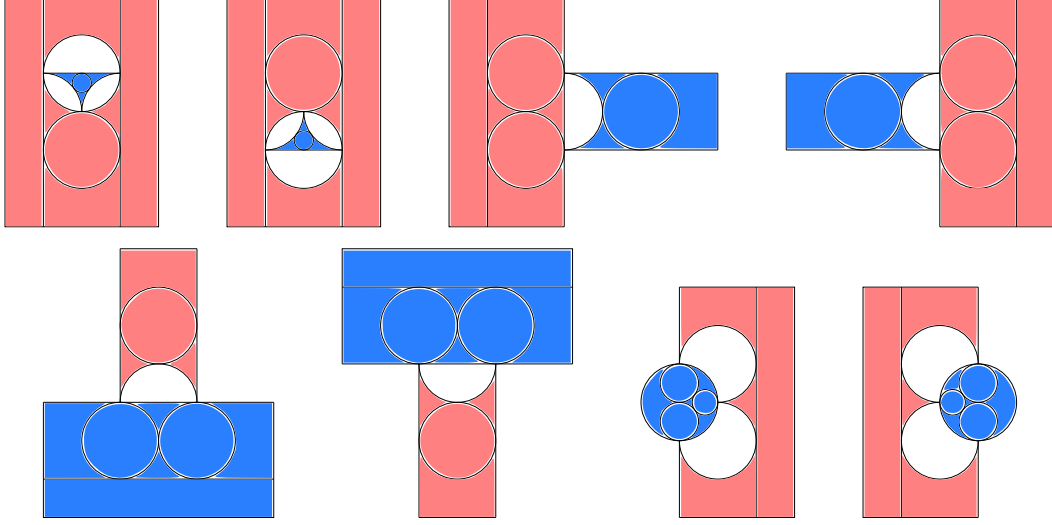


FIGURE 12. Top, left to right: regions $B'_i \times B'_i$, $i = 1, 2, 3, 4$. Bottom, left to right: regions $B_i \times B_i$, $i = 1, 2, 3, 4$. Shown red $\times$ blue with subdivisions.

Theorem 3.5 defines the functions f_A and f_B implicitly. We now compute what they are explicitly, starting with f_B . Computing the relevant integrals in Theorem 3.5 gives $\pi/4$ times hyperbolic area on the triangular regions B'_i :

$$f_B(x, y) = \begin{cases} \frac{\pi}{4(\frac{1}{4}-d^2)^2} & z \in B'_1, \quad d^2 = (x - \frac{1}{2})^2 + (y - 1)^2 \\ \frac{\pi}{4(\frac{1}{4}-d^2)^2} & z \in B'_2, \quad d^2 = (x - \frac{1}{2})^2 + y^2 \\ \frac{\pi}{4(1-x)^2} & z \in B'_3 \\ \frac{\pi}{4x^2} & z \in B'_4 \end{cases},$$

and on the circular regions B_i we have

$$f_B(x, y) = \begin{cases} H(x, y) & z \in B_1 \\ H(x, 1-y) & z \in B_2 \\ G(x, y) & z \in B_3 \\ G(1-x, y) & z \in B_4 \end{cases},$$

where

$$\begin{aligned} H(x, y) &= h(x, y) + h(1-x, y) + h(x^2 - x + y^2, y), \\ G(x, y) &= h(x, y^2 - y + x^2) + h(x^2 - x + y^2, y^2 - y + x^2) + h(x^2 - x + (1-y)^2, y^2 - y + x^2), \\ h(x, y) &= \frac{\arctan(x/y)}{4x^2} - \frac{1}{4xy}. \end{aligned}$$

Furthermore, we have the relationship

$$f_A(w) = f_B(\rho w) = f_B(\mathfrak{d}w)$$

where ρ is rotation by $\pi/2$ around $1/(1-i)$ and $\mathfrak{d}$ is the isometry switching opposite faces of the octahedron (the duality operator 2.2). The measures μ_A , μ_B have S_3 symmetry on each of the A_i , A'_i , B_i , B'_i . For instance the isometries permuting $\{0, 1, \infty\}$ on A'_1 , B_1 preserve the measure (generators shown for a transposition and three-cycle)

$$(0, 1) \sim -\bar{z} + 1, \quad (0, 1, \infty) \sim \frac{-1}{z-1}.$$

The measures also have S_4 symmetry on $\mathbb{C}$, permuting the pairs $\{A_i, A'_i\}$, $\{B_i, B'_i\}$ (transpositions $(i, i+1)$ shown)

$$(1, 2) \sim \bar{z} + i, \quad (2, 3) \sim \frac{1}{\bar{z}}, \quad (3, 4) \sim -\bar{z} + 1.$$

The total measure assigned to each of the A_i, A'_i, B_i, B'_i , is $\pi^2/4$, so to normalize μ_A, μ_B , we divide by $2\pi^2$.

All of the above should be compared with Nakada's extension of Schmidt's system, described in an appendix.

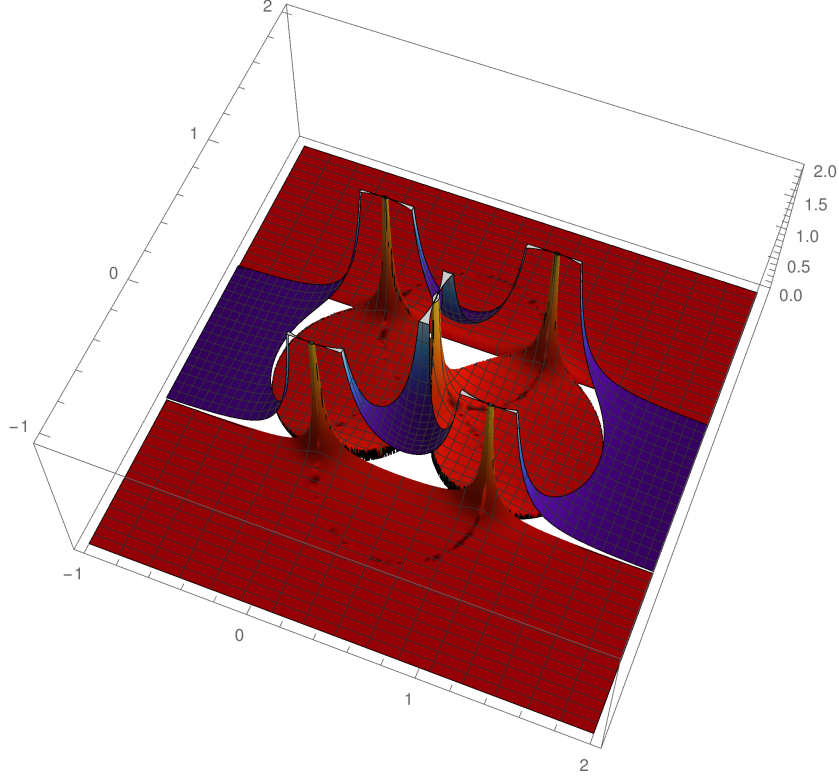


FIGURE 13. Graph of $f_B(z)$, the invariant density for T_B .

The following lemmas compare the measures of some Farey circles and triangles via the involutions $\perp$ and $^{-1}$, simplifying computations in Section 5. For use in the proofs, we note the equality of regions

$$(3.1) \quad \begin{aligned} \mathfrak{s}_i^\perp \mathcal{B}_i &= A'_i = \mathfrak{d} B'_i, \\ \mathfrak{s}_i \mathcal{B}'_i &= A_i = \mathfrak{d} B_i, \\ \mathfrak{s}_i \mathcal{A}_i &= B'_i = \mathfrak{d} A'_i, \\ \mathfrak{s}_i^\perp \mathcal{A}'_i &= B_i = \mathfrak{d} A_i, \end{aligned}$$

where $\mathfrak{d}$ is as in (2.2).

Lemma 3.6. *For $\mathfrak{m}$ in swap normal form, and $\mathfrak{n}$ in invert normal form, there are equalities of measure*

$$\mu_B(F_B(\mathfrak{m})) = \mu_B(F_B(\mathfrak{m}^\perp)), \quad \mu_A(F_A(\mathfrak{n})) = \mu_A(F_A(\mathfrak{n}^\perp)).$$

Proof. Consider the case where $\mathbf{m} = \mathfrak{s}_i \dots \mathfrak{s}_j$ so that $F_B(\mathbf{m}) = \mathbf{m}\mathfrak{s}_j B'_j \subseteq B'_i$ and $F_B(\mathbf{m}^\perp) = \mathbf{m}^\perp \mathfrak{s}_i^\perp B_i \subseteq B_j$ (all other cases are analogous) and let $\omega = |z - w|^{-4} du dv dx dy$ be the invariant form on the space of geodesics. Recall $\mathbf{m}^\perp = \mathfrak{d}\mathbf{m}^{-1}\mathfrak{d}$ from (2.2). We have

$$\mu_B(F_B(\mathbf{m})) = \int_{B'_i} \int_{F_B(\mathbf{m})} \omega = \int_{B'_i} \int_{\mathbf{m}\mathfrak{s}_j B'_j} \omega$$

while

$$\begin{aligned} \mu_B(F_B(\mathbf{m}^\perp)) &= \int_{B_j} \int_{F_B(\mathbf{m}^\perp)} \omega = \int_{B_j} \int_{\mathbf{m}^\perp \mathfrak{s}_i^\perp B_i} \omega = \int_{B_j} \int_{\mathfrak{d}\mathbf{m}^{-1}\mathfrak{d}\mathfrak{s}_i^\perp B_i} \omega = \int_{\mathbf{m}\mathfrak{d}B_j} \int_{\mathfrak{d}\mathfrak{s}_i^\perp B_i} \omega \\ &= \int_{\mathbf{m}\mathfrak{d}B_j} \int_{\mathfrak{s}_i \mathfrak{d}B_i} \omega = \int_{\mathbf{m}\mathfrak{s}_j B'_j} \int_{\mathfrak{s}_i A_i} \omega = \int_{\mathbf{m}\mathfrak{s}_j B'_j} \int_{B'_i} \omega. \end{aligned}$$

Here we are using the fact that $\omega = |z - w|^{-4} du dv dx dy$ is the invariant form and the relations in (3.1). $\square$

Lemma 3.7. *For $\mathbf{m}$ in swap normal form, and $\mathbf{n}$ in invert normal form, there are equalities of measure*

$$\mu_B(F_B(\mathbf{m})) = \mu_A(F_A(\mathbf{m}^{-1})), \quad \mu_A(F_A(\mathbf{n})) = \mu_B(F_B(\mathbf{n}^{-1})).$$

Proof. Consider the case where $\mathbf{m} = \mathfrak{s}_i \dots \mathfrak{s}_j$ so that $F_B(\mathbf{m}) = \mathbf{m}\mathfrak{s}_j B'_j \subseteq B'_i$ and $F_A(\mathbf{m}^{-1}) = \mathbf{m}^{-1} \mathfrak{s}_i A_i \subseteq A_j$ (all other cases are analogous) and let $\omega = |z - w|^{-4} du dv dx dy$ be the invariant form on the space of geodesics. Then

$$\mu_B(F_B(\mathbf{m})) = \int_{B'_i} \int_{F_B(\mathbf{m})} \omega = \int_{B'_i} \int_{\mathbf{m}\mathfrak{s}_j B'_j} \omega = \int_{\mathfrak{s}_i A_i} \int_{\mathbf{m} A_j} \omega = \int_{\mathbf{m}^{-1} \mathfrak{s}_i A_i} \int_{A_j} \omega$$

while

$$\mu_A(F_A(\mathbf{m}^{-1})) = \int_{F_A(\mathbf{m}^{-1})} \int_{A_j} \omega = \int_{\mathbf{m}^{-1} \mathfrak{s}_i A_i} \int_{A_j} \omega.$$

Here we are using the fact that $\omega = |z - w|^{-4} du dv dx dy$ is the invariant form and the relations in (3.1). $\square$

4. DYNAMICAL SYSTEMS ON LORENTZ AND DESCARTES QUADRUPLES

The dynamical systems introduced in the previous section can be translated to ones on Lorentz and Descartes quadruples. In this section, we explore these systems as well as how they interact with individual Apollonian packings.

4.1. Dynamics on Lorentz quadruples. In analogy to Romik [R], we now define a dynamical system on integer Lorentz quadruples $a^2 = b^2 + c^2 + d^2$ with $a > 0$ which decreases the value of a and terminates at one of the six quadruples

$$(g, \pm g, 0, 0), (g, 0, \pm g, 0), (g, 0, 0, \pm g), \quad g = \gcd(a, b, c, d).$$

Define the following dynamical system on $C = \{(a, b, c, d) : a^2 = b^2 + c^2 + d^2, a > 0\} \subset \mathbb{R}^4$:

$$T_L(a, b, c, d) = \begin{cases} L_1^\perp(a, b, c, d)^t & \text{if } 2a + b + c + d \leq a \\ L_2^\perp(a, b, c, d)^t & \text{if } 2a + b - c - d \leq a \\ L_3^\perp(a, b, c, d)^t & \text{if } 2a - b + c - d \leq a \\ L_4^\perp(a, b, c, d)^t & \text{if } 2a - b - c + d \leq a \\ \text{if none of the above then} & \\ L_1(a, b, c, d)^t & \text{if } 2a - b - c - d \leq a \\ L_2(a, b, c, d)^t & \text{if } 2a - b + c + d \leq a \\ L_3(a, b, c, d)^t & \text{if } 2a + b - c + d \leq a \\ L_4(a, b, c, d)^t & \text{if } 2a + b + c - d \leq a \end{cases}.$$

Iteration of T_L produces a word $W_1 W_2 \cdots W_n \cdots$ defined by $T_L^k(a, b, c, d) = W_k T_L^{k-1}(a, b, c, d)$.

Theorem 4.1. *The word $W_1 W_2 \cdots W_n$ produced by iteration of T_L on a primitive integer Lorentz quadruple (a, b, c, d) is in swap normal form and satisfies*

$$(a, b, c, d)^t = W_1 W_2 \cdots W_k \mathbf{b}^t$$

where $\mathbf{b}$ is one of the following simplest Lorentz quadruples:

$$(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (1, -1, 0, 0), (1, 0, -1, 0), (1, 0, 0, -1),$$

under iteration of T_L .

See Figure 4.

Proof. The map is well defined unless

$$a \pm b \pm c \pm d > 0$$

for all choices of signs. In this case $a < |b| + |c| + |d|$, which is impossible given $a^2 = b^2 + c^2 + d^2$.

Note that the map preserves $\gcd(a, b, c, d)$, so it is well-defined on $\tilde{C}$, the primitive integral Lorentz quadruples of C .

Therefore, it suffices to verify the following:

$$T_L(W_1 W_2 \cdots W_n \mathbf{b}^t) = W_2 W_3 \cdots W_n \mathbf{b}^t.$$

We do this by constructing an explicit intertwining map demonstrating that $(\mathbb{P}^1(\mathbb{Q}(i)), T_B)$ and $(\tilde{C}, T_L)$ are conjugate. Then the result will follow from Theorem 3.2.

In analogy to work of Romik on Pythagorean triples, it is possible to scale this system to act on a sphere. Scaling so that $a = 1$ ($X = b/a$, $Y = c/a$, $Z = d/a$) (call this projection π), we get a system on the sphere

$$T_{sph}(X, Y, Z) = \begin{cases} \frac{(-1-Y-Z, -1-X-Z, -1-X-Y)}{2+X+Y+Z} & \text{if } 1 + X + Y + Z < 0 \\ \frac{(-1+Y+Z, 1+X-Z, 1+X-Y)}{2+X-Y-Z} & \text{if } 1 + X - Y - Z < 0 \\ \frac{(1+Y-Z, -1+X+Z, 1-X+Y)}{2-X+Y-Z} & \text{if } 1 - X + Y - Z < 0 \\ \frac{(1-Y+Z, 1-X+Z, -1+X+Y)}{2-X-Y+Z} & \text{if } 1 - X - Y + Z < 0 \\ \text{if none of the above, then} & \\ \frac{(1-Y-Z, 1-X-Z, 1-X-Y)}{2-X-Y-Z} & \text{if } 1 - X - Y - Z < 0 \\ \frac{(1+Y+Z, -1+X-Z, -1+X-Y)}{2-X+Y+Z} & \text{if } 1 - X + Y + Z < 0 \\ \frac{(-1+Y-Z, 1+X+Z, -1-X+Y)}{2+X-Y+Z} & \text{if } 1 + X - Y + Z < 0 \\ \frac{(-1-Y+Z, -1-X+Z, 1+X+Y)}{2+X+Y-Z} & \text{if } 1 + X + Y - Z < 0 \end{cases}.$$

The regions defined on the sphere (Figure 14) are given by the intersection of the sphere with the hyper-ideal tetrahedron defined by the linear inequalities

$$1 + X + Y + Z \geq 0, 1 + X - Y - Z \geq 0, 1 - X + Y - Z \geq 0, 1 - X - Y + Z \geq 0$$

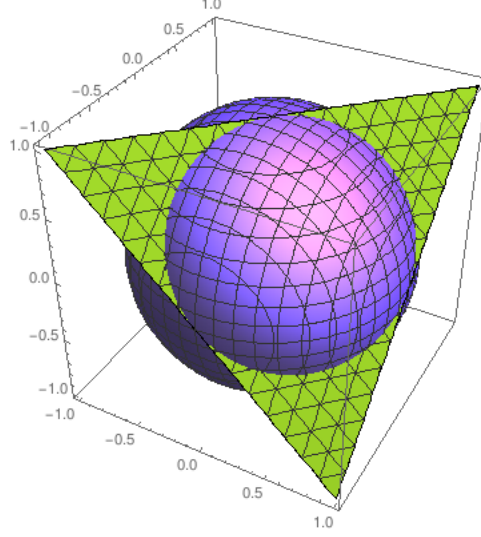


FIGURE 14. The hyper-ideal tetrahedron in the Klein model

in the Klein projective model of hyperbolic space, and the cases of T_{sph} are reflections in those geodesic planes and the planes of the dual tetrahedron.

The systems (S^2, T_{sph}) and $(\mathbb{P}^1(\mathbb{C}), T_B)$ are conjugate, moving between the projective and the upper half-space models of hyperbolic space. Specifically, after stereographic projection $(X, Y, Z) \mapsto (\frac{X}{1-Z}, \frac{Y}{1-Z}) = z$, rotating $(e^{-\pi i/4}z)$, scaling $(z/\sqrt{2})$, shifting $(z + \frac{1+i}{2})$, and switching two circles $(\frac{z}{-i\bar{z}+1})$, we obtain an intertwining map

$$\phi(X, Y, Z) = \frac{i\bar{z} + 1}{\bar{z} + 1} = \frac{(1 + Y - Z) + iX}{(1 + X - Z) - iY}, \quad T_B \circ \phi = \phi \circ T_{sph}.$$

Under the composition $\phi \circ \pi$, one can associate a Lorentz quadruple (a, b, c, d) with the complex point

$$z = \phi \circ \pi(a, b, c, d)^t := \frac{a + c - d + bi}{a + b - d - ci},$$

where π is the scaling projection from above, the correspondence between fixed points being

$$(1, \pm 1, 0, 0) \mapsto \frac{1}{1-i}, \infty, \quad (1, 0, \pm 1, 0) \mapsto 1 + i, 0, \quad (1, 0, 0, \pm 1) \mapsto i, 1.$$

Then,

$$\mathfrak{s}_i \circ \phi \circ \pi(a, b, c, d)^t = \phi \circ \pi \circ L_i(a, b, c, d)^t, \quad \mathfrak{s}_i^\perp \circ \phi \circ \pi(a, b, c, d)^t = \phi \circ \pi \circ L_i^\perp(a, b, c, d)^t,$$

and the conditions defining the cases of T_L and T_B correspond. Therefore, $T_B \circ \phi \circ \pi = \phi \circ \pi \circ T_L$. $\square$

To summarize the above: the following diagram commutes, ϕ is invertible, and there is a unique primitive representative in $\pi^{-1}(q)$ for rational $q \in S^2$.

$$\begin{array}{ccc}
 C & \xrightarrow{T_L} & C \\
 \pi \downarrow & & \downarrow \pi \\
 S^2 & \xrightarrow{T_{sph}} & S^2 \\
 \phi \downarrow & & \downarrow \phi \\
 \mathbb{P}^1(\mathbb{C}) & \xrightarrow{T_B} & \mathbb{P}^1(\mathbb{C})
 \end{array}$$

4.2. Dynamics on Descartes quadruples. Under the change of variables of Section 2.2, the dynamical system of the last section becomes a dynamical system on primitive integer Descartes quadruples. Define the following:

$$T_S(a, b, c, d) = \begin{cases} S_1^\perp(a, b, c, d)^t & \text{if } a < 0 \\ S_2^\perp(a, b, c, d)^t & \text{if } b < 0 \\ S_3^\perp(a, b, c, d)^t & \text{if } c < 0 \\ S_4^\perp(a, b, c, d)^t & \text{if } d < 0 \\ \text{if none of the above then} & . \\ S_1(a, b, c, d)^t & \text{if } b + c + d < a \\ S_2(a, b, c, d)^t & \text{if } a + c + d < b \\ S_3(a, b, c, d)^t & \text{if } a + b + d < c \\ S_4(a, b, c, d)^t & \text{if } a + b + c < d \end{cases}$$

Iteration of T_S produces a word $W_1 W_2 \cdots W_n$ defined by $T_S^k(a, b, c, d)^t = W_k T_S^{k-1}(a, b, c, d)^t$.

Theorem 4.2. *The word $W_1 W_2 \cdots W_n$ produced by iteration of T_S on a primitive integer Descartes quadruple (a, b, c, d) is in swap normal form and satisfies*

$$(a, b, c, d)^t = W_1 W_2 \cdots W_k \sigma(1, 1, 0, 0)^t$$

where σ is some permutation of the entries of the vector. In other words, any primitive integral Descartes quadruple (a, b, c, d) eventually reaches one of the so-called simplest Descartes quadruples

$$(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)$$

under iteration of T_S .

The proof is immediate from the previous sections, using conjugation by J defined in (2.1).

We now state an analogous system conjugate to T_A . The proof is similar.

Define

$$T_I(a, b, c, d) = \begin{cases} S_1(a, b, c, d)^t & \text{if } b + c + d < a \\ S_2(a, b, c, d)^t & \text{if } a + c + d < b \\ S_3(a, b, c, d)^t & \text{if } a + b + d < c \\ S_4(a, b, c, d)^t & \text{if } a + b + c < d \\ \text{if none of the above then} & . \\ S_1^\perp(a, b, c, d)^t & \text{if } a < 0 \\ S_2^\perp(a, b, c, d)^t & \text{if } b < 0 \\ S_3^\perp(a, b, c, d)^t & \text{if } c < 0 \\ S_4^\perp(a, b, c, d)^t & \text{if } d < 0 \end{cases}$$

Iteration of T_I produces a word $W_1 W_2 \cdots W_n$ defined by $T_I^k(a, b, c, d) = W_k T_I^{k-1}(a, b, c, d)$.

Theorem 4.3. *The word $W_1W_2\cdots W_n$ produced by iteration of T_I on a primitive integer Descartes quadruple (a, b, c, d) is in invert normal form and satisfies*

$$(a, b, c, d)^t = W_1W_2\cdots W_k\sigma(1, 1, 0, 0)^t$$

where σ is some permutation of the entries of the vector. In other words, any primitive integral Descartes quadruple (a, b, c, d) eventually reaches one of the so-called simplest Descartes quadruples

$$(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)$$

under iteration of T_I .

4.3. Dynamics on Apollonian circle packings. The Apollonian circle packist will be interested to consider how the dynamical systems interact with individual Apollonian circle packings. Each Apollonian circle packing has a root quadruple, i.e. the largest four pairwise tangent circles in the packing. The main result of this section is to show that the invert normal form word $W_1W_2\cdots W_n$ produced by the dynamical system T_I is such that the longest substring $W_1W_2\cdots W_k$ consisting only of swaps will end with the root quadruple of the packing containing the initial quadruple. In other words, the dynamical system T_I , or equivalently T_A , moves any quadruple to the root of its Apollonian circle packing, then inverts, then moves to the root, then inverts, etc.

Theorem 4.4. *The dynamical system T_I moves a quadruple to the root of its packing via swaps before inverting.*

In particular, while it remains in a single Apollonian circle packing, the dynamical system T_I agrees with the Reduction Algorithm for Descartes quadruples of [GLMWY1, Section 3], until the last step (when Graham, Lagarias, Mallows, Wilks and Yan reorder the quadruple).

The proof requires several lemmas. A Descartes quadruple (a, b, c, d) is a root quadruple of the Apollonian packing in which it resides if $a \leq 0 \leq b \leq c \leq d$, and $a + b + c \geq d$. Furthermore, it exists if the packing is integral, and is unique [GLMWY3, Section 3].

Lemma 4.5. *Let (x, y, z, w) be a Descartes quadruple such that the i -th coordinate is not maximal in the quadruple. Let $S_{ii} = S_i^\perp S_i$. Then $S_{ii}(x, y, z, w)^t$ is a root quadruple of the packing in which it resides, possibly after re-ordering the coordinates.*

Proof. We have

$$S_{11} = \begin{pmatrix} 1 & -2 & -2 & -2 \\ -2 & 5 & 4 & 4 \\ -2 & 4 & 5 & 4 \\ -2 & 4 & 4 & 5 \end{pmatrix} \quad S_{22} = \begin{pmatrix} 5 & -2 & 4 & 4 \\ -2 & 1 & -2 & -2 \\ 4 & -2 & 5 & 4 \\ 4 & -2 & 4 & 5 \end{pmatrix}$$

$$S_{33} = \begin{pmatrix} 5 & 4 & -2 & 4 \\ 4 & 5 & -2 & 4 \\ -2 & -2 & 1 & -2 \\ 4 & 4 & -2 & 4 \end{pmatrix} \quad S_{44} = \begin{pmatrix} 5 & 4 & 4 & -2 \\ 4 & 5 & 4 & -2 \\ 4 & 4 & 5 & -2 \\ -2 & -2 & -2 & 1 \end{pmatrix},$$

We prove the lemma for the case $i = 1$ and note that the other cases are identical. We have

$$S_{11}(x, y, z, w)^t = (x - 2y - 2z - 2w, -2x + 5y + 4z + 4w, -2x + 4y + 5z + 4w, -2x + 4y + 4z + 5w)^t.$$

Since x is not maximal among x, y, z, w , we have that the first coordinate of $S_1(x, y, z, w)^t$ is ≥ 0 (if x is negative, $-x + 2y + 2z + 2w$ is a sum of positive numbers and hence it is positive; if x is nonnegative, it is enough to argue that the first coordinate of $S_1(x, y, z, w)^t$ is at least x). Hence $x - 2y - 2z - 2w \leq 0 \leq -2x + 5y + 4z + 4w, -2x + 4y + 5z + 4w, -2x + 4y + 4z + 5w$. Without loss of generality, assume $y \leq z \leq w$, so that

$$x - 2y - 2z - 2w \leq 0 \leq -2x + 5y + 4z + 4w \leq -2x + 4y + 5z + 4w \leq -2x + 4y + 4z + 5w$$

In order to show that the above is a root quadruple, we need only to show that

$$(4.1) \quad x - 2y - 2z - 2w - 2x + 5y + 4z + 4w - 2x + 4y + 5z + 4w + 2x - 4y - 4z - 5w = -x + 3y + 3z + w \geq 0.$$

Using Descartes' theorem, and the fact that w is maximal, we have that

$$w = \frac{2x + 2y + 2z + \sqrt{16xy + 16xz + 16yz}}{2} = x + y + z + 2\sqrt{xy + xz + yz}.$$

So the expression in (4.1) can be rewritten as

$$4y + 4z + 2\sqrt{xy + xz + yz}.$$

If $0 \leq y \leq z$, then this is clearly nonnegative and we are done. Suppose $y < 0$. Then the circle corresponding to y in the Descartes quadruple (x, y, z, w) contains the one of curvature z , and so the circle of curvature z has smaller radius and hence larger curvature than the one of curvature y . Thus $4y + 4z > 0$ and the expression in (4.1) is nonnegative as desired. $\square$

Lemma 4.6. *Let (a, b, c, d) be a root quadruple of some packing. Then $S_i^\perp(a, b, c, d)^t$ is a root quadruple of another packing, after re-ordering, for any $2 \leq i \leq 4$.*

Proof. We consider the case where $i = 2$ and note that the other cases are identical. We have $S_2^t(a, b, c, d)^t = (2b + a, -b, 2b + c, 2b + d)^t$, and $-b \leq 0 \leq 2b + a \leq 2b + c \leq 2b + d$. We compute

$$-b + 2b + a + 2b + c - 2b - d = b + a + c - d \geq 0$$

since (a, b, c, d) is a root quadruple, and hence $(2b + a, -b, 2b + c, 2b + d)$ is also a root quadruple after reordering. $\square$

Proof of Theorem 4.4. The system T_I generates a word satisfying $(a, b, c, d)^t = W_1 W_2 \cdots W_n \mathbf{b}$ where $\mathbf{b}$ is a simplest Descartes quadruple. In this word, swaps are as far left as possible and inversions as far right as possible. Therefore if the leftmost inversion $S_i^\perp$ occurs at some position k , i.e. $W_k = S_i^\perp$, it is because it is followed by S_i , or because it occurs as the last letter ($k = n$), or else because it is followed by another inversion $S_j^\perp$.

We assume that the leftmost inversion is W_1 , and will show that $(a, b, c, d)^t$ is a root quadruple. The theorem will follow.

In the first case, $(a, b, c, d)^t = S_i^\perp S_i(x, y, z, w)^t$, where the i -th coordinate is not maximal (since it is a result of S_i in the application of T_S). Therefore, by the first lemma, (a, b, c, d) is a root quadruple.

In the second case, (a, b, c, d) is created by an inversion from a simplest Descartes quadruple, which is, in particular, a root quadruple. But it is not an inversion in a circle of curvature 0, since that would not change the Descartes quadruple. Therefore the second lemma applies.

In the third case, $(a, b, c, d)^t = S_i^\perp S_j^\perp(x, y, z, w)^t$, where $i \neq j$. Then, by induction (with the previous two cases as base cases and the second lemma as an inductive step), $(a, b, c, d)^t$ is a base quadruple. (Since $i \neq j$, we know the i -th circle is not the largest in $S_j^\perp(x, y, z, w)^t$.) $\square$

5. TYPICAL EXPANSIONS OF A POINT FROM TWO PERSPECTIVES

What can one say about the chain of swaps and inversions that are used in the rational approximation of a typical point under iteration? First, we consider the question as a limiting question on finite expansions. In the second section, we consider the question for all expansions. Finally, we provide some numerical data.

5.1. Digit probabilities in finite expansions. We consider the behavior of the expansions of rational points (those points with finite expansion). We use a measure of the *height* of a rational point. For example, in the case of Lorentz quadruples, we may define the set of points of height N to be

$$X_N := \{(a, b, c, d) \in \mathbb{Z}^4 : a^2 = b^2 + c^2 + d^2, 0 < a \leq N, \gcd(b, c, d) = 1\}.$$

We then consider X_N to be a discrete probability space with uniform probability measure. Under this measure, we can ask about the distribution of the n -th digit in the expansion, as $N \rightarrow \infty$. In this section, we prove the following theorem.

Write $\delta(X, Y, Z)$ for the letter produced by applying T_{Sph} to (X, Y, Z) . Let $d\mu = \frac{1}{4\pi}dA$ be the normalized uniform area measure on the sphere.

Theorem 5.1. *Under the uniform probability measure for X_N , the distribution of n -th digit in the expansion of a random Lorentz quadruple $(a, b, c, d) \in X_N$, converges to the distribution of δ under the measure $\mathcal{F}^{n-1}(\mathbf{I})d\mu$, where $\mathbf{I}$ is the constant function 1 and $\mathcal{F}$ is the transfer operator of T_{Sph} . In particular, the distributions of the 1st and 2nd digits converge to the distribution of δ under the measures*

$$d\mu, \quad \left(\sum \frac{1}{(2 \pm X \pm Y \pm Z)^4} \right) d\mu,$$

respectively. (The sum is over all choices of sign combinations.)

For example, the probabilities of possible first digits approach the proportional areas of the circular and triangular regions on the sphere in Figure 14. These are, respectively, the probability the first digit is an inversion,

$$\frac{1}{\pi} \int_{1/\sqrt{2}}^{\infty} \int_{-\infty}^{\infty} \frac{4dxdy}{(1+x^2+y^2)^2} = 2 \left(1 - \frac{1}{\sqrt{3}} \right) = 0.84529946 \dots$$

(the integrand is the pushforward of surface area on the sphere to the plane under stereographic projection); and the probability it is a swap,

$$\frac{1}{4\pi} \left(4\pi - 8\pi \left(1 - \frac{1}{\sqrt{3}} \right) \right) = \left(\frac{2}{\sqrt{3}} - 1 \right) = 0.15470053 \dots$$

Note that the same result applies, via application of the change of coordinates J , to Descartes quadruples chosen uniformly among those whose sum of curvatures is less than N . It is also worth remarking that this sort of result is limited in the sense that a slightly different way of measuring height, say the maximum of the curvatures is less than N , may potentially lead to different probabilities.

The transfer operator $\mathcal{F} : L^1(S^2, \mu) \rightarrow L^1(S^2, \mu)$ arising from the transformation T_{Sph} is defined in the following way: for $f \in L^1(S^2, \mu)$, we have

$$(\mathcal{F}f)(\tilde{X}) = \sum_{\tilde{Y} \in f^{-1}(\tilde{X})} g(\tilde{Y})f(\tilde{Y}),$$

where g is the inverse of the Jacobian of T_{sph} .

Proof of Theorem 5.1. By definition, the transfer operator is given by

$$\begin{aligned}
(\mathcal{F}f)(X, Y, Z) = & \frac{1}{(2+X+Y+Z)^4} f\left(\frac{-1-Y-Z}{2+X+Y+Z}, \frac{-1-X-Z}{2+X+Y+Z}, \frac{-1-X-Y}{2+X+Y+Z}\right) \\
& + \frac{1}{(2+X-Y-Z)^4} f\left(\frac{-1+Y+Z}{2+X-Y-Z}, \frac{1+X-Z}{2+X-Y-Z}, \frac{1+X-Y}{2+X-Y-Z}\right) \\
& + \frac{1}{(2-X+Y-Z)^4} f\left(\frac{1+Y-Z}{2-X+Y-Z}, \frac{-1+X+Z}{2-X+Y-Z}, \frac{1-X+Y}{2-X+Y-Z}\right) \\
& + \frac{1}{(2-X-Y+Z)^4} f\left(\frac{1-Y+Z}{2-X-Y+Z}, \frac{1-X+Z}{2-X-Y+Z}, \frac{-1+X+Y}{2-X-Y+Z}\right) \\
& + \frac{1}{(2-X-Y-Z)^4} f\left(\frac{1-Y-Z}{2-X-Y-Z}, \frac{1-X-Z}{2-X-Y-Z}, \frac{1-X-Y}{2-X-Y-Z}\right) \\
& + \frac{1}{(2-X+Y+Z)^4} f\left(\frac{1+Y+Z}{2-X+Y+Z}, \frac{-1+X-Z}{2-X+Y+Z}, \frac{-1+X-Y}{2-X-Y+Z}\right) \\
& + \frac{1}{(2+X-Y+Z)^4} f\left(\frac{-1+Y-Z}{2+X-Y+Z}, \frac{1+X+Z}{2+X-Y+Z}, \frac{-1-X+Y}{2+X-Y+Z}\right) \\
& + \frac{1}{(2+X+Y-Z)^4} f\left(\frac{-1-Y+Z}{2+X+Y-Z}, \frac{-1-X+Z}{2+X+Y-Z}, \frac{1+X+Y}{2+X+Y-Z}\right).
\end{aligned}$$

Therefore the second part of the theorem follows from the first.

The result follows if we show that for a given region R in S^2 ,

$$(5.1) \quad \lim_{N \rightarrow \infty} \frac{|\{(a, b, c, d) \in X_N : (b/a, c/a, d/a) \in R\}|}{|X_N|} = \mu(R).$$

This would show that the random vector $(b/a, c/a, d/a)$ converges in distribution to μ . Therefore, the transformation $T_{sph}(X, Y, Z)$ has distribution $(\mathcal{F}(\mathbf{I}))(X, Y, Z) d\mu$.

It suffices to consider regions R on the unit sphere given by

$$R = \{(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) : s_1 < \theta < s_2, t_1 < \phi < t_2\}.$$

We let S_{s_1, s_2, t_1, t_2} be the three dimensional region given by

$$S_{s_1, s_2, t_1, t_2} = \left\{ (x, y, z) : x^2 + y^2 + z^2 \leq 1, s_1 < \arccos \frac{z}{\sqrt{x^2 + y^2 + z^2}} < s_2, t_1 < \arctan \frac{y}{x} < t_2 \right\}.$$

Now

$$\begin{aligned}
& \frac{|\{(a, b, c, d) \in X_N : (b/a, c/a, d/a) \in R\}|}{|X_N|} \\
&= \frac{\#\{(a, b, c, d) \in \mathbb{Z}^4 : a > 0, \gcd(b, c, d) = 1, b^2 + c^2 + d^2 \leq N^2, (b/a, c/a, d/a) \in R\}}{\#\{(a, b, c, d) \in \mathbb{Z}^4 : a > 0, \gcd(b, c, d) = 1, b^2 + c^2 + d^2 \leq N^2\}} \\
&= \frac{\#\{(b, c, d) \in \mathbb{Z}^3 : \gcd(b, c, d) = 1, (b/N, c/N, d/N) \in S_{s_1, s_2, t_1, t_2}\}}{\#\{(b, c, d) \in \mathbb{Z}^3 : \gcd(b, c, d) = 1, (b/N, c/N, d/N) \in S_{0, 2\pi, 0, \pi}\}} \\
&= (1 + o(1)) \frac{N^3 \frac{1}{\zeta(3)} \text{ volume of } S_{s_1, s_2, t_1, t_2}}{N^3 \frac{1}{\zeta(3)} \text{ volume of } S^2} = (1 + o(1)) \frac{1/3(\cos s_1 - \cos s_2)(t_2 - t_1)}{4\pi/3} \\
&= (1 + o(1)) \frac{1}{4\pi} (\cos s_1 - \cos s_2)(t_2 - t_1) = (1 + o(1)) \frac{dA}{4\pi},
\end{aligned}$$

which proves (5.1). □

5.2. Expansion statistics assuming ergodicity. The results in this section depend on the conjectural ergodicity of the systems $(\mathbb{P}^1(\mathbb{C}), T_A, \mu_A)$ and $(\mathbb{P}^1(\mathbb{C}), T_B, \mu_B)$ constructed in the previous section. These systems are similar to the system constructed by Schmidt in [Sc1], which is ergodic (the main result of Schmidt's [Sc2]). Similarly, Romik constructs a measure-preserving system on the first quadrant of the unit circle in [R] which he proves to be ergodic. In both Romik's and Schmidt's work, ergodicity is then applied to provide information about the expansions of a typical point.

While we hope to prove ergodicity in a subsequent paper, we state it as a conjecture here.

Conjecture 5.2. *The systems $(\mathbb{P}^1(\mathbb{C}), T_A, \mu_A)$ and $(\mathbb{P}^1(\mathbb{C}), T_B, \mu_B)$ are ergodic.*

Then standard theorems of ergodic theory would imply that for almost all $z \in \mathbb{P}^1(\mathbb{C})$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(T_A^i z) = \int \phi \, d\mu$$

with ϕ an indicator function for various Farey circles and triangles (and similarly with T_B in place of T_A). In particular, by judicious choice of an indicator function, we can compute the frequencies of certain chains of swaps and inversions occurring in the typical expansion of a point.

5.2.1. Example: Two or three swaps in a row. The frequency of two swaps in a row $\{S_i S_j : i, j \in [1, 4], i \neq j\}$, as well as the frequency of two inversions in a row $\{S_i^\perp S_j^\perp : i, j \in [1, 4], i \neq j\}$ is $0.345299\dots$. The frequency of three swaps (or inversions) in a row, $\{S_i S_j S_k : i, j, k \in [1, 4], i \neq j, j \neq k\}$, is $0.246913\dots$.

With reference to Figure 6, there are twelve triangular regions corresponding to two swaps in a row. The total measure of these triangular regions, as a proportion of the total measure of the plane, is the frequency we wish to compute. As noted in Section 3.4, the total area with respect to μ_A or μ_B of all eight Farey circles and triangles (the entire plane) is $2\pi^2$.

The area of a region with respect to μ_A and μ_B is exactly hyperbolic area multiplied by a factor of $\pi/4$. The hyperbolic area of a Euclidean disk of radius r in the upper half-plane whose center has y -coordinate $b > 0$ is

$$(5.2) \quad I(\alpha) = \int_{-1}^1 \frac{2\sqrt{1-x^2}}{\alpha^2 - (1-x^2)} dx = 2\pi \left(\frac{\alpha}{\sqrt{\alpha^2 - 1}} - 1 \right) \text{ (cf. lemmas 2.2, 2.3 of [Sc2])},$$

where $\alpha = b/r > 1$. The hyperbolic area of a translate (in the y direction by α) of the ideal triangle with vertices $0, 1, \infty$ is

$$(5.3) \quad J(\alpha) = \int_0^1 \frac{dx}{\alpha + \sqrt{x(1-x)}} = \pi - \frac{2 \arccos(1/2\alpha)}{\sqrt{1 - (1/2\alpha)^2}}.$$

Note: this is $\Phi(1/2\alpha)$ where Φ is as in Lemma 3.3 from [Sc2].

Putting this together, we have that the frequency of two swaps in a row is

$$\frac{12 \cdot (\frac{\pi}{4} \cdot J(1))}{2\pi^2} = 0.345299\dots,$$

the frequency of three swaps in a row is

$$\frac{12}{8\pi} (J(1) - I(4)) = 0.246913\dots,$$

and so on.

We may use Lemmas 3.6 and 3.7 to translate integrals over Farey circles to integrals over Farey triangles, and vice versa. In particular, the frequency of n swaps will equal the frequency of n inversions.

string	exp 1	exp 2	conjecture	unif freq
2 swaps, $\{S_i S_j\}$	0.338...	0.328...	0.345299...	12/44=0.273...
2 inversions, $\{S_i^\perp S_j^\perp\}$	0.329...	0.347...	0.345299...	12/44=0.273...
3 swaps, $\{S_i S_j S_k\}$	0.235...	0.224...	0.246913...	36/224=0.1607...
3 inversions, $\{S_i^\perp S_j^\perp S_k^\perp\}$	0.220...	0.251...	0.246913...	36/224=0.1607...

TABLE 1. Experimental results for frequencies of random expansions. The column *exp 1* refers to the frequencies from bounded Lorentz quadruples; while *exp 2* refers to the frequencies from random point expansions. The column *conjecture* refers to the conjectured value for random point expansions (compare to *exp 2*). The column *unif freq* refers to the frequency expected if words were uniformly random.

digit	experiment	theorem
first swap	0.161...	0.154...
first inversion	0.838...	0.845...
second swap	0.365...	
second inversion	0.634...	

TABLE 2. Frequencies of digits or swaps in first or second digit of the expansions of bounded Lorentz quadruples.

5.2.2. *Example: Certain strings of Schmidt.* The frequency in the expansion of almost every $z \in \mathbb{C}$ of a string of alternating swaps or inversions of length n ,

$$S_{i/j} \dots S_j S_i S_j S_i, \quad S_{i/j}^\perp \dots S_j^\perp S_i^\perp S_j^\perp S_i^\perp \quad (i \neq j \text{ fixed}),$$

is $J(n-1)/8\pi$, where $J(\alpha)$ is as in (5.3). These are probabilities associated to a string of only swaps or only inversions having a common fixed point. Consider the probability that one of the vertices is fixed for *exactly* n iterations by a string of only swaps or only inversions, i.e. the frequency of strings

$$M S_{i/j} \dots S_j S_i S_j S_i N, \quad M S_{i/j}^\perp \dots S_j^\perp S_i^\perp S_j^\perp S_i^\perp N \quad (i \neq j \text{ fixed}),$$

where $M, N \neq S_i, S_j$ on the left and $M, N \neq S_i^\perp, S_j^\perp$ on the right. This frequency is (cf. [Sc2, Theorem 5.3A])

$$\frac{1}{8\pi} (J(n-1) - 2J(n) + J(n+1)).$$

In particular, the conjectured values for length 1, 2 and 3 Schmidt strings are

$$0.084117\dots, \quad 0.007180\dots, \quad 0.002249\dots$$

respectively.

5.3. Experimental data. We ran two experiments. In the first, we generated all Lorentz quadruples with $a < 200$ and computed the frequencies of certain substrings, as well as the frequencies of the first several digits. The former are shown in Table 1. The frequency of swaps and inversions in the first and second digits are shown in Table 2. These results are to be compared to Section 5.1.

In the second, we approximated the frequencies in a random expansion as follows: we selected 100 points uniformly at random from the $[0, 1] \times [0, 1]$ square and computed the first 100 letters of the expansion iterating T_A or T_B , and tabulated the frequency of certain substrings. These results are to be compared to Section 5.2.

6. RESTRICTION TO THE REAL LINE AND ITS ERGODIC INVARIANT MEASURE

6.1. The dynamical system and its ergodic invariant measure. In this section, we consider the action of T_A and T_B on the real line, thus obtaining a version of the Euclidean algorithm. On $\mathbb{P}^1(\mathbb{R})$, T_A and T_B agree, the map being

$$t(x) := \begin{cases} \mathbf{a}(x) = -x & x \in A = [-\infty, 0] \\ \mathbf{b}(x) = \frac{x}{2x-1} & x \in B = [0, 1] \\ \mathbf{c}(x) = 2-x & x \in C = [1, \infty] \end{cases}.$$

The Möbius transformations $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are reflections in the sides of the ideal hyperbolic triangle with vertices $\{0, 1, \infty\}$, generating a group $\Gamma_{\mathbb{R}}$ isomorphic to the free product $\mathbb{Z}/(2) * \mathbb{Z}/(2) * \mathbb{Z}/(2)$. The fixed points of t are $\{0, 1, \infty\}$ and iteration of t takes rationals to one of the fixed points in finite time, depending on the parity of the numerator and denominator, preserving parity as $\Gamma_{\mathbb{R}}$ is the kernel of the map $\mathrm{PGL}_2(\mathbb{Z}) \rightarrow \mathrm{SL}_2(\mathbb{Z}/(2))$ (index 6). See the next subsection for a proof that orbits are finite on rationals. Iteration of t with input x produces a word $\mathbf{r} = \mathbf{m}_1 \dots$ in $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$, $\mathbf{m}_i(t^{i-1}(x)) = t(t^{i-1}(x))$, such that $\mathbf{m}_i \neq \mathbf{m}_{i+1}$. We take the word to be finite if x is rational.

We obtain rational approximations to x by following the inverse orbit of $\{0, 1, \infty\}$:

$$\frac{p_{n,\alpha}}{q_{n,\alpha}} = \left(\prod_{i=1}^n \mathbf{r}_i \right) (\alpha), \quad \alpha \in \{0, 1, \infty\},$$

which include the usual convergents from simple continued fractions. The convergents can be constructed starting from $(1/1, \pm 1/0, 0/1)$, according as x is positive or negative, and taking mediants, moving through the Farey tree. Applying $\mathbf{a}, \mathbf{b}$, or $\mathbf{c}$ updates the first, second, or third position by taking the mediant $\frac{p}{q} \oplus \frac{r}{s} := \frac{p+q}{r+s}$ of the other two entries.

For example for a random number

$$x = 0.4189513796210592\dots, \quad \mathbf{r} = \mathbf{bacabcbcacacababac}\dots,$$

the first 20 convergents are (mediants in red, see Figure 15)

$$\begin{aligned} & (1/1, 1/0, 0/1), (1/1, \mathbf{1/2}, 0/1), (\mathbf{1/3}, 1/2, 0/1), (1/3, 1/2, \mathbf{2/5}), (\mathbf{3/7}, 1/2, 2/5), (3/7, \mathbf{5/12}, 2/5), \\ & (3/7, 5/12, \mathbf{8/19}), (\mathbf{13/31}, 5/12, 8/19), (13/31, 5/12, \mathbf{18/43}), (13/31, \mathbf{31/74}, 18/43), \\ & (13/31, 31/74, \mathbf{44/105}), (\mathbf{75/179}, 31/74, 44/105), (75/179, 31/74, \mathbf{106/253}), \\ & (\mathbf{137/327}, 31/74, 106/253), (137/327, 31/74, \mathbf{168/401}), (\mathbf{199/475}, 31/74, 168/401), \\ & (199/475, \mathbf{367/876}, 168/401), (\mathbf{535/1277}, 367/876, 168/401), (535/1277, \mathbf{703/1678}, 168/401), \\ & (\mathbf{871/2079}, 703/1678, 168/401), (871/2079, 703/1678, \mathbf{1574/3757}) \\ & = (0.4189514\dots, 0.4189511\dots, 0.4189512\dots). \end{aligned}$$

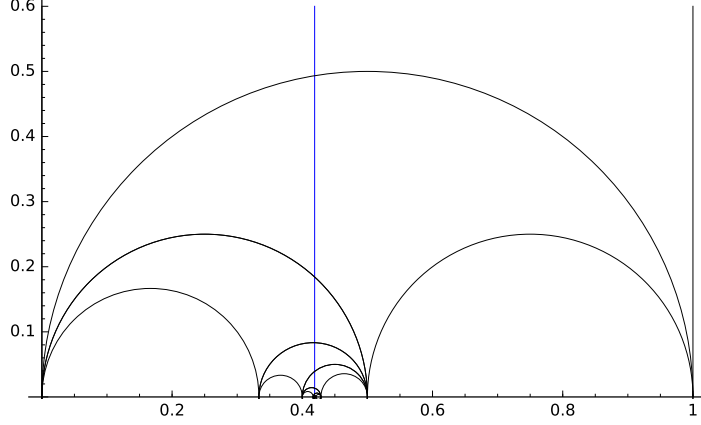
The map $T(y, x) = (\mathbf{m}(y), \mathbf{m}(x))$, $\mathbf{m}(x) = t(x)$, extending t in second coordinate, is a bijection on the space of geodesics

$$\mathcal{G}_{\mathbb{R}} = A \times B \cup A \times C \cup B \times C \cup B \times A \cup C \times A \cup C \times B \setminus \text{diag}.$$

where there is an isometry invariant measure $dx \, dy \, |x - y|^{-2}$. Pushing forward to the second coordinate gives the *infinite* t -invariant measure

$$f(x) \, dx = d\mu(x) = \begin{cases} \frac{dx}{-x} & x < 0, \\ \frac{dx}{x(1-x)} & 0 < x < 1, \\ \frac{dx}{x-1} & x > 1, \end{cases}.$$

This dynamical system is clearly a cross-section of billiards in the ideal hyperbolic triangle: the bi-infinite word $\mathbf{r}^{-1}\mathbf{r}$ in $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ corresponding to a geodesic (y, x) records the sequence of collisions

FIGURE 15. Approximating $x = 0.4189513796210592\dots$, $\mathfrak{x} = \text{bacabcacbcacacababac}\dots$.

with the walls. The return time, say for a geodesic $(y, x) \in [-\infty, 0] \times [1, \infty]$, is given by

$$r(y, x) = \frac{1}{2} \log \left(\frac{x(1-y)}{y(1-x)} \right).$$

The return time is integrable with respect to $(x-y)^{-2} dy dx$, for instance

$$\frac{1}{2} \int_{-\infty}^0 \int_1^\infty \log \left(\frac{x(1-y)}{y(1-x)} \right) \frac{dx dy}{(y-x)^2} = \frac{\pi^2}{6}.$$

Since this triangle reflection group has finite covolume, the system $(\mathcal{G}_{\mathbb{R}}, T, (x-y)^{-2} dy dx)$ is ergodic, implying ergodicity of $(P^1(\mathbb{R}), t, \mu)$.

6.2. A Euclidean algorithm. By homogenizing t above, we get a dynamical system on pairs of integers $(p, q) \in \mathbb{Z}^2$ which halts when $p = q$ or one of p, q is zero.

$$(p, q) \mapsto \begin{cases} a(p, q) = (-p, q) & q < 0 < p \text{ or } p < 0 < q, \text{ i.e. } p/q < 0, \\ b(p, q) = (p, 2p - q) & 0 < p < q \text{ or } q < p < 0, \text{ i.e. } 0 < p/q < 1, \\ c(p, q) = (2q - p, q) & 0 < q < p \text{ or } p < q < 0, \text{ i.e. } p/q > 1. \end{cases}$$

The b step reduces $|q|$, the c step reduces $|p|$, and after applying a , one of either b or c follows. Hence the algorithm terminates. If $(p, q) \neq (0, 0)$ then the non-zero entry of the output is $\pm \gcd(p, q)$, and working backwards allows us to write the gcd as a linear combination of p and q (using only the coefficients -1 and 2 at each step).

6.3. Dynamics on triples. In this section, we make some remarks relating the triangle reflection group to a “dual Apollonian group” on the line and conjugate this group to act on Pythagorean triples (obtaining a system on the circle conjugate to $(P^1(\mathbb{R}), t, \mu)$).

Here is a lower dimensional version of Descartes’ theorem on curvatures. Consider three real numbers, $a < b < c$ (one of which may be ∞ , ordered as on a circle). The inverses of the lengths of the intervals they define (infinite length if one endpoints is ∞ , negative length if the interval contains ∞) satisfy $(x + y + z)^2 - (x^2 + y^2 + z^2) = 0$. In analogy with the dual Apollonian group $\langle \mathfrak{s}_i^\perp : 1 \leq i \leq 4 \rangle$, we can define a group acting on ordered triples of mutually tangent oriented intervals as “inversions” in the equivalent of ACC coordinates using indefinite binary quadratic forms. If

$$F = \begin{pmatrix} a & -b \\ -b & c \end{pmatrix}, \quad a, b, c \in \mathbb{R}, \quad \det(F) = -1,$$

then the zero set of F , $(x - b/a)^2 = 1/a^2$, defines an oriented interval with curvature a , co-curvature c , and curvature-center b (if $a = 0$, then b is ± 1 depending on orientation). Three forms F_i (considered as geodesics of the upper half-plane model of H^2) are in “triangular” configuration if they define an ideal hyperbolic triangle with proper orientation. The analogous group has generators

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & 2 & 1 \end{pmatrix} C = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix}$$

each defining an inversion in the given interval. The group generated by A, B, C preserves the quadratic form $xy + xz + yz$, i.e.

$$M^t Q M = Q, \quad Q = \frac{1}{2} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad M \in \langle A, B, C \rangle.$$

Three intervals/forms/geodesics R are in triangular configuration if they satisfy

$$R^t Q R = \begin{pmatrix} 0 & -2 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The group $\Gamma_{\mathbb{R}}$ is then a geometric realization of this group with respect to the base triple representing the ideal hyperbolic triangle with vertices $0, 1, \infty$ (rows in (c, a, b) coordinates)

$$\begin{pmatrix} 0 & 0 & -1 \\ 0 & 2 & 1 \\ -2 & 0 & 1 \end{pmatrix}.$$

The form Q is rationally equivalent to the “Pythagorean” diagonal form. For instance

$$P = J^t Q J, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}.$$

Conjugating A, B, C by J gives

$$A^J = \begin{pmatrix} 3 & 2 & 2 \\ -2 & -1 & -2 \\ -2 & -2 & -1 \end{pmatrix}, \quad B^J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad C^J = \begin{pmatrix} 3 & 2 & -2 \\ -2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}.$$

We can define a dynamical system on triples of integers (a, b, c) such that $a^2 = b^2 + c^2$, $a > 0$ which reduces the value of a and terminates at one of $(g, -g, 0)$, $(g, 0, \pm g)$, where g is the GCD of a, b , and c :

$$(a, b, c) \mapsto \begin{cases} (3a + 2a + 2c, -2a - b - 2c, -2a - 2b - c) & b < 0, c < 0, \\ (a, -b, c) & b > 0, \\ (3a + 2b - 2c, -2a - b + 2c, 2a + 2b - c) & b < 0, c > 0. \end{cases}$$

Dividing by a^2 , we obtain a sytem on the circle $x^2 + y^2 = 1$, conjugate to t

$$(x, y) \mapsto \begin{cases} \left(\frac{-2-x-2y}{3+2x+2y}, \frac{-2-2x-y}{3+2x+2y} \right) & x < 0, y < 0, \\ (-x, y) & x > 0, \\ \left(\frac{-2-x+2y}{3+2x-2y}, \frac{2+2x-y}{3+2x-2y} \right) & x < 0, y > 0. \end{cases}$$

6.4. Comparing to Romik's Work. Romik's action on Pythagorean triples also gives a dynamical system on the real line with an ergodic invariant measure and another variation on the Euclidean algorithm [R]. For example, Romik's Euclidean algorithm in section 3.2 of [R] is a dynamical system on nonnegative pairs of integers (p, q) with $p > q$ where

$$(p, q) \mapsto \begin{cases} a(p, q) = (p - 2q, q) & p - 2q > q, \\ b(p, q) = (q, p - 2q) & q \geq p - 2q > 0, \\ c(p, q) = (q, 2q - p) & p - 2q \leq 0. \end{cases}$$

For the sake of comparison, here is the greatest common divisor of 246 and 113 in the system described in section 6.2:

$$\begin{aligned} (246, 113) &\rightarrow (-20, 113) \rightarrow (20, 113) \rightarrow (20, -73) \rightarrow (-20, -73) \rightarrow (-20, 33) \rightarrow (20, 33) \\ &\rightarrow (20, 7) \rightarrow (-6, 7) \rightarrow (6, 7) \rightarrow (6, 5) \rightarrow (4, 5) \rightarrow (4, 3) \rightarrow (2, 3) \rightarrow (2, 1) \rightarrow (0, 1), \end{aligned}$$

while Romik's algorithm runs more quickly as follows:

$$\begin{aligned} (246, 113) &\rightarrow (113, 20) \rightarrow (73, 20) \rightarrow (33, 20) \rightarrow (20, 7) \rightarrow (7, 6) \rightarrow (6, 5) \\ &\rightarrow (5, 4) \rightarrow (4, 3) \rightarrow (3, 2) \rightarrow (2, 1) \rightarrow (1, 0). \end{aligned}$$

7. LORENTZ QUADRUPLES BY SIZE: ANOTHER DYNAMICAL SYSTEM

In contrast to the approach taken so far, it may be desirable to prioritize the feature that the dynamical system moves to the root by travelling to the arithmetically simplest among the adjacent Lorentz quadruples. This leads to a different dynamical system.

We say a Lorentz quadruple (a, b, c, d) is normalized if $a \geq b \geq c \geq d \geq 0$. We can normalize a quadruple by changing signs of its entries and reordering entries. Order the normalized Lorentz quadruples lexicographically, so that the first, or least, is $(1, 1, 0, 0)$. We will refer to the position of the quadruples in this ordering as their *height*; quadruples with the same normalization have the same height. The height of a possibly non-normalized Lorentz quadruple is the height of its normalization.

The purpose of this section is to construct a dynamical system on quadruples which travels to the root by passing at each step to the adjacent quadruple of least height. The paths to the origin according to this system form a tree. Writing quadruples in normalized form, one obtains a tree organizing quadruples by height which is shown in Figure 16.

Let $\mathcal{L}_+ \subseteq \mathcal{L}$ represent the Lorentz quadruples (a, b, c, d) that satisfy $a, b, c, d \geq 0$. We define a dynamical system on $\mathcal{L}_+$. Write $A = 2a - b - c - d$, $B = a - c - d$, $C = a - b - d$ and $D = a - b - c$. Define

$$T_D(a, b, c, d) = (|A|, |B|, |C|, |D|) = (A, |B|, |C|, |D|).$$

Observe that $A > 0$, i.e. $2a > b + c + d$ for any Lorentz quadruple.

Define the seven matrices:

$$\begin{aligned} D_1 &:= \begin{pmatrix} 2 & 1 & -1 & -1 \\ -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & 1 & 0 \end{pmatrix}, & D_2 &:= \begin{pmatrix} 2 & -1 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & -1 & 0 \end{pmatrix}, & D_3 &:= \begin{pmatrix} 2 & -1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \\ & & D_4 &:= \begin{pmatrix} 2 & -1 & -1 & -1 \\ -1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \end{aligned}$$

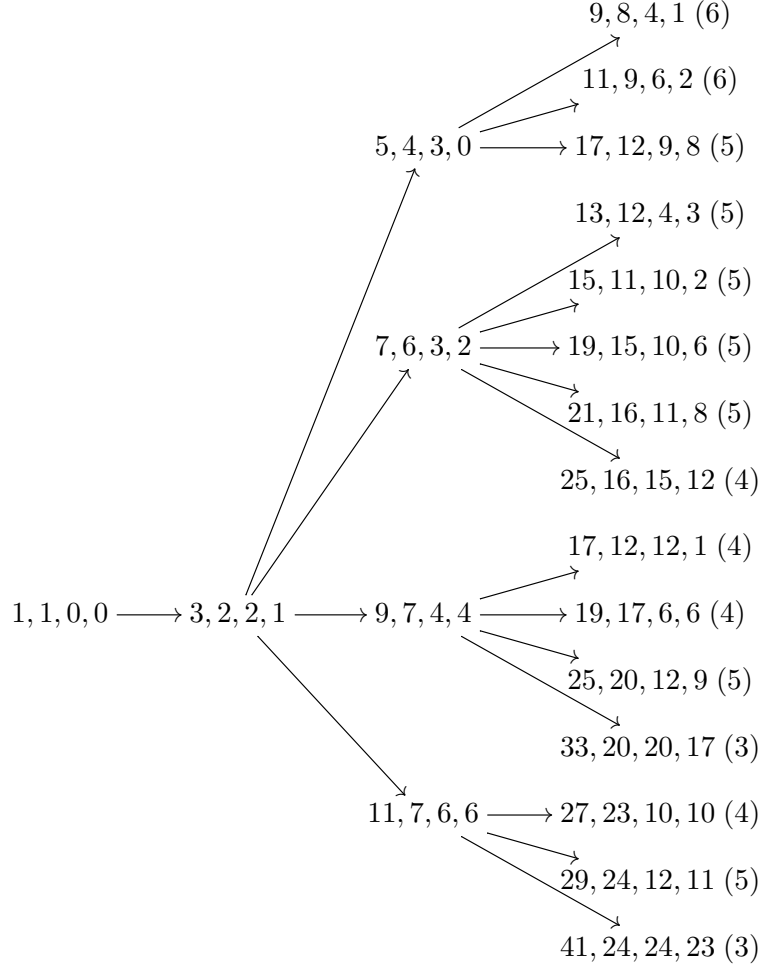


FIGURE 16. Quadruples are considered up to sign and ordering; we normalize them as all non-negative, with $a \geq b \geq c \geq d$.

$$D_5 := \begin{pmatrix} 2 & -1 & 1 & 1 \\ -1 & 0 & -1 & -1 \\ -1 & 1 & 0 & -1 \\ -1 & 1 & -1 & 0 \end{pmatrix} \quad D_6 := \begin{pmatrix} 2 & 1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & 1 & 0 \end{pmatrix} \quad D_7 := \begin{pmatrix} 2 & 1 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

The dynamical system generates a word $M_1 M_2 \dots$ from left-to-right in the D_i , since if $T_D(a, b, c, d) = (|A|, |B|, |C|, |D|)$, then $D_i(A, B, C, D)^t = (a, b, c, d)^t$ for exactly one i (it is not possible that $a > c + d, b > d, b > c$). In other words exactly one of the D_i undoes T_D . Explicitly, we let D_i be

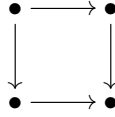
$$\begin{cases} D_1 & D, C < 0, B > 0 \\ D_2 & D, B < 0, C > 0 \\ D_3 & C, B < 0, D > 0 \\ D_4 & B, C, D < 0 \\ D_5 & D, C > 0, B < 0 \\ D_6 & D, B > 0, C < 0 \\ D_7 & B, C > 0, D < 0 \end{cases}$$

We give $\mathcal{L}_+$ the structure of a graph as follows: let two quadruples $\mathbf{b}_1$ and $\mathbf{b}_2$ be joined by an edge whenever $\mathbf{b}_2$ is obtained by $L_i \mathbf{b}_1$, followed by taking absolute values.

Theorem 7.1. *For any $(a, b, c, d) \in \mathcal{L}_+$, $T_D^{(n)}(a, b, c, d) = \mathbf{b}$ for some integer n , and some $\mathbf{b} \in \{(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1)\}$. Then the word $M = M_1 \cdots M_n$ generated by T_D satisfies $(a, b, c, d) = M\mathbf{b}$. Furthermore, the path $(T_D^{(k)}(a, b, c, d))_{k=0}^n$ to $\mathbf{b}$ on $\mathcal{L}_+$ is of minimal length in the graph $\mathcal{L}_+$. Finally, the path is the same as the path obtained by always travelling to the adjacent quadruple of smallest height.*

Lemma 7.2. *Label $\mathcal{L}^+$ according to direction of decrease of height. Then*

- (1) *All edges are directed.*
- (2) *the origin vertices $\mathbf{b} \in \{(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1)\}$ have no outward directed edges,*
- (3) *aside from the origins, every vertex has at least one outward directed edge,*
- (4) *the only minimal cycles in the graph are squares, and*
- (5) *any square is isomorphic to*



Proof. Of the adjacent quadruples, the smallest is that with first entry $2a - b - c - d$. We have $2a - b - c - d \geq a$ if and only if $a \geq b + c + d$. However, since $a^2 = b^2 + c^2 + d^2$, and we are assuming $b, c, d \geq 0$, this would entail $b^2 + c^2 + d^2 \geq (b + c + d)^2$, which can only occur if at least two of b, c, d are zero. Hence, in the primitive case, this only occurs when the vertex is an origin. Therefore the origins have no outward directed edges, but every other vertex does have an outward directed edge.

An edge is undirected if and only if the quadruples are the same up to reordering the non-negative quantities b, c, d . The adjacent quadruples have largest entry $2a \pm b \pm c \pm d$. Therefore an undirected edge can occur only if $a = \pm b \pm c \pm d$ for some choice of signs. But since $a \geq b, c, d \geq 0$, this is only possible with at most one negative sign. If $a = b + c + d$, we are in the case of the root, as above. Otherwise, if $a = b + c - d$, an undirected edge implies that $a \geq b = c \geq d$, and the two quadruples are the same, including ordering, hence the same vertex. This shows that the graph has no undirected edges.

Observe that $\mathcal{L}_+$ is by definition a union of images of the Cayley graph $\mathcal{C}_L$ under graph homomorphism. In particular, the presentation of the Super-Apollonian group implies that the only minimal cycles in $\mathcal{C}_L$ are squares. So the minimal cycles in $\mathcal{L}_+$ consist of images of squares.

Up to permuting the second through fourth coordinates, or changing their signs, the homomorphic images of a square in $\mathcal{C}_L$ is always labelled with first coordinates as follows, where edges are labelled with differences in the direction shown:

$$\begin{array}{ccc}
 2a - b + c + d & \xrightarrow{a-b+c-d} & 3a - 2b + 2c \\
 \uparrow a-b+c+d & & \uparrow a-b+c+d \\
 a & \xrightarrow{a-b+c-d} & 2a - b + c - d
 \end{array}$$

Since in a square of $\mathcal{C}_L$ the edge labels are non-zero, the direction of the edges is determined by these values, and it is evident that parallel edges must be directed in the same way, from which the assertion about directions on squares follows.

Finally, the non-existence of triangles in $\mathcal{L}_+$ follows from the diagram above, since such a triangle must be a homomorphic image of a square, and the edge labellings demonstrate that if one side of the square collapses, the opposite side also collapses. $\square$

Proof of Theorem 7.1. By items (2) and (3) of Lemma 7.2, and the well-ordering principle for heights, we know that from any quadruple there is a path to the origin which respects the direction

of the graph. The path of greatest decrease is unique since there is a unique adjacent quadruple of least height. This last fact arises as follows: suppose $2a - b - c - d = 2a \pm b \pm c \pm d$ for some other choice of signs. Then one of b, c, d is zero. If two entries are zero, we are at the root. If one entry is zero, comparing matrices L_i shows that the two quadruples are equal, so there is just one quadruple of least height.

Therefore the graph formed of fastest-dropping paths consists of three trees, which together span the graph.

It remains to show minimality of the path lengths. Compare the path of greatest descent to the image of the swap normal form path. Both descend the origin, and both respect the direction of the graph (as observed in the proof of Theorem 4.1). The moves that take one path to the other consist of square crossings (by Lemma 7.2 (4)). Let the weight of a path be given by the sum of 1 for each edge traversed respecting the graph direction, and -1 for each edge traversed against the graph direction. Then square crossings preserve weight, by Lemma 7.2 (5), so both paths have equal weight. Since both paths respect graph direction, this implies they are of the same length. $\square$

APPENDIX: SUMMARY OF A. L. SCHMIDT'S CONTINUED FRACTIONS

Here we give a quick summary of Asmus Schmidt's continued fraction algorithm [Sc1], its ergodic theory [Sc2], and further results of Hitoshi Nakada concerning these [N1], [N2], [N3]. We also explicitly relate Schmidt's system to ours.

Define the following matrices in $\text{PGL}(2, \mathbb{Z}[i])$:

$$V_1 = \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}, V_2 = \begin{pmatrix} 1 & 0 \\ -i & 1 \end{pmatrix}, V_3 = \begin{pmatrix} 1-i & i \\ -i & 1+i \end{pmatrix},$$

$$E_1 = \begin{pmatrix} 1 & 0 \\ 1-i & i \end{pmatrix}, E_2 = \begin{pmatrix} 1 & -1+i \\ 0 & i \end{pmatrix}, E_3 = \begin{pmatrix} i & 0 \\ 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & -1+i \\ 1-i & i \end{pmatrix}.$$

Note that

$$S^{-1}V_iS = V_{i+1}, S^{-1}E_iS = V_{i+1}, S^{-1}CS = C \text{ (indices modulo 3)}$$

where S is the order three elliptic element $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$, and that

$$\mathfrak{c} \circ m \circ \mathfrak{c} = m^{-1}$$

for the Möbius transformations m induced by $\{V_i, E_i, C\}$ (here $\mathfrak{c}$ is complex conjugation).

In [Sc1] Schmidt uses infinite words in these letters to represent complex numbers as infinite products $z = \prod_n T_n, T_n \in \{V_i, E_i, C\}$ in two different ways. Let $M_N = \prod_{n=1}^N T_n$. We have *regular chains*

$$\det M_N = \pm 1 \Rightarrow T_{n+1} \in \{V_i, E_i, C\}, \det M_N = \pm i \Rightarrow T_{n+1} \in \{V_i, C\},$$

representing z in the upper half-plane $\mathcal{I}$ (the model circle) and *dually regular chains*

$$\det M_N = \pm i \Rightarrow T_{n+1} \in \{V_i, E_i, C\}, \det M_N = \pm 1 \Rightarrow T_{n+1} \in \{V_i, C\},$$

representing $z \in \{0 \leq x \leq 1, y \geq 0, |z - 1/2| \geq 1/2\} =: \mathcal{I}^*$ (the model triangle). The model circle is a disjoint union of four triangles and three circles, and the model triangle is a disjoint union of three triangles and one circle (pictured in Figure 17):

$$\mathcal{I} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{C}$$

$$\mathcal{I}^* = \mathcal{V}_1^* \cup \mathcal{V}_2^* \cup \mathcal{V}_3^* \cup \mathcal{C}^*$$

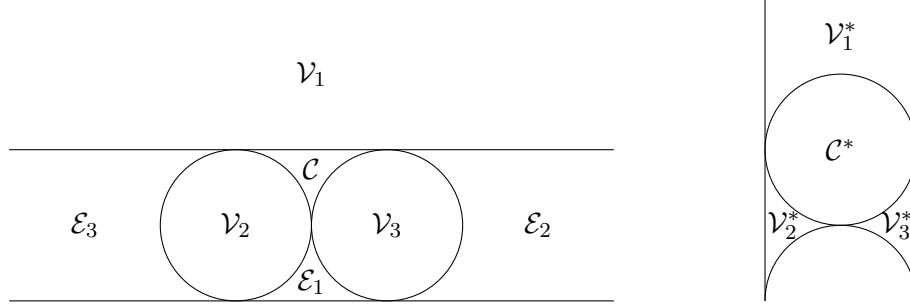


FIGURE 17. Model circle and model triangle of Schmidt.

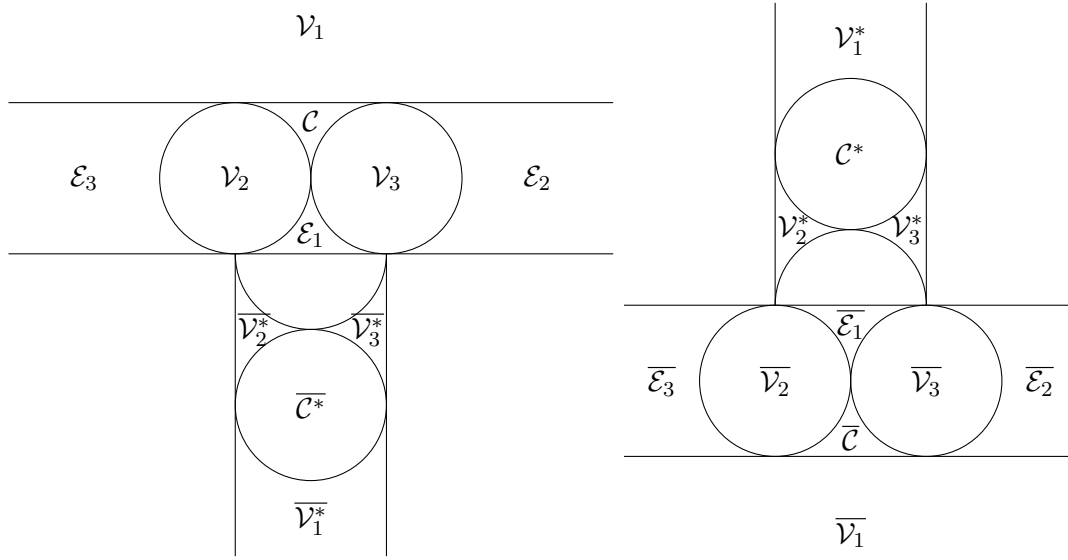


FIGURE 18. Regions for invertible extension of Nakada.

where

$$\mathcal{V}_i = v_i(\mathcal{I}), \mathcal{E}_i = e_i(\mathcal{I}^*), \mathcal{C} = c(\mathcal{I}^*), \mathcal{V}_i^* = v_i(\mathcal{I}^*), \mathcal{C}^* = c(\mathcal{I}),$$

(lowercase letters indicating the Möbius transformation associated to the corresponding matrix).

By considering $z = \prod_n T_n$ we obtain rational approximations $p_i^{(N)}/q_i^{(N)}$ to z by

$$M_N \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} p_1^{(N)} & p_2^{(N)} & p_3^{(N)} \\ q_1^{(N)} & q_2^{(N)} & q_3^{(N)} \end{pmatrix}$$

(which are the orbits of $\infty, 0, 1$ under the partial products $m_N = t_1 \circ \dots \circ t_N$). In [Sc1, Theorem 2.5] a quality of approximation is given: if $|z - p/q| < \frac{1}{(1+1/\sqrt{2})|q|^2}$, then p/q is a convergent to z .

The shift map T on $X = \mathcal{I} \cup \mathcal{I}^* = \{\text{chains, dual chains}\}$ maps X to itself via Möbius transformations, specifically (mapping $\mathcal{V}_i, \mathcal{C}^*$ onto $\mathcal{I}$ and $\mathcal{V}_i^*, \mathcal{E}_i, \mathcal{C}$ onto $\mathcal{I}^*$)

$$T(z) = \begin{cases} v_i^{-1}z & z \in \mathcal{V}_i \cup \mathcal{V}_i^* \\ e_i^{-1}z & z \in \mathcal{E}_i \\ c^{-1}z & z \in \mathcal{C} \cup \mathcal{C}^* \end{cases}.$$

The shift $T : X \rightarrow X$ is shown to be ergodic ([Sc2, Theorem 5.1]) with respect to the following probability measure

$$\tilde{f}(z) = \begin{cases} \frac{1}{2\pi^2}(h(z) + h(sz) + h(s^2z)) & z = x + yi \in \mathcal{I} \\ \frac{1}{2\pi} \frac{1}{y^2} & z = x + yi \in \mathcal{I}^* \end{cases}$$

where

$$h(z) = \frac{1}{xy} - \frac{1}{x^2} \arctan\left(\frac{x}{y}\right).$$

By inducing to $X \setminus \cup_i (\mathcal{V}_i \cup \mathcal{V}_i^*)$ Schmidt gives “faster” convergents $\hat{p}_\alpha^{(n)}/\hat{q}_\alpha^{(n)}$ and a sequence of exponents e_n (1 for E_i , C , and the return time k for V_i^k). He then gives results analogous to those of simple continued fractions via the pointwise ergodic theorem, including the arithmetic and geometric mean of the exponents which exist for almost every z ([Sc2, Theorem 5.3]):

$$\lim_{n \rightarrow \infty} \left(\prod_{i=1}^n e_i \right)^{1/n} = 1.26 \dots, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n e_i = 1.6667 \dots$$

In [N1], Nakada constructs an invertible extension of T on a space of geodesics in two copies of three-dimensional hyperbolic space. In one copy we take geodesics from $\overline{\mathcal{I}}^*$ to $\mathcal{I}$ and in the other the geodesics from $\overline{\mathcal{I}}$ to $\mathcal{I}^*$ where the overline indicates complex conjugation. The regions are pictured in Figure 18. The extension acts as Schmidt’s T depending on the second coordinate. Nakada doesn’t provide a second proof of ergodicity, but quotes Schmidt’s result. Also in [N1], results about the density of Gaussian rationals p/q that appear as convergents and satisfy $|z - p/q| < c/|q|^2$ are obtained. For instance according to [N1, Theorem 7.3], for almost every $z \in X$ and $0 < c < \frac{1}{1+\sqrt{2}}$, it holds that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{p/q \in \mathbb{Q}(i) : p/q = p_i^{(n)}/q_i^{(n)}, 1 \leq n \leq N, i = 1, 2, 3, |z - p/q| < c/|q|^2\} = \frac{c^2}{\pi}.$$

In [N2, Main Theorem] Nakada describes the rate of convergence of Schmidt’s convergents. Namely for almost every z ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |q_i^{(n)}| = \frac{E}{\pi}, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \left| z - \frac{p_i^{(n)}}{q_i^{(n)}} \right| = -\frac{2E}{\pi}, \quad E = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2}.$$

For reference, the relationship between the Super-Apollonian Möbius generators $\{\mathfrak{s}_i, \mathfrak{s}_i^\perp\}$ and Schmidt’s $\{v_i, e_i, c\}$ are

$$\begin{aligned} \mathfrak{s}_1 &= c^2 \circ \mathfrak{c}, \quad \mathfrak{s}_2 = e_1^2 \circ \mathfrak{c}, \quad \mathfrak{s}_3 = e_2^2 \circ \mathfrak{c}, \quad \mathfrak{s}_4 = e_3^2 \circ \mathfrak{c}, \\ \mathfrak{s}_1^\perp &= 1 \circ \mathfrak{c}, \quad \mathfrak{s}_2^\perp = v_1^2 \circ \mathfrak{c}, \quad \mathfrak{s}_3^\perp = v_2^2 \circ \mathfrak{c}, \quad \mathfrak{s}_4^\perp = v_3^2 \circ \mathfrak{c}. \end{aligned}$$

REFERENCES

- [P] H.L. Price, *The Pythagorean Tree: A New Species* (2008), <https://arxiv.org/abs/0809.4324>.
- [Be] B. Berggren, *Pytagoreiska triangelar*, Tidskrift for elementar matematik, fysik och kemi **17** (1934), 129–139.
- [Ba] 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. **ZW-011** (1963), 37 pp.
- [Bi] P. Billingsley, *Ergodic Theory and Information*, J. Wiley, 1965.
- [EW] M. Einsiedler and T. Ward, *Ergodic Theory with a view towards Number Theory*, Graduate Texts in Mathematics **259** (Springer-Verlag London Limited 2011).
- [F] L.R. Ford, *On the closeness of approach of complex rational fractions to a complex irrational number*, Trans. Amer. Math. Soc. **27** (1925), 146–154.
- [Fu] E. Fuchs, *Counting problems in Apollonian packings*, Bull. Amer. Math. Soc., **50** (2013), 229–266.
- [GLMWY1] R.L. Graham, J.C. Lagarias, C.L. Mallows, A.R. Wilks, C.H. Yan, *Apollonian circle packings: number theory*, Journal of Number Theory **100** (2003), 1–45.
- [GLMWY2] R.L. Graham, J.C. Lagarias, C.L. Mallows, A.R. Wilks, C.H. Yan, *Apollonian Circle Packings: Geometry and Group Theory I. The Apollonian Group*, Discrete Comput. Geom. **34** (2005), no. 4, 547–585.

- [GLMWY3] R.L. Graham, J.C. Lagarias, C.L. Mallows, A.R. Wilks, C.H. Yan, *Apollonian Circle Packings: Geometry and Group Theory II. Super-Apollonian Group and Integral Packings*, Discrete Comput. Geom. **35** (2006), no. 1, 1–36.
- [H1] A. Hurwitz, *Über die Entwicklung complexer Grössen in Kettenbrüche*(German), Acta. Math. **11** (1887), 187–200.
- [LMW] J.C. Lagarias, C.L. Mallows, A.R. Wilks, *Beyond the Descartes circle theorem*, Amer. Math. Monthly **109** (2002), no. 4, 338361.
- [H] A. Hall, *Genealogy of Pythagorean triads*, Math. Gaz. **54** (1970), 377–379.
- [Ka] A.R. Kanga, *The family tree of Pythagorean triples*, Bull. Inst. Math. Appl. **26** (1990), no. 1-2, 1517.
- [Ke] M. Keane, *A continued fraction titbit.*, Symposium in Honor of Benoit Mandelbrot (Curaao, 1995), Fractals 3 (1995), no. 4, 641650.
- [Kh] A. Ya. Khinchin, *Continued Fractions*, 1997, Dover.
- [Kon] A. Kontorovich, *From Apollonius to Zaremba: Local-global phenomena in thin orbits*, Bull. Amer. Math. Soc. **50** (2013), 187–228.
- [Koc] J. Kocik, *A theorem on circle configurations* (2007), <https://arxiv.org/abs/0706.0372>.
- [N1] H. Nakada, *On Ergodic Theory of A. Schmidt's Complex Continued Fractions over Gaussian Field*, Mh. Math. **105** (1988), 131–150.
- [N2] H. Nakada, *The metrical theory of complex continued fractions*, Acta Arith. **56** (1990), no. 4, 279–289.
- [N3] H. Nakada, *On metrical theory of Diophantine approximation over imaginary quadratic field*, Acta Arith. **51** (1988), no. 4, 399–403.
- [R] D. Romik, *The dynamics of Pythagorean triples*, Trans. Amer. Math. Soc. **360** (2008), 6045–6064.
- [Sa] SageMath, the Sage Mathematics Software System (Version 7.3), The Sage Developers, 2016, <http://www.sagemath.org>.
- [Sc1] A.L. Schmidt, *Diophantine Approximation of Complex Numbers*, Acta Math. **134** (1975), 1–85.
- [Sc2] A.L. Schmidt, *Ergodic Theory for Complex Continued Fractions*, Mh. Math. **93** (1985), 39–62.
- [Sc3] A.L. Schmidt, *Farey triangles and Farey quadrangles in the complex plane*, Math. Scand., **21** (1967), 241–295.
- [Sc4] A.L. Schmidt, *Diophantine approximation in the field $\mathbb{Q}(i\sqrt{2})$* , Journal of Number Theory **131** (2011), 1983–2012.
- [wS] W. Schmidt, *Diophantine Approximation*, Lecture Notes in Mathematics no. 785, 1996, Springer-Verlag.
- [St1] K.E. Stange, *Visualizing the arithmetic of imaginary quadratic fields*, Int. Math. Res. Not. (2017).
- [St2] K.E. Stange, *The Apollonian structure of Bianchi groups*, to appear in Trans. Amer. Math. Soc., <http://arxiv.org/abs/1505.03121>.
- [V] L.Y. Vulakh, *Diophantine approximation on Bianchi groups*, J. Number Theory **54** (1995), 73–80.
- [W] Wolfram Research, Inc., Mathematica, Version 10.0, Champaign, IL (2014).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, 1409 WEST GREEN STREET, URBANA, IL 61801
Email address: chaubey2@illinois.edu

DEPARTMENT OF MATHEMATICS, UC DAVIS, ONE SHIELDS AVENUE, DAVIS, CA 95616
Email address: efuchs@math.ucdavis.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, CAMPUS BOX 395, BOULDER, COLORADO 80309-0395
Email address: robert.hines@colorado.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO, CAMPUS BOX 395, BOULDER, COLORADO 80309-0395
Email address: kstange@math.colorado.edu