Ergod. Th. & Dynam. Sys. (2014), **34**, 1832–1848 © *Cambridge University Press, 2013*
doi:10.1017/etds.2013.23

Distribution of approximants and geodesic flows

ALBERT M. FISHER† and THOMAS A. SCHMIDT‡

† *Dept Mat IME-USP, Caixa Postal 66281, CEP 05315-970 São Paulo, Brazil*
(e-mail: afisher@ime.usp.br)
‡ *Oregon State University, Corvallis, OR 97331, USA*
(e-mail: toms@math.orst.edu)

(*Received* 11 *August* 2012 *and accepted in revised form* 6 *March* 2013)

Abstract. We give a new proof of Moeckel's result that for any finite index subgroup of the modular group, almost every real number has its regular continued fraction approximants equidistributed into the cusps of the subgroup according to the weighted cusp widths. Our proof uses a skew product over a cross-section for the geodesic flow on the modular surface. Our techniques show that the same result holds true for approximants found by Nakada's α-continued fractions, and also that the analogous result holds for approximants that are algebraic numbers given by any of Rosen's λ-continued fractions, related to the infinite family of Hecke triangle Fuchsian groups.

1. *Introduction*

For each non-zero real number, the regular continued fraction algorithm produces a sequence of rational numbers that are best approximations in the appropriate sense. It is natural to ask about the number-theoretic properties of the numerators and denominators in each such sequence. In particular, as we learned from Sheingorn [**She93**], going back to the 1940s there has been study of the distribution into congruence classes of these numerators and denominators. The situation was beautifully resolved in 1982 by Moeckel [**Moe82**], who used the celebrated connection between regular continued fractions and geodesics on the hyperbolic modular surface.

A fascinating aspect of the problem is that it does not directly involve observables of the dynamical system defined by the *Gauss map*, that is, by the interval map related to the regular continued fractions. Furthermore, the vocabulary of the result indicates some of its depth: (1) any finite index subgroup H of $\mathrm{SL}(2,\mathbb{Z})$ uniformizes the hyperbolic *surface* $\bar{H}\backslash\mathbb{H}$, where $\mathbb{H}$ is the Poincaré upper half-plane and $\bar{H}$ denotes the image of H in $\Gamma := \mathrm{PSL}(2,\mathbb{Z}) = \mathrm{SL}(2,\mathbb{Z})/\pm I$; (2) a *cusp* κ of H is an $\bar{H}$-orbit in $\mathbb{Q}\cup\{\infty\}$; (3) the *width* of a cusp is the index $w(\kappa)$ of the $\bar{H}$-stabilizer of a point in this $\bar{H}$-orbit inside the point's $\bar{\Gamma}$-stabilizer.

Combined with a later result of Nakanishi [**Nak89**], the Moeckel result, which we re-prove here and then generalize, is the following.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

THEOREM 1. (Moeckel, Nakanishi) *Suppose that H is a finite index subgroup $H \subset \mathrm{SL}(2, \mathbb{Z})$. Then, for almost every real x, the regular continued fraction approximants of x are distributed in the cusps of H according to the relative cusp widths. That is, for each cusp κ and almost all x,*

$$\lim_{N\to\infty} \frac{\#\{0 \le n \le N \mid p_n/q_n \in \kappa\}}{N} = w(\kappa)/[\Gamma : \bar{H}]. \tag{1}$$

Since for each positive integer m, the principal congruence subgroup $\Gamma(m)$, defined as the kernel of the homomorphism to the projective matrix group over the finite ring $\mathbb{Z}/m\mathbb{Z}$, is a normal subgroup, the result states that for almost all x, the approximants p_k/q_k are equidistributed in the cusps of $\Gamma(m)$. In particular, when $m = 2$ the result is that for almost all x, its approximants have limiting frequency of $1/3$ for each of the type `odd/odd`, `odd/even` and `even/odd`.

For his proof, Moeckel relied on the ergodicity of the geodesic flow on the unit tangent bundle of any finite cover of the modular surface. A few years later, Jager and Liardet [**JL88**] gave a new proof (in the main setting of the principal congruence subgroups) that was in a sense purely ergodic theoretic; in particular, they used a skew product over the Gauss map to avoid the use of the geodesic flow. The logical structure of both proofs is of two steps: an ergodicity argument and a counting argument. Whereas Moeckel's counting argument seems slightly artificial—indeed, it forced a restriction to a particular subset of the finite covers of the modular surface, a restriction later removed by Nakanishi but only by the addition of a second counting argument in the complementary case—Jager and Liardet's counting argument is straightforward. On the other hand, the ergodicity argument in Moeckel's approach is in some sense immediate, but Jager and Liardet must labor in the corresponding step.

We give a combined proof that unites the strengths of the two approaches. By using a skew product based over a cross-section of the geodesic flow on the unit tangent bundle of the modular surface, which is itself a cross-section for a lifted flow, ergodicity is immediate while counting is straightforward. Furthermore, our proof needs no adjustment to treat the setting of any finite cover of the modular surface. The cross-section of the geodesic flow on the unit tangent bundle of the modular surface that we use is that found by Arnoux [**Arn94**]; this has the Gauss map as a factor transformation. To be more exact, the first return map to Arnoux's cross-section gives a double covering of the natural extension of the Gauss map. Our approach is quite flexible, as we show by not only recovering the Moeckel–Nakanishi results for the approximants generated by the regular continued fractions, but also extending these results to Nakada's α-continued fractions [**Nak81**] and to Rosen's λ_m-continued fractions [**Ros54**]. The latter are related not to merely the modular group, but to each of an infinite family of triangle Fuchsian groups, known as the Hecke groups; the Rosen fractions give approximation by quotients of certain algebraic integers.

The study of the intertwined nature of continued fractions and the geodesic flow on the modular surface has a rich history. One of the first significant steps was in 1924, when Artin [**Art24**] gave a coding of geodesics in terms of the regular continued fraction expansions of the real endpoints of their lifts. A decade later, Hedlund [**Hed35**] used this to show the ergodicity of the geodesic flow on the modular surface. A few years thereafter, Hopf proved ergodicity for the geodesic flow on any hyperbolic surface of finite volume;

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

see his reprisal [**Hop71**]. Adler and Flatto (see especially their [**AF91**]) and Series [**Ser85**] (inspired by Moeckel's result) gave cross-sections for the geodesic flow on the modular surface such that the regular continued fraction map is given as a factor. Quite recently, Mayer and co-authors [**MS08, MM10**] have given explicit cross-sections for geodesic flow on the surfaces uniformized by the Hecke triangle groups such that a certain variant of the Rosen continued fractions [**Ros54**] occurs as a factor. (The second-named author first learned of such possibilities from Haas; see the related treatment of another variant of the Rosen fractions in [**GH96**].) For much of this history, further motivation, and also the work of S. Katok and co-authors using various means to code geodesics on the modular surface, see [**KU07**]. As stated above, for the regular continued fractions we use a cross-section given by Arnoux [**Arn94**]; for the other continued fractions we study, we use cross-sections found by Arnoux and Schmidt [**AS13**] with a method directly inspired by [**Arn94**]. (Arnoux's approach to constructing cross-sections is more algebraic in flavor than the others we have mentioned.)

Our approach does have the disadvantage of presenting consecutive approximants as first or second columns alternately (see Lemma 5), but this is natural when directly following the Gauss map. We address this by twisting from a conjugate of the subgroup of interest to the subgroup itself; see the end of the proof of Corollary 2.

After this work was completed, we learned from Pollicott's recent survey article [**Pol11**] that Chang and Mayer [**CM00**] had previously used the skew product approach to lift a cross-section for geodesic flow on the unit tangent bundle of the modular surface to give cross-sections for finite index covers; this allows them to study dynamical zeta functions for these covers.

2. *Background*

2.1. *Regular continued fractions.*

The regular continued fractions map, or the Gauss interval map, on $\mathbb{I} := [0, 1)$ is

$$T(x) := \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \quad \text{for } x \neq 0; \;\; T(0) := 0.$$

For $x \in \mathbb{I}$, put $d(x) := \lfloor 1/x \rfloor$, with $d(0) = \infty$, and set $d_n = d_n(x) := d(T^{n-1}(x))$. This yields the regular continued fraction expansion of $x \in \mathbb{R}$:

$$x = d_0 + \cfrac{1}{d_1 + \frac{1}{d_2 + \cdots}} =: [d_0; d_1, d_2, \ldots],$$

where $d_0 \in \mathbb{Z}$ is such that $x - d_0 \in \mathbb{I}$. (Standard convergence arguments justify equality of x and its expansion.) The *approximants* p_n/q_n of $x \in \mathbb{I}$ are given by

$$\begin{pmatrix} p_{-1} & p_0 \\ q_{-1} & q_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & d_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & d_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & d_n \end{pmatrix} \tag{2}$$

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

for $n \geq 1$. Of course, this is equivalent to $p_n/q_n = [d_1, d_2, \ldots, d_n]$. From the definition it is immediate that $|p_{n-1}q_n - q_{n-1}p_n| = 1$ and in particular that the rational numbers p_n/q_n are all in reduced form.

The standard number-theoretic planar map associated to continued fractions is

$$\mathcal{T}(x, y) := \left(T(x), \frac{1}{d(x) + y}\right), \quad (x, y) \in \Omega,$$

with $\Omega = \mathbb{I} \times [0, 1]$. The measure μ on $\mathbb{R}^2$ given by

$$d\mu = (1 + xy)^{-2}\, dx\, dy \tag{3}$$

is invariant for $\mathcal{T}$. Indeed, Nakada *et al* [**NIT77**] showed that this two-dimensional system gives a natural extension for the map T. In particular, the marginal measure obtained by integrating μ over the fibers $\{x\} \times \{y \mid (x, y) \in \Omega\}$ gives the invariant measure ν with $d\nu = (1 + x)^{-1}dx$. The probability measure on $\mathbb{I}$ obtained by division by $\log 2$ is known as Gauss measure.

2.2. *Geodesic flow.* Some of the following can be found in [**Man91**].

Using the Möbius action of $\mathrm{SL}_2(\mathbb{R})$ on the Poincaré upper-half plane $\mathbb{H}$, one can identify a matrix with the image of $z = i$ under the matrix. This results in $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$ being identified with the Poincaré upper half-plane $\mathbb{H}$. Similarly, $G := \mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\pm I$ can be identified with the *unit tangent bundle* of $\mathbb{H}$. Recall that a Fuchsian group Γ is any discrete subgroup of G, equivalently that Γ acts properly discontinuously on $\mathbb{H}$ and the quotient is a hyperbolic surface (to be precise, possibly an orbifold). The unit tangent bundle of this quotient can be identified with $\Gamma\backslash G$.

The geodesic flow in our setting acts on the surface's unit tangent bundle. Given a time t and a unit tangent vector v, since the unit tangent vector uniquely determines a geodesic passing through the vector's base point, we can follow that geodesic for arc length t in the direction of v; the unit vector that is tangent to the geodesic at the endpoint of the geodesic arc is the image, $g_t(v)$, under the geodesic flow. The hyperbolic metric on $\mathbb{H}$ corresponds to an element of arc length satisfying $ds^2 = (dx^2 + dy^2)/y^2$ with coordinates $z = x + iy$. In particular, for $t > 0$, the points $z = i$ and $w = e^t i$ are of distance t apart. Since G acts on the left by isometries on $\mathbb{H}$, the geodesic flow on its unit tangent bundle is given by sending $A \in G$ to AE_t, where $E_t = \left(\begin{smallmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{smallmatrix}\right)$. Similarly, on $\Gamma\backslash G$ one sends the class represented by A to that represented by AE_t; we sometimes denote this by $g_t([A]_\Gamma) = [AE_t]_\Gamma$.

There is a natural measure on the unit tangent bundle $T^1\mathbb{H}$: *Liouville measure* is given as the product of the hyperbolic area measure on $\mathbb{H}$ with the length measure on the circle of unit vectors at any point. This measure is (left and right) $\mathrm{SL}_2(\mathbb{R})$-invariant, and thus gives Haar measure on G. In particular, this measure is invariant for the geodesic flow. When discussing measure theory, we always consider the Borel σ-algebra or its μ-completion.

Let $(X, \mathcal{B}, \mu)$ be a measure space and Φ_t a measure-preserving flow on X, that is, $\Phi : X \times \mathbb{R} \to X$ is a measurable function such that for $\Phi_t(x) = \Phi(x, t)$, $\Phi_{s+t} = \Phi_s \circ \Phi_t$. Then $\Sigma \subset X$ is a *measurable cross-section* for the flow Φ_t if: (1) the flow orbit of almost every point meets Σ; (2) for almost every $x \in X$ the set of times t such that $\Phi_t(x) \in \Sigma$ is a discrete subset of $\mathbb{R}$; (3) for every $\tau > 0$, the *flow box* $A_{[0,\tau]} \equiv \{\Phi_t(A) \mid A \in \Sigma, t \in [0, \tau]\}$ is μ-measurable. This last defines the induced σ-algebra $\mathcal{B}_\Sigma$ on Σ.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

The *return-time function* $r = r_\Sigma$ is $r(x) = \inf\{t > 0 : \Phi_t(x) \in \Sigma\}$ and the *return map* $R : \Sigma \to \Sigma$ is defined by $R(x) = \Phi_{r(x)}(x)$. The *induced measure* μ_Σ on Σ is defined from flow boxes: one sets $\mu_\Sigma(A) = (1/\tau)\mu(A_{[0,\tau]})$ for (all) $0 < \tau < \inf_{x\in A}\{r(x)\}$. The flow $(X, \mathcal{B}, \mu)$ is then naturally isomorphic to the *special flow* built over the return map (Σ, μ_Σ, R) with return time r, as follows: define $\widehat{\Sigma} = \{(x, t) : 0 \le t \le r(x)\}/\sim$, where $\sim$ is the equivalence relation $(x, r(x)) \sim (R(x), 0)$, equipped with the measure $\widehat{\mu}$, the product of μ_Σ on Σ with Lebesgue measure on $\mathbb{R}$; the flow Φ^R on $\widehat{\Sigma}$ is defined by $\Phi^R_t : (x, s) \mapsto (x, s + t)$; this preserves $\widehat{\mu}$.

Convention. In all that follows, we will write *cross-section* to denote measurable cross-section.

A flow Φ_t is *ergodic* if for any invariant set either it or its complement is of measure zero. Of fundamental importance for us is Hopf's result that the geodesic flow is ergodic on the unit tangent bundle of any finite volume hyperbolic surface; see his reprisal in [**Hop71**]. A flow is *recurrent* if the Φ-orbit of almost every point meets any positive measure set infinitely often. By the Poincaré recurrence theorem, an ergodic flow on a finite measure space is recurrent. Given a cross-section, a return-time transformation $(\Sigma, \mathcal{B}_\Sigma, \mu_\Sigma, R_\Sigma)$ is ergodic and recurrent if and only if the flow is. A *semiconjugacy*, *homomorphism* or *factor map* of two measure-preserving flows or transformations is an almost surely onto map which preserves the measures and the dynamics; a *conjugacy* or *isomorphism* is in addition invertible, with the inverse map also a homomorphism.

The following summarizes some of the above from a perspective of central importance to us.

LEMMA 1. *Let G denote the group* $\mathrm{PSL}(2, \mathbb{R})$. *Suppose that Γ is a Fuchsian group of finite covolume and that $\mathcal{A} \subset G$ projects injectively to give a cross-section Σ for the geodesic flow on the surface uniformized by Γ. Then for almost every $A \in \mathcal{A}$ there exists a unique pair of first return time $t = t_A$ and $M \in \Gamma$ such that $MAE_t \in \mathcal{A}$.*

We shall also need the following result about lifts of cross-sections.

LEMMA 2. *Let $(\widetilde{X}, \widetilde{\mathcal{B}}, \widetilde{\mu}, \widetilde{\Phi}_t)$ and $(X, \mathcal{B}, \mu, \Phi_t)$ be two measure-preserving flows such that the second is a factor of the first, that is, there exists $\pi : \widetilde{X} \to X$ measure-preserving and onto such that $\pi \circ \widetilde{\Phi}_t = \Phi_t \circ \pi$. Let (Σ, μ_R, R) be a cross-section and return map for Φ_t. Then $\widetilde{\Sigma} \equiv \pi^{-1}(\Sigma)$ is a cross-section for $\widetilde{\Phi}_t$, with return function $\widetilde{r} = r \circ \pi$, return map $\widetilde{R} = \widetilde{\Phi}_{\widetilde{r}(\widetilde{x})}(\widetilde{x})$ and measure $\widetilde{\mu}$ on $\widetilde{\Sigma}$ the induced measure; the two return transformations are semiconjugate by way of the restriction of π to $\widetilde{\Sigma}$.*

The proof follows from the definitions. We call $(\widetilde{\Sigma}, \widetilde{\mathcal{B}}_\Sigma, \widetilde{\mu}, \widetilde{R})$ the *lifted cross-section*.

2.3. *Arnoux cross-sections.* Inspired by Veech's 'zippered rectangles' [**Vee84**] as a means to study the geodesic flow on Teichmüller spaces, Arnoux [**Arn94**] gave a coding of the geodesic flow on the unit tangent bundle of the modular surface $\mathrm{PSL}(2, \mathbb{Z})\backslash G$ by way of 'boxes' described in terms of two types of subsets of $\mathrm{SL}_2(\mathbb{R})$. Given $(x, y) \in \mathbb{R}^2$, let

$$A_{-1}(x, y) = \begin{pmatrix} 1 & y \\ -x & 1 - xy \end{pmatrix} \quad \text{and} \quad A_{+1}(x, y) = \begin{pmatrix} x & 1 - xy \\ -1 & y \end{pmatrix}. \tag{4}$$

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

Let

$$\mathcal{Z} : \mathbb{R}^2 \backslash \{(x, y) \mid y \neq -1/x\} \to \mathbb{R}^2 \qquad (x, y) \mapsto (x, y/(1 + xy)) \tag{5}$$

and define the following subsets of $\mathrm{SL}_2(\mathbb{R})$ (recall that Ω is the domain of the planar natural extension of the regular continued fractions map):

$$\mathcal{A}_{-1} = \{A_{-1}(x, y) \mid (x, y) \in \mathcal{Z}^{-1}(\Omega)\} \quad \text{and} \quad \mathcal{A}_{+1} = \{A_{+1}(x, y) \mid (x, y) \in \mathcal{Z}^{-1}(\Omega)\}.$$

Finally, let $\mathcal{A} = \mathcal{A}_{-1} \cup \mathcal{A}_{+1}$, considered as elements of G.

Arnoux [**Arn94**] shows the following theorem.

THEOREM 2. (Arnoux) *The set $\mathcal{A} \subset G$ projects injectively to give a cross-section for the geodesic flow on the unit tangent bundle of the modular surface, thus on $\Gamma\backslash G = \mathrm{PSL}(2, \mathbb{Z})\backslash\mathrm{PSL}(2, \mathbb{R})$. Furthermore, the first return map to this cross-section by the geodesic flow is given by the projection of*

$$A_\sigma(x, y) \mapsto M A_\sigma(x, y) E_t,$$

with $t = -2 \log x$ and

$$M = \begin{cases} \begin{pmatrix} 1 & \lfloor 1/x \rfloor \\ 0 & 1 \end{pmatrix} & \text{if } \sigma = -1, \\ \begin{pmatrix} 1 & 0 \\ \lfloor 1/x \rfloor & 1 \end{pmatrix} & \text{if } \sigma = +1. \end{cases}$$

Moreover, the dynamical system defined by this first return map has the Gauss map as a factor via the projection to the first coordinate.

2.4. *Cusps, cosets and covers.* An element of $G = \mathrm{PSL}(2, \mathbb{R})$ is parabolic if it fixes a unique element of $\mathbb{R} \cup \{\infty\}$. For the corresponding $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{R})$, this is equivalent to $|\mathrm{tr} A| = |a + d|$ being equal to 2. If Γ is a Fuchsian subgroup of G, then the Γ-orbit of any of its parabolic fixed points consists solely of parabolic fixed points for Γ; we call such an orbit a *cusp* of Γ. It is also a standard fact that the Γ-stabilizer of any parabolic fixed point is infinite cyclic. By conjugating in G as necessary, we can assume that the parabolic fixed point is at infinity, and thus the stabilizer is a cyclic group of translations; the cusp of the group corresponds to a finite area end of the hyperbolic surface $\Gamma\backslash\mathbb{H}$, traditionally also called a *cusp*—with our normalization, this geometric cusp is the projection of a hyperbolic triangle with one vertex at infinity, where a generator of the stabilizer acts so as to identify the two sides of the triangle.

Fix some finite index subgroup $H \subset \Gamma$. For any cusp κ of H, choose a parabolic fixed point p in this H-orbit, and define the *width* (with respect to Γ) of κ as the index $w(\kappa)$ of the H-stabilizer of p in the Γ-stabilizer of p. The geometric significance of this width is that one can glue together $w(\kappa)$ copies of the triangle in $\mathbb{H}$ giving the geometric cusp of the surface uniformized by Γ to construct the triangle giving the geometric cusp corresponding to κ.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

We denote a right H-coset in Γ as either $H\gamma$ or $\gamma \bmod H$, depending upon ease of typography.

Now, given a parabolic fixed point p of Γ and a coset $H\gamma$, since $\gamma \cdot p$ is a parabolic fixed point (also) for H, it lies in some cusp κ of H. But then so does its whole H-orbit; that is, the image of p under $H\gamma$ is the full H-orbit κ. There is thus an equivalence relation on the set of cosets $H\backslash\Gamma$, characterized by sending p to the same cusp. Suppose that Γ has exactly one cusp. Since Γ acts transitively on its single cusp (viewed as a Γ-orbit), and Γ is the union of the right cosets of H in Γ, we see that every H-cusp κ has a non-empty corresponding equivalence class. Let $w'(\kappa)$ be the number of elements in this class.

We now show the equality $w'(\kappa) = w(\kappa)$. Choose a parabolic fixed point q whose H-orbit is κ, and also a full set of right coset representatives γ_i for H in Γ. Each $H\gamma_i$ in the equivalence class of κ is then such that there is an $h_i \in H$ such that $h_i\gamma_i \cdot p = q$. For ease of argument, take each $h_i\gamma_i$ as the coset representative in place of γ_i. Now choose a convex connected fundamental domain with geodesic sides, $\mathcal{F}$, for Γ acting on $\mathbb{H}$ which reaches the boundary at p. Let P be a generator of the Γ-stabilizer of p (thus, P pairs the two sides of $\mathcal{F}$ meeting at p); by elementary group theory each γ_i conjugates P to some generator of the Γ-stabilizer of q. In fact, due to the Möbius action being orientation-preserving, each conjugates P to the same generator, say Q. Now, a fundamental domain for H, say $\mathcal{G}$, is given by taking the union of the images of $\mathcal{F}$ under each of the full set of coset representatives.

The union of the $w'(\kappa)$ images of $\mathcal{F}$ under the γ_i sending p to q is bounded by two sides of $\mathcal{G}$. These two sides are paired by $Q^{w'(\kappa)}$; furthermore, our choice of coset representatives is such that no other parabolic point on the boundary of $\mathcal{G}$ is in the H-orbit of q. Therefore, by classical arguments for Fuchsian groups (see, say, the discussion of 'parabolic cycle' in [**Leh66**]), the H-stabilizer of q is generated by $Q^{w'(\kappa)}$. Thus, the index of the H-stabilizer of q inside its Γ-stabilizer is $w'(\kappa)$, and the equality holds.

We summarize the above in the following lemma.

LEMMA 3. *Suppose that Γ is a Fuchsian group with a single cusp and that H is a finite index subgroup of Γ. Then upon choice of a parabolic fixed point of Γ, there is an equivalence relation on the right cosets $H\backslash\Gamma$, corresponding to cusps of H. The size of an equivalence class is the width of the corresponding cusp.*

We define notation for the equivalence classes on the right cosets.

Definition 1. Given Γ and H as in Lemma 3, and a parabolic fixed point p for Γ, let

$$[H\gamma]_{\cdot p}$$

denote the equivalence class of the right coset $H\gamma$ under the relation defined by equality of image of p, and

$$\{\kappa\}_{\cdot p}$$

denote the partition block of $H\backslash\Gamma$ corresponding to κ under this equivalence relation. We refer to the relation itself as the $\cdot p$ *relation* on $H\backslash\Gamma$.

Note that we will occasionally slur over distinctions and refer to the various equivalence classes of $H\backslash\Gamma$ defined by a $\cdot p$ relation as the cusps of H.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

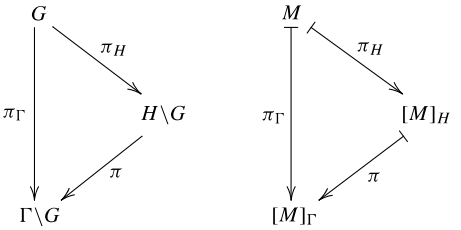


FIGURE 1. Basic notation of covers: sets and elementwise maps; here $G \supset \Gamma \supset H$.

We will have need to switch roles between a group H and its conjugate by

$$\iota = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in G. \tag{6}$$

Note that $\iota^{-1} = \iota$ in G.

LEMMA 4. *Suppose that Γ is a Fuchsian group with a single cusp, with $\iota \in \Gamma$, and that H is a finite index subgroup of Γ. If $\{\gamma_i\}_{1 \le i \le n}$ is a full set of coset representatives for $\iota H \iota \backslash \Gamma$, then the $\iota\gamma_i$ form a full set of coset representatives of $H\backslash\Gamma$. Furthermore, the map $p \mapsto \iota \cdot p$ places the $\iota H \iota$-cusps in one-to-one correspondence with the H-cusps. This correspondence preserves cusp widths.*

Proof. If Γ is partitioned by the $\iota H \iota \gamma_i$, then since $\iota \in \Gamma$ we have that $\Gamma = \iota\Gamma$ is also partitioned by the $H\iota\gamma_i$. Thus the $\iota\gamma_i$ do give a full set of H-cosets.

If p is a parabolic fixed point for Γ, then left multiplication by ι sends $\iota H \iota \cdot p$ to $H(\iota \cdot p)$. If $\iota H \iota \gamma_i \cdot p = \iota H \iota \gamma_j \cdot p$, then by left multiplication by $\iota^{-1} = \iota$, we immediately have $H(\iota\gamma_i) \cdot p = H(\iota\gamma_j) \cdot p$. Thus, the equivalence relation defined by p gives the same cusp widths for H as for $\iota H \iota$. □

For H a subgroup of G, we also use the notation for the associated projections shown in Figure 1—in particular, for ease of reading, we often use an expression of the form $[M]_\Gamma$ to denote $\pi_\Gamma(M)$.

3. *Lifting cross-sections as skew products*

Given a cross-section for the geodesic flow on a hyperbolic surface and a finite degree cover of the surface, from the general facts of Lemma 2 the cross-section can be lifted to give a cross-section for the geodesic flow on the cover; as we see here, in this case the resulting system can be nicely expressed as a skew product over the original cross-section.

THEOREM 3. *Suppose that Γ is a Fuchsian group of finite covolume and that the projection map π_Γ injectively maps a subset $\mathcal{A} \subset G$ to a cross-section Σ for the geodesic flow on $\Gamma\backslash G \cong T^1(\Gamma\backslash\mathbb{H})$. Let $H \subset \Gamma$ be any finite index subgroup of Γ, and $H\backslash\Gamma$ the set of right cosets of H in Γ. Then the skew product transformation*

$$\mathcal{S} : \mathcal{A} \times H\backslash\Gamma \to \mathcal{A} \times H\backslash\Gamma \quad \textit{defined by}$$
$$(A, \gamma \bmod H) \mapsto (MAE_t, \gamma M^{-1} \bmod H),$$

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

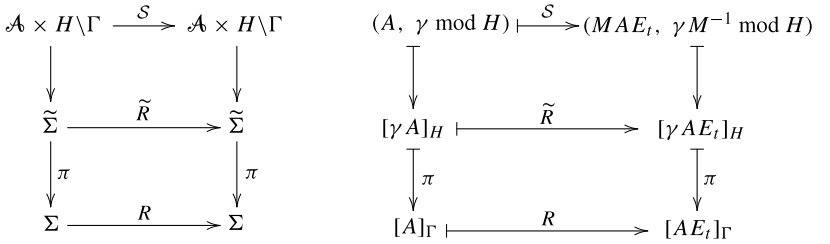


FIGURE 2. Lifting the cross-section $\Sigma = \pi_\Gamma(\mathcal{A})$. Diagrams of sets and elementwise maps.

where M and E_t are as in Lemma 1, defines a dynamical system naturally isomorphic to the first return map on the lifted cross-section of the geodesic flow on $H\backslash G \cong T^1(H\backslash\mathbb{H})$. The measures involved are these: on $\Gamma\backslash G$ and $H\backslash G$ we have Liouville measure, and their cross-sections Σ, $\widetilde{\Sigma}$ carry the natural induced measures; the measure on $\mathcal{A}$ is passed over via the bijection to Σ and on the finite set $H\backslash\Gamma$ we take counting measure, with the skew product carrying the product of these.

See Figure 2 for diagrams related to this proof. We use juxtaposition of sets of group elements to denote the set of all products of respective elements of these sets.

Proof. Choose a set of right coset representatives $\mathcal{C} \subset \Gamma$ for $H\backslash\Gamma$.

Consider $A_0 \in \mathcal{A}$; its projection $[A_0]_\Gamma \in \pi_\Gamma(\mathcal{A}) = \Sigma$ is sent to some $[A_1]_\Gamma$ by the first return map R on Σ. We claim that R is isomorphic to the map on $\mathcal{A}$ defined by $A_0 \mapsto M_0A_0E_0$. To check this, by Lemma 1 the geodesic flow on $\Gamma\backslash G$ takes $[A_0]_\Gamma$ to $[A_1]_\Gamma$, with $A_1 = M_0A_0E_0$, where as usual $M_0 \in \Gamma$ and E_0 is the diagonal matrix realizing the flow for a time t_0. Next we consider $\gamma_0 \in \mathcal{C}$ and $\gamma_0A_0 \in \mathcal{C}\mathcal{A}$; we claim that the return map $\widetilde{R}$ for the lifted cross-section $\widetilde{\Sigma}$ is isomorphic to the map on $\mathcal{C}\mathcal{A}$ defined by $\gamma_0A_0 \mapsto \gamma_1A_1$ where $\gamma_1 = \gamma_0M^{-1}$. But since Γ uniquely decomposes as $H\mathcal{C}$, there exists a unique pair $h \in H$, $\gamma_1 \in \mathcal{C}$ such that $h^{-1}\gamma_1 = \gamma_0M_0^{-1} \in \Gamma$. Equivalently, $h\gamma_0 = \gamma_1M_0$ and thus

$$h\gamma_0A_0E_0 = \gamma_1M_0A_0E_0 = \gamma_1A_1.$$

That is, $[\gamma_0A_0]_H$ returns to $\pi_H(\mathcal{C}\mathcal{A})$ at time t_0. This is, moreover, the *first* return to this set. For suppose that there are $h' \in H$, $\gamma \in \mathcal{C}$, $A \in \mathcal{A}$ and $t > 0$ such that $h'\gamma_0A_0E_t = \gamma A$. The element in Γ given by $M = \gamma^{-1}h'\gamma_0$ gives $MA_0E_t = A$ and thus t is also a Φ_Γ-return time for $[A_0]_\Gamma$, showing that $t \geq t_0$. Therefore, t_0 is indeed the first return time for $[\gamma_0A_0]_H$.

Since the map sending $(A, H\gamma)$ to $[\gamma A]_H$ is clearly surjective from $\mathcal{A} \times H\backslash\Gamma$ to $\pi_H(\mathcal{C}\mathcal{A})$, it remains to show it is also injective. Now suppose that $A, A' \in \mathcal{A}$ and $\gamma, \gamma' \in \Gamma$ such that $[\gamma A]_H = [\gamma' A']_H$. It follows that also $[\gamma A]_\Gamma = [\gamma' A']_\Gamma$ and hence $\pi_\Gamma(A) = \pi_\Gamma(A')$; by the injectivity of π_Γ on $\mathcal{A}$, we have $A = A'$. But $[\gamma A]_H = [\gamma' A']_H$ implies the existence of $h \in H$ such that $h\gamma A = \gamma' A'$; canceling $A = A'$, we find that γ and γ' are in the same right H-coset of Γ.

We have verified that our map is a bijection that conjugates the skew product transformation to the return map of $\widetilde{\Sigma}$. The measures on the cross-sections correspond as well, since the Liouville measure on $H\backslash G$ may be viewed as the restriction, or projection,

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

of Haar measure on G, while the induced measure on the cross-section $\widetilde{\Sigma}$ is the product of Haar measure on G restricted to $\mathcal{A}$ times counting measure on $H\backslash\Gamma$ (since a flow box for the lifted cross-section is a finite disjoint union of copies of the corresponding flow box for the section).

Finally, since the skew product is independent of choice of coset representatives, we find that the isomorphism is natural. This completes the proof. □

Since invariant subsets of a cross-section and of a flow correspond, and since we know from Hopf's theorem that the geodesic flow on a finite volume hyperbolic surface is ergodic, the following is evident.

COROLLARY 1. *The skew product transformation $\mathcal{S}$ is ergodic.*

We note the following connection between flow cross-sections and fundamental domains.

PROPOSITION 1. *If $\mathcal{A} \subset G$ projects injectively to give a cross-section with return time r for the geodesic flow on the surface uniformized by Γ, then a (measurable) fundamental domain for the action of Γ on G is given by the set*

$$\{AE_t \mid A \in \mathcal{A}, 0 \le t \le r\}.$$

4. *Distribution of regular continued fraction approximants into cusps*

Now we let $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$ be the modular group, and H be any of its finite index subgroups, and begin with the cross-section $\mathcal{A}$ for the geodesic flow on $\Gamma\backslash G$ given by Arnoux; see Theorem 2 above. The first return map to this cross-section maps elements of the $(\sigma = -1)$-component to the $(\sigma = +1)$-component and *vice versa*. This aspect of the map affects the presentation of the approximants p_k/q_k.

LEMMA 5. *Let $\mathcal{A}$ be the cross-section of the geodesic flow on the unit tangent bundle of the modular surface given in Theorem 2. For $A_{-1}(x, y) \in \mathcal{A}$, the second component of the k*th *composite of the skew product transformation $\mathcal{S}$ (defined in Theorem 3) with itself applied to $(A, I \bmod H)$ is*

$$\mathrm{proj}_2(\mathcal{S}^k(A, I \bmod H)) = \begin{cases} \begin{pmatrix} q_k & -q_{k-1} \\ -p_k & p_{k-1} \end{pmatrix} \bmod H & \text{if } k \text{ is even}, \\ \begin{pmatrix} q_{k-1} & -q_k \\ -p_{k-1} & p_k \end{pmatrix} \bmod H & \text{if } k \text{ is odd}. \end{cases} \tag{7}$$

Proof. Since the initial value of $\sigma = -1$ and this value alternates in sign with each application of $\mathcal{S}$,

$$\begin{aligned} \mathrm{proj}_2(\mathcal{S}^k(A, I \bmod H)) &= M_1^{-1} M_2^{-1} \cdots M_k^{-1} \mod H \\ &= \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ d_2 & 1 \end{pmatrix}^{-1} \cdots M_k^{-1} \mod H. \end{aligned}$$

Upon referring to equation (2), an elementary proof by induction verifies the statement. □

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

The following two results first address the fact that the matrices above are of two distinct types, and then that the transpose of (p_k, q_k) is not a column of either of the two matrices displayed above. These matrices act respectively so as to send the parabolic fixed points ∞ and 0 to $-q_k/p_k$; we thus consider the cusps identified as in Definition 1 with equivalence classes of $H\backslash\Gamma$ for the $\cdot p$ relation, where p is 0 or ∞. In the following, we denote the indicator function for a set E by $\mathbb{1}_E$.

LEMMA 6. *Fix a cusp κ of H, and let Ψ be the function*

$$\begin{aligned}\Psi : \mathcal{A} \times H\backslash\Gamma &\to \mathbb{R}\\ (A, \gamma \bmod H) &\mapsto \mathbb{1}_{\mathcal{A}_{-1}\times\{\kappa\}_{\cdot\infty}} + \mathbb{1}_{\mathcal{A}_{+1}\times\{\kappa\}_{\cdot 0}}.\end{aligned}$$

Then, for almost every $(A, \gamma \bmod H)$ in $\mathcal{A} \times H\backslash\Gamma$, the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Psi \circ \mathcal{S}^k(A, \gamma \bmod H)$$

is equal to the relative width of the cusp, $w(\kappa)/[\Gamma : H]$.

Proof. The function Ψ is the sum of indicator functions for two sets, each of whose measure relative to the measure of the full space $\mathcal{A} \times H\backslash\Gamma$ is the positive real number $w(\kappa)/(2[\Gamma : H])$. Therefore, the ergodicity of $\mathcal{S}$ and the Birkhoff ergodic theorem imply the result. □

In the next proof, we first easily find that for almost every x, the sequence of values $-q_k/p_k$ derived from its regular continued fraction approximants has the appropriate frequency. To obtain the Moeckel–Nakanishi result, we then focus on the fact that $\iota \cdot (-q_k/p_k) = p_k/q_k$. That is, we now finish our proof of that result. (When comparing with (1), recall that we now insist on H denoting a projective group.)

COROLLARY 2. *For each cusp κ of H and for almost every $x \in \mathbb{I}$,*

$$\lim_{N\to\infty} \frac{\#\{1 \le k \le N \mid p_k/q_k \in \kappa\}}{N} = w(\kappa)/[\Gamma : H].$$

Proof. The subset $\mathcal{A}_{-1} \times I \bmod H \subset \mathcal{A} \times H\backslash\Gamma$ has positive measure, and hence the limit of Lemma 6 holds for almost every element of it. But this then implies that the map defined as

$$\widetilde{\Psi}(A_\sigma(x, y), \gamma \bmod H) = \begin{cases} [H\gamma]_{\cdot\infty} & \text{if } \sigma = -1,\\ [H\gamma]_{\cdot 0} & \text{if } \sigma = +1,\end{cases}$$

sends the corresponding sequence $\mathcal{S}^k(A_{-1}(x, y), I \bmod H)$ into the cusp κ with limiting frequency equal to $w(\kappa)/[\Gamma : H]$. From Lemma 5 it follows that for almost every $x \in \mathbb{I}$, the sequence of $-q_k/p_k$ has this limiting frequency.

We now use the above with the role of H played by $\iota H \iota$—we have that $-q_k/p_k$ has this limiting frequency in any cusp of $\iota H \iota$. By Lemma 4, we find that the sequence $p_k/q_k = \iota \cdot (-q_k/p_k)$ has that same limiting frequency for corresponding cusps of H. By the same lemma, the cusp widths are the same for the two groups. But conjugate subgroups have the same index. Therefore, the result holds. □

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

5. *α-continued fractions*

Nakada [**Nak81**] defined the following family of interval maps. For $\alpha \in (0, 1]$, we let $\mathbb{I}_\alpha := [\alpha - 1, \alpha)$ and define the map $T_\alpha : \mathbb{I}_\alpha \to \mathbb{I}_\alpha$ by

$$T_\alpha(x) := \left|\frac{1}{x}\right| - \left\lfloor \left|\frac{1}{x}\right| + 1 - \alpha \right\rfloor \quad \text{for } x \neq 0;\ T_\alpha(0) := 0.$$

For $x \in \mathbb{I}_\alpha$, put

$$\varepsilon(x) := \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0, \end{cases} \qquad \text{and} \quad d_\alpha(x) := \left\lfloor \left|\frac{1}{x}\right| + 1 - \alpha \right\rfloor,$$

with $d_\alpha(0) = \infty$.

Furthermore, for $n \geq 1$, put

$$\varepsilon_n = \varepsilon_{\alpha,n}(x) := \varepsilon(T_\alpha^{n-1}(x)) \quad \text{and} \quad d_n = d_{\alpha,n}(x) := d_\alpha(T_\alpha^{n-1}(x)).$$

This yields the *α-continued fraction* expansion of $x \in \mathbb{R}$:

$$x = d_0 + \frac{\varepsilon_1}{d_1 + \frac{\varepsilon_2}{d_2 + \cdots}},$$

where $d_0 \in \mathbb{Z}$ is such that $x - d_0 \in \mathbb{I}_\alpha$. (Standard convergence arguments again justify equality of x and its expansion.) These include the regular continued fractions, given by $\alpha = 1$ and the nearest integer continued fractions, given by $\alpha = 1/2$.

We define the α-approximants to x so that the recurrence relations

$$\begin{aligned} p_{-1} = 1, \quad p_0 = 0, \quad p_n = d_n p_{n-1} + \varepsilon_n p_{n-2},\ n \geq 1, \\ q_{-1} = 0, \quad q_0 = 1, \quad q_n = d_n q_{n-1} + \varepsilon_n q_{n-2},\ n \geq 1, \end{aligned}$$

hold. One easily checks that $|p_{n-1}q_n - q_{n-1}p_n| = 1$.

We show in this section that the distribution of the α-continued fraction approximants p_k/q_k is the same as that for the regular continued fractions.

THEOREM 4. *Suppose that H is a finite index subgroup $H \subset \mathrm{PSL}(2, \mathbb{Z})$. Then, for each $\alpha \in (0, 1]$ and for almost every real x, the α-continued fraction approximants of x are distributed in the cusps of H according to the relative cusp widths.*

The standard number-theoretic planar map associated to these continued fractions is defined by

$$\mathcal{T}_\alpha(x, y) := \left(T_\alpha(x), \frac{1}{d_\alpha(x) + \varepsilon(x)\, y} \right), \quad (x, y) \in \Omega_\alpha;$$

see [**KSS12**] for a description of $\Omega_\alpha \subset \mathbb{I}_\alpha \times [0, 1]$. Here also, μ as given in (3) is an invariant measure.

Let μ_α be the probability measure given by normalizing μ on Ω_α, and ν_α the probability measure normalized from the marginal measure obtained by integrating μ_α over the fibers $\{x\} \times \{y \mid (x, y) \in \Omega_\alpha\}$, $\mathcal{B}_\alpha$ the Borel σ-algebra of $\mathbb{I}_\alpha$, and $\mathcal{B}'_\alpha$ the Borel σ-algebra of Ω_α. Kraaikamp *et al* [**KSS12**] showed that $(\Omega_\alpha, \mathcal{T}_\alpha, \mathcal{B}'_\alpha, \mu_\alpha)$ is a natural extension of $(\mathbb{I}_\alpha, T_\alpha, \mathcal{B}_\alpha, \nu_\alpha)$.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

Analogously to the regular continued fraction case of $\alpha = 1$, let

$$\mathcal{A}_{\alpha,-1} = \{A_{-1}(x, y) \mid (x, y) \in \mathcal{Z}^{-1}(\Omega_\alpha)\}$$

and

$$\mathcal{A}_{\alpha,+1} = \{A_{+1}(x, y) \mid (x, y) \in \mathcal{Z}^{-1}(\Omega\alpha)\},$$

where $\mathcal{Z}$ is given in (5); and, let $\mathcal{A}_\alpha = \mathcal{A}_{\alpha,-1} \cup \mathcal{A}_{\alpha,+1}$, considered as elements of G.

Arnoux and Schmidt [**AS13**] show the following theorem.

THEOREM 5. (Arnoux–Schmidt) *The set $\mathcal{A}_\alpha \subset G$ projects injectively (up to measure zero) to a cross-section for the geodesic flow on the unit tangent bundle of the modular surface. Furthermore, the first return map to this cross-section by the geodesic flow is given by the projection of*

$$A_\sigma(x, y) \mapsto M A_\sigma(x, y) E_t,$$

with $t = -2 \log |x|$ and M equaling

$$\begin{pmatrix} 0 & 1 \\ -1 & d \end{pmatrix}, \quad \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} d & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ d & 1 \end{pmatrix},$$

where $d = d_\alpha(x)$, as $(\sigma, \varepsilon(x)) = (-1, -1), (-1, +1), (+1, -1), (+1, +1)$, respectively. Moreover, the dynamical system defined by this first return map has the map T_α as a factor.

Note that when $\varepsilon(x) = +1$, M is of the same shape as in the regular continued fraction case given by Arnoux; see Theorem 2. In all cases, $\mathcal{A}_\alpha$ has two components, indexed by $\sigma = \pm 1$. In general, the map sends an element with $\varepsilon = +1$ in the component indexed by σ to the component indexed by $-\sigma$, whereas elements with $\varepsilon = -1$ are mapped to within the same component. This is key to showing that the analog of Lemma 5 holds with only two types of matrices, instead of four as the previous result might suggest.

LEMMA 7. *Let $\mathcal{A}_\alpha$ be the cross-section of the geodesic flow on the unit tangent bundle of the modular surface described in Theorem 5. Suppose that $A_{-1}(x, y) \in \mathcal{A}_\alpha$, then the second component of the kth composite of the skew product transformation $\mathcal{S}$ with itself applied to $(A, I \bmod H)$ is*

$$\begin{aligned} &\mathrm{proj}_2(\mathcal{S}^k(A, I \bmod H)) \\ &\quad = \begin{cases} \begin{pmatrix} q_k & -q_{k-1} \\ -p_k & p_{k-1} \end{pmatrix} \bmod H & \text{if } \mathrm{proj}_1(\mathcal{S}^k(A, I \bmod H)) \in \mathcal{A}_{\alpha,-1}, \\ \begin{pmatrix} q_{k-1} & -q_k \\ -p_{k-1} & p_k \end{pmatrix} \bmod H & \text{otherwise.} \end{cases} \end{aligned}$$

Note that the condition defining the two cases above becomes that of Lemma 5 upon restricting to the case of all $\varepsilon_i(x) = +1$.

Proof. For $A = A_{-1}(x, y) \in \mathcal{A}_{\alpha,-1}$, we have $\mathrm{proj}_1(\mathcal{S}^k(A, I \bmod H)) \in \mathcal{A}_{\alpha,-1}$ if and only if the number of $\varepsilon_i(x) = +1$ for $1 \le i \le k$ is even. The direct calculation verifying Lemma 5 shows that each application of $\mathcal{S}$ with $\varepsilon_i(x) = +1$ changes between the two shapes of matrices shown. Another direct calculation shows that each application of $\mathcal{S}$ with $\varepsilon_i(x) = -1$ preserves the respective shapes. The result follows. □

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

It is now immediate that the analogous statements to Lemma 6 and Corollary 2 hold, and we indeed conclude that the limiting distribution into cusps of the α-approximants is the same as that for the regular continued fraction approximants.

6. *Rosen continued fractions*

Let $\lambda = \lambda_m = 2\cos(\pi/m)$; then Γ_m, the Hecke group of index m, is the single-cusped Fuchsian group generated by the elements ι given above in (6), and

$$\gamma_m = \begin{pmatrix} 1 & \lambda_m \\ 0 & 1 \end{pmatrix}.$$

The group Γ_3 is the modular group $\mathrm{PSL}(2, \mathbb{Z})$. For $m \notin \{3, 4, 6\}$, the group Γ_m is non-arithmetic, (see [**Leu67**]) and thus, in particular, has no finite index subgroup that is $\mathrm{PSL}(2, \mathbb{R})$-conjugate to a finite index subgroup of $\mathrm{PSL}(2, \mathbb{Z})$.

In 1954, Rosen [**Ros54**] defined an infinite family of continued fraction algorithms to aid in the study of the Hecke groups. The Rosen continued fractions and variants have been of recent interest, leading to results especially about their dynamical and arithmetical properties (see [**BKS00, DKS09, Nak10**]) as well on their applications to the study of geodesics on related hyperbolic surfaces (see [**BS04, MS08, SS95**]) and to Teichmüller geodesics arising from (Veech) translation surfaces (see [**AH00, AS09, SU10, SU11**]). Several basic questions remain open, including that of arithmetically characterizing the real numbers having a finite Rosen continued fraction expansion (see [**AS09, HMTY08, Leu67**]).

We show in this section that the Rosen continued fraction approximants p_k/q_k are distributed analogously to those for the regular continued fractions. (We note that [**Nak10**] gives the following result in the setting of the distribution into cusps of the principal congruence subgroups of the Hecke groups, using an approach directly related to Moeckel's.)

THEOREM 6. *Fix an integer $m \geq 3$, and suppose that H is a finite index subgroup of the Hecke group of index m, $H \subset \Gamma_m$. Then, for almost every real x, the Rosen λ_m-continued fraction approximants of x are distributed in the cusps of H according to the relative (to Γ_m) cusp widths.*

Let $\mathbb{I}_m = [-\lambda/2, \lambda/2)$ for $m \geq 3$. For a fixed integer $m \geq 3$, the Rosen continued fraction map is defined by

$$T_m(x) = \begin{cases} \left|\frac{1}{x}\right| - \lambda\left\lfloor \left|\frac{1}{\lambda x}\right| + \frac{1}{2}\right\rfloor, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

for $x \in \mathbb{I}_m$; here and below, we omit the index m whenever it is clear from context. For $n \geq 1$, we define

$$\varepsilon_n(x) = \varepsilon(T_m^{n-1}x) \quad \text{and} \quad r_n(x) = r(T_m^{n-1}x)$$

with

$$\varepsilon(y) = \mathrm{sgn}(y) \quad \text{and} \quad r(y) = \left\lfloor \left|\frac{1}{\lambda y}\right| + \frac{1}{2}\right\rfloor.$$

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

Then, as Rosen showed in [**Ros54**], the Rosen continued fraction expansion of x is given by

$$[\varepsilon_1(x) : r_1(x),\, \varepsilon_2(x) : r_2(x),\, \ldots,\, \varepsilon_n(x) : r_n(x),\, \ldots] := \frac{\varepsilon_1}{r_1\lambda + \frac{\varepsilon_2}{r_2\lambda + \cdots}}.$$

As usual we define the *approximants* p_n/q_n of $x \in \mathbb{I}_m$ by

$$\begin{pmatrix} p_{-1} & p_0 \\ q_{-1} & q_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \begin{pmatrix} 0 & \varepsilon_1 \\ 1 & \lambda r_1 \end{pmatrix} \begin{pmatrix} 0 & \varepsilon_2 \\ 1 & \lambda r_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & \varepsilon_n \\ 1 & \lambda r_n \end{pmatrix}$$

for $n \geq 1$. From this definition it is immediate that $|p_{n-1}q_n - q_{n-1}p_n| = 1$, and that the following well-known recurrence relations hold:

$$\begin{aligned} p_{-1} = 1, \quad p_0 = 0, \quad p_n = \lambda r_n p_{n-1} + \varepsilon_n p_{n-2},\ n \geq 1; \\ q_{-1} = 0, \quad q_0 = 1, \quad q_n = \lambda r_n q_{n-1} + \varepsilon_n q_{n-2},\ n \geq 1. \end{aligned}$$

Sketch of proof of Theorem 6. For each m, Burton *et al* [**BKS00**] determined a planar natural extension on a region Ω_m with the measure μ as above. Using this, Arnoux and Schmidt [**AS13**] determine an $\mathcal{A}_m \subset G = \mathrm{PSL}(2, \mathbb{R})$ such that $\mathcal{A}_m$ projects to a cross-section of the geodesic flow on $\Gamma_m \backslash G$. The first return map to this cross-section is a double cover of the natural extension. Furthermore, for each index m an analog of Theorem 5 holds as shown in [**AS13**], where again when $\varepsilon_i(x) = +1$ there is a change of component, and when $\varepsilon_i(x) = -1$ there is not; the matrices giving the return to the cross-section here are

$$\begin{pmatrix} 0 & 1 \\ -1 & d\lambda_m \end{pmatrix}, \begin{pmatrix} 1 & d\lambda_m \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} d\lambda_m & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ d\lambda_m & 1 \end{pmatrix},$$

where $d = d_\alpha(x)$, as $(\sigma, \varepsilon(x)) = (-1, -1), (-1, +1), (+1, -1), (+1, +1)$, respectively.

Therefore, we easily find that the analog of Lemma 7 holds. Now, for each m we have both that γ_m fixes ∞, and $\iota \in \Gamma_m$; this allows the proof of Corollary 2 to apply. From this, one easily completes the proof of Theorem 6. □

As an example of an implication of this, we consider the case of $q = 5$. Note that λ_5 is the golden ratio, $(1 + \sqrt{5})/2$.

Example 1. Almost every real number has Rosen λ_5-continued fraction approximants with equal asymptotic frequency of the five types: `odd/even`, `even/odd`, $1/1$, $1/\lambda_5$, $1/(\lambda_5 + 1)$.

The proof of this relies on checking the data related to the principal congruence subgroup of Γ_5 for the $\mathbb{Z}[\lambda_5]$ ideal generated by 2: this (normal) subgroup has five cusps, each of width 2. We have chosen cusp representatives in the $(\gamma_5\iota)$-orbit of ∞.

We briefly sketch a non-geometric way to see that there should indeed be five cusps for this principal congruence subgroup. Note that $\mathbb{Z}[\lambda_5]$ is the full ring of integers of its quotient field, and that the rational prime ideal $\langle 2 \rangle$ remains prime (is inert to this number field extension). Thus, each element of a pair (p_k, q_k) reduces modulo $\langle 2 \rangle$ to some element

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

of the quotient field $\mathbb{Z}[\lambda_5]/\langle 2\rangle$, a finite field with four elements. However, each pair is relatively prime, thus there are obviously at most 12 combinations possible. Now, if either of p_k, q_k is a multiple of 2, then multiplying each by an inverse modulo 2 of the other reduces the quotient to the form of one of the cusps `odd/even`, `even/odd`. For the six remaining possibilities, one multiplies through by a multiplicative inverse of p_k to see membership in one of the three remaining cusps.

7. *Further remarks*

We note that Theorem 3 remains true upon replacing the geodesic flow by the flow defined by any one-parameter subgroup of G; if this new flow is represented by the right multiplication by matrices of the form F_t, then Lemma 1 and the theorem simply require the replacement of the E_t by these F_t. Of course the ergodicity of the skew product transformation, as announced in Corollary 1, depends on the flow at hand. For the (stable and unstable) horocycle flows, results of Hedlund [**Hed36**] and later of Dani and Smillie [**DS84**] can be invoked, so ergodicity (and even essential unique ergodicity; that is, unique ergodicity after the suppression of weights on the collection of periodic orbits) of the skew product again holds.

Acknowledgements. Fisher was partially supported by FAPESP and by CAPES-Cofecub 661/10. We thank the anonymous referee for a careful reading.

References

[AF91] R. L. Adler and L. Flatto. Geodesic flows, interval maps and symbolic dynamics. *Bull. Amer. Math. Soc.* **25** (1991), 229–334.

[AH00] P. Arnoux and P. Hubert. Fractions continues sur les surfaces de Veech. *J. Anal. Math.* **81** (2000), 35–64.

[Arn94] P. Arnoux. Le codage du flot géodésique sur la surface modulaire. *Enseign. Math.* **40** (1994), 29–48.

[Art24] E. Artin. Ein mechanisches System mit quasiergodischen Bahnen. *Abh. Math. Semin. Univ. Hambg.* **3** (1924), 170–175; reproduced in *Collected Papers*. Springer, New York, 1982, pp. 499–505.

[AS09] P. Arnoux and T. A. Schmidt. Veech surfaces with nonperiodic directions in the trace field. *J. Mod. Dyn.* **3**(4) (2009), 611–629.

[AS13] P. Arnoux and T. A. Schmidt. Cross sections for geodesic flows and α-continued fractions. *Nonlinearity* **26** (2013), 711–726.

[BKS00] R. M. Burton, C. Kraaikamp and T. A. Schmidt. Natural extensions for the Rosen fractions. *Trans. Amer. Math. Soc.* **352**(3) (2000), 1277–1298.

[BS04] E. Bogomolny and C. Schmit. Multiplicities of periodic orbit lengths for non-arithmetic models. *J. Phys. A: Math. Gen.* **37** (2004), 4501–4526.

[CM00] C.-H. Chang and D. Mayer. Thermodynamic formalism and Selberg's zeta function for modular groups. *Regul. Chaotic Dyn.* **5**(3) (2000), 281–312.

[DKS09] K. Dajani, C. Kraaikamp and W. Steiner. Metrical theory for α-Rosen fractions. *J. Eur. Math. Soc. (JEMS)* **11**(6) (2009), 1259–1283.

[DS84] S. G. Dani and J. Smillie. Uniform distribution of horocycle orbits for Fuchsian groups. *Duke Math. J.* **51**(1) (1984), 185–194.

[GH96] K. Gröchenig and A. Haas. Backward continued fractions, Hecke groups and invariant measures for transformations of the interval. *Ergod. Th. & Dynam. Sys.* **16**(6) (1996), 1241–1274.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press

[Hed35] G. A. Hedlund. A metrically transitive group defined by the modular groups. *Amer. J. Math.* **57**(3) (1935), 668–678.

[Hed36] G. A. Hedlund. Fuchsian groups and transitive horocycles. *Duke Math. J.* **2** (1936), 530–542.

[HMTY08] E. Hanson, A. Merberg, C. Towse and E. Yudovina. Generalized continued fractions and orbits under the action of Hecke triangle groups. *Acta Arith.* **134**(4) (2008), 337–348.

[Hop71] E. Hopf. Ergodic theory and the geodesic flow on surfaces of constant negative curvature. *Bull. Amer. Math. Soc.* **77** (1971), 863–877.

[JL88] H. Jager and P. Liardet. Distributions arithmétiques des dénominateurs de convergents de fractions continues. *Indag. Math.* **50**(2) (1988), 181–197.

[KSS12] C. Kraaikamp, T. A. Schmidt and W. Steiner. Natural extensions and entropy of α-continued fractions. *Nonlinearity* **25** (2012), 2207–2243.

[KU07] S. Katok and I. Ugarcovici. Symbolic dynamics for the modular surface and beyond. *Bull. Amer. Math. Soc.* **44**(1) (2007), 87–132.

[Leh66] J. Lehner. *A Short Course in Automorphic Functions*. Holt, Rinehart and Winston, New York, 1966.

[Leu67] A. Leutbecher. Über die Heckeschen Gruppen $\mathfrak{G}(\lambda)$. *Abh. Math. Semin. Univ. Hambg.* **31** (1967), 199–205.

[Man91] A. Manning. Dynamics of geodesic and horocycle flows on surfaces of constant negative curvature. *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces (Trieste, 1989)*. Oxford University Press, New York, 1991, pp. 71–91.

[MM10] D. Mayer and T. Mühlenbruch. Nearest λ_q-multiple fractions. *Spectrum and Dynamics (CRM Proceedings & Lecture Notes, 52)*. American Mathematical Society, Providence, RI, 2010, pp. 147–184.

[Moe82] R. Moeckel. Geodesics on modular surfaces and continued fractions. *Ergod. Th. & Dynam. Sys.* **2** (1982), 69–83.

[MS08] D. Mayer and F. Strömberg. Symbolic dynamics for the geodesic flow on Hecke surfaces. *J. Mod. Dyn.* **2**(4) (2008), 581–627.

[Nak81] H. Nakada. Metrical theory for a class of continued fraction transformations and their natural extensions. *Tokyo J. Math.* **4**(2) (1981), 399–426.

[Nak89] T. Nakanishi. A remark on R. Moeckel's paper: Geodesics on modular surfaces and continued fractions. *Ergod. Th. & Dynam. Sys.* **9**(3) (1989), 511–514.

[Nak10] H. Nakada. On the Lenstra constant associated to the Rosen continued fractions. *J. Eur. Math. Soc. (JEMS)* **12**(1) (2010), 55–70.

[NIT77] H. Nakada, S. Ito and S. Tanaka. On the invariant measure for the transformations associated with some real continued-fractions. *Keio Engrg. Rep.* **30**(13) (1977), 159–175.

[Pol11] M. Pollicott. Dynamical zeta functions. *Integers* **11B** (2011), article no. A11.

[Ros54] D. Rosen. A class of continued fractions associated with certain properly discontinuous groups. *Duke Math. J.* **21** (1954), 549–563.

[Ser85] C. Series. The modular surface and continued fractions. *J. Lond. Math. Soc.* **31** (1985), 69–80.

[She93] M. Sheingorn. Continued fractions and congruence subgroup geodesics. *Number Theory with an Emphasis on the Markoff Spectrum (Provo, UT, 1991) (Lecture Notes in Pure and Applied Mathematics, 147)*. Dekker, New York, 1993, pp. 239–254.

[SS95] T. A. Schmidt and M. Sheingorn. Length spectra of the Hecke triangle groups. *Math. Z.* **220**(3) (1995), 369–397.

[SU10] J. Smillie and C. Ulcigrai. Geodesic flow on the Teichmüller disk of the regular octagon, cutting sequences and octagon continued fractions maps. *Dynamical Numbers – Interplay between Dynamical Systems and Number Theory (Contemporary Mathematics, 532)*. American Mathematical Society, Providence, RI, 2010, pp. 29–65.

[SU11] J. Smillie and C. Ulcigrai. Beyond Sturmian sequences: coding linear trajectories in the regular octagon. *Proc. Lond. Math. Soc.* (3) **102**(2) (2011), 291–340.

[Vee84] W. A. Veech. The metric theory of interval exchange transformations I, II, III. *Amer. J. Math.* **106** (1984), 1331–1422.

https://doi.org/10.1017/etds.2013.23 Published online by Cambridge University Press