RIMS Kôkyûroku Bessatsu
B58 (2016), 1–18

Continued Fractions: Old and New

By

Thomas A. SCHMIDT*

Contents

§ 1. Overview

In this expository note, we attempt to point out some interesting recent developments in the theory and applications of continued fractions. The level of exposition is naturally uneven, with tendency towards informality a necessity.

One aspect of the Tambara lecture that we kept in these short notes is the use of the natural extension (without measure theory) to derive diophantine results for continued fractions. As far as the author knows, this approach was originated by Jager and Kraaikamp [14]. We also hint at the wide variety of settings for continued fractions by mentioning continued fractions

Received October 31, 2015. Revised February 16, 2016. Accepted May 1, 2016.
2010 Mathematics Subject Classification(s): 11K50, 37P30, 37M25
Key Words: *continued fractions, Diophantine approximation, cross-sections of geodesic flow.*
*Department of Mathematics, Oregon State University, Corvallis, OR 97331 USA
e-mail: toms@math.orst.edu

over function fields, semi-regular continued fractions, and continued fractions related to various Fuchsian groups. We end with a brief overview of an approach of Arnoux and co-author giving a heuristic method for determining invariant measures, by way of natural extensions, for certain interval maps. Here we emphasize the connection of this to cross-sections for the geodesic flow on the unit tangent bundle of a corresponding hyperbolic surface.

As in our Tambara lecture, we give a quick introduction to the basics of continued fractions; there are of course many excellent textbooks where this material is laid out in detail. Depending upon where one is beginning, besides the classic Hardy-Wright text, one can recommend [9, 11, 12, 26, 30, 32] and an expository article by Lachaud [17].

I thank the organizer of the Tambura workshop, Professor Shigeki Akiyama, the participants of the workshop, and also the various colleagues from whom I learned of the beauty of this part of mathematics.

§ 2. Basics of regular continued fractions

§ 2.1. Just the Basics

The Euclidean algorithm leads very naturally to the idea of continued fractions. An integral step in that algorithm is to invert and then take the integer part, the related interval map is the following.

$$T(x) = \begin{cases} \frac{1}{x} - \lfloor \frac{1}{x} \rfloor & \text{if } x \neq 0; \\ \\ 0 & \text{when } x = 0 \end{cases}$$

for $x \in \mathbb{I} := [0, 1)$. For $n \geq 1$, let $a_n(x) = \lfloor 1/T^{n-1}(x) \rfloor$.

Further define

$$[\, a_1,\, a_2, \ldots,\, a_n, \ldots] := [\, a_1(x),\, a_2(x), \ldots,\, a_n(x), \ldots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}\,.$$

The *convergents* 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

$$\begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}$$

for $n \geq 1$. From this definition it is immediate that $p_{n-1}q_n - q_{n-1}p_n = (-1)^n$ — from which, we clearly have that p_n/q_n is a reduced rational number, and that the recurrence relations

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

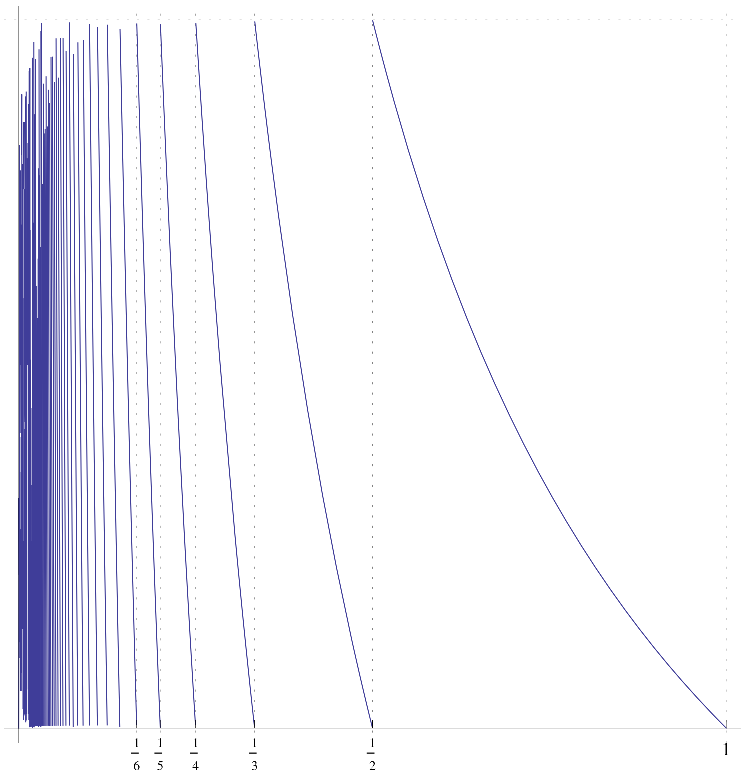

Figure 1. The continued fraction map T.

hold. It also follows that

$$\frac{p_n}{q_n} = [\,a_1,\, a_2, \ldots,\, a_n] \tag{2.1}$$

and, the *mirror formula*:

$$\frac{q_{n-1}}{q_n} = [\,a_n,\, a_{n-1}, \ldots,\, a_1]\,. \tag{2.2}$$

We define

$$M_n = \begin{pmatrix} p_{n-1} \, p_n \\ q_{n-1} \, q_n \end{pmatrix},$$

and, related to this find

$$x = \frac{p_{n-1}\,T^n(x) + p_n}{q_{n-1}\,T^n(x) + q_n}\,. \tag{2.3}$$

One calls the quantities denoted a_n above the *partial quotients*, and the quantities $x_n = 1/T^{n-1}(x)$ the *complete quotients* ($n \geq 2$). With an abuse of notation that we will take as standard, we thus have

$$x = [a_1, \ldots, a_{n-1} + T^{n-1}(x)] = [a_1, \ldots, a_{n-1}, x_n]\,. \tag{2.4}$$

Example 2.1. The definitions easily extend to give a notion of the simple continued fraction expansion of any real x by letting $\tilde{T} : \mathbb{R} \to [0, 1)$ agree with T on $[0, 1)$ and otherwise to act as $x \mapsto x - \lfloor x \rfloor$. One defines a notation

$$[a_0; a_1, \ldots, a_n, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} .$$

With this notation, if $x > 1$, then the nth convergent p_n/q_n of $1/x$ is the reciprocal of the $(n-1)$st convergent of x, the simple continued fraction expansion of $\pi = 3.14159265\ldots$ is computed as in Table 1.

n	a_n	p_n	q_n
0	3	3	1
1	7	22	7
2	15	333	106
3	1	355	113
4	292	103993	33102
5	1	104348	33215
6	1	208341	66317

Table 1. Approximating π

Thus, we find a sequence of rationals

$$\frac{3}{1}, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \ldots$$

whose decimal equivalents are

$$3, 3.14\cdots, 3.1415\cdots, 3.1415929\cdots, \ldots .$$

§ 2.2. The Legendre Constant

Using Equation (2.3),

$$x - \frac{p_n}{q_n} = \frac{1}{q_n} \, \frac{(-1)^n}{q_{n+1} + q_n \, T^{n+1}(x)} .$$

Since $0 \le T^{n+1}(x) \le 1$ and the various q_j are positive, we easily find that

$$\frac{1}{q_n(q_n + q_{n+1})} \le \left| x - \frac{p_n}{q_n} \right| \le \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2} . \tag{2.5}$$

Remark.

1. We find that for $x \notin \mathbb{Q}$ the convergents p_n/q_n do indeed converge to x, and that consecutive convergents lie on opposite sides of x itself.

2. Since $q_{n+1} = a_{n+1}q_n + q_{n-1}$, the first upper bound of (2.5) implies that if a_n is very large, then p_n/q_n will be a very good approximation of x. This is the case with the approximation of π by 355/113, as $a_4 = 292$.

We've just seen that in some sense the *convergents* p_n/q_n do very well at approximating the real number x.

Definition 2.2. A *best approximation* to a real number x is a rational number a/b such that whenever $0 < d \le b$ and $c/d \ne a/b$, one has

$$|\, dx - c \,| > |\, bx - a \,| .$$

Thus, a best approximation fraction a/b better approximates x than any fraction with denominator at most b. We prove that every convergent is a best approximation.

Proposition 2.3. *Fix $\xi \in (0,1)$, and suppose that $a/b \in \mathbb{Q}$ with $b > 0$. Then for each $n \ge 1$, $|\xi b - a| < |q_n\xi - p_n|$ implies $b \ge q_{n+1}$. Furthermore, $|\xi - a/b| < |\xi - p_n/q_n|$ implies $b > q_n$.*

The following is a key result and is proved in various textbooks.

Theorem 2.4 (Legendre). *If $x \in (0,1)$ is irrational and $p/q \in \mathbb{Q}$ with*

$$\left| x - \frac{p}{q} \right| < \frac{1}{2q^2} ,$$

then p/q is a convergent of x.

See the text of Rockett and Szüsz [27] for a discussion of Huygen's use of continued fractions for the calculations necessary in the construction of an early mechanical model of the solar system.

§ 2.3. Results of Lagrange and Galois

We return to the setting of $[a_0; a_1, \ldots]$. An infinite simple continued fraction is called *periodic* if it can be written in the form $[b_0; \ldots, b_j, \overline{a_0, a_1, \ldots, a_{n-1}}\,]$ where the bar indicates that the sequence of partial quotients is repeated.

Theorem 2.5 (Lagrange). *Each real quadratic number has a periodic simple continued fraction expansion.*

The condition characterizing numbers of purely periodic expansions was found by Galois.

Theorem 2.6 (Galois). *The simple continued fraction expansion of the real quadratic irrationality ξ is purely periodic if and only if $\xi > 1$ and $-1 < \xi' < 0$, where ξ' denotes the conjugate of ξ.*

§ 2.4. Equivalent numbers

Recall that the group $\mathrm{GL}_2(\mathbb{R})$ acts on $\mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$ by fractional linear transformations: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \dfrac{ax+b}{cx+d}$, with usual conventions when $x = -d/c$ or $x = \infty$. The action is projective: M and λM act the same for any $\lambda \neq 0$.

We say that two real numbers are *equivalent* if they are in the same $\mathrm{GL}_2(\mathbb{Z})$-orbit. The following result is classic, and lies behind many results about continued fractions.

Theorem 2.7. *Two irrational numbers are $GL_2(\mathbb{Z})$-equivalent if and only if their simple continued fraction expansions eventually agree.*

If we ask for the perhaps more natural equivalence under $\mathrm{SL}_2(\mathbb{Z})$ then the result is only true under the condition that the expansions eventually agree beginning at partial convergents whose indices have the same parity. This parity condition is the root cause of various phenomena, for instance that we find a *double* cover of the natural extension of the Gauss map as the cross-section for geodesic flow on the unit tangent bundle of the modular surface, see Remark 4.3. Recently Nakada-Natsui [25] have compared the group-orbits with the relation of eventual expansion agreement for a certain infinite family of continued fraction algorithms (see below for these α-continued fractions).

§ 2.5. Diophantine properties using an invertible map

The interval map $T : [a_1, a_2, \dots] \mapsto [a_2, a_3, \dots]$ shifts sequences to the "future" and is far from being invertible. We associate to it a map that also records the "past". Let

$$\mathcal{T}(x, y) = \left(T(x), \frac{1}{a_1(x) + y} \right).$$

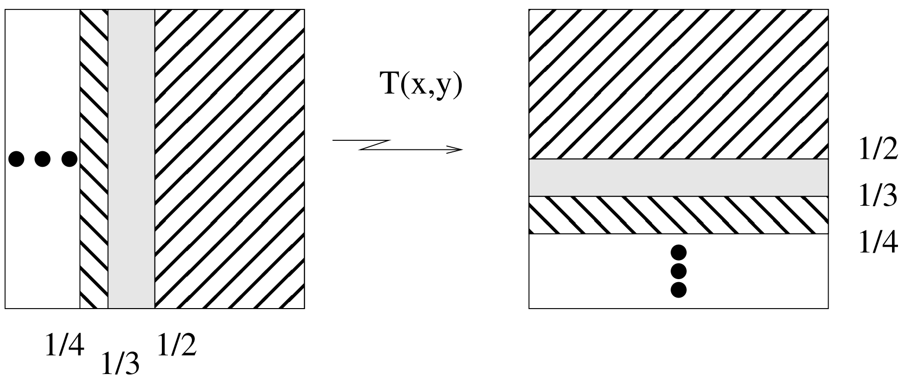

Figure 2. $\mathcal{T}(x, y) : (x, y) \mapsto \left(T(x), \dfrac{1}{a_1(x) + y} \right)$

Definition 2.8. For each $x \in (0, 1)$, define the sequence of points $(t_n, v_n) = \mathcal{T}^n(x, 0)$. Also, define the *coefficient of Diophantine approximation* to x for a rational p/q to be

$$\Theta(x, p/q) = q^2 \mid x - p/q \mid .$$

Then the n-th *coefficient of Diophantine approximation* for x is defined to be

$$\Theta_n = \Theta(x, p_n/q_n) = q_n^2 \mid x - p_n/q_n \mid \ .$$

The letters t_n, v_n are to make us think of "toekomst" and "verleden", Dutch words for "future" and "past". Note that $v_n = 1/(a_n + 1/(a_{n-1} + 1/(\cdots + 1/a_1)\ldots)) = q_{n-1}/q_n$.

Lemma 2.9. *Let $x \in (0,1)$, then*

$$\Theta_n = \frac{t_n}{1 + v_n t_n} = \frac{v_{n+1}}{1 + v_{n+1} t_{n+1}}$$

Proof. Recall from (2.3) that $M_n \cdot T^n(x) = x$. Substituting this into the definition of Θ_n easily gives the first equality. For the second, we substitute $M_{n+1} \cdot T^{n+1}(x) = x$ into this same definition. $\square$

Definition 2.10. Let $\Omega = [0,1] \times [0,1]$ and

$$\begin{aligned} \Psi : \quad & \Omega \to \mathbb{R}^2 \\ (x,y) \mapsto & \left(\frac{y}{1+xy}, \frac{x}{1+xy} \right) \end{aligned}$$

Note that this gives for any irrational x and all $n \geq 1$ that $\Psi(t_n, v_n) = (\Theta_{n-1}, \Theta_n)$.

Lemma 2.11. *(Vahlen 1895) Let $x \in (0,1)$, then for all $n \geq 1$ we have*

$$\min(\Theta_{n-1}, \Theta_n) < \frac{1}{2}$$

Proof. We easily find that $\Psi(\Omega)$ is the triangle with vertices $(0,0), (1,0)$ and $(0,1)$, and that this map is a bijection. Since every (Θ_{n-1}, Θ_n) defines a point in this region, the result clearly holds. $\square$

Theorem 2.12. *(Borel 1903) Let $x \in (0,1)$, then for all $n \geq 1$ we have*

$$\min(\Theta_{n-1}, \Theta_n, \Theta_{n+1}) \leq \frac{1}{\sqrt{5}},$$

and this constant is best possible.

Proof. Key here is the simple observation that since $\mathcal{T}(t_n, v_n) = (t_{n+1}, v_{n+1})$, we have $(\Theta_n, \Theta_{n+1}) = \Psi \circ \mathcal{T} \circ \Psi^{-1}(\Theta_{n-1}, \Theta_n)$. We thus aim to show that the image under $\Psi \circ \mathcal{T} \circ \Psi^{-1}$ of the triangle $\{(x,y) \in \Psi(\Omega) \mid x > 1/\sqrt{5}, y > 1/\sqrt{5}\}$ lies outside of itself. (This triangle is the complement of the region where both $\Theta_{n-1}, \Theta_n \leq 1/\sqrt{5}$.) Equivalently, we show that the image under $\mathcal{T}$ of

$$\mathcal{D} = \{(x,y) \in \Omega \mid \sqrt{5} - 1/x > y > 1/(\sqrt{5} - x)\}$$

lies outside of $\mathcal{D}$ itself. Letting $g = (-1 + \sqrt{5})/2$, one easily verifies that (g,g) is a fixed point of $\mathcal{T}$ and also that this point lies on the internal "vertex" of $\mathcal{D}$. Since $\mathcal{T}(x,y) = (T(x), 1/(a_1 + y))$

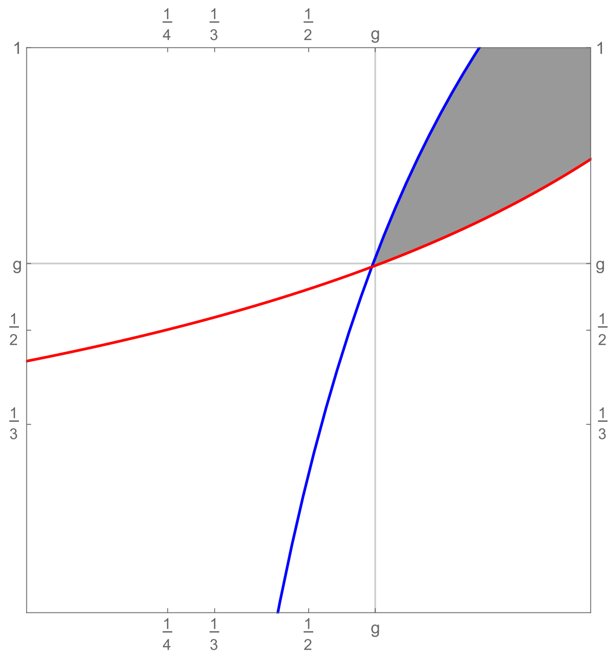

Figure 3. The region $\mathcal{D}$

is height reversing on each set with constant a_1, and as $\mathcal{D}$ lies completely in the region where $a_1 = a_1(x) = 1$, we find that $\mathcal{T}(\mathcal{D})$ lies both to the left of $x = g$ and below $y = g$, and thus has no intersection with $\mathcal{D}$. Thus, since no point $P \in \mathcal{D}$ is such that also $\mathcal{T}(P) \in \mathcal{D}$, we certainly have that no $\Theta_{n-1}, \Theta_n, \Theta_{n+1}$ can all be greater than $1/\sqrt{5}$.

To see that $1/\sqrt{5}$ is best possible, consider the coefficients of Diophantine approximation for $x = g$. One easily verifies that for $n \geq 1$, $(t_n, v_n) = \mathcal{T}^n(g, 0) = (g, F_{n-1}/F_n)$, where F_n is the nth Fibonacci number, indexed so that $F_0 = F_1 = 1$, $F_2 = 2, \ldots$. Thus, the points $\mathcal{T}^n(g, 0) = (g, F_{n-1}/F_n)$ have (g, g) as their unique limit point, and this is a limit point from both above and below (on the line $x = g$). But, a decrease in the bound $1/\sqrt{5}$ would result in replacing $\mathcal{D}$ by a larger region $\mathcal{D}'$, and in particular one that included (g, g) as an interior point. This fixed point would remain in the region $\mathcal{T}(\mathcal{D}')$ as would the preimages of (Θ_{n-1}, Θ_n), the (t_n, v_n), for n sufficiently large. □

Combining the result of Borel with Legendre's Theorem, we have a result of Hurwitz.

Corollary 2.13. *(Hurwitz) For every irrational number there exists infinitely many pairs of integers p and q such that*

$$\left|x - p/q\right| \leq \frac{1}{\sqrt{5}}\,\frac{1}{q^2}\,.$$

Furthermore, the constant $1/\sqrt{5}$ is best possible.

Proof. The Borel result immediately implies the first statement. To prove that $1/\sqrt{5}$ is optimal, we note that Legendre's result shows that any p/q that approximate x so well are indeed convergents; therefore, the sequence $\mathcal{T}^n(g, 0)$ also shows here the optimality of this value. □

§ 3. Other continued fractions

§ 3.1. Over function fields

The utility of continued fractions as a tool for investigating the arithmetic of real quadratic fields was well known to Gauss. They also provide a simple but effective tool in the setting of hyperelliptic function fields. We merely quickly state a result to hint at this; for background, see say the text [33].

$\mathbb{Z}$	$k[x]$
$\mathbb{Q}$	$k(x)$
infinite place	$\mathcal{O}_\infty = \{\frac{f}{g} \mid f(x), g(x) \in K[x], \deg f \leq \deg g\}$
$\mathbb{R}$	$k((x^{-1}))$
$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}$	$A_0(x) + \cfrac{1}{A_1(x) + \cfrac{1}{A_2(x) + \cdots}}$

A hint of the analogies that lead to continued fractions over function fields.

Set $\mid \alpha \mid := e^m$, when $\alpha = c_m x^m + c_{m-1}x^{m-1} + \cdots + c_0 + c_1 x^{-1} + \cdots$. The analog of the Legendre result is now.

Theorem 3.1. *Suppose that $\alpha \in K((x^{-1})) \setminus K(x)$ and $P, Q \in K[x]$ such that*

$$\mid \alpha - P/Q \mid < 1/ \mid Q \mid^2 .$$

Then P/Q is a convergent P_n/Q_n of α.

The arithmetic of these continued fractions have a geometric interpretation. The following result is attributable to, in part: Abel, Chebyshev, Adams-Razar, van der Poorten, W. Schmidt. See [1], [34], [29].

Theorem 3.2. *Suppose that F/K is a hyperelliptic function field given by $y^2 = D(x)$, and let P, Q be the two places lying over the place at infinity of $K(x)$. Then the divisor $P - Q$ has finite order in the divisor class group of F/K if and only if y has a periodic CF expansion. In this case, the divisor at infinity D_∞ is torsion of order*

$$N = g + 1 + \sum_{i=1}^{m-1} (\deg A_i) ,$$

where m is the quasi-period length of this expansion.

Furthermore, $1 \le \deg a_i \le g$ for each $1 \le i \le m-1$. In her 2013 Oregon State University Ph.D. thesis, K. Daowsud was able to improve this bound.

§ 3.2. Semi-regular CF

Semi-regular continued fractions:

$$[\,b_0; \varepsilon_1 b_1,\, \varepsilon_2 b_2, \ldots,\, \varepsilon_n b_n, \ldots] = b_0 + \cfrac{\varepsilon_1}{b_1 + \cfrac{\varepsilon_2}{b_2 + \cdots}}\,,$$

with $\varepsilon_i = \pm 1$; $b_0 \in \mathbb{Z}$; $b_n \in \mathbb{N}$ for $n > 0$, with

$$b_n + \varepsilon_{n+1} \ge 1 \;\text{ for }\; n \ge 1,$$

and $b_n + \varepsilon_{n+1} \ge 2$ infinitely often.

Among the semi-regular continued fractions are the *nearest integer* continued fractions, already studied in the 19th century, related to the interval map $T_{1/2}(x) = 1/|x| - \lfloor 1/|x| + 1/2 \rfloor$. For this map one finds that $b_n \ge 2$ and $b_n + \varepsilon_{n+1} \ge 2$ holds for all $n \ge 1$. In fact, this property characterizes the nearest integer expansions.

In 1981 Nakada [22] defined a 1-parameter family of semi-regular continued fractions. These α-CF, where α is a parameter in $[0,1]$ have underlying map

$$T_\alpha : x \mapsto \frac{1}{|x|} - \left\lfloor \frac{1}{|x|} + 1 - \alpha \right\rfloor,$$

with 0 sent to 0.

Kraaikamp [16] found a way to study a class of semi-regular continued fractions, including some of the α-CF. His notion of S-continued fractions is based upon the following.

Lemma 3.3. *Suppose that for some $\ell \ge 1$ one has*

$$b_{\ell+1} = 1 \;\text{ and }\; \varepsilon_{\ell+1} = \varepsilon_{\ell+2} = 1\,.$$

Then $[\,b_0; \varepsilon_1 b_1,\, \varepsilon_2 b_2, \ldots,\, \varepsilon_\ell \mathbf{b}_\ell, \mathbf{1}, \mathbf{b}_{\ell+2}, \ldots]$

$$= [\,b_0; \ldots,\, \varepsilon_{\ell-1} b_{\ell-1}, \varepsilon_\ell(\mathbf{b}_\ell + \mathbf{1}), -(\mathbf{b}_{\ell+2} + \mathbf{1}),\, \varepsilon_{\ell+3} b_{\ell+3},\, \varepsilon_{\ell+4} b_{\ell+4}, \ldots]$$

Proof. One trivially checks that following: For any a, b positive integers and $\xi \in (-1,1)$

$$a + \cfrac{1}{1 + \cfrac{1}{b+\xi}} = a + 1 + \frac{-1}{b+1+\xi}\,.$$

But, $[\,b_\ell; \varepsilon_{\ell+1} b_{\ell+1}, \varepsilon_{\ell+2} b_{\ell+2} + \xi]$ has the form of the left hand side of the identity. □

Definition 3.4. Given some semi-regular continued fraction expansion $[\,b_0; \varepsilon_1 b_1,\, \varepsilon_2 b_2, \ldots,\, \varepsilon_\ell b_\ell, \ldots]$, and some $\ell \ge 1$ such that $b_{\ell+1} = 1$ and $\varepsilon_{\ell+1} = \varepsilon_{\ell+2} = 1$, we call the expansion

$$[\,b_0; \varepsilon_1 b_1,\, \varepsilon_2 b_2, \ldots,\, \varepsilon_{\ell-1} b_{\ell-1}, \varepsilon_\ell(b_\ell + 1), -(b_{\ell+2} + 1),\, \varepsilon_{\ell+3} b_{\ell+3},\, \varepsilon_{\ell+4} b_{\ell+4}, \ldots]$$

the *singularization* of the partial quotient $b_{\ell+1} = 1$.

From the lemma, we see that singularization skips the regular continued fraction convergents that at the worst approximations (due to the *next* partial quotient being equal to 1). Thus, one can think of singularization as resulting in a sequence of better approximations.

Kraaikamp's insight was that one can control how to singularize by using the two dimensional system. Fix $S \subset [1/2, 1) \times [0, 1]$ such that $\mathcal{T}(S) \cap S = \emptyset$, where again $\mathcal{T}(x, y) = \left(T(x), \dfrac{1}{a_1(x) + y} \right)$ is the 2-D Gauss map. Then singularize a_{n+1} if and only if $\mathcal{T}^n(x, 0) \in S$. This results in a large class of CF-algorithms, including the Nakada α-CF for $g \leq \alpha < 1$. Both ergodic theoretic (in particular entropy values) and approximation properties are accessible with this approach.

Fairly recently, a family of continued fractions was introduced by S. Katok-Ugarcovici [15], whose ergodic theoretic properties they studied in a series of papers. Let $a, b \in \mathbb{R}$ with $a \leq 0 \leq b$, $b - a \geq 1$, $-ab \leq 1$, they define $\hat{f}_{a,b}$ as the first return map of $f_{a,b}$ to the interval $[a, b)$ of

$$f_{a,b}(x) = \begin{cases} x + 1 & \text{if } x < a \\ -1/x & \text{if } a \leq x < b \\ x - 1 & \text{if } x \geq b\,. \end{cases}$$

§ 3.3. Other groups

See works especially of Bowen-Series, Series and Adler-Flatto for continued fractions related to various Fuchsian groups. Here we mention a few other examples of these.

Let us first remark that already the basic results of continued fractions can be difficult to fully resolve in these more general settings. Whereas there are usually simple geometric interpretations of finite and periodic expansions (these usually correspond to orbits of cusps and classes of closed geodesics, respectively), it seems nearly impossible to arithmetically exactly describe these, for remarks on some of this, see the work of Leutbecher [18]. A very interesting result of Nakada [23] is that if an analog of the Legendre constant exists for such continued fraction algorithms, then it can actually be calculated using ergodic theory; see [8] for the antecedents of this. Diophantine approximation results using the analog of the $\Theta_n(x)$ are also usually available.

The Hecke triangle Fuchsian group of index q is

$$G_q = \langle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda_q \\ 0 & 1 \end{pmatrix} \rangle =: \langle S, T \rangle\,,$$

with $\lambda_q = 2\cos \pi/q$.

Determining whether a given element of $\mathrm{PSL}_2(\mathbb{Z}[\lambda_q]$ belongs to G_q is in general not straightforward. To address this word problem, in 1954 D. Rosen [28] introduced his λ-CF. The underlying interval map on $[-\lambda_q/2, \lambda_q/2)$ sends nonzero x to $1/x - \lfloor x/\lambda_q + 1/2 \rfloor \lambda_q$. This continued fraction expansion and variants have been studied from various perspectives over the years.

Nearly 50 years later, Arnoux and Hubert introduced a continued fraction expansion that is directly related to an index two subgroup of G_q, for each even $q \geq 6$. The expansion is in fact

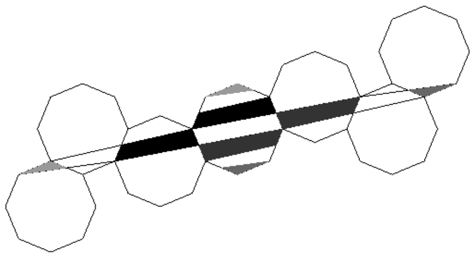

Figure 4. A "parabolic" affine diffeomorphism of the octagon surface.

in terms of directions for linear flow on the (so-called translation) surfaces formed by identifying the opposite sides of regular even-sided polygons. The underlying geometry of the situation was determined by Veech [35], see Figure 4 for a hint of some of the so-called affine diffeomorphisms involved. A comparison, using the method of Section 4, is made of these continued fractions with the Rosen continued fractions in [5].

For another large family of continued fractions, one can generalize the continued fractions introduced in [10] in the following fashion. Set $\mu = 2\cos\pi/m, \nu = -2\cos\pi/n, t = \mu + \nu$ and let

$$A = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \; B = \begin{pmatrix} \nu & 1 \\ -1 & 0 \end{pmatrix}, \; C = \begin{pmatrix} -\mu & 1 \\ -1 & 0 \end{pmatrix},$$

and note that $C = AB$.

$$\mathbb{I} := \mathbb{I}_{m,n,\alpha} = [(\alpha - 1)t, \alpha t]$$

Let

$$T = T_{m,n,\alpha} : x \mapsto A^k C^l \cdot x$$

where

- l is minimal such that $C^l \cdot x \notin \mathbb{I}$
- $$k = -\lfloor (C^l \cdot x)/\tau + 1 - \alpha \rfloor$$

As shown in [10], Diophantine results using the analog of the $\Theta_n(x)$ are reasonably reached for such families of continued fractions. Note that there one finds a setting in which the natural extension has infinite area if and only if the denominators of the convergents, the analogs of the q_n, are not always strictly increasing.

§ 4. Geodesic flow and continued fractions

In 1924, E. Artin used continued fractions to show that there exist dense paths under geodesic flow on the unit tangent bundle of the modular surface. In 1931, E. Hopf, see the reprisal [13], showed that the corresponding flow is ergodic for any finite covolume hyperbolic surface. Authors relating this back to continued fractions include: Moeckel [21], Adler-Flatto (see the bibliography of [2]), Series [31] and Arnoux [3].

Here we review work of Arnoux and joint work with Arnoux identifying a class of continued fractions maps each of which has an invariant measure so that the natural extension of this 1-dimensional dynamical system is a factor of a cross-section of the geodesic flow on the unit tangent space of a hyperbolic surface. For an introduction to the geodesic flow, see say Manning's contribution [20] to [7] .

§ 4.1. First return type

In [4], we give an example of an interval map with a model of its natural extension (double-covered by) a subset of the unit tangent bundle of a surface, but for which the first return map of the geodesic flow does *not* project so as to give the interval map. This is an example of a map that is not of first return type!

Definition 4.1. For $M \in \mathrm{SL}(2,\mathbb{R})$ and $x \in \mathbb{R}$ such that $M \cdot x \neq \infty$, let

$$\tau(M,x) := -2\,\log|cx + d\,|$$

where (c,d) is the bottom row of M as usual.

As usual, we consider the projective group $\mathrm{PSL}(2,\mathbb{R})$ in lieu of $\mathrm{SL}(2,\mathbb{R})$. Elementary calculation shows that τ induces on $\mathrm{PSL}(2,\mathbb{R})$ a cocycle, in the following sense:

$$\tau(MN,x) = \tau(M,Nx) + \tau(N,x)$$

whenever all terms are defined and we choose each projective representative such that the corresponding $cx+d$ is positive. (In all that follows, the set where such a $cx+d$ is zero can be avoided.)

Definition 4.2. Suppose that f is a piecewise Möbius interval map, say defined on an interval I, with $I = \bigcup I_\alpha$ and for each α, f on I_α given by $x \mapsto M_\alpha \cdot x$. For each $x \in I$, the *return time* of x is $\tau_f(x) := \tau(M_\alpha\,,\,x)$. Finally, let Γ_f be the group generated by the set of the M_α.

For $M \in \mathrm{SL}(2,\mathbb{R})$, let

$$\mathcal{T}_M : (x,y) \to (\,M \cdot x, (cx+d)^2 y - c(cx+d)\,)\,. \tag{4.1}$$

An elementary calculation shows that the Jacobian matrix of $\mathcal{T}_M$ has determinant one. Thus, $\mathcal{T}_M$ is Lebesque measure preserving on $\mathbb{R}^2$. Recall that a Fuchsian group is a discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$ (or of $\mathrm{PSL}(2,\mathbb{R})$, depending upon context).

Definition 4.3. For f as above, on $I \times \mathbb{R}$ we have the piecewise defined map $\mathcal{T}_f$ given by taking each transformation $\mathcal{T}_{M_\alpha}$ above I_α. We say that *f has a positive planar model* if there is a compact set $\Omega_f \subset I \times \mathbb{R}$ also fibering over I, and of positive Lebesgue measure, say c_f, such that $\mathcal{T}_f(\Omega_f) = \Omega_f$. Such a set is unique, and $\mathcal{T}_f$ is then a measurable automorphism of Ω_f. We then refer to the marginal measure of $(1/c_f)\,dx\,dy$ (that is, the measure on I obtained by integrating along the fibres) simply as *the marginal measure.*

Further, we say that f is of *first return type* if:

1. f has a positive planar model;
2. Γ_f is a Fuchsian group;
3. for almost every $x \in I$ we have $\tau_f(x) > 0$; and,
4. for almost every $(x, y) \in \Omega_f$ and every non-trivial $M \in \Gamma_f$ with $\mathcal{T}_M(x, y) \in \Omega_f$ and $\tau(M, x) \geq 0$, we have $\tau_f(x) \leq \tau(M, x)$.

Remark. Recall that, with our usual notation, the derivative of $M \cdot x$ at x is $(cx + d)^{-2}$. The positivity of the values $\log 1/(cx+d)$ shows that any (non-trivial) piecewise fractional linear map of first return type is expanding almost everywhere.

§ 4.2. Unit tangent bundles as quotient groups

The group $\mathrm{SL}_2(\mathbb{R})$ acts transitively on the upper half-plane $\mathcal{H} = \{z = x+iy \,|\, x, y \in \mathbb{R}, y > 0\}$, allowing us to identify $\mathcal{H}$ with $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2$, by considering the orbit of $z = i$. Similarly, one can show that $\mathrm{SL}_2(\mathbb{R})$ acts transitively on the unit tangent vectors of $\mathcal{H}$, with the unit tangent vector directed vertically at basepoint i being stabilized by $\pm\mathrm{Id}$; that is, the unit tangent bundle can be identified with $\mathrm{PSL}_2(\mathbb{R})$.

Haar measure on $\mathrm{SL}_2(\mathbb{R})$ descends to this quotient group, agreeing with the usual Liouville measure (the product measure given by hyperbolic area times Lebesgue measure on S^1) on the unit tangent bundle. For simplicity, here we will define geodesic flow on this unit tangent bundle as the map $\mathbb{R} \times \mathrm{PSL}_2(\mathbb{R}) \to \mathrm{PSL}_2(\mathbb{R})$ given by $(t, \pm A) \mapsto \pm A g_t$, where g_t is the diagonal matrix entries on the diagonal are $e^{t/2}, e^{-t/2}$.

The left multiplication by $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ on $\mathrm{SL}_2(\mathbb{R})$ leads to the modular surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2 = \Gamma\backslash\mathcal{H}$, with its unit tangent bundle identified with $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$. The geodesic flow here is analogously defined.

§ 4.3. Arnoux's transversal

Consider the subset of $\mathrm{SL}_2(\mathbb{R})$,

$$G_\gamma^+ = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \,|\, \gamma > 0 \right\}.$$

A fairly easy discussion, see [4], leads to the following key lemma.

Lemma 4.4. *Let $\Sigma \subset G_\gamma^+$ be defined by $\gamma = 1$. Consider local coordinates x, y by letting each $A \in \Sigma$ be given as*

$$A = \begin{pmatrix} x & xy-1 \\ 1 & y \end{pmatrix}.$$

Then G_γ^+ has local coordinates (x, y, t) by way of

$$M = A\, g_t$$

with $A \in \Sigma$ and g_t as above. Furthermore, $dx\, dy\, dt$ gives Haar measure restricted to G_γ^+.

For an element $M \in \mathrm{SL}_2(\mathbb{R})$, let $[M]$ denote the corresponding element of $\mathrm{PSL}_2(\mathbb{R})$, also let G_γ be defined as G_γ^+, but with γ non-zero. Then since Haar measure $\mathrm{SL}_2(\mathbb{R})$ induces Liouville measure on $T^1\mathbb{H}$, and G_γ^+ projects one-to-one to a set of full measure in $\mathrm{PSL}_2(\mathbb{R})$, we have the following.

Theorem 4.5. *Under the identification of $T^1\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$, Liouville measure on the full measure set $\{\, [M] \mid M \in G_\gamma \,\}$ is proportional to $dx\, dy\, dt$ where x, y, t are as above.*

We now state and sketch the proof of a result showing that a certain class of interval maps does indeed arise from cross-sections of geodesic flow. For f of positive planar model Ω_f, with Γ_f Fuchsian of finite covolume, let $\Sigma_f = \{\, [A] \mid A = \left(\begin{smallmatrix} x & xy-1 \\ 1 & y \end{smallmatrix}\right) \text{ for } (x, y) \in \Omega_f \,\} \subset \Gamma_f \backslash \mathrm{PSL}(2, \mathbb{R})$.

Theorem 4.6 ([5])**.** *Suppose that f is a piecewise Möbius interval map with a positive planar model of its natural extension, the group Γ_f is Fuchsian of finite covolume, and f is a Markov map that is ergodic with respect to the marginal measure. Then f is of first return type if and only if the volume of the unit tangent bundle of the surface uniformized by Γ_f equals the product of the entropy of f with the Lebesgue measure of Σ_f.*

Proof. For ease of typography, let $\Sigma = \Sigma_f$ and let λ be the Lebsgue measure coming from Lemma 4.4.

By Rohlin's formula for the entropy of an ergodic interval map, see say [12], and the fact that locally $f(x) = (ax + b)/(cx + d)$, we have

$$\begin{aligned} h(T) &= \int_I \log |f'(x)|\, d\nu \\ &= \int_I -2 \log |cx + d|\, d\nu \\ &= \int_I \tau_f(x)\, d\nu \\ &= \frac{1}{c_f} \int_\Sigma \tau_f(x)\, dx\, dy \\ &= \frac{\int_\Sigma \tau_f(x)\, dx\, dy}{\lambda(\Sigma)} \\ &\geq \frac{\mathrm{vol}(T^1(\Gamma\backslash\mathbb{H}))}{\lambda(\Sigma)}\,. \end{aligned}$$

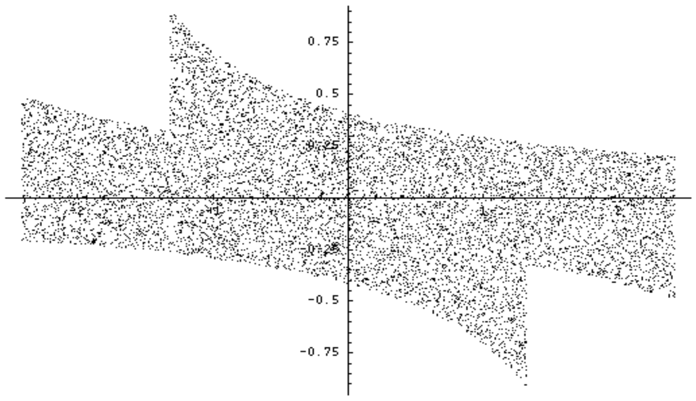

Figure 5. Plotting orbit to discover a natural extenion.

The final inequality holds since the *first* return map to Σ is given by following the unit tangent vectors along the geodesic arcs with unit tangent vectors in Σ until a first return to Σ. The fact that $\mathcal{A}$ is a transversal implies that these geodesic arcs sweep out the unit tangent bundle (up to measure zero). Hence, equality holds if and only if f is of first return type. $\square$

Remark. That the Gauss interval map of regular continued fractions is a factor of a cross-section of the geodesic flow on the unit tangent bundle of the modular surface does follow from the above result. However, one must use a *double cover*, since matrix acting as $x \mapsto 1/x - \lfloor x \rfloor$ has negative determinant and thus does not correspond to an element of $\mathrm{SL}(2,\mathbb{R})$. Similarly, the Nakada α-continued fractions are also given as factors of cross-sections; this was conjectured by Luzzi-Marmi [19] and proven in [4] by techniques similar to the above.

Remark. Given an interval map $f(x)$ that is piecewise Möbius, one can attempt to determine an invariant measure and then gain related ergodic theoretic information by the following heuristic method. Compute a fairly long orbit under $\mathcal{T}_f$ of Definition 4.3; this will often lie within a compact Ω_f, the boundaries of which one can quite likely guess from the orbit; such boundaries can be verified using the definition of $\mathcal{T}_f$ (often here it is easier to use the analog of the 2-D Gauss map of Subsection 2.5, as this will send rectangles to rectangles, at the cost of having an invariant measure other than 2-D Lebesgue measure). If this Ω_f is of positive measure (and its closure has this same measure), then it is a natural extension for f with the marginal measure on the interval (that is, with the measure induced by integrating out fibers).

As an indication of this approach, see Figure 5, giving the region Ω_f where f is as in Subsection 3.3 the Arnoux-Hubert map for the octagon.

References

[1] Adams, W. and Razar, M., Multiples of points on elliptic curves and continued fractions, *Proc. London Math. Soc.* (3) **41** (1980), no. 3, 481–498

[2] Adler, R. and Flatto, L., Geodesic flows, interval maps, and symbolic dynamics, *Bull. Amer. Math. Soc. (N.S.)* **25** (1991), no. 2, 229–334.

[3] Arnoux, P., Le codage du flot géodésique sur la surface modulaire, *Enseign. Math. (2)* **40** (1994), no. 1-2, 29–48.

[4] Arnoux, P. and Schmidt, T. A., Cross sections for geodesic flows and α-continued fractions, *Nonlinearity* **26** (2013), no. 3, 711–726.

[5] ______, Commensurable continued fractions, *Discrete Contin. Dyn. Syst.* **34** (2014), no. 11, 4389–4418.

[6] Artin, E., Ein mechanisches System mit quasi- ergodischen Bahnen, *Abh. Math. Sem. Hamburg* **3** (1924) 170–175 (and *Collected Papers*, Springer-Verlag, New York, 1982, 499–505).

[7] *Ergodic Theory, symbolic dynamics, and hyperbolic spaces*, Bedford, T., Keane, M. and Series, C., eds., Oxford Univ. Press, 1991.

[8] Bosma, W., Jager, H. and Wiedijk, F., Some metrical observations on the approximation by continued fractions, *Nederl. Akad. Wetensch. Indag. Math.* **45** (1983), no. 3, 281–299.

[9] Burger, E., *Exploring the Number Jungle*, Student Mathematical Library, **8**, AMS, 2000.

[10] Calta, K. and Schmidt, T. A., Continued fractions for a class of triangle groups, *J. Austral. Math. Soc.*, **93** (2012) 21–42.

[11] Chrystal, G. *Textbook of Algebra*, Black, Ltd., 1922.

[12] Dajani, K. and Kraaikamp, C., *Ergodic Theory of Numbers*, The Carus Mathematical Monographs, **29**, (2002) AMS.

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

[14] Jager, H. and Kraaikamp, C., On the approximation by continued fractions, *Nederl. Akad. Wetensch. Indag. Math.* **51** (1989), no. 3, 289–307.

[15] Katok, S. and Ugarcovici, I., Theory of (a, b)-continued fraction transformations and applications, *Electron. Res. Announc. Math. Sci.* **17** (2010), 20–33.

[16] Kraaikamp, C., A new class of continued fraction expansions, *Acta Arith.* **57** (1991), no. 1, 1–39.

[17] Lachaud, G., Continued fractions, binary quadratic forms, quadratic fields, and zeta functions, in: *Algebra and topology 1988* (Taejon, 1988), 1–56, Korea Inst. Tech., 1988.

[18] Leutbecher, A., Über die Heckeschen Gruppen $G(\lambda)$. II, *Math. Ann.* **211** (1974), 63–86.

[19] Luzzi, L. and Marmi, S., On the entropy of Japanese continued fractions, *Discrete Contin. Dyn. Syst.*, **20**, (2008), 673–711.

[20] Manning, A., Dynamics of geodesic and horocycle flows on surfaces of constant negative curvature, 71–91, in [7].

[21] Moeckel, R., Geodesics on modular surfaces and continued fractions, *Ergodic Theory Dynamical Systems* **2** (1982), no. 1, 69–83.

[22] Nakada, H., Metrical theory for a class of continued fraction transformations and their natural extensions, *Tokyo J. Math.*, **4**, no. 2, (1981), 399-426.

[23] ______, On the Lenstra constant associated to the Rosen continued fractions, *J. Eur. Math. Soc. (JEMS)* **12** (2010), no. 1, 55–70.

[24] Nakada, H., Ito, S. and Tanaka, S., On the invariant measure for the transformations associated with some real continued-fractions, *Keio Engrg. Rep.* **30** (1977), no. 13, 159–175.

[25] Nakada, H. and Natsui, R., On the equivalence relations of α-continued fractions, *Indag. Math. (N.S.)* **25** (2014), no. 4, 800–815.

[26] Niven, I., Zuckerman, H. and Montgomery, H., *Introduction to the theory of numbers*, 5th ed., Wiley (1991)

[27] Rockett, A. and Szüsz, P., *Continued fractions*, World Scientific, 1992.

[28] Rosen, D., A Class of Continued Fractions Associated with Certain Properly Discontinuous Groups, *Duke Math. J.* **21** (1954), 549–563.

[29] Schmidt, W., On continued fractions and diophantine approximation in power series fields, *Acta Arithm.* **95** (2000), 139–166.

[30] Schweiger, F., *Ergodic theory of fibred systems and metric number theory.* Oxford: Clarendon Press, 1995.

[31] Series, C., The modular surface and continued fractions, *J. London Math. Soc. (2)* **31** (1985), no. 1, 69–80.

[32] Stark, H., *An introduction to number theory*, Markham (1970)

[33] Stichtenoth, H., *Algebraic function fields and codes*, Springer, 1993.

[34] van der Poorten, A. and Tran, X., Quasi-elliptic integrals and periodic continued fraction, *Monatsh. Math.* **131** (2000), no. 2, 155–169.

[35] Veech, W., Teichmüller curves in moduli space, Eisenstein series, and an application to triangular billiards, *Inv. Math.* **97** (1989), 553 – 583.