ACTA ARITHMETICA
154.4 (2012)

Sequences and dynamical systems associated with canonical approximation by rationals

by

ANDREW HAAS (Storrs, CT)

1. Introduction. Each irrational $x \in (0,1)$ has a unique representation as an infinite regular continued fraction $[a_1, a_2, \ldots]$. By virtue of this representation, it is possible to associate to x the sequence of convergents $p_n/q_n = [a_1, \ldots, a_n]$ (cf. [12]). Define $\theta_n(x) = q_n|q_n x - p_n|$. We are interested in studying generic properties of certain subsequences of the sequences $\{p_n/q_n\}$, $\{\theta_n\}$ and $\{(\theta_n, \theta_{n+1})\}$. Our methods make use of ergodic theory and hyperbolic geometry and build upon Bosma, Jager and Wiedijk's proof of the Doeblin–Lenstra conjecture (referred to as D-L) [3, 17, 6] as well as Jager's subsequent treatment of the sequence of approximation pairs [13].

For $\alpha \in (0,1]$, define the subsequence $\bar{\Theta}(\alpha) = \{\theta_{n_k}\}$, where $\theta_{n_k} < \alpha$ and if $\theta_n < \alpha$ then $\theta_n = \theta_{n_k}$ for some k. There is of course the associated subsequence of convergents $\{p_{n_k}/q_{n_k}\}$. Denote these two subsequences by $\bar{\Theta}(\alpha) = \{\bar{\theta}_k\}$ and $\{\bar{p}_k/\bar{q}_k\}$. One interesting fact that is evident in our treatment is that, while for $\alpha \le 1/2$ and for almost all $x \in (0,1)$ the sequence $\bar{\Theta}(\alpha)$ is equidistributed in the interval $(0,\alpha)$, the corresponding sequence of pairs $\{\bar{\theta}_k, \bar{\theta}_{k+1}\}$ is far from evenly distributed in its natural domain. This was first observed by Kraaikamp [19] for $\alpha = 1/2$. For $\alpha \ge 1/2$ we derive the density function for the distribution. We also see how D-L and a famous theorem of Lévy on the growth of the numerator of convergents can be realized in this setting.

As in [3] and [13], the main tool is the natural automorphic extension $\mathbf{T}$ of the Gauss map. The subsequences are associated to a family of automorphisms defined by taking the first return to a region Ω_α. These first return maps are interesting in their own right and are related to Kraaikamp's theory of S-expansions and singularization [20, 6] as well as his treatment of Minkowski's diagonal continued fraction—the case when $\alpha = 1/2$. While $\mathbf{T}$ is Bernoulli and therefore qualifies as being chaotic, with decreasing α

2010 *Mathematics Subject Classification*: Primary 11K50, 37A45; Secondary 30F35.
Key words and phrases: Diophantine approximation, dynamical systems.

DOI: 10.4064/aa154-4-3

the first return maps exhibit increasing complexity and appear to do so in a manner reminiscent of the structure of the Markov spectrum [5]. Techniques from hyperbolic geometry will be employed to gain some understanding of the structure of the first return maps and to point the way for dealing with cases when $\alpha < 1/2$.

2. Basic properties of the sequences and the natural extension

2.1. The natural extension of the Gauss map. We begin by discussing the Gauss map, its natural automorphic extension $\mathbf{T}$ and a related group of first return maps. These maps are used to pinpoint the sequence of values n_k mentioned above as well as providing the framework for analyzing dynamical properties of the sequence of thetas and the pairs.

The classical Gauss map is defined on the open unit interval $I = (0,1)$ by $T(x) = 1/x - [1/x]$, where we use $[r]$ to denote the greatest integer less than or equal to r. The map has several nice properties. First, T acts as a shift on the continued fraction expansions: $T([a_1, a_2, \ldots]) = [a_2, a_3, \ldots]$. Secondly, T is ergodic with respect to Lebesgue measure and has the absolutely continuous invariant probability measure $(\log 2(1+x))^{-1}\,dx$ (see [6]).

There is a simple and useful realization of the natural automorphic extension $\mathbf{T}$ of T (cf. [15]) due to Nakada et al. [22, 6]. See also [1]. Let $J = (-\infty, -1)$. We shall use a closely related realization of $\mathbf{T}$, defined on $\Omega = I \times J$ by

$$\mathbf{T}(x,y) = (1/x - [1/x], 1/y - [1/x]),$$

with the ergodic invariant probability measure $\mu = (\log 2)^{-1}(x-y)^{-2}\,dA$ (cf. [7, 24, 23]).

Define $\mathbf{T}(x,\infty) = (1/x - [1/x], -[1/x]) = (x_0, y_0)$ and let $\Omega^* = I \times (J \cup \{\infty\})$. We shall use the convention of denoting the nth iterate of a map with an exponent. Let $\mathbf{T}^n(x_0, y_0) = (x_n, y_n)$. By induction we have: if $x = [a_1, a_2, \ldots]$ then

$$x_n = [a_{n+2}, a_{n+3}, \ldots] \quad \text{and} \quad y_n = -a_{n+1} - [a_n, \ldots, a_1].$$

Note that for x rational and n sufficiently large, $T^n(x)$ is not defined.

2.2. Definition of the first return maps and their relation to thetas. Set $\Omega_\alpha = \{(x,y) \in \Omega \mid (x-y)^{-1} < \alpha\}$. For $(x,y) \in \Omega^*$, let $\tau_\alpha(x,y) = \min\{n \geq 1 \mid \mathbf{T}^n(x,y) \in \Omega_\alpha\}$. Since $\mathbf{T}$ is a measure preserving ergodic transformation, with the exception of a set of measure zero, τ_α is defined and $\mathbf{T}^n(x,y) \in \Omega_\alpha$ for infinitely many positive integers [15, 4]. Moreover, for a.a. x, if $\mathbf{T}^n(x,y) \in \Omega_\alpha$ for infinitely many positive integers, then it follows from [9, Lemma 1], that for any $y' \in J \cup \{\infty\}$, $\mathbf{T}^n(x,y') \in \Omega_\alpha$ for infinitely many positive integers.

Henceforth we assume that the exceptional set is removed from Ω^*. Now define $\mathbf{T}_\alpha$ on Ω^* by

$$\mathbf{T}_\alpha(x,y) = \mathbf{T}^{\tau_\alpha(x,y)}(x,y).$$

$\mathbf{T}_\alpha$ is called the *first return map* when restricted to Ω_α. Note that $\Omega_1 = \Omega$ and $\mathbf{T}_1 = \mathbf{T}$.

For $(u,v) \in \mathbb{R}^2$ let $\|(u,v)\| = (u-v)^{-1}$. Given $x \in (0,1)$ set $(\bar{x}_k, \bar{y}_k) = \mathbf{T}_\alpha^k(x_0,y_0) = \mathbf{T}_\alpha^k(\mathbf{T}(x,\infty))$. Aside from having some inherent interest, the importance of $\mathbf{T}_\alpha$ stems from the way in which it ties the sequence $\bar{\Theta}(\alpha)$ to an ergodic dynamical system.

Lemma 2.1. *For $0 < \alpha \le 1$ and almost all irrational $x \in (0,1)$, $\theta_n < \alpha$ for infinitely many positive integers. Furthermore, $\mathbf{T}^n(x_0,y_0) \in \Omega_\alpha$ for infinitely many positive integers and $\bar{\theta}_k = \|\mathbf{T}_\alpha^k(x_0,y_0)\| = (\bar{x}_k - \bar{y}_k)^{-1}$.*

Proof. We first assume $\theta_n < \alpha$ for infinitely many positive integers and prove the final assertion of the lemma. It is an elaboration on $\theta_n = \|\mathbf{T}^n(x_0,y_0)\|$ (cf. [9, 11]), which is proved in the following sequence of equalities:

$$\frac{1}{x_n - y_n} = \frac{1}{[a_{n+2},\ldots] + a_{n+1} + [a_n,\ldots,a_1]} = \left(\frac{1}{T^n(x)} - \frac{q_{n-1}}{q_n}\right)^{-1} = \theta_n \tag{2.1}$$

where we refer to [12] and [18], respectively, for the second and third equalities. In particular, it follows from this that $\mathbf{T}^n(x_0,y_0) \in \Omega_\alpha$ for infinitely many positive integers.

Now we argue by induction. Let m be the smallest value so that $\mathbf{T}^m(x_0,y_0) = (x_m,y_m) \in \Omega_\alpha$. It follows that $\mathbf{T}^m(x_0,y_0) = \mathbf{T}_\alpha(x_0,y_0)$. But this is also equivalent to m being the smallest value for which $\theta_m = 1/(x_m - y_m) < \alpha$. Together these give $\bar{\theta}_1 = \theta_m = \|\mathbf{T}^m(x_0,y_0)\| = \|\mathbf{T}_\alpha(x_0,y_0)\|$.

Now to the inductive step. We suppose that $\bar{\theta}_k = \|\mathbf{T}_\alpha^k(x_0,y_0)\|$. In terms of the natural extension this is $\theta_{n_k} = \|\mathbf{T}^{n_k}(x_0,y_0)\|$. Let $\theta_m = \theta_{n_{k+1}} = \bar{\theta}_{k+1}$. Then $\|\mathbf{T}^m(x_0,y_0)\| = \theta_m < \alpha$ and for $n_k < r < m$, $\|\mathbf{T}^r(x_0,y_0)\| > \alpha$. Translating this says that $\mathbf{T}^m(x_0,y_0) \in \Omega_\alpha$ and $\mathbf{T}^r(x_0,y_0) \notin \Omega_\alpha$ for $n_k < r < m$. Thus, $\mathbf{T}^m(x_0,y_0) = \mathbf{T}_\alpha^{k+1}(x_0,y_0)$. Putting all this together we have

$$\bar{\theta}_{k+1} = \theta_m = \|\mathbf{T}^m(x_0,y_0)\| = \|\mathbf{T}_\alpha^{k+1}(x_0,y_0)\|.$$

Turning to the first sentence of the lemma, by a theorem of Khinchin, for almost all $x \in (0,1)$ the limiting average of the partial quotients a_i diverges to infinity [16]. Consequently, using the characterization of θ_n in (2.1), for a.a. x the values θ_n get arbitrarily small. It follows that for almost all x, $\theta_n < \alpha$ for infinitely many positive integers. ∎

2.3. Basic properties of the subsequences. Our main tool in this section is the following theorem, which will later be shown to hold when $\mathbf{T}$ is replaced by one of the maps $\mathbf{T}_\alpha$ and Ω is replaced by Ω_α. The theorem is originally due to Bosma, Jager and Wiedijk [3, 13] but we use a form that resembles Theorem 3 from [9].

THEOREM 2.2. *For almost all* $x \in (0,1)$ *and all* $y \in J \cup \{\infty\}$, *the sequence of points* $\{\mathbf{T}^n(x,y)\}$ *is distributed in* Ω *according to the density function* $f(x,y) = (\log 2)^{-1}(x-y)^{-2}$.

Observe that f is the density function for the $\mathbf{T}$-invariant probability measure defined earlier, but the Ergodic Theorem is not sufficient in itself to guarantee convergence for the particular values appearing in the theorem [3, 17]. Using Theorem 2.2 we can prove the following version of the well known theorem of Lévy [21]. It is interesting to see that the placement of the log term in the constant is a consequence of the same phenomena observed in D-L, where good approximations that are only first mediants are not accounted for by the continued fraction expansion when $\alpha > 1/2$ (cf. [2, 12]).

PROPOSITION 2.3. *Given* $\alpha \in (0,1]$, *for almost all* $x \in (0,1)$,

$$\lim_{n\to\infty} \frac{\log \bar{q}_n}{n} = \begin{cases} \pi^2(12(1-\alpha+\log 2+\log\alpha))^{-1} & \text{if } \alpha > 1/2, \\ \pi^2(12\alpha)^{-1} & \text{if } \alpha \le 1/2, \end{cases}$$

Proof. Suppose x satisfies the conclusions of Lemma 2.1 and Theorem 2.2. By definition n_k is the smallest value n such that $k = \#\{j \le n \mid \theta_j < \alpha\}$. By the lemma, k is precisely $\#\{j \le n_k \mid \mathbf{T}^j(x_0,y_0) \in \Omega_\alpha\}$. Thus, making use of Theorem 2.2,

$$\begin{aligned} \lim_{k\to\infty} \frac{k}{n_k} &= \lim_{k\to\infty} \frac{1}{n_k}\#\{j \le n_k \mid \mathbf{T}^j(x_0,y_0) \in \Omega_\alpha\} \\ &= \lim_{n\to\infty} \frac{1}{n}\#\{j \le n \mid \mathbf{T}^j(x_0,y_0) \in \Omega_\alpha\} = \mu(\Omega_\alpha). \end{aligned} \tag{2.2}$$

This last value is $(\log 2)^{-1}(1-\alpha+\log 2+\log\alpha)$ if $\alpha > 1/2$ and is $(\log 2)^{-1}\alpha$ if $\alpha \le 1/2$.

Using the above and the theorem of Lévy, for a.a. x we have

$$\begin{aligned} \frac{\pi^2}{12\log 2} &= \lim_{n\to\infty} \frac{\log q_n}{n} = \lim_{k\to\infty} \frac{\log q_{n_k}}{n_k} = \lim_{k\to\infty} \frac{\log \bar{q}_k}{n_k} \\ &= \Big(\lim_{k\to\infty} \frac{k}{n_k}\Big)\Big(\lim_{k\to\infty} \frac{\log \bar{q}_k}{k}\Big) = \mu(\Omega_\alpha) \lim_{k\to\infty} \frac{\log \bar{q}_k}{k}, \end{aligned}$$

which proves the proposition. ∎

Henceforth we shall write $c_\alpha = (\log 2\,\mu(\Omega_\alpha))^{-1}$. By a similar approach, one gets the following version of D-L.

PROPOSITION 2.4. *For $\alpha \in (0,1]$ and for almost all $x \in (0,1)$ the sequence $\bar{\Theta}(\alpha)$ is distributed in the interval $(0,\alpha)$ according to the density function $c_\alpha(2\log 2)^{-1}\zeta^{-1}(1-|1-2\zeta|)$.*

2.4. Ergodic theory of the first return maps. The $\mathbf{T}$-invariant measure μ restricts to an invariant measure for $\mathbf{T}_\alpha$ on Ω_α, with respect to which $\mathbf{T}_\alpha$ is ergodic [15]. We normalize to get the invariant probability measure $\mu_\alpha = c_\alpha(x-y)^{-2}\,dA$.

THEOREM 2.5. *For almost all $x \in (0,1)$ and all $y \in J \cup \{\infty\}$, the sequence of points $\{\mathbf{T}_\alpha^n(x,y)\}$ is distributed in Ω_α according to the density function $f_\alpha(x,y) = c_\alpha(x-y)^{-2}$.*

Proof. Let $\mathbf{B}$ be a Borel set in Ω_α with boundary of zero measure. Let n_k be the smallest value n such that $k = \#\{j \le n \mid \mathbf{T}^j(x,y) < \alpha\}$. Recall the notation of (2.2) with (x_0,y_0) replaced by $(x,y) \in \Omega^*$. Then for a.a. $x \in (0,1)$ and all $y \in J \cup \{\infty\}$ the following limits exists and we have

$$\left(\lim_{k\to\infty}\frac{n_k}{k}\right)\left(\lim_{k\to\infty}\frac{1}{n_k}\#\{0<j<n_k \mid \mathbf{T}^j(x,y)\in\mathbf{B}\}\right) \tag{2.3}$$

$$= \lim_{k\to\infty}\frac{1}{k}\#\{0<j<k \mid \mathbf{T}_\alpha^j(x,y)\in\mathbf{B}\}. \tag{2.4}$$

In view of (2.2), the value of the first limit in (2.3) is $c_\alpha \log 2$ and consequently the above is equal to

$$(c_\alpha\log 2)\lim_{n\to\infty}\frac{1}{n}\#\{0<j<n \mid \mathbf{T}^j(x,y)\in\mathbf{B}\} = (c_\alpha\log 2)\,\mu(B) = \mu_\alpha(B). \ \blacksquare$$

By viewing $\mathbf{T}_\alpha$ as an automorphism of Ω_α with invariant measure μ_α, the value for c_α with $\alpha > 1/2$ can be obtained from [20], as a special case of Theorem 5.9. The constant $c_{1/2}$ was found in [19].

3. The first return maps and the distribution of theta pairs.

In this section we turn to the space of pairs of the form $\{(\bar\theta_n,\bar\theta_{n+1})\}$ and see how Jager's approach can be modified to derive the distribution function for the generic sequence of pairs.

3.1. Theta pairs for $\alpha \ge 1/2$. The natural domain for the pairs, when α is taken to be greater than or equal to $1/2$, is the set

$$\Lambda_\alpha = \{(w,z)\in\mathbb{R}^2 \mid 0<w<\alpha,\ 0<z<\alpha,\ w+z<1\}$$

Define $\Lambda_\alpha^- = \{(w,z)\in\Lambda_\alpha \mid z < w-\alpha+\sqrt{1-4\alpha w}\}$ and $\Lambda_\alpha^+ = \Lambda_\alpha \setminus \Lambda_\alpha^-$.

On Λ_α we have the density function

$$\lambda_\alpha(w,z) = \begin{cases} c_\alpha(\sqrt{1-4\alpha zw})^{-1} & \text{if } (w,z)\in\Lambda_\alpha^+,\\ c_\alpha((\sqrt{1-4\alpha zw})^{-1}+(\sqrt{1+4\alpha zw})^{-1}) & \text{if } (w,z)\in\Lambda_\alpha^-. \end{cases}$$

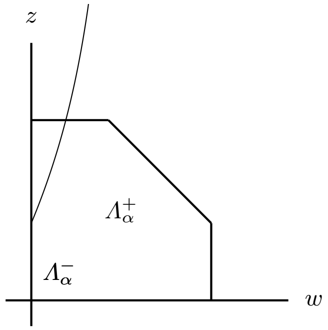


Fig. 1. Λ_α when $\alpha = 0.7$

Jager's description of the distribution of approximating pairs becomes

THEOREM 3.1. *For $\alpha \geq 1/2$ and almost all $x \in (0,1)$, the sequence $\{\bar\theta_k, \bar\theta_{k+1}\}$ is distributed in the region Λ_α according to the density function $\lambda_\alpha(w,z)$. In other words, for almost all $x \in (0,1)$ and for any Borel subset B of Λ_α with boundary of measure zero,*

$$\lim_{n\to\infty} \frac{1}{n} \#\{j \leq n \mid (\bar\theta_j, \bar\theta_{j+1}) \in B\} = \iint_B \lambda_\alpha(w,z)\, dw\, dz.$$

The proof is deferred until Section 3.4, where we prove the stronger Theorem 3.4.

3.2. The structure of the first return maps for $\alpha \geq 1/2$. Define the following sets:

$$\Omega_\alpha^- = \left\{(x,y) \in \Omega \;\middle|\; y \leq \frac{\alpha x}{\alpha - x}\right\},$$
$$\nabla_\alpha = \{(x,y) \in \Omega_\alpha \mid y \geq x - 1/\alpha\} = \Omega \setminus \Omega_\alpha, \qquad \Omega_\alpha^+ = \Omega \setminus (\Omega_\alpha^- \cup \nabla_\alpha).$$

As usual we denote the closure of a set with an overline. Let ∇_α^* denote the union of ∇_α and its boundary along the line $x = 0$.

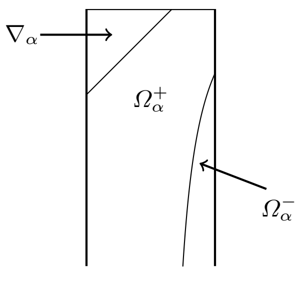


Fig. 2. Ω when $\alpha = 0.6$

LEMMA 3.2. *When $\alpha \geq 1/2$, $\mathbf{T}$ is a bijection, mapping $\overline{\Omega_\alpha^-}$ onto ∇_α^*. Consequently, there is a simple dichotomy describing $\mathbf{T}_\alpha$:*

$$\mathbf{T}_\alpha(x,y)=\begin{cases}\mathbf{T}(x,y) & \textit{if } (x,y)\in\Omega_\alpha^+\cup\nabla_\alpha,\\ \mathbf{T}^2(x,y) & \textit{if } (x,y)\in\Omega_\alpha^-.\end{cases}$$

The more general case, in terms of S-expansions, was addressed in Kraaikamp [20]. We include a simple, short proof.

Proof of Lemma 3.2. For $1/2 < x < 1$, $\mathbf{T}(x,y) = (1/x-1, 1/y-1)$. This function naturally extends to a homeomorphism of the strip $(1/2,1]\times(-\infty,-1]$ onto the rectangle $[0,1)\times[-2,-1)$, taking the ray $x=1$ to the segment $x=0$. Furthermore, $\mathbf{T}$ takes the curve $y=\alpha x/(\alpha-x)$ bounding Ω_α^- to the diagonal $y=x-1/\alpha$ bounding ∇_α. The first assertion of the lemma follows.

Given $(x,y)\in\Omega$, $\mathbf{T}(x,y)\notin\Omega_\alpha$ if and only if $\mathbf{T}(x,y)\in\nabla_\alpha$. It follows from the above that this occurs if and only if $(x,y)\in\Omega_\alpha^-$. We observe that ∇_α is disjoint from both Ω_α^+ and Ω_α^-. Consequently, if $(x,y)\in\Omega_\alpha^+\cup\nabla_\alpha$ then $\mathbf{T}(x,y)\in\Omega_\alpha$ and $\mathbf{T}(x,y)=\mathbf{T}_\alpha(x,y)$. If $(x,y)\in\Omega_\alpha^-$ than $\mathbf{T}(x,y)\in\nabla_\alpha$ and $\mathbf{T}_\alpha(x,y)=\mathbf{T}^2(x,y)\in\Omega_\alpha$, as asserted. ∎

3.3. From Ω_α to the domain of theta pairs. Define

$$F^+(x,y)=\frac{-xy}{x-y}\quad\text{and}\quad F^-(x,y)=\frac{(1-x)(1-y)}{x-y}$$

and set

$$\mathbf{F}^+(x,y)=(\|(x,y)\|,F^+(x,y)),\qquad \mathbf{F}^-(x,y)=(\|(x,y)\|,F^-(x,y)).$$

Then we let

$$\mathbf{F}(x,y)=\begin{cases}\mathbf{F}^+(x,y) & \text{if } (x,y)\in\Omega_\alpha^+,\\ \mathbf{F}^-(x,y) & \text{if } (x,y)\in\Omega_\alpha^-.\end{cases}$$

The next proposition describes the relationship between the space Ω_α and the natural domain for the theta pairs Λ_α. In effect it allows us to equate the $\mathbf{T}_\alpha$-orbit of a point in Ω_α with a sequence of pairs in Λ_α. It is an easy step from here to the proof of Theorem 3.1.

The case where $\alpha=1$ was proved in [14] and the result for $\alpha=1/2$ was obtained in [19].

PROPOSITION 3.3.

(a) $\mathbf{F}^+(x,y)=(w,z)$ *is a homeomorphism of* Ω_α^+ *onto* Λ_α. *Its inverse is given by*

$$\mathbf{H}^+(w,z)=\left(\frac{1-\sqrt{1-4wz}}{2w},\frac{-1-\sqrt{1-4wz}}{2w}\right).$$

(b) $\mathbf{F}^-(x,y)=(w,z)$ *is a homeomorphism of* Ω_α^- *onto* Λ_α^-. *Its inverse is given by*

$$\mathbf{H}^-(w,z)=\left(\frac{2w+1-\sqrt{1+4wz}}{2w},\frac{2w-1-\sqrt{1+4wz}}{2w}\right).$$

(c) *Furthermore, for almost all* $x \in (0,1)$ *and* $(x_0, y_0) = \mathbf{T}(x, \infty)$,

$$\mathbf{F}(\bar{x}_k, \bar{y}_k) = \mathbf{F}\big(\mathbf{T}_\alpha^k(x_0, y_0)\big) = (\bar{\theta}_k, \bar{\theta}_{k+1}).$$

Proof. One sees easily that if $1 - 4wz \geq 0$, then $\mathbf{F}^+ \circ \mathbf{H}^+(w,z) = (w,z)$. We show that $\mathbf{F}^+$ is injective in $A^+ = \{x + y < 0,\ y < x\}$. Suppose $(x,y) \neq (u,v)$ so that $\|(x,y)\| = \|(u,v)\|$ and $F^+(x,y) = F^+(u,v)$. These can be written

$$x - y = u - v \tag{3.1}$$

and

$$xy(u - v) = uv(x - y). \tag{3.2}$$

It follows that $xy = uv$ or $u = xy/v$. Substituting into (3.2) we get the equation $x^2 - x(y - v) - yv = 0$. Since $y < 0$ this implies $v = -x$ and also $u = -y$. But then $(u,v) \notin A^+$ proving that F^+ is injective.

Note that in general $F^+(x,y) = F^+(-y,-x)$ and the image of the ray $y = -x$ with $x > 0$ is piece of the hyperbola $z = 1/(4w)$ with $w > 0$. Thus $\mathbf{F}^+$ maps A^+ homeomorphically onto a subset of $B^+ = \{z < 1/(4w),\ w > 0\}$. But since $\mathbf{H}^+$ is defined in B^+, A^+ must map homeomorphically onto the entire region. One easily checks that the curves bounding Ω_α^+ in A^+ map to the four curves bounding Λ_α in the right half-plane. That completes the proof of part (a).

Part (b) follows the same script. One concludes first that $\mathbf{F}^-$ is injective on $A^- = \{x + y < 0, y < -x + 2\}$ with inverse $\mathbf{H}^-$. In this case we see that $\mathbf{F}^-(x,y) = \mathbf{F}^-(-y+2, -x+2)$ and $\mathbf{F}^-$ takes the ray $y = -x + 2$ with $x > 0$ to the piece of the hyperbola $z = -1/(4w)$ with $w > 0$. As above it follows that $\mathbf{F}^-$ maps A^- homeomorphically onto the region in the right half-plane with $z > -1/(4w)$. The positive x-axis maps to the positive w-axis. The other boundary arc of Ω_α^- in A^- is the curve $y = \alpha x/(\alpha - x)$. Its image under $\mathbf{F}^-$ has

$$w = \frac{x - \alpha}{x^2} \quad \text{and therefore} \quad x = \frac{1 - \sqrt{1 - 4w\alpha}}{2w}. \tag{3.3}$$

Now substituting in for y and using the first and second parts of (3.3) we have

$$\begin{aligned} z &= \frac{(1 - x)(1 - \alpha x/(\alpha - x))}{x - \alpha x/(\alpha - x)} = \frac{x - \alpha}{x^2} + \frac{2\alpha}{x} - 1 - \alpha \\ &= w + \frac{2\alpha}{x} - 1 - \alpha = w - \alpha + \sqrt{1 - 4w\alpha}. \end{aligned}$$

This curve, together with the w-axis, bounds Λ_α^- in A^-. That completes the proof of (b).

We still need to address part (c), which says that given $x \in (0,1)$, $\mathbf{F}$ takes the sequence $(\bar{x}_k, \bar{y}_k)$ to the corresponding sequence of theta pairs. Suppose x

belongs to the full measure set for which $\bar{\theta}_k = \|\mathbf{T}_\alpha^k(x_0, y_0)\|$, as in Lemma 2.1. Since $\bar{\theta}_k = \|(\bar{x}_k, \bar{y}_k)\|$, the result is clear for the w coordinate, so we turn to the second coordinate.

By Lemma 3.2 there are only two possibilities for the value of $\bar{\theta}_{k+1}$. First, if $(\bar{x}_k, \bar{y}_k) \in \Omega_\alpha^+$, then $\mathbf{T}_\alpha(\bar{x}_k, \bar{y}_k) = \mathbf{T}(\bar{x}_k, \bar{y}_k)$ and

$$\bar{\theta}_{k+1} = \|\mathbf{T}_\alpha(\bar{x}_k, \bar{y}_k)\| = \|\mathbf{T}(\bar{x}_k, \bar{y}_k)\| = \frac{-\bar{x}_k\bar{y}_k}{\bar{x}_k - \bar{y}_k} = F^+(\bar{x}_k, \bar{y}_k).$$

The second possibility occurs when $(\bar{x}_k, \bar{y}_k) \in \Omega_\alpha^-$. Then we must have $1/2 \le \alpha < \bar{x}_k < 1$ and consequently $[1/\bar{x}_k] = 1$. Using this fact and a little calculation gives

$$\bar{\theta}_{k+1} = \|\mathbf{T}_\alpha(\bar{x}_k, \bar{y}_k)\| = \|\mathbf{T}^2(\bar{x}_k, \bar{y}_k)\| = \frac{(1-\bar{x}_k)(1-\bar{y}_k)}{\bar{x}_k\bar{y}_k} = F^-(\bar{x}_k, \bar{y}_k).$$

That completes the proof of the proposition. ∎

3.4. The distribution. The following is a strengthened version of Theorem 3.1. Again the cases $\alpha = 1$ and $\alpha = 1/2$ were dealt with in [13] and [19], respectively. See also [14].

THEOREM 3.4. *For almost all $\eta \in (0,1)$ and all $\zeta \in J \cup \{\infty\}$, the sequence $(w_k, z_k) = \mathbf{F}(\mathbf{T}_\alpha^k(\eta, \zeta))$ is distributed in the region Λ_α according to the density function λ_α. In particular, for almost all $x \in (0,1)$ this holds with $(\eta, \zeta) = (x_0, y_0) = \mathbf{T}(x, \infty)$ and in this case $(w_k, z_k) = (\bar{\theta}_k, \bar{\theta}_{k+1})$.*

Proof. This is a modification of the proof in [13] along the lines of [10]. First off, we define the measure λ_α on Borel sets in Λ_α by setting $\lambda_\alpha(D) = \mu_\alpha(\mathbf{F}^{-1}(D))$. Let $D_\alpha^- = D \cap \Lambda_\alpha^-$. Then following a computation of Jacobians, we have

$$\begin{aligned}\lambda_\alpha(D) = &\iint_{F^{-1}(D)} f_\alpha\,dx\,dy = \iint_D f_\alpha(\mathbf{H}(w,z))|\mathrm{Jac}\,\mathbf{H}(w,z)|\,dw\,dz\\ &+ \iint_{D_\alpha^-} f_\alpha(\mathbf{H}^-(w,z))|\mathrm{Jac}\,\mathbf{H}^-(w,z)|\,dw\,dz = \iint_D \lambda_\alpha(w,z)\,dwdz.\end{aligned}$$

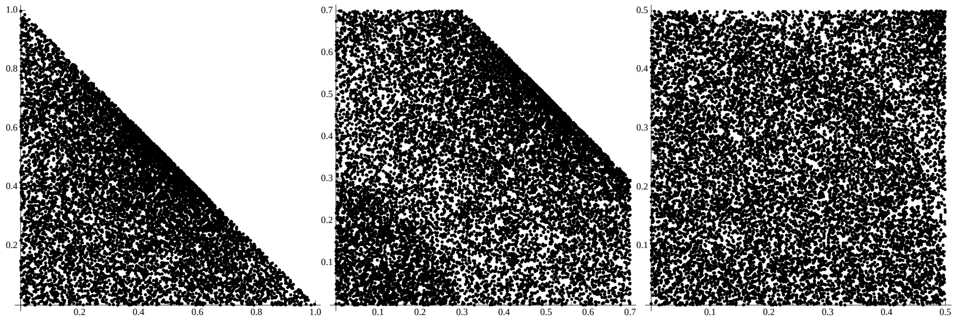
As a consequence of Theorem 2.5, for almost all $(\eta, \zeta) \in \Omega^*$,

$$\begin{aligned}(3.4)\qquad \lambda_\alpha(D) = \mu_\alpha(\mathbf{F}^{-1}(D)) &= \lim_{k\to\infty}\frac{1}{k}\#\{j \le k \mid \mathbf{T}_\alpha^k(\eta,\zeta) \in \mathbf{F}^{-1}(D)\}\\ &= \lim_{k\to\infty}\frac{1}{k}\#\{j \le k \mid (w_k, z_k) \in D\}.\end{aligned}$$

That proves the first assertion of the theorem.

Now suppose $x \in (0,1)$ is chosen from the full measure set guaranteed by Theorem 2.5, for which the sequence $(\bar{x}_k, \bar{y}_k)$ is distributed according to the density function f_α. Then as a consequence of Proposition 3.3 and (3.4) above we get

$$\lim_{n\to\infty}\frac{1}{n}\#\{j\le n \mid (\bar\theta_j,\bar\theta_{j+1})\in D\}$$
$$=\lim_{k\to\infty}\frac{1}{k}\#\{j\le k \mid (w_k,z_k)\in D\}=\lambda_\alpha(D). \quad \blacksquare$$



(a) $\alpha = 1$ (b) $\alpha = .7$ (c) $\alpha = .5$

Fig. 3

In the three graphs of Figure 3, the sets of pairs $(\bar\theta_j,\bar\theta_{j+1})$ are plotted for different values of α. The $\bar\theta$'s have been extracted from a sequence of θ's generated by taking $x\approx\pi^2+\sqrt{2}-1$. When $\alpha=1$ you see the distribution described by Jager's theorem [13]. Note the similarity to Figure 1 in [14]. When $\alpha=1/2$ compare to Figure 2 in [19]. For $\alpha<1/2$ it is difficult to see any interesting detail using this approach.

4. A geometric approach that addresses all values α. In this final section we revisit the first return maps using a more geometric approach.

4.1. The setup. Given a fraction p/q in lowest terms and a number $\kappa\in(0,1]$, $\mathbf{D}_{p/q}(\kappa)$ is the open disc of radius κ/q^2 which is tangent to the real line at the point p/q. Set $\mathbf{D}_\infty(\kappa)=\{\zeta=u+iv \mid v>1/(2\kappa)\}$. These regions are sometimes called *horocycles* or their boundaries are called *Ford circles*. Let $\mathcal{D}(\kappa)$ denote the union of all such discs. We will be particularly interested when p/q is in the closed unit interval.

Suppose x is irrational with continued fraction expansion $x=[a_1,a_2,\ldots]$. As in [4] define

$$\Delta=\Delta^{n+1}_{a_1,\ldots,a_{n+1}}=\{\zeta=[a_1,\ldots,a_{n+1}+r] \mid r\in(0,1),\ \text{irrational}\}.$$

This interval has endpoints

$$\frac{p_{n+1}}{q_{n+1}}=[a_1,\ldots,a_{n+1}] \quad\text{and}\quad \frac{p_{n+1}+p_n}{q_{n+1}+q_n}=[a_1,\ldots,a_{n+1}+1].$$

Notice that T^{n+1} restricted to the closure of Δ is a Möbius transformation g mapping $\bar{\Delta}$ one-one onto $[0,1]$ and taking p_n/q_n to ∞. It is easy to verify that

$$g_{n+1}(x) = \begin{pmatrix} q_{n+1} & -p_{n+1} \\ -q_n & p_n \end{pmatrix}(x) = \frac{q_{n+1}x - p_{n+1}}{-q_n x + p_n}. \tag{4.1}$$

Then for any $y < -1$, $\mathbf{T}^{n+1}(x,y) = (g_{n+1}(x), g_{n+1}(y))$ (cf. [9]).

The transformation g_n is an automorphism of the Riemann sphere. It will preserve the upper half-plane $\mathbb{H}$ when n is even and it will interchange the upper and lower half-planes when n is odd. We shall extend g_n to a self-map of $\mathbb{H}$ by setting

$$G_n(z) = \begin{cases} g_n(z) & \text{if } n \text{ even}, \\ \overline{g_n(z)} & \text{if } n \text{ odd}. \end{cases}$$

Given $x \neq y$, real numbers or infinity, let $\overline{xy}$ denote the arc of the circle in the upper half-plane $\mathbb{H}$ orthogonal to $\mathbb{R} \cup \{\infty\} = \hat{\mathbb{R}}$. This is a geodesic in the Poincaré model for the hyperbolic plane. We will have $\mathbf{T}^{n+1}$ act on geodesics by setting $\mathbf{T}^{n+1}(\overline{xy}) = \overline{g_{n+1}(x)g_{n+1}(y)} = G_{n+1}(\overline{xy})$.

One begins to see how this fits with the earlier material in the following.

PROPOSITION 4.1. *For x irrational in $(0,1)$ with convergents p_j/q_j, any $y < -1$ and $n \in \mathbb{N}$,*

$$\|\mathbf{T}^{n+1}(x,y)\| < \alpha \quad \textit{if and only if} \quad \overline{xy} \cap \mathbf{D}_{p_n/q_n}(\alpha) \neq \emptyset.$$

Proof. The result holds for $n = 0$ by taking $p_0/q_0 = 0/1 = 0$.

A simple calculation [8] shows that the transformations G_n permute the discs in $\mathcal{D}(\alpha)$. Then, since $g_{n+1}(p_n/q_n) = \infty$, $G_{n+1}(\mathbf{D}_{p_n/q_n}(\alpha)) = \mathbf{D}_\infty(\alpha)$. Therefore $\overline{xy} \cap \mathbf{D}_{p_n/q_n}(\alpha) \neq \emptyset$ if and only if $\overline{\mathbf{T}^{n+1}(x,y)} \cap \mathbf{D}_\infty(\alpha) \neq \emptyset$. But this last is equivalent to $(x-y)/2 > 1/(2\alpha)$ or $\|\mathbf{T}^{n+1}(x,y)\| < \alpha$. ∎

4.2. Geodesic-horocycle intersections and τ_α. Given $(x,y) \in \Omega^*$, let $\mathbf{D}(x,y) = \mathbf{D}_{p_n/q_n}(\alpha)$ be the horocycle intersecting $\overline{xy}$ with q_n minimal. If $q_n = 1$ and $\mathbf{D}_0(\alpha) \cap \overline{xy} \neq 0$ then set $\mathbf{D}(x,y) = \mathbf{D}_0(\alpha)$, otherwise set $D(x,y) = \mathbf{D}_1(\alpha)$.

THEOREM 4.2. *Suppose $\alpha \geq 1/2$. Then $\mathbf{D}(x,y)$ is either $\mathbf{D}_0(\alpha)$ or $\mathbf{D}_1(\alpha)$ and*

$$\tau_\alpha(x,y) = \begin{cases} 1 & \textit{if } \mathbf{D}(x,y) = \mathbf{D}_0(\alpha), \\ 2 & \textit{if } \mathbf{D}(x,y) = \mathbf{D}_1(\alpha). \end{cases} \tag{4.2}$$

Suppose $\alpha < 1/2$. Then either $\mathbf{D}(x,y)$ is one of $\mathbf{D}_0(\alpha)$ or $\mathbf{D}_1(\alpha)$ and the conclusion of (4.2) *holds, or else $\mathbf{D}(x,y) = \mathbf{D}_{p_n/q_n}(\alpha)$ with $q_n > 1$ and then*

$\tau_\alpha(x,y)$

$$= \begin{cases} n+1 & \textit{if } n \textit{ is odd and } x < p_n/q_n, \textit{ or if } n \textit{ is even and } x > p_n/q_n, \\ n+2 & \textit{if } n \textit{ is even and } x < p_n/q_n, \textit{ or if } n \textit{ is odd and } x > p_n/q_n. \end{cases}$$

One could use this point of view to characterize Ω_α^+ and Ω_α^- and reprove Lemma 3.2. In Example 1, we will see how one might address the next simplest case with $\alpha < 1/2$.

Proof of Theorem 4.2. First observe that if $\alpha \geq 1/2$, every $\overline{xy}$ must meet one of $\mathbf{D}_0(\alpha)$ or $\mathbf{D}_1(\alpha)$. Suppose $D(x,y) = \mathbf{D}_0(\alpha)$. Then $\overline{xy} \cap \mathbf{D}_0(\alpha) \neq \emptyset$, and by Proposition 4.1, $\|\mathbf{T}(x,y)\| < \alpha$. In other words, $\tau_\alpha(x,y) = 1$. If $\overline{xy} \cap \mathbf{D}_0(\alpha) = \emptyset$, then $\overline{xy} \cap \mathbf{D}_1(\alpha) \neq \emptyset$. Then $x > 1/2$ and consequently $p_1/q_1 = 1/1$. Again, it follows from the proposition that $\tau_\alpha(x,y) = 2$. Note that (4.2) remains true even when $\alpha \leq 1/2$.

Henceforth we take $\alpha \leq 1/2$. Suppose $\mathbf{D}(x,y) = \mathbf{D}_{1/a_1}(\alpha)$ for $a_1 > 1$. If $x > 1/a_1$ then an easy computation shows that $x < 1/(a_1-1)$ or, in other words, $x \in \Delta^1_{a_1-1}$. Observe that $T(x) = g(x) = 1/x - a_1 + 1$ maps the closure of $\Delta^1_{a_1-1}$ to $[0,1]$, taking $1/a_1$ to 1. Let $\mathbf{T}(x,y) = (x',y')$. Then $\mathbf{T}(\overline{xy}) = G(\overline{xy}) = \overline{x'y'}$ intersects $\mathbf{D}_1(\alpha) = G(\mathbf{D}_{1/a_1}(\alpha))$. Furthermore, since $\overline{xy}$ is disjoint from $\mathbf{D}_{1/(a_1-1)}(\alpha)$, $\overline{x'y'}$ does not intersect $\mathbf{D}_0(\alpha) = G(\mathbf{D}_{1/(a_1-a)}(\alpha))$. It follows from the previous paragraph that $\tau_\alpha(x',y') = 2$ and consequently that $\tau_\alpha(x,y) = 3$.

Similarly, if $x < 1/a_1$ then $x > 1/(a_1+1)$ or $x \in \Delta^1_{a_1}$. We notice that $T(x) = g(x) = 1/x - a_1$ maps the closure of $\Delta^1_{a_1}$ to $[0,1]$, taking $1/a_1$ to 0. Now, $\mathbf{T}(\overline{xy}) = G(\overline{xy}) = \overline{x'y'}$ intersects $\mathbf{D}_0(\alpha) = G(\mathbf{D}_{1/a_1}(\alpha))$. Then $\tau_\alpha(x',y') = 1$ and $\tau_\alpha(x,y) = 2$. The theorem follows for $k = 1$ where $p_1/q_1 = 1/a_1$.

The proof is completed by induction. Suppose the result holds for $k > 1$ and $\mathbf{D}(x,y) = \mathbf{D}_{p_{k+1}/q_{k+1}}(\alpha)$. The convergent p_{k+1}/q_{k+1} is contained in the interval $\Delta^1_{a_1}$. Thus, on the interval, and therefore in a neighborhood of the fraction, T is the transformation g which extends to $\mathbb{H}$ as $G(z) = 1/\bar{z} - a_1$. Then, as above, $\mathbf{T}$ maps the geodesic $\overline{xy}$ to $\overline{x'y'}$. Since g reverses orientation on $\mathbb{R}$, if x is greater than or less than p_{k+1}/q_{k+1} then the reverse is true for $g(x)$ relative to $g(p_{k+1}/q_{k+1})$. In particular, if $x > p_{k+1}/q_{k+1}$, then $g(x) = x' < g(p_{k+1}/q_{k+1})$. Now, note that the kth convergent to $g(x)$ is $g(p_{k+1}/q_{k+1})$. Then $\mathbf{D}(x',y') = G(\mathbf{D}_{p_{k+1}/q_{k+1}}(\alpha)) = \mathbf{D}_{g(p_k/q_k)}(\alpha)$, and by the inductive hypothesis, $\tau_\alpha(x',y')$ is $k+1$ if k is odd and it is $k+2$ if k is even. It follows that $\tau_\alpha(x,y)$ is $k+2$ if $k+1$ is even and it is $k+3$ if $k+1$ is odd. In the same manner the result follows when $x < p_{k+1}/q_{k+1}$. That completes the proof. ∎

Example 1. Choose $1/\sqrt{5} < \alpha < 1/2$. Since $\alpha < 1/2$ there exist $(x,y) \in I \times J$ so that $\mathbf{D}(x,y) = \mathbf{D}_{1/2}(\alpha)$. Moreover, because of the choice of lower bound, $\mathbf{D}(x,y)$ must be one of the horocycles $\mathbf{D}_0(\alpha)$, $\mathbf{D}_1(\alpha)$ or $\mathbf{D}_{1/2}(\alpha)$ for any $(x, y) \in \Omega$ and consequently in Ω_α. If the first or the second possibility holds, then $\tau_\alpha(x, y)$ is respectively 1 or 2. In the remaining case $\overline{xy}$ meets $\mathbf{D}_{1/2}(\alpha)$ but neither of the other horocycles. If $x < 1/2$, then by the theorem $\tau_\alpha(x, y) = 2$, whereas if $x > 1/2$, then $\tau_\alpha(x, y) = 3$.

Given such an α and some free time, it would be possible to numerically characterize the regions on which each of the behaviors holds. Then Ω_α will be divided into regions, and by specifying the appropriate power $\tau_\alpha(x, y)$ of $\mathbf{T}$ on each of them, one can describe $\mathbf{T}_\alpha$. One could then, in principle, compute the density function for the distribution of theta pairs as in Theorem 3.4.

It would appear that the dynamical systems determined by values α with $1/\sqrt{5} < \alpha < 1/2$ will be structurally identical, where the sets Ω_α are subdivided into regions of the same shape, on which $\tau_\alpha(x, y)$ is constant; much like the situation when $\alpha > 1/2$. We conjecture that there will be an infinite, discrete set of numbers α_i decreasing to some value α_∞ greater than zero, so that for $\alpha_{i+1} < \alpha < \alpha_i$ the dynamical systems $\{\mathbf{T}_\alpha, \Omega_\alpha\}$ are topologically conjugate. At each of the values α_i, the system will abruptly change, in particular $\sup \tau_\alpha(x, y)$ will increase. Below α_∞ the systems should attain a higher degree of complexity. All this might somehow relate to the Markov spectrum [5].

Example 2. It follows in the previous example that if $\alpha > 1/\sqrt{5}$ then $\tau_\alpha(x, y)$ is defined for all $(x, y) \in \Omega$. This is no longer true if $\alpha < 1/\sqrt{5}$. In the simplest case, consider what happens when we take $\zeta = \frac{1}{2}(\sqrt{5} - 1)$. By an old theorem of Hurwitz [12], for $\alpha < 1/\sqrt{5}$ there are only finitely many fractions p/q with

$$\left|\zeta - \frac{p}{q}\right| < \frac{\alpha}{q^2}.$$

This inequality can be read geometrically as implying that $\overline{\zeta\infty}$ meets only finitely many of the horocycles $\mathbf{D}_{p_n/q_n}(\alpha)$, where the fractions are the continued fraction convergents of ζ. It follows that $\mathbf{T}^n(\zeta, \infty) \in \Omega_{.4}$ for only finitely many values n.

Set $\alpha = .4$, and let $\eta = \frac{1}{2}(-\sqrt{5} - 1)$. It follows from Markov's theorem [5] that $\overline{\zeta\eta}$ is disjoint from all of the horocycles $D_{p/q}$. Consequently, $\tau_{.4}(\zeta, \eta)$ is not defined.

As in Example 1, one expects that as α decreases, the set on which τ_α is *not* defined will increase in complexity.

References

[1] R. L. Adler and L. Flatto, *Cross section maps for geodesic flows. I. The modular surface*, in: Ergodic Theory and Dynamical Systems II, A. Katok (ed.), Progr. Math. 21, Birkhäuser, Boston, 1982, 103–161.

[2] W. Bosma, *Approximation by mediants*, Math. Comp. 54 (1990), 421–434.
[3] W. Bosma, H. Jager and F. Wiedijk, *Some metrical observations on the approximation of continued factions*, Indag. Math. 45 (1983), 281–299.
[4] I. P. Cornfeld, S. V. Fomin and Ya. G. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.
[5] T. W. Cusick and M. E. Flahive, *The Markoff and Lagrange Spectra*, Math. Surveys Monogr. 30, Amer. Math. Soc., Providence, RI, 1989.
[6] K. Dajani and C. Kraaikamp, *Ergodic Theory of Numbers*, Carus Math. Monogr. 29, Math. Assoc. Amer., Washington, DC, 2002.
[7] A. Haas, *An ergodic sum related to the approximation by continued fractions*, New York J. Math. 11 (2005), 345–349.
[8] A. Haas, *The distribution of geodesic excursions out the end of a hyperbolic orbifold and approximation with respect to a Fuchsian group*, Geom. Dedicata 116 (2005), 129–155.
[9] A. Haas and D. Molnar, *Metrical diophantine approximation for continued fraction like maps of the interval*, Trans. Amer. Math. Soc. 356 (2004), 2851–2870.
[10] A. Haas and D. Molnar, *The distribution of Jager pairs for continued fraction like mappings of the interval*, Pacific J. Math. 217 (2004), 101–114.
[11] A. Haas and C. Series, *The Hurwitz constant and Diophantine approximation on Hecke groups*, J. London Math. Soc. 34 (1986), 219–234.
[12] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Univ. Press, Oxford, 1979.
[13] H. Jager, *The distribution of certain sequences connected with the continued fraction*, Indag. Math. 48 (1986), 61–69.
[14] H. Jager and C. Kraaikamp, *On the approximation by continued fractions*, Indag. Math. 51 (1989), 289–307.
[15] G. Keller, *Equilibrium States in Ergodic Theory*, London Math. Soc. Student Texts 42, Cambridge Univ. Press, Cambridge, 1998.
[16] A. Ya. Khinchin, *Continued Fractions*, Univ. of Chicago Press, Chicago, 1964.
[17] D. E. Knuth, *The distribution of continued fraction approximations*, J. Number Theory 19 (1984), 443–448.
[18] J. F. Koksma, *Diophantische Approximationen*, Springer, Berlin, 1936.
[19] C. Kraaikamp, *Statistic and ergodic properties of Minkowski's diagonal continued fraction*, Theoret. Comput. Sci. 65 (1989), 197–212.
[20] C. Kraaikamp, *A new class of continued fraction expansions*, Acta Arith. 57 (1991), 1–39.
[21] P. Lévy, *Sur le développement en fraction continue d'un nombre choisi au hasard*, Compos. Math. 3 (1936), 286–303.
[22] H. Nakada, *Metrical theory for a class of continued fraction transformations and their natural extensions*, Tokyo J. Math. 4 (1981), 399–426.
[23] H. Nakada, *Continued fractions, geodesic flows and Ford circles*, in: Algorithms, Fractals and Dynamics, Y. Takahashi (ed.), Plenum, New York, 1995, 179–191.
[24] C. Series, *The modular surface and continued fractions*, J. London Math. Soc. 31 (1985), 69–80.

Andrew Haas
Department of Mathematics
University of Connecticut
Storrs, CT 06269, U.S.A.
E-mail: haas@math.uconn.edu

Received on 12.2.2012
and in revised form on 10.3.2012 (6619)