NATURAL EXTENSIONS AND GAUSS MEASURES FOR PIECEWISE HOMOGRAPHIC CONTINUED FRACTIONS

BY PIERRE ARNOUX & THOMAS A. SCHMIDT

ABSTRACT. — We give a heuristic method to solve explicitly for an absolutely continuous invariant measure for a piecewise differentiable, expanding map of a compact subset I of Euclidean space $\mathbb{R}^d$. The method consists of constructing a skew product family of maps on $I \times \mathbb{R}^d$, which has an attractor. Lebesgue measure is invariant for the skew product family restricted to this attractor. Under reasonable measure-theoretic conditions, integration over the fibers gives the desired measure on I. Furthermore, the attractor system is then the natural extension of the original map with this measure. We illustrate this method by relating it to various results in the literature.

PIERRE ARNOUX, Université d'Aix-Marseille et CNRS, Institut de Mathématiques de Marseille (UMR7373), Site Sud, Campus de Luminy, Case 907, 13288 MARSEILLE Cedex 9 France   •   *E-mail :* pierre@pierrearnoux.fr

THOMAS A. SCHMIDT, Department of Mathematics, Oregon State University, Corvallis, OR 97331   •   *E-mail :* toms@math.orst.edu

Résumé (*Extensions naturelles et mesures de Gauss pour des algorithmes de fraction continues homographiques par morceaux*). — Nous donnons une méthode heuristique pour trouver explicitement une mesure invariante absolument continue pour une application différentiable par morceaux et expansive d'un sous-ensemble compact I de l'espace euclidien $\mathbb{R}^d$. La méthode consiste à construire une famille d'applications qui est un produit croisé sur $I \times \mathbb{R}^d$, et à montrer que cette famille possède un attracteur. La mesure de Lebesgue est par construction invariante pour la restriction à l'attracteur de cette famille d'applications. Sous des hypothèses raisonnables sur la mesure, l'intégration le long des fibres donne la mesure invariante cherchée sur I. De plus, le système construit sur l'attracteur est l'extension naturelle de l'application originale, munie de cette mesure invariante. Nous illustrons cette méthode par plusieurs exemples tirés de la littérature.

1. Introduction

In a letter to Laplace in 1812, see the Appendix, Gauss [19] gave the explicit formula for the invariant measure associated with regular continued fractions. This is the measure that we refer to as the standard Gauss measure. To this day, it remains unclear how Gauss (using "very simple reasoning") found this measure. Keane [25] suggests the possibility that Gauss used a version of the natural extension of the interval map that was not explicitly introduced into the literature until 1977, by Nakada-Ito-Tanaka [34]. This latter system was used for ergodic theoretic results by Bosma-Jager-Wiedijk (1983), [12].

Indirectly related to this were the works showing that the Gauss system is a factor of the geodesic flow on the unit tangent bundle of the modular surface: Artin [9] in 1924 seems to be the first to have shown this connection, followed among others by Adler-Flatto in 1984 (revisited in [1]), Series [39] in 1985, and Arnoux [2] in 1994 (see also [6, 4]). A large number of authors have pursued these and related matters, see for example [20, 30, 29, 21]. 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 [23].

We do not pretend to suggest that Gauss used the method we present here. Still, in the modern mathematical world this is an applicable, easily used, method. Indeed, in Section 6, we give various examples of interval maps and higher dimensional maps for which our method determines an invariant density (thus a Gauss measure) and the corresponding natural extension. In particular, one easily recovers the natural extension as given by Nakada-Ito-Tanaka and thus the invariant measure is rediscovered, see Section 6.1.

In brief, we show that if an expanding piecewise Möbius interval map is such that a naturally associated two dimensional system has a positive invariant Lebesgue measure, then this system is the natural extension of the interval map whose Gauss measure is the marginal measure of the two dimensional system. In fact, the approach is not limited in dimensions, see [6, 4].

It seems to be part of the folklore that natural extensions for certain interval maps arise as attractors. In particular, Katok and Ugarcovici [24] explicitly prove that the planar models that they determine for the natural extensions of the family of maps that they study are indeed attactors. Our approach is to view the Möbius functions as defining what can reasonably be called a parametrized iterated function system (pIFS) on the interval cross on the real line, and, generalizing Hutchinson's Theorem [22], we show that this pIFS has a fixed compact set projecting onto the line. This compact set gives the aforementioned two-dimensional system, and thus the Gauss measure for the interval map.

The pIFS arises in the following manner. Suppose that our expanding continued fraction map T is given locally by a Möbius transformation of matrix

$$M = \begin{pmatrix} a\ b \\ c\ d \end{pmatrix}$$

with determinant one, and consider the transformation $\widetilde{T}$ locally defined by

$$(1) \qquad \mathcal{T}_M : (x,y) \to (M \cdot x, (cx+d)^2 y - c(cx+d)).$$

This has a Jacobian matrix determinant whose value is one and thus each branch of $\mathcal{T}_M$ preserves Lebesque measure. Under a boundedness condition on the set of $c(cx+d)$, we prove that there is a unique compact set K, arising as an appropriate fixed point, such that K equals the closure of $\widetilde{T}(K)$. If furthermore, K and $\widetilde{T}(K)$ have the same strictly positive measure, then an invariant density for T is given by taking the measure of the fiber of K over any x in the domain of T, and $\widetilde{T}(K)$ on K defines a natural extension for T and this invariant measure.

As we have already shown in [8], in practice one can plot the orbit of almost any point under $\widetilde{T}(K)$ and use the resulting figure to compute equations for the boundary of K and the invariant density. Again, see Section 6 for examples.

1.1. Main results. — We state and prove our fixed point theorem in a more general setting.

THEOREM 1.1. — *Suppose that I is a compact Hausdorff metric space, and Y is a complete metric space. Let $\mathcal{H}$ be the set of non-empty compact subsets of Y with the usual Hausdorff metric, d_H. Let $\mathcal{S}$ be the space of upper semi-continuous maps from I to $\mathcal{H}$, $d_{\mathcal{S}}(K, K') := \sup\{d_H (K(x), K'(x)) \,|\, x \in I\}$.*
    Given

   (i.) *a countable cover $\{I_\alpha\}_{\alpha \in \mathcal{A}}$ of I by compact sets,*
   (ii.) *a family T of continuous maps $T_\alpha : I_\alpha \to I$ such that the family $\{T_\alpha (I_\alpha)\}_{\alpha \in \mathcal{A}}$ is a cover of I, and*
   (iii.) *a uniformly contracting family of continuous maps $S_\alpha : I_\alpha \times Y \to Y$, that is totally bounded,*

define
(a.) *for each α, $\widetilde{T}_\alpha : I_\alpha \times Y \to I \times Y$ by $\widetilde{T}_\alpha(x, y) = (T_\alpha(x), S_\alpha(x, y))$, and*
(b.) *$\Theta : \mathcal{S} \to \mathcal{S}$; by $\Theta(K) = \overline{\cup_{\alpha \in \mathcal{A}} \widetilde{T}_\alpha(K_\alpha)}$.*
Then the map Θ has a unique fixed point.

A restricted version of this theorem has already been applied in a paper by Arnoux, Mizutani, and Sellami [5].

The above leads to the calculation of explicit models of natural extensions in the following manner.

THEOREM 1.2. — *Let I be a compact subset of $\mathbb{R}^d$, which is the closure of its interior, and let $Y = \mathbb{R}^d$. Given (i–iii) as above, with*
(i′.) *each I_α the closure of its interior, and all non-trivial intersections $I_\alpha \cap I_\beta$ of Lebesgue measure zero,*
(ii′.) *each $T_\alpha : I_\alpha \to T_\alpha(I_\alpha) \subset I$ a diffeomorphism, and the collection of Jacobian matrices $T'_\alpha(x)$ is uniformly expanding,*
(iii′.) *each S_α given by $S_\alpha(x, y) = {T'_\alpha}^{*-1}(x).y + f_\alpha(x)$, for some family $\{f_\alpha\}$ of piecewise continuous and uniformly bounded maps $f_\alpha : I_\alpha \to \mathbb{R}^d$. Here ${T'_\alpha}^{*-1}$ denotes the inverse of the transpose of T'_α.*

Let K be the unique fixed point of Θ. Let μ be the marginal measure for m on K by way of $\pi : K \to I$. Denote by $\mathcal{B}, \widetilde{\mathcal{B}}$ the Borel σ-algebras of I, and K, respectively.

If the Lebesgue measure $m(K)$ of K is not zero and $K \setminus \widetilde{T}(K)$ has measure zero, then $(K, \widetilde{\mathcal{B}}, m, \widetilde{T})$ is a model for the natural extension of $(I, \mathcal{B}, \mu, T)$.

2. Basic definitions

2.1. Hausdorff distance on compact sets. — Let $(Y, d) = (Y, d_Y)$ be a complete metric space. If A is a nonempty subset of Y, and $y \in Y$, we define the distance of y to A as $d(y, A) := \inf\{d(y, z) \mid z \in A\}$. For two subsets A, B of Y, we define $\delta(A, B) := \sup_{y \in A} d(y, B)$. This is obviously not symmetric, since $\delta(A, B) = 0$ if $A \subset B$, but symmetrization allows us to define the Hausdorff distance.

DEFINITION 2.1. — The Hausdorff pseudo-distance between two nonempty subsets A, B of Y is defined by $d_H(A, B) := \max(\delta(A, B), \delta(B, A))$.

Alternatively, we introduce the following notation, which will be useful later.

NOTATION 2.2. — For any nonempty set $A \subset Y$, define

$$A_\epsilon = \{y \in Y \mid d(y, A) \le \epsilon\}.$$

Thus, $d_H(A, B)$ is the greatest lower bound of the ϵ such that $A \subset B_\epsilon$ and $B \subset A_\epsilon$.

LEMMA 2.3. — *Given $\epsilon, \delta > 0$ and nonempty sets A, B, C such that $A \subset B_\epsilon$ and $B \subset C_\delta$, one has $A \subset C_{\epsilon+\delta}$.*

Proof. — Let $y \in A$. Then we can find $z \in B$ such that $d(y, z) \leq \epsilon$, and $w \in C$ such that $d(z, w) \leq \delta$, hence $d(y, w) \leq \epsilon + \delta$. This proves that $A \subset C_{\epsilon+\delta}$. □

This implies that d_H satisfies the metric inequality $d_H(A, C) \leq d_H(A, B) + d_H(B, C)$; but the pseudo-distance $d_H(A, B)$ is not a true distance function, since it takes the value zero when evaluated at a set and its closure. However, it becomes a distance when restricted to nonempty compact subsets of Y.

NOTATION 2.4. — We denote by $\mathcal{H}$ the set of nonempty compact subsets of Y.

The following proposition is classical, see [10], Theorem 7.1.

PROPOSITION 2.5. — *The set $\mathcal{H}$ endowed with the Hausdorff pseudo-distance d_H is a complete metric space.*

We will make use of the following lemma, whose elementary proof we leave to the reader.

LEMMA 2.6. — *Let K be a compact set, and let $(A_\alpha)_{\alpha \in \mathcal{A}}$ and $(B_\alpha)_{\alpha \in \mathcal{A}}$ be two sequences of nonempty compact subsets of K indexed by the same countable (possibly finite) set $\mathcal{A}$, such that $d_H(A_\alpha, B_\alpha) \leq \epsilon$ for all $\alpha \in \mathcal{A}$. Let A (resp. B) be the closure of the union of the A_α (resp. B_α). Then $d_H(A, B) \leq \epsilon$.*

2.2. Upper semi-continuous functions. — Let I be a compact Hausdorff space, and let $\pi : I \times Y \to I$ be the projection on the first coordinate. We now consider a compact subset $K \subset I \times Y$ such that $\pi(K) = I$.

One can view K as a fiber bundle over I, whose fiber over x we denote by $K(x)$. Since each fiber is compact, we mainly view K as a map

$$K : I \to \mathcal{H}, \quad x \mapsto K(x) := \{y \in Y | (x, y) \in K\}.$$

This map has a remarkable property: it is upper semi-continuous. Let us recall the appropriate definition.

DEFINITION 2.7. — Let I be a topological space, and let $(\mathcal{H}, \leq)$ be an ordered space in which any subset which has an upper (resp. lower) bound admits a least upper (resp. greatest lower) bound. Let $f : I \to \mathcal{H}$ be a map, and suppose that f is bounded on a neighborhood $V \ni x_0$. The limit superior of f when x tends to x_0 is the greatest lower bound over all neighborhoods $U \subset V$ of x_0 of the least upper bound for all $x \in U$ of $f(x)$. In standard notation, we have $\limsup_{x \to x_0} f(x) = \wedge_{U \subset V; x_0 \in U} \vee_{x \in U} f(x)$.

Note that $f(x_0) \leq \limsup_{x \to x_0} f(x)$. We can now define upper semi-continuous functions.

DEFINITION 2.8. — Let I be a topological space, and let $\mathcal{H}$ be an ordered space in which any subset which has an upper (resp. lower) bound admits a least upper (resp. greatest lower) bound. We say that a map $f : I \to \mathcal{H}$ is upper semi-continuous if for every $x_0 \in I$ we have $\limsup_{x \to x_0} f(x) = f(x_0)$.

We return to the setting that $\mathcal{H}$ is the space of non-empty compact subsets of a complete metric space Y. The inclusion relation on sets gives an order on $\mathcal{H}$. The greatest lower bound of a subset of $\mathcal{H}$ is the intersection of the corresponding subsets of Y if it is not empty, and the least upper bound is given by the closure of the union, if it is compact. Thus, $\mathcal{H}$ satisfies the condition of the definition.

We restrict ourselves to the case that I is a metric space. Using the notation introduced in Notation 2.2, we can reformulate the definition in this setting.

LEMMA 2.9. — *A map $K : I \to \mathcal{H}$ is upper semi-continuous if and only if for any $x \in I$ and any $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, x') < \delta$, then $K(x') \subset K(x)_\epsilon$.*

The proof of the lemma is left to the reader. We can now characterize the functions corresponding to compact subsets of $I \times Y$ with its product topology in terms of the map sending points of I to the corresponding fiber as a subset of Y.

PROPOSITION 2.10. — *A subset $K \subset I \times Y$ with $\pi(K) = I$ is compact if and only if it defines an upper semi-continuous function $K : I \to \mathcal{H}$.*

Proof. — Suppose that K is compact, and that the corresponding function is not upper semi-continuous. Then we can find a point $x \in I$, $\epsilon > 0$, and a sequence $(x_i)_{i \in \mathbb{N}}$ converging to x such that, for all $i \in \mathbb{N}$, $K(x_i) \notin K(x)_\epsilon$. For each x_i, we can hence find y_i such that $d_Y(y_i, K(x)) > \epsilon$. Since $(x_i, y_i) \in K$ for all n, by compactness we can extract a subsequence which converges to a limit of the form (x, y). We must have $d(y, K(x)) \geq \epsilon$, hence (x, y) cannot be in K; but this is impossible, since K is closed.

Conversely, certainly if the map to fibers defined by K does not take all of its values in $\mathcal{H}$, then K is not compact. Now suppose that it does define a function $K : I \to \mathcal{H}$, and that this function is upper semi-continuous. Fix some number $\epsilon > 0$; for any x, we can find an open ball $B(x, \delta) \subset I$ such that, for all $x' \in B(x, \delta)$, $K(x') \subset K(x)_\epsilon$. By compactness, we can cover I with a finite number of balls $B(x_n, \delta_n)$, with $1 \leq n \leq N$. Now define the compact set $L = \cup_{n=1}^{N} K(x_n)_\epsilon$, and note that for any $x \in I$, we have $K(x) \subset L$. That is, all of the fibers of K are contained in a single compact set.

Furthermore, suppose that (x_i, y_i) is a sequence of points in K which converges to some $(x, y) \in I \times Y$. By semi-continuity, $d(y_i, K(x))$ tends to zero, hence $d(y, K(x)) = 0$. Since $K(x)$ is compact, this implies that $y \in K(x)$,

hence $(x, y) \in K$. Thus, K is closed. Since K is closed and contained in the compact set $I \times L$, it is compact. □

DEFINITION 2.11. — We denote by $\mathcal{S}$ the space of upper semi-continuous maps from I to $\mathcal{H}$, endowed with the uniform distance

$$d_{\mathcal{S}}(K, K') := \sup\{d_H\left(K(x), K'(x)\right) \mid x \in I\}.$$

This is well defined, because every $K(x)$ is contained in the compact subset of Y which is the projection of K on the second coordinate, hence the map $x \mapsto d_H(K(x), K'(x))$ is uniformly bounded.

EXAMPLE 2.12. — The space $\mathcal{S}$ is also the space of compact subsets of $I \times Y$ which project onto I, endowed with a particular distance. To aid the reader's intuition, in the case of $I = [0, 1]$, and $Y = \mathbb{R}$, consider the sets

$$A = ([0, 1/2] \times [0, 1]) \cup ([1/2, 1] \times [0, 1/2]),$$
$$B = ([0, 1/2] \times [0, 1 + \epsilon]) \cup ([1/2, 1] \times [0, 1/2]),$$
$$C = ([0, 1/2 + \epsilon] \times [0, 1]) \cup ([1/2 + \epsilon, 1] \times [0, 1/2]),$$

giving $d_H(A, B) = d_H(A, C) = d_{\mathcal{S}}(A, B) = \epsilon$, but $d_{\mathcal{S}}(A, C) = 1/2$.

We will need the following property.

PROPOSITION 2.13. — *The space $\mathcal{S}$ endowed with the uniform distance is complete.*

Proof. — Suppose that K_n is a Cauchy sequence in $\mathcal{S}$. Then for any $x \in I$, the sequence $K_n(x)$ is a Cauchy sequence in $\mathcal{H}$; by completeness of $\mathcal{H}$, it converges to a compact nonempty set $K(x)$. It remains to prove that the function associated to K is upper semi-continuous.

Let $\epsilon > 0$; by definition, we can find an integer N such that, for all $m, n > N$ and all $x \in I$, $d_H(K_m(x), K_n(x)) < \epsilon/3$. This implies in the limit that, for all $n > N$, $d_H(K_n(x), K(x)) \leq \epsilon/3$. Now, choose such an n, and $x \in I$. We can find δ such that, if $d_I(x', x) < \delta$, we have $K_n(x') \subset K_n(x)_{\epsilon/3}$. But we have $K(x') \subset K_n(x')_{\epsilon/3}$ and $K_n(x) \subset K(x)_{\epsilon/3}$. By Lemma 2.3, we have $K(x') \subset K(x)_{\epsilon}$ as soon as $d_I(x', x) < \delta$, hence the limit sequence is in $\mathcal{S}$. □

3. A fixed point theorem

3.1. The setting. — We consider a finite or countable covering $\{I_\alpha\}_{\alpha \in \mathcal{A}}$ of I by compact sets, and a family T of continuous maps $T_\alpha : I_\alpha \to I$ such that $\{T_\alpha(I_\alpha)\}_{\alpha \in \mathcal{A}}$ is a covering of I. We also consider a family of continuous maps $S_\alpha : I_\alpha \times Y \to Y$. For any $x \in I_\alpha$, we denote by $S_{\alpha,x}$ the map $S_{\alpha,x} : Y \to Y$, $y \mapsto S_\alpha(x, y)$.

We suppose that the maps $S_{\alpha,x}$ are uniformly contracting; that is, there exists a positive constant $c < 1$ such that, for all $\alpha \in \mathcal{A}$, for all $x \in I_\alpha$, and

for all $y, z \in Y$, $d_Y(S_{\alpha,x}(y), S_{\alpha,x}(z)) \leq c \, d_Y(y, z)$. We now define a family $\widetilde{T}$ of (skew-product) maps $\widetilde{T}_\alpha : I_\alpha \times Y \to I \times Y$ by $\widetilde{T}_\alpha(x, y) = (T_\alpha(x), S_\alpha(x, y))$.

NOTATION 3.1. — For $\alpha \in \mathcal{A}$ and $K \in \mathcal{S}$ we denote by $K_\alpha = \{(x, y) \in K \mid x \in I_\alpha\}$ the subset of K which projects to I_α. (Since α belongs to $\mathcal{A}$ and not $\mathbb{R}$, this should cause no confusion with K_ϵ.)

ASSUMPTION 3.2. — *If the covering is infinite, we add the condition that, for any $K \in \mathcal{S}$,*

$$\overline{\cup_{\alpha \in \mathcal{A}} \widetilde{T}_\alpha(K_\alpha)}$$

is compact. This is equivalent to assuming the boundedness condition*: that, for some (or equivalently, any) $y \in Y$, the set $\{S_\alpha(x, y) \mid \alpha \in \mathcal{A}, x \in I_\alpha\}$ is totally bounded; this condition is automatically satisfied if $\mathcal{A}$ is finite.*

DEFINITION 3.3. — The set map associated with the skew-product $\widetilde{T}$ is the map

$$\Theta : \mathcal{S} \to \mathcal{S}; \quad \Theta(K) = \overline{\cup \widetilde{T}_\alpha(K_\alpha)}$$

REMARK 3.4. — The set $\Theta(K)$ is in $\mathcal{S}$, because it is compact by hypothesis and it projects onto I since the $T_\alpha(I_\alpha)$ cover I.

3.2. The theorem. — Our first result is the following.

THEOREM 3.5. — *The map Θ has a unique fixed point.*

Proof. — Let K be an element of $\mathcal{S}$, and K' be its image by Θ. By definition, we have

$$K'(x') = \overline{\cup_{\alpha,x; T_\alpha(x)=x'} S_{\alpha,x}(K(x))}$$

Suppose that K_1, K_2 are two elements of $\mathcal{S}$, with $d_{\mathcal{S}}(K_1, K_2) = \delta$. Let $K'_1 = \Theta(K_1)$, $K'_2 = \Theta(K_2)$, and let $x' \in I$. Let α, x be such that $T_\alpha(x) = x'$; we have $d_H(K_1(x), K_2(x)) \leq \delta$. Since $S_{\alpha,x}$ is uniformly contracting by at least c, we have $d_H(S_{\alpha,x}(K_1(x)), S_{\alpha,x}(K_2(x))) \leq c\delta$. Using Lemma 2.6, we deduce that $d_H(K'_1(x'), K'_2(x')) \leq c\delta$. Since this is true for all x', we obtain $d_{\mathcal{S}}(K'_1, K'_2) \leq c\delta$.

We have proved that Θ is a strict contraction by at least c on $\mathcal{S}$; since this space is complete, by the Banach fixed point theorem, Θ has a unique fixed point. □

4. Heuristic models for the Gauss measure and natural extensions of continued fraction maps

We will now apply the result of the previous section, with $Y = \mathbb{R}^d$ and I a compact subset of $\mathbb{R}^d$. We consider a piecewise uniformly expansive surjective map $T : I \to I$, on a compact nonempty subset $I \subset \mathbb{R}^d$ which is the closure of its interior.

NOTATION 4.1. — More precisely, we consider a finite or countable covering $\{I_\alpha\}_{\alpha \in \mathcal{A}}$ of I by compact sets each the closure of its interior, such that the intersection $I_\alpha \cap I_\beta$ of two distinct elements of the covering is of Lebesgue measure zero. We then suppose that there is a constant $C > 1$ and a countable collection of diffeomorphisms $T_\alpha : I_\alpha \to T_\alpha(I_\alpha) \subset I$ with Jacobian matrix $T'_\alpha(x)$ uniformly expanding by at least C: for any $x \in I_\alpha$, and for any tangent vector v, $\|T'_\alpha(x).v\| \geq C\|v\|$, where we denote by $\|v\|$ the Euclidean norm. We also suppose that $\bigcup_{\alpha \in \mathcal{A}} T_\alpha(I_\alpha) = I$.

REMARK 4.2. — The map T is defined by $T(x) = T_\alpha(x)$ if $x \in I_\alpha$. Hence the map might be multiply defined on the intersection of two elements of the covering; however, this intersection has measure zero, and is irrelevant from the measurable viewpoint.

Our goal is to determine an absolutely continuous invariant measure (a Gauss measure) μ for T, and to build a geometric model for the natural extension of the dynamical system $(I, \mathcal{B}, \mu, T)$, where $\mathcal{B}$ is the Borel σ-algebra. (Similarly we will use $\widetilde{\mathcal{B}}$ for the Borel σ-algebra in the geometric model).

DEFINITION 4.3. — The *standard form with parameter f* for the natural extension of T, where f is a family of piecewise continuous and uniformly bounded maps $f_\alpha : I_\alpha \to \mathbb{R}^d$, is the map

$$\widetilde{T} : I \times \mathbb{R}^d \to I \times \mathbb{R}^d$$

$$(x, y) \mapsto \left(T_\alpha(x), T'^{*-1}_\alpha(x).y + f_\alpha(x) \right),$$

where T'^{*-1}_α denotes the inverse of the transpose of T'_α (and as usual we identify $\mathbb{R}^d$ with its dual real vector space).

PROPOSITION 4.4. — *There exists a unique compact set K which projects onto I and is invariant under $\widetilde{T}$.*

Proof. — We can apply the theorem of the previous section, with $S_{\alpha,x}(y) = T'^{*-1}_\alpha(x).y + f_\alpha(x)$; the maps $S_{\alpha,x}$ are by definition contractions of ratio at most $\frac{1}{C}$, and they satisfy the boundedness condition, hence there is a unique $K \in \mathcal{S}$ such that $\Theta(K) = K$. $\qquad\square$

Note that, for any given α, the map $\widetilde{T}_\alpha$ clearly has Jacobian determinant equal to 1, since its Jacobian matrix is a $2d \times 2d$ matrix of the form

$$\begin{pmatrix} T'_\alpha(x) & * \\ 0 & T'^{*-1}_\alpha(x) \end{pmatrix}.$$

Hence the Lebesgue measure of K_α is equal to the measure of its image by $\widetilde{T}_\alpha$.

REMARK 4.5. — See pp. 648–650 of [6] for a discussion of an antecedent of this approach.

DEFINITION 4.6. — Consider the measurable function $\phi : I \to \mathbb{R}$ given by $\phi(x) = m(K(x))$. We define the measure μ on I by

$$\mu(S) := \int_S \phi(x)dx,$$

for S any Borel set.

THEOREM 4.7. — *Let K be the unique fixed point of Θ. If the Lebesgue measure $m(K)$ of K is not zero and $K \setminus \widetilde{T}(K)$ has measure zero (which always holds when $\mathcal{A}$ is finite), then $(K, \widetilde{\mathcal{B}}, m, \widetilde{T})$ is a model for the natural extension of $(I, \mathcal{B}, \mu, T)$.*

Proof. — By construction, $\widetilde{T}_\alpha : K_\alpha \to \widetilde{T}_\alpha(K_\alpha)$ is a bijection which preserves the Lebesgue measure, hence $m(K_\alpha) = m(\widetilde{T}_\alpha(K_\alpha))$. We have $K = \bigcup K_\alpha$ and up to measure zero $K = \widetilde{T}(K) = \bigcup \widetilde{T}_\alpha(K_\alpha)$; hence we have

$$\sum_{\alpha \in \mathcal{A}} m(\widetilde{T}_\alpha(K_\alpha)) = \sum_{\alpha \in \mathcal{A}} m(K_\alpha) = m(K) = m(\widetilde{T}(K)) = m\left(\bigcup \widetilde{T}_\alpha(K_\alpha)\right),$$

and the sets $\widetilde{T}_\alpha(K_\alpha)$ must be disjoint in measure. This implies that $\widetilde{T}$ is a bijection (up to a set of measure zero) which preserves Lebesgue measure.

Since $\pi(K) = I$, the system $(I, \mathcal{B}, \mu, T)$ is a factor of $(K, m, \widetilde{\mathcal{B}}, \widetilde{T})$. It remains to prove that $(K, \widetilde{\mathcal{B}}, m, \widetilde{T})$ is isomorphic to the natural extension, and not a strict extension of it.

An element of the natural extension defines a sequence $(x_n)_{n \in \mathbb{Z}}$ such that $T(x_n) = x_{n+1}$ (this sequence is the projection of the well-defined bi-infinite orbit of the element in the natural extension). Suppose such a sequence is the image of two distinct elements of K. This means that we can find two elements (x_0, y_0) and (x_0, y_0') in K, and two sequences $(y_n)_{n \in \mathbb{Z}}$ and $(y_n')_{n \in \mathbb{Z}}$ such that, for all $n \in \mathbb{Z}$, $\widetilde{T}^n(x_0, y_0) = (x_n, y_n)$ and $\widetilde{T}^n(x_0, y_0') = (x_n, y_n')$. But since the first coordinate is the same for both sequences, at each step the same branch of T is used, and hence for any $n \in \mathbb{Z}$, we have $d(y_n, y_n') \geq Cd(y_{n+1}, y_{n+1}')$.

If y_0 is different from y_0', this implies that $\lim_{n \to -\infty} d(y_n, y_n') = \infty$, which is impossible since K is compact. Hence the projection to the natural extension is one-to-one. □

COROLLARY 4.8. — *If $K \setminus \widetilde{T}(K)$ has measure zero and K has nonempty interior, then $(K, \widetilde{\mathcal{B}}, m, \widetilde{T})$ is a model for the natural extension of T.*

Proof. — This is immediate, since any open subset of $\mathbb{R}^{2d}$ has nonzero Lebesgue measure. □

REMARK 4.9. — The knowledge of K gives us the density of an invariant measure. Indeed, for $x \in I$, and ϕ as in Definition 4.6, we have $\phi(x) = \sum_{x', T(x')=x} |\mathrm{Jac}_{x'} T|^{-1} \phi(x')$. That is, the function ϕ satisfies the classical Ruelle equation for an invariant density.

REMARK 4.10. — There are (obvious) examples where K has an empty interior: If the parameter f of $\widetilde{T}$ is identically 0, then it is immediate that $K = I \times \{0\}$, which has zero Lebesgue measure. Hence we need a good choice of the parameter in order to obtain a domain K with a nonzero measure. The next section will give examples of such a parameter in the case of Möbius transformations.

5. Piecewise homographies

5.1. A natural choice of parameter for the standard form. — We now restrict ourselves to a special case. Suppose that the map T is piecewise homographic and locally increasing on a compact interval $I \subset \mathbb{R}$. More precisely, suppose that we have a finite or countable partition (up to extremities) of I in compact subintervals I_α, and that T is given on I_α by an element $M_\alpha \in \mathrm{PSL}(2, \mathbb{R})$.

REMARK 5.1. — To compute, we must represent M_α by a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R});$$

this matrix is well-defined up to a sign, and we can suppose that $c \geq 0$; we will abuse notations by identifying M_α and this representative, which is uniquely defined if $c \neq 0$.

For this matrix M_α, we have

$$T(x) = \frac{ax + b}{cx + d} \quad \text{and} \quad T'(x) = \frac{1}{(cx + d)^2}.$$

We suppose that T is uniformly expansive, that is, there exists some constant $k < 1$ such that $(cx + d)^2 \leq k$. To use the previous section, we must consider a map

$$\widetilde{T}(x, y) = \left(\frac{ax + b}{cx + d}, (cx + d)^2 y + f(x) \right);$$

the problem is to choose a good function f; as we have seen, the constant zero function will not work.

We will find an heuristic for our function by supposing first that the elements M_α generate a lattice $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$, that is, a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$ such that $\Gamma \backslash \mathrm{PSL}(2, \mathbb{R})$ has finite volume.

We recall here some results from [7], Section 3. Let $A_\gamma \subset \mathrm{PSL}(2, \mathbb{R})$ be the set

$$A_\gamma = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mid \gamma > 0 \right\}.$$

The set A_γ has full Haar measure, as it is the complement of the co-dimension 1 set defined by $\gamma = 0$. On A_γ, we can take α, γ, δ as coordinates, and the Haar measure of $\mathrm{PSL}(2, \mathbb{R})$ is given on A_γ, up to a constant, by $\frac{d\alpha \, d\gamma \, d\delta}{|\gamma|}$.

The diagonal (or *geodesic*) flow is the action on $\mathrm{PSL}(2, \mathbb{R})$ on the right by the one-parameter diagonal group

$$g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}.$$

This preserves Haar measure, and commutes with the action of Γ on the left, hence acts on the quotient $\Gamma \backslash \mathrm{PSL}(2, \mathbb{R})$.

We will obtain $\widetilde{T}$ as a return map of the diagonal flow to a section. Any element in A_γ can be written

$$\begin{pmatrix} xe^t & (xy-1)e^{-t} \\ e^t & ye^{-t} \end{pmatrix}$$

An easy computation shows that, in these coordinates, the Haar measure is given by $dx \, dy \, dt$. The matrices of the form

$$\begin{pmatrix} x & xy-1 \\ 1 & y \end{pmatrix},$$

form a transversal of the diagonal flow on A_γ: any orbit of the diagonal flow on A_γ meets this set in one point. We can now compute the first return map to the corresponding section for the diagonal flow on the quotient space. The product of this matrix by M_α is

$$\begin{pmatrix} ax+b & a(xy-1)+by \\ cx+d & c(xy-1)+d \end{pmatrix};$$

its orbit under the diagonal flow returns to the section at the point

$$\begin{pmatrix} \frac{ax+b}{cx+d} & * \\ 1 & (cx+d)^2y - c(cx+d) \end{pmatrix},$$

and the transverse invariant measure is given by $dx \, dy$.

We thus are lead to define the function $\widetilde{T}$ by

$$\widetilde{T}(x, y) := \left(\frac{ax+b}{cx+d}, (cx+d)^2 y - c(cx+d) \right).$$

This map is a local diffeomorphism with a Jacobian of 1. If the function $c(cx+d)$ is uniformly bounded on I, the results of the previous section apply: There is a unique compact set K such that $\overline{\widetilde{T}(K)} = K$. If this compact set has nonzero measure and satisfies the condition $K \setminus \widetilde{T}(K)$ has measure zero, then

$\widetilde{T} : K \to K$ is a natural extension for T, which preserves Lebesgue measure on K. This allows one to compute an explicit formula for an invariant measure for T that is absolutely continuous with respect to Lebesgue measure.

We will show in the next section that this heuristic works in a large number of cases.

5.2. Some conjugate forms of the natural extension map. — The parametrization above has the nice feature that the invariant measure appears as Lebesgue measure on the invariant domain, with however the inconvenience that the geometric significance of the domain is less straightforward.

We can find other sections of the diagonal flow by considering its geometric interpretation as the geodesic flow on the hyperbolic plane. Recall that one can identify $\mathrm{PSL}(2, \mathbb{R})$ with the unit tangent bundle of the hyperbolic plane; any matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ determines a unique geodesic, with origin $u = \frac{\beta}{\delta}$ and extremity $x = \frac{\alpha}{\gamma}$. The highest point on this geodesic corresponds to a matrix such that $\gamma = \delta$; the endpoints of the geodesic, plus the distance to the highest point, determine another coordinate system on A_γ: each element can be written in a unique way

$$\begin{pmatrix} \frac{xe^s}{\sqrt{x-u}} & \frac{ue^{-s}}{\sqrt{x-u}} \\ \frac{e^s}{\sqrt{x-u}} & \frac{e^{-s}}{\sqrt{x-u}} \end{pmatrix}.$$

The complex coordinate of the highest point of the corresponding geodesic is

$$\frac{x+u}{2} + i\frac{x-u}{2},$$

and in these coordinates the Haar measure is given by

$$\frac{dx\,du\,ds}{(x-u)^2}.$$

Taking $s = 0$, we obtain another section parametrized as

$$\begin{pmatrix} \frac{x}{\sqrt{x-u}} & \frac{u}{\sqrt{x-u}} \\ \frac{1}{\sqrt{x-u}} & \frac{1}{\sqrt{x-u}} \end{pmatrix}$$

which can be identified with the hyperbolic plane; in these coordinates, the first return map to the section is given by

$$(x, u) \mapsto \left(\frac{ax+b}{cx+d}, \frac{au+b}{cu+d} \right),$$

and the invariant measure turns out to be the natural hyperbolic measure, with density $\frac{dx\,du}{(x-u)^2}$.

A variant quite common in number theory consists in replacing the coordinate u by $v = -\frac{1}{u}$. In that case, we consider the same section, but parametrized as

$$\begin{pmatrix} \dfrac{x}{\sqrt{x+\frac{1}{v}}} & \dfrac{-1}{v\sqrt{x+\frac{1}{v}}} \\ \dfrac{1}{\sqrt{x+\frac{1}{v}}} & \dfrac{1}{\sqrt{x+\frac{1}{v}}} \end{pmatrix}.$$

A small computation gives us another heuristic formula in this case:

$$\widehat{T}(x,v) := \left(\frac{ax+b}{cx+d}, \frac{dv-c}{a-bv} \right).$$

One can easily check that this formula can also be written as

$$\widehat{T}(x,v) := \left(M.x, M^{t^{-1}}.v \right);$$

$\widetilde{T}$ and $\widehat{T}$ are conjugate by the map $(x,y) \mapsto \left(x, \frac{y}{1+xy} \right)$, and the map $\widehat{T}$ leaves as invariant the density $\frac{dx\,dv}{(1+xv)^2}$.

Note that these two alternate forms of the natural extension map are locally products, as opposed to honest skew-product maps. As well, $\widehat{T}$ sends rectangles to rectangles, making it particularly convenient. As we will see below, the domain of $\widehat{T}$ has a local product structure, whereas that of $\widetilde{T}$ does not need to have this. On the other hand, the action of $\widetilde{T}$ on its second coordinate is not linear, which raises difficulties in proving that it is uniformly contracting on this coordinate.

6. Examples: classical and others

In this section we give a number of examples, some of which are already in the literature. When this is the case, we content ourselves with a brief presentation of the results with references for the proofs.

6.1. The Gauss map and the Farey map. — The classical Gauss map has been studied in many papers since the letter of Gauss, see for example [9, 1, 39, 2]; it is related to the classical Euclid algorithm, replacing a pair (a,b) with $a < b$ by $(b \mod a, a)$ and renormalizing at each step so that the largest number is 1. More precisely, if we denote the integer part of x by $[x] = \sup\{n \in \mathbb{N}|\, n \le x\}$, and the fractional part of x by $\{x\} = x - [x]$, the Gauss map is defined as

$$T : [0,1] \to [0,1]; \quad x \mapsto \left\{ \frac{1}{x} \right\}.$$

It is locally given by a homography $x \mapsto \frac{1}{x} - n$, related to the matrix $\begin{pmatrix} -n & 1 \\ 1 & 0 \end{pmatrix}$. The map itself is not strictly contracting, since the derivative at $x = 1$ is 1,
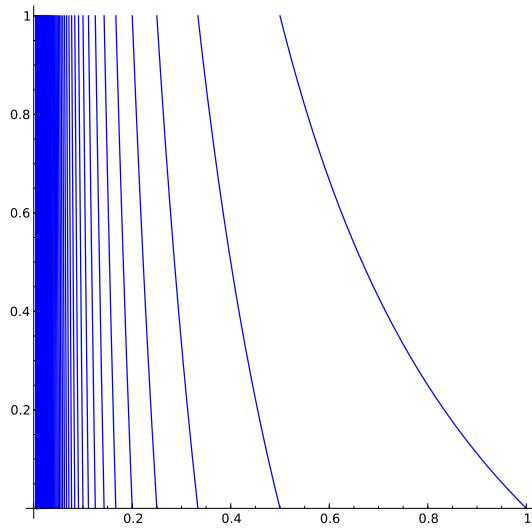
FIGURE 6.1. The graph of the Gauss map.

but its square is strictly contracting; it satisfies the conditions of the theorem, hence there is a unique compact set K invariant by the natural extension

$$\widetilde{T}(x, y) = \left(\left\{ \frac{1}{x} \right\}, x - x^2 y \right).$$

This invariant set is easily found by a numerical experiment, plotting the initial part of the orbit of a generic point: it is the set $K = \{(x, y) \mid 0 \le x \le 1, 0 \le y \le \frac{1}{1+x}\}$, see Figure 6.2. This set is decomposed in a countable Markov partition $K_n = \{(x, y) \mid \frac{1}{n+1} \le x \le \frac{1}{n}, 0 \le y \le \frac{1}{1+x}\}$, see Figure 6.3.

Alternatively, one can consider the other presentation of the natural extension; if we denote by $a(x) := [\frac{1}{x}]$ the first partial quotient, the map is defined by

$$\widehat{T} : [0, 1]^2 \to [0, 1]^2 \quad (x, u) \mapsto \left(\left\{ \frac{1}{x} \right\}, \frac{1}{a(x) + u} \right).$$

This map preserves the invariant density $\frac{dx\,du}{(1+xu)^2}$, and we can see from Figure 6.4 that the point distribution of a generic orbit is not uniform. The map is very natural from an arithmetic viewpoint: it has a countable Markov partition, $\left[\frac{1}{n+1}, \frac{1}{n} \right] \times [0, 1]$, the symbolic dynamic related to this partition is the full shift, and the bi-infinite symbolic dynamic is given by joining the continued fraction expansions of x and u.

Note that this map is locally decreasing, hence it is not defined by a map of $\text{PSL}(2, \mathbb{R})$ (but its square is). This is a problem which occurs frequently in
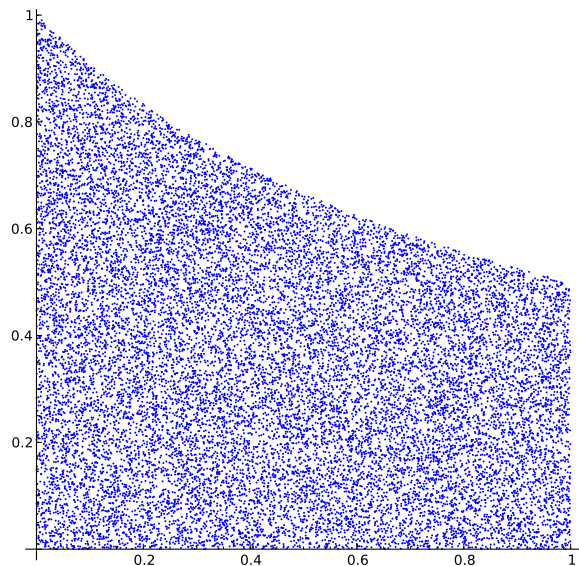
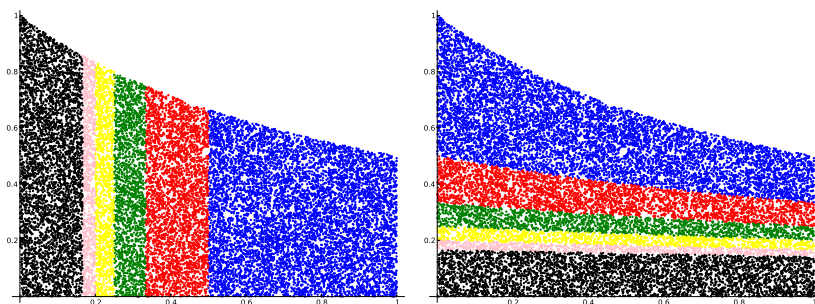FIGURE 6.2. The domain of $\widetilde{T}$: 20,000 points of an orbit.

FIGURE 6.3. The Markov partition of $\widetilde{T}$: five boxes and their image.

the classical theory of continued fractions, and which explains why, in many theorems, one must consider parity of indices. There is a way to get rid of this by symmetrizing the Gauss map: define the *Truncation map* on $\mathbb{R}$ by $\mathrm{trunc}(x) = \{x\}$ if $x > 0$ and $\mathrm{trunc}(x) = -\{-x\}$ if $x < 0$, and the *symmetrized Gauss map S* on $[-1,1]$ by $S(x) := \mathrm{trunc}(-1/x)$; this is a locally increasing function defined by elements of $SL(2, \mathbb{R})$. Using the formula given above, one
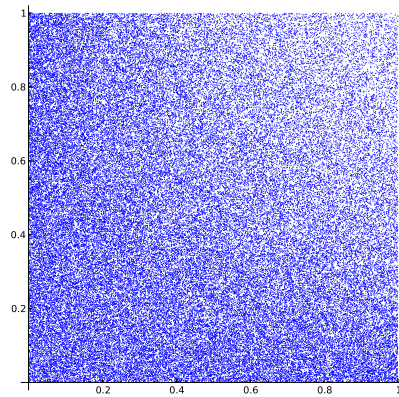
FIGURE 6.4. The domain of $\widehat{T}$: $20,000$ points of an orbit. Note that the point distribution of the orbit is not uniform, compare with Figure 6.2.
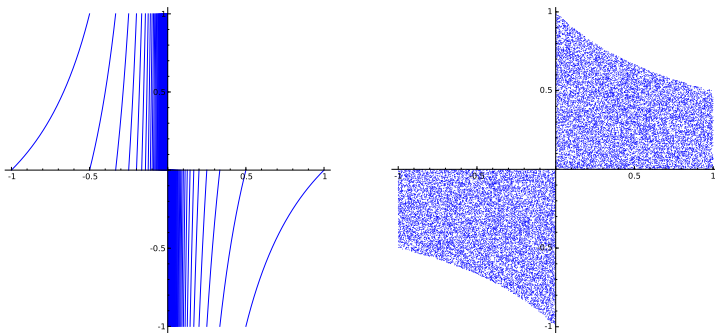

FIGURE 6.5. The symmetrized Gauss map, and 20,000 points of an orbit of its natural extension.

defines a model for the natural extension as $\tilde{S}(x,y) = (S(x), x^2 y - x)$; computation of a generic orbit immediately gives the domain for the natural extension of S, which factors by symmetry on the natural extension of T; see Figure 6.5

The Gauss map admits an additive version, related to the additive euclidean algorithm, of which it is an acceleration, the *Farey map F*, defined on $[0,1]$ by $F(x) = \frac{x}{x-1}$ if $x < \frac{1}{2}$, and $F(x) = \frac{1}{x} - 1$ if $x > \frac{1}{2}$. Using the above formula, one can find a natural extension given by $\tilde{F}(x,y) = (\frac{x}{x-1}, (1-x)^2 y + 1 - x)$ if $x < 1/2$, and $\tilde{F}(x,y) = (\frac{1}{x} - 1, x - x^2 y)$ if $x > \frac{1}{2}$. This map is no longer expanding: it has an indifferent point at the origin, so our theorem does not apply in this case, and
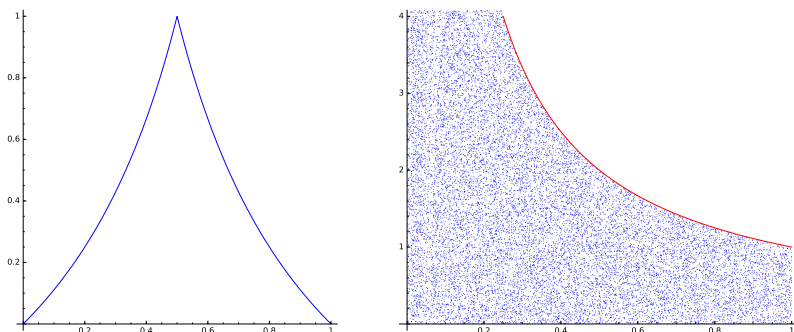
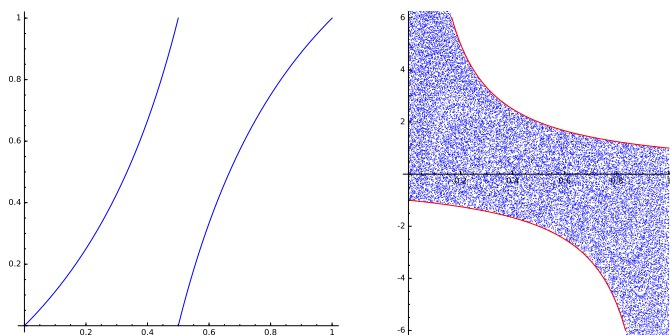FIGURE 6.6. The Farey map, and 20,000 points of an orbit of its natural extension.

FIGURE 6.7. The orientation preserving Farey map, and 20,000 points of an orbit of its natural extension.

there is no invariant compact set. However, computer experiments immediately show that the closed set $\{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq 1, 0 \leq y \leq \frac{1}{x}\}$ is invariant, and gives the infinite invariant measure $\frac{dx}{x}$, see Figure 6.6. The indifferent fixed point precludes the possibility of an absolutely continuous invariant measure.

The Farey maps also admits an orientation preserving version F^+, defined by the same formula as F for $x < \frac{1}{2}$, and by $F^+(x) = 2 - \frac{1}{x}$ for $x > \frac{1}{2}$; its natural extension is given as before by $\widetilde{F}^+(x, y) = (\frac{x}{x-1}, (1 - x)^2 y + 1 - x)$ for $x < \frac{1}{2}$, and by $\widetilde{F}^+(x, y) = (2 - \frac{1}{x}, x^2 y - x)$ for $x > \frac{1}{2}$. This map leaves invariant the domain $\{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq 1, \frac{1}{x-1} \leq y \leq \frac{1}{x}\}$, which gives for F^+ the invariant density $\frac{1}{x(1-x)}$: this density is infinite in 0 and 1, because these are now two indifferent fixed points, see Figure 6.7.
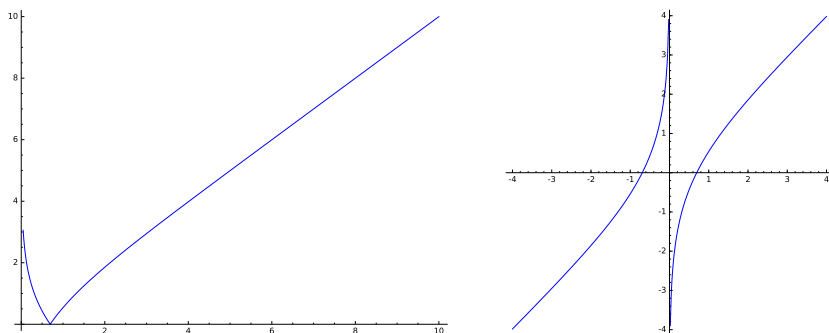
FIGURE 6.8. The graphs of the Lebesgue measure preserving Farey maps.

REMARK 6.1. — It is possible to conjugate the Farey map and the orientation preserving Farey map in such a way that the invariant density becomes constant; in this way, we obtain two curious maps; the first one is defined on $(0, +\infty)$ by

$$x \mapsto |\log(e^x - 1)|$$

and the second is defined on $\mathbb{R}$ by

$$x \mapsto \operatorname{sign}(x) \log(e^{|x|} - 1)$$

It is readily verified, using the Ruelle equation, that these two maps preserve the (infinite) Lebesgue measure on their domain, and it is a consequence of the dynamical definition that this measure is ergodic; the maps are very close to the identity on most of their domain, see Figure 6.8. One can easily give models for the natural extensions of these two maps, in the form of functions of two variables which leaves invariant $(0, +\infty) \times [-1, 1]$ for the first map and $\mathbb{R} \times [-1, 1]$ for the second.

6.2. Ralston continued fraction. — To study precisely the discrepancy of codings of rotations, D. Ralston [35, 36] introduced a variant of the Gauss map; if we denote as above the classical Gauss map by T and by $a(x) = [\frac{1}{x}]$ the first partial quotient of x, he defined a map $R : [0, 1] \to [0, 1]$ by $R(x) = T^2(x)$ if $a(x)$ is even, $R(x) = \frac{1}{1+T(x)}$ if $a(x) > 1$ is odd, and $R(x) = 1 - x$ if $a(x) = 1$.

Some of his results depend on the existence of an absolutely continuous invariant measure, which can be proved by classical techniques, but this is not trivial, since the map is quite complicated (see Figure 6.9); its set of discontinuity points has a countable set of accumulation points.

However, this map satisfies all the hypotheses needed to get an invariant compact set; a tentative model for the natural extension is readily computed,
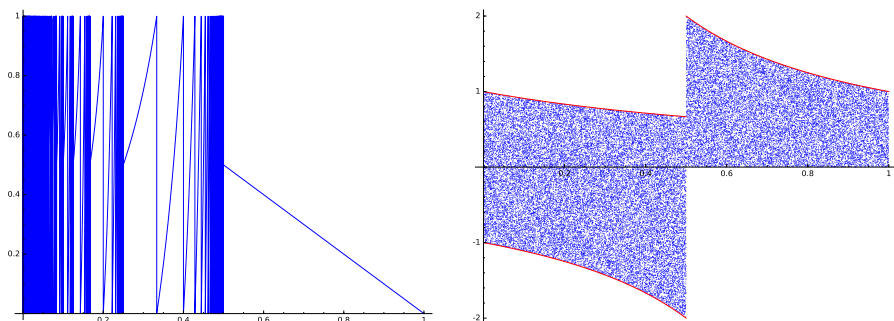
FIGURE 6.9. The Ralston map, and 20,000 points of an orbit
of its natural extension.

see the right part of Figure 6.9, and shows that the invariant set is defined by
the equations $\frac{1}{x-1} \leq y \leq \frac{1}{x+1}$ for $0 \leq x \leq \frac{1}{2}$, and $0 \leq y \leq \frac{1}{x}$ for $\frac{1}{2} \leq x \leq 1$.
This implies that the invariant probability density ϕ is given by $\phi(x) = \frac{2C}{1-x^2}$
for $0 \leq x < \frac{1}{2}$, and by $\phi(x) = \frac{C}{x}$ for $\frac{1}{2} \leq x \leq 1$, with $C = \frac{1}{\log(6)}$.

This could be explicitly proved, either by showing that the Ruelle equation
is satisfied (which would be rather tedious, since listing all the antecedents of
a point is not simple), or, in a more informative way, by understanding how
the images of the various elements of the Markov partition are situated with
respect to each other.

6.3. Nakada α-continued fractions. — There are many variants of the Gauss
map, including those arising from the so-called semi-regular continued frac-
tions. The best known of these are the nearest integer continued fractions
and the backwards continued fractions. Nakada [32] introduced a very inter-
esting family of continued fractions, depending on a real parameter $\alpha \in [0,1]$
and called the (Nakada) α-continued fractions. On $[\alpha - 1, \alpha)$, one defines
$T_\alpha(x) = 1/|x| - \lfloor 1/|x| + 1 - \alpha \rfloor$. The values $\alpha = 1, 1/2, 0$ give the regular,
the nearest integer, and a variant of the backwards continued fractions, respec-
tively.

Nakada used planar models of the natural extensions to determine invari-
ant absolutely continuous measures for T_α, $\alpha \geq 1/2$. Kraaikamp [26] later
confirmed this using his theory of S-fractions. Moussa, Cassa and Marmi [31]
extended this to $\alpha \geq \sqrt{2} - 1$. Luzzi and Marmi [28] extended this for certain
rational values of α in $[0,1]$, in a work that revived interest in these matters.
Carminati and Tiozzo [15] and Kraaikamp-Schmidt-Steiner [27] gave the result
for all α. Central to the study of this family of dynamical systems has been
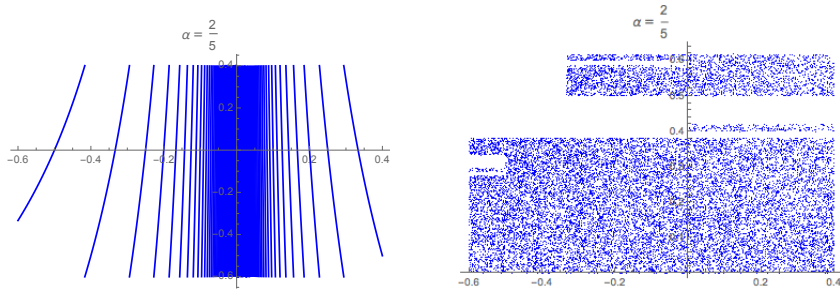
FIGURE 6.10. The Nakada α-CF map, when $\alpha = 2/5$, and 20,000 points of an orbit of its natural extension. Note that the domain is neither a product nor connected.

the relation of the entropy of T_α to the mass of a planar model of its natural extension. See the aforementioned papers and references for more on this matter.

Explicit presentations of (number theoretic versions of) planar models of the natural extensions occur in each of [32, 28, 15, 27]. Note that the map T_α is increasing on the negative portion of its domain and decreasing on the positive; this mixed nature seems to be the cause of the non-product structure of the natural extension. Indeed, it is possible to define an orientation preserving version, replacing $1/|x|$ by $-1/x$, which seems better behaved in this respect. For example, see the natural extensions of Katok-Ugarcovici's [24] a, b-continued fractions. (See also [14].)

6.4. Rosen and Veech continued fractions. — There are various ways to associate interval maps to a Fuchsian group. Rosen [37] defined a continued fraction for each Hecke triangle Fuchsian group. For $q \in \mathbb{N}, q \geq 3$ and $\lambda = \lambda_q = 2\cos\frac{\pi}{q}$, on the interval $[-\lambda/2, \lambda/2)$ the associated map is $T_q(x) = \left|\frac{1}{x}\right| - \lambda\left\lfloor\left|\frac{1}{\lambda x}\right| + 1/2\right\rfloor$. (Thus, when $q = 3$ one has the nearest integer map.) The metric properties of these continued fractions by way of planar natural extensions has seen a great deal of activity in the last 20 years, related work includes [38, 33, 13, 20, 29, 30, 7]. Several authors have also studied one parameter deformations of these and related continued fraction maps, (called α-Rosen maps), see in particular [17].

In these settings, the method of this paper can be used to at least gain insight into the possibilities for planar models of the natural extensions of the maps being studied.

Arnoux and Hubert [3] introduce continued fractions motivated by the study of Veech groups in the setting of translation surfaces. Their continued fractions are directly related to subgroups of the Hecke groups that underlie the Rosen
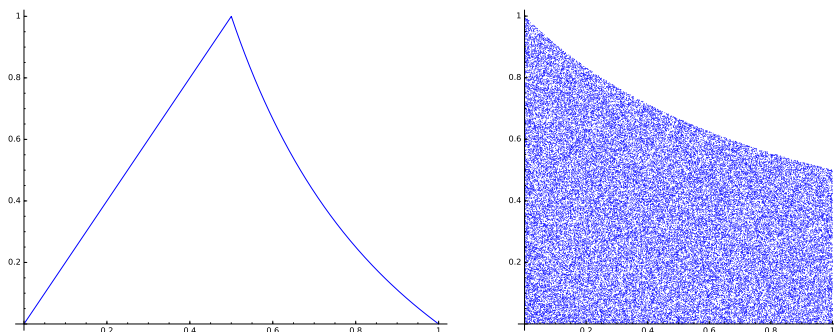
FIGURE 6.11. The additive Hei-Chi Chan map, and 20,000 points of an orbit of its natural extension.

continued fractions. In [8], we used natural extensions (informed by the technique of this paper) to compare these two families of continued fractions. Note that Smillie-Ulcigrai [40] introduced another continued fraction for the study of the geodesics of the translation surface arising from identifying opposite sides of the regular octagon.

6.5. Hei-Chi Chan map. — One can show that any irrational number in $[0, 1]$ can be written in a unique way:

$$x = \cfrac{2^{-a_1}}{1 + \cfrac{2^{-a_2}}{1 + \cdots}}$$

Using this particular continued fraction, Hei-Chi Chan, in [16], defines a generalized Gauss map with an invariant density of the form $\frac{c}{(1+t)(2+t)}$. This map admits an additive variant, for which the computations are particularly simple; we define S_a on $[0, 1]$ by $S_a(x) = 2x$ if $0 \leq x \leq \frac{1}{2}$, and $S_a(x) = \frac{1}{x} - 1$ if $\frac{1}{2} \leq x \leq 1$. The heuristic leads us to conjecture a natural extension of the form $\tilde{S}_a(x, y) = (2x, y/2)$ if $0 \leq x \leq \frac{1}{2}$, and $\tilde{S}_a(x, y) = (1/x - 1, x - x^2 y)$ if $\frac{1}{2} < x \leq 1$, and an experiment (see Figure 6.11) shows that the compact set $\{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq \frac{1}{1+x}\}$ is invariant by S_a.

By a nice coincidence, this additive map has the same invariant domain and the same invariant density as the Gauss map. Note that this map has a fixed point, but it is not indifferent, which explains why this additive map has a finite invariant measure.

The map actually studied by Hei-Chi Chan is an acceleration of S_a; define, for $x \in (0, 1)$, $n(x) = \sup\{k \in \mathbb{N} | 2^k x \leq 1\}$; the multiplicative Hei-Chi
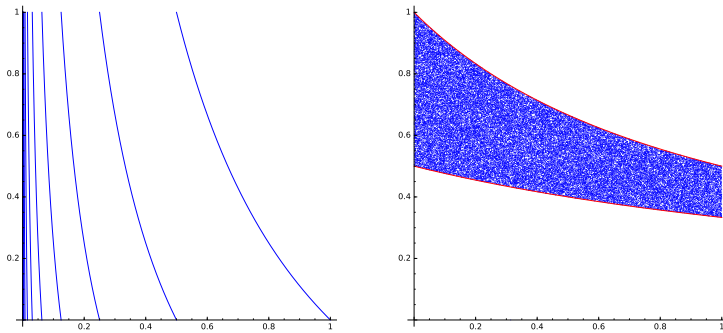
FIGURE 6.12. The multiplicative Hei-Chi Chan map, and 20,000 points of an orbit of its natural extension.

Chan map is given by $S_m(x) = \frac{1}{2^{n(x)}} - 1$. This map has infinitely many decreasing branches, which accumulate exponentially fast to the vertical axis, see Figure 6.11.

Numerical experiment shows that the invariant set is given by the equation $\frac{1}{2+x} \leq y \leq \frac{1}{1+x}$, giving the invariant density $\frac{c}{(1+x)(2+x)}$ found by Hei-Chi Chan in his paper. The domain of the map $\tilde{S}_a$ is in fact the image of the Markov box $\frac{1}{2} \leq x \leq 1$, which is natural: it consists of stopping when we have done a series of multiplication by 2 and an inversion, that is, each time we pass the rightmost Markov box.

REMARK 6.2. — These maps are linked to a curious binary GCD algorithm on integers, defined as follows. Consider two integers $0 < p < q$. First divide them by the largest power of 2 that divides both of them; hence we can suppose that one of them is odd, and their GCD is odd. If $2p < q$, and q is even, divide q by 2. If $2p < q$, and q is odd, multiply p by 2. If $2p \geq q$, substract p from q and reorder; continue until $p = 0$.

This algorithm preserves the GCD; the value of q is non increasing, and it is ultimately decreasing if $p > 0$, hence the algorithm terminates with $p = 0$ in finite time, the value of q giving the GCD.

This algorithm is a variant of the very efficient binary GCD algorithm studied by Vallée [41]; the maps S_a and S_m might prove interesting for studying the properties of this algorithm.

REMARK 6.3. — As opposed to the previous examples, this map does not come from elements of $SL(2, \mathbb{Z})$ or of a discrete group of $SL(2, \mathbb{R})$; this does not prevent the heuristics from working efficiently. However, the importance of powers of 2 in the map, as well as the related binary GCD algorithm on the
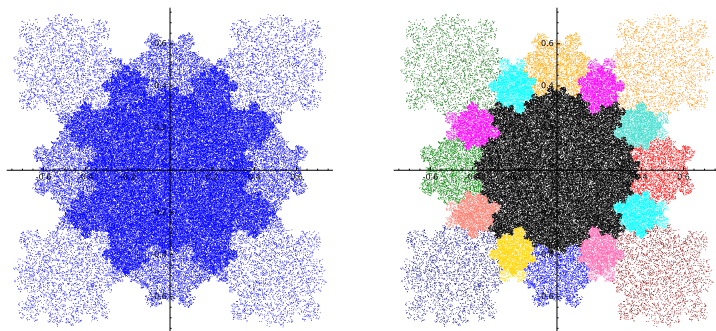
FIGURE 6.13. 100,000 points of an orbit of the natural extension of Hurwitz map (projected on the second coordinate), colored in the right figure according to the previous partial quotient.

integers, hints at the idea that there should be a 2-adic version of this map and its natural extension.

6.6. Hurwitz complex continued fraction. — The Hurwitz continued fraction is simply the extension to the complex plane of the nearest integer continued fraction. More precisely, we consider the Gauss integers $n + mi$, with n, m integers, and for any complex number z, following the notations of Nakada, we denote by $[z]_2$ the nearest Gaussian integer (we ignore the measure zero case when there is more than one such integer). The Hurwitz map H is defined on the unit square

$$\left\{ z \in \mathbb{C} \mid -\frac{1}{2} < \Im(z), \Re(z) < \frac{1}{2} \right\}$$

by

$$H(z) = \frac{1}{z} - \left[\frac{1}{z} \right]_2$$

(similar maps can be defined in an analogous way for other tilings of the complex plane).

This map and its natural extension has been recently the object of study by Ei, Ito, Nakada and Natsui, see [18]; the structure seems quite complicated. One can use the arithmetical version of the natural extension

$$\tilde{H}(z, w) = \left(\frac{1}{z} - \left[\frac{1}{z} \right]_2, \frac{1}{w + \left[\frac{1}{z} \right]_2} \right),$$

or the conjugate form which preserves Lebesgue measure, and our theorem applies, hence there is a compact invariant set K.
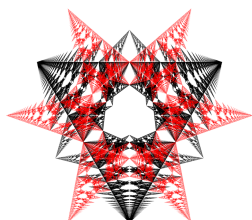
FIGURE 6.14. 2,000,000 points of an orbit of the conjectured natural extension of the modified Poincaré algorithm (projected on the dual coordinates), colored according to the coding of the preimage. (Courtesy of Sébastien Labbé.)

It is easy to do numerical experiments, but difficult to interpret their result because they are in four dimensions, and do not allow simple visualization. It seems however that the invariant compact set has a nonempty interior, hence gives a natural extension; but the characterization of the domain of the natural extension is not simple, as it seems to have a fractal boundary. Figure 6.13 shows 100,000 points of an orbit, projected on the second coordinate; it seems quite apparent that the density is not the same everywhere. This receives a first explanation when we color the orbit according to the previous partial quotient: the incomplete partial quotients of modulus at most $\sqrt{5}$ correspond to outer, and less frequent, pieces of the natural extension.

It seems probable that the family of sets $K_z = \{w \in \mathbb{C}|(z, w) \in K\}$ is locally constant on a finite (17) number of subsets of the unit square. It would be interesting, first to prove it, and then to understand if it still holds when the lattice varies. It would of course be more interesting to determine explicitly the sets K_z, their measure and their boundary.

6.7. Some examples in higher dimensions: Brun and Jacobi-Perron. — Our technique also works in higher dimensions. Some examples can be found in [6] and [4], where it was applied to recover the explicit formula for the invariant density of several multidimensional algorithms as Brun and Selmer, and to find this formula for other algorithms such as the reverse algorithm and the Cassaigne algorithm.

Note however that in some cases it fails spectacularly, for example for the classical Jacobi-Perron algorithm, or for the AR-Poincaré algorithm (see [11]); Figure 6.14, which is due to Sébastien Labbé shows two million points of an orbit. The picture, despite its undeniable aesthetic appeal, gives little clue as to a possible model for the natural extension, and its invariant measure. In particular, it can be seen that sets of points coming from the inner triangle (in

red) and from the outer triangles (in black) seem to intersect on a large set, which might explain the degeneracy in this case.

It also fails in the case of Rauzy induction of interval exchange maps; in that case, direct application of the heuristic gives a fixed point contained in a subspace $H(\pi)$ of lower dimension, and one needs, as was first done by Veech [42], to add new parameters to find the natural extension.

7. Further questions

This paper leaves many questions unanswered. The first, and most important, would be to give sufficient conditions so that the compact invariant set K given by the main theorem has positive measure. In particular, is there a simple verifiable condition guaranteeing that K and $\widetilde{T}(K)$, in the notation of Theorem 4.7, share the same positive measure? (One should expect a condition reminiscent of the open set condition for IFS.) A first step should be to consider the case when the matrices describing a piecewise homographic algorithm generate a lattice in $\mathrm{SL}(2, \mathbb{R})$; this case is similar to the Bowen-Series interval maps, and one might expect to find the natural extension as a return map of the geodesic flow which factors over the given interval map.

More precisely, if the piecewise homographic map is expanding, it is clear that $cx + d$ is bounded, because $(cx + d)^2 < k$ by hypotheses. But c has no reason to be bounded; can we prove that $c(cx + d)$ is always bounded? This is the case in all examples. It would also be interesting to find such a condition for positive measure in higher dimensions, and when this is the case, to find a classical object for which there is a flow corresponding to the suspension of the natural extension of the given map.

As we have seen, one also has convergence, albeit to a non-compact set, in some cases of weakly contracting maps, and the theorem should obviously be extendable to those cases, which are of interest since most additive types of continued fractions admit indifferent fixed points. Note also that these indifferent fixed points have different effects in dimension one, where they lead to unbounded invariant density and to infinite invariant measure, and in higher dimensions, where the invariant density can have a pole while the total measure remains finite.

Note that our basic example is related to matrices in $SL(2, \mathbb{Z})$, hence orientation preserving and unimodular. However, we have seen that the heuristic extends quite well to other situations, either orientation reversing (α-type continued fraction) or non-unimodular (Hei-Chi Chan continued fraction). Is there a more general setting which includes all these cases? In particular, can we extend this to a p-adic setting? Is it true that in the orientation preserving case, the domain of the natural extension is connected?

Finally, to use more efficiently this heuristic method, are there effective ways to visualize and understand the domains we obtain in higher dimensions, maybe by animated views of slices of the domain?

Appendix A. A letter from Gauss to Legendre on continued fractions, translated from French

January 30, 1812

Sir,

I tell you one thousand thanks for the two memoirs you made me the honor to send me, and that I received these past few days. The functions you discuss there, as well as the questions of probabilities on which you prepare a large work, have great appeal for me, although I myself have worked little on those last. I remember nevertheless a curious problem, which I considered 12 years ago, but that I could not then solve to my satisfaction. You might interest yourself with it for a little time, in which case I am sure that you will find a more complete solution. Here is the problem. Let M be an unknown quantity between the limits 0 and 1, for which all values are, either equally probable, or more or less following a given law. We suppose it converted in to a continued fraction

$$M = \cfrac{1}{a' + \cfrac{1}{a'' + \text{etc}}}$$

What is the probability that, stopping the expansion at a finite term $a^{(n)}$, the following fraction

$$\cfrac{1}{a^{(n+1)} + \cfrac{1}{a^{(n+2)} + \text{etc}}}$$

be between the limits 0 and x? I denote it by $P(n, x)$, and I have, supposing for M all values equally probable,

$$P(0, x) = x;$$

$P(1, x)$ is a transcendental function depending on the function

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{x}.$$

which Euler calls unexplainable and on which I just gave several researches in a memoir presented to our society of sciences which will soon be printed. But for the cases where n is larger, the exact value of $P(n, x)$ seems intractable. However, I found by very simple reasoning that, for infinite n, one has

$$P(n, x) = \frac{\log(1 + x)}{\log 2}.$$

But the efforts I made during my researches to assign

$$P(n, x) - \frac{\log(1 + x)}{\log 2}$$

for a very large, but not infinite, value of n have been fruitless.

$$(\cdots)$$

BIBLIOGRAPHY

[1] R. ADLER & L. FLATTO – "Geodesic flows, interval maps, and symbolic dynamics", *Bull. Amer. Math. Soc. (N.S.)* **25** (1991), no. 2, p. 229–334.

[2] P. ARNOUX – "Le codage du flot géodésique sur la surface modulaire. (French) [Coding of the geodesic flow on the modular surface]", *Enseign. Math. (2)* **40** (1994), no. 1–2, p. 29–48.

[3] P. ARNOUX & P. HUBERT – "Fractions continues sur les surfaces de Veech", *J. Anal. Math.* **81** (2000), p. 35–64.

[4] P. ARNOUX & S. LABBÉ – "On some symmetric multidimensional continued fraction algorithms", *Ergodic Theory and Dynamical Systems* (2017), p. 1–26.

[5] P. ARNOUX, M. MIZUTANI & T. SELLAMI – "Random product of substitutions with the same incidence matrix", *Theoret. Comput. Sci.* **543** (2014), p. 68–78.

[6] P. ARNOUX & A. NOGUEIRA – "Mesures de gauss pour des algorithmes de fractions continues multidimensionnelles", *Ann. Sci. École Norm. Sup. (4)* **26** (1993), no. 6, p. 645–664.

[7] P. ARNOUX & T. A. SCHMIDT – "Cross sections for geodesic flows and α-continued fractions", *Nonlinearity* **26** (2013), p. 711–726.

[8] _____ , "Commensurable continued fractions", *Discrete and Continuous Dynamical Systems – Series A (DCDS-A)* **34** (2014), no. 11, p. 4389–4418.

[9] E. ARTIN – "Ein mechanisches System mit quasi-ergodischen Bahnen", *Abh. Math. Sem. Hamburg* **1982** (1924), p. 170–175.

[10] M. F. BARNLEY – *Fractals everywhere, Second edition. revised with the assistance of and with a foreword by Hawley Rising, III*, Academic Press Professional, Boston, MA, 1993.

[11] V. BERTHÉ & S. LABBÉ – "Factor complexity of S-adic words generated by the Arnoux-Rauzy-Poincaré algorithm", *Adv. Appl. Math.* **63** (2015), p. 90–130.

[12] W. BOSMA, H. JAGER & F. WIEDIJK – "Some metrical observations on the approximation by continued fractions", *Nederl. Akad. Wetensch. Indag. Math.* **45** (1983), no. 3, p. 281–299.

[13] R. BURTON, C. KRAAIKAMP & T. SCHMIDT – "Natural extensions for the Rosen fractions", *Trans. Amer. Math. Soc.* **352** (2000), p. 1277–1298.

[14] K. Calta, C. Kraaikamp & T. A. Schmidt – "Synchronization is full measure for all α-deformations of an infinite class of continued fraction transformations", *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, to appear; 50 pp.

[15] C. Carminati & G. Tiozzo – "A canonical thickening of $\mathbb{Q}$ and the entropy of α-continued fraction transformations", *Ergodic Theory Dynam. Systems* **32** (2012), no. 4, p. 1249–1269.

[16] H. C. Chan – "A Gauss-Kuzmin-Lévy theorem for a certain continued fraction", *International Journal of Mathematics and Mathematical Sciences* **2004** (2004), no. 20, p. 1067–1076.

[17] K. Dajani, C. Kraaikamp & W. Steiner – "Metrical theory for α-Rosen fractions", *J. Eur. Math. Soc. (JEMS)* **11** (2009), no. 6, p. 1259–1283.

[18] H. Ei, S. Ito, H. Nakada & R. Natsui – "On the construction of the natural extension of the Hurwitz complex continued fraction map", *Monatsh. Math.* **188** (2019), no. 1, p. 37–86.

[19] C. F. Gauss – "Letter to Laplace", in *Gauss Werke*, no. X, part 1, 1917, p. 371–374.

[20] K. Gröchenig & A. Haas – "Backward continued fractions and their invariant measures", *Canad. Math. Bull.* **39** (1996), p. 186–198.

[21] J. Hilgert & A. Pohl – "Symbolic dynamics for the geodesic flow on locally symmetric orbifolds of rank one", in *Infinite dimensional harmonic analysis IV*, no. 97111, NJ, World Sci. Publ., Hackensack, 2009.

[22] J. Hutchinson – "Fractals and self similarity", *Indiana Univ. Math. J.* **30** (1981), no. 5, p. 713–747.

[23] S. Katok & I. Ugarcovici – "Symbolic dynamics for the modular surface and beyond", *Bull. Amer. Math. Soc. (N.S.)* **44** (2007), no. 1, p. 87–132.

[24] _____ , "Structure of attractors for (a,b)-continued fraction transformations", *J. Mod. Dyn.* **4** (2010), no. 4, p. 637–691.

[25] M. Keane – "*A continued fraction titbit*, symposium in honor of benoit mandelbrot (curaçao, 1995)", *Fractals* **3** (1995), no. 4, p. 641–650.

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

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

[28] L. Luzzi & S. Marmi – "On the entropy of japanese continued fractions", *Discrete Contin. Dyn. Syst.* **20** (2008), p. 673–711.

[29] D. Mayer & T. Mühlenbruch – "Nearest λ_q-multiple fractions", *Spectrum and dynamics* **52** (2010), p. 147–184.

[30] D. Mayer & F. Strömberg – "Symbolic dynamics for the geodesic flow on hecke surfaces", *J. Mod. Dyn.* **2** (2008), no. 4, p. 581–627.

[31] P. Moussa, A. Cassa & S. Marmi – "Continued fractions and Brjuno functions", *J. Comput. Appl. Math.* **105** (1999), no. 1–2, p. 403–415.

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

[33] _____, "Continued fractions, geodesic flows and Ford circles", *Algorithms, fractals, and dynamics: Okayama/Kyoto* **1992** (1995), p. 179–191.

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

[35] D. Ralston – "Substitutions and 1/2-discrepancy of $\{n\theta + x\}$", *Acta Arith.* **154** (2012), no. 1, p. 1–28.

[36] _____, "1/2 -Heavy sequences driven by rotation", *Monatsh. Math.* **175** (2014), p. 595–612.

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

[38] T. A. Schmidt – "Remarks on the Rosen λ-continued fractions", in *Number theory with an emphasis on the Markoff spectrum* (A. Pollington & W. Moran, éds.), Dekker, New York, 1993, p. 227–238.

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

[40] J. Smillie & C. Ulcigrai – "Beyond Sturmian sequences: coding linear trajectories in the regular octagon", *Proc. Lond. Math. Soc. (3)* **2** (2011), no. 102, p. 291–340.

[41] B. Vallée – "Euclidean dynamics", *Discrete and Continuous Dynamical Systems – Series S, American Institute of Mathematical Sciences* (2006), p. 281–352.

[42] W. Veech – "Gauss measures for transformations on the space of interval exchange maps", *Ann. Math.* **115** (1982), p. 201–242.