LEBESGUE ERGODICITY OF A DISSIPATIVE SUBTRACTIVE ALGORITHM

HENK BRUIN

Abstract. We prove Lebesgue ergodicity and exactness of a certain dissipative 2-dimensional subtractive algorithm, completing a partial answer by Fokkink et al. to a question by Schweiger. This implies for Meester's subtractive algorithm in dimension d, that there are $d-2$ parameters which completely determine the ergodic decomposition of Lebesgue measure.

1. Introduction

Consider a triple $x = x^{(0)} = (x_1, x_2, x_3)$ of positive reals, and form a sequence $(x^{(n)})_{n\geq 0}$, by repeatedly subtracting the smallest of the three from the other two. This dynamical system emerged from a percolation problem studied by Meester [M]. Although $(x^{(n)})_{n\geq 0}$ is clearly a decreasing sequence, $x^\infty = \lim_{n\to\infty} x^{(n)}$ is different from 0 for Lebesgue-a.e. initial position. Let us write this more formally as iterations of the subtractive map of increasing triples $0 \leq x_1 \leq x_2 \leq x_3$:

$$F(x_1, x_2, x_3) = \mathbf{sort}(x_1, x_2 - x_1, x_3 - x_1),$$

where **sort** stands for putting the coordinates in increasing order. It is obvious that $x_1^\infty = x_2^\infty = 0$, but also that if $x_3 > x_1 + x_2$, then $\eta := x_3 - (x_1 + x_2)$ is a preserved quantity. This means that once $x_3 > x_1 + x_2$, the third coordinate will always remain the largest, even under the unsorted subtractive algorithm, and in fact $x_3^\infty = \eta$. Meester and Nowicki [MN] showed that for Lebesgue-a.e. initial vector, there is indeed some $n \geq 0$ such that $\eta = x_3^{(n)} - (x_1^{(n)} + x_2^{(n)}) > 0$.

Therefore F is non-ergodic w.r.t. Lebesgue measure λ: triples with different non-negative values of η have disjoint orbits, and thus belong to 'carriers' of different ergodic components, which can be defined in the usual way even though λ is non-invariant and in fact dissipative. Let us recall these definitions.

Definition 1. *A transformation $(X, \mathcal{B}, \lambda; T)$ is*

- non-singular *if $\lambda(B) > 0$ implies $\lambda(T(B)) > 0$;*
- ergodic *if $T^{-1}(B) = B$ implies that $\lambda(B) = 0$ or $\lambda(X \setminus B) = 0$;*
- conservative *if for every set $B \in \mathcal{B}$ of positive measure, there is $n \geq 1$ such that $\lambda(T^n(B) \cap B) > 0$;*
- dissipative *if it is fails to be conservative, and* totally dissipative*, if there is no invariant subset $X_0 \subset X$ of positive measure on which T is conservative;*
- exact *if $T^{-n} \circ T^n(B) = B$ for all $n \geq 0$ implies that $\lambda(B) = 0$ or $\lambda(X \setminus B) = 0$.*

All of these properties can be defined even though λ is not T-invariant.

The hospitality of the Mittag-Leffler Institute in Stockholm (2010 Spring programme on Dynamics and PDEs) is gratefully acknowledged. This paper grew out of a project on subtractive algorithm for which I would like to thank Delft University of Technology for its hospitality over the summer of 2009.

The result of [MN] was generalised by Kraaikamp and Meester [KM] to dimension $d \geq 3$. They showed that for the map

$$F_d(x_1, \dots x_d) = \mathbf{sort}(x_1, x_2 - x_1, \dots, x_d - x_1),$$

and Lebesgue-a.e. initial vector x, the quantity $\eta_3 = x_3^{(n)} - (x_1^{(n)} + x_2^{(n)})$ is eventually positive, and so is $\eta_k := x_k^{(n)} - x_{k-1}^{(n)}$ for $k > 3$. Once $\eta_3 > 0$, all η_k are preserved, and, as observed in [FKN], Lebesgue measure is therefore not ergodic. This answers in the negative a question posed by Schweiger [S]. The natural question, however, is whether the level sets

$$\{x \in \mathbb{R}^d_{\geq 0} : x_k^\infty = \sum_{j=3}^k \eta_j \text{ for all } 3 \leq k \leq d\}$$

constitute the ergodic decomposition of Lebesgue measure.

We can rephrase this question by passing from projective space (on which F_d acts) to a fixed simplex $\Delta = \{x = (x_1, \dots, x_d) : 0 \leq x_1 \leq \dots \leq x_d = 1\}$, by scaling the largest coordinate to 1. The map F_d then becomes $f_d : \Delta \to \Delta$, defined as

$$\begin{cases} x' = F_d(x) = \mathbf{sort}(x_1, x_2 - x_1, \dots x_d - x_1), \\ f_d : x \mapsto \frac{1}{x'_d}\, x'. \end{cases}$$

For $d = 2$, the map F_d reduces to the Farey map

$$x \mapsto \begin{cases} \frac{x}{1-x} & \text{if } x \in [0, \frac{1}{2}]; \\ \frac{1-x}{x} & \text{if } x \in [\frac{1}{2}, 1]. \end{cases} \tag{1}$$

In the next simplest case $d = 3$, we know that $\lim_{n\to\infty} f_d^n(x_1, x_2, 1) = (0, 0, 1)$ as soon as $x_1 + x_2 < 1$, so f_d is totally dissipative on the simplex Δ.

Nogueira [N] used properties of $GL(2, \mathbb{Z})$ to prove that, although dissipative, the three-dimensional system is Lebesgue ergodic. In this paper we use a different method (based on a transient random walk argument with a Lebesgue typical speed of "convergence to 0", combined with distortion estimates) to reprove ergodicity. Our method also yields Lebesgue exactness, and is, we hope, adaptable to similar (higher-dimensional) systems as well, see also Remark 1.

Theorem 1. *Partition the triangle* $\Delta = \{(x, y) : 0 \leq x \leq y \leq 1\}$ *into* $\Delta_L = \{(x, y) : 0 \leq 2x \leq y \leq 1\}$, $\Delta_R = \{(x, y) : 0 < x \leq y < 2x \leq 1\}$ *and* $\Delta_T = \{(x, y) : \frac{1}{2} < x \leq y \leq 1\}$.

Then with respect to the map $f : \Delta \to \Delta$ *defined as*

$$f(x, y) = \begin{cases} (\frac{y-x}{x}, \frac{1-x}{x}) & \text{if } (x, y) \in \Delta_T, \\ (\frac{y-x}{1-x}, \frac{x}{1-x}) & \text{if } (x, y) \in \Delta_R, \\ (\frac{x}{1-x}, \frac{y-x}{1-x}) & \text{if } (x, y) \in \Delta_L, \end{cases}$$

see Figure 1, Lebesgue measure is totally dissipative, ergodic and exact.

It follows from [KM] that for $d \geq 3$ and Lebesgue-a.e. initial vector x, there is $n \in \mathbb{N}$ such that for $x_3^{(n)} > x_2^{(n)} + x_1^{(n)}$, so this case reduces to Theorem 1 as well. In fact, we have the corollary:

Corollary 1. *For each* $\eta_3 > 0$ *and* $\eta_4, \dots, \eta_d \geq 0$, *the map* F_d *restricted to the invariant set* $\{x \in \mathbb{R}^d_{\geq 0} : x_k^\infty = \sum_{j=3}^k \eta_j$ *for* $3 \leq k \leq d\}$ *is ergodic and exact w.r.t. Lebesgue measure.*

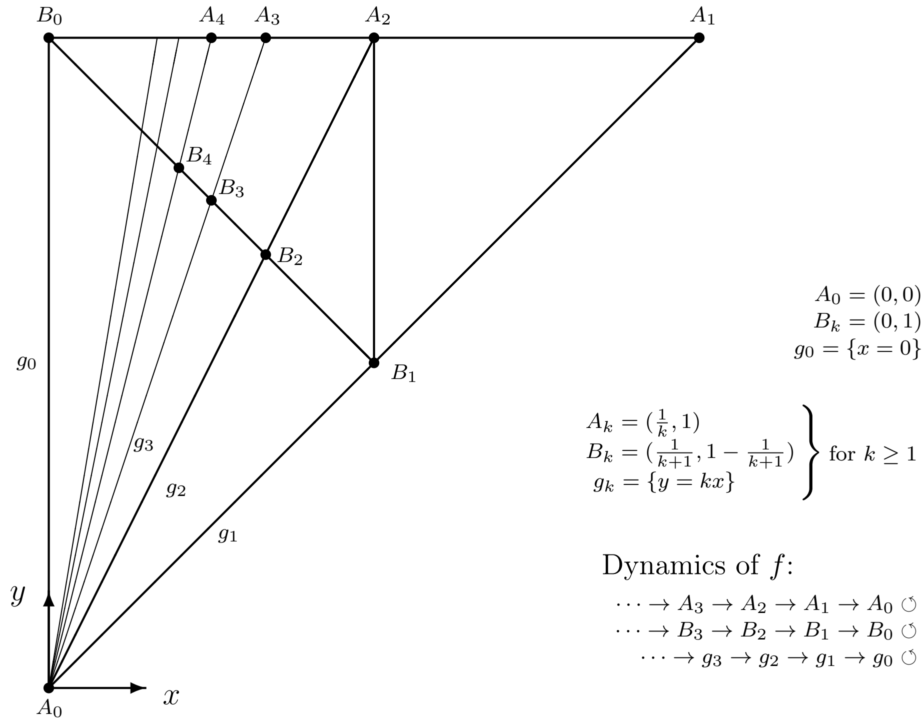

FIGURE 1. The Markov partition for partition $f : \Delta \to \Delta$ consists of the triangles Δ_L (to the left of the line $g_1 = \{y = 2x\}$), Δ_R (between g_1 and the line $\{x = \frac{1}{2}\}$) and Δ_T (to the right of $\{x = \frac{1}{2}\}$). Each of these triangles is mapped onto Δ by f. Further diagonal lines g_k bound the regions where the first return times to Δ_R are constant (namely k between g_k and g_{k+1}). The line $\{x + y = 1\}$ is invariant and separates the part where $\eta > 0$ and where η is not yet determined.

Proof. Since $\eta_3 > 0$, we can divide the space $\{x : x_3^\infty = \eta_3\}$ into a countable union $\cup_{\tau \geq 0} X_\tau$ where $\tau = \min\{n \geq 0 : F_d^n(x)_3 > F_d^n(x)_1 + F_d^n(x)_2\}$. That is, after τ iterations, the order of the coordinates $F_d^\tau(x)_k$ for $3 \leq k \leq d$ will not change anymore under further iteration. (In fact $F_d^\tau(x)_k = \sum_{j=3}^k \eta_j + F_d^\tau(x)_1 + F_d^\tau(x)_2$.) So from this iterate onwards, we can scale so that $F_d^\tau(x)_3 = 1$ and restrict our attention to the first two coordinates. Theorem 1 applies to them. □

Remark 1. *Meester and Nowicki's result was generalised by Fokkink et al.* [FKN] *to a two-parameter setting, called Schweiger's fully subtractive algorithm, see* [S, Chapter 9]*:*

$$F_{ad}(x_1, \dots x_d) = \mathbf{sort}(x_1, \dots, x_a, x_{a+1} - x_a, \dots, x_d - x_a).$$

Analogous quantities η_k for $k \geq a+2$ are still preserved as soon as $\eta_{a+2} \geq 0$, and [FKN] *shows that this happens almost surely. The present paper shows Lebesgue ergodicity and exactness of the level sets of $(\eta_2, \dots, \eta_d)$ for $F_{1,d}$ and all $d \geq 3$. It is hoped that the techniques will be useful to understand F_{ad} for general $a \in \{1, 2, \dots, d-2\}$.*

2. The proof of Theorem 1

2.1. **Finding convenient coordinates.** To start the proof, it helps to recall from [FKN] the Markov partition of Δ that f possesses, see Figure 1. The Markov partition $\Delta = \Delta_L \cup \Delta_R \cup \Delta_T$ consists of three full branches. In fact, f extends to a diffeomorphism

$f : \overline{\Delta_i} \to \overline{\Delta}$ for $i = L, R, T$. The region Y under the line $x + y = 1$ is invariant; it is here that $\eta = 1 - x - y > 0$, and $f^n(x, y) \to (0, 0)$ for every $(x, y) \in Y$. Clearly $f(\Delta_T) \supset Y$, and an additional distortion argument ensures that Lebesgue-a.e. (x, y) eventually falls into Y. Therefore f is totally dissipative.

The question is whether the convergence to $(0, 0)$ is so chaotic, that $\lambda|_Y$ is in fact ergodic or even exact. Let us restrict our Markov partition to

$$\{Y_L = \Delta_L \cap Y,\ Y_R = \Delta_R \cap Y\},$$

and study the first entry map $G : Y \to Y_R$ in a new set of coordinates. First note that the lines $g_k = \{(x, y) \in Y : y = kx\}$, $k \geq 1$, and $g_0 = \{(x, y) \in Y : x = 0\}$ satisfy $f(g_k) = g_{k-1}$ for $k \geq 1$ and g_0 consists of neutral fixed points. Hence the return time to Y_R on the region between g_{k+1} and g_{k+2} is exactly k for $k \geq 1$. For fixed $t \geq 0$, the lines $\ell(p, t) = \{(x, y) \in Y : y = p - tx\}$, $0 < p \leq 1$, foliate Y and

$$f(\ell(p, t) \cap Y_L) = \ell(p, t + 1 - p), \quad f(\ell(p, t) \cap Y_R) = \ell\left(\frac{p}{t + 1 - p}, \frac{1}{t + 1 - p}\right).$$

Therefore, if $A_n(p, t) \subset \ell(p, t) \cap Y_R$ is a maximal arc on which the first return time is n, then

$$G_n(p, t) := G(A_n(p, t)) = \ell\left(\frac{p}{t + 1 - p}, \frac{n + (n-1)t - 2(n-1)p}{t + 1 - p}\right) \cap Y_R.$$

Remark 2. *The point $(0, 0)$ is attracting under G, but not quite under f itself. Namely, on Y_L,*

$$Df|_{Y_L}(0, 0) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

which is a nilpotent shear, whereas on Y_R,

$$Df|_{Y_R}(0, 0) = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix},$$

which is hyperbolic with stable eigenvalue $\lambda_s = \frac{1}{2}(\sqrt{5} - 1)$ on stable eigenspace $E_s =$ span$(\lambda_s, 1)^T$ (where ^{T} stands for the transpose) and unstable eigenvalue $\lambda_u = -\frac{1}{2}(\sqrt{5}+1) < -1$ on unstable eigenspace $E_u =$ span$(\lambda_u, 1)^T$. Therefore, if

$$(p_k, t_k) = G_{n_1 \dots n_k} := G_{n_k} \circ G_{n_{k-1}} \circ \dots \circ G_{n_1}(p, t)$$

for successive return times $(n_k)_{k \in \mathbb{N}}$, then $t_k \to \frac{1}{2}(\sqrt{5} + 1)$ as $k \to \infty$ and $n_j = 1$ for all large j, whereas t_k immediately becomes large if n_k is large.

Remark 3. *For each (p, t), the length of $A_n(p, t)$ is $1/n(n+1)$ times the length of $\ell(p, t) \cap Y_R$. Let*

$$A_{n_1 \dots n_k}(p, t) = \{x \in \ell(p, t) \cap Y_R : \text{the first } k \text{ return times to } Y_R \text{ are } n_1, \dots, n_k\}.$$

Its length is approximately $\prod_{i=1}^k n_i^{-2}$. Each map $G^k : A_{n_1 \dots n_k}(p, t) \to Y_R$ acts as the Gauss map with corresponding uniform distortion control, see Lemma 2. Therefore, the conditional probability $\mathbb{P}(n_{k+1} = n \mid n_1 \dots n_k\}) \sim n^{-2}$, uniformly in k and the history $n_1, \dots, n_k$. The process $(S_k)_{k \in \mathbb{N}}$ given by $S_k(x) = n_1 + \dots + n_k$ if $x \in A_{n_1 \dots n_k}$ (which is a cone over $A_{n_1 \dots n_k}(1, 1)$) is a deterministic version of the one-sided discrete Cauchy walk. Taking the difference of two sample paths of such a walk, we obtain a symmetric two-sided Cauchy walk, i.e., *a random walk where the steps are distributed according to $\mathbb{P}(X_k = n) = \mathbb{P}(X_k = -n) \sim cn^{-2}$. This walk is recurrent, as follows from more general* *theory on stable laws (see* [D, Theorem 2.9][1]*), so for λ-a.e. pair $(z,z') \in Y_R^2$, there are infinitely many k, such that their respective sums $S_k = S'_k$,* i.e., *$f^k(z)$ and $f^k(z')$ both belong to Y_R. For our proof, however, it suffices to have the somewhat weaker result proved in Proposition 1.*

Let us write $p = \frac{p}{\alpha+\beta t+\gamma p}$ and $t = \frac{\hat\alpha+\hat\beta t+\hat\gamma p}{\alpha+\beta t+\gamma p}$, for integers $\alpha, \beta, \gamma, \hat\alpha, \hat\beta . \hat\gamma$, so the initial values are $\alpha = \hat\beta = 1$ and $\hat\alpha = \beta = \gamma = \hat\gamma = 0$. Direct computation gives:

$$G_n\left(\ell(\frac{p}{\alpha+\beta t+\gamma p}, \frac{\hat\alpha+\hat\beta t+\hat\gamma p}{\alpha+\beta t+\gamma p})\right) =$$

$$Y_R \cap \ell\left(\frac{p}{\alpha+\hat\alpha+(\beta+\hat\beta)t+(\gamma+\hat\gamma-1)p},\right.$$

$$\left.\frac{n\alpha+(n-1)\hat\alpha+(n\beta+(n-1)\hat\beta)p+(n\gamma+(n-1)\hat\gamma-2(n-1))p}{\alpha+\hat\alpha+(\beta+\hat\beta)t+(\gamma+\hat\gamma-1)p}\right).$$

This means that the iteration of G, for initial values $p \in (0,1]$ and $t \geq p$, we find that we can represent the iterations

$$(p_k, t_k) = G_{n_1 \dots n_k}(p,t) = \left(\frac{p}{\alpha_k+\beta_k t+\gamma_k p}, \frac{\hat\alpha_k+\hat\beta_k t+\hat\gamma_k p}{\alpha_k+\beta_k t+\gamma_k p}\right) \tag{2}$$

by affine transformations on the integer vectors $(\alpha, \hat\alpha, \beta, \hat\beta, \gamma, \hat\gamma)^T$:

$$\begin{pmatrix} \alpha \\ \hat\alpha \\ \beta \\ \hat\beta \\ \gamma \\ \hat\gamma \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ n & n-1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & n & n-1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & n & n-1 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \hat\alpha \\ \beta \\ \hat\beta \\ \gamma \\ \hat\gamma \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 2(n-1) \end{pmatrix}$$

with initial value $(1,0,0,1,0,0)^T$ mapping to $(1,n,1,n-1,-1,-2(n-1))^T$, etc. It is easy to check by induction that

$$\begin{cases} \alpha_k+\beta_k+\gamma_k = \hat\alpha_k+\hat\beta_k+\hat\gamma_k = 1 & \text{for all } k \geq 0, \\ \beta_k \leq \alpha_k \leq 2\beta_k, & \text{for all } k \geq 0, \\ \hat\beta_k \leq \hat\alpha_k \leq 2\hat\beta_k, & \text{except that } \hat\beta_1 = 0 \\ & \quad \text{when } n_1 = 1, \\ \alpha_k \leq \hat\alpha_k \leq 2\alpha_k & \text{when } n_k = 2. \end{cases} \tag{3}$$

Therefore, as far as asymptotics are concerned, it suffices to keep track of α_k and $\hat\alpha_k$ (or just of α_k whenever $n_k = 2$), cf. Proposition 1, so it makes sense to focus just on the recursive relation

$$\begin{cases} \alpha_{k+1} &= \alpha_k + \hat\alpha_k, \\ \hat\alpha_{k+1} &= n_{k+1}\alpha_k + (n_{k+1}-1)\hat\alpha_k, \end{cases} \qquad \alpha_0 = 1, \ \hat\alpha_0 = 0. \tag{4}$$

[1]In fact, the Cauchy distribution models the position on the horizontal axis where a standard random walk on $\mathbb{Z}^2$, starting from $(0,0)$ returns to the horizontal axis. Since the standard random walk on $\mathbb{Z}^2$ is recurrent, the Cauchy walk is recurrent as well.

In fact, there is $\Lambda = \Lambda(p,t)$, but independent of k, such that

$$1 \le \frac{\alpha_k + \beta_k t + \gamma_k p}{\alpha_k}, \frac{\hat\alpha_k + \hat\beta_k t + \hat\gamma_k p}{\hat\alpha_k} \le \Lambda, \tag{5}$$

whenever $t \ge p$ and $n_k = 2$.

2.2. **Distortion results.** Given intervals $J' \subset J$, we say that J is a δ-*scaled* neighbourhood of J' if both component of $J \setminus J'$ have length $\ge \delta|J'|$. The following Koebe distortion property is well-known, see [MS, Section IV.1]: If $g : I \to J$ is a diffeomorphism with *Schwarzian derivative* $Sg := g'''/g' - 3/2(g''/g')^2 \le 0$, then for every $I' \subset I$ such that J is a δ-scaled neighbourhood of $J' := g(I')$, the distortion

$$\sup_{x,y\in I'} \left|\frac{g'(x)}{g'(y)}\right| \le K(\delta) := \left(\frac{1+\delta}{\delta}\right)^2. \tag{6}$$

Möbius transformations g have zero Schwarzian derivative, so (6) holds for g and g^{-1} alike.

Lemma 1. *The foliation of Y into radial lines*

$$h_\theta = \{(r\cos\theta, r\sin\theta) : 0 \le r \le (\sin\theta + \cos\theta)^{-1}\}$$

with $\theta \in [\pi/4, \pi/2]$ is invariant. Moreover, the distortion of $G^k : h_\theta \to h_{\theta_k}$ is bounded in the sense of (6) *uniformly in $\theta \in [0, \pi/2]$ and $k \in \mathbb{N}$.*

Proof. Since f preserves lines and $(0,0)$ is fixed, the invariance of the foliation is immediate.

Let t_k be as in (2) and θ_k the angle of the image of h_θ under $G_{n_1\dots n_k}$. The line $\ell(1,t_k)$ and h_{θ_k} intersect at a point $(R_k\cos\theta_k, R_k\sin\theta_k)$ for $R_k = (\cos\theta_k + t_k\sin\theta_k)^{-1}$. Using (2) again, we see that $G_{n_1\dots n_k}$ acts on the parameter r as a Möbius transformation

$$M_k : r \mapsto R_k \, \frac{r}{1+\beta_k(1-r)},$$

which has zero Schwarzian derivative, and so has its inverse. Therefore, within an interval $J \Subset [0,R_0]$ such that both components of $[0,R_0] \setminus J$ have length $\delta|J|$, the distortion $\sup_{r_0,r_1\in J} |M_k'(r_0)|/|M_k'(r_1)|$ is bounded by $K(\delta)$ uniformly in k and $n_1,\dots,n_k$. □

The following lemma is straightforward, using $d = 1$ in (6).

Lemma 2. *The map f preserves the line $\ell(1,1) = \{(x,y) : x+y = 1\}$ and acts on it like the Farey map* (1)*. Hence the return map G acts like the Gauss map, and the distortion of every branch $G_{n_1\dots n_k} : \cup_{n\ge n_k} A_{n_1\dots n_k n} \to \ell(1,1)$ is uniformly bounded by $K = 4$.*

2.3. **Growth of α_k and $\hat\alpha_k$ at different points.** Let $\alpha_k(x)$ and $\hat\alpha_k(x)$ be as in (4). The first component of the expression (2), together with (5), shows that the $\alpha_k(x)$ roughly dictate the distance between $F^k(x)$ and the origin. Hence the following proposition should be interpreted as: typical pairs of points infinitely often visit regions of similar distance to the origin.

Proposition 1. *There is $L \ge 10$ such that for Lebesgue-a.e. $(x,y) \in Y_R^2$,*

$$\frac{1}{L} \le \frac{\alpha_k(x)}{\alpha_l(y)}, \frac{\hat\alpha_k(x)}{\hat\alpha_l(y)} \le L \qquad \text{for infinitely many } k,l \in \mathbb{N}. \tag{7}$$

Proof. The heuristics behind proving (7) is that the numbers $\log\alpha_k$ are dominated by random variables

$$X_k = \sum_{j=1}^k \lceil 3\log n_j \rceil.$$

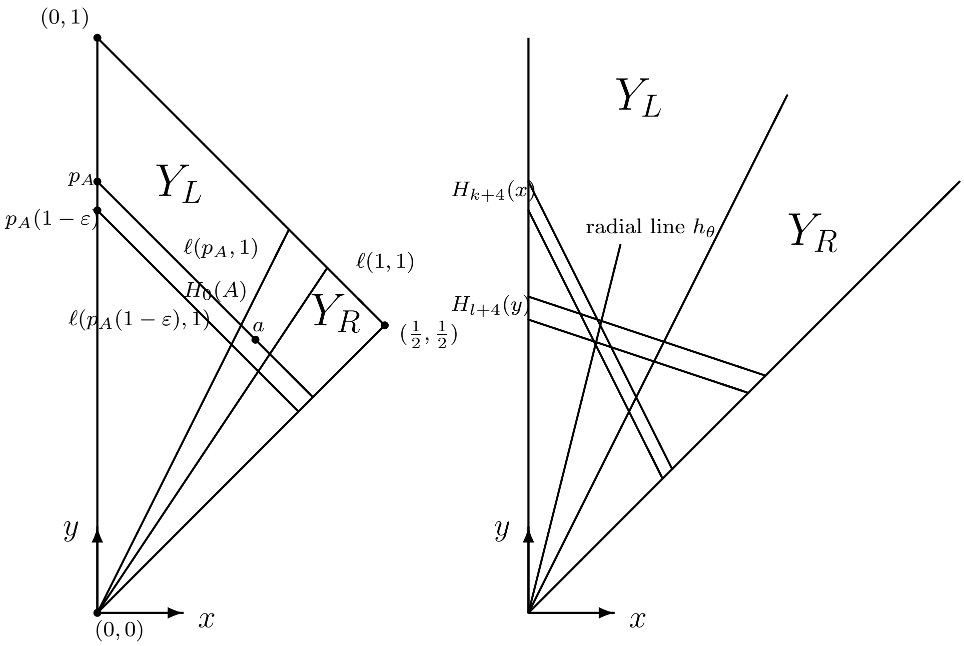

FIGURE 2. Left: The lines $\ell(p_A, 1)$ and $\ell(p_A(1-\varepsilon), 1)$ enclose$H_0(A)$ and the area of large density near a. Right: The strips $H_{k+4}(x)$ and $H_{l+4}(y)$ must intersect.

This follows immediately from (4). The probabilities $\mathbb{P}(\lceil 3 \log n_k \rceil = t) = O(e^{-t/3})$ for all k and t, so X_k is the sum of k random variables of finite expectation μ. Standard probability theory (see *e.g.* [D, Theorem 4.1]) gives that $\frac{1}{r}\#\{k : X_k \in [0, r]\} \to 1/\mu > 0$ as $r \to \infty$. Therefore, almost every sample path of $\{\Gamma_k\}_{k\in\mathbb{N}}$ is a sequence with positive density, and since $\log \alpha_k \le X_k$ also for λ-a.e. x, the sequence $(\log \alpha_k)_{k\in\mathbb{N}}$ has positive density. It follows that there is L_0 such that for $\lambda \times \lambda$-a.e. pair (x, y), there are infinitely many integers k, l such that $|\log \alpha_k(x) - \log \alpha_l(y)| \le L_0$. Taking the exponential function, we obtain the required result for α_k in (7). Since $\hat{\alpha}_k = n_k \alpha_{k-1} + (n_k - 1)\hat{\alpha}_{k-1}$ and the event $\{n_k = 2\}$ is basically independent of the previous choices of n_j, the result for $\hat{\alpha}_k$ in (7) follows as well. □

2.4. **The main proof.** The total dissipativity of f already follows from [FKN]; it is a direct consequence of $f^n(x, y) \to (0, 0)$ Lebesgue-a.e. We will now finish the proof of Theorem 1.

Proof. Assume that $A, A' \subset Y_R$ are sets of positive measure such that $f^{-1}(A) = A$ and $f^{-1}(A') = A'$. To prove ergodicity, we will find some $i, j \in \mathbb{N}$ such that $f^i(A) \cap f^j(A') \ne \emptyset$, so A and A' cannot be disjoint.

Use coordinates $u \in [0, 1]$, $v \in [0, p]$ to indicate points below the line $\ell(p, 1)$: $(x, y) = (uv, u(p-v))$. First take $a = (v_A, p_A - v_A)$ a density point of A, where it is not restrictive to assume that $p_A \in (0, 1)$. By Fubini's Theorem, we can find $\varepsilon \in (0, 1-p_A)$ such that, letting $H_0(A)$ be the strip between parallel lines $\ell(p_A, 1)$ and $\ell(p_A(1-\varepsilon), 1)$ (see Figure 2, left), there is a set $V_A \in [0, p_A]$ of positive measure such that $\{u \in [1-\varepsilon, 1] : (uv, u(p_A - v)) \notin A\}$ has measure $\le \varepsilon/(10KL)$ for every $v \in V_A$ and K as in (6) and L as in Proposition 1.

Since a 1-scaled neighbourhood of $[p_A(1-\varepsilon), p_A]$ is still contained in $[0, 1]$, we can choose $K = 4$ here as the common distortion bound in Lemmas 1 and 2. We can also assume that v_A is a density point of V_A.

We do the same for A', finding a point $p_{A'} \in (0,1)$, a set $V_{A'} \subset [0, p_{A'}]$ of positive measure and a density point $a' = (v_{A'}, p_{A'} - v_{A'})$ of $V_{A'}$.

By Proposition 1, it is not restrictive to assume that $a = (v_A, p_A - v_A)$ and $a' = (v_{A'}, p_{A'} - v_{A'})$ satisfy:

$$\frac{1}{L} \leq \frac{\alpha_k(a)}{\alpha_l(a')}, \ \frac{\hat{\alpha}_k(a)}{\hat{\alpha}_l(a')} \leq L \qquad \text{and} \qquad n_k = n'_l = 2$$

for infinitely many $k, l \in \mathbb{N}$. Let $Z_{n_1 \dots n_k} \ni a$ denote the k-cylinder set containing a, intersected with $H_0(A)$. Then $G^k(Z_{n_1 \dots n_k}) = H_k(A) \cap Y_R$, where $H_k(A)$ is the strip between the lines $G^k(\ell(p_A, 1)$ and $G^k(\ell(p_A(1-\varepsilon), 1))$. Due to the small difference between initial values p_A and $p_A(1-\varepsilon)$, formula (2) gives that these lines are roughly parallel.

Applying (4) twice we get

$$\begin{cases} \alpha_{k+2} &= (n_{k+1}+1)\alpha_k + n_{k+1}\hat{\alpha}_k, \\ \hat{\alpha}_{k+2} &= (n_{k+2}n_{k+1} + n_{k+2} - n_{k+1})\alpha_k + (n_{k+2}n_{k+1} - n_{k+1} + 1)\hat{\alpha}_k. \end{cases} \tag{8}$$

For $x \in Z_{n_1 \dots n_k}$, the variables $\alpha_k(x), \hat{\alpha}_k(x), \beta_k(x), \hat{\beta}_k(x), \gamma(x)$ and $\hat{\gamma}_k(x)$ are all well-defined and constant. By choosing $x \in Z_{n_1 \dots n_k}(a)$ so that $n_{k+2}(x) = n_{k+1}(x) = 1$ (which corresponds to choosing a $k+2$-subcylinder $Z_{n_1 \dots n_k 11}(x)$), formula (8) simplifies to

$$\begin{cases} \alpha_{k+2} &= 2\alpha_k + \hat{\alpha}_k, \\ \hat{\alpha}_{k+2} &= \alpha_k + \hat{\alpha}_k, \end{cases}$$

and we have $\hat{\alpha}_k(x) \leq \alpha_{k+2}(x) \leq 2\alpha_{k+2}(x)$ for each x in this subcylinder. In view of (2) and (5), this means that the slope of the strip $H_{k+2}(a)$ is between Λ and $1/\Lambda$. More precisely:

$$\frac{1}{\Lambda} \leq t_{k+2}(x) \leq \Lambda \qquad \text{for each } x \in Z_{n_1 \dots n_k 11}.$$

Similarly for cylinder $Z_{n'_1 \dots n'_l} \ni a'$, choosing also $n'_{l+1} = n'_{l+2} = 1$ and taking a similar $l+2$-subcylinder $Z_{n'_1 \dots n'_l}$, we find $\Lambda' = \Lambda'(p_{A'}, \varepsilon)$ such that $\frac{1}{\Lambda'} \leq t_{k+2}(y) \leq \Lambda'$ for each $y \in Z_{n_1 \dots n_l 11}$.

Furthermore,

$$\frac{1}{L} \leq \frac{\alpha_{k+2}(x)}{\alpha_{l+2}(y)}, \ \frac{\hat{\alpha}_{k+2}(x)}{\hat{\alpha}_{l+2}(y)} \leq L,$$

which implies that

$$\frac{1}{\Lambda L} \leq \frac{p_{k+2}(x)}{p_{l+2}(y)} \leq \Lambda L \qquad \text{for all } x \in Z_{n_1 \dots n_k 11} \text{ and } y \in Z_{n'_1 \dots n'_k 11}.$$

In other words, $H_{k+2}(x)$ and $H_{l+2}(y)$ are two strips of roughly the same slope and ordinates $p_{k+2}(x)$ and $p_{l+2}(y)$ differing by no more than a uniform factor ΛL.

The next step is to choose a $k+4$-subcylinder of $Z_{n_1 \dots n_k 11}$ and a $l+4$-subcylinder of $Z_{n'_1 \dots n'_l 11}$ so that their images $H_{k+2}(x)$ and $H_{l+2}(y)$ must intersect. We use (3) and (8) for $k+4$ instead of $k+2$ to find

$$\begin{aligned} p_{k+4} &= \frac{p_{k+4}}{p_{k+2}}\, p_{k+2} = \frac{\alpha_{k+2} + \beta_{k+2}t + \gamma_{k+2}p}{\alpha_{k+4} + \beta_{k+4}t + \gamma_{k+4}p}\, p_{k+2} \\ &= \frac{\alpha_{k+2} + \beta_{k+2}t + \gamma_{k+2}p}{(n_{k+3}+1)(\alpha_{k+2} + \beta_{k+2}t + \gamma_{k+4}p) + n_{k+3}(\hat{\alpha}_{k+2} + \hat{\beta}_{k+2} + \hat{\gamma}_{k+2}p)}\, p_{k+2} \\ &\sim \frac{p_{k+2}}{n_{k+3}} \end{aligned}$$

and

$$\begin{aligned} t_{k+4} &= \frac{t_{k+4}}{t_{k+2}}\, t_{k+2} = \frac{\hat\alpha_{k+4}+\hat\beta_{k+4}t+\hat\gamma_{k+4}p}{\alpha_{k+4}+\beta_{k+4}t+\gamma_{k+4}p}\ \frac{\alpha_{k+2}+\beta_{k+2}t+\gamma_{k+2}p}{\hat\alpha_{k+2}+\hat\beta_{k+2}t+\hat\gamma_{k+2}p}\ t_{k+2} \\ &= \frac{\begin{array}{r}(n_{k+4}n_{k+3}+n_{k+4}-n_{k+3})(\alpha_{k+2}+\beta_{k+2}t+\gamma_{k+2}p)+\\ (n_{k+4}n_{k+3}-n_{k+3}+1)(\hat\alpha_{k+2}+\hat\beta_{k+2}t+\hat\gamma_{k+2}p\end{array}}{(n_{k+3}+1)(\alpha_{k+2}+\beta_{k+2}t+\gamma_{k+4}p)+n_{k+3}(\hat\alpha_{k+2}+\hat\beta_{k+2}t+\hat\gamma_{k+2}p)} \\ &\quad\cdot \frac{\alpha_{k+2}+\beta_{k+2}t+\gamma_{k+2}p}{\hat\alpha_{k+2}+\hat\beta_{k+2}t+\hat\gamma_{k+2}p}\cdot t_{k+2} \\ &\sim \frac{n_{k+4}}{n_{k+3}} \end{aligned}$$

By interchanging the role of A and A', we can assume that

$$p_{l+2}(y) \le p_{k+2}(x) \le \Lambda L p_{l+2}(y)$$

for and $x \in Z_{n_1\dots n_{k+4}}$, $y \in Z_{n'_1\dots n'_{l+4}}$. Next choose $10 < M < 2\Lambda L$ and

$$n'_{l+3} = n_{k+4} = 2, \quad n'_{l+4} = n_{k+3} = 8M,$$

so that $\sum_{j=1}^4 n_{k+j} = \sum_{j=1}^4 n'_{l+j} = 4+8M$. Furthermore, for x in the corresponding $k+4$-subcylinder of $Z_{n_1\dots n_k}$, and y in the corresponding $l+4$-subcylinder of $Z_{n'_1\dots n'_l}$, we have

$$t_{k+4}(x) \sim \frac{1}{4M}, \quad t_{l+4}(y) \sim 4M.$$

Let $H_{k+4}(x)$ be entire strip between $\ell(p_{k+4}(x), t_{k+4}(x))$ and $\ell(p_{k+4}(x(1-\varepsilon)), t_{k+4}(x(1-\varepsilon)))$, and similarly for $H_{l+4}(y)$. By the above estimates on p_{k+4} and t_{k+4}, we see that $H_{k+4}(x)$ and $H_{l+4}(y)$ intersect, see Figure 2, right.

The foliation of Y into radial lines h_θ is invariant, see Lemma 1. There is an interval Θ, depending only on ε and M, such that if $\theta \in \Theta$ then the radial line h_θ intersects $H_{k+4}(x) \cap H_{l+4}(y)$. More precisely, the length

$$|h_\theta \cap H_{k+4}(x) \cap H_{l+4}(y)| \ge \frac{p_{k+4}(x)\varepsilon}{4M} \ge \frac{1}{8L}\min\left\{|h_\theta \cap H_{l+4}(y)|,\ |h_\theta \cap H_{k+4}(x)|\right\}.$$

Write $h(v)$ for the radial line intersecting the point $(v, p_A - v)$, and similarly for $h(w)$. If these lines are chosen such that both $G^{k+4}(h(v))$ and $G^{l+4}(h(w))$ are subset of h_θ, and $v \in V_A$, $w \in V_{A'}$, then we derive from the definition of V_A and $V_{A'}$, using the distortion bound K in Lemma 1, that

$$G^{k+4}(h(v) \cap A) \cap G^{l+4}(h(w) \cap A') \ne \emptyset.$$

Since v_A and $v_{A'}$ are density points of V_A and $V_{A'}$ respectively, we can assume that k and l are so large that the relative measure of $V_{A'}$ in $\cup_{n'\ge M} Z_{n'_1\dots n'_{l+2}n'} \cap \ell(p_{A'}, 1)$ is at least $1 - |\Theta|/2K$ and similarly, the relative measure of V_A in $\cup_{n\ge 1} Z_{n_1\dots n_{k+2}n} \cap \ell(p_A, 1)$ is at least $1 - |\Theta|/2K$.

Recall that $K = 4$ is also the uniform distortion bound for iterates of the Gauss map in Lemma 2, and that $G|_{\ell(1,1)}$ acts as the Gauss map. Thus expressed in terms of polar angle $\theta \in [\pi/4, \pi/2]$, the distortion bound is similar.

From this we can conclude that for each θ in a subset of Θ of positive measure, h_θ indeed intersects both $G^{k+4}(H_0(A) \cap h(v))$ for some $v \in V_A$ and $G^{l+4}(H_0(A') \cap h(w))$ for some

$w \in V_{A'}$. Therefore $h_\theta \cap G^k(A) \cap G^l(A') \neq \emptyset$, proving that $f^i(A) \cap f^j(A') \neq \emptyset$ for some $i, j \geq 0$. This concludes the ergodicity proof.

Now to prove exactness, we invoke [BH, Proposition 2.1], which states that a non-singular ergodic transformation $(X, \mathcal{B}, \lambda; T)$ is exact if and only if for every set $A \in \mathcal{B}$ of positive measure there is $n \in \mathbb{N}$ such that $\lambda(T^{n+1}(A) \cap T^n(A)) > 0$. Choosing $a = (v_A, p_A - v_A)$ for density point $v_A \in V_A$ and $\varepsilon \in (0, 1 - p_A)$ as before, we can assume that $(n_i(a))_{i \in \mathbb{N}}$ contain infinitely many k such that $n_k(a) = n_{k+1}(a) = 1$. Let us consider the $k+2$-subcylinder $Z_{n_1 \dots n_{k+1} 1}$ as the set A' with $a' = (v_{A'}, p_{A'} - v_{A'})$ for $p_{A'} = p_B$ and $v_{A'}$ a density point of $V_{A'} = V_A$. Also set $l = k+1$. Then the above methods show that $G^{k+4}(A') \cap G^{l+4}(A)$ intersect, and since $A' \subset A$, we have verified the above condition for exactness with $n = k+4$, $n+1 = l+4$. □

References

[BH] H. Bruin, J. Hawkins, *Exactness and maximal automorphic factors of unimodal interval maps,* Ergod. Th. & Dyn. Sys. **21** (2001), 1009–1034.

[D] T. Durrett, *Probability. Theory and examples,* The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1991.

[FKN] R. Fokkink, C. Kraaikamp, H. Nakada, *On Schweiger's problems on fully subtractive algorithms,* Israel J. Math. **186** (2011), 285–296.

[KM] C. Kraaikamp, R. Meester, *Ergodic properties of a dynamical system arising from percolation theory,* Ergod. Th. & Dyn. Sys. **15** (1995), no. 4, 653–661.

[M] R. Meester, *An algorithm for calculating critical probabilities and percolation functions in percolation models defined by rotations,* Ergod. Th. & Dyn. Sys. **9** (1989), 495–509.

[MN] R. Meester, T. Nowicki, *Infinite clusters and critical values in two-dimensional circle percolation,* Israel J. Math. **68** (1989), no. 1, 63–81.

[MS] W. de Melo and S. van Strien, *One-Dimensional Dynamics,* Springer, Berlin Heidelberg New York, (1993).

[N] A. Nogueira, *The three-dimensional Poincaré continued fraction algorithm,* Israel J. Math. **90** (1995), 373–401.

[S] F. Schweiger, *Multidimensional continued fractions,* Oxford University Press, Oxford, 2000.

Faculty of Mathematics
University of Vienna
Oskar Morgensternplatz 1
A-1090 Vienna, Austria
henk.bruin@univie.ac.at