THE CONVERGENCE OF THE GENERALISED SELMER ALGORITHM

HENK BRUIN, ROBBERT FOKKINK, COR KRAAIKAMP

Abstract. Schweiger introduced the notion of a subtractive algorithm, to classify certain types of multidimensional continued fractions. We study the limit behaviour of one particular subtractive algorithm, which generalises a continued fraction algorithm that was originally proposed by Selmer. The algorithm that we study depends on two parameters a and b. We first find a Markov partition if $a \geq b$. Using inducing techniques, we then prove the existence of an ergodic absolutely continuous invariant probability measure if $a \geq b$. Finally, a theorem of Lagarias is shown to give estimates for the quality of the rational approximations for Lebesgue-typical multidimensional vectors.

1. Introduction

1.1. **Subtractive algorithms.** For positive integers a, b with sum $d = a + b$, let X_d be the space of sorted $a + b$-tuples $\vec{x} = (x_1, \ldots, x_d)$ with $0 \leq x_1 \leq x_2 \leq \cdots \leq x_d$. Obviously, X_d is a d-dimensional simplicial cone. The map $F_{a,b} : X_d \to X_d$ is defined by

$$F_{a,b}(x_1, \ldots, x_a, x_{a+1}, \ldots, x_{a+b}) = \texttt{sort}(x_1, \ldots, x_a, x_{a+1} - x_1, \ldots, x_{a+b} - x_1),$$

where the sorting rearranges the coordinates into non-decreasing order. The map $F_{a,b}$ is piecewise linear, and the simplicial subcones on which it is linear form the time-1-partition of X_d. If we iterate $F_{a,b}$ at an arbitrary initial point $\vec{x} \in X_d$ then the limit $\vec{x}^{\infty} := \lim_{k\to\infty} F_{a,b}^k(\vec{x})$ exists by monotonicity and it is a fixed point of $F_{a,b}$ by continuity. Therefore, the first coordinate of $\vec{x}^{\infty}$ is equal to zero. How many more zero coordinates can we expect? This is one of the motivating questions for our paper.

If $a = b = 1$ then $F_{a,b}$ is a homogeneous version of the Farey map, and maps like $F_{a,b}$ have emerged predominantly from number theory. Fritz Schweiger [11] has coined the term *subtractive algorithm* for maps like $F_{a,b}$. A more general subtractive algorithm would be

$$F_{a,b,c}(x_1, \ldots, x_a, x_{a+1}, \ldots, x_{a+b}) = \texttt{sort}(x_1, \ldots, x_a, x_{a+1} - x_c, \ldots, x_{a+b} - x_c),$$

for $c \leq a$. In this paper we restrict ourselves to $c = 1$. If $b = c = 1$, then the map is known as Selmer's algorithm since it was first considered by Selmer [14]. This is why we call $F_{a,b,1}$ the *generalised Selmer algorithm*. We denote it simply by $F_{a,b}$.

Subtractive algorithms $F_{a,b,c}$ have been studied for special values of a, b, c. If $c = a$, then $F_{a,b,a}$ is called the *fully subtractive algorithm*, which has been studied in [7, 8, 4], [11, Ch 9] and in [3, 10]. In the latter papers it is shown that for Lebesgue almost every $\vec{x}$ the limit $\vec{x}^{\infty}$ has its first $a + 1$ coordinates equal to zero, but its $a + 2$-nd coordinate is

Date: Version of June 16, 2014.

2000 *Mathematics Subject Classification.* 11A55, 11J70, 11K50, 11K55, 28D05.

Key words and phrases. Selmer's algorithm, subtractive algorithm, Diophantine approximation, invariant measure .

positive. See also [12] for variations on the Poincaré map studied in [10]. The subtractive algorithm $F_{a,1,c}$ is considered in [11, Ch 8], where it is shown that $\vec{x}^{\infty}$ is equal to the origin for Lebesgue almost every $x \in X_d$. This analysis has recently been extended for the special case $a = 3, c = 2$ in [13]. For these parameters, Schweiger has shown that the cone X_d contains invariant simplicial subcones that are different from the time-1-partition. Finally, we mention that the definitive analysis of Selmer's algorithm $F_{a,1}$ is contained in [11, Ch 7].

Our first main result concerns the number of zeroes in the limit $\vec{x}^{\infty}$:

Theorem 1. *Let $\vec{x}^{\infty} = \lim_{n\to\infty} F_{a,b}(\vec{x})$ for an ordered d-tuple $\vec{x} \in X_d$. The ascending chain $\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_d$ is defined by*

$$\mathcal{P}_r = \{\vec{x} \in X_d \colon \vec{x}_r^{\infty} > 0\}.$$

Then $\mathcal{P}_1 = \emptyset$ and $\mathcal{P}_r$ is a null set if $r \le a+1$. The first element of the chain that has positive measure is $\mathcal{P}_{a+2}$. If $r \le \min\{d, 2a\}$, then $\mathcal{P}_r$ is not full: its complement has positive measure.

We believe that $X_d \setminus \mathcal{P}_r$ is a null set if $r > 2a$, and this is supported by numerical experiments, but we have no proof. The proof of Theorem 1 relies on the existence of certain simplicial subsets that we call *trapping regions*

$$\mathcal{T}_r := \{\vec{x} \in X_d : \frac{1}{r-a}\sum_{j\le r} x_j < x_r\},$$

for $r \ge a+1$, and where by convention $\mathcal{T}_r = X_d$ if $r > d$. Trapping regions are invariant sets. Once an orbit enters $\mathcal{T}_r$ it remains there, as we will show in the proof of Lemma 2. It turns out that $\mathcal{T}_j = \mathcal{P}_j$ up to a set of Lebesgue measure zero if $a+1 \le j \le 2a$, as we will see in Lemma 4.

In Section 1.2 we consider the subtractive algorithm on the projective cone, by scaling the vectors so that the last coordinate is constant one. For initial vectors $\vec{x}$ in $X_d \setminus \mathcal{T}_d$ (which according to Theorem 1 it has positive measure), the dynamics of the scaled algorithm can be chaotic, even though $F^k(\vec{x}) \to \vec{0}$. Theorem 2 below states that for $a \ge \min\{2, b\}$, the scaled algorithm is Lebesgue ergodic on this set and admits an absolutely continuous invariant probability measure.

Although subtractive algorithms have been studied in percolation theory models (cf. [7, 8]), they are rooted in number theory, so our paper would not be complete without considering the accuracy of the convergents of the generalised Selmer algorithm. We discuss this in section 1.3. A theorem of Lagarias [5] can readily be applied to establish the order of approximation of generic convergents.

1.2. **The scaled algorithm.** Since $F_{a,b}(\lambda\vec{x}) = \lambda F_{a,b}(\vec{x})$ for every $\lambda > 0$ (*i.e.*, $F_{a,b}$ is a *homogeneous* algorithm), the generalised Selmer algorithm remains well-defined on the projective cone of non-negative d-tuples. If we consider $F_{a,b}$ up to projective equivalence, then we say that we consider the *scaled algorithm* $f_{a,b}$. The sets $\mathcal{P}_r$ remain well defined and invariant under the scaled algorithm. Their elements converge to a fixed point under iteration if $r \le d$. However, if $\vec{x} \in \mathcal{V}_d := \Delta_d \setminus \mathcal{P}_d$ then the ω-limit of the scaled algorithm no longer needs to be a singleton. Here,

$$\Delta_d := \{x = (x_1, \ldots, x_d) : 0 \le x_1 \le \cdots \le x_{d-1} \le x_d = 1\}.$$

contains one element from each equivalence class of the projective cone. The compact simplex Δ_d is a convex hull spanned by d vertices. As such, it has finite $d-1$-dimensional

Lebesgue measure, so we can express properties of the scaled algorithm $f_{a,b}$ in terms of probabilities. Now that we have fixed the representatives of the projective equivalence classes, we can define the scaled algorithm explicitly. If we denote $\hat{x} = F_{a,b}(\vec{x})$, then the scaled Selmer algorithm is given by

$$f_{a,b}(\vec{x}) = \frac{1}{\hat{x}_d} F_{a,b}(x_1, \ldots, x_d).$$

If $a = b = 1$ the first coordinate of $f_{a,b}$ is equivalent to the Farey map $x \mapsto \frac{x}{1-x}$ if $x \in [0, \frac{1}{2}]$ and $x \mapsto \frac{1-x}{x}$ if $x \in [\frac{1}{2}, 1]$.

As we will demonstrate in Theorem 4 in Section 3, under our assumption $b \leq a$ the dynamical system $(X_d, F_{a,b})$ possesses a *Markov partition*, *i.e.,* a partition $\{P_i\}$ such that for each k, $F_{a,b}(P_k)$ is the union of elements of $\{P_i\}$, up to a set of measure zero. The scaled version of $\{P_i\}$ to Δ_d is a Markov partition for $f_{a,b}$. This simplifies the study of $f_{a,b}$ considerably, but Δ_d has a boundary plane $\{x_1 = 0\}$ of neutral fixed points of quadratic tangency. Also the Farey map has such a fixed point, as do many other systems related to continued fractions and interval exchange transformations. The existence of neutral fixed points can cause absolutely continuous invariant measures to be infinite (*e.g.* the Farey map), but not always, see [11, pages 50 and 60] for invariant measures of Brun's and Selmer's algorithms.

Theorem 2. *Assume that $a \geq \max\{b, 2\}$. The restriction $f_{a,b} : \mathcal{V}_d \to \mathcal{V}_d$ preserves a probability measure μ_d, which is equivalent to the restriction of Lebesgue measure to $\mathcal{V}_d$ with density bounded away from 0. The measure μ_d is ergodic and exact (and hence mixing).*

Remark 1. *The condition $a \geq b$ is important to obtain the Markov partition; without it, the behaviour of the map can be quite different. For the map $f_{1,2}$, Miernowski & Nogueira* [9] *and* [10, Corollary 8.2] *showed although totally dissipative, Lebesgue measure is ergodic. (In* [2]*, the same result as well as Lebesgue exactness was shown.)*

1.3. **Accuracy of the algorithm.** We focus on how Theorem 2 helps in estimating how well the algorithm performs in providing rational approximations. The estimate depends on Oseledec's Ergodic Theorem, applied to the invariant set $\mathcal{V}_d$, so it applies only to rational approximations for elements of this set.

For each $k \geq 1$, there is a partition of Δ_d into maximal polytopes on which $f_{a,b}^k$ is diffeomorphic. This is the time-1-partition of $f_{a,b}^k$. The polytopes are called *k-cylinder sets*, or simply *cylinders*. If Z is such a cylinder then $f_{a,b}^k|_Z$ is the composition of a linear map, represented by a matrix $A_k = A_k(Z)$, and a scaling. The inverse matrix A_k^{-1} is integral and non-negative. Each column of A_k^{-1}, after scaling, gives a rational approximation of $\vec{x} \in Z$:

$$\vec{w}_{k,i} = \left(\frac{p_{1,k-i}}{q_{k-i}}, \ldots, \frac{p_{d-1,k-i}}{q_{k-i}}, 1 \right), \text{ for } 0 \leq i \leq d-1.$$

Following Lagarias [6], we interpret the quantity $\eta(\vec{w}, \vec{x}) = \frac{-\log \|\vec{w} - \vec{x}\|}{\log q}$ (where $\| \ \|$ is the sum-norm) as a measure for the error relative to the denominator q. The *best approximation exponent* is defined as

$$\eta(\vec{x}) = \limsup_{k \to \infty} \sup_{0 \leq i < d} \eta(\vec{w}_{k,i}, \vec{x}).$$

Dirichlet's Theorem (see *e.g.* [11, page 147]) states that every vector $\vec{x} \in \Delta_d$ has infinitely many rational approximations $\vec{w}$ such that $\|\vec{w}-\vec{x}\| \le q^{-(1+1/(d-1))}$, so $\eta(\vec{x}) \ge 1+1/(d-1)$ for every $\vec{x}$, provided the algorithm finds infinitely many of such approximations. However, for many $\vec{x}$, the best approximation exponent can be larger. On the other hand, there is no a priori way of telling which of the approximations are good, so it can be more useful to have a **uniform** approximation exponent. Each step of the algorithm produces a d-tuple of rational approximations $\vec{w}_{k,i}$ of $\vec{x}$, namely the columns of A_k^{-1} divided by their bottom element, the denominator $q_{k,i}$. For $0 \le i < d$, the worst error of the approximation at the k-th step is $\max_{0\le i<d} \|\vec{w}_{k,i} - \vec{x}\|$ and it is achieved at an "expense" of $\max_{0\le i<d} q_{k-i}$. Therefore it is reasonable to accept the *uniform approximation exponent*

$$\eta^*(\vec{x}) = \inf_k \frac{\min_{0\le i<d} -\log \|\vec{w}_{k,i} - \vec{x}\|}{\max_{0\le i<d} \log q_{k-i}}$$

as a guaranteed measure of the error. Therefore,

$$\|\vec{w}_{k,i} - \vec{x}\| \le \begin{cases} q_{k-i}^{-\eta(\vec{x})} & \text{infinitely often} \\ q_{k-i}^{-\eta^*(\vec{x})} & \text{for all } k, i. \end{cases}$$

We can interpret $A_k^{-1} = A_k^{-1}(\vec{x})$ as part of a cocycle

$$(\vec{x}, A) \mapsto (f_{a,b}(\vec{x}) \ , \ B^{-1} \cdot \Pi^{-1}(\vec{x}) \cdot A)$$

where B is a fixed matrix representing the subtractions and $\Pi(\vec{x})$ is the permutation matrix expressing the reordering after subtraction. The k-th iterate of this cocycle, starting with the identity matrix, is $(\vec{x}, I) \mapsto (f_{a,b}^k(\vec{x}) \ , \ A_k^{-1}(\vec{x}))$. Oseledec's Theorem asserts that $(A_k^{-1})_{k\ge 0}$ has Lebesgue typical Lyapunov exponents $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_d$. Lagarias uses this to compute the approximation exponents for continued fraction algorithms; in our setting this results in the following:

Theorem 3. *Let $a \ge \max\{b, 2\}$. For Lebesgue-a.e. initial vector $\vec{x} \in \mathcal{V}_d$, we have*

$$\eta(\vec{x}) = \eta^*(\vec{x}) = 1 - \frac{\lambda_2}{\lambda_1} > 1,$$

where $\lambda_1 > 0 > \lambda_2$ are the largest two typical Lyapunov exponents of the cocycle.

Arnaldo Nogueira has pointed out to us that this theorem can also be obtained through work of Broise & Guivarc'h [1].

1.4. **Organisation of this paper.** Section 2 describes the trapping regions. Once a point $\vec{x}$ gets trapped in the j-th trapping region, its limit $\vec{x}^\infty$ has j-th coordinate equal to zero. Section 3 shows that F has a Markov partition provided $b \le a$. In Section 4, we turn to the scaled algorithm f, and define a region Y for which an associated induced map avoids the non-hyperbolicity created by the neutral fixed points of f. Section 4 discusses the properties of a "first passage" induced map G which is used to produce the absolutely continuous $f_{a,b}$-invariant probability measures μ_d supported on $\mathcal{V}_d$. Here Theorem 2 and Theorem 3 are proved.

Notation: Henceforth, we will drop subscripts in $F_{a,b}$ and $f_{a,b}$. In other words, $F = F_{a,b}$ and $f = f_{a,b}$ throughout. For vectors $\vec{x}$, we use a subscript for coordinates and a superscript for iterations, so x_j^k is the j-th coordinate of $F^k(\vec{x})$.

Acknowledgement: We thank Arnaldo Nogueira and an anonymous referee for useful comments, which have helped us improve an earlier version of this paper.

2. The trapping regions

We study the limit of $\vec{x}^k$ in measure theoretic terms and we often restrict ourselves to the subset of d-tuples with rationally independent coordinates. This is allowed, since this subset has full Lebesgue measure.

Lemma 1. *Let $\vec{x}^\infty = \lim_{i\to\infty} F^i(\vec{x})$. Then $x^\infty_{a+1} = 0$ if the coordinates of $\vec{x}$ are rationally independent.*

Proof. The limit $\vec{x}^\infty$ exists by monotonicity: all coordinates are non-increasing under iteration. The limit $\vec{x}^\infty$ is a fixed point of F so $x^\infty_1 = 0$. By rational independence all coordinates of $\vec{x}^i$ are positive. In particular, the sequence of first coordinates x^i_1 is positive and descends to zero. Now x^{i+1}_1 is either equal to x^i_1 or to $x^i_{a+1} - x^i_1$. In order to descend to zero, it is required that $x^i_{a+1} - x^i_1 < x^i_1$ infinitely often. Take the limit on both sides of this inequality to find that $x^\infty_{a+1} = 0$. □

Recall that the *r-trapping region* for $r \geq a+1$ is defined as

$$\mathcal{T}_r = \{\vec{x} \in X_d : \frac{1}{r-a}\sum_{j\leq r} x_j < x_r\}.$$

Clearly $\mathcal{T}_{a+1} = \emptyset$ and if $r > d$ then we put $\mathcal{T}_r = X_d$. Since $x_r \leq x_{r+1}$ it follows that $\mathcal{T}_r \subset \mathcal{T}_{r+1}$ and that the sequence

$$\emptyset = \mathcal{T}_{a+1} \subset \cdots \subset \mathcal{T}_{d+1} = X_d$$

is strictly increasing.

For subtractive algorithms the sorting operation is similar to a rifle shuffle. Sorting $(x_1, \ldots, x_a, x_{a+1} - x_1, \ldots, x_d - x_1)$ is like shuffling a deck of a cards with a deck of b cards.

Lemma 2. *For $r \geq a+1$, $\mathcal{T}_r$ is F-invariant and if $\vec{x} \in \mathcal{T}_r$ then $x^\infty_r > 0$.*

Proof. Take $\vec{x} \in \mathcal{T}_d$ and write $\hat{x} = F(\vec{x})$.
Case 1: $x_r - x_1 \geq x_a$. Observe that x_r is the $r-a$-th card of the b-deck. In this case, the rifle shuffle affects only the first $r-a-1$ cards in the b-deck. Therefore $\hat{x}_j = x_j - x_1$ for $j \geq r$. The first $r-1$ coordinates of $\hat{x}$ are a permutation of $x_1, \ldots, x_a, x_{a+1} - x_1, \ldots, x_{r-1} - x_1$. This observation is used in the final equality in the computation below:

$$\begin{aligned}
\hat{x}_r = x_r - x_1 &> \frac{1}{r-a}\sum_{j\leq r} x_j \quad - \ x_1 \\
&= \frac{1}{r-a}\sum_{j=a+1}^{r} (x_j - x_1) + \frac{1}{r-a}\sum_{j\leq a} x_j \\
&= \frac{1}{r-a}\sum_{j\leq r} \hat{x}_j,
\end{aligned}$$

and we see that $\hat{x} \in \mathcal{T}_r$. More precisely, the difference between x_r and $\frac{1}{r-a}\sum_{j\leq r} x_j$ *is the same for* $\hat{x}$.
Case 2: $x_r - x_1 < x_a$. In this case the rifle shuffle places $x_r - x_1$ before the r-th coordinate in $\hat{x}$, so $\hat{x}_r > x_r - x_1$. Now the first and the last equality sign in the computation above change to inequalities.
Consequently, the difference between x_r and $\frac{1}{r-a}\sum_{j\leq r} x_j$ *grows for* $\hat{x}$. We conclude that

$\mathcal{T}_r$ is invariant and that the difference between x_r and $\frac{1}{r-a}\sum_{j\le r} x_j$ is non-decreasing under iteration. Therefore

$$x_r^\infty \ge x_r^\infty - \frac{1}{r-a}\sum_{j\le r} x_j^\infty \ge x_r - \frac{1}{r-a}\sum_{j\le r} x_j > 0.$$

□

Lemma 3. *If $r \le 2a$, then $\mathcal{T}_{r+1} \setminus \mathcal{T}_r$ is F-invariant. In particular, since $\mathcal{T}_{d+1} = X_d$, the complement of $\mathcal{T}_d$ is invariant if $d \le 2a$.*

Proof. Suppose that $\vec{x} \in \mathcal{T}_{r+1} \setminus \mathcal{T}_r$. If $r > d$ then this set is empty and there is nothing to prove. So we may assume that $r \le d$. If $x_r - x_1 \ge x_a$ then by the same argument as in case 1 in the proof of Lemma 2, but now with the $>$ sign replaced by a $<$ sign, we find that $F(\vec{x}) \notin \mathcal{T}_r$. So we may restrict our attention to $x_r - x_1 < x_a$. In this case x_a is the maximal value among the first r coordinates before the sorting, so $\hat{x}_r \le x_a$ and

$$\begin{aligned}
\hat{x}_r \;\le\; x_a \;&=\; \frac{1}{r-a}\sum_{j=a+1}^{r} x_a \le \frac{1}{r-a}\sum_{j=a+1}^{r} x_j \\
&=\; \frac{1}{r-a}\left(\sum_{j=a+1}^{r}(x_j - x_1)\right) + x_1 \\
&\le\; \frac{1}{r-a}\left(\sum_{j=a+1}^{r}(x_j - x_1)\right) + \frac{1}{a}\sum_{j=1}^{a} x_1 \\
&\le\; \frac{1}{r-a}\left(\sum_{j=a+1}^{r}(x_j - x_1)\right) + \frac{1}{r-a}\sum_{j=1}^{a} x_j = \frac{1}{r-a}\sum_{j\le r}\hat{x}_j,
\end{aligned}$$

so $\hat{x} \notin \mathcal{T}_r$. This proves the invariance of $\mathcal{T}_{r+1} \setminus \mathcal{T}_r$. □

The restriction $r \le 2a$ cannot be avoided, and to illustrate this we give an example of $\vec{x} \in \mathcal{T}_{d+1} \setminus \mathcal{T}_d$ such that $F(\vec{x}) \in \mathcal{T}_d$. Take $d = 2a+1$ and consider the d-tuple that has $2a$ coordinates equal to 1 and its largest coordinate is equal to $2-\varepsilon$. Since $\mathcal{T}_{d+1} = X_d$ we have $\vec{x} \in \mathcal{T}_{d+1}$ by definition. To see that $\vec{x} \notin \mathcal{T}_d$ compute

$$\frac{1}{d-a}\sum_{j\le d} x_j = \frac{d+1-\varepsilon}{d-a} = \frac{2a+2-\varepsilon}{a+1} > x_d.$$

Now $\hat{x} = F(\vec{x})$ has a coordinates equal to 0 and a coordinates that are equal to 1 while one coordinate is equal to $1-\varepsilon$. To see that $\hat{x} \in \mathcal{T}_d$, we compute

$$\frac{1}{d-a}\sum_{j\le d} \hat{x}_j = \frac{a+1-\varepsilon}{a+1} < 1 = \hat{x}_d.$$

Lemma 4. *Suppose that $\vec{x} \in \mathcal{T}_{2a+1}$ has rationally independent coordinates. Then $x \in \mathcal{T}_j$ if and only if x_j^∞ is the first non-zero coordinate of $\vec{x}^\infty$. In other words, $\mathcal{T}_j$ is equal a.e. to $\mathcal{P}_j$.*

Proof. Assume that $\vec{x} \in \mathcal{T}_{2a+1}$ and that its coordinates are rationally independent. Then $\vec{x} \in \mathcal{T}_j \setminus \mathcal{T}_{j-1}$ for a unique $j \le 2a+1$. By the previous lemmas, $\vec{x}^\infty \in \mathcal{T}_j \setminus \mathcal{T}_{j-1}$ and $x_i^\infty = 0$ for $i \le a+1$. Therefore the first non-zero coordinate of $\vec{x}^\infty$ must be the j-th coordinate. We conclude that $x_j^\infty > 0$ if and only if $\vec{x} \in \mathcal{T}_j$. □

The proof of Theorem 1 is now straightforward.

Proof of Theorem 1. If $x \in \mathcal{T}_{2a+1}$ then $x_j^\infty > 0$ if and only if $\vec{x} \in \mathcal{T}_j \setminus \mathcal{T}_{j-1}$. Now $\mathcal{T}_j$ is strictly contained in Δ_d if $j \leq d$ and it has non-empty interior if $j > a+1$. Therefore the probability that $\vec{x}$ is in $\mathcal{T}_j$ is strictly between 0 and 1 if $j \leq \min\{d, 2a\}$, and $j > a+1$. □

3. A Markov partition for X_d if $b \leq a$

We consider a simplicial partition of X_d by certain hyperplanes that we call folding planes. In this section we show that this division induces a Markov partition if $b \leq a$. The boundary ∂X_d is the union of d hyperplanes

$$\{x_1 = 0\} \cup \{x_2 = x_1\} \cup \cdots \cup \{x_d = x_{d-1}\}.$$

In other words, $\vec{x} \in \partial X_d$ if and only if its first coordinate is zero or if it has two equal coordinates. For $1 \leq j \leq a < k \leq d$ we say that

$$L_{j,k} = \{x_1 + x_j = x_k\} \subset X_d$$

is a *folding plane*; this is a plane where the sorting operation folds the image of X_d. We say that the partition of X_d by the folding planes is the time-1-partition.

Lemma 5. *The map F is a local diffeomorphism at all interior points of X_d that are not in a folding plane.*

Proof. If $\vec{x}$ is not in a folding plane, then all coordinates of $F(\vec{x})$ are unequal, *i.e.,* all elements of $\{x_1, \ldots, x_a, x_{a+1} - x_1, \ldots, x_d - x_1\}$ are unequal. Therefore, if we change the coordinates of $\vec{x}$ by ε then for the perturbed point $\tilde{x}$ the coordinates of $F(\tilde{x})$ will remain unequal and will be sorted in the same way. It follows that F is not only homeomorphic: it is even linear on $B_\varepsilon(\vec{x})$. For $\vec{x}$ in the interior of X_d, $F(\vec{x}) \neq \vec{0}$, so the scaling operation is also a local diffeomorphism. □

If $b \geq a$ then we say that

$$\Lambda = \{\vec{x} \in X_d : x_1 + x_{a+1} = x_{2a}\}$$

is the *pre-folding plane*. To make the definition independent of b, we say that Λ is empty if $b < a$. If $a = 1$ then $x_1 = 0$ and Λ is an invariant plane in the boundary of the simplex.

Lemma 6. *If $\vec{x} \in \Lambda$ then $F^n(\vec{x}) \in \partial X_d$ for some $n > 1$. Furthermore, $f^k(\vec{x}) \in \Lambda$ for $0 \leq k \leq n-2$ and $F^{n-1}(\vec{x})$ lies in a folding plane.*

Proof. The result is trivial if $a = 1$, so suppose that $a > 1$. Put $\hat{x} = F(\vec{x})$. The relation $\hat{x}_{2a} = \hat{x}_1 + \hat{x}_{a+1}$ continues to hold, until the $a+1$-th coordinate gets sorted to the first a coordinates. Without loss of generality we may suppose that such a sorting occurs immediately, so $x_{a+1} - x_1 < x_a$. Note that $x_{2a} - x_1 = x_{a+1} \geq x_a$ so the $2a$-th coordinate remains in place; it is not affected by the left hand deck in the rifle shuffle. Therefore $\hat{x}_j = x_{a+1} - x_1$ for some $j \leq a$ and $\hat{x}_{2a} = x_{2a} - x_1$. If $1 < j$ then $\hat{x} \in L_{j,2a}$ which is a folding plane which implies that $F(\hat{x}) \in \partial\Delta_d$. If $j = 1$ then $\hat{x}_1 = x_{a+1} - x_1$ and $\hat{x}_m = x_1$ for some $m \leq a$. But then $\hat{x} \in L_{m,2a}$, which is another folding plane. □

Lemma 7. *If $b \leq a$ then for every $\vec{x} \in \partial X_d$, there exists an $n \geq 1$ such that $F^n(\vec{x}) \in \partial X_d$.*

Proof. It suffices to prove that $F^n(\vec{x})$ is in a folding plane for some $n \geq 1$. If $x_1 = 0$ then $F(\vec{x}) = \vec{x}$ so immediately $F(\vec{x}) \in \partial X_d$. If $x_j = x_{j+1}$ for $j \neq a$ then under iteration of F either x_1 is subtracted from both coordinates or from neither. Either way, the coordinates remain equal to each under when F is applied, which implies that $F(\vec{x}) \in \partial X_d$

immediately. Therefore, in the interesting case when $\vec{x}$ is in $\partial\Delta_d$ and $F(\vec{x})$ is not, we have that $\vec{x} \in \{x_a = x_{a+1}\}$. In this case $x_{a+1} - x_1$ is sorted to the first a coordinates of $\hat{x}$ and x_a is sorted to the last b coordinates. Therefore $\hat{x}_j = x_{a+1} - x_1$ for some $j \leq a$ and $\hat{x}_k = x_a$ for some $k > a$. If $j > 1$ then $\hat{x} \in L_{j,k}$ and we are done. If $j = 1$ then $\hat{x}_1 = x_{a+1} - x_1$ and $\hat{x}_m = x_1$ for some $m > 1$. If $m \leq a$ then $\hat{x} \in L_{m,k}$ and we are done. If $m > a$ then it can be at most $d - a + 1 \leq a + 1$. (This is the only point where the assumption $b \leq a$ is used!) Therefore $m = a + 1$ and $\hat{x}$ is in the pre-folding plane, and so it eventually gets mapped into the boundary of the simplex. □

The folding planes induce a simplicial subdivision of X_d. We have seen that the union of the boundary and folding planes together with the pre-folding plane Λ is forward invariant if $b \leq a$ and that F is a simplicial map with respect to this partition. Therefore, it forms a Markov partition if $b \leq a$. The simplices of a subdivision are usually called the *cylinders*. A cylinder is *full* if its image is equal to the entire set.

Theorem 4. *If $b \leq a$ then F admits a Markov partition. If a cylinder of the Markov partition has a boundary that is disjoint from $\{x_a = x_{a+1}\} \cup \Lambda$, then it is full.*

Proof. We only need to prove that the specified cylinder sets are full, *i.e.,* it maps to the entire space. A cylinder set Z is convex and bounded by a collection of hyperplanes, that are either folding or pre-folding or in the boundary. Suppose that the boundary of Z does not intersect $\{x_a = x_{a+1}\}$ or the pre-folding plane. Since these are the only two planes that are not mapped inside $\partial\Lambda$, the boundary of $F(Z)$ must be contained in ∂X_d. The only convex set that has its boundary in ∂X_d is the entire set. □

We can summarise this theorem as follows. If $b < a$ then the partition by folding planes is Markov because $\{x_a = x_{a+1}\}$ is mapped to a folding plane. If $b = a$ then $\{\vec{x} \in \mathcal{V}_d : x_d + x_1 \leq x_a = x_{a+1}\}$ is mapped to Λ, which eventually gets mapped to a folding plane. If we extend the partition by the pre-folding plane, then we still have a Markov partition. If $b > a$, then the dynamics of the boundary plane $\{x_a = x_{a+1}\}$ becomes intractable. While it is no longer possible to keep track of the iterated images $f^k(Z)$ of cylinders Z, it is possible to say something about the images $f(Z)$. We do this in the next lemma.

For $1 < j < k$ and $k > a$ the hyperplane $\{x_1 + x_j = x_k\}$ divides X_d into two half spaces, and we denote one of these halves by $U_{jk} = \{x_1 + x_j \geq x_k\}$.

Lemma 8. *Let $b \geq 1$ be arbitrary. The partition of X_d by folding planes has the property that each cylinder is either full or it is mapped onto a half space U_{jk} for $1 \leq j < k$. If $j > a$ then $k - j \geq a - 1$.*

In particular, if $b \leq a$ as in the previous theorem, then $j \leq a + 1$ and a cylinder is either full, or it is mapped onto a half space that is bounded by a (pre-)folding plane.

Proof. Let Z be a cylinder. If Z does not intersect the boundary, then all hyperplanes in ∂Z are folding. In this case $\partial F(Z) \subset \partial X_d$ and since $F(Z)$ is convex, this implies that the cylinder is full. Now suppose that Z does intersect the boundary. If it does not intersect $\{x_a = x_{a+1}\}$ then Z is full by the same argument. Suppose therefore that Z intersects $\{x_a = x_{a+1}\}$. All points in Z are subject to the same rifle shuffle, sorting a boundary triple $\{x_1, x_a, x_{a+1} - x_1\}$ to $\{\hat{x}_1, \hat{x}_j, \hat{x}_k\}$ for $1 \leq j \leq k$ and $k > a$. We have that $\hat{x}_1 + \hat{x}_j = \hat{x}_k$. All the other hyperplanes in ∂Z are folding, so $F(Z)$ is a half space of the hyperplane $\{x_1 + x_j = x_k\}$. The cylinder set contains interior points for which $x_a < x_{a+1}$ and these points are mapped to $\hat{x}$ for which $\hat{x}_1 + \hat{x}_j > \hat{x}_k$. That is why $F(Z) = U_{jk}$. To see that

there are special conditions on j, note that if $j > a$ then $\hat{x}_1 = x_{a+1} - x_1$. The coordinates x_m for $1 < m < a$ remain in between $\hat{x}_j$ and $\hat{x}_k$, so if $j > a$ then $k - j \geq a - 1$. □

Note that each $U_{jk} \cap \Delta_d$ contains the *top* vertex $\vec{e} := (1, \dots, 1)$ of Δ_d. If $b > a$ then its image is contained in the first non-trivial trapping region $\mathcal{T}_{a+2}$.

4. The first passage map.

From this point onwards, we are exclusively interested in the scaled algorithm (Δ_d, f). All the previous results on trapping regions and Markov partition readily carry over to the scaled algorithm, and we will therefore keep the same notation.

4.1. **The inducing region Y.** We will consider the first passage map of the scaled algorithm $f : \Delta_d \to \Delta_d$ to a region Y. We first describe this region and then we show that the first passage map has bounded distortion. Let

$$L_1 := \{\vec{x} \in \Delta_d : \frac{1}{d-a} \sum_{j \leq d} x_j = x_d\}$$

be the 'upper' boundary of the trapping region $\mathcal{T}_d$, see Lemma 2. If $b = 1$, i.e., if we have Selmer's original algorithm, then the trapping region is empty, and L_1 reduces to $\{\vec{x} \in \Delta_d : x_1 = \cdots = x_{d-1} = 0\}$. The plane L_1 divides the simplex into the two disjoint components $\mathcal{T}_d \cap \Delta_d$ and $\mathcal{V}_d := \Delta_d \setminus \mathcal{T}_d$, which are invariant by Lemma 3, since $b \leq a$. We will study the dynamics on $\mathcal{T}_d^c$, which is the component containing the vertex $e = (1, \dots, 1)$. From now on, we will denote $\mathcal{T}_d^c$ by $\mathcal{V}_d$, which is equivalent up to a null set by Lemma 4. We say that $\mathcal{V}_d$ is the *top* of the simplex. We also say that the *coordinate a is overtaken* if it changes under the unscaled map F. In particular

$$L_2 := \{\vec{x} \in \Delta_d : x_{a+1} = 2x_1\}$$

bounds the region where the first coordinate is overtaken. It divides $\mathcal{V}_d$ into two components. We define the *inducing region* Y to be the component that contains the top. Note that L_2 is equal to the folding plane $L_{1,a+1}$, so the inducing region Y is a union of elements of the Markov partition. For the smallest parameters $a = b = 1$, when f is equivalent to the Farey map, the inducing region corresponds to $[\frac{1}{2}, 1]$.

Lemma 9. *If $b > 1$, then $\Delta_d \setminus (L_1 \cup L_2)$ consists of four regions and if $b = 1$, then $\Delta_d \setminus L_2$ consists of two regions. In either case, let Y be the closure of the region containing the* top *$e = (1, \dots, 1)$. Then Y does not intersect the boundary plane $\{\vec{x} \in \Delta_d : x_1 = 0\}$. Indeed, if $\vec{x} \in Y$ then $x_1 \geq 1/(2a + 1)$.*

Proof. If $b = 1$ then Y is equal to the region $\{x_d = x_{a+1} \leq 2x_1\}$. We have scaled $x_d = 1$ so we conclude that $x_1 \geq 1/2 > 1/(2a + 1)$. If $b > 1$ then Y is equal to the region $\{x_{a+1} \leq 2x_1, \frac{1}{d-a} \sum_{j \leq d} x_j \geq x_d\}$. For any $\vec{x} \in Y$ we have that $(d - a)x_d \leq \sum_{j \leq d} x_j \leq x_1 + ax_{a+1} + (d - a - 1)x_d$, and it follows that $x_d \leq x_1 + ax_{a+1} \leq (2a + 1)x_1$. We have scaled $x_d = 1$ and we conclude that $x_1 \geq 1/(2a + 1)$. □

The limit $\vec{x}^\infty$ of the unscaled map is equal to $\vec{0}$ for Lebesgue-a.e. element of $X_d \setminus \mathcal{T}_d$. Therefore, the first coordinate is overtaken infinitely often for a.e. $\vec{x} \in Y$. Since the orbit of a.e. element of $\mathcal{V}_d$ visits Y infinitely often, the passage time $\tau(\vec{x}) = 1 + \min\{n \geq 0 : f^n(\vec{x}) \in Y\}$ is well defined. Note that $\tau(\vec{x})$ counts the number of times that we subtract the same coordinate. We define the *first passage map*

$$G(\vec{x}) = f^{\tau(x)}(\vec{x}) \quad \text{for} \quad \vec{x} \in \mathcal{V}_d.$$

If $a = b = 1$ then f is equivalent to the Farey map and Y is equivalent to the half interval $[\frac{1}{2}, 1]$. The first passage map is equivalent to the Gauss map, in this case. For general a, b, accelerating f to the first passage map G renders the Markov partition of $\mathcal{V}_d$ infinite, but leaves the image partition unchanged. Denote the matrix representation of the (unscaled) first passage map F^τ on a τ-cylinder by A_τ. By definition of Y, the first coordinate doesn't changes over these τ iterates of F. Hence, A_τ^{-1} can have negative coordinates only in the first column, and its inverse is a non-negative matrix of the form:

$$A_\tau^{-1} = \begin{pmatrix} 1 & 0 & \dots & & \dots & 0 \\ m_2 & 1 & & & & \vdots \\ \vdots & & \ddots & & & \\ \vdots & & & & & \\ m_{d-1} & & & & \ddots & \vdots \\ m_d & 0 & \dots & & \dots & 1 \end{pmatrix}. \tag{1}$$

where $0 \le m_2 \le \cdots \le m_d \le \tau$ and $m_{a+1} > 0$. The local action of the first passage map is given by ΠA_τ for a permutation matrix Π which permutes the first coordinate to a higher coordinate.

4.2. **Distortion of the first passage map.** The first passage map divides $\mathcal{V}_d$ into (principal) cylinders, which we denote by

$$\Delta_{A_k,\Pi} = \{\vec{x} \in \mathcal{V}_d \colon G(\vec{x}) = \Pi A_k(\vec{x})\}.$$

There are infinitely many cylinders, but they have finitely many images by Theorem 4. A cylinder of rank n is a maximal subset on which G^m acts linearly and injectively:

$$\Delta_{(A_{k_1},\Pi_1),\dots,(A_{k_n},\Pi_n)} = \left\{\vec{x} \in \mathcal{V}_d \colon G^{j-1}(\vec{x}) \in \Delta_{A_{k_j},\Pi_j}, 1 \le j \le n\right\}.$$

An n-cylinder is a maximal subset on which G^n acts linearly. It is the image of a subset of Y under G^{-n}, which is a composition of linear maps with non-negative coefficients

$$B_n = A_{k_1}^{-1}\Pi_1^{-1} \cdots A_{k_n}^{-1}\Pi_n^{-1}. \tag{2}$$

More precisely, the unscaled action of G^{-n} is given by B_n.

Proposition 1. *The map G has bounded distortion wherever it is defined,* i.e., *there is an absolute constant K such that*

$$\left|\frac{|J_{G^n}(\vec{x})|}{|J_{G^n}(\vec{y})|} - 1\right| \le K|G^n(\vec{x}) - G^n(\vec{y})|, \tag{3}$$

for all $n \ge 0$ whenever $\vec{x}, \vec{y}$ are in the same cylinder of G^n.

Remark 2. *Obviously, the estimate* (3) *implies that there is $K > 0$ such that*

$$\left|\frac{J_{G^n}(\vec{x})}{J_{G^n}(\vec{y})}\right| \le K \tag{4}$$

for all $n \ge 0$ and $\vec{x}, \vec{y}$ in the same cylinder of G^n. It is this estimate that we need in the proof of Theorem 2 later on, but the stronger version (3) *is standard and useful for different applications.*

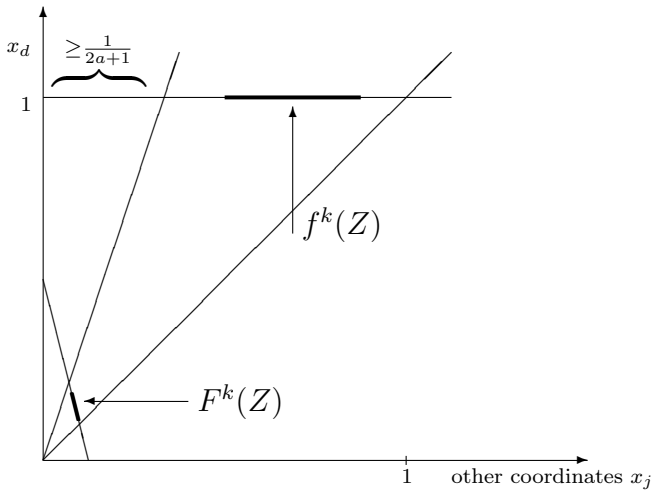

FIGURE 1. The images $F^k(Z)$ and $f^k(Z)$ obtained from $F^k(Z)$ by dividing by the largest coordinate. Note that $f^k(Z) \subset Y$ where $x_d \geq x_1 \geq 1/(2a+1)$ by Lemma 9.

Proof. Let $Z = \Delta_{(A_{k_1},\Pi_1),\ldots,(A_{k_n},\Pi_n)}$ be any n-cylinder. It is the image of a branch of G^{-n}. Take k such that $G^n = f^{k+1}$, so $f^k : Z \subset f(Y) \to Y$. From (2) we know that G^{-n} is given by B_n, and analogously, we can find a non-negative matrix B'_n representing the iterate f^{-k} on the subset $f^k(Z) \subset Y$. Straightforward computation shows that $J_{f^k}(\vec{x}) = (F^k(\vec{x})_d)^{-(d+1)}$, see also [11, Prop. 2], and therefore

$$\sup_{\vec{x},\vec{y}\in Z} \frac{J_{f^k}(\vec{x})}{J_{f^k}(\vec{y})} = \left(\frac{F^k(\vec{y})_d}{F^k(\vec{x})_d}\right)^{d+1} = \left(1 + \frac{F^k(\vec{y})_d - F^k(\vec{x})_d}{F^k(\vec{x})_d}\right)^{d+1}.$$

Since $F^k(Z)$ is not in the trapping region $\mathcal{T}_d$, Lemma 9 gives that the smallest coordinate of any $F^k(\vec{x})$ is at least $1/(2a+1)$ times the largest coordinate. Scaling (*i.e.,* dividing by its largest coordinate) maps $F^k_{a,b}(Z)$ into Y, so this largest coordinate varies on $F^k(Z)$ by no more than the multiplicative constant $2a+1$, see Figure 1.

Therefore there is $\tilde{K} = \tilde{K}(d,a)$ by which the above estimate can be reworked to

$$\sup_{\vec{x},\vec{y}\in U} \left|\frac{J_{f^k}(\vec{x})}{J_{f^k}(\vec{y})} - 1\right| \leq \tilde{K}|f^k(\vec{x}) - f^k(\vec{y})|.$$

Now for the last iterate, we use again that the smallest coordinate of any vector in Y is at least $1/(2a+1)$, *i.e.,* the largest coordinate necessary in the scaling within $f : Y \to f(Y)$ is at least $1/(2a+1)$, so bounded away from zero. Hence the remaining single iterate f has bounded distortion, and the proposition follows. □

4.3. **Proof of Theorem 2.** Note that this distortion result applies to the Jacobian determinant only and not necessarily to expansion factors, because G is non-conformal and not necessarily uniformly expanding for any iterate. Yet it suffices to prove Theorem 2.

Proof of Theorem 2. Since G has a Markov partition with finite image partition, ergodicity and exactness follow in the standard way, see [15, Theorem 1]. Let $\mathcal{P}$ be the time-1 partition for G, and $\mathcal{P}^n = \bigvee_{j=0}^{n-1} G^{-j}\mathcal{P}$ the time-n partition. Let $\mathcal{Q}$ be the finite partition generated by $G(\mathcal{P})$. Therefore for each $Z \in \mathcal{P}^{n-1}$, $G^n(Z) \in \mathcal{Q}$ is one out of a finite set. Let

$B \subset f(Y)$ be a Lebesgue measurable set and $Z \in \mathcal{P}^{n-1}$. Then for K from Proposition 1 we have

$$\begin{aligned} m(B) &\geq \int_{G^{-n}B\cap Z} J_{G^n}(\vec{x})\, dm(\vec{x}) \geq \frac{1}{K}\int_{G^{-n}B\cap Z} \sup\{J_{G^n}(\vec{y}) : \vec{y}\in Z\}\, dm(\vec{x}) \\ &\geq \frac{1}{K}\frac{m(G^{-n}B\cap Z)}{m(Z)}\int_Z J_{G^n}(\vec{x})\, dm(\vec{x}) = \frac{1}{K}\frac{m(G^{-n}B\cap Z)}{m(Z)} m(G^n(Z)). \end{aligned} \quad (5)$$

Summing over all $Z \in \mathcal{P}^{n-1}$, we find

$$m(G^{-n}B) \leq \frac{K}{\inf\{m(Q) : Q\in\mathcal{Q}\}} m(B) =: K' m(B).$$

Let ν be a weak accumulation point of the sequence $(\frac{1}{n}\sum_{i=0}^{n-1} m(G^{-i}(\cdot)))_n$. Since ν is a limit of Cesaro means, it is a G-invariant probability measure on $f(Y)$.

By construction, the measure ν has the property $\nu(B) \leq K' m(B)$, so it has a bounded density $h = \frac{d\nu}{dm}$. We need to show that h is bounded away from zero as well. It suffices to consider a Lebesgue measurable set B contained in a single $Q \in \mathcal{Q}$. A similar computation as (5), but with reversed inequalities, gives

$$\begin{aligned} m(B) &= \int_{G^{-n}B\cap Z} J_{G^n}(\vec{x})\, dm(\vec{x}) \leq K\int_{G^{-n}B\cap Z} \inf\{J_{G^n}(\vec{y}) : \vec{y}\in Z\}\, dm(\vec{x}) \\ &\leq K\frac{m(G^{-n}B\cap Z)}{m(Z)}\int_Z J_{G^n}(\vec{x})\, dm(\vec{x}) = K\frac{m(G^{-n}B\cap Z)}{m(Z)} m(G^n(Z)). \end{aligned}$$

Summing over all such $Z \in \mathcal{P}^n$ such that $G^n(Z) \supset Q$, we obtain

$$m(G^{-n}B) \geq \frac{1}{K}\cdot m(B)\cdot m(G^{-n}(Q)), \quad (6)$$

and therefore we need to find a lower bound for $m(G^{-n}(Q))$. Assume that $G(\mathcal{P})$ is a transitive aperiodic partition (otherwise we restrict to a transitive subset and an iterate of G). Fix N such that $G^N(Q') \supset Q$ for every $Q, Q' \in \mathcal{Q}$. By the finiteness of $\mathcal{Q}$, we can find $\eta > 0$ such that $m(G^{-N}(Q)\cap Q') > \eta$ for all $Q, Q' \in \mathcal{Q}$. Since for each $n \in \mathbb{N}$, the whole space $f(Y)$ is disjointly covered by $\#\mathcal{Q}$ sets $G^{-(n-N)}(Q')$, we get $\max\{m(G^{-(n-N)}(Q')) : Q' \in \mathcal{Q}\} \geq 1/\#\mathcal{Q}$. Using distortion bound (6) for $B = G^{-N}(Q)\cap Q'$ where $Q' \in \mathcal{Q}$ is chosen such that $m(G^{-(n-N)}(Q'))$ is maximal, we obtain

$$m(G^{-n}(Q)) \geq m(G^{-(n-N)}(B)) \geq \frac{1}{K}\cdot m(B)\cdot m(G^{-(n-N)}(Q')) \geq \frac{\eta}{K\#\mathcal{Q}} > 0.$$

Hence $h = \frac{d\nu}{dm}$ is indeed bounded away from 0.

An f-invariant measure μ_d is obtained by pulling back ν according to the standard formula

$$\mu_d(B) = \frac{1}{\int \tau d\nu}\sum_{t\in\mathbb{N}}\sum_{k=0}^{t-1} \nu(f^{-k}(B)\cap U_t), \quad (7)$$

where $U_t = \{\vec{y} \in f(Y) : \tau(\vec{y}) = t\}$. The normalising constant

$$\int \tau d\nu = \sum_{t\in\mathbb{N}} t\nu(\{x \in Y : \tau(\vec{x}) = t\}) = \sum_{t\in\mathbb{N}} \nu(\{\vec{x} \in Y : \tau(\vec{x}) \geq t\})$$

is finite for $a \geq 2$ due to the following:

Claim: The ν-measure of the tail $\nu(\{\vec{y} \in f(Y) : \tau(\vec{y}) \geq t\}) = O(t^{-a})$.

Note that the case $a = 1$ is treated by Kraaikamp & Meester [4]: Lebesgue a.e. orbit enters all trapping regions $\mathcal{T}_r$, $r = 3, \ldots, d$, and Lebesgue is totally dissipative.

To prove the claim, assume that Lebesgue measure has a conservative part $Y' \subset Y$. In particular, the return time τ is well-defined on a subset $f(Y')$ of $f(Y)$ of positive Lebesgue measure. The set $\{\vec{x} \in \Delta_d : x_1 = 0\}$ has dimension $d-1$. Since $f|_{Y'}$ is a piecewise diffeomorphism, $\overline{f(Y')} \cap \{\hat{x} \in \Delta_d : \hat{x}_1 = 0\}$ has dimension $\leq d - a$, because if $f(x)_1 = 0$, then $x_1 = x_{a+1}$ whence also $x_j = x_1$ for $1 \leq j \leq a+1$.

Until $\hat{x} := f(\vec{x})$ returns to Y under iteration of f, the coordinate $\hat{x}_1$ is not overtaken, and hence will increase by a factor $1/(1 - d \cdot \hat{x}_1)$ every q iterates for some $q \leq d$. Therefore, if the first entry time into Y is $n-1$, then $\frac{q}{d \cdot n} \leq \hat{x}_1 = x_{a+1} - x_1 \leq \frac{q}{d \cdot (n-1)}$, so that $0 \leq x_j - x_1 < \frac{q}{dn}$ for $2 \leq j \leq a$. This shows that $m(\{x \in Y' : \tau(x) = n\}) = O((\frac{q}{d})^a \cdot \frac{1}{n^{a+1}})$. Summing over $t \geq n$ then proves the claim.

Finiteness of $\int \tau \, d\nu$ follows, and hence μ_d is the required f-invariant probability measure absolutely continuous w.r.t. (and in fact equivalent to) Lebesgue measure on $\mathcal{V}_d$. Moreover, since there is a finite N such that $\cup_{k=0}^{N} f^k(f(Y)) \supset \mathcal{V}_d$, the lower bound of the density $\frac{d\nu}{dm}$ carries over to a lower bound of the density $\frac{d\mu}{dm}$

Finally, the measure μ inherits ergodicity from ν and, since the greatest common divisor of the induce times is one, also exactness. □

4.4. **Proof of Theorem 3.** This proof follows directly from Theorem 4.1 in [6, page 314] which relies on five conditions, called (H1)-(H5) in that paper.

Proof. We will check state and check conditions (H1)-(H5) in our setting.

(H1) The map f has an ergodic absolutely continuous probability measure.
This is clear from Theorem 2.

(H2) The map f is piecewise continuous with non-vanishing Jacobian Lebesgue-a.e.
This is obviously true in our setting.

(H3) There is $c_0 > 1$ such that for Lebesgue-a.e. $\vec{x} \in \mathcal{V}_d$, there is $k_0(\vec{x})$ such that

$$\max_{0 \leq m < d-1} \|\vec{w}_{k,m} - \vec{x}\| \leq c_0^{-k} \text{ for all } k \geq k_0(\vec{x}).$$

Uniform expansion of G is achieved within a certain region of $f(Y)$. Indeed there is a subset E compactly contained in $f(Y)$ such that every return to E contract the projective metric by a fixed factor $\rho \in (0,1)$. Since $\Lambda := \int \tau d\nu < \infty$, vectors in Δ_d visit E on average once every $\nu(E)/\Lambda$ iterates, so the contraction after k iterates is $\rho^{k\nu(E)/\Lambda}$. The set E is compactly contained in $f(Y)$, so that projective and Euclidean metric are equivalent on E. Hence (H3) holds for any $1 < c_0 < \rho^{-\nu(E)/\Lambda}$.

(H4) The integral $\int_{\mathcal{V}_d} \log \max\{1, \|B^{-1}\Pi\|\} d\mu_d < \infty$.
Here the permutation Π in $B^{-1}\Pi$ depends on the point $x \in \Delta_d$, but since $\|B^{-1}\Pi\|$ is uniformly bounded, condition (H4) is trivially fulfilled.

(H5) The Lebesgue integral $\int_{\mathcal{V}_d} \tau^* \, dm < \infty$ where τ^* is defined as

$$\tau^*(\vec{x}) := \min\{k : A_k^{-1}(\vec{x}) \text{ is strictly positive}\}.$$

The matrix A_k^{-1} becomes strictly positive once every coordinate has become the smallest in the iteration of F. We can choose $M \in \mathbb{N}$ and an M-cylinder $E \subset f(Y)$ of positive measure such that A_M^{-1} is strictly positive on E, $f^M(E) = \mathcal{V}_d$ and $f^k(E) \cap E = \emptyset$ for $0 < k < M$. Boundedness of distortion give some constant

$C = C(E)$ such that $\frac{d\mu \circ f^M}{d\mu} \leq C$ on E. Let $\tau_E(\vec{x}) := \min\{k \geq 0 : f^k(\vec{x}) \in E\}$ denote the first sojourn time in E. Then, by Kac' Lemma,

$$\begin{aligned}\int_{\mathcal{V}_d} \tau^* \, d\mu &\leq \int_{\mathcal{V}_d} (\tau_E + M) \, d\mu \leq \int_E (\tau_E + M) \cdot \frac{d\mu \circ f^M}{d\mu} \, d\mu \\ &\leq C \int_E (\tau_E + M) \, d\mu \leq C(1 + M\mu(E)).\end{aligned}$$

Finally, because $\frac{d\mu}{dm}$ is bounded away from zero, $\int_{\mathcal{V}_d} \tau^* dm$ is finite as well. □

Hence we know the asymptotic errors of approximation of Lebesgue typical initial vectors $\vec{x}$. But typical $\vec{x}$ is not all $\vec{x}$. What can be said about non-typical $\vec{x}$?

References

[1] A. Broise-Alamichel, Y. Guivarc'h, *Characteristic exponents of the Jacobi-Perron algorithm and the associated transformation* [French], Ann. Institute Fourier **51**, no. 3 (2001), 565–686.

[2] H. Bruin, *Lebesgue ergodicity of a dissipative subtractive algorithm,* Preprint 2012, to appear in Springer Proceedings in Mathematics.

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

[4] C. Kraaikamp, R. Meester, *Ergodic properties of a dynamical system arising from percolation theory,* Ergodic Theory Dynam. Systems **15** (1995), no. 4, 653–661.

[5] J. C. Lagarias, *The computational complexity of simultaneous Diophantine approximation problems,* SIAM. J. Comput. **14** (1985), 196–209.

[6] J. C. Lagarias, *The quality of the Diophantine approximations found by the Jacobi-Perron algorithm and related algorithms,* Monatsh. Math. **115** (1993), 299–328.

[7] R. Meester, *An algorithm for calculating critical probabilities and percolation functions in percolation models defined by rotations,* Ergodic Theory Dynam. Systems **9** (1989), no. 3, 495-509.

[8] R. Meester, T. Nowicki, *Infinite clusters and critical values in two-dimensional circle percolation,* Israel J. Math. **68** (1989), 63–81.

[9] T. Miernowski, A. Nogueira, *Absorbing sets of homogeneous subtractive algorithms,* Monatsh. Math. **167** (2012), 547–569.

[10] A. Nogueira, *The three-dimensional Poincaré continued fraction algorithm,* Israel J. Math. **99** (1995) 373–401.

[11] F. Schweiger, *Multidimensional continued fractions,* Oxford Science Publ. 2000.

[12] F. Schweiger, *Variations of the Poincaré map,* Integers **12** (2012), 167–178.

[13] F. Schweiger, *Invariant simplices for subtractive algorithms,* J. Number Theory, **133**(2013), 2182–2185.

[14] E. S. Selmer, *Om Flerdimensjonaler Kjedebrøk*, Nord. Mat. Tidskr. **9** (1961), 37-43.

[15] L.-S. Young, *Recurrence times and rates of mixing,* Israel J. Math. **110** (1999), 153–188.

Faculty of Mathematics, University of Vienna,
Oskar Morgensternplatz 1, 1090 Vienna, Austria
henk.bruin@univie.ac.at
http://www.mat.univie.ac.at/∼bruin

Institute of Applied Mathematics, Delft University of Technology,
Mekelweg 4, 2628 CD Delft, The Netherlands
R.J.Fokkink@tudelft.nl

Institute of Applied Mathematics, Delft University of Technology,
Mekelweg 4, 2628 CD Delft, The Netherlands
c.kraaikamp@tudelft.nl