The *d*-dimensional Gauss Transformation: Strong Convergence and Lyapunov Exponents

D.M. Hardcastle[1], K.Khanin
Basic Research Institute in the Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-2000-26
December 18th , 2001*

E-mail: D.M. Hardcastle@ma.hw.ac.uk, K.Khanin@newton.cam.ac.uk

multi-dimensional continued fractions, computer assisted proof

We discuss a method of producing computer assisted proofs of almost everywhere strong convergence of the *d*-dimensional Gauss algorithm. This algorithm is equivalent to Brun's algorithm and to the modified Jacobi-Perron algorithm considered by Podsypanin and Schweiger. In this paper we focus on the reduction of the problem to a finite number of calculations. These calculations have been carried out for the three-dimensional algorithm and the results, which prove almost everywhere strong convergence, will be published separately.

* Internal Accession Date Only Approved for External Publication
[1] Department of Mathematics, Heriot-Watt University, Edinburgh, EH14 4AS, UK

The d-dimensional Gauss transformation: strong convergence and Lyapunov exponents

D. M. Hardcastle[1] and K. Khanin[1−4]

[1] Department of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, UK. E-mail: D.M.Hardcastle@ma.hw.ac.uk

[2] Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, Cambridge CB3 0EH, UK. E-mail: K.Khanin@newton.cam.ac.uk

[3] BRIMS, Hewlett-Packard Laboratories, Stoke Gifford, Bristol BS34 8QZ, UK.

[4] Landau Institute for Theoretical Physics, Kosygina Str.,2, Moscow 117332, Russia.

Abstract

We discuss a method of producing computer assisted proofs of almost everywhere strong convergence of the d-dimensional Gauss algorithm. This algorithm is equivalent to Brun's algorithm and to the modified Jacobi-Perron algorithm considered by Podsypanin and Schweiger. In this paper we focus on the reduction of the problem to a finite number of calculations. These calculations have been carried out for the three-dimensional algorithm and the results, which prove almost everywhere strong convergence, will be published separately.

1 Introduction

Multidimensional continued fraction algorithms produce a sequence of simultaneous rational approximations to a given irrational vector. The best known of these algorithms is perhaps the Jacobi-Perron algorithm (JPA) [8, 14], but other algorithms such as Brun's [2], Selmer's [19] and Podsypanin's modified Jacobi-Perron algorithm [15, 16] have also been widely studied. These algorithms are believed to be strongly convergent almost everywhere, i.e. for Lebesgue almost all $\boldsymbol{\omega} \in [0,1]^d \setminus \mathbb{Q}^d$ the sequence $(p_1(n)/q(n), \ldots, p_d(n)/q(n))$ produced by the algorithm is believed to satisfy

$$\lim_{n\to\infty} \|q(n)\boldsymbol{\omega} - (p_1(n), \ldots, p_d(n))\| = 0.$$

However, rigorous proofs of almost everywhere strong convergence currently only exist in two dimensions. Strong convergence of the two-dimensional JPA follows from a paper of Paley and Ursell [13]; this fact was first noticed by Khanin [9] (see also [18]). In 1993, Ito, Keane and Ohtsuki produced a computer assisted proof of strong convergence of the two-dimensional modified JPA [7, 3]. Since then, Meester has provided a proof which does not involve the use of computers [12].

In [4], we discussed a method which, in principle, can be used to produce a computer assisted proof of almost everywhere strong convergence in arbitrary dimension. We illustrated our scheme by discussing the ordered (or modified) Jacobi-Perron algorithm, which is equivalent to Brun's algorithm. It is particularly suitable for numerical study since an explicit formula for the invariant density is known. From the theoretical viewpoint suggested in [4], this scheme is quite simple. However, it is currently impossible to use in practice since the number of calculations required is vast.

In this paper, we discuss in detail a numerical scheme which is perhaps more complicated than the previously mentioned scheme. However, this scheme can be used to obtain new results: the number of calculations required to prove almost everywhere strong convergence in the three-dimensional case is large, but not so large as to be impractical. This scheme was used by Ito, Keane and Ohtsuki in [7] to prove almost everywhere strong convergence of the two-dimensional modified (ordered) Jacobi-Perron algorithm, but they did not discuss it in higher dimensions. We discuss the scheme in arbitrary dimension, and show how the error terms can be estimated explicitly.

Throughout this paper we call the endomorphism corresponding to the modified (ordered) Jacobi-Perron algorithm, the d-dimensional Gauss transformation. This name was suggested in [5] where it was shown that the algorithm has many remarkable properties of the one-dimensional case. We understand that the use of different names for the same object is unfortunate, but we hope that it will not confuse our readers.

Any computer assisted proof consists of two parts: a description of how to reduce the problem to a finite number of calculations, and the actual performance of those calculations. We consider the first part in this paper. Numerical results which prove almost everywhere strong convergence of the three-dimensional Gauss algorithm will be published separately (see [6]).

Acknowledgements. DMH is grateful to the Engineering and Physical Sciences Research Council of the UK for financial support.

2 The d-dimensional Gauss transformation

In this section we give the definitions which will be needed for the rest of the paper. We begin by defining the d-dimensional Gauss algorithm, and then we formally define strong convergence and discuss its relationship with the Lyapunov exponents of the algorithm.

Let $\Delta^d = \{(\omega_1, \ldots, \omega_d) \in [0,1]^d : \omega_1 \geq \omega_2 \geq \cdots \geq \omega_d\}$. Define $T : \Delta^d \to \Delta^d$ by

$$T(\omega_1, \ldots, \omega_d) = \begin{cases} \left(\left\{\frac{1}{\omega_1}\right\}, \frac{\omega_2}{\omega_1}, \ldots, \frac{\omega_d}{\omega_1}\right) & \text{if } \left\{\frac{1}{\omega_1}\right\} > \frac{\omega_2}{\omega_1}; \\ \left(\frac{\omega_2}{\omega_1}, \ldots, \frac{\omega_j}{\omega_1}, \left\{\frac{1}{\omega_1}\right\}, \frac{\omega_{j+1}}{\omega_1}, \ldots, \frac{\omega_d}{\omega_1}\right) & \text{if } \frac{\omega_j}{\omega_1} > \left\{\frac{1}{\omega_1}\right\} > \frac{\omega_{j+1}}{\omega_1}; \\ \left(\frac{\omega_2}{\omega_1}, \ldots, \frac{\omega_d}{\omega_1}, \left\{\frac{1}{\omega_1}\right\}\right) & \text{if } \frac{\omega_d}{\omega_1} > \left\{\frac{1}{\omega_1}\right\}. \end{cases} \tag{1}$$

In this formula, $\{x\}$ denotes the fractional part of a real number x, i.e. $\{x\} = x - [x]$ where $[x]$ is the integer part of x.

Definition 1. The transformation $T : \Delta^d \to \Delta^d$ is called the *d-dimensional Gauss transformation.*

Notice that the image of an ordered vector $\boldsymbol{\omega}$ under T is formed by placing $\{1/\omega_1\}$ in the correct position in the sequence $\omega_2/\omega_1, \dots, \omega_d/\omega_1$. The transformation T naturally arises from a geometrical scheme for approximating $\boldsymbol{\omega}$ (see [5]). Here we just give a formal description of how it can be used to produce approximations to $\boldsymbol{\omega}$.

To each $\boldsymbol{\omega} \in \Delta^d$ we associate $m(\boldsymbol{\omega}) = [1/\omega_1] \in \mathbb{N}$ and $j(\boldsymbol{\omega}) \in \{1, 2, \dots, d\}$, which gives the position of $\{1/\omega_1\}$ in the vector $T(\boldsymbol{\omega})$. Each pair (m, j) labels a particular branch of T^{-1}, which we denote $T^{-1}_{(m,j)}$. This branch is given by

$$T^{-1}_{(m,j)}(\omega_1, \dots, \omega_d) = \left(\frac{1}{m + \omega_j}, \frac{\omega_1}{m + \omega_j}, \dots, \frac{\omega_{j-1}}{m + \omega_j}, \frac{\omega_{j+1}}{m + \omega_j}, \dots, \frac{\omega_d}{m + \omega_j} \right).$$

We now define a matrix $\widetilde{A}_{(m,j)} = (\widetilde{a}_{i,l})_{1 \le i,l \le d+1} \in GL(d+1, \mathbb{Z})$. The first row of $\widetilde{A}_{(m,j)}$ has only two nonzero entries:

$$\widetilde{a}_{1,1} = m, \quad \widetilde{a}_{1,j+1} = 1.$$

The other rows of $\widetilde{A}_{(m,j)}$ have only one nonzero entry: $\widetilde{a}_{i,i-1} = 1$ for $i = 2, \dots, j+1$, and $\widetilde{a}_{i,i} = 1$ for $i = j+2, \dots, d+1$. So

$$\widetilde{A}_{(m,j)} = \begin{pmatrix} m & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

The matrices $\widetilde{A}_{(m,j)}$ give the action of $T^{-1}_{(m,j)}$ on rational vectors:

$$T^{-1}_{(m,j)}\left(\frac{p_1}{q}, \dots, \frac{p_d}{q}\right) = \left(\frac{\widetilde{p}_1}{\widetilde{q}}, \dots, \frac{\widetilde{p}_d}{\widetilde{q}}\right) \text{ if and only if } \begin{pmatrix} \widetilde{q} \\ \widetilde{p}_1 \\ \vdots \\ \widetilde{p}_d \end{pmatrix} = \widetilde{A}_{(m,j)} \begin{pmatrix} q \\ p_1 \\ \vdots \\ p_d \end{pmatrix}. \tag{2}$$

Let $A_{(m,j)} = (\widetilde{A}_{(m,j)})^t$ where A^t denotes the transpose of a matrix A. Define a matrix-valued function on Δ^d by

$$A(\boldsymbol{\omega}) = A_{(m(\boldsymbol{\omega}), j(\boldsymbol{\omega}))}.$$

Define also

$$C_n(\boldsymbol{\omega}) = A(T^{n-1}\boldsymbol{\omega}) \cdots A(T\boldsymbol{\omega})A(\boldsymbol{\omega}). \tag{3}$$

Let

$$C_n(\boldsymbol{\omega}) = \begin{pmatrix} q(n,0) & p_1(n,0) & \dots & p_d(n,0) \\ q(n,1) & p_1(n,1) & \dots & p_d(n,1) \\ \vdots & \vdots & & \vdots \\ q(n,d) & p_1(n,d) & \dots & p_d(n,d) \end{pmatrix} \tag{4}$$

and

$$\frac{\boldsymbol{p}(n,i)}{q(n,i)} = \left(\frac{p_1(n,i)}{q(n,i)}, \dots, \frac{p_d(n,i)}{q(n,i)}\right), \quad 0 \le i \le d. \tag{5}$$

Denote

$$\boldsymbol{\tau}_n = \left(\frac{p_1(n)}{q(n)}, \dots, \frac{p_d(n)}{q(n)}\right) = \frac{\boldsymbol{p}(n,0)}{q(n,0)}.$$

We shall consider $\boldsymbol{\tau}_n$ as the n^{th} approximation to $\boldsymbol{\omega}$ given by the Gauss algorithm. It is easy to see that $\boldsymbol{\tau}_n$ is one of the vertices of the simplex $\Delta_n(\boldsymbol{\omega})$ which is the element of the n^{th} level of the Markov partition containing $\boldsymbol{\omega}$ (see Section 4). In some sense the whole simplex $\Delta_n(\boldsymbol{\omega})$, and not only the vertex $\boldsymbol{\tau}_n$, can be considered as the n^{th} approximation to $\boldsymbol{\omega}$.

Definition 2. A sequence of rational vectors $\boldsymbol{\tau}_n = \left(\frac{p_1(n)}{q(n)}, \dots, \frac{p_d(n)}{q(n)}\right)$ is *exponentially strongly convergent to* $\boldsymbol{\omega}$ if there exist constants $k > 0$, $\alpha > 0$ such that

$$\|q(n)\boldsymbol{\omega} - (p_1(n), \dots, p_d(n))\| \le k q(n)^{-\alpha}.$$

Our aim is to prove that the d-dimensional Gauss algorithm is exponentially strongly convergent almost everywhere, i.e. for almost all $\boldsymbol{\omega} \in \Delta^d$ the sequence of rational vectors $\boldsymbol{\tau}_n = \boldsymbol{p}(n,0)/q(n,0)$ defined by (5) is exponentially strongly convergent to $\boldsymbol{\omega}$.

We first relate the strong convergence of the algorithm to its Lyapunov exponents. In order to do this we briefly discuss the ergodic properties of the map T (see [16]).

Define

$$\rho(\boldsymbol{\omega}) = \sum_{\pi \in S_d} \frac{1}{1+\omega_{\pi(1)}} \frac{1}{1+\omega_{\pi(1)}+\omega_{\pi(2)}} \cdots \frac{1}{1+\omega_{\pi(1)}+\omega_{\pi(2)}+\dots+\omega_{\pi(d)}} \tag{6}$$

where S_d is the group of permutations of $\{1, 2, \dots, d\}$. Let $K = \int_{\Delta^d} \rho(\boldsymbol{\omega})\, d\boldsymbol{\omega}$. Then the probability measure μ defined by

$$\mu(X) = \frac{1}{K} \int_X \rho(\boldsymbol{\omega})\, d\boldsymbol{\omega}, \quad X \text{ a Borel subset of } \Delta^d,$$

is invariant under T and ergodic. This measure is the unique absolutely continuous invariant probability measure.

The endomorphism T together with the matrix-valued function A and the invariant measure μ form a cocycle which we denote (T, A, μ). It is easy to show that this cocycle is integrable. Let $\lambda_1(A) \ge \lambda_2(A) \ge \dots \ge \lambda_{d+1}(A)$ be the corresponding Lyapunov exponents.

The following theorem, which is based on the work of Lagarias [11], was proved in [4].

Theorem 1. (i) *The largest Lyapunov exponent $\lambda_1(A)$ is strictly positive and simple.*

(ii) *For almost all* $\boldsymbol{\omega} \in \Delta^d$

$$\lim_{n\to\infty} \frac{1}{n} \log q(n) = \lambda_1(A).$$

(iii) *The sequence* $\boldsymbol{\tau}_n$ *is exponentially strongly convergent to* $\boldsymbol{\omega}$ *for almost all* $\boldsymbol{\omega}$ *if and only if* $\lambda_2(A) < 0$.

Remark. In fact it follows from [1] that the Lyapunov spectrum is simple, i.e. $\lambda_1(A) > \lambda_2(A) > \cdots > \lambda_{d+1}(A)$, but we will not use this fact in this paper.

3 Analysis of Lyapunov exponents

The calculation of Lyapunov exponents is, in general, a very hard problem. However, the Subadditive Ergodic Theorem allows one to obtain an upper bound for the largest Lyapunov exponent. It is for this reason that we replace the cocycle (T, A, μ) by another cocycle which has largest Lyapunov exponent $\lambda_2(A)$.

The construction of the new cocycle (T, D, μ) is based on the following observation which was made by Lagarias [11]. Let

$$E_2(\boldsymbol{\omega}) = \{\boldsymbol{v} \in \mathbb{R}^{d+1} : \langle \boldsymbol{v}, (1, \omega_1, \ldots, \omega_d)\rangle = 0\}.$$

Then $A(\boldsymbol{\omega})E_2(\boldsymbol{\omega}) = E_2(T\boldsymbol{\omega})$ and, for all $\boldsymbol{v} \in E_2(\boldsymbol{\omega})$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|C_n(\boldsymbol{\omega})\boldsymbol{v}\| \le \lambda_2(A).$$

Hence the cocycle corresponding to $A(\boldsymbol{\omega}) : E_2(\boldsymbol{\omega}) \to E_2(T\boldsymbol{\omega})$ has Lyapunov exponents $\lambda_2(A) \ge \lambda_3(A) \ge \cdots \ge \lambda_{d+1}(A)$. Denote

$$\boldsymbol{e}_1(\boldsymbol{\omega}) = \begin{pmatrix} -\omega_1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{e}_2(\boldsymbol{\omega}) = \begin{pmatrix} -\omega_2 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \boldsymbol{e}_d(\boldsymbol{\omega}) = \begin{pmatrix} -\omega_d \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{7}$$

Then $\{\boldsymbol{e}_1(\boldsymbol{\omega}), \ldots, \boldsymbol{e}_d(\boldsymbol{\omega})\}$ is a basis for $E_2(\boldsymbol{\omega})$. Let $D(\boldsymbol{\omega})$ be the $d \times d$ matrix which gives the action of $A(\boldsymbol{\omega})$ on $E_2(\boldsymbol{\omega})$ in terms of this basis. If $\boldsymbol{v} = \sum_{j=1}^{d} v_j \boldsymbol{e}_j(\boldsymbol{\omega})$, $v_1, \ldots, v_d \in \mathbb{R}$ then

$$A(\boldsymbol{\omega})\boldsymbol{v} = \sum_{j=1}^{d} v_j A(\boldsymbol{\omega})\boldsymbol{e}_j(\boldsymbol{\omega}) = \sum_{i=1}^{d} \left(\sum_{j=1}^{d} d_{ij}(\boldsymbol{\omega}) v_j \right) \boldsymbol{e}_i(T\boldsymbol{\omega})$$

where $d_{ij}(\boldsymbol{\omega})$ is equal to the $(i+1)^{\text{th}}$ coordinate of $A(\boldsymbol{\omega})\boldsymbol{e}_j(\boldsymbol{\omega})$. Denote the matrix $(d_{ij}(\boldsymbol{\omega}))_{1\le i,j\le d}$ by $D(\boldsymbol{\omega})$. One can give the explicit form of the matrices $D(\boldsymbol{\omega})$ which follows easily from the definition of $d_{ij}(\boldsymbol{\omega})$. Let

$$\widetilde{D} = (\widetilde{D}_{i,j})_{1\le i,j\le d} = \begin{pmatrix} -\omega_1 & -\omega_2 & -\omega_3 & \dots & -\omega_d \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{8}$$

Then $D(\boldsymbol{\omega})$ is obtained from $\widetilde{D}$ by the permutation of the rows

$$(1, 2, \dots, d) \mapsto (2, \dots, j(\boldsymbol{\omega}), 1, j(\boldsymbol{\omega}) + 1, \dots, d).$$

Namely we just put the first row in the $j(\boldsymbol{\omega})^{\text{th}}$ position.

Let $\lambda_1(D)$ be the largest Lyapunov exponent of the cocycle (T, D, μ). The construction described above implies the following lemma.

Lemma 2.

$$\lambda_1(D) = \lambda_2(A).$$

Combining (iii) of Theorem 1 and Lemma 2 we get the following corollary.

Corollary 3. $\lambda_1(D) < 0$ *is equivalent to exponentially strong convergence almost everywhere.*

Denote

$$D_n(\boldsymbol{\omega}) = D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega})D(\boldsymbol{\omega}).$$

The next lemma is an immediate consequence of the Subadditive Ergodic Theorem [10].

Lemma 4. $\lambda_1(D) < 0$ *if and only if there exists* $n \in \mathbb{N}$ *such that*

$$\frac{1}{n}\int_{\Delta^d} \log \|D_n(\boldsymbol{\omega})\| \, \mu(d\boldsymbol{\omega}) < 0. \tag{9}$$

Although the matrices $D(\boldsymbol{\omega})$ and the density of the invariant measure are given explicitly it is not easy to estimate the integral in (9). This is mainly due to the fact that the matrix product $D_n(\boldsymbol{\omega})$ is a smooth function only when $\boldsymbol{\omega}$ belongs to a particular element of the n^{th} level of the Markov partition (see Section 4), so that the function being integrated has infinitely many singularities.

In the next section we will show how we can rigorously prove inequality (9). Notice that (9) implies that for almost all $\boldsymbol{\omega}$ the matrix elements of $D_n(\boldsymbol{\omega})$ decay exponentially fast as $n \to \infty$. It follows from (8) that each matrix element of the product $D_n(\boldsymbol{\omega}) = D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega})D(\boldsymbol{\omega})$ is equal to the sum of many positive and negative terms. In fact the matrix elements of $D_n(\boldsymbol{\omega})$ decay exponentially only due to the cancellation of these positive and negative terms. To see this consider the "positive" cocycle $D_+(\boldsymbol{\omega})$ which is obtained by the replacement of all $-\omega_i$ in (8) by $+\omega_i$. It is obvious that $\lambda_1(D) \le \lambda_1(D_+)$. However, numerics show that $\lambda_1(D_+) < 0$ only if $d \le 2$ (see Table 1). This is another manifestation of the well-known fact that the case $d \ge 3$ is much harder than $d = 2$.

d	Largest Lyapunov exponent of (T, D_+, μ)	Largest Lyapunov exponent of (T, D, μ)
2	-0.088	-0.25
3	0.081	-0.11
4	0.14	-0.059
5	0.17	-0.036

Table 1: The largest Lyapunov exponents of the cocycles (T, D_+, μ) and (T, D, μ) for dimensions 2, 3, 4 and 5.

We finish this section with a discussion of the connection between the cocycle (T, D, μ) and the natural Jacobi cocycle of the map T. Denote by $J(\boldsymbol{\omega})$ the Jacobi matrix of the map T:

$$J(\boldsymbol{\omega}) = \frac{dT(\boldsymbol{\omega})}{d\boldsymbol{\omega}}.$$

Then the Jacobian cocycle (T, J, μ) is formed by the product of the Jacobi matrices along the cocycle (T, D, μ), i.e.

$$J_n(\boldsymbol{\omega}) = \frac{d(T^n)(\boldsymbol{\omega})}{d\boldsymbol{\omega}} = J(T^{n-1}\boldsymbol{\omega}) \cdots J(T\boldsymbol{\omega}) J(\boldsymbol{\omega}).$$

Proposition 5. (i)

$$D(\boldsymbol{\omega}) = \frac{1}{\omega_1}(J^{-1}(\boldsymbol{\omega}))^t \tag{10}$$

(ii)

$$D_n(\boldsymbol{\omega}) = \left(\prod_{i=0}^{n-1} \frac{1}{(T^i\boldsymbol{\omega})_1}\right)\left(J_n^{-1}(\boldsymbol{\omega})\right)^t \tag{11}$$

Proof. The first statement can be proved by an easy calculation. To prove (11) one iterates (10) n times. $\square$

Denote by $\lambda_1(J) \geq \lambda_2(J) \geq \cdots \geq \lambda_d(J)$ the Lyapunov exponents of the Jacobi cocycle. Then the following statement holds.

Proposition 6.

$$\lambda_1(D) = \lambda_1(A) - \lambda_d(J) \tag{12}$$

Proof. It is easy to see that

$$q(n, 0) \leq \prod_{i=0}^{n-1} \frac{1}{(T^i\boldsymbol{\omega})_1} \leq (d+1)q(n, 0) \tag{13}$$

(see [4]). Using (13), (11) and statement (ii) of Theorem 1 one immediately gets (12). $\square$

In fact Proposition 6 can be used to give an independent proof of Corollary 3. Denote $\boldsymbol{\omega}^{(i)} = T^i(\boldsymbol{\omega})$, $i \geq 0$. Since

$$\frac{\boldsymbol{p}(n,0)}{q(n,0)} = T^{-1}_{(m_1,j_1)} \circ \cdots \circ T^{-1}_{(m_n,j_n)} \left(\frac{0}{1}, \ldots, \frac{0}{1}\right)$$

and

$$\boldsymbol{\omega} = T^{-1}_{(m_1,j_1)} \circ \cdots \circ T^{-1}_{(m_n,j_n)} (\boldsymbol{\omega}^{(n)}),$$

the distance between $\boldsymbol{\omega}$ and $\boldsymbol{p}(n,0)/q(n,0)$ can be estimated through the minimum average expansion of the map T, which is given by $\lambda_d(J)$. Namely, $\|\boldsymbol{\omega} - \boldsymbol{p}(n,0)/q(n,0)\|$ is of the order $e^{-\lambda_d(J)n}$ so that after expansion by T^n, resulting in the rescaling of length by the factor $e^{\lambda_d(J)n}$, we get two points $\boldsymbol{\omega}^{(n)}$ and $\mathbf{0}$ of distance of order constant apart. As a result we get, in the limit as $n \to \infty$,

$$\frac{1}{n} \log \|q(n,0)\boldsymbol{\omega} - \boldsymbol{p}(n,0)\| = \frac{1}{n} \log q(n,0) + \frac{1}{n} \log \left\| \boldsymbol{\omega} - \frac{\boldsymbol{p}(n,0)}{q(n,0)} \right\| \to \lambda_1(A) - \lambda_d(J) = \lambda_1(D).$$

4 Analysis of the matrices $D(\boldsymbol{\omega})$ and the invariant measure

It was shown in the previous section that the proof of almost everywhere strong convergence is based on the estimation of a particular integral. We will estimate this integral by splitting the phase space Δ^d into the elements of the Markov partition, and then calculating an upper bound for the integral over each piece separately. We first show that the matrix product $D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega})D(\boldsymbol{\omega})$ can be expressed in terms of $\boldsymbol{\omega}$ and its approximations.

For $n \geq 1$, consider $\boldsymbol{\omega}^{(n)} = (\omega_1^{(n)}, \ldots, \omega_d^{(n)}) = T^n \boldsymbol{\omega}$, $\boldsymbol{\omega} \in \Delta^d$.

Proposition 7. *For all $n \in \mathbb{N}$*

(i)

$$\begin{aligned}(\omega_1^{(n)}, \omega_2^{(n)}, \ldots, \omega_d^{(n)}) &D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega})D(\boldsymbol{\omega}) \\ &= q(n,0)(\omega_1, \omega_2, \ldots, \omega_d) - (p_1(n,0), p_2(n,0), \ldots, p_d(n,0))\end{aligned}$$

(ii)

$$D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega})D(\boldsymbol{\omega})$$
$$= - \begin{pmatrix} q(n,1)\omega_1 - p_1(n,1) & q(n,1)\omega_2 - p_2(n,1) & \ldots & q(n,1)\omega_d - p_d(n,1) \\ q(n,2)\omega_1 - p_1(n,2) & q(n,2)\omega_2 - p_2(n,2) & \ldots & q(n,2)\omega_d - p_d(n,2) \\ \vdots & \vdots & & \vdots \\ q(n,d)\omega_1 - p_1(n,d) & q(n,d)\omega_2 - p_2(n,d) & \ldots & q(n,d)\omega_d - p_d(n,d) \end{pmatrix}.$$

Proof. Let $M(\boldsymbol{\omega})$ be the $(d+1)\times d$ matrix with columns $\boldsymbol{e}_1(\boldsymbol{\omega}), \boldsymbol{e}_2(\boldsymbol{\omega}), \dots, \boldsymbol{e}_d(\boldsymbol{\omega})$ (see (7)). Then obviously

$$A(\boldsymbol{\omega})M(\boldsymbol{\omega}) = M(T\boldsymbol{\omega})D(\boldsymbol{\omega}).$$

Multiplying this by $A(T^i\boldsymbol{\omega})$, $i = 1, \dots, n-1$, one gets

$$C_n(\boldsymbol{\omega})M(\boldsymbol{\omega}) = A(T^{n-1}\boldsymbol{\omega})\cdots A(T\boldsymbol{\omega})A(\boldsymbol{\omega})M(\boldsymbol{\omega}) = M(T^n\boldsymbol{\omega})D_n(\boldsymbol{\omega}), \tag{14}$$

which immediately implies the proposition. □

Remark. The matrices $D(\boldsymbol{\omega})$ have been considered in the literature before. In [7], Ito, Keane and Ohtsuki defined $D(\boldsymbol{\omega})$ (for the two-dimensional Modified Jacobi-Perron algorithm) by a formula similar to (8) and then observed that a formula such as that in Proposition 7 could be proven by induction. In [17] Schweiger defined the matrices $D(\boldsymbol{\omega})$ in arbitrary dimension. In short, the significance of the matrices $D(\boldsymbol{\omega})$ has been known for some time, in general they are treated separately to the matrices $A(\boldsymbol{\omega})$. The description we give, in particular equation (14), yields a trivial proof of Proposition 7.

The d-dimensional Gauss transformation has a natural Markov partition associated to it. Namely, for $m \in \mathbb{N}$ and $1 \le j \le d$ define

$$\Delta_{(m,j)} = \{\boldsymbol{\omega} \in \Delta^d : m(\boldsymbol{\omega}) = m, j(\boldsymbol{\omega}) = j\}.$$

Then $\{\Delta_{(m,j)} : m \in \mathbb{N}, 1 \le j \le d\}$ is a Markov partition for T and in fact

$$T(\Delta_{(m,j)}) = \Delta^d, \quad \forall (m,j) \in \mathbb{N} \times \{1, 2, \dots, d\}.$$

Let

$$\Delta_{(m_1,j_1),\dots,(m_n,j_n)} = \{\boldsymbol{\omega} \in \Delta^d : m(T^{i-1}\boldsymbol{\omega}) = m_i, j(T^{i-1}\boldsymbol{\omega}) = j_i \text{ for } 1 \le i \le n\}.$$

In the following proposition we use the notion of the Farey sum of two rational vectors: the Farey sum of $\boldsymbol{p}_1/q_1$ and $\boldsymbol{p}_2/q_2$ is defined by

$$\frac{\boldsymbol{p}_1}{q_1} \oplus \frac{\boldsymbol{p}_2}{q_2} = \frac{\boldsymbol{p}_1 + \boldsymbol{p}_2}{q_1 + q_2}.$$

Proposition 8. $\Delta_{(m_1,j_1),\dots,(m_n,j_n)}$ *is the simplex with vertices*

$$\frac{\boldsymbol{p}(n,0)}{q(n,0)}, \frac{\boldsymbol{p}(n,0)}{q(n,0)} \oplus \frac{\boldsymbol{p}(n,1)}{q(n,1)}, \frac{\boldsymbol{p}(n,0)}{q(n,0)} \oplus \frac{\boldsymbol{p}(n,1)}{q(n,1)} \oplus \frac{\boldsymbol{p}(n,2)}{q(n,2)}, \dots,$$
$$\frac{\boldsymbol{p}(n,0)}{q(n,0)} \oplus \frac{\boldsymbol{p}(n,1)}{q(n,1)} \oplus \frac{\boldsymbol{p}(n,2)}{q(n,2)} \oplus \cdots \oplus \frac{\boldsymbol{p}(n,d)}{q(n,d)}$$

where $\boldsymbol{p}(n,i)/q(n,i)$ are the vectors defined by (5).

Proof. Let $(m_1,j_1),\dots,(m_n,j_n)$ be arbitrary. The corresponding element of the Markov partition is given by

$$\Delta_{(m_1,j_1),\dots,(m_n,j_n)} = T^{-1}_{(m_1,j_1)} \circ \cdots \circ T^{-1}_{(m_n,j_n)} \Delta^d.$$

Denote the vertices of the original simplex Δ^d by

$$\boldsymbol{h}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{h}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \boldsymbol{h}_{d+1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Then

$$\boldsymbol{v}_i = \frac{\boldsymbol{p}_i}{q_i} = T^{-1}_{(m_1,j_1)} \circ \cdots \circ T^{-1}_{(m_n,j_n)}(\boldsymbol{h}_i), \quad 1 \le i \le d+1,$$

are the vertices of $\Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$. Let V_n be the matrix with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{d+1}$. It follows from (2) and (3) that

$$V_n = \widetilde{A}_{(m_1,j_1)} \cdots \widetilde{A}_{(m_n,j_n)} V_0 = (A_{(m_n,j_n)} \cdots A_{(m_1,j_1)})^t V_0 = C_n^t V_0$$

where

$$V_0 = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 0 & 1 & \ldots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

This implies the statement of the proposition. $\square$

We now prove that the maximum value of $\|D_n(\boldsymbol{\omega})\|$ over a simplex $\Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$ is attained at one of the vertices. We will do this by showing that the function $\boldsymbol{\omega} \mapsto \|D_n(\boldsymbol{\omega})\|$ is convex on each element of the n^{th} level of the Markov partition.

The norm that we will use is defined as follows. Let $\|\cdot\|$ denote the standard Euclidean norm on $\mathbb{R}^d$, i.e.

$$\|(v_1, \ldots, v_d)\| = \left(\sum_{i=1}^d v_i^2 \right)^{\frac{1}{2}}.$$

Then the corresponding norm of a linear operator $D : \mathbb{R}^d \to \mathbb{R}^d$ is given by

$$\|D\| = \sup_{\boldsymbol{v} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}} \frac{\|D\boldsymbol{v}\|}{\|\boldsymbol{v}\|}.$$

It is well known that

$$\|D\|^2 = \max\{\gamma : \gamma \text{ is an eigenvalue of } D^t D\}.$$

Lemma 9. *If $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$ then for all $\alpha \in [0,1]$*

$$D_n\big(\alpha\boldsymbol{\omega} + (1-\alpha)\boldsymbol{\omega}'\big) = \alpha D_n(\boldsymbol{\omega}) + (1-\alpha) D_n(\boldsymbol{\omega}').$$

Proof. This follows immediately from statement (ii) of Proposition 7. $\square$

This lemma implies the next statement.

Lemma 10. *The function* $\boldsymbol{\omega} \mapsto \|D_n(\boldsymbol{\omega})\|$ *is convex on each simplex* $\Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$*, i.e. for any* $\alpha \in [0,1]$ *and any* $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$

$$\big\|D_n\big(\alpha\boldsymbol{\omega} + (1-\alpha)\boldsymbol{\omega}'\big)\big\| \leq \alpha\|D_n(\boldsymbol{\omega})\| + (1-\alpha)\|D_n(\boldsymbol{\omega}')\|.$$

The following lemma is well known.

Lemma 11. *Let* $\Delta \subset \mathbb{R}^d$ *be a simplex. If* $f : \Delta \to \mathbb{R}$ *is continuous and convex then* f *attains a global maximum at a vertex of* Δ*.*

The next corollary is an easy consequence of Lemmas 10 and 11.

Corollary 12. *The maximum value of* $\log\|D_n(\boldsymbol{\omega})\|$ *over any simplex*

$$\Delta \subseteq \Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$$

is attained at a vertex of Δ*.*

We next show that the invariant density ρ, defined by (6), is a convex function. The following lemma gives another expression for the invariant density.

Lemma 13. *For all* $\boldsymbol{\omega} \in \Delta^d$*,*

$$\rho(\boldsymbol{\omega}) = d! \sum_{\pi \in S_d} \int_{\Delta^d} \frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+1}}\, d\boldsymbol{\psi}.$$

Proof. This can be checked by a direct calculation. Alternatively, it follows from [5]. □

For each $\boldsymbol{\psi} \in \Delta^d$ and $\pi \in S_d$ define $f_{\boldsymbol{\psi},\pi} : \Delta^d \to \mathbb{R}$ by

$$f_{\boldsymbol{\psi},\pi}(\boldsymbol{\omega}) = \frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+1}}.$$

Lemma 14. *For each* $\boldsymbol{\psi} \in \Delta^d$ *and* $\pi \in S_d$ *the function* $f_{\boldsymbol{\psi},\pi}$ *is convex on* Δ^d*.*

Proof. For $1 \leq k, l \leq d$,

$$\frac{\partial^2 f_{\boldsymbol{\psi},\pi}}{\partial\omega_k \partial\omega_l} = (d+1)(d+2)\frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+3}}\psi_{\pi(k)}\psi_{\pi(l)}.$$

We have

$$\begin{aligned}
\sum_{k,l=1}^d \frac{\partial^2 f_{\boldsymbol{\psi},\pi}}{\partial\omega_k \partial\omega_l} u_k u_l &= (d+1)(d+2) \sum_{k,l=1}^d \frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+3}}\psi_{\pi(k)}\psi_{\pi(l)} u_k u_l \\
&= (d+1)(d+2)\frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+3}} \sum_{k=1}^d \sum_{l=1}^d \psi_{\pi(k)}\psi_{\pi(l)} u_k u_l \\
&= (d+1)(d+2)\frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+3}} \sum_{k=1}^d \left(\psi_{\pi(k)} u_k \sum_{l=1}^d \psi_{\pi(l)} u_l\right) \\
&= (d+1)(d+2)\frac{1}{(1 + \sum_{i=1}^d \omega_i \psi_{\pi(i)})^{d+3}} \left(\sum_{l=1}^d \psi_{\pi(l)} u_l\right)^2 \geq 0.
\end{aligned}$$

Hence $f_{\boldsymbol{\psi},\pi}$ is convex on Δ^d. □

Proposition 15. *ρ is convex on Δ^d.*

Proof. Since $f_{\boldsymbol{\psi},\pi}$ is convex we have

$$f_{\boldsymbol{\psi},\pi}(\alpha\boldsymbol{\omega}+(1-\alpha)\boldsymbol{\omega}') \leq \alpha f_{\boldsymbol{\psi},\pi}(\boldsymbol{\omega})+(1-\alpha)f_{\boldsymbol{\psi},\pi}(\boldsymbol{\omega}') \quad \forall \boldsymbol{\omega},\boldsymbol{\omega}' \in \Delta^d, \alpha \in [0,1].$$

Integrating this inequality over all $\boldsymbol{\psi} \in \Delta^d$ and taking the sum over $\pi \in S_d$ gives the statement of the proposition. □

Corollary 16. (i) *The maximum value of ρ over a simplex $\Delta \subseteq \Delta^d$ occurs at one of the vertices of Δ, i.e.*

$$\max_{\boldsymbol{\omega}\in\Delta} \rho(\boldsymbol{\omega}) = \rho(\boldsymbol{v})$$

where $\boldsymbol{v}$ is a vertex of Δ.
(ii) *If $\Delta \subseteq \Delta^d$ is a simplex then*

$$\mu(\Delta) \leq \frac{1}{K}\left(\max_{1\leq i\leq d+1} \rho(\boldsymbol{v}_i)\right) \mathrm{vol}_d(\Delta)$$

where $\boldsymbol{v}_1,\dots,\boldsymbol{v}_{d+1}$ are the vertices of Δ and vol_d denotes d-dimensional Lebesgue measure.

Proof. This follows immediately from Lemma 11 and Proposition 15. □

We will also need an estimate for the lower bound of the density over a simplex. The following estimate is rather crude but it is enough for our purposes.

For a permutation $\pi \in S_d$ and $1 \leq i \leq d$ define $g_{\pi,i} : \Delta^d \to \mathbb{R}$ by

$$g_{\pi,i}(\boldsymbol{\omega}) = g_{\pi,i}(\omega_1,\dots,\omega_d) = \frac{1}{1+\omega_{\pi(1)}+\cdots+\omega_{\pi(i)}}. \tag{15}$$

Then

$$\rho(\boldsymbol{\omega}) = \sum_{\pi\in S_d}\prod_{i=1}^{d} g_{\pi,i}(\boldsymbol{\omega}).$$

Lemma 17. *For an arbitrary simplex $\Delta \subseteq \Delta^d$ and any $\pi \in S_d$, $1 \leq i \leq d$, the minimum of $g_{\pi,i}$ over Δ occurs at one of the vertices of Δ.*

Proof. This statement is obvious since $1/g_{\pi,i}$ is an affine function. □

Let $\boldsymbol{v}_1,\dots,\boldsymbol{v}_{d+1}$ denote the vertices of Δ. For each function $g_{\pi,i}$ let $\boldsymbol{v}_{k(\pi,i)}$ be the vertex at which the minimum of $g_{\pi,i}$ over Δ is attained, i.e.

$$g_{\pi,i}(\boldsymbol{v}_{k(\pi,i)}) = \min_{\boldsymbol{\omega}\in\Delta} g_{\pi,i}(\boldsymbol{\omega}). \tag{16}$$

Corollary 18. *For any $\boldsymbol{\omega} \in \Delta$*

$$\rho(\boldsymbol{\omega}) \geq \sum_{\pi\in S_d}\prod_{i=1}^{d} g_{\pi,i}(\boldsymbol{v}_{k(\pi,i)}).$$

5 Scheme of numerical estimation

In Section 2 it was shown (see Lemma 4) that the d-dimensional Gauss algorithm is strongly convergent almost everywhere if and only if there exists $n \in \mathbb{N}$ such that

$$\frac{1}{n} \int_{\Delta^d} \log \|D(T^{n-1}\boldsymbol{\omega}) \cdots D(T\boldsymbol{\omega}) D(\boldsymbol{\omega})\| \, \mu(d\boldsymbol{\omega}) < 0. \tag{17}$$

In this section we will describe how to find an upper bound for this integral numerically.

1. The first step is to find a plausible value of n for which (17) holds and this is normally done by Monte-Carlo estimation of the integral. Obviously it is preferable to have n as small as possible, although one also wants to have a "negative enough" value of the integral. For example, a rigorous negative upper bound for the integral (17) in the case $d = 3$ was obtained for $n = 8$ (see [6]), although Monte-Carlo estimations suggest that the integral is negative even for $n = 5$. Unfortunately, when $n = 5$ the integral is too small in absolute value to be used for rigorous estimates. For $d = 2$, Monte-Carlo estimates indicate that (17) is satisfied for $n \geq 2$ and for $d = 4$ it is satisfied for $n \geq 12$.

2. For fixed n, we have to perform the integration by splitting Δ^d into the elements of the Markov partition and integrating over each element separately. Since the number of elements of the Markov partition is infinite, numerically one has to divide the elements of the Markov partition into a finite part where the integration is really performed and an infinite part where one uses a crude upper bound for the value of the integral. Let Ξ_n denote the set of all elements of the n^{th} level of the Markov partition, i.e.

$$\Xi_n = \{\Delta_{(m_1,j_1),\ldots,(m_n,j_n)} : m_1, \ldots, m_n \in \mathbb{N}, 1 \leq j_1, \ldots, j_n \leq d\}.$$

Denote by Z_n the finite subset of Ξ_n which is used for integration. Certainly, for some m, $Z_n \subset Z_n(m)$ where $Z_n(m)$ is the set of elements $\Delta_{(m_1,j_1),\ldots,(m_n,j_n)}$ where all $m_i \leq m$, i.e.

$$Z_n(m) = \{\Delta_{(m_1,j_1),\ldots,(m_n,j_n)} \in \Xi_n : m_1, \ldots, m_n \leq m\}.$$

It would be easiest to consider the whole set $Z_n(m)$ but, because the number of elements can be huge, one may be forced to consider only those simplices whose invariant measure is not too small. We will see that the final set $Z_n(m)$ is determined in the course of the numerical estimation.

3. The set Z_n is divided into two parts: $Z_n = Z_n^{(1)} \cup Z_n^{(2)}$, where $Z_n^{(1)}$ consists of those $\overline{\Delta} \in Z_n$ whose diameter is small enough, namely

$$Z_n^{(1)} = \{\overline{\Delta} \in Z_n : \operatorname{diam}(\overline{\Delta}) \leq \alpha_n\}.$$

Of course the threshold value α_n has to be specified in advance. The elements of $Z_n^{(2)}$ are then subdivided into smaller simplices whose diameters are less than α_n. The subdivision is performed in an arbitrary way, and the simplices obtained are not necessarily the elements of the Markov partition. As a result we get a splitting of the whole set of integration

$$\Omega_n = \bigcup_{\overline{\Delta}\in Z_n} \overline{\Delta}$$

into non-intersecting simplices Δ of diameter smaller than α_n. Denote the set of these simplices Δ by $\overline{Z}_n$ so that

$$\Omega_n = \bigcup_{\Delta\in\overline{Z}_n} \Delta.$$

4. For each $\Delta \in \overline{Z}_n$ we estimate the integral over Δ

$$I_\Delta = \frac{1}{n}\int_\Delta \log\|D_n(\boldsymbol{\omega})\|\,\mu(d\boldsymbol{\omega})$$

from above and the invariant measure $\mu(\Delta)$ of the simplex Δ from below. Denote the vertices of Δ by $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{d+1}$. Let

$$d_n(\Delta) = \frac{1}{n}\max_{1\le i\le d+1} \log\|D_n(\boldsymbol{v}_i)\|$$

and

$$\overline{\rho}(\Delta) = \max_{1\le i\le d+1} \rho(\boldsymbol{v}_i), \quad \underline{\rho}(\Delta) = \sum_{\pi\in S_d}\prod_{i=1}^{d} g_{\pi,i}(\boldsymbol{v}_{k(\pi,i)}),$$

where $g_{\pi,i}$ is defined by (15) and $\boldsymbol{v}_{k(\pi,i)}$ is defined by (16).

If $d_n(\Delta) > 0$ then

$$I_\Delta \le d_n(\Delta)\frac{1}{K}\int_\Delta \rho(\boldsymbol{\omega})\,d\boldsymbol{\omega} \le \frac{1}{K}d_n(\Delta)\overline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta).$$

For $d_n(\Delta) < 0$

$$I_\Delta \le d_n(\Delta)\frac{1}{K}\int_\Delta \rho(\boldsymbol{\omega})\,d\boldsymbol{\omega} \le \frac{1}{K}d_n(\Delta)\underline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta).$$

Also

$$\mu(\Delta) = \frac{1}{K}\int_\Delta \rho(\boldsymbol{\omega})\,d\boldsymbol{\omega} \ge \frac{1}{K}\underline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta).$$

Denote

$$\overline{I}_\Delta = \begin{cases} \dfrac{1}{K}d_n(\Delta)\underline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta) & \text{if } d_n(\Delta) < 0; \\ \dfrac{1}{K}d_n(\Delta)\overline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta) & \text{if } d_n(\Delta) > 0; \end{cases}$$

and

$$\underline{\mu}(\Delta) = \frac{1}{K}\underline{\rho}(\Delta)\,\mathrm{vol}_d(\Delta).$$

Then

$$\frac{1}{n}\int_{\Omega_n} \log\|D_n(\boldsymbol{\omega})\|\,\mu(d\boldsymbol{\omega}) \le \sum_{\Delta\in\overline{Z}_n} \overline{I}_\Delta$$

and

$$\mu(\Omega_n) \geq \sum_{\Delta \in \overline{Z}_n} \underline{\mu}(\Delta).$$

Denote

$$\delta_n = 1 - \sum_{\Delta \in \overline{Z}_n} \underline{\mu}(\Delta).$$

Then obviously

$$\mu(\Delta^d \setminus \Omega_n) \leq \delta_n.$$

5. We now estimate the integral over $\Delta^d \setminus \Omega_n$. Notice that

$$\frac{1}{n} \log \|D_n(\boldsymbol{\omega})\| \leq \max_{\boldsymbol{\omega} \in \Delta^d} \log \|D(\boldsymbol{\omega})\| \leq \log \sqrt{d+1} = \frac{1}{2} \log(d+1).$$

Hence

$$\frac{1}{n} \int_{\Delta^d \setminus \Omega_n} \log \|D_n(\boldsymbol{\omega})\| \, \mu(d\boldsymbol{\omega}) \leq \frac{1}{2} \log(d+1) \delta_n.$$

6. Finally we get the following estimate

$$\frac{1}{n} \int_{\Delta^d} \log \|D_n(\boldsymbol{\omega})\| \, \mu(d\boldsymbol{\omega}) \leq \left(\sum_{\Delta \in \overline{Z}_n} \overline{I}_\Delta \right) + \frac{1}{2} \log(d+1) \delta_n. \tag{18}$$

Including more simplices in the set Z_n makes the first sum more negative. It also decreases δ_n and hence the second term. One has to stop when the right hand side of (18) is negative. This condition essentially determines how big the set Z_n is. The above procedure is relatively easy to implement on a computer. As was mentioned above, the computer assisted proof in the case $d = 3$ was performed for $n = 8$. The program was run for 750 hours on a SUN Ultra 5 and this produced the rigorous estimate $\lambda_1(D) < -0.005329$.

In reality, the calculation differs from the scheme above in just a few technical details. Basically, in some parts of the complementary set $\Delta^d \setminus \Omega_n$, we used better estimates for $\frac{1}{n} \log \|D_n(\boldsymbol{\omega})\|$ than the crude estimate $\frac{1}{2} \log(d+1)$ (see [6] for more details).

6 Discussion

We have described a scheme which can be used to give a computer assisted proof of almost everywhere strong convergence of the d-dimensional Gauss algorithm for any particular dimension d. It is easy to carry out the scheme in two dimensions. The three-dimensional case is significantly harder, although one can obtain the desired results (see [6]). In higher dimensions, it becomes even harder to implement. However, there is no reason to doubt that the result is true in all dimensions.

Although the scheme can in principle be used for other algorithms of Jacobi-Perron type, it will require very good approximations of the invariant measure. Producing such approximations seems to be a hard problem. Here, of course, we fully use the advantage of knowing an explicit formula for the invariant density, which doesn't exist for many other algorithms including the Jacobi-Perron algorithm itself.

There remains the challenging open problem of finding a conceptual proof of strong convergence of continued fraction algorithms in arbitrary dimension.

References

[1] A. Broise-Alamichel and Y. Guivarc'h, Exposants caractéristiques de l'algorithme de Jacobi-Perron et de la transformation associée, preprint (1999).

[2] V. Brun, Algorithmes euclidiens pour trois et quatre nombres, in: *13 ième Congre. Math. Scand., Helsinki* (1957), 45-64.

[3] T. Fujita, S. Ito, M. Keane and M. Ohtsuki, On almost everywhere exponential convergence of the modified Jacobi-Perron algorithm: a corrected proof, *Ergod. Th. and Dyn. Sys.* **16** (1996), 1345-1352.

[4] D. M. Hardcastle and K. Khanin, On almost everywhere strong convergence of multidimensional continued fraction algorithms, to appear in *Ergodic Theory and Dynamical Systems.*

[5] D. M. Hardcastle and K. Khanin, Continued fractions and the d-dimensional Gauss transformation, to appear in *Communications in Mathematical Physics.*

[6] D. M. Hardcastle, The three-dimensional Gauss algorithm is strongly convergent almost everywhere, to be published.

[7] S. Ito, M. Keane and M. Ohtsuki, Almost everywhere exponential convergence of the modified Jacobi-Perron algorithm, *Ergod. Th. and Dyn. Sys.* **13** (1993), 319-334.

[8] C. G. J. Jacobi, Allgemeine theorie der kettenbruchähnlichen algorithmen, in welchen jede Zahl aus drei vorhergehenden gebildet wird, *J. Reine Angew. Math* **69** (1868), 29-64.

[9] K. Khanin, Talk at the International Workshop on Dynamical Systems, Porto, August 1992.

[10] J. F. C. Kingman, The ergodic theory of subadditive stochastic processes, *J. Royal Stat. Soc. B* **30** (1968), 499-510.

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

[12] R. Meester, A simple proof of the exponential convergence of the modified Jacobi-Perron algorithm, *Ergod. Th. and Dyn. Sys.* **19** (1999), 1077-1083.

[13] R. E. A. C. Paley and H. D. Ursell, Continued fractions in several dimensions, *Proc. Cambridge Philos. Soc* **26** (1930), 127-144.

[14] O. Perron, Grundlagen für eine theorie des Jacobischen kettenbruchalgorithmus, *Math. Ann.* **64** (1907), 1-76.

[15] E. V. Podsypanin, A generalization of the algorithm for continued fractions related to the algorithm of Viggo Brunn, *Zap. Naucn. Sem. Leningrad Otdel. Mat. Inst. Steklov* **67** (1977), 184-194. English translation: *Journal of Soviet Math.* **16** (1981), 885-893.

[16] F. Schweiger, A modified Jacobi-Perron algorithm with explicitly given invariant measure, in: *Ergodic Theory, Proceedings Oberwolfach, Germany 1978, Lecture Notes in Mathematics* **729**, 199-202, Springer-Verlag (1979).

[17] F. Schweiger, Ergodic Theory of Fibred Systems and Metric Number Theory, Oxford University Press (1995).

[18] F. Schweiger, The exponent of convergence for the 2-dimensional Jacobi-Perron algorithm, *Proceedings of the Conference on Analytic and Elementary Number Theory in Vienna 1996 (Ed.: W. G. Nowak and J. Schoissengeier)*, 207-213.

[19] E. Selmer, Om flerdimensjonal Kjede brøk, *Nordisk Mat. Tidskr.* **9** (1961), 37-43.