Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 59, No. 2, June 15, 1977

A Jacobi Algorithm and Metric Theory for Greatest Common Divisors*

M. S. WATERMAN

Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

Submitted by Gian-Carlo Rota

Greatest common divisor algorithms are used to provide a natural motivation for considering a class of Jacobi–Perron algorithms which includes the original Jacobi algorithm. This work proves convergence and establishes metric properties for one of these algorithms. The proofs generalize to the larger class of algorithms. Full connections with the calculation of greatest common divisors will be treated elsewhere.

1. INTRODUCTION

In 1868 in a posthumous paper, Jacobi [3] generalized continued fractions to two dimensions. One of Jacobi's motivations was to characterize real algebraic irrationalities of degree higher than two, a problem that is still unsolved in the framework of Jacobi's algorithm. Of course Minkowski, proceeding along different lines, solved the characterization problem in 1899 [1, p. 7].

Perron [6] in 1907 extended Jacobi's algorithm to n-dimensions ($n \geqslant 1$) and proved many important results including convergence. It is useful for this discussion to present a version of the Jacobi–Perron algorithm. Let $\mathbf{x} \in \{(t_1, t_2, \ldots, t_n): 0 \leqslant t_i < 1 \text{ for } i = 1, 2, \ldots, n\}$. Then define, for $x_1 \neq 0$,

$$T(\mathbf{x}) = \left(\frac{x_2}{x_1} - \left[\frac{x_2}{x_1}\right], \ldots, \frac{x_n}{x_1} - \left[\frac{x_n}{x_1}\right], \frac{1}{x_1} - \left[\frac{1}{x_1}\right]\right),$$

$$\mathbf{a}^1(\mathbf{x}) = \left(\left[\frac{x_2}{x_1}\right], \ldots, \left[\frac{x_n}{x_1}\right], \left[\frac{1}{x_1}\right]\right),$$

and

$$\mathbf{a}^\nu(\mathbf{x}) = \mathbf{a}^1(T^{\nu-1}(\mathbf{x})) = (a_1^\nu, a_2^\nu, \ldots, a_n^\nu).$$

* This work was partially supported by the National Science Foundation under Grant DCR 75-07070 and also performed under the auspices of the USERDA.

ISSN 0022-247X

Next define

$$\Lambda^0 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\Lambda^\nu = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & a_1^\nu \\ 0 & 1 & \cdots & 0 & a_2^\nu \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & a_n^\nu \end{pmatrix},$$

and

$$\Omega_\nu = \Lambda^0 \Lambda^1 \cdots \Lambda^{\nu-1} = (\omega_{ij}^\nu),$$

where i and j belong to $\{0, 1,..., n\}$. These matrices imply

$$\omega_{jn}^{\nu+1} = w_{j0}^\nu + a_1^\nu \omega_{j1}^\nu + \cdots + a_n^\nu \omega_{jn}^\nu .$$

The convergence result of Perron states

$$\lim_{\nu\to\infty}(\omega_{in}^\nu/\omega_{0n}^\nu) = x_i .$$

Much of the modern work on the Jacobi–Perron algorithm can be found in the books of Bernstein [1], who considers periodicity and algebraic number fields, and Schweiger [7], who considers metric theory. Bernstein's work contains many generalized Jacobi–Perron algorithms.

The Jacobi algorithm (actually a class of algorithms) presented in this paper is not motivated by periodicity or algebraic fields but by greatest common divisors (g.c.d.'s). It is clear by an examinaton of two of Jacobi's papers [2, 3] that he was aware of the connection between his algorithm and greatest common divisors and in fact the Jacobi–Perron algorithm can be naturally motivated by making this connection clear.

To begin, consider Euclid's algorithm for greatest common divisors. Given a pair of integers (m, l), with $m \leqslant l$, the algorithm is to perform

$$Q(m, l) = (l \bmod m, m)$$

until $l \bmod m = 0$. Since

$$\text{g.c.d.}(m, l) = \text{g.c.d.}\, Q(m, l),$$

the final value of the second coordinate is the greatest common divisors of m and l. The connection between Euclid's algorithm and continued fractions is well known and can be stated in the following fashion. Let the relation

$$(m, l) \sim m/l$$

associate each pair of integers $0 \leqslant m < l$ with a point in $[0, 1)$. Then

$$Q(m, l) = (l - [l/m]\, m, m) \sim (l/m) - [l/m] = T(m/l),$$

where

$$T(x) = (1/x) - [1/x]$$

is the shift on the digits of a continued fraction. That is, if $x = [a_1, a_2, \ldots]$, then $T(x) = [a_2, a_3, \ldots]$.

To generalize this consideration, let an $n + 1$ tuple of integers be given $\mathbf{m} = (m_1, m_2, \ldots, m_{n+1})$ where $m_i < m_{n+1}$ for all i. Then define

$$Q(\mathbf{m}) = (m_2 \bmod m_1, \ldots, m_{n+1} \bmod m_1, m_1).$$

Of course the greatest common divisor of $\mathbf{m}$ is equal to the greatest common divisor of $Q(\mathbf{m})$. Now define

$$\mathbf{m} \sim (m_1/m_{n+1}, m_2/m_{n+1}, \ldots, m_n/m_{n+1}),$$

which associates each $\mathbf{m}$ with a point in $([0, 1))^n$. Also

$$Q(\mathbf{m}) \sim \left(\frac{m_2}{m_1} - \left[\frac{m_2}{m_1}\right], \ldots, \frac{m_{n+1}}{m_1} - \left[\frac{m_{n+1}}{m_1}\right]\right)$$

$$= T\left(\frac{m_1}{m_{n+1}}, \frac{m_2}{m_{n+1}}, \ldots, \frac{m_n}{m_{n+1}}\right),$$

where T is the transformation associated with the Jacobi–Perron algorithm above.

If Q is examined from the point of view of computing greatest common divisors, however, it is clear that m_1 should be the smallest of all the m_i. In fact Knuth [4, p. 300] states this algorithm in his book "Seminumerical Algorithms." In this way the algorithm should converge faster. Therefore let $\mathbf{m}$ be a vector such that $0 < m_1 \leqslant m_2 < \cdots \leqslant m_{n+1}$. The relation $\sim$ associates $\mathbf{m}$ with a vector in $I^* = \{(t_1, \ldots, t_n): 0 < t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n < 1\}$. If $O(s_1, \ldots, s_n) = (s_{i_1}, s_{i_2}, \ldots, s_{i_n}) \in I^*$, then it is natural to define, for $\mathbf{t} \in I^*$,

$$S(\mathbf{t}) = O(T(\mathbf{t})).$$

The object of this paper is to state explicitly the Jacobi algorithm associated with the transformation S and to prove convergence and metric properties of the algorithm. The work closely follows Schweiger's treatment [7] of the Jacobi–Perron algorithm and previous work on multidimensional F-expansions [9]. It turns out that any permutation will also define a Jacobi algorithm as well as O and this point will be returned to in the last section. It should be noted that Paley and Ursell [5] consider a similar class of continued fractions which they treat in a quite different fashion. The explicit motivation of greatest common divisors does not appear there and the result of this paper that seems to have an analog in [5] is Lemma 3.1(b).

2. Definition and Convergence of the Algorithm

In this section the Jacobi algorithm associated with S is defined and convergence of the algorithm is shown. Let $I^* = \{\mathbf{x} \in (0, 1)^n : 0 < x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n < 1\}$. For $\mathbf{x} \in I^*$ define

$$F(\mathbf{x}) = \left(\frac{1}{x_n}, \frac{x_1}{x_n}, \ldots, \frac{x_{n-1}}{x_n}\right),$$

$$\mathbf{a}(\mathbf{x}) = \left(\left[\frac{x_2}{x_1}\right], \ldots, \left[\frac{x_n}{x_1}\right], \left[\frac{1}{x_1}\right]\right),$$

and

$$T(\mathbf{x}) = \left(\frac{x_2}{x_1} - \left[\frac{x_2}{x_1}\right], \ldots, \frac{x_n}{x_1} - \left[\frac{x_n}{x_1}\right], \frac{1}{x_1} - \left[\frac{1}{x_1}\right]\right).$$

If $\mathbf{t} \in (0, 1)^n$, let $O(\mathbf{t}) = (t_{i_1}, t_{i_2}, \ldots, t_{i_n}) \in I^*$. Then, S is defined by

$$S(\mathbf{x}) = O(T(\mathbf{x})). \tag{2.1}$$

Next let $\sigma(\mathbf{x})$ be a permutation such that

$$\sigma(\mathbf{x})\, S(\mathbf{x}) = T(\mathbf{x}).$$

That is, we require $s_{\sigma_1} = t_1, \ldots, s_{\sigma_n} = t_n$. Finally, define σ^0 to be the identity permutation and

$$\mathbf{a}^1(\mathbf{x}) = \mathbf{a}(\mathbf{x}), \qquad \sigma^1(\mathbf{x}) = \sigma(\mathbf{x}),$$

$$\mathbf{a}^i(\mathbf{x}) = \mathbf{a}(S^{i-1}(\mathbf{x})), \qquad \sigma^i(\mathbf{x}) = \sigma(S^{i-1}(\mathbf{x})) \quad \text{for } i > 1.$$

Let $N = \{\mathbf{x} \in I^* : (S^k(\mathbf{x}))_1 = 0 \text{ for some } k \geqslant 0\}$. The set N is contained in the intersection of I^* with a countable union of hyperplanes, and consequently N has n-dimensional Lebesgue measure equal to 0. This assertion about N can be shown as in [7, Lemma 1.1]. Consequently the set of interest will be

$$I = I^* \sim N.$$

The first observation is that, for $\nu \geqslant 1$,

$$\begin{aligned}
\mathbf{x} &= F(\mathbf{a}(\mathbf{x}) + T(\mathbf{x})) \\
&= F(\mathbf{a}^1 + \boldsymbol{\sigma}^1 S(\mathbf{x})) \\
&= F(\mathbf{a}^1 + \boldsymbol{\sigma}^1 F(\mathbf{a}^2 + \boldsymbol{\sigma}^2 S^2(\mathbf{x}))) \qquad (2.2)\\
&\vdots \\
&= F(\mathbf{a}^1 + \boldsymbol{\sigma}^1 F(\mathbf{a}^2 + \boldsymbol{\sigma}^2 F(\cdots + \boldsymbol{\sigma}^{\nu-1} F(\mathbf{a}^\nu + \boldsymbol{\sigma}^\nu S^\nu(\mathbf{x}) \cdots)))).
\end{aligned}$$

Next, martices analogous to those of Perron are defined. This step is the key to the results for the algorithm. Let

$$A_0 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$A_i = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & a_1{}^i \\ 0 & 1 & \cdots & 0 & a_2{}^i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & a_n{}^i \end{pmatrix}, \qquad i \geqslant 1.$$

These matrices are exactly those defined by Perron and correspond to $\mathbf{a}(\mathbf{x})$. Next the matrices for $\boldsymbol{\sigma}(\mathbf{x})$ are defined. Let E be the $(n+1) \times (n+1)$ identity matrix. Define

$$\Sigma_{-1} = \Sigma_0 = E,$$

and

$$\Sigma_\nu = (\alpha_{ij}), \qquad i, j = 0, 1,..., n,$$

where, for $i < n$,

$$\begin{aligned} \alpha_{ij} &= 1, \qquad \text{if} \quad j = \sigma_{i+1}^\nu - 1, \\ &= 0 \qquad \text{otherwise;} \end{aligned}$$

and

$$\begin{aligned} \alpha_{nj} &= 1, \qquad \text{if} \quad j = n, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Finally, define

$$\Omega_\nu = (A_{ij}^{(\nu)}), \qquad i, j = 0, 1,..., n,$$

by

$$\Omega_0 = E$$

and

$$\Omega_{\nu+1} = \Sigma_{-1}\Lambda_0\Sigma_0\Lambda_1 \cdots \Sigma_{\nu-1}\Lambda_\nu$$
$$= \Omega_\nu \Sigma_{\nu-1}\Lambda_\nu\,, \qquad \nu \geqslant 0.$$

Now $\Sigma_{\nu-1}\Lambda_\nu$ is a row permutation of the matrix Λ_ν with the nth row left unchanged. The ith row ($i < n$) of $\Sigma_{\nu-1}\Lambda_\nu$ is zero except for a 1 in the $j = \sigma_{i+1}^{\nu-1} - 2$ column and $a^{\nu}_{\sigma_{i+1}^{\nu-1}-1}$ in the nth column. Therefore

$$A_{i,l}^{(\nu+1)} = A_{ij}^{(\nu)}, \qquad \text{where} \qquad l = \sigma_{j+1}^{\nu-1} - 2,$$

$$A_{i,n-1}^{(\nu+1)} = A_{i,n}^{(\nu)}\,,$$

and

$$A_{i,n}^{(\nu+1)} = a_n{}^\nu A_{i,n}^{(\nu)} + \sum_{j=0}^{n-1} a^{\nu}_{\sigma_{j+1}^{\nu-1}-1} A_{i,j}^{(\nu)}\,.$$

Notice the relationship between the multiplication of matrices in the definition of Ω_ν and Eq. (2.2). Matrices corresponding to the permutations have been inserted between the Λ_i.

The next theorem represents the components of $\mathbf{x}$ and will be used to show convergence.

THEOREM 2.1. *If* $1 \leqslant i \leqslant n$, $0 \leqslant \nu$, *and* $S^\nu(\mathbf{x}) = \mathbf{y}$, *then*

$$x_i = \frac{A_{i,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu+1)} y_{\sigma_{j+1}^{\nu}}}{A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^{\nu}}}\,.$$

Proof. Let $\nu = 0$ and $\boldsymbol{\sigma}^0$ be the identity permutation. Then $y = S^0(\mathbf{x}) = \mathbf{x}$ and, for $1 \leqslant i \leqslant n$,

$$\frac{A_{i,n}^{(1)} + \sum_{j=0}^{n-1} A_{i,j}^{(1)} x_{j+1}}{A_{0,n}^{(1)} + \sum_{j=0}^{n-1} A_{0,j}^{(1)} x_{j+1}} = \frac{0 + x_i}{1} = x_i\,.$$

Next let $\nu \geqslant 1$, $S^{\nu-1}(\mathbf{x}) = \mathbf{y}$, $S^\nu(\mathbf{x}) = S(\mathbf{y}) = \mathbf{z}$ and assume

$$x_i = \frac{A_{i,n}^{(\nu)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu)} y_{\sigma_{j+1}^{\nu-1}}}{A_{0,n}^{(\nu)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu)} y_{\sigma_{j+1}^{\nu-1}}}\,, \qquad 1 \leqslant i \leqslant n.$$

Since $S(y) = \mathbf{z}$,

$$y_2/y_1 = a_1{}^\nu + z_{\sigma_1{}^\nu},$$
$$\vdots$$
$$y_n/y_1 = a_{n-1}^\nu + z_{\sigma_{n-1}^\nu},$$
$$1/y_1 = a_n{}^\nu + z_{\sigma_n{}^\nu}.$$

If $a_0{}^\nu = 1$ and $z_{\sigma_0{}^\nu} = 0$, then

$$y_j = (a_{j-1}^\nu + z_{\sigma_{j-1}})/(a_n{}^\nu + z_{\sigma_n{}^\nu}).$$

Therefore, for $1 \leqslant i \leqslant n$,

$$(a_n{}^\nu + z_{\sigma_n{}^\nu})\left(A_{i,n}^{(\nu)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu)} y_{\sigma_{j+1}^{\nu-1}}\right)$$

$$= a_n A_{i,n}^{(\nu)} + z_{\sigma_n{}^\nu} A_{i,j}^{(\nu)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu)} (a^\nu_{\sigma_{j+1}^{\nu-1}-1} + z_{\sigma^\nu_{\sigma_{j+1}^{\nu-1}-1}})$$

$$= a_n{}^\nu A_{i,n}^{(\nu)} + \sum_{j=0}^{n-1} A_{i,j}^\nu a^\nu_{\sigma_{j+1}^{\nu-1}-1} + z_{\sigma_n{}^\nu} A_{i,n}^{(\nu)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu)} z_{\sigma^\nu_{\sigma_{j+1}^{\nu-1}-1}}$$

$$= A_{i,n}^{(\nu+1)} + \sum_{j=0}^{n-1} z_{\sigma_{j+1}^\nu} A_{i,j}^{(\nu+1)}.$$

The last step follows from the remarks about $A_{i,j}^{(\nu+1)}$ preceding the theorem. The theorem follows by taking ratios of this last equation.

The following corollary follows immediately from Theorem 2.1.

COROLLARY 2.1. *If* $\mathbf{x} = F(\mathbf{a}^1 + \boldsymbol{\sigma}^1 F(\mathbf{a}^2 + \cdots + \boldsymbol{\sigma}^{\nu-1} F(\mathbf{a}^\nu) \cdots)$, *then, for* $1 \leqslant i \leqslant n$,

$$x_i = A_{i,n}^{(\nu+1)} / A_{0,n}^{(\nu+1)}$$

To prove convergence of the algorithm, a definition of a cylinder of order ν is required. Let

$$\mathbf{b}^i(\mathbf{x}) = (\boldsymbol{\sigma}^{i-1}(x), \mathbf{a}^i(\mathbf{x})),$$

and

$$B_\nu(\mathbf{b}^1, \ldots, \mathbf{b}^\nu) = \{\mathbf{x} : \mathbf{b}^i(\mathbf{x}) = \mathbf{b}^i, 1 \leqslant i \leqslant \nu\}.$$

Sometimes $B_\nu(\mathbf{b}^1, \ldots, \mathbf{b}^\nu)$ will be abbreviated to B_ν. Next let

$$B_\nu(\mathbf{x}) = \{\mathbf{y} : \mathbf{b}^i(\mathbf{y}) = \mathbf{b}^i(\mathbf{x}), 1 \leqslant i \leqslant \nu\},$$

and define

$$\xi(\nu) = \sup_{\mathbf{x} \in I} \operatorname{diam} B_\nu(\mathbf{x}).$$

Clearly $\xi(\nu + 1) \leqslant \xi(\nu) \leqslant 1$. The algorithm will converge if $\lim_{\nu \to \infty} \xi(\nu) = 0$. The next theorem, following a result of Fischer [7, p. 47], shows the convergence is geometric.

THEOREM 2.2. *For* $\nu < 1$, $\xi(\nu) = O(\theta)^\nu$, *where*

$$\theta = (1 - (n+1)^{-n})^{1/n}.$$

Proof. Since $A_{i,n}^{(\nu)}/A_{0,n}^{(\nu)} \in B(\mathbf{b}^1, ..., \mathbf{b}^\nu)$, it is sufficient to show

$$\left| x_i - \frac{A_{i,n}^{(\nu)}}{A_{0,n}^{(\nu)}} \right| = O(\theta^\nu).$$

By the definition of $A_{i,n}^{(\nu+1)}$, for some permutation π,

$$A_{i,n}^{(\nu+1)}/A_{0,n}^{(\nu+1)} = \sum_{j=0}^{n} \lambda_j (A_{i,n}^{(\nu+n-j)}/A_{0,n}^{(\nu+n-j)})$$

where

$$\lambda_n = a_n A_{0,n}^{(\nu)}/A_{0,n}^{(\nu+1)} \qquad \text{and} \qquad \lambda_j = a_{\pi_j}^{\nu} A_{0,n}^{(\nu+n-j)}/A_{0,n}^{(\nu+1)}.$$

It is easy to see that $\lambda_j \geqslant 0$, $\sum_{j=1}^{n} \lambda_j = 1$, and, since $1 \leqslant a_j^\nu \leqslant a_n^\nu$ (see Lemma 3.1(c)), $\lambda_n \geqslant 1/(n+1)$. It is easy to show by induction that

$$A_{i,n}^{(\nu+g)}/A_{0,n}^{(\nu+g)} = \sum_{j=0}^{n} \lambda_j^{(g)} (A_{i,n}^{(\nu+n-j)}/A_{0,n}^{(\nu+n-j)}),$$

where $g \geqslant 1$, $\lambda_j^{(g)} \geqslant 0$, $\lambda_n^{(g)} \geqslant (n+1)^{-g}$, and $\sum_{j=0}^{n} \lambda_j^{(g)} = 1$. Then

$$\left| \frac{A_{i,n}^{(\nu+g)}}{A_{0,n}^{(\nu+g)}} - \frac{A_{i,n}^{(\nu)}}{A_{0,n}^{(\nu)}} \right| = \left| \sum_{j=0}^{n-1} \lambda_j^{(g)} \left(\frac{A_{i,n}^{(\nu+n-j)}}{A_{0,n}^{(\nu+n-j)}} - \frac{A_{i,n}^{(\nu)}}{A_{0,n}^{(\nu)}} \right) \right| \leqslant (1 - (n+1)^{-g})\, \xi(\nu).$$

Next, for $1 \leqslant g < h \leqslant n$,

$$\left| \frac{A_{i,n}^{(\nu+g)}}{A_{0,n}^{(\nu+g)}} - \frac{A_{i,n}^{(\nu+h)}}{A_{0,n}^{(\nu+h)}} \right| \leqslant (1 - (n+1)^{-(h-g)})\, \xi(\nu+g) \leqslant (1 - (n+1)^{-n})\, \xi(\nu).$$

Adding and subtracting x inside the leftmost member of the above inequality and using a form of the triangle inequality show that

$$\xi(\nu + n) \leqslant \theta \xi(\nu).$$

3. Metric Theory

The theorems of this section will follow from the theory presented in [7] or [9] as soon as some preliminary results are established. Let m be Lebesgue measure on $(0, 1)^n$. Then m normalized on I will be defined by

$$\lambda(A) = n!m(A).$$

Part (c) of the next lemma was used in the proof of Theorem 2.2.

Lemma 3.1. *For some* $\mathbf{t} \in I$, *let* $B_\nu = B_\nu(\mathbf{t})$ *for* $\nu \geqslant 1$. *Then*

(a) $S^\nu(B_\nu) = I$ *so that* $\lambda(S^\nu B_\nu) = 1$,

(b) $\det \Omega_\nu = \pm 1$,

(c) $1 \leqslant a_1{}^i \leqslant a_2{}^i \leqslant \cdots \leqslant a_n{}^i$ *for* $i \geqslant 1$,

(d) $A_{0,j}^{(\nu+1)} \leqslant A_{0,n}^{(\nu+1)}$ *for* $1 \leqslant j \leqslant n$.

Proof. (a) If $B_\nu = B_\nu(\mathbf{b}^1,..., \mathbf{b}^\nu)$, then $F(\mathbf{a}^1 + \boldsymbol{\sigma}^1 F(\cdots + F(\mathbf{a}^\nu + \mathbf{t}) \cdots))$ is defined for all $\mathbf{t} \in I$.

(b) $\det \Omega_\nu = \det(\Omega_{\nu-1}) \det(\Sigma_{\nu-1}) \det(\Lambda_\nu) = \det(\Omega_{\nu-1}) (\pm 1) (-1)^n$.

(c) Since $S^{i-1}(\mathbf{x}) = \mathbf{y}$ satisfies $0 < y_1 < \cdots < y_n < 1$ for $i \geqslant 1$, then $1 < y_2/y_1 < y_3/y_1 < \cdots < 1/y_1$ and the result follows.

(d) Since $\Omega_1 = \Lambda_0$ and $0 \leqslant 1$, $A_{0,j}^{(1)} \leqslant A_{0,n}^{(1)}$ and the result holds for $\nu = 0$. Assume that (d) holds for $\nu \leqslant m$. If $j < n$, then

$$A_{0,l}^{(\nu+1)} = A_{0,j}^{(\nu)}, \qquad \text{where} \qquad l = \sigma_{j+1}^{\nu-1} - 2,$$

and

$$A_{0,n-1}^{(\nu+1)} = A_{0,n}^{(\nu)}.$$

But

$$A_{0,n}^{(\nu)} \leqslant a_n A_{0,n}^{(\nu)} + \sum_{j=0}^{n-1} a^\nu_{\sigma_{j+1}^{\nu-1}-1} A_{0,j}^{(\nu)} = A_{0,n}^{(\nu+1)},$$

and (d) follows by induction.

Next, following [9] for $\mathbf{t} \in I$ let

$$f_{\mathbf{b}^i}(\mathbf{t}) = \boldsymbol{\sigma}^{i-1} F(\mathbf{a}^i + \mathbf{t})$$

Then, if $B_\nu = B(\mathbf{b}^1,..., \mathbf{b}^\nu)$,

$$B_\nu = \prod_{i=1}^{\nu} of_s{}^i(I).$$

The next theorem follows [7, Lemma 2.4] and is a key result in establishing the metric theory.

THEOREM 3.1. *For* $\nu \geqslant 1$ *let* $B_\nu = B(\mathbf{b}^1, \ldots, \mathbf{b}^\nu)$ *and* $f_\nu = \prod_{i=1}^{\nu} of_{\mathbf{b}^i}$. *Then the absolute value of the Jacobian of* f_ν, J_ν, *satisfies*

$$| J_\nu(\mathbf{y})| = \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)^{-n-1}.$$

Proof. Note that, if $\mathbf{x} = \prod_{i=1}^{\nu} of_{\mathbf{b}^i}(y)$, then $S^\nu \mathbf{x} = \mathbf{y}$. Therefore Theorem 2.1 implies

$$x_i = \frac{A_{i,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}}{A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}}.$$

Thus, for $1 \leqslant i,\ j \leqslant n$,

$$\begin{aligned}\frac{\partial x_i}{\partial y_{\sigma_{j+1}^\nu}} &= \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)^{-2} \left\{A_{i,j-1}^{(\nu+1)} \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)\right.\\ &\quad \left. - A_{0,j-1}^{(\nu+1)} \left(A_{i,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{i,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)\right\}\\ &= \frac{A_{i,j-1}^{(\nu+1)} - x_i A_{0,j-1}^{(\nu+1)}}{A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}}.\end{aligned}$$

Now

$$\begin{aligned}\det\left(\frac{\partial x_i}{\partial y_j}\right) &= \pm \det\left(\frac{\partial x_i}{\partial y_{\sigma_j{}^\nu}}\right)\\ &= \pm \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)^{-n} \det(A_{i,j-1}^{(\nu+1)} - x_i A_{0,j-1}^{(\nu+1)}).\end{aligned}$$

The determinant on the right-hand side of the last equation is equal to

$$\begin{aligned}&\det \begin{pmatrix} 1 & A_{0,0}^{(\nu+1)} & \cdots & A_{0,n-1}^{(\nu+1)} \\ 0 & A_{1,0}^{(\nu+1)} - x_1 A_{0,0}^{(\nu+1)} & \cdots & A_{1,n-1}^{(\nu+1)} - x_1 A_{0,n-1}^{(\nu+1)} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & A_{n,0}^{(\nu+1)} - x_n A_{0,0}^{(\nu+1)} & \cdots & A_{n,n-1}^{(\nu-1)} - x_n A_{0,n-1}^{(\nu+1)} \end{pmatrix}\\ &= \det \begin{pmatrix} 1 & A_{0,0}^{(\nu+1)} & \cdots & A_{0,n-1}^{(\nu+1)} \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & A_{1,0}^{(\nu+1)} & \cdots & A_{1,n-1}^{(\nu+1)} \\ x_n & A_{n,0}^{(\nu+1)} & \cdots & A_{n,n-1}^{(\nu+1)} \end{pmatrix}\\ &= \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)} y_{\sigma_{j+1}^\nu}\right)^{-1} (\pm \det(A_{i,j}^{(\nu+1)})).\end{aligned}$$

The last equality follows by Theorem 2.1 and the proof is easily completed.

The next corollary establishes condition (C) for S.

COROLLARY 3.1. *If* $B_\nu = B_\nu(\mathbf{b}^1, \ldots, \mathbf{b}^\nu)$, *then, for* $\mathbf{t} \in I$,

$$\frac{\sup |J_\nu(\mathbf{t})|}{\inf |J_\nu(\mathbf{t})|} \leqslant C = (n+1)^{n+1}.$$

Proof. By Theorem 3.1,

$$\sup |J_\nu(\mathbf{t})| \leqslant (A_{0,n}^{\nu+1})^{-n-1}$$

and

$$\begin{aligned}\inf |J_\nu(\mathbf{t})| &\geqslant \left(A_{0,n}^{(\nu+1)} + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)}\right)^{-n-1} \\ &= (A_{0,n}^{(\nu+1)})^{-n-1}\left(1 + \sum_{j=0}^{n-1} A_{0,j}^{(\nu+1)}/A_{0,n}^{(\nu+1)}\right)^{-n-1} \\ &\geqslant (A_{0,n}^{(\nu+1)})^{-n-1}(1+n)^{-n-1}.\end{aligned}$$

The last inequality follows from Lemma 3.1(d).

The next theorem follows from [9] or as in [7] and establishes the ergodic theory of S.

THEOREM 3.2. *There exists a probability measure* μ *on* I *such that* $\mu \sim \lambda$, *S is a measure-preserving transformation for μ, and S is ergodic under μ or λ. This implies that the ergodic theorem holds and, for any g which is Lebesgue integrable on I,*

$$\lim_{\nu\to\infty} \nu^{-1} \sum_{j=0}^{\nu-1} g(S^j(\mathbf{x})) = \int g\, d\mu \qquad \textit{for } \text{a.a. } \mathbf{x}.$$

The ergodic theorem implies that digit frequencies exist for almost all $\mathbf{x} \in I$. Therefore not only does $\mathbf{a} = (2, 3)$ have a limiting frequency (for $n = 2$) but $\boldsymbol{\sigma} = (2, 1)$ also does.

Next, some conclusions related to Rohlin's formula are stated. See [9, Sec. 5] for a discussion of these concepts and for general proofs.

THEOREM 3.3. *The transformation S is an exact endomorphism and is mixing of all degrees. Moreover, for* a.a. $\mathbf{x}$,

$$\begin{aligned}+\lim_{\nu\to\infty} \nu^{-1} \log \lambda(B_\nu(\mathbf{x})) &= +\lim_{\nu\to\infty} \nu^{-1} \log \mu B_\nu(\mathbf{x}) \\ &= (n+1) \int \log(t_1)\, d\mu(\mathbf{t}).\end{aligned}$$

The entropy of S, $h(S)$, is the negative of the last quantity above.

In exactly the same manner as that of Schweiger [7, Lemma 7.10], the entropy can be related to $A_{0,n}^{(\nu)}$.

COROLLARY 3.2. $\lim_{\nu\to\infty}((n+1)/\nu)\, A_{0,n}^{(\nu)} = h(S)$.

Nest Kuzmin's theorem is given, which states the rate of convergence of a sequence of functions to the (unknown) density of μ. For a proof see [8].

THEOREM 3.4. *Let Ψ_0 satisfy $0 < m \leqslant \Psi_0 \leqslant M$ and $|\Psi_0(\mathbf{x}) - \Psi_0(y)| \leqslant K |\mathbf{x} - \mathbf{y}|$ for $\mathbf{x}, \mathbf{y} \in I$. Then define for $\nu \geqslant 0$,*

$$\Psi_{\nu+1}(\mathbf{x}) = \sum_{(\sigma,a)=b} \Psi_\nu(f_b(\mathbf{x}))\, |J_{f_b}(\mathbf{x})|\,.$$

It then follows that

$$|\Psi_\nu(\mathbf{x}) - A \frac{d\mu}{d\lambda}(\mathbf{x})| < B\xi(\nu),$$

where $A = \int \Psi_0\, d\lambda$ and B are constants independent of $\mathbf{x}$.

Results on the mixing of S follow from Theorem 3.4.

THEOREM 3.5. *Let $E \subset I$ be a Borel set.*

(a) $|\lambda(S^{-\nu}(E)) - \mu(E)| < b\lambda(E)\, \xi(\nu)$.

(b) *For $F = B_i(\mathbf{b}^1,..., \mathbf{b}^i)$,*

$$|\mu(F \cap S^{-\nu-1-i}(E)) - \mu(F)\,\mu(E)| < K_2\mu(E)\,\mu(F)\,\xi(\nu).$$

4. CONCLUSION

It is clear from the proofs of Theorems 2.1 and 2.2 that attention need not be restricted to the specific order permutation considered here. In fact the entire Section 2 holds with *any* sequence of permutations chosen in any fashion. This observation joined with the generalized Jacobi–Perron algorithms of Bernstein [1] makes a very general class of Jacobi algorithms. We hope to study these algorithms in later work.

Since the motivation for this paper was greatest common divisors, it is also of interest to consider the set $I = \{\mathbf{x}: x_1 \leqslant x_i \text{ for } 1 \leqslant i \leqslant n\}$. That is, the appropriate permutation would be to make the smallest x_i the first component, leaving all other orders unchanged. The observation from computing suggests that this operation is faster than an ordering of all n components and should be used. The metric theory for this transformation is essentially harder but it is possible. The normalized Lebesgue measure is $\lambda(A) = nm(A)$, $C = (n+1)^{n+1}$, and $L = 1/(n-1)!$ (See [8] for a discussion of these terms.)

The implications of this work for computing greatest common divisors with the transformation $Q(\mathbf{m})$ will be treated elsewhere. Several associated results will be obtained, the relationship with expansions of rationals will be treated [10], and some numerical experiments including estimates of $d\mu/d\lambda$ and the entropy of S will be given.

References

1. L. Bernstein, "The Jacobi–Perron Algorithm—Its Theory and Application," *Lecture Notes in Mathematics* No. 207, Springer–Verlag, Berlin/New Nork, 1971.
2. C. G. J. Jacobi, Über die Auflösung der Gleichung $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + a_n x_n = fu$, *J. Reine Angew. Math.* **69** (1868), 1–28.
3. C. G. J. Jacobi, Allgemeine Theorie der Kettenbruchahnlichen Algorithmen, in welchen jede Zahl aus drei vorhergehenden gebildet wird, *J. Reine Angew. Math.* **69** (1968), 29–64.
4. D. E. Knuth, "The Art of Computer Programming," Vol. 2, "Seminumerical Algorithms," Addison–Wesley, Reading, Mass./London, 1969.
5. R. E. A. C. Paley and H. D. Ursell, Continued fraction in several dimensions, *Proc. Cambridge Philos. Soc.* **26** (1930), 127–144.
6. O. Perron, Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus. *Math. Ann.* **64** (1907), 1–76.
7. F. Schweiger, "The Metrical Theory of Jacobi–Perron Algorithm," Lecture Notes in Mathematics No. 334, Springer–Verlag, Berlin/New York, 1973.
8. F. Schweiger and M. S. Waterman, Some remarks on Kuzmin's theorem for F-expansions, *J. Number Theory*, **5** (1973), 123–131.
9. M. S. Waterman, Some ergodic properties of multidimensional F-expansions, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **16** (1970), 77-103.
10. M. S. Waterman, On F-expansions of rationals, *Aequationes Math.* **13** (1975), 263–268.

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium