

Frequencies of factors in Arnoux–Rauzy sequences

by

NATHAN WOZNY and LUCA Q. ZAMBONI (Denton, TX)

1. Introduction. Let $\omega = \omega_1\omega_2\dots$ be a sequence with values in a finite alphabet A . For each $m \geq 0$ let $L_m(\omega) = \{\omega_k\omega_{k+1}\dots\omega_{k+m-1} : k \geq 1\}$ denote the set of factors of ω of length m , and set $L(\omega) = \bigcup_{m \geq 0} L_m(\omega)$. The *complexity function* $p_\omega : \mathbb{N} \rightarrow \mathbb{N}$ assigns to each m the cardinality of $L_m(\omega)$. A fundamental result due to Morse and Hedlund states that a sequence ω is ultimately periodic if and only if for some m the complexity satisfies $p_\omega(m) \leq m$ (see [20]). Sequences of complexity $p(m) = m + 1$ are called *Sturmian sequences* or *Sturmian words*. The best known example is the so-called Fibonacci sequence

12112121121121211211212112112121121211211212112112121121121121\dots

fixed by the morphism $1 \mapsto 12$ and $2 \mapsto 1$. It is well known that all Sturmian words can be realized geometrically by an irrational rotation on the circle (see [11, 20]). More precisely, every Sturmian word is obtained by coding the symbolic orbit of a point x on the circle (of circumference one) under a rotation by an irrational angle α where the circle is partitioned into two complementary intervals, one of length α and the other of length $1 - \alpha$. And conversely, every such coding gives rise to a Sturmian word. The irrational number α is called the *slope*. In [4] V. Berthé showed that the frequencies of factors of the same length in a Sturmian word assume at most three values. Moreover it is shown in [4] that each allowable frequency, as well as the number of factors with a given frequency, can be described in terms of the continued fraction expansion of the slope α of the Sturmian word ⁽¹⁾. More precisely:

THEOREM 1.1 (V. Berthé [4]). *Let ω be a Sturmian word with slope α and $m \geq 1$. Then ω contains*

2000 *Mathematics Subject Classification*: 11J70, 37B10.

⁽¹⁾ Berthé's result extended a prior result of F. M. Dekking in [12] on the frequencies of factors of the Fibonacci word.

- $m - q_2 + 1$ factors of length m with frequency $p_2 - q_2\alpha$,
- $m - q_1 + 1$ factors of length m with frequency $q_1\alpha - p_1$,
- $(q_1 + q_2) - m - 1$ factors of length m with frequency $(q_1 - q_2)\alpha + p_2 - p_1$

where p_1/q_1 and p_2/q_2 are consecutive m -Farey numbers with $p_1/q_1 < \alpha < p_2/q_2$.

In [4] Berthé showed that Theorem 1.1 is intimately connected with the so-called *three distance theorem* in diophantine analysis [23]. Subsequently Alessandri and Berthé [1] exhibited broader connections between more general distance theorems and combinatorics of words arising from rotations on the circle. In this paper we extend Berthé's result to a class of sequences of complexity $p(m) = (k - 1)m + 1$ originally defined and studied by P. Arnoux and G. Rauzy in [3]:

DEFINITION 1.2. Let $A_k = \{1, \dots, k\}$ with $k \geq 2$. A sequence ω in the alphabet A_k is called an *Arnoux–Rauzy sequence* if it satisfies the following four conditions:

- ω is uniformly recurrent, i.e., each (finite) factor v of ω occurs in ω with bounded gaps,
- the complexity function satisfies $p(m) = (k - 1)m + 1$,
- each factor in $L_m(\omega)$ is a prefix of exactly one factor in $L_{m+1}(\omega)$ except for one which is a prefix of k factors in $L_{m+1}(\omega)$,
- each factor in $L_m(\omega)$ is a suffix of exactly one factor in $L_{m+1}(\omega)$ except for one which is a suffix of k factors in $L_{m+1}(\omega)$.

Arnoux–Rauzy sequences are a natural generalization of Sturmian words; in fact Sturmian words correspond to taking $k = 2$ in the above definition. For $k = 3$ the combinatorial conditions listed in Definition 1.2 distinguish them from other sequences of complexity $2n + 1$ such as those obtained by coding trajectories of 3-interval exchange transformations [15, 16, 17] or those of *Chacon type*, i.e., topologically isomorphic to the subshift generated by the Chacon sequence [7, 14].

Arnoux and Rauzy showed that each such sequence may be geometrically realized by an interval exchange on $2k$ intervals on the circle, and is uniquely ergodic. Perhaps the best known example on three letters is the so-called *Tribonacci sequence* defined as the fixed point of the morphism $\tau(1) = 12$, $\tau(2) = 13$ and $\tau(3) = 1$. In [21] Rauzy showed that the subshift generated by τ is isomorphic (in measure) to an exchange of three fractal domains in \mathbb{R}^2 which generate a tiling of the plane. Arnoux–Rauzy sequences have since been extensively studied from many different points of view including dynamical systems (see [2], [3], [8], [9]), number theory (see [8], [10], [18], [22], [24]) and combinatorics (see [5], [6], [8], [13], [22]).

In order to extend Theorem 1.1 to Arnoux–Rauzy sequences we use a vectorial division algorithm originally defined by the second author in [24]. This algorithm is a generalization of the continued fraction process and associates with each Arnoux–Rauzy sequence ω on A_k , an infinite array of $k \times k$ rational numbers. In the special case $k = 2$ these fractions are consecutive Farey numbers arising from the continued fraction expansion of the frequencies of the letters $\{1, 2\}$. In this paper we give a reformulation of the construction in [24] by organizing the sequence of $k \times k$ rational numbers into two sequences of matrices (P_n) and (Q_n) where P_n is the matrix of numerators and Q_n the matrix of denominators. We show that for each m , the frequencies of factors in $L_m(\omega)$ assume at most $k + 1$ values, each of which can be expressed as an integral linear combination of the base frequencies of the letters $\{1, \dots, k\}$ where the coefficients are computed from the matrix P_n . We use the matrices Q_n to determine the number of factors with a given frequency (see Theorem 4.5).

In addition to the classical theory of continued fractions, Berthé’s result relied on a careful analysis of the evolution of the Rauzy graphs of words. A similar analysis was undertaken by N. Chekhova in [8] and [9] in the context of Arnoux–Rauzy sequences. In particular in [9] the author also studies frequencies of factors of Arnoux–Rauzy sequences in connection with covering numbers. Our approach does not make use of Rauzy graphs. Instead we rely on the combinatorial approach developed by R. Risley and the second author in [22] for generating the bispecial factors.

2. Preliminaries. Let ω be an Arnoux–Rauzy sequence on the alphabet $A_k = \{1, \dots, k\}$. According to Definition 1.2 there is exactly one factor in $L_m(\omega)$ which is a prefix (resp. suffix) of more than one factor in $L_{m+1}(\omega)$. We call this factor *right special* (resp. *left special*). A factor which is both right special and left special is called *bispecial*. Clearly any suffix of a right special factor of ω is also a right special factor of ω and similarly any prefix of a left special factor is also left special. In [22] it is shown that $v = v_1 v_2 \dots v_n \in L(\omega)$ if and only if $\bar{v} \in L(\omega)$ where $\bar{v} = v_n v_{n-1} \dots v_1$ denotes the reverse word of v . Hence, $w \in L(\omega)$ is right special (resp. left special) if and only if \bar{w} is left special (resp. right special). If w is bispecial, then $w = \bar{w}$.

Let $(w_n)_{n \geq 1}$ denote the set of bispecial factors of ω ordered by increasing length, i.e., $|w_n| < |w_{n+1}|$ for all n . We set w_0 equal to the empty word. For each $n \geq 1$ there exists a unique letter $i_n \in \{1, \dots, k\}$ so that $i_n w_{n-1}$ is right special. The sequence $(i_n)_{n \geq 1}$ is called the *coding sequence* of ω (see [3], [22] and [13] in the context of episturmian sequences). Any sequence $(i_n) \in \{1, \dots, k\}^{\mathbb{N}}$ with the property that each letter occurs an infinite number of times is the coding sequence of an Arnoux–Rauzy sequence (see Theorem III.5 of [22]). In the case of the Fibonacci sequence, the coding sequence is

the periodic sequence 121212... while in the case of Tribonacci it is given by the periodic sequence 123123123... In the Sturmian case, the slope α of a Sturmian word ω completely determines the coding sequence of ω . In fact, if $\alpha = [0, a_1, a_2, \dots]$ is the continued fraction expansion of the slope, then the coding sequence of ω is given by $0^{a_1-1}1^{a_2}0^{a_3}1^{a_4} \dots$ (see [3]). More generally, given an Arnoux–Rauzy sequence ω on $\{1, \dots, k\}$, the coding sequence is determined by the multi-dimensional continued fraction expansion (defined in [3] for $k = 3$ and for all $k \geq 2$ in [24]) of the frequencies of the first $k - 1$ letters (see also the algorithms of Sections 3 and 4). Unfortunately, this multi-dimensional continued fraction expansion is only defined on a set of measure zero in \mathbb{R}^{k-1} .

If $u, v \in L(\omega)$ with u a prefix (resp. suffix) of v we write $u \vdash v$ (resp. $v \dashv u$) if for all $w \in L(\omega)$ with $|w| = |v|$, if u is a prefix (resp. suffix) of w then $w = v$. Thus for instance we have $w_{n-1}i_n \vdash w_n$ and $w_n \dashv i_nw_{n-1}$ for all $n \geq 1$.

For each $v \in L(\omega)$ we denote by $f(v)$ the frequency of v in ω , i.e.,

$$f(v) = \lim_{n \rightarrow \infty} \frac{|\omega_1\omega_2 \dots \omega_n|_v}{n}$$

where $|w|_v$ denotes the number of occurrences of v in w . Frequencies of factors exist in all Arnoux–Rauzy sequences and by minimality are all positive; in fact subshifts generated by Arnoux–Rauzy sequences are uniquely ergodic [3]. For all $u, v \in L(\omega)$ if $u \vdash v$ or if $v \dashv u$ then $f(u) = f(v)$. If v is right special, then $f(v) = \sum_{a \in A_k} f(va)$ and similarly if v is left special then $f(v) = \sum_{a \in A_k} f(av)$. Set $F_m(\omega) = \{f(v) : v \in L_m(\omega)\}$.

LEMMA 2.1. *Suppose $v \in L(\omega)$. Then $f(v) = f(\bar{v})$.*

PROOF. We proceed by induction on $|v|$. If $|v| = 1$, then $v = \bar{v}$. Assume the property holds for all factors of length m . Let $v \in L_{m+1}(\omega)$. We write $v = aub$, with $a, b \in A_k$ and $u \in L_m(\omega)$. If au is not right special, then

$$f(aub) = f(au) = f(\bar{u}a) = f(b\bar{u}a).$$

Similarly if ub is not left special, then

$$f(aub) = f(ub) = f(b\bar{u}) = f(b\bar{u}a).$$

Finally if au is right special and ub is left special, then u is bispecial or empty. Thus $u = \bar{u}$ and $a = b$ whence $aub = b\bar{u}a$. ■

LEMMA 2.2. *Suppose $v \in L_m(\omega)$ is either right special or left special. Then $f(v) \geq f(u)$ for all $u \in L_m(\omega)$. If v is bispecial, then $f(v) > f(u)$ for all $u \neq v$ in $L_m(\omega)$.*

PROOF. Suppose $v \in L_m(\omega)$ is left special and let $u \neq v$ be in $L_m(\omega)$. Let z be a first return word to u , that is $zu \in L(\omega)$ and u occurs exactly

twice in zu , once as a prefix and once as a suffix. We will see that v appears at least once in zu . Let z' be a first return word to u distinct from z . (If the only first return word to u were z , then ω would be periodic.) Since zu and $z'u$ are distinct and both end in u , they must have a common left special suffix y with $|y| > |u|$. Since v is left special and $|v| < |y|$, v is a prefix of y . Thus v appears at least once in zu . Hence between any two consecutive occurrences of u in ω there is an occurrence of v , and therefore $f(v) \geq f(u)$. A similar argument applies in case v is right special.

Now suppose $v \in L_m(\omega)$ is bispecial, and $u \in L_m(\omega)$ is distinct from v . Then $u \vdash ua$ for some $a \in A_k$, and hence $f(u) = f(ua)$. Since v is right special, we have $f(v) = f(v1) + \dots + f(vk)$. Fix $b \in A_k$ so that vb is left special. Then $f(v) = f(v1) + \dots + f(vk) > f(vb) \geq f(ua) = f(u)$. ■

LEMMA 2.3. *Let $v \in L(\omega)$ and let $a, b \in A_k$. Then if va and vb are in $L(\omega)$ and $f(va) = f(vb)$ then $a = b$.*

PROOF. Suppose va and vb are in $L(\omega)$ and $f(va) = f(vb)$. Then v is either right special or the empty factor. We suppose to the contrary that $a \neq b$ and consider two cases:

CASE 1: *va or vb is left special.* Without loss of generality, we may assume that va is left special. Thus v is either bispecial or empty. Thus $v = \bar{v}$. Since all frequencies are positive, we have $f(va) > f(bva)$. Since $avb \vdash vb$, we have $f(vb) = f(avb)$, and by Lemma 2.1, $f(avb) = f(bva)$. But this implies that $f(va) > f(vb)$, a contradiction.

CASE 2: *Neither va nor vb is left special.* Let $u \in L(\omega)$ be such that $uva \vdash va$ and $uvb \vdash vb$ and either uva or uvb is left special. Since $uva \vdash va$ and $uvb \vdash vb$, we have $f(uva) = f(va) = f(vb) = f(uvb)$. Thus we are back in Case 1 and obtain a contradiction. ■

3. An algorithm for computing the frequencies of factors. Let ω be an Arnoux–Rauzy sequence on the alphabet $A_k = \{1, \dots, k\}$ with coding sequence (i_n) . Let $(w_n)_{n \geq 1}$ be the sequence of bispecial factors of ω arranged in increasing length, i.e., $|w_n| < |w_{n+1}|$ for all n . For each $1 \leq p \leq k + 1$ we associate with ω a sequence of the form $(f_{(m,p)}, s_{(m,p)})_{m \geq 1}$ where $f_{(m,p)} \in \mathbb{R}$ and $s_{(m,p)} \in \mathbb{Z}$. The $f_{(m,p)}$ and $s_{(m,p)}$ are defined recursively as follows: For $m = 1$ and $1 \leq p \leq k$ we set $f_{(1,p)} = f(p)$ and $s_{(1,p)} = 1$ while for $p = k + 1$ we set

$$f_{(1,k+1)} = f_{(1,i_1)} - \sum_{1 \leq r \leq k, r \neq i_1} f_{(1,r)}$$

and $s_{(1,k+1)} = 0$. For $m \geq 1$ we consider two cases:

CASE 1: *$m \neq |w_n|$ for all n .* In this case let n be the greatest integer such that $m > |w_n|$. Then for all $1 \leq p \leq k + 1$ we set

$$(f_{(m+1,p)}, s_{(m+1,p)}) = \begin{cases} (f_{(m,p)}, s_{(m,p)} + 1) & \text{if } p \neq i_{n+1}, \\ (f_{(m,p)}, s_{(m,p)} - 1) & \text{if } p = i_{n+1}. \end{cases}$$

CASE 2 (critical case): $m = |w_n|$ for some n . For $1 \leq p \leq k$ set

$$(f_{(m+1,p)}, s_{(m+1,p)}) = \begin{cases} (f_{(m,p)}, s_{(m,p)} + 1) & \text{if } p \neq i_n, \\ (f_{(m,k+1)}, s_{(m,k+1)} + 1) & \text{if } p = i_n \end{cases}$$

and for $p = k + 1$ set

$$f_{(m+1,k+1)} = f_{(m+1,i_{n+1})} - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f_{(m+1,r)}$$

and $s_{(m+1,k+1)} = 0$.

Note that for all $1 \leq p \leq k + 1$ we have

$$(3.1) \quad f_{(|w_n|+1,p)} = f_{(|w_n|+2,p)} = f_{(|w_n|+3,p)} = \dots = f_{(|w_{n+1}|,p)}.$$

We will prove (see Theorem 3.2) that for each $w \in L_m(\omega)$ the frequency $f(w)$ is equal to $f_{(m,p)}$ for some $1 \leq p \leq k + 1$. We will also show that for each $1 \leq p \leq k + 1$ there are exactly $s_{(m,p)}$ factors of length m with frequency $f_{(m,p)}$. In particular it will follow that $f_{(m,p)}$ is a positive real number and $s_{(m,p)}$ a nonnegative integer for each m and p . We begin by illustrating the above algorithm with an example:

EXAMPLE 3.1. Let ω be an Arnoux–Rauzy sequence on the alphabet $A_3 = \{1, 2, 3\}$ with coding sequence (i_n) beginning in 1213... Set $\alpha = f(1)$ and $\beta = f(2)$ so that $f(3) = 1 - (\alpha + \beta)$. The steps defined as critical in the recursive algorithm are marked with a C on the left. Note that in the passage from a critical step to the next, one $f_{(m,p)}$ disappears and a new one is introduced.

	m	$(f_{(m,1)}, s_{(m,1)})$	$(f_{(m,2)}, s_{(m,2)})$	$(f_{(m,3)}, s_{(m,3)})$	$(f_{(m,4)}, s_{(m,4)})$
C	1	$(\alpha, 1)$	$(\beta, 1)$	$(1 - (\alpha + \beta), 1)$	$(2\alpha - 1, 0)$
	2	$(2\alpha - 1, 1)$	$(\beta, 2)$	$(1 - (\alpha + \beta), 2)$	$(2\beta - \alpha, 0)$
C	3	$(2\alpha - 1, 2)$	$(\beta, 1)$	$(1 - (\alpha + \beta), 3)$	$(2\beta - \alpha, 1)$
(3.2)	4	$(2\alpha - 1, 3)$	$(2\beta - \alpha, 2)$	$(1 - (\alpha + \beta), 4)$	$(4\alpha - \beta - 2, 0)$
	5	$(2\alpha - 1, 2)$	$(2\beta - \alpha, 3)$	$(1 - (\alpha + \beta), 5)$	$(4\alpha - \beta - 2, 1)$
C	6	$(2\alpha - 1, 1)$	$(2\beta - \alpha, 4)$	$(1 - (\alpha + \beta), 6)$	$(4\alpha - \beta - 2, 2)$
	7	$(4\alpha - \beta - 2, 3)$	$(2\beta - \alpha, 5)$	$(1 - (\alpha + \beta), 7)$	$(3 - 4\alpha - 2\beta, 0)$
	8	$(4\alpha - \beta - 2, 4)$	$(2\beta - \alpha, 6)$	$(1 - (\alpha + \beta), 6)$	$(3 - 4\alpha - 2\beta, 1)$
	\vdots	\vdots	\vdots	\vdots	\vdots

It will follow from Theorem 3.2 below that of the eleven factors of length five, two have frequency $2\alpha - 1$, three have frequency $2\beta - \alpha$, five have frequency $1 - (\alpha + \beta)$ and one has frequency $4\alpha - \beta - 2$.

THEOREM 3.2. *Suppose ω is an Arnoux–Rauzy sequence on the alphabet $A_k = \{1, \dots, k\}$ with coding sequence (i_n) . Set $F_m = F_m(\omega)$. Define for each $1 \leq p \leq k + 1$ the sequence $(f_{(m,p)}, s_{(m,p)})_{m \geq 1}$ as above and set $F'_m = \{f_{(m,p)} : 1 \leq p \leq k + 1\}$. Then for each m we have $F_m \subseteq F'_m$ and for all $1 \leq p \leq k + 1$ the sequence ω contains exactly $s_{(m,p)}$ factors of length m with frequency $f_{(m,p)}$. In particular $\text{Card}(F_m) \leq k + 1$ for each m .*

As an immediate consequence of Theorem 3.2 we have ⁽²⁾:

COROLLARY 3.3. *For each $n \geq 0$ we have*

$$|w_n| = \frac{\sum_{j=1}^k s_{(|w_n|+1,j)} - k}{k - 1}.$$

PROOF. Since $s_{(|w_n|+1,k+1)} = 0$, the sum of $s_{(|w_n|+1,j)}$ over all $1 \leq j \leq k$ is equal to the cardinality of $L_{|w_n|+1}(\omega)$, which equals $(k - 1)(|w_n| + 1) + 1$. ■

To prove Theorem 3.2 we will need the following lemma:

LEMMA 3.4. *Let $p \in A_k$ and w_n denote the n th bispecial factor of ω . Then $f(pw_n) = f_{(|w_n|+1,p)}$. Hence, $f_{(|w_n|+1,i_{n+1})}$ is the greatest frequency in $F_{|w_n|+1}(\omega)$.*

PROOF. The second claim of the lemma follows immediately from the first using Lemma 2.2 since $i_{n+1}w_n$ is the right special factor of length $|w_n| + 1$. We prove the first claim by induction on n . For $n = 1$ we have $i_1 = w_1$. If $p \neq i_1$, then $p \vdash pi_1$, so

$$f(pw_1) = f(pi_1) = f(p) = f_{(1,p)} = f_{(|w_1|+1,p)}.$$

On the other hand, for $p = i_1$ we have

$$\begin{aligned} f(i_1w_1) &= f(w_1) - \sum_{1 \leq r \leq k, r \neq i_1} f(rw_1) = f_{(1,i_1)} - \sum_{1 \leq r \leq k, r \neq i_1} f_{(1,r)} \\ &= f_{(1,k+1)} = f_{(|w_1|+1,i_1)}. \end{aligned}$$

This proves the case $n = 1$. We now suppose $f(pw_n) = f_{(|w_n|+1,p)}$ for all p and show that $f(pw_{n+1}) = f_{(|w_{n+1}|+1,p)}$. We consider two cases:

CASE 1: $p \neq i_{n+1}$. Then $f_{(|w_{n+1}|+1,p)} = f_{(|w_{n+1}|,p)} = f_{(|w_n|+1,p)}$. So we must show that $f(pw_{n+1}) = f_{(|w_n|+1,p)}$. Since $p \neq i_{n+1}$, we have $pw_n \vdash pw_{n+1}$. Thus $f(pw_{n+1}) = f(pw_n) = f_{(|w_n|+1,p)}$ as required.

⁽²⁾ For $k = 2$ Corollary 3.3 should be compared with Proposition 3 of [4], while for $k = 3$ Corollary 3.3 should be compared with the definition of 3-PER in [6] and with the evolution of the Rauzy graphs of words developed in [9]. See also Corollary 4.2 in the next section of this paper.

CASE 2: $p = i_{n+1}$. In this case pw_n is right special and we have

$$\begin{aligned}
 f_{(|w_{n+1}|+1, i_{n+1})} &= f_{(|w_{n+1}|, k+1)} = f_{(|w_n|+1, k+1)} \\
 &= f_{(|w_n|+1, i_{n+1})} - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f_{(|w_n|+1, r)} \\
 &= f(i_{n+1}w_n) - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f(rw_n) \\
 &= f(w_n i_{n+1}) - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f(rw_n i_{n+1}) \\
 &= f(i_{n+1}w_n i_{n+1}) = f(i_{n+1}w_{n+1}).
 \end{aligned}$$

The last equality follows since $w_n i_{n+1} \vdash w_{n+1}$. ■

Proof of Theorem 3.2. We proceed by induction on m . The result is clear for $m = 1$; in fact it is readily verified that for distinct letters a and b , if the first occurrence of a in the coding sequence (i_n) precedes the first occurrence of b in (i_n) , then $f(a) > f(b)$ (see the proof of Theorem III.5 in [22]). In particular the frequencies of the letters are all distinct. We suppose the theorem holds for all positive integers less than or equal to m and we will show it holds for $m + 1$.

CASE 1: $m \neq |w_n|$ for each $n \geq 1$. Let n be the greatest integer so that $m > |w_n|$. Since $m < |w_{n+1}|$ it follows from (3.1) that $f_{(m+1, p)} = f_{(|w_n|+1, p)}$ for all $1 \leq p \leq k + 1$. Let $v_0 \in L_m(\omega)$ be right special. Then $v_0 \vdash i_{n+1}w_n$. Thus

$$f(v_0) = f(i_{n+1}w_n) = f_{(|w_n|+1, i_{n+1})} = f_{(m+1, i_{n+1})}.$$

If $p \neq i_{n+1}$ then

$$f(v_0 p) = f(i_{n+1}w_n p) = f(w_n p) = f(pw_n) = f_{(|w_n|+1, p)} = f_{(m+1, p)} \in F'_{m+1}.$$

On the other hand,

$$\begin{aligned}
 f(v_0 i_{n+1}) &= f(i_{n+1}w_n i_{n+1}) = f(i_{n+1}w_n) - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f(i_{n+1}w_n r) \\
 &= f(i_{n+1}w_n) - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f(w_n r) \\
 &= f_{(|w_n|+1, i_{n+1})} - \sum_{1 \leq r \leq k, r \neq i_{n+1}} f_{(|w_n|+1, r)} \\
 &= f_{(|w_n|+1, k+1)} = f_{(m+1, k+1)} \in F'_{m+1}.
 \end{aligned}$$

Thus for all $p \in A_k$ we have $f(v_0 p) \in F'_{m+1} \setminus \{f_{(m+1, i_{n+1})}\}$.

On the other hand, if $v \in L_m(\omega)$ is not right special then $v \vdash va$ for some $a \in A_k$. By induction hypothesis $f(v) = f_{(m, p)}$ for some $1 \leq p \leq k + 1$. Hence $f(va) = f(v) = f_{(m, p)} = f_{(m+1, p)} \in F'_{m+1}$. Having exhausted all

factors of length $m + 1$ we have shown that $F_{m+1} \subseteq F'_{m+1}$ and that the number of factors of ω of length $m + 1$ having frequency $f_{(m,p)} = f_{(m+1,p)}$ is $s_{(m,p)} + 1$ if $p \neq i_{n+1}$ and $s_{(m,p)} - 1$ if $p = i_{n+1}$.

CASE 2: $m = |w_n|$ for some $n \geq 1$. In this case we show that $F_{m+1} \subseteq F'_m \setminus \{f_{(m,i_n)}\} \subset F'_{m+1}$. Let $w \neq w_n$ be a factor of length m . Then $w \vdash wa$ for some $a \in A_k$. Hence $f(wa) = f(w)$, which, by induction hypothesis, is in F'_m . But

$$f(w) < f(w_n) = f(i_n w_{n-1}) = f_{(|w_{n-1}|+1, i_n)} = f_{(|w_n|, i_n)}.$$

Thus $f(wa) \in F'_m \setminus \{f_{(m,i_n)}\}$. Also, for each $1 \leq p \leq k$ we have $f(w_n p) = f(i_n w_{n-1} p)$. If $p \neq i_n$ then

$$f(w_n p) = f(i_n w_{n-1} p) = f(w_{n-1} p) = f_{(|w_{n-1}|+1, p)} = f_{(|w_n|, p)}.$$

On the other hand,

$$\begin{aligned} f(w_n i_n) &= f(i_n w_{n-1} i_n) = f(i_n w_{n-1}) - \sum_{1 \leq r \leq k, r \neq i_n} f(i_n w_{n-1} r) \\ &= f(i_n w_{n-1}) - \sum_{1 \leq r \leq k, r \neq i_n} f(w_{n-1} r) \\ &= f_{(|w_{n-1}|+1, i_n)} - \sum_{1 \leq r \leq k, r \neq i_n} f_{(|w_{n-1}|+1, r)} \\ &= f_{(|w_{n-1}|+1, k+1)} = f_{(|w_n|, k+1)}. \end{aligned}$$

Hence for each $p \in A_k$ we have $f(w_n p) \in F'_m \setminus \{f_{(m,i_n)}\}$. Having exhausted all factors of length $m+1$ we have shown that $F_{m+1} \subseteq F'_m \setminus \{f_{(m,i_n)}\} \subset F'_{m+1}$ and that for all $1 \leq p \leq k + 1$ different from i_n the number of factors of ω of length $m + 1$ with frequency $f_{(m,p)}$ is $s_{(m,p)} + 1$ as required. ■

Note that $F_{m+1}(\omega) \neq F_m(\omega)$ only when $m = |w_n|$ for some n . Thus to recover the allowable frequencies of factors it suffices to consider only the sets F_m where $m = |w_n| + 1$, that is, to restrict the algorithm to those values of m immediately following the critical case. We will refer to this restriction as the *multiplicative form of the algorithm*. Applied to Example 3.1 where the coding sequence begins in 12133... it gives:

n	m	$(f_{(m,1)}, s_{(m,1)})$	$(f_{(m,2)}, s_{(m,2)})$	$(f_{(m,3)}, s_{(m,3)})$	$(f_{(m,4)}, s_{(m,4)})$
	0	$(\alpha, 1)$	$(\beta, 1)$	$(1 - (\alpha + \beta), 1)$	$(2\alpha - 1, 0)$
	1	$(2\alpha - 1, 1)$	$(\beta, 2)$	$(1 - (\alpha + \beta), 2)$	$(2\beta - \alpha, 0)$
(3.3)	2	$(2\alpha - 1, 3)$	$(2\beta - \alpha, 2)$	$(1 - (\alpha + \beta), 4)$	$(4\alpha - \beta - 2, 0)$
	3	$(4\alpha - \beta - 2, 3)$	$(2\beta - \alpha, 5)$	$(1 - (\alpha + \beta), 7)$	$(3 - 4\alpha - 2\beta, 0)$
	4	$(4\alpha - \beta - 2, 10)$	$(2\beta - \alpha, 12)$	$(3 - 4\alpha - 2\beta, 7)$	$(5 - 7\alpha - 3\beta, 0)$
	\vdots	\vdots	\vdots	\vdots	\vdots

We may represent the multiplicative version of the algorithm in matrix form as follows: Define a sequence of $k \times k$ matrices (B_n) recursively by setting

$$(3.4) \quad B_0 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

and for all $n \geq 1$,

$$(3.5) \quad B_n = B_{n-1} - \begin{pmatrix} \vec{0} \\ \vdots \\ \sum_{r \neq i_n} B_{n-1}(r) \\ \vdots \\ \vec{0} \end{pmatrix}_{i_n \text{ th row}} \quad k \times k$$

where $B_{n-1}(r)$ denotes the r th row vector of the matrix B_{n-1} , and the nonzero row of the second matrix on the right hand side of (3.5) is the i_n th row. It is evident that

$$(3.6) \quad \begin{bmatrix} f_{(|w_n|+1,1)} \\ f_{(|w_n|+1,2)} \\ \vdots \\ f_{(|w_n|+1,k)} \end{bmatrix} = B_n \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(k) \end{bmatrix}.$$

In the matrix formulation, we have omitted the frequency $f_{(|w_n|+1,k+1)}$. But this frequency is determined in the next step when $m = |w_{n+1}| + 1$ since $f_{(|w_n|+1,k+1)} = f_{(|w_{n+1}|+1,i_{n+1})}$. In fact the multiplicative formulation contains the same information as the preceding additive version. To determine the set of frequencies of factors of length m , where $m \neq |w_n| + 1$ for all n , we choose the smallest n such that $m < |w_n| + 1$. Then

$$F_m(\omega) = \left\{ f_{(|w_n|+1,1)}, f_{(|w_n|+1,2)}, \dots, f_{(|w_n|+1,k)}, \sum_{p=1}^k f_{(|w_n|+1,p)} \right\}.$$

In the multiplicative formulation of the algorithm, the $s_{(|w_n|+1,p)}$ satisfy the following recursive relation:

LEMMA 3.5. *For each $n \geq 0$ and $1 \leq p \leq k$ we have*

$$s_{(|w_{n+1}|+1,p)} = \begin{cases} s_{(|w_n|+1,p)} + s_{(|w_n|+1,i_{n+1})} & \text{if } p \neq i_{n+1}, \\ s_{(|w_n|+1,i_{n+1})} & \text{if } p = i_{n+1}. \end{cases}$$

Proof. If $p \neq i_{n+1}$ then

$$\begin{aligned} s_{(|w_{n+1}|+1,p)} &= s_{(|w_{n+1}|,p)} + 1 = s_{(|w_n|+1,p)} + |w_{n+1}| - |w_n| \\ &= s_{(|w_n|+1,p)} + s_{(|w_n|+1,i_{n+1})} \end{aligned}$$

while if $p = i_{n+1}$ then

$$s_{(|w_{n+1}|+1,p)} = s_{(|w_{n+1}|,k+1)} + 1 = |w_{n+1}| - |w_n| = s_{(|w_n|+1,i_{n+1})}. \blacksquare$$

4. Simultaneous rational approximations of the frequencies.

Equation (3.6) may be used to obtain simultaneous rational approximations of the frequencies $f(1), \dots, f(k)$. First we rewrite (3.6) in the form ⁽³⁾

$$(4.1) \quad \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(k) \end{bmatrix} = B_n^{-1} \begin{bmatrix} f_{(|w_n|+1,1)} \\ f_{(|w_n|+1,2)} \\ \vdots \\ f_{(|w_n|+1,k)} \end{bmatrix}.$$

Then for each $1 \leq p \leq k$ we set the $k - 1$ frequencies $\{f_{(|w_n|+1,j)} : j \neq p\}$ in (4.1) equal to zero and $f_{(|w_n|+1,p)}$ equal to $1/s_{(|w_n|+1,p)}$. This gives a rational approximation of each frequency $f(1), \dots, f(k)$ by unreduced fractions having a common denominator equal to $s_{(|w_n|+1,p)}$. This method generates an infinite array of $k \times k$ (unreduced) rational numbers. In the special case $k = 2$, for each $i \in \{1, 2\}$ the two rational numbers $p_1/q_1, p_2/q_2$ approximating $f(i)$ are consecutive m -Farey numbers where

$$m = q_1 + q_2 - 1 = s_{(|w_n|+1,1)} + s_{(|w_n|+1,2)} - 1$$

and $p_1/q_1 < f(i) < p_2/q_2$ (see [4]). In general, the $k \times k$ array of fractions we obtain from (4.1) coincides with the fraction generating algorithm associated with Arnoux–Rauzy sequences in [24].

In this section we will be concerned with the reverse problem of expressing the quantities $f_{(m,p)}$ and $s_{(m,p)}$ in terms of the base frequencies $f(1), \dots, f(k)$ and the generalized Farey numbers defined in [24]. In fact this was the point of view taken by Berthé in [4] where the frequencies of factors in a Sturmian word, as well as the number of factors with a given frequency, were each expressed in terms of the Farey numbers arising from the continued fraction expansion of the slope α of the Sturmian word.

We begin with a reformulation of the construction in [24] by organizing the sequence of $k \times k$ rational numbers into two sequences of matrices (P_n) and (Q_n) where P_n is the matrix of numerators and Q_n the matrix of denominators. We then give a formula relating the matrices B_n and P_n . Via (3.6) we can then express each of the frequencies $f_{(|w_n|+1,p)}$ as integer linear combinations of the base frequencies $f(1), \dots, f(k)$ where the coefficients are determined from the entries of P_n . On the other hand, the matrices Q_n are used to determine the quantities $s_{(m,p)}$.

⁽³⁾ In a slightly different formulation, a similar expression was obtained by N. Chekhova in [9] for $k = 3$.

Let ω be an Arnoux–Rauzy sequence on A_k with coding sequence (i_n) . Following [24], for all $1 \leq i, j \leq k$ and $n \geq 0$ we define integers $p_{ij}^{(n)}$ and $q_{ij}^{(n)}$ recursively using the coding sequence (i_n) . Set

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

and $q_{ij}^{(0)} = 1$. For $n \geq 1$, set

$$p_{ij}^{(n)} = \begin{cases} p_{ij}^{(n-1)} + p_{i_n j}^{(n-1)} & \text{if } i \neq i_n, \\ p_{i_n j}^{(n-1)} & \text{otherwise} \end{cases}$$

and

$$q_{ij}^{(n)} = \begin{cases} q_{ij}^{(n-1)} + q_{i_n j}^{(n-1)} & \text{if } i \neq i_n, \\ q_{i_n j}^{(n-1)} & \text{otherwise.} \end{cases}$$

We represent this algorithm by fixing a value of j and writing the unreduced fraction $p_{ij}^{(n)}/q_{ij}^{(n)}$ in the i th row and n th column of an infinite array. For example if $j = 1$, and the coding sequence (i_n) begins in 121332... then we have

$$(4.2) \quad \begin{array}{cccccccc} \frac{1}{1} & \rightarrow & \frac{1}{1} & & \frac{2}{3} & \rightarrow & \frac{2}{3} & & \frac{6}{10} & & \frac{10}{17} & & \frac{21}{36} & \dots \\ \frac{0}{1} & & \frac{1}{2} & \rightarrow & \frac{1}{2} & & \frac{3}{5} & & \frac{7}{12} & & \frac{11}{19} & \rightarrow & \frac{11}{19} & \dots \\ \frac{0}{1} & & \frac{1}{2} & & \frac{2}{4} & & \frac{4}{7} & \rightarrow & \frac{4}{7} & \rightarrow & \frac{4}{7} & & \frac{15}{26} & \dots \end{array}$$

The positions of the arrows correspond to the values of i_n . If ω is a Sturmian sequence, then $p_{1j}^{(n)}/q_{1j}^{(n)}$ and $p_{2j}^{(n)}/q_{2j}^{(n)}$ are consecutive m -Farey numbers where $m = q_{1j}^{(n)} + q_{2j}^{(n)} - 1$ and the subsequence of fractions followed by an arrow are precisely the convergents of the continued fraction expansion of $f(j)$.

For each $n \geq 0$ define matrices P_n and Q_n by $P_n = (p_{ij}^{(n)})$ and $Q_n = (q_{ij}^{(n)})$. Thus

$$(4.3) \quad P_0 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

and for $n \geq 1$,

$$(4.4) \quad P_n = P_{n-1} + \begin{pmatrix} P_{n-1}(i_n) \\ \vdots \\ \vec{0} \\ \vdots \\ P_{n-1}(i_n) \end{pmatrix}_{k \times k} \text{ } i_n \text{th row}$$

while

$$(4.5) \quad Q_0 = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

and for $n \geq 1$,

$$(4.6) \quad Q_n = Q_{n-1} + \begin{pmatrix} Q_{n-1}(i_n) \\ \vdots \\ \vec{0} \\ \vdots \\ Q_{n-1}(i_n) \end{pmatrix}_{k \times k} \text{ } i_n \text{th row}$$

We note that all entries in a given row in Q_n are equal. We denote by $q_n(i)$ the value of each entry in row i of Q_n . It is also evident by induction that the sum of the entries in each row i of P_n is equal to $q_n(i)$.

LEMMA 4.1. *For each $n \geq 0$ and $1 \leq i \leq k$ we have $q_n(i) = s_{(|w_n|+1,i)}$.*

PROOF. For $n = 0$ we have $q_0(i) = 1 = s_{(1,i)}$ for all $1 \leq i \leq k$. For $n \geq 0$ it follows from (4.6) that

$$q_{n+1}(i) = \begin{cases} q_n(i) + q_n(i_{n+1}) & \text{if } i \neq i_{n+1}, \\ q_n(i_{n+1}) & \text{if } i = i_{n+1} \end{cases}$$

which is the same recursive relation in Lemma 3.5 satisfied by the $s_{(|w_n|+1,i)}$. ■

Combining Lemma 4.1 and Corollary 3.3 gives:

COROLLARY 4.2. *For each $n \geq 0$ we have*

$$|w_n| = \frac{\sum_{j=1}^k q_n(j) - k}{k - 1}.$$

LEMMA 4.3. *For each $n \geq 0$ we have $(B_n)^{-1} = (P_n)^t$.*

PROOF. We prove this by induction on n . Let I denote the $k \times k$ identity matrix. For $n = 0$ we have $B_0 = P_0 = I$, so $(B_0)(P_0)^t = I$. Now suppose that $(B_n)(P_n)^t = I$. Then the dot product $B_n(r) \cdot P_n(r) = 1$ and $B_n(r) \cdot P_n(s) = 0$

if $r \neq s$. We will show that $(B_{n+1})(P_{n+1})^t = I$. By equations (3.5) and (4.4), $(B_{n+1})(P_{n+1})^t$ is equal to

$$\begin{aligned}
 (B_n)(P_n)^t + B_n \begin{pmatrix} P_n(i_{n+1}) \\ \vdots \\ \vec{0} \\ \vdots \\ P_n(i_{n+1}) \end{pmatrix}^t &= \begin{pmatrix} \vec{0} \\ \vdots \\ \sum_{r \neq i_{n+1}} B_n(r) \\ \vdots \\ \vec{0} \end{pmatrix} P_n^t \\
 &= \begin{pmatrix} \vec{0} \\ \vdots \\ \sum_{r \neq i_{n+1}} B_n(r) \\ \vdots \\ \vec{0} \end{pmatrix} \begin{pmatrix} P_n(i_{n+1}) \\ \vdots \\ \vec{0} \\ \vdots \\ P_n(i_{n+1}) \end{pmatrix}^t.
 \end{aligned}$$

The first product is I by induction hypothesis. We claim that the second and third products are equal and the fourth product is the zero matrix. In fact

$$\begin{aligned}
 B_n \begin{pmatrix} P_n(i_{n+1}) \\ \vdots \\ \vec{0} \\ \vdots \\ P_n(i_{n+1}) \end{pmatrix}^t &= \begin{pmatrix} B_n(1) \cdot P_n(i_{n+1}) & \dots & 0 & \dots & B_n(1) \cdot P_n(i_{n+1}) \\ B_n(2) \cdot P_n(i_{n+1}) & \dots & 0 & \dots & B_n(2) \cdot P_n(i_{n+1}) \\ \vdots & & \vdots & & \vdots \\ B_n(k) \cdot P_n(i_{n+1}) & \dots & 0 & \dots & B_n(k) \cdot P_n(i_{n+1}) \end{pmatrix} \\
 &= \begin{pmatrix} 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & 1 & \dots & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}
 \end{aligned}$$

where the nonzero row is the i_{n+1} th row and the column of all 0's is the i_{n+1} th column. This is equal to the third matrix product:

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(1) & \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(2) & \dots & \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(k) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Finally the fourth matrix product is given by

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(i_{n+1}) & \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(i_{n+1}) & \dots & \sum_{r \neq i_{n+1}} B_n(r) \cdot P_n(i_{n+1}) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

which is the zero matrix as required. ■

COROLLARY 4.4. *For each n we have $B_n = \text{cof}(P_n)$ and $P_n = \text{cof}(B_n)$ where cof denotes the cofactor matrix, i.e., the transpose of the adjoint.*

PROOF. This follows immediately from Lemma 4.3 and the fact that the determinant of B_n and P_n are each equal to one. ■

We summarize our results as follows:

THEOREM 4.5. *Suppose ω is an Arnoux–Rauzy sequence on A_k . Fix $m \geq 1$ and let n be the smallest positive integer such that $m \leq |w_n| + 1$. Let B_n be defined recursively according to the coding sequence of ω as in (3.5) and set $P_n = \text{cof}(B_n)$. Let Q_n be the $k \times k$ matrix whose i th row is of the form (a, a, \dots, a) where a is the sum of the entries of the i th row of P_n . Then*

$$F_m(\omega) \subseteq \left\{ f_{(|w_n|+1,1)}, f_{(|w_n|+1,2)}, \dots, f_{(|w_n|+1,k)}, \sum_{i=1}^k f_{(|w_n|+1,i)} \right\},$$

where

$$(4.7) \quad \begin{bmatrix} f_{(|w_n|+1,1)} \\ f_{(|w_n|+1,2)} \\ \vdots \\ f_{(|w_n|+1,k)} \end{bmatrix} = \text{cof}(P_n) \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(k) \end{bmatrix}.$$

Moreover for each $1 \leq i \leq k$, $L_m(\omega)$ contains $s'_{(m,i)}$ factors with frequency $f_{(|w_n|+1,i)}$ where

$$s'_{(m,i)} = q_n(i) + m - \left(\frac{[\sum_{j=1}^k q_n(j)] - k}{k - 1} + 1 \right)$$

and $s'_{(m,k+1)}$ factors with frequency $\sum_{i=1}^k f_{(|w_n|+1,i)}$ where

$$s'_{(m,k+1)} = (k - 1)m + 1 - \sum_{i=1}^k s'_{(m,i)}.$$

EXAMPLE 4.6 (the Sturmian case). We show how Theorem 4.5 may be used to recover the main result of Berthé in [4] on frequencies of factors

in Sturmian words. Let ω be a Sturmian word on $A_2 = \{1, 2\}$, and set α equal to the frequency of 2 in ω . For each $m \geq 1$ we compute the set F_m of frequencies of factors of length m and the number of factors $s_{(m,p)}$ with frequency $f_{(m,p)}$. Let n be the smallest integer such that $m \leq |w_n| + 1$. The matrices P_n and Q_n are of the form

$$P_n = \begin{pmatrix} q_1 - p_1 & p_1 \\ q_2 - p_2 & p_2 \end{pmatrix}, \quad Q_n = \begin{pmatrix} q_1 & q_1 \\ q_2 & q_2 \end{pmatrix}$$

where

$$\frac{p_1}{q_1} \leq \alpha \leq \frac{p_2}{q_2} \quad \text{and} \quad \frac{q_1 - p_1}{q_1} \leq 1 - \alpha \leq \frac{q_2 - p_2}{q_2}.$$

Moreover, p_1/q_1 and p_2/q_2 are consecutive q -Farey numbers where $q = q_1 + q_2 - 1$, as are $(q_1 - p_1)/q_1$ and $(q_2 - p_2)/q_2$. Then

$$B_n = \text{cof } P_n = \begin{pmatrix} p_2 & p_2 - q_2 \\ -p_1 & q_1 - p_1 \end{pmatrix}$$

and hence

$$\begin{pmatrix} f_{(|w_n|+1,1)} \\ f_{(|w_n|+1,2)} \end{pmatrix} = \begin{pmatrix} p_2 & p_2 - q_2 \\ -p_1 & q_1 - p_1 \end{pmatrix} \begin{pmatrix} 1 - \alpha \\ \alpha \end{pmatrix} = \begin{pmatrix} p_2 - q_2\alpha \\ q_1\alpha - p_1 \end{pmatrix}.$$

Thus by Theorem 4.5 we obtain:

$$F_m \subseteq \{p_2 - q_2\alpha, q_1\alpha - p_1, (q_1 - q_2)\alpha + p_2 - p_1\}.$$

Also by Theorem 4.5 we deduce that $s'_{(m,1)} = q_1 + m - (q_1 + q_2 - 1) = m - q_2 + 1$ and $s'_{(m,2)} = q_2 + m - (q_1 + q_2 - 1) = m - q_1 + 1$ and $s'_{(m,3)} = m + 1 - s'_{(m,1)} - s'_{(m,2)} = (q_1 + q_2) - m - 1$.

In other words, $L_m(\omega)$ has

- $m - q_2 + 1$ factors with frequency $p_2 - q_2\alpha$,
- $m - q_1 + 1$ factors with frequency $q_1\alpha - p_1$,
- $(q_1 + q_2) - m - 1$ factors with frequency $(q_1 - q_2)\alpha + p_2 - p_1$.

5. Concluding remarks. It would be interesting to investigate the quality of simultaneous approximation of the frequencies obtained from the algorithms of Sections 3 and 4. Partial results in this direction were obtained by Chekhova, Hubert, and Messaoudi in [10] in the case of Tribonacci. They showed that there is natural quadratic form associated with the incidence matrix of the Tribonacci substitution with respect to which the approximation is optimal.

In the Sturmian case, Berthé [4] discovered a connection between Theorem 1.1 and the three distance theorem in diophantine analysis [23]. Subsequently Alessandri and Berthé [1] investigated connections between generalizations of the three distance theorem and combinatorics on words for sequences defined by codings of irrational rotations on the unit circle. It would

be interesting to explore possible connections between Arnoux–Rauzy sequences and distance type theorems arising from the geometric realizations given in [2] and [3]. It was conjectured that all Arnoux–Rauzy sequences on a 3-letter alphabet code the trajectory of a point on the 2-torus under a rotation. A counterexample to this conjecture was recently obtained by Cassaigne, Ferenczi, and the second author [5] by constructing an Arnoux–Rauzy sequence ω on $\{0, 1, 2\}$ which is unbalanced in the following sense: For each N there exist two factors of ω of equal length with one having at least N more occurrences of the symbol 0 than the other. In itself this sequence also provides a counterexample to a balancing property conjectured by Droubay, Justin and Pirillo [13] for episturmian sequences.

Finally, it would be nice to find a formula for the number of blocks of length n which occur as a factor of some Arnoux–Rauzy sequence. This was resolved by Mignosi in the Sturmian case [19].

Acknowledgements. The first author was supported in part by the Texas Academy of Mathematics and Science. The second author was supported in part by NSF grant INT-9726708 and from a grant from the Texas Advanced Research Program. We wish to thank the referee of this paper for taking the time to carefully read the manuscript and for providing us with many helpful suggestions which have improved the overall presentation of the paper.

References

- [1] P. Alessandri and V. Berthé, *Three distance theorems and combinatorics on words*, Enseign. Math. 44 (1998), 103–132.
- [2] P. Arnoux and S. Ito, *Pisot substitutions and Rauzy fractals*, preprint, 1999.
- [3] P. Arnoux et G. Rauzy, *Représentation géométrique de suites de complexité $2n+1$* , Bull. Soc. Math. France 119 (1991), 199–215.
- [4] V. Berthé, *Fréquences des facteurs des suites sturmiennes*, Theoret. Comput. Sci. 165 (1996), 295–309.
- [5] J. Cassaigne, S. Ferenczi and L. Q. Zamponi, *Imbalances in Arnoux–Rauzy sequences*, Ann. Inst. Fourier (Grenoble) 50 (2000), 1265–1276.
- [6] M. G. Castelli, F. Mignosi and A. Restivo, *Fine and Wilf’s theorem for three periods and a generalization of sturmian words*, Theoret. Comput. Sci. 218 (1999), 83–94.
- [7] R. V. Chacon, *Weakly mixing transformations which are not strongly mixing*, Proc. Amer. Math. Soc. 22 (1969), 559–562.
- [8] N. Chekhova, *Propriétés combinatoires des suites d’Arnoux–Rauzy et réponse à une question d’Hedlund et Morse*, preprint, 1999.
- [9] —, *Algorithme d’approximation et propriétés ergodiques des suites d’Arnoux–Rauzy*, preprint, 1999.

- [10] N. Chekhova, P. Hubert et A. Messaoudi, *Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci*, J. Théor. Nombres Bordeaux, to appear.
- [11] E. M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory 7 (1973), 138–153.
- [12] F. M. Dekking, *On the Prouhet–Thue–Morse measure*, Acta Univ. Carolin. Math. Phys. 33 (1992), 35–40.
- [13] X. Droubay, J. Justin and G. Pirillo, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci., to appear.
- [14] S. Ferenczi, *Les transformations de Chacon : combinatoire, structure géométrique, lien avec les systèmes de complexité $2n + 1$* , Bull. Soc. Math. France 123 (1995), 271–292.
- [15] S. Ferenczi, C. Holton and L. Q. Zamboni, *The structure of 3-interval exchange transformations I: An arithmetic study*, preprint, 2000.
- [16] —, —, —, *The structure of 3-interval exchange transformations II: Combinatorial properties*, preprint, 2000.
- [17] —, —, —, *The structure of 3-interval exchange transformations III: Ergodic and spectral properties*, preprint, 2000.
- [18] S. Ferenczi and C. Mauduit, *Transcendence of numbers with a low complexity expansion*, J. Number Theory 67 (1997), 146–161.
- [19] F. Mignosi, *On the number of factors of sturmian words*, Theoret. Comput. Sci. 82 (1991), 71–84.
- [20] M. Morse and G. A. Hedlund, *Symbolic dynamics II: Sturmian sequences*, Amer. J. Math. 62 (1940), 1–42.
- [21] G. Rauzy, *Nombres algébriques et substitutions*, Bull. Soc. Math. France 110 (1982), 147–178.
- [22] R. N. Risley and L. Q. Zamboni, *A generalization of Sturmian sequences: Combinatorial structure and transcendence*, Acta Arith. 95 (2000), 167–184.
- [23] N. B. Slater, *Gaps and steps for the sequence $n\theta \pmod{1}$* , Proc. Cambridge Philos. Soc. 63 (1967), 1115–1123.
- [24] L. Q. Zamboni, *Une généralisation du théorème de Lagrange sur le développement en fraction continue*, C. R. Acad. Sci. Paris Sér. I 327 (1998), 527–530.

Department of Mathematics
 University of North Texas
 Denton, TX 76203-5116, U.S.A.
 E-mail: nwozny@caltech.edu
 luca@unt.edu

Received on 28.9.1999
 and in revised form 9.2.2000

(3693)