Jason P. BELL

Logarithmic frequency in morphic sequences

Logarithmic frequency in morphic sequences

par Jason P. BELL

RÉSUMÉ. Nous répondons affirmativement à une question d'Allouche et Shallit en montrant l'existence de la fréquence logarithmique des lettres et des mots dans une suite morphique.

ABSTRACT. We study the logarithmic frequency of letters and words in morphic sequences and show that this frequency must always exist, answering a question of Allouche and Shallit.

1. Introduction

Let Σ be a finite alphabet. Given a sequence $s = s(1), s(2), s(3), \ldots$ taking values in Σ it is natural to ask the question: With what frequency does a letter/word $a \in \Sigma^*$ appear in the sequence s?

To make this question precise, we must carefully define what we mean by frequency. Given a sequence s, the *ordinary frequency* of a letter a in s is simply the limit

$$\lim_{x \to \infty} \frac{\#\{n \le x \mid s(n) = a\}}{x},$$

if it exists. More generally, if w is a word of length d on Σ, then the ordinary frequency of w in s is the limit

$$\lim_{x \to \infty} \frac{\#\{n \le x \mid s(n)s(n+1) \cdots s(n+d-1) = w\}}{x},$$

if it exists.

The main problem with the ordinary frequency is that for many interesting clases of sequences, it cannot be guaranteed that it exists. For this reason, the *logarithmic frequency* is often a better measure of frequency.

Definition 1. Given a finite alphabet Σ and a sequence $s = s(1), s(2), s(3), \ldots$ taking values in Σ, we define the *logarithmic frequency* of a letter $a \in \Sigma$ to be the limit

$$\lim_{x \to \infty} \frac{1}{\log x} \sum_{\{n \le x \mid s(n) = a\}} \frac{1}{n}$$

provided the limit exists. Similarly, we define the *logarithmic frequency* of a word $w \in \Sigma^*$ to be the limit

$$\lim_{x \to \infty} \frac{1}{\log x} \sum_{\{n \leq x \; \mid \; s(n) \cdots s(n+d-1)=w\}} \frac{1}{n}$$

provided the limit exists, where d is the length of w.

The main advantage of logarithmic frequency over ordinary frequency is that it exists for large classes of sequences; for example, Cobham showed that the logarithmic frequency of letters in automatic sequences always exists [2], while the ordinary frequency need not exist for automatic sequences. Logarithmic frequency generalizes ordinary frequency in the following sense: if the ordinary frequency of a letter/word in a sequence exists, then the logarithmic frequency will also exist and be equal to the ordinary frequency. Automatic sequences are a subset of the collection of *morphic* sequences. We define morphic sequences here.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a morphism. We say that a letter $a_i \in \Sigma$ is *mortal* if $\phi^j(a_i) = \epsilon$, the empty word, for some $j \geq 1$. Suppose that $\phi(a_1) = a_1 x$ for some non-empty word x containing a non-mortal letter. Then we say that ϕ is prolongable on a_1. Then we can form the right-infinite word

$$\phi^\omega(a_1) := a_1 x \phi(x) \phi^2(x) \cdots .$$

Note that the right-infinite word $\phi^\omega(a_1)$ is a fixed point of ϕ. Words defined in this manner are called *pure morphic words*. In general, a word u on a finite alphabet Δ is *morphic* if there is a set map $f : \Sigma \to \Delta$ such that $u = f(w)$ for some pure morphic word w on Σ. We note that a word can be thought of as an infinite sequence of letters, and it is this interpretation that we use when we talk of letter frequency.

Since automatic sequences form a subset of the collection of morphic sequences, it is natural to ask whether the logarithmic frequency of a letter in a morphic sequence exists. This is a question of Allouche and Shallit [1, Section 8.8, p. 282]. Our main theorem is the following.

Theorem 1.1. *Let s be a morphic sequence on a finite alphabet Σ. Then the logarithmic frequency of a word $w \in \Sigma^*$ in s exists.*

We note that the ordinary frequency in morphic sequences has been studied by many authors [3, 4, 5, 6].

Our approach to proving Theorem 1.1 is to obtain how many occurrences of a word w appear in certain "blocks" of $\phi^\omega(a_1)$. We then use the fact that ϕ is a morphism to obtain sufficiently small blocks that we can in fact compute the logarithmic frequency.

The outline of this paper is as follows. In Section 2, we discuss incidence matrices and basic results about nonnegative matrices. In Section 3, we give

notation that will be used in obtaining the proof of our main theorem. In Section 4, we give asymptotic results that will be necessary for the proof. In Section 5 we prove Theorem 1.1.

2. Incidence matrices

In this section we discuss the relationship between matrices and morphisms and Perron-Frobenius theory.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a morphism. We define the *incidence matrix* $M(\phi)$ of ϕ to be the $d \times d$ matrix whose (i, j) entry is the number of occurrences of a_i in $\phi(a_j)$. We note that $M(\phi^n) = M(\phi)^n$. From this we get the following fact:

Proposition 2.1. *Let* $\Sigma = \{a_1, \ldots, a_d\}$ *be a finite alphabet and let* $\phi : \Sigma^* \to \Sigma^*$ *be a morphism. If w is a finite word on Σ then*

$$|\phi^n(w)| = [1 \ 1 \ \cdots \ 1]M(\phi)^n \mathbf{v},$$

where $\mathbf{v}$ is the Parikh vector of the word w; that is, $\mathbf{v}$ is the $d \times 1$ column vector whose ith coordinate is the number of occurrences of the letter a_i in the word w.

Proof. See Corollary 8.2.4 of Allouche and Shallit [1]. $\qquad\square$

We note that if Σ is a finite alphabet and $\phi : \Sigma^* \to \Sigma^*$ is a morphism, then $M(\phi)$ is a *nonnegative matrix*; that is, its entries are all nonnegative real numbers. Of fundamental importance in the study of nonnegative matrices is the Perron-Frobenius theorem.

Theorem 2.2. *(Perron-Frobenius) Let A be a nonnegative matrix. Then there is a nonnegative real eigenvalue α of A such that if λ is an eigenvalue of A then either $|\lambda| < \alpha$ or $\lambda = \alpha\omega$ for some root of unity ω.*

The eigenvalue α is called a *Perron-Frobenius* eigenvalue of A. We denote the Perron-Frobenius eigenvalue of a nonnegative matrix A by $\alpha(A)$. We note that any other eigenvalue of A on the circle $|z| = \alpha(A)$ satisfies $\lambda^d = \alpha^d$ for some d. We thus obtain the following remark.

Remark 1. *If A is a nonnegative matrix then there is some number d such that for every eigenvalue λ of A^d, if $\lambda \neq \alpha(A^d)$ then $|\lambda| < \alpha(A^d)$.*

Definition 2. Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a morphism. We say that ϕ is *strongly Perron-Frobenius* if for every eigenvalue λ of the incidence matrix $M(\phi)$, if $\lambda \neq \alpha(M(\phi))$ then $\lambda < \alpha(M(\phi))$.

Lemma 2.3. *Let A be a $d \times d$ nonnegative matrix and let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$ be two vectors with nonnegative entries. Suppose α is the Perron-Frobenius eigenvalue of A and $\beta < \alpha$ is such that all other eigenvalues of A are strictly*

*less than β in absolute value. Then there is a (possibly zero) polynomial $p(x)$
such that*

$$\mathbf{w}^{\mathrm{T}} A^n \mathbf{v} = p(n)\alpha^n + \mathrm{O}(\beta^n).$$

Proof. There exists an invertible complex $d \times d$ matrix S such that

$$S^{-1} A S = J$$

is in Jordan form. The i, j entry of J^n can be expressed as $q_{i,j}(n)\alpha^n + \mathrm{O}(\beta^n)$
for some polynomial $q_{i,j}$, possibly zero. Let $\mathbf{w}_0 = \mathbf{w}S$ and $\mathbf{v}_0 = JS^{-1}\mathbf{v}$.
Then

$$\mathbf{w}^{\mathrm{T}} A^n \mathbf{v} = \mathbf{w}_0^{\mathrm{T}} J^n \mathbf{v}_0.$$

Since the entries of J^n are all of the form $q(n)\alpha^n + \mathrm{O}(\beta^n)$, we obtain the
desired result. $\qquad\square$

We are considering nonnegative matrices with integer entries. In this
case an interesting trichotomy arises: the Perron-Frobenius eigenvalue is
either strictly larger than 1, is equal to 1, or is 0. (It cannot be strictly
between 0 and 1 since the product of the nonzero roots of the characteristic
polynomial of M is an integer.) If the Perron-Frobenius eigenvalue is zero,
then M is nilpotent. In particular, if $M = M(\phi)$ for some morphism ϕ of a
finite alphabet, then every letter is mortal with respect to this morphism. If
the Perron-Frobenius eigenvalue is 1 and M is strongly Perron-Frobenius,
the only eigenvalues of A are 0 and 1. In particular, the $\mathrm{O}(\beta^n)$ term can
be eliminated in the statement of Lemma 2.3 and all quantities involved
become polynomials in n. In this case, the logarithmic frequency of words
always exists. In fact, the ordinary frequency of words even exists [6], [7,
Theorem 3.8]! The case that the Perron-Frobenius eigenvalue is strictly
larger than 1 is the case we consider.

3. Notation

In this section we introduce notation that will be used throughout our
proof that logarithmic frequency of words in morphic sequences exists.

Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and suppose that $\phi : \Sigma^* \to \Sigma^*$
is a strongly Perron-Frobenius morphism prolongable on a_1. We keep the
morphism ϕ fixed throughout this section. As is standard, if v and w are
words on the same alphabet, we let $|v|_w$ denote the number of occurrences
of w in the word v

Write $\phi(a_1) = a_1 x$ with $x \in \Sigma^*$ and let m be a natural number. Write

$$\phi^m(x) = a_{i_1} a_{i_2} \cdots a_{i_{t(m)}},$$

where

(1) $$t(n) = |\phi^{n+1}(a_1)| - |\phi^n(a_1)|.$$

We note that if

$$\phi^m(x) = a_{i_1} a_{i_2} \cdots a_{i_{t(m)}},$$

then

$$\phi^{n+m}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{t(m)}})$$

and

$$|\phi^{n+m}(x)| = \sum_{j=1}^{t(m)} |\phi^n(a_{i_j})|.$$

We define $t(m)$ functions

$$(2) \qquad h_{j,m}(n) = |\phi^{n+m}(a_1)| + \sum_{k=1}^{j} |\phi^n(a_{i_k})| \qquad \text{for } 0 \le j \le t(m).$$

Note that in the expression of $h_{j,m}(n)$, $|\phi^{n+m}(a_1)|$ is just the length of $\phi^{n+m}(a_1)$, and the sum is the length of $\phi^n(a_{i_1} \cdots a_{i_j})$. Thus $h_{j,m}(n)$ is the length of an initial subword of $\phi^{n+m+1}(a_1)$ that contains $\phi^{n+m}(a_1)$. Let $w \in \Sigma^*$. We now define two new functions, which will be important in getting upper and lower bounds when estimating the logarithmic frequency.

$$(3) \qquad A_{m,w}(n) = \sum_{j=1}^{t(m)} \Big(\log(h_{j,m}(n) + |\phi^n(a_{i_j})|_w - 1) - \log(h_{j,m}(n) - 1) \Big)$$

and

$$(4) \qquad B_{m,w}(n) = \sum_{j=1}^{t(m)} \Big(\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w) \Big).$$

In the study of logarithmic frequency of a word $w \in \Sigma^*$ in a morphic sequence, it is of course necessary to know the positions at which w occur.

Definition 3. Let $\Sigma = \{a_1, \dots, a_d\}$ be a finite alphabet and suppose that $\phi : \Sigma^* \to \Sigma^*$ is a strongly Perron-Frobenius morphism prolongable on a_1. We define

$$s_w(n) = s_w(n) := \begin{cases} 1 & \text{if there is an occurrence of } w \text{ at} \\ & \quad \text{the } n\text{th position of } \phi^\omega(a_1), \\ 0 & \text{otherwise.} \end{cases}$$

4. Asymptotics

In this section, we prove several asymptotic results that we will need to prove Theorem 1.1

Proposition 4.1. *Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a strongly Perron-Frobenius morphism prolongable on a_1. Suppose that every letter $a \in \Sigma$ appears in $\phi^k(a_1)$ for some k. Then there are polynomials $p_1, \ldots, p_d$ with $p_1 \neq 0$ and real numbers α, β with $\alpha > \beta > 0$ such that*

$$|\phi^n(a_j)| = p_j(n)\alpha^n + \mathrm{O}(\beta^n).$$

Proof. Let α denote the largest eigenvalue of the matrix $M(\phi)$. Choose $\beta < \alpha$ such that the remaining eigenvalues of $M(\phi)$ are strictly less than β in absolute value. Then by Lemma 2.3, there exist polynomials $p_1, \ldots, p_d$ such that

$$(5) \qquad |\phi^n(a_j)| = p_j(n)\alpha^n + \mathrm{O}(\beta^n).$$

To see why $p_1 \neq 0$, note that $M(\phi)$ has α as an eigenvalue. Hence there is some nonzero vector $\mathbf{v}$ such that $M(\phi)\mathbf{v} = \alpha\mathbf{v}$. Note that $\mathbf{v}$ is a linear combination of the Parikh vectors of $a_1, a_2, \ldots, a_d$. By Proposition 2.1, it follows that $\alpha^n[1, 1, \ldots, 1] \cdot \mathbf{v}$ is the same linear combination of $|\phi^n(a_1)|, |\phi^n(a_2)|, \ldots, |\phi^n(a_d)|$. Hence p_j is nonzero for some j. By hypothesis, $|\phi^n(a_1)| \geq |\phi^{n-k}(a_j)|$, and so p_1 must also be nonzero. $\qquad \square$

We note that by the definition of the function $h_{j,m}(n)$, they must also have asymptotics as described in the statement of Proposition 4.1.

Proposition 4.2. *Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a strongly Perron-Frobenius morphism prolongable on a_1. If $w \in \Sigma^*$, and α is the Perron-Frobenius eigenvalue of $M(\phi)$ then there are polynomials $q_1, \ldots, q_d$ and a nonnegative real number β with $\beta < \alpha$ such that $|\phi^n(a_j)|_w = q_j(n)\alpha^n(1 + \mathrm{o}(1))$.*

Proof. Let m denote the length of w. By relabeling if necessary, we may assume that $\{a_e, \ldots, a_d\}$ are the mortal letters with respect to ϕ. We note that in any word of the form $\phi^n(a_j)$ there is a uniform bound on the size of the "gaps" between consecutive appearances of non-mortal letters. Hence there are distinct natural numbers b and c with $b < c$ such that

    (1) the shortest prefix of $\phi^b(a_j)$ containing ℓ non-mortal letters is the same as the shortest prefix of $\phi^c(a_j)$ containing ℓ non-mortal letters for $1 \leq j \leq d$;

    (2) the shortest suffix of $\phi^b(a_j)$ containing ℓ non-mortal letters is the same as the shortest suffix of $\phi^c(a_j)$ containing ℓ non-mortal letters for $1 \leq j \leq d$.

(We note that this prefix and suffix need not exist if a_j is mortal, or $|\phi^n(a_j)|$ is bounded.) Pick k such that $\phi^k(a_j) = \varepsilon$ for $e \leq j \leq d$. Then by construction, the first ℓ letters of $\phi^n(a_j)$ are the same as the first ℓ letters of $\phi^{n+c-b}(a_j)$ for $1 \leq j \leq d$ and $n \geq b + k$.

Write

$$\phi(a_j) = a_{i_1,j} \cdots a_{i_{m_j},j},$$

for $1 \leq j \leq d$. Then

$$\phi^{n+1}(a_j) = \phi^n(a_{i_1,j}) \cdots \phi^n(a_{i_{m_j},j}).$$

Looking at both sides and counting occurrences of w, we see

$$|\phi^{n+1}(a_j)|_w = |\phi^n(a_{i_1})|_w + \cdots |\phi^n(a_{i_{m_j}})|_w + C(j, n),$$

where $C(j, n)$ counts the number of occurrences of w that "overlap" at least two words of the form $\phi^n(a_{i,j})$. By the above remarks, we have $C(j, n) = C(j, n + c - b)$ for $n \geq b + k$. Let

(6) $$G_{j,w}(n) = |\phi^n(a_j)|_w - |\phi^{n-c+b}(a_j)|_w.$$

Then we obtain the vector equation:

$$\begin{bmatrix} G_{1,w}(n+1) \\ \vdots \\ G_{d,w}(n+1) \end{bmatrix} = M(\phi) \begin{bmatrix} G_{1,w}(n) \\ \vdots \\ G_{d,w}(n) \end{bmatrix}$$

for $n \geq b + k$. It follows from Lemma 2.3 that there are polynomials $Q_1, \ldots, Q_d$ such that

$$G_{j,w}(n) = Q_j(n)\alpha^n + O(\beta^n).$$

We note that if $\alpha = 1$, then by the remarks at the end of Section 2, we can ignore the $O(\beta^n)$ term and by telescoping we see that

$$|\phi^n(a_j)|_w = \left(\sum_{k=0}^{\lfloor (n-b-k)/(c-b) \rfloor} Q_j(n - k(c+b)) \right),$$

which is asymptotic to some polynomial $q_j(n)$. If $\alpha > 1$, then we can pick $\beta > 1$. Then

$$|\phi^n(a_j)|_w = \left(\sum_{k=0}^{\lfloor (n-b-k)/(c-b) \rfloor} Q_j(n - k(c+b)) \right) \alpha^n + O(\beta^n),$$

which is again asymptotic to $q_j(n)\alpha^n$ for some polynomial $q_j(n)$. $\qquad\square$

We note that we do not say anything about the error term. We could get a stronger error term, however, by replacing ϕ by a suitable iterate.

5. Estimates

In this section we prove Theorem 1.1. To this end we need the following estimates.

Estimate 1. *Let a and b be integers with $b > a > 2$. Then*

$$\log(b) - \log(a) \le \sum_{n=a}^{b-1} 1/n \le \log(b-1) - \log(a-1).$$

Proof. This is a straightforward consequence of the fact that the function $1/x$ is decreasing on $[1, \infty)$, using upper and lower Riemann sums. $\square$

Estimate 2. *Let a, b, c be positive real numbers with $b > a$ and let $F(x) = \log(x+c) - \log(x)$. Then*

$$|F(b) - F(a)| \le (b-a)c/a^2.$$

Proof. This is an easy consequence of the Mean Value Theorem. $\square$

Lemma 5.1. *Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a strongly Perron-Frobenius morphism prolongable on a_1. Then*

$$B_{m,w}(n) \le \sum_{j=|\phi^{n+m}(a_1)|}^{|\phi^{n+m+1}(a_1)|-1} s_w(j)/j \le t(m)|w|/|\phi^{n+m}(a_1)| + A_{m,w}(n)$$

for $m, n \ge 1$.

Proof. Write $\phi(a_1) = a_1 x$ and $\phi^m(x) = a_{i_1} \cdots a_{i_{t(m)}}$. Then

$$(7) \qquad\qquad \phi^{m+n}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{t(m)}}).$$

Using the notation of Section 3, we have

$$\sum_{j=|\phi^{n+m}(a_1)|}^{|\phi^{n+m+1}(a_1)|-1} s_w(j)/j = \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j+1,m}(n)-1} s_w(k)/k.$$

By assumption, there are exactly $|\phi^n(a_{i_j})|_w$ occurrences of the word w in $\phi^n(a_{i_j})$. This accounts for all occurrences of w in $\phi^{m+n}(x)$ except for those which intersect at least two words of the form $\phi^n(a_{i_k})$ that appear in the right-hand side of equation (7). Since there are $t(m)$ such words, this leaves at most

$$t(m)|w|$$

unaccounted for occurrences of w. Therefore, we see that

$$t(m)|w|/|\phi^{n+m}(a_1)| + \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j,m}(n+1)+|\phi^n(a_{i_j})|_w-1} 1/k$$

$$\geq \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j+1,m}(n)-1} s_w(k)/k$$

$$\geq \sum_{j=1}^{t(m)} \sum_{k=h_{j+1,m}(n)-|\phi^n(a_{i_j})|_w}^{h_{j+1,m}(n)-1} 1/k.$$

We now use the estimates from Remark 1 to obtain:

$$\sum_{j=1}^{t(m)} \sum_{k=h_{j+1,m}(n)-|\phi^n(a_{i_j})|_w}^{h_{j+1,m}(n)-1} 1/k$$

$$\geq \sum_{j=1}^{t(m)} \log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w)$$

$$= B_{m,w}(n).$$

Similarly,

$$\sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j,m}(n)+|\phi^n(a_{i_j})|_w-1} 1/k$$

$$\leq \sum_{j=1}^{t(m)} \log(h_{j,m}(n) + |\phi^n(a_{i_j})|_w - 1) - \log(h_{j,m}(n) - 1)$$

$$= A_{m,w}(n).$$

The result follows. $\qquad\square$

Lemma 5.2. *Let* $\Sigma = \{a_1, \ldots, a_d\}$ *be a finite alphabet and let* $\phi : \Sigma^* \to \Sigma^*$ *be a strongly Perron-Frobenius morphism prolongable on* a_1. *Let* m *be a natural number and* $w \in \Sigma^*$, *then*

$$\lim_{n\to\infty} A_{m,w}(n)$$

and

$$\lim_{n\to\infty} B_{m,w}(n)$$

exist and are finite.

Proof. Fix m and write $\phi(a_1) = a_1 x$ and $\phi^m(x) = a_{i_1} \cdots a_{i_{t(m)}}$. Consider $A_{m,w}(n)$. Since $t(m)$ is fixed it is sufficient to show that

$$\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w)$$

and

$$\log(h_{j,m}(n+1) + |\phi^n(a_{i_j})|_w - 1) - \log(h_{j,m}(n+1) - 1)$$

tend to a finite limit for each j. By the remarks immediately following Proposition 4.1 and by Proposition 4.2, both $h_{j+1,m}(n)$ and $h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w$ can be expressed in the form $p(n)\alpha^n(1 + o(1))$, for suitable polynomials $p(x)$, where α is a Perron-Frobenius eigenvalue of $M(\phi)$. Thus

$$\log\left(h_{j+1,m}(n)/(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w\right) = \log p(n)/q(n) + o(1),$$

for some polynomials p and q. We note that p and q are nonzero and satisfy $p(n) \geq q(n)$ for sufficiently large n, since

$$h_{j+1} \geq h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w \geq h_0(n) \geq |\phi^{n+m}(a_1)|,$$

and $|\phi^n(a_1)|$ is asymptotic to $Cn^k\alpha^n$ for some positive constant C by Proposition 4.1.

We note that if p and q have the same degree then the limit exists and is finite, so the only way a problem can arise is if the degree of p is strictly greater than the degree of q. But if this occurs the limit must be infinite, and by construction we have

$$\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w) \leq \sum_{j=|\phi^{n+m}(a_1)|}^{|\phi^{n+m+1}(a_1)|} 1/j \ \rightarrow \ \log\alpha,$$

and so the limsup cannot exceed α. A similar argument works for $B_{m,w}(n)$. The result follows.   □

Lemma 5.3. *Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \to \Sigma^*$ be a strongly Perron-Frobenius morphism prolongable on a_1. Then there is a positive constant K and a natural number k such that for all m sufficiently large,*

$$\limsup_{n \to \infty} |A_{m,w}(n) - B_{m,w}(n)| \leq \begin{cases} Km^k\alpha^{-m/2} & \text{if } \alpha > 1 \\ 0 & \text{if } \alpha = 1. \end{cases}$$

Proof. Using Estimate 2, we have

$$A_{m,w}(n) - B_{m,w}(n)$$

$$= \sum_{j=1}^{t(m)} \Big(\log(h_{j+1,m}(n)) - \log(h_{j+1,m}(n) - |\phi^n(a_{i_j})|_w$$

$$\qquad - \log(h_{j,m}(n+1) + |\phi^n(a_{i_j})|_w + \log(h_{j,m}(n+1)) \Big)$$

$$\leq \sum_{j=1}^{t(m)} |h_{j+1,m}(n) - h_{j,m}(n) - |\phi^n(a_{i_j})|_w| |\phi^n(a_{i_j})|_w| / h_{j,m}(n)^2$$

$$\leq \sum_{j=1}^{t(m)} |h_{j+1,m}(n)/h_{j,m}(n) - 1|^2$$

$$\leq \sum_{j=1}^{t(m)} \Big| |\phi^n(a_{i_j})| / h_{j,m}(n) \Big|^2 .$$

We now divide the proof into cases.

CASE I : $\alpha > 1$.

It is no loss of generality to assume that for each $i \leq d$, there is some number m such that a_i appears as a letter in $\phi^m(a_1)$, otherwise we can just delete the letter a_i from Σ. That is,

(8) For each i, there exists e_i such that $\phi^{e_i}(a_1)$ has an occurrence of a_i.

Thus we may assume that there exist natural numbers $e_1, \ldots, e_d$ such that

$$|\phi^n(a_j)| \geq |\phi^{n-e_j}(a_1)|$$

for all applicable n; furthermore, $h_{j,m}(n) \geq h_{0,m}(n) = |\phi^{n+m}(a_1)|$. Pick $e \geq e_1, \ldots, e_d$. By Proposition 4.1 there exist positive constants C_0 and C_1 and a natural number k such that

$$C_0 n^k \alpha^n \leq |\phi^n(a_1)| \leq C_1 n^k \alpha^n$$

for all sufficiently large n. Hence

$$|A_{m,w}(n) - B_{m,w}(n)| \leq t(m) |\phi^{n+e}(a_1)|^2 / |\phi^{n+m}(a_1)|^2$$

$$\leq t(m) \left(\frac{C_1(n+e)^k \alpha^{n+k}}{C_0(n+m)^k \alpha^{n+m}} \right)^2$$

for all n sufficiently large, Thus

$$|A_{m,w}(n) - B_{m,w}(n)| \leq t(m) C_1^2 C_0^{-2} \alpha^{-2m+e}$$

for n sufficiently large. Since $t(m) \leq |\phi^m(a_1)| \leq C_1 m^k \alpha^m$ for m sufficiently large, we see that

$$|A_{m,w}(n) - B_{m,w}(n)| \leq C_1^3 C_0^{-2} m^k \alpha^{-m+e}$$

whenever m and n are both sufficiently large. The result follows.   □

CASE II : $\alpha = 1$. In this case, note that

$$|\phi^n(a_{i_j})| \leq h_{j+1,m}(n) - h_{j,m}(n) \leq |\phi^{n+m+1}(a_1)| - |\phi^{n+m}(a_1)|$$

and so

$$|\phi^n(a_{i_j})|/h_{j,m}(n) \leq |\phi^{n+m+1}(a_1)|/|\phi^{n+m}(a_1)| - 1 \leq K/(n+m)$$

as $n \to \infty$ for some positive constant K, since $|\phi^n(a_1)|$ is a polynomial in n. Thus

$$\sum_{j=1}^{t(m)} \left| |\phi^n(a_{i_j})|/h_{j,m}(n) \right|^2 \leq Kt(m)/(n+m)^2.$$

Since $t(m) = O(m^k)$, the result follows.   □

We are almost ready to prove Theorem 1.1. We first make a simple remark.

Remark 2. *Let Σ and Δ be two finite alphabets and let $f : \Sigma \to \Delta$. Suppose w is a pure morphic word on Σ such that the logarithmic frequency of every subword exists. Then the logarithmic frequency of every subword of $f(w)$ exists.*

This is a consequence of the fact that if x is a subword of $f(w)$ then $f^{-1}(x)$ is a finite collection of subwords of w and the logarithmic frequency of x in $f(w)$ is simply the sum of the logarithmic frequencies of the subwords in $f^{-1}(x)$.

Proof of Theorem 1.1: By Remark 2, we may assume that our morphic sequence is pure. Let $\Sigma = \{a_1, \ldots, a_d\}$ be a finite alphabet and let ϕ be a morphism from Σ^* to itself that is prolongable on a_1. By Remark 1, we can assume that ϕ is strongly Perron-Frobenius by replacing it with a suitable iterate. We let α denote the Perron-Frobenius eigenvalue of the incidence matrix of ϕ. By a result of Saari [7, Theorem 3.8], the ordinary frequency of all words exists when $\alpha = 1$, and hence we may assume $\alpha > 1$.

Write $\phi(a_1) = a_1 x$ with $x \in \Sigma^*$. Let m be a natural number and write

$$\phi^m(x) = a_{i_1} a_{i_2} \cdots a_{i_{t(m)}}.$$

We note that the word

$$\phi^{n+m}(x) = \phi^n(a_{i_1}) \cdots \phi^n(a_{i_{t(m)}})$$

has length

$$\sum_{j=1}^{t(m)} |\phi^n(a_{i_j})|.$$

Since for fixed m, $t(m)|w|/|\phi^{n+m}(a_1)| \to 0$ as $n \to \infty$, we see by Lemma 5.1 that

$$\left| \limsup_{n\to\infty} \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j+1,m}(n)-1} s_w(k)/k - \liminf_{n\to\infty} \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j+1,m}(n)-1} s_w(k)/k \right|$$

$$\leq \left| \limsup_{n\to\infty} A_{m,w}(n) - \liminf_{n\to\infty} B_{m,w}(n) \right|$$

for all m. By Lemmas 5.2 and 5.3, $\lim_{n\to\infty} A_{m,w}(n)$ and $\lim_{n\to\infty} B_{m,w}(n)$ exist, and

$$\lim_{m\to\infty} \lim_{n\to\infty} |A_{m,w}(n) - B_{m,w}(n)| = 0.$$

Thus we see that the limit

$$\lim_{m\to\infty} \lim_{n\to\infty} \sum_{j=1}^{t(m)} \sum_{k=h_{j,m}(n)}^{h_{j+1,m}(n)-1} s_w(k)/k$$

exists. But this is the same as saying

$$\lim_{m\to\infty} \lim_{n\to\infty} \sum_{|\phi^{n+m}(a_1)|\leq j\leq|\phi^{n+m+1}(a_1)|} s_w(j)/j$$

exists, which implies

$$\lim_{n\to\infty} \sum_{|\phi^n(a_1)|\leq j\leq|\phi^{n+1}(a_1)|} s_w(j)/j$$

exists. We let γ denote this limit. We claim that the logarithmic frequency of the subword w in the pure morphic sequence $\mathbf{s}$ is $\gamma/\log(\alpha)$. Let $\varepsilon > 0$ and pick $N > 0$ such that

$$(9) \qquad \left| \sum_{j=|\phi^n(a_1)|+1}^{|\phi^{n+1}(a_1)|} s_w(j)/j - \gamma \right| < \varepsilon$$

for all $n \geq N$. Let M be a natural number that is greater than N. Then for $|\phi^M(a_1)| \leq x < |\phi^{M+1}(a_1)|$, we use Proposition 4.1 to infer that there are positive constants C_0 and C_1 such that

$$(10) \qquad C_0 n^k \alpha^n \leq |\phi^n(a_1)| \leq C_1 n^k \alpha^n$$

for n sufficiently large. We next obtain the estimate

$$\frac{1}{\log x} \sum_{j \leq x} s_w(j)/j$$

$$\leq \frac{1}{\log x} \left(\sum_{i=1}^{|\phi^N(a_1)|} 1/i + \sum_{i=N}^{M} \sum_{j=|\phi^i(a_1)|+1}^{|\phi^{i+1}(a_1)|} s_w(j)/j \right)$$

$$\leq \frac{1}{\log x} \left(\log(|\phi^N(a_1)|) + 1 + (M - N + 1)(\gamma + \varepsilon) \right),$$

where we use Estimate 1 to estimate the first sum and equation (9) for the second. Using this expression and equation (10) to estimate $\log x$, we see

$$\frac{1}{\log x} \sum_{j \leq x} s_w(j)/j$$

$$\leq \frac{\log(|\phi^N(a_1)|) + 1 + (M - N + 1)(\gamma + \varepsilon)}{M \log \alpha + \log(C_0) + k \log M}$$

$$\rightarrow \frac{\gamma + \varepsilon}{\log \alpha}$$

as $M \rightarrow \infty$. Similarly,

$$\liminf_{x \rightarrow \infty} \frac{1}{\log x} \sum_{j \leq x} s_w(j)/j \geq (\gamma - \varepsilon)/\log(\alpha).$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Acknowledgments

References

[1] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, Cambridge, 2003.

[2] A. Cobham, *Uniform tag sequences.* Math. Systems Theory. **6** (1972), 164–192.

[3] P. Michel, *Sur les ensembles minimaux engendrés par les substitutions de longueur non constante.* Thèse, Université de Rennes, 1975.

[4] P. Michel, *Stricte ergodicité dÔensembles minimaux de substitution.* Théorie Ergodique: Actes des Journées Ergodiques, Rennes, 1973/1974, Lecture Notes in Mathematics **532**, Springer-Verlag, 1976.

[5] S. Nicolay, M. Rigo, *About frequencies of letters in generalized automatic sequences.* Theoret. Comput. Sci. **374** (2007), no. 1-3, 25–40.

[6] K. Saari, *On the frequency of letters in morphic sequences.* Computer science—theory and applications, 334–345, Lecture Notes in Comput. Sci. **3967**, Springer, Berlin, 2006.

[7] K. Saari, *On the frequency and periodicity of infinite words.* PhD thesis, University of Turku, 2008.

Jason P. Bell
Department of Mathematics
Simon Fraser University
Burnaby, BC, V5A 1S6, CANADA
E-mail: jpb@math.sfu.ca