ABOUT FREQUENCIES OF LETTERS IN GENERALIZED AUTOMATIC SEQUENCES

S. NICOLAY, M. RIGO

Abstract. We present some asymptotic results about the frequency of a letter appearing in a generalized unidimensional automatic sequence. Next, we study multidimensional generalized automatic sequences and the corresponding frequencies.

1. Introduction

An infinite sequence which is the image under a letter-to-letter morphism of the fixed point of a prolongable morphism μ is said to be *morphic*. If all images under μ of letters have same length $k \geq 2$ then the sequence is said to be *k-automatic*. In the seminal paper [2] A. Cobham shows that if the frequency of a symbol appearing in a k-automatic sequence exists then it is rational. Extended results about the frequency of a symbol appearing in a k-automatic sequence have been obtained recently in [10]. For a morphic sequence, a criterion for the existence of the frequency of a letter has been obtained in [14], and if this frequency exists then it is an algebraic number (see for instance [1, Theorem 8.4.5]).

Here, we consider generalized automatic sequences as introduced in [12, 13]. We say that a sequence $(x_n)_{n\in\mathbb{N}}$ is *S-automatic* if it can be constructed as follows. In all this paper, we consider an *abstract numeration system* $S = (L, \Sigma, <)$ consisting of an infinite regular language L over the totally ordered alphabet $(\Sigma, <)$. Enumerating the words of L by increasing genealogical ordering (also called radix order) gives a one-to-one correspondence val_S between L and $\mathbb{N}$. Otherwise stated, $\mathrm{val}_S(w) = n$ if w is the $(n+1)$th word in the ordered language L (for an introduction to abstract numeration systems, see for instance [7]). In this paper, $\mathcal{M} = (Q, q_0, \Sigma, \delta, F)$ will always refer to the minimal automaton of L (for details about automata theory, we refer to [3]). As usual, Q is the finite set of states of $\mathcal{M}$, q_0 is its initial state, $\delta : Q \times \Sigma \to Q$ is the transition function and $F \subseteq Q$ is the set of final states. The transition function can be extended to $\delta : Q \times \Sigma^* \to Q$ by $\delta(q, \varepsilon) = q$ and $\delta(q, \sigma w) = \delta(\delta(q, \sigma), w)$, where ε is the empty word, $q \in Q$ and $w \in \Sigma^*$. We will denote by $\mathcal{A} = (Q', q_0', \Sigma, \delta', \Gamma, \tau)$ a given deterministic finite automaton with output (DFAO) where Q', q_0', δ' are defined as in $\mathcal{M}$, Γ is the output alphabet and $\tau : Q \to \Gamma$ is the output function of $\mathcal{A}$. Using the terminology of [1], given a word $w \in \Sigma^*$ the *output* of $\mathcal{A}$ for the input w is denoted $f_{\mathcal{A}}(w)$ or simply $f(w)$ and is defined by

$$f_{\mathcal{A}}(w) := \tau(\delta'(q_0', w)).$$

To shorten notation, we often write $q.w$ or $q'.w$ instead of $\delta(q, w)$ and $\delta'(q', w)$ respectively. A sequence $(x_n)_{n\in\mathbb{N}} \in \Gamma^{\mathbb{N}}$ is said to be *S-automatic*, if it can be constructed as follows,

$$x_n = f_{\mathcal{A}}(w_n)$$

where $w_n \in L$ is the word such that $\mathrm{val}_S(w_n) = n$. Otherwise stated, the n-th symbol of $(x_n)_{n\in\mathbb{N}}$ is obtained as the output of $\mathcal{A}$ fed with the $(n+1)$-st word of L, $n \geq 0$. So an S-automatic sequence is completely determined by the abstract numeration system $S = (L, \Sigma, <)$ and a DFAO $\mathcal{A}$. It is shown in [13] that the set of generalized automatic sequences and the set of morphic sequences are the same. Moreover, if the language L is equal to $\{0, \ldots, k-1\}^*$, $k \geq 2$, then a sequence is S-automatic for the abstract numeration system built over L and the usual ordering of the digits if and only if it is k-automatic [2]. Let $a \in \Gamma$ and $x = (x_n)_{n\in\mathbb{N}}$ be an infinite word over Γ, the function counting the number of a's among the first n symbols of x, $n \geq 1$, is denoted by

$$\pi(n, a, x) = \#\{i \in [0, n-1] \mid x_i = a\} = \sum_{i=0}^{n-1} \mathbf{1}_a(x_i),$$

where $\mathbf{1}_a(x_i) = 1$ if and only if $x_i = a$. If the limit

$$\lim_{n\to\infty} \frac{\pi(n, a, x)}{n}$$

exists then its value $d(a, x)$ is called the *frequency* of a.

Our main result for unidimensional S-automatic sequences (Theorem 2) explains the asymptotic behaviour of the function $\pi(n, a, x)$ under some natural hypothesis developed later. To obtain these results, we follow basically the same scheme as in [6] (in fact, this allows us to present the main differences with [6] and to avoid some technical developments) where the summatory function of a function satisfying an additive property, $f(\sigma_1 \cdots \sigma_k) = \sum_{i=1}^{k} f(\sigma_i)$, is investigated. But notice that if the sequence is S-automatic, the function $\mathbf{1}_a(x_i) = \mathbf{1}_a(f_{\mathcal{A}}(w_i))$ related to the summatory function $\pi(n, a, x)$ does not have such an additive property: $\mathbf{1}_a(f_{\mathcal{A}}(\sigma_1 \cdots \sigma_k))$ is not necessarily equal to $\sum_{i=1}^{k} \mathbf{1}_a(f_{\mathcal{A}}(\sigma_i))$.

This paper is organized in the following way. In Section 2, we present the working hypothesis, we state the results for unidimensional S-automatic sequences and spectral properties of incidence matrices related to $\mathcal{M}$. Sections 3 and 4 are devoted respectively to the proof of Theorem 2 and its corollary. In Section 5, we introduce the frequency of an m-dimensional automatic sequence. By enumerating m-tuples of words in genealogical ordering, we can view this m-dimensional sequence as a unidimensional one. It is interesting to notice that we produce a new enumeration of $\mathbb{N}^m$ analogous to the primitive recursive enumeration of Peano. It is therefore sufficient to show that the two notions of frequency for m-dimensional and unidimensional sequences coincide. In order to obtain the existence of a frequency, we develop a sufficient framework to be able to apply the same construction as Peyrière in [11]. In the last section, we show that the frequency of a letter appearing in a sequence is independent of the total ordering of the alphabet. (This result has to be mentioned because it is well-known that recognizability of a set of integers usually depends on the ordering of the alphabet, see [7].)

2. Working hypothesis and consequences

Let us be more precise. We assume that the set Σ^ω of infinite words over Σ is equipped with the usual distance t defined as follows. Let $v = v_0 v_1 \cdots$ and $w = w_0 w_1 \cdots$ be in Σ^ω. If $v \neq w$ then we set $t(v, w) = 2^{-i}$ where i is the smallest integer such that $v_i \neq w_i$. Otherwise, $v = w$ and we set $t(v, w) = 0$. This notion can be extended to $\Sigma^\infty = \Sigma^\omega \cup \Sigma^*$ by adding an extra symbol ζ to the alphabet Σ. Namely, if v belongs to Σ^* then consider the word $v\zeta^\omega$ belonging to the metric space $(\Sigma \cup \{\zeta\})^\omega$. In this setting, we can therefore speak of converging sequences of (finite or infinite) words. In this paper, we consider converging sequences of words in L and we introduce the following notation

$$\mathcal{L}_\infty = \{w \in \Sigma^\omega \mid \exists (w^{(n)})_{n\in\mathbb{N}} \in L^{\mathbb{N}} : \lim_{n\to\infty} w^{(n)} = w\}.$$

Recall from the first section that the automata $\mathcal{M}$ and $\mathcal{A}$ are given. We denote by L_q the language accepted by $\mathcal{M}$ from state q, i.e.,

$$L_q = \{w \in \Sigma^* \mid \delta(q, w) \in F\}.$$

We set $\mathbf{u}_q(n) = \#(L_q \cap \Sigma^n)$ and $\mathbf{v}_q(n) = \#(L_q \cap \Sigma^{\leq n})$. We assume that the states of Q (resp. Q') are ordered as follows

$$Q = \{q_0 < q_1 < \cdots < q_r\} \text{ and } Q' = \{q'_0 < q'_1 < \cdots < q'_s\}.$$

Therefore, we order $Q \times Q'$ by

$$(q_0, q'_0) < (q_1, q'_0) < \cdots < (q_r, q'_0) < (q_0, q'_1) < \cdots < (q_r, q'_s).$$

When dealing with vectors and matrices whose elements are indexed by Q, Q' or $Q \times Q'$, we will implicitly use these orderings. Let $\mathcal{P}$ be the product automaton defined by

$$\mathcal{P} = (Q \times Q', (q_0, q'_0), \Sigma, \Delta)$$

where the transition function Δ is such that

$$\Delta((q, q'), \sigma) = (\delta(q, \sigma), \delta'(q', \sigma)).$$

Let M (resp. A, P) be the incidence matrix of $\mathcal{M}$ (resp. $\mathcal{A}$, $\mathcal{P}$), i.e.,

$$M_{q_i,q_j} = \#\{\sigma \in \Sigma \mid \delta(q_i, \sigma) = q_j\}$$

(A and P being defined in the same way). As stated above, we use the orderings of Q, Q' and $Q \times Q'$ to order the elements of those incidence matrices. In order to relate the eigenvalues of M to the growth of the language L, we assume that $\mathcal{M}$ is *trim* (i.e., it is *accessible* — any state can be reached from the initial state, and *coaccessible* — any state can reach a final state, [3]) and that $\mathcal{A}$ is accessible and complete. Therefore the functions δ and Δ could be partial but δ' is a total function. By definition of M and P, we have

$$(1) \qquad \sum_{\ell=0}^{s} P_{(q_i,q'_k),(q_j,q'_\ell)} = M_{q_i,q_j}, \quad 0 \leq i, j \leq r, \quad 0 \leq k \leq s.$$

We will consider the following hypothesis:

(H) The matrix P has only one dominating eigenvalue $\lambda > 1$ (i.e., if $\gamma \neq \lambda$ is an eigenvalue of P, then $|\gamma| < \lambda$).

Remark 1. Assuming **(H)** is a usual consideration in the framework of substitutive sequences and a large class of S-automatic sequences fulfills **(H)**. Indeed, primitive substitutions have been widely studied [4]. (A substitution $\phi : \Gamma \to \Gamma^+$ is *primitive* if there exists k such that for any $\gamma, \gamma' \in \Gamma$, γ' appears in $\phi^k(\gamma)$.) In this case, the matrix associated with the substitution is primitive and Perron's theorem is used (see for instance [9]). In particular, any pure morphic sequence generated by a primitive substitution is clearly a special case of S-automatic sequence satisfying **(H)**.

As for Perron's theorem in the substitutive case, adopting **(H)** gives us various asymptotic estimates like, for instance, the expression of $\mathbf{u}_q(n)$ as $P_q(n)\lambda^n + o(\lambda^n)$ for some polynomial P_q. Without **(H)**, we would have to deal with several dominating eigenvalues of same modulus which can compensate each other.

Finally let us notice that **(H)** is also considered in [8] for representing real numbers in abstract numeration systems.

We can now state our main two results for unidimensional automatic sequences.

Theorem 2. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and let $x = (x_n)_{n\in\mathbb{N}} \in \Gamma^{\mathbb{N}}$ be an S-automatic sequence generated by a DFAO $\mathcal{A}$ such that* **(H)** *is satisfied. For every $a \in \Gamma$, there exists a bounded function $G_a : L \to \mathbb{R}$ such that*

$$\pi(N, a, x) = N\, G_a(W) + \mathcal{O}\left(\frac{N}{|W|}\right)$$

where $W \in L$ is such that $\mathrm{val}_S(W) = N$. Moreover, if $(W_n)_{n\in\mathbb{N}} \in L^{\mathbb{N}}$ tends to a limit $\omega \in \mathcal{L}_\infty$ then $G_a(W_n)$ also tends to a limit $G_a(\omega) := \lim_{W_n\to\omega} G_a(W_n)$ and the function $\omega \mapsto G(\omega)$ is continuous on $\mathcal{L}_\infty$.

Definition 3. Let $h : \mathbb{R}^+ \to \mathbb{R}^+$ be defined by

$$h(y) = n + \frac{\log y - \log \mathbf{v}_{q_0}(n)}{\log \mathbf{v}_{q_0}(n+1) - \log \mathbf{v}_{q_0}(n)} \quad \text{for } \mathbf{v}_{q_0}(n) \le y \le \mathbf{v}_{q_0}(n+1),\ n \in \mathbb{N}.$$

Roughly speaking, $\{h(\mathrm{val}_S(w))\}$ gives the relative position of w amongst the words of length $|w|$ inside L.

Corollary 4. *With the setting of Theorem 2 and with the function h defined above, one has*

$$\pi(N, a, x) = N\, \mathcal{G}_a(h(N)) + \mathcal{O}\left(\frac{N}{|W|}\right)$$

where $\mathcal{G}_a : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous periodic function of period 1.

Example 5. We use notation of Theorem 2. Consider the numeration system $S = (L, \{a, b, c\}, a < b < c)$ where L is the language over $\{a, b, c\}$ of the words u or uc where $u \in \{a, b\}^*$. The corresponding trim minimal automaton $\mathcal{M}$ is depicted in Figure 1 together with the DFAO $\mathcal{A}$ used to built the S-automatic sequence x. This example is inspired from the one given in [14]. We want an example where the frequency does not exist. Indeed, if the frequency of $a \in \Gamma$ exists then $G_a(\omega) = d(a, x)$ for all $\omega \in \mathcal{L}_\infty$ and the function $\mathcal{G}_a$ is clearly periodic.

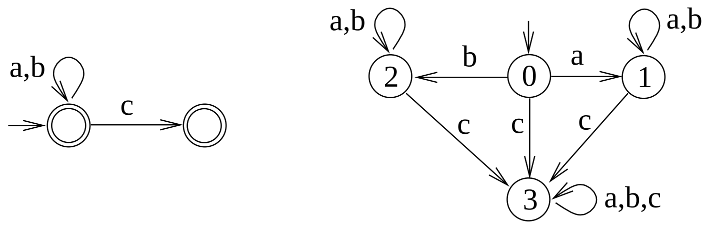

FIGURE 1. A trim minimal automaton $\mathcal{M}$ and a DFAO $\mathcal{A}$.

One can easily check that the product automaton $\mathcal{P}$ has the same structure as $\mathcal{A}$ except that the loop on state 3 has to be removed. Therefore the corresponding matrix P has only 0 and 2 as eigenvalues (both of algebraic multiplicity 2) and **(H)** is satisfied. Feeding $\mathcal{A}$ with the words of L in genealogical ordering:

$$\varepsilon < a < b < c < aa < ab < ac < ba < bb < bc < aaa < aab < aac$$
$$< aba < abb < abc < baa < bab < bac < bba < bbb < bbc < \dots$$

gives the sequence $x = 0123113223113113223223\cdots$. Let $n \geq 2$. For this numeration system,

$$\mathrm{val}_S(a^n) = 3.2^{n-1} - 2 \quad \text{and} \quad \mathrm{val}_S(ba^{n-1}) = 9.2^{n-2} - 2;$$

if $i_n := \mathrm{val}_S(a^n)$ then the factor $x_{i_n} \cdots x_{i_n+3.2^{n-2}} = (113)^{2^{n-2}}2$ and if $j_n := \mathrm{val}_S(ba^{n-1})$ then $x_{j_n} \cdots x_{j_n+3.2^{n-2}} = (223)^{2^{n-2}}1$. Furthermore, we have $\pi(i_2, 1, x) = 1$ and for $n \geq 3$,

$$\pi(i_n, 1, x) = \pi(i_{n-1}, 1, x) + 2^{n-2} \quad \text{and} \quad \pi(j_{n-1}, 1, x) = \pi(i_n, 1, x).$$

It follows easily that $\pi(i_n, 1, x) = 2^{n-1} - 1$ and

$$\lim_{n\to\infty} \frac{\pi(i_n, 1, x)}{i_n} = \frac{1}{3} \text{ but } \lim_{n\to\infty} \frac{\pi(j_n, 1, x)}{j_n} = \lim_{n\to\infty} \frac{\pi(i_{n+1}, 1, x)}{j_n} = \frac{4}{9}.$$

So $d(1, x)$ does not exist. The sequence $(a^n)_{n\in\mathbb{N}}$ (resp. $(ba^n)_{n\in\mathbb{N}}$) converges to $a^\omega \in \mathcal{L}_\infty$ (resp. $ba^\omega \in \mathcal{L}_\infty$) and $G_1(a^n)$ (resp. $G_1(ba^n)$) converges to $G_1(a^\omega) = 1/3$ (resp. $G_1(ba^\omega) = 4/9$). Figure 2 gives an approximation of the graph of G_1, the dash lines have equation $y = 1/3$ and $y = 4/9$ respectively. On the left, we have plotted points $(n, \pi(n, 1, x)/n)$ and on the right, points $(h(n), \pi(n, 1, x)/n)$ with h given in Definition 3. The periodicity

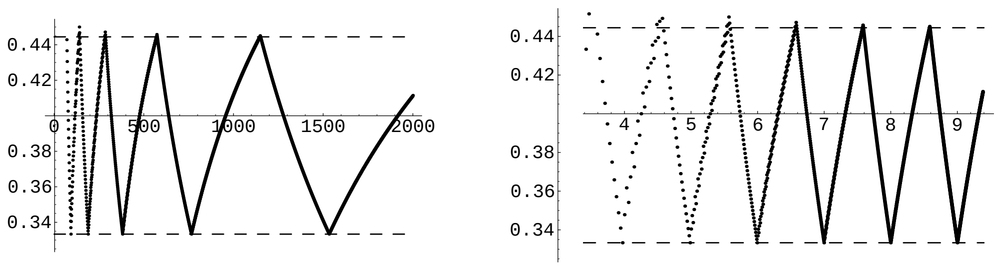

FIGURE 2. Graph of $G_1(W)$ for $\mathrm{val}_S(W) \leq 2000$.

of $\mathcal{G}_1$ follows from the fact that $\mathrm{val}_S(b^{n-1}c) + 1 = \mathrm{val}_S(a^{n+1})$ and thus the sequences $(G_1(b^{n-1}c))_{n\in\mathbb{N}}$ and $(G_1(a^{n+1}))_{n\in\mathbb{N}}$ converge to the same limit 1/3.

Let us now make some comments about the eigenvalues of M and P.

Remark 6. Any eigenvalue of M is also an eigenvalue of P. Indeed, if the vector $\overrightarrow{x}$ of size r is such that $M\overrightarrow{x} = \alpha\overrightarrow{x}$ then by formula (1) the vector $\overrightarrow{x_{(s)}}$ obtained as s consecutive copies of $\overrightarrow{x}$ is such that $P\overrightarrow{x_{(s)}} = \alpha\overrightarrow{x_{(s)}}$. In particular, this shows that the geometric multiplicity of any eigenvalue of M is less than or equal to the corresponding one of P.

Proposition 7. *The spectral radii of M and P are equal.*

Proof. Assume first that M and P are irreducible matrices. It is well-known (see for instance [5, Chap. XIII]) that the spectral radius r_X of a square matrix X of size $n+1$ is given by

$$r_X = \max_{\overrightarrow{y}>0} \min_{0\le i\le n} \frac{(X\overrightarrow{y})_i}{y_i}$$

where notation like $\overrightarrow{y} > 0$ is interpreted component-wise and y_i denotes the ith component of $\overrightarrow{y}$. To stick to our notation introduced earlier, vectors related to M (resp. P) are indexed by states (resp. pairs of states). If $\overrightarrow{y}$ is such a vector then one of its components is denoted by $(\overrightarrow{y})_{q_i}$ or simply y_{q_i} (resp. $(\overrightarrow{y})_{(q_i,q'_j)}$ or simply $y_{(q_i,q'_j)}$).

It is therefore sufficient to show that for any vector $\overrightarrow{c} > 0$ in $\mathbb{R}^{r+1}$, there exists a vector $\overrightarrow{d} > 0$ in $\mathbb{R}^{(r+1)(s+1)}$ such that

$$\min_{q\in Q} \frac{(M\overrightarrow{c})_q}{c_q} \le \min_{(q,q')\in Q\times Q'} \frac{(P\overrightarrow{d})_{(q,q')}}{d_{(q,q')}}, \tag{2}$$

and conversely, that for any vector $\overrightarrow{c} > 0$ in $\mathbb{R}^{(r+1)(s+1)}$, there exists a vector $\overrightarrow{d} > 0$ in $\mathbb{R}^{r+1}$ such that

$$\min_{(q,q')\in Q\times Q'} \frac{(P\overrightarrow{c})_{(q,q')}}{c_{(q,q')}} \le \min_{q\in Q} \frac{(M\overrightarrow{d})_q}{d_q}. \tag{3}$$

Indeed (2) and (3) imply respectively that $r_M \le r_P$ and $r_P \le r_M$.

For a given vector $\overrightarrow{c} > 0$ belonging to $\mathbb{R}^{r+1}$, let us define $\overrightarrow{d} \in \mathbb{R}^{(r+1)(s+1)}$ such that $d_{(q,q')} := c_q$ for all $(q,q') \in Q \times Q'$. With these two vectors and using (1), we have

$$\sum_{k=0}^{r} M_{q,q_k} c_{q_k} = \sum_{k=0}^{r} \left(\sum_{\ell=0}^{s} P_{(q,q'),(q_k,q'_\ell)} \right) c_{q_k} = \sum_{k=0}^{r} \sum_{\ell=0}^{s} P_{(q,q'),(q_k,q'_\ell)} d_{(q_k,q'_\ell)}$$

for any $(q,q') \in Q \times Q'$. This relation implies (2). For the second inequality, given a vector $\overrightarrow{c} > 0$ of $\mathbb{R}^{(r+1)(s+1)}$, we can define a vector $\overrightarrow{d} \in \mathbb{R}^{r+1}$ by setting for all $q \in Q$,

$$d_q := \max_{q'\in Q'} c_{(q,q')}.$$

For such a vector, we have

$$\begin{aligned}\min_{(q,q')\in Q\times Q'} \frac{(P\overrightarrow{c})_{(q,q')}}{c_{(q,q')}} &= \min_{(q,q')\in Q\times Q'} \frac{1}{c_{(q,q')}} \sum_{k=0}^{r}\sum_{\ell=0}^{s} P_{(q,q'),(q_k,q'_\ell)} c_{(q_k,q'_\ell)} \\ &\leq \min_{(q,q')\in Q\times Q'} \frac{1}{c_{(q,q')}} \underbrace{\sum_{k=0}^{r}\left(\sum_{\ell=0}^{s} P_{(q,q'),(q_k,q'_\ell)}\right) d_{q_k}}_{:=\Theta}.\end{aligned}$$

Since Θ does not depend on the state q', the minimum is reached for a state q such that $c_{q,q'} = d_q$ and inequality (3) follows.

If M or P is reducible then the result still holds. If a matrix X is reducible, we can construct a sequence $(X_m)_{m\in\mathbb{N}}$ of irreducible matrices converging to X as m goes to $+\infty$ by replacing zero entries of X with terms of the form α/m, α being a constant. If the property holds for each matrix of the sequence $(X_m)_{m\in\mathbb{N}}$, it also holds for X. When adding such terms α/m to entries of P or M, we just have to be careful that the modified matrices M_m and P_m still satisfy (1). If $P_{(q_i,q'_k),(q_j,q'_\ell)}$ is zero and replaced by $1/m$, then $1/m$ is added to $M_{(q_i,q_j)}$ and to $P_{(q_i,q'_h),(q_j,q'_\ell)}$ for all $h \neq k$. Moreover, if $M_{(q_i,q_j)}$ is zero and replaced by $(s+1)/m$ then $1/m$ is added to $P_{(q_i,q'_k),(q_j,q'_\ell)}$ for all $k,\ell \in \{0,\dots,s\}$. □

As a consequence of Remark 6 and Proposition 7, the following result is then obvious.

Corollary 8. *Under Hypothesis* **(H)**, *M has only one dominating eigenvalue equal to the one of P.*

Remark 9. Without Hypothesis **(H)**, if we assume that M has only one dominating eigenvalue λ then P has naturally λ as eigenvalue but it could also have other eigenvalues of maximal modulus. Indeed, let us consider the following example given by the automata $\mathcal{M}$, $\mathcal{A}$ and $\mathcal{P}$ represented in Figure 3. In this situation, it is easy to show that the golden ratio $\tau = \frac{1+\sqrt{5}}{2}$ is the

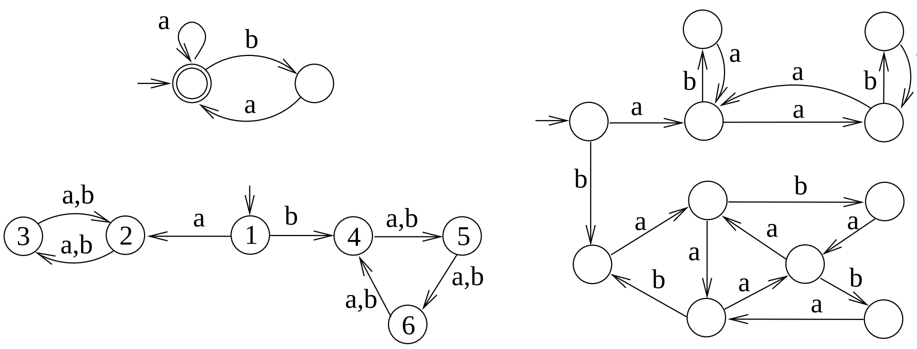

FIGURE 3. The automata $\mathcal{M}$, $\mathcal{A}$ and $\mathcal{P}$.

dominating eigenvalue of $\mathcal{M}$. But P has $\tau, -\tau, \tau e^{2i\pi/3}, \tau e^{4i\pi/3}$ as eigenvalues of modulus τ.

Under Hypothesis **(H)**, for any $q \in Q$, we can write

$$\mathbf{u}_q(n) = P_q(n)\lambda^n + o(\lambda^n)$$

for (possibly zero) polynomials P_q. Moreover, we can assume that P_{q_0} is non-zero and that $\deg P_{q_0} = d \geq 0$. We can therefore split Q into three subsets

$$Q_1 = \{q \mid \mathbf{u}_q(n) = P_q(n)\lambda^n + o(\lambda^n), \deg P_q = d\},$$

$Q_2 = \{q \mid \mathbf{u}_q(n) = P_q(n)\lambda^n + o(\lambda^n), \deg P_q = d-1\}$ and $Q_3 = Q \setminus (Q_1 \cup Q_2)$.

3. Proof of Theorem 2

In what follows, we are only interested in a given letter $a \in \Gamma$. So we do not write this letter in the forthcoming notation. For any $(q, q') \in Q \times Q'$, we define

$$F_{q,q'}(n) := \sum_{\substack{w \in L_q \\ |w|=n}} \mathbf{1}_a(\tau(\delta'(q', w))).$$

Clearly, if $n > 0$ then

$$F_{q,q'}(n) = \sum_{\sigma \in \Sigma} F_{q.\sigma,q'.\sigma}(n-1). \tag{4}$$

We set

$$\overrightarrow{F_{q'}}(n) = (F_{q_0,q'}(n), \ldots, F_{q_r,q'}(n))^T \in \mathbb{N}^{r+1}$$

and

$$\overrightarrow{F}(n) = (\overrightarrow{F_{q_0'}}^T(n), \ldots, \overrightarrow{F_{q_s'}}^T(n))^T \in \mathbb{N}^{(r+1)(s+1)}.$$

From (4), it is obvious that for $n \geq 1$, we have

$$\overrightarrow{F}(n) = P\overrightarrow{F}(n-1)$$

and thus, $\overrightarrow{F}(n) = P^n \overrightarrow{F}(0)$. Moreover, $F_{q,q'}(0) = \mathbf{1}_a(\tau(q'))\,\mathbf{u}_q(0)$. If we set

$$\overrightarrow{\mathbf{u}}(n) = (\mathbf{u}_{q_0}(n), \ldots, \mathbf{u}_{q_r}(n))^T$$

then $\overrightarrow{F_{q'}}(0) = \mathbf{1}_a(\tau(q'))\overrightarrow{\mathbf{u}}(0)$. The matrix P^n can be written as a block matrix

$$P^n = \left(P^{(n)}_{q_k',q_\ell'}\right)_{0 \leq k,\ell \leq s}$$

where each block is a square matrix of size $r+1$. Clearly,

$$\left(P^{(n)}_{q_k',q_\ell'}\right)_{q_i,q_j} := (P^n)_{(q_i,q_k'),(q_j,q_\ell')}$$

counts the number of words w of length n such that $\Delta((q_i, q_k'), w) = (q_j, q_\ell')$ and since $(M^n)_{q_i,q_j}$ is the number of words v of length n such that $\delta(q_i, v) = q_j$ then, as in formula (1),

$$\sum_{\ell=0}^{s} \left(P^{(n)}_{q_k',q_\ell'}\right)_{q_i,q_j} = (M^n)_{q_i,q_j}, \quad 0 \leq i,j \leq r, \quad 0 \leq k \leq s. \tag{5}$$

Let $q' \in Q'$, we have

$$\overrightarrow{F_{q'}}(n) = \sum_{i=0}^{s} P^{(n)}_{q_i',q'}\,\overrightarrow{F_{q_i'}}(0) = \sum_{i=0}^{s} \mathbf{1}_a(\tau(q_i'))P^{(n)}_{q_i',q'}\,\overrightarrow{\mathbf{u}}(0).$$

Therefore, we have obtained the following result.

Lemma 10. *For any state $q' \in Q'$, there exists a constant $C_{q'}$ such that*

$$\overrightarrow{F_{q'}}(n) \leq C_{q'}\, \overrightarrow{\mathbf{u}}(n)$$

where the inequality is interpreted component-wise.

Proof. This is a consequence of our last computation and (5). □

Let $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_t$ be the eigenvalues of P. With Hypothesis **(H)**, we have $\lambda > |\lambda_2| \geq \cdots \geq |\lambda_t|$. Since $\overrightarrow{F}(n) = P^n \overrightarrow{F}(0)$, from the general theory of matrix recurrences [5], we deduce easily that

$$F_{q,q'}(n) = \sum_{\ell=1}^{t} R^{(\ell)}_{q,q'}(n)\lambda_\ell^n$$

for polynomials $R^{(\ell)}_{q,q'}$. If $q \in Q$ is a state such that $\lim_{n\to\infty} \lambda^{-n}\mathbf{u}_q(n) \neq 0$ then $\mathbf{u}_q(n) = P_q(n)\lambda^n + o(\lambda^n)$ and from Lemma 10 we obtain that for such a state q, $\deg R^{(1)}_{q,q'} \leq \deg P_q$ and there exists a real constant $D_{q,q'} \geq 0$ such that

$$F_{q,q'}(n) = D_{q,q'}\mathbf{u}_q(n) + \mathcal{O}(\mathbf{u}_q(n)n^{-1}). \tag{6}$$

Proof of Theorem 2. Let $W = W_1 \cdots W_m$ be a word belonging to L (for all i, $W_i \in \Sigma$), $N = \mathrm{val}_S(W)$ and x be the morphic sequence generated by $S = (L, \Sigma, <)$ and $\mathcal{A}$. We now turn our attention to $\pi(N, a, x)$. One has

$$\begin{aligned}
\pi(N,a,x) &= \sum_{\substack{w\in L\\ w<W}} \mathbf{1}_a(f(w)) = \sum_{\substack{w\in L\\ |w|<|W|}} \mathbf{1}_a(f(w)) + \sum_{\substack{w\in L\\ |w|=|W|, w<W}} \mathbf{1}_a(f(w))\\
&= \mathbf{1}_a(\tau(q_0'))\,\mathbf{u}_{q_0}(0) + \sum_{k=1}^{|W|-1}\sum_{\sigma\in\Sigma}\sum_{\substack{w\in L_{q_0.\sigma}\\ |w|=k-1}} \mathbf{1}_a(\tau(\delta'(q_0'.\sigma, w)))\\
&\quad + \sum_{k=1}^{|W|}\sum_{\substack{\sigma\in\Sigma\\ \sigma<W_k}}\sum_{\substack{w\in L_{q_0.W_1\cdots W_{k-1}\sigma}\\ |w|=|W|-k}} \mathbf{1}_a(\tau(\delta'(q_0'.W_1\cdots W_{k-1}\sigma, w)))\\
&= \mathbf{1}_a(\tau(q_0'))\,\mathbf{u}_{q_0}(0) + \sum_{k=1}^{|W|-1}\sum_{\sigma\in\Sigma} F_{q_0.\sigma, q_0'.\sigma}(k-1)\\
&\quad + \sum_{k=1}^{|W|}\sum_{\substack{\sigma\in\Sigma\\ \sigma<W_k}} F_{q_0.W_1\cdots W_{k-1}\sigma, q_0'.W_1\cdots W_{k-1}\sigma}(|W|-k).
\end{aligned}$$

By introducing two new coefficients, we can replace the summation over the alphabet Σ with a sum over the states. We set

$$\gamma_{q,q'} = \#\{\sigma \in \Sigma \mid \Delta((q_0, q_0'), \sigma) = (q, q')\}, \quad (q, q') \in Q \times Q'$$

and for $(q, q') \in Q \times Q', 1 \leq i \leq |W|$,

$$\beta_{q,q',i}(W) = \#\{\sigma < W_i \mid \Delta((q_0.W_1\cdots W_{i-1}\sigma, q_0'.W_1\cdots W_{i-1}\sigma) = (q, q')\}.$$

Therefore, we obtain

$$\begin{aligned}\pi(N,a,x) &= \mathbf{1}_a(\tau(q_0'))\,\mathbf{u}_{q_0}(0)+\sum_{k=1}^{|W|-1}\sum_{(q,q')\in Q\times Q'}\gamma_{q,q'}F_{q,q'}(k-1)\\ &\quad+\sum_{k=1}^{|W|}\sum_{(q,q')\in Q\times Q'}\beta_{q,q',k}(W)F_{q,q'}(|W|-k)\\ &= \mathbf{1}_a(\tau(q_0'))\,\mathbf{u}_{q_0}(0)\\ &\quad+\sum_{(q,q')\in Q\times Q'}\Bigg(\sum_{k=2}^{|W|}(\gamma_{q,q'}+\beta_{q,q',k}(W))F_{q,q'}(|W|-k)\\ &\quad+\beta_{q,q',1}(W)F_{q,q'}(|W|-1)\Bigg).\end{aligned}$$

Let us set

$$\alpha_{q,q',i}(W)=\beta_{q,q',i}(W)+(1-\mathbf{1}_1(i))\gamma_{q,q'},\quad 1\le i\le |W|$$

so,

$$\pi(N,a,x)=\mathbf{1}_a(\tau(q_0'))\,\mathbf{u}_{q_0}(0)+\sum_{(q,q')\in Q\times Q'}\underbrace{\sum_{k=1}^{|W|}\alpha_{q,q',k}(W)\,F_{q,q'}(|W|-k)}_{:=S_{q,q'}}.$$

We need some asymptotic information about $F_{q,q'}(N-n)$. We proceed exactly as in [6] and for the sake of completeness we recall the main facts of this paper. We introduce an increasing continuous function $g:\mathbb{R}^+\to\mathbb{R}^+$ by

$$g(n+x)=\mathbf{v}_{q_0}(n)^{1-x}\mathbf{v}_{q_0}(n+1)^x \quad \text{for } 0\le x\le 1 \quad \text{and } n\in\mathbb{N}. \tag{7}$$

This function has the property $g(n)=\mathbf{v}_{q_0}(n)$ for all $n\in\mathbb{N}$. Notice that the inverse function of g is the function $h:\mathbb{R}^+\to\mathbb{R}^+$ introduced in Definition 3. By Hypothesis **(H)** $\mathbf{v}_{q_0}(n)=\sum_{\ell=0}^n\mathbf{u}_{q_0}(\ell)$ can be written as $\mathbf{v}_{q_0}(n)=T(n)\lambda^n+o(\lambda^n)$ for a polynomial T of degree $d=\deg P_{q_0}$. Furthermore, we have

$$\frac{g(n+x)}{g(n)}=\lambda^x\left(1+\frac{xd}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)\qquad \text{for } x\in\mathbb{R} \tag{8}$$

and $\lim_{x\to\infty}\frac{g(x)}{T(x)\lambda^x}=1$. From all this and (6), we compute the following asymptotic expansion

$$\frac{F_{q,q'}(N-n)}{g(N)}= \tag{9}$$
$$\begin{cases}\lambda^{-n}\left(f_q-df_q\frac{n}{N}+\frac{g_q}{N}+\mathcal{O}\left(\frac{n^2}{N^2}\right)\right) & \text{for } q\in Q_1 \quad \text{and } n=o(\sqrt{N})\\ \lambda^{-n}\left(\frac{f_q}{N}+\mathcal{O}\left(\frac{n}{N^2}\right)\right) & \text{for } q\in Q_2 \quad \text{and } n=o(N)\\ \mathcal{O}\left(\frac{1}{N^2\lambda^n}\right) & \text{for } q\in Q_3 \quad \text{and } n\le N\end{cases}$$

where f_q and g_q can be computed from the two leading coefficients of the polynomials T, P_q and $R_q^{(1)}$. (Actually, we have the same kind of asymptotic expansion for $F_{q,q'}$ and $\mathbf{u}_q$.) Having at our disposal all the needed asymptotic information, we can consider again $S_{q,q'}$. Technically, we would have to split summation at indices of order $o(\sqrt{|W|})$ and $o(|W|)$, but we will omit this, since the contribution to the error term is negligible compared to the other error terms.

If $q \in Q_1$, then

$$S_{q,q'} = g(|W|) f_q \sum_{k=1}^{|W|} \alpha_{q,q',k}(W) \lambda^{-k} + \mathcal{O}\Big(\frac{g(|W|)}{|W|}\Big).$$

If q belongs to $Q_2 \cup Q_3$ then

$$S_{q,q'} = \mathcal{O}\Big(\frac{g(|W|)}{|W|}\Big).$$

If we set

$$\Psi(W) = \sum_{\substack{(q,q')\in Q\times Q' \\ q\in Q_1}} f_q \sum_{k=1}^{|W|} \alpha_{q,q',k}(W) \lambda^{-k},$$

then

$$\pi(N,a,x) = g(|W|)\Psi(W) + \mathcal{O}\Big(\frac{g(|W|)}{|W|}\Big)$$

where we recall that $\mathrm{val}_S(W) = N$. Notice that since the $\alpha_{q,q',k}$'s are bounded, the function Ψ extends to a continuous function on $\mathcal{L}_\infty$ by $\Psi(\omega) = \lim_{W\to\omega} \Psi(W)$. To conclude the proof, one can proceed exactly as in [6]. Since the function g depends only on the abstract system S, we have

$$\frac{\mathrm{val}_S(W)}{g(|W|)} = Y(W) + \frac{1}{|W|} Z(W) + \mathcal{O}(|W|^{-2})$$

with the same functions Y and Z as in [6]. To obtain the expected result, we set $G(W) = \Psi(W)/Y(W)$. □

4. Proof of Corollary 4

First we introduce essential words, then we give the proof of the corollary and finally we give an example illustrating the concepts involved in the proof. Consider all the strongly connected components $C_1, \ldots, C_k$ of $\mathcal{M}$. To each C_j, $j = 1, \ldots, k$, corresponds an irreducible matrix M_j of dominating eigenvalue $\lambda_j \geq 1$. (We say that λ_j is the dominating eigenvalue of C_j.) Thanks to Corollary 8, if $\lambda_j = \lambda$, then λ_j is the unique eigenvalue of M_j of modulus λ and the matrix M_j is primitive. Otherwise, we would have other eigenvalues of modulus λ which contradict our Hypothesis **(H)**.

A path in $\mathcal{M}$ is *essential* if starting from the initial state q_0 it goes through a maximal number α of strongly connected components having $\lambda > 1$ as dominating eigenvalue. Since $\mathcal{M}$ is trim, it is clear that

$$\mathbf{u}_{q_0}(n) \asymp n^{\alpha-1}\lambda^n \quad \text{and} \quad \mathbf{v}_{q_0}(n) \asymp n^{\alpha-1}\lambda^n.$$

A word $w \in \Sigma^*$ is *essential* if the path $\mathfrak{p}_w$ in $\mathcal{M}$ starting from the initial state and corresponding to the reading of w is the prefix of an essential path $\mathfrak{e}$ and if $\mathfrak{p}_w$ ends before or inside the first strongly connected component encountered on $\mathfrak{c}$ having λ as dominating eigenvalue. Otherwise the word is said to be *inessential*. Consequently, if a word x is inessential then

$$\lim_{n\to\infty} \frac{\mathbf{u}_{q_0.x}(n)}{n^{\alpha-1}\lambda^n} = 0. \tag{10}$$

Furthermore, if u is inessential, so is uv, for any $u, v \in \Sigma^*$.

Remark 11. Let v be an essential word and $\widetilde{v}$ be a prefix of infinitely many words in L. Then there exist words z and $\widetilde{z}$ such that $|vz| = |\widetilde{v}\widetilde{z}|$ and $vz, \widetilde{v}\widetilde{z} \in L$. Indeed, if the matrix associated with a strongly connected component is primitive, then there exists N_0 such that for all $n \geq N_0$, there exists a path of length n connecting any two states of the component [9, Theorem 4.5.8]. This result is enough to obtain a suffix z having the expected properties for any long enough suffix $\widetilde{z}$.

Proof of Corollary 4. From Theorem 2, we know that

$$\pi(N, a, x) = NG_a(W) + \mathcal{O}\left(\frac{N}{|W|}\right)$$

where $\mathrm{val}_S(W) = N$. Again, the proof is based on the one in [6] but here, since we do not have additive functions, we will have to consider other inequalities. Let $\omega \in \mathcal{L}_\infty$ be such that the sequence $(v_k)_{k\in\mathbb{N}} \in L^{\mathbb{N}}$ converges to ω. We denote by $\mathrm{val}_\infty(\omega)$ the real number represented by the infinite word ω (see [8] for details),

$$\mathrm{val}_\infty(\omega) = \lim_{k\to\infty} \frac{\mathrm{val}_S(v_k)}{g(|v_k|)}.$$

First we prove that $G_a(\omega) := \lim_{W\to\omega} G_a(W)$ does not depend on $\omega \in \mathcal{L}_\infty$ but depends only on $\mathrm{val}_\infty(\omega)$. Let $(v_k)_{k\in\mathbb{N}}$ and $(\widetilde{v_k})_{k\in\mathbb{N}}$ be two sequences of words in L converging respectively to ω and $\widetilde{\omega}$ such that $\mathrm{val}_\infty(\omega) = \mathrm{val}_\infty(\widetilde{\omega})$.

Assume first that infinitely many words in at least one of the sequences $(v_k)_{k\in\mathbb{N}}$ or $(\widetilde{v_k})_{k\in\mathbb{N}}$ are essential. Thanks to Remark 11 we may furthermore assume that $|v_k| = |\widetilde{v_k}|$ for all $k \geq 0$. We have

$$\begin{aligned} \mathrm{val}_S(\widetilde{v_k})G_a(\widetilde{v_k}) - \mathrm{val}_S(v_k)G_a(v_k) + \mathcal{O}\left(\frac{\mathrm{val}_S(v_k) + \mathrm{val}_S(\widetilde{v_k})}{|v_k|}\right) \\ = \sum_{v_k \leq w < \widetilde{v_k}} \mathbf{1}_a(f(w)) \leq \mathrm{val}_S(\widetilde{v_k}) - \mathrm{val}_S(v_k) \end{aligned} \tag{11}$$

and dividing this by $g(|v_k|)$ and letting k tends to ∞ we get $G_a(\omega) = G_a(\widetilde{\omega})$.

Assume now that ω has a shortest prefix y of length $\ell \geq 1$ which is inessential (therefore any prefix of ω longer than ℓ is inessential and only finitely many elements in $(v_k)_{k\in\mathbb{N}}$ are essential). Consider the lexicographical ordering of $\mathcal{L}_\infty$. Let ω' (resp. ω'') be the largest (resp. smallest) infinite word in $\mathcal{L}_\infty$ whose prefixes are essential and which is less (resp. greater) than ω. At least one of the two words ω' or ω'' exists. Assume that ω' exists, the arguments are similar for ω''. Let $T_n(\omega) = \mathrm{Pref}_n\{z \in \mathcal{L}_\infty \mid \omega' < z < \omega\}$, where we denote by $\mathrm{Pref}_n(X)$ the set of prefixes of length n of the words in the set X. From the definition of y, $\mathrm{Pref}_{\ell-1}(\omega') = \mathrm{Pref}_{\ell-1}(\omega)$. Obviously, for $n \geq \ell$, any element in $T_n(\omega)$ distinct from $\mathrm{Pref}_n(\omega')$ is inessential. Consequently, any word $w = ps$ in $T_n(\omega)$ having a shortest inessential prefix p

of length $k+1$ has a common prefix of length k with ω', $\ell \le k+1 \le n$ and thanks to (10), the number of the admissible suffixes s of length $n-k-1$ is in $\mathcal{O}((n-k-1)^{\alpha-2}\lambda^{n-k-1})$. Therefore, there exists a constant K such that

$$\#T_n(\omega) \le K \sum_{i=0}^{n-\ell} i^{\alpha-2}\lambda^i \quad \text{and} \quad \lim_{n\to\infty} \frac{\#T_n(\omega)}{n^{\alpha-1}\lambda^n} = 0.$$

From [8], we have $\mathrm{val}_\infty(\omega) = \mathrm{val}_\infty(\omega')$. Let w in $L \cap \Sigma^n$ having y as prefix. With the same arguments about the asymptotic behaviour of $\#T_n(\omega)$, if n tends to infinity then $|\pi(\mathrm{val}_S(w), a, x)/\mathrm{val}_S(w) - G_a(\omega')|$ tends to zero. Therefore, one can replace ω with ω' and consider the first case.

To show that $\mathcal{G}_a(\log_\lambda \mathrm{val}_\infty(\omega)) := F(\omega)$ can be written as a function of $\{h(N)\}$, the Lipschitz continuity of $\mathcal{G}_a$ and its periodicity, one can proceed as in [6] using inequality (11). □

Example 12. We focus here on essential words, other computational details are mainly the same as in Example 5. Clearly words over $\{a,b\}^*$ are essential

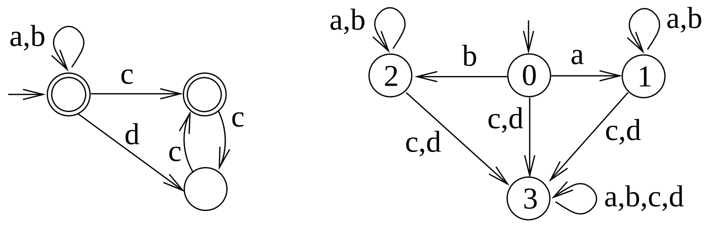

FIGURE 4. A trim minimal automaton $\mathcal{M}$ and a DFAO $\mathcal{A}$.

but not words containing c or d. The words $x = ac$ and $y = ad$ are inessential and it is obvious that the length of any word in L having x (resp. y) as prefix is even (resp. odd). So it is not possible to obtain a property similar to the one given in Remark 11 and to consider two sequences of words $(xv_k)_{k\in\mathbb{N}}$ and $(y\widetilde{v_k})_{k\in\mathbb{N}}$ in L such that $|xv_k| = |y\widetilde{v_k}|$ for all k. For this numeration system,

$$\mathrm{val}_S(a^n) = 2^{n+1}-n-2,\ \mathrm{val}_S(ab^{n-1}) = 3.2^n-2n-3,\ \mathrm{val}_S(ba^{n-1}) = 3.2^n-n-3$$

and for n even, $\mathrm{val}_S(ac^{n-1}) = 3.2^n - n - 4 = \mathrm{val}_S(adc^{n-2})$. We are interested in $\pi(N,1,x)$. Amongst the words of length $i \ge 1$, 2^{i-1} words belong to $a\{a,b\}^*$ (only these words contribute to the letter 1), so

$$G_1(a^\omega) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n-1} 2^{i-1}}{\mathrm{val}_S(a^n)} = \frac{1}{4}.$$

In the same way, between a^n and ab^{n-1}, 2^{n-1} words belong to $a\{a,b\}^*$ and the same constatation holds for the words between a^n and ba^{n-1}, so

$$G_1(ab^\omega) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n-1} 2^{i-1} + 2^{n-1}}{\mathrm{val}_S(ab^{n-1})} = \frac{1}{3} = G_1(ba^\omega).$$

These computations show once again that the frequency of 1 does not exist. Let us now illustrate the case of inessential prefixes. For $N_n = \mathrm{val}_S(ac^{2n-1})$, we will show that

$$\lim_{n\to\infty} \frac{\pi(N_n, 1, x)}{N_n} = \frac{1}{3} = G_1(ab^\omega) = G_1(ba^\omega).$$

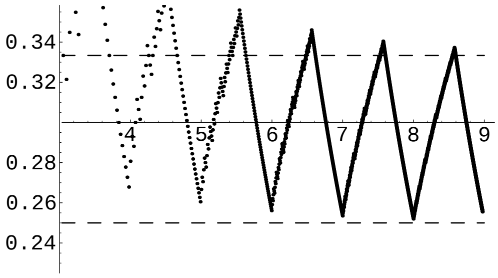

FIGURE 5. Graph of $G_1(W)$ for $\mathrm{val}_S(W) \le 5084$.

Indeed, with the notation of the previous proof, if $\omega = ac^\omega$ then $\omega' = ab^\omega$. The number of words in L between ab^{2n-1} and ac^{2n-1} is exactly $\#T_n(\omega) = 2n$. Then, there exists $C_n \in \mathcal{O}(\#T_n(\omega))$ such that

$$\begin{aligned}\frac{\pi(N_n,1,x)}{N_n} &= \frac{\pi(\mathrm{val}_S(ab^{2n-1}),1,x)}{N_n} + \frac{C_n}{N_n}\\ &= \underbrace{\frac{\pi(\mathrm{val}_S(ab^{2n-1}),1,x)}{\mathrm{val}_S(ab^{2n-1})}}_{\to G_1(ab^\omega)} \underbrace{\frac{\mathrm{val}_S(ab^{2n-1})}{N_n}}_{\to 1} + \underbrace{\frac{C_n}{N_n}}_{\to 0}.\end{aligned}$$

We can therefore replace ω with ω' or ω'' when dealing with G_a, $a \in \Gamma$.

5. Frequency of multidimensional sequences

For the sake of simplicity, we restrict mainly ourselves to the case of bidimensional sequences. Let $(x_{i,j})_{i,j\in\mathbb{N}}$ be a bidimensional sequence over Γ. If $a \in \Gamma$, we denote the function counting the number of a's by

$$\pi_2(n,a,x) = \#\{(i,j) \in [0,n-1] \times [0,n-1] \mid x_{i,j} = a\}.$$

Multidimensional automatic sequences have been considered in [15] and are also presented in [1]. Generalization to abstract numeration systems are considered in [13]. If $S = (L,\Sigma,<)$ is an abstract numeration system, then we consider the alphabet $\Sigma_\$ = \Sigma \cup \{\$\}$ where the symbol \$ does not belong to Σ. If x and y are two words over Σ then we define

$$(u,v)^\$:= \begin{cases} (\$^{|v|-|u|}u, v), & \text{if } |u| \le |v|;\\ (u, \$^{|u|-|v|}v), & \text{if } |u| > |v|.\end{cases}$$

If $\mathcal{A} = (Q', q_0', \Sigma_\$ \times \Sigma_\$, \delta', \Gamma, \tau)$ is a DFAO over the alphabet $\Sigma_\$ \times \Sigma_\$$ then the element $x_{i,j}$ of the bidimensional S-automatic sequence generated by $\mathcal{A}$ is given by

$$\tau(\delta'(q_0', [\mathrm{val}_S^{-1}(i), \mathrm{val}_S^{-1}(j)]^\$)).$$

With this definition, it is therefore quite natural to be interested in the following limit

$$\lim_{n\to\infty} \frac{\pi_2(n,a,x)}{n^2}. \tag{12}$$

To obtain such kind of information, we first show how the sequence $(x_{i,j})_{i,j\in\mathbb{N}}$ can be roughly seen as a unidimensional one.

If L is a regular language over the alphabet $\Sigma = \{\sigma_1 < \cdots < \sigma_k\}$ then the language

$$L^{(2)} = \{(u,v)^{\$} \mid u, v \in L\}$$

is a regular language over $\Sigma_{\$} \times \Sigma_{\$}$. Indeed, since the set of regular languages is closed under inverse morphism, intersection and complementation [3] then the language

$$L^{(2)} = \left[p_1^{-1}(\$^* L) \cap p_2^{-1}(\$^* L) \cap (\Sigma_{\$} \times \Sigma_{\$})^*\right] \setminus (\$,\$)(\Sigma_{\$} \times \Sigma_{\$})^*$$

is regular, where

$$p_1 : \Sigma_{\$}^* \times \Sigma_{\$}^* \to \Sigma_{\$}^* : (u,v) \mapsto u \text{ and } p_2 : \Sigma_{\$}^* \times \Sigma_{\$}^* \to \Sigma_{\$}^* : (u,v) \mapsto v$$

are the canonical projection morphisms. If we assume that $\$$ is less than σ for all $\sigma \in \Sigma$ then the alphabet $\Sigma_{\$} \times \Sigma_{\$}$ can be lexicographically ordered using the total ordering $\$ < \sigma_1 < \cdots < \sigma_k$ of $\Sigma_{\$}$:

$$(\$,\$) < (\$,\sigma_1) < \cdots < (\$,\sigma_k) < (\sigma_1,\$) < \cdots < (\sigma_k,\sigma_k).$$

Using this ordering of $\Sigma_{\$} \times \Sigma_{\$}$, the words of $L^{(2)}$ can be genealogically ordered. Let $(i,j) \in \mathbb{N}^2$ and $u, v \in L$ be such that $\mathrm{val}_S(u) = i$ and $\mathrm{val}_S(v) = j$. We denote by

$$\rho_{L^{(2)}}(i,j),$$

or simply $\rho(i,j)$ if the context is clear, the position of the word $(u,v)^{\$}$ within the genealogically ordered language $L^{(2)}$. (Remember that positions are counted from zero.) The reader familiar with the Peano function (see for instance [16]) will not be surprised by our developments. The function $\rho(i,j)$ is just another way of enumerating the elements of $\mathbb{N}^2$.

In the following, we will be interested only in the number of words in language L, so we simply write $\mathbf{u}(n)$ and $\mathbf{v}(n)$ instead of $\mathbf{u}_{q_0}(n)$ and $\mathbf{v}_{q_0}(n)$ respectively.

Example 13. Consider the language $L = \{b, ab\}^*\{a, \varepsilon\}$ of the words which do not contain the factor aa. The first words of the language are

0	1	2	3	4	5	6	7	8	9	10	11	
ε	a	b	ab	ba	bb	aba	abb	bab	bba	bbb	$abab$	$\ldots$

The following table lists the first value of $\rho(i,j)$.

i	j $\rightarrow$										
$\downarrow$	0	1	2	3	4	5	6	7	8	9	10
0	**0**	1	2	9	10	11	36	37	38	39	40
$\mathbf{v}(0) = 1$	3	4	5	12	13	14	41	42	43	44	45
2	6	7	**8**	15	16	17	46	47	48	49	50
$\mathbf{v}(1) = 3$	18	19	20	21	22	23	51	52	53	54	55
4	24	25	26	27	28	29	56	57	58	59	60
5	30	31	32	33	34	**35**	61	62	63	64	65
$\mathbf{v}(2) = 6$	66	67	68	69	70	71	72	73	74	75	76
7	77	78	79	80	81	82	83	84	85	86	87
8	88	89	90	91	92	93	94	95	96	97	98
9	99	100	101	102	103	104	105	106	107	108	109
10	110	111	112	113	114	115	116	117	118	119	**120**

Indeed, we have in the ordered language $L^{(2)}$

$$(\varepsilon,\varepsilon) < (\$,a) < (\$,b) < (a,\$) < (a,a) < (a,b) < (b,\$) < (b,a) < (b,b) < (\$\$,aa) < \cdots .$$

Enumerating the pairs (i,j) with increasing values of the function ρ coincides with enumerating the words of $L^{(2)}$ in genealogical ordering. Since the number of words $(u,v)^\$$ of length exactly n in $L^{(2)}$ is

$$\underbrace{2\mathbf{u}(n)\mathbf{v}(n-1)}_{|u|\neq|v|}+\underbrace{\mathbf{u}(n)^2}_{|u|=|v|}=\mathbf{v}(n)^2-\mathbf{v}(n-1)^2$$

then

$$\{(i,j)\mid \rho(i,j)<\mathbf{v}(n)^2\}=[0,\mathbf{v}(n)-1]\times[0,\mathbf{v}(n)-1]. \tag{13}$$

It is an easy exercise to obtain a formula for computing $\rho(i,j)$. Let us set $M=\max(i,j)$. There exists a unique integer n such that

$$\mathbf{v}(n)\le M<\mathbf{v}(n+1)$$

and

$$\rho(i,j)=\begin{cases}\mathbf{v}(n)^2+i\,\mathbf{u}(n+1)+j-\mathbf{v}(n) & \text{, if } i<\mathbf{v}(n);\\ \mathbf{v}(n)^2+\mathbf{v}(n)\mathbf{u}(n+1)+(i-\mathbf{v}(n))\,\mathbf{v}(n+1)+j & \text{, if } i\ge\mathbf{v}(n).\end{cases}$$

Conversely, let us define $\kappa_1:\mathbb{N}\to\mathbb{N}$ and $\kappa_2:\mathbb{N}\to\mathbb{N}$ such that, for all $k\in\mathbb{N}$ $\rho(\kappa_1(k),\kappa_2(k))=k$. For each k, there exists a unique n such that

$$\mathbf{v}(n)^2\le k<\mathbf{v}(n+1)^2$$

and we set $T_n=\mathbf{v}(n)^2+\mathbf{v}(n)\mathbf{u}(n+1)$. If $k\ge T_n$ then

$$k-T_n=i\,\mathbf{v}(n+1)+j,\quad \text{with } 0\le j<\mathbf{v}(n+1),$$

$\kappa_1(k)=i+\mathbf{v}(n)$ and $\kappa_2(k)=j$. Otherwise $k<T_n$ and we have

$$k-\mathbf{v}(n)^2=i\,\mathbf{u}(n+1)+j,\quad \text{with } 0\le j<\mathbf{u}(n+1),$$

$\kappa_1(k)=i$ and $\kappa_2(k)=j+\mathbf{v}(n)$.

Remark 14. In the case of an m-dimensional sequence, $m\ge 2$, we can define in the same way a language $L^{(m)}$ over the totally ordered alphabet $(\Sigma_\$)^m$ and a function $\rho_{L^{(m)}}:\mathbb{N}^m\to\mathbb{N}$ counting positions of the elements of $L^{(m)}$. Due to the genealogical ordering of this latter language, the formula obtained previously can be extended as follows. There exists a unique n such that $\mathbf{v}(n)\le\max(i_1,\ldots,i_m)<\mathbf{v}(n+1)$. If $\max(i_1,\ldots,i_{m-1})<\mathbf{v}(n)$ then

$$\rho_{L^{(m)}}(i_1,\ldots,i_m)=\mathbf{v}(n)^m+\rho_{L^{(m-1)}}(i_1,\ldots,i_{m-1})\,\mathbf{u}(n+1)+i_m-\mathbf{v}(n).$$

Otherwise,

$$\begin{aligned}\rho_{L^{(m)}}(i_1,\ldots,i_m) &= \mathbf{v}(n)^m+\mathbf{v}(n)^{m-1}\,\mathbf{u}(n+1)\\ &\quad+\left(\rho_{L^{(m-1)}}(i_1,\ldots,i_{m-1})-\mathbf{v}(n)^{m-1}\right)\,\mathbf{v}(n+1)+i_m.\end{aligned}$$

Let us consider back case $m=2$ and set

$$W_2(n,x,a)=\#\{(i,j)\in\mathbb{N}^2\mid \rho(i,j)<n \text{ and } x_{i,j}=a\}.$$

In view of (13), we clearly have

$$W_2(\mathbf{v}(t)^2,x,a)=\pi_2(\mathbf{v}(t),x,a) \tag{14}$$

for any $t \in \mathbb{N}$. The sequence $y = (x_{\kappa_1(n),\kappa_2(n)})_{n\in\mathbb{N}}$ is a unidimensional automatic sequence generated by the regular language $L^{(2)}$ and the DFAO $\mathcal{A}$. If the matrix P associated with $L^{(2)}$ and $\mathcal{A}$ is such that **(H)** is satisfied then the limit

$$\lim_{n\to\infty} \frac{W_2(n,a,y)}{n}$$

exists and is denoted by $\mathbf{d}$. Our aim is now to show (under some hypotheses) that the sequence $(\pi_2(n,a,x)/n^2)_{n\in\mathbb{N}}$ is converging. If it is converging, it converges to the same limit $\mathbf{d}$. Indeed, from (14), we know that a subsequence of this sequence is converging to $\mathbf{d}$. Since a and x are given, we will omit them in notation $\pi_2(n)$.

To obtain the convergence of $(\pi_2(n)/n^2)_{n\in\mathbb{N}}$, we shall assume in what follows that

- the incidence matrix P of the product automaton $\mathcal{P}$ constructed on the minimal automaton of $L^{(2)}$ and on the DFAO $\mathcal{A}$ is primitive (in particular, notice that **(H)** will therefore be satisfied),
- the language L is a *prefix language*, i.e., if $w\sigma \in L$, $w \in \Sigma^*$, $\sigma \in \Sigma$ then $w \in L$.

Our task is now to present the right setting in which the same kind of construction as in [11] can be applied. We define a function $\mathcal{S} : L\times L \to 2^{L\times L}$ in the following way. Let $(u, v) \in L \times L$.

- If $u, v \neq \varepsilon$ then

$$\mathcal{S}(u, v) = \{(u\sigma, v\tau) \in L \times L \mid \sigma, \tau \in \Sigma\},$$

- if $u = \varepsilon$, $v \neq \varepsilon$ then

$$\mathcal{S}(u, v) = \{(\sigma, v\tau) \in L \times L \mid \sigma, \tau \in \Sigma\} \cup \{(\varepsilon, v\tau) \in L \times L \mid \tau \in \Sigma\},$$

- if $u \neq \varepsilon$, $v = \varepsilon$ then

$$\mathcal{S}(u, v) = \{(u\sigma, \tau) \in L \times L \mid \sigma, \tau \in \Sigma\} \cup \{(u\sigma, \varepsilon) \in L \times L \mid \sigma \in \Sigma\},$$

- $u, v = \varepsilon$ then

$$\mathcal{S}(u, v) = \{(\sigma, \tau) \in L \times L \mid \sigma, \tau \in \Sigma \cup \{\varepsilon\}\}.$$

This function can naturally be extended to $2^{L\times L}$. Using $\mathcal{S}$, for any $j \in \mathbb{N}$, any finite set of $\mathbb{N}^2$ can be included into $\mathcal{S}^j(E)$ for some minimal set $E \in 2^{L\times L}$ (we have a one-to-one correspondence between $\mathbb{N}^2$ and $L \times L$ through the use of the numeration system). This fact is a consequence of the definition of $\mathcal{S}$ and is enlightened by the following example. Moreover, since L is a prefix language, for each $(u, v) \in L \times L$ there exists a unique $(x, y) \in L \times L$ such that $(u, v) \in \mathcal{S}(x, y)$.

Example 15. Consider the same language as in Example 13. We represent a partition of $\mathbb{N}^2 = L \times L$ in terms of sets of the form $\mathcal{S}(u,v)$,

u	v	$\rightarrow$							
$\downarrow$	ε	a	b	ab	ba	bb	aba	abb	$\cdots$
ε	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
a	$\cdot$	$\mathcal{S}(\varepsilon,\varepsilon)$	$\cdot$	$\mathcal{S}(\varepsilon,a)$	$\mathcal{S}(\varepsilon,b)$	$\cdot$	$\mathcal{S}(\varepsilon,ab)$	$\cdot$	
b	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
ab	$\cdot$	$\mathcal{S}(a,\varepsilon)$	$\cdot$	$\mathcal{S}(a,a)$	$\mathcal{S}(a,b)$	$\cdot$	$\mathcal{S}(a,ab)$	$\cdot$	
ba	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
bb	$\cdot$	$\mathcal{S}(b,\varepsilon)$	$\cdot$	$\mathcal{S}(b,a)$	$\mathcal{S}(b,b)$	$\cdot$	$\mathcal{S}(b,ab)$	$\cdot$	
aba	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	
abb	$\cdot$	$\mathcal{S}(ab,\varepsilon)$	$\cdot$	$\mathcal{S}(ab,a)$	$\mathcal{S}(ab,b)$	$\cdot$	$\mathcal{S}(ab,ab)$	$\cdot$	
$\vdots$	$\vdots$								$\ddots$

For instance, if we want a partition with sets of the form $\mathcal{S}^2(u,v)$, then we have $\mathcal{S}^2(\varepsilon,\varepsilon) = (\{\varepsilon, a, \ldots, bb\} \cap L)^2$, $\mathcal{S}^2(a,\varepsilon) = \mathcal{S}(ab,\varepsilon) \cup \mathcal{S}(ab,a) \cup \mathcal{S}(ab,b)$ and $\mathcal{S}^2(a,a) = \mathcal{S}(ab,ab)$. In particular, it is easy to show that for $u, v \neq \varepsilon$, we have

$$\mathcal{S}^n(u,v) = \{(uw_1, vw_2) \in L \times L \mid w_1, w_2 \in \Sigma^n\},\ \forall n \in \mathbb{N}.$$

Remark 16. Since L is a prefix language, then it is also the case for $L^{(2)}$ and all states of the trim minimal automaton of $L^{(2)}$ are therefore final. Therefore, we avoid words "without output".

Assume that $\mathcal{P}$ has t states denoted $\{q_1, \ldots, q_t\}$ with q_1 as initial state. Let $(i,j) \in \mathbb{N}^2$. To this pair corresponds a pair $w = (\mathrm{val}_S^{-1}(i), \mathrm{val}_S^{-1}(j))^{\$}$ and therefore a state of $\mathcal{P}$ obtained by reading w in $\mathcal{P}$ starting from q_1. This state is denoted $q_{(i,j)}$. Let E be a finite set of $\mathbb{N}^2$. To this set corresponds a row vector χ_E of size t such that its kth component is given by

$$(\chi_E)_k = \#\{(i,j) \in E \mid q_{(i,j)} = q_k\}.$$

Using Remark 16, the following relations are therefore obvious

$$\chi_{\mathcal{S}(u,v)} = \chi_{\{(u,v)\}} P \quad \text{and} \quad \chi_{\mathcal{S}^n(u,v)} = \chi_{\{(u,v)\}} P^n,\ \forall n.$$

Since P is primitive, there exists a unique eigenvector ξ such that the sum of its components is 1, having the following property

$$\forall (u,v) \in L \times L : \ \frac{\chi_{\mathcal{S}^n(u,v)}}{\#\mathcal{S}^n(u,v)} \rightarrow \xi, \text{ if } n \rightarrow \infty.$$

Let E be a finite set of $\mathbb{N}^2$. If we denote by ∂E the elements in E having a neighbour not in E, then with the same reasoning as in [11], we can obtain that

$$\left| \frac{\chi_E}{\#E} - \xi \right| \leq C \left(\frac{\#\partial E}{\#E} \right)^h$$

for some positive constants C and h. Consequently, the limit (12) exists and equals the algebraic number $\mathbf{d}$. Indeed, consider $E = [0, n-1] \times [0, n-1]$ and in the vector χ_E sum together all the components corresponding to states with a same output in the DFAO $\mathcal{A}$. Doing this, you will replace the frequency vector of the states reached in $\mathcal{P}$ with the frequency vector of the letters of output alphabet Γ.

6. Remark on the ordering

Let us make some comments about the independence of the frequency with respect to the total ordering of the alphabet.

Remark 17. Assume that an automatic sequence x is generated by a numeration system $S = (L, \Sigma, <)$ and a DFAO $\mathcal{A}$ satisfying assumption **(H)**. If we consider another total ordering of Σ, say $\prec$, then obviously **(H)** is still satisfied. Indeed, the ordering does not appear in the definition of the incidence matrices. Considering the numeration system $S = (L, \Sigma, \prec)$ instead of $S = (L, \Sigma, <)$ affects the sequence x by permuting the elements of x within the range $\{\mathbf{v}(n), \ldots, \mathbf{v}(n) + \mathbf{u}(n+1) - 1\}$. Considering the automata $\mathcal{M}$ and $\mathcal{A}$ given in Figure 3, if $a < b$ the first words of L are

$$\varepsilon < a < aa < ba < aaa < aba < baa < aaaa < aaba < abaa < baaa < baba$$

giving rise to the output sequence

$$x_< = 1|2|35|226|33344| \cdots .$$

If $b \prec a$ then

$$\varepsilon \prec a \prec ba \prec aa \prec baa \prec aba \prec aaa \prec baba \prec baaa \prec abaa \prec aaba \prec aaaa$$

giving rise to the output sequence

$$x_\prec = 1|2|53|622|44333| \cdots .$$

Under assumption **(H)**, the frequency of a letter a exists for both sequences $x_<$ and $x_\prec$. But since,

$$\forall n \in \mathbb{N}, \quad \pi(\mathbf{v}(n), a, x_<) = \pi(\mathbf{v}(n), a, x_\prec)$$

the converging sequences $(\pi(n, a, x_<)/n)_{n \in \mathbb{N}}$ and $(\pi(n, a, x_\prec)/n)_{n \in \mathbb{N}}$ have a common infinite subsequence. So the two sequences are converging to the same limit, i.e., the frequency is independent of the ordering of the alphabet.

Acknowledgments

We thank Valérie Berthé who pointed out the useful reference [11] and the anonymous referees who helped us to improve the paper.

References

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

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

[3] S. Eilenberg, *Automata, Languages and Machines* vol. A, Academic Press, New York, (1974).

[4] Fogg, N. Pytheas, *Substitutions in Dynamics, Arithmetics and Combinatorics*, V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel Editors, *Lect. Notes in Math.* **1794**, Springer-Verlag, Berlin, (2002).

[5] F.R. Gantmacher, *The Theory of Matrices* vol. 2, Chelsea publishing company, New York, (1959).

[6] P.J. Grabner, M. Rigo, Additive functions with respect to numeration systems on regular languages, *Monatsh. Math.* **139** (2003), 205–219.

[7] P. Lecomte, M. Rigo, Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001), 27–44.

[8] P. Lecomte, M. Rigo, On the representation of real numbers using regular languages, *Theory Comput. Syst.* **35** (2002), 13–38.

[9] D. Lind, B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, (1995).

[10] M. Peter, The asymptotic distribution of elements in automatic sequences, *Theoret. Comput. Sci.* **301** (2003), 285–312.

[11] J. Peyrière, Fréquence des motifs dans les suites doubles invariantes par une substitution, *Ann. Sci. Math. Québec* **11** (1987), 133–138.

[12] M. Rigo, Generalization of automatic sequences for numeration systems on a regular language, *Theoret. Comput. Sci.* **244** (2000), 271–281.

[13] M. Rigo and A. Maes, More on generalized automatic sequences, *J. Autom. Lang. Comb.* **7** (2002), 351–376.

[14] K. Saari, On the frequency of letters in pure binary morphic sequences, in Developments in Language Theory, *Lect. Notes Comput. Sci.* **3572**, Springer, (2005), 397–408

[15] O. Salon, Suites automatiques à multi-indices, *Séminaire de théorie des nombres de Bordeaux*, Exp. 4 (1986–1987), **4**.01–**4**.27; followed by an Appendix by J. Shallit, 4-29A–4-36A.

[16] S. Yablonski, *Introduction aux mathématiques discrètes*, Éd. Mir, Moscou, (1983).

Samuel Nicolay, Michel Rigo
Institute of Mathematics
University of Liège
Grande Traverse 12 (B 37)
4000 Liège
Belgium
e-mail: {S.Nicolay,M.Rigo}@ulg.ac.be