# ODOMETERS ON REGULAR LANGUAGES

VALÉRIE BERTHÉ AND MICHEL RIGO†

Abstract. Odometers or “adding machines” are usually introduced in the context of positional numeration systems built on a strictly increasing sequence of integers. We generalize this notion to systems defined on an arbitrary infinite regular language. In this latter situation, if $(A, <)$ is a totally ordered alphabet, then enumerating the words of a regular language $L$ over $A$ with respect to the induced genealogical ordering gives a one-to-one correspondence between $\mathbb{N}$ and $L$. In this general setting, the odometer is not defined on a set of sequences of digits but on a set of pairs of sequences where the first (resp. the second) component of the pair is an infinite word over $A$ (resp. an infinite sequence of states of the minimal automaton of $L$). We study some properties of the odometer like continuity, injectivity, surjectivity, minimality,... We then study some particular cases: we show the equivalence of this new function with the classical odometer built upon a sequence of integers whenever the set of greedy representations of all the integers is a regular language; we also consider substitution numeration systems as well as the connection with $\beta$-numerations.

## 1. Introduction

To any infinite regular language $L$ over a totally ordered alphabet $(A, <)$, an *abstract numeration system* $S = (L, A, <)$ is associated in the following way [24]. Enumerating the words of $L$ by increasing genealogical order gives a one-to-one correspondence between $\mathbb{N}$ and $L$, the non-negative integer $n$ being represented by the $(n+1)$-th word of the ordered language $L$. In particular, these systems generalize classical positional systems like the $k$-ary systems, the Fibonacci system or more generally the numeration systems built on a sequence of integers satisfying a linear recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number [8].

In this framework of abstract numeration systems, the properties of the so-called $S$-recognizable sets of integers have been extensively studied (see for instance [24, 34, 35]). Moreover, these abstract systems have been extended to allow not only the representation of integers but also of real numbers [25]. In this latter situation, a real number is represented by an infinite word which is the limit of a converging sequence of words in $L$. Clearly, these systems lead to the generalization of various concepts related to the representation of integers like the automatic sequences or the summatory functions of additive functions [22, 36].

In this paper, we want to define and study the properties of odometers (also called adding machines) in the framework of abstract numeration systems built on an infinite regular language. In [21] odometers for positional numeration systems defined on a strictly increasing sequence $(U_n)_{n\in\mathbb{N}}$ of integers such that $U_0 = 1$ are investigated. In this latter situation, the odometer function — conventionally defined on a set a right-infinite words — can be defined on the set $\mathcal{R}$ of left-infinite words $\cdots \alpha_2\alpha_1\alpha_0$ satisfying a *greedy property* [16], i.e., for all $\ell \geq 0$,

$$\sum_{i=0}^{\ell} \alpha_i U_i < U_{\ell+1}. \tag{1}$$

†Corresponding author.

(We will also consider the greedy property (1) for finite words in the following.) We have chosen to consider left-infinite words (as in [17, 18], and contrarily for instance to the choice made in [21]) to avoid technical difficulties arising in the general situation of abstract numeration systems: working with right-infinite words would have required for instance to consider the mirror image of the regular language $L$ whose minimal automaton is not necessarily the automaton obtained by reversing the edges in the minimal automaton of $L$. Furthermore, this choice of notation is consistent with the one commonly used for finite words. Indeed the most significant digits are usually written first: the prefix $\alpha_k \cdots \alpha_0$ of a word in $\mathcal{R}$ has value $\alpha_k U_k + \cdots + \alpha_0 U_0$. The odometer is thus defined as the infinite extension of the successor function acting on the finite words of the form $\alpha_k \cdots \alpha_0$.

As an example, let us consider a finite word: the usual decimal representation of one hundred seventy-two is the word "172" and adding one leads to the word "173". Moreover, the language of the finite words representing all the integers can be embedded into $\mathcal{R}$ by concatenating ${}^\omega 0$ to the left of a greedy representation starting with the least significant digit as the rightmost element. So, one hundred seventy-two gives the element ${}^\omega(0)172 \in \mathcal{R}$. Adding one to an infinite word in $\mathcal{R}$ can produce a carry propagating to the left. As an example, the application of the odometer to ${}^\omega(0)2999$ gives ${}^\omega(0)3000$. In the case of the Fibonacci system where $U_0 = 1$, $U_1 = 2$ and $U_{n+2} = U_{n+1} + U_n$ applying the odometer to ${}^\omega(0)101010$ gives ${}^\omega(0)1000000$ (indeed, to be in $\mathcal{R}$ the greedy condition (1), i.e., the pattern "11" does not occur, must be satisfied).

One of the pioneering works concerning odometers is due to Vershik [46, 47] who introduced the seminal notion of adic transformation based on Bratteli diagrams [6], one motivation being the question of the approximation of ergodic systems. An adic transformation acts as a successor map on a Markov compactum (which might be considered as an analog of a Markov chain) defined as a lexicographically ordered set of infinite paths in an infinite labeled graph whose transitions are provided by an infinite sequence of transition matrices. In the stationary case (the transition matrices coincide, the infinite graph is a tree whose levels all have the same structure), (generalized) adic transformations correspond to substitutions and stationary odometers (see [26, 48] and [13]); for a survey of the relations between stationary adic transformations and substitutions, see [31], Chap. 7; see also [41]. Our approach is naturally inspired by the adic formalism but our framework is more general since we cannot represent our language $L$ just by considering the transitions in its minimal automaton: in particular, final (acceptance) states play a crucial role in our study (see for instance Example 22), and primitivity of the adjacency matrix of the minimal automaton is no more a sufficient condition for unique ergodicity, contrary to the adic case.

There is an important literature devoted to the study of odometers. Let us briefly quote [4] which continues the study of [21] from a combinatorial and topological point of view, and [3] for a metrical approach. Odometers can also be defined for two-sided dynamical systems as investigated in [17, 39]; we refer to [42] for the golden ratio case. See also [7] for an ergodic application of this notion in the framework of unimodal maps and wild attractors. Lastly, let us mention [18] which studies the sequential properties of the successor function for positional numeration systems.

One key-point in our study is the fact that we need to take into account not only infinite words but also the corresponding sequence of states: hence, the odometer is not only defined on a set of sequences of digits but on a set of pairs of sequences where the first (resp. the second) component of the pair is an infinite word over $A$ (resp. an infinite sequence of states of the minimal automaton of $L$). We illustrate this situation in Example 9 continued in Example 23, where a sequence of digits is associated to two sequences of states, each of the two corresponding pairs having a different image under the action of the odometer. Nevertheless, in the particular case of positional numeration, then Proposition 34 implies that we can forget the sequence of states, the reason being that we have in this case an underlying greedy algorithm. We stress the fact that in the general case of an abstract numeration system we cannot determine the successor without the information provided by the sequence of states; this is due in particular to the role played by the final states in the minimal automaton of the language.

We take as a definition for the set $\mathcal{K}$ on which the odometer $\tau$ is defined the set of left-infinite sequences $(x, y)$ over the alphabet $A \times Q$, where $A$ is the alphabet on which the language is defined, and $Q$ is the set of states of the minimal automaton recognizing $L$, with natural admissibility conditions provided by the automaton: the sequence $x$ is a limit of suffixes of words in the language $L$, whereas $y$ denotes the corresponding path in the automaton. For odometers defined upon classical numeration systems as studied in [21], the sequences of digits for which the carry can propagate to infinity when adding one play a special role (as illustrated and studied in details in [4, 3]). In our framework, the corresponding role will be played by the maximal words of fixed length and the set of their limit points in $\mathcal{K}$, that we denote $Max(\mathcal{K})$. We prove in particular that any element of $Max(\mathcal{K})$ is ultimately periodic and that $Max(\mathcal{K})$ is finite (Theorem 21). The set $Max(\mathcal{K})$ plays a special role concerning the continuity of the odometer $\tau$: indeed the continuity cannot be ensured on $Max(\mathcal{K})$ whatever is the value taken by $\tau$ for the points in this set. Moreover, we prove that $\tau$ is surjective onto $\mathcal{K} \setminus Min(\mathcal{K})$, injective on $\mathcal{K} \setminus Max(\mathcal{K})$, one-to-one from $\mathcal{K} \setminus Max(\mathcal{K})$ onto $\mathcal{K} \setminus Min(\mathcal{K})$, and continuous on $\mathcal{K} \setminus Max(\mathcal{K})$; furthermore, if the odometer is continuous then the dynamical system $(\mathcal{K}, \tau)$ is minimal. Let us observe that the set of discontinuity points is finite since it is included in $Max(\mathcal{K})$ which is finite.

This paper is organized as follows. After recalling the basic notions required in this paper, we define in Section 2 the set $\mathcal{K}$ on which the odometer acts, and state a few preliminary properties. Special focus is given on its extremal elements in Section 3, which allows us to define the odometer in Section 4. Its first properties (continuity, injectivity, surjectivity and minimality) are then stated in Section 5. We illustrate this study by making explicit the connection with a few well-known situations where the odometer is perfectly described: we consider the case of positional number systems in Section 6, the case of substitution numeration systems (with special focus on Pisot substitutions) in Section 7, and the case of $\beta$-numeration in Section 8. We consider the possibility of getting a real representation of the odometer in Section 9 and end this paper by considering some special cases in Section 10.

## 2. Preliminaries

Let $A = \{a_0 < a_1 < \cdots < a_k\}$ be a finite and totally ordered alphabet. In this paper $L \subset A^*$ will always denote a regular language such that

(2) the minimal automaton of $L$ has a loop of label $a_0$ in the initial state

In particular, this implies that $a_0^* L \subseteq L$ or otherwise stated, that $w \in L \Leftrightarrow \forall n \geq 0,\ a_0^n w \in L$. In some sense, property (2) can be related to the property of numeration systems built on a sequence of integers $(U_n)_{n\in\mathbb{N}}$ such that if $w = w_k \cdots w_0$ is the greedy representation of an integer $w_k\, U_k + \cdots + w_0\, U_0$ then $0^n w$, $n \in \mathbb{N}$, still satisfies the greedy condition (1) and represents the same integer. Here, we will be able to write an arbitrary number of $a_0$'s on the left of any word in $L$ and still obtain words belonging to $L$. Property (2) will therefore ensure the embedding of the finite words of $L$ representing the non-negative integers into some set of infinite sequences that will be precised later (Definition 1 below).

The minimal automaton of $L$ is denoted $\mathcal{M}_L = (Q, q_0, A, \delta, F)$ where $Q$ is the set of states, $q_0$ is the initial state, $F \subseteq Q$ is the set of final states and $\delta : Q \times A \to Q$ is the transition function. We assume that $\delta$ is *total* or equivalently that $\mathcal{M}_L$ is *complete*, i.e., $\delta$ is defined for all pairs $(q, a) \in Q \times A$ (notice that even with this assumption, $\mathcal{M}_L$ might contain a *sink*, i.e., a non-final state $s$ such that for any $a \in A$, $\delta(s, a) = s$). As usual, $\delta$ can be extended to $Q \times A^*$. Using these notations, property (2) is expressed as $\mathcal{M}_L$ has a loop in $q_0$ of label $a_0$. For the properties of the minimal automaton, see for instance [14, 38].

For any state $q \in Q$, we denote by $L_q$ the regular language accepted by $\mathcal{M}_L$ from state $q$,

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

and by $\mathbf{u}_q(n)$ the number of words of length $n$ in $L_q$. In particular, $L = L_{q_0}$.

Since $A$ is totally ordered, we can order the words of $A^*$ using the *genealogical ordering*. Let $u, v \in A^*$. We say that $u < v$ if $|u| < |v|$ or if $|u| = |v|$ and there exist $p, u', v' \in A^*$, $a, b \in A$, $a < b$ such that $u = pau'$ and $v = pbv'$. If $M$ is a language over $A$, we define $Max(M)$ as the set of the greatest words of each length in $M$, i.e.,

$$Max(M) = \{u \in M \mid \forall v \in M, |u| = |v| \Rightarrow v \leq u\}.$$

Observe that for all $n \geq 0$, $\#(Max(M) \cap A^n) \in \{0, 1\}$. In the same way, we can also define the set $Min(M)$ containing the smallest word of each length in $M$. It is well-known that if $M$ is regular then $Max(M)$ and $Min(M)$ are also regular [40].

If $w = w_0 \cdots w_\ell$ is a word over $A$ then the *reversal* (or *mirror*) of $w$ is $w_\ell \cdots w_0$ and is denoted $\widetilde{w}$. If $M$ is a language, then $\widetilde{M}$ is the language $\{\widetilde{w} \mid w \in M\}$. If $x = \cdots x_2 x_1 x_0$ is a left-infinite word then $\widetilde{x}$ is the right-infinite word $x_0 x_1 x_2 \cdots$.

Let $x = \cdots x_2 x_1 x_0$ be a left-infinite word. We say that $x_k \cdots x_0$ is a *prefix* of $x$ of length $k + 1$. For a finite word $y = y_1 \cdots y_p \cdots y_{p+\ell}$, we say that $y_p \cdots y_{p+\ell}$ is a *suffix* of $y$ of length $\ell + 1$.

A left-infinite word $(x_i)_{i \in \mathbb{N}}$ belongs to $\widetilde{\mathcal{L}}$ if there exists a sequence $(w_n)_{n \in \mathbb{N}}$ of words in $L$ such that for all $\ell > 0$ there exist $N_\ell > 0$ such that for all $n \geq N_\ell$, a suffix of $w_n$ of length at least $\ell$ is a prefix of $x$. Otherwise stated, if we use the topology induced by the infinite product topology on $A^{\mathbb{N}}$, a left-infinite word $x = (x_i)_{i \in \mathbb{N}}$ belongs to $\widetilde{\mathcal{L}}$ if and only if if and only if the right-infinite word $\widetilde{x} = x_0 x_1 x_2 \cdots$ is the limit of a converging sequence of words in $\widetilde{L}$ (or equivalently, if and only if there exists a sequence of words in $L$ whose suffixes are converging to $x$).

Abstract numeration systems being more general than positional numeration systems, we do not have a property equivalent to the greedy condition (1). In the framework of positional numeration systems, this greedy condition is enough to be able to define the odometer function on the set of left-infinite words $\mathcal{R}$ satisfying (1). This is no more true for abstract numeration systems and therefore we will not only consider words but also the extra information given by the sequence of reached states in $\mathcal{M}_L$. This is the reason of the introduction of the set $\mathcal{K}$ defined below.

**Definition 1.** We define the set $\mathcal{K} \subseteq {}^\omega(A \times Q)$ by $(x, y) = (\cdots x_2 x_1 x_0, \cdots y_2 y_1 y_0)$ belongs to $\mathcal{K}$ if and only if the following conditions hold

(1) $x$ belongs to $\widetilde{\mathcal{L}}$,
(2) $y_0$ belongs to $F$,
(3) for all $i \geq 0$, $\delta(y_{i+1}, x_i) = y_i$.

As we shall see in Example 9, a left-infinite word $x$ in $\widetilde{\mathcal{L}}$ can give rise to more than one sequence of states. So this extra information cannot be retrieved from $x$.

**Remark 2.** If $(x, y) = (\cdots x_2x_1x_0, \cdots y_2y_1y_0)$ belongs to $\mathcal{K}$ then for all $k \in \mathbb{N}$, $x_k \cdots x_0$ belongs to $L_{y_{k+1}}$.

**Remark 3.** Let $k \geq 0$. If $(x, y) = (\cdots x_2x_1x_0, \cdots y_2y_1y_0)$ belongs to $\mathcal{K}$ then $y_k \cdots y_0$ is completely determined by $x_k \cdots x_0$ and $y_{k+1}$. This is due to the third condition in the definition of $\mathcal{K}$ and because $\mathcal{M}_L$ is deterministic.

**Definition 4.** Let $j \geq 0$. A finite word $(x, y) = (x_kx_{k-1} \cdots x_0, y_ky_{k-1} \cdots y_0) \in (A \times Q)^{k+1}$, $k > j$, (resp. an infinite word $(x, y) = (\cdots x_2x_1x_0, \cdots y_2y_1y_0) \in {}^{\omega}(A \times Q)$), is said to have the property $\mathbf{max}_j$ and we write $(x, y) \in \mathbf{max}_j$ if $x_j \cdots x_0$ belongs to $Max(L_{y_{j+1}})$. In the same way, $(x, y)$ has the property $\mathbf{min}_j$ if $x_j \cdots x_0 \in Min(L_{y_{j+1}})$.

**Lemma 5.** *Let $(x, y) = (\cdots x_2x_1x_0, \cdots y_2y_1y_0) \in \mathcal{K}$ and $j \geq 0$. If $(x, y)$ has property $\mathbf{max}_j$ (resp. $\mathbf{min}_j$) then for all $k < j$, $(x, y)$ has also the property $\mathbf{max}_k$ (resp. $\mathbf{min}_k$).*

*Proof.* Assume that $(x, y) \in \mathbf{max}_j$ but $(x, y) \notin \mathbf{max_k}$, $k < j$. Therefore there exists $x'_k \cdots x'_0$ accepted from $y_{k+1}$ and genealogically greater than $x_k \cdots x_0$. So $x_j \cdots x_{k+1}x'_k \cdots x'_0$ belongs to $L_{y_{j+1}}$ and is greater than $x_j \cdots x_0$. This is a contradiction. □

**Corollary 6.** *Let $(x, y) = (\cdots x_2x_1x_0, \cdots y_2y_1y_0) \in \mathcal{K}$ and $j \geq 0$. If $(x, y) \notin \mathbf{max}_j$ (resp. $(x, y) \notin \mathbf{min}_j$) then for all $i \geq j$, $(x, y) \notin \mathbf{max}_i$ (resp. $(x, y) \notin \mathbf{min}_i$).*

**Notation 7.** Let $q$ be a given state in $Q$. For any word $w = w_\ell \cdots w_1$ in $L$, we denote by $p_q(w)$ the word over $A \times Q$ defined by

$$p_q(w) := (w_\ell, \delta(q, w_\ell))(w_{\ell-1}, (\delta(q, w_\ell w_{\ell-1})) \cdots (w_1, \delta(q, w_\ell \cdots w_1)) \in (A \times Q)^\ell$$

which represents simultaneously the label and the path given by the states reached consecutively in $\mathcal{M}_L$ by reading $w$.

Let us now present some other properties of this set $\mathcal{K}$.

**Proposition 8.** *For each $x \in \widetilde{\mathcal{L}}$, there exists $y \in {}^{\omega}Q$ such that $(x, y)$ belongs to $\mathcal{K}$.*

*Proof.* Since $x$ belongs to $\widetilde{\mathcal{L}}$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of words in $L$ whose suffixes are converging to $x$. For an infinite number of $n \in \mathbb{N}$, the last letter of $p_{q_0}(x_n)$ is a same element in $A \times F$. We take the corresponding subsequence $(x_{k_1(n)})_{n\in\mathbb{N}}$. For an infinite number of $n$, the words $p_{q_0}(x_{k_1(n)})$ have the same suffix of length two. So we consider the corresponding subsequence $(x_{k_2(n)})_{n\in\mathbb{N}}$. If we iterate this process, the suffixes of the words in the sequence $(x_{k_n(1)})_{n\in\mathbb{N}}$ are converging to $x$ and the sequence $(p_{q_0}(x_{k_n(1)}))_{n\in\mathbb{N}}$ is converging to a word $(x, y)$ in ${}^{\omega}(A \times Q)$ such that $(x, y)$ belongs to $\mathcal{K}$. □

**Example 9.** In this example, we consider a regular language $L \subset \{a < b < c\}^*$ satisfying the hypothesis $a^*L \subset L$ and given by its minimal automaton depicted in Figure 1. We just present some elements belonging to $\mathcal{K}$:

$$({}^{\omega}(abb), {}^{\omega}(012)),\ ({}^{\omega}a, {}^{\omega}0),\ ({}^{\omega}(a)b, {}^{\omega}(1)2),$$

$$({}^{\omega}(cbb)ab, {}^{\omega}(120201012)01) \text{ and } ({}^{\omega}(cbb)ab, {}^{\omega}(201012120)01).$$

Notice that the last two elements have the same first component but have different sequences of states. This situation can happen in very simple situations like $({}^{\omega}b, {}^{\omega}(012))$ and $({}^{\omega}b, {}^{\omega}(201))$. We shall see that the odometer function will act differently on these two elements of $\mathcal{K}$ even if they have the same first component.

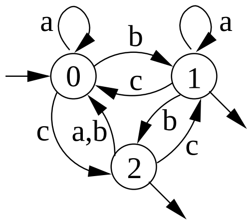


FIGURE 1. The minimal automaton of a language $L$.

As shown in the previous example, to one infinite word $x \in \widetilde{\mathcal{L}}$, it may correspond more than one sequence of states. If two such sequences give rise to elements in $\mathcal{K}$ then they differ almost everywhere.

**Lemma 10.** *Let $(x, y)$ and $(x, y')$ be two elements of $\mathcal{K}$ such that $y \neq y'$. Then there exists an index $i$ such that $y_i \neq y'_i$ and for all $n \geq 0$, $y_{i+n} \neq y'_{i+n}$.*

*Proof.* This is a direct consequence of Remark 3. □

The next proposition shows that to any finite word in $L$ corresponds at least one element in $\mathcal{K}$. The same kind of properties holds in the case of numeration systems built on a sequence of integers. If $w$ is the greedy representation of an integer (most significant digit on the left), then ${}^{\omega}0w$ belongs to the set $\mathcal{R}$ of left-infinite words satisfying the greedy property (1).

**Proposition 11.** *If $w = w_k \cdots w_0$ belongs to $L$ then there exists $y_k \cdots y_0 \in Q^{k+1}$ such that $({}^{\omega}(a_0)\, w_k \cdots w_0, {}^{\omega}(q_0)\, y_k \cdots y_0)$ belongs to $\mathcal{K}$.*

*Proof.* By our assumption (2) on $L$, if $w$ belongs to $L$ then $a_0^n w$ also belongs to $L$, $n \geq 0$. According to Notation 7, if $(w_k, y_k) \cdots (w_0, y_0) = p_{q_0}(w_k \cdots w_0)$ then $(a_0, q_0)^n\, (y_k, w_k) \cdots (y_1, w_1) = p_{q_0}(a_0^n\, w_k \cdots w_1)$. The result follows easily. □

## 3. PROPERTIES OF $Max(\mathcal{K})$ AND $Min(\mathcal{K})$

For odometers defined upon classical positional systems related to a sequence $(U_n)_{n\in\mathbb{N}}$ of integers, some sequences of digits play a special role (see in particular [4, 3]). Namely, they are the sequences for which the carry when adding one can propagate to infinity. A sequence $\cdots \alpha_2\alpha_1\alpha_0$ is of this kind if

$$\sum_{i=0}^{\ell_j} \alpha_i U_i = U_{\ell_j+1} - 1$$

for a strictly increasing infinite sequence $(\ell_j)_{j\in\mathbb{N}}$ of indices. In our framework, the corresponding elements in $\mathcal{K}$ belong to the set $Max(\mathcal{K})$ that we define below. The elements which have the dual property belong to $Min(\mathcal{K})$. In this section, we concentrate on the structural properties of the sets $Max(\mathcal{K})$ and $Min(\mathcal{K})$.

**Definition 12.** Let us define two particular subsets of $\mathcal{K}$,

$$Max(\mathcal{K}) = \{(x, y) \in \mathcal{K} \mid \forall i \geq 0, (x, y) \in \mathbf{max}_i\}$$

and

$$Min(\mathcal{K}) = \{(x, y) \in \mathcal{K} \mid \forall i \geq 0, (x, y) \in \mathbf{min}_i\}.$$

Let us observe that following Lemma 5, then it is sufficient in the definition of $Max(\mathcal{K})$ (resp. $Min(\mathcal{K})$)) that there exist infinitely many $i$ such that $(x, y) \in \mathbf{max}_i$ (resp. $(x, y) \in \mathbf{min}_i$). The following lemma is obvious.

**Lemma 13.** *Let $L$ be a regular language satisfying our assumption* (2).

- *A word $w$ belongs to $Min(L)$ if and only if for all $n \geq 0$, $a_0^n w$ belongs to $Min(L)$ (assuming that $a_0$ is the smallest letter in the ordered alphabet $A$).*
- *Let $q$ be a state of $\mathcal{M}_L$. If $vw$ belongs to $Max(L_q)$ then the word $w$ belongs to $Max(L_{\delta(q,v)})$.*

**Definition 14.** Let $w$ be the smallest word in $Min(L)$, i.e., $w$ is the first word in the ordered language $L$. If $w = w_1 \cdots w_\ell$ is not the empty word $\varepsilon$ (i.e., if $q_0 \notin F$) then we have a path in $\mathcal{M}_L$ of the form

$$q_0 \xrightarrow{w_1} q_1 \xrightarrow{w_2} \cdots \xrightarrow{w_\ell} q_\ell \in F.$$

We set $\mathbf{0} = ({}^\omega(a_0)\, w, {}^\omega(q_0)\, q_1 \cdots q_\ell)$. Otherwise $w = \varepsilon$ and we set $\mathbf{0} = ({}^\omega(a_0), {}^\omega(q_0))$.

**Proposition 15.** *The sets $Max(\mathcal{K})$ and $Min(\mathcal{K})$ are non-empty.*

*Proof.* As a consequence of Lemma 5 and Lemma 13, the element $\mathbf{0}$ given in Definition 14 belongs to $Min(\mathcal{K})$.

We use the same idea as in the proof of Proposition 8. Let $w_i$ be the $i$-th word of $Max(L)$ (clearly, $|w_i| < |w_{i+1}|$ for all $i \geq 1$). An infinite number of $w_i$'s have the same last letter $a_{k_1}$ and lead in $\mathcal{M}_L$ from $q_0$ to a same final state $q_{k_1}$. We therefore consider the corresponding subsequence $(w_{k_1(n)})_{n\in\mathbb{N}}$ built upon those $w_i$'s. We iterate this process: an infinite number of words among the $w_{k_1(n)}$'s have the same suffix $a_{k_2}a_{k_1}$ and finally lead in $\mathcal{M}_L$ to the states $q_{k_2}$ followed by $q_{k_1}$. Therefore we build a sequence converging to

$$(\cdots a_{k_2}a_{k_1}, \cdots q_{k_2}q_{k_1}).$$

Thanks to Lemma 13, this element belongs to $Max(\mathcal{K})$. □

**Example 16.** We consider the language and the automaton given in Example 9. It is easy to check that $({}^\omega(c)b, {}^\omega(021)2)$, $({}^\omega c, {}^\omega(021))$ and $({}^\omega c, {}^\omega(102))$ belong to $Max(\mathcal{K})$. We also have $\mathbf{0} = ({}^\omega(a)b, {}^\omega(0)1)$ and $({}^\omega a, {}^\omega 1)$ as elements of $Min(\mathcal{K})$. To show that these elements are the only ones, we will need some more results about the structure of $Max(L_q)$ and $Min(L_q)$.

3.1. **Structure of $Max(L_q)$.** In some particular cases, the structure of $Max(L_q)$ is easy to obtain. Recall that a state $s$ is a sink if for any $a \in A$, $\delta(s,a) = s$ and $s$ is not a final state.

**Notation 17.** Let $q$ be a state in $\mathcal{M}_L$. If there exists $a \in A$ such that $\delta(q,a)$ is not the sink then we denote by $m(q)$ the largest letter having this property, otherwise we set $m(q) = \varepsilon$.

Let us introduce a small algorithm to detect what we will call the *maximal cycles* in $\mathcal{M}_L$.

**Algorithm 18.** Let $q \in Q$.

- Set $y_0 \leftarrow q$ and $i \leftarrow 0$.
- If $m(y_i) \neq \varepsilon$ then set $y_{i+1} \leftarrow \delta(y_i, m(y_i))$ and $i \leftarrow i+1$.
  Otherwise the algorithm halts.
- If $y_0, \ldots, y_i$ are all different, repeat the previous step.
  Otherwise, a cycle is found and the algorithm halts.

After applying this algorithm to a state $q \in Q$ which is not the sink, we can have two kinds of situations. If we encounter some state $y_k$ such that $m(y_k) = \varepsilon$ then we have obtained something like

$$y_0 \xrightarrow{m(y_0)} y_1 \xrightarrow{m(y_1)} \cdots \longrightarrow y_{k-1} \xrightarrow{m(y_{k-1})} y_k$$

where all the $y_i$'s are different and $y_k$ belongs necessarily to $F$ (because $\mathcal{M}_L$ is minimal). Or we have the situation

$$y_0 \xrightarrow{m(y_0)} y_1 \xrightarrow{m(y_1)} \cdots \longrightarrow \mathbf{y_k} \xrightarrow{m(y_k)} \cdots \longrightarrow y_{k+n} \xrightarrow{m(y_{k+n})} \mathbf{y_k}$$

where $y_0, \ldots, y_{k+n}$ are all different; we say that $(y_k, m(y_k), \ldots, y_{k+n}, m(y_{k+n}), y_k)$ is a *maximal cycle* starting in $y_k$ and the word $m(y_k) \cdots m(y_{k+n})$ is the *label* of this cycle. Notice that two maximal cycles have no state in common or share exactly the same states. In this latter case, the label of one of the two cycles is a cyclic permutation of the other one.

**Example 19.** Considering the automaton of Example 9, we have three maximal cycles: $(0, c, 2, c, 1, c, 0)$, $(2, c, 1, c, 0, c, 2)$ and $(1, c, 0, c, 2, c, 1)$ all having the same label $ccc$ and sharing the same states.

**Lemma 20.** *If $\mathcal{C}$ is a maximal cycle of label $w$ starting in $q$, then there exist an integer $k \le |w|$ depending only on $\mathcal{C}$ and $k$ words $u_1, \ldots, u_k$ of minimal length such that $|u_i| \not\equiv |u_j| \mod |w|$ if $i \neq j$ and*

$$Max(L_q) = w^*\{u_1, \ldots, u_k\}.$$

*Proof.* Let $w$ be the label of a maximal cycle $\mathcal{C}$ starting in $q$. If $v$ belongs to $Max(L_q)$ then by construction of the maximal cycle, $wv$ also belongs to $Max(L_q)$. Assume now that $u, v \in Max(L_q)$ are such that $|u| \equiv |v| \mod |w|$ and $|u| < |v|$. Therefore, there exists $i$ such that $w^i u$ belongs to $Max(L_q)$ and $|w^i u| = |v|$. But $Max(L_q)$ contains at most one word of each length, so $w^i u = v$. Consequently, if $v$ belongs to $Max(L_q)$ then it is of the form $w^n u$ for some $n \ge 0$ and $w$ is not a prefix of $u$. For each $j \in \{0, \ldots, k-1\}$ there is at most one $u$ of this kind such that $|u| \equiv j \mod |w|$ (actually $u$ is the smallest word of length $j + n|w|$ possibly belonging to $Max(L_q)$, $n \ge 0$). Notice that it does not mean that $|u| < |w|$. Clearly two states in the same maximal cycle give rise to the same kind of maximal set. □

It is more difficult to express the form of $Max(L_q)$ when this set is infinite and $q$ does not belong to a maximal cycle. But hopefully we have a more general result extending Lemma 20 which holds even if $q$ does not belong to a maximal cycle. Indeed, since $\#(Max(L_q) \cap A^n) \le 1$ for all $n \in \mathbb{N}$ then it is well-known (see [30] or [40]) that there exists a finite set $R$ of words, an integer $k \ge 0$ and words $u_i, w_i \in A^*$, $v_i \in A^+$, $i = 0, \ldots, k$ such that

$$Max(L_q) = \bigcup_{i=0}^{k} u_i\, v_i^*\, w_i \cup R \tag{3}$$

where the languages $u_i\, v_i^*\, w_i$ are pairwise disjoint and also disjoint from $R$. Otherwise stated, if $i \neq j$ then

$$\{|u_i w_i| + n|v_i| : n \in \mathbb{N}\} \cap \{|u_j w_j| + n|v_j| : n \in \mathbb{N}\} = \emptyset$$

and $\{|u_i w_i| + n|v_i| : n \in \mathbb{N}\} \cap |R| = \emptyset$, for all $i$ ($|R|$ denotes the set of lengths of elements of $R$). One can observe that the form of $Max(L_q)$ given in Lemma 20 is a special case of (3).

3.2. **Structure of $Max(\mathcal{K})$.** We now have gathered all the required elements to be able to deduce information on the elements of $Max(\mathcal{K})$ from the languages $Max(L_q)$.

**Theorem 21.** *Any element in $Max(\mathcal{K})$ is ultimately periodic and $Max(\mathcal{K})$ is finite.*

*Proof.* (a) The ideas of the first part of this proof are the same as in [25, Lemma 7]. Let $q$ be such that $\#Max(L_q) = \infty$. If $x$ is a word in $Max(L_q)$ of length large enough then thanks to (3) there exist unique words $u, v, w$ (depending on $x$) such that $x = uv^n w$. Among

$$\delta(q, u),\ \delta(q, u\,v), \ldots,\ \delta(q, u\,v^{\#Q}) \tag{4}$$

a same state appears at least twice. Let $t$ be the first state appearing twice in this list. Let $i < j$ be the smallest integers such that $\delta(q, u\,v^i) = \delta(q, u\,v^j) = t$. We set $P = (j - i)|v|$. We can already notice that $P$ is bounded by $\#Q \cdot |v|$. In what follows, we write simply $t$, $i$, $j$, $P$ assuming that the word $x$ is understood from the context.
(b) We use Notation 7. Consider again the word $x = uv^n w \in Max(L_q)$ introduced in (a). For $n$ large enough, $p_q(uv^n w)$ is a word over $A \times Q$ having

i) a non-periodic prefix $p_q(u\,v^i)$ of length bounded by $|u| + \#Q \cdot |v|$;
ii) a maximal periodic factor having a period of length $P$; actually the Euclidean division of $n - i$ by $P/|v|$ gives
$$n - i = m\frac{P}{|v|} + r \quad \text{with } r < P/|v|;$$
the periodic factor corresponding to $v^{mP/|v|}$ is $p_t(v^{mP/|v|})$ and the period corresponding to $v^{P/|v|}$ is $p_t(v^{P/|v|})$ where $t$ is as in (a) the first state appearing twice in the list (4);
iii) a non-periodic suffix of length bounded by $|w| + P$; indeed this factor corresponds to $v^r w$ and is of the form $p_t(v^r w)$.

For a better understanding, the situation is sketched in Figure 2.

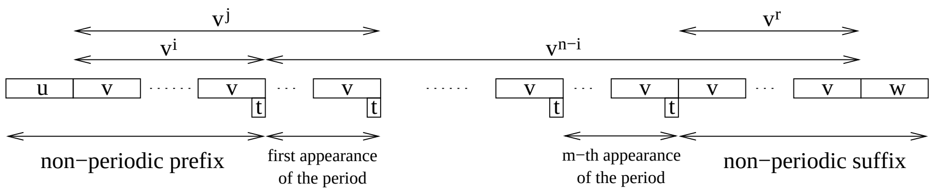


FIGURE 2. A schematic representation of $p_q(uv^n w)$.

(c) Let $n' > n$ and $x' = uv^{n'} w$. Then $p_q(uv^n w)$ and $p_q(uv^{n'} w)$ have the same prefix corresponding to $u\,v^i$. The periodic factors have the same period of length $P$ but the number of repetitions could be larger for $x'$. Finally, if $n$ and $n'$ are not congruent modulo $P/|v|$ then the corresponding suffixes could be different, otherwise the suffixes are the same. Notice that there are only finitely many possible suffixes corresponding to the words of the form $v^r w$ for $r = 0, \ldots, P/|v| - 1$.
(d) From the previous observations, we can easily exhibit elements in $Max(\mathcal{K})$. Let $n_0$ be large enough and set $z_m = p_q(u\,v^{n_0 + mP/|v|}\,w)$ for $m \geq 0$. From the previous point, the sequence $(z_m)_{m \in \mathbb{N}} \in {}^{\omega}(A \times Q)$ admits a sequence of suffixes which converges to an ultimately periodic element in $\mathcal{K}$. From Lemma 13, this element belongs to $Max(\mathcal{K})$.
(e) Clearly, any element $(x, y) = (\cdots x_1 x_0, \cdots y_1 y_0)$ in $Max(\mathcal{K})$ is ultimately periodic. Indeed, since $Q$ is finite, a state $q$ must appear infinitely often in $y$, say, in strictly increasing positions $k(n)$. For each $n$, $x_{k(n-1)} \cdots x_0$ belongs to $Max(L_q)$ and the words of this kind have a longer and longer common suffix when $n$ is increasing. As a consequence of (a), $(x, y)$ is ultimately periodic with $x$ of the form ${}^{\omega}(v)w$, for some finite words $v$ and $w$.
(f) In (d), we have obtained elements of $Max(\mathcal{K})$ of a special form but in (e) we

have shown that any element in $Max(\mathcal{K})$ is of this kind. To conclude, we have to show that $Max(\mathcal{K})$ is finite. First from (3), for each state $q$ the number of words $u_i$, $v_i$, $w_i$ used to obtain the structure of $Max(L_q)$ is finite. For each of these 3-tuples $(u_i, v_i, w_i)$ of words, we can obtain ultimately periodic elements in $Max(\mathcal{K})$ but the period of such an element is bounded by $\#Q \cdot |v_i|$ (see (a)) and the length of its prefix is bounded by $|w_i| + \#Q \cdot |v_i|$ (see (b)). In other words, we have a finite number of 3-tuples $(u_i, v_i, w_i)$ each one giving at most a finite number of elements in $Max(\mathcal{K})$. □

This proof shows that the elements of $Max(\mathcal{K})$ can be determined by the knowledge of the languages $Max(L_q)$. As we will see in the following example, obtaining the decomposition of the form (3) for the languages $Max(L_q)$ gives rise to all the elements in $Max(\mathcal{K})$. Moreover, observe that these languages $Max(L_q)$ can be efficiently obtained from $\mathcal{M}_L$.

Naturally, Algorithm 18, Lemma 20 and Theorem 21 are easily adapted to the set $Min(\mathcal{K})$. In this case, similarly as in Notation 17, if there exists $a \in A$ such that $\delta(q, a)$ is not the sink then we denote by $m(q)$ the smallest letter having this property.

**Example 22.** Continuing again Example 9. We are now able to show that $Max(\mathcal{K})$ contains exactly the elements $({}^{\omega}(c)b, {}^{\omega}(021)2)$, $({}^{\omega}c, {}^{\omega}(021))$ and $({}^{\omega}c, {}^{\omega}(102))$. We have a maximal cycle of label $ccc$ containing the three states of $\mathcal{M}_L$, so using Lemma 20 we obtain $Max(L_0) = (ccc)^*\{c, cc, ccb\}$, $Max(L_1) = (ccc)^*\{\varepsilon, b, cc\}$ and $Max(L_2) = (ccc)^*\{\varepsilon, c, cb\}$. Let us first see which elements in $Max(\mathcal{K})$ come from the words in $Max(L_0)$. The word $(ccc)^n c$ read from the state 0 gives in $\mathcal{M}_L$ the path

$$0 \xrightarrow{c} 2 \xrightarrow{c} 1 \xrightarrow{c} 0 \cdots \xrightarrow{c} 2 \xrightarrow{c} 1 \xrightarrow{c} 0 \xrightarrow{c} 2.$$

With Notation 7, we have

$$p_0((ccc)^n c) = ((ccc)^n c, (210)^n 2).$$

Reading this path to the left and letting $n$ tend to infinity gives the element $({}^{\omega}c, {}^{\omega}(210)2) = ({}^{\omega}c, {}^{\omega}(102))$. In the same way, the word $(ccc)^n cc$ gives $({}^{\omega}c, {}^{\omega}(021))$ and finally $(ccc)^n ccb$ gives $({}^{\omega}(c)b, {}^{\omega}(021)2)$. If we do the same for the words in $Max(L_i)$, $i = 1, 2$, then we consider paths starting in $i$ and we obtain exactly the same three elements of $\mathcal{K}$. It is clear that each set $Max(L_q)$ produces the same elements of $Max(\mathcal{K})$ because all the states are in the same maximal cycle.

Let us now show that $Min(\mathcal{K})$ contains exactly $\mathbf{0} = ({}^{\omega}(a)b, {}^{\omega}(0)1)$ and $({}^{\omega}a, {}^{\omega}1)$. Here we have two minimal cycles: $(0, a, 0)$ and $(1, a, 1)$. So thanks to the analogue of Lemma 20, we have $Min(L_0) = a^*b$ and $Min(L_1) = a^*$. From the analogue of (3), one finds $Min(L_2) = a^*ab \cup \{\varepsilon, c\}$. For instance, starting in state 2 and reading $a^n ab$ gives the path

$$2 \xrightarrow{a} 0 \xrightarrow{a} 0 \cdots \xrightarrow{a} 0 \xrightarrow{a} 0 \xrightarrow{b} 1.$$

Reading this path to the left and letting $n$ tend to infinity gives the element $\mathbf{0}$. Starting in 0 with $a^n b$ also leads to the same element $\mathbf{0}$. Finally starting in 1 with $a^n$ gives $({}^{\omega}a, {}^{\omega}1)$. Obviously, if two states $q$ and $q'$ belong to two different minimal cycles then the sets $Min(L_q)$ and $Min(L_{q'})$ will never lead to a same element in $Min(\mathcal{K})$ because the two cycles have no state in common.

## 4. Defining the odometer

In [21], if a left-infinite word $\cdots \alpha_2\alpha_1\alpha_0$ of digits belonging to the set $\mathcal{R}$ of left-infinite words satisfying the greedy condition (1) is such that there exists $M$ such that for all $\ell \ge M$

$$[\alpha_\ell \cdots \alpha_0] := \sum_{i=0}^{\ell} \alpha_i U_i < U_{\ell+1} - 1$$

then the odometer maps $\cdots \alpha_2\alpha_1\alpha_0$ onto $\cdots \alpha_{k+2}\alpha_{k+1}\alpha'_k \cdots \alpha'_0 \in \mathcal{R}$ where $\alpha'_k \cdots \alpha'_0$ is the representation of $[\alpha_k \cdots \alpha_0]+1$ computed through the greedy algorithm (and it is shown that the result is independent of the choice of the index $k \ge M$). Obviously, the representations of $[\alpha_k \cdots \alpha_0]$ and $[\alpha_k \cdots \alpha_0]+1$ have the same length. Otherwise, infinitely often the situation $[\alpha_\ell \cdots \alpha_0] = U_{\ell+1} - 1$ occurs and then the odometer is defined to map $\cdots \alpha_2\alpha_1\alpha_0$ onto ${}^\omega 0$.

Here we want to do the same in the context of abstract numeration systems and define a function $\tau_L : \mathcal{K} \to \mathcal{K}$, or simply $\tau$ if $L$ is clearly understood, having the expected adding behavior for an odometer. First we define $\tau$ on $\mathcal{K} \setminus Max(\mathcal{K})$. Assume that for $(x, y) = (\cdots x_1x_0, \cdots y_1y_0) \in \mathcal{K}$ there exists $i \ge 0$ such that $(x, y) \notin \mathbf{max}_i$. For each state $q$ of $\mathcal{M}_L$, we define the function

$$\mathrm{Succ}_q : L_q \to L_q$$

mapping the $k$-th word in the genealogically ordered language $L_q$ to the $(k+1)$-th one in the same language (if $L_q$ is finite, we decide that $\mathrm{Succ}_q$ maps the largest word in $L_q$ onto the smallest one). Since $(x, y) \notin \mathbf{max}_i$, it is clear that $x_i \cdots x_0$ and $\mathrm{Succ}_{y_{i+1}}(x_i \cdots x_0)$ have the same length. Let us denote this latter word belonging to $L_{y_{i+1}}$ by $x'_i \cdots x'_0$. We set $y'_i = \delta(y_{i+1}, x'_i)$ and $y'_j = \delta(y'_{j+1}, x'_j)$ for $j = i-1, \ldots, 0$. In other words, $y'_i, \ldots, y'_0$ are the states reached in $\mathcal{M}_L$ when reading $x'_i \cdots x'_0$ from $y_{i+1}$. In particular, observe that $y'_0$ belongs to $F$. Hence $(\cdots x_{i+1}\, x'_i \cdots x'_0, \cdots y_{i+1}\, y'_i \cdots y'_0)$ belong to $\mathcal{K}$. The function $\tau$ is defined by

$$\tau(\cdots x_{i+1}\, x_i \cdots x_0, \cdots y_{i+1}\, y_i \cdots y_0) = (\cdots x_{i+1}\, x'_i \cdots x'_0, \cdots y_{i+1}\, y'_i \cdots y'_0).$$

We have to show that $\tau$ is well-defined. Assume that there exist $i < j$ such that $(x, y) \notin \mathbf{max}_i$ and $(x, y) \notin \mathbf{max}_j$. (Notice that in view of Corollary 6, if $(x, y) \notin \mathbf{max}_i$ then for all $k \ge i$, $(x, y) \notin \mathbf{max}_k$.) Then the previous construction does not depend on the choice of the index. Indeed, notice that by definition of $\mathcal{K}$, $\delta(y_{j+1}, x_j \cdots x_{i+1}) = y_{i+1}$ and as a consequence of the genealogical ordering,

$$\mathrm{Succ}_{y_{j+1}}(x_j \cdots x_{i+1}x_i \cdots x_0) = x_j \cdots x_{i+1}\, \mathrm{Succ}_{y_{i+1}}(x_i \cdots x_0).$$

Therefore, as a consequence of Remark 3, the corresponding sequences of states are the same: if $y''_j \cdots y''_0$ are the states reached in $\mathcal{M}_L$ when reading $\mathrm{Succ}_{y_{j+1}}(x_j \cdots x_0)$ from $y_{j+1}$, we have

$$y''_j \cdots y''_{i+1}y''_i \cdots y''_0 = y_j \cdots y_{i+1}y'_i \cdots y'_0.$$

Thus, the value of $\tau$ does not depend on the index $i$ such that $(x, y) \notin \mathbf{max}_i$.

**Example 23.** We still consider the language and the automaton given in Example 9. For instance, $(x, y) = ({}^\omega(a)bccabb, {}^\omega(0)102012)$ belongs to $\mathcal{K}$. The word $b$ belongs to $Max(L_1)$ so $(x, y) \in \mathbf{max}_1$ but $bb$ belongs to $L_0 \setminus Max(L_0)$ so $(x, y) \notin \mathbf{max}_2$. It is easy to see that the next word accepted from 0 is $cc$ and the path is $0 \xrightarrow{c} 2 \xrightarrow{c} 1$, thus

$$\tau({}^\omega(a)bcca|bb, {}^\omega(0)1020|12) = ({}^\omega(a)bcca|cc, {}^\omega(0)1020|21).$$

If we had considered the word $ccabb$ accepted from state 1 (because $(x, y) \notin \mathbf{max}_5$), the next word in $L_1$ is $ccacc$ and this would have lead to the same element in $\mathcal{K}$:

$$\tau({}^\omega(a)b|ccabb, {}^\omega(0)1|02012) = ({}^\omega(a)b|ccacc, {}^\omega(0)1|02021).$$

In the next section, we will see that in general, the continuity of the odometer cannot be ensured on $Max(\mathcal{K})$ whatever is the value taken by $\tau$ for the points in this set (see Example 31). Therefore, we decide that for all $(x, y) \in Max(\mathcal{K})$, $\tau(x, y) = \mathbf{0}$, where $\mathbf{0}$ is the canonical element of $Min(\mathcal{K})$ given in Definition 14.

**Remark 24.** When applying $\tau$ to two elements $(x, y)$ and $(x, y')$ in $\mathcal{K}$ having the same first component $x$ but two distinct sequences of states $y$ and $y'$, we can obtain elements for which the first components are different. This new phenomenon does not appear in positional numeration systems due to the greedy property (1) that must be satisfied by any element in $\mathcal{R}$. To convince the reader that such a phenomenom can appear in our general framework, consider once again the language and the automaton given in Example 9. We already know that $({}^\omega b, {}^\omega(012))$ and $({}^\omega b, {}^\omega(201))$ belong to $\mathcal{K}$. If we compute $\tau$, we obtain

$$\tau({}^\omega b, {}^\omega(012)) = ({}^\omega(b)cc, {}^\omega(120)21) \text{ and } \tau({}^\omega b, {}^\omega(201)) = ({}^\omega(b)c, {}^\omega(120)2).$$

Indeed, $b \in Max(L_1)$, $\mathrm{Succ}_0(bb) = cc$ and for the second computation, $\mathrm{Succ}_0(b) = c$.

Naturally, since positional numeration systems are particular cases of abstract systems, there are situations where $\tau$ can be computed without the use of the sequence of states. This very strict case is studied in Section 6 and in particular in Proposition 34. Nevertheless, in our general framework, the use of pairs cannot be avoided.

**Remark 25.** We can as in [21] or [46] define a partial ordering on $\mathcal{K}$, called *antipodal order*, in the following way. We have $(x, y) \prec (x', y')$ if $(x, y) = (x', y')$ or there exists some index $k$ such that $x_k < x'_k$ and for all $j > k$, $(x_j, y_j) = (x'_j, y'_j)$. The elements in $Max(\mathcal{K})$ are therefore the maximal elements in $(\mathcal{K}, \prec)$. For any $(x, y) \notin Max(\mathcal{K})$, then its image under $\tau$ is the smallest (with respect to $\prec$) of all the elements in $\mathcal{K}$ which are larger than $(x, y)$. Hence the map $\tau$ is a successor function which can be considered as an adic transformation following [46, 47].

## 5. Properties of the odometers

The aim of this section is to state some general properties of the odometer like continuity, injectivity, surjectivity, minimality, ...

**Proposition 26.** *The map $\tau$ is surjective onto $\mathcal{K} \setminus Min(\mathcal{K})$.*

*Proof.* The proof is immediate. Let $(x, y) = (\cdots x_1 x_0, \cdots y_1 y_0)$ be such that $(x, y)$ is not in $\mathbf{min}_i$ for some $i$. Therefore, there exists a word $x' = x'_i \cdots x'_0$ of length $i + 1$ such that $\mathrm{Succ}_{y_{i+1}}(x'_i \cdots x'_0) = x_i \cdots x_0$. As usual, if $y' = y'_i \cdots y'_0$ is the path followed in $\mathcal{M}_L$ from $y_{i+1}$ when reading $x'$ then

$$\tau(\cdots x_{i+2} x_{i+1} x', \cdots y_{i+2} y_{i+1} y') = (x, y).$$

□

**Remark 27.** A similar result holds in the framework of positional number systems: the odometer is proved to be surjective if and only if ${}^\omega 0$ admits an antecedent (see [21]).

**Proposition 28.** *The map $\tau$ is injective on $\mathcal{K} \setminus Max(\mathcal{K})$.*

*Proof.* Let $(x, y) = (\cdots x_1 x_0, \cdots y_1 y_0)$ and $(x', y') = (\cdots x'_1 x'_0, \cdots y'_1 y'_0)$ be in $\mathcal{K} \setminus Max(\mathcal{K})$ and such that $\tau(x, y) = \tau(x', y')$. Let $i$ and $i'$ be such that $x_i \cdots x_0 \notin Max(L_{y_{i+1}})$ and $x'_{i'} \cdots x'_0 \notin Max(L_{y'_{i'+1}})$. Pose $I = \sup\{i, i'\}$. Thanks to Lemma 5, $(x, y)$ and $(x', y')$ do not belong to $\mathbf{max}_I$ so the application of $\tau$ will at most affect their suffix of length $I + 1$. Since $\tau(x, y) = \tau(x', y')$, we have $x_j = x'_j$ and $y_j = y'_j$ for all $j > I$. Therefore, $x_I \cdots x_0$ and $x'_I \cdots x'_0$ belongs to $L_{y_{I+1}} = L_{y'_{I+1}}$ and have the same successor. So these two words are the same. The conclusion that $y_I \cdots y_0$ and $y'_I \cdots y'_0$ are the same comes from Remark 3. □

**Corollary 29.** *The map $\tau$ is one-to-one from $\mathcal{K} \setminus Max(\mathcal{K})$ onto $\mathcal{K} \setminus Min(\mathcal{K})$.*

*Proof.* It is a direct consequence of the fact that $\tau(\mathcal{K} \setminus Max(\mathcal{K})) \subset \mathcal{K} \setminus Min(\mathcal{K})$. Indeed the restriction of $\tau$ on $\mathcal{K} \setminus Max(\mathcal{K})$ is surjective onto $\mathcal{K} \setminus Min(\mathcal{K})$ since the image of $Max(\mathcal{K})$ equals $\{\mathbf{0}\} \subset Min(\mathcal{K})$. $\square$

The topology on ${}^{\omega}(A \times Q)$ is as usual induced by the distance $d$ defined by

$$d((x, y), (x', y')) = 2^{-k} \text{ where } k = \inf\{i \mid (x_i, y_i) \neq (x'_i, y'_i)\}.$$

**Proposition 30.** *The map $\tau$ is continuous on $\mathcal{K} \setminus Max(\mathcal{K})$.*

*Proof.* Let $(u, v) \in \mathcal{K} \setminus Max(\mathcal{K})$ and $\epsilon > 0$. We show that there exists $\eta > 0$ such that if $d((u, v), (x, y)) < \eta$ then $d(\tau(u, v), \tau(x, y)) < \epsilon$. Let $j$ be the smallest index such that $(u, v) \notin \mathbf{max}_j$. If there exists $i > j$ such that $(u, v)$ and $(x, y)$ have the same prefix of length $i + 1$ then $\tau(u, v)$ and $\tau(x, y)$ also have the same suffix of length $i + 1$. Clearly, one can take

$$\eta = 2^{-\sup\{1-\log_2 \epsilon, j+1\}}.$$

$\square$

The following example shows that $\tau$ is generally not continuous on the points of $Max(\mathcal{K})$.

**Example 31.** Consider the regular language $L$ accepted by the automaton depicted in Figure 3 (where the sink is not represented). For instance, $(u, v) = ({}^{\omega}d, {}^{\omega}(21))$

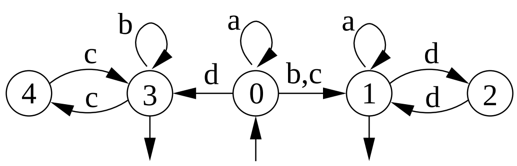


FIGURE 3. The minimal automaton of a language $L$.

belongs to $Max(\mathcal{K})$. The points

$$({}^{\omega}(a)\, b\, (dd)^n, {}^{\omega}(0)\, 1\, (21)^n) \text{ and } ({}^{\omega}(a)\, c\, (dd)^n, {}^{\omega}(0)\, 1\, (21)^n)$$

can be chosen arbitrarily close of $(u, v)$ for $n$ large enough. Whatever is the value of $\tau(u, v)$, the map $\tau$ is not continuous at $(u, v)$. Indeed,

$$\tau({}^{\omega}(a)\, b\, (dd)^n, {}^{\omega}(0)\, 1\, (21)^n) = ({}^{\omega}(a)\, c\, (aa)^n, {}^{\omega}(0)\, 1\, (11)^n)$$

but

$$\tau({}^{\omega}(a)\, c\, (dd)^n, {}^{\omega}(0)\, 1\, (21)^n) = ({}^{\omega}(a)\, d\, (bb)^n, {}^{\omega}(0)\, 3\, (33)^n).$$

So clearly, if a point $(x, y)$ is close from an element in $Max(\mathcal{K})$ then its image $\tau(x, y)$ is close from an element in $Min(\mathcal{K})$ but nothing more can be said.

**Proposition 32.** *The set $\mathcal{K}$ is a compact subset of ${}^{\omega}(A \times Q)$. If the odometer $\tau$ is continuous, then the dynamical system $(\mathcal{K}, \tau)$ is minimal, that is, every non-empty closed subset of $\mathcal{K}$ invariant under the action of $\tau$ is equal to $\mathcal{K}$.*

*Proof.* We follow here the proof of [21] adapted to our situation. The compactness of $\mathcal{K}$ is immediate as a closed subset of ${}^{\omega}(A \times Q)$.

We assume that $\tau$ is continuous. Let us prove that the closure of the orbit $\{\tau^n(x, y) \mid n \in \mathbb{N}\}$ of any point $(x, y) \in \mathcal{K}$ is equal to $\mathcal{K}$, which immediately implies the minimality.

Let us first observe that the orbit $\{\tau^n(\mathbf{0}) \mid n \in \mathbb{N}\}$ of $\mathbf{0}$ is dense in $\mathcal{K}$. Indeed, let $(x, y) \in \mathcal{K}$. Let $(w_n)_{n\in\mathbb{N}}$ be a sequence of words in $L$ such that $x$ is the limit of suffixes of words of the sequence $(w_n)_{n\in\mathbb{N}}$. Let $l_n$ denote the $n$-th word in the ordered language $L$. The point $\tau^n(\mathbf{0})$ is by definition equal to ${}^{\omega}(a_0, q_0)p_{q_0}(l_n)$, according to Notation 7. Hence $(x, y)$ is a limit of elements of $\{\tau^n(\mathbf{0}) \mid n \in \mathbb{N}\}$, and $\mathcal{K}$, which is a closed set, is the closure of $\{\tau^n(\mathbf{0}) \mid n \in \mathbb{N}\}$.

Now if $(x, y) \in Max(\mathcal{K})$, then $\tau(x, y) = \mathbf{0}$ and $\mathbf{0}$ belongs to the orbit of $(x, y)$, which implies that the closure of the orbit of $(x, y)$ is equal to $\mathcal{K}$.

Let us suppose that $x \notin Max(\mathcal{K})$. Let

$$D : \mathcal{K} \setminus Max(\mathcal{K}) \to \mathbb{N} : x \mapsto \sup\{k \mid \ (x, y) \in \mathbf{max}_k\}.$$

Let us prove that $D$ does not take bounded values on the orbit of $(x, y)$. Suppose on the contrary that there exists $C$ such that $D(\tau^n(x, y)) < C$, for every $n$; in particular, $(x, y)$ does not have the property $\mathbf{max}_C$; by definition of the odometer, after a suitable number of iterations of $\tau$, say, $n$, then $\tau^n(x, y)$ is easily seen to belong to $\mathbf{max}_C$, hence a contradiction. There thus exists an increasing sequence of integers $(n_k)_{k\in\mathbb{N}}$ such that $\tau^{n_k}(x, y) \in \mathbf{max}_k$. By compactness of $\mathcal{K}$, one can extract from $(n_k)_{k\in\mathbb{N}}$ an increasing sequence $(m_k)_{k\in\mathbb{N}}$ such that the sequence $(\tau^{m_k}(x, y))_{k\in\mathbb{N}}$ converges; its limit belongs to $Max(\mathcal{K})$, according to Lemma 5. By continuity of $\tau$, $(\tau^{m_k+1}(x, y))_{k\in\mathbb{N}}$ converges toward $\mathbf{0}$, which implies that the closure of the orbit of $(x, y)$ contains $\mathbf{0}$ and thus equals $\mathcal{K}$. □

## 6. Equivalence with positional number systems

Let $(U_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence of integers such that $U_0 = 1$. Such a sequence is called a *positional number system*. We assume furthermore that the set $L = 0^* \operatorname{rep}_U(\mathbb{N})$ of all the greedy representations of the integers is a regular language over a finite alphabet $A_U$ (from now on $\operatorname{rep}_U(n)$ denotes the $U$-representation of $n$ computed by the greedy algorithm with the most significant digit on the left). The finiteness of $A_U$ implies that the ratio $U_{n+1}/U_n$ is bounded. In particular, for $L = 0^* \operatorname{rep}_U(\mathbb{N})$ (or equivalently for $\operatorname{rep}_U(\mathbb{N})$) to be regular, it is shown in [40] that the sequence $(U_n)_{n\in\mathbb{N}}$ must satisfy a linear recurrence relation with constant coefficients. In [23], a sufficient condition is given in terms of the polynomial of the recurrence that $(U_n)_{n\in\mathbb{N}}$ satisfies. The reader can also see the special case treated in [27]. As an example, the set $\operatorname{rep}_U(\mathbb{N})$ is regular whenever the sequence $(U_n)_{n\in\mathbb{N}}$ satisfies a linear recurrence relation whose characteristic polynomial is the minimal polynomial of a Pisot number [8].

In this small section, we study the link between the odometer $\tau_L$ built over the language $L$ and the odometer $\tau_U$ presented in [21], with left-infinite sequences for the sake of consistency in notation, contrarily to the convention used in [21]. Notice that we allow leading zeroes in the greedy representations to obtain a language satisfying hypothesis (2).

**Remark 33.** Notice that, in this particular setting, as a consequence of the greedy algorithm, if $uv$ belongs to $L$ then $v$ belongs also to $L$.

**Proposition 34.** *Let $(U_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence of integers such that $U_0 = 1$, $\tau_U$ be the odometer associated to this sequence according to* [21]*, and let us assume that the language $L = 0^* \operatorname{rep}_U(\mathbb{N})$ associated to the numeration system built upon the sequence $(U_n)_{n\in\mathbb{N}}$ is regular. Let $p_1 : \mathcal{K} \to \widetilde{\mathcal{L}}$ be the projection mapping $(x, y)$ onto $x$. Then the following relation holds on $\mathcal{K}$:*

$$p_1 \circ \tau_L = \tau_U \circ p_1.$$

*Proof.* Let us first observe that the set on which $\tau_U$ is defined and acts, which is the set of left-infinite words satisfying the greedy property (1), is exactly $\widetilde{\mathcal{L}}$, following Remark 33.

Let $(x, y) = (\cdots x_1 x_0, \cdots y_1 y_0)$ be an element in $\mathcal{K} \setminus Max(\mathcal{K})$. Thus there exists $i$ such that $(x, y) \notin \mathbf{max}_i$.

Notice that if $u$ belongs to $L_q$, since $\mathcal{M}_L$ is accessible, then there exists $v$ such that $vu$ belongs to $L$. So thanks to Remark 33, $u$ also belongs to $L$.

Therefore $x_i \cdots x_0$ belongs to both $L_{y_{i+1}}$ and $L$. Since $x_i \cdots x_0$ does not belong to $Max(L_{y_{i+1}})$, then it does not belong to $Max(L)$ which means that $x_i U_i + \cdots + x_0 U_0$ is strictly less than $U_{i+1} - 1$. We set $x'_i \cdots x'_0 = \mathrm{Succ}_{y_{i+1}}(x_i \cdots x_0)$, so with our notation

$$\tau_L(\cdots x_{i+1}\, x_i \cdots x_0, y) = (\cdots x_{i+1}\, x'_i \cdots x'_0, y')$$

for some $y' \in {}^{\omega}Q$. We have to show that the successor of the word $x_i \cdots x_0$ in the genealogically ordered language $L$ is $x'_i \cdots x'_0$ which means therefore that

$$x_i U_i + \cdots + x_0 U_0 + 1 = x'_i U_i + \cdots + x'_0 U_0$$

and thus $\tau_U(\cdots x_{i+1}\, x_i \cdots x_0\,) = (\cdots x_{i+1}\, x'_i \cdots x'_0)$. To the contrary, assume that there exists $z_i \cdots z_0 \in L$ such that $x_i \cdots x_0 < z_i \cdots z_0 < x'_i \cdots x'_0$. Let $v$ be such that $\delta(q_0, v) = y_{i+1}$. The words $v x_i \cdots x_0$ and $v x'_i \cdots x'_0$ are accepted from $q_0$ and satisfy therefore the greedy condition (1). Since $z_i \cdots z_0 < x'_i \cdots x'_0$, $v z_i \cdots z_0$ satisfies the greedy condition and so it belongs to $L$. Since $\mathcal{M}_L$ is deterministic, $z_i \cdots z_0$ is also accepted from $y_{i+1}$. Therefore $x'_i \cdots x'_0 \neq \mathrm{Succ}_{y_{i+1}}(x_i \cdots x_0)$ which is a contradiction.

Consequently if $p_1$ is the projection mapping $(x, y)$ onto $x$, then we have shown that on $\mathcal{K} \setminus Max(\mathcal{K})$, the following holds

$$p_1 \circ \tau_L = \tau_U \circ p_1. \tag{5}$$

Observe that here, $\mathbf{0}$ is $({}^{\omega}0, {}^{\omega}q_0)$ because the empty word $\varepsilon$ is the representation of 0 and belongs to $L$. If $(x, y)$ belongs to $Max(\mathcal{K})$ then $\tau_L(x, y) = \mathbf{0}$ and it is clear that $\sum_{\ell=0}^{i} x_\ell\, U_\ell = U_{i+1} - 1$ for an infinite number of indices $i$. Therefore from [21], $\tau_U(x) = {}^{\omega}0$ and the relation (5) holds on the whole set $\mathcal{K}$. ☐

**Remark 35.** A characterization of the continuity of the odometer for positional number system is given in [18], in terms of the right subsequentiality of the successor function on $0^*L$. We will see in Proposition 46 that we can have $\tau_L$ continuous whereas $\tau_U$ is not continuous.

## 7. Substitution numeration systems

After having established in the previous section the connection with the classical odometer for positional number systems, we consider in the present section a second type of classical systems, namely the substitution numeration systems.

7.1. **Definition.** Let $\Sigma = \{a_1, \cdots, a_d\}$ be an alphabet (here, $\Sigma$ does not need to be totally ordered). Let $\sigma : \Sigma \to \Sigma^+$ be a *substitution*, i.e., a morphism of the free monoid $\Sigma^*$ such that $\sigma(a_1) \in a_1\Sigma^+$. To this substitution, we associate a deterministic automaton $\mathcal{M}_\sigma = (Q, a_1, A, \delta, F)$ in the classical way. The set of states is $Q = \Sigma \cup \{s\}$ where a sink state $s \notin Q$ is possibly added to $Q$ in order to make $\mathcal{M}_\sigma$ complete when $\sigma$ is not a uniform substitution (a substitution is said to be *uniform* if the images of all the letters have the same length). The alphabet of the automaton is

$$A = \{0, \ldots, \sup_{a \in \Sigma} |\sigma(a)| - 1\}.$$

There is an edge of label $i \in A$ between two states $a$ and $b$, that is, $\delta(a,i) = b$, if and only if the $(i+1)$-th letter in $\sigma(a)$ is $b$. The initial state is $a_1$ and all the states are final, i.e., $F = \Sigma$.

In the literature [9, 10, 11, 12, 31, 33] the notion of *prefix automaton* (respectively *prefix-suffix automaton*) can also be found: in this latter case, the label $i$ between $a$ and $b$ is replaced by the prefix of length $i$ of $\sigma(a)$ (respectively, the prefix of length $i$ of $\sigma(a)$ and the suffix of length $|\sigma(a)| - i - 1$ of $\sigma(a)$) (if $i = 0$ then the prefix is $\varepsilon$). It is well-known (see for instance [11, 12]) that each integer $n \geq 1$ has a unique decomposition of the form

$$n = \sum_{i=1}^{\ell} |\sigma^{i-1}(m_i)| \tag{6}$$

where $m_\ell \cdots m_1$ is the label of a path read in the prefix automaton from the initial state $a_1$ with $m_\ell \neq \varepsilon$.

Let us recall that $d$ denotes the cardinal of the alphabet $\Sigma$. The *incidence matrix* of the substitution $\sigma$ is defined as the $d \times d$ matrix whose entry of index $(a,b)$ counts the number of occurrences of the letter $a$ in $\sigma(b)$. The incidence matrix of $\sigma$ coincides with the transpose of the adjacency matrix of the automaton $\mathcal{M}_\sigma$.

Since the alphabet $A = \{0, 1, \ldots, \sup_{a\in\Sigma} |\sigma(a)| - 1\}$ is totally ordered by the usual ordering on $\mathbb{N}$, we can order the words of the language $L \subset A^*$ accepted by $\mathcal{M}_\sigma$ using the genealogical ordering. This leads to an abstract numeration system $S = (L, A, <)$ built upon $L$.

7.2. **Equivalence between substitution and abstract numeration systems.** In this section, we give a new interpretation of the numeration systems built upon a substitution according to (6).

Let $\mathcal{M}'_\sigma$ be the automaton built upon $\mathcal{M}_\sigma$ but having an extra state $a_0$ which is the initial state of this new automaton. For $k = 2, \ldots, |\sigma(a_1)|$, we add an edge labeled by $k-1$ from $a_0$ to the $k$-th letter of $\sigma(a_1)$. Observe that if $L$ is the language accepted by $\mathcal{M}_\sigma$ then $L \setminus 0A^*$ is the language accepted by $\mathcal{M}'_\sigma$. This kind of construction is also classical and was for instance used in [36]; the point here is to realize a one-to-one correspondence between the $n$-th word of the ordered language and the non-negative integer $n$; allowing leading zeroes introduces perturbations in the genealogical ordering, and thus in this one-to-one correspondence. We denote from now on by $L'$ the language accepted by $\mathcal{M}'_\sigma$.

**Example 36.** Consider the substitution on $\Sigma = \{a_1, a_2\}$ defined by $\sigma(a_1) = a_1a_2a_1$ and $\sigma(a_2) = a_1$. We have the following automata $\mathcal{M}_\sigma$ and $\mathcal{M}'_\sigma$ sketched in Figure 4. Here $A = \{0, 1, 2\}$ and the sink has not been represented.

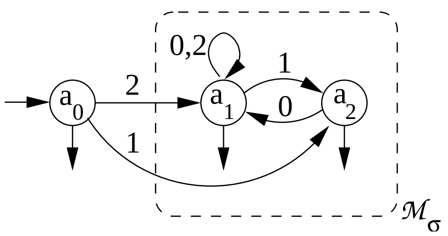


FIGURE 4. The automata $\mathcal{M}_\sigma$ and $\mathcal{M}'_\sigma$.

Naturally, we can also order the words of the language $L' \subset A^*$ accepted by $\mathcal{M}'_\sigma$ using the genealogical ordering. This leads to an abstract numeration system

$S' = (L', A, <)$ built upon $L'$. The representation of the integer $n$ is defined as the $(n+1)$-th word $w$ of $L'$ and we write $\mathrm{val}_{S'}(w) = n$ (let us recall that the first word of $L'$ is the empty word).

The following proposition allows us to make the connection with the substitutive numeration system as expressed in (6).

**Proposition 37.** *The $(n+1)$th word $w_1 \cdots w_\ell$ of the genealogically ordered language $L'$ generates the prefix $u_0 \cdots u_{n-1}$ of length $n$ of $\sigma^\omega(a_1)$ as follows: $u_0 \cdots u_{n-1}$ is equal to the concatenation of $\sigma^{\ell-i}[\delta(a_1, w_1 \cdots w_{i-1}0) \cdots \delta(a_1, w_1 \cdots w_{i-1}(w_i - 1))]$ in decreasing order of indices $1 \leq i \leq \ell$, where $w_1 \cdots w_{i-1}$ is understood as $\varepsilon$ if $i = 1$, as well as $\delta(a_1, w_1 \cdots w_i 0) \cdots \delta(a_1, w_1 \cdots w_i(w_i - 1))$ if $w_i = 0$. In other words, $u_0 \cdots u_{n-1}$ is equal to*

$$\sigma^{\ell-1}[\delta(a_1, 0) \cdots \delta(a_1, (w_1-1))] \cdots \sigma^0[\delta(a_1, w_1 \cdots w_{\ell-1}0) \cdots \delta(a_1, w_1 \cdots w_{l-1}(w_\ell-1))]$$

*and*

$$n = \sum_{i=0}^{\ell-1} \left|\sigma^i[\delta(a_1, w_1 \cdots w_{i-1}0) \cdots \delta(a_1, w_1 \cdots w_{i-1}(w_i - 1))]\right|.$$

*Proof.* The proof is based on the fact that the prefix of $\sigma(q)$ of length $t \leq |\sigma(q)|$ read from the state $q \in \Sigma$ is equal to $\delta(q, 0) \cdots \delta(q, t - 1)$.

Let us recall that for a state $q$, $\mathbf{u}_q(n)$ denotes the cardinal of the set of the words of length $n$ accepted from $q$. If $w = w_1 \cdots w_\ell \in L'$ (this means in particular that $w_1 > 0$), then with respect to the automaton $\mathcal{M}'_\sigma$ the following formula holds (see [24, 25])

$$\begin{aligned} \mathrm{val}_{S'}(w) = & \sum_{i=0}^{\ell-1} \mathbf{u}_{a_0}(i) + \sum_{b<w_1} \mathbf{u}_{\delta(a_0,b)}(\ell - 1) \\ & + \sum_{b<w_2} \mathbf{u}_{\delta(a_0,w_1 b)}(\ell - 2) + \cdots + \sum_{b<w_\ell} \mathbf{u}_{\delta(a_0,w_1\cdots w_{\ell-1}b)}(0). \end{aligned} \tag{7}$$

The interested reader can find a combinatorial interpretation of this formula in [22]. We have two immediate observations

$$\sum_{i=0}^{\ell-1} \mathbf{u}_{a_0}(i) = \mathbf{u}_{a_1}(\ell - 1) \quad \text{ and } \quad \forall q \in \Sigma, \forall n \in \mathbb{N}, \mathbf{u}_q(n) = |\sigma^n(q)|.$$

We are now able to prove the equivalence of the two formulas (6) and (7). First notice that $\delta(a_0, 0)$ is the sink $s$ of $\mathcal{M}'_\sigma$. Therefore $\mathbf{u}_{\delta(a_0,0)}(n) = 0$ for all $n$. If $b \neq 0$ then $\delta(a_0, b) = \delta(a_1, b)$. The first two terms in (7) can be written as

$$|\sigma^{\ell-1}(a_1)| + \sum_{0<b<w_1} \left|\sigma^{\ell-1}[\delta(a_1, b)]\right| = \left|\sigma^{\ell-1}[\delta(a_1, 0) \cdots \delta(a_1, w_1 - 1)]\right|.$$

Notice that for the latter equality, we have used the fact that $\delta(a_1, 0) = a_1$ and that $\sigma$ is a morphism. Consequently, (7) can be written as

$$\mathrm{val}_{S'}(w) = \sum_{i=1}^{\ell} \left|\sigma^{\ell-i}[\delta(a_1, w_1 \cdots w_{i-1}0) \cdots \delta(a_1, w_1 \cdots w_{i-1}(w_i - 1))]\right|.$$

This gives another interpretation of (6).

□

7.3. **First properties of the odometer.** Since the infinite language $L$ accepted by $\mathcal{M}_\sigma$ satisfies property (2), then one can consider the set $\mathcal{K}$ built upon $(L, A, <)$. Let us observe that $L'$ does not satisfy (2), but that $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{L}'}$ do coincide. The set $\mathcal{K}$ is a subshift of finite type of ${}^\omega(A \times Q)$ since every state (except the sink) in $\mathcal{M}_\sigma$ is a final state. Let us observe

**Proposition 38.** *The set $\mathcal{K}$ corresponds to the subset of ${}^\omega(A \times \Sigma)$ of sequences $(\cdots x_1 x_0, \cdots y_1 y_0)$ satisfying $\forall i \geq 0$, $y_{i+1}$ is the $(x_i + 1)$-th letter of $\sigma(y_i)$. Furthermore,*

$$Max(\mathcal{K}) = \{(x, y) \in \mathcal{K} |\ \forall i \geq 0,\ x_i = |\sigma(y_i)| - 1\},$$
$$Min(\mathcal{K}) = \{(x, y) \in \mathcal{K} |\ \forall i \geq 0,\ x_i = 0\}.$$

*Proof.* There is an edge in the automaton $\mathcal{M}_\sigma$ of label $i$ between two states $u$ and $v$ if and only if the $(i+1)$-th letter in $\sigma(u)$ is $v$; furthermore, all the states are initial and final. Hence $\widetilde{\mathcal{L}}$ is equal to the set of mirror images of labels of right-infinite paths in the automaton $\widetilde{\mathcal{M}_\sigma}$ obtained by reversing the transition relations in $\mathcal{M}_\sigma$, which implies the desired description of $\mathcal{K}$. The characterization of $Max(\mathcal{K})$ and $Min(\mathcal{K})$ is immediate.

□

Let us recall that the odometer $\tau$ is one-to-one from $\mathcal{K} \setminus Max(\mathcal{K})$ onto $\mathcal{K} \setminus Min(\mathcal{K})$ following Corollary 29.

**Remark 39.** The automaton $\mathcal{M}_\sigma$ associated to a substitution $\sigma$ is not necessarily minimal. Indeed, if you consider the Thue-Morse substitution defined by $\sigma(a_1) = a_1 a_2$ and $\sigma(a_2) = a_2 a_1$ then it is easy to see that $\mathcal{M}_\sigma$ accept $\{0, 1\}^*$ and is not minimal. To obtain unambiguous constructions, we have always considered the minimal automaton of a language but clearly, we could define a set $\mathcal{K}$ and an odometer $\tau$ depending on the choice of a finite deterministic automaton which is not necessarily minimal. We just need a loop in the initial state $a_1$ labeled by the smallest letter 0 of the alphabet (it is always the case for the automaton associated to a substitution $\sigma$ which satisfies $\sigma(a_1) \in a_1 A^+$).

7.4. **The Pisot case.** Nevertheless, there are some cases for which the automaton $\mathcal{M}_\sigma$ can be proved to be minimal. A substitution is said of *Pisot type* if the eigenvalues of its incidence matrix satisfy the following: there exists a dominant eigenvalue $\alpha$ such that for every other eigenvalue $\lambda$, one gets $0 < |\lambda| < 1 < |\alpha|$. A substitution of Pisot type is primitive and the characteristic polynomial $\chi_\sigma$ of its incidence matrix is irreducible over $\mathbb{Q}$ [10]. Let us recall that $\chi_\sigma$ is also the minimal polynomial of the adjacency matrix of $\mathcal{M}_\sigma$.

**Proposition 40.** *Let $\sigma$ be a Pisot substitution. The automaton $\mathcal{M}_\sigma$ is minimal.*

*Proof.* Let $\sigma$ be a substitution of Pisot type. The automaton $\mathcal{M}_\sigma$ is accessible since $\sigma$ is primitive, that is, all its states can be reached from its initial state $a_1$. Hence the minimal polynomial $\chi_\sigma$ of its adjacency matrix is dividable by the minimal polynomial of the minimal automaton recognizing the language $M_\sigma$. Since $\chi_\sigma$ is irreducible, this implies that both polynomials do coincide, and thus that $\mathcal{M}_\sigma$, which is deterministic, is the minimal automaton recognizing the language $M_\sigma$. □

In the particular case of a Pisot substitution, we are now able to give a dynamical interpretation of $(\mathcal{K}, \tau)$. Let $S$ denote the shift map on $\Sigma^{\mathbb{Z}}$: $S((w_i)_{i\in\mathbb{Z}}) = (w_{i+1})_{i\in\mathbb{Z}}$. A word $u \in \Sigma^{\mathbb{Z}}$ such that there exists a positive integer $k$ with $S^k(u) = u$ is called a *periodic point* under the action of $\sigma$. Let us recall that the (two-sided) *symbolic dynamical system* generated by a primitive substitution $\sigma$ is the pair $(X_\sigma, S)$, where

$X_\sigma$ is the set of two-sided sequences in $\Sigma^{\mathbb{Z}}$ with the same set of factors of any periodic point $u$ of $\sigma$; this definition does not depend on the choice of $u$ by primitivity of $\sigma$.

We use here the notation and results of [9, 10] adapted to our framework. Following [29], every two-sided sequence $v$ in $X_\sigma$ has a unique decomposition

$$v = S^k(\sigma(w)), \text{ with } w \in X_\sigma \text{ and } 0 \le k < |\sigma(w_0)|$$

($w_0$ denotes here the 0-th coordinate of $w$).

Let

$$\begin{cases} \theta: \ X_\sigma \to X_\sigma : \ v \mapsto w, \\ \text{where } v = S^k(\sigma(w)), \text{ with } 0 \le k < |\sigma(w_0)|. \end{cases}$$

The map $\theta$ is called the *desubstitution map*.

Let

$$\begin{cases} \gamma: \ X_\sigma \to (A \times \Sigma) : \ v \mapsto (k+1, w_0), \\ \text{where } v = S^k(\sigma(w)), \text{ with } w \in X_\sigma \text{ and } 0 \le k < |\sigma(w_0)|. \end{cases}$$

In other words, if $\gamma(v) = (k, q)$, then $v_0$ is the $(k+1)$-th letter of $\sigma(q)$. Hence, for every $v \in X_\sigma$, the mirror image of the sequence $(\gamma \circ \theta^i(v))_{i \in \mathbb{N}}$ is easily seen to belong to $\mathcal{K}$. Let us now define

$$\Gamma : X_\sigma \to \mathcal{K} : \ v \mapsto \widetilde{(\gamma \circ \theta^i(v))}_{i \in \mathbb{N}}.$$

The following theorem is a direct consequence of [9, 10].

**Theorem 41.** [9, 10] *Let $\sigma$ be a Pisot substitution. The map $\Gamma$ is continuous and onto $\mathcal{K}$; it is one-to-one except on the orbit of periodic points of $\sigma$. Furthermore,*

$$\Gamma \circ S = \tau \circ \Gamma \text{ and } \Gamma \circ \theta = S_{\mathcal{K}} \circ \Gamma,$$

*where $S_{\mathcal{K}}$ denotes the shift map acting on elements of ${}^{\omega}(A \times \Sigma)$.*

*Proof.* We know from [9] that $(X_\sigma, S)$ is measure-theoretically isomorphic with the subshift of finite type $\mathcal{D}$ defined as the set of labels of infinite paths $\mathcal{D}$ in the mirror image of the prefix-suffix automaton defined as follows: there is an edge from $a$ to $b$ of label $(p, a, s)$ if $\sigma(b) = a$, and all the states (which are the letters of $\Sigma$) are both initial and final. Let us prove that $\mathcal{K}$ and $\mathcal{D}$ are in one-to-one correspondence. This comes from the fact that the following map is one-to-one:

$$\mathcal{K} \to \mathcal{D}, \ (x, y) \mapsto (\delta(y_{i+1}, 0) \cdots \delta(y_{i+1}, x_i - 1), y_i, s_i)_{i \in \mathbb{N}},$$

where $s_i$ is the suffix of size $|\sigma(y_{i+1})| - x_i - 1$ of $\sigma(y_{i+1})$. Now from Remark 25, the map $\tau$ coincides with the adic transformation acting on $\mathcal{D}$. It just remains to apply the results of [9, 10]. □

**Remark 42.** Two dynamical systems can be built over $\mathcal{K}$, i.e., $(\mathcal{K}, \tau)$ and $(\mathcal{K}, S_{\mathcal{K}})$. Theorem 41 gives us two combinatorial interpretations for these systems: the action of the desubstitution map $\theta$ (the "inverse" of $\sigma$) on $X_\sigma$ corresponds to the action of the shift $S_{\mathcal{K}}$ on $\mathcal{K}$, whereas the action of the shift $S$ on $X_\sigma$ corresponds to action of the odometer $\tau$ on $\mathcal{K}$.

**Remark 43.** There exists furthermore a unique shift invariant measure on the dynamical system $(X_\sigma, S)$ since $\sigma$ is primitive ($(X_\sigma, S)$ is said uniquely ergodic); for more details see for instance [32]. This measure can be naturally carried on $(\mathcal{K}, \tau)$ via the map $\Gamma$ (which is one-to-one except on a countable number of points). Theorem 41 means that $(\mathcal{K}, \tau)$ endowed with this measure is measure-theoretically isomorphic with $(X_\sigma, S)$. One interest of this approach is that it provides us some insight on a metrical study of $(\mathcal{K}, \tau)$, according to [3].

**Remark 44.** It is possible to give a combinatorial interpretation of $Min(\mathcal{K})$ and $Max(\mathcal{K})$ in this framework. Following [9], $Min(\mathcal{K})$ and $Max(\mathcal{K})$ correspond respectively to the periodic points (under the action of $\sigma$) of $X_\sigma$ (we denote this set $\mathrm{Per}(X_\sigma)$) and to the preimages $S^{-1}(\mathrm{Per}(X_\sigma))$ under the shift $S$ of those periodic points. Both sets do not have necessarily the same cardinal as illustrated for instance in Section 8, Proposition 46, below.

## 8. The case of sofic beta-numerations

This section gathers results of Section 6 and 7 within the framework of $\beta$-numeration. Let $U = (U_n)_{n\in\mathbb{N}}$ be a positional numeration system such that the ratio $U_{n+1}/U_n$ is bounded, as defined in Section 6. We add now the following extra hypothesis of right-extendibility: $\mathrm{rep}_U(\mathbb{N})0^*$ is included in $\mathrm{rep}_U(\mathbb{N})$; the positional number system $U$ is then said to be a *Bertrand numeration system*. Bertrand numeration systems are closely related to $\beta$-expansions as recalled below.

Let $\beta > 1$ be a positive real number. The Rényi $\beta$-expansion of a real number $x \in [0,1]$ is defined as the sequence $(x_i)_{i\geq 1}$ with values in $\{0,1,\ldots,\lceil\beta\rceil - 1\}$ produced by the $\beta$-transformation $T_\beta : [0,1] \to [0,1] : \ x \mapsto \beta x \pmod 1$ as follows

$$\forall i \geq 1,\ x_i = \lfloor \beta T_\beta^{i-1}(x)\rfloor, \text{ and thus } x = \sum_{i\geq 1} x_i \beta^{-i}.$$

Let $d_\beta(1) = (t_i)_{i\geq 1}$ denote the $\beta$-expansion of 1. Let $d_\beta^*(1) = d_\beta(1)$, if $d_\beta(1)$ is infinite, and $d_\beta^*(1) = (t_1 \ldots t_{m-1}(t_m - 1))^\omega$, if $d_\beta(1) = (t_1 \ldots t_{m-1} t_m)$ is finite (with $t_m \neq 0$). The set $D_\beta$ of $\beta$-expansions of numbers in $[0,1)$ is exactly the set of sequences $(c_i)_{i\geq 1}$ that satisfy:

$$\forall k \in \mathbb{Z},\ (c_i)_{i\geq k} <_{\mathrm{lex}} d_\beta^*(1).$$

For more details, see for instance [28]. We denote by $F(D_\beta)$ the set of finite factors of the sequences in $D_\beta$.

Numbers $\beta$ such that $d_\beta(1)$ is ultimately periodic are called *Parry numbers* and those such that $d_\beta(1)$ is finite are called *simple Parry numbers*. If $\beta$ is a Parry number (simple or not), the minimal automaton $\mathcal{M}_\beta$ recognizing the set of factors of $F(D_\beta)$ can easily be constructed (representations of this classical automaton $\mathcal{M}_\beta$ can be found in [20] or [25]). Furthermore, let us recall that when $\beta$ is assumed to be Pisot, then $\beta$ is either a Parry number or a simple Parry number, and $(X_\beta, S)$ is sofic.

Bertrand numeration systems are characterized by the following theorem:

**Theorem 45.** [5] *Let $U$ be a positional number system over a finite alphabet. Then $U$ is a Bertrand numeration system if and only if there exists a real number $\beta > 1$ such that $L = 0^* \mathrm{rep}_U(\mathbb{N}) = F(D_\beta)$. Furthermore, $L$ is regular if and only if $\beta$ is a Parry number.*

There is a natural way to associate a substitution $\sigma_\beta$ with the $\beta$-numeration when $\beta$ is a Parry number (simple or not). These substitutions will be called in all what follows *$\beta$-substitutions*. The automaton $\mathcal{M}_{\sigma_\beta}$ associated with $\sigma_\beta$, as defined in Section 7, coincides with the minimal automaton $\mathcal{M}_\beta$ which recognizes $F(D_\beta)$. For more details, see [15, 45]. Let us note that $d_\beta(1)$ cannot be purely periodic, hence one has either $d_\beta^*(1) = (t_1 \cdots t_{n-1}(t_n - 1))^\omega$ with $t_n \neq 0$ or $d_\beta^*(1) = t_1 \cdots t_n (t_{n+1} \cdots t_{n+p})^\omega$, with $t_n \neq t_{n+p}$ and $n \geq 1$.

- Assume $d_\beta(1) = (t_1 \cdots t_{n-1} t_n)$ with $t_n \neq 0$ and thus $d_\beta^*(1) = (t_1 \cdots t_{n-1}(t_n - 1))^\omega$. Consider the substitution $\sigma_\beta$ defined over the alphabet $\{1, 2, \ldots, n\}$

by:

$$\sigma_\beta : \begin{cases} 1 & \mapsto 1^{t_1}2 \\ 2 & \mapsto 1^{t_2}3 \\ \vdots & \vdots \\ n-1 & \mapsto 1^{t_{n-1}}n \\ n & \mapsto 1^{t_n}. \end{cases}$$

- Assume $d_\beta(1) = d^*_\beta(1) = t_1 \cdots t_n(t_{n+1} \cdots t_{n+p})^\omega$, with $t_{n+1} \cdots t_{n+p} \neq 0^p$ and $t_n \neq t_{n+p}$. Furthermore $n \geq 1$. Consider the substitution $\sigma_\beta$ defined over the alphabet $\{1, 2, \ldots, n+p\}$ by:

$$\sigma_\beta : \begin{cases} 1 & \mapsto 1^{t_1}2 \\ 2 & \mapsto 1^{t_2}3 \\ \vdots & \vdots \\ n+p-1 & \mapsto 1^{t_{n+p-1}}(n+p) \\ n+p & \mapsto 1^{t_{n+p}}(n+1). \end{cases}$$

We assume from now on that the positional number system $U$ is a Bertrand numeration associated with $\beta$ Pisot number; thus $L = 0^* \operatorname{rep}_U(\mathbb{N}) = F(D_\beta)$ is an infinite regular language which satisfies (2). We still denote $\tau_L$ the odometer acting on $\mathcal{K}$. When $\beta$ is a simple Parry number such that the length of $d_\beta(1)$ coincides with its degree, then the substitution $\sigma_\beta$ is of Pisot type since the characteristic polynomial of its incidence matrix coincides with the minimal polynomial of $\beta$. Hence, the results of Section 7.4 do apply.

We end now this section by proving that the odometer $\tau_L$ is continuous when $\beta$ is a Pisot number, contrarily to the positional number systems case where continuity holds if and only if $\beta$ is a simple Parry number (see [21] and [18]).

**Proposition 46.** *Let $\beta$ be a Pisot number and let $L = F(D_\beta)$. Then the odometer $\tau_L$ is continuous on $\mathcal{K}$.*

*Proof.* We know from Proposition 30 that $\tau_L$ is continuous on $\mathcal{K} \setminus Max(\mathcal{K})$.

Let $(x, y) \in Max(\mathcal{K})$. Let us prove that for any sequence $(x^{(n)}, y^{(n)})_{n\in\mathbb{N}}$ with values in $\mathcal{K}$ which converges toward $(x, y)$, then $\tau_L(x^{(n)}, y^{(n)})_{n\in\mathbb{N}}$ converges toward $\mathbf{0} = ({}^\omega 0, {}^\omega q_0)$. Let $(x^{(n)}, y^{(n)})_{n\in\mathbb{N}}$ be such a sequence. We assume furthermore that for $n$ large enough, then $(x^{(n)}, y^{(n)}) \notin Max(\mathcal{K})$. There exists a state $q$ for which there exist infinitely many integers $k$ such that $x_k x_{k-1} \cdots x_0 \in Max(L_q)$. Let $N$ be fixed. Let $k \geq N$ such that $x_k x_{k-1} \cdots x_0 \in Max(L_q)$ with $q = y_{k+1}$. For $n$ large enough, $(x^{(n)}, y^{(n)})$ coincides on its first $N$ values with $(x, y)$ and $(x^{(n)}, y^{(n)}) \notin Max(\mathcal{K})$. In particular, $x^{(n)}_k x^{(n)}_{k-1} \cdots x^{(n)}_0 \in Max(L_q)$, with $q = y^{(n)}_{k+1}$. Since $(x^{(n)}, y^{(n)}) \notin Max(\mathcal{K})$, there exists a non-negative integer $l > k$ such that $x^{(n)}_l x^{(n)}_{l-1} \cdots x^{(n)}_0 \notin Max(L_{y^{(n)}_{l+1}})$. Let $l_0$ denote the smallest of these integers. The successor in $L_{y^{(n)}_{l_0+1}}$ of $x^{(n)}_{l_0} x^{(n)}_{l_0-1} \cdots x^{(n)}_0$ is $(x^{(n)}_{l_0} + 1)0^{l_0}$. Furthermore, any edge labeled by 0 in $\mathcal{M}_\beta$ leads to the initial state $q_0$ (recall that the interested reader can find a representation of $\mathcal{M}_\beta$ in [20, 25]). Hence $\tau_L(x^{(n)}, y^{(n)})$ admits as a suffix $(0^N, y^N_0)$ for $n$ large enough, which ends the proof. □

Let us illustrate on an example the main difference between the present situation and the positional number system one, for what concerns continuity. We consider Example 5 in [21].

**Example 47.** Let $\beta$ be the Pisot number $\frac{3+\sqrt{5}}{2}$. One has $d_\beta(1) = 2(1)^\omega$, hence $\beta$ is a Parry number. The corresponding substitution $\sigma_\beta$ is defined as $\sigma_\beta(a) = aab$, $\sigma_\beta(b) = ab$. The minimal automaton of the language $L = F(D_\beta)$ is depicted in

Figure 5. One has $Max(L_a) = \{21^n : n \in \mathbb{N}\}$ and $Max(L_b) = \{1^n : n \in \mathbb{N}\}$, hence

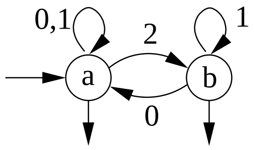


FIGURE 5. The automaton accepting $F(D_\beta)$, with $\beta = \frac{3+\sqrt{5}}{2}$.

$Max(\mathcal{K}) = \{({}^\omega 1, {}^\omega b)\}$, hence $\tau_L({}^\omega 1, {}^\omega b) = \mathbf{0} = ({}^\omega 0, {}^\omega a)$.

The odometer $\tau_U$ based on the associated positional number system $U$ is not continuous on ${}^\omega 1$. Indeed, let $x^{(n)} := {}^\omega(0)21^n$, for $n \in \mathbb{N}$; the sequence $(x^{(n)})_{n\in\mathbb{N}}$ takes its values in $\widetilde{\mathcal{L}}$ and converges to ${}^\omega 1$. One has $\tau_U(x^{(n)}) = {}^\omega(0)10^{n+1}$, which converges to ${}^\omega 0$, whereas $\tau_U({}^\omega 1) = {}^\omega 12$, hence the discontinuity of $\tau_U$ on ${}^\omega 1$. Nevertheless, for every given $n$, $x^{(n)}$ has a unique antecedent in $\mathcal{K}$ with respect to the projection $p_1$: one has $p_1({}^\omega(0)21^n, {}^\omega(a)b^n) = x^{(n)}$; ${}^\omega 1$ admits exactly two antecedents: one has $p_1({}^\omega 1, {}^\omega a) = p_1({}^\omega 1, {}^\omega b) = {}^\omega 1$. The sequence $({}^\omega(0)21^n, {}^\omega(a)b^n)_{n\in\mathbb{N}}$ converges to $({}^\omega 1, {}^\omega b)$; one checks that $\tau_L({}^\omega(0)21^n, {}^\omega(a)b^n) = ({}^\omega(0)10^{n+1}, {}^\omega a)$ which tends to $({}^\omega 0, {}^\omega a) = \mathbf{0} = \tau_L(({}^\omega 1, {}^\omega b))$. Notice furthermore that $\tau_L({}^\omega 1, {}^\omega a) = ({}^\omega 12, {}^\omega ab)$.

**Remark 48.** The sets $\mathcal{K}$ and $\widetilde{\mathcal{L}}$ are not in one-to-one correspondence. Indeed the word ${}^\omega 0$ admits several representations in $\mathcal{K}$ (see for instance Example 47 above where the word ${}^\omega 0$ admits as representations $({}^\omega 0, {}^\omega a)$ or $({}^\omega 0, {}^\omega ba^n)$ for every $n$). Hence we cannot deduce directly continuity results from Proposition 34. Let us observe nevertheless that there is at most a countable number of antecedents to elements of $\widetilde{\mathcal{L}}$ according to the projection $p_1 : \mathcal{K} \to \widetilde{\mathcal{L}}$, $(x, y) \mapsto x$, in the particular situation described in the present section.

## 9. Real representation of the odometer

The aim of this section is to outline the first steps of a study of a geometric representation of the dynamical system $(\mathcal{K}, \tau)$. A *geometric representation* of the dynamical system $(\mathcal{K}, \tau)$ is a continuous map $\varphi$ from $\mathcal{K}$ onto a geometric dynamical system $(Y, T)$ such that $\varphi \circ \tau = T \circ \varphi$, and on which there exists a partition indexed by the alphabet $A \times Q$ such that every word $(x, y)$ is the itinerary (represented as a left-infinite sequence) of a point of $(Y, T)$ with respect to the partition. This question is natural in the framework of arithmetics dynamics [41] : we will mainly consider here situations where we "encode" rotations of the torus by producing explicit arithmetic expansions of real numbers based on our abstract numeration systems; see also see [31].

Let $L$ be an arbitrary regular language satisfying (2). A first representation which might be possible consists in extending the work of [24, 25, 37], where a real value is attributed to limits of finite words for abstract numeration systems built on an exponential regular language satisfying the following condition: there exist $\beta > 1$ and $P \in \mathbb{R}[X]$ such that for all states $q \in Q$, there exist some non-negative real numbers $a_q$ such that $\lim_{n\to\infty} \frac{\mathbf{u}_q(n)}{P(n)\beta^n} = a_q$. (We recall that $\mathbf{u}_q(n)$ represents the number of words of length $n$ in $L_q$.) We assume now that $\widetilde{L}$ satisfies this condition. (Clearly, if $L$ is exponential then $\widetilde{L}$ is also exponential because $\#(L \cap \Sigma^n) = \#(\widetilde{L} \cap \Sigma^n)$.) The main assumptions for building a representation map rely therefore on the asymptotic behavior of the sequences $\frac{\mathbf{u}_{q'}(n)}{P(n)\beta^n}$ for all the states $q'$ of the minimal automaton of $\widetilde{L}$. Let $\mathbf{v}(n)$ denote the number of words of length at most $n$ in $L$ (or in $\widetilde{L}$), and $\mathrm{val}_{\widetilde{L}}(w)$ the numerical value of $w \in \widetilde{L}$, i.e., if

$\mathrm{val}_{\tilde{L}}(w) = n$, then $w$ is the $(n+1)$-th word of $\widetilde{L}$. Let $(x,y) \in \mathcal{K}$. Since $x \in \widetilde{\mathcal{L}}$, there exists a sequence $(w_n)_{n\in\mathbb{N}}$ of words in $\widetilde{L}$ which converges to $\tilde{x}$. The limit

$$\lim_{n\to\infty} \frac{\mathrm{val}_{\widetilde{L}}(w_n)}{\mathbf{v}(|w_n|)}$$

does not depend on the choice of the sequence $(w_n)_{n\in\mathbb{N}}$ (see [24, Corollary 8]) and is denoted by $\mathrm{val}_\infty(x)$. We thus can define a map $r : \mathcal{K} \to \mathbb{R} : (x,y) \mapsto \mathrm{val}_\infty(x)$. It remains to explore the properties of such a representation map $r$.

We propose now a second possible geometric representation in the framework of languages associated with substitutions. We continue here with the notation of Section 7. A substitution is said *unimodular* if the determinant of its incidence matrix equals $\pm 1$. It is conjectured that for a Pisot unimodular substitution, the dynamical system $(X_\sigma, S)$, and hence $(\mathcal{K}, \tau)$ according to Remark 43, is measure-theoretically isomorphic to a rotation on the torus $\mathbb{T}^{d-1}$, where $d$ denotes the cardinal of the alphabet $\Sigma$. There are numerous families of substitutions for which this result is known to hold true. For more details, see for instance Chap. 7 of [31]. One simple way to exhibit this rotation is to give a geometrical representation of $(\mathcal{K}, \tau)$ as explained in the next paragraph. We follow the formalism of [9, 10].

We assume now that $\sigma$ is either a Pisot unimodular substitution (over the alphabet $\Sigma$ of cardinal $d$), or that $\sigma$ is a $\beta$-substitution associated to $\beta$ Pisot unit (of degree $d$), according to Section 8. Now $L$ denotes either, in the first case, the language associated with $\sigma$, as explained in Section 7, or $L = F(D_\beta)$, in the second case. Let $\alpha_1, \ldots, \alpha_r$ denote the $r$ real conjugates of $\beta$, and $\alpha_{r+1}, \ldots, \alpha_{r+s}$, $\overline{\alpha_{r+1}}, \ldots, \overline{\alpha_{r+s}}$, denote its $2s$ complex conjugates ($r+2s = d$). Let us assume $\alpha_1 > 1$, hence, $|\alpha_i| < 1$, for $i \geq 2$.

We consider now eigenvectors either of the incidence matrix of $\sigma$ for the substitution case, or of the transpose of the adjacency matrix of the minimal automaton $\mathcal{M}_L$ recognizing $L$, in the beta-numeration case. Let $\vec{v}^{(1)}$ be a left eigenvector associated with the eigenvalue $\alpha_1$ with coefficients in the field $\mathbb{Q}(\alpha_1)$. Let $\alpha_k$ be an eigenvalue and let $\rho_k$ be the canonical morphism from $\mathbb{Q}(\alpha_1)$ onto $\mathbb{Q}(\alpha_k)$, extended to $\mathbb{Q}(\alpha_1)^d$. Let $\vec{v}^{(k)} = \rho_k(\vec{v}^{(1)})$. For $k, j = 1, \cdots, d$, $v_j^{(k)}$ denotes the $j$-th coordinate of $\vec{v}^{(k)}$. We propose as a geometric representation of the set $\mathcal{K}$ in this framework the following map $\varphi : \mathcal{K} \to \mathbb{R}^{r-1} \times \mathbb{C}^s$

$$(x,y) \mapsto \left( \begin{array}{ll} \sum_{i\geq 0} \alpha_2^i (v^{(2)}_{\delta(y_{i+1},0)} & + \cdots + v^{(2)}_{\delta(y_{i+1},x_i-1)}), \cdots, \\ & , \cdots, \alpha_{r+s}^i (v^{(r+s)}_{\delta(y_{i+1},0)} + \cdots + v^{(r+s)}_{\delta(y_{i+1},x_i-1)}) \end{array} \right).$$

The series involved are easily seen to converge. This map can be factorized as a map on the torus. Indeed, let $\mathbb{L}$ denote the lattice

$$\left\{ \sum_{k=1}^{d} n_k \vec{w}^{(k)} \mid n_k \in \mathbb{Z}, \sum_{k=1}^{d} n_k = 0 \right\},$$

where $\vec{w}^{(k)}$ is the vector of $\mathbb{R}^{r-1} \times \mathbb{C}^s$ whose coordinates are $(v_k^{(1)}, \cdots, v_k^{(r+s)})$. Following [9, 10], the map

$$\varphi_{\mathbb{L}} : \mathcal{K} \to (\mathbb{R}^{r-1} \times \mathbb{C}^s)/\mathbb{L} \equiv \mathbb{T}^{d-1},\ (x,y) \mapsto \varphi(x,y) \bmod \mathbb{L}$$

is well defined and continuous. Consider now the toral translation

$$T :\ (\mathbb{R}^{r-1} \times \mathbb{C}^s)/\mathbb{L} \to (\mathbb{R}^{r-1} \times \mathbb{C}^s)/\mathbb{L} : z \mapsto z + \vec{w}^{(1)} \bmod \mathbb{L}.$$

One checks that $\varphi_{\mathbb{L}} \circ \tau = T \circ \varphi_{\mathbb{L}}$, and that $\varphi_{\mathbb{L}}$ is continuous and onto.

In particular, for some families of $\beta$-substitutions, this map is known to provide a measure-theoretical isomorphism (this is the case in particular for numbers $\beta$ having the finiteness property (F) introduced in [19], which states that the set of non-negative real numbers with finite $\beta$-expansion coincides with the set of non-negative elements of $\mathbb{Z}[1/\beta]$). We deduce the following proposition from the results of [19, 1] (see also [43, 44] for connected results):

**Proposition 49.** *Assume that $L = F(D_\beta)$, where $\beta > 1$ is either*

- *the positive root of the polynomial $X^m - t_1 x^{m-1} - \cdots - 1$, where $t_i \in \mathbb{Z}$, and $t_1 \geq t_2 \geq \cdots \geq t_m > 0$,*
- *the dominant root of the polynomial $X^m - t_1 x^{m-1} - \cdots - 1$, where $t_i \in \mathbb{N}$, and $t_1 > \sum_{i=2}^{d} |t_i| > 0$, and $(t_1, t_2) \neq (2, -1)$,*
- *a cubic Pisot unit.*

*Then the map $\varphi_{\mathbb{L}}$ is one-to-one except on an at most countable number of points and is a geometrical representation of $(\mathcal{K}, \tau)$, the partition being given by the sets $\varphi_{\mathbb{L}}(\{(x,y) |\ (x,y) \in \mathcal{K},\ y_0 = q\})$, $q \in Q$.*

*Proof.* The fact that $\varphi_{\mathbb{L}}$ is one-to-one except on an at most countable number of points comes from [19] for the first case, and from [1] for the last two points. It remains to prove that the sets $\varphi_{\mathbb{L}}\{(x,y) |\ (x,y) \in \mathcal{K},\ y_0 = q\}$, $q \in Q$ are disjoint up to sets of zero measure. This is a direct consequence of the fact that $\beta$-substitutions satisfy the so-called strong coincidence condition, according to [2]. □

## 10. Some special cases

In [21], the odometer is defined on a set $\mathcal{R}$ of sequences of digits. Here, we have introduced an odometer on a set $\mathcal{K}$ of pairs of infinite words. In this section, we show that in some particular situations, we can restrict ourselves to unidimensional sequences. So we exhibit hypothesis where the extra information given by the sequence of states is useless. The interest relies on the fact that the odometer can be directly defined on $\widetilde{\mathcal{L}}$, similarly to what occurs in the case of positional number systems (Proposition 34).

**Definition 50.** Let $d \geq 1$. A regular language $L$ is said to be *$d$-synchronizing* if there exists a function $f : A^d \to Q$ such that for any word $w \in A^*$ of length $d$ and any $q \in Q$, $\delta(q, w)$ is equal to $f(w)$ (let us recall that $\delta$ denotes the transition function of the minimal automaton of $L$). In other words, for any element $(x, y) = (x_0 x_1 \cdots, y_0 y_1 \cdots)$ in $\mathcal{K}$, for all $i \geq 0$ the state $y_i$ is completely determined by $x_i \cdots x_{i+d-1}$. A language is *synchronizing* if there exists a positive integer $d$ such that $L$ is $d$-synchronizing. Otherwise stated, this means that $y$ can be deduced from $x$.

As a consequence, we obtain that the projection map $p_1 : \mathcal{K} \to \widetilde{\mathcal{L}}\ (x, y) \mapsto x$ is injective, which implies, following Proposition 11, that both sets $\mathcal{K}$ and $\widetilde{\mathcal{L}}$ are in one-to-one correspondence.

**Example 51.** Consider the language accepted by the automaton $\mathcal{M}_\sigma$ depicted in Figure 4 of Example 36. Here, we represent in Figure 6 the automaton $\mathcal{M}_\sigma$. This

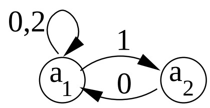


Figure 6. The automaton $\mathcal{M}_\sigma$.

language is 1-synchronizing. Indeed, assume that $(x, y)$ is an element in $\mathcal{K}$. The factors possibly appearing in $x$ are 00, 01, 02, 10, 20, 21 and 22. Actually, 11 and 12 cannot occur in $x$ because no infinite path in the automaton depicted in Figure 6 contains such a factor. Clearly, if $x_i \in \{0,2\}$ then $y_i = a_1$ and if $x_i = 1$ then $y_i = a_2$.

**Example 52.** Continuing Examples 36 and 51, the language accepted by $\mathcal{M}_\sigma$ depicted in Figure 4 is 1-synchronizing and we have the function

$$f : 0 \mapsto a_1,\ 1 \mapsto a_2,\ 2 \mapsto a_1.$$

For this automaton $\mathcal{M}_\sigma$, we have $\delta(a_1,0) = \delta(a_2,0) = f(0) = a_1$, $\delta(a_2,0) = f(1) = a_2$, $\delta(a_2,1)$ is the sink, $\delta(a_1,2) = f(2) = a_1$ and $\delta(a_2,2)$ is also the sink.

## Acknowledgment

We would like to warmly thank the referees who helped us to improve significantly the readability of this paper.

## References

[1] S. Akiyama, H. Rao, W. Steiner, A certain finiteness property of Pisot number systems, *J. Number Theory* **107** (2004), 135–160.
[2] P. Arnoux, S. Ito, Pisot substitutions and Rauzy fractals, *Bull. Belg. Math. Soc. Simon Stevin* **8** (2001), 181–207.
[3] G. Barat, T. Downarowicz, A. Iwanik, P. Liardet, Propriétés topologiques et combinatoires des échelles de numération, *Colloq. Math.* **84/85** (2000), 285–306.
[4] G. Barat, T. Downarowicz, P. Liardet, Dynamiques associées à une échelle de numération, *Acta Arith.* **103** (2002), 41–78.
[5] A. Bertrand-Mathis, Comment écrire les nombres entiers dans une base qui n'est pas entière, *Acta Math. Acad. Sci. Hungar.* **54** (1989), 237–241.
[6] O. Bratteli, Inductive limits of finite-dimensional $C^*$-algebras, *Trans. Amer. Math. Soc.* **171** (1972), 195–234.
[7] H. Bruin, K. Heller, M. St. Pierre, Adding machines and wild attractors, *Ergodic Theory Dynam. Systems* **17** (1997), 1267–1268.
[8] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, Latin American Theoretical INformatics (Valparaíso, 1995), *Theoret. Comput. Sci.* **181** (1997), 17–43.
[9] V. Canterini, A. Siegel, Automate des préfixes-suffixes associé à une substitution primitive, *J. Théor. Nombres Bordeaux* **13** (2001), 353–369.
[10] V. Canterini, A. Siegel, Geometric representation of substitutions of Pisot type, *Trans. Amer. Math. Soc.* **353** (2001), 5121-5144.
[11] J.-M. Dumont, A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theoret. Comput. Sci.* **65** (1989), 153–169.
[12] J.-M. Dumont, A. Thomas, Digital sum moments and substitutions, *Acta Arith.* **64** (1993), 205–225.
[13] F. Durand, B. Host, C. Skau, Substitutional dynamical systems, Bratteli diagrams and dimension groups, *Ergodic Theory Dynam. Systems* **19** (1999), 953–993.
[14] S. Eilenberg, *Automata, Languages and Machines*, Vol. A, Academic Press, New York, (1974).
[15] S. Fabre, Substitutions et $\beta$-systèmes de numération, *Theoret. Comput. Sci.* **137** (1995), 219–236.
[16] A. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985), 105–114.
[17] C. Frougny, On-line odometers for two-sided symbolic dynamical systems, *Proceedings of DLT 2002, Lect. Notes in Comput. Sci.* **2450** (2002), 405–416.
[18] C. Frougny, On the sequentiality of the successor function, *Inform. and Comput.* **139** (1997), 17–38.
[19] C. Frougny and B. Solomyak, Finite beta-expansions, *Ergodic Theory Dynam. Systems* **12**, (1992), 713–723.
[20] C. Frougny and B. Solomyak, On representation of integers in linear numeration systems, *Ergodic theory of $Z_d$ actions* (Warwick, 1993–1994), 345–368, London Math. Soc. Lecture Note Ser. **228**, Cambridge Univ. Press, Cambridge, 1996.
[21] P. J. Grabner, P. Liardet, R. F. Tichy, Odometers and systems of numeration, *Acta Arith.* **70** (1995), 103–123.
[22] P. J. Grabner, M. Rigo, Additive functions with respect to numeration systems on regular languages, *Monatsh. Math.* **139** (2003), 205–219.
[23] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Syst.* **31** (1998), 111–133.

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

[25] P. Lecomte, M. Rigo, Real numbers having ultimately periodic representations in abstract numeration systems, *Inform. and Comput.* **192** (2004), 57–83.

[26] A. N. Livshits, A sufficient condition for weak mixing of substitutions and stationary adic transformations, *Mat. Zametki* **44** (1988), 785–793; English transl. in *Math. Notes* **44** (1988), 920–925.

[27] N. Loraud, $\beta$-shift, systèmes de numération et automates, *J. Théor. Nombres Bordeaux* **7** (1995), 473–498.

[28] M. Lothaire, *Algebraic Combinatorics on words*, Cambridge University Press, Cambridge, (2002).

[29] B. Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques, *Bull. Soc. Math. France* **124** (1996), 329–346.

[30] G. Păun, A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995), 257–270.

[31] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lect. Notes in Math. **1794**, Springer-Verlag, Berlin, (2002).

[32] M. Queffélec, *Substitution dynamical systems. Spectral analysis*, Lect. Notes in Math. **1294**, Springer-Verlag, (1987).

[33] G. Rauzy, Sequences defined by iterated morphisms, *Sequences (Naples/Positano, 1988)*, 275–286, Springer, New York, 1990.

[34] M. Rigo, Numeration systems on a regular language: arithmetic operations, recognizability and formal power series, *Theoret. Comput. Sci.* **269** (2001), 469–498.

[35] M. Rigo, Construction of regular languages and recognizability of polynomials, *Discrete Math.* **254** (2002), 485–496.

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

[37] M. Rigo, W. Steiner, Abstract $\beta$-expansions and ultimately periodic representations, to appear in *J. Théor. Nombres Bordeaux*.

[38] J. Sakarovitch, *Éléments de théorie des automates*, Vuibert, (2003).

[39] K. Schmidt, Algebraic coding of expansive group automorphisms and two-sided beta-shifts, *Monatsh. Math.* **129** (2000), 37–61.

[40] J. Shallit, Numeration systems, Linear recurrences, and Regular sets, *Inform. and Comput.* **113** (1994), 331–347.

[41] N. Sidorov, Arithmetic dynamics, in *Topics in dynamics and ergodic theory*, S. Bezuglyi et al. (ed.), Cambridge University Press, Lond. Math. Soc. Lect. Note Ser. **310** (2003), 145–189.

[42] N. Sidorov, A. Vershik, Ergodic properties of the Erdös measure, the entropy of the golden shift, and related problems, *Monatsh. Math.* **126** (1998), 215–261.

[43] B. Solomyak, On the spectral theory of adic transformations, *Adv. Soviet Math.* **9** (1992), 217–230.

[44] B. Solomyak, Substitutions, adic transformations and beta-expansions, in *Symbolic Dynamics and its applications*, P. Walters Editor, Contemporary Mathematics **135** (1992), 361–372.

[45] W. P. Thurston, Groups, tilings and finite state automata, Lectures notes distributed in conjunction with the Colloquium Series, in *AMS Colloquium lectures*, (1989).

[46] A. M. Vershik, Uniform algebraic approximation of shift and multiplication operators, *Soviet Math. Dokl.* **24** (1981), 97–100.

[47] A. M. Vershik, A theorem on the Markov periodical approximation in ergodic theory, *J. Sov. Math.* **28** (1985), 667–674.

[48] A. M. Vershik, A. N. Livshits, Adic models of ergodic transformations, spectral theory, substiutions, and related topics, *Adv. Soviet Math.* **9** (1992), 185–204.



(V. Berthé)
LIRMM, Université Montpellier II
161 rue Ada
34392 Montpellier Cedex 5
France.
*E-mail address*: `berthe@lirmm.fr`

(M. Rigo)
Université de Liège
Institut de Mathématiques
Grande Traverse 12 (B 37)
B-4000 Liège
Belgium.
*E-mail address*: M.Rigo@ulg.ac.be