Automatic sequences based on Parry or Bertrand numeration systems

Adeline Massuir[1], Jarkko Peltomäki[2,3,4], and Michel Rigo[1]

[1]University of Liège, Department of Mathematics, Liège, Belgium
[2]The Turku Collegium for Science and Medicine TCSM, University of Turku, Turku, Finland
[3]Turku Centre for Computer Science TUCS, Turku, Finland
[4]University of Turku, Department of Mathematics and Statistics, Turku, Finland

Abstract

We study the factor complexity and closure properties of automatic sequences based on Parry or Bertrand numeration systems. These automatic sequences can be viewed as generalizations of the more typical k-automatic sequences and Pisot-automatic sequences. We show that, like k-automatic sequences, Parry-automatic sequences have sublinear factor complexity while there exist Bertrand-automatic sequences with superlinear factor complexity. We prove that the set of Parry-automatic sequences with respect to a fixed Parry numeration system is not closed under taking images by uniform substitutions or periodic deletion of letters. These closure properties hold for k-automatic sequences and Pisot-automatic sequences, so our result shows that these properties are lost when generalizing to Parry numeration systems and beyond. Moreover, we show that a multidimensional sequence is U-automatic with respect to a positional numeration system U with regular language of numeration if and only if its U-kernel is finite.

1 Introduction

Roughly speaking, an automatic sequence is an infinite word over a finite alphabet such that its nth symbol is obtained as the output given by a deterministic finite automaton fed with the representation of n in a suitable numeration system. Precise definitions are given in Subsection 2.2.

If we consider the usual base-k numeration systems, then we get the family of k-automatic sequences [1]. These words are images under a coding of a fixed point of a substitution of constant length. On a larger scale, if one considers abstract numeration systems based on a regular language (see for instance [4, Chap. 3] or [20]), then we get exactly the family of morphic words. Morphic words are images under a coding of a fixed point of an arbitrary substitution. Between these two extremes, we have the automatic sequences based on Pisot, Parry, and Bertrand numeration systems (the definitions are given in Subsection 2.1), and we have the following hierarchy:

$$\begin{aligned}&\text{Integer base systems} \subsetneq \text{Pisot systems} \subsetneq \text{Parry systems}\\ \subsetneq\ &\text{Bertrand systems with a regular numeration language} \subsetneq \text{Abstract numeration systems.}\end{aligned}$$

Abstract numeration systems are uniquely based on the genealogical ordering of the words belonging to a regular language. This is contrasting with the more restricted case, treated in this paper, of positional numeration systems based on an increasing sequence of integers: a digit occurring in nth position is multiplied by the nth element of the underlying sequence.

E-mail addresses: a.massuir@uliege.be (A. Massuir), r@turambar.org (J. Peltomäki), m.rigo@uliege.be (M. Rigo).

The Pisot-automatic sequences behave in many respects like k-automatic sequences. Most importantly, in both cases automatic sets have a characterization in terms of first-order logic. This characterization in the k-automatic case is due to Büchi [6], and it was generalized to the Pisot case by Bruyère and Hansel [5]; also see [20, Chap. 3] and the references therein. Now by using the logical characterization, it is particularly straightforward to show that both the class of k-automatic sequences and the class of Pisot-automatic sequences enjoy many closure properties. For instance, both classes are closed under taking images by a uniform substitution and under periodic deletion of letters. For k-automatic sequences, these are classical results of Cobham [7]. The proofs of these results presented in [1, Chap. 6.8] are straightforward to generalize to Pisot-automatic sequences given the logical characterization of [5]. For more closure properties, see [1, Chap. 6.8].

In this paper, we study if some properties common to k-automatic sequences and Pisot-automatic sequences also hold for Parry-automatic sequences or more general automatic sequences. In a sense, we show that the generalization to Pisot numeration systems is the broadest possible generalization if the goal is to preserve the many good properties of k-automatic sequences.

It has been known before that a logical characterization no longer exists for Parry-automatic sequences. This is follows from [11, Example 3]; we shall return to this matter in Section 4. We show that the closure properties mentioned above break when generalizing from Pisot to Parry and obtain as a corollary yet another proof showing that no logical characterization indeed exists for these sequences.

In combinatorics on words and in symbolic dynamics, the factor complexity of infinite words is often of interest. It was famously shown by Pansiot [17] that the factor complexity of an infinite word generated by a substitution is in one of the following five classes: $\mathcal{O}(1)$, $\Theta(n)$, $\Theta(n \log\log n)$, $\Theta(n \log n)$, or $\Theta(n^2)$. Previously, it has been known that the factor complexity of a k-automatic sequence is sublinear (that is, it is in $\mathcal{O}(1)$ or $\Theta(n)$) [7], [1, Thm. 10.3.1]. We extend this result and show that the factor complexity of any Parry-automatic sequence is sublinear. In contrast, we show by an explicit example that there exists a Bertrand-automatic sequence of superlinear complexity.

A well-known result concerning k-automatic sequences is their characterization in terms of the k-kernel originally due to Eilenberg [8]. This was generalized in [19] for all sequences associated with abstract numeration systems. The multidimensional version of this generalization [19, Prop. 32] however needs an additional assumption that is not required in the k-automatic case. We show in this paper that this additional assumption is unnecessary also for positional numeration systems with a regular numeration language.

This paper is organized as follows. In Section 2, we recall needed results and notation on numeration systems and automatic sequences. Then in Section 3 we study the factor complexity of Parry-automatic sequences, and in Section 4, we show that the closure properties of Pisot-automatic sequences do not hold for Parry-automatic sequences. The paper is concluded by Section 5, where the relationship of U-automaticity and the finiteness of the U-kernel is studied in the multidimensional setting.

2 Basics

2.1 Background on Numeration Systems

For general references on numeration systems and words, we refer the reader to [4, 15, 20]. Let us first consider the representation of integers. A *positional numeration system*, or simply, a *numeration system*, is an increasing sequence $U = (U_n)_{n\geq 0}$ of integers such that $U_0 = 1$ and $C_U := \sup_{n\geq 0}\lceil U_{n+1}/U_n\rceil < +\infty$. We let A_U be the integer alphabet $\{0, \ldots, C_U - 1\}$. The *greedy representation* of the positive integer n is the word $\operatorname{rep}_U(n) = w_{\ell-1} \cdots w_0$ over A_U satisfying

$$\sum_{i=0}^{\ell-1} w_i\, U_i = n,\ w_{\ell-1} \neq 0, \quad \text{and} \quad \forall j \in \{1, \ldots, \ell\}, \quad \sum_{i=0}^{j-1} w_i\, U_i < U_j.$$

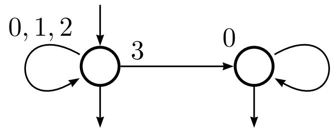

Figure 1: The canonical automaton accepting $\{0,1,2\}^*(\{\varepsilon\}\cup 3\,0^*)$.

We set $\mathrm{rep}_U(0)$ to be the empty word ε. The language $\mathrm{rep}_U(\mathbb{N})$ is called the *numeration language*. A set X of integers is *U-recognizable* if $\mathrm{rep}_U(X)$ is regular, i.e., accepted by a finite automaton. The *numerical value* $\mathrm{val}_U\colon \mathbb{Z}^* \to \mathbb{N}$ maps a word $d_{\ell-1}\cdots d_0$ over any alphabet of integers to the number $\sum_{i=0}^{\ell-1} d_i U_i$.

Recall that the *genealogical ordering* orders words from a language first by length and then by the lexicographic ordering (induced, in this paper, typically by the natural order of the digits).

Definition 2.1. A numeration system U is a *Bertrand numeration system* if, for all $w \in A_U^+$, $w \in \mathrm{rep}_U(\mathbb{N})$ if and only if $w0 \in \mathrm{rep}_U(\mathbb{N})$.

Example 1. The usual base-k numeration system $(k^n)_{n\geq 0}$ is a Bertrand numeration system. Taking $F_0 = 1$, $F_1 = 2$, and $F_{n+2} = F_{n+1} + F_n$ for $n \geq 0$ gives the Fibonacci numeration system $F = (F_n)_{n\geq 0}$, which is a Bertrand numeration system: $\mathrm{rep}_F(\mathbb{N}) = 1\{0,01\}^* \cup \{\varepsilon\}$. If we slightly modify the Fibonacci system by taking the initial conditions $U_0 = 1$, $U_1 = 3$, we get a numeration system $(U_n)_{n\geq 0} = (1,3,4,7,11,18,29,47,\ldots)$, which is no longer a Bertrand system. Indeed, 2 is the greedy representation of an integer but 20 is not because $\mathrm{rep}_U(\mathrm{val}_U(20)) = 102$.

Example 2. The numeration system B given by the recurrence $B_n = 3B_{n-1} + 1$ for all $n \geq 1$ and $B_0 = 1$ is such that $0^*\,\mathrm{rep}_B(\mathbb{N}) = \{0,1,2\}^*(\{\varepsilon\}\cup 3\,0^*)$; see [13, p. 131]. The automaton accepting the language $0^*\,\mathrm{rep}_B(\mathbb{N})$ is depicted in Figure 1. By its simple form, it is obvious that it is a Bertrand numeration system. Notice that the sequence $(B_n)_{n\geq 0}$ also satisfies the homogeneous linear recurrence $B_n = 4B_{n-1} - 3B_{n-2}$.

There is a link between the representation of integers and the representation of real numbers. Let $\beta > 1$ be a real number. The *β-expansion* of a real number $x \in [0,1]$ is the sequence $d_\beta(x) = (x_i)_{i\geq 1} \in \mathbb{N}^\omega$ that satisfies

$$x = \sum_{i=1}^{+\infty} x_i \beta^{-i}$$

and which is the maximal element in $\mathbb{N}^\omega$ having this property with respect to the lexicographic order over $\mathbb{N}$. Notice that β-expansions can be obtained by a greedy algorithm and they only contain letters (digits) over the alphabet $A_\beta = \{0,\ldots,\lceil\beta\rceil - 1\}$. By $\mathrm{Fact}(D_\beta)$, we denote the set of finite factors occurring in the β-expansions of the real numbers in $[0,1)$.

Definition 2.2. If $d_\beta(1) = t_1\cdots t_m 0^\omega$, with $t_1,\ldots,t_m \in A_\beta$ and $t_m \neq 0$, then we say that $d_\beta(1)$ is *finite* and we set $d^*_\beta(1) = (t_1\cdots t_{m-1}(t_m - 1))^\omega$. Otherwise, we set $d^*_\beta(1) = d_\beta(1)$. An equivalent definition is to set $d^*_\beta(1) = \lim_{x\to 1^-} d_\beta(x)$. When $d^*_\beta(1)$ is (ultimately) periodic, then β is said to be a *Parry number*.

Definition 2.3. Let $\beta > 1$ be a real number such that $d^*_\beta(1) = (t_i)_{i\geq 1}$. The numeration system $U_\beta = (U_n)_{n\geq 0}$ canonically associated with β is defined by

$$U_n = t_1 U_{n-1} + \cdots + t_n U_0 + 1, \quad \forall n \geq 0. \tag{1}$$

As a consequence of Bertrand's theorem, see [3] or [4, Chap. 2], the numeration system U_β associated with β satisfies

$$0^*\,\mathrm{rep}_{U_\beta}(\mathbb{N}) = \mathrm{Fact}(D_\beta). \tag{2}$$

Thus for all $\beta > 1$, the canonical numeration system U_β associated with β is a Bertrand numeration system because $w \in \mathrm{Fact}(D_\beta)$ if and only if $w0 \in \mathrm{Fact}(D_\beta)$ for all $w \in A_\beta^*$.

Remark 1. If β is a Parry number, then the canonical numeration system U_β satisfies a linear recurrence equation with integer coefficients which can be obtained from (1).

Definition 2.4. A numeration system U is a *Parry numeration system* if there exists a Parry number β such that $U = U_\beta$.

Example 3. Consider the Golden mean $\beta = (1+\sqrt{5})/2$. We have $d_\beta(1) = 11$ and $d_\beta^*(1) = (10)^\omega$, so the Golden mean is a Parry number. It is straightforward to deduce from (1) that the associated Parry numeration system is the Fibonacci system of Example 1 defined by the recurrence $F_{n+2} = F_{n+1} + F_n$ and the initial conditions $F_0 = 1$, $F_1 = 2$.

Lemma 2.5. *The set of Parry numeration systems is a strict subset of the set of Bertrand numeration systems.*

Proof. We have already deduced from (2) that Parry numeration systems are Bertrand numeration systems. Now consider the Bertrand system $B = (B_n)_{n\ge 0}$ of Example 2. We will show that there is no $\beta > 1$ such that $B = U_\beta$. Proceed by contradiction. Assume that there exists β such that B is the numeration system canonically associated with β. The greatest word of length n for the lexicographical order in $0^* \mathrm{rep}_B(\mathbb{N})$ is 30^{n-1}. Consequently, we have $1 = 3/\beta$. The Parry numeration system U_3 is the classical base-3 system and $0^* \mathrm{rep}_{U_3}(\mathbb{N}) = \{0, 1, 2\}^*$, which differs from $0^* \mathrm{rep}_B(\mathbb{N})$. This is a contradiction. □

Theorem 2.6 (Parry [16])**.** *A sequence $x = (x_i)_{i\ge 1}$ over $\mathbb{N}$ is the β-expansion of a real number in $[0, 1)$ if and only if $(x_{n+i})_{i\ge 1}$ is lexicographically less than $d_\beta^*(1)$ for all $n \ge 0$.*

As a consequence of this result, with any Parry number β is canonically associated a deterministic finite automaton $\mathcal{A}_\beta = (Q_\beta, q_{\beta,0}, F_\beta, A_\beta, \delta_\beta)$ accepting the language $\mathrm{Fact}(D_\beta)$. Otherwise stated, the numeration language of a Parry numeration system is regular. This automaton $\mathcal{A}_\beta$ has a special form. Let $d_\beta^*(1) = t_1 \cdots t_i (t_{i+1} \cdots t_{i+p})^\omega$ where $i \ge 0$ and $p \ge 1$ are the minimal preperiod and period respectively. The set of states of $\mathcal{A}_\beta$ is $Q_\beta = \{q_{\beta,0}, \ldots, q_{\beta,i+p-1}\}$. All states are final. For every $j \in \{1, \ldots, i + p\}$, we have t_j edges $q_{\beta,j-1} \to q_{\beta,0}$ labeled by $0, \ldots, t_j - 1$ and, for $j < i + p$, one edge $q_{\beta,j-1} \to q_{\beta,j}$ labeled by t_j. There is also an edge $q_{\beta,i+p-1} \to q_{\beta,i}$ labeled by t_{i+p}. See, for instance, [12, 20]. Note that in [15, Thm. 7.2.13], $\mathcal{A}_\beta$ is shown to be the trim minimal automaton of $\mathrm{Fact}(D_\beta)$.

Remark 2. Let β be a Parry number. The automaton $\mathcal{A}_\beta$ is well-known to be primitive (see, e.g., [14, Chap. 3]). Indeed, the periodic part of $d_\beta^*(1)$ contains at least a nonzero digit. Consequently, there is a path from every state of $\mathcal{A}_\beta$ to the initial state $q_{\beta,0}$. Moreover, there is a loop on $q_{\beta,0}$ with label 0. Hence $\mathcal{A}_\beta$ is irreducible (i.e., strongly connected) and aperiodic (the gcd of length of the cycles going through any state is 1). The conclusion follows.

A *Pisot number* is an algebraic integer $\beta > 1$ whose conjugates have modulus less than 1. A *Salem number* is an is an algebraic integer $\beta > 1$ whose conjugates have modulus less than or equal to 1 and at least one has modulus equal to 1. If β is a Pisot number, then U_β has many interesting properties [5, 10, 20]: $\mathrm{rep}_{U_\beta}(\mathbb{N})$ is regular, normalization $w \mapsto \mathrm{rep}_{U_\beta}(\mathrm{val}_{U_\beta}(w))$ (and thus addition) is computable by a finite automaton, and U_β-recognizable sets are characterized in terms of first order logic.

Definition 2.7. A numeration system U is a *Pisot numeration system* if there exists a Pisot number β such that $U = U_\beta$.

Remark 3. Pisot numbers are Parry numbers [2, 21], so Pisot numeration systems belong to the set of Parry numeration systems. The numeration system (4) studied in Section 4 is Parry but not Pisot, so this inclusion is strict. Further, the Fibonacci numeration system of Example 1 is a Pisot numeration system because the Golden mean (the largest root of the polynomial $X^2 - X - 1$) is a Pisot number. As this numeration system is clearly not a base-k numeration system, we see that base-k numeration systems are strictly included in Pisot numeration systems.

Remark 4. Let $\mathcal{B}$ be the set of Bertrand numeration systems. Let $\mathcal{R}$ be the set of numeration systems U whose numeration language $\text{rep}_U(\mathbb{N})$ is regular. The three sets $\mathcal{B} \cap \mathcal{R}$, $\mathcal{B} \setminus \mathcal{R}$ and $\mathcal{R} \setminus \mathcal{B}$ are nonempty. For instance, the modified Fibonacci system of Example 1 belongs to $\mathcal{R} \setminus \mathcal{B}$. All Parry numeration systems and the Bertrand system of Example 2 belong to $\mathcal{B} \cap \mathcal{R}$.

If β is not a Parry number, for instance when β is transcendental, then the numeration language $\text{rep}_{U_\beta}(\mathbb{N})$ is not regular even though U_β is a Bertrand system. Hence $\mathcal{B} \setminus \mathcal{R}$ is nonempty.

We will often make the assumption that we are dealing with positional numeration systems U such that $\text{rep}_U(\mathbb{N})$ is regular. This is particularly important when we will deal with finite U-kernels in Section 5. This assumption is in fact somewhat restrictive. In [22], Shallit proves that if $\text{rep}_U(\mathbb{N})$ is regular, then $(U_n)_{n\geq 0}$ must satisfy a homogeneous linear recurrence.

2.2 Automatic Sequences

Definition 2.8. Let U be a numeration system. An infinite word $\mathbf{x} = (x_n)_{n\geq 0}$ over an alphabet B is *U-automatic*, i.e., it is an *U-automatic sequence*, if there exists a complete DFAO (deterministic finite automaton with output) $(Q, q_0, A_U, \delta, \tau)$ with transition function $\delta\colon Q \times A_U \to Q$ and output function $\tau\colon Q \to B$ such that $\delta(q_0, 0) = q_0$ and

$$x_n = \tau(\delta(q_0, \text{rep}_U(n))), \quad \forall n \geq 0.$$

The infinite word $\mathbf{x}$ is *k-automatic* (resp. *Parry-automatic*, *Bertrand-automatic*) if $U = (k^n)_{n\geq 0}$ for an integer $k \geq 2$ (resp. U is a Parry numeration system, resp. U is a Bertrand numeration system). Properties of Parry-automatic sequences are discussed in [9].

The next result is classical, see for instance [18].

Theorem 2.9. *Let U be a numeration system such that* $\text{rep}_U(\mathbb{N})$ *is regular. An infinite word* $\mathbf{x} = (x_n)_{n\geq 0}$ *over A is U-automatic if and only if, for all $a \in A$, the set $\{j \geq 0 \mid x_j = a\}$ is U-recognizable.*

Let $k \geq 2$ be an integer. The *k-kernel* of an infinite word $\mathbf{x} = (x_n)_{n\geq 0}$ over A is the set of its subsequences of the form

$$\{(x_{k^e n+d})_{n\geq 0} \in A^{\mathbb{N}} \mid e \geq 0, 0 \leq d < k^e\}.$$

Observe that an element of the k-kernel is obtained by considering those indices whose base-k expansions end with $\text{rep}_k(d)$ (possibly preceded by some zeroes to get a suffix of length e). With this in mind, we introduce the more general U-kernel of an infinite word.

Definition 2.10. Let U be a numeration system and $s \in A_U^*$ be a finite word. Define the ordered set of integers

$$\mathcal{I}_s := \text{val}_U(0^* \text{rep}_U(\mathbb{N}) \cap A_U^* s) = \{i(s,0) < i(s,1) < \cdots\}. \tag{3}$$

Depending on s, it is possible for this set to be finite or empty. The *U-kernel* of an infinite word $\mathbf{x} = (x_n)_{n\geq 0}$ over B is the set

$$\ker_U(\mathbf{x}) := \{(x_{i(s,n)})_{n\geq 0} \mid s \in A_U^*\}.$$

With the above remark, this set can contain finite or even empty subsequences.

The next two results have been obtained in the general framework of abstract numeration systems; see [19, Prop. 7] and [19, Prop. 9].

Proposition 2.11. *Let U be a numeration system such that* $\text{rep}_U(\mathbb{N})$ *is regular. A word $\mathbf{x}$ is U-automatic if and only if its U-kernel is finite.*

Proposition 2.12. *Let U be a numeration system. If an infinite word is U-automatic, then it is reversal-U-automatic, i.e., its nth term is obtained by reading the reversal of* $\text{rep}_U(n)$ *in a DFAO.*

Notice that the proof of the latter result only relies on classical constructions on automata defined from the DFAO generating the U-automatic sequence. The same construction applies in multidimensional setting, and we shall make use of this in Section 5.

3 Factor Complexity

The *factor complexity function* $p_{\mathbf{x}}(n)$ of an infinite word $\mathbf{x}$ counts the number of factors of length n occurring in $\mathbf{x}$. For more on factor complexity, see [4, Chap. 3].

Let us recall the following result of Cobham.

Proposition 3.1. *[7], [1, Thm. 10.3.1] The factor complexity function of a k-automatic sequence is sublinear.*

Next we generalize this result. For the proof, we need the following definition and result.

Definition 3.2. Let $\sigma\colon A^* \to A^*$ be a substitution. If there exists a number α, $\alpha \geq 1$, such that $|\sigma^n(a)| = \Theta(\alpha^n)$ for all $a \in A$, then we say that σ is *quasi-uniform.*

Proposition 3.3. *[17], [4, Thm. 4.7.47] The factor complexity of a fixed point of a quasi-uniform substitution is sublinear.*

Theorem 3.4. *The factor complexity function of a Parry-automatic sequence is sublinear.*

Proof. Let U be a Parry numeration system having canonical automaton $\mathcal{A}$, and let $\mathbf{x}$ be an U-automatic sequence generated by a DFAO $\mathcal{B}$. The product automaton $\mathcal{A}\times\mathcal{B}$ has $Q_{\mathcal{A}\times\mathcal{B}} := Q_{\mathcal{A}}\times Q_{\mathcal{B}}$ as set of states, the initial state q_0 is the pair made of the initial states of $\mathcal{A}$ and $\mathcal{B}$, and the transition function is given by

$$\delta_{\mathcal{A}\times\mathcal{B}}((q,q'),i) = (\delta_{\mathcal{A}}(q,i), \delta_{\mathcal{B}}(q',i)).$$

We consider the automaton $\mathcal{A}\times\mathcal{B}$ as a DFAO by setting that the output function τ maps a state $(q_{\mathcal{A}}, q_{\mathcal{B}})$ of $\mathcal{A}\times\mathcal{B}$ to the output of the state $q_{\mathcal{B}}$ of $\mathcal{B}$. It is clear that $\mathbf{x}$ is generated by $\mathcal{A}\times\mathcal{B}$.

Based on the automaton $\mathcal{A}\times\mathcal{B}$, we can build a substitution σ and consider the output function τ as a coding such that $\mathbf{x} = \tau(\sigma^\omega(q))$ for some $q \in Q_{\mathcal{A}\times\mathcal{B}}$. The construction is classical, see for instance [20, Lemma 2.28]. The substitution σ is defined as follows

$$\sigma((q_{\mathcal{A}},q_{\mathcal{B}})) = (\delta_{\mathcal{A}}(q_{\mathcal{A}},0),\delta_{\mathcal{B}}(q_{\mathcal{B}},0))(\delta_{\mathcal{A}}(q_{\mathcal{A}},1),\delta_{\mathcal{B}}(q_{\mathcal{B}},1))\cdots(\delta_{\mathcal{A}}(q_{\mathcal{A}},C_U-1),\delta_{\mathcal{B}}(q_{\mathcal{B}},C_U-1)).$$

In the latter expression, since $\mathcal{A}$ is in general not complete, if $\delta_{\mathcal{A}}(q_{\mathcal{A}},j)$ is undefined, then $(\delta_{\mathcal{A}}(q_{\mathcal{A}},j),\delta_{\mathcal{B}}(q_{\mathcal{B}},j))$ is replaced by ε. Notice that the substitution σ is defined over the alphabet $Q_{\mathcal{A}\times\mathcal{B}}$. Since $\mathcal{A}\times\mathcal{B}$ has a loop with label $(0,0)$ on its initial state q_0, iterating σ on this state generates the sequence of states $\sigma^\omega(q_0)$ in $\mathcal{A}\times\mathcal{B}$ reached from the initial state by the words of $\mathrm{rep}_U(\mathbb{N})$ in genealogical order.

Since every state of a canonical automaton of a Parry numeration system is final, the coding τ is non-erasing. Then by [4, Lemma 4.6.6], the factor complexity of $\mathbf{x}$ is at most the factor complexity of $\sigma^\omega(q_0)$, so it is sufficient to show that a fixed point of σ has sublinear complexity. This is accomplished as follows. First we establish that there exists a number α such that $|\sigma^n(q)| = \Theta(\alpha^n)$ for every state $q \in Q_{\mathcal{A}\times\mathcal{B}}$. In other words, we show that the substitution σ is *quasi-uniform.* It then follows from Proposition 3.3 that the factor complexity of a fixed point of σ is sublinear.

Let us define a projection mapping $\varphi\colon Q_{\mathcal{A}\times\mathcal{B}} \to Q_{\mathcal{A}}$ by setting $\varphi((q_{\mathcal{A}},q_{\mathcal{B}})) = q_{\mathcal{A}}$ if $(q_{\mathcal{A}},q_{\mathcal{B}})$ is a state of $\mathcal{A}\times\mathcal{B}$. By the definition of the product automaton $\mathcal{A}\times\mathcal{B}$, we have $\varphi(\delta_{\mathcal{A}\times\mathcal{B}}((q_{\mathcal{A}},q_{\mathcal{B}}),a)) = \delta_{\mathcal{A}}(\varphi((q_{\mathcal{A}},q_{\mathcal{B}})),a)$ for all letters a and all states $q_{\mathcal{A}}$ and $q_{\mathcal{B}}$.

Recall that given an automaton $\mathcal{C}$ with adjacency matrix $Adj(\mathcal{C})$, the entry $(Adj(\mathcal{C}))^n_{i,j}$ counts the number of distinct paths of length n from state i to state j; see [14, Chap. 2]. Let $(q_{\mathcal{A}},q_{\mathcal{B}})$ be a state of $\mathcal{A}\times\mathcal{B}$ and consider all paths of length n starting from this state. These paths can be identified with their edge labels. Given such a path with edge label w, we find by applying the projection mapping φ a path in $\mathcal{A}$ with edge label w starting at the state $q_{\mathcal{A}}$. Conversely, given a path of length n in $\mathcal{A}$ with edge label w starting at state $q_{\mathcal{A}}$, there exists a path with edge label w in $\mathcal{A}\times\mathcal{B}$ starting at the state $(q_{\mathcal{A}},q_{\mathcal{B}})$ because the automaton $\mathcal{B}$ is complete (see Definition 2.8).

Denoting the total number of paths of length n starting at a state q of $\mathcal{A} \times \mathcal{B}$ by $K_q(n)$, we have thus argued that

$$K_q(n) = \sum_{r \in Q_{\mathcal{A} \times \mathcal{B}}} (Adj(\mathcal{A} \times \mathcal{B}))^n_{q,r} = \sum_{s \in Q_{\mathcal{A}}} (Adj(\mathcal{A}))^n_{\varphi(q),s}.$$

The canonical automaton of a Parry numeration system is primitive (recall Remark 2), we have for each i and j that $((Adj(\mathcal{A}))^n_{i,j} = \Theta(\alpha^n)$, where α is the Perron–Frobenius eigenvalue of $\mathcal{A}$. Thus $K_q(n) = \Theta(\alpha^n)$. By rephrasing the number $K_q(n)$ in terms of substitutions, we have $|\sigma^n(q)| = K_q(n)$. Hence $|\sigma^n(q)| = \Theta(\alpha^n)$, and we are done. □

Notice that in fact we showed in the proof of Theorem 3.4 that for each Parry-automatic sequence $\mathbf{x}$ there exists a coding τ and a quasi-uniform substitution σ such that $\mathbf{x} = \tau(\sigma^\omega(a))$ for a letter a. This should be contrasted with the fact that k-automatic sequences are codings of fixed points of *uniform* substitutions.

From Lemma 2.5, there exist Bertrand numeration systems that are not Parry numeration systems. We show that Theorem 3.4 does not generalize to Bertrand-automatic sequences.

Theorem 3.5. *There exists a Bertrand-automatic sequence with superlinear factor complexity.*

Proof. Consider the numeration system given by the recurrence $B_n = 3B_{n-1} + 1$ with $B_0 = 1$. In Example 2, it was shown that this numeration system is a Bertrand numeration system.

The substitution associated with the canonical automaton, depicted in Figure 1, is $\sigma\colon a \mapsto aaab$, $b \mapsto b$; see in the proof of Theorem 3.4 how this substitution is defined. Let $\mathbf{x}$ be the infinite fixed point of σ. Observe that $\mathbf{x}$ is a Bertrand-automatic sequence. It is easy to see that ab^na occurs in $\mathbf{x}$ for all $n \geq 0$. Thus $\mathbf{x}$ is aperiodic and there exists infinitely many bounded factors occurring in $\mathbf{x}$ (a factor is w *bounded* if the sequence $(|\sigma^n(w)|)_{n\geq 0}$ is bounded). It follows by Pansiot's theorem [17] that the factor complexity of $\mathbf{x}$ is quadratic; see also [4, Thm. 4.7.66]. □

We do not have examples of Bertrand-automatic sequences with factor complexities $\Theta(n \log\log n)$ or $\Theta(n \log n)$.

4 Closure Properties

It is easy to see that the image of a k-automatic sequence $\mathbf{x} \in A^\omega$ under a substitution $\mu\colon A \to B^*$ of constant length ℓ is again a k-automatic sequence. Indeed, Theorem 2.9 implies that for all $a \in A$ there exists a first-order formula $\varphi_a(n)$ in $\langle \mathbb{N}, +, V_k \rangle$ which holds if and only if $\mathbf{x}[n] = a$. Let us then define for each $b \in B$ a formula $\psi_b(n)$ that holds if and only if $\mu(\mathbf{x})[n] = b$. For each n there exist unique q and r such that $0 \leq r < \ell$ and $n = \ell q + r$. For each $a \in A$, we can construct a formula $\sigma_a(r)$ that holds if and only if $\mu(a)$ contains the letter b at position r (indexing from 0). Setting

$$\psi_b(n) = (\exists q)(\exists r < \ell)(n = \ell q + r \wedge \bigvee_{a \in A} (\varphi_a(q) \wedge \sigma_a(r)))$$

certainly has the desired effect. Notice that this is indeed a formula in $\langle \mathbb{N}, +, V_k \rangle$ since ℓ is constant. Therefore it follows from Theorem 2.9 that $\mu(\mathbf{x})$ is k-automatic. For a proof not based on logic, see [1, Cor. 6.8.3].

Example 4. Assume $A = \{a, b\}$, $B = \{c, d\}$, $\ell = 3$ and set $\mu(a) = ccd$, $\mu(b) = dcd$. In this case, the formula $\psi_c(n)$ is given by

$$(\exists q)(\exists r < 3)(n = 3q + r \wedge [(\varphi_a(q) \wedge (r = 0 \vee r = 1)) \vee (\varphi_b(q) \wedge r = 1)]).$$

The same construction can be applied to numeration systems canonically associated with a Pisot number [5]. Here, we show that this closure property does not hold for Parry-automatic sequences.

Theorem 4.1. *There exists a Parry numeration system U such that the class of U-automatic sequences is not closed under taking image by a uniform substitution.*

Throughout this section, we shall consider a specific numeration system U given by the recurrence

$$U_n = 3U_{n-1} + 2U_{n-2} + 3U_{n-4} \tag{4}$$

with initial values $U_0 = 1$, $U_1 = 4$, $U_2 = 15$, and $U_3 = 54$ (it is from [11, Example 3]). The characteristic polynomial has two real roots β and γ and two complex roots with modulus less than 1. We have $\beta \approx 3.61645$ and $\gamma \approx -1.09685$. Thus from the basic theory of linear recurrences, we obtain that $U_n \sim c\beta^n$ for some constant c. The characteristic polynomial of the recurrence is the minimal polynomial of β so, in particular, γ is an algebraic conjugate of β. Since $|\gamma| > 1$, the number β is neither a Pisot number nor a Salem number. It is however a Parry number, as it is readily checked that $d_\beta(1) = 3203$. Thus U is a Parry numeration system. Moreover, we have $U = U_\beta$. Recall that $\mathrm{rep}_U(\mathbb{N})$ is regular as this holds for all Parry numeration systems.

Consider the characteristic sequence $\mathbf{x}$ of the set $\{U_n \mid n \geq 0\}$:

$$\mathbf{x} = 010010000000000010000000000000\cdots.$$

From Theorem 2.9, this sequence is U-automatic. We consider the constant length substitution $\mu\colon 0 \mapsto 0^t, 1 \mapsto 10^{t-1}$ with $t \geq 4$. Observe that $\mu(\mathbf{x})$ is the characteristic sequence of $\{tU_n \mid n \geq 0\}$. The multiplier 4 is the first interesting value to consider because for $j = 2, 3$ we have $\mathrm{rep}_U(\{jU_n \mid n \geq 0\}) = j0^*$, and we trivially get U-recognizable sets. Our aim is to show that $\mu(\mathbf{x})$ is not U-automatic (see Corollary 4.4). This will prove Theorem 4.1.

We begin with an auxiliary result that is of independent interest.

Proposition 4.2. *Let $r \geq 2$ be an integer. If t is an integer such that $4 \leq t \leq \lfloor \beta^r \rfloor$, then the β-expansion of the number t/β^r is aperiodic.*

For the proof, we need the following technical lemma, which is obtained by adapting [21, Lemma 2.2] to our situation. Since β is an algebraic number of degree 4, it is well-known that every element in $\mathbb{Q}(\beta)$ can be expressed as a polynomial in β of degree at most 3 with coefficients in $\mathbb{Q}$.

Lemma 4.3. *Let $x \in [0, 1) \cap \mathbb{Q}(\beta)$, and write*

$$x = q^{-1} \sum_{i=0}^{3} p_i \beta^i$$

for integers q and p_i. If $d_\beta(x)$ is ultimately periodic, then

$$q^{-1} \sum_{i=0}^{3} p_i \gamma^i = \sum_{i=1}^{\infty} d_\beta(x)[i] \gamma^{-i}.$$

Proof of Proposition 4.2. Let us first make the additional assumption that $t \geq \lceil \beta^{r-1} \rceil$ and prove the result in this case. Set $x = t/\beta^r = q^{-1} \sum_{i=0}^{3} p_i \beta^i$, and assume for a contradiction that $d_\beta(x)$ is ultimately periodic. Write $d_\beta(x) = d_1 d_2 \cdots$. Since β and γ are conjugates,

$$\frac{t}{\gamma^r} = q^{-1} \sum_{i=0}^{3} p_i \gamma^i$$

and it follows from Lemma 4.3 that

$$\frac{t}{\gamma^r} = \sum_{i=1}^{\infty} d_i \gamma^{-i}.$$

In other words, for any positive integer k, we have

$$t = \sum_{i=1}^{\infty} d_i \gamma^{-i+r} = S_{1,k} + S_{k+1,\infty}, \tag{5}$$

where $S_{m,n} := \sum_{i=m}^{n} d_i \gamma^{-i+r}$. Since γ is negative and $d_i \leq 3$ for all $i \geq 1$, we have

$$S_{r+1,\infty} = \sum_{i=1}^{\infty} d_{i+r} \gamma^{-i} \leq 3 \sum_{i=1}^{\infty} \gamma^{-2i} = \frac{3\gamma^{-2}}{1-\gamma^{-2}} < 15. \tag{6}$$

Similarly by discarding the odd terms and estimating $d_i \leq 3$, we obtain

$$S_{1,r} = \sum_{i=0}^{r-1} d_{r-i} \gamma^{i} \leq \frac{3(1-\gamma^{2(k+1)})}{1-\gamma^2}, \tag{7}$$

where k is the largest integer such that $2k \leq r-1$. Combining (5), (6), and (7) with our assumption $t \geq \lceil \beta^{r-1} \rceil$, we obtain that

$$\beta^{r-1} < \frac{3(1-\gamma^{2(k+1)})}{1-\gamma^2} + 15 \tag{8}$$

The left side of (8) clearly increases faster than the right side when $r \to \infty$ since $\beta \approx 3.62$ and $\gamma^2 \approx 1.20$. Using these approximations, it is straightforward to compute that for $r = 4$ the left side of (8) is approximately 47 while the right side is only approximately 22. Hence it must be that $r \leq 3$.

We are thus left with a few cases we have to deal with separately. The idea is the same, but we need to actually compute some digits d_i. Suppose first that $r = 3$. Like previously, we see that $S_{4,\infty} \leq 3\gamma^{-4}/(1-\gamma^{-2}) < 12.28$. Since $14 = \lceil \beta^2 \rceil \leq t \leq \lfloor \beta^3 \rfloor = 47$, by enumerating all possibilities for the word $d_1 d_2 d_3$ (within the given range for t), we see that $f(t) = t - S_{1,3}$ is minimized when $t = \lceil \beta^2 \rceil = 14$.

t	$d_\beta(t/\beta^3)$	$t - S_{1,3}$
14	$100\cdots$	12.797
15	$101\cdots$	12.797
16	$102\cdots$	12.797
17	$110\cdots$	16.894
18	$111\cdots$	16.894
$\vdots$	$\vdots$	$\vdots$
44	$311\cdots$	40.488
45	$312\cdots$	40.488
46	$313\cdots$	40.488
47	$320\cdots$	45.584

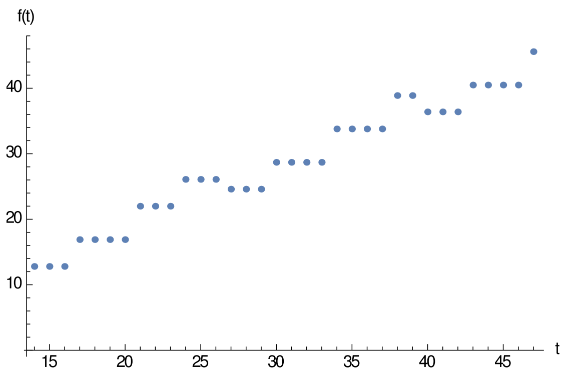

Figure 2: Values of $t - S_{1,3}$

In this case, $d_1 d_2 d_3 = 100$ and $t - S_{1,3} > 12.79$. This contradicts (5). Suppose then that $r = 2$. We proceed as above, but now we are interested in the number $t - S_{1,12}$ instead. By enumerating all possibilities, we see that $t - S_{1,12}$ is minimized for $t = \lceil \beta \rceil = 4$. Then $d_1 \cdots d_{12} = 101111202300$ and $t - S_{1,12} > 5.38$. Since $S_{13,\infty} < 5$, we get a contradiction.

Suppose finally that $4 \leq t < \lceil \beta^{r-1} \rceil$. For $r = 2$, we have already proven the claim because $\lceil \beta \rceil = 4$, so we may suppose that $r > 2$. As $t < \beta^{r-1}$, we see that $\beta \cdot t/\beta^r < 1$ meaning that $d_1 = 0$. Thus $d_2 d_3 \cdots$ is the β-expansion of t/β^{r-1}. Inductively it follows that this expansion is aperiodic, and we are done. $\square$

1	1	0									
2	1	0	1								
3	1	0	1	1							
4	1	0	1	1	1						
5	1	0	1	1	1	1					
6	1	0	1	1	1	1	2				
7	1	0	1	1	1	1	2	0			
8	1	0	1	1	1	1	2	0	3		
9	1	0	1	1	1	1	2	0	2	3	
10	1	0	1	1	1	1	2	0	2	3	0

Table 1: Representations of the first $4U_n$.

We now have tools to prove that the set $\{tU_n \mid n \geq 0\}$ is not U-recognizable for $t \geq 4$.

Corollary 4.4. *The set $\{tU_n \mid n \geq 0\}$ is not U-recognizable for $t \geq 4$. In other words, its characteristic sequence $\mu(\mathbf{x})$ is not U-automatic.*

Proof. Let $t \geq 4$, and suppose that $\lceil \beta^{r-1} \rceil \leq t \leq \lfloor \beta^r \rfloor$ for some $r \geq 2$. Recall that $U_n \sim c\beta^n$ for some positive constant c. By some simple asymptotic analysis, it is easy to see that $U_{n+r-1} < tU_n < U_{n+r}$ for n large enough. Hence, for n large enough, $\mathrm{rep}_U(tU_n)$ is a word of length $n + r$ (starting with a nonzero digit). Let $k > 0$. We show that, for large enough n, $\mathrm{rep}_U(tU_n)$ and $d_\beta(t/\beta^r)$ have the same prefix of length k. See Table 1 for an example. Assume that

$$\mathrm{rep}_U(tU_n) = d_1 \cdots d_k d_{k+1} \cdots d_{n+r}.$$

The extremal values for $d_{k+1} \cdots d_{n+r}$ are 0^{n+r-k} and (possibly) $\mathrm{rep}_U(U_{n+r-k} - 1)$ due to the greediness of the representations. Hence

$$0 \leq tU_n - d_1 U_{n+r-1} - \cdots - d_k U_{n+r-k} < U_{n+r-k}.$$

Dividing by U_{n+r} and letting n tend to infinity, we get

$$0 \leq \frac{t}{\beta^r} - \frac{d_1}{\beta} - \cdots - \frac{d_k}{\beta^k} < \frac{1}{\beta^k}.$$

Otherwise stated, the first k digits of $d_\beta(t/\beta^r)$ are $d_1 \cdots d_k$.

Now proceed by contradiction and assume that $\mathrm{rep}_U(\{tU_n \mid n \geq 0\})$ is accepted by a finite deterministic automaton. By a classical pumping argument, there exist $x, y, z \in A_U^*$, with y nonempty, such that $xy^j z$ is accepted by this automaton for all $j \geq 0$. Hence, $d_\beta(t/\beta^r)$ should be of the form xy^ω contradicting Proposition 4.2. □

Corollary 4.4 is interesting because it shows that addition in U is not computable by a finite automaton, i.e., its graph is not a regular language. Indeed, if this was the case, then surely multiplication by any constant would be computable by a finite automaton contrary to Corollary 4.4. This result is not new; it already appears in [11, Example 3]. The conclusion is that addition in a Parry numeration system is not necessarily computable by a finite automaton. This shows in particular that Parry-recognizable sets do not have a characterization based on first-order logic like Pisot-recognizable sets have. This is a considerable defect of Parry numeration systems that are not Pisot.

Let us then describe why a word obtained from a k-automatic sequence by periodically deleting letters is still k-automatic. Suppose that $\mathbf{x}$ is a k-automatic sequence over A, and let $\mathbf{y}$ be the word obtained from $\mathbf{x}$ by keeping only the letters at positions 0, t, $2t$, $3t$, $\ldots$ for a fixed integer $t \geq 2$. In other words, we have $\mathbf{y}[n] = \mathbf{x}[tn]$. As mentioned at the beginning of this section, for each $a \in A$, there exists a first-order formula $\varphi_a(n)$ in $\langle \mathbb{N}, +, V_k \rangle$ such that it holds if and only if

$\mathbf{x}[n] = a$. By substituting n by tn in $\varphi_a(n)$, we obtain a new first-order formula in $\langle \mathbb{N}, +, V_k \rangle$ such that it holds if and only if $\mathbf{y}[n] = a$. It follows from Theorem 2.9 that $\mathbf{y}$ is k-automatic. Again, a similar construction works in the Pisot case. See also [1, Thm. 6.8.1].

Let us next show that the class of U-automatic sequences is not closed under periodic deletion. Consider the characteristic sequence $\mathbf{y}$ of the set $\{U_n/2 \mid n \geq 0 \text{ and } U_n \text{ is even}\}$:

$$\mathbf{y} = 00100000000000000000000000001 \cdots .$$

This sequence $\mathbf{y}$ is obtained from the characteristic sequence $\mathbf{x}$ of the set $\{U_n \mid n \geq 0\}$ by removing its every second letter. Indeed, $\mathbf{y}[n] = \mathbf{x}[2n]$ hence $\mathbf{y}[n] = 1$ if and only if $2n$ belongs to $\{U_j \mid j \geq 0\}$. We will show that $\mathbf{y}$ is not U-automatic, which will prove the following theorem.

Theorem 4.5. *There exists a Parry numeration system U such that the class of U-automatic sequences is not closed under periodic deletion.*

Let us begin with the following result.

Proposition 4.6. *The β-expansion of $1/2$ is aperiodic.*

Proof. Assume for a contradiction that $d_\beta(1/2) = d_1 d_2 \cdots$ is ultimately periodic. As in the proof of Proposition 4.2, we obtain that

$$\frac{1}{2} = \sum_{i=1}^{\infty} d_i \gamma^{-i} = S_{1,k} + S_{k+1,\infty},$$

where $S_{r,s} = \sum_{i=r}^{s} d_i \gamma^{-i}$. It can be computed that $d_1 \cdots d_{21} = 123102303001010220123$. This computation actually needs some extra accuracy. It is sufficient to know that 3.61645454325 are correct initial digits for β. Using this information on $d_1 \cdots d_{21}$, it is computed that

$$S_{1,21} < -2.20.$$

Since γ is negative and $d_i \leq 3$ for all $i \geq 1$, we obtain that

$$S_{22,\infty} \leq \frac{3\gamma^{-22}}{1-\gamma^{-2}} < 2.33.$$

The two preceding inequalities show that $1/2 < -2.20 + 2.33 = 0.13$, which is obviously absurd. ☐

Interestingly the β-expansion of $1/3$ is ultimately periodic. Indeed, it can be shown that $d_\beta(1/3) = 10(2212)^\omega$.

Corollary 4.7. *The set $\{U_n/2 \mid n \geq 0$ and U_n is even$\}$ is not U-recognizable. In other words, its characteristic sequence $\mathbf{y}$ is not U-automatic.*

Proof. We follow steps similar to those of the proof of Corollary 4.4. From (4), it is clear that $U_{n-1} < \lfloor U_n/2 \rfloor < U_n$ for $n > 1$, so that $\mathrm{rep}_U(\lfloor U_n/2 \rfloor)$ is a word of length n. Let $k > 0$. We show that, for large enough n, $\mathrm{rep}_U(\lfloor U_n/2 \rfloor)$ and $d_\beta(1/2)$ have the same prefix of length k. Assume that

$$\mathrm{rep}_U(\lfloor U_n/2 \rfloor) = d_1 \cdots d_k d_{k+1} \cdots d_n.$$

Again, the extremal values for $d_{k+1} \cdots d_n$ are 0^{n-k} and (possibly) $\mathrm{rep}_U(U_{n-k} - 1)$ due to the greediness of the representations. Therefore

$$0 \leq \lfloor U_n/2 \rfloor - d_1 U_{n-1} - \cdots - d_k U_{n-k} < U_{n-k}.$$

Clearly $\lfloor U_n/2 \rfloor / U_n \xrightarrow{n\to\infty} 1/2$ so, dividing by U_n and letting n tend to infinity, we obtain

$$0 \leq \frac{1}{2} - \frac{d_1}{\beta} - \cdots - \frac{d_k}{\beta^k} < \frac{1}{\beta^k}.$$

Thus the first k digits of $d_\beta(\lfloor U_n/2\rfloor)$ are $d_1\cdots d_k$. This means that the words of the language $\operatorname{rep}_U(\{U_n/2 \mid n\ge 0 \text{ and } U_n \text{ is even}\})$ share longer and longer prefixes with $d_\beta(1/2)$.

The results follows by an argument similar to the final paragraph of the proof of Corollary 4.4: if $\operatorname{rep}_U(\{U_n/2 \mid n\ge 0 \text{ and } U_n \text{ is even}\})$ is accepted by a finite deterministic automaton, then $d_\beta(1/2)$ is ultimately periodic, and this is impossible by Proposition 4.6. □

Notice that the proof in fact shows that the set $\{\lfloor U_n/2\rfloor \mid n\ge 0\}$ is not U-recognizable. Even though $\mathbf{y}$ is not U-automatic, we suspect that the word obtained from $\mathbf{x}$, the characteristic sequence of $\{U_n \mid n\ge 0\}$, by keeping only the letters at indices that are divisible by 3 is U-automatic. This would follow from our conjecture that $\operatorname{rep}_U(\{U_n/3 \mid n\ge 0 \text{ and } U_n\equiv 0 \pmod 3\})$ equals $11+10(2212)^*(3+23+222+2213)$, but we have not attempted to prove this rigorously. Notice that U_n is divisible by 3 when $n\ge 2$.

5 Multidimensional sequences

By Proposition 2.11, an infinite word is U-automatic with respect to a numeration system U with $\operatorname{rep}_U(\mathbb{N})$ regular if and only its U-kernel is finite. Moreover, this is true more generally for abstract numeration systems. The generalization of this result to multidimensional sequences $\mathbf{x}=(x_{m,n})_{m,n\ge 0}$ [19, Prop. 32] is however slightly problematic as an extra assumption on the projections $(x_{k,n})_{n\ge 0}$ and $(x_{m,k})_{m\ge 0}$ is required. This extra assumption is however unnecessary for positional numeration systems considered in this paper. We did not find this fact in the literature, and this section is devoted to filling this gap.

For the sake of simplicity of presentation, we limit our presentation to two-dimensional sequences. We will consider finite automata reading pairs of digits. In particular, a pair of words can be read only if the two components have the same length. With positional numeration systems, when considering two representations of different length, then the shorter is padded with leading zeroes. For general abstract numeration systems an additional padding letter needs to be added, and this causes some complications.

Definition 5.1. Let U be a numeration system. A 2-dimensional word $\mathbf{x}=(x_{m,n})_{m,n\ge 0}$ over an alphabet B is *U-automatic* if there exists a complete DFAO $(Q,q_0,A_U\times A_U,\delta,\tau)$ with transition function $\delta\colon Q\times(A_U\times A_U)^*\to Q$ and output function $\tau\colon Q\to B$ such that $\delta(q_0,(0,0))=q_0$ and

$$x_{m,n}=\tau(\delta(q_0,(0^{\ell-|\operatorname{rep}_U(m)|}\operatorname{rep}_U(m),0^{\ell-|\operatorname{rep}_U(n)|}\operatorname{rep}_U(n)))),\quad \forall m,n\ge 0,$$

where $\ell=\max\{|\operatorname{rep}_U(m)|,|\operatorname{rep}_U(n)|\}$. The 2-dimensional word $\mathbf{x}$ is *k-automatic* (resp. *Parry-automatic*, *Bertrand-automatic*) if $U=(k^n)_{n\ge 0}$ for an integer $k\ge 2$ (resp. U is a Parry numeration system, U is a Bertrand numeration system).

Definition 2.10 is extended as follows (we make use of the notation $i(s,n)$ introduced therein).

Definition 5.2. The *U-kernel* of a 2-dimensional word $\mathbf{x}=(x_{m,n})_{m,n\ge 0}$ over B is the set

$$\ker_U(\mathbf{x}):=\{(x_{i(s,m),i(t,n)})_{m,n\ge 0}\in B^{\mathbb{N}^2} \mid s,t\in A_U^*,\ |s|=|t|\}.$$

Let us then state and prove the result mentioned above.

Proposition 5.3. *Let U be a numeration system such that the numeration language $\operatorname{rep}_U(\mathbb{N})$ is regular. A 2-dimensional word $\mathbf{x}=(x_{m,n})_{m,n\ge 0}$ is U-automatic if and only if its U-kernel is finite.*

Proof. Let $\mathbf{x}=(x_{m,n})_{m,n\ge 0}$ be a 2-dimensional word. From [19, Prop. 32], we already know that if $\mathbf{x}$ is U-automatic, then its U-kernel is finite because the result holds for all abstract numeration systems. We only need to prove the converse.

We let K denote the U-kernel of $\mathbf{x}$ and suppose that it is finite. For $s\in A_U^*$, define

$$\mathcal{L}(s):=0^*\operatorname{rep}_U(\mathbb{N})\cdot s^{-1}=\{w\in A_U^* \mid ws\in 0^*\operatorname{rep}_U(\mathbb{N})\}.$$

By assumption $0^* \operatorname{rep}_U(\mathbb{N})$ is regular, so it follows from the Myhill–Nerode theorem that the set of right quotients

$$J := \{\mathcal{L}(s) \mid s \in A_U^*\}$$

is finite. Let us define a DFAO $\mathcal{M}$ with state set

$$Q := J \times J \times K,$$

transition function δ, output function τ, and initial state

$$q_0 := (0^* \operatorname{rep}_U(\mathbb{N}), 0^* \operatorname{rep}_U(\mathbb{N}), (x_{m,n})_{m,n\geq 0}) = \left(\mathcal{L}(\varepsilon), \mathcal{L}(\varepsilon), (x_{i(\varepsilon,m),i(\varepsilon,n)})_{m,n\geq 0}\right).$$

For a state $q = (\mathcal{L}(s), \mathcal{L}(t), (x_{i(s,m),i(t,n)})_{m,n\geq 0})$ in Q, with $|s| = |t|$, and each pair (a, b) of digits in $A_U \times A_U$, we set

$$\delta(q, (a, b)) = (\mathcal{L}(as), \mathcal{L}(bt), (x_{i(as,m),i(bt,n)})_{m,n\geq 0}).$$

For other types of states, i.e., $(\mathcal{L}(s), \mathcal{L}(t), (x_{i(s',m),i(t',n)})_{m,n\geq 0})$ with $s \neq s'$ or $t \neq t'$, we leave the transition function undefined as it is clear that such states are not reachable from the initial state q_0.

We have to check that the transition function δ is well-defined. Assume that

$$(\mathcal{L}(s), \mathcal{L}(t), (x_{i(s,m),i(t,n)})_{m,n\geq 0}) = (\mathcal{L}(s'), \mathcal{L}(t'), (x_{i(s',m),i(t',n)})_{m,n\geq 0})$$

with $|s| = |t|$ and $|s'| = |t'|$. For all $(a, b) \in A_U \times A_U$, we need to show that

$$(\mathcal{L}(as), \mathcal{L}(bt), (x_{i(as,m),i(bt,n)})_{m,n\geq 0}) = (\mathcal{L}(as'), \mathcal{L}(bt'), (x_{i(as',m),i(bt',n)})_{m,n\geq 0}).$$

For the first two components, the result follows from the definition: $\mathcal{L}(as) = \mathcal{L}(s) \cdot a^{-1}$ for any letter a. For the third component, we want to prove that $x_{i(as,m),i(bt,n)} = x_{i(as',m),i(bt',n)}$ for all $m, n \geq 0$. We know that $\mathcal{L}(s) = \mathcal{L}(s')$, $\mathcal{L}(t) = \mathcal{L}(t')$, and $x_{i(s,m),i(t,n)} = x_{i(s',m),i(t',n)}$ for all $m, n \geq 0$. Enumerate the words of $\mathcal{L}(s) \setminus 0A_U^*$ in genealogical order $\prec$:

$$\mathcal{L}(s) \setminus 0A_U^* = \{r_{s,0} \prec r_{s,1} \prec r_{s,2} \prec \cdots\}.$$

Similarly, we write

$$\mathcal{L}(t) \setminus 0A_U^* = \{r_{t,0} \prec r_{t,1} \prec r_{t,2} \prec \cdots\}.$$

Note that if s is a suffix occurring in a valid U-representation, then $r_{s,0} = \varepsilon$; similarly for $r_{t,0}$. Let $j, k \geq 0$. Since $r_{s,j}$ and $r_{s,k}$ do not start with a zero digit, we have

$$r_{s,j} \prec r_{s,k} \Leftrightarrow \operatorname{val}_U(r_{s,j}0^{|s|}) < \operatorname{val}_U(r_{s,k}0^{|s|}),$$

and an analogous equivalence holds for $r_{t,j}$ and $r_{t,k}$. The subsequence $(x_{i(s,m),i(t,n)})_{m,n\geq 0}$ is the same as the sequence

$$(x_{\operatorname{val}_U(r_{s,m}s),\operatorname{val}_U(r_{t,n}t)})_{m,n\geq 0}$$

because by definition (3), $i(s, m)$ (resp. $i(t, n)$) is the mth (resp. nth) integer belonging to $\mathcal{I}_s = \operatorname{val}_U(0^* \operatorname{rep}_U(\mathbb{N}) \cap A_U^* s)$ (resp. $\mathcal{I}_t$). Notice that words in $\mathcal{L}(s)$ (resp. $\mathcal{L}(t)$) starting with 0 do not provide any new indices. So when building the subsequence, we can limit ourselves to words not starting with 0. If we select in $\mathcal{L}(s) \setminus 0A_U^*$ all words ending with a, we get exactly $(\mathcal{L}(as) \setminus 0A_U^*)a$, which is equal to $(\mathcal{L}(as') \setminus 0A_U^*)a$ because $\mathcal{L}(as) = \mathcal{L}(as')$. Let $m \geq 0$ and $r_{as,m}$ be the mth word in $\mathcal{L}(as) \setminus 0A_U^*$. Suppose that the mth word in $(\mathcal{L}(as) \setminus 0A_U^*)a$, which is $r_{as,m}a$, occurs as the kth word $r_{s,k}$ in $\mathcal{L}(s) \setminus 0A_U^*$. Then $r_{s,k}$ also occurs as the kth word $r_{s',k}$ in $\mathcal{L}(s') \setminus 0A_U^*$. With our notation, we have

$$r_{as,m}a = r_{s,k}, \ \operatorname{val}_U(r_{as,m}as) = \operatorname{val}_U(r_{s,k}s), \text{ and } i(as, m) = i(s, k) = i(s', k).$$

We can make similar observations for the other component. Supposing that $r_{bt,n} = r_{t,\ell}$ for some ℓ, we thus have

$$x_{i(as,m),i(bt,n)} = x_{i(s,k),i(t,\ell)} = x_{i(s',k),i(t',\ell)} = x_{i(as',m),i(bt',n)},$$

where the central equality comes from our initial assumption. Therefore we have shown that δ is well-defined.

From our definition of the transition function δ, the accessible part of $\mathcal{M}$ is limited to states q of the form

$$(\mathcal{L}(s), \mathcal{L}(t), (x_{i(s,m),i(t,n)})_{m,n\geq 0})$$

with $|s| = |t|$. For such a state q, we set

$$\tau(q) = x_{i(s,0),i(t,0)}.$$

Notice that the preceding arguments show that τ is also well-defined. To conclude the proof, let us show that if s, t are two words of the same length in $0^* \operatorname{rep}_U(\mathbb{N})$, then

$$\tau(\delta(q_0, (s^R, t^R))) = x_{\operatorname{val}_U(s),\operatorname{val}_U(t)},$$

where s^R and t^R respectively denote the reversals of the words s and t. Reading (s^R, t^R) from q_0 leads to the state $(\mathcal{L}(s), \mathcal{L}(t), (x_{i(s,m),i(t,n)})_{m,n\geq 0})$. Since $s, t \in 0^* \operatorname{rep}_U(\mathbb{N})$, we have that ε belongs to $\mathcal{L}(s)$ and $\mathcal{L}(t)$. It is clear that $i(s, 0) = \operatorname{val}_U(s)$ and $i(t, 0) = \operatorname{val}_U(t)$.

We have thus proved that $\mathbf{x}$ is reversal-U-automatic. It follows from Proposition 2.12 (which also holds in the multidimensional setting) that $\mathbf{x}$ is U-automatic. □

Acknowledgments

This work was initiated when the second author visited his coauthors at the Department of Mathematics of the University of Liège during June 2018. He thanks the department for their hospitality.

We thank the referee for a careful reading of the paper, which improved the presentation.

References

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

[2] A. Bertrand. Développements en base de Pisot et répartition modulo 1. *C. R. Acad. Sci. Paris*, Sér. A-B **285** (1977), A419–A421.

[3] A. Bertrand-Mathis. Comment écrire les nombres entiers dans une base qui n'est pas entière. *Acta Math. Hungar.* **54** (1989), 237–241. DOI:10.1007/BF01952053.

[4] V. Berthé, M. Rigo, eds. *Combinatorics, Automata and Number Theory.* Encyclopedia of Mathematics and Its Applications, vol. **135**. Cambridge University Press (2010).

[5] V. Bruyère, G. Hansel. Bertrand numeration systems and recognizability. *Theoret. Comput. Sci.* **181** (1997), 17-43. DOI:10.1016/S0304-3975(96)00260-5.

[6] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlag. Math.* **6** (1960), 66–92.

[7] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192. DOI:10.1007/BF01706087.

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

[9] S. Fabre. Substitutions et β-systèmes de numération. *Theoret. Comput. Sci.* **137** (1995), 219–236.

[10] Ch. Frougny. Representations of numbers and finite automata. *Math. Systems Theory* **25** (1992), 37–60. DOI:10.1007/BF01368783.

[11] Ch. Frougny. On the sequentiality of the successor function. *Inform. and Comput.*, **139.1** (1997), 17–38. DOI:10.1006/inco.1997.2650.

[12] Ch. Frougny, B. Solomyak. On representation of integers in linear numeration systems. In *Ergodic Theory of $\mathbb{Z}^d$-Actions (Warwick, 1993–1994)*, 345–368, London Math. Soc. Lecture Note Ser. 228, Cambridge University Press (1996).

[13] M. Hollander. Greedy numeration systems and regularity. *Theory Comput. Syst.* **31.2** (1998), 111–133. DOI:10.1007/s002240000082.

[14] D. Lind, B. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press (1995).

[15] M. Lothaire. *Algebraic Combinatorics on Words.* Encyclopedia of Math. and Its Applications, vol. 90, Cambridge University Press (2002).

[16] W. Parry. On the β-expansions of real numbers. *Acta Math. Acad. Sci. Hungar.* **11** (1960), 401–416.

[17] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In J. Paredaens, ed., *Proc. 11th Int'l Conf. on Automata, Languages, and Programming (ICALP).* Lecture Notes in Computer Science, Vol. 172, Springer-Verlag, 1984, pp. 380–389.

[18] M. Rigo. Generalization of automatic sequences for numeration systems on a regular language. *Theoret. Comput. Sci.* **244** (2000), 271–281. DOI:10.1016/S0304-3975(00)00163-8.

[19] M. Rigo, A. Maes. More on generalized automatic sequences. *J. of Automata, Languages and Combinatorics* **7** (2002), 351–376.

[20] M. Rigo. *Formal Languages, Automata and Numeration Systems*, vol. 2. Applications to recognizability and decidability. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, (2014).

[21] K. Schmidt. On periodic expansions of Pisot numbers and Salem numbers. *Bull. London Math. Soc.*, 12:269–278, 1980. DOI:10.1112/blms/12.4.269.

[22] J. Shallit. Numeration systems, linear recurrences, and regular sets. *Inform. and Comput.* **113.2** (1994), 331–347. DOI:10.1006/inco.1994.1076.