THE GROWTH FUNCTION OF S-RECOGNIZABLE SETS

ÉMILIE CHARLIER AND NARAD RAMPERSAD

ABSTRACT. A set $X \subseteq \mathbb{N}$ is S-recognizable for an abstract numeration system S if the set $\mathrm{rep}_S(X)$ of its representations is accepted by a finite automaton. We show that the growth function of an S-recognizable set is always either $\Theta((\log(n))^{c-df}n^f)$ where $c, d \in \mathbb{N}$ and $f \geq 1$, or $\Theta(n^r \theta^{\Theta(n^q)})$, where $r, q \in \mathbb{Q}$ with $q \leq 1$. If the number of words of length n in the numeration language is bounded by a polynomial, then the growth function of an S-recognizable set is $\Theta(n^r)$, where $r \in \mathbb{Q}$ with $r \geq 1$. Furthermore, for every $r \in \mathbb{Q}$ with $r \geq 1$, we can provide an abstract numeration system S built on a polynomial language and an S-recognizable set such that the growth function of X is $\Theta(n^r)$. For all positive integers k and ℓ, we can also provide an abstract numeration system S built on a exponential language and an S-recognizable set such that the growth function of X is $\Theta((\log(n))^k n^\ell)$.

1. INTRODUCTION

A set $X \subseteq \mathbb{N}$ is *b-recognizable* if the set of representations of the elements of X in base b is accepted by a finite automaton. The class of b-recognizable sets is a very well-studied class (the main results can be found in the book of Allouche and Shallit [AS03]). It is therefore somewhat surprising that, to the best of our knowledge, no characterization of the possible growths of b-recognizable sets is currently known. In this paper we provide such a characterization for the much more general class of S-recognizable sets.

Let L be a language over an alphabet Σ. We assume that the letters $a_1, \ldots, a_\ell$ of Σ are ordered by $a_1 < \cdots < a_\ell$. This order on the alphabet Σ induces an order $<$ on the language L called the *genealogic order* or the *radix order*: words are ordered length by length, and for a given length, we use the lexicographic order. This leads to the definition of an abstract numeration system.

Definition 1. An *abstract numeration system* is a triple $S = (L, \Sigma, <)$ where L is an infinite language over a totally ordered finite alphabet $(\Sigma, <)$. The language L is called the *numeration language*. The map $\mathrm{rep}_S \colon \mathbb{N} \to L$ is a bijection mapping $n \in \mathbb{N}$ to the $(n+1)$-th word of L ordered genealogically. The inverse map is denoted by $\mathrm{val}_S \colon L \to \mathbb{N}$.

Most of the time, we assume that L is a regular language. Nevertheless, some results hold true for arbitrary numeration languages.

The integer base numeration systems are particular cases of abstract numeration systems since for all $b, n, m \in \mathbb{N}$ with $b \geq 2$, we have $n < m \Leftrightarrow \mathrm{rep}_b(n) < \mathrm{rep}_b(m)$, where $\mathrm{rep}_b(x)$ designates the usual greedy representation of x in base b. In this case, the numeration language is $\mathrm{rep}_b(\mathbb{N}) = \{1, 2, \ldots, b-1\}\{0, 1, 2, \ldots, b-1\}^* \cup \{\varepsilon\}$, which is regular.

Definition 2. Let S be an abstract numeration system. Let $X \subseteq \mathbb{N}$. The set X is S-*recognizable* if the language $\mathrm{rep}_S(X) = \{\mathrm{rep}_S(n) \colon n \in X\}$ is regular. Let $b \geq 2$ be an integer. The set X is b-*recognizable* if it is S-recognizable for the abstract numeration system S built on the language $\mathrm{rep}_b(\mathbb{N})$ consisting of the base-b representations of the elements of X. The

set X is 1-*recognizable* if it is S-recognizable for the abstract numeration system S built on a^*.

Note that $\mathbb{N}$ is S-recognizable if and only if the numeration language L is regular.

Rigo [Rig02] stated the following two fundamental questions regarding S-recognizable sets:

- For a given numeration system S, what are the S-recognizable subsets of $\mathbb{N}$?
- For a given subset X of $\mathbb{N}$, is it S-recognizable for some numeration system S?

Of course, both of these questions are quite difficult to address in full generality. In this paper, we consider these questions in terms of the growth functions of both the set X and the numeration language.

Definition 3. For a subset X of $\mathbb{N}$, we let $t_X(n)$ denote the $(n+1)$-th term of X. The map $t_X \colon \mathbb{N} \to \mathbb{N}$ is called the *growth function* of X.

We address the following problem.

Problem 1. Let S be an abstract numeration system built on a regular language. What do the growth functions of S-recognizable sets look like? Of course, this question has to be answered in terms of the growth function of the numeration language.

Apart from the following result, known as *Eilenberg's gap theorem*, we are not aware of any result in the direction of answering the above problem.

Theorem 4. [Eil74] *Let $b \geq 2$ be an integer. A b-recognizable set X of nonnegative integers satisfies either* $\limsup_{n \to +\infty}(t_X(n+1) - t_X(n)) < +\infty$ *or* $\limsup_{n \to +\infty} \frac{t_X(n+1)}{t_X(n)} > 1$.

Thanks to this result, examples of sets that are not b-recognizable for any b have been exhibited. The set $\{n^2 : n \in \mathbb{N}\}$ of squares is such an example. However, the set of squares is S-recognizable for the abstract numeration system

$$S = (a^*b^* \cup a^*c^*, \{a, b, c\}, a < b < c).$$

More generally, Rigo [Rig02] and Strogalov [Str90] showed that for any polynomial $P \in \mathbb{Q}[x]$ such that $P(\mathbb{N}) \subseteq \mathbb{N}$, there exists S such that P is S-recognizable. Observe that in the case of an integer base numeration system, the number of words of each length in the numeration language grows exponentially, whereas in the case of the numeration system S, this number grows polynomially. This leads to the natural question: Can a set of the form $P(\mathbb{N})$ ever be recognized in a numeration system where the numeration language is exponential? In Section 6, we show that the answer to this question is no for all polynomials P of degree 2 or more.

Let us fix some asymptotic notation.

Definition 5. Let f and g be functions with domain $\mathbb{N}$. We say that f is $\Theta(g)$, and we write $f = \Theta(g)$, if there exist positive constants c and d and a nonnegative integer N such that, for all integers $n \geq N$, we have $cg(n) \leq f(n) \leq dg(n)$. We say that f and g have *equivalent behaviors at infinity*, which is denoted by $f(n) \sim g(n)$ $(n \to +\infty)$ (or simply $f \sim g$ when the context is clear), if we have $\lim_{n \to +\infty} \frac{f(n)}{g(n)} = 1$. Finally, we write $f = o(g)$ if if we have $\lim_{n \to +\infty} \frac{f(n)}{g(n)} = 0$.

Definition 6. For any language L over an alphabet Σ and any non-negative integer n, we let

$$\mathbf{u}_L(n) = |L \cap \Sigma^n|$$

denote the number of words of length n in L and

$$\mathbf{v}_L(n) = \sum_{i=0}^{n} \mathbf{u}_L(i) = |L \cap \Sigma^{\leq n}|$$

denote the number of words of length less than or equal to n in L. The maps $\mathbf{u}_L \colon \mathbb{N} \to \mathbb{N}$ and $\mathbf{v}_L \colon \mathbb{N} \to \mathbb{N}$ are called the *counting (or growth) functions of L*.

When L is a regular language, the sequence $(\mathbf{u}_L(n))_{n \geq 0}$ satisfies a linear recurrence relation with integer coefficients (for instance, see [BR88]): there exist a positive integer k and $a_1, \ldots, a_k \in \mathbb{Z}$ such that we have

$$\forall n \in \mathbb{N}, \ \mathbf{u}_L(n + k) = a_1 \mathbf{u}_L(n + k - 1) + \cdots + a_k \mathbf{u}_L(n).$$

Then, since we have $\mathbf{v}_L(n) - \mathbf{v}_L(n - 1) = \mathbf{u}_L(n)$, the sequence $(\mathbf{v}_L(n))_{n \geq 0}$ satisfies the linear recurrence relation of length $k + 1$ whose characteristic polynomial is $(x - 1)(x^k - a_1 x^{k-1} - \cdots - a_k)$.

Our main result can be stated as follows.

Theorem 7. *Let $S = (L, \Sigma, <)$ be an abstract numeration system built on a regular language and let X be an infinite S-recognizable set of nonnegative integers. Suppose*

$$\forall i \in \{0, \ldots, p - 1\}, \ \mathbf{v}_L(np + i) \sim a_i n^c \alpha^n \ (n \to +\infty),$$

for some $p, c \in \mathbb{N}$ with $p \geq 1$, some $\alpha \geq 1$ and some positive constants $a_0, \ldots, a_{p-1}$, and

$$\forall j \in \{0, \ldots, q - 1\}, \ \mathbf{v}_{\mathrm{rep}_S(X)}(nq + j) \sim b_j n^d \beta^n \ (n \to +\infty),$$

for some $q, d \in \mathbb{N}$ with $q \geq 1$, some $\beta \geq 1$ and some positive constants $b_0, \ldots, b_{q-1}$. Then we have

- $t_X(n) = \Theta \left((\log(n))^{c - d \frac{\log(\sqrt[p]{\alpha})}{\log(\sqrt[q]{\beta})}} n^{\frac{\log(\sqrt[p]{\alpha})}{\log(\sqrt[q]{\beta})}} \right)$ *if $\beta > 1$;*
- $t_X(n) = \Theta \left(n^{\frac{c}{d}} (\sqrt[p]{\alpha})^{\Theta(n^{\frac{1}{d}})} \right)$ *if $\beta = 1$.*

Furthermore we have

- $t_X(n) = \Theta \left(n^{\frac{c}{d}} (\sqrt[p]{\alpha})^{(1 + o(1))(\frac{n}{b})^{\frac{1}{d}}} \right)$ *if $\beta = 1$, $q = 1$, and $\mathbf{v}_{\mathrm{rep}_S(X)}(n) \sim bn^d \ (n \to +\infty)$;*
- $t_X(n) = \Theta \left(n^{\frac{c}{d}} (\sqrt[p]{\alpha})^{(\frac{n}{b})^{\frac{1}{d}}} \right)$ *if $\beta = 1$, $q = 1$, and $\mathbf{v}_{\mathrm{rep}_S(X)}(n) = bn^d$ for all $n \in \mathbb{N}$.*

This paper has the following organization. In Section 2, we recall some necessary background concerning S-recognizable sets. In Section 3, we provide a proof of Theorem 7. In Section 4, we describe some constructions of abstract numeration systems S that effectively realize some particular behaviors of S-recognizable sets. As a consequence of our main result, we also show that certain behaviors of S-recognizable sets are never achieved by any abstract numeration systems S. Then, in Section 5, we propose examples illustrating every possible behavior of $t_X(n)$ when X is an S-recognizable set. Finally, in Section 6, we discuss a few additional results. Namely, we show that no polynomial can be recognized in an abstract numeration system built on an exponential regular language. This result extends a well-known result in the integer base numeration systems.

2. Background on S-recognizable sets

Definition 8. Let S be an abstract numeration system. An infinite word x over an alphabet Γ is S-automatic if, for all non-negative integers n, its $(n+1)$st letter $x[n]$ is obtained by "feeding" a DFAO $\mathcal{A} = (Q, \Sigma, \delta, q_0, \Gamma, \tau)$ with the S-representation of n:

$$\forall n \in \mathbb{N}, \ \tau(\delta(q_0, \mathrm{rep}_S(n))) = x[n].$$

Definition 9. Let X be a set of nonnegative integers. Its characteristic sequence is the sequence $\chi_X = (\chi_X(n))_{n \geq 0}$ defined by

$$\chi_X(n) = \left\{ \begin{array}{ll} 1, & \text{if } n \in X; \\ 0, & \text{otherwise.} \end{array} \right.$$

The following result is self-evident.

Proposition 10. *Let S be an abstract numeration system. A set is S-recognizable if and only if its characteristic sequence is S-automatic.*

Definition 11. If μ is a morphism over an alphabet Σ and a is a letter in Σ such that the image $\mu(a)$ begins with a, then we say that μ is *prolongable on a*.

If a morphism μ is prolongable on a letter a, then the limit $\lim_{n \to +\infty} \mu^n(a)$ is well defined. As usual, we denote this limit by $\mu^\omega(a)$. Furthermore, this limit word is a fixed point of μ. Observe that it is an infinite word if and only if there is a letter b occurring in $\mu(a)$ that satisfies $\mu^n(b) \neq \varepsilon$ for all non-negative integers n.

Definition 12. An infinite word is said to be *pure morphic* if it can be written as $\mu^\omega(a)$ for some morphism μ prolongable on a letter a. It is said to be *morphic* if it is the image under a morphism of some pure morphic word.

Theorem 13. [Rig00, RM02] *An infinite word is S-automatic for some abstract numeration system S if and only if it is morphic.*

Example 14. Consider the morphism $h \colon \{0,1\}^* \to \{0,1\}^*$ defined by $h(1) = 1010$ and $h(0) = 00$. Let X be the set of nonnegative integers whose characteristic sequence is $h^\omega(1)$. The first element in X are $0, 2, 6, 8, 16, 18, 22, 24, 40, 42, 46, \ldots$ From Theorem 13 and Proposition 10, the set X is S-recognizable for some abstract numeration system S. The associated DFAO is depicted in Figure 1 (details on how to build this DFAO are given in Section 3). For all $k \in \mathbb{N}$, we have $|h^k(1)| = (k+1)2^k$. For all $n \in \mathbb{N}$, there is a unique $k := k(n) \in \mathbb{N}$ such that $(k+1)2^k \leq t_X(n) < (k+2)2^{k+1}$. Since the number of occurrences of 1 in the prefix $h^k(1)$ is equal to 2^k, we also have $(k+1)2^k \leq t_X(n) < (k+2)2^{k+1} \Leftrightarrow 2^k \leq n < 2^{k+1}$. This means that $k(n) = \log_2(n)$. Consequently, $t_X(n)$ is $\Theta(n \log(n))$.
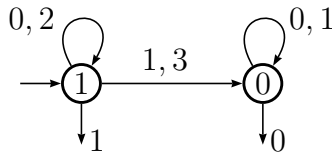

FIGURE 1. A DFAO generating X.

3. Proof of the main result

Definition 15. Let $\mathcal{A} = (Q, \{a_1, \ldots, a_r\}, \delta, q_0, F)$ be a DFA whose alphabet is ordered by $a_1 < \cdots < a_r$. The *morphism associated with $\mathcal{A}$* is the morphism $\mu_{\mathcal{A}} \colon (Q \cup \{\alpha\})^* \to (Q \cup \{\alpha\})^*$ defined by

$$\forall q \in Q, \ \mu_{\mathcal{A}}(q) = \delta(q, a_1) \cdots \delta(q, a_r) \ \text{ and } \ \mu_{\mathcal{A}}(\alpha) = \alpha \mu_{\mathcal{A}}(q_0).$$

where α is a new letter, not belonging to Q.

In [Rig00], it was shown that any S-automatic word is morphic (see Theorem 13 above). Here we only need to consider the particular case where the S-automatic word is the characteristic word of an S-recognizable set of nonnegative integers. We follow the construction of [Rig00] in the following definition. Nevertheless, notice that our definitions of $\mathcal{A}$ and g are slightly different.

Definition 16. Let $S = (L, \Sigma, <)$ be an abstract numeration system and let X be an S-recognizable set of nonnegative integers. Let $\mathcal{A}_L$ be the trim minimal automaton of L and let $\mathcal{A}_X$ be the (complete) minimal automaton of $\mathrm{rep}_S(X)$. Now define $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ to be the product of these two automata (take only the accessible part) and consider the canonically associated morphism $\mu_{\mathcal{A}}$. Let $g \colon (Q \cup \{\alpha\})^* \to \{0, 1\}^*$ be the morphism defined by $g(\alpha) = g(q_0)$ and for all states (p, q) of $\mathcal{A}$,

$$g(p, q) = \begin{cases} 1, & \text{if } p \text{ is final in } \mathcal{A}_L \text{ and } q \text{ is final in } \mathcal{A}_X; \\ 0, & \text{if } p \text{ is final in } \mathcal{A}_L \text{ and } q \text{ is not final in } \mathcal{A}_X; \\ \varepsilon, & \text{if } p \text{ is not final in } \mathcal{A}_L. \end{cases}$$

Furthermore, for each $n \in \mathbb{N}$, we define $F(n)$ to be the number of occurrences of the letter 1 in $g(\mu_{\mathcal{A}}^n(\alpha))$.

The following lemma is self-evident.

Lemma 17. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and let X be an S-recognizable set of nonnegative integers. With the notation of Definition 16, we have*

$$\forall n \in \mathbb{N}, \forall k \in \mathbb{N}, \ |g(\mu_{\mathcal{A}}^k(\alpha))| \leq t_X(n) < |g(\mu_{\mathcal{A}}^{k+1}(\alpha))| \Leftrightarrow F(k) \leq n < F(k+1).$$

Lemma 18. *Let $S = (L, \Sigma, <)$ be an abstract numeration system and let X be an S-recognizable set of nonnegative integers. With the notation of Definition 16, we have*

$$\forall n \in \mathbb{N}, \ |g(\mu_{\mathcal{A}}^n(\alpha))| = \mathbf{v}_L(n) \ \text{ and } \ F(n) = \mathbf{v}_{\mathrm{rep}_S(X)}(n).$$

Proof. Define K to be the language accepted by the trim automaton $\mathcal{A}_K$ whose graph is the same as $\mathcal{A}_L$ but where all states are final. Observe that the following inclusions hold: $\mathrm{rep}_S(X) \subseteq L \subseteq K$. By construction, for all $n \in \mathbb{N}$, the $(n+2)$-th letter of the infinite word $\mu_{\mathcal{A}}^\omega(\alpha)$ is the state in $\mathcal{A}$ reached by reading the $(n+2)$-th word in K. Define the morphism $f \colon (Q \cup \{\alpha\})^* \to Q^*$ by $f(\alpha) = f(q_0)$ and for all states (p, q) of $\mathcal{A}$,

$$f(p, q) = \begin{cases} (p, q), & \text{if } p \text{ is final in } \mathcal{A}_L; \\ \varepsilon, & \text{if } p \text{ is not final in } \mathcal{A}_L. \end{cases}$$

Then, for all $n \in \mathbb{N}$, the $(n+1)$-th letter of the infinite word $f(\mu_{\mathcal{A}}^\omega(\alpha))$ is the state in $\mathcal{A}$ reached by reading the $(n+1)$-th word in L. Now define the morphism $h \colon Q^* \to \{0, 1\}^*$ be the morphism defined by

$$h(p, q) = \begin{cases} 1, & \text{if } q \text{ is final in } \mathcal{A}_X; \\ 0, & \text{if } q \text{ is not final in } \mathcal{A}_X, \end{cases}$$

for all states (p, q) of $\mathcal{A}$. We clearly have $g = h \circ f$.

We claim that $|\mu_{\mathcal{A}}^n(\alpha))| = \mathbf{v}_K(n)$ for all $n \in \mathbb{N}$. If so, since f erases the words not belonging to L and h is a letter-to-letter morphism, then we obtain $|f(\mu_{\mathcal{A}}^n(\alpha))| = |g(\mu_{\mathcal{A}}^n(\alpha))| = \mathbf{v}_L(n)$ for all $n \in \mathbb{N}$. Then, by definition of F, we also obtain $F(n) = \mathbf{v}_{\mathrm{rep}_S(X)}(n)$ for all $n \in \mathbb{N}$.

It is thus sufficient to prove the claim. We define a new automaton $\mathcal{B} = (Q \cup \{\alpha\}, \Sigma \cup \{a_0\}, \delta', \alpha, Q \cup \{\alpha\})$ by slightly modifying the automaton $\mathcal{A}$. The initial state q_0 of $\mathcal{A}$ is no longer initial in $\mathcal{B}$ and we add a new initial state α with a loop labeled by a new letter a_0 (not belonging to Σ). All states are final. The (partial) transition function $\delta' \colon Q \cup \{\alpha\} \times \Sigma \cup \{a_0\} \to Q \cup \{\alpha\}$ is defined by $\delta'(\alpha, a) = \delta(q_0, a)$ for all $a \in \Sigma$, $\delta'(\alpha, a_0) = \alpha$, and $\delta'(q, a) = \delta(q, a)$ for all $q \in Q$ and all $a \in \Sigma$. Observe that K is the language of the words accepted by $\mathcal{B}$ from which we remove the words starting with α. With the terminology of [CKR10], this automaton $\mathcal{B}$ is the *automaton canonically associated with the morphism $\mu_{\mathcal{A}}$ and the letter α* and the corresponding *directive language* is K. We thus can apply [CKR10, Lemma 28] to obtain the claim. This completes the proof. $\qquad\square$

Lemma 19. *For all regular languages L, there exist $p, c \in \mathbb{N}$ with $p \geq 1$ and $\alpha \geq 1$ such that for all $i \in \{0, \ldots, p-1\}$, we have*

$$\mathbf{v}_L(np + i) \sim a_i n^c \alpha^n \ (n \to +\infty)$$

where $a_0, \ldots, a_{p-1}$ are some positive constants.

Proof. It is well-known that $\sum_{n \geq 0} \mathbf{v}_L(n)$ is an $\mathbb{N}$-rational series for all regular languages L (see for instance [BR88]). Since $(\mathbf{v}_L(n))_{n \geq 0}$ is a non-decreasing sequence, the lemma follows from [SS78, Theorem II.10.2]. $\qquad\square$

We are ready for the proof of Theorem 7.

Proof of Theorem 7. Let $p, q, a_0, \ldots, a_{p-1}, b_0, \ldots, b_{q-1}, c, d, \alpha$, and β be numbers like in the statement. The hypotheses imply $\mathbf{v}_L(n) = \Theta(n^c(\sqrt[p]{\alpha})^n)$ and $\mathbf{v}_{\mathrm{rep}_S(X)}(n) = \Theta(n^d(\sqrt[q]{\beta})^n)$. Consider the notation of Definition 16. For all $n \in \mathbb{N}$, there exists a unique $k := k(n) \in \mathbb{N}$ such that we have $|g(\mu_{\mathcal{A}}^k(\alpha))| \leq t_X(n) < |g(\mu_{\mathcal{A}}^{k+1}(\alpha))|$. From Lemma 17, this number k is also the only nonnegative integer that satisfies $F(k) \leq n < F(k+1)$.

First consider the case where $\beta > 1$. From Lemma 18 we obtain $F(k) = \Theta(k^d(\sqrt[q]{\beta})^k)$. From [BR10, Lemma 4.7.14] we find

$$k(n) = \frac{1}{\log(\sqrt[q]{\beta})}(\log(n) - d\log(\log(n))) + O(1).$$

Using Lemma 18, this gives the announced asymptotic behavior.

Now consider the case where $\beta = 1$. From Lemma 18 we obtain $F(k) = \Theta(k^d)$. This gives $k(n) = \Theta(n^{\frac{1}{d}})$. Hence, from Lemma 18, the announced asymptotic behavior holds. If moreover $q = 1$, we can be more precise. In this case, we have $F(k) = \mathbf{v}_{\mathrm{rep}_S(X)}(k) \sim bk^d \ (k \to +\infty)$ for some b. This gives $k(n) \sim (\frac{n}{b})^{\frac{1}{d}} \ (k \to +\infty)$ and from Lemma 18, the announced asymptotic behavior holds. If moreover $\mathbf{v}_{\mathrm{rep}_S(X)}(n) = bn^d$ for all $n \in \mathbb{N}$ and for some b, then we have $F(k) = bk^d$ for all $k \in \mathbb{N}$. This gives $k(n) = \lfloor \frac{n}{b}^{\frac{1}{d}} \rfloor$. Therefore, the announced asymptotic behavior holds in this case as well. This concludes the proof. $\qquad\square$

Remark 20. Note that the hypotheses of Theorem 7 implies that either $\sqrt[q]{\beta} < \sqrt[p]{\alpha}$ or ($\sqrt[q]{\beta} = \sqrt[p]{\alpha}$ and $d \leq c$). Since $\mathrm{rep}_S(X)$ is assumed to be an infinite language, we also always have $\beta \geq 1$.

The following corollary is a particular case of Theorem 7 where $\alpha = \beta^r$ for some $r \geq 1$. Recall that α and β are said to be *multiplicatively dependent* if $\alpha = \beta^r$ for some non-null $r \in \mathbb{Q}$.

Corollary 21. *Let $S = (L, \Sigma, <)$ be an abstract numeration system built on a regular language and let X be an infinite S-recognizable set of nonnegative integers. Suppose*

$$\forall i \in \{0, \ldots, p-1\}, \ \mathbf{v}_L(np+i) \sim a_i n^c \beta^{rn} \ (n \to +\infty),$$

for some $p, c \in \mathbb{N}$ with $p \geq 1$, some $\beta \geq 1$, some $r \geq 1$, and some positive constants $a_0, \ldots, a_{p-1}$, and

$$\forall j \in \{0, \ldots, q-1\}, \ \mathbf{v}_{\mathrm{rep}_S(X)}(nq+j) \sim b_j n^d \beta^n \ (n \to +\infty),$$

for some $q, d \in \mathbb{N}$ with $q \geq 1$ and some positive constants $b_0, \ldots, b_{q-1}$. Then we have

- *$t_X(n) = \Theta\left((\log(n))^{c-dr\frac{q}{p}} n^{r\frac{q}{p}}\right)$ if $\beta > 1$;*
- *$t_X(n) = \Theta\left(n^{\frac{c}{d}}\right)$ if $\beta = 1$.*

Proof. The case where $\beta > 1$ is simply a rewriting of Theorem 7 with $\alpha = \beta^r$ and $r = \frac{\log(\alpha)}{\log(\beta)}$. For the case $\beta = 1$, observe that the hypotheses imply $\beta^r = 1$. Hence, the conclusion follows directly from Theorem 7 with $\alpha = \beta = 1$. $\square$

4. Achieving particular behaviors

Recall that a language L is *polynomial* if $\mathbf{u}_L(n)$ is $O(n^d)$ for some $d \in \mathbb{N}$ and is *exponential* if there exist $c > 0$ and $\theta > 1$ such that the inequality $\mathbf{u}_L(n) \geq c\theta^n$ holds for infinitely many integers n. Let us recall now the following gap theorem.

Theorem 22. [SYZS92] *Any regular language is either polynomial or exponential.*

Proposition 23.
- *For all $k, \ell \in \mathbb{N}$ with $\ell > 0$, there exists an abstract numeration system S built on an exponential regular language and an infinite S-recognizable set $X \subseteq \mathbb{N}$ such that $t_X(n) = \Theta((\log(n))^k n^\ell)$.*
- *For all $k, \ell \in \mathbb{N}$ with $\ell > 1$, there exists an abstract numeration system S built on an exponential regular language and an infinite S-recognizable set $X \subseteq \mathbb{N}$ such that $t_X(n) = \Theta\left(\frac{n^\ell}{(\log(n))^k}\right)$.*
- *For all positive integer k and for all abstract numeration systems S, there is no S-recognizable set $X \subseteq \mathbb{N}$ such that $t_X(n) = \Theta\left(\frac{n}{(\log(n))^k}\right)$.*

Proof. Let ℓ be a positive integer and let $k \in \mathbb{N}$.

From [Rig02, Proposition 17], for all $c \in \mathbb{N}$, there exists a regular language L having $\mathbf{v}_L(n) = (n+1)^c 2^{\ell(n+1)}$ as growth function (for $n \geq 1$). This language L is obtained by considering unions and shuffles of regular languages over distinct alphabets, i.e, alphabets having empty pairwise intersections. In particular, for all nonnegative integers $b \leq c$, it contains a regular sublanguage $M^{(b)}$ such that $\mathbf{u}_{M^{(b)}}(n) = n^b 2^n$ for all $n \in \mathbb{N}$.

Let $S = (L, \Sigma, <)$ be an abstract numeration system built on a regular language L satisfying $\mathbf{v}_L(n) \sim n^k 2^{\ell(n+1)}$ $(n \to +\infty)$. This language is defined as in the previous paragraph. Then the set $X = \mathrm{val}_S(M^{(0)})$ is an infinite S-recognizable set such that $\mathbf{v}_{\mathrm{rep}_S(X)}(n) \sim 2^{n+1}$

$(n \to +\infty)$. From Theorem 7, we obtain $t_X(n) = \Theta((\log(n))^k n^\ell)$. This proves the first assertion.

Now we assume $\ell \geq 2$. Choose any integer $d \geq \frac{k}{\ell-1}$ and let $c = \ell d - k$. From the choice of d, we have $c \geq d$. Let $S = (L, \Sigma, <)$ be an abstract numeration system built on a regular language L satisfying $\mathbf{v}_L(n) \sim n^c 2^{\ell(n+1)}$ $(n \to +\infty)$. This language is defined as in the first paragraph. Then the set $X = \mathrm{val}_S(M^{(d)})$ is an infinite S-recognizable set such that $\mathbf{v}_{\mathrm{rep}_S(X)}(n) \sim n^d 2^{n+1}$ $(n \to +\infty)$. From Theorem 7, we obtain $t_X(n) = \Theta\left(\frac{n^\ell}{(\log(n))^k}\right)$. This proves the second assertion.

Consider now the third assertion and assume $k > 0$. Let $S = (L, \Sigma, <)$ be an abstract numeration system. Suppose that such a set X exists. In view of Lemma 19 and Theorem 7, we should have

$$\forall j \in \{0, \ldots, q-1\}, \; \mathbf{v}_{\mathrm{rep}_S(X)}(nq + j) \sim b_j n^d \beta^n \; (n \to +\infty),$$

for some $q, d \in \mathbb{N}$ with $q \geq 1$, some $\beta > 1$ and some positive constants $b_0, \ldots, b_{q-1}$. In the same way, we can write

$$\forall i \in \{0, \ldots, p-1\}, \; \mathbf{v}_L(np + i) \sim a_i n^c \alpha^n \; (n \to +\infty),$$

for some $p, c \in \mathbb{N}$ with $p \geq 1$, some $\alpha > 1$ and some positive constants $a_0, \ldots, a_{p-1} > 0$. Then we must have $\sqrt[p]{\alpha} = \sqrt[q]{\beta}$. From Theorem 7 and Remark 20, we then obtain $t_X(n) = \Theta((\log(n))^{c-d} n)$ with $c \geq d$, a contradiction. This ends the proof. $\qquad\square$

The following corollary of Theorem 7 considers the case of a polynomial numeration language.

Corollary 24. *Let $S = (L, \Sigma, <)$ be an abstract numeration system built on a polynomial regular language and let X be an infinite S-recognizable set of nonnegative integers. Then we have $t_X(n) = \Theta(n^r)$ for some rational number $r \geq 1$.*

Proof. Since L is an infinite polynomial regular language, its growth function $\mathbf{v}_L(n)$ must satisfy

$$\forall i \in \{0, \ldots, p-1\}, \; \mathbf{v}_L(np + i) \sim a_i n^c \; (n \to +\infty),$$

for some $p, c \in \mathbb{N}$ with $p \geq 1$ and some positive constants $a_0, \ldots, a_{p-1}$. The sublanguage $\mathrm{rep}_S(X)$ of L is necessarily polynomial too, and since X is an infinite S-recognizable set, the growth function $\mathbf{v}_{\mathrm{rep}_S(X)}(n)$ must satisfy

$$\forall j \in \{0, \ldots, q-1\}, \; \mathbf{v}_{\mathrm{rep}_S(X)}(nq + j) \sim b_j n^d \; (n \to +\infty),$$

for some $q, d \in \mathbb{N}$ with $q \geq 1$ and $d \leq c$, and some positive constants $b_0, \ldots, b_{q-1}$. Then from Theorem 7, we obtain $t_X(n) = \Theta(n^{\frac{c}{d}})$. $\qquad\square$

Proposition 25. *For every rational number $r \geq 1$, there exists an abstract numeration system S built on a polynomial regular language and an infinite S-recognizable set of nonnegative integers X such that $t_X(n) = \Theta(n^r)$.*

Proof. Fix a rational number $r \geq 1$. Write $r = \frac{c}{d}$ where c and d are positive integers. Define $\mathcal{B}_\ell$ to be the *bounded language* $a_1^* a_2^* \cdots a_\ell^*$. We have $\mathbf{v}_{\mathcal{B}_\ell}(n) = \binom{n+\ell}{\ell}$ for all $\ell \geq 1$ and $n \in \mathbb{N}$ (for example see [CRS08, Lemma 1]). Let S be the abstract numeration system built on $\mathcal{B}_c$ with the order $a_1 < a_2 < \cdots < a_c$ and let $X = \mathrm{val}_S(\mathcal{B}_d)$ (since $c \geq d$, we have $\mathcal{B}_d \subseteq \mathcal{B}_c$). Hence we have $\mathbf{v}_{\mathcal{B}_c}(n) = \binom{n+c}{c}$ and $\mathbf{v}_{\mathrm{rep}_S(X)}(n) = \binom{n+d}{d}$ for all $n \in \mathbb{N}$. Then from Theorem 7, we obtain $t_X(n) = \Theta(n^{\frac{c}{d}}) = \Theta(n^r)$. $\qquad\square$

5. EXAMPLES

In this section we provide several examples to illustrate the constructions of Sections 3 and 4.

Example 26. Let us continue Example 14. Consider the abstract numeration system S built on the language L accepted by the automaton of Figure 1 from which are removed the words beginning with 0 and the alphabet order $0 < 1 < 2 < 3$. The trim minimal automaton of L is depicted in Figure 2. The set X is S-recognizable since $\mathrm{rep}_S(X) = 2\{0,2\}^* \cup \{\varepsilon\}$.
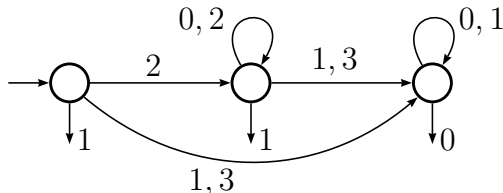

FIGURE 2. The trim minimal automaton of L.

We have $\mathbf{u}_L(n) = (n+2)2^{n-1}$ if $n \geq 1$ and $\mathbf{u}_L(0) = 1$ and $\mathbf{u}_{\mathrm{rep}_S(X)}(n) = 2^{n-1}$ if $n \geq 1$ and $\mathbf{u}_{\mathrm{rep}_S(X)}(0) = 1$. This gives $\mathbf{v}_L(n) = (n+1)2^n$ and $\mathbf{v}_{\mathrm{rep}_S(X)}(n) = 2^n$ for all $n \in \mathbb{N}$. Observe that accordingly to Lemma 18, we have $|h^n(1)| = \mathbf{v}_L(n) = (n+1)2^n$ and $F(n) = \mathbf{v}_{\mathrm{rep}_S(X)}(n) = 2^n$ for all $n \in \mathbb{N}$, where $F(n)$ is the number of occurrences of the letter 1 in $|h^n(1)|$. Recall that we found

$$t_X(n) = \Theta(n \log(n)),$$

which is consistent with Theorem 7.

Example 27. Consider the base 4 numeration system, that is, the abstract numeration system built on $\mathcal{L}_4 = \{\varepsilon\} \cup \{1,2,3\}\{0,1,2,3\}^*$ with the natural order on the digits. Let $X = \mathrm{val}_4(L) = \{0,2,8,10,32,34,40,42,128,130,136,138,160,\ldots\}$ where L is the language accepted by the automaton of Figure 2. It is 4-recognizable. From Example 26 and from Theorem 7, we obtain

$$t_X(n) = \Theta\left(\left(\frac{n}{\log(n)}\right)^2\right).$$

It is well known that the set of squares $\{n^2 \mid n \in \mathbb{N}\}$ is not b-recognizable for all integer bases $b \geq 2$ (for instance see [Eil74]). In [Str90, Rig02] (also see Theorem 33 below), it was shown that any set of the form $\{n^k \mid n \in \mathbb{N}\}$, with $k \in \mathbb{N}$, is S-recognizable for some S. In those constructions, the exhibited abstract numeration systems are built on polynomial languages. In the following example, we exhibit 4-recognizable sets of nonnegative integers having their n-th terms in $\Theta(n^k)$ for $k = 2$ and some $k \notin \mathbb{N}$. These considerations have to be compared with Proposition 23 above and Proposition 35 in the next section.

Example 28. Consider again the base 4 numeration system. Let $X = \mathrm{val}_4(\{1,3\}^*) = \{1,3,5,7,13,15,21,23,29,31,\ldots\}$. It is clearly 4-recognizable. We have $\mathbf{v}_{\mathcal{L}_4}(n) = 4^n$ and $\mathbf{v}_{\{1,3\}^*}(n) = 2^{n+1} - 1$ for all $n \in \mathbb{N}$. From Theorem 7, we obtain

$$t_X(n) = \Theta(n^2).$$

This also illustrates Proposition 23.

Now let $Y = \mathrm{val}_4(\{1,2,3\}^*) = \{0,1,3,5,6,7,9,10,11,13,14,15,21,22,\ldots\}$. It is clearly 4-recognizable. We have $\mathbf{v}_{\{1,2,3\}^*}(n) = \frac{1}{2}(3^{n+1}-1)$ for all $n \in \mathbb{N}$. From Theorem 7, we obtain

$$t_Y(n) = \Theta\left(n^{\frac{\log(4)}{\log(3)}}\right).$$

This also illustrates Corollary 21.

Define $L_F = \{\varepsilon\} \cup 1(0+01)^*$ to be the language of the Fibonacci numeration system and let $Z = \mathrm{val}_4(L_F) = \{0,1,4,16,17,64,70,256,257,260,272,273,1024,\ldots\}$. Again it is 4-recognizable. We have $\mathbf{v}_{L_F}(n) = \frac{5+3\sqrt{5}}{10}(\frac{1+\sqrt{5}}{2})^n + \frac{5-3\sqrt{5}}{10}(\frac{1-\sqrt{5}}{2})^n$ for all $n \in \mathbb{N}$. Therefore, we find

$$\mathbf{v}_{L_F}(n) \sim \frac{5+3\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{2}\right)^n \quad (n \to +\infty)$$

and by Theorem 7, we obtain

$$t_Z(n) = \Theta\left(n^{\frac{\log(4)}{\log(\varphi)}}\right)$$

where $\varphi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

Now we illustrate Theorem 7 when $p = 1$ and $q = 2$. Define K to be the language accepted by the automaton depicted in Figure 3. Let $V = \mathrm{val}_4(K)$. The first values of V are
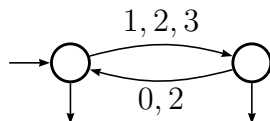

FIGURE 3. The trim minimal automaton of K.

$0,1,2,3,4,6,8,10,12,14,17,18,19,25,26,27,33$. It is 4-recognizable. We have $\mathbf{u}_K(2n) = 6^n$ and $\mathbf{u}_K(2n+1) = 3 \cdot 6^n$ for all $n \in \mathbb{N}$. Then $\mathbf{v}_K(2n) \sim \frac{9}{5}6^n$ and $\mathbf{v}_K(2n+1) \sim \frac{24}{5}6^n$ $(n \to +\infty)$. From Theorem 7, we obtain

$$t_V(n) = \Theta\left(n^{\frac{\log(4)}{\log(\sqrt{6})}}\right).$$

Example 29. Consider the base 2 numeration system, that is, the abstract numeration system built on $\mathcal{L}_2 = \{\varepsilon\} \cup 1\{0,1\}^*$ with the natural order on the digits. Let $X = \mathrm{val}_2(1^*0^*) = \{0,1,2,4,6,7,8,12,15,16,24,28,30,31,\ldots\}$. It is 2-recognizable. We have $\mathbf{v}_{1^*0^*}(n) = \binom{n+2}{2}$ for all $n \in \mathbb{N}$. From Theorem 7, we obtain

$$t_X(n) = 2^{(1+o(1))\sqrt{2n}}.$$

We can also use our main result to show that certain sets of integers are not S-recognizable for any abstract numeration system S.

Example 30. Let $C = (C_n)_{n\geq 0}$ denote the set of *Catalan numbers* [Slo]: i.e.,

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

These numbers occur in many counting problems; for example, they count the number of *Dyck words* of length $2n$. Asymptotically, we have

$$C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}},$$

which does not correspond to any of the forms described by Theorem 7. Hence, for all S, the set C is not S-recognizable.

6. ADDITIONAL RESULTS

Ultimately periodic sets of integers play a special role. On the one hand, such infinite sets are coded by a finite amount of information. On the other hand, the famous theorem of Cobham asserts that these sets are the only ones that are recognizable in all integer base numeration systems. The following result shows that this property extends to abstract numeration systems.

Theorem 31. [LR01] *Any ultimately periodic set is S-recognizable for all abstract numeration systems S built on a regular language.*

In [CLR], the latter result is extended to the multidimensional case.

Ultimately periodic sets are polynomial sets of degree 1. The following example shows that there exists non-ultimately periodic polynomial sets of degree 1 that are recognized in some abstract numeration systems.

Example 32. Consider the integer base 2 numeration system. Let $t = (t_n)_{n \geq 0} \in \{0,1\}^{\mathbb{N}}$ be the Thue-Morse sequence defined as follows: $t_n = (s_2(n) \mod 2)$ where $s_2(n)$ is the number of 1's in the 2-representation $\text{rep}_{S_2}(n)$ of n. Let $T \subseteq \mathbb{N}$ be the characteristic set of t: $n \in T$ if and only if $t_n = 1$. It is well-known that t is 2-automatic, which is equivalent to the fact that T is S_2-recognizable. It is easily seen that the growth function $t_T(n)$ of T is bounded by $2n$. Furthermore, T is not an ultimately periodic set.

Theorem 33. [Rig02] *Let k be a positive integer and for all $i \in \{1, \ldots, k\}$, let $P_i \in \mathbb{Q}[x]$ be such that $P_i(\mathbb{N}) \subseteq \mathbb{N}$ and let $\alpha_i \in \mathbb{N}$. Define $f(x) = \sum_{i=1}^{k} P_i(x)\alpha_i^x$. Then an abstract numeration system S built on a regular language L such that $f(\mathbb{N})$ is S-recognizable can be effectively provided. Furthermore, if $f(\mathbb{N})$ is polynomial (resp. exponential), then the numeration language of the provided abstract numeration system is polynomial (resp. exponential).*

Let us recall the following result of Durand and Rigo.

Theorem 34. [DR09] *Let S be an abstract numeration system built on a polynomial regular language and let T be an abstract numeration system built on an exponential regular language. If a subset of $\mathbb{N}$ is both S-recognizable and T-recognizable, then it is ultimately periodic.*

Proposition 35. *Let S be an abstract numeration system built on an exponential regular language. If $f \in \mathbb{Q}[x]$ is a polynomial of degree greater than 1 such that $f(\mathbb{N}) \subseteq \mathbb{N}$, then the set $f(\mathbb{N})$ is not S-recognizable.*

Proof. It follows directly from Theorems 33 and 34. $\qquad \square$

The latter result has to be compared with Proposition 23.

7. ACKNOWLEDGMENTS

References

[AS03]   J.-P. Allouche and J. Shallit. *Automatic Sequences. Theory, Applications, Generalizations*. Cambridge University Press, Cambridge, 2003.

[BR88]   J. Berstel and Ch. Reutenauer. *Rational series and their languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Berlin, 1988.

[BR10]   V. Berthé and M. Rigo, editors. *Combinatarics, Automata and Number Theory*, volume 135 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[CKR10]   E. Charlier, T. Kärki, and M. Rigo. Multidimensional generalized automatic sequences and shape-symmetric morphic words. *Discrete Math.*, 310:1238–1252, 2010.

[CLR]   E. Charlier, A. Lacroix, and N. Rampersad. Multi-dimensional sets recognizable in all abstract numeration systems. Submitted.

[CRS08]   E. Charlier, M. Rigo, and W. Steiner. Abstract numeration systems on bounded languages and multiplication by a constant. *Integers*, 8:A35, 19, 2008.

[DR09]   Fabien Durand and Michel Rigo. Syndeticity and independent substitutions. *Adv. in Appl. Math.*, 42(1):1–22, 2009.

[Eil74]   S. Eilenberg. *Automata, Languages, and Machines*, volume A. Academic Press, New York, 1974. Pure and Applied Mathematics, Vol. 58.

[LR01]   P. B. A. Lecomte and M. Rigo. Numeration systems on a regular language. *Theory Comput. Syst.*, 34(1):27–44, 2001.

[Rig00]   M. Rigo. Generalization of automatic sequences for numeration systems on a regular language. *Theoret. Comput. Sci.*, 244(1-2):271–281, 2000.

[Rig02]   M. Rigo. Construction of regular languages and recognizability of polynomials. *Discrete Math.*, 254(1-3):485–496, 2002.

[RM02]   M. Rigo and A. Maes. More on generalized automatic sequences. *J. Autom., Lang. and Comb.*, 7(3):351–376, 2002.

[Slo]   N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. `http://oeis.org/A000108`. Catalan numbers.

[SS78]   A. Salomaa and M. Soittola. *Automata-theoretic aspects of formal power series*. Springer-Verlag, New York, 1978. Texts and Monographs in Computer Science.

[Str90]   A. S. Strogalov. Regular languages with polynomial growth in the number of words. *Diskret. Mat.*, 2(3):146–152, 1990.

[SYZS92] A. Szilard, S. Yu, K. Zhang, and J. Shallit. Characterizing regular languages with polynomial densities. In *Mathematical Foundations of Computer Science 1992 (Prague, 1992)*, volume 629 of *Lecture Notes in Comput. Sci.*, pages 494–503. Springer, Berlin, 1992.

(É. Charlier) David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada

E-mail address, É. Charlier: `echarlier@uwaterloo.ca`

(N. Rampersad) Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37), B-4000 Liège, Belgium

E-mail address, N. Rampersad: `nrampersad@ulg.ac.be`