Syntactic Complexity of Ultimately Periodic Sets of Integers

Michel Rigo and Élise Vandomme*

Institute of Mathematics, University of Liège, Grande Traverse 12 (B 37),
B-4000 Liège, Belgium. {e.vandomme,M.Rigo}@ulg.ac.be

Abstract. We compute the cardinality of the syntactic monoid of the language $0^* \operatorname{rep}_b(m\mathbb{N})$ made of base b expansions of the multiples of the integer m. We also give lower bounds for the syntactic complexity of any (ultimately) periodic set of integers written in base b. We apply our results to some well studied problem: decide whether or not a b-recognizable sets of integers is ultimately periodic.

1 Introduction

Syntactic complexity has received some recent and renewed interest. See for instance [6] for some background and we quote "*in spite of suggestions that syntactic semigroups deserve to be studied further, relatively little has been done on the syntactic complexity of a regular language*". In this paper syntactic complexity is introduced in the framework of numeration systems.

We compute the syntactic complexity of the set $m\mathbb{N}$ written in base b, i.e., the cardinality $M_{b,m}$ of the syntactic monoid of the language $0^* \operatorname{rep}_b(m\mathbb{N})$ made of base b expansions of the multiples of the integer m. A similar problem was solved for the state complexity of the language $0^* \operatorname{rep}_b(m\mathbb{N})$, i.e., the number of states of its minimal automaton. As usual (m, n) denotes the GCD of m and n.

Theorem 1 (B. Alexeev [1]). *Let $b, m \geq 2$ be integers. Let N, M be such that $b^N < m \leq b^{N+1}$ and $(m,1) < (m,b) < \cdots < (m,b^M) = (m,b^{M+1}) = (m,b^{M+2}) = \cdots$. The minimal automaton of $0^* \operatorname{rep}_b(m\mathbb{N})$ has exactly*

$$\frac{m}{(m,b^{N+1})} + \sum_{t=0}^{\inf\{N,M-1\}} \frac{b^t}{(m,b^t)} \text{ states.}$$

For the binary system, the first few values of $M_{2,m}$ are given below. Let $b \geq 2$. In this paper, we obtain an explicit formula for $M_{b,m}$ as a consequence of Theorems 2, 3 and 4 where we discuss three cases: the constant m and the base b are coprime or, m is a power of b or, $m = b^n q$ with $(q,b) = 1$, $q \geq 2$ and $n \geq 1$. Furthermore, we provide lower bounds for the syntactic complexity of any ultimately periodic set of integers written in base b, i.e., any finite union

* with the support of a grant CGRI-FRS/CNRS Wallonie-Bruxelles International.

of arithmetic progressions. In the framework of numeration systems, syntactic complexity has an advantage in comparison to left or right quotients, we have the opportunity to work simultaneously on prefixes and suffixes of base b expansions, that is on most and least significant digits.

m	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
$M_{2,m}$	3	6	5	20	13	21	7	54	41	110	20	156	43	60	9	136	109	342	62	126	221

A motivation for this work comes from the following decision problem. Let S be an abstract numeration system built on a regular language. See [4, Chap. 3] for background. It is well-known that any ultimately periodic set is S-recognizable, i.e., it has a regular language of representations within the system S. An instance of the decision problem is given by an abstract numeration system S and a DFA accepting some S-recognizable set $X \subseteq \mathbb{N}$. The question is therefore to decide whether X is ultimately periodic or not. This problem was settled positively for integer base systems by Honkala in [9]. See also [2] and in particular [5] for a first order logic approach. Recently this decision problem was settled positively in [3] for a large class of numeration systems based on linear recurrence sequences. Considering this decision problem for any abstract numeration system turns out to be equivalent to the so-called ω-HD0L ultimate periodicity decision problem, see again [4], or [10]. In its full generality, this problem is still open.

Since syntactic complexity provides an alternative measure for the complexity of a regular language, one could try to develop new decision procedures based on the syntactic complexity instead of the state complexity of the corresponding languages. A step in that direction is to consider first integer base numeration systems. As a consequence of our results, we present such a procedure restricted to a prime base in Section 4.

In the next section, we recall basic definitions and fix notation. Section 3 contains our main results: Theorems 2, 3, 4 are about the particular sets $m\mathbb{N}$ and Propositions 2, 3 as well as Theorem 2 are about any periodic set. We end the paper with a procedure for the decision problem described above and we present some directions for future work.

2 Definitions

For $i \leq j$, we denote by $[\![i,j]\!]$ the interval of integers $\{i, i+1, \ldots, j-1, j\}$. A *deterministic finite automaton* (or DFA) over the alphabet A is a 5-tuple $\mathcal{A} = (Q, q_0, F, A, \delta)$ where Q is the set of states, q_0 is the initial state, F is the set of final states and $\delta : Q \times A^* \to Q$ is the (extended) transition function. We denote by $|u|$ the length of the word $u \in A^*$ and by $\#P$ the cardinality of P.

Integer Base Numeration Systems Let $b \geq 2$ be an integer. We denote by A_b the canonical alphabet of digits $[\![0, b-1]\!]$. For any word $u = u_\ell \cdots u_0 \in A_b^*$, we define the *numerical value* of u as

$$\mathrm{val}_b(u) = \sum_{i=0}^{\ell} u_i\, b^i.$$

Note that $\mathrm{val}_b(uv) = \mathrm{val}_b(u)\, b^{|v|} + \mathrm{val}_b(v)$ for all $u, v \in A_b^*$. For any integer $n > 0$, we denote the usual base b expansion of n by $\mathrm{rep}_b(n)$. We assume that such a greedy expansion does not start with 0. By convention, $\mathrm{rep}_b(0)$ is the empty word ε. A set X of integers is said to be *b-recognizable* if the language $\mathrm{rep}_b(X) \subseteq A_b^*$ is a regular language accepted by some DFA.

A set $X \subseteq \mathbb{N}$ is *periodic* of *period* p if for all $n \in \mathbb{N}$, $n \in X \Leftrightarrow n + p \in X$. The period is always understood to be the minimal period of X. In particular, if $X \subseteq \mathbb{N}$ is periodic of period p, then for all $i, j \in \mathbb{N}$,

$$i \not\equiv j \bmod p \Rightarrow \exists r \in [\![0, p-1]\!] : (i + r \in X, j + r \notin X) \text{ or } (i + r \notin X, j + r \in X). \quad (1)$$

A set $X \subseteq \mathbb{N}$ is *ultimately periodic* of *period* p and *index* $I > 0$ if for all $n \geq I$, $n \in X \Leftrightarrow n + p \in X$ and exactly one of the two elements $I - 1, I + p - 1$ is in X. Again, index and period are always understood to be minimal. It is easy to see that any ultimately periodic set is b-recognizable for all bases $b \geq 2$.

Syntactic Complexity Let L be a language over the finite alphabet A. The *context* of a word $u \in A^*$ with respect to L is given by the set of pairs

$$\mathtt{C}_L(u) = \{(x, y) \in A^* \times A^* \mid xuy \in L\}.$$

If L is clearly understood, we will simply write $\mathtt{C}(u)$. Define the *Myhill congruence* [11] of L by $u \leftrightarrow_L v$ if and only if, for all $x, y \in A^*$, $xuy \in L \Leftrightarrow xvy \in L$. In other words, $u \leftrightarrow_L v$ if and only if $\mathtt{C}_L(u) = \mathtt{C}_L(v)$. This congruence is also known as the *syntactic congruence* of L. The monoid $A^*/{\leftrightarrow_L}$ made of the equivalence classes of the relation $\leftrightarrow_L$, is the *syntactic monoid* of L. It is well-known that L is a regular language if and only if $A^*/{\leftrightarrow_L}$ is finite. The *syntactic complexity* of L is the cardinality of its syntactic monoid. If $X \subseteq \mathbb{N}$ is a b-recognizable set of integers, by extension we define the *syntactic complexity* of X (w.r.t. b) as the syntactic complexity of the language $0^* \mathrm{rep}_b(X)$.

Proposition 1. *Let L be a language over A. Two words $u, v \in A^*$ are such that $u \leftrightarrow_L v$ if and only if they perform the same transformation on the set of states of the minimal automaton $\mathcal{M} = (Q_L, q_{0,L}, F_L, A, \delta_L)$ of L, i.e., for all $r \in Q_L$, $\delta_L(r, u) = \delta_L(r, v)$. In particular, if u, v are such that $\delta_L(q_{0,L}, u) \neq \delta_L(q_{0,L}, v)$, then $u \not\leftrightarrow_L v$.*

Definition 1. *A language $L \subseteq A^*$ is* weakly n-definite, *if for any $x, y \in A^*$ satisfying $|x| \geq n$, $|y| \geq n$ and having the same suffix of length n, $x \in L$ if and only if $y \in L$ [12, 7]. In other words, L can be written as $G \cup A^*F$ where F (resp. G) is finite and contains only words of length n (resp. less than n). Let $n \geq 1$. A language is* n-definite *if it is weakly n-definite and not weakly $(n-1)$-definite. One also finds the terminology* suffix testable *in the literature, see [13].*

3 Main Results

Let $m, x \geq 2$ be integers such that $(m, x) = 1$. We denote by $\mathrm{ord}_m(x)$ the order of x in the multiplicative group $U(\mathbb{Z}/m\mathbb{Z})$ made of the invertible elements in $\mathbb{Z}/m\mathbb{Z}$.

That is $\mathrm{ord}_m(x)$ is the smallest positive integer j such that $x^j \equiv 1 \bmod m$. In particular, $\mathrm{ord}_m(x)$ is the period of the sequence $(x^n \bmod m)_{n\geq 0}$.

We first consider the case where the base and the period are coprime. Interestingly, the syntactic complexity depends only on the period and not on the structure of the periodic set.

Theorem 2. *Let $m, b \geq 2$ be integers such that $(m, b) = 1$. If $X \subseteq \mathbb{N}$ is periodic of (minimal) period m, then its syntactic complexity is given by $m.\,\mathrm{ord}_m(b)$. In particular, this result holds for $X = m\mathbb{N}$.*

Now consider the case where the period is a power of the base.

Theorem 3. *Let $b \geq 2$ and $m = b^n$ with $n \geq 1$. Then the syntactic complexity of $0^* \mathrm{rep}_b(m\mathbb{N})$ is given by $M_{b,m} = 2n + 1$.*

Proposition 2. *Let $b \geq 2$. If $X \subseteq \mathbb{N}$ is a periodic set of (minimal) period $m = b^n$ with $n \geq 1$, then the syntactic complexity of $L = 0^* \mathrm{rep}_b(X)$ is greater than or equal to $n+1$. Moreover there exist arbitrarily large integers $t_1, \ldots, t_{n+1}$ such that the $n+1$ words $\mathrm{rep}_b(t_1), \ldots, \mathrm{rep}_b(t_{n+1})$ belong to different equivalence classes of $\leftrightarrow_L$.*

Remark 1. The bound in Proposition 2 is tight. One can for instance consider the set $5 + 8\mathbb{N}$. The corresponding syntactic monoid has exactly four infinite equivalence classes.

We now turn to the case where $m = b^n q$ with $(q, b) = 1$ and $n \geq 1$.

Theorem 4. *Let $b \geq 2$ and $m = b^n q$ where $n \geq 1$ and $(q, b) = 1$ and $q \geq 2$. Then the syntactic complexity of $0^* \mathrm{rep}_b(m\mathbb{N})$ is given by*

$$M_{b,m} = (n + 1).M_{b,q} + n = (n + 1).q.\,\mathrm{ord}_q(b) + n.$$

For the next proposition, we restrict ourselves in this abstract to the case where b is a prime number.

Proposition 3. *Let b be a prime number and $m = b^n q$ where $n \geq 1$ and $(q, b) = 1$ and $q \geq 2$. If $X \subseteq \mathbb{N}$ is periodic of (minimal) period m, then the syntactic complexity of $0^* \mathrm{rep}_b(X)$ is greater than or equal to $(n + 1).q$. Moreover there exist arbitrarily large integers $t_1, \ldots, t_{(n+1)q}$ such that $\mathrm{rep}_b(t_1), \ldots, \mathrm{rep}_b(t_{(n+1)q})$ belong to different equivalence classes of $\leftrightarrow_{0^* \mathrm{rep}_b(X)}$.*

Note that if b is not a prime number, there are integers of the kind $m = b^n q$ where n is maximal and $(b, q) > 1$, as an example take $b = 4$ and $m = 72 = 4.18$. Such a situation is not taken into account by Theorem 2, Propositions 2 and 3.

4 Application to a Decision Procedure

Let $X \subseteq \mathbb{N}$ be a b-recognizable set of integers such that $0^* \operatorname{rep}_b(X)$ is accepted by some DFA $\mathcal{A}$. A usual technique for deciding whether or not X is ultimately periodic is to prove that whenever X is ultimately periodic, then its period and its preperiod must be bounded by some quantities depending only on the size of the DFA $\mathcal{A}$. Therefore, one has a finite number of admissible periods and preperiods to test leading to a decision procedure. For details, see [3]. In particular, the following result [3, Prop. 44] stated in full generality for any abstract numeration system (i.e., the language of numeration is a regular language) shows that we have only to obtain an upper bound on the admissible periods.

Proposition 4. *Let $S = (L, \Sigma, <)$ be an abstract numeration system. If $X \subseteq \mathbb{N}$ is an ultimately periodic set of period p_X such that $\operatorname{rep}_S(X)$ is accepted by a DFA with d states, then the preperiod of X is bounded by an effectively computable constant C depending only on d and p_X.*

The following result is a consequence of Theorem 2, Propositions 2 and 3.

Theorem 5. *If $X \subseteq \mathbb{N}$ is an ultimately periodic set of period $p_X = b^n q$ with $(q, b) = 1$ and $n \geq 0$, then the syntactic complexity of $0^* \operatorname{rep}_b(X)$ is greater than or equal to $(n + 1)q$.*

Assume that $b \geq 2$ is a prime number. Therefore, giving a DFA $\mathcal{A}$ accepting $0^* \operatorname{rep}_b(X)$ and so the corresponding syntactic monoid, if X is ultimately periodic, then we get an upper bound on its period.

5 Further Work

We will try to extend the present work to a wider class of numeration systems. For instance, for the Fibonacci numeration system (defined by the sequence $F_{n+2} = F_{n+1} + F_n$, $F_0 = 1$, $F_1 = 2$) where integers are represented using the greedy algorithm, the syntactic complexity of $0^* \operatorname{rep}_F(m\mathbb{N})$ is given by $M_{F,m} = 4.m^2.P_F(m) + 2$ where $P_F(m)$ is the period of $(F_i \bmod m)_{i \geq 0}$. The proof essentially follows the same lines as in the proof of Theorem 2 and [8]. Recall that a word is a valid representation if it does not contain a factor 11. This later fact explains the factor 4 in the expression $M_{F,m}$. Let $u = u_k \cdots u_0, v = v_\ell \cdots v_0 \in \{0, 1\}^*$. We have $u \leftrightarrow_{0^* \operatorname{rep}_F(m\mathbb{N})} v$ if and only if

$$(\operatorname{val}_F(u) \equiv \operatorname{val}_F(v) \bmod m) \wedge (\operatorname{val}_F(u0) \equiv \operatorname{val}_F(v0) \bmod m) \wedge \\ (|u| \equiv |v| \bmod P_F(m)) \wedge (u_k = v_\ell) \wedge (u_0 = v_0).$$

For the Tribonacci numeration system, the syntactic complexity of $0^* \operatorname{rep}_T(m\mathbb{N})$ is given by $M_{T,m} = 9.m^3.P_T(m) + 3$.

References

1. Alexeev, B.: Minimal DFA for testing divisibility. J. Comput. System Sci. **69**(2) (2004) 235–243
2. Allouche, J.P., Rampersad, N., Shallit, J.: Periodicity, repetitions, and orbits of an automatic sequence. Theoret. Comput. Sci. **410**(30-32) (2009) 2795–2803
3. Bell, J.P., Charlier, E., Fraenkel, A.S., Rigo, M.: A decision problem for ultimately periodic sets in non-standard numeration systems. Int. J. Algebra and Computation **19** (2009) 809–839
4. Berthé, V., Rigo, M., eds.: Combinatorics, Automata and Number Theory. Volume 135 of Encyclopedia of Mathematics and its Applications. Cambridge University Press (2010)
5. Bruyère, V., Hansel, G., Michaux, C., Villemaire, R.: Logic and p-recognizable sets of integers. Bull. Belg. Math. Soc. **1** (1994) 191–238
6. Brzozowski, J., Ye, Y.: Syntactic complexity of ideal and closed languages. preprint `arXiv:1010.3263v1` (2010)
7. Brzozowski, J.A.: Canonical regular expressions and minimal state graphs for definite events. In Fox, J., ed.: Proceedings of the Symposium on Mathematical Theory of Automata New York. Volume 12 of MRI Symposia Series, Polytechnic Press of the Polytechnic Institute of Brooklyn (1963) 529–561
8. Charlier, E., Rampersad, N., Rigo, M., Waxweiler, L.: The minimal automaton recognizing $m\mathbb{N}$ in a linear numeration system. Integers (to appear).
9. Honkala, J.: A decision method for the recognizability of sets defined by number systems. Theor. Inform. Appl. **20** (1986) 395–403
10. Honkala, J., Rigo, M.: A note on decidability questions related to abstract numeration systems. Discrete Math. **285** (2004) 329–333
11. Myhill, J.: Finite automata and representation of events. Technical Report 57-624, Wright Air Development Center Technical Report (1957)
12. Perles, M., Rabin, M.O., Shamir, E.: The theory of definite automata. IEEE Trans. Electron. Comp. (1963) 233–243
13. Pin, J.E.: Syntactic semigroups. In Rozenberg, G., Salomaa, A., eds.: Handbook of language theory. Volume I. Springer Verlag (1997) 679–746