ULTIMATE PERIODICITY PROBLEM FOR LINEAR NUMERATION SYSTEMS

É. CHARLIER[1], A. MASSUIR[1], M. RIGO[1] AND E. ROWLAND[2]

Abstract. We address the following decision problem. Given a numeration system U and a U-recognizable set $X \subseteq \mathbb{N}$, i.e. the set of its greedy U-representations is recognized by a finite automaton, decide whether or not X is ultimately periodic. We prove that this problem is decidable for a large class of numeration systems built on linear recurrence sequences. Based on arithmetical considerations about the recurrence equation and on p-adic methods, the DFA given as input provides a bound on the admissible periods to test.

1. Introduction

Let us first recall the general setting of linear numeration systems that are used to represent, in a greedy way, non-negative integers by words over a finite alphabet of digits. See, for instance, [12]. Let $\mathbb{N} = \{0, 1, 2, \dots\}$.

Definition 1. A *numeration system* is given by an increasing sequence $U = (U_i)_{i\geq 0}$ of integers such that $U_0 = 1$ and $C_U := \sup_{i\geq 0}\lceil \frac{U_{i+1}}{U_i}\rceil$ is finite. Let $A_U = \{0, \dots, C_U - 1\}$ be the canonical alphabet of digits. The *greedy U-representation* of a positive integer n is the unique finite word $\mathrm{rep}_U(n) = w_\ell \cdots w_0$ over A_U satisfying

$$n = \sum_{i=0}^{\ell} w_i\, U_i,\ w_\ell \neq 0 \text{ and } \sum_{i=0}^{t} w_i\, U_i < U_{t+1},\ t = 0, \dots, \ell.$$

We set $\mathrm{rep}_U(0)$ to be the empty word ε. A set $X \subseteq \mathbb{N}$ of integers is *U-recognizable* if the language $\mathrm{rep}_U(X)$ over A_U is regular (i.e. accepted by a finite automaton).

Recognizable sets of integers are considered as particularly simple because membership can be decided by a deterministic finite automaton in linear time with respect to the length of the representation. It is well

2000 *Mathematics Subject Classification.* 68Q45, 11U05, 11B85, 11S85.

Key words and phrases. Decision problem ; numeration system ; automata theory ; linear recurrent sequence ; p-adic valuation.

known that such a property for a subset of $\mathbb{N}$ depends on the choice of the numeration system. For a survey on integer base systems, see [7]. For generalized numeration systems, see [28]. For basic results in automata theory, see, for instance [28, 30].

Definition 2. If $x = x_\ell \cdots x_0$ is a word over an alphabet of integers, then the *U-numerical value* of x is

$$\mathrm{val}_U(x) = \sum_{i=0}^{\ell} x_i\, U_i.$$

From the point of view of formal languages, it is quite desirable that $\mathrm{rep}_U(\mathbb{N})$ is regular; we want to be able to check whether or not a word is a valid greedy U-representation. This implies that U satisfies a linear recurrence relation. See, for instance, [32] or [3, Prop. 3.1.5].

Definition 3. A numeration system U is said to be *linear* if it ultimately satisfies a homogeneous linear recurrence relation with integer coefficients. There exist $k \geq 1$, $a_{k-1}, \ldots, a_0 \in \mathbb{Z}$ such that $a_0 \neq 0$ and $N \geq 0$ such that for all $i \geq N$,

$$U_{i+k} = a_{k-1} U_{i+k-1} + \cdots + a_0 U_i. \tag{1.1}$$

The polynomial $X^N(X^k - a_{k-1}X^{k-1} - \cdots - a_0)$ is called the *characteristic polynomial* of the system (where it is assumed that k and then N are chosen to be minimal).

The regularity of $\mathrm{rep}_U(\mathbb{N})$ is also important for another reason. The language $\mathrm{rep}_U(\mathbb{N})$ is regular if and only if every ultimately periodic set of integers is U-recognizable [19, Thm. 4]. In particular, as recalled in Proposition 17, if an ultimately periodic set X is given, then a DFA accepting $\mathrm{rep}_U(X)$ can effectively be obtained.

In this paper, we address the following decidability question. Our aim is to prove that this problem is decidable for a large class of numeration systems.

Problem 1. Given a linear numeration system U and a (deterministic) finite automaton $\mathcal{A}$ whose accepted language is contained in the numeration language $\mathrm{rep}_U(\mathbb{N})$, decide whether the subset X of $\mathbb{N}$ that is recognized by $\mathcal{A}$ is ultimately periodic, i.e. whether or not X is a finite union of arithmetic progressions (along a finite set).

This question about ultimately periodic sets is motivated by the celebrated theorem of Cobham. Let $p, q \geq 2$ be integers. If p and q are multiplicatively independent, i.e. $\frac{\log(p)}{\log(q)}$ is irrational, then the ultimately periodic sets are the only sets that are both p-recognizable and

q-recognizable [9]. These are exactly the sets definable by a first-order formula in the Presburger arithmetic $\langle \mathbb{N}, + \rangle$. Cobham's result has been extended to various settings; see [10, 24] for an application to morphic words. See [11] for a survey.

In this paper, we write greedy U-representations with most significant digit first (MSDF convention): the leftmost digit is associated with the largest U_ℓ occurring in the decomposition. Considering least significant digit first would not affect decidability (a language is regular if and only if its reversal is) but this could have some importance in terms of complexity issues not discussed here.

What is known. Let us quickly review cases where the decision problem is known to be decidable. Relying on number theoretic results, the problem was first solved by Honkala for integer base systems [16]. An alternative approach bounding the syntactic complexity of ultimately periodic sets of integers written in base b was studied in [18]. Recently a deep analysis of the structure of the automata accepting ultimately periodic sets has led to an efficient decision procedure for integer base systems [22, 5, 21]. An integer base system is a particular case of a Pisot system, i.e. a linear numeration system whose characteristic polynomial is the minimal polynomial of a Pisot number (an algebraic integer larger than 1 whose conjugates all have modulus less than one). For these systems, one can make use of first-order logic and the decidable extension $\langle \mathbb{N}, +, V_U \rangle$ of Presburger arithmetic [6]. For an integer base p, $V_p(n)$ is the largest power of p dividing n. A typical example of Pisot system is given by the Zeckendorf system based on the Fibonacci sequence $1, 2, 3, 5, 8, \ldots$. Given a U-recognizable set X, there exists a first-order formula $\varphi(n)$ in $\langle \mathbb{N}, +, V_U \rangle$ describing X. The formula

$$(\exists N)(\exists p)(\forall n \geq N)(\varphi(n) \Leftrightarrow \varphi(n+p))$$

thus expresses when X is ultimately periodic, N being a preperiod and p a period of X. The logical formalism can be applied to systems such that the addition is U-recognizable by an automaton, i.e. the set $\{(x, y, z) \in \mathbb{N}^3 : x + y = z\}$ is U-recognizable. This is the case for Pisot systems [13].

When addition is not known to be U-recognizable, other techniques must be sought. Hence the problem was also shown to be decidable for some non-Pisot linear numeration systems satisfying a gap condition $\lim_{i \to +\infty} U_{i+1} - U_i = +\infty$ and a more technical condition $\lim_{m \to +\infty} \mathcal{N}_U(m) = +\infty$ where $\mathcal{N}_U(m)$ is the number of residue classes that appear infinitely often in the sequence $(U_i \bmod m)_{i \geq 0}$; see [2]. An example of such a system is built on the relation $U_i = 3\,U_{i-1} + 2\,U_{i-2} +$

$3\,U_{i-3}$ [14]. For extra pointers to the literature (such as an extension to a multidimensional setting), the reader can follow the introduction in [2].

Our contribution. In view of the above summary, we are looking for a decision procedure that may be applied to non-Pisot linear numeration systems such that $\mathcal{N}_U(m) \not\to \infty$ when m tends to infinity. Hence we want to take into account systems where we are not able to apply a decision procedure based on first-order logic nor on the technique from [2]. We follow Honkala's original scheme: if a DFA $\mathcal{A}$ is given as input (the question being whether the corresponding recognized subset of $\mathbb{N}$ is ultimately periodic), the number of states of $\mathcal{A}$ should provide an upper bound on the admissible preperiods s and periods p. If there is a finite number of such pairs to test, then we build a DFA $\mathcal{A}_{s,p}$ for each pair (s, p) and one can test whether or not the two automata $\mathcal{A}$ and $\mathcal{A}_{s,p}$ accept the same language. This provides us with a decision procedure. Roughly speaking, if the given DFA is "small", then it cannot accept an ultimately periodic set with a minimal period being "overly complicated", i.e. "quite large".

Example 4. Here is an example of a numeration system based on a Parry (the β-expansion of 1 is finite or ultimately periodic, see [3, Chap. 2]) non-Pisot number β:

$$U_{i+4} = 2\,U_{i+3} + 2\,U_{i+2} + 2\,U_i.$$

Indeed, the largest root β of the characteristic polynomial is roughly 2.804, and -1.134 is another root of modulus larger than one. With the initial conditions $1, 3, 9, 23$, $\mathrm{rep}_U(\mathbb{N})$ is the regular language over $\{0, 1, 2\}$ of words avoiding factors 2202, 221 and 222. For details, see [3, Ex. 2.3.37] or [23]. When m is a power of 2, there is a unique congruence class visited infinitely often by the sequence $(U_i \bmod m)_{i\geq 0}$ because $U_i \equiv 0 \pmod{2^r}$ for large enough i. For such an example, $\mathcal{N}_U(m)$ does not tend to infinity and thus the previously known decision procedures may not be applied. This is a perfect candidate for which no decision procedures are known.

This paper is organized as follows. In Section 2, we make clear our assumptions on the numeration system. In Section 3, we collect several known results on periodic sets and U-representations. In particular, we relate the length of the U-representation of an integer to its value. The core of the paper is made of Section 4 where we discuss cases to bound the admissible periods. In particular, we consider two kinds of prime factors of the admissible periods: those that divide all the coefficients of the recurrence and those that don't, see (4.1). In Section 5, we apply the discussion of the previous section. First, we obtain a decision procedure when the gcd of the coefficients of the recurrence relation is 1, see Theorem 37. This extends the scope of results from [2]. On the other hand, if there exist primes dividing all the coefficients, our approach heavily relies on quite general arithmetic properties of linear recurrence relations. It has therefore inherent limitations because of notoriously difficult problems in p-adic analysis such as finding bounds on the growth rate of blocks of zeroes in the digit sequences of p-adic numbers of a special logarithmic form. We discuss the question and give illustrations of these p-adic techniques in Section 6. The paper ends with some concluding remarks.

2. Our setting

We have minimal assumptions on the considered linear numeration system U.

(H1) $\mathbb{N}$ is U-recognizable.

(H2) There are arbitrarily large gaps between consecutive terms:

$$\limsup_{i\to+\infty}(U_{i+1}-U_i)=+\infty.$$

(H3) The gap sequence $(U_{i+1}-U_i)_{i\geq 0}$ is ultimately non-decreasing: there exists $G\geq 0$ such that for all $i\geq G$,

$$U_{i+1}-U_i\leq U_{i+2}-U_{i+1}.$$

Note that Example 4 satisfies all the above assumptions. Let us make a few comments.

Remark 5. (H1) gives sense and meaning to our decision problem; under that assumption, ultimately periodic sets are U-recognizable. As recalled in the introduction, it is a well known result that (H1) implies that the numeration system $(U_i)_{i\geq 0}$ satisfies a linear recurrence relation with integer coefficients as in (1.1). Some sufficient conditions that guarantee $\mathbb{N}$ to be U-recognizable are given in [20, 17]. However, the general case remains open, see [17, Section 8.2].

Remark 6. The assumptions (H2) and (H3) imply that $\lim_{i\to+\infty}(U_{i+1}-U_i)=+\infty$. However, in many cases, even if $\lim_{i\to+\infty}(U_{i+1}-U_i)=+\infty$, the gap sequence may decrease from time to time. So, even a stronger assumption than (H2) does not imply (H3). The main reason why we introduce (H3) is the following one. Let $10^\ell w$ be a greedy U-representation for some $\ell\geq 0$. Assume (H3) and $i=|w|+\ell\geq G$. Then for all $\ell'\geq\ell$, $10^{\ell'}w$ is a greedy U-representation as well. Indeed, if n is a non-negative integer such that $U_i+n<U_{i+1}$, then

$U_{i+1} + n = U_{i+1} - U_i + U_i + n \le U_{i+2} - U_{i+1} + U_i + n < U_{i+2}$. Hence $U_{i'} + n < U_{i'+1}$ for all $i' \ge i$, meaning that as soon as the greediness property is fulfilled, one can shift the leading 1 at every larger index. This is not always the case, as seen in Example 15. This property will be used in Lemma 14, which in turn will be crucial in the proofs of Propositions 27 as well as Theorem 36, where we construct U-representations with leading 1's in convenient positions.

Remark 7. If $(U_i)_{i\ge 0}$ is a linear recurrence sequence, so are the first and second differences $(V_i)_{i\ge 0} := (U_{i+1} - U_i)_{i\ge 0}$ and $(W_i)_{i\ge 0} := (V_{i+1} - V_i)_{i\ge 0}$. (H3) can be restated as follows. There exists G such that $W_i \ge 0$ for all $i \ge G$. It relates to the *Ultimate Positivity Problem*: given a linear recurrence sequence $(W_i)_{i\ge 0}$, are all but finitely many terms of $(W)_{i\ge 0}$ non-negative? This problem is known to be decidable for integer linear recurrence sequences of order at most 5 in polynomial time [26]. It is also decidable whenever the characteristic polynomial only has simple roots [27]. However, in a general setting, it remains a longstanding open problem [31].

Remark 8. The following deep result due independently to Evertse and to van der Poorten and Schlickewei is discussed in [1], see the terminology and the references therein: For any non-degenerate algebraic linear recurrence sequence $(V_i)_{i\ge 0}$ of dominant modulus $\rho > 1$, and any $\varepsilon > 0$, there exists a constant H such that, for all $i \ge H$, we have $|V_i| \ge \rho^{(1-\varepsilon)i}$. As noticed in [27], any degenerate linear recurrence sequence can be effectively decomposed into a finite number of non-degenerate linear recurrence sequences. In our setting of numeration systems, the sequence $(U_i)_{i\ge 0}$ is increasing so the gap sequence $(V_i)_{i\ge 0} := (U_{i+1} - U_i)_{i\ge 0}$ is positive and (H2) is thus satisfied whenever the associated dominant modulus is larger than 1. The presence of a root of modulus larger than 1 can be tested with the Lehmer–Schur algorithm, see [25, Chap. 10].

Example 9. Our toy example that will be treated all along the paper is given by the recurrence $U_{i+3} = 12\,U_{i+2} + 6\,U_{i+1} + 12\,U_i$. Even though the system is associated with a Pisot number, it is still interesting because $\mathcal{N}_U(m)$ does not tend to infinity (so we cannot follow the decision procedure from [2]) and the gcd of the coefficients of the recurrence is larger than 1. Let $r \ge 1$. If the modulus is a power of 2 or 3, then $U_i \equiv 0 \pmod{2^r}$ (resp. $U_i \equiv 0 \pmod{3^r}$) for large enough i. By taking the initial conditions $1, 13, 163$, the language of greedy U-representations is regular. For the reader aware of β-numeration systems, let us mention that this choice of initial conditions corresponds to the Bertrand initial conditions, in which case the language $\mathrm{rep}_U(\mathbb{N})$ is equal to the set of factors (with no leading zeroes) occurring in the β-expansions of real numbers where β is the dominant root of the characteristic polynomial $X^3 - 12X^2 - 6X - 12$ of the recurrence relation of the system U [4].

3. Some classical lemmas

A set $X \subseteq \mathbb{N}$ is *ultimately periodic* if its characteristic sequence $\mathbf{1}_X \in \{0,1\}^{\mathbb{N}}$ is of the form uv^ω where u, v are two finite words over $\{0,1\}$. It is assumed that u, v are chosen of minimal length. Hence the *period* of X denoted by π_X is the length $|v|$ and its preperiod is the length $|u|$. We say that X is *(purely) periodic* whenever the preperiod is zero. The following lemma is a simple consequence of the minimality of the period chosen to represent an ultimately periodic set.

Lemma 10. *Let $X \subseteq \mathbb{N}$ be an ultimately periodic set of period π_X and let i, j be integers greater than or equal to the preperiod of X. If $i \not\equiv j \pmod{\pi_X}$ then there exists $r < \pi_X$ such that either $i + r \in X$ and $j + r \notin X$ or, $i + r \notin X$ and $j + r \in X$.*

Our assumption (H2) permits us to extend greedy U-representations with some extra leading digits. See [2, Lemma 7] for a proof.

Lemma 11. *Let U be a numeration system satisfying (H2). For all greedy U-representations w, there exists arbitrarily large r such that the word $10^r w$ is also a greedy U-representation.*

When $\mathbb{N}$ is U-recognizable, using a pumping-like argument, we can give an upper bound on the number of zeroes to be inserted.

Lemma 12. *Let U be a numeration system satisfying (H1) and (H2). Then there is an integer constant $C > 0$ such that if w is a greedy U-representation, then for some $\ell < C$, $10^\ell w$ is also a greedy U-representation.*

Proof. By assumption (H1), there is a DFA, say with C states, accepting the numeration language $\mathrm{rep}_U(\mathbb{N})$. Let w be a greedy U-representation. Then from Lemma 11, there is $r \geq C$ such that $10^r w \in \mathrm{rep}_U(\mathbb{N})$. The path of label $10^r w$ starting from the initial state is accepting. Since $r \geq C$, a state is visited at least twice when reading the block 0^r. Thus there is an accepting path of label $10^\ell w$ with $\ell < C$. $\square$

Let us introduce a constant Z.

Definition 13. Let U be a numeration system satisfying (H1), (H2) and (H3). Thanks to (H2), there exist infinitely many R such that

$$U_{R+1} - U_R \geq U_{i+1} - U_i$$

for all $i \leq R$. We may choose the least R with this property and such that $R \geq G$ where G is the constant given in (H3). We set

$$Z = \max\{R, C\}$$

where C is the constant given in Lemma 12.

In view of Remark 7 about the status of the general decision problem about (H3), we assume that G is given as an input with the numeration system. Hence the constants C, R, Z can be effectively computed. Indeed, C can be deduced from the automaton accepting the language of the numeration. Then R can be computed by an exhaustive search and finally, one has to choose $Z = \max\{R, C\}$.

Thanks to (H3), we have more flexibility about the inserted zeroes: we can add as many zeroes as needed to greedy representations and obtain again greedy representations.

Lemma 14. *Let U be a numeration system satisfying (H1), (H2) and (H3). If w is a greedy U-representation, then for all $z \geq Z$, $10^z w$ is also a greedy U-representation.*

Proof. Let w be a greedy U-representation. By Lemma 12, there is $\ell < C$ such that $10^\ell w$ is a greedy U-representation. Let $i = \ell + |w|$. Let $n = \mathrm{val}_U(w)$. We have $U_i + n < U_{i+1}$.

- If $i \geq Z$, similarly as in Remark 6, since $Z \geq G$,

$$U_{i+1} + n = U_{i+1} - U_i + U_i + n \leq U_{i+2} - U_{i+1} + U_i + n < U_{i+2}.$$

  Hence $U_j + n < U_{j+1}$ for all $j \geq i$. Otherwise stated, $10^{\ell'} w$ is a greedy U-representation for all $\ell' \geq \ell$. In particular, since $\ell < Z$, for all $z \geq Z$, $10^z w$ is a greedy U-representation.
- If $i < Z$, then

$$\begin{aligned} U_Z + n = U_Z - U_i + U_i + n &< U_Z - U_i + U_{i+1} \\ &\leq U_{Z+1} - U_{i+1} + U_{i+1} \leq U_{Z+1} \end{aligned}$$

  because $Z \geq R \geq G$. Hence $10^{Z-|w|} w$ is a greedy U-representation. We conclude by applying the first part of the proof.

□

Example 15. The sequence $1, 2, 4, 5, 16, 17, 64, 65, \ldots$ is a solution of the linear recurrence $U_{i+4} = 5U_{i+2} - 4U_i$ but it does not satisfy (H3). The property stated in Lemma 14 does not hold: only some shifts to the left of the leading coefficient 1 lead to valid greedy expansions. The word 1001 is the greedy representation of 6 but for all $t \geq 1$, $1(00)^t 1001$ is not a greedy representation.

Example 16. The sequence $1, 2, 3, 4, 8, 12, 16, 32, 48, 64, 128, \dots$ is a solution of the linear recurrence $U_{i+3} = 4U_i$. The numeration language $0^* \operatorname{rep}_U(\mathbb{N})$ is the set of suffixes of $\{000, 001, 010, 100\}^*$, hence (H1) holds. For all $i \geq 0$, $U_{i+1} - U_i = 4^{\lfloor i/3 \rfloor}$. Therefore, (H2) and (H3) are also verified.

We will also make use of the following folklore result. See, for instance, [3, Prop. 3.1.9]. It relies on the fact that a linear recurrence sequence is ultimately periodic modulo Q.

Proposition 17. *Let $Q, r \geq 0$. Let $A \subseteq \mathbb{N}$ be a finite alphabet. If U is a linear numeration system, then*

$$\{w \in A^* \mid \operatorname{val}_U(w) \in Q\,\mathbb{N} + r\}$$

is accepted by a DFA that can be effectively constructed. In particular, whenever $\mathbb{N}$ is U-recognizable, i.e. under (H1), then any ultimately periodic set is U-recognizable.

Under assumption (H1) the formal series $\sum_{i \geq 0} U_i\, X^i$ is $\mathbb{N}$-rational because U_i is the number of words of length less than or equal to i in the regular language $\operatorname{rep}_U(\mathbb{N})$. One can therefore make use of Soittola's theorem [31, Thm. 10.2]: The series is the merge of rational series with dominating eigenvalues and polynomials. We thus define the following quantities.

Definition 18. We introduce an integer u and a real number β depending only on the numeration system. From Soittola's theorem, there exist an integer $u \geq 1$, real numbers $\beta_0, \dots, \beta_{u-1} \geq 1$ and non-zero polynomials $P_0, \dots, P_{u-1}$ such that for $r \in \{0, \dots, u-1\}$ and large enough i, say $i \geq I_1$,

$$U_{ui+r} = P_r(i)\, \beta_r^i + Q_r(i) \tag{3.1}$$

where $\frac{Q_r(i)}{\beta_r^i} \to 0$ when $i \to \infty$. Since $(U_i)_{i \geq 0}$ is increasing, for $r < s < u$, for all $i \geq I_1$, we have

$$U_{ui+r} < U_{ui+s} < U_{u(i+1)+r}.$$

By letting i tend to infinity, this shows that we must have $\beta_0 = \dots = \beta_{u-1}$ which we denote by β and $\deg(P_0) = \dots = \deg(P_{u-1})$ which we denote by d. Otherwise stated, $U_{ui+r} \sim c_r i^d \beta^i$ for some constant c_r. Finally, let T be such that $c_T = \max_{0 \leq r < u} c_r$. Otherwise stated, we highlight with T a subsequence $(U_{ui+T})_{i \geq 0}$ with the maximal dominant coefficient. Since $(U_i)_{i \in \mathbb{N}}$ is increasing and $\frac{Q_T(i)}{\beta^i} \to 0$ when $i \to +\infty$, there is $I_2 > 0$ such that $P_T(i) > 0$, for all $i \geq I_2$. Moreover, there is $I_3 > 0$ such that P_T is non-decreasing "after I_3", i.e. $P_T(n) \leq P_T(n+1)$ for all $n \geq I_3$. Finally, let I be the positive integer $\max\{I_1, I_2, I_3\}$.

Note that if a numeration system has a dominant root, i.e. the minimal recurrence relation satisfied by $(U_i)_{i\geq 0}$ has a unique root $\beta > 1$, possibly with multiplicity greater than 1, of maximum modulus, then $u = 1$.

Lemma 19. *With the notation of Definition 18, if $\beta > 1$ then there exist non-negative constants K and L such that for all n,*

$$|\mathrm{rep}_U(n)| < u \log_\beta(n) + K$$

and

$$|\mathrm{rep}_U(n)| > u \log_\beta(n) - u \log_\beta(P_T(\log_\beta(n) + K/u)) - L.$$

This lemma shows that the length of the greedy U-representation of n grows at most like $\log_{\beta^{1/u}}(n)$. If P_T is a constant polynomial, the lower bound is of the form $u \log_\beta(n) - L'$ for some non-negative constant L'. From this result, we may express the weaker information (on ratios instead of differences) that $|\mathrm{rep}_U(n)| \sim u \log_\beta(n)$. The intricate form of the lower bound can be seen on an example such as $(U_i)_{i\geq 0} = (i^d\, 2^i)_{i\geq 0}$. In such a case, we get $\log_2(n) < |\mathrm{rep}_U(n)| + d\, \log_2(|\mathrm{rep}_U(n)|)$. Hence a lower bound for $|\mathrm{rep}_U(n)|$ is less than $\log_2(n)$.

Proof. We have $|\mathrm{rep}_U(n)| = \ell$ if and only if $U_{\ell-1} \leq n < U_\ell$. We make use of Definition 18 for u, β, T and I. Let $j = \lfloor \frac{\ell-1-T}{u} \rfloor$.

Suppose that n is large enough so that $j \geq I$. Since U is increasing and $j \geq I$, (3.1) gives

$$U_{\ell-1} \geq U_{uj+T} = P_T(j)\beta^j + Q_T(j).$$

We get

$$\log_\beta(n) \geq \log_\beta(U_{\ell-1}) \geq j + \log_\beta(P_T(j)) + \log_\beta\left(1 + \frac{Q_T(j)}{P_T(j)\beta^j}\right).$$

Note that, for large enough n, we can suppose that $1 + \frac{Q_T(j)}{P_T(j)\beta^j} > 0$ (since $\frac{Q_T(i)}{\beta^i} \to 0$ when $i \to +\infty$ and P_T is non-decreasing after I), so that the last logarithm in the above inequality is well defined.
Hence

$$j \leq \log_\beta(n) - \log_\beta(P_T(j)) - \log_\beta\left(1 + \frac{Q_T(j)}{P_T(j)\beta^j}\right).$$

Moreover, $j > \frac{\ell-1-T}{u} - 1 \geq \frac{\ell-u}{u} - 1 \geq \frac{\ell}{u} - 2$. We obtain

$$\ell < u(j+2) \leq u \log_\beta(n) + 2u - u \log_\beta(P_T(j)) - u \log_\beta\left(1 + \frac{Q_T(j)}{P_T(j)\beta^j}\right).$$

Since $j \geq I$ and P_T is non-decreasing after I, we get

$$\ell < u(j+2) \leq u \log_\beta(n) + 2u - u \log_\beta(P_T(I)) - u \log_\beta \left(1 + \frac{Q_T(j)}{P_T(j)\beta^j}\right).$$

Finally, since $\frac{Q_T(i)}{\beta^i} \to 0$ when $i \to +\infty$, there is a constant $K \geq 0$ such that

$$\ell < u(j+2) \leq u \log_\beta(n) + K.$$

We have supposed n to be large enough so that $j \geq I$ and $1 + \frac{Q_T(j)}{P_T(j)\beta^j} > 0$. There is only a finite number of integers not fulfilling these conditions. Hence, possibly increasing the value of the constant K, we can assume that the above inequality holds for all integers n.

We proceed similarly to get a lower bound for ℓ. Let $k = \lfloor \frac{\ell - T}{u} \rfloor$. Observe that $j \leq k$, hence $k \geq I$. Since U is increasing, we have

$$U_\ell < U_{u(k+1)+T} = P_T(k+1)\beta^{k+1} + Q_T(k+1).$$

We obtain

$$\log_\beta(n) < \log_\beta(U_\ell) < k+1+\log_\beta(P_T(k+1))+\log_\beta \left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right).$$

As in the first part of the proof, we can suppose that n is large enough to get $1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}} > 0$.
Observe that $k \leq j+1$. Hence, from the first part, we get

$$k + 1 \leq j + 2 \leq \log_\beta(n) + \frac{K}{u}.$$

Since $k \leq \frac{\ell - T}{u} \leq \frac{\ell}{u}$, we have

$$\ell \geq uk > u \log_\beta(n) - u - u \log_\beta(P_T(k+1)) - u \log_\beta \left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right).$$

We have $k+1 > k \geq I$ and recall that P_T is non-decreasing after I, hence

$$P_T(k+1) \leq P_T\left(\log_\beta(n) + \frac{K}{u}\right).$$

Hence

$$\ell > u \log_\beta(n) - u - u \log_\beta \left(P_T\left(\log_\beta(n) + \frac{K}{u}\right)\right) - u \log_\beta \left(1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}}\right).$$

Furthermore, since $\frac{Q_T(i)}{\beta^i} \to 0$ when $i \to +\infty$ and P_T is non-decreasing after I, there is a constant $L \geq 0$ such that

$$\ell > u \log_\beta(n) - u \log_\beta \left(P_T \left(\log_\beta(n) + \frac{K}{u} \right) \right) - L.$$

As in the first part of the proof, we only considered those n such that $j \geq I$ and $1 + \frac{Q_T(k+1)}{P_T(k+1)\beta^{k+1}} > 0$. Possibly increasing the value of L, we can assume that the above inequality is satisfied for all integers n. $\square$

Example 20. Consider the sequence $1, 2, 6, 12, 36, 72, \ldots$ defined by $U_0 = 1$, $U_{2i+1} = 2U_{2i}$ and $U_{2i+2} = 3U_{2i+1}$. Then for all $i \geq 0$, $U_{i+2} = 6U_i$. It is easily seen that $U_{2i} = 6^i$ and $U_{2i+1} = 2 \cdot 6^i$. With the notation of Definition 18, $u = 2$, $\beta = 6$, $d = 0$ and $P_T = c_T = 2$. The language $0^* \operatorname{rep}_U(\mathbb{N})$ is made of words where in even (resp. odd) positions digits belong to $0, 1$ (resp. $0, 1, 2$), i.e.

$$0^* \operatorname{rep}_U(\mathbb{N}) = (\varepsilon + 0 + 1)((0 + 1 + 2)(0 + 1))^*.$$

If $|\operatorname{rep}_U(n)| = 2\ell + 1$ then $U_{2\ell} = 6^\ell \leq n < U_{2\ell+1} = 2 \cdot 6^\ell$, so $|\operatorname{rep}_U(n)| \leq 2\log_6(n) + 1$ and $|\operatorname{rep}_U(n)| > 2\log_6(\frac{n}{2}) + 1 = 2\log_6(n) - 2\log_6(2) + 1$. If $|\operatorname{rep}_U(n)| = 2\ell$ then $U_{2\ell-1} = 2 \cdot 6^{\ell-1} \leq n < U_{2\ell} = 6^\ell$, so $|\operatorname{rep}_U(n)| \leq 2\log_6(3n) = 2\log_6(n) + 2\log_6(3)$ and $|\operatorname{rep}_U(n)| > 2\log_6(n)$.

Example 21. Consider the sequence $1, 3, 8, 20, 48, 112, \ldots$ defined by $U_0 = 1$, $U_1 = 3$ and $U_{i+2} = 4U_{i+1} - 4U_i$. Then $U_i = (\frac{i}{2} + 1)2^i$. With the notation of Definition 18, $u = 1$, $\beta = 2$, $d = 1$ and $P_T(n) = \frac{n}{2} + 1$. If $|\operatorname{rep}_U(n)| = \ell$ then $U_{\ell-1} = (\frac{\ell-1}{2} + 1)2^{\ell-1} \leq n < U_\ell = (\frac{\ell}{2} + 1)2^\ell$, so $|\operatorname{rep}_U(n)| < \log_2(n) + 1$ and $|\operatorname{rep}_U(n)| > \log_2(n) - \log_2(\frac{\ell}{2} + 1) > \log_2(n) - \log_2(\frac{1}{2}\log_2(n) + \frac{3}{2})$. With the notation of Lemma 19, $K = 1$ and $P_T(\log_2(n) + K) = \frac{1}{2}\log_2(n) + \frac{3}{2}$.

As shown by the next result. It is enough to obtain a bound on the possible periods of X. In [2, Prop. 44], the result is given in a more general setting (i.e. for abstract numeration systems) and we restate it in our context.

Proposition 22. *Let U be a numeration system satisfying (H1), let $X \subseteq \mathbb{N}$ be an ultimately periodic set and let $\mathcal{A}_X$ be a DFA accepting* $\operatorname{rep}_U(X)$. *Then the preperiod of X is bounded by a computable constant depending only on the size of $\mathcal{A}_X$ and the period π_X of X.*

Thus, our aim is to bound the period π_X only in terms of the given automaton recognizing X.

4. Number of states

We follow Honkala's strategy introduced in [16]. A DFA $\mathcal{A}$ accepting $\mathrm{rep}_U(X)$ is given as input. Assuming that X is ultimately periodic, the number of states of $\mathcal{A}$ should provide an upper bound on the possible period and preperiod of X. Roughly speaking, the minimal preperiod/period should not be too large compared with the size of $\mathcal{A}$. This should leave us with a finite number of candidates to test. Thanks to Proposition 17, one therefore builds a DFA for each pair of admissible preperiod/period. Equality of regular languages being decidable, we compare the language accepted by this DFA and the one accepted by $\mathcal{A}$. If an agreement is found, then X is ultimately periodic, otherwise it is not. As a consequence of Proposition 22, we only focus on the admissible periods.

For an ultimately periodic set $X \subseteq \mathbb{N}$, we consider the prime decomposition of its period π_X. There are two types of prime factors.

(T1) Those that do not simultaneously divide all the coefficients of the recurrence relation.
(T2) The primes dividing all the coefficients of the recurrence relation.

Our strategy is to bound those two types of factors separately. We depart from the strategy developed in [2] because we have to deal with the case of what we call a zero period discussed below.

4.1. Prime factors of the period that do not divide all the coefficients of the recurrence relation.

If a prime factor p of the candidate period for X does not divide all the coefficients of the recurrence relation, we will show that, for some integer $\mu \geq 1$, the periodic part of the sequence $(U_i \bmod p^\mu)_{i\geq 0}$ contains a non-zero element. This fact will provide us with an upper bound on p and its exponent in the prime decomposition of the candidate period.

Definition 23. We say that an ultimately periodic sequence has a *zero period* (or, zero periodic part) if it has period 1 and the repeated element is 0. Otherwise stated, the sequence has a tail of zeroes.

Remark 24. Let $\mu \geq 1$. Observe that if the periodic part of $(U_i \bmod p^\mu)_{i\geq 0}$ contains a non-zero element, then the same property holds for all sequences $(U_i \bmod p^{\mu'})_{i\geq 0}$ with $\mu' \geq \mu$.

Furthermore, assume that for infinitely many μ, $(U_i \bmod p^\mu)_{i\geq 0}$ has a zero period. Then from the previous paragraph, we conclude that $(U_i \bmod p^\mu)_{i\geq 0}$ has a zero period for all $\mu \geq 1$.

Example 25. We give a sequence where only finitely many sequences modulo p^μ have a zero period. Take the sequence $U_0 = 1$, $U_1 = 4$, $U_2 = 8$ and $U_{i+2} = U_{i+1} + U_i$ for $i \geq 1$. Then the sequence $(U_i \bmod 2^\mu)_{i\geq 0}$ has a zero period for $\mu = 1, 2$ because of the particular initial conditions. But it is easily checked that it has a non-zero period for all $\mu \geq 3$.

The next result is a special instance of [2, Thm. 32] and its proof turns out to be much simpler.

Theorem 26. *Let p be a prime. The sequence $(U_i \bmod p^\mu)_{i\geq 0}$ has a zero period for all $\mu \geq 1$ if and only if all the coefficients $a_0, \ldots, a_{k-1}$ of the linear relation* (1.1) *are divisible by p.*

Proof. Let N be given in Definition 3. It is clear that if $a_0, \ldots, a_{k-1}$ are divisible by p, then for any choice of initial conditions $U_0, \ldots, U_{N+k-1}$, the elements $U_{N+k}, \ldots, U_{N+2k-1}$ are divisible by p, hence the elements $U_{N+2k}, \ldots, U_{N+3k-1}$ are divisible by p^2, and so on and so forth. Otherwise stated, for all $\mu \geq 1$ and all $i \geq N + \mu k$, U_i is divisible by p^μ.

We turn to the converse. Since the sequence $(U_i)_{i\geq 0}$ ultimately satisfies a linear recurrence relation, the power series

$$\mathsf{U}(x) := \sum_{i\geq 0} U_i\, x^i$$

is rational. By assumption, $(U_i \bmod p^\mu)_{i\geq 0}$ has a zero period for all $\mu \geq 1$. Otherwise stated, with the p-adic absolute value notation, $|U_i|_p \leq p^{-\mu}$ for large enough i, i.e. $|U_i|_p \to 0$ as $i \to +\infty$. Recall that a series $\sum_{i\geq 0} \gamma_i$ converges in $\mathbb{Q}_p$ if and only if $\lim_{i\to+\infty} |\gamma_i|_p = 0$. Hence the series $\mathsf{U}(x)$ converges in $\mathbb{Q}_p$ in the closed unit disc. Therefore, the poles $\rho_1, \ldots, \rho_r \in \mathbb{C}_p$ of $\mathsf{U}(x)$ must satisfy $|\rho_j|_p > 1$ for $1 \leq j \leq r$.

Let $P(x) = 1 - a_{k-1}x - \ldots - a_0 x^k$ be the reciprocal polynomial of the linear recurrence relation (1.1). By minimality of the order k of the recurrence, the roots of P are precisely the poles of $\mathsf{U}(x)$ with the same multiplicities. If we factor

$$P(x) = (1 - \delta_1 x) \cdots (1 - \delta_k x)$$

each of the δ_j is one of the $\frac{1}{\rho_1}, \ldots, \frac{1}{\rho_r}$. For $n > 0$, the coefficient of x^n in $P(x)$ is an integer equal to a sum of product of elements of p-adic absolute value less than 1. Since $|a + b|_p \leq \max\{|a|_p, |b|_p\}$ and $|ab|_p = |a|_p |b|_p$, this coefficient is an integer with a p-adic absolute value less than 1, i.e. a multiple of p. □

Thanks to Theorem 26, if p is a prime not dividing all the coefficients of the recurrence relation (1.1) then there exists an integer $\lambda \geq 1$ such that the periodic part of $(U_i \bmod p^\lambda)_{i\geq 0}$ contains a non-zero element.

Proposition 27. *Assume (H1), (H2) and (H3). Let p be a prime not dividing all the coefficients of the recurrence relation* (1.1) *and let $\lambda \geq 1$ be an integer such that the periodic part of $(U_i \bmod p^\lambda)_{i\geq 0}$ contains a non-zero element. If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set with period $\pi_X = p^\mu \cdot r$ where $\mu \geq \lambda$ and r is not divisible by p, then the minimal automaton of $\operatorname{rep}_U(X)$ has at least $p^{\mu-\lambda+1}$ states.*

Proof. We will make use of the following observation. Let $n \geq 1$. In the additive group $(\mathbb{Z}/p^n\mathbb{Z}, +)$, an element a has order p^s with $0 < s \leq n$ if and only if $a = p^{n-s} \cdot m$ where m is not divisible by p.

By assumption the periodic part of $(U_i \bmod p^\lambda)_{i\geq 0}$ contains a non-zero element R of order $\operatorname{ord}_{p^\lambda}(R) = p^\theta$ for some θ such that $0 < \theta \leq \lambda$. Using the above observation twice, $R = p^{\lambda-\theta} \cdot m$ for some m coprime with p, the order of R modulo p^μ is $s := \operatorname{ord}_{p^\mu}(R) = p^{\mu-\lambda+\theta}$.

Let us define s integers $k_1, \ldots, k_s \geq 0$ and thus s words $w_1, \ldots, w_s \in \{0,1\}^*$ of the following form

$$w_j := 10^{k_j} 10^{k_{j-1}} \cdots 10^{k_1} 0^{|\operatorname{rep}_U(\pi_X)|}.$$

Thanks to Lemma 14, we may impose the following conditions.

- First, k_1 is taken large enough to ensure that $\operatorname{val}_U(w_1)$ is larger than the preperiod of X.
- Second, $k_1, \ldots, k_s$ are taken large enough to ensure that $w_j \in \operatorname{rep}_U(\mathbb{N})$ for all j. Simply choose $k_j \geq Z$ for all j.
- Third, we can choose $k_1, \ldots, k_s$ so that the 1's occur at indices m such that $U_m \equiv R \pmod{p^\mu}$.

Observe that $\operatorname{val}_U(w_j) \equiv j \cdot R \pmod{p^\mu}$. Since p^μ divides π_X, the words $w_1, \ldots, w_s$ have pairwise distinct values modulo π_X.

Let $i, j \in \{1, \ldots, s\}$ such that $i \neq j$. By Lemma 10, we can assume that there exists $r_{i,j} < \pi_X$ such that $\operatorname{val}_U(w_i) + r_{i,j} \in X$ and $\operatorname{val}_U(w_j) + r_{i,j} \notin X$ (the symmetric situation is handled similarly). In particular, $|\operatorname{rep}_U(r_{i,j})| \leq |\operatorname{rep}_U(\pi_X)|$. Consider the two words

$$10^{k_i} 10^{k_{i-1}} \cdots 10^{k_1} x_{i,j} \quad \text{and} \quad 10^{k_j} 10^{k_{j-1}} \cdots 10^{k_1} x_{i,j}$$

where

$$x_{i,j} = 0^{|\operatorname{rep}_U(\pi_X)| - |\operatorname{rep}_U(r_{i,j})|} \operatorname{rep}_U(r_{i,j}).$$

The first word belongs to $\operatorname{rep}_U(X)$ and the second does not. Consequently, the number of states of the minimal automaton of $\operatorname{rep}_U(X)$ is at least $s = p^{\mu-\lambda+\theta}$. The conclusion follows since $\theta \geq 1$. □

From the above proposition, we immediately get the following.

Corollary 28. *Assume (H1), (H2) and (H3). Let $p > \max\{|a_0|, U_N\}$ be a prime. If $X \subseteq \mathbb{N}$ is an ultimately periodic U-recognizable set with* *period $\pi_X = p^\mu \cdot r$ where $\mu \geq 1$ and r is not divisible by p, then the minimal automaton of $\mathrm{rep}_U(X)$ has at least p^μ states.*

Proof. The sequence $(U_i)_{i\geq 0}$ satisfies the recurrence relation (1.1) for all $i \geq N$. Since $p > |a_0|$, p does not divide a_0 and $(U_{N+i} \bmod p)_{i\geq 0}$ is purely periodic. By assumption $p > U_N$, hence the first element of the periodic part equals U_N and is non-zero modulo p. We conclude that the non-zero element U_N occurs infinitely often in the sequence $(U_i \bmod p)_{i\geq 0}$. Hence we may apply Proposition 27 with $\lambda = 1$. □

4.2. **Prime factors of the period that divide all the coefficients of the recurrence relation.** We can factor the period π_X as

$$\pi_X = Q_X \cdot p_1^{\mu_{X,1}} \cdots p_t^{\mu_{X,t}} \tag{4.1}$$

where every p_j divides all the coefficients of the recurrence relation (1.1) and, for every prime factor q of Q_X, at least one of the coefficients of the recurrence relation (1.1) is not divisible by q. Otherwise stated, the factor Q_X collects the prime factor of type (T1). Note that the primes p_j depend only on the numeration system U (i.e. the coefficients of the recurrence) and their exponents depend on π_X thus, on X.

Remark 29. There is a finite number of primes dividing all the coefficients of the recurrence relation. Thus, we only have to obtain an upper bound on the corresponding exponents $\mu_{X,1}, \dots, \mu_{X,t}$ that may appear in (4.1).

Definition 30. Let $j \in \{1, \dots, t\}$ and $\mu \geq 1$. From Theorem 26, the sequence $(U_i \bmod p_j^\mu)_{i\geq 0}$ has a zero period. We let $\mathsf{f}_{p_j}(\mu)$ denote the length of the preperiod, i.e. $U_{\mathsf{f}_{p_j}(\mu)-1} \not\equiv 0 \pmod{p_j^\mu}$ and $U_i \equiv 0 \pmod{p_j^\mu}$ for all $i \geq \mathsf{f}_{p_j}(\mu)$.

Example 31. Let us consider the numeration system from Example 4. The sequence $(U_i \bmod 2)_{i\geq 0}$ is $1, 1, 1, 1, 0^\omega$. Hence $\mathsf{f}_2(1) = 4$. The sequence $(U_i \bmod 4)_{i\geq 0}$ is $1, 3, 1, 3, 2, 0, 2, 2, 0^\omega$. Hence $\mathsf{f}_2(2) = 8$. Continuing this way, we have $\mathsf{f}_2(3) = 12$ and $\mathsf{f}_2(4) = 16$.

Note that f_{p_j} is non-decreasing: $\mathsf{f}_{p_j}(\mu + 1) \geq \mathsf{f}_{p_j}(\mu)$ and

$$\lim_{\mu\to+\infty} \mathsf{f}_{p_j}(\mu) = +\infty. \tag{4.2}$$

Lemma 32. [16, Lemma 6] *Let X be an ultimately periodic set with period (4.1). There exists $r \in \{0, \dots, Q_X - 1\}$ such that $X \cap (Q_X\mathbb{N} + r)$ is ultimately periodic of period $Q_X \cdot p_1^{\nu_1} \cdots p_t^{\nu_t}$ with*

$$\max_{1\leq j\leq t} \mu_{X,j} = \max_{1\leq j\leq t} \nu_j.$$

Definition 33. The quantity r in the previous lemma is not necessarily unique. To avoid ambiguity, we always consider the smallest possible such r denoted by r_X and the associated exponents $\nu_{X,1},\dots,\nu_{X,t}$. We therefore let ρ_X denote the corresponding quantity $Q_X \cdot p_1^{\nu_{X,1}} \cdots p_t^{\nu_{X,t}}$.

We let $M_{\boldsymbol{\mu},X}$ denote the maximum of the values $\mathsf{f}_{p_j}(\nu_{X,j})$ for $j \in \{1,\dots,t\}$:

$$M_{\boldsymbol{\mu},X} = \max_{1\le j\le t} \mathsf{f}_{p_j}(\nu_{X,j}).$$

Thus, $M_{\boldsymbol{\mu},X}$ is the least index such that for all $i \ge M_{\boldsymbol{\mu},X}$ and all $j \in \{1,\dots,t\}$, $U_i \equiv 0 \pmod{p_j^{\nu_{X,j}}}$. By the Chinese remainder theorem, $M_{\boldsymbol{\mu},X}$ is also the least index such that for all $i \ge M_{\boldsymbol{\mu},X}$,

$$U_i \equiv 0 \pmod{\frac{\rho_X}{Q_X}}.$$

The reader may notice that $M_{\boldsymbol{\mu},X}$ only depends on the exponents $\boldsymbol{\mu} = (\mu_{X,1},\dots,\mu_{X,t})$ occurring in (4.1).

From Lemma 32 and (4.2), for each $j \in \{1,\dots,t\}$, $\lim_{\mu_{X,j}\to+\infty} M_{\boldsymbol{\mu},X} = +\infty$.

Example 34. Let us consider the numeration system from Example 9. Here we have two prime factors 2 and 3 to take into account. Computations show that $\mathsf{f}_2(1) = 3$, $\mathsf{f}_2(2) = 5$, $\mathsf{f}_2(3) = 7$ and $\mathsf{f}_3(1) = 3$, $\mathsf{f}_3(2) = 6$, $\mathsf{f}_3(3) = 9$. Assume that we are interested in a period $\rho_X/Q_X = 72 = 2^3 \cdot 3^2$. With the above definition, $M_{\boldsymbol{\mu},X} = \max\{\mathsf{f}_2(3), \mathsf{f}_3(2)\} = 7$. One can check that $(U_i \bmod 72)_{i\ge 0}$ is $1, 13, 19, 30, 54, 48, 36, 0^\omega$.

We introduce a quantity γ_{Q_X} which only depends on the numeration system U and the number Q_X defined in (4.1). Since we are only interested in decidability issues, there is no need to find a sharp estimate on this quantity.

Definition 35. Let $Q \ge 1$ be an integer. Under (H1), for each $r \in \{0,\dots,Q-1\}$, a DFA accepting the language $\mathrm{rep}_U(Q\,\mathbb{N} + r)$ can be effectively built (see Proposition 17 or the construction in [3, Prop. 3.1.9]). We let γ_Q denote the maximum of the numbers of states of these DFAs for $r \in \{0,\dots,Q-1\}$.

The crucial point in the next statement is that the most significant digit 1 occurs for $U_{M_{\boldsymbol{\mu},X}-1}$ in a specific word. The proof makes use of the same kind of arguments built for definite languages as in [18, Lemma 2.1].

Theorem 36. *Assume (H1), (H2) and (H3). Let $X \subseteq \mathbb{N}$ be an ultimately periodic U-recognizable set with period π_X factored as in* (4.1). *Assume that $M_{\boldsymbol{\mu},X} - 1 - |\,\mathrm{rep}_U(\rho_X - 1)| \ge Z$, where Z is the constant* *given in Definition 13 and* $M_{\boldsymbol{\mu},X}$ *and* ρ_X *are given in Definition 33. Also assume that* $M_{\boldsymbol{\mu},X}$ *is greater than the preperiod of* $(U_i \bmod Q_X)_{i\in\mathbb{N}}$*. Then the minimal automaton of* $0^* \operatorname{rep}_U(X)$ *has at least* $\frac{|\operatorname{rep}_U(\rho_X-1)|+1}{\gamma_{Q_X}}$ *states.*

This result will provide us with an upper bound on $\mu_{X,1},\ldots,\mu_{X,t}$ (details are given in Section 5.2). If $\max_j \mu_{X,j} = \max_j \nu_{X,j} \to \infty$, then $\rho_X \to \infty$ and since Q_X has been bounded in the first part of this paper, the number of states of the minimal automaton of $\operatorname{rep}_U(X)$ should increase.

Proof. We may apply Lemma 14: if w is a greedy U-representation, then, for all $z \geq Z$, $10^z w$ also belongs to $\operatorname{rep}_U(\mathbb{N})$. Let r_X be the quantity given in Definition 33. The set $X \cap (Q_X\mathbb{N} + r_X)$ has period ρ_X. Let $\mathcal{B}_X$ be the minimal automaton of $0^* \operatorname{rep}_U(X \cap (Q_X\mathbb{N}+r_X))$. We will provide a lower bound on the number of states of this automaton. Let g be large enough so that

- $g \geq Z$
- $U_{M_{\boldsymbol{\mu},X}+g}$ is larger than the preperiod of $X \cap (Q_X\mathbb{N} + r_X)$
- $g + 1$ is a multiple of the period of $(U_i \bmod Q_X)_{i\in\mathbb{N}}$.

Consider

$$n_1 = \operatorname{val}_U((10^g)^{Q_X} 10^{M_{\boldsymbol{\mu},X}-1}) = \sum_{i=0}^{Q_X} U_{M_{\boldsymbol{\mu},X}-1+i(g+1)}$$
$$n_2 = \operatorname{val}_U(10^{M_{\boldsymbol{\mu},X}+g}) = U_{M_{\boldsymbol{\mu},X}+g}.$$

Observe that n_1 and n_2 are both congruent to $U_{M_{\boldsymbol{\mu},X}-1}$ modulo Q_X (we make use of the assumption that $M_{\boldsymbol{\mu},X}$ is greater than the preperiod of $(U_i \bmod Q_X)_{i\in\mathbb{N}}$). However, by definition of $M_{\boldsymbol{\mu},X}$,

$$n_1 \bmod \frac{\rho_X}{Q_X} = U_{M_{\boldsymbol{\mu},X}-1} \bmod \frac{\rho_X}{Q_X} \neq 0$$

but n_2 is congruent to 0 modulo $\frac{\rho_X}{Q_X}$. Consequently, n_1 and n_2 are not congruent modulo ρ_X. By Lemma 10 applied to the set $X\cap(Q_X\mathbb{N}+r_X)$, we may suppose that there exists $s < \rho_X$ such that

$$n_1 + s \in X \cap (Q_X\mathbb{N} + r_X) \quad \text{and} \quad n_2 + s \notin X \cap (Q_X\mathbb{N} + r_X)$$

(the symmetrical situation can be treated in the same way). By assumption, $M_{\boldsymbol{\mu},X} - 1 - |\operatorname{rep}_U(s)| \geq M_{\boldsymbol{\mu},X} - 1 - |\operatorname{rep}_U(\rho_X-1)| \geq Z$. Thanks to Lemma 14, both words

$$u = (10^g)^{Q_X} 10^{M_{\boldsymbol{\mu},X}-1-|\operatorname{rep}_U(s)|} \operatorname{rep}_U(s)$$

and

$$v = 10^g 00^{M_{\boldsymbol{\mu},X}-1-|\operatorname{rep}_U(s)|} \operatorname{rep}_U(s)$$

are greedy U-representations. For all $\ell \geq 0$, define an equivalence relation E_ℓ on the set of states of $\mathcal{B}_X$:

$$E_\ell(q,q') \Leftrightarrow (\forall x \in A_U^*)\big[|x| \geq \ell \Rightarrow (\delta(q,x) \in \mathcal{F} \Leftrightarrow \delta(q',x) \in \mathcal{F})\big]$$

where δ (resp. $\mathcal{F}$) is the transition function (resp. the set of final states) of $\mathcal{B}_X$. Let us denote the number of equivalence classes of E_ℓ by P_ℓ. Clearly, $E_\ell(q,q')$ implies $E_{\ell+1}(q,q')$, and thus $P_\ell \geq P_{\ell+1}$. One can already observe that P_0 is the number of states of $\mathcal{B}_X$.

Let $i \in \{0,\ldots,|\operatorname{rep}_U(\rho_X-1)|\}$. By assumption, $|\operatorname{rep}_U(\rho_X-1)| < M_{\boldsymbol{\mu},X}$. Since u and v have the same suffix of length $M_{\boldsymbol{\mu},X}-1$, we can factorize these words as

$$u = u_i w_i \quad \text{and} \quad v = v_i w_i$$

where $|w_i| = i$. Let q_0 be the initial state of $\mathcal{B}_X$. By construction, $\delta(q_0, u_i w_i) \in \mathcal{F}$ whereas $\delta(q_0, v_i w_i) \notin \mathcal{F}$, hence the states $\delta(q_0,u_i)$ and $\delta(q_0,v_i)$ are not in relation with respect to E_i. Let us show that, for all $j > i$, they satisfy E_j. It is enough to show that

$$E_{i+1}(\delta(q_0,u_i),\delta(q_0,v_i)). \tag{4.3}$$

Figures 1 and 2 can help the reader. Let x be such that $|x| = i+t$, with $t \geq 1$. Let p be the prefix of $\operatorname{rep}_U(s)$ of length $|\operatorname{rep}_U(s)| - i$, this prefix p being empty whenever this difference is negative. If we replace w_i by x in u and v, we get

$$u_i x = (10^g)^{Q_X} 10^{M_{\boldsymbol{\mu},X}-1-|px|+t} px \quad \text{and} \quad v_i x = 10^g 00^{M_{\boldsymbol{\mu},X}-1-|px|+t} px.$$

Then

$$\operatorname{val}_U(u_i x) - \operatorname{val}_U(v_i x) = U_{M_{\boldsymbol{\mu},X}+t-1} + \sum_{i=2}^{Q_X} U_{M_{\boldsymbol{\mu},X}-1+i(g+1)+t}.$$

Since by assumption, $M_{\boldsymbol{\mu},X}$ is larger than the preperiod of $(U_i \bmod Q_X)_{i\in\mathbb{N}}$, this quantity is congruent to 0 modulo Q_X and by definition of $M_{\boldsymbol{\mu},X}$, it is also congruent to 0 modulo $\frac{\rho_X}{Q_X}$. Hence, $\operatorname{val}_U(u_i x)$ and $\operatorname{val}_U(v_i x)$ belong to the periodic part of $X \cap (Q_X\mathbb{N} + r_X)$ and they differ by a multiple of the period ρ_X. Therefore, $\operatorname{val}_U(u_i x)$ belongs to $X \cap (Q_X\mathbb{N}+r_X)$ if and only if $\operatorname{val}_U(v_i x)$ also does.

In order to obtain (4.3), it remains to show that either both $u_i x$ and $v_i x$ are valid greedy U-representations or both are not. If the word px is not a greedy U-representation then neither $u_i x$ nor $v_i x$ can be valid. Assume now that px is a greedy U-representation. Note that in both situations described in Figures 1 and 2, $|px| \leq |\operatorname{rep}_U(\rho_X-1)| + t$. Thanks to the assumption, we obtain $M_{\boldsymbol{\mu},X} - 1 - |px| + t \geq M_{\boldsymbol{\mu},X} - 1 - |\operatorname{rep}_U(\rho_X-1)| \geq Z$. The greediness of px and Lemma 14 imply that

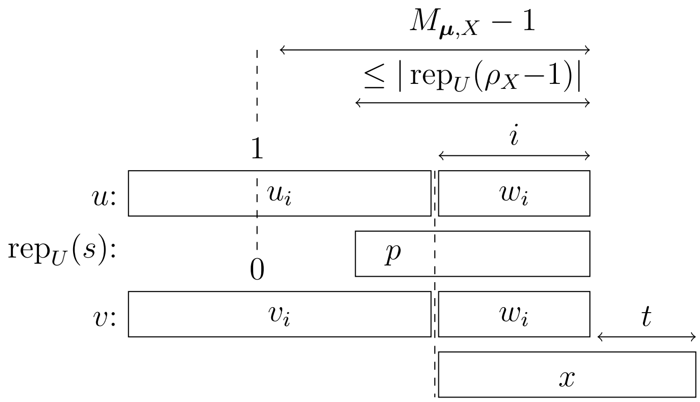

FIGURE 1. The different words (case where $i \leq |\operatorname{rep}_U(s)|$).

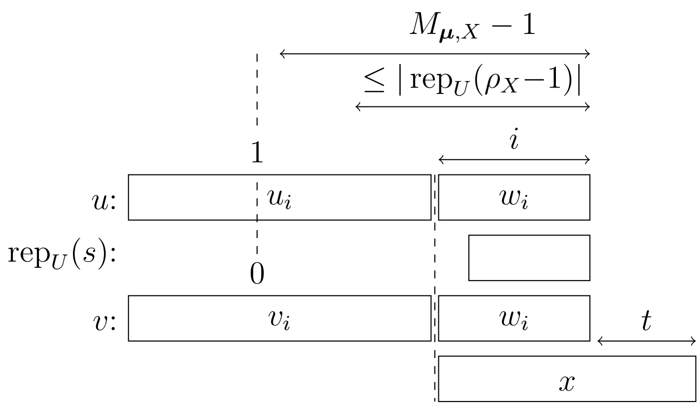

FIGURE 2. The different words (case where $i > |\operatorname{rep}_U(s)|$).

$10^{M_{\boldsymbol{\mu},X}-1-|px|+t}px$ is a greedy U-representation. Since $g \geq Z$, $u_i x$ is also a greedy U-representation and the same observation trivially holds for $v_i x$.

We conclude that

$$P_0 > P_1 > \cdots > P_{|\operatorname{rep}_U(\rho_X-1)|} \geq 1.$$

Since P_0 is the number of states of $\mathcal{B}_X$, the automaton $\mathcal{B}_X$ has at least $|\operatorname{rep}_U(\rho_X-1)|+1$ states.

Finally, let $\mathcal{A}_X$ and $\mathcal{A}_r$ be the minimal automata of $0^* \operatorname{rep}_U(X)$ and $0^* \operatorname{rep}_U(Q_X\mathbb{N}+r_X)$ respectively. The number of states of $\mathcal{A}_r$ is bounded by γ_{Q_X}. The DFA $\mathcal{B}_X$ is a quotient of the product automaton $\mathcal{A}_X \times \mathcal{A}_r$, hence the number of states of $\mathcal{B}_X$ is at most the number of states of $\mathcal{A}_X$ times γ_{Q_X}. We thus obtain that the number of states of $\mathcal{A}_X$ is at least $\frac{|\operatorname{rep}_U(\rho_X-1)|+1}{\gamma_{Q_X}}$. $\square$

5. Cases we can deal with

5.1. The gcd of the coefficients of the recurrence relation is 1.

In this case, for any ultimately periodic set X, the factorization of the period π_X given in (4.1) has the special form $\pi_X = Q_X$ and the addressed decision problem turns out to be decidable.

Theorem 37. *Let U be a linear numeration system satisfying (H1), (H2) and (H3), and such that the gcd of the coefficients of the recurrence relation* (1.1) *is* 1*. Given a DFA accepting a language contained in the numeration language* $\mathrm{rep}_U(\mathbb{N})$*, it is decidable whether this DFA recognizes an ultimately periodic set.*

Proof. Let $\mathcal{A}$ be a DFA accepting a language contained in the numeration language. Let X be the set of integers recognized by $\mathcal{A}$.

Assume that X is an ultimately periodic set with period π_X. Let p be a prime that divides π_X. Either $p \leq \max\{|a_0|, U_N\}$ or $p > \max\{|a_0|, U_N\}$.

In the former case, there is only a finite number of such primes. By assumption, p does not divide all the coefficients of the recurrence relation. Then thanks to Theorem 26, there exists $\lambda \geq 1$ such that the periodic part of the sequence $(U_i \bmod p^\lambda)_{i\geq 0}$ contains a non-zero element. By an exhaustive search, one can determine the value of the least such λ: one finds the period of a sequence $(U_i \bmod p^\lambda)_{i\geq 0}$ as soon as two k-tuples $(U_i \bmod p^\lambda, \ldots, U_{i+k-1} \bmod p^\lambda)$ are identical (where k is the order of the recurrence). We then apply Proposition 27. For any $\mu \geq 1$, if p^μ divides π_X then either $\mu < \lambda$ or $p^{\mu-\lambda+1}$ is bounded by the number S of states of $\mathcal{A}$. So we have bounded the exponent μ of those primes that may occur in π_X by $\max\{\lambda, \log_p(S) + \lambda - 1\}$.

In the latter case, thanks to Corollary 28, for any $\mu \geq 1$, if p^μ divides π_X then p^μ is bounded by the number of states of $\mathcal{A}$.

The previous discussion provides us with an upper bound on π_X, i.e. on the admissible periods for X. Then from Proposition 22, associated with each admissible period, there is a computable bound for the corresponding admissible preperiods for X. We conclude that there is a finite number of pairs of candidates for the preperiod and period of X. Similar to Honkala's scheme, we therefore have a decision procedure by enumerating a finite number of candidates. For each pair (a, b) of possible preperiods and periods, there are $2^a 2^b$ corresponding ultimately periodic sets X. For each such candidate X, we build a DFA accepting $\mathrm{rep}_U(X)$ and compare it with $\mathcal{A}$. We can conclude since equality of regular languages is decidable. $\square$

There exist recurrence relations satisfying the assumptions of the above theorem but that were not handled in [2]. Take [2, Example 35]

$$U_{i+5} = 6U_{i+4} + 3U_{i+3} - U_{i+2} + 6U_{i+1} + 3U_i,\ \forall i \geq 0.$$

For this recurrence relation, $\mathcal{N}_U(3^i) \not\to \infty$. The characteristic polynomial has the dominant root $3+2\sqrt{3}$ and it also has three roots of modulus 1. Therefore, no decision procedure was known. But thanks to Theorem 37, we can handle such new cases under our mild assumptions (H1), (H2) and (H3). Indeed, by applying Bertrand's theorem with the initial conditions $1, 7, 45, 291, 1881$, the numeration language $0^* \operatorname{rep}_U(\mathbb{N})$ is the set of words over $\{0, 1, \ldots, 6\}$ avoiding the factors $63, 64, 65, 66$, hence (H1) holds. Moreover, it is easily checked that for all $i \geq 0$, $U_{i+1} - U_i \geq 5U_i$. Therefore, the system U also satisfies (H2) and (H3).

5.2. **The gcd of the coefficients of the recurrence relation is larger than 1.** If X is an ultimately periodic set with period $\pi_X = Q_X \cdot p_1^{\mu_{X,1}} \cdots p_t^{\mu_{X,t}}$ with $t \geq 1$ as in (4.1), then the quantity $M_{\boldsymbol{\mu},X}$ is well defined. Theorem 36 has a major assumption. The quantity

$$n_X = M_{\boldsymbol{\mu},X} - 1 - |\operatorname{rep}_U(\rho_X - 1)|$$

should be larger than some positive constant Z, which only depends on the numeration system U.

Theorem 38. *Let U be a linear numeration system satisfying (H1), (H2) and (H3), and such that the gcd of the coefficients of the recurrence relation* (1.1) *is larger than 1. Let Z be the constant given in Definition 13. Assume there exists a computable positive integer D such that for all ultimately periodic sets X of period $\pi_X = Q_X \cdot p_1^{\mu_{X,1}} \cdots p_t^{\mu_{X,t}}$ as in* (4.1) *with $t \geq 1$, if $\max(\mu_{X,1}, \ldots, \mu_{X,t}) \geq D$ then $n_X \geq Z$. Then, given a DFA accepting a language contained in the numeration language $\operatorname{rep}_U(\mathbb{N})$, it is decidable whether this DFA recognizes an ultimately periodic set.*

Proof. Let $\mathcal{A}$ be a DFA accepting a language contained in the numeration language. Let X be the set of integers recognized by $\mathcal{A}$.

Assume that X is an ultimately periodic set with period $\pi_X = Q_X \cdot p_1^{\mu_{X,1}} \cdots p_t^{\mu_{X,t}}$ as in (4.1). Note that there are only finitely many primes dividing all the coefficients of the recurrence relation (1.1), hence the possible $p_1, \ldots, p_t$ belong to a finite set depending only on the numeration system U.

Applying the same reasoning as in the proof of Theorem 37, Q_X is bounded by a constant B deduced from $\mathcal{A}$. So the quantity γ_{Q_X} introduced in Definition 35 is also bounded.

Compute the greatest preperiod P of the sequences $(U_i \bmod b)_{i\in\mathbb{N}}$, for $b \in \{1,\dots,B\}$. Then by definition of $M_{\boldsymbol{\mu},X}$, there exists a computable constant D' such that if $\max(\mu_{X,1},\cdots,\mu_{X,t}) \geq D'$, then $M_{\boldsymbol{\mu},X}$ is greater than P.

By hypothesis, there is a computable positive integer constant D such that if $\max(\mu_{X,1},\cdots,\mu_{X,t}) \geq D$ then $n_X \geq Z$. Let $E = \max(D,D')$. The number of t-uples $(\mu_{X,1},\cdots,\mu_{X,t})$ in $\{0,\dots,E-1\}^t$ is finite. Hence there is a finite number of periods π_X of the form $Q_X \cdot p_1^{\mu_{X,1}} \cdots p_t^{\mu_{X,t}}$ with Q_X bounded by B and $(\mu_{X,1},\cdots,\mu_{X,t})$ in this set. We can enumerate them and proceed as in the last paragraph of the proof of Theorem 37.

We may now assume that $\max(\mu_{X,1},\cdots,\mu_{X,t}) \geq E$. In this case, $n_X \geq Z$. Moreover, $M_{\boldsymbol{\mu},X}$ is greater than P. We are thus able to apply Theorem 36[1]: it provides a bound on ρ_X and thus on the possible exponents $\mu_{X,1},\dots,\mu_{X,t}$ depending only on $\mathcal{A}$. We conclude in the same way as in the proof of Theorem 37. $\square$

In the last part of this section, we present a possible way to tackle new examples of numeration systems by applying Theorem 38. We stress the fact that when π_X is increasing then potentially both terms $M_{\boldsymbol{\mu},X}$ and $|\mathrm{rep}_U(\rho_X - 1)|$ are increasing. If $\beta > 1$ (see Definition 18), then the growth of $|\mathrm{rep}_U(\rho_X - 1)|$ has a logarithmic bound thanks to Lemma 19, so we need insight on $\mathsf{f}_{p_j}(\mu)$ to be able to guarantee $n_X \geq Z$. In the next few pages we therefore try to obtain conditions allowing us to apply the decision procedure of Theorem 38 and, facing non-trivial number theoretic problems, we discuss how far it is possible to go.

The *p-adic valuation* of an integer n, denoted $\nu_p(n)$, is the exponent of the highest power of p dividing n. There is a clear link between ν_{p_j} and f_{p_j}: for all non-negative integers μ and N,

$$\mathsf{f}_{p_j}(\mu) = N \iff (\nu_{p_j}(U_{N-1}) < \mu \ \wedge \ \forall i \geq N,\, \nu_{p_j}(U_i) \geq \mu).$$

Remark 39. With our Example 9 and initial conditions $1,2,3$, computing the first few values of $\nu_2(U_i)$, as shown in Figure 3, might suggest that it is bounded by a function of the form $\frac{i}{2} + c$, for some constant c. Nevertheless, computing more terms we get the following pairs $(i,\nu_2(U_i))$: $(67,44)$, $(2115,1070)$, $(10307,5172)$, $(534595,267318)$, $(2631747,1315896)$. The constant c suggested by each of these points is respectively $\frac{21}{2}$, $\frac{25}{2}$, $\frac{37}{2}$, $\frac{41}{2}$, $\frac{45}{2}$, which is increasing. This example explains the second term $g(i)$ in the function bounding $\nu_{p_j}(U_i)$ in the next statement.

[1]Considering leading zeroes or not does not change the reasoning.

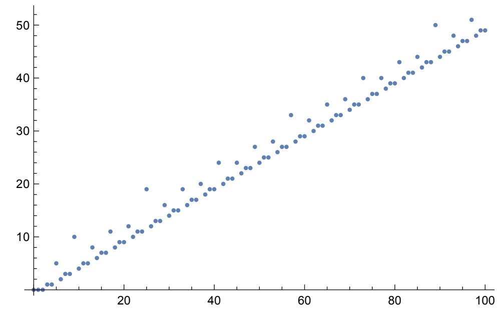

FIGURE 3. Plot of the 2-adic valuation of the sequence in Example 9.

In the next statement, the reader can think about logarithm function instead of a general function g. Indeed, for any $\epsilon > 0$, for large enough i, $\log(i) < \epsilon\, i$. We also keep context and notation from (4.1).

Lemma 40. *Let $j \in \{1, \dots, t\}$ and let β as in Definition 18. Assume that $\beta > 1$ and that there exist $\alpha, \epsilon \in \mathbb{R}_{>0}$ and a non-decreasing function g such that*

$$\nu_{p_j}(U_i) < \lfloor \alpha i \rfloor + g(i)$$

and there exists N such that $g(i) < \epsilon\, i$ for all $i > N$. Then, for large enough μ,

$$\mathsf{f}_{p_j}(\mu) > \frac{\mu}{\alpha + \epsilon}.$$

Proof. By definition of the p-adic valuation, $p_j^{\nu_{p_j}(U_i)} \mid U_i$ and $p_j^{\nu_{p_j}(U_i)+1} \nmid U_i$. Thus, by definition of f_{p_j}, for all i,

$$\mathsf{f}_{p_j}(\nu_{p_j}(U_i) + 1) \geq i + 1.$$

For all μ, there exists i such that

$$\lfloor \alpha i \rfloor + g(i) \leq \mu < \lfloor \alpha(i + 1) \rfloor + g(i + 1).$$

Take μ large enough so that $i \geq N$. Using the right-hand side inequality, $\mu < \alpha(i + 1) + \epsilon(i + 1)$ and we get

$$i > \frac{\mu}{\alpha + \epsilon} - 1.$$

Using the left-hand side inequality, $\mu \geq \lfloor \alpha i \rfloor + g(i) > \nu_{p_j}(U_i)$. Since we have integers on both sides, $\mu \geq \nu_{p_j}(U_i)+1$. Since f_{p_j} is non-decreasing, for all large enough μ,

$$\mathsf{f}_{p_j}(\mu) \geq \mathsf{f}_{p_j}(\nu_{p_j}(U_i) + 1) \geq i + 1 > \frac{\mu}{\alpha + \epsilon}. \qquad \square$$

We look for a lower bound for n_X. Suppose that for each $j \in \{1,\ldots,t\}$, there exists $\alpha_j, \epsilon_j, g_j$ and N_j as in the above lemma. Then

$$M_{\boldsymbol{\mu},X} = \max_j \mathsf{f}_{p_j}(\nu_{X,j}) > \max_j \left(\frac{\nu_{X,j}}{\alpha_j+\epsilon_j}\right) \geq \frac{\max_j \nu_{X,j}}{\max_j(\alpha_j+\epsilon_j)}.$$

Second, let u and β as in Definition 18. By hypothesis, $\beta > 1$. Applying Lemma 19, there exists a constant K such that

$$|\mathrm{rep}_U(\frac{\pi_X}{Q_X}-1)| \leq u \log_\beta\left(\prod_j p_j^{\mu_{X,j}}\right) + K.$$

The right hand side is

$$u\sum_j \mu_{X,j}\log_\beta(p_j) + K \leq u(\max_j \mu_{X,j})\sum_j \log_\beta p_j + K.$$

Recall that $\max_j \nu_{X,j} = \max_j \mu_{X,j}$ (see Lemma 32). Consequently,

$$n_X \geq \max_j \mu_{X,j}\left(\frac{1}{\max_j(\alpha_j+\epsilon_j)} - u\sum_j \log_\beta p_j\right) - K - 1.$$

If π_X tends to infinity (and assuming that the corresponding factor Q_X remains bounded as explained in the proof of Theorem 38), then $\max_j \mu_{X,j}$ must also tend to infinity. So we are able to conclude, i.e. n_X tends to infinity and in particular, n_X will become larger than Z (the constant from Definition 13) whenever

$$\frac{1}{\max_j(\alpha_j+\epsilon_j)} > u\sum_j \log_\beta p_j. \tag{5.1}$$

Actually, we don't need n_X tending to infinity, we have the weaker requirement $n_X \geq Z$. The constant D from Theorem 38 can be obtained as follows. To ensure that $n_X \geq Z$, it is enough to have

$$\max_j \mu_{X,j} \geq \frac{Z+K+1}{\frac{1}{\max_j(\alpha_j+\epsilon_j)} - u\sum_j \log_\beta p_j} \tag{5.2}$$

and the right hand side only depends on the numeration system U.

As a conclusion, we simply define the constant D as the right hand side in (5.2) and, under the assumption of Lemma 40 about the behavior of the p_j-adic valuations of $(U_i)_{i\geq 0}$, the decision procedure of Theorem 38 may thus be applied. From a practical point of view, even though n_X tending to infinity is not required, trying to make a conjecture on (5.1) is relatively easy as seen in the following remark. This is not a formal proof, simply rough computations suggesting what could be the value of α in Lemma 40.

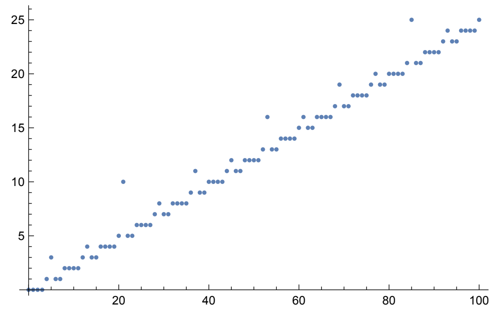

FIGURE 4. Plot of the 2-adic valuation of the sequence in Example 4.

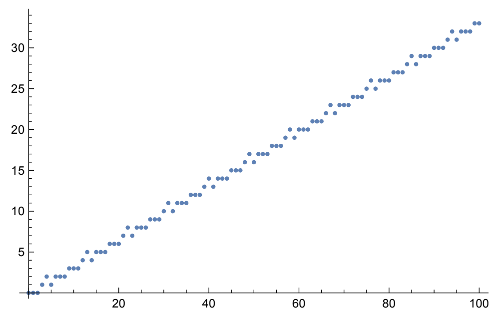

FIGURE 5. Plot of the 3-adic valuation of the sequence in Example 9.

Remark 41. One can first make some computational experiments. Take the numeration system of Example 4. If we compute $\nu_2(U_i)$, the values for $41 \le i \le 60$ are given by

$$10, 10, 10, 11, 12, 11, 11, 12, 12, 12, 12, 13, 16, 13, 13, 14, 14, 14, 14, 15.$$

This sequence is plotted in Figure 4. Hence, one can conjecture that $\alpha_1 = \frac{1}{4}$ and, assuming ϵ_1 to be negligible, the above condition (5.1) (with $u = 1$) becomes

$$4 > \log_{2.804}(2) \simeq 0.672.$$

Take the numeration system of Example 9. If we compute $\nu_2(U_i)$, the values for $41 \le i \le 60$ are given by

$$24, 20, 21, 21, 24, 22, 23, 23, 27, 24, 25, 25, 28, 26, 27, 27, 33, 28, 29, 29$$

and, similarly, for $\nu_3(U_i)$

$$13, 14, 14, 14, 15, 15, 15, 16, 17, 16, 17, 17, 17, 18, 18, 18, 19, 20, 19, 20.$$

These sequences are plotted in Figures 3 and 5. Hence, one can conjecture that $\alpha_1 = \frac{1}{2}$ and $\alpha_2 = \frac{1}{3}$. The recurrence has a real dominant root

$\beta \simeq 12.554$. Assuming ϵ_1 and ϵ_2 to be negligible, the condition (5.1) is therefore

$$2 > \log_{12.554}(2) + \log_{12.554}(3) \simeq 0.708.$$

6. An incursion into p-adic analysis

In this section, we discuss the requirement on the p-adic valuation given in Lemma 40. We are able to show that this condition holds in certain cases. In other cases, obtaining this condition requires information about the blocks of zeroes in the digit sequences of certain p-adic numbers, and in general it is not known how to get this information.

6.1. **A third-order sequence.** We reconsider our toy example. Throughout this section, let $U_{i+3} = 12U_{i+2}+6U_{i+1}+12U_i$ with initial conditions $U_0 = 1, U_1 = 13, U_2 = 163$ be the sequence of Example 9. The 3-adic valuation of U_i, shown in Figure 5, has a simple structure.

Theorem 42. *For all $i \geq 0$,*

$$\nu_3(U_i) = \left\lfloor \frac{i}{3} \right\rfloor + \begin{cases} 1 & \text{if } i \equiv 4 \pmod 9 \\ 0 & \text{if } i \not\equiv 4 \pmod 9. \end{cases}$$

Proof. Let $T_i = U_i/3^{\frac{i-2}{3}}$. Since $U_{i+3} = 12U_{i+2} + 6U_{i+1} + 12U_i$, the sequence $(T_i)_{i\geq 0}$ satisfies the recurrence $T_{i+3} = 4\cdot 3^{2/3}T_{i+2}+2\cdot 3^{1/3}T_{i+1}+4T_i$. The initial terms are $T_0 = 3^{2/3}, T_1 = 13\cdot 3^{1/3}, T_2 = 163$, so it follows that $T_i \in \mathbb{Z}[3^{1/3}]$ for all $i \geq 0$. Modulo $9\mathbb{Z}[3^{1/3}]$, one computes that the sequence $(T_i)_{i\geq 0}$ is periodic with period length 27 and period

$$\begin{array}{lllllllll} 3^{2/3}, & 4\cdot 3^{1/3}, & 1, & 7\cdot 3^{2/3}, & 3\cdot 3^{1/3}, & 1, & 2\cdot 3^{2/3}, & 2\cdot 3^{1/3}, & 4, \\ 3^{2/3}, & 3^{1/3}, & 7, & 7\cdot 3^{2/3}, & 3\cdot 3^{1/3}, & 7, & 8\cdot 3^{2/3}, & 5\cdot 3^{1/3}, & 1, \\ 3^{2/3}, & 7\cdot 3^{1/3}, & 4, & 7\cdot 3^{2/3}, & 3\cdot 3^{1/3}, & 4, & 5\cdot 3^{2/3}, & 8\cdot 3^{1/3}, & 7. \end{array}$$

Therefore the sequence $(\nu_3(T_i))_{i\geq 0}$ of 3-adic valuations is

$$\frac{2}{3},\ \frac{1}{3},\ 0,\ \frac{2}{3},\ \frac{4}{3},\ 0,\ \frac{2}{3},\ \frac{1}{3},\ 0,\ \ldots$$

with period length 9. (Here we use the natural extension of ν_3 to a function $\nu_3 \colon \mathbb{Z}[3^{1/3}] \to \frac{1}{3}\mathbb{Z}$.) Equivalently,

$$\nu_3(T_i) = \left\lfloor \frac{i}{3} \right\rfloor - \frac{i-2}{3} + \begin{cases} 1 & \text{if } i \equiv 4 \pmod 9 \\ 0 & \text{if } i \not\equiv 4 \pmod 9. \end{cases}$$

It follows that

$$\nu_3(U_i) = \frac{i-2}{3} + \nu_3(T_i) = \left\lfloor \frac{i}{3} \right\rfloor + \begin{cases} 1 & \text{if } i \equiv 4 \pmod 9 \\ 0 & \text{if } i \not\equiv 4 \pmod 9) \end{cases}$$

for all $i \geq 0$. □

Theorem 42 implies $\frac{i-2}{3} \leq \nu_3(U_i) \leq \frac{i+2}{3}$ for all $i \geq 0$. In particular, $\nu_3(U_i) < \lfloor \frac{i}{3} \rfloor + 2$, so the condition of Lemma 40 is satisfied, and therefore for every $\epsilon > 0$ we have

$$\mathsf{f}_3(\mu) > \frac{\mu}{\frac{1}{3} + \epsilon}$$

for large enough μ. This takes care of one of the two primes dividing 6, the gcd of the coefficients of the recurrence relation. To apply Theorem 38, it remains to bound $\nu_2(U_i)$.

However, Theorem 42 is not representative of the behavior of $\nu_p(s_i)$ for a general sequence $(s_i)_{i\geq 0}$ satisfying a linear recurrence with constant coefficients. For instance, the 2-adic valuation of $(U_i)_{i\geq 0}$ is (much) more complicated. To study the more general setting, we will make use of the field $\mathbb{Q}_p$ of p-adic numbers and its ring of integers $\mathbb{Z}_p$. The p-adic valuation $\nu_p(x)$ of an element $x \in \mathbb{Q}_p$ is related to its p-adic absolute value $|x|_p$ by $|x|_p = p^{-\nu_p(x)}$. For an introduction to p-adic analysis, see [15].

Let $|\mathrm{rep}_p(n)|$ be the number of digits in the standard base-p representation of n. For all $n \geq 1$, we can bound $\nu_p(n)$ as

$$\nu_p(n) \leq |\mathrm{rep}_p(n)| - 1 = \left\lfloor \frac{1}{\log(p)} \log(n) \right\rfloor \leq \frac{1}{\log(p)} \log(n).$$

(We avoid writing "$\log_p(n)$" here to reserve $\log_p$ for the p-adic logarithm, which will come into play shortly.) Proposition 43 below gives the analogous upper bound on $\nu_p(n - \zeta)$ when ζ is a p-adic integer whose sequence of base-p digits does not have blocks of consecutive 0s that grow too quickly.

Notation. Let p be a prime, and let $\zeta \in \mathbb{Z}_p \setminus \mathbb{N}$. Write $\zeta = \sum_{i\geq 0} d_i p^i$, where each $d_i \in \{0, 1, \ldots, p-1\}$. For each $a \geq 0$, let $\ell_\zeta(a) \geq 0$ be maximal such that $0 = d_a = d_{a+1} = \cdots = d_{a+\ell_\zeta(a)-1}$.

Proposition 43. *Let p be a prime, and let $\zeta \in \mathbb{Z}_p \setminus \mathbb{N}$. If there exist real numbers C, D such that $C > 0$, $D \geq -(C+1)$, and $\ell_\zeta(a) \leq Ca + D$ for all $a \geq 2$, then $\nu_p(n - \zeta) \leq \frac{2C+D+2}{\log(p)} \log(n)$ for all $n \geq p$.*

Proof. Write $\zeta = \sum_{i\geq 0} d_i p^i$, where each $d_i \in \{0, 1, \ldots, p-1\}$. For each $a \geq 0$, define the integer $\zeta_a := (\zeta \bmod p^a) = \sum_{i=0}^{a-1} d_i p^i$. Then $\nu_p(\zeta_a - \zeta) = a + \ell_\zeta(a)$.

Let $n \geq p$, and let $a := |\mathrm{rep}_p(n)| \geq 2$. Since $\zeta \notin \mathbb{N}$, the p-adic valuation $b := \nu_p(n - \zeta)$ is an integer. There are two cases.

If $n \leq \zeta_b$, then in fact $n = \zeta_b$; this is because $n \leq \zeta_b < p^b$, so $n \neq \zeta_b$ implies $n - \zeta_b \not\equiv 0 \pmod{p^b}$, which contradicts $b = \nu_p(n - \zeta)$. Since

$|\mathrm{rep}_p(n)| = a$ and $n = \zeta_b$, we have $0 = d_a = \cdots = d_{b-1}$. Therefore $\zeta_a = \zeta_b = n \geq p^{a-1}$, and

$$\frac{\nu_p(n-\zeta)}{\log(n)} = \frac{\nu_p(\zeta_a-\zeta)}{\log(\zeta_a)} \leq \frac{a+\ell_\zeta(a)}{\log(p^{a-1})} \leq \frac{a+Ca+D}{(a-1)\log(p)} \leq \frac{2+2C+D}{\log(p)},$$

where the final inequality follows from $1 + C + D \geq 0$.

If $n > \zeta_b$, then $n = \zeta_b + p^b m$ for some positive integer m. Therefore $n \geq p^b$, so

$$\frac{\nu_p(n-\zeta)}{\log(n)} \leq \frac{b}{\log(p^b)} = \frac{1}{\log(p)} < \frac{1+C}{\log(p)} \leq \frac{2+2C+D}{\log(p)}$$

if $b \geq 1$ and $\frac{\nu_p(n-\zeta)}{\log(n)} = 0 < \frac{2+2C+D}{\log(p)}$ if $b = 0$. □

We now turn our attention to the sequence of 2-adic valuations $\nu_2(U_i)$. The following result concerns the local peaks in Figure 3.

Theorem 44. *There exists a unique 2-adic integer ζ with the property that if $(i_n)_{n\geq 0}$ is a sequence of non-negative integers such that $\nu_2(U_{i_n}) - \frac{i_n}{2} \to \infty$ then $i_n \to \zeta$ in $\mathbb{Z}_2$.*

A formula for ζ is given by Equation (6.2) in the proof. In particular, ζ is a computable number, and one computes $\zeta \equiv 660098850944665 \pmod{2^{50}}$.

Proof of Theorem 44. Let $p = 2$. To analyze the 2-adic behavior of $(U_i)_{i\geq 0}$, we construct a piecewise interpolation of U_i to $\mathbb{Z}_2$ using the method described by Rowland and Yassawi [29]. Let $P(x) = x^3 - 12x^2 - 6x - 12$ be the characteristic polynomial of $(U_i)_{i\geq 0}$. The polynomial $P(x)$ has a unique root $\beta_1 \in \mathbb{Z}_2$ satisfying $\beta_1 \equiv 2 \pmod 4$; this can be shown by an application of Hensel's lemma (checking $|P(2)|_2 < |P'(2)|_2^2$). Polynomial division shows that $P(x)$ factors in $\mathbb{Z}_2[x]$ as

$$P(x) = (x-\beta_1)\left(x^2 + (\beta_1 - 12)x + (\beta_1^2 - 12\beta_1 - 6)\right).$$

One checks that $P(x)$ has no roots in $\mathbb{Z}_2$ congruent to 0, 1, 3, 4, 5, or 7 modulo 8. Since β_1 has multiplicity 1, this implies that the splitting field K of $P(x)$ is a quadratic extension of $\mathbb{Q}_2$. Let β_2 and β_3 be the other two roots of $P(x)$ in $K = \mathbb{Q}_2(\beta_2)$. Since $\beta_1 \equiv 2 \pmod 4$, the 2-adic absolute value of β_1 is $|\beta_1|_2 = \frac{1}{2}$. Using the quadratic factor of $P(x)$ and an approximation to β_1, one computes $|\beta_2|_2 = |\beta_3|_2 = \frac{1}{\sqrt{2}}$.

Let $c_1, c_2, c_3 \in K$ be such that

$$U_i = c_1\beta_1^i + c_2\beta_2^i + c_3\beta_3^i$$

for all $i \geq 0$. Using the initial conditions, we solve for c_1, c_2, c_3 to find

$$c_1 = \frac{-U_0\beta_2\beta_3 + U_1(\beta_2+\beta_3) - U_2}{(\beta_2-\beta_1)(\beta_1-\beta_3)}$$
$$c_2 = \frac{-U_0\beta_3\beta_1 + U_1(\beta_3+\beta_1) - U_2}{(\beta_3-\beta_2)(\beta_2-\beta_1)}$$
$$c_3 = \frac{-U_0\beta_1\beta_2 + U_1(\beta_1+\beta_2) - U_2}{(\beta_1-\beta_3)(\beta_3-\beta_2)},$$

where $U_0 = 1, U_1 = 13, U_2 = 163$. One computes $|c_1|_2 = 2$ and $|c_2|_2 = 2\sqrt{2} = |c_3|_2$. Factoring out β_2^i gives

$$U_i = \beta_2^i \left(c_1 \left(\tfrac{\beta_1}{\beta_2}\right)^i + c_2 + c_3 \left(\tfrac{\beta_3}{\beta_2}\right)^i \right). \tag{6.1}$$

Since $|\frac{\beta_1}{\beta_2}|_2 = \frac{1}{\sqrt{2}}$ and $|\frac{\beta_3}{\beta_2}|_2 = 1$, the power $(\frac{\beta_1}{\beta_2})^i$ approaches 0 as $i \to \infty$, while $(\frac{\beta_3}{\beta_2})^i$ does not. Therefore the size of $\nu_2(U_i/\beta_2^i) = \nu_2(U_i) - \frac{i}{2}$ for large i is determined by the proximity of $c_2 + c_3 (\frac{\beta_3}{\beta_2})^i$ to 0.

To analyze the size of $c_2 + c_3 (\frac{\beta_3}{\beta_2})^i$, we interpret $(\frac{\beta_3}{\beta_2})^i$ as a function of a p-adic variable. For this we need the p-adic exponential and logarithm, which are defined on extensions of $\mathbb{Q}_p$ by their usual power series; $\log_p(1+x)$ converges if $|x|_p < 1$, and $\exp_p x$ converges if $|x|_p < p^{-1/(p-1)}$. Moreover, $\log_p$ is an isomorphism from the multiplicative group $\{x : |x-1|_p < p^{-1/(p-1)}\}$ to the additive group $\{x : |x|_p < p^{-1/(p-1)}\}$, and its inverse map is $\exp_p$ [15, Proposition 4.5.9 and Section 6.1]. Direct computation shows $|(\frac{\beta_3}{\beta_2})^4 - 1|_2 = \frac{1}{8} < \frac{1}{2} = p^{-1/(p-1)}$. Therefore, for all $m \geq 0$ and $r \in \{0,1,2,3\}$,

$$\begin{aligned}
\left(\tfrac{\beta_3}{\beta_2}\right)^{r+4m} &= \left(\tfrac{\beta_3}{\beta_2}\right)^r \left(\tfrac{\beta_3}{\beta_2}\right)^{4m} \\
&= \left(\tfrac{\beta_3}{\beta_2}\right)^r \exp_2 \log_2\left(\left(\tfrac{\beta_3}{\beta_2}\right)^{4m}\right) \\
&= \left(\tfrac{\beta_3}{\beta_2}\right)^r \exp_2\left(m \log_2\left(\left(\tfrac{\beta_3}{\beta_2}\right)^4\right)\right).
\end{aligned}$$

Denote $L := \log_2((\frac{\beta_3}{\beta_2})^4)$. Using the power series for $\log_2$, one computes $|L|_2 = \frac{1}{8}$. For each $x \in \mathbb{Z}_2[\beta_2]$ and $r \in \{0,1,2,3\}$, define

$$f_r(r+4x) := c_2 + c_3 \left(\tfrac{\beta_3}{\beta_2}\right)^r \exp_2(Lx).$$

For all $x \in \mathbb{Z}_2$, we have $|Lx|_2 = \frac{1}{8}|x|_2 \leq \frac{1}{8} < \frac{1}{2} = p^{-1/(p-1)}$, so f_r is well defined on $r + 4\mathbb{Z}_2$. The four functions f_0, f_1, f_2, f_3 comprise a piecewise interpolation of $c_2 + c_3 (\frac{\beta_3}{\beta_2})^i$. Namely, $c_2 + c_3 (\frac{\beta_3}{\beta_2})^i = f_{i \bmod 4}(i)$ for all $i \geq 0$.

Since each f_r is a continuous function, from Equation (6.1) we see that $\nu_2(U_i/\beta_2^i) = \nu_2(U_i) - \frac{i}{2}$ is large when i is close to a zero of $f_{i \bmod 4}$. The equation $f_r(r+4x) = 0$ is equivalent to

$$\exp_2(Lx) = -\tfrac{c_2}{c_3}(\tfrac{\beta_2}{\beta_3})^r.$$

For $r \in \{0,2,3\}$, one computes $\left|-\frac{c_2}{c_3}(\frac{\beta_2}{\beta_3})^r - 1\right|_2 \geq \frac{1}{2}$, so there is no solution x for these values of r. For $r = 1$, $\left|-\frac{c_2}{c_3}(\frac{\beta_2}{\beta_3})^r - 1\right|_2 = \frac{1}{16} < \frac{1}{2}$, so there is a unique solution, namely $x = \frac{1}{L}\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)$, which has size $|x|_2 = \frac{1}{2}$. Let

$$\zeta := 1 + 4\tfrac{1}{L}\log_2\left(-\tfrac{c_2\beta_2}{c_3\beta_3}\right), \tag{6.2}$$

so that $f_1(\zeta) = 0$ and $|\zeta|_2 = 1$. It follows that every sequence $(i_n)_{n\geq 0}$ of non-negative integers with $\nu_2(U_{i_n}) - \frac{i_n}{2} \to \infty$ satisfies $i_n \to \zeta$. (If $\zeta \notin \mathbb{Z}_2$, then such sequences do not exist.)

It remains to show that $\zeta \in \mathbb{Z}_2$. Let $\sigma : K \to K$ be the Galois automorphism that non-trivially permutes β_2 and β_3. The formulas for c_2 and c_3 imply $\frac{c_2}{c_3} \cdot \frac{\sigma(c_2)}{\sigma(c_3)} = 1$; this implies

$$\begin{aligned}\log_2\left(-\tfrac{c_2\beta_2}{c_3\beta_3}\right) + \sigma\left(\log_2\left(-\tfrac{c_2\beta_2}{c_3\beta_3}\right)\right) &= \log_2\left(\tfrac{c_2\beta_2}{c_3\beta_3}\cdot\tfrac{\sigma(c_2)\beta_3}{\sigma(c_3)\beta_2}\right)\\ &= \log_2(1) = 0.\end{aligned}$$

Similarly,

$$\log_2((\tfrac{\beta_3}{\beta_2})^4) + \sigma\left(\log_2((\tfrac{\beta_3}{\beta_2})^4)\right) = \log_2(1) = 0.$$

Therefore

$$\frac{\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)}{\log_2((\frac{\beta_3}{\beta_2})^4)} = \frac{-\sigma\left(\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)\right)}{-\sigma\left(\log_2((\frac{\beta_3}{\beta_2})^4)\right)} = \sigma\left(\frac{\log_2\left(-\frac{c_2\beta_2}{c_3\beta_3}\right)}{\log_2((\frac{\beta_3}{\beta_2})^4)}\right)$$

is invariant under σ and thus is an element of $\mathbb{Q}_2$. It follows from $|\zeta|_2 = 1$ that $\zeta \in \mathbb{Z}_2$. □

Remark. The interpolation in the previous proof depends on appropriate powers of $\frac{\beta_3}{\beta_2}$ satisfying $x = \exp_2(\log_2(x))$. We verified this by directly checking $|(\frac{\beta_3}{\beta_2})^4 - 1|_2 < \frac{1}{2}$. In general, an appropriate exponent is given by [29, Lemma 6], namely

$$\begin{cases} 1 & \text{if } e < p-1 \\ p^{\lceil \log(e+1)/\log p\rceil} & \text{if } e \geq p-1, \end{cases}$$

where e is the ramification index of the field extension. The ramification index of the extension K in the proof of Theorem 44 is $e = 2$; this follows from the fact that e is a divisor of the degree of the extension and that $e \neq 1$ since we identified an element $\beta_2 \in K$ with 2-adic valuation $\nu_2(\beta_2) = \frac{1}{2}$. Therefore the exponent $2^{\lceil \log(3)/\log(2) \rceil} = 4$ suffices. Since $|\frac{\beta_3}{\beta_2}|_2 = 1$, [29, Lemma 6] implies $|(\frac{\beta_3}{\beta_2})^4 - 1|_2 < \frac{1}{2}$. (In general, one must divide by a root of unity before raising to the appropriate exponent, but this root of unity is 1 for $\frac{\beta_3}{\beta_2}$ since the ramification index of K is equal to its degree.)

By Proposition 43, the growth rate of $\nu_2(U_i)$ is determined by the approximability of

$$\zeta = \cdots 100101100001011011001111011000011011110110100110 01_2$$

by non-negative integers.

Conjecture 45. *Let $\zeta \in \mathbb{Z}_2$ be defined as in Equation* (6.2)*. The lengths of the* 0 *blocks of the* 2*-adic digits of ζ satisfy $\ell_\zeta(a) \le \frac{2}{95}a + \frac{18}{5}$ for all $a \ge 0$.*

Conjecture 45 is weak in the sense that it is almost certainly far from sharp. One expects the digits of ζ to be randomly distributed, in which case $\ell_\zeta(a) = \frac{1}{\log(2)} \log(a) + O(1)$. Indeed, among the first 1000 base-2 digits of ζ, the longest block of 0s has length 10. However, results concerning digits of irrational numbers are notoriously difficult to prove. Bugeaud and Kekeç [8, Theorem 1.6] give a lower bound on the number of non-zero digits among the first a digits of an irrational algebraic number in $\mathbb{Q}_p$; see also Theorem 2.1 in the same paper. However, there are no known results of this form for transcendental numbers.

The conjectural bound was obtained by computing the line through $\ell_\zeta(19) = 4$ and $\ell_\zeta(304) = 10$. If Conjecture 45 is true, then an explicit formula for $\nu_2(U_i)$ is given by the following theorem. In particular, the approximation $\zeta \equiv 660098850944665 \pmod{2^{50}}$ is sufficient to compute $\nu_2(U_i)$ for all $i \le 2^{49}$.

Theorem 46. *Let $\zeta \in \mathbb{Z}_2$ be defined as in Equation* (6.2)*. Conjecture 45 implies that, for all $i \ge 10$,*

$$\nu_2(U_i) = \left\lfloor \frac{i-1}{2} \right\rfloor + \begin{cases} \nu_2(i - \zeta) & \text{if } i \equiv 1 \pmod 4 \\ 0 & \text{if } i \not\equiv 1 \pmod 4. \end{cases}$$

Proof. We start as in the proof of Theorem 42. Let $T_i = U_i / 2^{\frac{i}{2}-1}$. Since $U_{i+3} = 12U_{i+2} + 6U_{i+1} + 12U_i$, the sequence $(T_i)_{i\ge 0}$ satisfies the recurrence $T_{i+3} = 6\sqrt{2}T_{i+2} + 3T_{i+1} + 3\sqrt{2}T_i$. The initial terms are

$T_0 = 2, T_1 = 13\sqrt{2}, T_2 = 163$, so it follows that $T_i \in \mathbb{Z}[\sqrt{2}]$ for all $i \geq 0$. Modulo $2\mathbb{Z}[\sqrt{2}]$, the sequence $(T_i)_{i\geq 2}$ is periodic with period length 4: $1, \sqrt{2}, 1, 0, 1, \sqrt{2}, 1, 0, \ldots$. It follows that if $i \geq 2$ and $i \not\equiv 1 \pmod 4$ then

$$\begin{aligned}\nu_2(U_i) = \frac{i}{2} - 1 + \nu_2(T_i) &= \frac{i}{2} - 1 + \begin{cases} 0 & \text{if } i \equiv 0 \pmod 4 \\ 0 & \text{if } i \equiv 2 \pmod 4 \\ \frac{1}{2} & \text{if } i \equiv 3 \pmod 4 \end{cases} \\ &= \left\lfloor \frac{i-1}{2} \right\rfloor.\end{aligned}$$

It remains to determine $\nu_2(U_i)$ when $i \equiv 1 \pmod 4$. We continue to use the 2-adic numbers $\beta_1, \beta_2, \beta_3, c_1, c_2, c_3$ and the function f_1 defined in the proof of Theorem 44. When $i \equiv 1 \pmod 4$, Equation (6.1) gives

$$|U_i|_2 = 2^{-\frac{i}{2}} \left| c_1 \left(\tfrac{\beta_1}{\beta_2}\right)^i + f_1(i) \right|_2.$$

To obtain a simpler formula for $|U_i|_2$, we compare the sizes of the two terms being added and use the fact that $|x + y|_p = \max\{|x|_p, |y|_p\}$ if $|x|_p \neq |y|_p$. For the first, we have $\left| c_1 \left(\tfrac{\beta_1}{\beta_2}\right)^i \right|_2 = 2^{1-\frac{i}{2}}$. For the second,

$$|f_1(i)|_2 = \left| c_2 + \tfrac{c_3\beta_3}{\beta_2} \exp_2\!\left(L \cdot \tfrac{i-1}{4}\right) \right|_2.$$

Since the function $f_1(1 + 4x) = c_2 + \frac{c_3\beta_3}{\beta_2} \exp_2(Lx)$ has a unique zero $\frac{\zeta-1}{4}$, the p-adic Weierstrass preparation theorem [15, Theorem 6.2.6] implies the existence of a power series $h(x) \in K[\![x]\!]$ such that $h(0) = 1$, $|h(x)|_2 = 1$ for all $x \in \mathbb{Z}_2[\beta_2]$, and

$$f_1(1 + 4x) = \frac{c_2 + \frac{c_3\beta_3}{\beta_2}}{-\frac{\zeta-1}{4}} \left(x - \tfrac{\zeta-1}{4}\right) h(x).$$

Therefore

$$\begin{aligned}|f_1(i)|_2 &= \left| \frac{c_2 + \frac{c_3\beta_3}{\beta_2}}{-\frac{\zeta-1}{4}} \right|_2 \left| \tfrac{i-1}{4} - \tfrac{\zeta-1}{4} \right|_2 \\ &= \sqrt{2}\, |i - \zeta|_2 .\end{aligned}$$

Conjecture 45 and Proposition 43 imply $|i - \zeta|_2 \geq i^{-536/95}$ for all $i \geq 2$. The functions $2^{1-\frac{i}{2}}$ and $\sqrt{2} i^{-536/95}$ intersect at $i \approx 70.21$. For all $i \geq 73$ such that $i \equiv 1 \pmod 4$,

$$\left| c_1 \left(\tfrac{\beta_1}{\beta_2}\right)^i \right|_2 = 2^{1-\frac{i}{2}} < \sqrt{2} i^{-536/95} \leq |f_1(i)|_2$$

and therefore

$$|U_i|_2 = 2^{-\frac{i}{2}} \left| c_1 \left(\tfrac{\beta_1}{\beta_2}\right)^i + f_1(i) \right|_2 = 2^{-\frac{i}{2}} |f_1(i)|_2 = 2^{\frac{1-i}{2}} |i - \zeta|_2 .$$

Moreover, explicit computation shows that $2^{1-\frac{i}{2}} < \sqrt{2}|i-\zeta|_2$ for all $i \equiv 1 \pmod 4$ satisfying $13 \le i \le 69$, so $|U_i|_2 = 2^{\frac{1-i}{2}} |i-\zeta|_2$ for these values as well. Therefore $\nu_2(U_i) = \frac{i-1}{2} + \nu_2(i-\zeta)$ for all $i \ge 13$ such that $i \equiv 1 \pmod 4$. $\square$

Corollary 47. *Conjecture 45 implies that* $\nu_2(U_i) \le \frac{i}{2} + \frac{536}{95\log(2)} \log(i)$ *for all* $i \ge 10$*.*

Proof. Since $U_i \ne 0$ for all $i \ge 0$, we have $|U_i|_2 \ne 0$ for all $i \ge 0$. Since $|f_1(\zeta)|_2 = 0$, this implies $\zeta \notin \mathbb{N}$. Conjecture 45 and Proposition 43 imply $\nu_2(i-\zeta) \le \frac{536}{95\log(2)} \log(i)$ for all $i \ge 2$. By Theorem 46, $\nu_2(U_i) \le \frac{i}{2} + \frac{536}{95\log(2)} \log(i)$ for all $i \ge 10$. $\square$

This is sufficient to apply Lemma 40. Assuming Conjecture 45, we have the right behavior for both $\nu_2(U_i)$ and $\nu_3(U_i)$, and therefore we may apply the decision procedure of Theorem 38.

6.2. **A fourth-order sequence.** Bounding the p-adic valuation of a sequence satisfying a recurrence of higher order is even more complicated than the proof of Theorem 44. For example, let $p = 2$ and consider the sequence $(U_i)_{i\ge 0}$ satisfying the recurrence $U_{i+4} = 2U_{i+3} + 2U_{i+2} + 2U_n$ with initial conditions $U_0 = 1, U_1 = 3, U_2 = 9, U_2 = 23$ from Example 4. The 2-adic valuation is shown in Figure 4. By the Eisenstein criterion, the characteristic polynomial $P(x) = x^4 - 2x^3 - 2x^2 - 2$ is irreducible over $\mathbb{Q}_2$. Let K be the splitting field of $P(x)$ over $\mathbb{Q}_2$. Let $\beta_1, \beta_2, \beta_3, \beta_4$ be the four roots of $P(x)$ in K, and let c_1, c_2, c_3, c_4 be the elements of K such that $U_i = \sum_{j=1}^4 c_j \beta_j^i$ for all $i \ge 0$.

To compute with the roots β_i, we would want to write K as a simple extension $\mathbb{Q}_2(\alpha)$. For this, we need to determine the degree d of the extension and a polynomial $Q(x) \in \mathbb{Q}_2[x]$ of degree d such that $Q(x)$ is irreducible over $\mathbb{Q}_2$ and $Q(\alpha) = 0$. Then we could compare the sizes $|\beta_j|_2$ of the roots to each other. Experiments suggest that $|\beta_1|_2 = |\beta_2|_2 = |\beta_3|_2 = |\beta_4|_2 = 2^{-1/4}$ and $|(\frac{\beta_j}{\beta_1})^8 - 1|_2 = \frac{1}{4} < \frac{1}{2} = p^{-1/(p-1)}$ for each $j \in \{2,3,4\}$. Assuming this is the case, $U_i/\beta_1^i = \sum_{j=1}^4 c_j (\frac{\beta_j}{\beta_1})^i$ can be interpolated piecewise to $\mathbb{Z}_2$ using 8 analytic functions. However, we cannot solve $c_1 + b_2 \exp_2(L_2 x) + b_3 \exp_2(L_3 x) + b_4 \exp_2(L_4 x) = 0$ explicitly, as we solved $c_2 + c_3 \left(\frac{\beta_3}{\beta_2}\right)^r \exp_2(Lx) = 0$ in the proof of Theorem 44. Instead, we could use the p-adic Weierstrass preparation theorem [15, Theorem 6.2.6] to determine the number of solutions and compute approximations to them. However, we would also need to determine which of these solutions belong to $\mathbb{Z}_2$. We do not carry out this step here, but this would give an analogue of Theorem 44, with some finite set Z of 2-adic integers such that every sequence $(i_n)_{n\geq 0}$ of non-negative integers with $\nu_2(U_{i_n}) - \frac{i_n}{4} \to \infty$ satisfies $i_n \to \zeta$ for some $\zeta \in Z$. If the blocks of zeroes in the digit sequences of each $\zeta \in Z$ satisfy $\ell_\zeta(a) \leq Ca + D$ for some C, D as in Conjecture 45, then Proposition 43 gives an upper bound on $\nu_2(U_i)$. This same approach applies to a general constant-recursive sequence and a general prime p.

7. Concluding remarks

The case of integer base b numeration systems is not treated in this paper. Let $b \geq 2$. Assume first for the sake of simplicity that b is a prime. Consider the sequence $U = (b^i)_{i\geq 0}$. If X is an ultimately periodic set with period $\pi_X = b^\lambda$ for some λ, then with our notation $Q_X = 1$ and $|\mathrm{rep}_U(\pi_X - 1)| = \lambda$. The sequence $(b^i \bmod b^\lambda)_{i\geq 0}$ has a zero period and $\mathsf{f}_b(\lambda) = \lambda$. Hence we don't have the required assumption to apply Theorem 36: for every such set X, $n_X = 0$. Let us also point out that the technique of Proposition 27 cannot be applied: adding 1 as a most significant digit will not change the value of a representation modulo π_X when words are too long, $U_i \equiv 0 \pmod{b^\lambda}$ for large enough i. Of course, integer base systems can be handled with other decision procedures [5, 6, 16, 18, 21, 22]. If the base b is now a composite number of the form $p_1^{s_1} \cdots p_t^{s_t}$, the same observation holds. The length of the non-zero preperiod of $(b^i \bmod p_j^\mu)_{i\geq 0}$ is $\lfloor \frac{\mu}{s_j} \rfloor$. Taking again an ultimately periodic set with period $\pi_X = b^\lambda$, we get $Q_X = 1$ and $\mathsf{f}_{p_j}(\lambda s_j) = \lambda$, hence $M_{\boldsymbol{\mu},X} = \lambda$ and we still have $|\mathrm{rep}_U(\pi_X - 1)| = \lambda$, so $n_X = 0$.

A similar situation occurs in a slightly more general setting: the merge of r sequences that ultimately behave like b^i. Let $b \geq 2$, $u \geq 1$, $N \geq 0$. If the recurrence relation is of the form $U_{i+u} = bU_i$ for $i \geq N$ (as for instance in Example 16), then again $n_X \not\to \infty$ as $\pi_X \to \infty$. Indeed, if X is an ultimately periodic set with period $\pi_X = b^\lambda$, then $Q_X = 1$ and applying Lemma 19 (here the polynomial P_T with the notation of Definition 18 is just a constant), $|\mathrm{rep}_U(\pi_X - 1)| \geq u\lambda - L$, for some constant L, and with the same reasoning as for a composite integer base, $M_{\boldsymbol{\mu},X} \leq N + u\lambda$. Thus n_X remains bounded for all λ. So there is no way to ensure that n_X can be larger than Z.

Trying to figure out the limitations of our decision procedure and assuming that we are under the assumption of Lemma 40, this type of linear numeration systems is the only one that we were able to find where our procedure cannot be applied. Moreover, as shown by the following proposition, these systems are sufficiently close to the classical base-b system so usual decision procedures can still be applied. It is an open problem to determine if there exist linear numeration systems satisfying (H1), (H2) and (H3) where the decision procedure may not be applied and not of the above type.

Example 48. Take $b = 4$, $u = 2$ and $N = 0$. Start with the first two values 1 and 3. We get the sequence $1, 3, 4, 12, 16, 48, 64, \ldots$. We have $\mathsf{f}_2(\mu) = \mu$ if μ is even and $\mathsf{f}_2(\mu) = \mu + 1$ if μ is odd. Hence, for a set of period $\pi_X = 4^\lambda$, $M_{\boldsymbol{\mu},X} = \mathsf{f}_2(2\lambda) = 2\lambda$. Moreover, $|\mathrm{rep}_U(4^\lambda - 1)| = 2\lambda$. So, $n_X = 0$ for all λ.

Proposition 49. *Let $b \geq 2$, $u \geq 1$, $N \geq 0$. Let U be a linear numeration system $U = (U_i)_{i\geq 0}$ such that $U_{i+u} = bU_i$ for all $i \geq N$. If a set is U-recognizable then it is b-recognizable. Moreover, given a DFA accepting* $\mathrm{rep}_U(X)$ *for some set X, we can compute a DFA accepting* $\mathrm{rep}_b(X)$.

Proof. We build in two steps a sequence of transducers reading least significant digit first that maps any U-representation $c_{\ell-1} \cdots c_1 c_0 \in A_U^*$ (here written with the usual convention that the most significant digit is on the left) to the corresponding b-ary representation. Adding leading zeroes, we may assume that the length ℓ of the U-representation is of the form $N + mu$. The idea is to read the first $N + u$ (least significant) digits and to output a single digit (over a finite alphabet in $\mathbb{N}$) equal to

$$d_0 = \mathrm{val}_U(c_{N+u-1} \cdots c_0).$$

Then we process blocks of size u, each such block of the form

$$c_{N+(j+1)u-1} \cdots c_{N+ju}$$

gives as output a single digit equal to

$$d_j = c_{N+(j+1)u-1} U_{N+u-1} + \cdots + c_{N+ju} U_N.$$

So the digits $d_0, d_1, \ldots, d_{m-1}$ all belong to the finite set

$$\{\mathrm{val}_U(w) \colon w \in A_U^* \text{ and } |w| \leq N + u\}.$$

From the form of the recurrence, we have

$$\mathrm{val}_U(c_{N+mu-1} \cdots c_0) = \sum_{j=0}^{m-1} d_j b^j = \mathrm{val}_b(d_{m-1} \cdots d_0).$$

So this transducer $\mathcal{T}$ maps any U-representation to a non-classical b-ary representation of the same integer. Precisely, when a DFA accepting $\mathrm{rep}_U(X)$ is given, we build a DFA accepting the language

$$L = 0^* \,\mathrm{rep}_U(X) \cap \{w \in A_U^* \colon |w| \equiv N \pmod{u},\ |w| \geq N\}.$$

Recall that if L is a regular language then its image $\mathcal{T}(L)$ by a transducer is again regular. Moreover, $\mathrm{val}_b(\mathcal{T}(L)) = X$.

Then, it is a classical result that normalization in base b, i.e. mapping a representation over a non-canonical finite set of digits to the canonical expansion over $\{0, \dots, b-1\}$ can be achieved by a transducer $\mathcal{N}$ [13] (or [28, p. 104]). To conclude with the proof, we compose these two transducers and consider the image $\mathcal{N}(0^*\mathcal{T}(L)) = 0^* \operatorname{rep}_b(X)$. $\square$

With the above proposition, the decision problem for the merge of sequences ultimately behaving like b^i (such as the numeration systems of Examples 16 and 20) can be reduced to the usual decision problem for integer bases.

Acknowledgments

We thank Yann Bugeaud for pointing out relevant theorems in [8]. We thank Juha Honkala, Victor Marsault (who served as external reviewers for [23]) and the anonymous referee for their careful reading and their suggestions leading to many improvements along the text. We also thank Joël Ouaknine for pointing out [1] and the reference to the absolutely divergent problem.

References

[1] S. Almagor, B. Chapman, M. Hosseini, J. Ouaknine, J. Worrell, Effective divergence analysis for linear recurrence sequences, S. Schewe et al. (Ed.), 29th international conference on concurrency theory CONCUR 2018, *LIPIcs. Leibniz Int. Proc. Inform.* **118**, Art. 42, (2018).

[2] J. Bell, É. Charlier, A. Fraenkel, M. Rigo, A decision problem for ultimately periodic sets in nonstandard numeration system, *Internat. J. Algebra Comput.* **19** (2009), no. 6, 809–839.

[3] V. Berthé, M. Rigo (Eds), *Combinatorics, Automata, and Number Theory*, Encycl. Math. and its Appl. **135**, Cambridge Univ. Press (2010).

[4] A. Bertand-Mathis, Comment écrire les nombres entiers dans une base qui n'est pas entière, *Acta Math. Hungar.* **54** (1989), 237–241.

[5] B. Boigelot, I. Mainz, V. Marsault, M. Rigo, An efficient algorithm to decide periodicity of b-recognisable sets using MSDF convention, 44th International Colloquium on Automata, Languages, and Programming, Art. 118, 14 pp., *LIPIcs. Leibniz Int. Proc. Inform.* **80**, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, (2017).

[6] V. Bruyère, G. Hansel, Bertrand numeration systems and recognizability, *Theoret. Comput. Sci.* **181** (1997), 17–43.

[7] V. Bruyère, G. Hansel, Ch. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994), 191–238.

[8] Y. Bugeaud, G. Kekeç, On Mahler's classification of p-adic numbers, *Bulletin of the Australian Mathematical Society* **98** (2018) 203–211.

[9] A. Cobham, On the base-dependence of sets of numbers recognizable by finite automata, *Math. Systems Theory* **3** (1969), 186–192.

[10] F. Durand, Decidability of the HD0L ultimate periodicity problem, *RAIRO Theor. Inform. Appl.* **47** (2013), 201–214.

[11] F. Durand, M. Rigo, On Cobham's theorem, in *Handbook of Automata*, vol. II, Automata in Mathematics and Selected Applications, J.-É. Pin Ed., EMS Press, Berlin (2021).

[12] A. S. Fraenkel, Systems of numeration, *Amer. Math. Monthly* **92** (1985), 105–114.

[13] Ch. Frougny, Representations of numbers and finite automata, Math. Systems Theory **25** (1992), 37–60.

[14] Ch. Frougny, On the sequentiality of the successor function, *Inform. Comput.* **139** (1997), 17–38.

[15] F. Q. Gouvêa, *p-adic Numbers: An Introduction* second edition, Universitext, Springer–Verlag, Berlin, 1997.

[16] J. Honkala, A decision method for the recognizability of sets defined by number systems, *Theoret. Inform. Appl.* **20** (1986), 395–403.

[17] M. Hollander, Greedy numeration systems and regularity, *Theory Comput. Syst.* **31** (1998), 111–133.

[18] A. Lacroix, N. Rampersad, M. Rigo, É. Vandomme, Syntactic complexity of ultimately periodic sets of integers and application to a decision procedure, *Fund. Infor.* **116** (2012), 175–187.

[19] P. B. A. Lecomte, M. Rigo, Numeration Systems on a Regular Language, *Theory Comput. Systems* **34** (2001), 27–44.

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

[21] V. Marsault, An efficient algorithm to decide periodicity of b-recognisable sets using LSDF convention, *Log. Methods Comput. Sci.* **15** (2019), Paper No. 8, 30.

[22] V. Marsault, J. Sakarovitch, Ultimate periodicity of b-recognisable sets: a quasilinear procedure, Developments in language theory, 362–373, *Lect. Notes in Comput. Sci.* **7907**, Springer, Heidelberg, (2013).

[23] A. Massuir, *Positional Numeration Systems: Ultimate Periodicity, Complexity and Automatic Sequences*, Ph.D. thesis, University of Liège, (2021), `https://orbi.uliege.be/handle/2268/258492`

[24] I. V. Mitrofanov, Almost periodicity of morphic words, *Dokl. Math.* **93** (2016), 207–210.

[25] M. Marden, *The geometry of the zeros of a polynomial in a complex variable*, Mathematical Surveys, No. 3. American Mathematical Society, New York, N. Y., (1949).

[26] J. Ouaknine, J. Worrell, Positivity problems for low-order linear recurrence sequences. In Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 366–379, ACM, New York (2014).

[27] J. Ouaknine, J. Worrell, Ultimate positivity is decidable for simple linear recurrence sequences, in Automata, languages, and programming. Part II, 318–329, *Lect. Notes in Comput. Sci.* **8573**, Springer, Heidelberg, (2014).

[28] M. Rigo, *Formal Languages, Automata and Numeration Systems: Applications to Recognizability and Decidability*, vol. 2, Networks and Telecommunications series, ISTE-Wiley, 2014.

[29] E. Rowland, R. Yassawi, p-adic asymptotic properties of constant-recursive sequences, *Indag. Math.* **28** (2017), 205–220.

[30] J. Sakarovitch, *Elements of automata theory*, Cambridge University Press, Cambridge, (2009).

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

[32] J. Shallit, Numeration systems, linear recurrences and regular sets, *Inform. Comput.* **113** (1994), 331–347.

[1] University of Liège, Department of Mathematics, Allée de la découverte 12 (B37), B-4000 Liège, Belgium

Email address: echarlier@uliege.be ; A.Massuir@uliege.be ; M.Rigo@uliege.be

[2] Department of Mathematics, Hofstra University, Hempstead, NY, USA

Email address: Eric.Rowland@hofstra.edu