On Sets of Numbers Rationally Represented in a Rational Base Number System.

Victor Marsault* and Jacques Sakarovitch

Telecom-ParisTech and CNRS, 46 rue Barrault 75013 Paris, France

Abstract. In this work, it is proved that a set of numbers closed under addition and whose representations in a rational base numeration system is a rational language is not a finitely generated additive monoid.
A key to the proof is the definition of a strong combinatorial property on languages : the bounded left iteration property. It is both an unnatural property in usual formal language theory (as it contradicts any kind of pumping lemma) and an ideal fit to the languages defined through rational base number systems.

1 Introduction

The numeration systems in which the base is a rational number have been introduced and studied in [1]. It appeared there that the language of representations of all integers in such a system is "complicated", by reference to the classical Chomsky hierarchy and its usual iteration properties. This work is a contribution to a better understanding of the structure of this language. It consists in a result whose statement first requires some basic facts about number systems.

Given an integer p as a base, the set of non-negative integers $\mathbb{N}$ is represented by the set of words on the alphabet $A_p = \{0, 1, \ldots, (p-1)\}$ which do not begin with a 0. This set $L_p = (A_p \backslash 0)A_p^*$ is rational, that is, accepted by a finite automaton. This representation of integers has another property related to finite automata: the addition is realised by a finite 3-tape automaton.

This addition algorithm can be broken down into two steps : first a digit-wise addition which outputs a word on the double alphabet A_{2p-1} whose value in base p is the sum of the two input words; second a transformation of a word of $(A_{2p-1})^*$ into a word of A_p^* without modifying its value. This second step can be done by a finite transducer called the *converter* (see Section 2.2.2 of [3]).

Many non-standard numeration systems that have been studied so far have the property that the set of representations of the integers is a rational language. It is even *the* property that is retained in the study of the *abstract numeration systems*, even if it is not the case that addition can be realised by a finite automaton (*cf.* [6]).

In the rational base numeration systems, as defined and studied in [1], the situation is reverse: the set of integers is not represented by a rational language

* Corresponding author, victor.marsault@telecom-paristech.fr

(not even a context-free one), but nevertheless the addition is realised by a finite automaton. More precisely, let p and q be two coprime integers, with $p > q$. In the $\frac{p}{q}$-numeration system, the digit alphabet is again A_p, and the value of a word $u = a_n \cdots a_2 a_1$ in A_p^* is $\pi(u) = \frac{1}{q}\sum_{i=0}^{n} a_i (\frac{p}{q})^i$. In this system, every integer has a unique finite representation, but the set $L_{\frac{p}{q}}$ of the $\frac{p}{q}$-representations of the integers is not a rational language. The set $V_{\frac{p}{q}}$ of all numbers that can be represented in this system, $V_{\frac{p}{q}} = \pi(A_p^*)$, is closed under addition but is not finitely generated (as an additive monoid).

In this work, we establish the contradiction between being a finitely generated additive monoid and having a rational set of representations in a rational base number system.

Theorem 1. *The set of the $\frac{p}{q}$-representations of any finitely generated additive submonoid of $V_{\frac{p}{q}}$ is not a rational language.*

The proof of this statement relies on three ingredients. The first one is the description of a weak iteration property whose negation is satisfied by the language $L_{\frac{p}{q}}$. The second one is the construction of a sequential letter-to-letter right transducer that realises, on the $\frac{p}{q}$-representations, the addition of a fixed value to the elements of $V_{\frac{p}{q}}$. Finally, the third one is a characterisation of a finitely generated additive submonoid of $V_{\frac{p}{q}}$ as a finite union of translates of the set of the integers.

The paper is organised as follows: after the preliminaries, where we essentially recall the definition of transducers, we present with more details in Section 3 the numeration system in base $\frac{p}{q}$. In Section 4, we describe the Bounded Left Iteration Property (BLIP) and in Section 5, we build a transducer called incrementer. In the last section, we give the proof of a much stronger statement than Theorem 1, expressed with the BLIP property.

2 Preliminaries

We essentially follow notations and definitions of [8] for automata and transducers. An *alphabet* is a finite set of *letters*, the *free monoid* generated by A, and denoted by A^*, is the set of finite *words* over A. The *concatenation* of two words u and v of A^* is denoted by $u\,v$, or by $u.v$ when the dot adds hopefully to readability. A *language* (over A) is any subset of A^*.

A language is said to be *rational* (resp. *context-free*) if it is accepted by a finite automaton (resp. a pushdown automaton). The precise definitions of these classes of automata are however irrelevant to the present work, and can be found in [5]. Similarly, we are only considering (and thus defining) a very restricted class of transducers, namely the *sequential letter-to-letter* transducer.

Given two alphabets A and B, a sequential letter-to-letter (left) transducer $\mathcal{T}$ from A^* to B^* is a directed graph whose edges are labelled in $A \times B$. More precisely, $\mathcal{T}$ is defined by a 6-tuple $\mathcal{T} = \langle\, Q, A, B, \delta, \eta, i, \omega \,\rangle$ where Q is the set of *states*; A is the *input alphabet*; B is the *output alphabet*; $\delta : Q \times A \to Q$ is the *transition function*; $\eta : Q \times A \to B$ is the *output function*; i is the *initial state* and $\omega : Q \to B^*$ is the *final function*.

Moreover, we call *final* any state in the definition domain of ω. As usual, the function δ (resp. η) is extended to $Q \times A^* \to Q$ (resp. $Q \times A^* \to B^*$) by $\delta(p, \varepsilon) = p$ (resp. $\eta(p, \varepsilon) = \varepsilon$) and $\delta(p, a.u) = \delta(\delta(p, a), u)$ (resp. $\eta(p, a.u) = \eta(p, a).\eta(\delta(p, a), u)$).

Given $\mathcal{T}$, we write $p \xrightarrow[\mathcal{T}]{u \mid v} q$ if, and only if, $\delta(p, u) = q$ and $\eta(p, u) = v$. By analogy, we denote by $p \xrightarrow[\mathcal{T}]{w}$ the fact that p is a final state and that $\omega(p) = w$. The *image* by $\mathcal{T}$ of a word u, denoted by $\mathcal{T}(u)$, is the word $v.w$, if $i \xrightarrow[\mathcal{T}]{u \mid v} p \xrightarrow[\mathcal{T}]{w}$.

Finally, a transducer is said to be a *right transducer*, if it reads the words from right to left; and to be *complete* if both the transition function and the output function are total functions.

In the following, every considered transducer will be complete, letter-to-letter, right and sequential.

3 Rational Base Number System

We recall here the definitions, notations and constructions of [1]. Let p and q be two coprime integers such that $p > q > 1$. Given a positive integer N, let us define $N_0 = N$ and for all $i > 0$:

$$q N_i = pN_{i+1} + a_i \tag{1}$$

where a_i is the remainder of the Euclidean division of $q N_i$ by p, hence in A_p. Since $p > q$, the sequence $(N_i)_i$ is strictly decreasing and eventually stops at $N_{k+1} = 0$. Moreover the equation

$$N = \sum_{i=0}^{k} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \tag{2}$$

holds. The evaluation function π is derived from this formula. The value of a word $u = a_n a_{n-1} \cdots a_0$ over A_p is defined as

$$\pi(a_n a_{n-1} \cdots a_0) = \sum_{i=0}^{n} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \tag{3}$$

Conversely, a word u is called a $\frac{p}{q}$-representation of a number x if $\pi(u) = x$. Since the representation is unique up to leading 0's (see [1, Theorem 1]), u is denoted by $\langle x \rangle_{\frac{p}{q}}$ (or $\langle x \rangle$ for short), and in the case of integers, can be computed with the modified Euclidean division algorithm above. By convention, the representation of 0 is the empty word ε.

It should be noted that a rational base number systems is **not** a β-numeration (*cf.* [7, Chapter 7]) in the special case where β is rational. In the latter, the digit set is $\{0, 1, \ldots, \lceil \frac{p}{q} \rceil\}$ and the weight of the i-th leftmost digit is $(\frac{p}{q})^i$; whereas in rational base number systems, they respectively are $\{0, 1, \ldots, (p-1)\}$ and $\frac{1}{q}(\frac{p}{q})^i$.

Definition 1. *The representations of integers in the $\frac{p}{q}$-system form a language over A_p, which is denoted by $L_{\frac{p}{q}}$.*

It is immediate that $L_{\frac{p}{q}}$ is prefix-closed (since, in the modified Euclidean division algorithm $\langle N \rangle = \langle N_1 \rangle . a_0$) and prolongable (there exists an a such that q divides $(np + a)$ and then $\langle \frac{np+a}{q} \rangle = \langle n \rangle . a$). As a consequence, $L_{\frac{p}{q}}$ can be represented as a tree whose branches are all infinite (*cf.* Figure 1). On the

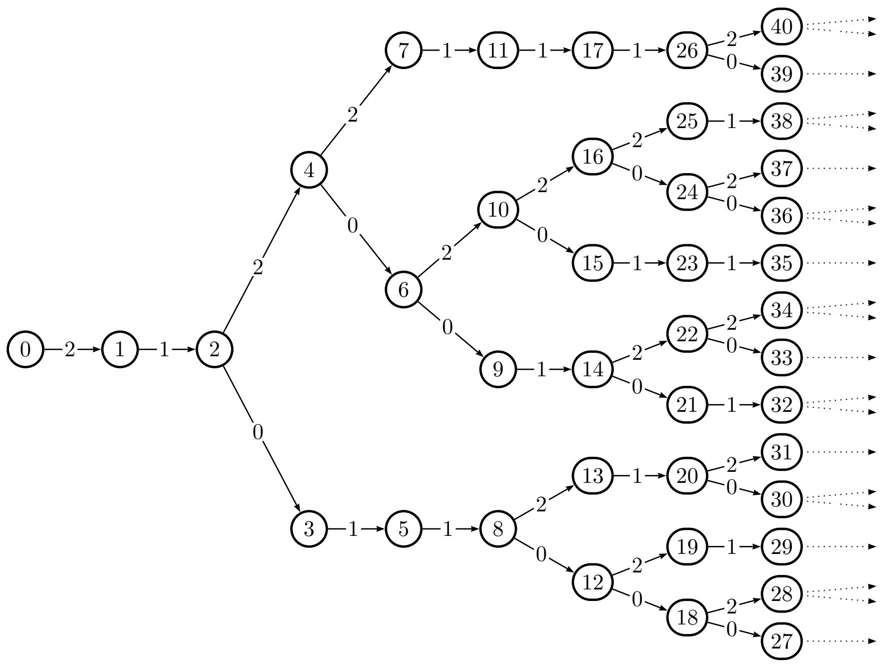

Fig. 1: The tree representation of the language $L_{\frac{3}{2}}$

other hand, the suffix language of $L_{\frac{p}{q}}$ is all A_p^*, and, moreover, every suffix appears periodically as established by the following:

Proposition 1 ([1, Proposition 10]). *For every word u over A_p of length k, there exists an integer $n < p^k$ such that u is a suffix of $\langle m \rangle$ if, and only if, m is congruent to n modulo p^k.*

In short, the congruence modulo p^k of n determines the suffix of length k of $\langle n \rangle$. In contrast, the congruence modulo q^k of n determines the words of length k appendable to $\langle n\rangle$ in order to stay in $L_{\frac{p}{q}}$, as is stated in the next lemma.

Lemma 1 ([1, Lemma 6]). *Given two integers n, m and a word u over A_p:*

(i) if both $\langle n\rangle.u$ and $\langle m\rangle.u$ are in $L_{\frac{p}{q}}$, then $n \equiv m\ [q^{|u|}]$
(ii) if $n \equiv m\ [q^{|u|}]$, $\langle n\rangle.u$ is in $L_{\frac{p}{q}}$ implies $\langle m\rangle.u$ is in $L_{\frac{p}{q}}$.

Proof. (i). The word $\langle n\rangle.u$ is in $L_{\frac{p}{q}}$ if, and only, if $(n(\frac{p}{q})^{|u|} + \pi(u))$ is an integer, and similarly for m. It follows that $(n-m)(\frac{p}{q})^{|u|}$ is equal to some integer z, and then $(p^{|u|})(n-m) = zq^{|u|}$, hence $n \equiv m\ [q^{|u|}]$.

(ii). Analogous to (i).

A direct consequence of this lemma is that given any two distinct words u and v of $L_{\frac{p}{q}}$, there exists a word w such that uw is in $L_{\frac{p}{q}}$ but vw is not. Hence, the set $\{u^{-1}L_{\frac{p}{q}} \mid u \in A_p^*\}$ of left quotients of $L_{\frac{p}{q}}$ is infinite, or equivalently:

Corollary 1. *The language $L_{\frac{p}{q}}$ is not rational.*

Definition 2 (The value set). *We denote by $V_{\frac{p}{q}}$ the set of numbers representable in base $\frac{p}{q}$, namely:*

$$V_{\frac{p}{q}} = \{x \mid \exists u \in A_p^*, \pi(u) = x\} \tag{4}$$

or equivalently $V_{\frac{p}{q}} = \pi\left(A_p^*\right)$

The most notable property of $V_{\frac{p}{q}}$ is that it is closed under addition, or more precisely that the addition is realised by a transducer, described in Section 5 (a full proof can be found in [1, Section 3.3]).

Secondly, from the definition of π, one derives easily that $V_{\frac{p}{q}} \subseteq \mathbb{Q}$. More precisely $V_{\frac{p}{q}}$ contains only numbers of the form $\frac{x}{y}$ where y divides a power of q, and conversely, for all k, $V_{\frac{p}{q}}$ contains almost every number $\frac{x}{q^k}$.

Lemma 2. *For every integer k, there exits an integer m_k such that, for every integer n greater than m_k, $\frac{n}{q^k}$ belongs to $V_{\frac{p}{q}}$.*

Proof. If $k = 0$, then one can take $m = 0$ since $\mathbb{N}$ is contained in $V_{\frac{p}{q}}$.

For $k \geqslant 1$, the words 1 and $1.0^{(k-1)}$ have for respective value $\frac{1}{q}$ and $\frac{p^{k-1}}{q^k}$. For every integer i and j, the number $(\frac{i\times p^{(k-1)} + j \times q^{(k-1)}}{q^k})$ is in $V_{\frac{p}{q}}$, since $V_{\frac{p}{q}}$ is closed under addition, and this can be rewritten as $(p^{(k-1)}\mathbb{N} + q^{(k-1)}\mathbb{N})\frac{1}{q^k} \subseteq V_{\frac{p}{q}}$. Since $p^{(k-1)}$ and $q^{(k-1)}$ are coprime, $(p^{(k-1)}\mathbb{N} + q^{(k-1)}\mathbb{N})$ ultimately covers $\mathbb{N}$.

Experimentally, the bound m_k is increasing with k but the expression resulting from this Lemma is far from being tight. As a consequence, it proves to be difficult to define $V_{\frac{p}{q}}$ without using the $\frac{p}{q}$-rational base number system.

4 BLIP Languages

In the previous section, an insight is given about why $L_{\frac{p}{q}}$ is not rational. It is additionally proven in [1] that $L_{\frac{p}{q}}$ is not context-free either. However, being context sensitive doesn't seem to accurately describe $L_{\frac{p}{q}}$. This section depicts a very strong language property, taylored to capture the structural complexity of $L_{\frac{p}{q}}$.

Let us first define a (very) weak iteration property for languages:

Definition 3. *A language L of A^* is said to be* left-iterable *if there exist two words u and v in A^* such that uv^i is a prefix of words in L for an infinite number of exponents i.*

Of course, every rational or context-free language is left-iterable. The definition is indeed designed above all for stating its negation.

Definition 4. *A language L which is not left-iterable is said to have* the Bounded Left-Iteration Property*, or, for short, to be* BLIP*.*

Example 1. A very simple way of building BLIP languages is to consider infinitely many prefixes of an infinite and aperiodic word. For instance the language $\{u_i\}$, where $u_0 = \varepsilon$ and $u_{i+1} = u_i.1.0^i$; or the language of the finite powers of the Fibonacci morphism $\{\sigma^i(0)\}$ where $\sigma(0) = 01$ and $\sigma(1) = 0$.

In order to build a less trivial example let us define the following family of functions f_i:

$$f_i \ : \ \begin{array}{l} n \mapsto n \text{ if } n \neq i \\ n \mapsto 0 \text{ if } n = i. \end{array}$$

The language $\{u_{i,j}\}$, where $u_{i,0} = 1$ and $u_{i,j+1} = u_{i,j}.1.0^{f_i(j)}$, is BLIP as can be easily checked.

Since Definition 4 was taylored for the study of $L_{\frac{p}{q}}$, the following holds, as essentially established in [1, Lemma 8].

Proposition 2. *The language $L_{\frac{p}{q}}$ is BLIP.*

Proof. If $L_{\frac{p}{q}}$ were left iterable, there would exist two nonempty words u and v such that uv^i is prefix of a word of $L_{\frac{p}{q}}$ for infinitely many i. Since $L_{\frac{p}{q}}$ is prefix-closed, the word uv^i would be itself in $L_{\frac{p}{q}}$, for all i. From Lemma 1, it follows that the integers $\pi(u)$ and $\pi(uv)$ are congruent modulo q^k, for all k, a contradiction.

Being BLIP is a very stable property for languages, as expressed by the following properties.

Lemma 3. *(i) Every finite language is BLIP.*
(ii) Any finite union of BLIP languages is BLIP.
(iii) Any intersection of BLIP languages is BLIP.
(iv) Any sublanguage of a BLIP language is BLIP.

Of course, BLIP languages are not closed under complementation, star or transposition.

The bounded left iteration property can be expressed with the more classical notion of IRS language (for Infinite Regular Subset) that has been introduced by Sheila Greibach in her study of the family of context-free languages ([4], *cf.* also [2]). A language is IRS if it does not contain any infinite rational sublanguage. For instance, the language $\{a^n \mid n \text{ is a prime number}\}$ is IRS (but not BLIP).

It is immediate that a BLIP language is IRS; even that a BLIP language contains no infinite context-free sublanguage. However the converse is not true as seen with the above example. More precisely, the following statement holds:

Proposition 3. *A language L is BLIP if, and only if, Pref(L) is IRS.*

Proof.

$$\begin{aligned}
\mathrm{Pref}(L) \text{ is not IRS} &\iff \mathrm{Pref}(L) \text{ contains a sublanguage of the form } u\,v^*w \\
&\iff uv^* \text{ is a sublanguage of } \mathrm{Pref}(L) \\
&\iff \text{for infinitely many } i,\ u\,v^i \text{ is prefix of a word of } L \\
&\iff L \text{ is not BLIP}
\end{aligned}$$

Proposition 3 shows that BLIP and IRS are equivalent properties on prefix-closed languages, which means that IRS is indeed a very strong property for prefix-closed languages.

Even though the purpose of this work is to prove Theorem 1, we actually prove a stronger version of it:

Theorem 2. *The set of the $\frac{p}{q}$-representations of any finitely generated additive submonoid of $V_{\frac{p}{q}}$ is a BLIP language.*

This is not a minor improvement, as it shows that every language representing a finitely generated monoid is basically as complex as $L_{\frac{p}{q}}$.

5 The Incrementer

The purpose of this section is to build a letter-to-letter sequential right transducer $A_p \to A_p$ realising a constant addition: given as *parameter* a word w of A_p^* it would perform the application $u \mapsto v$, such that $\pi(v) = \pi(u)+\pi(w)$. This transducer is based on the converter defined in [3] that we recall in Definition 5, below.

Theorem 3 ([1],[3]). *Given any digit alphabet A_n, there exists a finite letter-to-letter right sequential transducer $\mathcal{C}_{\frac{p}{q},n}$ from A_n to A_p such that for every w in ${A_n}^*$, $\pi\left(\mathcal{C}_{\frac{p}{q},n}(w)\right) = \pi(w)$.*

Definition 5. *For every integer n, the* converter $\mathcal{C}_{\frac{p}{q},n} = \langle\, \mathbb{N}, A_n, A_p, 0, \delta, \eta, \omega\, \rangle$, *is the right transducer with input alphabet A_n, output alphabet A_p, and whose transition and output functions are defined by:*

$$\forall s \in \mathbb{N}\,,\ \forall a \in A_n \qquad s \xrightarrow{a|c} s' \iff q\,s + a = p\,s' + c \quad ,$$

and final function by: $\omega(s) = \langle s \rangle_{\frac{p}{q}}$, for every state s in $\mathbb{N}$.

Definition 5 describes a transducer with an infinite number of states, but its reachable part is finite (cf [1, Proposition 13] or [3, Section 2.2.2]). In particular, if $n = 2p-1$, the converter is in fact an additioner: given two words $u = a_n \cdots a_2\, a_1$ and $v = b_n \cdots b_2\, b_1$ over A_p, the digit-wise addition yields the word $(a_n + b_n) \cdots (a_1 + b_1)$ over A_{2p-1} which is transformed by $\mathcal{C}_{\frac{p}{q},2p-1}$ into $\langle \pi(u) + \pi(v) \rangle_{\frac{p}{q}}$. The converter from A_5 to A_3 in base $\frac{3}{2}$ is shown at Figure 2.

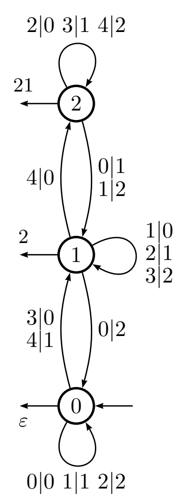

Fig. 2: The converter $\mathcal{C}_{\frac{3}{2},5}$

For every word w of A_p^*, we define a letter-to-letter sequential right transducer $\mathcal{R}_{\frac{p}{q},w}$ which increments the input by w, that is, given a word u as input, it outputs the $\frac{p}{q}$-representation $\langle \pi(u) + \pi(w) \rangle_{\frac{p}{q}}$. It is obtained as a specialisation of $\mathcal{C}_{\frac{p}{q},2p-1}$.

Definition 6. *For every $w = b_{n-1} \cdots b_1\, b_0$ in A_p^*, the* incrementer

$$\mathcal{R}_{\frac{p}{q},w} = \langle\, \mathbb{N} \times \{0, 1, \ldots, n\}, A_p, A_p, (0,0), \delta', \eta', \psi\, \rangle$$

is the (right) transducer with input and output alphabet A_p, and whose transition and output functions are defined by:

$$\forall s \in \mathbb{N}\,,\ \forall a \in A_p\,,$$

$$\forall i < n \qquad (s,i) \xrightarrow{a|c} (s', i+1) \quad \Longleftrightarrow q\,s + (a + b_i) = p\,s' + c$$

$$(s,n) \xrightarrow{a|c} (s', n) \quad \Longleftrightarrow \quad q\,s + a = p\,s' + c$$

and whose final function is defined by:

$$\forall s \in \mathbb{N} \qquad \psi((s,n)) = \langle s \rangle_{\frac{p}{q}}\,,$$

$$\psi((s,i)) = \psi((s', i+1)).c \quad \textit{if} \quad i < n \quad \textit{and} \quad (s,i) \xrightarrow{0|c} (s', i+1)$$

This last line means that if the input word is shorter than w, then the final function behaves as if the input word ended with enough 0's (on the left, since we read from right to left). Definition 6 describes a transducer with an infinite number of states but, as in the case of the converter, it is easy to verify that its reachable part is finite. The incrementer $\mathcal{R}_{\frac{3}{2},121}$ is shown at Figure 3.

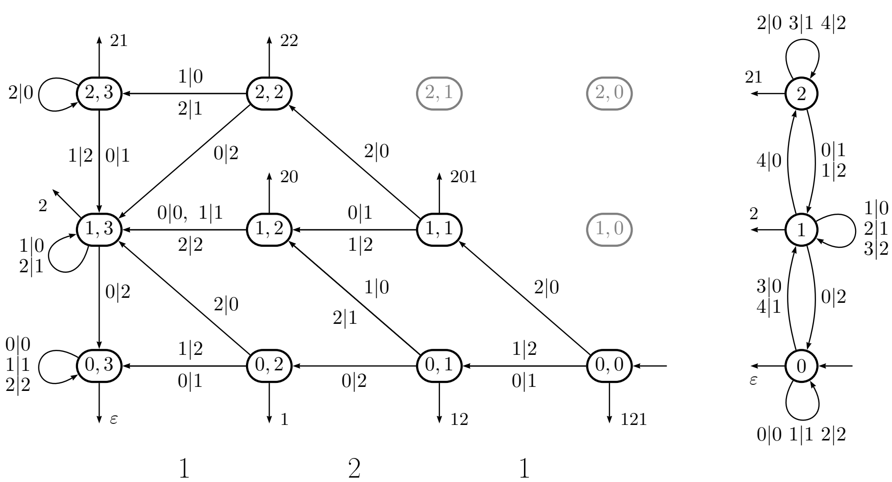

Fig. 3: The incrementer $\mathcal{R}_{\frac{3}{2},121}$

It is a simple verification that the incrementer has the expected behaviour.

Proposition 4. *For every u and w in A_p^*, $v = \mathcal{R}_{\frac{p}{q},w}(u)$ is a word in A_p^* such that $\pi(v) = \pi(u) + \pi(w)$ holds.*

6 Proof of Theorem 2

The core of the proof lies in the next statement.

Proposition 5. *For every w in A_p^*, the image of a left-iterable language by $\mathcal{R}_{\frac{p}{q},w}$ is left-iterable.*

Proof. Let u and v be in A_p^*, $I \subseteq \mathbb{N}$ an infinite set of indexes and $\{y_i\}_{i\in I}$ an infinite family of words in A_p^*. The proof consists in showing that $\left\{\mathcal{R}_{\frac{p}{q},w}(u\,v^i y_i) \,\middle|\, i \in I\right\}$ is left-iterable.

Since I is infinite, we may assume, without loss of generality, that the length of the y_i's is strictly increasing hence, that all y_i's have a length greater than $n = |w|$ but also that the reading of every y_i leads $\mathcal{R}_{\frac{p}{q},w}$ to a *same* state $(s, 0)$:

$$\forall s \in \mathbb{N}\,,\ \forall i \in I \qquad (0, n) \xrightarrow[\mathcal{R}_{\frac{p}{q},w}]{y_i|y'_i} (s, 0) \ .$$

From the definition of the transitions of $\mathcal{R}_{\frac{p}{q},w}$:

$$(s, 0) \xrightarrow{a|c} (s', 0) \qquad \Longleftrightarrow \qquad q\,s + a = p\,s' + c \ ,$$

follows, since $a < p$ and $q < p$, that $s \geqslant s'$. Hence, the sequence of (first component of) states of $\mathcal{R}_{\frac{p}{q},w}$ in a computation starting in $(s, 0)$ and with input v^i, with unbounded i, is ultimately stationary at state $(t, 0)$.

Without loss of generality, we thus may assume that $(0, n) \xrightarrow{y_i|y'_i} (t, 0)$ for every i in I and, since $(t, 0) \xrightarrow{v|v'} (t, 0)$, it holds that $\mathcal{R}_{\frac{p}{q},w}(u\,v^i y_i) = u'\,v'^i y'_i$, where u' is the output of a computation starting in $(t, 0)$ and with input u.

The special case of additive submonoids of $V_{\frac{p}{q}}$ allows us to reverse the condition from left-iterable to BLIP:

Proposition 6. *Let w be a word of A_p^*, and L be a BLIP language such that $\pi(L)$ is an additive submonoid of $V_{\frac{p}{q}}$. The language $\mathcal{R}_{\frac{p}{q},w}(L)$ is BLIP.*

Proof. Since $\pi(L)$ is an additive submonoid of $V_{\frac{p}{q}}$, it contains $m\mathbb{N}$ for some m (as it must contains some number $\frac{m}{q^l}$ for some m and l).

Let n and k be the integers such that $\pi(w) = \frac{n}{q^k} = x$. From Lemma 2, it follows that there exists m_k such that for every $j > m_k$, $\frac{j}{q_k}$ is in $V_{\frac{p}{q}}$. In particular, there exists j such that $n + j \equiv 0 \bmod (m\,q^k)$ and $\frac{j}{q^k}$ is in $V_{\frac{p}{q}}$. If we denote by $y = \frac{j}{q^k}$, it means that $(x + y)$ is in $m\mathbb{N}$. Hence, $\pi(L) + x + y$ is contained in $\pi(L)$.

Let us denote by $u = \langle y \rangle_{\frac{p}{q}}$, and $L' = \mathcal{R}_{\frac{p}{q},w}(L)$.

It follows that $\pi\left(\mathcal{R}_{\frac{p}{q},u}(L')\right) = (\pi(L)+x+y) \subseteq \pi(L)$, hence that $\mathcal{R}_{\frac{p}{q},u}(L')$ is an infinite subset of L, and as such BLIP (from Lemma 3). If L' were left-iterable, so would be $\mathcal{R}_{\frac{p}{q},u}(L')$ by Proposition 5, a contradiction.

Finally we prove a property of finitely generated submonoids of $V_{\frac{p}{q}}$.

Proposition 7. *Let M be a finitely generated additive submonoid of $V_{\frac{p}{q}}$. There exists a finite family $\{g_i\}_{i\in I}$ of elements of $V_{\frac{p}{q}}$ such that M is contained in $\bigcup_{i\in I}(g_i+\mathbb{N})$.*

Proof. Let $\{y_1, y_2, \ldots, y_h\}$ be a generating family of M. Every y_j is in $V_{\frac{p}{q}}$ and it is then a rational number $\frac{n_j}{q^{k_j}}$ for some integers n_j and k_j. Let k be the largest of the k_j. Hence, every element in M is a rational number whose denominator is a divisor of q^k, and thus $M \subseteq V_{\frac{p}{q}} \cap \left(\frac{1}{q^k}\mathbb{N}\right)$.

Since every number in $\frac{1}{q^k}\mathbb{N}$ can be written as $n + \frac{i}{q^k}$ for some n in $\mathbb{N}$ and some i in $\{0, 1, \ldots, q^k - 1\}$, it follows that $\frac{1}{q^k}\mathbb{N} = \bigcup_{0\leqslant i<q^k}(\mathbb{N} + \frac{i}{q^k})$, hence $M \subseteq \bigcup_{0\leqslant i<q^k}(V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{i}{q^k}))$. Besides, for every i in $\{0, 1, \ldots, q^k - 1\}$, we denote by g_i the smallest number in $V_{\frac{p}{q}} \cap (\mathbb{N}+\frac{i}{q^k})$. Then, and since $V_{\frac{p}{q}}+\mathbb{N} = V_{\frac{p}{q}}$, for every i, $V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{i}{q^k}) = m_i + \mathbb{N}$. Hence $M \subseteq \bigcup_{0\leqslant i<q^k}(\mathbb{N} + m_i)$.

Even though this proposition seems rather weak (it is a poor approximation from above), it is enough: it indeed reduces Theorem 2 to proving that $\langle n+\mathbb{N}\rangle$ (or equivalently $\mathcal{R}_{\frac{p}{q},w}(L_{\frac{p}{q}})$) is BLIP for any n, which was proven in Proposition 6.

Proof (of Theorem 2). Let M be a finitely generated additive submonoid of $V_{\frac{p}{q}}$. By Proposition 7, there exists a finite family $\{m_i\}_{i\in I}$ of elements of $V_{\frac{p}{q}}$ such that $M \subseteq \bigcup_{i\in I}(m_i + \mathbb{N})$.

Let $L = \langle M\rangle_{\frac{p}{q}}$ the language of the $\frac{p}{q}$-representations of the elements of M and write $w_i = \langle m_i\rangle_{\frac{p}{q}}$. Hence, L is contained in $(\bigcup_i \mathcal{R}_{\frac{p}{q},w_i}(L_{\frac{p}{q}}))$, and thus BLIP by Lemma 3.

7 Conclusion and Future Work

In this work, we have defined a new property, in an effort to capture the structural complexity of $L_{\frac{p}{q}}$. This property contradicts any form of pumping lemma, placing $L_{\frac{p}{q}}$ outside the scope of classical language theory. Even more so that every other example of BLIP languages we describe seem to be purely artificial (*cf.* Example 1)

Paradoxically, Theorem 2 shows that such examples are very common within a rational base number system. It seems that every reasonable number set is represented by a BLIP language and that every simple language represents a complicated set of numbers.

This work led us to a conjecture about rational approximations of $L_{\frac{p}{q}}$:

Conjecture 1. Let L be a rational language closed by addition and containing $L_{\frac{p}{q}}$. Then L contains $X.A_p^*$ where $X = L_{\frac{p}{q}} \cap A_p^{\leqslant k}$, for some k.

Any approximation of $L_{\frac{p}{q}}$ by a rational language L, would only keep a finite part of the structure: the automaton accepting L would be the subtree of depth k of $L_{\frac{p}{q}}$ whose leaves are all-accepting states. Figure 4 gives two examples of rational approximation of $L_{\frac{3}{2}}$, respectively when the $L_{\frac{p}{q}}$ is cut at depth $k = 2$ and $k = 5$.

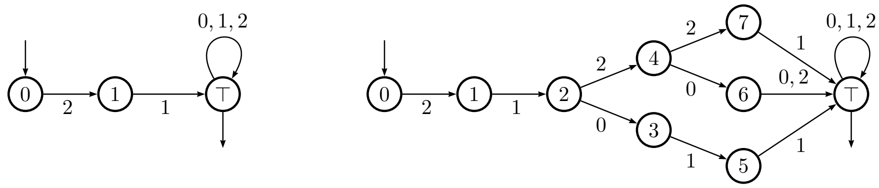

Fig. 4: Two rational approximations of $L_{\frac{3}{2}}$

References

1. Shigeki Akiyama, Christiane Frougny, and Jacques Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math.*, 168:53–91, 2008.
2. Jean-Michel Autebert, Joffroy Beauquier, Luc Boasson, and Michel Latteux. Indécidabilité de la condition IRS. *ITA*, 16(2):129–138, 1982.
3. Christiane Frougny and Jacques Sakarovitch. Number representation and finite automata. in *Combinatorics, Automata and Number Theory*, V. Berthé, M. Rigo (Eds), Encyclopedia of Mathematics and its Applications 135, Cambridge Univ. Press (2010) 34–107.
4. Sheila A. Greibach. One counter languages and the IRS condition. *J. Comput. Syst. Sci.*, 10(2):237–247, 1975.
5. John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 2000.
6. Pierre Lecomte and Michel Rigo. Abstract numeration systems. in *Combinatorics, Automata and Number Theory*, V. Berthé, M. Rigo (Eds), Encyclopedia of Mathematics and its Applications 135, Cambridge Univ. Press (2010) 108–162.
7. M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
8. Jacques Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, 2009. Corrected English translation of *Éléments de théorie des automates*, Vuibert, 2003.