On sets of numbers rationally represented in a rational base number system

V. Marsault∗ and J. Sakarovitch†

August 28, 2012

Abstract

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.

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 \( A_p = \{0, 1, \cdots (p - 1)\} \) which do not begin with a 0. This set \( L_p = (A_p \setminus \{0\})A_p^* \) is rational, that is, recognised 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 can be done by a finite transducer called the converter.

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 (cf. [3]), even if it is not the case that addition can be realised by a finite automaton.

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 (not even a context-free one), but nevertheless the addition is realised by a finite automaton. The purpose of this work is to further investigate the relationship between these two features in the case of rational base numeration systems.

Let \( p \) and \( q \) be two coprime integers, with \( p > q \). In the \( \frac{p}{q} \)-numeration system, the digit alphabet is \( A_p \), 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 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 obviously not finitely generated (as an additive monoid).

In this note 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.

∗Corresponding author, LTCI, Telecom ParisTech
†LTCI, CNRS / Telecom ParisTech
This statement relies on three properties: the first 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 second is the description of an iteration property that is not satisfied by the language $L_{\frac{p}{q}}$ of $\frac{p}{q}$-representations of the integers; and finally the construction of a letter-to-letter sequential right transducer that realises on the $\frac{p}{q}$-representations of numbers the addition of a fixed value to the elements of $V_{\frac{p}{q}}$.

1 Finitely generated additive submonoids

**Proposition 2.** Let $M$ be a finitely generated additive submonoid of $V_{\frac{p}{q}}$. There exists a finite family $\{m_i\}_{i \in I}$ of elements of $V_{\frac{p}{q}}$ such that $M$ is contained in $\bigcup_{i \in I}(m_i + \mathbb{N})$.

**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 $\frac{n_j}{q_j}$ 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 $M \subseteq V_{\frac{p}{q}} \cap \left(\frac{1}{q} \mathbb{N}\right)$.

Since every number in $\frac{n}{q} \mathbb{N}$ can be written as $n + \frac{i}{q}$ with $n$ in $\mathbb{N}$ and $i$ in $\{0, 1, \ldots, q^k - 1\}$, then $\frac{1}{q} \mathbb{N} = \bigcup_{0 \leq i < q^k} (\mathbb{N} + \frac{i}{q})$, and then $M \subseteq \bigcup_{0 \leq i < q^k} (V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{i}{q}))$ holds. For every $i$ in $\{0, 1, \ldots, q^k - 1\}$, let $m_i$ be the smallest number in $V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{i}{q})$. Then, and since $V_{\frac{p}{q}} + \mathbb{N} = V_{\frac{p}{q}}$, for every $i$, $V_{\frac{p}{q}} \cap (\mathbb{N} + \frac{i}{q}) = m_i + \mathbb{N}$. Hence $M \subseteq \bigcup_{0 \leq i < q^k} (\mathbb{N} + m_i)$.

Another property of the set $V_{\frac{p}{q}}$ will be useful later on.

**Lemma 3.** For every integer $k$, there exists an integer $m$ such that, for every integer $n$ greater than $m$, $\frac{n}{q^k}$ 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 $k \geq 1$, the words "1" and "$10^{k-1}$" have for respective value $\frac{1}{q}$ and $\frac{q^{k-1}}{q^k}$. 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 $\frac{p^{k-1}N + q^{k-1}N}{q^k} \subseteq V_{\frac{p}{q}}$. Since $p^{k-1}$ and $q^{k-1}$ are coprime integers, $(p^{k-1}N + q^{k-1}N)$ ultimately covers $\mathbb{N}$.

2 BLIP languages

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

**Definition 4.** A language $L$ of $A^*$ is said to be left-iterable if there exists 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 5.** A language $L$ which is not left-iterable if said to have the bounded left-iteration property, or, for short, to be BLIP.

This definition is almost tailor-made for the study of $L_{\frac{p}{q}}$ since the following holds, as essentially established in [1], Lemma 8].
Proposition 6. The language $L_{\frac{p}{q}}$ is BLIP.

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

Property 7.

(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.

3 The incrementer

Proposition 2 gives the intuition that studying finitely generated submonoids of $V_{\frac{p}{q}}$ is closely linked to the operation $n \mapsto n + x$ where $x$ is a fixed value in $V_{\frac{p}{q}}$ and $n$ any integer. The incrementation by a constant value $x$ is realised by a finite transducer. This construction is based on the converter defined in [2].

Theorem 8 ([1], [2]). Given any digit alphabet $A_n$, there exists a finite letter-to-letter right sequential transducer $C_{\frac{p}{q},n}$ from $A_n$ to $A_p$ such that for every $w$ in $A_n^*$, $\pi(C_{\frac{p}{q},n}(w)) = \pi(w)$.

Definition 9. For every integer $n$, the converter $C_{\frac{p}{q},n} = \langle N, A_n, A_p, 0, \delta, \omega \rangle$, is the (right) transducer with input alphabet $A_n$, output alphabet $A_p$, and whose transition function is defined by:

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

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

Definition 9 describes a transducer with an infinite number of states, but its reachable part is finite (cf [1, Proposition 13] or [2, section 2.2.2]). In particular, if $n = 2p - 1$, the converter is in fact an additioner: given two words $a = a_n \cdots a_1$ and $b = b_n \cdots b_1$ 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 $C_{\frac{p}{q},2p-1}$ into $\langle \pi(a) + \pi(b) \rangle_{\frac{p}{q}}$.

Example 10. The converter from $A_5$ to $A_3$ in base $\frac{3}{2}$ is shown at Figure 1.

For every word $w$ of $A_p^*$, we define a transducer $I_{\frac{p}{q},w}$ which increments the input by $w$, i.e. 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 $C_{\frac{p}{q},2p-1}$.
Definition 11. For every $w = b_1 b_2 \cdots b_n$ in $A_p^*$, the incrementer
$$I_{p,w} = \langle \mathbb{N} \times \{0, \ldots, n\}, A_p, A_p, (0, n), \eta, \psi \rangle$$
is the (right) transducer with input and output alphabet $A_p$, and whose transition function is defined by:
$$\forall s \in \mathbb{N}, \forall a \in A_p, (s, i) \xrightarrow{a|c} (s', i - 1) \iff qs + (a + b_i) = ps' + c \quad \text{if} \quad i > 0$$
$$\quad (s, 0) \xrightarrow{a|c} (s', 0) \iff qs + a = ps' + c$$
and whose final function is defined by:
$$\forall s \in \mathbb{N} \quad \psi((s, 0)) = \langle s \rangle_{\#},$$
$$\psi((s, i)) = \psi((s', i - 1))c \quad \text{for} \quad i > 0 \quad \text{and} \quad (s, i) \xrightarrow{0|c} (s', i - 1).$$

This last line means that if the input word is smaller than $w$, then the final function behaves as if the input word ended with enough 0's. Definition 11 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. It is also a simple verification that the incrementer has the expected behaviour.

Proposition 12. For every $u$ and $w$ in $A_p^*$, $v = I_{p,w}(u)$ is a word in $A_p^*$ such that $\pi(v) = \pi(u) + \pi(w)$ holds.

Example 13. The incrementer $I_{2,121}$ is shown at Figure 2.

The next proposition is the core of the proof of Theorem 1. It strongly links the left-iterable property with the incrementer.

Proposition 14. For every $w$ in $A_p^*$, the image of a left-iterable language by $I_{p,w}$ is left-iterable.

Proof. Let $u$ and $v$ be in $A_p^*$, $I \subseteq \mathbb{N}$ an infinite set of indices and $\{y_i\}_{i \in I}$ an infinite sequence of words in $A_p^*$. The proof consists in showing that $\{I_{p,w}(uv^iy_i) \mid i \in I\}$ 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 and thus not only that all $y_i$’s have a length greater than $n = |w|$ but also that the reading of every $y_i$ leads $I_{e,w}$ to a same state $(s, 0)$:

$$
\forall s \in \mathbb{N}, \forall i \in I \quad (0, n) \quad \xrightarrow{y_i y'_i} (s, 0).
$$

From the definition of the transitions of $I_{e,w}$:

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

follows, since $a < p$ and $q < p$, that $s \geq s'$. Hence the sequence of (first component of) states of $I_{e,w}$ 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 $(t, 0) \xrightarrow{v^i v'} (t, 0)$, it holds $I_{e,w}(u v^i y_i) = u' v^i y'_i$, where $u'$ is the output of a computation starting in $(t, 0)$ and with input $u$.

In order to complete the proof, we deduce from Proposition 14 a stronger property.

**Proposition 15.** For every $w$ in $A^*_p$, the language $L = I_{e,w}(L_{e,w})$ is BLIP.

**Proof.** Let $n$ and $k$ be the integers such that $\pi(w) = \frac{n}{k} = x$. By Lemma 3 there exists $j$ such that $n + j \equiv 0 \mod k$ and $y = \frac{j}{k}$ is in $V_{e,w}$. Hence, $N + x + y$ is contained in $\mathbb{N}$. Write $u = \langle y \rangle_{e,w}$.

Then, $\pi(I_{e,w}(L)) = (N + x + y)$ and $I_{e,w}(L)$ is an infinite subset of $L_{e,w}$. If $L$ were left-iterable, so would be $I_{e,w}(L)$ by Proposition 14, a contradiction.

**Proof of Theorem 7.** Let $M$ be a finitely generated additive submonoid of $V_{e,w}$. By Proposition 2 there exists a finite family $\{m_i\}_{i \in I}$ of elements of $V_{e,w}$ 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 $\bigcup_i I_{\frac{p}{q}} I_{\frac{p}{q}} (L_{\frac{p}{q}})$, and thus BLIP by Property 7, and thus certainly not rational.

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] 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.

[3] 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.

[4] Jacques Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, 2009. Corrected English translation of *Éléments de théorie des automates*, Vuibert, 2003.