Auto-Similarity in Rational Base Number Systems

Shigeki Akiyama[1] , Victor Marsault[2⋆], and Jacques Sakarovitch[2]

[1] University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 350-8571 Japan
[2] Telecom-ParisTech and CNRS, 46 rue Barrault, 75013 Paris, France

Abstract. This work is a contribution to the study of set of the representations of integers in a rational base number system. This prefix-closed subset of the free monoid is naturally represented as a highly non regular tree whose nodes are the integers and whose subtrees are all distinct. With every node of that tree is then associated a minimal infinite word (and a maximal infinite word).
The main result is that a sequential transducer which computes for all n the minimal word associated with $n+1$ from the one associated with n, has essentially the same underlying graph as the tree itself.
These infinite words are then interpreted as representations of real numbers; the difference between the numbers represented by the maximal and minimal word associated with n is called the span of n. The preceding construction allows to characterise the topological closure of the set of spans.

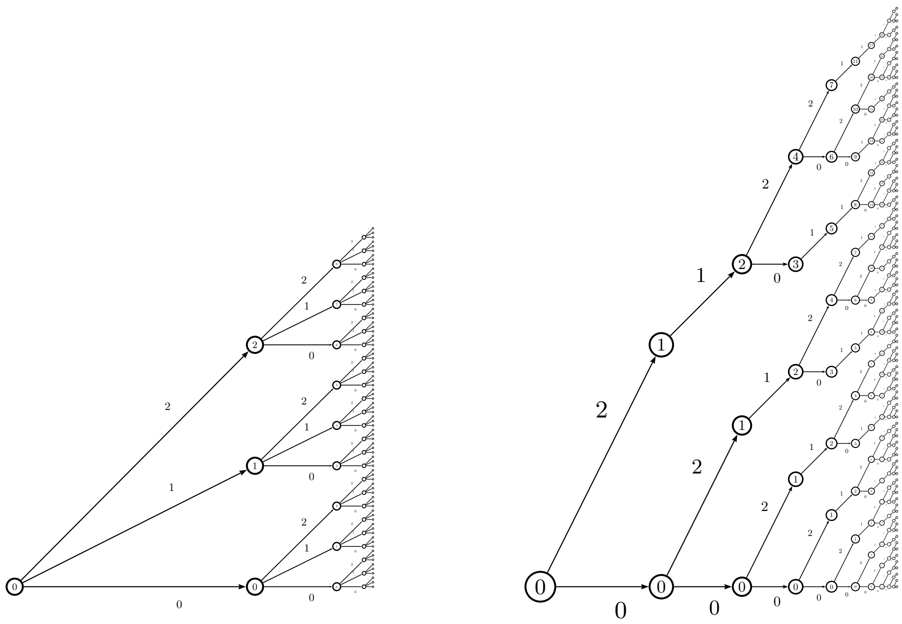
1 Introduction

The purpose of this work is a further exploration and a better understanding of the set of *words* that represent integers in a rational base number systems. These numeration systems have been introduced and studied in [1], leading to some progress in the results around the so-called Malher's problem (*cf.* [5]). We give below a precise definition of rational base number systems and of the representation of numbers in such a system. But one can hint at the results established in this paper by just looking at the figure showing the 'representation tree' of the integers – that is, the compact way of describing the words that represent the integers – in a rational base number system (Figure 1b for the base $\frac{3}{2}$) and by comparison with the representation tree (or trie) in an analogous integer base number system (Figure 1a for the base 3).

In the latter, all subtrees are the same and equal to the full ternary tree, whereas in the former, all subtrees are different. As a result, the language of the representations of the integers is not a regular language. It may even be shown that the language satisfies no iteration lemma of any kind ([6]). With the hope of finding some order or regularity within what seems to be closer to complete randomness (which, on the other hand, is not established either) we consider the *minimal words* originating from every node of the tree.

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

In the case of an integer base, this is meaningless: all these minimal words are equal to 0^ω. In the case of a rational base these words are on the contrary all distinct and none are even ultimately periodic (as no ultimately periodic word can be found in this tree). In order to find some invariant of all these distinct words, or at least a relationship between them, we have studied the function that maps the minimal word w_n^- associated with n onto the one associated with $n+1$, and tried to describe this function by a (possibly infinite) transducer.



(a) Integer base 3 (b) Rational base $\frac{3}{2}$

Fig. 1: Representation trees in two number systems

The computation of such a transducer in the case of the base $\frac{3}{2}$, and more generally in the case of a base $\frac{p}{q}$ with $p = 2q - 1$, leads to a surprising and unexpected result. If $\mathcal{T}_{\frac{p}{q}}$ denotes the representation tree – viewed as an infinite automaton, – the transducer, denoted by $\mathcal{D}_{\frac{p}{q}}$, is obtained by replacing the label of every transition of $\mathcal{T}_{\frac{p}{q}}$ by a set of pairs of letters that depends upon this label only. In other words, the *underlying graphs* of $\mathcal{T}_{\frac{p}{q}}$ and $\mathcal{D}_{\frac{p}{q}}$ *coincide*, and $\mathcal{D}_{\frac{p}{q}}$ is obtained from $\mathcal{T}_{\frac{p}{q}}$ by a *substitution* from the alphabet of digits into the alphabet of pairs of digits, in this special and remarkable case.

The general case is hardly more difficult to describe, once it has been understood. Let $B_{p,q}$ be the digit alphabet with $2q - 1$ (consecutive) elementsand whose greatest element is $p-1$. If $p>2q-1$, then $B_{p,q}$ is contained in A_p; it consists of A_p, enlarged with enough *negative* digits otherwise.

From $\mathcal{T}_{\frac{p}{q}}$ and with the digit alphabet $B_{p,q}$, we first define another 'automaton' denoted by $\widehat{\mathcal{T}_{\frac{p}{q}}}$: either by *deleting the transitions* of $\mathcal{T}_{\frac{p}{q}}$ whose labels do not belong to $B_{p,q}$ in the case where $p>2q-1$ or, in the case where $p<2q-1$ by *adding transitions* labelled with the new negative digits. Then, $\mathcal{D}_{\frac{p}{q}}$ is obtained from $\widehat{\mathcal{T}_{\frac{p}{q}}}$ exactly as above, by a *substitution* from the alphabet of digits into the alphabet of pairs of digits. This construction of $\mathcal{D}_{\frac{p}{q}}$, which we call the *derived transducer*, and the proof of its correctness are presented in Section 3. In the following Section 4, we turn to a problem seems to be of different nature.

In [1], the tree $\mathcal{T}_{\frac{p}{q}}$, which is built from the representations of integers, is used to *define* the representations of real numbers: the label of an infinite branch of the tree is the development 'after the decimal point' of a real number and the drawing of the tree as a fractal object — like in Figure 1b — is fully justified by this point of view. The same idea leads to the definition of the (renormalised) *span* of a node n of the representation tree: it is the difference between the reals represented respectively by the maximal and the minimal words originating in the node n (see Remark 1, page 11).

Again, this notion is meaningless in the case of an integer base p: the span of node n is always 1. And again, the notion is far more richer and complex in the case of a rational base $\frac{p}{q}$. The trivial relationship between the minimal word originating at node $n+1$ and the maximal word originating at node n leads to the connexion between the construction of the derived transducer $\mathcal{D}_{\frac{p}{q}}$ and the description of the set of spans $\mathtt{S}_{\frac{p}{q}}$. Not only the *digit-wise difference* between maximal and minimal words is written on the alphabet $B_{p,q}$, but all these 'difference words' are infinite branches in the tree $\widehat{\mathcal{T}_{\frac{p}{q}}}$. This is explained in Section 4. From the structure of $\widehat{\mathcal{T}_{\frac{p}{q}}}$, it then follows (Theorem 3) that the topological closure of $\mathtt{S}_{\frac{p}{q}}$ is an interval in the case where $p\leqslant 2q-1$, and a set with empty interior in the case where $p>2q-1$.

In conclusion, we have shown that a straightforward computation of w_{n+1}^{-} from w_n^{-} requires the same structure as $\mathcal{T}_{\frac{p}{q}}$ itself – despite the fact that every minimal word looks as complex as the whole tree – whether it be performed directly on the words, or indirectly via the span of the nodes. It is this phenomenon that we call *auto-similarity* of the structure $\mathcal{T}_{\frac{p}{q}}$. In this process, the number systems where $p=2q-1$ appear to mark the boundary between two different behaviours, in a more deeper way than that was described in the first study of rational base number systems [1].

This paper is meant to be self-contained and gives, in particular, all necessary definitions concerning rational base number systems. However, the reference [1] where these systems have been defined and the sets of representations first studied will probably be useful. In order to meet the space constraints, all proofs and even some figures have been removed. The reader may find them in a complete version downloadable from arXiv [2].

2 Preliminaries and Notations

2.1 Numbers and Words

Given two *real numbers* x and y, we denote by $\frac{x}{y}$ their division in $\mathbb{R}$ (even if x or y happens to be integers), by $[x, y]$ the corresponding interval of $\mathbb{R}$ and by $\lceil x \rceil$ the integer n such that $(n-1) < x \leqslant n$. On the other hand, given two *positive integers* n and m, we denote by $n \div m$ and $n \,\%\, m$ respectively the quotient and the remainder of the Euclidean division of n by m, that is, $n = (n \div m)\, m + (n \,\%\, m)$ and $0 \leqslant (n \,\%\, m) < m$. Additionally, we denote by $[\![n, m]\!]$ the integer interval $\{n, (n+1), \ldots, m\}$.

An *alphabet* is a finite set of symbols called *letters* or *digits* when they are integers. Given an alphabet A, we consider both the sets of *finite* and *infinite* words over A respectively denoted by A^* and A^ω and we denote the *empty word* by ε. For every positive integer p, we denote by A_p the canonical digit alphabet of the base p number system: $A_p = \{0, 1, \ldots, p-1\}$. For clarity, we denote finite words by u, v and infinite words by w. The concatenation of two words u, v is either explicitly denoted by a low dot, as in $u.v$, or implicitly when there is no ambiguity, as in uv. A finite word u is said to be a prefix of a finite word v (resp. an infinite word w) if there exists a finite word v' (resp. an infinite word w') such that $v = uv'$ (resp. $w = uw'$). The set of subsets of an alphabet A is denoted by $\mathfrak{P}(A)$.

2.2 Automata and Transducers

We deal here with a very special class of automata and transducers only: they are infinite, their state set is $\mathbb{N}$, they are *deterministic* (or *letter-to-letter* and *sequential*), the initial state is 0, and all states are final.

As usual, an *automaton* $\mathcal{X}$ over A is denoted by a 5-tuple $\mathcal{X} = \langle\, \mathbb{N}, A, \delta, 0, \mathbb{N}\,\rangle$, where $\delta : \mathbb{N} \times A \to \mathbb{N}$ is the *transition function*. The partial function δ is extended to $\mathbb{N} \times A^*$, and $\delta(n, u) = m$ is also denoted by $n \cdot u = m$ or by $n \xrightarrow{u} m$. Given an integer n, every state $n \cdot a$ for some a in A is called a *successor* of n. A word u in A^* (resp. a word w in A^ω) is *accepted* by $\mathcal{X}$ if $0 \cdot u$ exists (resp. if $0 \cdot v$ exists for every finite prefix v of w). The *language* of *finite* words (resp. of *infinite* words) accepted by $\mathcal{X}$ is denoted by $L(\mathcal{X})$ (resp. by $\mathcal{L}(\mathcal{X})$).

For transducers, we essentially use the notation of [3], adapted for the infinite case. A *transducer* is an automaton whose transitions are labelled by (set of) pairs of letters. Formally, it is represented by a tuple $\mathcal{Y} = \langle\, \mathbb{N}, A \times B, \delta, \eta, 0, \mathbb{N}\,\rangle$ where $\langle\, \mathbb{N}, A, \delta, 0, \mathbb{N}\,\rangle$ is an automaton, called *the underlying input automaton* of $\mathcal{Y}$, A is called *the input alphabet*, B is the *output alphabet* and $\eta \colon \mathbb{N} \times A \to B$ is the *output function*. The transition function δ is extended as in automata, and η is extended to $\mathbb{N} \times A^* \to B^*$ by $\eta(n, \varepsilon) = \varepsilon$ and $\eta(n, ua) = \eta(n, u).\eta(n \cdot u, a)$, and $\eta(n, u)$ is also denoted by $n * u$ for short.

Moreover, given two finite words u and v, we denote by $n \xrightarrow{u\,|\,v} m$ the combination of $n \cdot u = m$ and $n * u = v$. We say that the *image* of a finite word u by $\mathcal{Y}$, denoted by $\mathcal{Y}(u)$, is the word v, if it exists, such that $0 \xrightarrow{u\,|\,v} k$

for some k. Similarly, the image of the infinite word w is w' if, for every finite prefix u of w, $\mathcal{Y}(u)$ is a prefix of w'.

2.3 Rational Base Number System

Let p and q be two co-prime integers such that $p > q > 1$. Given a positive ieger N, let us write $N_0 = N$ and define the sequence $(N_i)_{i\in\mathbb{N}}$ by all $i > 0$,

$$q\,N_i = p\,N_{(i+1)} + a_i \qquad \text{for all } i > 0 \ \ ,$$

where a_i is the remainder of the Euclidean division of $q\,\mathbb{N}_i$ by p, hence in the alphabet $A_p = [\![0, p-1]\!]$. Since $p > q$, the sequence $(N_i)_{i\in\mathbb{N}}$ is strictly decreasing and eventually stops at $N_{k+1} = 0$. Moreover, it holds that

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

The *evaluation map* π is derived from this formula. Given a word $a_n a_{n-1} \cdots a_0$ over A_p, and indeed over any alphabet of digits, its *value* is defined by

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

Conversely, a word u in A_p^* is called a $\frac{p}{q}$*-representation* of an integer 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 can be computed with the modified Euclidean division algorithm above. By convention, the representation of 0 is the empty word ε. The set of all $\frac{p}{q}$-representations of integers is denoted by $L_{\frac{p}{q}}$:

$$L_{\frac{p}{q}} = \left\{ \langle n\rangle_{\frac{p}{q}} \;\middle|\; n \in \mathbb{N} \right\} \ \ .$$

It should be noted that a rational base number system is *not* a β-numeration — where the representation of a number is computed by the (greedy) Rényi algorithm (cf. [4, Chapter 7]) — in the special case where β is a rational number. In such a system, 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 the rational base number system, they are $\{0, 1\ \ldots (p-1)\}$ and $\frac{1}{q}(\frac{p}{q})^i$ respectively.

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 right-extendable (for every representation $\langle n\rangle$, there exists (at least) an a in A_p 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 an infinite tree, or 'trie' (cf. Figure 2).

It is known that $L_{\frac{p}{q}}$ is not rational (not even context-free), and the following automaton (accepting indeed the language $0^* L_{\frac{p}{q}}$) is infinite.

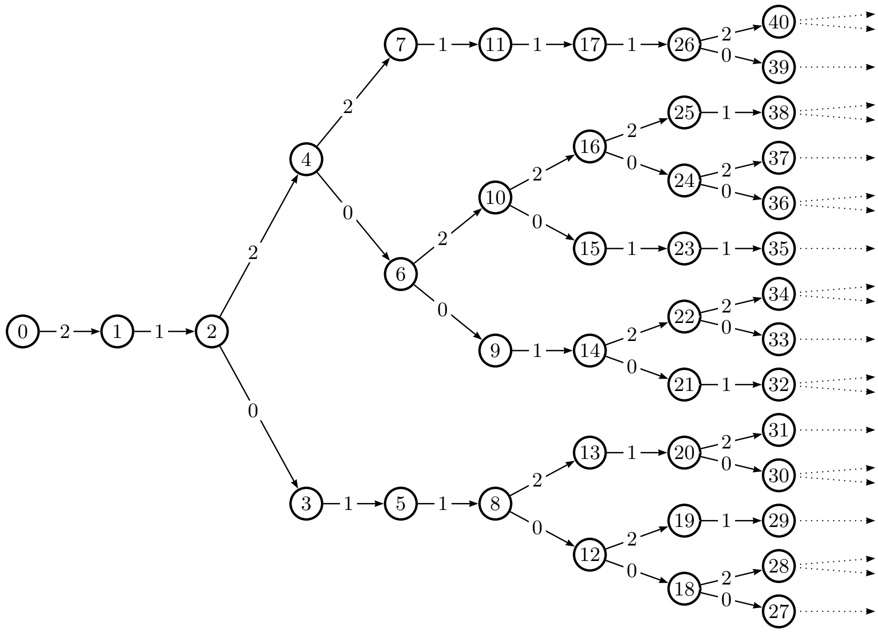


Fig. 2: The tree representation of the language $L_{\frac{3}{2}}$

Definition 1. *Let $\tau_{\frac{p}{q}} : \mathbb{N} \times \mathbb{Z} \to \mathbb{N}$ be the (partial) function defined[3] by:*

$$\forall n \in \mathbb{N}\,,\ \forall a \in \mathbb{Z} \qquad \tau_{\frac{p}{q}}(n,a) = \left(\frac{n\,p+a}{q}\right) \quad \textit{if } (n\,p+a) \textit{ is divisible by } q. \quad (2)$$

We denote[4] by $\mathcal{T}_{\frac{p}{q}}$ the automaton $\mathcal{T}_{\frac{p}{q}} = \left\langle \mathbb{N}, A_p, \tau_{\frac{p}{q}}, 0, \mathbb{N} \right\rangle$.

In $\mathcal{T}_{\frac{p}{q}}$, we then have the transitions $n \xrightarrow{a} \left(\frac{n\,p+a}{q}\right)$ for every n in $\mathbb{N}$, and every a in A_p such that $(np+a)$ is divisible by q. The tree representation of $L_{\frac{p}{q}}$, as in Figure 2 augmented by an additional loop labelled by 0 on the state 0 becomes a representation of $\mathcal{T}_{\frac{3}{2}}$.

We call *minimal alphabet* the subalphabet $A_q = [\![0, q-1]\!]$ of A_p and respectively *maximal alphabet* the subalphabet $[\![(p-q),(p-1)]\!]$. Any letter of A_q is then called a *minimal letter*, *maximal letter* being defined analogously. The definition of $\tau_{\frac{p}{q}}$ implies that every state of $\mathcal{T}_{\frac{p}{q}}$ has a successor by a *unique* minimal (resp. maximal) letter.

[3] The function $\tau_{\frac{p}{q}}$ is defined on $\mathbb{N} \times \mathbb{Z}$ instead of $\mathbb{N} \times A_p$ in anticipation of future developments.

[4] In [1], $\mathcal{T}_{\frac{p}{q}}$ is denoted an infinite directed tree. The labels of the (finite) paths starting from the root precisely formed the language $0^* L_{\frac{p}{q}}$, as is $L\left(\mathcal{T}_{\frac{p}{q}}\right)$ in our case.

Definition 2 (minimal word). *A minimal word (in the* $\frac{p}{q}$*-system) is an infinite word in* A_q^ω *labelling an (infinite) path of* $\mathcal{T}_{\frac{p}{q}}$ *(not necessarily starting from the initial state* 0*).*

It is immediate that, for every n in $\mathbb{N}$, there exists a unique infinite word in A_q^ω starting from the state n in $\mathcal{T}_{\frac{p}{q}}$. We call this word *the* minimal word associated with n and denote it by w_n^-. Additionally, we will use the term *minimal outgoing label* of n, to designate the first letter of w_n^- and *minimal successor* of n the unique successor of n by a minimal letter.

We define in a similar way the *maximal word* w_n^+ associated with n.

3 The Derived Transducer

The goal of this section is to build a sequential letter-to-letter transducer $A_q \times A_q$ realising the function $w_n^- \mapsto w_{(n+1)}^-$. We call this transducer the *derived transducer* and denote it by $\mathcal{D}_{\frac{p}{q}}$. It will be obtained from $\mathcal{T}_{\frac{p}{q}}$ by a local transformation and this is the subject of Section 3.1.

3.1 From $\mathcal{T}_{\frac{p}{q}}$ to $\mathcal{D}_{\frac{p}{q}}$

The transformation of $\mathcal{T}_{\frac{p}{q}}$ into $\mathcal{D}_{\frac{p}{q}}$ is a two-step process. First, the structure of $\mathcal{T}_{\frac{p}{q}}$ is locally modified, by changing the alphabet, and a new automaton $\widehat{\mathcal{T}_{\frac{p}{q}}}$ is obtained. The second step consists in replacing the labels in $\widehat{\mathcal{T}_{\frac{p}{q}}}$ by a subset of $A_q \times A_q$ by means of a *substitution* (meaning that two transitions of $\widehat{\mathcal{T}_{\frac{p}{q}}}$ labelled by the same letter will be replaced by the same set of pair of letters) and produces $\mathcal{D}_{\frac{p}{q}}$.

Changing the Alphabet We write $B_{p,q} = [\![p-(2q-1),(p-1)]\!]$, that is $B_{p,q}$ is the alphabet whose maximal element is $p-1$ and containing $2q-1$ consecutive digits. In particular, if $p = (2q-1)$, $B_{p,q} = A_p$; if $p < (2q-1)$, $B_{p,q}$ contains negative digits; and if $p > (2q-1)$, $B_{p,q}$ is an uppermost subset of A_p. Note that $B_{p,q}$ is always of cardinal $(2q-1)$, an *odd number*, that the digit $(p-q)$ is then the *centre* of $B_{p,q}$ and that its maximal element $p-1$ coincides with the one of A_p.

The automaton $\widehat{\mathcal{T}_{\frac{p}{q}}}$ is then defined by:

$$\widehat{\mathcal{T}_{\frac{p}{q}}} = \left\langle \mathbb{N}, B_{p,q}, \tau_{\frac{p}{q}}, 0, \mathbb{N} \right\rangle \ .$$

This is possible, even if $B_{p,q}$ is larger than A_p because, in Equation 2, $\tau_{\frac{p}{q}}$ is defined on $\mathbb{N} \times \mathbb{Z}$, hence on $\mathbb{N} \times B_{p,q}$.

Figure 3 shows an example of the case when p is strictly smaller than $2q-1$, that is, transitions are added (thick arrows in the figure). The resulting automaton is a DAG (more complex than a tree with one loop).

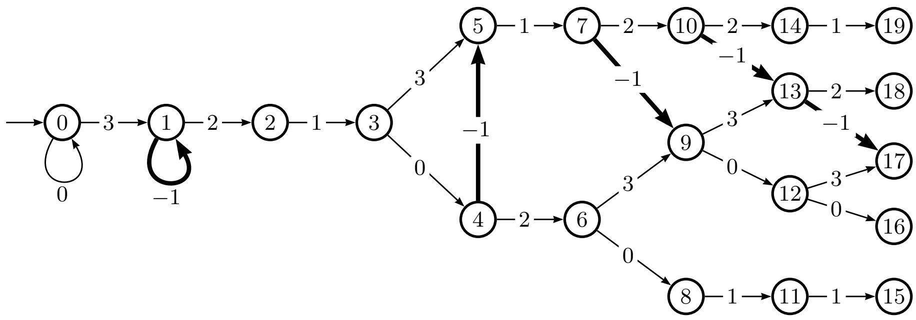
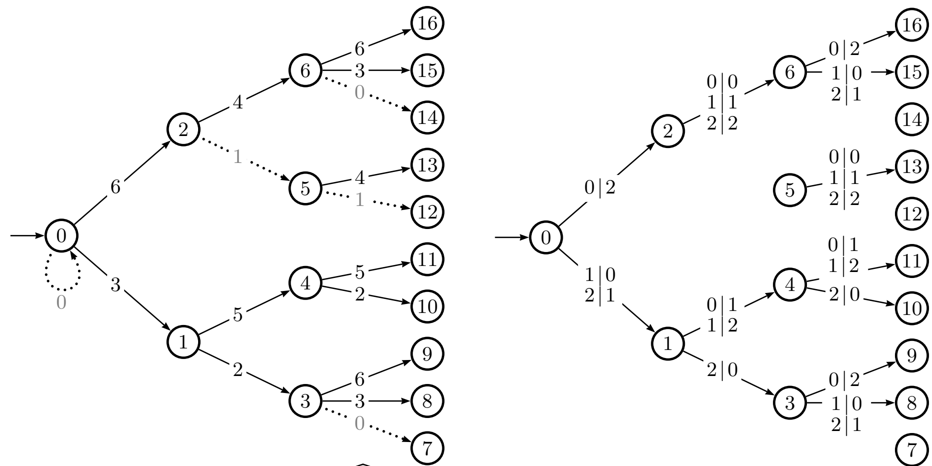


Fig. 3: Transforming $\mathcal{T}_{\frac{4}{3}}$ into $\widehat{\mathcal{T}_{\frac{4}{3}}}$

Figure 4a shows an example of the case when p is strictly greater than $2q-1$, that is, transitions are removed (dotted arrows in the figure). In this case, the resulting automaton is a forest (that is, an infinite union of trees). The accessible part is the tree rooted in 0. The other trees of the forest are not accessible; they are kept in $\widehat{\mathcal{T}_{\frac{p}{q}}}$, as they will come into play at Section 4. Furthermore, as already noted, if $p = (2q-1)$, $B_{p,q} = A_p$ and $\mathcal{T}_{\frac{p}{q}} = \widehat{\mathcal{T}_{\frac{p}{q}}}$.



(a) Transforming $\mathcal{T}_{\frac{7}{3}}$ into $\widehat{\mathcal{T}_{\frac{7}{3}}}$

(b) The derived transducer $\mathcal{D}_{\frac{7}{3}}$

Fig. 4: From $\mathcal{T}_{\frac{7}{3}}$ to $\mathcal{D}_{\frac{7}{3}}$

This construction ensures that every state of $\widehat{\mathcal{T}_{\frac{p}{q}}}$ congruent to -1 modulo q has a unique successor and that every other state has exactly two successors.

Changing the Labels Every label of $\widehat{\mathcal{T}_{\frac{p}{q}}}$ (which is a letter of $B_{p,q}$) is replaced by a *set* of pairs of digits in $A_q \times A_q$. The *label replacement function* $\omega_{\frac{p}{q}} : B_{p,q} \to \mathfrak{P}(A_q \times A_q)$ (or ω for short), is more easily defined in two steps. First, the function $\overline{\omega}$ computes the distance of the input to the centre of $B_{p,q}$: $\overline{\omega}(a) = a - (p - q)$, for every a in $B_{p,q}$. Then, the image of a by ω is the set of pairs of letters in A_q whose difference is $\overline{\omega}(a)$:

$$\forall a \in B_{p,q} \qquad \omega(a) \quad = \quad \{(b\,|\,c) \in A_q \times A_q \mid c - b = \overline{\omega}(a)\} \quad . \tag{3}$$

Example 1 (Case $\frac{3}{2}$). The functions $\overline{\omega}_{\frac{3}{2}}$ and $\omega_{\frac{3}{2}}$ are as follows:

$$\begin{array}{l|l} \overline{\omega}_{\frac{3}{2}} :\ 0 \longmapsto -1 & \omega_{\frac{3}{2}} :\ 0 \longmapsto \{\ 1\,|\,0\ \} \\ \overline{\omega}_{\frac{3}{2}} :\ 1 \longmapsto 0 & \omega_{\frac{3}{2}} :\ 1 \longmapsto \{\ 0\,|\,0,\ 1\,|\,1\ \} \\ \overline{\omega}_{\frac{3}{2}} :\ 2 \longmapsto 1 & \omega_{\frac{3}{2}} :\ 2 \longmapsto \{\ 0\,|\,1\ \} \end{array}$$

and Fig. 5 shows $\mathcal{D}_{\frac{3}{2}}$ ($D_{\frac{7}{3}}$ has been placed at Figure 4b in anticipation).

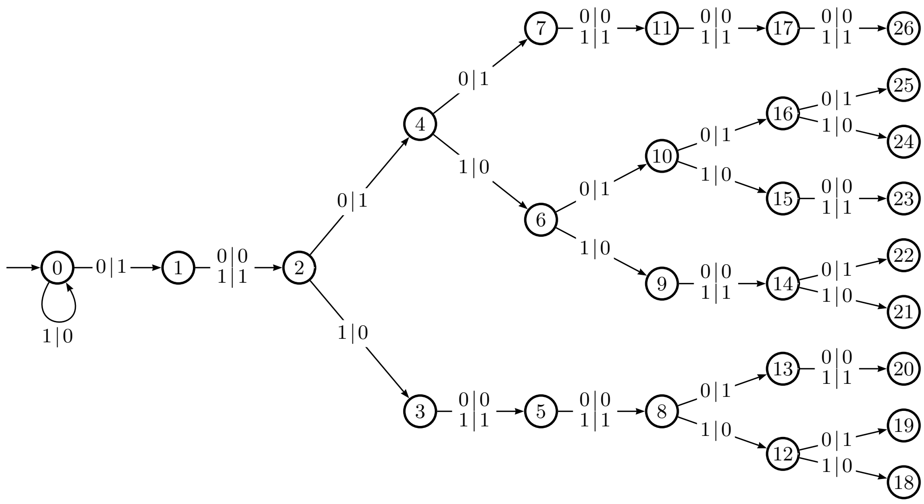


Fig. 5: The derived transducer $\mathcal{D}_{\frac{3}{2}}$

Formally, the transducer $\mathcal{D}_{\frac{p}{q}} = \langle \mathbb{N}, A_q \times A_q, \delta, \eta, 0, \mathbb{N} \rangle$ is defined *implicitly* or, more precisely, the transition function δ and the output function η are implicit functions defined by the following statement:

$$\begin{aligned} &\forall n \in \mathbb{N}\,,\ \forall a \in B_{p,q}\,,\ \forall (b,c) \in \omega(a) \\ &\quad \tau_{\frac{p}{q}}(n,a) \text{ defined} \qquad \Longrightarrow \qquad n \xrightarrow{b\,|\,c} \tau_{\frac{p}{q}}(n,a) \quad \text{is a transition of } \mathcal{D}_{\frac{p}{q}}\,, \\ &\qquad\qquad\qquad\qquad \text{that is, } \delta(n,b) = \tau_{\frac{p}{q}}(n,a) \text{ and } \eta(n,b) = c. \end{aligned} \tag{4}$$

In other words, the transitions of $\mathcal{D}_{\frac{p}{q}}$ are labelled as follows: if $n \equiv -1$ $[q]$, the state n has exactly one outgoing transition with labels $0\,|\,0$, $1\,|\,1$, $\ldots$, $q-1\,|\,q-1$. Otherwise, the state n has two outgoing transitions. If we write $k = a - (p-q)$ where a is the maximal outgoing label of n in $\mathcal{T}_{\frac{p}{q}}$: the label of the upper transition is $0\,|\,k$, $1\,|\,k+1$, $\ldots$, $(q-1-k)\,|\,q-1$; while the label of the lower transition is $q-k\,|\,0$, $(q-k+1)\,|\,1, \ldots, q-1\,|\,k-1$.

The transducer constructed in this manner is sequential and input-complete, as stated by the following lemma.

Lemma 1. *For every state n of $\mathcal{D}_{\frac{p}{q}}$ and every letter b of A_q, there exists a unique state m and a unique letter c such that $n \xrightarrow{b\,|\,c} m$.*

Corollary 1. *For every infinite word w in A_q^ω, $\mathcal{D}_{\frac{p}{q}}(w)$ exists and is unique.*

3.2 Correctness of $\mathcal{D}_{\frac{p}{q}}$

It remains to establish that $\mathcal{D}_{\frac{p}{q}}$ has the expected behaviour, as stated below.

Theorem 1. *For every n in $\mathbb{N}$, $\mathcal{D}_{\frac{p}{q}}(w_n^-) = w_{(n+1)}^-$.*

The proof of Theorem 1 relies on the equivalent (and more explicit) definition of the transitions of $\mathcal{D}_{\frac{p}{q}}$, stated in the following proposition.

Proposition 1. *If $n \xrightarrow{b\,|\,c} m$ is a transition of $\mathcal{D}_{\frac{p}{q}}$, then*

$$c = (b - (n+1)p) \,\%\, q \qquad \text{and} \qquad m = \left\lceil \frac{(n+1)p - b}{q} - 1 \right\rceil \quad .$$

In the case of finite words, a stronger version can be stated.

Theorem 2. *Given a base $\frac{p}{q}$ and two words u, v in A_q^*, the image of u by $\mathcal{D}_{\frac{p}{q}}$ is v if and only if there exists an integer n such that u is a prefix of w_n^- and v is a prefix of w_{n+1}^-.*

Theorem 2 is purposely stated on finite words and a similar statement for infinite words would be false: for every infinite word w of A_q^ω, $\mathcal{D}_{\frac{p}{q}}(w)$ exists, hence there are uncountably many pairs of infinite words $w\,|\,\mathcal{D}_{\frac{p}{q}}(w)$ accepted by $\mathcal{D}_{\frac{p}{q}}$ while there are only countably many pairs $w_n^-\,|\,w_{n+1}^-$.

4 Span of a Node

Lets us consider now the real value of infinite words. We denote by $\rho : A_p^\omega \to \mathbb{R}$, the *real evaluation function*, defined as follows:

$$\rho(a_1 a_2 \cdots a_n \cdots) \quad = \quad \sum_{i \geqslant 0} \frac{a_i}{q} \left(\frac{p}{q}\right)^{-i} \quad . \tag{5}$$

We denote by $W_{\frac{p}{q}}$ the language of infinite words $\mathcal{L}(\mathcal{T}_{\frac{p}{q}})$. It is proven in [1, Theorem 2] that $\rho(W_{\frac{p}{q}})$ is the interval $[0,\ \rho(w_0^+)]$. By extension, we denote by $W_{\frac{p}{q},n}$ (or, for short, W_n) the language of infinite words $\langle n\rangle^{-1}W_{\frac{p}{q}}$. Intuitively, an infinite word w over A_p is in W_n if $n \cdot u$ exists in $\mathcal{T}_{\frac{p}{q}}$ for every finite prefix u of w. Analogously to $W_{\frac{p}{q}}$, the following holds.

Lemma 2. *For every integer n, $\rho(W_{\frac{p}{q},n})$ is the interval $[\rho(w_n^-)\,,\ \rho(w_n^+)]$.*

Definition 3. *For every integer n, the span of n, denoted by $\mathtt{span}(n)$, is the size of $\rho(W_n)$: $\mathtt{span}(n) = (\rho(w_n^+) - \rho(w_n^-))$.*

Remark 1. Let us stress that what we call the span of the node n is not, in the fractal drawing (Figure 1b), the width of the subtree rooted in n. This quantity is obviously decreasing exponentially with the depth of the node n and the set of these has 0 as unique accumulation point. What we call span is this quantity *renormalised* by multiplication by $(\frac{p}{q})^k$, where k is the depth of the node n.

Let a be a letter from the minimal alphabet $A_q = [\![0,(q-1)]\!]$ and b a letter from the maximal alphabet $[\![(p-q),(p-1)]\!]$. The integer $(b-a)$ is necessarily in $[\![p-(2q-1),\ p-1]\!] = B_{p,q}$. Hence, through this digit-wise subtraction, denoted by '$\ominus$', $(w_n^+ \ominus w_n^-)$ is a word over $B_{p,q}$, and is called *the span-word* of n. It is routine to check that the following statement is true.

Lemma 3. *For all integer n, $\mathtt{span}(n) = \rho(w_n^+ \ominus w_n^-)$.*

Let $\mathtt{S}_{\frac{p}{q}}$ be the set of real numbers $\mathtt{S}_{\frac{p}{q}} = \{\mathtt{span}(n) \mid n \in \mathbb{N}\}$; the following statement holds.

Theorem 3.
(i) *If $p \leqslant 2q-1$, $\mathtt{S}_{\frac{p}{q}}$ is dense in $[0,\ \rho(w_0^+)]$.*
(ii) *If $p > 2q-1$, $\mathtt{S}_{\frac{p}{q}}$ is nowhere dense.*

The key to Theorem 3 is the connexion between the span-words and $\widehat{\mathcal{T}_{\frac{p}{q}}}$, achieved by Theorem 4 and Proposition 3.

Theorem 4. *All span-words are accepted by $\widehat{\mathcal{T}_{\frac{p}{q}}}$.*

The proof of this theorem is a direct consequence of Proposition 2, below and requires more definitions. There exists a (trivial) map m from the minimal alphabet to the maximal alphabet, such that, for all integer n, $m(w_{n+1}^-) = w_n^+$.

$$\begin{array}{rccl} m: & A_q & \longrightarrow & [\![(p-q),(p-1)]\!] \\ & a & \longmapsto & m(a) = \mathtt{maxLetter}(a+p) \end{array}$$

where $\mathtt{maxLetter}(x)$ is the greatest integer congruent to x modulo q and strictly smaller than p. By extending m to A_q^ω, Theorem 4 reduces to the statement that $\widehat{\mathcal{T}_{\frac{p}{q}}}$ accepts $(m(w_{n+1}^-) \ominus w_n^-)$ for every n:

Proposition 2. *If $w\,|\,w'$ is a pair of infinite words accepted by $\mathcal{D}_{\frac{p}{q}}$ then $\widehat{\mathcal{T}_{\frac{p}{q}}}$ accepts the word $(m(w') \ominus w)$.*

Analogously to the case of $\mathcal{D}_{\frac{p}{q}}$, $\widehat{\mathcal{T}_{\frac{p}{q}}}$ accepts uncountably many infinite words, therefore words that are not $(w_n^+ \ominus w_n^-)$ for any n. That being said, it seems to be the best result we can hope for, as the following two statements hold.

Proposition 3. *Every finite word accepted by $\widehat{\mathcal{T}_{\frac{p}{q}}}$ is the prefix of a span-word.*

Corollary 2. *The language of infinite words of $\widehat{\mathcal{T}_{\frac{p}{q}}}$ is the topological closure of the span-words.*

5 Conclusion

In the search of elucidating the structure of the set of representations of integers in a rational base number system, we have shown that the correspondence between two consecutive minimal words is achieved by a transducer that exhibits essentially the same structure as the one of the set of representations we started with. We have called this property an "*auto-similarity*" of the structure, as the structure is indeed not *self-similar*.

Let us note that the infinite transducer we have built realises the correspondence for *all minimal words*. It does not contradict the following conjecture that would express that each minimal word contains the complexity of the whole tree.

Conjecture 1. For every integer n, there exists a finite transducer that transforms w_n^- into w_{n+1}^-.

It is also remarkable that in this construction, the case $p = 2q - 1$ appears as the frontier between two completely different behaviours of the systems, in a much stronger way than it was described in our first work [1] on rational base number systems. It was hinted that there might be structural differences between two classes of rational base number systems. Indeed, those where $p \geqslant 2q - 1$ have an additional property, namely that, for every integer n, the span of n is never equal to 0. It was however never proved that this property was false when $p < 2q - 1$.

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. Shigeki Akiyama, Victor Marsault, and Jacques Sakarovitch. Auto-similarity in rational base number systems (full version). http://arxiv.org/abs/1305.6757.
3. Jean Berstel. *Transductions and Context-Free Languages.* Teubner, 1979.
4. M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge University Press, 2002.
5. K. Mahler. An unsolved problem on the powers of 3/2. *J. Austral. Math. Soc.*, 8:313–321, 1968.
6. Victor Marsault and Jacques Sakarovitch. On sets of numbers rationally represented in a rational base number system. In T. Muntean, D. Poulakis, and R. Rolland, editors, *Proceedings of CAI 2013.* Springer, 2013. to appear.