On subtrees of the representation tree in rational base numeration systems

Discrete Mathematics and Theoretical Computer Science DMTCS vol. **20**:1, 2018, #10

On subtrees of the representation tree in rational base numeration systems

Shigeki Akiyama[1] Victor Marsault[2,3] Jacques Sakarovitch[4]

[1] *University of Tsukuba, Ibaraki, Japan.*
[2] *University of Edinburgh, United Kingdom.*
[3] *University of Liège, Belgium.*
[4] *IRIF, CNRS/Paris Diderot University, and LTCI, Telecom-ParisTech, France.*

received 27th June 2017, *revised 27th Jan. 2018*, *accepted 23rd Feb. 2018*.

Every rational number $\frac{p}{q}$ defines a rational base numeration system in which every integer has a unique finite representation, up to leading zeroes. This work is a contribution to the study of the set of the representations of integers. This prefix-closed subset of the free monoid is naturally represented as a highly non-regular tree. Its nodes are the integers, its edges bear labels taken in $\{0, 1, \ldots, p-1\}$, and its subtrees are all distinct.

We associate with each subtree (or with its root n) three infinite words. The bottom word of n is the lexicographically smallest word that is the label of a branch of the subtree. The top word of n is defined similarly. The span-word of n is the digitwise difference between the latter and the former.

First, we show that the set of all the span-words is accepted by an infinite automaton whose underlying graph is essentially the same as the tree itself. Second, we study the function that computes for all n the bottom word associated with $n+1$ from the one associated with n, and show that it is realised by an infinite sequential transducer whose underlying graph is once again essentially the same as the tree itself.

An infinite word may be interpreted as an expansion in base $\frac{p}{q}$ after the radix point, hence evaluated to a real number. If T is a subtree whose root is n, then the evaluations of the labels of the branches of T form an interval of $\mathbb{R}$. The length of this interval is called the span of n and is equal to the evaluation of the span-word of n. The set of all spans is then a subset of $\mathbb{R}$ and we use the preceding construction to study its topological closure. We show that it is an interval when $p \leqslant 2q-1$, and a Cantor set of measure zero otherwise.

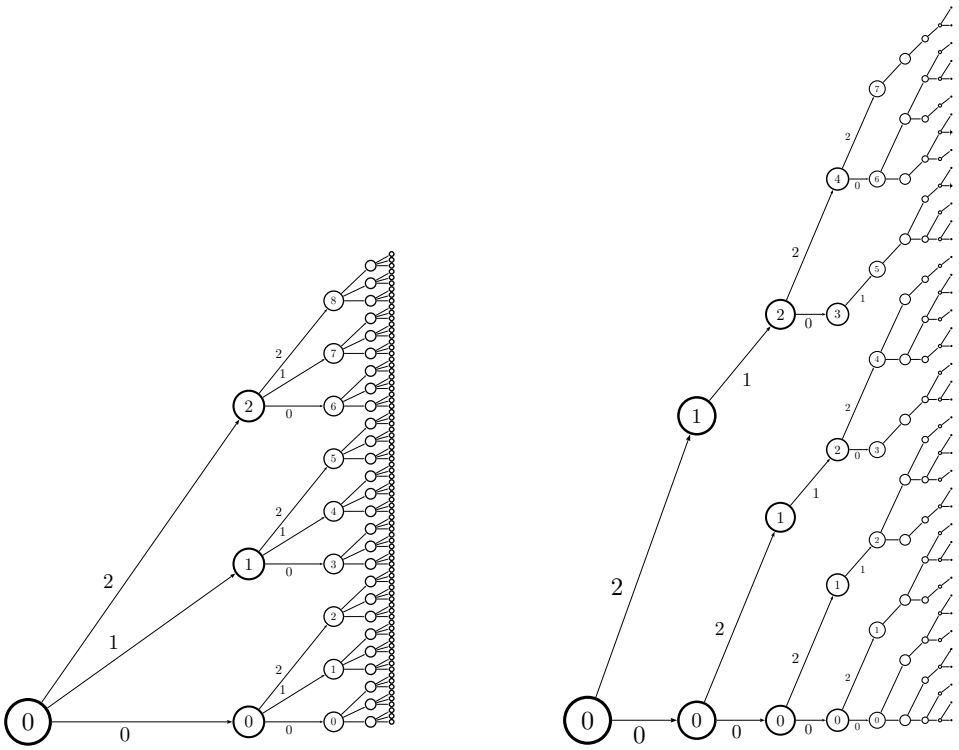
Keywords: Rational base numeration systems, Real-representation tree, Infinite words, Infinite transducers, Cantor sets, Hausdorff measure

ISSN 1365–8050

1 Introduction

The purpose of this work is a further exploration and a better understanding of the set of *infinite words* that appear in the definition of rational base numeration systems. These numeration systems have been introduced and studied by Akiyama, Frougny, and Sakarovitch (2008), leading to some progress and results in a number theoretic problem related to the distribution modulo 1 of the powers of rational numbers and usually known as Mahler's problem (Mahler, 1968). Besides these results, these systems raise many new and fascinating problems.

We give later the precise definition of rational base numeration systems and of the representation of numbers (integers and reals) in such systems. But one can hint at the results established in this paper by just looking at the figure showing the 'representation tree' in a rational base numeration system (Figure 1 (b) for the base $\frac{3}{2}$) and by comparison with the representation tree in a integer base numeration system (Figure 1 (a) for the base 3). In these trees, nodes are the natural integers, and the label of the path from the root to an integer n is the *representation* of n in the system, whereas the label of an infinite branch gives the representation in the system of a real number, indeed, and because the trees are drawn in a fractal way, of *the* real number which is the ordinate of the point where the branch ends.



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

Fig. 1: Representation trees in two number systems

The first striking fact is that the *representation language*, that is, the set of representations of integers, in a rational base numeration system does not fit at all in the usual classifications of formal language theory. It looks very chaotic and defeats any kind of iteration lemma. Nevertheless, these representation languages hide a certain kind of regularity and we have shown (Marsault and Sakarovitch, 2017) that they are so to speak characterized by their *periodic signatures*, that is, if one of these languages is drawn as a tree and traversed breadth-first, the degrees of the nodes are periodic.

If we now turn to the infinite branches of the trees, we first find that every subtree in the tree of Figure 1 (a) is the full ternary tree, whereas every subtree in the tree of Figure 1 (b) is different from all other subtrees. With the hope of finding some order or regularity within what seems to be close to complete randomness (which, on the other hand, is not established either and would be a very interesting result) we consider the *minimal words*, that we rather call *bottom words*, originating from every node of the tree.

In the case of an integer base, this is perfectly uninteresting: all these bottom words are equal to 0^ω. In the case of a rational base these words are on the contrary all distinct, none are even ultimately periodic (as the other infinite words in the representation tree). In order to find some invariant of all these distinct words, or at least a relationship between them, we have studied the function ξ that maps the bottom word $\mathbf{w}_n^-$ associated with n onto $\mathbf{w}_{n+1}^-$, the one associated with $n+1$. This function ξ is easily seen to be *online* and *realtime*, that is, the knowledge of the first i digits of the input is enough to compute the first i digits of the output, and hence ξ is computable by an infinite sequential letter-to-letter transducer.

The computation of such a transducer in the case the base $\frac{3}{2}$, and more generally in the case of a base $z = \frac{p}{q}$ with $p = 2q-1$, leads to a surprising and unexpected result. The transducer, denoted by $\mathcal{D}_z$, is obtained by replacing in the representation tree, denoted by $\mathcal{T}_z$, the label of every edge by a set of pairs of letters that depends upon this label only. In other words, the *underlying graphs* of $\mathcal{T}_z$ and $\mathcal{D}_z$ *coincide*, and $\mathcal{D}_z$ is obtained from $\mathcal{T}_z$ 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. In the special case, the canonical digit alphabet has $p = 2q-1$ elements; in the general case, we still consider a digit alphabet with $2q-1$ elements denoted by D_z, either by keeping the larger $2q-1$ elements of the canonical digit alphabet, when p is is greater than $2q-1$, or by enlarging the canonical alphabet with enough negative digits, when p is smaller than $2q-1$; in both cases, $p-1$ is the largest digit.

From $\mathcal{T}_z$ and with the digit alphabet D_z, we then define another 'representation graph' denoted by $\mathcal{S}_z$: either by *deleting the edges* of $\mathcal{T}_z$ labelled by digits that do not belong to D_z in the case where $p > 2q-1$ or, in the case where $p < 2q-1$ by *adding edges* labelled with the new negative digits. Then, $\mathcal{D}_z$ is obtained from $\mathcal{S}_z$ exactly as above, by a *substitution* from the alphabet of digits into the alphabet of pairs of digits. This construction of $\mathcal{D}_z$, and the proof of its correctness yields:

Theorem I.
Let p, q be two coprime integers such that $p > q > 1$ and $z = \frac{p}{q}$. Then $\mathcal{D}_z$ realises ξ.

In the original article (Akiyama et al., 2008), the tree $\mathcal{T}_z$, 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 radix point' of a real number and the drawing of the tree as a fractal object — like in Figure 1 — is fully justified by this point of view. The same idea leads to the definition of the (normalised[(i)]) *span* of a node n of the representation tree: it is the difference between the real numbers

[(i)] The classical definition of span of the node n is, in the fractal drawing, the width of the subtree rooted in n. This value is obviously

represented respectively by the top and the bottom words originating in the node n and let us denote by $\mathbf{Span}_z$ the set of spans for all integers and by $c\ell(\mathbf{Span}_z)$ its topological closure.

Again, this notion is totally uninteresting in the case of a numeration system with an integer base p: the span of every node n is always 1. And again, the notion is far more richer and complex in the case of a rational base $\frac{p}{q}$ since we establish the following.

Theorem II. *Let p, q be two coprime integers such that $p > q > 1$ and $z = \frac{p}{q}$.*

(a) *If $p \leqslant 2q - 1$, then $c\ell(\mathbf{Span}_z)$ is an interval.*

(b) *If $p > 2q - 1$, then $c\ell(\mathbf{Span}_z)$ is a Cantor set of measure zero.*

As different they may look, Theorems I and II have a common root in the construction of the automaton $\mathcal{S}_z$. The trivial relationship between the bottom word originating at node $n + 1$ and the top word originating at node n leads to the connexion between the construction of the transducer $\mathcal{D}_z$ and the description of the set of spans $\mathbf{Span}_z$. The *digitwise difference* between top and bottom words is written on the alphabet D_z, and all these 'difference words' are infinite branches in the automaton $\mathcal{S}_z$. This is explained in Section 4. Theorem I is then established in Section 5 and Theorem II in Section 6. The second case of Theorem II is completed with an upper bound for the Hausdorff dimension of $c\ell(\mathbf{Span}_z)$. This paper is meant to be self-contained and starts, in particular, with all necessary definitions concerning rational base number systems in Section 3. We conclude the paper with an open problem on minimal words which indeed was the motivating force of all this work, and with a conjecture on the Hausdorff dimension of $c\ell(\mathbf{Span}_z)$.

The present article is a long version of a work (Akiyama et al., 2013) presented at the 9th International Conference on Words. Most of the results are also part of the thesis of the second author (Marsault, 2016).

2 Preliminaries and notation

2.1 On words and numbers

An *alphabet* is a finite set of symbols, called *letters*. A *word* (resp. an *ω-word*) is a finite (resp. infinite) sequence of letters and a language (resp. an ω-language) is a set of words (resp. ω-words). The set of the words (resp. ω-words) over an alphabet A is denoted by A^* (resp. A^ω). Subsets of A^* are called *languages* over A and those of A^ω are called *ω-languages* over A. For the sake of clarity, we use the standard math font for letters and words: $a, b, c, d, u, v, w\ldots$ and a bold sans-serif font for ω-words: $\mathbf{u}, \mathbf{v}, \mathbf{w}\ldots$ The *length* of a word u is denoted by $|u|$ and the *concatenation* of two words u and v is denoted simply by uv.

If $w = uv$ (resp. $\mathbf{w} = u\mathbf{v}$), then u is called a *prefix* of w (resp. of $\mathbf{w}$); note that the prefixes of word or of ω-words always are words. We denote by PRE the function $A^* \cup A^\omega \to \mathfrak{P}(A^*)$ that maps a word or an ω-word to the set of all its prefixes; PRE is naturally lifted to languages and ω-languages, that is, to a function $\mathfrak{P}(A^*) \cup \mathfrak{P}(A^\omega) \to \mathfrak{P}(A^*)$. A language L is said *prefix-closed* if $\text{PRE}(L) = L$.

Words and ω-words will later be evaluated using a rational base numeration system (defined in Section 3). It is then convenient to have a different index convention for words and ω-words: we index (finite) words *from right to left* and use 0 as the rightmost index (as in $a_k \cdots a_1 a_0$), while ω-words are indexed from left to right, starting with index 1 (as in $a_1 a_2 \cdots$).

decreasing (exponentially) with the depth of the node n, hence the span of two nodes cannot be easily compared. In this work, we only consider the *normalised span* which is the span multiplied by $(\frac{p}{q})^k$, where k is the depth of the node n.

In this article, letters always are (relative) integers and we use *digit* as a synonym for letter. Moreover, alphabets always are integer intervals, that is, sets of consecutive integers. In particular, our alphabets are totally ordered, which implies that any set of words is equipped with two total orders: the *radix order* and the *lexicographic order*:

Definition 1. *Let u and v be words over A and w their longest common prefix.*

(a) $u \leqslant_{lex} v$ *if*
- *either $u = w$, that is, u is a prefix of v,*
- *or $u = wax$ and $v = wby$ with a, b in A and $a < b$.*

(b) $u \leqslant_{rad} v$ *if*
- *either $|u| < |v|$*
- *or $|u| = |v|$ and $u \leqslant_{lex} v$.*

Let $\mathbf{u}$ *and* $\mathbf{v}$ *be ω-words over A.*

(c) $\mathbf{u} \leqslant_{lex} \mathbf{v}$ *if*
- *either* $\mathbf{u} = \mathbf{v}$*,*
- *or, if w (in A^*) is their longest common prefix,* $\mathbf{u} = wa\mathbf{x}$ *and* $\mathbf{v} = wb\mathbf{y}$ *with* $\mathbf{x}, \mathbf{y}$ *in A^ω and a, b in A such that $a < b$.*

The set of ω-words is classically equipped with the product topology which can also be defined with a distance.

Definition 2. *Let* $\mathbf{u}, \mathbf{v}$ *be two infinite words. The distance between these two words is*

$$d(\mathbf{u}, \mathbf{v}) = \begin{cases} 0 & \text{if } \mathbf{u} = \mathbf{v} \\ 2^{-|w|} & \text{where } w \text{ is the longest common prefix of } \mathbf{u} \text{ and } \mathbf{v}, \text{ otherwise.} \end{cases}$$

2.2 On trees, automata and transducers

In this article we consider infinite, directed graphs of a special form. First, there is a special *initial* vertex called the *root* and indicated by an incoming arrow in figures. Second, the edges are labelled over a finite alphabet. Third, they are *deterministic*: there is never two different edges originating from the same vertex and labelled by the same letter. Such graphs are represented by quadruple $\langle A, V, i, \delta \rangle$ where A is the finite alphabet, V is the (infinite) vertex-set, δ is a function $V \times A \to V$ is the set of edges. We call such graphs *automata* and we use terminology of automata theory; in particular we use *state* rather than *vertex*, and *transition* rather than *edge*.

A transition is denoted by $s \xrightarrow{a} s'$, where s, s' are states and a is a letter. We will consider *finite* and *infinite* paths in these graphs. We refer to infinite paths as *branches* and refer to finite paths simply as *paths*. A branch is thus denoted by $s \xrightarrow{\mathbf{w}} \cdots$ and a path by $s \xrightarrow{u} s'$, where s, s' denote states, $\mathbf{w}$ an ω-word and u a word. We call *dead-end* a state with no outgoing transitions; in this article, automata will have no dead-end.

A *run* refers to a path starting from the root. *The run of u* is the unique run labelled by u as a label, if it exists; in which case u is said to be *accepted* by the automaton. The language accepted by $\mathcal{A}$, denoted

by $L(\mathcal{A})$ is the set of the words accepted by $\mathcal{A}$. The notions of *ω-run* and *accepted ω-language* (denoted by $\Lambda(\mathcal{A})$) are defined similarly. If $\mathcal{A}$ has no dead-end, then $L(\mathcal{A}) = \textsc{Pre}(\Lambda(\mathcal{A}))$.

We call *tree* an automaton in which every state is reached by exactly one run.

A *transducer* is an automaton where the labels are taken in a product alphabet $A \times B$; A is the *input alphabet* and B the *output alphabet*. All the transducers we consider are *input-deterministic*: if $s \xrightarrow{(a,b)} t$ and $s \xrightarrow{(a,b')} t'$ then $b = b'$ and $t = t'$. They are interpreted as computing functions: the first component is the input and the second is the output. If (u, v) labels a run of a transducer $\mathcal{T}$, then we say that v is the *image by $\mathcal{T}$ of u*; by abuse of language, this run will be called *the run of u*.

With the usual definition of automata and transducers (as for instance in Sakarovitch, 2009) what we call automaton is indeed an *infinite deterministic automaton with all states final* and what we call transducer is indeed an *infinite letter-to-letter pure-sequential transducer*.

Let us conclude this section with a statement linking the language and the ω-language accepted by an automaton (more details on the subject in Perrin and Pin, 2004).

Lemma 3. *Let $\mathcal{A}$ be an automaton with no dead-end and S an ω-language. It holds $L(\mathcal{A}) = \textsc{Pre}(\mathsf{S})$ if and only if $\Lambda(\mathcal{A}) = c\ell(\mathsf{S})$.*

Proof: Forward direction. Let $\mathbf{w}$ be an ω-word. The following sequence of equivalences holds.

$$\begin{aligned}
\mathbf{w} \in \Lambda(\mathcal{A}) \quad &\iff \quad \textsc{Pre}(\mathbf{w}) \subseteq L(\mathcal{A}) \quad \iff \quad \textsc{Pre}(\mathbf{w}) \subseteq \textsc{Pre}(\mathsf{S}) \\
&\iff \quad \forall u \in \textsc{Pre}(\mathbf{w}),\ \exists \mathbf{s}_u \in \mathbf{S} \qquad u \in \textsc{Pre}(\mathbf{s}_u) \quad \iff \quad \mathbf{w} \in c\ell(\mathsf{S}) \ .
\end{aligned}$$

Backward direction. Let u be a word. The following sequence of equivalences holds.

$$\begin{aligned}
u \in L(\mathcal{A}) \quad &\iff \quad \exists \mathbf{w} \in \Lambda(\mathcal{A}) \qquad u \in \textsc{Pre}(\mathbf{w}) \quad \text{(no-dead-end hypothesis)} \\
&\iff \quad \exists \mathbf{w} \in c\ell(\mathsf{S}) \qquad u \in \textsc{Pre}(\mathbf{w}) \quad \text{(backward-dir. hypothesis)} \\
&\iff \quad \exists \mathbf{w'} \in \mathsf{S} \qquad u \in \textsc{Pre}(\mathbf{w'}) \quad \text{(closure definition)} \\
&\iff \quad u \in \textsc{Pre}(\mathsf{S}) \ .
\end{aligned}$$

□

3 Rational base numeration systems

In this section, we recall the definition of rational base numeration systems that have been introduced by Akiyama, Frougny, and Sakarovitch (2008), and the properties of the representation trees that were established in this paper.

Notation 4. *We denote by p and q two co-prime integers such that $p > q > 1$, and by z the rational number $z = \frac{p}{q}$. They will be fixed throughout the article.*

Note that the numeration system in base $\frac{p}{q}$ we are about to describe is *not* the β-numeration where $\beta = \frac{p}{q}$. Indeed, in the latter, the representation of a number is computed by a left-to-right algorithm (called *greedy*, *cf.* Lothaire, 2002, Chapter 7), the digit set is $\left\{0, 1, \ldots, \left\lfloor \frac{p}{q} \right\rfloor\right\}$ and the weight of the i-th leftmost digit

is $(\frac{p}{q})^i$. Meanwhile, in base $\frac{p}{q}$, the representations are computed by a right-to-left algorithm (Equation (1)), digits are taken in $\{0, 1, \ldots, (p-1)\}$ and the weight of the i-th digits is $\frac{1}{q}(\frac{p}{q})^i$.

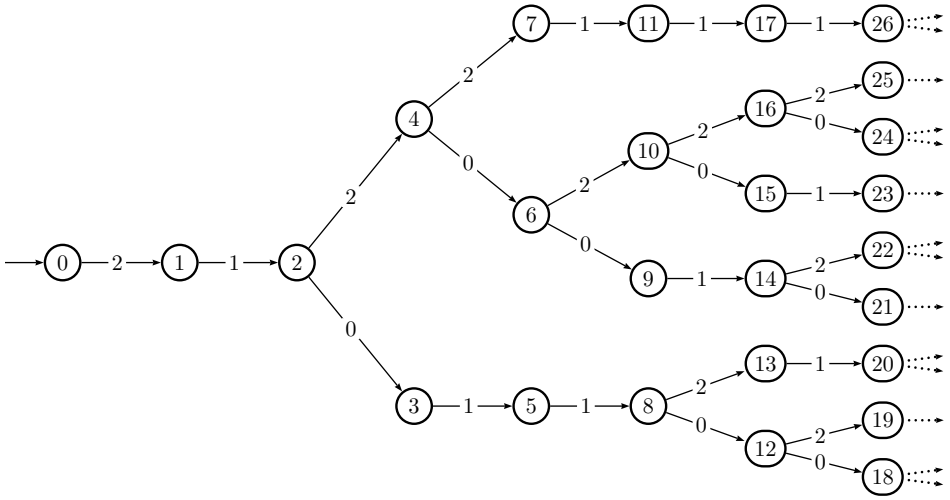


Fig. 2: The language $L_{\frac{3}{2}}$ represented as a tree

3.1 Representation of integers

Given a positive integer N, let us define $N_0 = N$ and, for all $i > 0$,

$$qN_i \;=\; pN_{(i+1)} + a_i \;,$$

where a_i and $N_{(i+1)}$ are the remainder and the quotient of the Euclidean division of qN_i by p. Hence a_i belongs to the alphabet $A_p = \{0, 1, \ldots, p-1\}$. Since $p > q$, the sequence $(N_i)_{i\in\mathbb{N}}$ is first strictly decreasing until it reaches 0: there is an integer k such that $N_0 > N_1 > \cdots > N_k > N_{k+1} = 0$. The word $a_k \cdots a_1 a_0$ of ${A_p}^*$ is denoted by $\langle N \rangle_z$. Equation (1), below, gives a compact definition of the same algorithm.

$$\langle 0 \rangle_z \;=\; \varepsilon \tag{1a}$$

$$\forall m > 0 \qquad \langle m \rangle_z \;=\; \langle n \rangle_z \, a \qquad \text{where} \quad n \in \mathbb{N}\,, \quad a \in A_p \quad \text{and} \quad qm \;=\; pn + a \tag{1b}$$

If $\langle N \rangle_z = a_k a_{k-1} \cdots a_0$, then it holds

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

The *evaluation function* π_z is derived from this formula. The *value* of any word $a_k a_{k-1} \cdots a_0$ over A_p, and indeed over any alphabet of digits, is defined by

$$\pi_z(a_k a_{k-1} \cdots a_0) \;=\; \sum_{i=0}^{k} \frac{a_i}{q} \left(\frac{p}{q}\right)^i \quad . \tag{2}$$

A word u in A_p^* is called a $\frac{p}{q}$*-expansion* of an integer n, if $\pi_z(u) = n$. Since $\frac{p}{q}$-expansions are unique up to leading 0's (*cf.* Akiyama et al. 2008, Theorem 1), u is equal to $0^i \langle n \rangle_z$ for some integer i and $\langle n \rangle_z$ is called *the* $\frac{p}{q}$*-representation of* n. The set of the $\frac{p}{q}$-representations of integers is denoted by L_z:

$$L_z \;=\; \{\langle n \rangle_z \mid n \in \mathbb{N}\} \quad . \tag{3}$$

It follows from (1b) that L_z is prefix-closed and right-extendable. As a consequence, L_z can be represented as a tree with no dead-end (cf. Figures 2, 3 and later on 6). The node set is $\mathbb{N}$, the root is 0, and there is an arc $n \xrightarrow{a} m$ if $\langle n \rangle_z a = \langle m \rangle_z$.

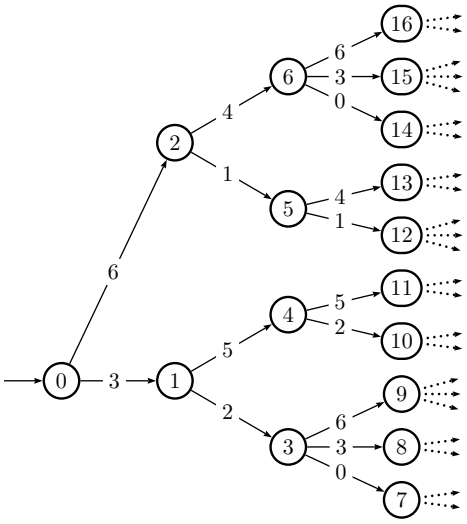


Fig. 3: The language $L_{\frac{7}{3}}$ represented as a tree

Moreover, the base $\frac{p}{q}$ is the "abstract numeration system" (*cf.* Lecomte and Rigo, 2001, 2010) built from L_z, a property that may be stated as follows:

Proposition 5 (Akiyama et al., 2008, Proposition 11)**.** $\forall n, m \in \mathbb{N} \qquad n \leqslant m \quad \iff \quad \langle n \rangle_z \leqslant_{rad} \langle m \rangle_z$.

or, equivalently as:

$$\forall u, v \in L_z \qquad \pi_z(u) \;\leqslant\; \pi_z(v) \quad \iff \quad u \;\leqslant_{\text{rad}}\; v \quad . \tag{4}$$

It is known that L_z is not a regular language (not even a context-free language). In fact, it even possesses a "Finite Left Iteration Property" which essentially says that L_z cannot satisfy any kind of pumping lemma. Lemma 11, later on, is a consequence of this fact.

Definition 6. (a) *Let $\tau_z : \mathbb{N} \times \mathbb{Z} \to \mathbb{N}$ be the (partial) function defined by:*

$$\forall n \in \mathbb{N}\,, \quad \forall a \in \mathbb{Z} \qquad \tau_z(n,a) \;=\; \left(\frac{np+a}{q}\right) \qquad \textit{if } (np+a) \textit{ is divisible by } q. \tag{5}$$

(b) *We denote* [(ii)] *by $\mathcal{T}_z$ the* infinite *automaton: $\mathcal{T}_z \;=\; \langle\, A_p,\, \mathbb{N},\, 0,\, \tau_z \,\rangle$.*

Remark 7.
- *The function τ_z is defined on $\mathbb{N} \times \mathbb{Z}$ instead of $\mathbb{N} \times A_p$ in anticipation of future developments.*
- *The automaton $\mathcal{T}_z$ is not quite a tree. Indeed, the state 0 (that is, the root) holds a loop labelled by the digit 0 since $\tau_z(0,0) = 0$.*

The transitions of $\mathcal{T}_z$ are characterised by the following.

$$\forall n, m \in \mathbb{N}\,, \quad \forall a \in A_p \qquad n \xrightarrow[\mathcal{T}_z]{a} m \quad \Longleftrightarrow \quad qm \;=\; pn + a \tag{6}$$

Comparing (1) and (6) shows how the difference between L_z and $\mathcal{T}_z$ is mostly a question of formalism. It holds $L(\mathcal{T}_z) = \; 0^* L_z$ and next lemma gives a more precise statement.

Lemma 8. *Let u be in $L(\mathcal{T}_z)$. Then, $\pi_z(u)$ is in $\mathbb{N}$ and $0 \xrightarrow[\mathcal{T}_z]{u} \pi_z(u)$.*

Lemma 8 implies that the tree representation of L_z, as in Figures 2, 3 and 6, augmented by an additional loop labelled by 0 onto the root 0 becomes a representation of $\mathcal{T}_z$. Moreover, since L_z is right-extendable, the next statement holds.

Lemma 9. *$\mathcal{T}_z$ has no dead-end.*

We now state a few properties of $\mathcal{T}_z$. They are the translations of results due to Akiyama et al. (2008) into the formalism we use here.

Lemma 10 (Akiyama et al., 2008, Lemma 6)**.** *Let n, n' be two integers. Let k be another integer.*

(a) *If n and n' are congruent modulo q^k, then for every word u of length k, the following are equivalent.*
- *There exists an integer m such that $n \xrightarrow{u} m$.*
- *There exists an integer m' such that $n' \xrightarrow{u} m'$.*

(b) *If there exist two integers m, m' and a word u of length k such that $n \xrightarrow{u} m$ and $n' \xrightarrow{u} m'$, then n et n' are congruent modulo q^k.*

Lemma 11. *Let $n \xrightarrow{\mathbf{w}} \cdots$ be a branch of $\mathcal{T}_z$. If $\mathbf{w}$ is periodic, then $n = 0$ and $\mathbf{w} = 0^\omega$.*

[(ii)] In Akiyama et al. (2008), $\mathcal{T}_z$ denotes an infinite directed tree. The labels of the (finite) paths starting from the root precisely formed the language $0^* L_z$, as is $L(\mathcal{T}_z)$ in our case.

Proof: The hypothesis implies that there is a word u such that $n \xrightarrow{u} m \xrightarrow{u^\omega} \cdots$ is a branch of $\mathcal{T}_z$. From Lemma 10(b), n and m are congruent modulo $q^{|u|\times i}$ for every integer i. Hence $n = m$. The only circuit in $\mathcal{T}_z$ is $0 \xrightarrow{0} 0$, hence the statement. □

For every integer k let us define the (total) function $f_k \colon \mathbb{N} \to {A_p}^k$ in the following way. Equation (6) implies that every state of $\mathcal{T}_z$ (incuding 0) has exactly one incoming transition, hence, by induction on k, exactly one incoming path of length k: for every integer m, $f_k(m) = u$ where u is the label of this unique path of length k ending in m.

Lemma 12 (Akiyama et al., 2008, Proposition 10)**.** *Let m, m' be two integers. For every integer k, m and m' are congruent modulo p^k if and only if $f_k(m) = f_k(m')$.*

Lemma 13. *For every integer k, f_k is a bijection between any integer interval S of cardinal p^k and ${A_p}^k$.*

Proof: Two integers m, m' in S are necessarily in different residue classes modulo p^k, hence from Lemma 12, satisfy $f_k(m) \neq f_k(m')$. It follows that $f_k(S)$ is of cardinal p^k. □

Applying Lemma 13 to every integer k yields the following.

Lemma 14. *Every word in ${A_p}^*$ is the label of some path of $\mathcal{T}_z$.*

3.2 Representation of real numbers

Let us define a second *evaluation function* ρ_z. It evaluates an ω-word after the radix point (for short a.r.p.) hence computes a real number. The a.r.p. value of an ω-word $\mathbf{w} = a_1\, a_2 \cdots$ over the alphabet A_p, or indeed *over any digit alphabet*, is

$$\rho_z(a_1\, a_2 \cdots) \;=\; \sum_{i \geqslant 1} \frac{a_i}{q} \left(\frac{p}{q}\right)^{-i} \quad . \tag{7}$$

Proposition 15. *The function ρ_z is uniformly continuous.*

Let us stress that the function ρ_z is *not* order-preserving. Since for every (non integer) rational base $\frac{p}{q}$, $q \geqslant 2$ and $p \geqslant 3$ hold, the following inequalities hold

$$0\,(p-1)\,0^\omega \;<_{\text{lex}}\; 1\,0^\omega \qquad \text{and} \qquad \rho_z(0\,(p-1)\,0^\omega) \;=\; \frac{q(p-1)}{p^2} \;>\; \frac{1}{p} \;=\; \rho_z(10^\omega) \;.$$

However, ρ_z is order-preserving on the ω-language accepted by $\mathcal{T}_z$ (Proposition 17 below).

Definition 16. *We denote by W_z the ω-language accepted by $\mathcal{T}_z$, that is, $\mathsf{W}_z = \Lambda(\mathcal{T}_z)$.*

For instance, Figures 10 and 11(a) (pages 21 and 28) are representations of $\mathsf{W}_{\frac{3}{2}}$ and $\mathsf{W}_{\frac{7}{3}}$ as fractal trees. In these figures, consider a path from the root to a node X labelled by a word u. The node X is then at the ordinate $\rho_z(u0^\omega)$ and is labelled by $\pi_z(u)$. The abscissa has no particular meaning except that it grows with the length of u. For example, in Figure 10, there is a path starting from the root and labelled by $u = 21$; the endpoint of this path is a node labelled by $\pi_z(21) = 2$ and positioned at the

ordinate $\rho_{\frac{3}{2}}(210^\omega) = 0.888\cdots$. Similarly, the run of $u = 210$ reaches a node labelled by $\pi_z(210) = 3$ and whose ordinate is also $\rho_{\frac{3}{2}}(2100^\omega) = \rho_{\frac{3}{2}}(210^\omega) = 0.888\cdots$.

Proposition 17 (Akiyama et al., 2008, Lemma 34)**.** $\forall \mathbf{u}, \mathbf{v} \in \mathsf{W}_z \qquad \rho_z(\mathbf{u}) \leqslant \rho_z(\mathbf{v}) \quad \Longleftrightarrow \quad \mathbf{u} \leqslant_{lex} \mathbf{v}$.

As figures suggest, the set W_z, when projected to $\mathbb{R}$ by ρ_z, produces an interval, as stated below.

Theorem 18 (Akiyama et al., 2008, Theorem 2)**.** *The image of* W_z *by* ρ_z *is an interval.*

3.3 Bottom and top words

Lemma 9 states that every state n of $\mathcal{T}_z$ is the root of an infinite subtree. We now turn our attention to the ω-words that are the frontiers of these subtrees. Let us first call *lower alphabet*, and denote by B_z, the set of the smallest q integers: $B_z = \{0, 1, \ldots, q-1\}$.

Definition 19. (a) *We call* bottom word[iii] of n*, and denote by* $\mathbf{w}_n^-$*, the smallest* ω*-word that labels a branch of* $\mathcal{T}_z$ *originating from* n*.*

(b) *Let* Bot_z *denote the set of the bottom words:* $\mathsf{Bot}_z = \{\mathbf{w}_n^- \mid n \in \mathbb{N}\}$.

Example 20. *One reads on Figure 2 some bottom words in base* $\frac{3}{2}$*:*

$$\mathbf{w}_1^- = 1011000\cdots \quad , \qquad \mathbf{w}_3^- = 11000\cdots \qquad \textit{and} \qquad \mathbf{w}_4^- = 00101\cdots$$

Bottom words are characterised by the alphabet they are written on:

Property 21. $\mathsf{Bot}_z = \mathsf{W}_z \cap B_z^\omega$.

This property will be used under the following form.

Property 22. *Let* n *be in* $\mathbb{N}$ *and* u *in* B_z^**. If* $n \xrightarrow[\mathcal{T}_z]{u} m$*, then* u *is a prefix of* $\mathbf{w}_n^-$*.*

From Lemma 14 and Property 21 follows the next statement.

Lemma 23. *The set* Bot_z *is dense in* ${B_z}^\omega$*.*

Symmetrically, we denote by $\mathbf{w}_n^+$ the *top word* [iv] *of* n, by Top_z the set of the top words and call *upper alphabet* the alphabet $C_z = \{p-q, p-q+1, \ldots, p-1\}$. Statements much similar to Property 21, Property 22 and Lemma 23 could be made about the top words and the upper alphabet.

Example 24. *One reads on Figure 2 some top words in base* $\frac{3}{2}$*:*

$$\mathbf{w}_1^+ = 1221112\cdots \quad , \qquad \mathbf{w}_3^+ = 11212\cdots \qquad \textit{and} \qquad \mathbf{w}_4^+ = 21112\cdots$$

[iii] *Bottom words* were called *minimal words* in Akiyama et al. (2008).
[iv] *Top words* were called *maximal words* in Akiyama et al. (2008).

The bottom word of $(n+1)$ and the top word of n are related by the function $\mu\colon C_z \to B_z$ defined by

$$\mu(c) \;=\; c - (p-q)\ , \tag{8}$$

and extended to a (letter-to-letter) morphism from C_z^* to B_z^*, and from C_z^ω to B_z^ω.

Lemma 25. *For every integer n, $\mathbf{w}^-_{n+1} = \mu(\mathbf{w}^+_n)$.*

Proposition 26. *Let n, m be two integers and let a be a letter of A_p such that $n \xrightarrow[\mathcal{T}_z]{a} m$ and $n \xrightarrow[\mathcal{T}_z]{a+q} m+1$. Then, $\rho_z\big((a+q)\mathbf{w}^-_{m+1}\big) = \rho_z(a\mathbf{w}^+_m)$.*

4 Span-words

The notion of span-word will be central in the proof of both Theorems I and II via the construction of a new automaton denoted by $\mathcal{S}_z$ and obtained from $\mathcal{T}_z$ by enlarging, or restricting, the alphabet.

Definition 27. *Let D_z denote the set of the differences between letters from the upper alphabet and letters from the lower alphabet:*

$$D_z \;=\; C_z - B_z \;=\; \{d \in \mathbb{Z} \mid \exists c \in C_z,\, \exists b \in B_z \quad d \;=\; c - b\}\ .$$

The alphabet D_z is the integer interval whose cardinal is the odd integer $(2q-1)$, whose largest element is $(p-1)$. Its 'central element', called *middle-point*, is $p-q$:

$$D_z \;=\; \{\,p-(2q-1), \ldots, (p-1)\,\}\ .$$

Property 28. (a) $C_z \subseteq D_z$.
(b) *If $p = (2q-1)$, then $D_z = A_p$.*
(c) *If $p < (2q-1)$, then $D_z \supsetneq A_p$ and contains **negative** digits.*
(d) *If $p > (2q-1)$, then $D_z \subsetneq A_p$; more precisely, D_z is the set of the largest $(2q-1)$ digits of A_p.*

Definition 29. *We denote by $\oplus$ and $\ominus$ the digitwise addition and subtraction of words of the same length respectively, that is,*

$$\begin{aligned}(a_k \cdots a_1\, a_0) \oplus (b_k \cdots b_1\, b_0) \;&=\; (a_k + b_k) \cdots (a_1 + b_1)(a_0 + b_0) \quad ;\\ (a_k \cdots a_1\, a_0) \ominus (b_k \cdots b_1\, b_0) \;&=\; (a_k - b_k) \cdots (a_1 - b_1)(a_0 - b_0) \quad .\end{aligned}$$

Digitwise addition and subtraction of ω-words are defined similarly.

Property 30. *For any w in D_z^*, there exist u in B_z^* and v in C_z^* such that $w = v \ominus u$.*

Definition 31. (a) *We call* span-word[(v)] of n, *and denote by* $\mathsf{s}(n)$, *the ω-word* $\mathbf{w}^+_n \ominus \mathbf{w}^-_n$.
(b) *We denote by* Spw_z *the set of all span-words:* $\mathsf{Spw}_z = \{\mathsf{s}(n) \mid n \in \mathbb{N}\}$.

[(v)] The denomination *span-word* comes from the a.r.p. value of those ω-words, and will be explained in Section 6 (Definition 55).

Example 32. *In base $\frac{3}{2}$, it reads:*

$$\begin{aligned}
\mathsf{s}(1) &= \mathbf{w}_1^+ \ominus \mathbf{w}_1^- = (1221112\cdots) \ominus (1011000\cdots) = 0210112\cdots \\
\mathsf{s}(3) &= \mathbf{w}_3^+ \ominus \mathbf{w}_3^- = (11212\cdots) \ominus (11000\cdots) = 00212\cdots \\
\mathsf{s}(4) &= \mathbf{w}_4^+ \ominus \mathbf{w}_4^- = (21112\cdots) \ominus (00101\cdots) = 21011\cdots
\end{aligned}$$

Since bottom words belong to ${B_z}^\omega$ and top words to ${C_z}^\omega$, it follows:

Property 33. $\mathsf{Spw}_z \subseteq {D_z}^\omega$.

Definition 34. *Let $\mathcal{S}_z$ be the automaton defined by*

$$\mathcal{S}_z = \langle\, D_z,\, \mathbb{N},\, 0,\, \tau_z \,\rangle\ ,$$

where τ_z is defined by Equation (5) with domain restricted to $\mathbb{N} \times D_z$.

The transitions of $\mathcal{S}_z$ are characterised by:

$$\forall n, m \in \mathbb{N}\,,\ \ \forall a \in D_z \qquad n \xrightarrow[\mathcal{S}_z]{a} m \quad \Longleftrightarrow \quad qm = pn + a\ . \tag{9}$$

Using (9), it is a routine to show that Lemma 8 extends to $\mathcal{S}_z$.

Lemma 35. *Let u be in $L(\mathcal{S}_z)$. Then, $\pi_z(u)$ is in $\mathbb{N}$ and $0 \xrightarrow[\mathcal{S}_z]{u} \pi_z(u)$.*

Example 36. (a) *The base $\frac{3}{2}$ satisfies $p = (2q-1)$, hence $D_{\frac{3}{2}} = A_3$. In this case, $\mathcal{S}_{\frac{3}{2}}$ is simply equal to $\mathcal{T}_{\frac{3}{2}}$.*

(b) *The base $\frac{4}{3}$ satisfies $p < (2q-1)$, hence $D_{\frac{4}{3}}$ contains A_4 plus some negative digits (here only one: -1). Transitions are added to $\mathcal{T}_{\frac{4}{3}}$ in order to build $\mathcal{S}_{\frac{4}{3}}$. These transitions are drawn with a thick line in Figure 7 (page 18).*

(c) *The base $\frac{7}{3}$ satisfies $p > (2q-1)$, hence $D_{\frac{7}{3}}$ is a strict subset of A_4. The transitions labelled by the smallest two letters of A_4 are deleted from $\mathcal{T}_{\frac{7}{3}}$ in order to produce $\mathcal{S}_{\frac{7}{3}}$. These transitions are dashed in Figure 4.*

The main result of the section states that $\mathcal{S}_z$ accepts the span-words, and more precisely reads as follows.

Theorem 37. $\Lambda(\mathcal{S}_z) = c\ell(\mathsf{Spw}_z)$

The proof essentially boils down to the linearity of τ_z (the transition function of $\mathcal{T}_z$ and $\mathcal{S}_z$) as expressed by the next lemma, which follows immediately from (6) and (9).

Lemma 38. *Let n, m in $\mathbb{N}$ and x, y in $\mathbb{Z}$ and suppose that $\tau_z(n, x)$ is defined. Then, $\tau_z(m, y)$ is defined if and only if $\tau_z(n+m,\ x+y)$ is defined.*

In this case moreover, $\tau_z(n+m,\ x+y) = \tau_z(n, x) + \tau_z(m, y)$.

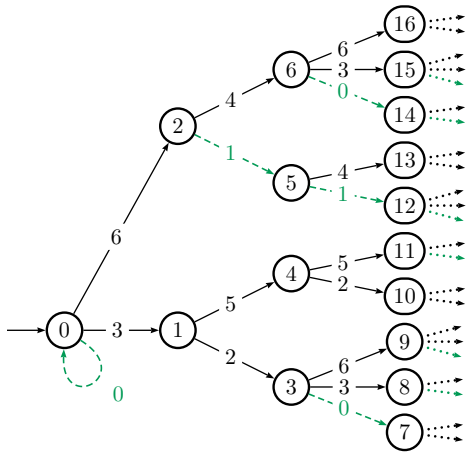


Fig. 4: Construction of $\mathcal{S}_{\frac{7}{3}}$ from $\mathcal{T}_{\frac{7}{3}}$, that is, deletion of the transitions labelled by the letters 0 and 1

Proposition 39. *Let u be in B_z^* and n and m in $\mathbb{N}$ such that $n \xrightarrow{u} m$ in $\mathcal{T}_z$. Let v be in C_z^* of the same length as u and i and j in $\mathbb{N}$. Then:*

$$(n+i) \xrightarrow[\mathcal{T}_z]{v} (m+j) \quad \Longleftrightarrow \quad i \xrightarrow[\mathcal{S}_z]{v \ominus u} j \ . \tag{10}$$

Proof: First, the statement holds if $|u| = |v| = 1$: u is then reduced to one letter b of B_z, v to one letter c of C_z, and $v \ominus u$ to the letter $(c-b)$ which belongs to D_z. By hypothesis, $\tau_z(n,b)$ is defined and equal to m, and Lemma 38 yields exactly Equation (10).

The case $u = v = \varepsilon$ is trivial. Let us suppose that $u = bu'$, $v = cv'$ and that

$$n \xrightarrow[\mathcal{T}_z]{b} n' \xrightarrow[\mathcal{T}_z]{u'} m \ .$$

If $i \xrightarrow[\mathcal{S}_z]{c-b} i' \xrightarrow[\mathcal{S}_z]{v' \ominus u'} j$ then $n+i \xrightarrow[\mathcal{T}_z]{c} n'+i'$ and $n'+i' \xrightarrow[\mathcal{T}_z]{v} m+j$, and hence $n+i \xrightarrow[\mathcal{T}_z]{cv'} m+j$. And Conversely, if $n+i \xrightarrow[\mathcal{T}_z]{c} n'+i' \xrightarrow[\mathcal{T}_z]{v'} m+j$ then $i \xrightarrow[\mathcal{S}_z]{c-b} i'$ and $i' \xrightarrow[\mathcal{S}_z]{v \ominus u} j$, and hence $i \xrightarrow[\mathcal{S}_z]{cv' \ominus bu} j$. □

Theorem 40. *Let i be an integer and w a word in D_z^*. The following are equivalent.*

(a) *There exists an integer j such that $i \xrightarrow{w} j$ is a path of $\mathcal{S}_z$.*

(b) *There exists an integer n such that w is a prefix of $\mathbf{w}^+_{n+i} \ominus \mathbf{w}^-_n$.*

Proof: (a)$\Rightarrow$(b). Let u in B_z^* and v in C_z^* such that $w = v \ominus u$ (Property 30).

Since every word in A_p^* labels a path of $\mathcal{T}_z$ (Lemma 14), there exist n and m in $\mathbb{N}$ such that $n \xrightarrow{u} m$. By hypothesis, the path $i \xrightarrow{w} j$ is in $\mathcal{S}_z$, and by the choice of u and v, Proposition 39 yields that $(n+i) \xrightarrow{v} (m+j)$. Since u is in B_z^*, it is a prefix of $\mathbf{w}^-_n$ (Property 22). Similarly, v is a prefix of $\mathbf{w}^+_{n+i}$. Hence, $w = v \ominus u$ is a prefix of $\mathbf{w}^+_{n+i} \ominus \mathbf{w}^-_n$.

(b) $\Rightarrow$ (a). Let w be a prefix of $\mathbf{w}^+_{n+i} \ominus \mathbf{w}^-_n$. We write u and v for the prefixes of length $|w|$ of $\mathbf{w}^-_n$ and $\mathbf{w}^+_{n+i}$ respectively. Hence it holds $w = v \ominus u$ (and $v = u \oplus w$). We denote by m and m' the endpoints of the paths $n \xrightarrow{u} m$ and $(n+i) \xrightarrow{v} m'$ of $\mathcal{T}_z$. Since $(n+i) \geqslant n$, it holds $m' \geqslant m$ and we write $j = m' - m$. Proposition 39 yields the existence of the path $i \xrightarrow{w} j$ in $\mathcal{S}_z$. □

Corollary 41. *For every n and i in $\mathbb{N}$, the ω-word $\mathbf{u} = \mathbf{w}^+_{n+i} \ominus \mathbf{w}^-_n$ is the label of a branch of $\mathcal{S}_z$ originating in state i.*

Theorem 37 is the direct consequence of Theorem 40 with $i = 0$, together with Lemma 3.

5 On the successor function for bottom words

We now consider the function ξ that maps the bottom word of n to the bottom word of $n+1$. This function is related to span-words by the following.

- The span-word of n is the digitwise difference of the top word of n and bottom word of n. In some sense, it is a way to transform the later into the former.
- The letter-to-letter morphism μ (previously defined in (8)) maps, for all n, the top word of n to the bottom word of $n+1$.

Using these facts, we define in Section 5.2 a label-replacement function ψ, which we apply to $\mathcal{S}_z$ and obtain a transducer $\mathcal{D}_z$. Finally we show Theorem I, restated below.

Theorem I. *Let p, q be two coprime integers such that $p > q > 1$. The infinite transducer $\mathcal{D}_z$ realises the continuous extension of ξ.*

5.1 The function ξ

Definition 42. *Let $\xi\colon \mathsf{Bot}_z \to \mathsf{Bot}_z$ be the function that maps $\mathbf{w}^-_n$ onto $\mathbf{w}^-_{n+1}$ for every n.*

The function ξ is "letter-to-letter", or "on-line" and "real-time", as stated by the following.

Lemma 43. *Let n and m be two integers. For every integer i, the prefixes of length i of $\mathbf{w}^-_n$ and of $\mathbf{w}^-_m$ are equal if and only if the prefixes of length i of $\xi(\mathbf{w}^-_n)$ and of $\xi(\mathbf{w}^-_m)$ are.*

Proof: Let u and v be the prefixes of length i of $\mathbf{w}^-_n$ and $\mathbf{w}^-_m$ respectively, and u' and v' those of $\mathbf{w}^-_{(n+1)}$ and $\mathbf{w}^-_{(m+1)}$. These four words belong to B_z^*.

If $u = v$, then $(n \cdot u)$ and $(m \cdot u)$ both exist (in $\mathcal{T}_z$). It follows from Lemma 10(b) that $n \equiv m\ [q^i]$, hence also $(n+1) \equiv (m+1)\ [q^i]$. Moreover, by definition of u', $((n+1) \cdot u')$ exists. Applying Lemma 10(a) then yields that $((m+1) \cdot u')$ exists as well. Since u' is over the lower alphabet B_z, it is a prefix of $\mathbf{w}^-_{n+1}$ (Property 22) hence $u' = v'$

Showing that $u' = v'$ implies $u = v$ is similar. □

Recall that Bot_z is dense in B_z^ω (Lemma 23). Then, it follows from Lemma 43 that ξ may be extended by continuity to a bijection $B_z^\omega \to B_z^\omega$. We still denote this function by ξ. Lemma 43 states that the

knowledge of the first i letters of an ω-word $\mathbf{w}$ is enough to compute the first i letters of $\xi(\mathbf{w})$. In other words, ξ is realised by an (infinite, letter-to-letter and sequential) transducer.

5.2 Definition of the transducer $\mathcal{D}_z$

Recall that $\mu\colon C_z \to B_z$ is the function defined by $\mu(c) = c - (p-q)$, for every c in C_z.

Definition 44. *We denote by ψ the function from D_z into $\mathfrak{P}(B_z \times B_z)$ defined by:*

$$\psi(d) \;=\; \left\{ \big(b, \mu(c)\big) \;\middle|\; b \in B_z \,,\;\; c \in C_z \,,\;\; c - b \;=\; d \right\} .$$

The function ψ may be given a more self-contained definition: the function μ extended to D_z computes the (signed) distance $\mu(d) = d - (p-q)$ of d to the middle-point of D_z and the set $\psi(d)$ is the set of all pairs (b, b') in $B_z \times B_z$ whose difference, $b' - b$, is equal to this distance.

Property 45. $\forall d \in D_z \qquad \psi(d) \;=\; \left\{\; (b, b') \;\middle|\; b' - b = d - (p-q) \;\right\}.$

The next property follows immediately.

Property 46. *For every pair of distinct d and d' in D_z, $\psi(d) \cap \psi(d') = \emptyset$.*

Definition 47. *Let $\mathcal{D}_z$ be the transducer*

$$\mathcal{D}_z \;=\; \langle\, B_z \times B_z,\, \mathbb{N},\, 0,\, \delta \,\rangle \;\;,$$

defined by $\delta(n, (b, b')) = \tau_z(n, ((b' - b) + (p-q)))$ for every n in $\mathbb{N}$ and letters b, b' of B_z. In other words,

$$\forall n, m \in \mathbb{N}\,,\;\; \forall b, b' \in B_z \qquad n \xrightarrow[\mathcal{D}_z]{(b,b')} m \qquad \Longleftrightarrow \qquad n \xrightarrow[\mathcal{S}_z]{d} m \quad \textit{and} \quad (b, b') \in \psi(d) \;\;,$$

that is, $\mathcal{D}_z$ is obtained from $\mathcal{S}_z$ by substituting every label d by $\psi(d)$.

The transitions of $\mathcal{D}_z$ are then also characterised by:

$$\forall n, m \in \mathbb{N}\,,\;\; \forall b, b' \in B_z \qquad n \xrightarrow[\mathcal{D}_z]{(b,b')} m \qquad \Longleftrightarrow \qquad q\,m \;=\; p\,n + (b' - b) + (p-q) \tag{11}$$

Example 48. (a) *In base $\frac{3}{2}$, the middle-point of D_z is $(p-q) = 1$ and it reads:*

$$\begin{aligned} \mu(0) &= -1 & \qquad \psi(0) &= \{\; 1\,|\,0 \;\} \\ \mu(1) &= 0 & \qquad \psi(1) &= \{\; 1\,|\,1,\; 0\,|\,0 \;\} \\ \mu(2) &= 1 & \qquad \psi(2) &= \{\; 0\,|\,1 \;\} \end{aligned}$$

The transducer $\mathcal{D}_{\frac{3}{2}}$ is shown in Figure 5. Since $p = 2q - 1$, it has the same underlying graph as $\mathcal{T}_z$.

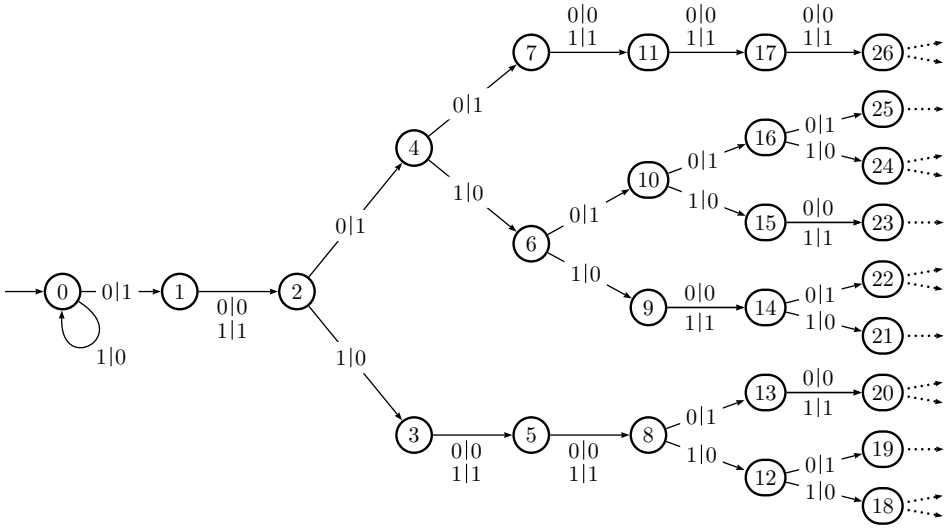


Fig. 5: $\mathcal{D}_{\frac{3}{2}}$

(b) *In base* $\frac{4}{3}$*, the middle-point is* 1 *as well and it reads:*

$$\begin{aligned} \mu(-1) &= -2 & \psi(-1) &= \{\ 2|0\ \} \\ \mu(0) &= -1 & \psi(0) &= \{\ 2|1,\ 1|0\ \} \\ \mu(1) &= 0 & \psi(1) &= \{\ 2|2,\ 1|1,\ 0|0\ \} \\ \mu(2) &= 1 & \psi(2) &= \{\ 1|2,\ 0|1\ \} \\ \mu(3) &= 2 & \psi(3) &= \{\ 0|2\ \} \end{aligned}$$

Figures 6, 7 and 8 sum up the construction of $\mathcal{D}_{\frac{4}{3}}$.

(c) *In base* $\frac{7}{3}$*,* $D_{\frac{7}{3}} = \{2, 3, 4, 5, 6\}$*, its middle-point is* 4 *and it reads:*

$$\begin{aligned} \mu(2) &= -2 & \psi(2) &= \{\ 2|0\ \} \\ \mu(3) &= -1 & \psi(3) &= \{\ 2|1,\ 1|0\ \} \\ \mu(4) &= 0 & \psi(4) &= \{\ 2|2,\ 1|1,\ 0|0\ \} \\ \mu(5) &= 1 & \psi(5) &= \{\ 1|2,\ 0|1\ \} \\ \mu(6) &= 2 & \psi(6) &= \{\ 0|2\ \} \end{aligned}$$

The transducer $\mathcal{D}_{\frac{7}{3}}$ *is shown in Figure 9; its inaccessible part is dashed out.*

5.3 Behaviour of $\mathcal{D}_{\mathcal{Z}}$

The transducer $\mathcal{D}_{\mathcal{Z}}$ is locally bijective, as both the underlying input and the underlying output automata are complete deterministic automata. More precisely:

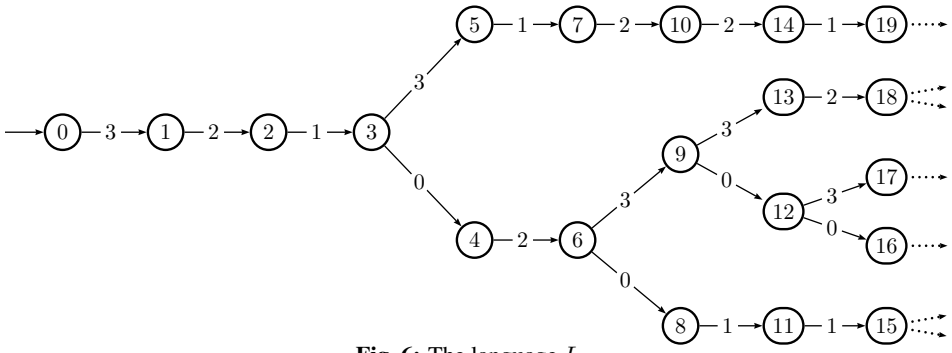


Fig. 6: The language $L_{\frac{4}{3}}$

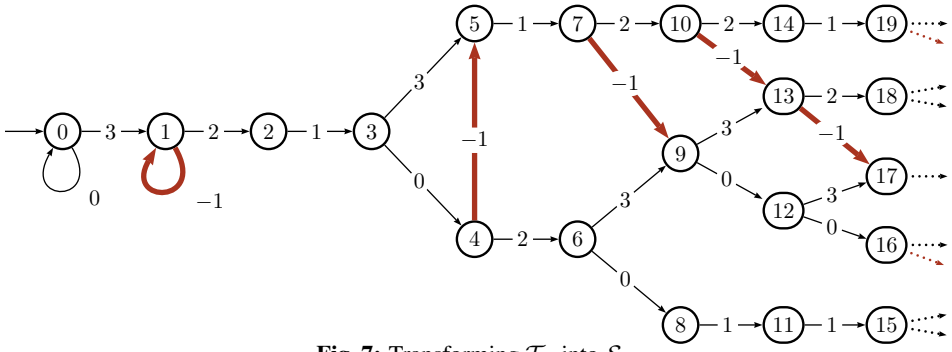


Fig. 7: Transforming $\mathcal{T}_{\frac{4}{3}}$ into $\mathcal{S}_{\frac{4}{3}}$

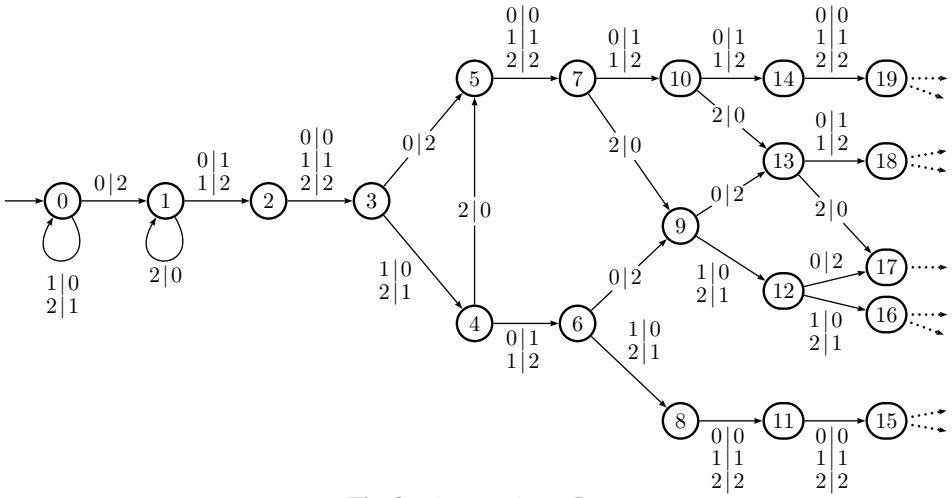


Fig. 8: The transducer $\mathcal{D}_{\frac{4}{3}}$

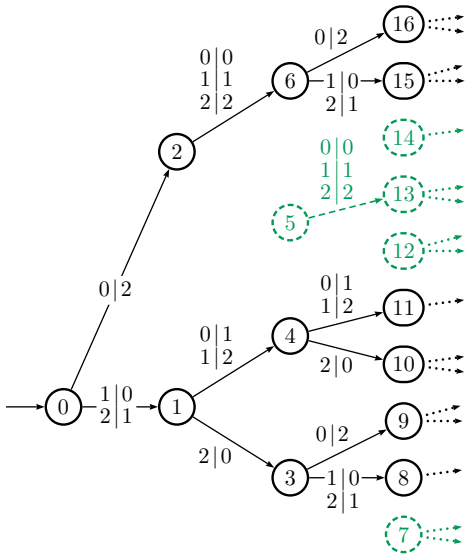


Fig. 9: The transducer $\mathcal{D}_{\frac{7}{3}}$

Lemma 49. *For every state n of $\mathcal{D}_z$ and every letter x in B_z, there exist:*

(a) *a unique transition* $n \xrightarrow[\mathcal{D}_z]{(b,x)} m$, *and* (b) *a unique transition* $n \xrightarrow[\mathcal{D}_z]{(x,b')} m'$.

Proof: (a) From (11), $n \xrightarrow[\mathcal{D}_z]{(b,x)} m$ exists if and only if $qm = pn + x - b + p - q$, that is, if and only if

$$qm + b \;=\; pn + x + p - q \;. \tag{12}$$

The unicity of the pair (m, b) in (12) follows, since b is in $B_z = \{0, 1, \ldots, q-1\}$.

A similar reasoning yields (b). □

Corollary 50. *For every state n of $\mathcal{D}_z$ and every ω-word $\mathbf{w}$ in ${B_z}^\omega$, there exist:*

(a) *a unique ω-word $\mathbf{u}$ in ${B_z}^\omega$ such that* $n \xrightarrow[\mathcal{D}_z]{(\mathbf{u},\mathbf{w})} \cdots$, *and*

(b) *a unique ω-word $\mathbf{v}$ in ${B_z}^\omega$ such that* $n \xrightarrow[\mathcal{D}_z]{(\mathbf{w},\mathbf{v})} \cdots$.

Corollary 51. *The transducer $\mathcal{D}_z$ realises a bijection: ${B_z}^\omega \to {B_z}^\omega$.*

For every i in $\mathbb{N}$, we define the transducer $\mathcal{D}_{z,\,i}$ obtained from $\mathcal{D}_z$ by changing the initial state 0 into the state i:

$$\mathcal{D}_{z,\,i} \;=\; \langle\, B_z \times B_z,\, \mathbb{N},\, i,\, \delta \,\rangle \;.$$

Theorem I is the direct consequence of the following more general statement.

Theorem 52. *For every integer n, $\mathcal{D}_{z,\,i}$ accepts the pair $(\mathbf{w}_n^-, \mathbf{w}_{n+i+1}^-)$.*

Proof: Let us write:

$$\begin{aligned} \mathbf{w}_n^- &= b_1\, b_2 \cdots && \text{(an } \omega\text{-word over } B_z) \\ \mathbf{w}_{n+i}^+ &= c_1\, c_2 \cdots && \text{(an } \omega\text{-word over } C_z) \\ \mathbf{w}_{n+i}^+ \ominus \mathbf{w}_n^- = \mathbf{u} &= a_1\, a_2 \cdots && \text{(an } \omega\text{-word over } D_z) \\ \mathbf{w}_{n+i+1}^- &= b'_1\, b'_2 \cdots && \text{(an } \omega\text{-word over } B_z) \end{aligned}$$

By Corollary 41, the ω-word $\mathbf{u}$ is the label of a branch of $\mathcal{S}_z$ originating from the state i. We write:

$$i \xrightarrow[\mathcal{S}_z]{a_1} m_1 \xrightarrow[\mathcal{S}_z]{a_2} m_2 \xrightarrow[\mathcal{S}_z]{a_3} \cdots$$

For every index k, $\mu(c_k) = b'_k$ (Lemma 25). Hence $(b_k, b'_k) = (b_i, \mu(c_k))$ satisfies the three conditions: $b_k \in B_z$, $c_k \in C_z$ and $a_k = c_k - b_k$; in other words, (b_k, b'_k) belongs to $\psi(a_k)$ (Definition 44).

It then follows from Definition 47 of $\mathcal{D}_z$ that the following branch exists in $\mathcal{D}_z$:

$$i \xrightarrow[\mathcal{D}_z]{(b_1,b'_1)} m_1 \xrightarrow[\mathcal{D}_z]{(b_2,b'_2)} m_2 \xrightarrow[\mathcal{D}_z]{(b_3,b'_3)} \cdots$$

In other words, $\mathcal{D}_{z,i}$ accepts the pair $(\mathbf{w}_n^-, \mathbf{w}_{n+i+1}^-)$. □

In particular, Theorem 52 implies, for $i = 0$, that $\mathcal{D}_z$ accepts every pair $(\mathbf{w}_n^-, \mathbf{w}_{n+1}^-)$, for n in $\mathbb{N}$. Since $\mathcal{D}_z$ is letter-to-letter (Definition 47), it realises a *continuous function*; since its domain is ${B_z}^\omega$ (Corollary 51) and since Bot_z is dense in ${B_z}^\omega$ (Lemma 23), $\mathcal{D}_z$ realises $\xi\colon {B_z}^\omega \to {B_z}^\omega$. This concludes the proof of Theorem I.

6 The set of spans

The proof of Theorem I draws the attention to the ω-words $\mathbf{s}(n) = \mathbf{w}_n^+ \ominus \mathbf{w}_n^-$ and naturally to their evaluation by the function ρ_z. For every integer n, let us write $u_n = \langle n \rangle_z$; the real number $\rho_z(u_n\, \mathbf{s}(n))$ is the length of the interval of the real line delimited, so to speak, by the 'end-points' of the ω-words $u_n \mathbf{w}_n^-$ and $u_n \mathbf{w}_n^+$ when the representation trees are drawn in a *fractal way*, as in the first Figure 1 or in the following Figure 10.

Of course, this value will decrease exponentially with the length ℓ of u_n and a reasonable 'renormalisation' consists in considering the value $\rho_z(\mathbf{s}(n))$ instead, which we call the *span of* n. In the case of a classical integer base numeration system, this notion is obviously uninteresting as this value is 1 for every n. And it is as easy to observe, for instance on Figure 10, that in a rational base numeration system, distinct integers may have distinct spans.

In this section we study the topological structure of the set of spans in a given system, and show that it depends upon whether $z = \frac{p}{q}$ is larger than 2 or not (Theorem II).

6.1 Span of a node

Notation 53. *For every integer n, we denote by V_n the set of all ω-words $\mathbf{w}$ such that $n \xrightarrow{\mathbf{w}} \cdots$ is a branch of $\mathcal{T}_z$:*

$$\mathsf{V}_n \;=\; \langle n \rangle_z^{-1} \mathsf{W}_z \;=\; \left\{ \mathbf{w} \;\middle|\; \left(\langle n \rangle_z \mathbf{w}\right) \in \mathsf{W}_z \right\} \quad .$$

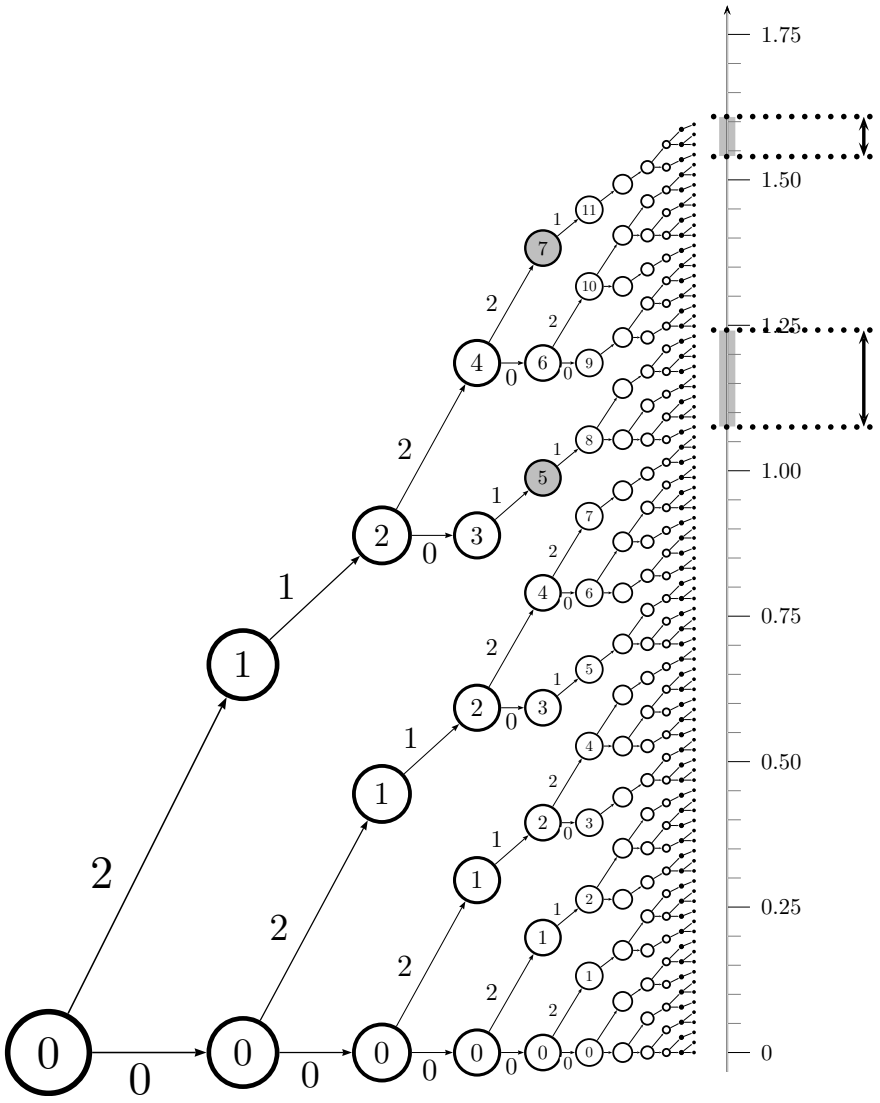


Fig. 10: Fractal drawing of real number representations in base $\frac{3}{2}$

Note that $\mathsf{V}_0 = \mathsf{W}_z$ and that for every integer n, the ω-words $\mathbf{w}_n^-$ and $\mathbf{w}_n^+$ belong to V_n. Theorem 18 states that $\rho_z(\mathsf{V}_0)$ is an interval, and next proposition extends it to any V_n.

Proposition 54. $\rho_z(\mathsf{V}_n) \;=\; \Big[\, \rho_z(\mathbf{w}_n^-),\ \rho_z(\mathbf{w}_n^+) \,\Big]$

Proof: For readability, we write $u_n = \langle n \rangle_z$ and let $\ell = |u_n|$. From the Definition 19 of bottom and top

words, every word $\mathbf{w}$ in V_n satisfies

$$\mathbf{w}_n^- \leqslant_{\text{lex}} \mathbf{w} \leqslant_{\text{lex}} \mathbf{w}_n^+ \qquad \text{hence} \qquad u_n\mathbf{w}_n^- \leqslant_{\text{lex}} u_n\mathbf{w} \leqslant_{\text{lex}} u_n\mathbf{w}_n^+ \ .$$

Conversely, since the prefix of length ℓ of any ω-words $\mathbf{v}$ such that $u_n\mathbf{w}_n^- \leqslant_{\text{lex}} \mathbf{v} \leqslant_{\text{lex}} u_n\mathbf{w}_n^+$ is u_n, it holds:

$$u_n\mathsf{V}_n = \left\{ \mathbf{v} \in \mathsf{W}_z \;\middle|\; u_n\mathbf{w}_n^- \leqslant_{\text{lex}} \mathbf{v} \leqslant_{\text{lex}} u_n\mathbf{w}_n^+ \right\} .$$

Since ρ_z preserves order on W_z (Proposition 17), it follows that $\rho_z(u_n\mathsf{V}_n)$ is an interval since $\rho_z(\mathsf{W}_z)$ is an interval.

For any ω-word $\mathbf{v}$, it holds:

$$\rho_z(u_n\mathbf{v}) = \rho_z(u_n0^\omega) + \left(\frac{p}{q}\right)^{-\ell} \rho_z(\mathbf{v}) \quad \text{hence} \quad \rho_z(\mathbf{v}) = \left(\frac{p}{q}\right)^{\ell} \rho_z(u_n\mathbf{v}) - \left(\frac{p}{q}\right)^{\ell} \rho_z(u_n0^\omega) \ .$$

It follows that $\rho_z(\mathsf{V}_n)$ is the image of the interval $\rho_z(u_n\mathsf{V}_n)$ by an affine transformation, hence an interval. □

Definition 55. (a) *For every integer n, we call* span of n*, and denote by $\sigma(n)$, the length of the interval $\rho_z(\mathsf{V}_n)$: $\sigma(n) = \rho_z(\mathbf{w}_n^+) - \rho_z(\mathbf{w}_n^-) = \rho_z(\mathbf{s}(n))$.*

(b) *We denote by $\mathbf{Span}_z$ the set of spans: $\mathbf{Span}_z = \{\sigma(n) \mid n \in \mathbb{N}\} = \{\rho_z(\mathbf{s}(n)) \mid n \in \mathbb{N}\}$.*

Since the function ρ_z is continuous, and $\Lambda(\mathcal{S}_z) = c\ell(\mathsf{Spw}_z)$ (Theorem 37), the next statement holds.

Theorem 56. $\rho_z(\Lambda(\mathcal{S}_z)) = c\ell(\mathbf{Span}_z)$.

The topological properties of the set $c\ell(\mathbf{Span}_z) \subset \mathbb{R}$ depend on whether p is smaller or greater than $2q-1$.[vi] Before stating the result, let us recall a definition. A bounded closed set that is nowhere dense and has no isolated point is called a *Cantor set*. The classical ternary Cantor set is of measure zero, but it is not necessarily the case of all Cantor sets (*cf.* Kechris, 1995).

Theorem II. *Let p, q be two coprime integers such that $p > q > 1$ and $z = \frac{p}{q}$.*

(a) *If $p \leqslant 2q-1$, then $c\ell(\mathbf{Span}_z)$ is equal to* the interval $\rho_z(\mathsf{W}_z)$.

(b) *If $p > 2q-1$, then $c\ell(\mathbf{Span}_z)$ is a Cantor set of measure zero.*

The two parts of Theorem II are shown independently in Section 6.2 and Section 6.3.

Beforehand, we give a characterisation of $c\ell(\mathbf{Span}_z)$ that holds in all cases but the status of which lies in between the two parts of Theorem II. For small bases, its proof uses a result from the **next** Section 6.2 and, this part of the statement is never applied in the following. For large bases, the proof is easy but will be used in the proof of Theorem II(b) later on. Recall that D_z is the integer interval whose length is $2q-1$ and whose largest element is $p-1$ (Definition 27).

[vi] It could seem simpler to write: 'whether z is smaller or greater than 2' which is logically equivalent since $z = 2$ defines an integer base rather than a rational base. But this would hide that the true *border case* is when $p = 2q-1$ and this case behaves sometimes like $p < 2q-1$ — as here in Theorem II — and sometimes like $p > 2q-1$ — as in Theorem 3 in Akiyama et al. (2008). Note also that p and q coprime and $p > 2q-1$ imply $p \geqslant 2q+1$.

Proposition 57. $c\ell(\mathbf{Span}_z) \;=\; \rho_z(\mathsf{W}_z \cap {D_z}^\omega)$

Proof: If $p \leqslant 2q-1$, then $A_p \subseteq D_z$ (Properties 28(c) and (b)), hence $\mathsf{W}_z \subseteq {A_p}^\omega \subseteq {D_z}^\omega$. It follows that $\mathsf{W}_z \cap {D_z}^\omega = \mathsf{W}_z$. We will see in the next Section 6.2 (Proposition 60) that $\rho_z(\Lambda(\mathcal{T}_z)) = \rho_z(\Lambda(\mathcal{S}_z))$. Finally, Theorem 56 concludes the proof in this case:

$$c\ell(\mathbf{Span}_z) \;=\; \rho_z(\Lambda(\mathcal{S}_z)) \;=\; \rho_z(\Lambda(\mathcal{T}_z)) \;=\; \rho_z(\mathsf{W}_z) \;=\; \rho_z(\mathsf{W}_z \cap {D_z}^\omega) \;.$$

If $p > 2q-1$, $\mathcal{S}_z$ is built from $\mathcal{T}_z$ by deleting the transitions labelled by $A_p \setminus D_z$. An ω-word $\mathbf{w}$ of ${A_p}^\omega$ is accepted by $\mathcal{S}_z$ if and only if 1) it is accepted by $\mathcal{T}_z$ and 2) every digit of $\mathbf{w}$ belongs to D_z. In other words:

$$\Lambda(\mathcal{S}_z) \;=\; \mathsf{W}_z \cap {D_z}^\omega \;.$$

Since by Definition 16, $\mathsf{W}_z = \Lambda(\mathcal{T}_z)$, Theorem 56 concludes the proof. □

6.2 The span-set in small bases ($p \leqslant 2q-1$)

First, we show that the shortest run reaching a given state n has the same length in $\mathcal{T}_z$ and in $\mathcal{S}_z$, that is, the fact that in this case $\mathcal{S}_z$ is obtained from $\mathcal{T}_z$ by adding new transitions does not allow nevertheless any 'shortcuts'.

Lemma 58. *Let u be in $L(\mathcal{S}_z)$ and $m = \pi_z(u)$. If $p \leqslant 2q-1$, then $|\langle m\rangle_z| \leqslant |u|$.*

Proof: By induction over the length of u. The case $u = \varepsilon$ is trivial. Let $u = u'd$ be a non-empty word over D_z that is accepted by $\mathcal{S}_z$. If $\pi_z(u) = 0$, then the lemma holds; we assume in the following that $\pi_z(u) > 0$.

We denote the run of u as follows:

$$0 \xrightarrow[\mathcal{S}_z]{u'} n \xrightarrow[\mathcal{S}_z]{d} m \quad .$$

Lemma 35 yields that $n = \pi_z(u')$ and $m = \pi_z(u) > 0$. From induction hypothesis, it holds

$$|\langle n'\rangle_z| \;\leqslant\; |u'| \quad . \tag{13}$$

Since z is a small base, A_p is included in D_z. The remainder of the proof depends on whether d belongs to A_p or to $D_z \setminus A_p$.

Case 1: $d \in A_p$. Then, the transition $n \xrightarrow{d} m$ exists in $\mathcal{T}_z$ (in addition to existing in $\mathcal{S}_z$). Since moreover $m \neq 0$, it follow that

$$\langle m\rangle_z \;=\; \langle n\rangle_z d \quad \text{and hence} \quad |\langle m\rangle_z| \;=\; |\langle n\rangle_z d| \;\leqslant\; |u'd| \;=\; |u| \;.$$

Case 2: $d \notin A_p$. The digit d belongs to $D_z \setminus A_p$, hence is negative (Property 28(c)). We apply the Euclidean division algorithm to m (Equation (1b) since $m > 0$): there exists a unique pair (n', a) in $(\mathbb{N} \times A_p)$ such that $\langle m\rangle_z = \langle n'\rangle_z a$. Thus, the state m has in $\mathcal{S}_z$ the two incoming transitions $n' \xrightarrow{b} m$ and

$n \xrightarrow{a} m$. Hence from Equation (9), qm is both equal to $n'p+b$ and $np+a$. Since a is negative and b is not, $n' < n$. Moreover, since representation in base $\frac{p}{q}$ preserves order (Proposition 5)

$$\langle n' \rangle_z \leqslant_{\text{rad}} \langle n \rangle_z \qquad \text{hence,} \qquad |\langle n' \rangle_z| \leqslant |\langle n \rangle_z| \quad . \tag{14}$$

Finally, we conclude Case 2 by applying in succession the definition of (n', a), and Equations (14) and (13):

$$|\langle m \rangle_z| = |\langle n' \rangle_z a| = |\langle n' \rangle_z| + 1 \leqslant |\langle n \rangle_z| + 1 \leqslant |u'| + 1 = |u| \quad .$$

□

Corollary 59. *For every u in $L(\mathcal{S}_z)$, there exists v in $L(\mathcal{T}_z)$ such that*

$$\pi_z(u) = \pi_z(v) \qquad \textit{and} \qquad |u| = |v| \quad .$$

Proof: With notation of Lemma 58, let $v = 0^i \langle m \rangle_z$ with the suitable number i of 0's. □

Next, we show that although $\mathcal{S}_z$ accepts more ω-words than $\mathcal{T}_z$, the extra accepted ω-words do not bring new a.r.p. values.

Proposition 60. *If $p \leqslant 2q-1$, then $\rho_z(\Lambda(\mathcal{T}_z)) = \rho_z(\Lambda(\mathcal{S}_z))$.*

Proof: Since $p \leqslant 2q-1$, A_p is included in D_z. It follows that every transition of $\mathcal{T}_z$ also appears in $\mathcal{S}_z$ and every ω-word of $\Lambda(\mathcal{T}_z)$ thus belongs to $\Lambda(\mathcal{S}_z)$ hence $\rho_z(\Lambda(\mathcal{T}_z)) \subseteq \rho_z(\Lambda(\mathcal{S}_z))$.

Let $\mathbf{w}$ be an ω-word in ${D_z}^\omega$ that is accepted by $\mathcal{S}_z$. For every integer i, we denote by w_i the prefix of $\mathbf{w}$ of length i. From Corollary 59, there exists a finite word v_i accepted by $\mathcal{T}_z$ such that $|v_i| = i$ and $\pi_z(v_i) = \pi_z(w_i)$. Since $\mathcal{T}_z$ has no dead-end (Lemma 9), there exists an ω-word $\mathbf{u}_i \in \Lambda(\mathcal{T}_z)$ that features v_i as prefix.

For every integer i, the ω-words $\mathbf{w}$ and $\mathbf{u}_i$ have respective prefixes of length i with the same value. It follows that

$$\mathsf{Abs}\left(\rho_z(\mathbf{w}) - \rho_z(\mathbf{u}_i)\right) < \sum_{n=i+1}^{\infty} \frac{\mathtt{Card}(A_p) + \mathtt{Card}(D_z)}{q} \left(\frac{p}{q}\right)^{-i} \quad .$$

Hence, $(\rho_z(\mathbf{u}_i))$ tends to $\rho_z(\mathbf{w})$ when i tends to infinity. Besides, since $\rho_z(\Lambda(\mathcal{T}_z))$ is a closed set (Theorem 18), $\rho_z(\mathbf{w})$ belongs to $\rho_z(\Lambda(\mathcal{T}_z))$. In other words, there exists an ω-word $\mathbf{v}$ in $\Lambda(\mathcal{T}_z)$ such that $\rho_z(\mathbf{v}) = \rho_z(\mathbf{w})$. Hence, $\rho_z(\Lambda(\mathcal{T}_z)) \supseteq \rho_z(\Lambda(\mathcal{S}_z))$. □

Proof of Theorem II(a)**:** Theorem 56 and Proposition 60 imply

$$c\ell(\mathbf{Span}_z) = \rho_z(\Lambda(\mathcal{S}_z)) = \rho_z(\Lambda(\mathcal{T}_z)) \ , \tag{15}$$

and $\rho_z(\Lambda(\mathcal{T}_z))$ is a closed interval by Theorem 18. □

6.3 The span-set in large bases ($p > 2q-1$)

In order to prevent any misinterpretation in case of cursory reading, we repeat the hypothesis $p > 2q-1$ in every statement. The proof is divided in two parts: Proposition 64 and Proposition 69. Let us recall first that the set $c\ell(\mathbf{Span}_z)$ is closed and bounded and then the following two properties that hold in large bases.

Property 61. *We assume $p > 2q-1$.*

(a) *Every state n of $\mathcal{T}_z$ has at least $\left\lfloor \frac{p}{q} \right\rfloor$ outgoing transitions.*

(b) *The digits of D_z are strictly positive.*

Lemma 62. *We assume $p > 2q-1$. For every integer n, it holds $0 < \gamma_z \leqslant \sigma(n) \leqslant \omega_z$, where $\omega_z = \rho_z(\mathbf{w}_0^+)$ and $\gamma_z = \rho_z(q\mathbf{w}_1^-)$.*

Proof: We denote by X the set consisting of the ω-words of W_z that do not start with the digit 0. Hence $\mathbf{w}_0^+$ and $q\mathbf{w}_1^-$ are respectively the greatest and the least ω-word of X in the lexicographic ordering. From Proposition 17 then follows that $\rho_z(X)$ is a subset of $[\gamma_z, \omega_z]$.

On the other hand, it follows from Proposition 57 that $\sigma(n)$ belongs to $\rho_z(\mathsf{W}_z \cap {D_z}^\omega)$. From Property 61 (b), D_z does not contain the digit 0, hence $\Lambda(\mathcal{S}_z) \subseteq X$ and it holds: $\sigma(n) \in \rho_z(X) \subseteq [\gamma_z, \omega_z]$. □

Lemma 63. *We assume $p > 2q-1$. For every integer n, there exist in $\mathcal{S}_z$ two branches originating from n that are labelled by ω-words with distinct a.r.p. values.*

Proof: We write $\mathbf{w} = \mathbf{w}_n^+$. Since C_z is included in D_z (Property 28 (a)), all the transitions of the branch $n \xrightarrow{\mathbf{w}} \cdots$ of $\mathcal{T}_z$ also exists in $\mathcal{S}_z$.

Since $\mathbf{w}$ is the label of a branch of $\mathcal{T}_z$, Lemma 11 yields that it is not equal to $(p-q)^\omega$. (Recall that $p-q$ is the smallest letter of C_z.) Thus, there exists a digit $a \in C_z$, $a > p-q$, a prefix u of $\mathbf{w}$ and two states n', m such that

$$n \xrightarrow[\mathcal{S}_z]{u} n' \xrightarrow[\mathcal{S}_z]{a} m \ .$$

The integer $(a-q)$ is greater than $p-(2q-1)$ (and smaller than $p-1$), hence a letter of D_z. Then, the definition of $\mathcal{S}_z$ (Equation (9)) implies that

$$n' \xrightarrow[\mathcal{S}_z]{a-q} (m-1)$$

We denote by $\mathbf{v}$ the word $\mathbf{v} = u(a-q)\mathbf{w}_{m-1}^+$, which labels a branch originating from n.

Proposition 26 (page 12) implies that the words $\mathbf{v} = u(a-q)\mathbf{w}_{m-1}^+$ and $ua\mathbf{w}_m^-$ have the same a.r.p. value. Hence it holds

$$\rho_z(\mathbf{w}) - \rho_z(\mathbf{v}) \ = \ \rho_z(ua\mathbf{w}_m^+) - \rho_z(ua\mathbf{w}_m^-) \ = \ \left(\frac{p}{q}\right)^{-|ua|} \sigma(m) \ .$$

Since z is a large base, every span is positive (Lemma 62) and the lemma holds. □

Proposition 64. *If $p > 2q-1$, the set $c\ell(\mathbf{Span}_z)$ contains no isolated point.*

Proof: Let x be a real number in $c\ell(\mathbf{Span}_z)$. There exists an ω-word $\mathbf{w} = a_0\,a_1\cdots a_i\cdots$ accepted by $\mathcal{S}_z$ such that $\rho_z(\mathbf{w}) = x$. We denote its ω-run in $\mathcal{S}_z$ as follows:

$$0 \;=\; n_0 \xrightarrow[\mathcal{S}_z]{a_1} n_1 \xrightarrow[\mathcal{S}_z]{a_2} n_2 \xrightarrow[\mathcal{S}_z]{a_3} \cdots$$

Let k be a positive integer. We apply the previous Lemma 63 to n_k : there exist two ω-words that label branches originating from n_k and that have different a.r.p. values. One of them must have a value distinct from $\rho_z(a_{k+1}a_{k+2}\cdots)$; we denote this ω-word by $\mathbf{v}$. We moreover write $\mathbf{v}_k = a_1\,a_2\cdots a_k\,\mathbf{v}$ which then satisfies the following.

$$\mathbf{v}_k \in \Lambda(\mathcal{S}_z) \tag{16}$$

$$\rho_z(\mathbf{v}_k) \;\neq\; \rho_z(\mathbf{w}) \tag{17}$$

$$\mathbf{v}_k \text{ and } \mathbf{w} \text{ have the same prefix of length } k \tag{18}$$

Theorem 56 yields that $\rho_z(\Lambda(\mathcal{S}_z)) = c\ell(\mathbf{Span}_z)$ and Equation (16) that $\rho_z(\mathbf{v}_k)$ belongs to $c\ell(\mathbf{Span}_z)$. From (17), $\rho_z(\mathbf{v}_k)$ indeed belongs to $\big(c\ell(\mathbf{Span}_z) \setminus \{x\}\big)$.

From (18), the sequence $(\mathbf{v}_k)_{k\in\mathbb{N}}$ tends to $\mathbf{w}$. Finally, since ρ_z is continuous, $(\rho_z(\mathbf{v}_k))_{k\in\mathbb{N}}$ is a sequence of $c\ell(\mathbf{Span}_z) \setminus \{x\}$ which tends to $\rho_z(\mathbf{w}) = x$. □

It remains to show that $c\ell(\mathbf{Span}_z)$ is of measure zero. Let us first recall the classical proof that the Ternary Cantor set K_3 has measure 0. The set K_3 is obtained from the interval $I_0 = [0,1]$ by successive refinements. At step n, I_n is a finite union of intervals $J_{n,j}$, every $J_{n,j}$ is divided in three intervals of equal length, and I_{n+1} is obtained by subtracting from each $J_{n,j}$ the (open) middle interval. The measure of I_n, that is, the sum of the lengths of the disjoint $J_{n,j}$ is $\left(\frac{2}{3}\right)^n$. The I_n form an infinite decreasing sequence of sets, $K_3 = \bigcap_{n\in\mathbb{N}} I_n$ and its measure is the limit of the sequence $\left(\frac{2}{3}\right)^n$, hence 0. The proof of part (b) of Theorem II follows the same scheme, loaded with some technicalities.

Lemma 65. *We assume $p > 2q-1$. Let i be an integer such that $\lfloor z \rfloor^i \geqslant 2q$. Then, for every integer n, there exists an integer m and a path $n \xrightarrow[\mathcal{T}_z]{} m$ in $\mathcal{T}_z$ of length i that does not exists in $\mathcal{S}_z$.*

Proof: Property 61 (a) states that every state has at least $\lfloor z \rfloor$ outgoing transitions in $\mathcal{T}_z$. Hence every state is the origin of at least $\lfloor z \rfloor^i$ distinct paths of length i.

Let n be a state and S the set of the states reachable from n in i steps. The cardinal of S is greater than $2q$ (previous paragraph) and S is an integer interval. Hence $S \bmod p$ visits at least $2q$ different residue classes modulo p. Since the function mapping the residue classes of a state s and the label of the unique incoming transition of s in $\mathcal{T}_z$. The incoming transitions of the states of S are labelled by at least $2q$ distinct letters. At least one of these letters does not belong to $\mathcal{D}_z$ (since it is of cardinal $2q-1$); we denote by a this letter and by m a state of S the incoming transition of which is labelled by a. The last transition of the path from n to m in $\mathcal{T}_z$ is deleted in $\mathcal{S}_z$. □

For every finite word u in $\textsc{Pre}(\mathsf{W}_z) = 0^*L_z$, we denote by Z_u the set of the ω-words that are accepted by $\mathcal{T}_z$ and that start with u: $\mathsf{Z}_u = u(u^{-1}\mathsf{W}_z)$. It is related to the sets V_n (Notation 53) by the following:

$$\forall u \in \textsc{Pre}(\mathsf{W}_z) \qquad \mathsf{Z}_u \;=\; u\mathsf{V}_n \qquad \text{where} \quad n \;=\; \pi_z(u)\ .$$

Moreover, we denote by I_u the set of the a.r.p. values of these words: $I_u = \rho_z(\mathsf{Z}_u)$. It then follows from the previous equation and Proposition 54 that

$$\forall u \in \textsc{Pre}(\mathsf{W}_z) \qquad I_u \;=\; \left[\rho_z(u\mathbf{w}_n^-),\, \rho_z(u\mathbf{w}_n^+)\right] \qquad \text{where} \quad n \;=\; \pi_z(u)\ . \tag{19}$$

When the base $\frac{p}{q}$ is large, I_u is never reduced to a single element since $(\rho_z(u\mathbf{w}_n^-) - \rho_z(u\mathbf{w}_n^-))$ is equal to $(\frac{p}{q})^{-|u|}\sigma(n)$, a positive real from Lemma 62. Note also the following properties satisfied by these intervals:

$$\forall u, v \in \textsc{Pre}(\mathsf{W}_z) \qquad u \text{ is a prefix of } v \quad\Longrightarrow\quad I_u \;\subseteq\; I_v\ . \tag{20}$$

$$\forall u, v \in \textsc{Pre}(\mathsf{W}_z) \qquad I_u \cap I_v \text{ is non-trivial} \quad\Longrightarrow\quad \left\{\begin{array}{ll} \text{either} & u \text{ is a prefix of } v \\ \text{or} & v \text{ is a prefix of } u \end{array}\right. \tag{21}$$

$$\forall u \in \textsc{Pre}(\mathsf{W}_z) \qquad I_u \;=\; \bigcup_{\substack{a \in A_p \\ ua \in \textsc{Pre}(W_z)}} I_{ua}\ . \tag{22}$$

We denote by $\mathfrak{I}$ the set of all the intervals I_u,

$$\mathfrak{I} \;=\; \left\{I_u \mid u \in \textsc{Pre}(\mathsf{W}_z)\right\} \quad , \tag{23}$$

and by refine the function $\mathfrak{P}(\mathfrak{I}) \to \mathfrak{P}(\mathfrak{I})$ defined as follows.

$$\forall \mathbb{S} \in \mathfrak{P}(\mathfrak{I}) \qquad \mathsf{refine}(\mathbb{S}) \;=\; \left\{\, I_{ud} \;\middle|\; I_u \in \mathbb{S}\ ,\ \ ud \in \textsc{Pre}(\mathsf{W}_z) \text{ and } d \in D_z \,\right\} \tag{24}$$

In (24), the variable d is taken in D_z whereas in (22) the variable a is taken in A_p. When z is a large base, D_z is strictly included A_p, hence refine is a refinement function:

$$\forall \mathbb{S} \in \mathfrak{P}(\mathfrak{I}) \qquad \left(\bigcup_{I \in \mathsf{refine}(\mathbb{S})} I\right) \;\subseteq\; \bigcup_{I \in \mathbb{S}} I\ .$$

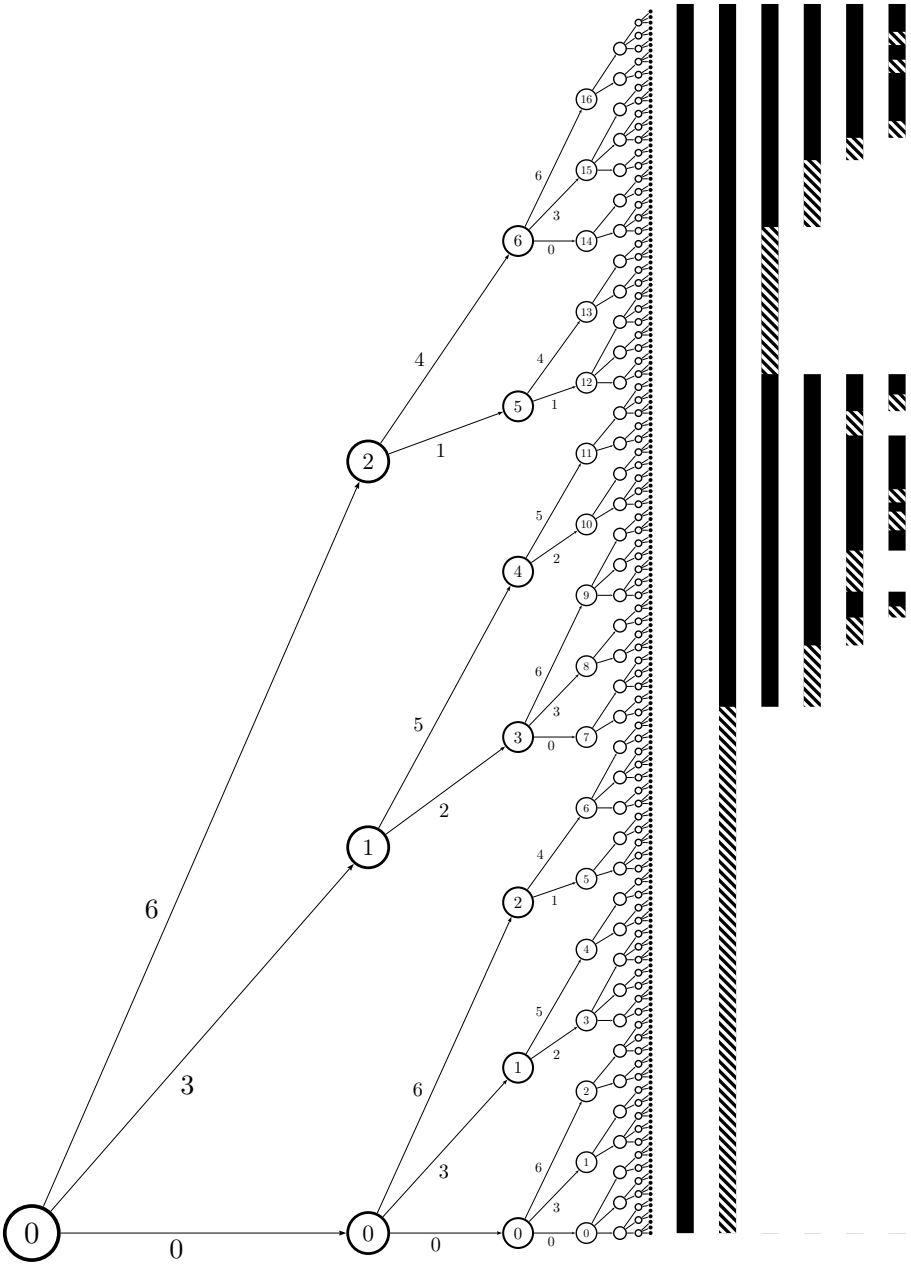
Figure 11(b) shows the successive applications of function refine to I_ε in the large base $z = \frac{7}{3}$. Hashed segments contain the points that are removed by the last application of refine.

Lemma 66. *We assume that $p > 2q - 1$. Let $\mathbb{S}_0 = \{I_\varepsilon\}$ and for every integer j, $\mathbb{S}_{j+1} = \mathsf{refine}(\mathbb{S}_j)$. Moreover, for every integer j we write $U_j = \left(\bigcup_{I \in \mathbb{S}_j} I\right)$. Then, it holds*

$$\bigcap_{j \geqslant 0} U_j \;=\; c\ell(\mathbf{Span}_z)\ .$$

Proof: Right inclusion. Let x be in $c\ell(\mathbf{Span}_z)$ and $\mathbf{w}$ a word in $c\ell(\mathsf{Spw}_z)$ such that $\rho_z(\mathbf{w}) = x$. From Theorem 37, $w \in \Lambda(\mathcal{S}_z)$. We fix an integer j and denote by u the prefix of length j of $\mathbf{w}$, hence u belongs to $D_z{}^*$. Inductively applying Equation (24) yields that

$$\mathbb{S}_j \;=\; \mathsf{refine}^j(\mathbb{S}_0) \;=\; \{I_v \mid v \in X_j\}\ , \qquad \text{with } X_j \;=\; \left\{v \in D_z{}^* \cap \textsc{Pre}(\mathsf{W}_z) \;\middle|\; |v| \;=\; j\right\} \quad .$$



(a) $\mathsf{W}_{\frac{7}{3}}$

(b) From left to right, I_ε, $\mathsf{refine}(I_\varepsilon), \ldots, \mathsf{refine}^5(I_\varepsilon)$

Fig. 11: Construction of $c\ell(\mathbf{Span}_{\frac{7}{3}})$ by successive interval deletions

It may be verified that u belongs to X_j, hence I_u to $\mathbb{S}_j$. Since by definition $I_u = \rho_z(\mathsf{Z}_u)$ and $\mathbf{w} \in \mathsf{Z}_u$, the number $x = \rho_z(\mathbf{w})$ belongs to I_u hence to U_j. Since j was taken arbitrarily, it follows that x belongs to $\bigcap_{j\geqslant 0} U_j$. Hence, it holds

$$\bigcap_{j\geqslant 0} U_j \;\supseteq\; c\ell(\mathbf{Span}_z) \ .$$

Left inclusion. Let x be a real number of $\bigcap_{j\geqslant 0} U_j$. For every integer j, the number x belongs to U_j, hence to some interval I_{u_j} of $\mathbb{S}_j$, where $u_j \in D_z{}^*$ and $|u_j| = j$. Therefore, there exists an ω-word $\mathbf{w}_j$ in W_z that starts with u_j and that evaluates a.r.p. to x. In particular, note that the first j letters of $\mathbf{w}_j$ belong to D_z.

The topology on $A_p{}^\omega$ implies that every infinite sequence has a convergent sub-sequence. We denote by $\mathbf{w}$ the limit of an arbitrary convergent sub-sequence of $(\mathbf{w}_j)_{j\in\mathbb{N}}$. Since W_z is closed, $\mathbf{w}$ is a word of W_z. Since for every integer j, the first j letters of $\mathbf{w}_j$ belong to D_z, $\mathbf{w}$ also belongs to $D_z{}^\omega$. Since ρ_z is continuous, $\rho_z(\mathbf{w}) = x$ and then, x belongs to $\rho_z(D_z{}^\omega \cap \mathsf{W}_z)$. Finally, Proposition 57 yields that

$$\bigcap_{j\geqslant 0} U_j \;\subseteq\; \rho_z(D_z{}^\omega \cap \mathsf{W}_z) \;=\; c\ell(\mathbf{Span}_z) \ .$$

□

We denote by ℓ the Lebesgue measure on $\mathbb{R}$. Then, from (19), it holds

$$\forall u \in \textsc{Pre}(\mathsf{W}_z) \quad \ell(I_u) \;=\; \left(\frac{p}{q}\right)^{-|u|} \sigma(n) \qquad \text{where } n \text{ is such that } \quad 0 \xrightarrow[\mathcal{T}_z]{u} n \ . \tag{25}$$

Applying Lemma 62 then implies the following.

Lemma 67. *If $p > 2q-1$, then for every $u \in \textsc{Pre}(\mathsf{W}_z)$, it holds*

$$0 \;<\; \gamma_z\, z^{-|u|} \;\leqslant\; \ell(I_u) \;\leqslant\; \omega_z\, z^{-|u|} \ ;$$

(recall that $\gamma_z = \rho_z(q\mathbf{w}_1^-)$ and $\omega_z = \rho_z(\mathbf{w}_0^+)$).

By abuse of notation, we use ℓ on elements of $\mathfrak{P}(\mathfrak{I})$ with the following meaning

$$\forall \mathbb{S} \in \mathfrak{P}(\mathfrak{I}) \quad \ell(\mathbb{S}) \;=\; \ell\left(\bigcup_{I\in\mathbb{S}} I\right) \ . \tag{26}$$

Lemma 68. *If $p > 2q-1$, there exists a positive integer i and a real number α, $0 < \alpha < 1$, such that for every word $u \in \textsc{Pre}(\mathsf{W}_z)$,*

$$\ell\left(\mathsf{refine}^i(I_u)\right) \;\leqslant\; \alpha\, \ell(I_u) \ .$$

Proof: We choose i as in Lemma 65. Let u be a word of $\textsc{Pre}(\mathsf{W}_z)$. We denote by k the length of u and by n the state reached by the run of u in $\mathcal{T}_z$. Then, from (24) and the contrapositive of (21), it holds

$$\ell\left(\mathsf{refine}^{\,i}(I_u)\right) \;=\; \ell\Big(\bigcup_{w\in X} I_{uw}\Big) \;=\; \sum_{w\in X} \ell(I_{uw})$$
$$\text{where}\quad X \;=\; \left\{w \in {D_z}^* \;\middle|\; |w| \;=\; i,\; uw \in \textsc{Pre}(\mathsf{W}_z)\right\} .$$

Then, Equation (22) yields that

$$\ell\left(\mathsf{refine}^{\,i}(I_u)\right) \;=\; \ell(I_u) - \sum_{w\in Y} \ell(I_{uw})$$
$$\text{where}\quad Y \;=\; \left\{w \in {A_p}^* \;\middle|\; w \notin {D_z}^*,\; |w| \;=\; i,\; uw \in \textsc{Pre}(\mathsf{W}_z)\right\} .$$

Now, we apply Lemma 65 to n : there exists a state m and a word v of length $i > 0$ such that the path $n \xrightarrow{\;v\;} m$ exists in $\mathcal{T}_z$ but does not exists in $\mathcal{S}_z$. Hence, v features a digit that belongs to $A_p \setminus D_z$. It follows that v belongs to Y, hence that the following holds.

$$\begin{aligned}
\ell\left(\mathsf{refine}^{\,i}(I_u)\right) &\leqslant \ell(I_u) - \ell(I_{uv}) \\
&\leqslant \ell(I_u)\left(1 - \frac{\ell(I_{uv})}{\ell(I_u)}\right) \\
&\leqslant \ell(I_u)\left(1 - \frac{z^{-|uv|}\gamma_z}{z^{-|u|}\omega_z}\right) \qquad \text{(Using Lemma 67)} \\
&\leqslant \alpha\,\ell(I_u) \qquad \text{with} \quad \alpha \;=\; 1 - z^{-i}\frac{\gamma_z}{\omega_z}
\end{aligned}$$

Since γ_z and ω_z are positive, then $\alpha < 1$. Since i is positive and $\gamma_z \leqslant \omega_z$, then $\alpha > 0$. □

Proposition 69. *If $p > 2q-1$, then $c\ell(\mathbf{Span}_z)$ is of measure zero.*

Proof: Let $(U_j)_{j\in\mathbb{N}}$ be the sequence defined in Lemma 66. Let i, α be the two parameters from Lemma 68. Applying the later yields

$$\forall k \in \mathbb{N} \qquad \ell(U_{ki}) \;<\; \alpha^k\,\ell(U_0) \;. \tag{27}$$

Since the sequence $\big(U_j\big)_{j\in\mathbb{N}}$ is decreasing by inclusion, the sequence $\big(\ell(U_j)\big)_{j\in\mathbb{N}}$ is decreasing. Then, (27) implies that the later sequence tends to 0 when j tends to infinity. Finally, the set $c\ell(\mathbf{Span}_z)$, which is the limit of the sequence $\big(U_j\big)_{j\in\mathbb{N}}$ (Lemma 66), is of measure zero. □

Hausdorff dimension One can go further in the comparison between the Cantor sets and the span-sets, and investigate their *Hausdorff dimension* which give more accurate information on their topological structure (*cf.* Falconer, 2014). It is known that the Hausdorff dimension of the Ternary Cantor set K_3 is $\frac{\ln 2}{\ln 3}$. Generalization of the construction that yields K_3, in which k parts out of n are kept usually results in sets with Hausdoff dimension $\frac{\ln k}{\ln n}$. In the case of $c\ell(\mathbf{Span}_z)$, we keep "in average" $2q-1$ parts out

of p, and one could expect a Hausdorff dimension of $\frac{\ln(2q-1)}{\ln p}$. We show below that this dimension is indeed strictly smaller.

Given a set F, the *d-dimensional Hausdorff measure* of F is defined by

$$\mathcal{H}^d(F) \;=\; \lim_{\varepsilon\to 0}\;\inf\left\{\;\sum_i {r_i}^d \;\middle|\; \begin{array}{l}\text{there is a countable cover of } F \text{ by balls } B_0, B_1, \dots\\ \text{such that for every } i,\ B_i \text{ has radius } r_i \text{ and } r_i < \varepsilon.\end{array}\right\}\ .$$

Then, the Hausdorff dimension of F is defined by:

$$\mathsf{dim}_{\mathcal{H}}(F) \;=\; \inf\left\{d \geqslant 0 \;\middle|\; \mathcal{H}^d(F) = 0\right\}\ .$$

Proposition 70. *If $p > 2q-1$, then $\dfrac{\ln 2}{\ln p - \ln q}$ is an upper bound for the Hausdorff dimension of $c\ell(\mathbf{Span}_z)$.*

Proof: We compute indeed an upper bound for the *Minkovski*, or *box-counting* dimension, which is known to be an upper bound for the Hausdorff dimension. Let r be a positive real number. We denote by $N(r)$ the minimal number of interval of length r required to cover $c\ell(\mathbf{Span}_z)$. Let d be a positive real number. The remainder of the proof consists in majoring $N(r)\,r^d$.

In the process of deleting edges from $\mathcal{T}_z$ to build $\mathcal{S}_z$, there are at most two surviving edges coming out from every node. Hence, at the depth i of $\mathcal{S}_z$, there are at most 2^i nodes accessible from the root. We fix i as follows:

$$i \;=\; \left\lceil \frac{\ln \omega_z - \ln r}{\ln z} \right\rceil \tag{28}$$

Note in particular that from Lemma 67, it holds:

$$\forall u \in \textsc{Pre}(\mathsf{W}_z)\,,\quad |u| \;=\; i \qquad \ell(I_u) \;<\; \omega_z z^{-i} \;<\; r\ .$$

Hence, one interval of length r is enough to cover I_u and then

$$N(r) \;<\; 2^i \;\leqslant\; 2^{\left(\frac{\ln \omega_z - \ln r}{\ln z} + 1\right)} \;=\; \eta\,(r)^{-\frac{\ln 2}{\ln z}} \qquad \text{with} \quad \eta \;=\; 2^{\frac{\omega_z + \ln z}{\ln z}}$$

Hence, $r^d N(r)$ is smaller than a constant times $r^{\left(d - \frac{\ln 2}{\ln z}\right)}$. If moreover $d > \frac{\ln 2}{\ln z}$, then $N(r)\,r^d$ tends to 0 when r tends to 0. Since for every real r, we may cover $c\ell(\mathbf{Span}_z)$ with $N(r)$ intervals of length r, it holds

$$\forall d \;>\; \frac{\ln 2}{\ln z} \qquad \mathcal{H}^d(c\ell(\mathbf{Span}_z)) \;\leqslant\; \lim_{r\to 0} N(r)\,r^d \;=\; 0\ .$$

□

In all cases different from $z = \frac{5}{2}$, the bound $\frac{\ln 2}{\ln z}$ is better (smaller) than the bound $\frac{\ln(2q-1)}{\ln p}$ that was inspired by the example of Cantor sets. This can be seen by means of some classical (though sometimes tedious) computations. The case $z = \frac{5}{2}$ is dealt with in a very similar way. In this case, every node in $\mathcal{T}_z$ possesses at most 3 surviving paths of length 2 that remains in $\mathcal{S}_z$. This yields a bound $\frac{\ln 3}{2\ln\frac{5}{2}}$ which is easily checked to be smaller than the corresponding bound $\frac{\ln 3}{\ln 5}$.

7 Conclusion

We have seen with Theorem I that the function ξ, which transforms a bottom word of $\mathcal{T}_z$ into another one, is realised by a transducer which is so to speak built upon $\mathcal{T}_z$ itself. To tell the truth, we had in mind a stronger property when we began this work.

All bottom words of $\mathcal{T}_z$ are distinct. But we conjecture that they all share something in common, that they are all of the 'same kind'. Two infinite words would be considered very naturally to be of the same kind if they can be mapped one to the other by a finite state machine. It is obviously the case for $\mathbf{w}_n^-$ and $\mathbf{w}_m^-$ if one is a suffix of the other, that is, if m is a node that is reached from n by its bottom word. We conjectured it is the case for every pair of integers n and m but were not able to prove it. We thus leave it as an open problem:

Problem 71. *Prove, or disprove, the following statement:*

Let p, q be two coprime integers such that $p > q > 1$ and $z = \frac{p}{q}$. For every integer n, there exists a finite *letter-to-letter and cosequential transducer $\mathcal{E}_n$ (which depends also on z of course) such that $\mathcal{E}_n(\mathbf{w}_n^-) = \mathbf{w}_{n+1}^-$.*

Another problem that is left open by this work is the computation of the Hausdorff dimension of the set $c\ell(\mathbf{Span}_z)$ in the cases where $p > 2q - 1$, along the line of Proposition 70. We have seen that in this cases the set $c\ell(\mathbf{Span}_z)$ may be described in a way comparable to the construction of the classical ternary Cantor set. As a result, both sets have similar topological properties (closed, bounded, empty interior, no isolated point, Lebesgue-measure zero). This comparison hence suggests that the Hausdorff dimension of $c\ell(\mathbf{Span}_z)$ could be $\frac{\ln(2q-1)}{\ln p}$. We showed an upper bound that is strictly smaller than this last value. The exact computation of the Hausdorff dimension seems to be more difficult and is the subject of ongoing work by the authors. The first attempts lead to the following conjecture.

Conjecture 72. *If $p > 2q - 1$, then the Hausdorff dimension of $c\ell(\mathbf{Span}_z)$ is equal to* $\frac{\ln(2q-1)-\ln q}{\ln p-\ln q}$.

Acknowledgments

The authors are very grateful to the unknown referee who suggested them to study the Hausdorff dimension of the span-sets and hinted at the bound from which they began to work.

The second author gratefully acknowledges the support of a Marie Skłodowska–Curie post-doctoral fellowship, co-funded by the European Union and the University of Liège (Belgium), while he was completing this work during the academic year 2016/2017.

References

S. Akiyama, C. Frougny, and J. Sakarovitch. Powers of rationals modulo 1 and rational base number systems. *Israel J. Math.*, 168:53–91, 2008.

S. Akiyama, V. Marsault, and J. Sakarovitch. Auto-similarity in Rational Base Number Systems. In *WORDS 2013*, number 8079 in Lect. Notes Comput. Sci., pages 34–45, 2013.

V. Berthé and M. Rigo, editors. *Combinatorics, Automata and Number Theory*. Number 135 in Encyclopedia Math. Appl. Cambridge University Press, 2010.

K. J. Falconer. *Fractal geometry: mathematical foundations and applications*. Wiley, third edition edition, 2014. ISBN 978-1-119-94239-9.

A. S. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.

P. Lecomte and M. Rigo. Numeration systems on a regular language. *Theory Comput. Syst.*, 34:27–44, 2001.

P. Lecomte and M. Rigo. *Abstract numeration systems*, chapter 3, pages 108–162. In Berthé and Rigo (eds.), 2010.

M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

K. Mahler. An unsolved problem on the powers of $3/2$. *J. Austral. Math. Soc.*, 8:313–321, 1968.

V. Marsault. *Énumération et numération*. PhD thesis, Télécom–ParisTech, 2016.

V. Marsault and J. Sakarovitch. The signature of rational languages. *Theoret. Computer Sci.*, 658:216–234, 2017.

D. Perrin and J.-É. Pin. *Infinite words: automata, semigroups, logic and games*, volume 141. Academic Press, 2004.

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