Super-state automata and rational trees

Frédérique Bassino, Marie-Pierre Béal, Dominique Perrin

Super-state automata and rational trees

Frédérique Bassino
Institut Gaspard Monge,
Université de Marne-la-Vallée

Marie-Pierre Béal
Institut Gaspard Monge,
Université Paris 7 et CNRS

Dominique Perrin
Institut Gaspard Monge,
Université de Marne-la-Vallée
http://www-igm.univ-mlv.fr/~{bassino,beal,perrin}

December 15, 1997

Abstract

We introduce the notion of super-state automaton constructed from another automaton. This construction is used to solve an open question about enumerative sequences of leaves of rational trees. We prove that any $\mathbb{N}$-rational sequence $s = (s_n)_{n\geq 0}$ of nonnegative numbers satisfying the Kraft inequality $\sum_{n\geq 0} s_n k^{-n} \leq 1$ is the enumerative sequence of leaves by height of a k-ary rational tree. This result had been conjectured and was known only in the case of strict inequality. We also give new proofs, based on the notion of super-state automata, to the following known result about enumerative sequences of nodes in trees: any $\mathbb{N}$-rational series t that has a primitive linear representation, such that $t_0 = 1$, $\forall n \geq 1, t_n \leq kt_{n-1}$, and whose convergence radius is strictly greater than $1/k$, is the enumerative sequence of nodes by height in a k-ary rational tree.

1 Introduction

We introduce in this paper the notion of super-state automata, which can informally be stated as follows. Let $\mathcal{A}$ be a finite automaton or a multigraph (we forget the labeling). A super-state automaton, constructed from the automaton $\mathcal{A}$, has states composed of unordered lists of states of $\mathcal{A}$ such that the list of followers of all states of a super-state can be partitioned in super-states. Compared to the automaton $\mathcal{A}$, a super-state automaton often appears to be a loss of information. Let us now assume that $\mathcal{A}$ has an initial

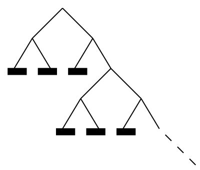

Figure 1: Tree associated to $3z^2(z^2)^*$

state i. We consider the tree that is a development of $\mathcal{A}$ from the initial state: each node of this tree is associate with one state of $\mathcal{A}$, and the sons of a node are associated to the followers of the state associated to their father, the root being associated to the initial state of $\mathcal{A}$. This tree is rational as it has only a finite number of non-isomorphic subtrees. The tree developed from the super-automaton also appears to be a loss of information compared to the previous one. Nevertheless, it keeps some interesting properties of the ordinary tree like the number of leaves or the number of nodes at each height. As its subtrees are identified to super-states, it can moreover have a more compact representation.

We use these notions of super-states to solve an open question about enumerative sequences of integers that can be realized as the enumerative sequences of leaves in a rational tree. We also give an alternative proof to a result proved in [4] about enumerative sequences of integers that can be realized as the enumerative sequences of nodes in a rational tree. These problems are linked with coding and symbolic dynamics. They can be considered as extensions of results of Huffman, Kraft, McMillan and Shannon on source coding.

Let s be an $\mathbb{N}$-rational sequence of nonnegative numbers, that is a sequence $s = (s_n)_{n \geq 1}$ such that s_n is the number of paths of length n going from an initial state to a final state in a finite multigraph or a finite automaton. We say that s satisfies the Kraft inequality for a positive integer k if $\sum_{n \geq 1} s_n k^{-n} \leq 1$. If s is the enumerative sequence of leaves of a rational k-ary tree, then s satisfies Kraft's inequality for the integer k.

In the first part of this paper, we study the converse of the above property. Consider for example the series $s(z) = 3z^2/(1-z^2)$. We have $s(1/2) = 1$ and we can obtain s as the enumerative sequence of the tree of Figure 1 associated with the prefix code $X = (aa)^*(ab + ba + bb)$ on the binary alphabet $\{a, b\}$.

Known constructions allow one to obtain a sequence s satisfying Kraft's inequality as the enumerative sequence of leaves of a k-ary tree, or as the enumerative sequence of leaves of a (perhaps not k-ary) rational tree. These two constructions lead in a natural way to the problem of building a tree both rational and k-ary. This question was already considered in [12], where it was conjectured that any $\mathbb{N}$-rational sequence satisfying Kraft's inequality is the enumerative sequence of leaves of a k-ary rational tree. The case of strict inequality was solved in [4]. In this paper, we completly settle the conjecture and the proof which we give works in both cases.

Proofs and algorithms used to establish the results are based on automata theory and on the theory of nonnegative matrices. Unlike in [3], we do not use any symbolic dynamic construction like state-splitting. But we use basic results of the Perron-Frobenius theory, and a very simple lemma, that we call the "weight lemma", due to B. Marcus in [9] (see also [8]), and already used by R. Adler, D. Coppersmith and M. Hassner in [1] to construct some finite-state codes with sliding block decoders for constrained channels.

A variant of the problem consists in replacing the enumerative sequence of leaves by the enumerative sequence of all nodes. Soittola ([15]) has characterized the series which are the enumerative sequence of nodes in a rational tree. The problem of a similar characterization for rational k-ary trees remains open in the general case. In [3], this question was solved for $\mathbb{N}$-rational series t that satisfy some necessary conditions, two trivial ones: $t_0 = 1$, $\forall n \geq 1, t_n \leq kt_{n-1}$, and a less trivial one, but proved to be necessary in [3]: the convergence radius of t is strictly greater than $1/k$, and another condition: t has a primitive linear representation. In this case there is a k-ary rational tree whose enumerative sequence of nodes by height is t.

In the second part of this paper, we give two new proofs of this result. Again, the proofs are no more based on state-splitting, but on the notion of super-state automata. With this new method, the trees obtained in a lot of examples have smaller representations.

2 Super-state automata

Let $\mathcal{A}$ be a finite state automaton (Q, E), where Q is the set of states and E the set of edges. In this paper, the labeling alphabet will always be reduced to one letter, say z, but some definitions can be extended to more general automata. So the labeling will not be represented on pictures. Automata can hence be seen as multigraphs, since several edges, (equally labeled), going from a state p to state q, may exist. Some initial or final states may also be sometimes specified.

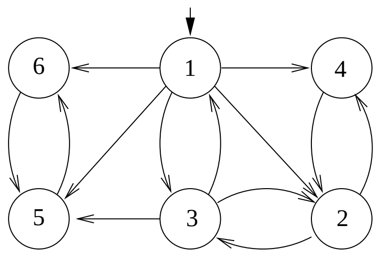

Figure 2: A 6-state automaton $\mathcal{A}$

We now give some basic definitions about trees. A *tree* T on a set of nodes N with a root r is a function $T : N - \{r\} \longrightarrow N$ which associates to each node distinct from the root its father $T(n)$, in such a way that, for each node n, there is a nonnegative integer h such that $T^h(n) = r$. The integer h is the height of the node n. A tree is k-ary if each node has at most k sons. A leaf is a node without son. A tree is said to be *rational* if it admits only a finite number of non-isomorphic subtrees.

Let $\mathcal{A}$ be an automaton with one initial state. We say that a tree is the *development* of the automaton $\mathcal{A}$ if it is constructed as follows: its root corresponds to the initial state of the graph. If a node of the tree at height n corresponds to a state i in the graph which has r outgoing edges ending in states $j_1, j_2, \ldots, j_r$, it admits r sons at height $n+1$, each of them corresponding respectively to the states $j_1, j_2, \ldots, j_r$ of the graph. The development of an automaton is a rational tree. We label the nodes with their corresponding state in $\mathcal{A}$.

Example Let us consider the 6-state automaton $\mathcal{A}$ of Figure 2, with state 1 as initial state. The development T of $\mathcal{A}$ is represented in Figure 3. If we now put in 3 boxes, respectively, the unordered sequences of states (1), (2, 5) and (4, 3, 6), we get the tree T' represented in Figure 4, that admits only 3 non-isomorphic subtrees.

This example introduces the notion of super-states and super-state automaton. In the previous example, the tree T' is a loss of information compared to tree T. But it is possible to keep in it informations like the number of nodes, or leaves at each height, or, more generally, the number of nodes that have a particular property, at each height. The gain can be, like here, a more compact representation, since we have transformed a 6-state automaton into a 3-super-state one. It can also be, as we shall see later, a way to construct rational trees that satisfy some properties.

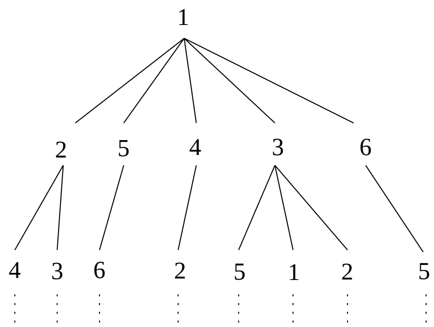

Figure 3: Development T of $\mathcal{A}$

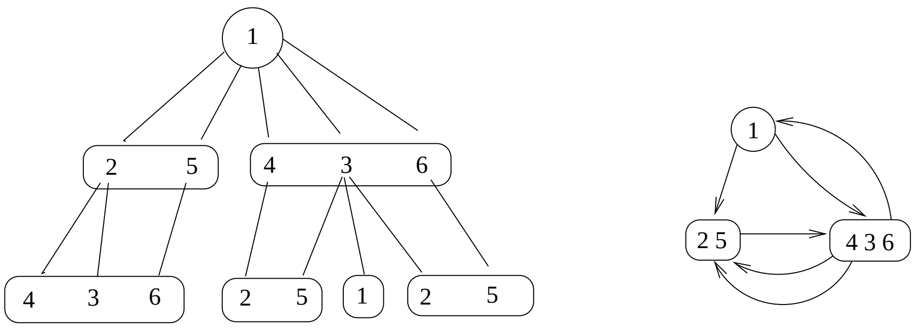

Figure 4: The tree T' and a super-automaton

In general, a *super-state automaton* associated to an automaton $\mathcal{A}$ is an automaton $\mathcal{B}$, whose states, called super-states, are chosen among unordered (or commutative) t-uples ($t \geq 1$) $(q_1, q_2, \ldots q_t)$ of states of $\mathcal{A}$, and where the edges are obtained as follows. If q is a state of $\mathcal{A}$, we denote by u_q the unordered uple obtained by concatenation of the ending states of edges of $\mathcal{A}$ going out of state q. If $(q_1, q_2, \ldots q_t)$ is a super-state, we denote by $u_{(q_1,q_2,\ldots q_t)}$ the unordered concatenation of all $u_{q_1}, u_{q_2}, \ldots, u_{q_t}$. We then partition $u_{q_1}, u_{q_2}, \ldots, u_{q_t}$ into unordered uples that are super-states, provided that such a partition exists. After the choice of the partition, we define the output edges of state $(q_1, q_2, \ldots q_t)$ in $\mathcal{B}$ as the edges of a multigraph ending in the super-states of the partition. If a super-state u appears r times in the decomposition, we have r edges from $(q_1, q_2, \ldots q_t)$ to u in the multigraph. Note that these edges are defined up to the choice of the partition.

In order to establish the first result, we shall use a particular class of super-state automata, constructed from an automaton $\mathcal{A}$ whose states have a positive *integral valuation.* We denote by $v(q)$ the valuation of a state q. We also choose and fix a positive integer m.

A super-state automaton, according to the valuation v and the integer m, is an automaton $\mathcal{B} = (Q', E')$ whose super-states are unordered (or commutative) t-uples $(q_1, q_2, \ldots q_t)$ of states of $\mathcal{A}$, with $1 \leq t \leq m$. We extend the definition of the valuation to the super-states, and, more generally, to any t-uple of states, as being the sum of the valuations of their components:

$$v((q_1, q_2, \ldots, q_t)) = \sum_{j=1}^{t} v(q_j).$$

Let $(q_1, q_2, \ldots q_t)$ be a super-state. With the previous notations, $u_{(q_1,q_2,\ldots q_t)}$ denotes the commutative concatenation of $u_{q_1}, u_{q_2}, \ldots, u_{q_t}$, where u_q is the unordered list of all followers of state q. Now we partition $u_{(q_1,q_2,\ldots q_t)}$ in several unordered t-uples ($1 \leq t \leq m$), in such a way that all parts, but possibly one, have a valuation divisible by m. Such a partition can be obtained by applying the following simple lemma, which is a key point in the state-splitting process used to construct coding schemes for constrained channels (see [8] and [5]):

Lemma 1 *(weight lemma) Let $v_1, v_2, \ldots, v_m$ be positive integers. Then there is a subset $S \subset \{1, 2, \ldots, m\}$ such that $\sum_{q \in S} v_q$ is divisible by m.*

Proof: The partial sums v_1, v_1+v_2, $v_1+v_2+v_3$, $\ldots$, $v_1+v_2+\cdots+v_m$ either are all distinct (mod m), or two are congruent (mod m). In the former case, at least one partial sum must be congruent to 0 (mod m). In the latter, there are $1 \leq p < r \leq m$ such that

$$v_1 + v_2 + \cdots + v_p \equiv v_1 + v_2 + \cdots + v_r (\bmod m)$$

Hence $v_{p+1} + v_{p+2} + \cdots + v_r \equiv 0 \pmod{m}$. □

The partition in super-states can be then obtained as follows: if $u_{(q_1,q_2,\ldots q_t)}$ has less than or exactly m (unordered) components, there is nothing to do. If not, consider the first m ones $(r_1, r_2, \ldots, r_m)$. By the weight lemma, there is a subset S of $\{1, 2, \ldots, m\}$ such that $\sum_{i \in S} v(r_i)$ is divisible by m. The t-uple composed of the r_i, with $i \in S$, is a super-state that is the first part of the partition. The process is iterated with the remaining components of $u_{(q_1,q_2,\ldots q_t)}$. We either get a decomposition in super-states whose valuation are all equal to zero modulo m, or a decomposition in super-states whose all but one valuations have this property, the last one being equal to a non-null value modulo m. After the choice of such a partition, we define the output edges in $\mathcal{B}$ of state $(q_1, q_2, \ldots q_t)$ as the edges of a multigraph ending in the super-states of the partition.

One can here remark that the automaton $\mathcal{B}$ is a finite state automaton since there is only a finite number of super-states. The t-uples are always unordered. This means that all components commute. A state of $\mathcal{A}$ can also appear several times in a same super-state as different components.

Example The super-state automaton $\mathcal{B}$ in Figure 6 is associated to the automaton $\mathcal{A}$ of Figure 5. (We only represent the part accessible from state 1). The valuation of states are represented in squares and the integer m is equal to the valuation of state 1, that is 3.

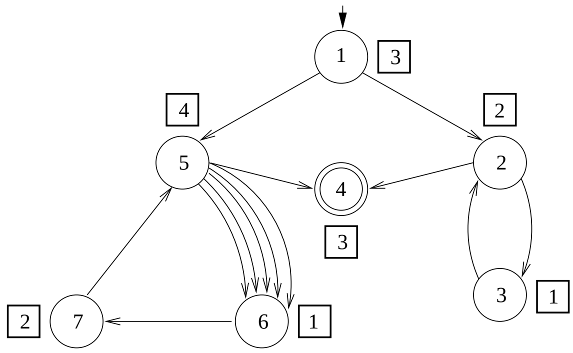

Figure 5: Automaton $\mathcal{A}$

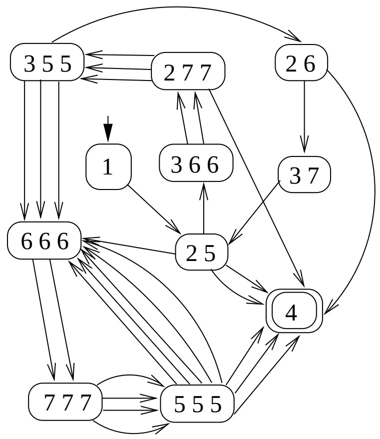

Figure 6: The super-state automaton $\mathcal{B}$

3 Rational sequences of nonnegative numbers

This section is devoted to basic definitions about rational sequences. A minimal background about the Perron-Frobenius theory of nonnegative matrices is also given.

3.1 Definitions and background

We consider sequences of nonnegative numbers. Such a sequence $s = (s_n)_{n\geq 0}$ will be said to be $\mathbb{N}$-*rational* if s_n is the number of paths of length n going from a state in I to a state in F in a finite directed graph G, where I and F are two special subsets of states, the initial and final states respectively. We say that the triple (G, I, F) is a *representation* of the sequence s. This definition is usually given for the series $\sum_{n\geq 0} s_n z^n$ instead of the sequence s. Any $\mathbb{N}$-rational sequence s satisfies a recurrence relation with integral coefficients. It is however not true that a sequence of nonnegative integers satisfying a linear recurrence relation is $\mathbb{N}$-rational (see [7] page 93).

A well known result in automata theory allows us to use a particular representation of an $\mathbb{N}$-rational sequence s. One can choose a representation (G, i, F) of s with a unique initial state i, such that no edge is entering state i and no edge is going out from any state of F. Such a representation is called a *normalized representation*. Moreover, it is possible to choose the set of final states reduced to one state (see for example [13] page 14).

If T is a tree, we denote by $l(T)$ its enumerative sequence of leaves by height, that is, the sequence of numbers s_n, where s_n is the number of leaves at height n. If T is a rational tree, this sequence is $\mathbb{N}$-rational.

The sequence $s = l(T)$ of a k-ary tree is the length distribution of a prefix code over a k-letter alphabet. The associate series $s(z) = \sum_{n \geq 0} s_n z^n$ satisfies then the Kraft inequality: $s(1/k) \leq 1$. We shall say that the Kraft strict inequality is satisfied when $s(1/k) < 1$. The equality is reached when each node of the tree has exactly zero or k sons. Conversely, the McMillan construction establishes that for any series s satisfying the Kraft inequality, there is a k-ary tree such that $s = l(T)$. Moreover, if the series satisfies the Kraft equality, then the internal nodes will have exactly k sons. But the tree obtained is not rational in general.

It is easy to see that an $\mathbb{N}$-rational sequence s is the enumerative sequence of leaves of a rational tree. This one can be obtained by developing a normalized representation of s (see section 2). The leaves of this tree correspond to the final states of the normalized representation. The maximal number of sons of a node is then equal to the maximal number of edges going out from any state of the graph of this representation.

Even if the sequence s satisfies the Kraft inequality, the above construction does not lead in general to a k-ary rational tree. The aim of the first result of this paper is to get a k-ary rational tree T such that $s = l(T)$. This result was conjectured in [12] and proved in [4] in the case of strict inequality. We shall settle it here in all cases by making use of super-state automata. Unlike the construction we gave in [4] and [3] to solve the case of strict inequality, this new method does not use any state splitting process or any symbolic dynamic construction. This proof appears to be better than the previous one for two reasons. First, it allows to solve the case of equality. Second, the rational tree obtained has, in a lot of cases, a more compact rational representation.

3.2 Approximate eigenvector

Let s be an $\mathbb{N}$-rational sequence and let (G, i, F) be a normalized representation of s. If we identify the initial state i and all final states of F in a single state still denoted i, we get a new graph denoted $\overline{G}$, which is strongly connected. The sequence s is then the length distribution of the paths of first returns to state i, that is of finite paths going from i to i without going through state i. Using the terminology of symbolic dynamics, the graph $\overline{G}$ can be seen as an irreducible shift of finite type (see, for example, [5], [6] or [8]).

We denote by M the adjacency matrix associated to the graph $\overline{G}$, that is the matrix $M = (m_{ij})_{1\leq i,j\leq n}$, where n is the number of nodes of $\overline{G}$ and where m_{ij} is the number of edges going from state i to state j. By the Perron-Frobenius theorem (see [8]), the nonnegative matrix M associated to the strongly connected graph $\overline{G}$ has a positive eigenvalue of maximal modulus denoted by λ, also called the spectral radius of the matrix. Actually, λ only depends on the series s, since $1/\lambda$ is the minimal modulus of the poles of $1/(1-s)$. It is known that the series s satisfies Kraft's strict inequality $s(1/k) < 1$ (resp. equality $s(1/k) = 1$) if and only if $\lambda < k$ (resp. $\lambda = k$).

The dimension of the eigenspace of λ is equal to one. There is a positive eigenvector (componentwise) associated to λ. When λ is an integer, the matrix admits a *positive integral* eigenvector. When $\lambda < k$, where k is an integer, the matrix admits a *k-approximate* eigenvector, that is, by definition, a positive integral vector $\mathbf{v}$ with $M\mathbf{v} \leq k\mathbf{v}$.

We shall compute approximate eigenvectors for the irreducible graphs $\overline{G}$ associated to normalized representations (G, i, T) of sequences. We associate to each node of G a value equal to the corresponding component of the approximate eigenvector of the graph $\overline{G}$. The initial and the final states will have same value since they correspond to the same state of $\overline{G}$. The computation of an approximate eigenvector can be obtained by the use of Franaszek algorithm (see for example [4]).

4 Enumerative sequence of leaves

We now state and prove, by the use of super-state automata, the first result about the enumerative sequences of leaves of rational trees:

Theorem 1 *Let $s = (s_n)_{n\geq 1}$ be an $\mathbb{N}$-rational sequence of nonnegative integers et let k be an integer such that $\sum_{n\geq 1} s_n k^{-n} \leq 1$. Then there is a k-ary rational tree such that s is the enumerative sequence by height of its leaves.*

Proof: We consider an $\mathbb{N}$-rational sequence s and an integer k such that $\sum_{n\geq 1} s_n k^{-n} \leq 1$. We begin with an automaton $\mathcal{A} = (G, i, F)$, which is a normalized representation of s. We denote by M the adjacency matrix of $\overline{G}$, and by λ its spectral radius. Hence $\lambda \leq k$. We compute a k-approximate eigenvector $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ of the graph $\overline{G}$. By definition, we have $M\mathbf{v} \leq k\mathbf{v}$. We consider $\mathbf{v}$ as a valuation, denoted by v, of the states of $\mathcal{A}$.

We define a super-state automaton $\mathcal{B}$ associated to the automaton $\mathcal{A}$, the valuation v, and the integer (used for the congruence) $m = v_i$, where i is the initial state of $\mathcal{A}$. We consider now the part of $\mathcal{B}$ accessible from the initial super-state which has, as unique component, the initial state of $\mathcal{A}$.

Let u be a super-state. Recall that $v(u)$ is the sum of the valuations of all components of u. If u is composed of n_j states j of $\mathcal{A}$, we have $v(u) = \sum_{1 \leq j \leq n} n_j v_j$. We associate to each super-state u another integer, denoted by $w(u)$, and defined by:

$$w(u) = \lceil v(u)/m \rceil.$$

Note that $w(i) = 1$.

Let us now suppose that u has t outgoing edges ending in the super-states $u_1, \ldots, u_t$. The sum of the valuations of $u_1, \ldots, u_t$ is equal to $\sum_{1 \leq j \leq n} n_j (M\mathbf{v})_j$. As $M\mathbf{v} \leq k\mathbf{v}$, we have $(M\mathbf{v})_j \leq kv_j, \forall j$. We get:

$$\begin{aligned} \sum_{j=1}^{t} v(u_j) &\leq kv(u), \\ \sum_{j=1}^{t-1} v(u_j)/m + v(u_t)/m &\leq kv(u)/m, \end{aligned}$$

By construction of the super-state automaton, $v(u_j)/m$ is an integer for $1 \leq j \leq (t-1)$. Hence we have:

$$\sum_{j=1}^{t-1} v(u_j)/m + \lceil v(u_t)/m \rceil \leq k \lceil v(u)/m \rceil.$$

Finally, we obtain:

$$\sum_{j=1}^{t} w(u_i) \leq kw(u)$$

We now consider the development of the multigraph $\mathcal{B}$. In order to get a k-ary rational tree, admitting s as enumerative sequence of leaves, we associate to each super-state u, at any height, $r = w(u)$ nodes. Since r nodes at height l have at most kr sons at height $l+1$, corresponding to the nodes associated to the super-states followers of u, it is possible to associate to each one k sons at the next height. The initial super-state itself corresponds to one node, the root of the tree. The tree is then k-ary.

The case of equality in the Kraft inequality appears to be just a particular case of the above construction. It means that we can only consider super-states whose valuation is divisible by m. The vector $\mathbf{v}$ is then an eigenvector: $M\mathbf{v} = k\mathbf{v}$. If a super-state u has t outgoing edges ending in the super-states $u_1, \ldots, u_t$, by construction of the super-state automaton, the valuations $v(u_1)$, $v(u_2), \ldots, v(u_{t-1})$ of the super-states are equal to zero modulo m.

As $M\mathbf{v} = k\mathbf{v}$, $v(u_t)$ also is divisible by m. Therefore the valuations of all super-states in the tree are divisible by m.

We choose to always leave alone the final states of $\mathcal{A}$ in a super-state. This is possible since their valuation is equal to $v(i) = m$. The leaves of the tree are then the nodes corresponding to a final state of $\mathcal{A}$.

As there is only a finite number of super-states, the tree is rational. □

Example Let s be the series defined by:

$$s(z) = \frac{z^2}{(1 - z^2)} + \frac{z^2}{(1 - 5z^3)}.$$

A normalized representation of s is given by the automaton $\mathcal{A}$ of Figure 5 (p. 7). In this figure, the valuation $v(q)$ of a state q is given in the square besides the representation of the state. Note that the final state 4 has same valuation ($v(4) = 3$) as the initial state 1.

A k-ary rational tree T, whose enumerative sequence of leaves is s, is given in Figure 7. In this figure, the components of the super-states are given inside the states. The number of small black balls above a super-state u is the number $w(u) = \lceil v(u)/3 \rceil$ of nodes of the tree represented by u. The final state 4 corresponds to the leaves of the tree.

5 Enumerative sequence of nodes

In this section, we give two new proofs of the existence of a k-ary rational tree whose enumerative sequence of nodes by height is an $\mathbb{N}$-rational sequence t that satisfies some necessary conditions like $t_0 = 1$ and $\forall n \geq 1$, $t_n \leq kt_{n-1}$, its convergence radius strictly greater than $1/k$, and another one: t has a primitive linear representation. This result has been obtained in [3] by making use of dynamic operations as an extended notion of state-splitting. The alternative proofs we give here are based on the construction of a super-state automaton. The first one does not lead, in general, to an easier construction, but it appears to be very efficient in a lot of cases. The construction which we obtain with the second proof is always simpler than the one obtained with the proof given in [3]. The trees which we have obtained in the examples with this new method have very compact representations.

Let t be an $\mathbb{N}$-rational series. A *linear representation* of t is a triple $(\mathbf{l}, M, \mathbf{c})$, where $\mathbf{l}$ is a nonnegative integral row vector, $\mathbf{c}$ is a nonnegative integral column vector, and M is a nonnegative integral matrix, with:

$$\forall n \geq 0, \qquad t_n = \mathbf{l}M^n\mathbf{c}.$$

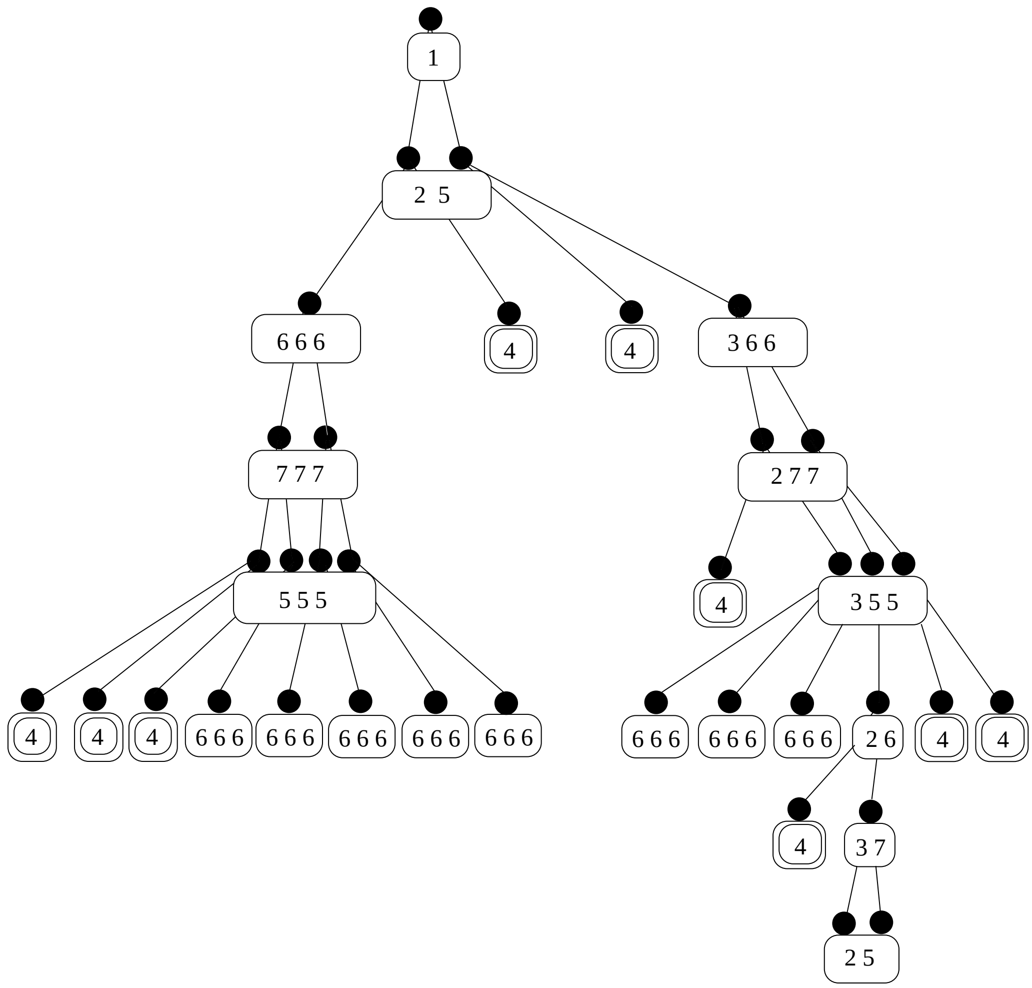

Figure 7: Tree T

The linear representation is said to be *irreducible* (resp. *primitive*) if M is an irreducible (resp. primitive) matrix. Recall that a nonnegative matrix is irreducible if for all indices i, j, there is an integer m such that $(M^m)_{ij} > 0$. The matrix is primitive if there is an integer m such that $M^m > 0$. Equivalently, the adjacency matrix of a strongly connected graph G is irreducible, and it is primitive if, moreover, the *g.c.d* of lengths of cycles in G is 1.

Let M be a $n \times n$ primitive matrix whose spectral radius is $\lambda < k$, where k is a positive integer. Let μ a positive real such that $\lambda < \mu < k$. We will denote by S_μ and S_k the following sets of *nonnegative real* vectors:

$$S_\mu = \{\mathbf{v} \in (\mathbb{R}^+)^n \mid M\mathbf{v} \leq \mu\mathbf{v}\}$$

$$S_k = \{\mathbf{v} \in (\mathbb{R}^+)^n \mid M\mathbf{v} \leq k\mathbf{v}\}.$$

We have $S_\mu \subset S_k$. Furthermore, the two sets are two simplex cones. As a consequence, they both satisfy the following properties:

$$\mathbf{v} \in S \quad \Rightarrow \quad \forall \rho \in \mathbb{R}^+, \rho\mathbf{v} \in S, \tag{1}$$

$$\mathbf{v}, \mathbf{v}' \in S \quad \Rightarrow \quad \mathbf{v} + \mathbf{v}' \in S. \tag{2}$$

Using the above notations, we state and prove the following lemma:

Lemma 2 *There is a finite subset P of integral vectors of the greater cone S_k such that all integral vectors of the smaller one S_μ is the sum of vectors of P.*

This means that the integral vectors of the big cone are finitely generated by integral vectors of the small one. The set P constitutes a Petri net for which all integral points of S_μ are accessible (see [14] for these notions).

In the geometrical proof below, we shall denote by $\mathbf{v}$ a point of $(\mathbb{R}^+)^n$. If $\mathbf{v}$ and $\mathbf{w}$ are two points, $(\mathbf{w} - \mathbf{v})$ can be seen either as a point or as the vector going from $\mathbf{v}$ to $\mathbf{w}$.

Proof: Let r be a positive integer. We denote by H_r the hyperplane of points $\mathbf{v}$ such that $v_n = r$. As the simplex S_μ and S_k are cones, the hyperplanes that limit them are not parallel to H_r.

As $\mu < k$, there is a large enough r such that for each (real) point $\mathbf{p}$ in $H_r \cap S_\mu$, one can find an *integral* point $\mathbf{u}$ in $H_r \cap S_k$, such that $\mathbf{u}' = \mathbf{p} + (\mathbf{p} - \mathbf{u})$ belongs to $H_r \cap S_\mu$.

We denote by P the finite set of all integral points $\mathbf{v} = (v_1, \ldots, v_n)$ of S_k located under the hyperplane H_{2r}, that is such that $v_n \leq 2r$. We are going to show that all integral points of S_μ are finitely generated by P.

Let us assume that the property is false, and denote by $\mathbf{w}$ an integral point of S_μ which is not the sum of integral points of P. Suppose that it is one of the closest points to the hyperplane H_r that has this property. Then $\mathbf{w}$ does not belong to P, and its last component is greater then $2r$. Let $\mathbf{p}$ be the intersection of H_r and the semi-line defined by the vector $\mathbf{w}$ and the origin. Let $\mathbf{u}$ and $\mathbf{u'}$ be defined as previously. We set $\mathbf{w'} = \mathbf{w} - \mathbf{u}$. Then $\mathbf{w'}$ is an integral point which is closer to H_r than $\mathbf{w}$. As $\mathbf{p}$ and $\mathbf{w}$ are two colinear vectors with $||\mathbf{p}|| < ||\mathbf{w}||$, and as S_μ satisfies the above properties (1) and (2), we have that $\mathbf{w'} = \mathbf{w} - \mathbf{p} + (\mathbf{p} - \mathbf{u})$ belongs to S_μ. This contradicts the hypothesis, concluding the proof of this lemma. □

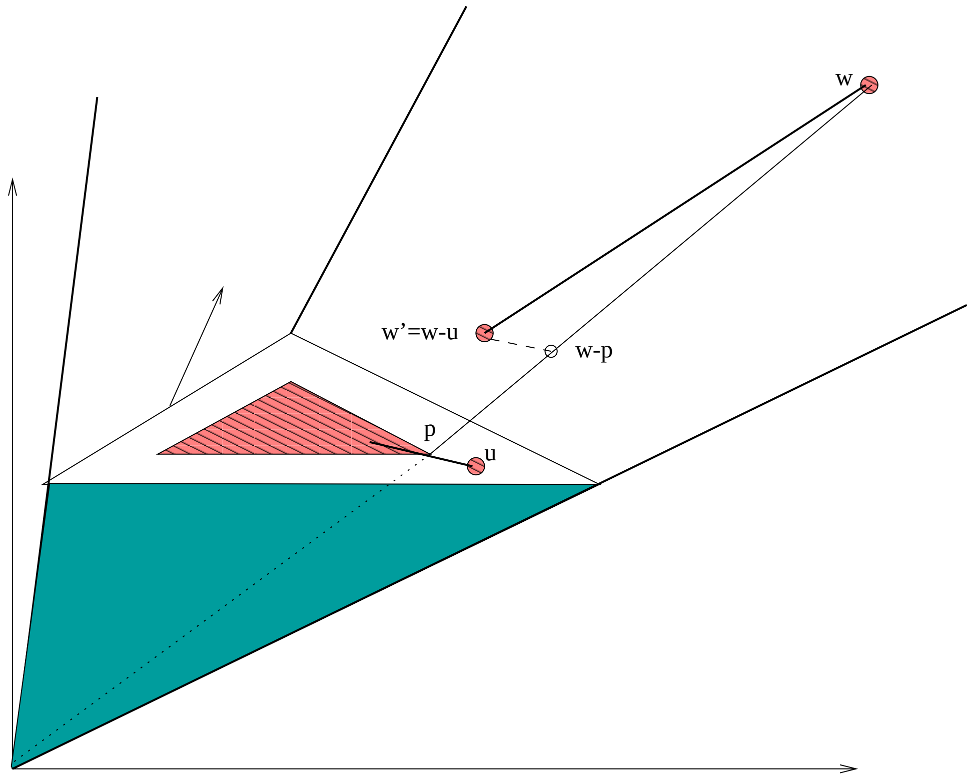

Figure 8: The geometrical lemma

We now prove the result about the enumerative sequences of nodes of rational trees. Let T be a tree. We define the enumerative sequence t of nodes by height of the tree T by $t = (t_n)_{n \geq 0}$, where t_n is the number of nodes of T at height n.

Theorem 2 *Let* $t(z) = \sum_{n\geq 0} t_n z^n$ *be an* $\mathbb{N}$*-rational series such that:*

- $t_0 = 1$.
- $\forall n \geq 1, t_n \leq k t_{n-1}$.
- *the convergence radius of* t *is strictly greater than* $1/k$ *(*$k \in \mathbb{N}^*$*).*
- t *has a primitive linear representation.*

Then $(t_n)_{n\geq 0}$ *is the enumerative sequence of nodes by height in a* k*-ary rational tree.*

First proof: We denote by $1/\lambda$ the convergence radius of t. Let $\mathbf{i}, M, \mathbf{c}$ be matrices with nonnegative integral entries such that $(\mathbf{i}, M, \mathbf{c})$ is a primitive linear representation of t, *i.e.*

$$\forall n \geq 0, \qquad t_n = \mathbf{i} M^n \mathbf{c}.$$

This representation defines an automaton $\mathcal{A}$ with n states, where M is the adjacency matrix of the multigraph.

As the matrix M is primitive with spectral radius λ, the sequence $((M/\lambda)^n)_{n\geq 1}$ tends towards a positive matrix N. Let μ be a real such that $\lambda < \mu < k$. Let S_μ and S_k be the simplex cones defined as follows:

$$S_\mu = \{\mathbf{x} \in (\mathbb{R}^+)^n \mid \mathbf{x}M \leq \mu \mathbf{x}\}$$

$$S_k = \{\mathbf{x} \in (\mathbb{R}^+)^n \mid \mathbf{x}M \leq k \mathbf{x}\}.$$

Note for example that $S_k = \{\mathbf{x} \in (\mathbb{R}^+)^n \mid M^t \mathbf{x}^t \leq \mu \mathbf{x}^t\}$. We obtain by the geometrical lemma a finite set P of integral points of S_k generating the integral points of S_μ. Since P is a finite set, there is an integer n_0 such that:

$$\forall \mathbf{x} \in P \cup \{\mathbf{i}\}, \qquad \mathbf{x} M^{n_0} \in S_\mu.$$

We define a super-state automaton $\mathcal{B}$ associated to the automaton $\mathcal{A}$. In order to do that, we identify a nonnegative integral vector $\mathbf{x} = (x_1, \ldots, x_n)$ to a t-uple composed of x_j states j of $\mathcal{A}$, for all $1 \leq j \leq n$. We now define the super-states of $\mathcal{B}$ as the integral points of $\bigcup_{j=0}^{n_0-1} \{\mathbf{x}M^j, \mathbf{x} \in P \cup \{\mathbf{i}\}\}$. Note that there is a finite number of such points. If $\mathbf{u}$ is a super-state in $\bigcup_{j=0}^{n_0-2} \{\mathbf{x}M^j, \mathbf{x} \in P \cup \{\mathbf{i}\}\}$, we define the list of its followers in $\mathcal{B}$ as the unique super-state $\mathbf{u}M$. If $\mathbf{u}$ is a super-state in $\{\mathbf{x}M^{n_0-1}, \mathbf{x} \in P \cup \{\mathbf{i}\}\}$, $\mathbf{u}M$ belongs to S_μ. As a consequence of the geometrical lemma, it is a sum of points $\mathbf{u_1}, \ldots, \mathbf{u_t} \in P$. We define the list of the super-states followers of $\mathbf{u}$ as $\mathbf{u_1}, \ldots, \mathbf{u_t}$.

If $\mathbf{u}$ is a super-state, we claim that either $\mathbf{u}$ is one of the points of $\{\mathbf{i}M^j, 0 \leq j \leq (n_0 - 1)\}$, or $\mathbf{u} \in S_k$. Actually if $\mathbf{u} = \mathbf{i}M^{n_0}$, it belongs to $S_\mu \subset S_k$. And if $\mathbf{u} \in S_k$, $\mathbf{u}M^j \in S_k, \forall j \geq 1$.

We define a tree rooted by the initial super-state $\mathbf{i}$, by developing the super-state automaton $\mathcal{B}$. We associate to each super-state $\mathbf{u}$ an integer $w(\mathbf{u})$ defined as the weighted number of final states contained in the super-state $\mathbf{u}$:

$$w(\mathbf{u}) = \mathbf{u} \cdot \mathbf{c},$$

where $\mathbf{c}$ is the column vector of the linear representation of t. If $\mathbf{u} \in S_k$, $\mathbf{u}M \leq k\mathbf{u}$, and $\mathbf{u}M \cdot \mathbf{c} \leq k\mathbf{u} \cdot \mathbf{c}$. If $\mathbf{u} = \mathbf{i}M^j, 0 \leq j \leq (n_0 - 1)$, $w(\mathbf{u}) = \mathbf{i}M^j\mathbf{c} = t_j$. As $\forall j \geq 1$, $t_j \leq kt_{j-1}$, $\mathbf{u}M \cdot \mathbf{c} \leq k\mathbf{u} \cdot \mathbf{c}$. Hence we get that for any super-state $\mathbf{u}$ whose followers in $\mathcal{B}$ are the super-states $\mathbf{u_1}, \ldots, \mathbf{u_t}$:

$$\mathbf{u}M\mathbf{c} = \sum_{j=1}^{t} \mathbf{u_j} \cdot \mathbf{c} \leq k\mathbf{u} \cdot \mathbf{c}$$

or equivalently:

$$\sum_{j=1}^{t} w(\mathbf{u_j}) \leq kw(\mathbf{u}).$$

Thus to each super-state $\mathbf{u}$ is associated $w(\mathbf{u})$ nodes. Since r nodes at a height l have at most kr sons at height $l+1$, corresponding to the nodes associated to the super-states followers of $\mathbf{u}$, it is possible to associate to each one at most k sons at the next height. The initial super-state itself ($\mathbf{i}$) corresponds to one node, the root of the tree, since $\mathbf{i} \cdot \mathbf{c} = t_0 = 1$. This defines a k-ary rational tree T admitting t as enumerative sequence of nodes by height. □

Example Let t be the series, which has the automaton of Figure 9 as

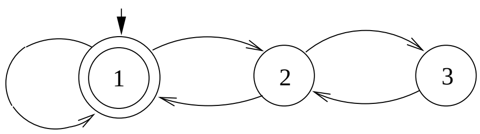

Figure 9: Primitive representation of t

primitive representation. Its convergence radius is greater than $1/k$, where

$k = 2$. We have $\mathbf{i} = (1, 0, 0)$ and $\mathbf{c} = (1, 0, 0)^t$. The adjacency matrix of the graph is

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The simplex cone $S_k = \{\mathbf{x} \in (\mathbb{R}^+)^n \mid \mathbf{x}M \leq k\mathbf{x}\}$ is the set of points $\mathbf{x} = (x_1, x_2, x_3)$ with:

$$\begin{aligned} x_2 &\leq x_1 \\ x_2 &\leq 2x_3 \\ x_1 + x_3 &\leq 2x_2 \end{aligned}$$

A tree, whose enumerative sequence of nodes is t, is given in the left part of Figure 10. Note that the super-states (1) and (1, 2) are not in S_k. Another one is given in the right part of Figure 10. This tree has only 3 super-states and also only 3 non-isomorphic subtrees. We can remark that such a compact representation could not be obtained on this example with the alternative state-splitting proof given in [3].

We now give a second proof, based on super-state automata, for the same result.

Second proof: We denote by $1/\lambda$ the convergence radius of t. Let $\mathbf{i}, M, \mathbf{c}$ be matrices with nonnegative integral entries such that $(\mathbf{i}, M, \mathbf{c})$ is a primitive linear representation of t, *i.e.*

$$\forall n \geq 0, \qquad t_n = \mathbf{i}M^n\mathbf{c}.$$

As M is primitive, there is an integer n_0 such that $M^{n_0}\mathbf{c}$ belongs to $S_k = \{\mathbf{x} \in (\mathbb{R}^+)^n \mid M\mathbf{x} \leq k\mathbf{x}\}$:

$$MM^{n_0}\mathbf{c} \leq kM^{n_0}\mathbf{c}$$

Let us denote by $\mathbf{d} = M^{n_0}\mathbf{c}$, and by t' the sequence obtained from t by n_0 shifts:

$$\forall n \geq 0, \qquad t'_n = t_{n+n_0} = \mathbf{i}M^nM^{n_0}\mathbf{c} = \mathbf{i}M^n\mathbf{d}.$$

The sequence t' admits $(\mathbf{i}, M, \mathbf{d})$ as primitive linear representation. We call $\mathcal{A}$ the multigraph whose adjacency matrix is M. We have $t'_0 = t_{n_0} = \mathbf{i} \cdot \mathbf{d}$.

We define a super-state automaton $\mathcal{B}$ whose states contain only one occurence of one state of $\mathcal{A}$. Note that the number of super-states is equal

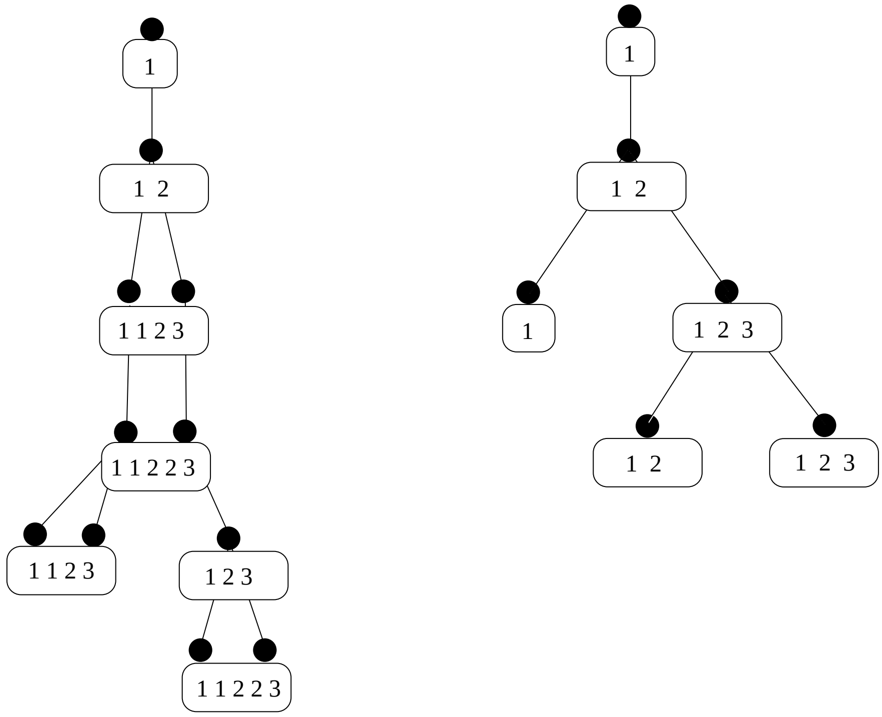

Figure 10: Two trees solution

to the number of states of $\mathcal{A}$. The followers of a super-state is then the list of its followers in $\mathcal{A}$.

We define a tree rooted by the initial super-state $\mathbf{i}$, by developing the super-state automaton $\mathcal{B}$. We associate to each super-state $\mathbf{u}$ an integer $w(\mathbf{u})$ defined as the weighted number of final states contained in the super-state $\mathbf{u}$:

$$w(\mathbf{u}) = \mathbf{u} \cdot \mathbf{d}.$$

As $M\mathbf{d} \leq k\mathbf{d}$, we get for each super-state $\mathbf{u}$:

$$\mathbf{u}M\mathbf{d} \leq k\mathbf{u} \cdot \mathbf{d}.$$

Hence we get that for any super-state $\mathbf{u}$ whose followers in $\mathcal{B}$ are the super-states $\mathbf{u_1}, \ldots, \mathbf{u_t}$:

$$\mathbf{u}M\mathbf{d} = \sum_{j=1}^{t} \mathbf{u_j} \cdot \mathbf{d} \leq k\mathbf{u} \cdot \mathbf{d}$$

or equivalently:

$$\sum_{j=1}^{t} w(\mathbf{u_j}) \leq kw(\mathbf{u}).$$

We associate to each super-state $\mathbf{u}$ $w(\mathbf{u})$ nodes. The initial super-state itself ($\mathbf{i}$) corresponds to t_{n_0} nodes, since $\mathbf{i} \cdot \mathbf{d} = t_{n_0}$. As r nodes at a height l have at most kr sons at height $l+1$, corresponding to the nodes associated to the super-states followers of $\mathbf{u}$, we associate to each one at most k sons in such a way that any node at the next height has one father. We finally complete the first n_0 levels to get a k-ary rational tree T admitting t as enumerative sequence of nodes by height. □

References

[1] R. L. Adler, D. Coppersmith, and M. Hassner. Algorithms for sliding block codes. *I.E.E.E. Trans. Inform. Theory*, IT-29:5–22, 1983.

[2] F. Bassino. *Séries rationnelles et distributions de longueurs.* Thèse, Université de Marne-la-Vallée, 1996.

[3] F. Bassino, M.-P. Béal, and D. Perrin. Enumerative sequences of leaves and nodes in rational trees. submitted, 1997.

[4] F. Bassino, M.-P. Béal, and D. Perrin. Enumerative sequences of leaves in rational trees. In *L.N.C.S. ICALP'97.* Springer-Verlag, 1997.

[5] M.-P. Béal. *Codage Symbolique.* Masson, 1993.

[6] M.-P. Béal and D. Perrin. Symbolic dynamics and finite automata. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 2, chapter 10. Springer-Verlag, 1997.

[7] J. Berstel and C. Reutenauer. *Rational Series and their Languages.* Springer-Verlag, 1988.

[8] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge, 1995.

[9] B. Marcus. Factors and extensions of full shifts. *Monatsh. Math*, 88:239–247, 1979.

[10] B. Marcus. Sofic systems and encoding data. *I.E.E.E. Trans. Inform. Theory*, IT-31:366–377, 1985.

[11] D. Perrin. Arbres et séries rationnelles. *C.R.A.S. Paris, Série I*, 309:713–716, 1989.

[12] D. Perrin. A conjecture on rational sequences. In R. Capocelli, editor, *Sequences*, pages 267–274. Springer-Verlag, 1990.

[13] D. Perrin. Finite automata. In J. Van Leeuven, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 1. Elsevier, 1990.

[14] C. Reutenauer. *Aspects Mathématiques des Réseaux de Petri.* Masson, 1989.

[15] A. Salomaa and M. Soittola. *Automata-theoretic Aspect of Formal Power Series.* Springer-Verlag, Berlin, 1978.