Enumerative sequences of leaves and nodes in rational trees

Frédérique Bassino, Marie-Pierre Béal, Dominique Perrin

Enumerative sequences of leaves and nodes in 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}`

Abstract

We prove that any $\mathbb{N}$-rational sequence $s = (s_n)_{n\geq 1}$ of nonnegative integers satisfying the Kraft strict inequality $\sum_{n\geq 1} s_n k^{-n} < 1$ is the enumerative sequence of leaves by height of a rational k-ary tree. We give an efficient algorithm to get a k-ary rational tree. Particular cases of this result had been previously proven. We give some partial results in the case of equality. Especially we solve this question when the associated sequence of internal nodes has a primitive linear representation.

1 Introduction

This paper is a study of problems linked with coding and symbolic dynamics. The results can be considered as an extension of the old results of Huffman, Kraft, McMillan and Shannon on source coding. We actually prove results on rational sequences of integers that can be realized as the enumerative sequence of leaves or nodes in a rational tree.

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 $\sum_{n\geq 1} s_n k^{-n} \leq 1$.

A rational tree is a tree which has only a finite number of non-isomorphic subtrees. 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 this paper, we study the converse of the above property. Consider for example the series $s(z) = \frac{3z^2}{1-z^2}$. We have $s(1/2) = 1$ and we can obtain s as the enumerative sequence of the tree of the figure below associated with the prefix code $X = (aa)^*(ab + ba + bb)$ on the binary alphabet $\{a, b\}$.

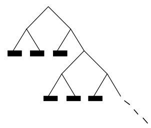

Fig. 1. Tree associated to $3z^2(z^2)^*$

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 [10], 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.

In this paper, we prove this conjecture in the case where the sequence satisfies Kraft's inequality with a strict inequality. Proofs and algorithms used to establish this result are based on automata theory and symbolic dynamics. In particular, we use the state splitting algorithm which has been introduced by R. Adler, D. Coppersmith and M. Hassner in [1] to solve coding problems for constrained channels by constructing finite-state codes with sliding block decoders. This was partly based on earlier work of B. Marcus in [7]. A variant of Franaszek's algorithm of computation of an approximate eigenvector makes the algorithms practical.

A variant of the problem considered here consists in replacing the enumerative sequence of leaves by the enumerative sequence of all nodes. Soittola ([12]) has characterized the series which are the enumerative sequence of nodes in a rational tree. We prove that any $\mathbb{N}$-rational sequence t that satisfies some necessary conditions: $t_0 = 1$, $\forall n \geq 1, t_n \leq kt_{n-1}$, the convergence radius of t is strictly greater than $1/k$, and such that t has a primitive linear representation, is the enumerative sequence of nodes by height of a k-ary rational tree. The proof of this result is based on an extended notion of state splitting that allows us to output split states without outgoing edges. The problem of a similar characterization for rational k-ary trees remains open in the general case.

The paper is organized as follows. We first give basic definitions and properties of rational objects, sequences and trees. We then give some definitions coming from the theory of symbolic dynamics. We define the notions of state splitting, approximate eigenvector and recall the algorithm of [1]. In Section 3, we establish the announced result (Theorem 1) concerning enumerative sequences of leaves and an give example for the construction. Next we give an efficient algorithm, that is a variant of Franaszek's algorithm, to find approximate eigenvectors. Section 5 devoted to enumerative sequences of nodes gives an extension (Theorem 2) of the main result of Section 3.

A preliminary shorter version of this paper was presented at the ICALP'97.

2 Definitions and background

2.1 Rational sequences of nonnegative numbers

We denote by G a directed graph with E as its set of edges. We actually use multigraphs instead of ordinary graphs in order to be able to have several distinct edges with the same origin and end. Formally a multigraph is given by two sets E (the edges) and V (the vertices) and two functions from E to V which define the origin and the end of an edge. An edge in a multigraph going from p to q will be noted (p, x, q) where $x \in \mathbb{N}$. This is equivalent to number the edges going from p to q in order to distinguish them. We shall always say "graph" instead of "multigraph".

In this paper, 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 integer coefficients. However, it is not true that a sequence of nonnegative integers satisfying a linear recurrence relation is $\mathbb{N}$-rational. An example can be found in [5] p. 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 and such that :

- no edge is coming in state i
- no edge is going out of any state of F.

Such a representation is called a *normalized representation*. Moreover, it is possible to reduce to one state the set of final states (see for example [11] p. 14).

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. We denote by $l(T)$ the enumerative sequence of its leaves by height, that is the sequence of numbers s_n, where s_n is the number of leaves of T at height n. A tree is said to be *rational* if it admits only a finite number of non isomorphic subtrees. If T is a rational tree, the sequence $l(T)$ is an $\mathbb{N}$-rational sequence.

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 $s(z) = \sum_{n\geq 1} s_n z^n$ satisfies then Kraft's inequality : $s(1/k) \leq 1$. We shall say that Kraft's 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 Kraft's inequality, there is a k-ary tree such that $s = l(T)$. Moreover, if the series satisfies Kraft's equality, then the internal nodes will have exactly k sons. But the tree obtained is not rational in general.

It is also easy to see that an $\mathbb{N}$-rational sequence is the enumerative sequence of the leaves of a rational tree. A normalized representation can be used to do that by "developing" the tree. The root will correspond 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 will admit 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 leaves of the tree will correspond to the final states of the normalized representation. The maximal number of sons of a node we get is then equal to the maximal number of edges going out of any state of the graph of this representation.

If s satisfies Kraft's inequality, the above construction does not lead in general to a k-ary rational tree. The aim of this paper is to get a k-ary rational tree T such that $s = l(T)$. This result was conjectured in [10]. We solve it for all $\mathbb{N}$-rational sequences satisfying Kraft's strict inequality and give a weaker result for the case of equality.

2.2 Approximate eigenvector and state splitting

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, [3], [4] or [6]).

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 [6]), the positive 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, $1/\lambda$ is the minimal modulus of the poles of $\frac{1}{1-s}$. The dimension of the eigenspace of λ is equal to one. There is a positive eigenvector (componentwise) associated to λ. Moreover, if there is a positive eigenvector associated to an eigenvalue ρ, then $\rho = \lambda$.

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 $M\mathbf{v} \leq k\mathbf{v}$.

For example the left side of the figure below gives a representation (G, i, F) of the serie $s(z) = 3z^2/(1 - z^2)$, and the right side gives the associated graph $\overline{G}$. The adjacency matrix of $\overline{G}$ is

$$M = \begin{pmatrix} 0 & 3 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Its maximal eigenvalue is $\lambda = 2$. The components of a positive integral eigenvector are written on the nodes.

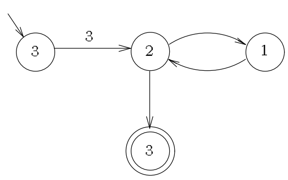

Fig. 2. Representation (G, i, F)

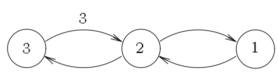

Fig. 3. Graph $\overline{G}$

Proposition 1 *If s satisfies Kraft's inequality $s(1/k) \leq 1$, then $\lambda \leq k$. In the case of equality where $s(1/k) = 1$ we have $\lambda = k$.*

For a proof, we refer the reader to [3], [4] or [6].

We now define the operation of *output state splitting* in a graph $G = (V, E)$. Let q be a vertex of G and let I (resp. O) be the set of edges coming in q (resp. going out of q). Let $O = O' + O''$ be a partition of O. The operation of *(output) state splitting* relative to (O', O'') transforms G into the graph $G' = (V', E')$ where $V' = (V \setminus \{q\}) \cup q' \cup q''$ is obtained from V by splitting state q into two states q' and q'', and where E' is defined as follows:

1. All edges of E that are not incident to q are left unchanged.
2. The both states q' and q'' have the same input edges as q.
3. The output edges of q are distributed between q' and q'' according to the partition of O into O' and O''. We denote U' and U'' the sets of output edges of q' and q'' respectively :
$U' = \{(q', x, p) \mid (q, x, p) \in O'\}$ and $U'' = \{(q'', x, p) \mid (q, x, p) \in O''\}$.

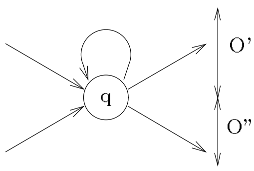

Fig. 4. Graph G

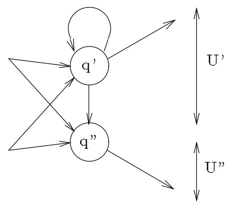

Fig. 5. Graph G'

Let us now assume that $\mathbf{v}$ is a k-approximate eigenvector for the graph G. We denote by v_p the component of index p of $\mathbf{v}$. All components v_p are positive integers. A state splitting of a state q is said to be *admissible* according to k, if the partition in O' and O'' is such that O' and O'' are not empty and:

$$k \text{ divides } \sum_{(q,x,r) \in O'} v_r$$

If the state splitting is admissible (according to k), the vector $\mathbf{v}'$ defined as follows will be a k-approximate eigenvector for the new graph G'. If p is a state distinct from q' and q'' then $v'_p = v_p$. For states q' and q'' we have:

$$v'_{q'} = \frac{1}{k} \sum_{(q,x,r) \in O'} v_r \qquad \text{and} \qquad v'_{q''} = v_q - v'_{q'}.$$

By the state splitting construction, one can check that $M'\mathbf{v}' \leq k\mathbf{v}'$, where M' is the adjacency matrix of G'.

The state splitting algorithm of [1] ensures that there is a finite number of state splittings leading to a k-ary graph, that is a graph such that at most k edges are going out of any state. For the sake of completeness, we briefly recall the proof. If there is a state q which admits more than k edges going out of it, we choose k of them and denote by $r_1, r_2, \ldots, r_k$ the sequence of end states of these edges. We then choose a subset O' of these k edges such that k divides $\sum_{(q,x,r)\in O'} v_r$. This is always possible. Indeed, by considering the $k+1$ numbers $v_{r_1}, v_{r_1}+v_{r_2}, \ldots, v_{r_1}+v_{r_2}+\cdots v_{r_k}$, we can see that at least two of them are equal modulo k, and then their difference is equal to zero modulo k. The partition of the output edges of q in O' and O'' leads to an admissible state splitting and v'_q is strictly less than v_q. This point ensures that the process stops after a finite number of splits, the final number of states being bounded by the sum of the components of the initial approximate eigenvector. The final graph obtained is k-ary.

We shall compute approximate eigenvectors for the strongly connected graphs $\overline{G}$ associated to normalized representations (G, i, F) of sequences. We shall then perform admissible state splittings that can be seen either on the graph G or on the graph $\overline{G}$. To do that, we shall 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 $\overline{G}$.

3 Enumerative sequences of leaves

We now state the result in the case of Kraft strict inequality.

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 $\sum_{n\geq 1} s_n k^{-n} < 1$. Then there is a k-ary rational tree such that s is the enumerative sequence of its leaves.*

In order to prove this result, we first prove one lemma that remains true in the case of equality. We therefore 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 a normalized representation (G, i, F) of the $\mathbb{N}$-rational sequence s. We denote by M the adjacency matrix of G and by λ its spectral radius. Then $\lambda \leq k$. We then compute a k-approximate eigenvector $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ of the graph $\overline{G}$. By definition, we have $M\mathbf{v} \leq k\mathbf{v}$. Without loss of generality, we can assume that state 1 is the initial state in all normalized representations.

Lemma 1 *If k divides v_1, then there is another normalized representation for s and a new corresponding approximate eigenvector v' with $v'_1 = v_1$ div k.*

Proof : We denote by P the set of states q such that there is in G an edge denoted (q, x, t) going from q to a final state t of F. Remark that, as state t is equal to state 1 in $\overline{G}$, the value of state t is equal to the value of state 1.

Let us first suppose that the initial state 1 does not belong to P. If there is in P a state q which admits more than one (say n) outgoing edges, we split q in q' and q'' according to partition (O', O'') where $O' = \{(q, x, t)\}$ contains exactly one edge. Since k divides v_1, this state splitting is admissible and $v'_{q'} = v_1$ div k. Moreover, in the new graph G', q' admits only one outgoing edge (going to t) and q'' is either not in P or admits less than n outgoing edges. By successive state splittings of all states in P having more than one outgoing edges, we shall get, in a finite number of steps, a representation such that all states with one outgoing edge ending in F have no other outgoing edges. Under the hypothesis that state 1 does not belong to P, the initial state has not been split during this process and so each new computed graph is still a normalized representation of the sequence. We denote again by $(G, 1, F)$ the final representation obtained for s and by P_{last} the set of states having one outgoing edge ending in F in this graph. Remark that the values of states of P_{last} are greater than or equal to v_1 div k. We turn all values of states of P_{last} greater than v_1 div k into v_1 div k; the vector $\mathbf{v}$ remains a k-approximate eigenvector.

We then transform the representation $(G, 1, F)$ in a new one, (H, i, P_{last}), where H is the graph obtained from G by adding a state i, an edge from i to 1 and by removing all edges of G going out of a state of P_{last}. If we look at paths in G going from 1 to F, we have just cut the last edge and added one at the beginning. We assign to state i the value v_1 div k, and the values of all states correspond now to a new k-approximate eigenvector for $\overline{H}$. We call this tranformation the "shift" transformation.

Let us now suppose that the initial state 1 belongs to P. We first split, as explained above, all states of P having more that one outgoing edge. In this case, state 1 may have been split. We denote by $1_{(1)}, 1_{(2)}, 1_{(3)}, \ldots 1_{(r)}$ the copies of state 1 obtained by successive state splittings of the initial state 1. We still denote by G the graph obtained by this transformation and by P_{last} the set of states having one outgoing edge ending in F in this graph. We then transform the representation $(G, 1, F)$ into a new one, (H, i, P_{last}), where H is the graph obtained from G by adding a state i, an edge from i to each $1_{(j)}, 1 \leq j \leq r$ and by removing all edges of G going out of a state of P_{last}. Remark that $(r-1)$ states among $1_{(1)}, 1_{(2)}, 1_{(3)}, \ldots 1_{(r)}$ belong to P_{last}. We again assign to the state i the value v_1 div k, and the values of all states correspond now to a new k-approximate eigenvector for $\overline{H}$. □

Corollary 1 *If v_1 is a power of k, then there is another normalized representation and a new corresponding approximate eigenvector v' with $v'_1 = 1$.*

Proof : If $v_1 = k^m$, we iterate the construction given in previous lemma and get $v_1^{'} = 1$ in m steps. $\square$

Example
Let s be the following series:

$$s(z) = 2z^3 + 2z^2 \left(z^2(z^2)^*\right)^*$$

Here, $k = 2$ and $s(1/2) = 1$.
In the following pictures, the nodes are labeled with their value.

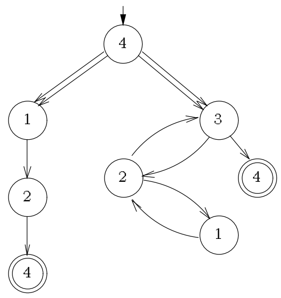

Fig. 6. Initial normalized representation

First step

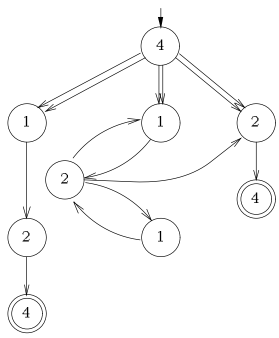

Fig. 7. First state splitting

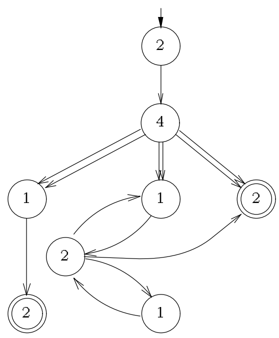

Fig. 8. First "shift"

Second step

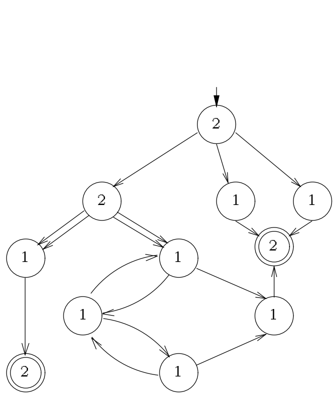

Fig. 9. Other state splittings

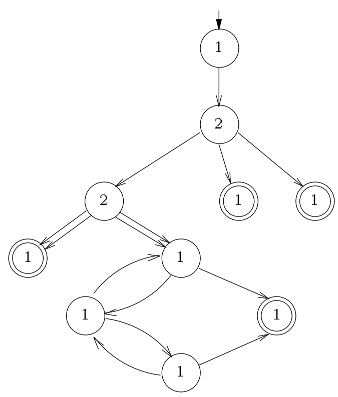

Fig. 10. Second "shift"

The *last step* is described in the proof of Theorem 1.
It corresponds here to a state splitting of all states of the graph of value different from 1.

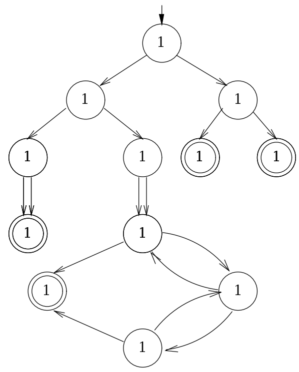

Fig. 11. Last representation

We now prove another lemma which is true only in the case of Kraft strict inequality.

Lemma 2 *Let M be a nonnegative integral matrix. If its spectral radius is strictly less than k, then there is a k-approximate eigenvector* $\mathbf{w}$ *of M such that w_1 is a power of k.*

Proof : Let λ ($\lambda < k$) be the positive real eigenvalue of maximal modulus of M and let $\mathbf{v}$ be an eigenvector associated to λ. We denote by P the set of positive vectors $\mathbf{w}$ such that $M\mathbf{w} < k\mathbf{w}$. The set P is an open set and $\mathbf{v}$ belongs to P. By dividing all components of $\mathbf{v}$ par v_1, we can assume that v_1 is equal to 1. As P is open, there is a positive real ϵ such that $B(v, \epsilon) \subset P$, where $B(v, \epsilon) = \{\mathbf{w} \mid v_i - \epsilon \leq w_i \leq v_i + \epsilon\}$. Let us now choose an integer m such that $1/k^m < \epsilon$. As $B(v, 1/k^m) \subset P$, we have $\{k^m\mathbf{w} \mid \mathbf{w} \in B(v, 1/k^m)\} \subset P$. This set is $\{\mathbf{w} \mid k^m v_i - 1 \leq w_i \leq k^m v_i + 1\}$ and contains $\mathbf{w}$ where $w_i = \lceil k^m v_i \rceil$. This vector is a positive integer vector $\mathbf{w}$ with $M\mathbf{w} < k\mathbf{w}$: it is a k-approximate eigenvector. Moreover $w_1 = k^m$.
□

We now prove Theorem 1.

Proof : (Theorem 1) We begin with a normalized representation of s and compute, by Lemma 2, a k-approximate eigenvector whose component for the initial state is a power of k. We then compute, by Corollary 1, a normalized representation $(G, 1, F)$ of s which admits a k-approximate eigenvector of component 1 for the initial state. Finally, we apply to G the state splitting algorithm described in the previous section to obtain a k-ary graph. As the component of the approximate eigenvector on the initial state is 1 and as the state splittings have to be admissible, this state will never be split during the process. A state splitting of a state of G different from state 1 leads by construction to a graph G' still representing the same sequence. The result follows then from the fact that the final normalized representation has a k-ary graph. □

We can apply the construction given above to the case of Kraft equality when it is possible to find a representation of s which admits a k-eigenvector with a power of k as component on the initial state. This may perhaps not always be the case. We can consider the example : $s(z) = (2z + 2z^2)(z^2)^* z$. It has a representation given by the graph G with an eigenvector such that $v_1 = 3$ and it admits also the representation given by graph H with an eigenvector such that $v_1 = 1$. The series is also equal to $2z^2 z^*$. There is no hope to succeed in transforming G into H by state splitting and merging since $\overline{H}$ and $\overline{G}$ are non isomorphic symbolic subshifts.

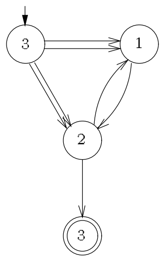

Fig. 12. Root of value 3 (G)

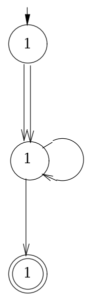

Fig. 13. Root of value 1 (H)

4 Computation of an approximate eigenvector

Although Lemma 2 tells us that, for an irreducible matrix with a spectral radius strictly less than an integer k, an approximate eigenvector with a power of k as first component, exists, it does not provide a good way to find one. Fortunately, there is an efficient algorithm to find such small approximate eigenvectors. The algorithm we give is a variant of Franaszek's algorithm to compute an upper approximate eigenvector (see [6] p.153). This makes the algorithm of Theorem 1 practical.

Let M be a nonnegative *irreducible* integral matrix and λ its spectral radius. If $\lambda \geq k$, a k-upper approximate eigenvector is a positive integral vector $\mathbf{v}$ such that $M\mathbf{v} \geq k\mathbf{v}$. If $\lambda \leq k$, a k-lower approximate eigenvector is a positive integral vector $\mathbf{v}$ such that $M\mathbf{v} \leq k\mathbf{v}$. In the case where $\lambda < k$, we want to compute a k-lower approximate eigenvector with a first component which is a power of k. In the sequel, we shall omit the word lower.

We use the following notation to state the algorithms. If $\mathbf{u}$ and $\mathbf{v}$ are vectors, let $\mathbf{w} = \max\{\mathbf{u}, \mathbf{v}\}$ (resp. $\mathbf{w} = \min\{\mathbf{u}, \mathbf{v}\}$) denote the componentwise maximum (resp. minimum), so that $w_i = \max\{u_i, v_i\}$ (resp.

$w_i = \min\{u_i, v_i\}$) for each index i. For a real number r, let $\lceil r \rceil$ denote the integer top of r, i.e., the smallest integer greater than or equal to r.

Proposition 2 *Let M be an integral irreducible matrix with a spectral radius less than or equal to an integer k. A smallest k-approximate eigenvector exists. If the spectral radius is strictly less than k, a smallest k-approximate eigenvector, whose first component is a power of k, exists.*

Proof : The proof of the existence of a k-(lower) approximate eigenvector is the same as the proof of the existence of a k-upper approximate eigenvector (for irreducible matrices with a spectral radius greater than or equal to k). Let now $\mathbf{u}$ and $\mathbf{u}'$ be two k-approximate eigenvectors. Then it is straightforward that $\mathbf{v} = \min\{\mathbf{u}, \mathbf{u}'\}$ is a k-approximate eigenvector.

When the spectral radius is strictly less than k, the proof of the existence of a k-approximate eigenvector, whose first component is a power of k, is given in Lemma 2. If $\mathbf{u}$ and $\mathbf{u}'$ are two k-approximate eigenvectors, whose first component is a power of k, then this is true also of $\min\{\mathbf{u}, \mathbf{u}'\}$. □

The lower version of Franaszek's algorithm that computes the smallest k-approximate eigenvector is the following:

FRANASZEK'S ALGORITHM:
begin
  $\mathbf{v}' := (1, 1, \ldots, 1)^t$;
  repeat
    begin
    $\mathbf{v} := \mathbf{v}'$;
    $\mathbf{v}' := \max\left\{\mathbf{v}, \lceil \frac{1}{k} M\mathbf{v} \rceil\right\}$;
    end
  until $\mathbf{v}' = \mathbf{v}$;
end

The proof that the algorithm is correct is the following:
Proof : The vectors computed are monotonically nondecreasing in each component. Let $\mathbf{u}$ be a k-approximate eigenvector. As all components of $\mathbf{u}$ are positive integers, we have $\mathbf{v}' \leq \mathbf{u}$. In fact, this is true for the first vector $\mathbf{v}' = (1, 1, \ldots, 1)^t$. If we assume that $\mathbf{v} \leq \mathbf{u}$ in the repeat loop, then either $\mathbf{v}' = \mathbf{v}$, or $\mathbf{v}' = \lceil \frac{1}{k} M\mathbf{v} \rceil$. In the first case, $\mathbf{v}' \leq \mathbf{u}$. In the second case, $\mathbf{v}' \leq \lceil \frac{1}{k} M\mathbf{u} \rceil \leq \lceil \mathbf{u} \rceil$. Since $\mathbf{u}$ has integer components, we also get $\mathbf{v}' \leq \mathbf{u}$. This proves that the process eventually stops.

The final vector $\mathbf{v}$ obtained is then a k-approximate eigenvector. In fact, as $\mathbf{v}' = \mathbf{v}$, $\lceil \frac{1}{k} M\mathbf{v} \rceil \leq \mathbf{v}$. Then $M\mathbf{v} \leq k\mathbf{v}$. As $\mathbf{v}$ is less than or equal to any other k-approximate eigenvector, it is the smallest. □

We now suppose that the spectral radius of M is *strictly less* than k. We are interested in computing a k-approximate eigenvector whose first component is a power of k. We give a variant of Franaszek's algorithm that finds the smallest k-approximate eigenvector whose first component is a power of k.

VARIANT OF FRANASZEK'S ALGORITHM:

begin

  $\mathbf{v}' := (1, 1, \ldots, 1)^t$;

  repeat

    begin

    $\mathbf{v} := \mathbf{v}'$;

    $v'_1 := \max\left\{v_1, k^{\lceil \log_k(\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil)\rceil}\right\}$;

    $v'_i := \max\left\{v_i, \lceil \frac{1}{k}(M\mathbf{v})_i \rceil\right\}$ for $i \geq 2$;

    end

  until $\mathbf{v}' = \mathbf{v}$;

end

The proof that the algorithm is correct is the following:

Proof : The vectors computed are monotonically nondecreasing in each component. Let $\mathbf{u}$ be now a k-approximate eigenvector whose first component is a power of k. We assume that $u_1 = k^s$, where s is a nonnegative integer. We still have $\mathbf{v}' \leq \mathbf{u}$. In fact, this is true at the beginning of the repeat loop since $\mathbf{u}$ is a positive integral vector. Let us suppose that $\mathbf{v} \leq \mathbf{u}$ in the loop, we have $M\mathbf{v} \leq M\mathbf{u} \leq k\mathbf{u}$. Then $\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil \leq u_1 = k^s$ and $\lceil \frac{1}{k}(M\mathbf{v})_i \rceil \leq u_i$ for $i \geq 2$, since $\mathbf{u}$ is integer. We get $v'_i \leq u_i$ for $i \geq 2$, and $\log_k(\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil) \leq s$. Finally, we have $\lceil \log_k(\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil)\rceil \leq s$ and $v'_1 \leq k^s = u_1$. This proves that $\mathbf{v}' \leq \mathbf{u}$ and that the process stops.

Let $\mathbf{v}$ be the final vector computed. We still have $M\mathbf{v} \leq k\mathbf{v}$. This is easily verified for all components greater than 1 as in the previous proof. For the first component, we successively get, as $\mathbf{v} = \mathbf{v}'$:

$$\begin{aligned} k^{\lceil \log_k(\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil)\rceil} &\leq v_1, \\ k^{\log_k(\lceil \frac{1}{k}(M\mathbf{v})_1 \rceil)} &\leq v_1, \end{aligned}$$

$$\begin{aligned} \lceil \frac{1}{k}(M\mathbf{v})_1 \rceil &\leq v_1, \\ \frac{1}{k}(M\mathbf{v})_1 &\leq v_1, \\ (M\mathbf{v})_1 &\leq v_1. \end{aligned}$$

The last vector computed is also smaller than any other k-approximate eigenvector whose first component is a power of k. This concludes the proof. □

5 Link with enumerative sequences of nodes

In this section, we are going to extend Theorem 1 to the case of Kraft equality when, moreover, the enumerative sequence of nodes corresponding to the sequence of leaves has a primitive linear representation.

Let s be an $\mathbb{N}$-rational series. A *linear representation* of s 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:

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

The triple $(\mathbf{l}, M, \mathbf{c})$ corresponds to a representation (G, I, F) of s defined in such a way that M is the adjacency matrix of G. The linear representation is said to be *primitive* if M is primitive. Recall that a matrix is primitive if there exists an integer n such that $M^n > 0$. Equivalently, the adjacency matrix of a strongly connected graph G is primitive if the *g.c.d* of lengths of cycles in G is 1.

Let T be a tree. We define the enumerative sequence t of internal nodes by height of the tree T by $t = (t_n)_{n \geq 0}$, where t_n is the number of internal nodes of T at height n. An internal node is a node which is not a leaf. Let s (resp. t) be the enumerative sequence of leaves (resp. of nodes) by height of a complete k-ary tree. The link between s and t is the following:

$$t(z) = \frac{1 - s(z)}{1 - kz} \quad \text{with} \quad s(1/k) = 1.$$

We also have $s(z) = P(z)/Q(z)$, where $P(z)$ and $Q(z)$ are polynomials with nonnegative integral coefficients and $Q(0) = 1$. Therefore the series t associated to s satifies:

$$t(z) = \frac{Q(z) - P(z)}{(1 - kz)Q(z)}.$$

As $s(1/k) = 1$, $P(1/k) = Q(1/k)$. Thus the series s and t have the same poles and especially the same convergence radius $1/\lambda > 1/k$.

We can now state the following result that we are going to prove:

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.

The construction that we give is partly based on a proof by Perrin ([9]).

Let $\mathbf{l}, M, \mathbf{c}$ be matrices with nonnegative entries such that $(\mathbf{l}, M, \mathbf{c})$ is a linear representation of t, *i.e.*

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

We denote by $1/\lambda$ the convergence radius of t. As the matrix M has $\lambda < k$ as spectral radius and is primitive, the sequence $((M/\lambda)^n)_{n\geq 1}$ tends toward a positive matrix N such that $(M/\lambda)N = N$. Thus, the sequence $((M/\lambda)^n \mathbf{c})_{n\geq 1}$ tends toward a positive vector $\mathbf{w} = N\mathbf{c}$. The vector $\mathbf{w}$ is an eigenvector associated to 1 for M/λ since $(M/\lambda)N\mathbf{c} = N\mathbf{c}$. Since $M\mathbf{w} < k\mathbf{w}$, for all large enough n, we get

$$M\left(\frac{M}{\lambda}\right)^n \mathbf{c} < k\left(\frac{M}{\lambda}\right)^n \mathbf{c},$$

and so, we have the inequality

$$\forall n \geq n_0, \qquad M^{n+1}\mathbf{c} < kM^n\mathbf{c}.$$

We shall give a construction of paths in the tree beginning at the nodes of height n_0, the construction of the remaining part of the tree being obvious. In order to do that, we transform the linear representation of $(t_n)_{n\geq 0}$ into the one of $(t_n)_{n\geq n_0}$. The latter is $(\mathbf{l}, M, \mathbf{v})$ with $\mathbf{v} = M^{n_0}\mathbf{c}$ since

$$\forall n \geq 0, \qquad t_{n+n_0} = \mathbf{l}M^n(M^{n_0}\mathbf{c}).$$

The vector $\mathbf{v}$ is then a k-(lower) approximate eigenvector of the matrix M, since we have $M\mathbf{v} \leq k\mathbf{v}$.

Now we shall prove that we can replace the triple $(\mathbf{l}, M, \mathbf{v})$ by a triple $(\mathbf{l}', M', \mathbf{v}')$, which is also a linear representation of $(t_n)_{n \geq n_0}$, such that $\mathbf{v}'$ has all its entries equal to 1, and such that the entries of each row of M' are at most k. The construction of this new linear representation basically makes use of the state splitting.

We shall use the following *extended notion of state splitting*, that we call again a state splitting. Let $G = (V, E)$ be a graph and let q be a vertex of G. We denote by I (resp. O) be the set of edges coming in q (resp. going out of q). Let $(O_1, O_2, \cdots, O_r)$ be a partition of O in r parts. Main difference with the definition of state splitting in Section 2 is that some parts O_i of the partition may be here empty. The operation of *(output) state splitting* relative to this partition transforms G into the graph $G' = (V', E')$ where $V' = (V \setminus \{q\}) \cup q_1 \cup q_2 \ldots \cup q_r$ is obtained from V by splitting state q into r states $q_1, \ldots, q_r$, and where E' is defined as follows:

1. All edges of E that are not incident to q are left unchanged.
2. All states q_i have the same input edges as q.
3. The output edges of q are distributed between $q_1, \ldots, q_r$ according to the partition of O into $(O_1, O_2, \ldots, O_r)$. We denote by U_i the set of output edges of q_i: $U_i = \{(q_i, x, p) \mid (q, x, p) \in O_i\}$. Note that some U_i may be empty.

Let us now assume that a k-approximate eigenvector $\mathbf{v}$ for the graph G, that is a k-approximate eigenvector of the adjacency matrix of G, exists. We denote by v_p the component of index p of $\mathbf{v}$, if p is a state of G. It is also called the *value* of state p (before the state splitting). All components v_p are positive integers. An *admissible* vector $\mathbf{v}'$ for the new graph G' is a positive integral vector satisfying the following conditions:

- If p is a state distinct from q_i then $v'_p = v_p$.
- $\sum_{1 \leq i \leq r} v'_{q_i} = v_q$.
- The vector $\mathbf{v}'$ is a k-approximate eigenvector of the adjacency matrix M' of the new graph G'.

Note the difference with the definition of an admissible state splitting given in Section 2.

Finally, let $\mathbf{l}$ be the row vector of a linear representation $(\mathbf{l}, M, \mathbf{v})$. We transform $\mathbf{l}$ into $\mathbf{l}'$ defined in the following way:

$$l'_{q_1} = l'_{q_2} = \cdots = l'_{q_r} = l_q,$$

$$\text{if} \quad p \neq q_i \qquad l'_p = l_p.$$

If $(\mathbf{l}, M, \mathbf{v})$ is a linear representation of a given series, then $(\mathbf{l}', M', \mathbf{v}')$ is a representation of the same series.

Proposition 3 *Let M be a nonnegative integral matrix whose spectral radius is less than a positive integer k. Let $\mathbf{v}$ be a k-(lower) approximate eigenvector for M. If $(\mathbf{l}, M, \mathbf{v})$ is the linear representation of an $\mathbb{N}$-rational series, there is another representation $(\mathbf{l}', M', \mathbf{v}')$, with matrices having entries in $\mathbb{N}$, of the same series, such that the sum of the entries of each row of M' is at most k, and all components of $\mathbf{v}'$ are equal to 1.*

Note that the matrix M is not necessarily irreducible. We define the graph G associated to $M \in \mathbb{N}^{n\times n}$ as the n-state graph having M_{ij} edges from i to j.
Proof : (Proposition 3) We may suppose that the matrix M is a $0-1$ matrix. Otherwise, we obtain such a matrix by successive output state splittings of all states having more than one output edges going to a same state.

The method consists in successive output state splittings of the graph G associated to M, in such a way that a new graph G' and a new admissible k-approximate eigenvector $\mathbf{v}'$ are obtained at each step. The inequality $M\mathbf{v} \leq k\mathbf{v}$ is then an invariant of the iteration of state splittings.

More precisely, we choose a state q of the graph G such that v_q is maximal and has not exactly k outgoing edges all ending in states of maximal value. Such a state of maximal value exists, since otherwise G would admit a subgraph whose states all have k outgoing edges, and the spectral radius of the matrix M would be then greater than or equal to k. We denote by O the set of outgoing edges of q, and we denote by $m = v_q$ the maximal value of the components of $\mathbf{v}$. We can assume that $m > 1$ since otherwise the process immediately stops. We denote by J the ending states of edges of O. We consider several cases to define the state splitting:

1$^{\text{st}}$ case We suppose that there is a non empty and strict subset J' of J such that $\sum_{p\in J'} v_p \equiv 0 \bmod k$. This is true especially when $\text{Card}(O) > k$ (see Section 3).

In this case, we do an output state splitting of q relative to the partition (O_1, O_2), where O_1 are the edges of O ending in J' and $O_2 = O \setminus O_1$. The graph G is transformed into a graph G' having M' as adjacency matrix. We define the vector $\mathbf{v}'$ by

$$v'_{q_1} = \frac{1}{k} \sum_{p \in J'} v_p, \qquad v'_{q_2} = v_q - v'_{q_1},$$

$$\text{and if} \quad p \neq q_1, q_2 \qquad v'_p = v_p.$$

The vector $\mathbf{v}'$ is positive integral vector that satisfies the inequality:

$$M'\mathbf{v}' \leq k\mathbf{v}'.$$

Indeed, setting $J'' = J \setminus J'$, we get

$$\text{if} \quad p \neq q_1, q_2, \qquad (M'\mathbf{v}')_p = (M\mathbf{v})_p.$$

$$\begin{aligned} (M'\mathbf{v}')_{q_1} &= \begin{cases} \sum_{p \in J'} v'_p & \text{if} \quad q \notin J' \\ \sum_{p \in J' \setminus \{q\}} v'_p + v'_{q_1} + v'_{q_2} & \text{if} \quad q \in J' \end{cases} \\ &= \sum_{p \in J'} v_p = kv'_{q_1} \end{aligned}$$

$$\begin{aligned} (M'\mathbf{v}')_{q_2} &= \begin{cases} \sum_{p \in J''} v'_p & \text{if} \quad q \notin J'' \\ \sum_{p \in J'' \setminus \{q\}} v'_p + v'_{q_1} + v'_{q_2} & \text{if} \quad q \in J'' \end{cases} \\ &= \sum_{p \in J''} v_p = \sum_{p \in J} v_p - \sum_{p \in J'} v_p \leq kv_q - kv'_{q_1} \\ &\leq kv'_{q_2}. \end{aligned}$$

Therefore $\mathbf{v}'$ is an admissible k-approximate eigenvector.

2$^{\mathbf{nd}}$ **case** We suppose that $\mathrm{Card}(O) = k$. Then we can also suppose that $\sum_{p \in J} v_p \equiv 0 \bmod k$. Otherwise we would be in the first case.

Since, by hypothesis, q has at least one outgoing edge ending in a state of value strictly less than the maximal value $m = v_q$, we have $\sum_{p \in J} v_p < km$. As the left member of the inequality is a multiple of k, we obtain

$$\sum_{p \in J} v_p \leq k(m-1). \tag{1}$$

We split state q according to the partition $(O_1 = O, O_2 = \emptyset)$ and define a k-approximate eigenvector $\mathbf{v}'$ by

$$v'_{q_1} = m - 1, \quad v'_{q_2} = 1$$

We still have $M'\mathbf{v}' \leq k\mathbf{v}'$ and $\mathbf{v}'$ is admissible.

We mention here another possibility of splitting according to a partition of O into two non empty subsets. If J' denotes the ending states of edges of O_1, the admissible vector $\mathbf{v}'$ is in this case defined by $v'_{q_1} = \lceil \frac{1}{k} \sum_{p \in J'} v_p \rceil$.

The remaining cases are devoted to the case where $\mathrm{Card}(O) = r < k$.

3$^{\text{rd}}$ case We suppose that $\mathrm{Card}(O) = r < k$ and $r \leq m \leq k$.

We split q in $q_1, q_2, \ldots, q_m$ according to $(O_1, O_2, \ldots, O_m)$, where O_i contains exactly one edge from O, for $1 \leq i \leq r$, and $O_i = \emptyset$ for $r+1 \leq i \leq m$. We define a k-approximate eigenvector $\mathbf{v}'$ by

$$v'_{q_i} = 1, \quad 1 \leq i \leq m$$

As $v'_p \leq m \leq k$ for any state p, $M'\mathbf{v}' \leq k\mathbf{v}'$ and $\mathbf{v}'$ is admissible.

4$^{\text{th}}$ case We suppose that $m < \mathrm{Card}(O) = r < k$.

We number $1, 2 \ldots, r$ the states of J. Let i be the greatest index such that

$$v_1 + v_2 + \cdots + v_i < k - m,$$

if it does not exist, we set $i = 0$.
Let us first assume that $0 \leq i < r - 1$.
Then

$$v_1 + v_2 + \cdots + v_{i+1} \geq rm - k(m-1).$$

In fact, otherwise we would have:

$$v_1 + v_2 + \cdots + v_{i+1} < rm - k(m-1) \leq (k-1)m - k(m-1) = k - m.$$

We then also get:

$$v_1 + v_2 + \cdots + v_i + v_{i+1} < k - m + m = k$$

and

$$\begin{aligned} v_{i+2} + \cdots + v_r &= (v_1 + v_2 \cdots + v_r) - (v_1 + v_2 + \cdots + v_{i+1}) \\ &\leq rm - (rm - k(m-1)) = k(m-1). \end{aligned}$$

As a consequence we define the splitting of q according to the partition $(O_1, O_2 = O \setminus O_1)$, where O_1 is the set of outgoing edges of q ending in states 1 to $i+1$. The vector $\mathbf{v}'$ defined by

$$v'_{q_1} = 1 \quad \text{and} \quad v'_{q_2} = m - 1,$$

is an admissible k-approximate eigenvector of M'.
Let us now assume that $r-1 \leq i \leq r$.
We define the splitting of q according to the partition $(O_1, \ldots, O_m)$, where $O_1 = O$ and $O_i = \emptyset$ for $2 \leq i \leq m$, and $\mathbf{v}'$ by

$$v'_{q_i} = 1 \quad \text{for} \quad 1 \leq i \leq m.$$

The vector $\mathbf{v}'$ is still admissible.

5^{th} case We suppose that $\text{Card}(O) = r < k < m$.

If $r > 1$, we define the splitting of q according to the partition $(O_1, O_2 = O \setminus O_1)$, where O_1 contains exactly one edge of O. We number $1, 2, \ldots, r$ the states of J, and we can suppose that state 1 is the ending state of the edge of O_1. The vector $\mathbf{v}'$ is defined by

$$v'_{q_1} = \lceil \frac{v_1}{k} \rceil \quad \text{and} \quad v'_{q_2} = v_q - v'_{q_1}.$$

We have:

$$v_1 + v_2 + \cdots + v_r \leq rm \leq (k-1)m,$$

and as $v_1 > k(\lceil \frac{v_1}{k} \rceil - 1)$,

$$\begin{aligned} v_2 + \cdots + v_r &< (k-1)m - k(\lceil \frac{v_1}{k} \rceil - 1), \\ &\leq k(m - \lceil \frac{v_1}{k} \rceil) + k - m \\ &\leq k(m - \lceil \frac{v_1}{k} \rceil). \end{aligned}$$

This proves that $\mathbf{v}'$ is again an admissible vector.

If $r \leq 1$, we then define the splitting of q according to the partition $(O_1, \ldots, O_{m - \lceil \frac{v_1}{k} \rceil})$, where O_1 contains the only edge of O and $O_i = \emptyset$ for $2 \leq i \leq (m - \lceil \frac{v_1}{k} \rceil)$. The vector $\mathbf{v}'$ defined by

$$v'_{q_1} = \lceil \frac{v_1}{k} \rceil \quad \text{and} \quad v'_i = 1 \quad \text{for} \quad 2 \leq i \leq (m - \lceil \frac{v_1}{k} \rceil)$$

is again an admissible vector.

In all cases, the state q is split in states with new values stricly less than the maximal value m. Moreover each state without output edge obtained by state splitting has value 1. This transformation is iterated until all components of the vector $\mathbf{v}'$ are equal to 1, concluding the proof of Proposition 3. □

We then obtain, by Proposition 3, a new representation $(\mathbf{l}', M', \mathbf{v}')$ of the sequence $(t_n)_{n \geq n_0}$, where the sum of each row of M' is at most k (since $M'\mathbf{v}' \leq k\mathbf{v}'$), and all components of $\mathbf{v}'$ are equal to 1. The sum of the components of $\mathbf{l}'$ is then t_{n_0}.

The interpretation of this new representation shows that t is the generating series of internal nodes of a k-ary rational tree, concluding the proof.

Example Let t be the following series: $t(z) = (z + z^3)^*(1 + z)$. We get that $M^2\mathbf{c} \leq kM\mathbf{c}$, so $M\mathbf{c} = (2, 1, 2)^t$ is a k approximate eigenvector and after splittings we obtain the tree of the last picture.

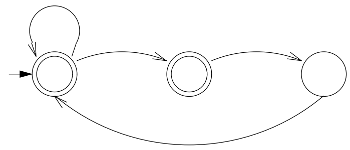

Fig. 14. The sequence t

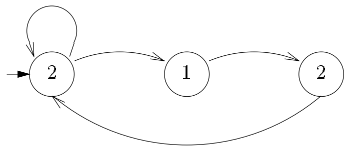

Fig. 15. The k-approximate eigenvector $M\mathbf{v}$.

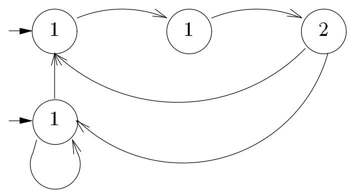

Fig. 16. First state splitting (1^{st} case)

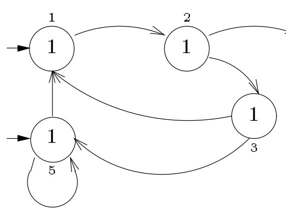

Fig. 17. Another state splitting (2^{nd} case)

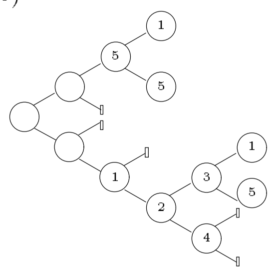

Fig. 18. The associated tree (nodes are labelled by state numbers)

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

[4] 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.

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

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

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

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

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

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

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

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