A FINITE STATE VERSION OF THE KRAFT-MCMILLAN THEOREM

FRÉDÉRIQUE BASSINO *, MARIE-PIERRE BÉAL * , AND DOMINIQUE PERRIN *

Abstract. The main result is a finite-state version of the Kraft-McMillan theorem characterizing the generating sequence of a k-ary regular tree. The proof uses a new construction called the *multiset construction* which is a version with multiplicities of the well-known subset construction of automata theory.

Key words. generating series, regular trees, nonnegative matrices.

AMS subject classifications. 68Q45, 68R10, 94A45, 37B10

1. Introduction.

The Kraft inequality $\sum_{n\geq 0} s_n k^{-n} \leq 1$ characterizes the generating sequences $(s_n)_{n\geq 0}$ of leaves in a k-ary tree. It is used in connexion with Huffman algorithm to build prefix codes or search trees and usually restricted to the case of finite trees. We are interested here in the case of infinite sequences corresponding to infinite trees. These infinite trees arise for example as search trees in infinite sets. They also appear in the context of finite automata having nested loops to represent the set of first returns to a given state. The tree thus obtained is called a *regular* tree. It has only a finite number of non-isomorphic subtrees since two subtrees corresponding to the same state of the automaton are isomorphic. The generating sequences of such infinite trees are of interest in the applications of finite automata to text compression or channel coding.

Our main result is a characterization of the generating sequences of leaves of regular k-ary trees. Its essence is that the two conditions of being the generating sequence of

(i) a k-ary tree
(ii) a regular tree

are independent in the sense that their conjunction is enough to guarantee that a sequence is the generating sequence of a regular k-ary tree.

The proof uses a new construction on graphs called the *multiset construction* which is a counterpart for automata with multiplicities of the well-known subset construction of automata theory.

Our results have a connexion with symbolic dynamics. Actually, in both cases, the emphasis is on the space of paths in a finite graph. Even if we do not use results from symbolic dynamics, some of the methods used, like state-splitting or the Perron theory are similar. Using an expression of Lind and Marcus [15], our treatment is "dynamical in spirit". The relationship with symbolic dynamics is discussed more closely in [7] and [8].

The paper is organized as follows. Section 2 contains preliminary results and definitions on graphs, trees, regular sequences and the Perron-Frobenius theory. In Section 3, we present the multiset construction. Section 4 contains the proof of our main result (Theorem 4.2). The following section (Section 5) treats a similar problem, with the set of leaves replaced by the set of all nodes.

The results contained in this paper represent the terminal point of a series of steps. In a previous paper [7] (with a preliminary version in [5]), we proved Theorem

*Institut Gaspard Monge, Université de Marne-la-Vallée,77454 Marne-la-Vallée Cedex 2 France. {bassino,beal,perrin}@univ-mlv.fr.

4.2 in the particular case of a strict inequality. The proof uses the technique of state-splitting from symbolic dynamics. In the same paper, we also give a proof of Theorem 5.3 which is different from the proof given here, which is based on the multiset construction and is more simple. Part of the results of the present paper was presented at the conference LATIN'98 [6]. Finally, the survey paper [8] gives an overview of length distributions and regular sequences.

2. Definitions and background. In this section, we fix our notation concerning graphs, trees and regular sequences. We also recall some notions concerning positive matrices.

A word on the terminology used here. We constantly use the term *regular* where a richer terminology is often used. In particular, what we call here a regular sequence is, in Eilenberg's terminology, an $\mathbb{N}$-rational sequence (see [11], [19] or [10]).

2.1. Graphs and trees. In this paper, we use directed multigraphs i.e. graphs with possibly several edges with the same origin and the same end. We simply call them graphs in all what follows. We denote $G = (Q, E)$ a graph with Q as set of vertices and E as set of edges. We also say that G is a graph on the set Q.

A *tree* T on a set of nodes N with a *root* $r \in N$ 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 children. A node without children is called a *leaf*. A node which is not a leaf is called *internal*. A node n is a *descendant* of a node m if $m = T^h(n)$ for some $h \geq 0$. A k-ary tree is *complete* if all internal nodes have exactly k children and have at least one descendant which is a leaf.

For each node n of a tree T, the *subtree* rooted at n, denoted T_n is the tree obtained by restricting the set of nodes to the descendants of n.

Two trees S, T are isomorphic, denoted $S \equiv T$, if there is a map which transforms S into T by permuting the children of each node. Equivalently, $S \equiv T$ if there is a bijective map $f : N \to M$ from the set of nodes of S onto the set of nodes of T such that $f \circ S = T \circ f$. Such a map f is called an isomorphism.

If T is a tree with N as set of nodes, the *quotient graph* of T is the graph $G = (Q, E)$ where Q and E are defined as follows. The set Q is the quotient of N by the equivalence $n \equiv m$ if $T_n \equiv T_m$. Let $\bar{m}$ denote the class of a node m. The number of edges from $\bar{m}$ to $\bar{n}$ is the number of children of m equivalent to n.

Conversely, the set of paths in a graph with given origin is a tree. Indeed, let $G = (Q, E)$ be a graph. Let $r \in Q$ be a particular vertex and let N be the set of paths in G starting at r. The tree T having N as set of nodes and such that $T(p_0, p_1, \ldots, p_n) = (p_0, p_1, \ldots, p_{n-1})$ is called the *covering tree* of G starting at r.

Both constructions are mutually inverse in the sense that any tree T is isomorphic to the covering tree of its quotient graph starting at the image of the root.

PROPOSITION 2.1. *Let T be a tree with root r. Let G be its quotient graph and let i be the vertex of G which is the class of the root of T. For each vertex q of G and for each $n \geq 0$, the number of paths of length n from i to q is equal to the number of nodes of T at height n in the class of q.*

A tree is said to be *regular* if it admits only a finite number of non-isomorphic subtrees, i.e. if its quotient graph is finite.

For example, the infinite tree represented on Figure 2.1 is a regular tree. Its quotient graph is represented on Figure 2.2.

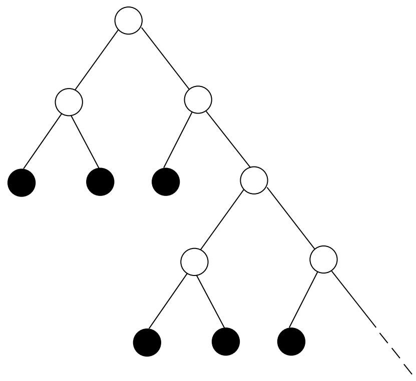

FIG. 2.1. *A regular tree.*

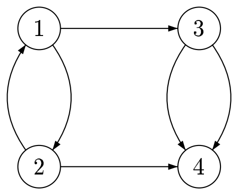

FIG. 2.2. *And its quotient graph.*

There is also a close connexion between trees and sets of words on an alphabet. Let X be a set of words on the alphabet $\{0, 1, \ldots, k-1\}$. The set X is said to be *prefix-closed* if any prefix of an element of X is also in X. When X is prefix closed, we can build a tree $T(X)$ as follows. The set of nodes is X, the root is the empty word ϵ and $T(a_1 a_2 \cdots a_n) = a_1 a_2 \cdots a_{n-1}$.

Let for example $X = \{\epsilon, 0, 1, 10, 11\}$. The tree $T(X)$ is represented on Figure 2.3.

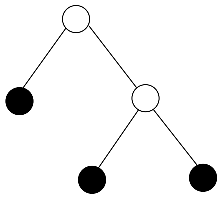

FIG. 2.3. *The tree $T(X)$.*

2.2. Regular sequences. We consider sequences of natural integers $s = (s_n)_{n \geq 0}$. We shall not distinguish between such a sequence and the formal series $s(z) = \sum_{n \geq 0} s_n z^n$.

We usually denote a vector indexed by elements of a set Q, also called a Q-vector, with boldface symbols. For $\mathbf{v} = (v_q)_{q \in Q}$ we say that $\mathbf{v}$ is nonnegative, denoted $\mathbf{v} \geq 0$, (resp. positive, denoted $\mathbf{v} > 0$) if $v_q \geq 0$ (resp. $v_q > 0$) for all $q \in Q$. The same conventions are used for matrices. A nonnegative $Q \times Q$-matrix M is said to be *irreducible* if, for all indices p, q, there is an integer m such that $(M^m)_{p,q} > 0$. The matrix is *primitive* if there is an integer m such that $M^m > 0$.

The *adjacency matrix* of a graph $G = (Q, E)$ is the $Q \times Q$-matrix M such that for each $p, q \in Q$, the integer $M_{p,q}$ is the number of edges from p to q. The adjacency matrix of a graph G is irreducible iff the graph is strongly connected. It is primitive if, moreover, the *g.c.d* of lengths of cycles in G is 1.

Let G be a finite graph and let I, T be two sets of vertices. For each $n \geq 0$, let s_n be the number of distinct paths of length n from a vertex of I to a vertex of T. The sequence $s = (s_n)_{n\geq 0}$ is called the sequence *recognized* by (G, I, T) or also by G if I and T are already specified. When $I = \{i\}$ and $T = \{t\}$, we simply denote (G, i, t) instead of $(G, \{i\}, \{t\})$.

A sequence $s = (s_n)_{n\geq 0}$ of nonnegative integers is said to be *regular* if it is recognized by such a triple (G, I, T), where G is finite. We say that the triple (G, I, T) is a *representation* of the sequence s. The vertices of I are called *initial* and those of T *terminal*. Two representations are said to be *equivalent* if they recognize the same sequence.

A representation (G, I, T) is said to be *trim* if every vertex of G is on some path from I to T. It is clear that any representation is equivalent to a trim one.

A well known result in theory of finite automata allows one to use a particular representation of any regular sequence s such that $s_0 = 0$. One can always choose in this case a representation (G, i, t) of s with a unique initial vertex i, a unique final vertex $t \neq i$ such that no edge is entering vertex i and no edge is going out of vertex t. Such a representation is called a *normalized representation* (see for example [17] page 14).

Let (G, i, t) be a trim normalized representation. If we merge the initial vertex i and the final vertex t in a single vertex still denoted by i, we obtain a new graph denoted by $\overline{G}$, which is strongly connected. The triple $(\overline{G}, i, i)$ is called the *closure* of (G, i, t).

Let s be a regular sequence such that $s_0 = 0$. The *star* s^* of the sequence s is defined by

$$s^*(z) = \frac{1}{1 - s(z)}.$$

PROPOSITION 2.2. *If (G,i,t) is a normalized representation of s, its closure $(\overline{G}, i, i)$ recognizes the sequence s^*.*

Proof. The sequence s is the length distribution of the paths of first returns to vertex i in $\overline{G}$, that is of finite paths going from i to i without going through vertex i. The length distribution of the set of all returns to i is thus $1 + s(z) + s^2(z) + \ldots = 1/(1 - s(z))$. □

An equivalent definition of regular sequences uses vectors instead of sets I, F. Let $\mathbf{i}$ be a Q-row vector of nonnegative integers and let $\mathbf{t}$ be a Q-column vector of nonnegative integers. We say that $(G, \mathbf{i}, \mathbf{t})$ recognizes the sequence $s = (s_n)_{n\geq 0}$ if for each integer $n \geq 0$

$$s_n = \mathbf{i} M^n \mathbf{t},$$

where M is the adjacency matrix of G. The proof that both definitions are equivalent follows from the fact that the family of regular sequences is closed under addition (see [11]). A triple $(G, \mathbf{i}, \mathbf{t})$ recognizing a sequence s is also called a representation of s and two representations are called equivalent if they recognize the same sequence.

A sequence $s = (s_n)_{n \geq 0}$ of nonnegative integers is *rational* if it satisfies a recurrence relation with integral coefficients. Equivalently, s is rational if there exist two polynomials $p(z), q(z)$ with integral coefficients and with $q(0) = 1$ such that

$$s(z) = \frac{p(z)}{q(z)}.$$

Any regular sequence is rational. The converse is however not true (see Section 5). For example, the sequence s defined by $s(z) = \frac{z}{1-z-z^2}$ is the sequence of Fibonacci

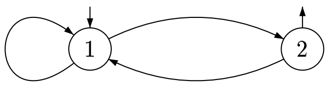

FIG. 2.4. *The Fibonacci graph.*

numbers also defined by $s_0 = 0, s_1 = 1$ and $s_{n+1} = s_n + s_{n-1}$. It is recognized by the graph of Figure 2.4 with $I = \{1\}$ and $T = \{2\}$.

2.3. Regular sequences and trees. If T is a tree, its *generating sequence of leaves* is the sequence of numbers $s = (s_n)_{n \geq 0}$, where s_n is the number of leaves at height n. We also simply say that s is the *generating sequence* of T.

The following result is a direct consequence of the definitions.

THEOREM 2.3. *The generating sequence of a regular tree is a regular sequence.*

Proof. Let T be a regular tree and let G be its quotient graph. Since T is regular, G is finite. The leaves of T form an equivalence class t. By Proposition 2.1, the generating sequence of T is recognized by (G, i, t) where i is the class of the root of T. □

We say that a sequence $s = (s_n)_{n \geq 1}$ satisfies the Kraft inequality for the integer k if

$$\sum_{n \geq 0} s_n k^{-n} \leq 1,$$

i.e. using the formal series $s(z) = \sum_{n \geq 0} s_n z^n$, if

$$s(1/k) \leq 1.$$

We say that s satisfies the strict Kraft inequality for k if $s(1/k) < 1$. The following result is well-known (see [3] page 35 for example).

THEOREM 2.4. *A sequence s is the generating sequence of a k-ary tree iff it satisfies the Kraft inequality for the integer k.*

Proof. Let first T be a k-ary tree and let s be its generating sequence. It is enough to prove that, for each $n \geq 0$, the sequence $(s_0, \ldots, s_n)$ satisfies the Kraft inequality. It is the generating sequence of the finite tree obtained by restricting T to the nodes at height at most n. We may thus suppose T to be a finite tree. We have

$$s(z) = zt_1(z) + \ldots + zt_k(z)$$

where $t_1, \ldots, t_k$ are the generating sequences of leaves of the (possibly empty) subtrees rooted at the children of the root of T. By induction on the number of nodes, we have $t_i(1/k) \leq 1$ whence the desired result.

Conversely, we use an induction on n to prove that there exists a k-ary tree with generating sequence $(s_0, \ldots, s_n)$. For $n = 0$, we have $s_0 \leq 1$ and T is either empty or reduced to one node. Suppose by induction hypothesis to have already built a tree T with generating sequence $(s_0, s_1, \ldots, s_{n-1})$. We have

$$\sum_{i=0}^{n} s_i k^{-i} \leq 1,$$

then

$$\sum_{i=0}^{n} s_i k^{n-i} \leq k^n,$$

and thus

$$s_n \leq k^n - \sum_{i=0}^{n-1} s_i k^{n-i}.$$

This allows us to add s_n leaves at height n to the tree T. $\square$

Let us consider the Kraft's equality case. If $s(1/k) = 1$, then any tree T having s as generating sequence is complete. The converse property is not true in general (see [11] p. 231). However, it is a classical result that when T is a complete regular tree, its generating sequence satisfies $s(1/k) = 1$ (see Proposition 2.8).

For the sake of a complete description of the construction described above in the proof of Theorem 2.4, we have to specify the choice made at each step among the leaves at height n. A possible policy is to choose to give as many children as possible to the nodes which are not leaves and of maximal height.

If we start with a finite sequence s satisfying Kraft's inequality, the above method builds a finite tree with generating sequence equal to s. It is not true that this incremental method gives a regular tree when we start with a regular sequence, as shown in the following example.

Let $s(z) = z^2/(1-2z^2)$. Since $s(1/2) = 1/2$, we may apply the Kraft construction to build a binary tree with length distribution s. The result is the tree $T(X)$ where X is the set of prefixes of the set

$$Y = \bigcup_{n \geq 0} 01^n 0\{0,1\}^n.$$

which is not regular.

If s is a regular sequence such that $s_0 = 0$, there exists a regular tree T having s as generating sequence. Indeed, let (G, i, t) be a normalized representation of s. The generating sequence of the covering tree of G starting at i is s. If s satisfies moreover the Kraft inequality for an integer k, it is however not true that the regular covering tree obtained is k-ary, as shown in the following example.

Let s be the regular sequence recognized by the graph of Figure 2.5 on the left with $i = 1$ and $t = 4$. We have $s(z) = 3z^2/(1-z^2)$. Furthermore $s(1/2) = 1$ and thus s satisfies Kraft's equality for $k = 2$. However there are four edges going out of vertex 2 and its regular covering tree starting at 1 is 4-ary. A solution for this example is given by the graph of Figure 2.5 on the right. It recognizes s and its covering tree starting at 1 is the regular binary tree of Figure 2.1.

The aim of Section 4 is to build from a regular sequence s that satisfies the Kraft inequality for an integer k a tree with generating sequence s which is both regular and k-ary.

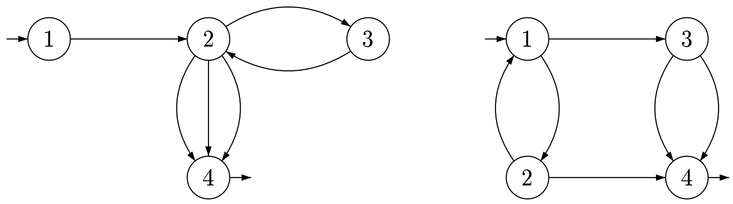

FIG. 2.5. *Graphs recognizing* $s(z) = 3z^2/(1-z^2)$.

2.4. Approximate eigenvector. Let M be the adjacency matrix of a graph G. By the Perron-Frobenius theorem (see [12], for a general presentation and [15], [14] or [9] for the link with graphs and regular sequences), the nonnegative matrix M has a nonnegative real eigenvalue of maximal modulus denoted by λ, also called the spectral radius of the matrix.

When G is strongly connected, the matrix is irreducible and the Perron-Frobenius theorem asserts that the dimension of the eigenspace of the matrix M corresponding to λ is equal to one, and that there is a positive eigenvector associated to λ.

Let k be an integer. A *k-approximate* eigenvector of a nonnegative matrix M is, by definition, an integral vector $\mathbf{v} \geq 0$ such that

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

One has the following result (see [15] p. 152).

PROPOSITION 2.5. *An irreducible nonnegative matrix M with spectral radius λ admits a positive k-approximate eigenvector iff $k \geq \lambda$.*

For a proof, see [15] p. 152. When M is the adjacency matrix of a graph G, we also say that $\mathbf{v}$ is a k-approximate eigenvector of G. The computation of an approximate eigenvector can be obtained by the use of Franaszek's algorithm (see for example [15]). It can be shown that there exists a k-approximate eigenvector with elements bounded above by k^{2n} where n is the dimension of M [4]. Thus the size of the coefficients of a k-approximate eigenvector is bounded above by an exponential in n and can be in the worst case of this order of magnitude.

The following result is well-known. It links the radius of convergence of a sequence with the spectral radius of the associated matrix.

PROPOSITION 2.6. *Let s be a regular sequence recognized by a trim representation (G, I, T). Let M be the adjacency matrix of G. The radius of convergence of s is the inverse of the maximal eigenvalue of M.*

Proof. The maximal eigenvalue λ of M is $\lambda = \limsup_{n\geq 0} \sqrt[n]{\|M^n\|}$, where $\|\ \|$ is any of the equivalent matrix norms. Let ρ be the radius of convergence of s and, for each $p, q \in Q$, let ρ_{pq} be the radius of convergence of the sequence $u_{pq} = (M^n_{pq})_{n\geq 0}$. Then $1/\lambda = \min \rho_{pq}$. Since (G, I, T) is trim, we have $\rho_{pq} \geq \rho$ for all $p, q \in Q$. On the other hand, $\rho \geq \min \rho_{pq}$ since s is a sum of some of the sequences u_{pq}. Thus $\rho_s = \min \rho_{pq}$ which concludes the proof. □

As a consequence of this result, the radius of convergence ρ of a regular sequence s is a pole. Indeed, with the above notation, $s(z) = \mathbf{i}(1 - Mz)^{-1}\mathbf{t}$. Then $\det(I - Mz)$ is a denominator of the rational fraction s, the poles of s are among the inverses of the eigenvalues of M. And since $1/\lambda$ is the radius of convergence of s, it has to be a pole of s. In particular, s diverges for $z = \rho$.

The following result, due to Berstel, is also well-known. It allows one to compute the radius of convergence of the star of a sequence.

PROPOSITION 2.7. *Let s be a regular sequence. The radius of convergence of the series $s^*(z) = 1/(1 - s(z))$ is the unique real number r such that $s(r) = 1$.* For a proof, see [11] pp 211-214, [10] p. 82 or [9] p. 84. As a consequence, we obtain the following result.

PROPOSITION 2.8. *Let s be a regular sequence and let λ be the inverse of the radius of convergence of s^*. The sequence s satisfies the Kraft strict inequality $s(1/k) < 1$ (resp. equality $s(1/k) = 1$) if and only if $\lambda < k$ (resp. $\lambda = k$).*

We have thus proved the following result, which is the basis of the constructions of the next sections.

PROPOSITION 2.9. *Let s be a regular sequence satisfying Kraft's inequality $s(1/k) \leq 1$. Let (G, i, t) be a normalized representation of s and let $(\overline{G}, i, i)$ be the closure of (G, i, t). The adjacency matrix M of $\overline{G}$ admits a k-approximate eigenvector.*

Actually, under the hypothesis of Proposition 2.9, the graph G itself also admits a k-approximate eigenvector. Indeed, let $\overline{\mathbf{w}} = (\overline{w}_q)_{q \in Q-t}$ be a k-approximate eigenvector of $\overline{G}$. Then the vector $\mathbf{w} = (w_q)_{q \in Q}$ defined by $w_q = \overline{w}_q$ for $q \neq t$ and $w_t = \overline{w}_i$ is a k-approximate eigenvector of G. This is illustrated in the following example.

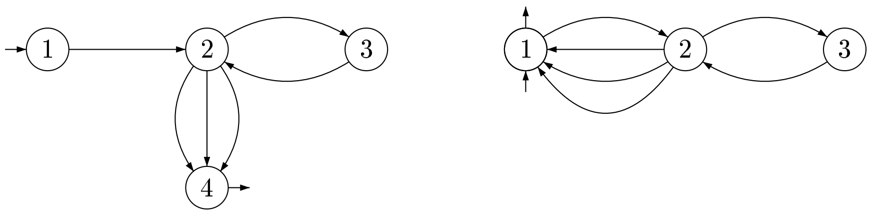

FIG. 2.6. *The graphs G and $\overline{G}$.*

Let us for example consider again $s(z) = 3z^2/(1 - z^2)$ (see Figure 2.5). The sequence s is recognized by the normalized representation $(G, 1, 4)$ where G is the graph represented on the left of Figure 2.6. The graph $\overline{G}$ is represented on the right. The vectors

$$\mathbf{w} = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \overline{\mathbf{w}} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

are 2-approximate eigenvectors of G and $\overline{G}$ respectively.

3. The multiset construction. In this section, we present the main construction used in this paper. It can be considered as a version with multiplicities of the subset construction used in automata theory to replace a finite automaton by an equivalent deterministic one. We use only unlabeled graphs but the construction can be easily generalized to graphs with edges labeled by symbols from an alphabet.

Our construction is also linked with one used by D. Lind to build a positive matrix with given spectral radius (see [15], especially Lemma 11.1.9).

We use for convenience the term *multiset* of elements of a set Q as a synonym of Q-vector. If $\mathbf{u} = (u_q)_{q\in Q}$ is such a multiset, the coefficient u_q is also called the *multiplicity* of q. The *degree* of $\mathbf{u}$ is the sum $\sum_{q\in Q} u_q$ of all multiplicities.

We start with a triple $(G, \mathbf{i}, \mathbf{t})$ where $G = (Q, E)$ is a finite graph and $\mathbf{i}$ (resp. $\mathbf{t}$) is a row (resp. column) Q-vector. We denote by M the adjacency matrix of G.

Let m be a positive integer. We define another triple $(H, \mathbf{J}, \mathbf{X})$ which is said to be obtained by the *multiset construction.* The graph H is called an *extension* of the graph G. The extension is not unique and depends as we shall see on some arbitrary choices. The set S of vertices of H is formed of multisets of elements of Q of total degree at most m. Thus, an element of S is a nonnegative vector $\mathbf{u} = (u_q)_{q\in Q}$ with indices in Q such that $\sum_{q\in Q} u_q \le m$. This condition ensures that H is a finite graph.

We now describe the set of edges of the graph H by defining its adjacency matrix N. Let U be the $S \times Q$-matrix defined by $U_{\mathbf{u},q} = u_q$. Then N is any nonnegative $S \times S$-matrix which satisfies

$$NU = UM.$$

Equivalently, for all $\mathbf{u} \in S$,

$$\sum_{\mathbf{v}\in S} N_{\mathbf{u},\mathbf{v}}\mathbf{v} = \mathbf{u}M.$$

Let us comment informally the above formula. We can describe the construction of the graph H as a sequence of choices. If we reach a vertex $\mathbf{u}$ of H, we partition the multiset $\mathbf{u}M$ of vertices reachable from the vertices composing $\mathbf{u}$ into multisets of degree at most m to define the vertices reachable from $\mathbf{u}$ in H. The integer $N_{\mathbf{u},\mathbf{v}}$ is the multiplicity of $\mathbf{v}$ in the partition. The formula simply expresses the fact that the result is indeed a partition. In general, there are several possible partitions. The matrix U is called the *transfer matrix* of the extension.

We further define the S-row vector $\mathbf{J}$ and the S-column vector $\mathbf{X}$. Let $\mathbf{J}$ be the S-row vector such that $J_{\mathbf{i}} = 1$ and $J_{\mathbf{u}} = 0$ for $\mathbf{u} \neq \mathbf{i}$. Let $\mathbf{X}$ be the S-column vector such that $X_{\mathbf{u}} = \mathbf{u} \cdot \mathbf{t}$.

Thus

$$\mathbf{J}U = \mathbf{i}, \quad \mathbf{X} = U\mathbf{t}.$$

To avoid unnecessary complexity, we only keep in S the vertices reachable from $\mathbf{i}$. Thus, we replace the set S by the set of elements $\mathbf{u}$ of S such that there is a path from $\mathbf{i}$ to $\mathbf{u}$.

The number of multisets of degree at most m on a set Q with n elements is $\frac{n^{m+1}-1}{n-1}$. Thus the number of vertices of a multiset extension is of order n^m. It is polynomial in n if m is taken as a constant.

Let for example G be the graph represented on Figure 3.1 on the left. The graph H represented on the right is a multiset extension of G with

$$\mathbf{i} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The matrices M, N and U are

$$M = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, N = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{J} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

FIG. 3.1. *The graphs G and H.*

In this case, the matrix U is invertible and the matrices M, N are conjugate.

The basic property of an extension is the following one.

PROPOSITION 3.1. *Let H be an extension of G. The triple $(H, \mathbf{J}, \mathbf{X})$ is equivalent to $(G, \mathbf{i}, \mathbf{t})$.*

Proof. For each $n \geq 0$, we have

$$UM^n = N^n U.$$

Consequently, for each integer $n \geq 0$,

$$\begin{aligned}\mathbf{J}N^n\mathbf{X} &= \mathbf{J}N^n U\mathbf{t}\\ &= \mathbf{J}UM^n\mathbf{t}\\ &= \mathbf{i}M^n\mathbf{t}.\end{aligned}$$

This shows that $(H, \mathbf{J}, \mathbf{X})$ recognizes s. □

We will also make use of the following additional property of extensions.

PROPOSITION 3.2. *Let H be an extension of G. Let M (resp. N) be the adjacency matrix of G (resp. H) and let U be the transfer matrix. If $\mathbf{w}$ is a k-approximate eigenvector of M, the vector $\mathbf{W} = U\mathbf{w}$ is a k-approximate eigenvector of N. If $\mathbf{w}$ is positive, then $\mathbf{W}$ is also positive.*

Proof. We have

$$N\mathbf{W} = NU\mathbf{w} = UM\mathbf{w} \leq kU\mathbf{w} = k\mathbf{W}.$$

Since all rows of U are distinct from $\mathbf{0}$, the vector $\mathbf{W}$ is positive whenever $\mathbf{w}$ is positive. □

In the next section, we will choose a particular extension of the graph G called admissible and which is defined as follows. Let $\mathbf{w}$ be a positive Q-vector and let m be a positive integer. Let H be an extension of G, let U be the transfer matrix, and let $\mathbf{W} = U\mathbf{w}$. We say that H is *admissible* with respect to $\mathbf{w}$ and m if for each $\mathbf{u} \in S$, all but possibly one of the vertices $\mathbf{v}$ such that $(\mathbf{u}, \mathbf{v})$ is an edge of H satisfy $W_{\mathbf{v}} \equiv 0 \bmod m$.

THEOREM 3.3. *For any graph G on Q, any positive Q-vector $\mathbf{w}$ and any integer $m > 0$, the graph G admits an admissible extension with respect to $\mathbf{w}$ and m.*

The proof relies on the following combinatorial lemma. This lemma is also used in a similar context by Adler et al. and Marcus [16],[1]. It is actually presented in [2] as a nice variant of the pigeon-hole principle.

LEMMA 3.4. *Let $w_1, w_2, \ldots, w_m$ be positive integers. Then there is a non-empty subset $S \subset \{1, 2, \ldots, m\}$ such that $\sum_{q \in S} w_q$ is divisible by m.*

Proof. The partial sums $w_1, w_1+w_2, w_1+w_2+w_3, \ldots, w_1+w_2+\cdots+w_m$ 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

$$w_1 + w_2 + \cdots + w_p \equiv w_1 + w_2 + \cdots + w_r (\bmod m)$$

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

Proof. of Theorem 3.3. We build progressively the set of edges of H. Let $\mathbf{u}$ be an element of S. We prove by induction on the degree $d(\mathbf{u}M) = \sum_{q \in Q} (\mathbf{u}M)_q$ of $\mathbf{u}M$ that there exists $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$ such that $\mathbf{u}M = \sum_{i=1}^n \mathbf{v}_i$ and $W_{\mathbf{v}_i} \equiv 0 \bmod m$ for $1 \leq i \leq n-1$. If $\mathbf{u}M \in S$, i.e. if $d(\mathbf{u}M) \leq m$, we choose $n = 1$ and $\mathbf{v}_1 = \mathbf{u}M$. Otherwise, there exists a decomposition $\mathbf{u}M = \mathbf{v} + \mathbf{u}'$ such that $d(\mathbf{v}) = m$. Let $w_1, w_2, \ldots, w_m$ be the sequence of integers formed by the w_q repeated v_q times. By Lemma 3.4 applied to the sequence of integers w_i, there is a decomposition $\mathbf{v} = \mathbf{v}' + \mathbf{r}$ with $\mathbf{v}' \neq 0$ such that $W_{\mathbf{v}'} \equiv 0 \bmod m$. We have $\mathbf{u}M = \mathbf{v}' + \mathbf{w}'$ with $\mathbf{w}' = \mathbf{r} + \mathbf{u}'$. Since $d(\mathbf{w}') < d(\mathbf{u}M)$, we can apply the induction hypothesis to $\mathbf{w}'$, giving the desired result. □

For an S-vector $\mathbf{W}$, we denote by $\lceil \frac{\mathbf{W}}{m} \rceil$ the S-vector $\mathbf{Z}$ such that for each $\mathbf{u}$ in S,

$$Z_{\mathbf{u}} = \lceil \frac{W_{\mathbf{u}}}{m} \rceil.$$

Summing up the previous results, we obtain the following statement.

PROPOSITION 3.5. *Let H be an admissible extension of G with respect to $\mathbf{w}$ and m. Let M (resp. N) be the adjacency matrix of G (resp. H), let U be the transfer matrix and let $\mathbf{W} = U\mathbf{w}$. If $\mathbf{w}$ is a positive k-approximate eigenvector of M, then $\lceil \frac{\mathbf{W}}{m} \rceil$ is a positive k-approximate eigenvector of N.*

Proof. By Proposition 3, the vector $\mathbf{W}$ is a positive k-approximate eigenvector of N. Thus

$$N\mathbf{W} \leq k\mathbf{W}.$$

Let $\mathbf{u}$ be an element of S. We have $W_{\mathbf{v}} \equiv 0 \bmod m$ for all indices $\mathbf{v}$ such that $N_{\mathbf{u},\mathbf{v}} > 0$ except possibly for an index $\mathbf{v}_0$. The previous inequality implies that

$$\sum_{\mathbf{v} \in S - \{\mathbf{v}_0\}} N_{\mathbf{u},\mathbf{v}} \frac{W_{\mathbf{v}}}{m} + N_{\mathbf{u},\mathbf{v}_0} \frac{W_{\mathbf{v}_0}}{m} \leq k \ \frac{W_{\mathbf{u}}}{m}.$$

Since $\frac{W_{\mathbf{v}}}{m}$ is a nonnegative integer for $\mathbf{v} \in Q - \{\mathbf{v}_0\}$, we get

$$\sum_{\mathbf{v} \in S - \{\mathbf{v}_0\}} N_{\mathbf{u},\mathbf{v}} \frac{W_{\mathbf{v}}}{m} + N_{\mathbf{u},\mathbf{v}_0} \lceil \frac{W_{\mathbf{v}_0}}{m} \rceil \leq k \ \lceil \frac{W_{\mathbf{u}}}{m} \rceil.$$

This proves that

$$N \lceil \frac{\mathbf{W}}{m} \rceil \leq k \lceil \frac{\mathbf{W}}{m} \rceil.$$

□

4. Generating sequence of leaves. In what follows, we state and prove, using the multiset construction, our main result concerning the generating sequences of regular trees. We begin with the following lemma, which is also used in the next section. We use the term leaf for a vertex of a graph without outgoing edges.

LEMMA 4.1. *Let G be a graph on a set Q of vertices. Let $i \in Q$ and $T \subset Q$. If G admits a k-approximate eigenvector $\mathbf{w}$, there is a graph G' and a set of vertices I' of G' such that*

1. *G' admits the k-approximate eigenvector $\mathbf{w}'$ with all components equal to 1.*
2. *the triple $(G, i, \mathbf{w})$ is equivalent to the triple $(G', I', \mathbf{w}')$;*
3. *If $w_p = 1$ for all $p \in T$, there is a set of vertices T' of G' such that the triple (G, i, T) is equivalent to the triple (G', I', T'). Moreover, if T is the set of leaves of G, we can choose for T' the set of leaves of G'.*

Proof. We first show that one can replace G by a graph without multiplicities, i.e. such that the adjacency matrix has coefficients 0 or 1.

For this, let n be the maximal value of the coefficients of M. Let Q' be the set of all pairs (p, j) for $p \in Q$ and $1 \le j \le n$. Let E' be the set of all pairs $((p, j), (q, h)) \in Q' \times Q'$ such that $1 \le j \le n$ and $1 \le h \le M_{p,q}$. Let $i' = (i, 1)$ and $T' = \{(t, j) \mid t \in T, 1 \le j \le n\}$. Let $G' = (Q', E')$. The triple (G', i', T') recognizes the same sequence as (G, i, T). Let $w'_{(p,j)} = w_p$ for all $p \in Q$ and all $1 \le j \le n$. The triple $(G', i', \mathbf{w}')$ recognizes the same sequence as $(G, i, \mathbf{w})$. The vector $\mathbf{w}'$ is a k-approximate eigenvector of M'.

We may thus suppose that all coefficients of M are 0 or 1, i.e. that the set E of edges of G can be identified with a subset of $Q \times Q$. We now transform the graph G into a graph G' such that there are at most k edges going out of every vertex. For this, let Q' be the set of pairs (q, j) with $q \in Q$ and $1 \le j \le w_q$. For each $p \in Q$, we have

$$\sum_{q \mid (p,q) \in E} w_q \le k w_p.$$

We may thus partition the pairs $(q, h) \in Q'$ in such a way that $(p, q) \in E$, in w_p groups $X_1, X_2, \ldots, X_{w_p}$ of at most k elements. The edges going out of (p, j) are all the pairs $((p, j), (q, h))$ such that $(q, h) \in X_j$. One can actually identify G with a multiset extension of G', where the set of multisets is $\{\bigcup_{1 \le j \le w_p} (p, j) \mid p \in Q\}$ that we identify to Q. Let $I' = \{(i, j) \mid 1 \le j \le w_i\}$. Let $w'_{(p,j)} = 1$ for all (p, j) with $p \in Q$ and $1 \le j \le w_p$. Then, according to Proposition 3.1, the triple $(G', I', \mathbf{w}')$ recognizes the same sequence as $(G, i, \mathbf{w})$. Moreover, if $w_p = 1$ for all $p \in T$, let T' be set of all $(p, 1) \in Q'$ with $p \in T$. Then the triple (G', I', T') recognizes the same sequence as (G, i, T). If T is the set of vertices that have no outgoing edges, it is clear that the same holds for T'. □

We now come to our main result.

THEOREM 4.2. *Let $s = (s_n)_{n \ge 0}$ be a regular sequence of nonnegative integers and let k be a positive integer such that $\sum_{n \ge 0} s_n k^{-n} \le 1$. Then there is a k-ary rational tree having s as its generating sequence.*

Proof. Let us consider a regular sequence s and an integer k such that $\sum_{n \ge 0} s_n k^{-n} \le 1$. Since the result holds trivially for $s(z) = 1$, we may suppose that $s_0 = 0$. Let (G, i, t) be a normalized representation of s and let $\overline{G}$ be the closure of G as defined at the beginning of Section 2.2. We denote by M (resp. $\overline{M}$) the adjacency matrix of G (resp. $\overline{G}$). Let $\overline{Q} = Q - \{t\}$ be the vertex set of $\overline{G}$. Let λ be the spectral radius of

$\overline{M}$. By Proposition 2.8, the matrix $\overline{M}$ admits a positive k-approximate eigenvector $\overline{\mathbf{w}}$. By definition, we have $\overline{M}\overline{\mathbf{w}} \leq k\overline{\mathbf{w}}$.

Let $\mathbf{w}$ be the Q-vector defined by $w_q = \overline{w}_q$ for all $q \in \overline{Q}$ and $w_t = \overline{w}_i$. Then, since there is no edge going out of t in G, $\mathbf{w}$ is a positive k-approximate eigenvector of M. Let $\mathbf{t}$ be the Q-vector which is the characteristic vector of the vertex t. Let $m = w_i$.

By Theorem 3.3 there exists an admissible extension H of G with respect to $\mathbf{w}$ and m. Let U be the transfer matrix and let $\mathbf{W} = U\mathbf{w}$. Since $w_t \equiv 0 \bmod m$, we may choose H with the following additional property. For all $\mathbf{u} \in S$ either $u_t = 0$ or $\mathbf{u} = \mathbf{t}$.

According to Proposition 3.1, the sequence s is recognized by $(H, \mathbf{J}, \mathbf{X})$ where $\mathbf{J}$ is the characteristic row vector of $\mathbf{i}$ and $\mathbf{X}$ is the characteristic column vector of $\mathbf{t}$. This means that s is recognized by the normalized representation consisting in the graph H, the initial vertex $\mathbf{i}$, that we identify to i, and the terminal vertex $\mathbf{t}$, that we identify to t.

Let N be the adjacency matrix of H. By Proposition 3.5, the vector $\lceil \frac{\mathbf{W}}{m} \rceil$ is a positive k-approximate eigenvector of N. Remark that $\lceil \frac{\mathbf{W}}{m} \rceil_i = \lceil \frac{\mathbf{W}}{m} \rceil_t = 1$.

We may now apply Lemma 4.1 to construct a triple (H', I', T') equivalent to (H, i, t). The set T' is the set of leaves of H'. Since $\lceil \frac{\mathbf{W}}{m} \rceil_i = 1$, I' is reduced to one vertex i'. Since H' admits a k-approximate eigenvector with all components equal to one, the graph H' is of outdegree at most k. Finally s is the generating sequence of the covering tree of H' starting at i'. This tree is k-ary and regular. □

Let us consider the above constructions in the particular case of the equality in Kraft's inequality. In this case, the result is a complete k-ary tree. Indeed, by Proposition 2.8, the matrix $\overline{M}$ admits a positive integral eigenvector $\mathbf{w}$ for the eigenvalue k. We have for all $p \in \overline{Q}$,

$$\sum_{q \in \overline{Q}} M_{p,q} w_q = k w_p.$$

As a consequence, for any $\mathbf{u} \neq \mathbf{t}$, we have

$$\sum_{\mathbf{v} \in S} N_{\mathbf{u},\mathbf{v}} W_{\mathbf{v}} = k W_{\mathbf{u}}.$$

Then the graph constructed in Lemma 4.1 is of constant outdegree k. Thus the k-ary tree obtained is complete.

Let us consider the complexity of the construction used in the proof of Theorem 4.2. Let n be the number of vertices of the graph G giving a normalized representation of s. The size of the integer $m = w_i$ is exponential in n (see Section 2.4). Thus the number of vertices of the graph H is bounded by a double exponential in n. The final regular tree is the covering tree of a graph whose set of vertices has the same size in order of magnitude.

Let for example s be the sequence defined by

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

Since $s(1/2) = 1$, it satisfies the Kraft equality for $k = 2$. The sequence s is recognized by (G, i, t) where $G = (Q, E)$ is the graph given in Figure 4 with $Q = \{1, 2, 3, 4, 5, 6, 7\}$, $i = 1$, $t = 4$. The adjacency matrix of G admits the 2-approximate eigenvector

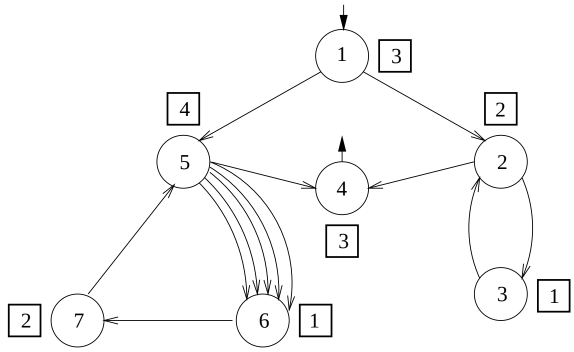

FIG. 4.1. *A normalized representation of s*

represented on Figure 4, where the coefficients of $\mathbf{w}$ are represented in squares beside the vertices. Thus $m = 3$.

An admissible extension H of G with respect to $\mathbf{w}$ and m is given in Figure 4.2. In this figure, each multiset of S is represented by a sequence of vertices with repetitions corresponding to the multiplicity. For example, the multiset $\mathbf{u} = (0,0,1,0,0,2,0)$ is represented by $(3,6,6)$. The sequence s is recognized by the normalized representation $(H,1,4)$, where the initial and final vertices are named as they appear on Figure 4.2. The coefficients of $\lceil \frac{\mathbf{W}}{m} \rceil$ are represented in squares beside the vertices.

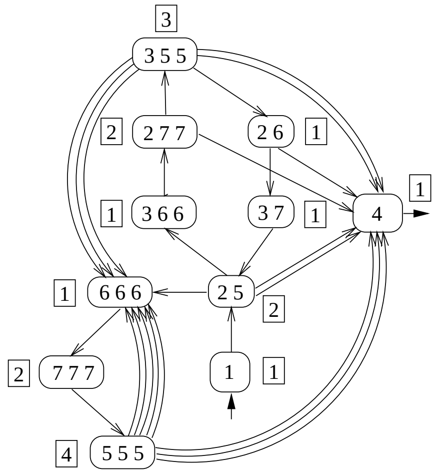

FIG. 4.2. *An admissible extension H.*

A regular binary tree T having s as generating sequence of leaves, is given in Figure 4.3. In this figure, the nodes have been renumbered, with the children of a node with a given label represented only once. The leaves of the tree are indicated by black boxes. The tree itself is obtained from the graph of Figure 4.2 by application of the construction of Lemma 4.1. For example, the vertex $(2,5)$, which has coefficient 6 in $\mathbf{W}$, is split into two vertices named 2 and 3 in the tree.

This example was suggested to us by Christophe Reutenauer [18].

5. Generating sequence of nodes. In this section, we consider the generating sequence of the set of all nodes in a tree instead of just the set of leaves. This is

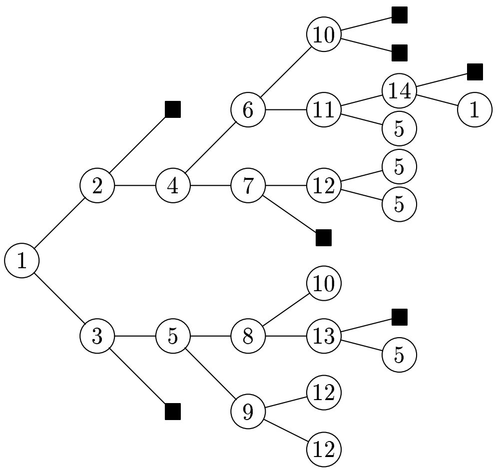

FIG. 4.3. *A regular binary tree with length distribution s.*

motivated by the fact that in search trees, the information can either be carried by the leaves or by all the nodes of the tree. We will see that the complete characterization of the generating sequences of nodes in regular trees (Theorem 5.1) is more complicated than the one for leaves.

Soittola (see [19] p. 104) has characterized the series which are the generating sequences of nodes in a regular tree. We characterize the ones that correspond to k-ary trees (Theorem 5.1). We also give a more direct construction in a particular case (Theorem 5.3).

Let T be a tree. The generating sequence of nodes of the tree T is the sequence $t = (t_n)_{n\geq 0}$, where t_n is the number of nodes of T at height n. The sequence t satisfies $t_0 \leq 1$ and, moreover, if T is a k-ary tree, the condition

$$t_n \leq kt_{n-1}$$

for all $n \geq 1$. If T is a regular tree, then t is a regular sequence. We now completely characterize the regular sequences t that are the generating sequences of nodes of a k-ary regular tree.

THEOREM 5.1. *Let $t = (t_n)_{n\geq 0}$ be a regular sequence and let k be a positive integer. The sequence $(t_n)_{n\geq 0}$ is the generating sequence of nodes of a k-ary regular tree iff it satisfies the following conditions.*

(i) *the convergence radius of t is strictly greater than $1/k$,*

(ii) *the sequence $s(z) = t(z)(kz-1)+1$ is regular.*

Proof. Let us first show that the conditions are necessary. Let $\overline{T}$ be the complete k-ary tree obtained by adding i new leaves to each node that has $k-i$ children. Since T is a regular tree, $\overline{T}$ is also regular.

Let s be the generating sequence of leaves of $\overline{T}$. Since $\overline{T}$ is complete, $s(1/k) = 1$. Since $kt_n = s_{n+1} + t_{n+1}$ for all $n \geq 0$, we have

$$1 - s(z) = t(z)(1 - kz).$$

Since s is a regular sequence, its radius of convergence is strictly larger than $1/k$ (see

Section 2.4). Since the value of the derivative of s at $z = 1/k$ is $kt(1/k)$, the same holds for t. This proves the necessity of the conditions.

Conversely, if t satisfies the conditions of the theorem, the regular series $s(z) = t(z)(kz - 1) + 1$ satisfies $s(1/k) = 1$. Thus, by Theorem 4.2, s is the generating sequence of leaves of a complete k-ary regular tree. The internal nodes of this tree form a k-ary regular tree whose generating sequence of nodes is t. □

The sequence s defined by condition (ii) is rational as soon as t is regular and therefore rational. Given a regular sequence t, condition (ii) is decidable in view of a theorem of Soittola [19], also found independently in [13] and recalled below. We say that a rational sequence has a *dominating root*, either if it is a polynomial or if it has a real positive pole which is strictly smaller than the modulus of any other one. A sequence r is a *merge* of the sequences r_i if there is an integer p such that

$$r(z) = \sum_{i=0}^{p-1} z^i r_i(z^p).$$

THEOREM 5.2 (Soittola). *A sequence of nonnegative integers $r = (r_n)_{n \geq 0}$ is regular if and only if it is a merge of rational sequences having a dominating root.*

This result shows that it is decidable if a rational series is regular (see [19]). In the positive case, there is an algorithm computing a representation of the sequence.

We may observe that condition (ii) of the theorem implies the non-negativity of the coefficients of the series s and thus the inequality $\forall n \geq 1, t_n \leq kt_{n-1}$. It also implies that $t_0 \leq 1$.

We now show that there are regular sequences t satisfying $t_n \leq kt_{n-1}$ for all $n \geq 1$, and condition (i) of the theorem and such that the sequence $s(z) = t(z)(kz - 1) + 1$ is not regular. The example is based on an example of a rational sequence with nonnegative coefficients and which is not regular (see [10] page 95). Let

$$r_n = b^{2n} cos^2(n\theta)$$

with $cos(\theta) = \frac{a}{b}$ where the integers a, b are such that $b \neq 2a$ and $0 < a < b$. The sequence r is rational, has nonnegative integer coefficients and is not regular. Its poles are $\frac{1}{b^2}$, $\frac{1}{b^2}e^{2i\theta}$ and $\frac{1}{b^2}e^{-2i\theta}$. We now define the sequence t as follows:

$$\begin{aligned} t_{2h} &= k^h, \\ t_{2h+1} &= k^h + r_h. \end{aligned}$$

We also assume that $b^2 < k$. By Soittola's theorem, the sequence t is regular since it is a merge of rational sequences having a dominating root. The convergence radius of t is $\frac{1}{\sqrt{k}} > \frac{1}{k}$. Therefore the sequence t satisfies the first condition of Theorem 5.1. Let s be the sequence defined by $s(z) = t(z)(kz - 1) + 1$. If $h = 2p$ is even,

$$\begin{aligned} s_h &= kt_{h-1} - t_h \\ &= kk^{p-1} + kr_{p-1} - k^p + 1 = kr_{p-1} + 1. \end{aligned}$$

Thus the sequence s is not regular.

The above example does not work for the small values of k (the least value is $k = 10$). We do not know of similar examples for $2 \leq k \leq 9$.

We finally describe a particular case of Theorem 5.1 in which one has a relatively simple method, based on the multiset construction, to build the regular tree with a given generating sequence of nodes. This avoids the use of Soittola's characterization which leads to a method of higher complexity.

A *primitive* representation of a regular sequence s is a representation $(G, \mathbf{i}, \mathbf{t})$ such that the adjacency matrix of G is primitive. The following result is proved in [7] with a different proof using the state-splitting method of symbolic dynamics. The proof given here relies on a simpler construction.

THEOREM 5.3. *Let $t = (t_n)_{n\geq 0}$ be a regular sequence and let k be a positive integer such that $t_0 = 1$, $t_n \leq kt_{n-1}$ for all $n \geq 1$ and such that*

(i) *the convergence radius of t is strictly greater than $1/k$,*

(ii) *t has a primitive representation.*

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

We are going to give a proof of the theorem which uses the multiset construction. We shall use the following lemma that we establish first.

LEMMA 5.4. *Let M be a primitive matrix with spectral radius λ. Let $\mathbf{v}$ be a non-null and nonnegative integral vector and let k be an integer such that $\lambda < k$. Then there is a positive integer n such that $M^n\mathbf{v}$ is a positive k-approximate eigenvector of M.*

Proof. For a primitive matrix M with spectral radius λ, it is known that the sequence $((\frac{M}{\lambda})^n)_{n\geq 0}$ converges to $\mathbf{r}.\mathbf{l}$ where $\mathbf{r}$ is a positive right eigenvector and $\mathbf{l}$ a positive left eigenvector of M for the eigenvalue λ with $\mathbf{l}\cdot\mathbf{r} = 1$ (see for example [15] p. 130). Thus $(\frac{M^n}{\lambda^n}\mathbf{v})_{n\geq 0}$ converges to $\mathbf{r}.\mathbf{l}.\mathbf{v}$ which is equal to $\rho\mathbf{r}$ where ρ is a nonnegative real number. Since $M\mathbf{r} = \lambda\mathbf{r}$, we get, for a large enough integer n,

$$M\frac{M^n}{\lambda^n}\mathbf{v} \leq k\frac{M^n}{\lambda^n}\mathbf{v}$$

or equivalently $MM^n\mathbf{v} \leq kM^n\mathbf{v}$. If n is large enough, we moreover have $M^n\mathbf{v} > 0$ since M is primitive. ◻

We now give the proof of Theorem 5.3. It uses a shift of indices of the sequence to obtain a new sequence to which a simple application of the multiset construction can be applied.

Proof. Since t is regular, it is recognized by a triple $(G, \mathbf{i}, \mathbf{t})$, where $G = (Q, E)$ is a finite graph. Let M be the adjacency matrix of G.

For each $n \geq 0$, we have

$$t_n = \mathbf{i}M^n\mathbf{t}.$$

We denote by λ the spectral radius of M. By Proposition 2.6 the positive real number $1/\lambda$ is the radius of convergence of t. Thus $\lambda < k$ by hypothesis (i). Since M is a primitive matrix, by Lemma 5.4, there exists a positive integer n_0 such that $M^{n_0}\mathbf{t}$ is a positive k-approximate eigenvector of M.

Let $\mathbf{w} = M^{n_0}\mathbf{t}$, and let t' be the sequence defined by $t'_n = t_{n+n_0}$ for $n \geq 0$. Thus, for each $n \geq 0$,

$$t'_n = \mathbf{i}M^n\mathbf{w}.$$

The sequence t' is thus recognized by the triple $(G, \mathbf{i}, \mathbf{w})$. Note that $t'_0 = \mathbf{i}\cdot\mathbf{w}$.

Let $H = (S, R)$ be the extension of G obtained by the multiset construction in the following way. When we reach a vertex $\mathbf{u}$ of H, we partition $\mathbf{u}M$ in multisets $\mathbf{v}$ of degree 1, i.e. such that $\mathbf{v}$ is a $0,1$-vector with $v_q = 0$ for all $q \in Q$ except one. All elements of S are thus elements of Q except perhaps the initial vertex $\mathbf{i}$. If $\mathbf{i}$ is of degree 1, the number of elements of S is then equal to the number of elements of Q.

Let U be the transfer matrix of the extension. Since $\mathbf{w}$ is a positive k-approximate eigenvector of M, by Proposition 3.2, the vector $\mathbf{W} = U\mathbf{w}$ is a positive k-approximate eigenvector of the adjacency matrix of H. By Proposition 3.1, the triple $(H, \mathbf{i}, \mathbf{W})$ is equivalent to $(G, \mathbf{i}, \mathbf{w})$.

We now apply Lemma 4.1 to the graph H. We use $\mathbf{i}$ as initial vertex and the k-approximate eigenvector $\mathbf{W}$. Since we only use the first assertion of the lemma, we will not use any set T of terminal states. According to the lemma, we construct a graph H' and a set of vertices I' of H' such that H' admits the k-approximate vector $\mathbf{W}'$ with all components equal to 1, and $(H', I', \mathbf{W}')$ is equivalent to $(H, \mathbf{i}, \mathbf{W})$. Thus H' is k-ary. Note that I' has $W_{\mathbf{i}} = \mathbf{i} \cdot \mathbf{w} = t_{n_0}$ elements.

Let T_p be the covering tree of H' starting at the state p of I'. Each T_p is a regular k-ary tree. Then t' is the sum the generating sequences of nodes of the trees T_p for $p \in I'$.

Finally, we build a finite k-ary tree T' whose generating sequence of nodes is $(t_0, t_1, \ldots, t_{n_0})$. This can actually be done since $t_0 = 1$ and $t_n \leq kt_{n-1}$ for $n \geq 1$. We then identify bijectively each leaf at height n_0 of T' to the root of a tree T_j. We get a regular k-ary binary tree whose generating sequence of nodes is t. ☐

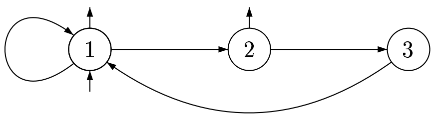

FIG. 5.1. *A primitive representation G of t.*

Let for example t be the series recognized by the graph G of Figure 5.1 with

$$\mathbf{i} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{t} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

The adjacency matrix M of G is the primitive matrix

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Its spectral radius is less than 2. The hypothesis of Theorem 5.3 are thus satisfied. We apply the method described above. We have

$$M^2\mathbf{t} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \text{ and } M^3\mathbf{t} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}.$$

Since $M^3\mathbf{t} \leq 2M^2\mathbf{t}$, $M^2\mathbf{t}$ is an approximate eigenvector of M. We thus set $n_0 = 2$ and $\mathbf{w} = M^2\mathbf{t}$. The graph H is the same as the graph G of Figure 5.1. The vector $\mathbf{W}$ is thus

$$\mathbf{W} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}.$$

The graph H' is represented on the left side of Figure 5.2. We finally obtain the binary regular tree T represented on the right side of Figure 5.2 (the nodes of the tree have been renumbered).

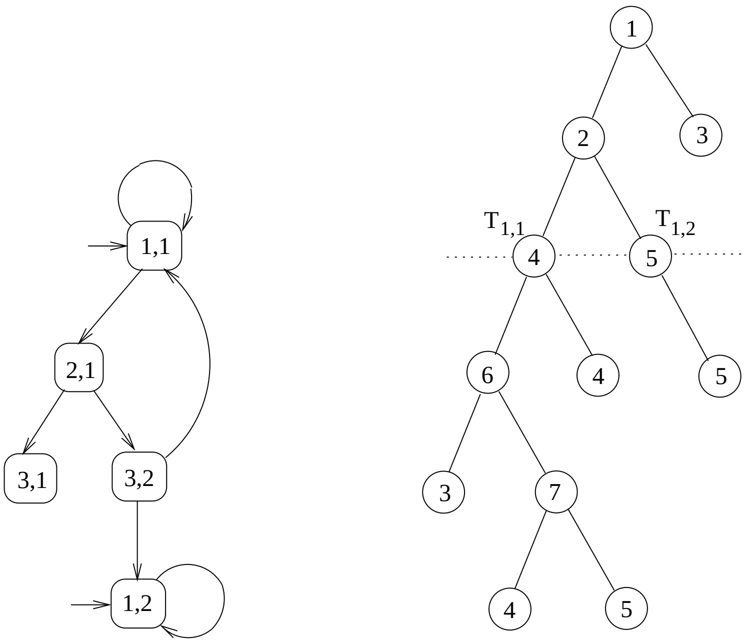

FIG. 5.2. *The graph H' and the tree T.*

Acknowledgements. The authors would like to thank Jean Berstel, Christophe Reutenauer and the anonymous referees for their help to improve the presentation of our paper.

REFERENCES

[1] R. L. ADLER, D. COPPERSMITH, AND M. HASSNER, *Algorithms for sliding block codes*, IEEE Trans. Inform. Theory, IT-29 (1983), pp. 5–22.
[2] M. AIGNER AND G. M. ZIEGLER, *Proofs from The Book*, Springer-Verlag, 1998.
[3] R. B. ASH, *Information Theory*, Dover Publications, Inc, New-York, 1990.
[4] J. J. ASHLEY, *A linear bound for sliding-block decoder window size*, IEEE Trans. Inform. Theory, (1988).
[5] F. BASSINO, M.-P. BÉAL, AND D. PERRIN, *Enumerative sequences of leaves in rational trees*, in ICALP'97, no. 1256 in Lecture Notes in Computer Science, Springer-Verlag, 1997, pp. 76–86.
[6] ———, *Super-state automata and rational trees*, in LATIN'98, C. L. Lucchesi and A. V. Moura, eds., no. 1380 in Lecture Notes in Computer Science, Springer-Verlag, 1998, pp. 42–52.
[7] ———, *Enumerative sequences of leaves and nodes in rational trees*, Theoret. Comput. Sci., (1999), pp. 41–60.
[8] ———, *Length distributions and regular sequences*, in Codes, Systems and Graphical Models, J. Rosenthal and B. Marcus, eds., IMA Volumes in Mathematics and its Applications, Springer-Verlag, 1999. To appear.
[9] M.-P. BÉAL, *Codage Symbolique*, Masson, 1993.
[10] J. BERSTEL AND C. REUTENAUER, *Rational Series and their Languages*, Springer-Verlag, 1988.
[11] S. EILENBERG, *Automata, Languages and Machines*, vol. A, Academic Press, 1974.
[12] F. R. GANTMATCHER, *Matrix Theory, Volume II*, Chelsea Publishing Company, New-York, 1960.

[13] T. Katayama, M. Okamoto, and H. Enomoto, *Characterization of the structure-generating of regular sets and DOL growth functions*, Information and Control, 36 (1978), pp. 85–101.

[14] B. P. Kitchens, *Symbolic Dynamics*, Springer-Verlag, 1997.

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

[16] B. H. Marcus, *Factors and extensions of full shifts*, Monats. Math, 88 (1979), pp. 239–247.

[17] D. Perrin, *Finite automata*, in Handbook of Theoretical Computer Science, J. V. Leeuven, ed., vol. B, Elsevier, 1990, ch. 1.

[18] C. Reutenauer, *personal communication.* 1997.

[19] A. Salomaa and M. Soittola, *Automata Theoretic Properties of Formal Power Series*, Springer-Verlag, 1978.