Advances in Applied Mathematics **26,** 129–153 (2001)
doi:10.1006/aama.2000.0711, available online at http://www.idealibrary.com on IDEAL®

A Technology for Reverse-Engineering a Combinatorial Problem from a Rational Generating Function

E. Barcucci

Dipartimento di Sistemi e Informatica, Università di Firenze, Via Lombroso 6/17, 50134 Firenze, Italy
E-mail: barcucci@dsi.unifi.it

A. Del Lungo and A. Frosini

Dipartimento di Matematica, Università di Siena, Via del Capitano 15, 53100 Siena, Italy
E-mail: dellungo@unisi.it, frosini@unisi.it

and

S. Rinaldi

Dipartimento di Sistemi e Informatica, Università di Firenze, Via Lombroso 6/17, 50134 Firenze, Italy
E-mail: rinaldi@dsi.unifi.it

Received June 4, 2000; accepted September 15, 2000

In this paper, we tackle the problem of giving, by means of a regular language, a combinatorial interpretation of a positive sequence (f_n) defined by a linear recurrence with integer coefficients. We propose two algorithms able to determine if the rational generating function of (f_n), $f(x)$, is the generating function of some regular language, and, in the affirmative case, to find it. We illustrate some applications of this method to combinatorial object enumeration problems and bijective combinatorics and discuss an open problem regarding languages having a rational generating function.

INTRODUCTION

In several papers of enumerative and bijective combinatorics a combinatorial interpretation is requested of some positive sequence (f_n) defined by a linear recurrence with integer coefficients, that is, positive sequences

0196-8858/01 $35.00

having a rational generating function $f(x) = \sum_{n \geq 0} f_n x^n$. Many papers are indeed devoted to finding classes of combinatorial objects which are enumerated by the given sequence to an enumerative parameter. To provide interesting examples we only quote some of them:

(1) $f_n = 6f_{n-1} - f_{n-2}$, with $f_0 = 1$, $f_1 = 7$. This recurrence defines the numbers known as NSW numbers (sequence M4423 of [20]). They count the total area of the region under elevated Schröder paths [1, 16, 22].

(2) $f_n = 2f_{n-1} + f_{n-2}$, with $f_0 = 0$, $f_1 = 1$. These are the well-known Pell numbers (sequence M1413 of [20]).

(3) $f_n = 3f_{n-1} - f_{n-2}$, with $f_0 = 1$, $f_1 = 2$. These are the Fibonacci numbers having an odd index (sequence M1439 of [20]). They enumerate *dcc*-polyominoes according to the cell's number [3].

(4) $f_n = f_{n-1} + 3f_{n-2} + 2f_{n-3} - 3f_{n-4} + f_{n-5} + f_{n-6}$ for $n > 7$, with $f_0 = 1$, $f_1 = 3$, $f_2 = 6$, $f_3 = 12$, $f_4 = 20$, $f_5 = 36$, $f_6 = 58$, $f_7 = 100$. These numbers do not appear in [20]. They enumerate the n-step self avoiding walks contained in the strip $\{0, 1\} \times [-\infty, \infty]$ [24].

In October 1999, Jim Propp posted on the "domino" mailing list the problem of determining a combinatorial interpretation of a sequence of positive numbers defined by a linear recurrence, and then he asked: "Is there a technology for reverse-engineering a combinatorial problem from a rational generating function?" This means that, given a sequence (f_n) of integer numbers defined by a linear recurrence, we have to determine classes of combinatorial objects which are enumerated by it. Our aim is to provide such a technology based on regular languages and on the theory of their generating functions. The generating function $f_{\mathscr{L}}(x)$ of a language $\mathscr{L}$ is the formal power series $f_{\mathscr{L}}(x) = \sum_{n \geq 0} f_n x^n$, such that $\forall n \in \mathbb{N}$, $f_n = \|\{w \in \mathscr{L} : \lg(w) = n\}\|$, where $\lg(w)$ denotes the length of w.

We reach our goal by determining the "inverse" of the so-called *Schützenberger methodology*. In an earlier paper, Chomsky and Schützenberger [7] proposed a methodology for determining the generating function of an unambiguous context free (u.c.f.) language $\mathscr{L}$ from an unambiguous grammar G, such that $L(G) = \mathscr{L}$. Let $G = \langle V, \Sigma, P, S \rangle$ be an u.c.f. grammar (where V denotes the set of variables and Σ the alphabet); the following morphism Θ on $(\Sigma \cup V)^*$ is defined:

$$
\begin{aligned}
\Theta(a) &= x && \text{if } a \in \Sigma, \\
\Theta(\lambda) &= 1, \\
\Theta(\alpha) &= \alpha(x) && \text{if } \alpha \in V.
\end{aligned}
$$

By applying Θ to the set P of G's productions, we obtain an algebraic system of equations in $\alpha(x)$, for all α in V,

$$\Theta(\alpha_i) = \sum_{j_i=1}^{n_i} \Theta\left(\beta_{i_{j_i}}\right), \qquad \text{where } \alpha_i \to \beta_{i_1}|\dots|\beta_{i_{n_i}} \text{ is a production of } P.$$

Now let $\Theta(L) = \sum_{w \in L} \Theta(w) = \sum_{n \geq 0} f_n x^n$. One can prove that there is a unique solution $S(x)$ which represents the generating function for $\mathscr{L}$. We usually use $f_G(x)$ instead of $f_{L(G)}(x)$ to indicate this function. We quote [7] the following theorems:

(i) If G is an u.c.f. grammar, then $f_G(x)$ is algebraic.

(ii) If G is an unambiguous regular (u.r.) grammar, then $f_G(x)$ is rational.

This technique is called the Schützenberger methodology and represents a very powerful tool for solving many combinatorial objects enumeration problems, when it is possible to describe the class of objects by means of a context free grammar (see also [8, 9]). An example of this is the Dyck walks, bijectively corresponding to the Dyck words, generated by the grammar $G = S \to aSbS \mid \lambda$.

By applying the Schützenberger methodology we obtain

$$S(x) = x^2 S^2(x) + 1,$$

$$S(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x^2} = 1 + x^2 + 2x^4 + 5x^6 + 14x^8 \dots .$$

The problem concerning the "inverse" of the Schützenberger methodology can be defined as follows: given a function $F(x)$, we want to establish if there exists an unambiguous context-free language $\mathscr{L}$ such that $F_{\mathscr{L}}(x) = F(x)$. This problem on rational functions was previously raised by Cori during a talk given by Del Ristoro (Labri, 1997, "Une interprétation combinatoire de la recurrence $f_{n+1} = 6f_n - f_{n-1}$").

Providing the "inverse" of the Schützenberger methodology for rational functions (i.e., $F(x)$ is rational), we solve both the problems proposed by Cori and by Propp (i.e., we give a reverse technology for the combinatorial problems that can be solved on the class of regular languages).

The first non-trivial problem when working with a rational function $F(x) = \sum_{n \geq 0} f_n x^n$ is that of deciding if every coefficient f_n is non-negative. This problem was proposed originally in [19]. The S-unit Theorem [10, Sect. 4.3] guarantees that it suffices to look at sufficiently many initial terms of the sequence (f_n) to determine positivity.

Going further in determining the inverse methodology, we give a characterization of the set $\mathscr{R}$ of regular languages' generating functions. A powerful tool for obtaining an analytic characterization of $\mathscr{R}$ is the theory of formal power series, and, in particular, the $\mathbb{N}$-rational series, proposed in the 1970s by Eilenberg and Soittola and Salomaa [11, 19, 21], and most recently treated by Berstel and Reutenauer [6] and Perrin [17]. Many important theorems of these authors helped us to determine two algorithms for:

(i) deciding if a given rational function lies in $\mathscr{R}$ (Algorithm 1, Section 3);

(ii) building an unambiguous regular expression for a function in $\mathscr{R}$ (Algorithm 2, Section 3).

These algorithms define the inverse of the Schützenberger methodology for rational functions, thus solving Propp's problem in terms of languages. This, in turn, means that given an integer positive sequence (f_n) defined by a linear recurrence with integer coefficients, we can decide if a regular language enumerated by this sequence exists and, if so, we can find it.

Now the question concerning our reverse technology is how "large" is the class of combinatorial problems that can be solved on regular languages, with respect to the class of all integer positive sequences defined by linear recurrences with integer coefficients. Unfortunately, there are classes of rational functions giving positive integer sequences but not deriving from any regular language.

We wish to point out that considering combinatorial problems on the class of unambiguous context-free languages does not improve the reverse technology. In fact, it is possible to prove that if an unambiguous context-free grammar has a rational generating function, then it must be $\mathbb{N}$-rational; i.e., a regular language having the same generating function does exist (see Open Problems).

We conclude that the problem of discovering a language, or another combinatorial object, whose generating function is $\mathbb{Z}$-rational but not $\mathbb{N}$-rational is still open.

1. FROM REGULAR LANGUAGES TO RATIONAL FUNCTIONS

In this section we deal with regular languages and give a characterization of the set of generating functions for a regular language $\mathscr{L}$. By Kleene's Theorem [13], we can work on regular expressions instead of languages. It is known that for every regular language, $\mathscr{L}$, it is possible to find an unambiguous regular expression (u.r.e.), such that $L(E) = \mathscr{L}$ (with regard to the concept of unambiguity of regular expressions, see [14]). Furthermore, given an unambiguous regular expression E, we use a simple method for obtaining its generating function $f_E(x) = f_{L(E)}(x)$:

(1) if $E = a, \emptyset, \lambda$ with $a \in \Sigma$, then $f_E(x) = x, 0, 1$, respectively;

(2) if $E = F \vee G$, FG, $E = F^*$ then $f_E(x) = f_F(x) + f_G(x)$, $f_F(x)f_G(x)$, $f_E(x) = 1/(1 - f_F(x))$, respectively.

We denote the set of the generating functions of a regular expression by $\mathscr{F}_{\mathscr{R}}$. Obviously, this set coincides with the set of the generating functions of a u.r.e. Now we define a set of rational functions $\mathscr{R}$ as follows:

(1) $0, 1, kx \in \mathscr{R}$, with $k \in \mathbb{N}$;

(2) if $f(x), g(x) \in \mathscr{R}$, and $f(0) + g(0) \leq 1$, then $f(x) + g(x) \in \mathscr{R}$;

(3) if $f(x), g(x) \in \mathscr{R}$, then $f(x)g(x) \in \mathscr{R}$;

(4) $f(x) \in \mathscr{R}$, and $f(0) = 0$, then $\frac{1}{1 - f(x)} \in \mathscr{R}$.

Let h be a function from the class of regular expressions into $\mathbb{N}$ such that:

(1) $h(\emptyset) = h(a) = 0$ with $a \in \Sigma^*$;

(2) $h(F \vee G) = h(FG) = \max\{h(G), h(F)\}$;

(3) $h(F^*) = h(F) + 1$.

$h(E)$ is called the *star-height* of a regular expression E and represents the maximal number of nested stars in the regular expression. The star-height of the generating function $f_E(x) = \sum_{n \geq 0} f_n x^n$ is the minimum among the star-heights of the regular expressions describing the language $L(E)$ (see [11]).

We now inductively define a procedure $ren : r.e. \to r.e.$ that maps a r.e. into a new r.e. such that all symbols are distinct and preserves unambiguity, by simply extending its alphabet.

Let E be the input r.e.; we initialize an alphabet Σ as empty. We define $ren(E)$ in the following way:

—if $E = \lambda$, then $ren(E) := \lambda$;

—if $E = a$ and $a \notin \Sigma$, then $ren(E) := a$ and $\Sigma := \Sigma \cup \{a\}$;

—if $E = a$ and $a \in \Sigma$, then $ren(E) := b$ with $b \notin \Sigma$ and $\Sigma := \Sigma \cup \{b\}$;

—if $E = y \vee z$, $E = yz$, $E = y^*$ (with y, z r.e.), then $ren(E) := ren(y) \vee ren(z)$, $ren(y)ren(z)$, $(ren(y))^*$.

EXAMPLE 1.1. Let $E = (((ab \vee aa)^* a)^* cd)$ and $\Sigma = \emptyset$:

$$\begin{aligned} ren(E) &= ren(((ab \vee aa)^* a)^*)ren(cd) \\ &= (ren((ab \vee aa)^*)ren(a))^* ren(c)ren(d) \\ &= ((ren(ab) \vee ren(aa))^* a)^* cd \\ &= ((ren(a)ren(b) \vee ren(a)ren(a))^* a)^* cd \\ &= ((eb \vee fg)^* a)^* cd \qquad (\text{with } \Sigma = \{a, b, c, d, e, f, g\}). \end{aligned}$$

LEMMA 1.1. *Let E be a u.r.e.; we have:*

(i) *$ren(E)$ is u.r.e.;*

(ii) $f_E(x) = f_{ren(E)}(x)$;

(iii) *if $f_E(0) = 0$, then $[ren(E)]^*$ is an unambiguous regular expression.*

Proof. Parts (i), (ii) trivially follow from the definition of the *ren* procedure.

(iii) We only have to prove that $ren(E)$ is a code. It follows from the fact that $ren(E)$ is constructed on an arbitrarily large alphabet (see also the theory of codes [5]). ∎

THEOREM 1.1. $\mathscr{R} = \mathscr{F}_{\mathscr{R}}$.

Proof. $(\supseteq)$ By induction on the number of steps to get $f(x) = \mathscr{F}_{\mathscr{R}}$:

Base. If $f(x) = 0, 1, kx$, then $f(x) \in \mathscr{R}$.

Step.

—$f(x) = g(x) + h(x)$. $g(x), h(x) \in \mathscr{R}$ by inductive hypothesis; moreover, there are two u.r.e. G, H, such that $G \cap H = \emptyset$, $G \vee H = F$, $f(x) = f_F(x)$, and F u.r.e. It is clear that $g(0) + h(0) \leq 1$; otherwise, $\lambda \in G \cap H$ would go against our hypothesis. Finally, $g(x) + h(x) = f(x) \in \mathscr{R}$.

—$f(x) = g(x)h(x)$. $g(x), h(x) \in \mathscr{R}$ by inductive hypothesis, so $g(x)h(x) = f(x) \in \mathscr{R}$.

—$f(x) = \frac{1}{1 - g(x)}$. By inductive hypothesis $g(x) \in \mathscr{R}$; let F and G be two u.r.e., such that $f_F(x) = f(x)$ and $f_G(x) = g(x)$. We therefore get $g(0) = 0$; otherwise, $\lambda \in G^0 \cap G^1$, with F being a code. Consequently, $f(x) \in \mathscr{R}$.

($\subseteq$) By induction on $f(x) \in \mathscr{R}$:

Base. $0, 1, kx \in \mathscr{F}_{\mathscr{R}}$ by definition.

Step.

—$f(x) = g(x) + h(x)$ and $g(0) + h(0) \leq 1$. The proof is obvious if $g(0) + h(0) = 0$. Let us then assume that $g(0) + h(0) = 1$. By inductive hypothesis two u.r.e. G, H exist, such that $\lambda \notin G \cap H$ and $f_G(x) = g(x)$, $f_H(x) = h(x)$. By Lemma 1.1, we know that $f_{ren(G)}(x) = g(x)$ and $f_{ren(H)}(x) = h(x)$ and $ren(G) \cap ren(H) = \emptyset$. It trivially follows that $F_{ren(G) \vee ren(H)}(x) = F_{ren(G)}(x) + F_{ren(H)}(x) = g(x) + h(x) = f(x)$ and $f(x) \in \mathscr{F}_{\mathscr{R}}$.

—$f(x) = g(x)h(x)$. By inductive hypothesis, two u.r.e. G, H exist, such that $f_G(x) = g(x)$, $f_H(x) = h(x)$. By Lemma 1.1 we obtain $f_{ren(G)}(x) = g(x)$ and $f_{ren(H)}(x) = h(x)$. It follows that $f_{ren(G)ren(H)}(x) = f_{ren(G)}(x) f_{ren(H)}(x) = g(x)h(x) = f(x)$ and $f(x) \in \mathscr{F}_{\mathscr{R}}$.

—$f(x) = \frac{1}{1 - g(x)}$. We know that $g(x) \in \mathscr{R}$ and $g(0) = 0$. By inductive hypothesis $\exists G$ u.r.e., such that $f_G(x) = g(x)$ and $\lambda \notin G$. By Lemma 1.1, we get $f_{ren(G)}(x) = g(x)$ and $ren(G)^*$ u.r.e., so $f_{ren(G)^*}(x) = \frac{1}{1 - g(x)} = f(x)$ and $f(x) \in \mathscr{F}_{\mathscr{R}}$. ∎

2. AN ANALYTIC CHARACTERIZATION OF ℕ-RATIONAL SERIES

In this section, we quote some definitions and well-known properties of ℕ-rational series (for further details, see [6, 19]). Let K be a sub-semi-ring contained in $\mathbb{R}$, and $K\langle\langle x \rangle\rangle$ be the semi-ring of formal power series in one variable over K. A series $f(x) = \sum_{n \geq 0} f_n x^n$ of $K\langle\langle x \rangle\rangle$ is K-rational if $f(x)$ can be obtained from polynomials whose coefficients are in K by means of the rational operations of sum, convolution product, and *quasi-inversion*, defined as

$$\text{if } f(0) = 0, \text{ then } f^*(x) = \frac{1}{1 - f(x)}.$$

A sequence $(f_n) \subset K$ is called K-rational if its generating function $f(x) = \sum_{n \geq 0} f_n x^n$ is K-rational. In this paper, we deal with $K = \mathbb{Z}$ and $K = \mathbb{N}$, and all the definitions and results refer to this case. The following statements are equivalent:

(i) (f_n) is $\mathbb{Z}$-rational;

(ii) the generating function $f(x) = \sum_{n \geq 0} f_n x^n$ is $f(x) = \frac{P(x)}{Q(x)}$, with $P(x), Q(x) \in \mathbb{Z}[x]$, $Q(0) = 1$ (i.e., $f(x) = (a_0 + a_1 x + \cdots + a_l x^l)/(1 - b_1 x - \cdots - b_m x^m)$, $a_i, b_j \in \mathbb{Z}$);

(iii) (f_n) satisfies the linear recurrence relation

$$f_n = b_1 f_{n-1} + \cdots + b_m f_{n-m}, \qquad n \geq \max\{l+1, m\}. \tag{1}$$

Let $f(x) = P(x)/Q(x) = (a_0 + a_1 x + \cdots + a_l x^l)/(1 - b_1 x - \cdots - b_m x^m)$ be a $\mathbb{Z}$-rational function, where $P(x)$ and $Q(x)$ have no common factor. The roots of $Q(x)$ are called the poles of $f(x)$. The only monic polynomial

$$\overline{Q}(x) = x^m - \sum_{i=1}^{m} b_i x^{m-i}$$

is called the *reciprocal polynomial* of $Q(x)$, and its roots are called $f(x)$'s *roots*. A polynomial $Q(x)$ has a *dominating root* if it has a real positive root α such that $\alpha > |\beta_i|$ for any other root $\beta_i \neq \alpha$. A series $f(x)$ has a dominating root if its minimal polynomial does.

A sequence (r_n) is a *merge* of sequences $(s_n^i)_{0 \leq i \leq p-1}$ if, for $0 \leq i \leq p-1$, $r_{np+i} = s_n^i$, $\forall n \geq 0$. We denote the generating functions of (r_n) and (s_n^i) as $f(x)$ and $F_i(x)$.

It can be proved that:

—$f(x) = \sum_{i=0}^{p-1} x^i F_i(x^p)$;

—if $\alpha_1, \ldots, \alpha_n$ are $f(x)$'s roots, then $F_i(x)$'s roots are among $\alpha_1^p, \ldots, \alpha_m^p$.

The following results bring us to an analytic characterization of $\mathbb{N}$-rational series:

THEOREM 2.1 [19, p. 57]. *Let K be the ring $\mathbb{Z}$ or a subfield of $\mathbb{C}$, (r_n) a nonterminating sequence of elements of K having $f(x)$ as generating function. Then the following conditions are equivalent*:

(i) *(r_n) is K-rational*;

(ii) *for large values of n we have*

$$r_n = \sum_i P_i(n)\, \alpha_i^n, \tag{2}$$

where the numbers α_i's are the roots of $f(x)$, the P_i's are algebraic over A, and the degree of each P_i is equal to the multiplicity of the root α_i minus one.

Consider the terms of the sum (2). Clearly the term where the absolute value of α_i is greatest determines the behavior of $f(x)$ for large values of n. For $\mathbb{R}_+$-rational series the α coming from such terms must be obtained by multiplying a positive number by a root of unity. More precisely, we have:

THEOREM 2.2 (Berstel [19, p. 61]). *If* $r(x) \in \mathbb{R}_+\langle\langle x \rangle\rangle$, *then the roots of its generating function having maximal modulus* (*if any*) *are of the form* $\varrho\xi$, *where* $\varrho > 0$, *and* ξ *is a root of unity*; ϱ *is one of these roots*.

By means of this theorem, Berstel and Reutenauer prove in [6] that for a $\mathbb{R}_+$-rational sequence (r_n) we can always get a decomposition showing the dominant term:

THEOREM 2.3. *If* $r(x) \in \mathbb{R}_+\langle\langle x \rangle\rangle$, *then there are positive integers* m *and* p *such that if* $0 \le j \le p-1$, *then*

$$r_{m+j+np} = P_j(n)\,\alpha_j^n + \sum_i P_{ij}(n)\,\alpha_{ij}^n, \tag{3}$$

where, for every i, j, $\alpha_j \ge 0$, *and* $\alpha_j > \max\{|\alpha_{ij}|\}$, *and* P_j, P_{ij} *are polynomials*.

Indeed the decomposition (3) is obtained by choosing a p such that the roots of unity involved become 1. The property characterizing $\mathbb{R}_+$-rational series within the family of $\mathbb{R}$-rational sequences is the existence of a decomposition (3).

THEOREM 2.4 (Soittola [19, p. 64]). *Let* K *be the ring* $\mathbb{Z}$ *or a subfield of* $\mathbb{R}$. *A* K*-rational sequence of positive numbers* (r_n), *such that for large values of* n,

$$r_n = P(n)\,\alpha^n + \sum_i P_i(n)\,\alpha_i^n$$

(*where* $\alpha > \max\{|\alpha_i|\}$, *and* $P \ne 0$, P, P_i *are polynomials*) *is* K_+*-rational.*

The following corollary is mostly used in the algorithms of the next section. To this purpose, we now give a sketch of a proof which relies on the previously cited theorems and on a technical lemma [6, Lemma 2.5, p. 84] which we do not quote for brevity's sake.

COROLLARY 2.1 [6, p. 90]. *A* $\mathbb{Z}$*-rational series* $f(x)$ *having positive coefficients is* $\mathbb{N}$*-rational if and only if it is a merge of* $\mathbb{Z}$*-rational series having a dominating root*.

Proof. ($\Leftarrow$) Let $f(x), F_0(x), \dots, F_{p-1}(x)$ be generating functions of $(r_n), (s_n^0), \dots, (s_n^{p-1})$, respectively. Let (r_n) be merge of $(s_n^0), \dots, (s_n^{p-1})$ and for all $i < p$, let $\alpha_{i1}, \dots, \alpha_{im}$ be $F_i(x)$'s roots with α_{i1} dominating.

Each $F_i(x)$ is $\mathbb{Z}$-rational, so by Theorem 2.1 for large values of n we have $(s_n^i) = P_{i1}(n)\alpha_{i1}^n + \sum_{j=2}^m P_{ij}(n)\alpha_{ij}^n$. By Theorem 2.4, each (s_n^i) is $\mathbb{N}$-rational and consequently $f(x)$ is $\mathbb{N}$-rational too.

($\Rightarrow$) Let $f(x)$ be a $\mathbb{N}$-rational series. By Theorem 2.2, any root of $f(x)$, with maximal modulus has the form $\varrho\xi$ with ξ root of unity and $\varrho > 0$. Let p be a common order of all these roots of unity and let $F_0(x), F_1(x), \dots, F_{p-1}(x)$ be the series whose merge is $f(x)$. By [6, Lemma 2.5], each $F_i(x)$ is $\mathbb{N}$-rational, and thus has a dominating root. ∎

Soittola's Theorem allows us to decide if a $\mathbb{Z}$-rational series is $\mathbb{N}$-rational [21]. It also establishes that any $\mathbb{N}$-rational series is of star-height 2 at most. The series whose star-height is equal to zero are polynomials of $\mathbb{N}[x]$. Bassino [4] proved that if an $\mathbb{N}$-rational series of star-height 1 then it has exclusively Handelmann numbers as real positive roots; if the series has also a dominating root the condition is also sufficient.

3. TWO ALGORITHMS

The following algorithm allows us to establish if a rational function $f(x)$ is $\mathbb{N}$-rational.

ALGORITHM 1.

Input: A $\mathbb{Z}$-rational function $f(x) = P(x)/Q(x) = (a_0 + a_2 x + \cdots + a_l x^l)/(1 - b_1 x - \cdots - b_m x^m)$, where $a_i, b_i \in \mathbb{Z}$, and $l < m$, and $P(x)$ and $Q(x)$ have no common factor. The function is not a polynomial, and we denote its sequence by $(f_n)_{n \geq 0}$.

Output: $f(x)$ is $\mathbb{N}$-rational or $f(x)$ is not $\mathbb{N}$-rational.

1. **Compute** the roots $\alpha_1, \dots, \alpha_k$ of $f(x)$.
2. **Compute** the least p so that the generating functions $F_i(x)$ of the subsequences $(f_{np+i})_{n \geq 0}$, $0 \leq i \leq p - 1$ have no roots of the form $\varrho\xi$, where $\varrho > 0$, and ξ is a root of unity different from one.
3. **Check** if $F_i(x)$ has a unique root of maximal modulus, for $i = 0, \dots, p - 1$.
4. **Check** if the coefficients of $F_i(x)$ are all positive, for $i = 0 \dots p - 1$.
5. If **3.** and **4.** hold, each $F_i(x)$ is $\mathbb{N}$-rational, and $f(x)$ is $\mathbb{N}$-rational; otherwise $f(x)$ is not $\mathbb{N}$-rational.

We now propose some methods for performing the steps indicated in the algorithm.

1. We consider a root-finding algorithm based on numerical approximation and which works in polynomial time. The following statement holds [15]: *If $G \in K[x]$ is a monic polynomial, $n = \deg G > 0$, $|G| = \Sigma_i |a_i|$, then all G's zeros can be computed with absolute error $\epsilon > 0$ using $O(n^2 \log n(n$*

$\log n) + \log \frac{|G|}{\epsilon})$ *arithmetical operations*. Moreover, in [12] a proof is given of the fact that two different roots of a polynomial having integer coefficients cannot be too close; that is, for such roots $|\alpha_i| \neq |\alpha_j|$, implies $||\alpha_i| - |\alpha_j|| \geq k(G)$, where $k(G)$ entirely depends on G. A precise expression for $k(G)$ can be found in [12, p. 156].

2. For this we need the following sub-steps:

(1) we find the $\varrho\xi$ roots of $f(x)$, with $\varrho > 0$ and ξ root of unity, and we denote them as $\alpha_{i_1}, \dots, \alpha_{i_h}$, among the roots $\alpha_1, \dots, \alpha_k$ of $f(x)$:

(2) we find the smallest integer p for which each $\alpha_{i_j}^p$ is not of the form $\varrho\xi$, with $\varrho > 0$, and ξ a root of unity different from one;

(3) we go on to examine separately the series $F_i(x) = \sum_n f_{np+i}x^n$, $0 \leq i \leq p-1$, whose roots are among $\alpha_1^p, \dots, \alpha_k^p$.

Sub-step (1) can be carried on by taking the roots $\alpha_1, \dots, \alpha_k$ of $f(x)$ and constructing the symmetrical polynomial having integer coefficients,

$$R(x) = \prod_{i,j} (\alpha_i x - \alpha_j).$$

Let $n_Q < \deg R$, such that, for each $n > n_Q$, we have $\varphi(n) > \deg R$ (where $\varphi(n)$ is the Euler function and indicates the degree of a primitive root of unity); we then divide R by the first n_Q cyclotomic polynomials and find the roots of unity of $R(x)$.

The generating function $F_i(x)$ of the sequence $(f_{np+i})_n$ can be calculated by means of the k-section formulas [18]: if s is a primitive p-root of unity, then

$$F_i(x) = \sum_{n \geq 0} f_{np+i}x^n = \frac{1}{px^{1/p}} \sum_{j=1}^{p} s^{p-ij} f(s^j, x^{1/p}), \qquad i = 0, \dots, p-1.$$

Now, $f(x)$ is a merge of the functions $F_i(x)$.

EXAMPLE 3.1. Apply step 2 to the function $f(x) = 1/(1 - x + x^2)(1 - 2x^2)(1 - 5x)$.

(1) The roots of $(1 - 2x^2)(1 - 5x)(1 - x + x^2)$ are $\alpha_1 = \sqrt{2}$, $\alpha_2 = -\sqrt{2}$, $\alpha_3 = 5$, $\alpha_4 = e^{(\pi/3)i}$, $\alpha_5 = e^{(2\pi/3)i}$.

(2) α_1, α_3 have the form desired for $p = 1$, while $(\alpha_4)^6 = (\alpha_5)^3 = 1$ and $(\alpha_2)^2 = 2$; so we take $p = 6$.

(3) It now suffices to study separately the sequences $(f_{6n+i})_n$, $i = 0, \ldots, 5$, whose generating functions are

$$F_0(x) = \sum_{n\geq 0} f_{6n}x^n = \frac{1 + 4583x - 15000x^2}{(1-8x)(1-x)(1-5^6x)},$$

$$F_1(x) = \sum_{n\geq 0} f_{6n+1}x^n = \frac{6 + 7286x - 12500x^2}{(1-8x)(1-x)(1-5^6x)},$$

$$F_2(x) = \sum_{n\geq 0} f_{6n+2}x^n = \frac{32 + 5176x}{(1-8x)(1-x)(1-5^6x)},$$

$$F_3(x) = \sum_{n\geq 0} f_{6n+3}x^n = \frac{161 + 10255x}{(1-8x)(1-x)(1-5^6x)},$$

$$F_4(x) = \sum_{n\geq 0} f_{6n+4}x^n = \frac{808 + 4400x}{(1-8x)(1-x)(1-5^6x)},$$

$$F_5(x) = \sum_{n\geq 0} f_{6n+5}x^n = \frac{4042 - 9250x}{(1-8x)(1-x)(1-5^6x)}.$$

The roots of the functions $F_0(x), \ldots, F_5(x)$ now have the desired form.

3. For simplicity's sake, let us assume we are working with the generic series $F(x) = \sum_{n\geq 0} F_n x^n = (a'_0 + a'_1 x + \cdots + a'_l x^l)/(1 - b'_1 x - \cdots - b'_m x^m)$, having roots of the desired form. Now the roots of $F(x)$ are among $\alpha_1^p, \ldots, \alpha_k^p$. Then a unique root of maximal modulus must be real and positive. Now Soittola's Theorem ensures that $F(x)$ is $\mathbb{N}$-rational if and only if the following two conditions hold:

(1) $F(x)$ has a positive dominating root;

(2) all $F(x)$ coefficients are non-negative.

4. We assume $F(x)$ has a positive dominating root $\alpha > 0$, $\alpha > \max\{|\alpha_i|\}$. It is therefore possible to establish if the coefficients F_n are all nonnegative. For n sufficiently large, we have

$$F_n = P(n)\alpha^n + \sum_i P_i(n)\alpha_i^n.$$

Let $u = \deg P - 1$ and $u_i = \deg P_i - 1$, for all i, $u_0 = \max\{u_i\}$, and a, b be rational numbers such that $\alpha > a > b > |\alpha_i|$ for all i. By standard techniques we can find a positive rational number A_0 smaller than the leading coefficient A of P and a rational number $C \geq \max\{$moduli of coefficients of $P_i\}$. Since $F_n \to A_0 n^u \alpha^n$ as $n \to \infty$, it follows that

$$\chi(n) = \frac{|\sum_i P_i(n)\alpha_i^n|}{A_0 n^u \alpha^n} \to 0;$$

that is, for all $\epsilon > 0$, N exists such that for all $n > N$, $\chi(n) < \epsilon$. Since $iu_0 C n^{u_0} b^n / A_0 n^u a_n > \chi(n)$,

$$\frac{iu_0 C n^{u_0} b^n}{A_0 n^u a^n} < \epsilon \to \chi(n) < \epsilon.$$

We thus get a inequality $\delta n^{\rho} \sigma^n < \epsilon$, where $\sigma < 1$. By solving this inequality for $\epsilon = 1$, we find an index n_+ such that if $F_n \geq 0$ for all $n \leq n_+$, then $F_n \geq 0$ for all $n \in \mathbb{N}$. The first n_+ coefficients can be computed by using the recurrence relation (1).

EXAMPLE 3.2. The rational function $f(x) = 1/(1 - 3x + 5x^2 - 8x^3)$ is $\mathbb{N}$-*rational*. Indeed, $x^3 - 3x^2 + 5x - 8$ has three roots, $\alpha_1 \approx 2,32$, $\alpha_2 \approx 0,33 + 1,82i$ and $\alpha_3 \approx 0,33 - 1,82i$, and α_1 is the dominating root. By some simple calculations, we find that $n_+ = 3$ works: since $f_0 = 1$, $f_1 = 3$, $f_2 = 4$, $f_3 = 5$, we can conclude that $f_n > 0$ for all n.

The following theorem states that the generating functions of regular languages are the $\mathbb{N}$-rational functions whose first coefficient is 0 or 1. Therefore, we have a criterion for establishing if there is a u.r.e. corresponding to a $\mathbb{Z}$-rational series.

THEOREM 3.1. *$f(x) \in \mathscr{R}$ if and only if $f(x)$ is $\mathbb{N}$-rational and $f(0) \in \{0, 1\}$.*

Proof. The necessary condition is obvious. To show the other direction it suffices to use induction on the $\mathbb{N}$-rational function $f(x)$, having $f(0) \in \{0, 1\}$. ∎

Let us now assume that $f(x) \in \mathscr{R}$. Our task is to determine an algorithm able to provide an unambiguous regular expression E of star-height at most 2, having $f(x)$ as generating function. Algorithm 2 relies on the original proof of Soittola's Theorem (Theorem 2.4) and the proof given by Perrin in [17], which uses positive matrices.

DEFINITION 3.1. Let $f(x)$ be a $\mathbb{N}$-rational function, $f(x) = P(x)/Q(x) = (a_0 + a_1 x + \cdots + a_l x^l)/(1 - b_1 x - \cdots - b_m x^m)$ having roots $\alpha_1 > |\alpha_2| \geq \cdots \geq |\alpha_m|$; $f(x)$ satisfies the K-condition if there is a positive integer k such that $\alpha_i < k < \alpha_1$, and

$$\begin{aligned}
G_1 &= b_1 - k \geq 0 \\
G_2 &= b_2 + b_1 k - k^2 \geq 0 \\
G_3 &= b_3 + b_2 k + b_1 k^2 - k^3 \geq 0 \\
&\ \vdots \\
G_m &= b_m + b_{m-1} k + \cdots + b_1 k^{m-1} - k^m \geq 0.
\end{aligned}$$

The last request is equivalent to asking that $(1 - (1 - b_1 x - \cdots - b_m x^m)/(1 - kx)) \in \mathscr{R}$.

ALGORITHM 2.

Input: A $\mathbb{N}$-rational function $f(x) = P(x)/Q(x) = (a_0 + a_1 x + \cdots + a_l x^l)/(1 - b_1 x - \cdots - b_m x^m)$ (by Algorithm 1, with no loss of generality, we assume that $f(x)$ has no roots of the form $\rho\xi$, with $\rho > 0$ and ξ a root of unity different from one), such that $a_0 \in \{0, 1\}$. We denote its sequence by (f_n).

Output: An unambiguous regular expression of star-height at most 2 and having generating function $f(x)$.

1. **Compute** the smallest integer p such that all the generating functions $F_i(x)$ of the subsequences $(f_{np+i})_{n>0}$, with $0 \le i \le p - 1$, satisfy the K-condition for a certain k.
2. **Find** an expression for each $F_i(x)$ involving rational operators $+, \cdot, *$ and polynomials in $\mathbb{N}[x]$, by means of k, determined in step **1**.
3. **Construct** the unambiguous regular expression E having generating function $f(x)$.

1. The proof of the existence of such a p is given in Soittola's Theorem and it can be used to determine the final expression of $f(x)$ in terms of rational operations (step **2**). For every i, let $F_i(x) = P_i(x)/Q_i(x)$ and $Q_i(x) = (1 - G_1 x - \cdots - G_{m_i - 1} x^{m_i - 1})(1 - kx) - G_m, x^{m_i}$, where the G_i are defined in Definition 3.1, and let $\alpha_{i1}^p, \ldots, \alpha_{ik_i}^p$ be $F_i(x)$'s roots with α_{i1}^p dominating. We find an upper bound for p such that each $F_i(x)$ satisfies the K-condition:

—From

$$Q_i(x) = 1 - b_{i,1} x - \cdots - b_{i,m_i} x^{m_i},$$

$$b_{i,j} = \sum_{t_1 < \cdots < t_j \le m_i} (-1)^{j+1} \alpha_{t_1}^p \cdot \cdots \cdot \alpha_{t_j}^p,$$

$$Q_i(x) = \left(1 - G_1 x - \cdots - G_{m_i - 1} x^{m_i - 1}\right)(1 - kx) - G_{m_i} x^{m_i},$$

and after defining $u_j = \sum_{1 < t_1 < \cdots < t_j \le m_i} \alpha_{t_1}^p \cdot \cdots \cdot \alpha_{t_j}^p$, we obtain

$$G_j = \alpha_1^p\left(k^{j-1} - k^{j-2} u_1 + \cdots + (-1)^j u_{j-1}\right) - k^j + k^{j-1} u_1 - \cdots + (-1)^{j+1} u_j. \tag{4}$$

Equation (4) provides the same expression for G_j as in [17, p. 360].

—Let $|\overline{\alpha}| = \max_{1<i\le k_i}\{|\alpha_i|\}$.

$$\begin{aligned}\frac{G_j}{k^j} &\ge \frac{\alpha_1^p}{k}\left(1 - m_i\frac{|\overline{\alpha}^p|}{k} - m_i^2\frac{|\overline{\alpha}^p|^2}{k^2} - \cdots\right)\\ &\quad - 1 - m_i\frac{|\overline{\alpha}^p|}{k} - \cdots - m_i^j\frac{|\overline{\alpha}^p|^j}{k^j}\\ &\ge \frac{\alpha_1^p}{k}\left(1 - \sum_{n=1}^{\infty}\left(m_i\frac{|\overline{\alpha}^p|}{k}\right)^n\right) - 1 - \sum_{n=1}^{\infty}\left(m_i\frac{|\overline{\alpha}^p|}{k}\right)^n. \qquad (5)\end{aligned}$$

In order that G_j be non-negative, it is sufficient to impose $(5) > 0$: $\sum_{n=1}^{\infty}(m_i(|\overline{\alpha}^p|/k))^n$ needs to be a small enough convergent series (less than 1). For this we set, for example, $m_i(|\overline{\alpha}^p|/k) \le \frac{1}{6}$, obtaining $a \ge 6m_i|\overline{\alpha}^p|$ and $(5) \ge (\alpha_1^p/k)(1-\frac{1}{5}) - 1 - \frac{1}{5} \ge 0$.

Finally, solving the last inequality with $k \ge 6m_i(|\overline{\alpha}^p| + 1)$, we get $p > \ln(9m_i)/\ln((\alpha_1+1)/|\overline{\alpha}|)$.

2. Let us consider a generic $F(x) = \frac{T(x)}{O(x)} = \sum_{n\ge 0} A_n x^n$ satisfying the K-condition for $k \in \mathbb{N}$; then there are $P_0, P_1, P_2, Q_1, Q_2 \in \mathbb{N}[x]$, such that

$$F(x) = P_0(x) + \frac{P_1(x) + P_2(x)/(1-kx)}{1 - (Q_1(x) + Q_2(x)/(1-kx))}.$$

Soittola's Theorem ensures that $F(x)$ can be written as

$$\begin{aligned}F(x) &= A_0 + A_1x + \cdots + A_{u-1}x^{u-1} + \frac{A_u x^u}{1-kx}\\ &\quad + \frac{1}{1-kx}\left(\frac{N_1(x) + N_2(x)/(1-kx)}{1 - \left(G_1x + G_2x^2 + \cdots + G_m x^m/(1-kx)\right)}\right),\end{aligned}$$

where $u \ge \max\{\deg O, \deg T + 1\}$ and $N_1, N_2 \in \mathbb{N}[x]$. $A_0, \ldots, A_u u$ can be easily computed by means of the well-known recurrence relation (1); N_1 and N_2 can be determined by some simple operations on polynomials. We set aside the first u terms of the sequence in order to eliminate the initial conditions.

3. Since $f(x)$ is a merge of $F_0(x), \ldots, F_{-1}(x)$, by applying step **2** we are able to write $f(x)$ in terms of some polynomials in $\mathbb{N}[x]$ and rational operations. It should be also remarked that this way of writing the function $f(x)$ is not unique. We now define a procedure Exp which works recur-

sively on $f(x)$ (written as above), and provides a regular expression on the alphabet $\Sigma = \{a\}$,

$$\mathsf{Exp}(x) = a, \qquad \mathsf{Exp}(1) = \lambda, \qquad \mathsf{Exp}(0) = \emptyset$$

$$\mathsf{Exp}(f(x) + g(x)) = \mathsf{Exp}(f(x)) \vee \mathsf{Exp}(g(x))$$

$$\mathsf{Exp}(f(x) \cdot g(x)) = \mathsf{Exp}(f(x)) \cdot \mathsf{Exp}(g(x))$$

$$\mathsf{Exp}\left(\frac{1}{1 - f(x)}\right) = (\mathsf{Exp}(f(x)))^*.$$

$E = ren(\mathsf{Exp}(f(x)))$ is an unambiguous regular expression whose generating function is $f(x)$. ∎

EXAMPLE 3.3. Find a u.r.e. for $f(x) = (1 - 2x + 3x^2 - 3x^3)/(1 - 3x + 2x^2)$. The roots of $1 - 3x + 2x^2 = 0$ are $\alpha_1 = 2$ and $\alpha_2 = 1$. $f(x)$ satisfies the K-condition for $k = 1$. By applying Algorithm 2 step by step, we get

$$f(x) = 1 + x + 4x^2 + 7x^3 + \frac{13x^4}{1 - x} + \frac{12x^5}{(1 - x)(1 - 2x)}$$

$$= 1 + x + \frac{4x^2}{1 - x} + \frac{3x^3}{(1 - x)(1 - 2x)}.$$

$$\mathsf{Exp}(f(x)) = \lambda \vee a \vee [aa \vee aa \vee aa \vee aa]a^*$$
$$\vee [aaa \vee aaa \vee aaa]a^*[a \vee a]^*.$$

After applying the *ren* procedure and minimalizing the alphabet Σ, we get the u.r.e.,

$$E = \lambda \vee a \vee [bb \vee cc \vee dd \vee ee]a^* \vee [bbb \vee ccc \vee ddd]e^*[a \vee b]^*$$

with the alphabet $\Sigma = \{a, b, c, d, e\}$.

Algorithm 2 may often not work so quickly. For instance, consider the following cases:

(i) $f(x) = \frac{P(x)}{1 - Q(x)}$ and $P(x), Q(x) \in \mathbb{N}[x]$, where $f(x) = P(x)[Q(x)]^*$;

(ii) $f(x) = P(x)/\prod_i(1 - \alpha_i x)$ and $\alpha_i \in \mathbb{N}$, where $f(x)$ has star-height 1.

Now we examine the case $\deg Q = 2$, to which Algorithms 1 and 2 can be very easily applied.

COROLLARY 3.1. *Let* $f(x) = \frac{P(x)}{Q(x)}$ *be* $\mathbb{Z}$*-rational and* $\deg Q = 2$. *Then* $f(x) \in \mathscr{R}$ *if and only if* $f_n \geq 0$ *for all* $n \in \mathbb{N}$.

Proof. The sufficient condition follows from the fact that if a $\mathbb{Z}$-rational series has non-negative coefficients, then it must have a real pole of minimum modulus. Given $\deg Q = 2$, the two roots of Q must be real numbers. ∎

COROLLARY 3.2. *Let $f(x) = P(x)/(1 - ax - bx^2) \in \mathscr{R}$, and let α_1, α_2 be the real roots of $f(x)$ (assume $\alpha_1 \leq \alpha_2$). We distinguish the following cases*:

(1) *if $a = 0$, then b must be positive and $f(x)$ can be easily expressed in terms of rational operations*;

(2) *if $a > 0$, then*

(2.1) *if $b \geq 0$ or $\Delta = 0$, then $f(x)$ has star height* 1;

(2.2) *if $b < 0$, then $f(x)$ satisfies the K-condition for all integers $k \in \mathbb{N}$, $\alpha_1 \leq k \leq \alpha_2$; $f(x)$ has star-height* 2 *if and only if its roots are not integers*;

(3) *if $a < 0$, then $f(x) \notin \mathscr{R}$.*

Proof. Parts (1), (2.1), and (3) are obvious since $f_n \geq 0$ for all $n \in \mathbb{N}$. In case (2.2) we have $f(x) = P(x)/(1 - ax + bx^2)$ with $a, b > 0$. Therefore $f(x)$ has two distinct positive real roots (because $a = \alpha_1 + \alpha_2$ and $b = \alpha_1\alpha_2$); $\alpha_2 - \alpha_1 = \Delta \geq 1$, so there is at least one positive integer k such that $\alpha_1 \leq k \leq \alpha_2$. We prove that $f(x)$ satisfies the K-condition for k. That is,

$$\begin{cases} a - k \geq 0 \\ ak - b - k^2 \geq 0. \end{cases}$$

Clearly $a \geq k$. The second condition is equivalent to $(ak - k^2)/b \geq 1$. Let O_1, O_2 be positive numbers, such that $\alpha_1 = k - O_1$, $\alpha_2 = k + O_2$. Then

$$\frac{ak - k^2}{b} = 1 + \frac{O_1 O_2}{b} > 1.$$

If the roots are not integers, then $f(x)$ has star-height 2, because it has a real dominating root and its roots α_1, α_2 are not Handelmann numbers [4].

EXAMPLE 3.4. Let $f(x) = 1/(1 - 6x + 6x^2) \in \mathscr{R}$. The roots of $x^2 - 6x + 6$ are $\alpha_1 = 3 - \sqrt{3}$ and $\alpha_2 = 3 + \sqrt{3}$; the integers between the two roots are 2, 3, 4, so we can use $k = 2, 3, 4$. Corollary 3.2 establishes that $f(x)$ can be written as

$$f(x) = \frac{1/(1 - 2x)}{1 - (3x + x/(1 - 2x))} = \frac{1/(1 - 3x)}{1 - (2x + x/(1 - 3x))}$$

$$= \frac{1/(1 - 4x)}{1 - \left(2x + 2x^2/(1 - 4x)\right)}.$$

Finally, we take a harder example into consideration.

EXAMPLE 3.5. Let $f(x) = 1/(1 - 3x + 5x^2 - 8x^3)$. It is easy to verify that, for $p = 1$ and $p = 2$, there is no k satisfying the K-condition; for $p = 3$, the generating functions of the sequences $f_0(x), f_1(x), f_2(x)$ are

$$f_0(x) = \sum_n f_{3n}x^n = \frac{1 - x + 64x^2}{1 - 6x - 43x^2 - 512x^3},$$

$$f_1(x) = \frac{3 + x}{1 - 6x - 43x^2 - 512x^3}, \qquad f_2(x) = \frac{4 + x}{1 - 6x - 43x^2 - 512x^3}.$$

In order to establish that there is a $k \in \mathbb{N}$ satisfying the K-condition, we only have to examine $f_0(x)$ ($f_1(x)$ and $f_2(x)$ have the same denominator), and find out that $f(x)$ has star-height equal to 1,

$$\begin{aligned} f(x) &= f_0(x^3) + xf(x^3) + x^2f(x^3) \\ &= 1 + \frac{3x + 4x^2 + 5x^3 + x^4 + 40x^5 + 107x^6 + 512x^9}{1 - (6x^3 + 43x^6 + 512x^9)}. \end{aligned}$$

4. THE APPLICATION OF THIS METHOD TO THE ENUMERATION OF RECURRENCES

By applying Algorithm 2, we answer some questions regarding the enumeration of combinatorial objects.

4.1. *Regular Languages and Recurrence Relations*

It is common knowledge that every regular language is enumerated by a linear recurrence relation whose first term is 0 or 1; thanks to our algorithms we are now able to invert this statement when it is possible. This means that we can solve the problem of giving a combinatorial interpretation of a linear recurrence relation with integer coefficients, in almost all cases. This can be done as follows.

Let (f_n) be a sequence of positive integer numbers ($f_0 = 0, 1$) described by a linear recurrence;

(1) determine the generating function $f(x) = \sum_{n \geq 0} f_n x^n$ of the sequence by standard methods;

(2) by means of Algorithm 1, state if $f(x)$ is $\mathbb{N}$-rational;

(3) if $f(x)$ is $\mathbb{N}$-rational, apply Algorithm 2 to $f(x)$ and find a regular language enumerated by (f_n) (a combinatorial interpretation); otherwise, there is no regular language enumerated by the sequence (f_n).

EXAMPLE 4.1. Let $(f_n) = 0, 1, 2, 4, 9, 21, 49, \dots$ be the sequence described by the recurrence relation

$$\begin{cases} f_0 = 0, \quad f_1 = 1, \quad f_2 = 2 \\ f_n = 3f_{n-1} - 2f_{n-2} + f_{n-3}. \end{cases}$$

The generating function of the sequence (f_n) is $f(x) = (x - x^2)/(1 - 3x + 2x^2 - x^3)$. Algorithm 1 establishes that $f(x)$ is $\mathbb{N}$-rational. By means of Algorithm 2, we find an unambiguous regular expression enumerated by (f_n). We write $f(x)$ as $x \cdot (1/(1 - (2x + x^3/(1 - x))))$ and build E on the alphabet $\Sigma = \{a, b, c\}$: $E = a(a \vee b \vee cacc^*)^*$.

The following example shows how our reverse technology can be applied to the already mentioned problem raised by Propp.

EXAMPLE 4.2. In October 1999, Jim Propp posted on the "domino" mailing list, domino@math.wisc.edu, the mail "1, 1, 1, 2, 3, 7, 11, 26, ... *and reverse-engineered combinatorics*," with the request for a combinatorial interpretation of the sequence of positive numbers defined by the recurrence relation

$$\begin{cases} f_0 = 1, \quad f_1 = 1, \quad f_2 = 1, \quad f_3 = 2 \\ f_n = 4f_{n-2} - f_{n-4}. \end{cases}$$

In particular he asked: "Is there a technology for reverse-engineering a combinatorial problem from a rational generating function?"

Our way to approach the problem easily answers these two questions. Indeed we calculate the generating function for (f_n),

$$f(x) = \frac{1 + x - 3x^2 - 2x^3}{1 - 4x^2 + x^4}.$$

By means of our method we easily see that $f(x) \in \mathscr{R}$, so there is a rational language having $f(x)$ as its generating function. Moreover we can write $f(x)$ in terms of rational operations:

$$f(x) = \frac{1 + x\left(1 + x^2/(1 - 3x^2)\right)}{1 - \left[x^2\left(1 + 2x^2/(1 - 3x^2)\right)\right]}.$$

Then we can construct a regular expression E enumerated by the given sequence,

$$E = (\epsilon \vee a[\epsilon \vee aa[bb \vee cc \vee dd]^*])(cc(\epsilon \vee bb \vee cc)(ee \vee aa \vee ff)^*)^*$$

on the alphabet $\Sigma = \{a, b, c, d, e, f\}$.

EXAMPLE 4.3. On the bisection of Fibonacci numbers, let us take into consideration the well-known sequence of integer numbers $1, 2, 5, 13, 34, \ldots,$ defined by the recurrence relation,

$$\begin{cases} f_0 = 1, & f_1 = 2 \\ f_n = 3f_{n-1} - f_{n-2}, \end{cases}$$

having $f(x) = (1 - x)/(1 - 3x + x^2)$ as its generating function. The application of the previous methods enables us to write $f(x)$ in terms of rational operations and polynomials in $\mathbb{N}[x]$:

$$f(x) = \frac{1}{1 - (x + x/(1 - x))}.$$

The corresponding u.r.e. is $E = (a \vee bc^*)^*$ on the alphabet $\Sigma = \{a, b, c\}$. We also point out that each language having $f(x)$ as its generating function must have at least three nonterminal symbols, since $f_2 = 5 > f_1^2$. Now we take into consideration two classes of objects which are proved to be enumerated by the sequence f_n, and in both cases we find an explicit bijection with the words of the language $L(E)$, where $L(E)$ is the regular language represented by E.

Directed Column-Convex Polyominoes. A polyomino is a finite-connected union of cells in $\mathbb{Z} \times \mathbb{Z}$ having no cut point; polyominoes are defined up to translations. In the general setting of polyominoes, the qualification of *directed column-convex* (briefly, *dcc*) refers to the following characteristics:

(i) they are directed in the sense that they can be built by starting with a single cell (the origin) and then by adding new cells on the right or on the top of an existing cell;

(ii) every column must be formed by contiguous cells.

The *area* of a dcc-polyomino is the number of its cells. Figure 1 shows the 5 dcc-polyominoes having area 3. In several papers it is proved that the number of dcc-polyominoes having area $n + 1$ is f_n, with $n \geq 0$ (for example, see [3]). We prove this statement by establishing a bijection

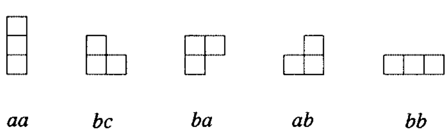

FIG. 1. The 5 dcc-polyominoes having area 2 and the corresponding words of $L(E)$.

between the dcc-polyominoes of area $n + 1$ and the words of length n of $L(E)$. To do this, we observe that each dcc-polyomino of area $n + 2$ can be obtained in a unique way by a dcc-polyomino having area $n + 1$ by means of the following recursive construction:

(1) if the first two columns of the polyomino start at the same level, we add a cell under the first column, a cell next to the first column, and a cell above the first column (see Fig. 2a);

(2) otherwise, we add a cell under the first column and a cell next to the first column (see Fig. 2b).

This construction can be translated into a construction on the words of $L(E)$. Let w be a word having length $n - 1$; a word of length n can be obtained:

—by adding a final a, b, or c to w, if w ends by b or c (this operation corresponds to the construction of step (1) on *dcc*-polyominoes);

—by adding a final a, or b to w, if w ends by a (step (2)).

$\mathscr{D}^{(2)}$ *Polyominoes*. The class of $\mathscr{D}^{(2)}$ polyominoes was studied in [2]. We denote by $l = (l_1, \dots, l_w)$ and $u = (u_1, \dots, u_w)$ the two vectors whose elements l_i and u_i are the level of the lowest and uppermost cells in the ith column, respectively (see Fig. 3). Let $\mathscr{D}^{(2)}$ be the class of parallelogram polyominoes satisfying the conditions,

$$u_i \leq l_{i+2}, \qquad 1 \leq i \leq w - 1.$$

Thus, each element of $\mathscr{D}^{(2)}$ has no more than two columns of height greater than or equal to 2 starting at the same level.

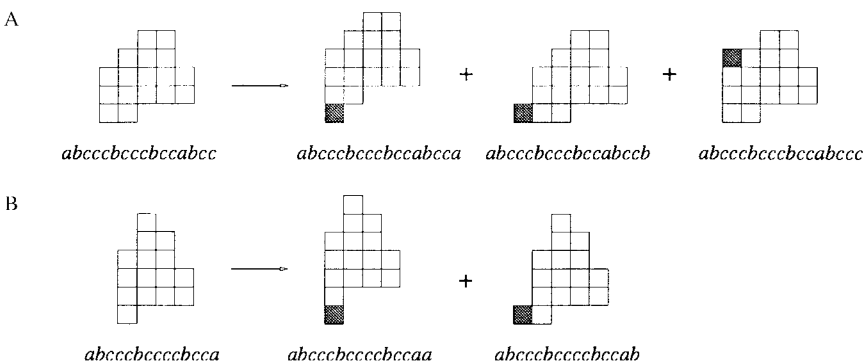

FIG. 2. The construction for *dcc*-polyominoes and the corresponding words of $L(E)$.

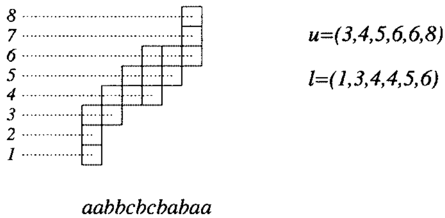

FIG. 3. A $\mathscr{D}^{(2)}$ parallelogram polyomino.

The number of $\mathscr{D}^{(2)}$ polyominoes having semi-perimeter $n + 2$ is f_n. Figure 4 shows that 5-polyominoes of $\mathscr{D}^{(2)}$ having semiperimeter equal to 4. Each $\mathscr{D}^{(2)}$ polyomino of semi-perimeter $n + 3$ can be obtained in a unique way by a $\mathscr{D}^{(2)}$ polyomino having semi-perimeter $n + 2$ by means of the following recursive construction:

(1) if the first two columns of the polyomino start at the same level, we add a cell under the first column, a cell next to the first column, and a row made up of two cells under the first row (see Fig. 5a); this operation corresponds to adding a final a, b, or c to the word w, ending by b or c, corresponding to the polyomino;

(2) otherwise, we add a cell under the first column and a cell next to the first column (see Fig. 5b); this operation corresponds to adding a final a or b to the corresponding word w, ending by a.

5. OPEN PROBLEMS

There are several open problems related to our initial purpose of providing a combinatorial interpretation, in terms of languages, of a given recurrence relation:

(1) *Determine the minimal number of alphabetical symbols for a language enumerated by a rational function* $f(x) \in \mathscr{R}$. It would be useful to improve Algorithm 2 in such a way that, given a rational function $f(x)$, the constructed unambiguous regular expression uses the least number of

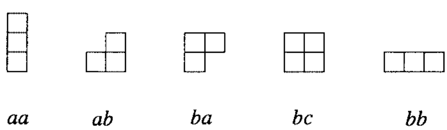

FIG. 4. $\mathscr{D}^{(2)}$ polyominoes having semi-perimeter 4.

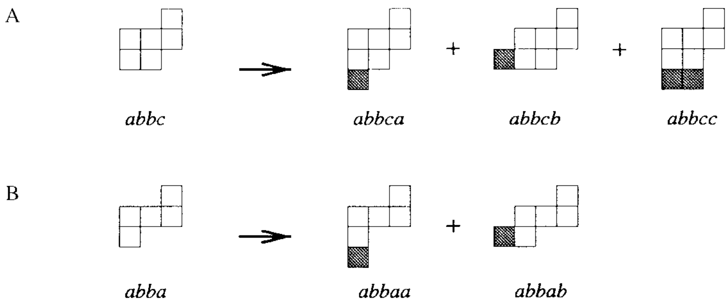

FIG. 5. The construction for $\mathscr{D}^{(2)}$ polyominoes and the corresponding words of $L(E)$.

terminal symbols as possible. To the authors' knowledge this can be done only in a few cases.

(2) *Linear recurrences that do not count any regular language*. Eilenberg [11] gives an example of a $\mathbb{Z}$-rational series with non-negative coefficients, which is not $\mathbb{N}$-rational. Let $f_n = c^{2n}\cos^2(n\theta)$, where $\cos\theta = \frac{a}{c}$, and $0 < a < c$, $a, c \in \mathbb{N}$ with $\frac{a}{c} \neq \frac{1}{2}$. Consider $f(x) = \sum_{n\geq 0} f_n x^n$. By using Euler's functions, we get

$$f(x) = \frac{1}{4}\left(\frac{1}{1 - xc^2e^{2i\theta}}\right) + \frac{1}{4}\left(\frac{1}{1 - xc^2e^{-2i\theta}}\right) + \frac{1}{2}\left(\frac{1}{1 - xc^2}\right)$$
$$= \frac{1 + (c^2 - 3a^2)x + a^2c^2x^2}{1 - (4a^2 - c^2)x + c^2(4a^2 - c^2)x^2 - c^6x^3}. \tag{6}$$

It follows that the poles of $f(x)$ are $(1/c^2)e^{2i\theta}$, $(1/c^2)e^{-2i\theta}$, $1/c^2$, and $f_n \geq 0$ for all n. The $\mathbb{Z}$-rationality of $f(x)$ is guaranteed. Berstel's Theorem states that if $e^{2i\theta}$ is not a root of unity, then $f(x)$ is not $\mathbb{N}$-rational; this is so whenever $c \neq 2a$. For example, if $\frac{a}{c} = \frac{2}{3}$ in (6), then we get

$$f(x) = \frac{1 - 3x + 36x^2}{1 - 7x + 63x^2 - 729x^3}. \tag{7}$$

The linear recurrence associated to (7) is not enumerated by any regular language. Moreover, it can be proved that no unambiguous context-free language can be enumerated by this recurrence. On the other hand, it is still unknown if a language (or a class of combinatorial objects) exists which can be enumerated by these recurrences.

(3) *The inversion of Schützenberger methodology for some larger classes of algebraic functions*. This means to provide a combinatorial interpretation of some polynomial recurrence (P-recurrences). Are there classes for which this can be easily done? On the other side, are there classes of algebraic functions for which the inversion problem is undecidable?

ACKNOWLEDGMENT

The authors thank C. Reutenauer for helpful suggestions.

REFERENCES

1. E. Barcucci, S. Brunetti, A. Del Lungo, and F. Del Ristoro, A combinatorial interpretation of the recurrence $f_{n+1} = 6f_n - f_{n-1}$, *Discrete Math*. **190** (1998), 235–240.
2. E. Barcucci, S. Brunetti, and F. Del Ristoro, Succession rules and Deco polyominoes, *Theoret. Inform. Appl*. **34** (2000), 1–14.
3. E. Barcucci, R. Pinzani, and R. Sprugnoli, Directed column-convex polyominoes by recurrence relations, *in* "TAPSOFT'93" (M.-C. Gaudel and J.-P. Jouannaud, Eds.), LLNCS, Vol. 668, pp. 282–298, Springer-Verlag, New York/Berlin, 1993.
4. F. Bassino, Star-height of an $\mathbb{N}$-rational series, *in* "STACS 96" (C. Puech and R. Reischuk, Eds.), LLNCS, Vol. 1046, pp. 125–135, Springer-Verlag, New York/Berlin, 1996.
5. J. Berstel and D. Perrin, "Theory of Codes," Academic Press, San Diego, 1985.
6. J. Berstel and C. Reutenauer, "Rational Series and Their Languages," Springer-Verlag, Berlin, 1988.
7. N. Chomsky and M. P. Schützenberger, The algebraic theory of context-free languages, *in* "Computer Programming and Formal Systems" (P. Braffort and D. Hirschberg, Eds.), pp. 118–161, North-Holland, Amsterdam, 1963.
8. R. Cori and B. Vauquelin, Planar maps are well labeled trees, *Canad. J. Math*. **33** (1981), 1023–1042.
9. M. P. Delest and G. Viennot, Algebraic languages and polyominoes enumeration, *Theoret. Comput. Sci*. **34** (1984), 169–206.
10. A. van der Poorten, Some facts should better be known; especially about rational functions, *in* "Number Theory and Applications" (R. A. Mollin, Ed.), pp. 497–528, Kluwer Academic, Dordrecht, 1989.
11. S. Eilenberg, "Automa, Languages, and Machines," Vol. A, Academic Press, New York, 1974.
12. X. Gourdon and B. Salvy, Effective asymptotics of linear recurrences with rational coefficients, *Discrete Math*. **153** (1996), 145–163.
13. J. E. Hopcroft and J. D. Ullman, "Introduction to Automata Theory, Languages, and Computation," Addison–Wesley, Reading, MA, 1979.
14. D. Krob, "Eléments de combinatoire," Magistère 1-ère année, Ecole Normale Superieure, 1995.
15. V. Pan, Algebraic complexity of computing polynomial zeroes, *Comput. Math. Appl*. (1987), 285–304.

16. E. Pergola and R. Pinzani, A combinatorial interpretation of the area of Schröder paths, *Electron. J. Combin.* **6** (1999).
17. D. Perrin, On positive matrices, *Theoret. Comput. Sci.* **94** (1992), 357–366.
18. J. Riordan, "Combinatorial Identities," Wiley, New York, 1958.
19. A. Salomaa and M. Soittola, "Automata-Theoretic Aspects of Formal Power Series," Springer-Verlag, New York, 1978.
20. N. J. A. Sloane and S. Plouffe, "The Encyclopedia of Integer Sequences," Academic Press, San Diego, 1996.
21. M. Soittola, Positive rational sequences, *Theoret. Comput. Sci.* **2** (1976), 317–322.
22. R. A. Sulanke, Bijective recurrences concerning Schröder paths, *Electron. J. Combin.*, http://www.combinatorics.org/volume5/v5iltoc.html.
23. J. West, Generating trees and the Catalan and Schröder numbers, *Discrete Math.* **146** (1995), 247–262.
24. D. Zeilberger, Self avoiding walks, the language of science, and Fibonacci numbers, *J. Statist. Plann. Inference* **54** (1996), 135–138.