On the enumerating series of an abstract numeration system

Pierre-Yves Angrand* Jacques Sakarovitch†

Abstract

It is known that any rational abstract numeration system is faithfully, and effectively, represented by an ℕ-rational series. A simple proof of this result is given which yields a representation of this series which in turn allows a simple computation of the value of words in this system and easy constructions for the recognition of recognisable sets of numbers.

It is also shown that conversely it is decidable whether an ℕ-rational series corresponds to a rational abstract numeration system.

1 Introduction

In order to state our result, we have first to recall the definition — due to Lecomte and Rigo [14] — of an *abstract numeration system* and, in order to motivate it, the more common one of numeration systems.

Numbers do exist independently of the way we represent them, and operations on numbers are defined independently of the way they are computed. The role of a numeration system is to set a framework in which numbers are represented by *words* (over a suitable alphabet) allowing to describe operations on numbers as algorithms on the representations, that is, on words.

The most common numeration system — in our modern times — is the k-ary system where numbers are given their representation *in base* k, that is, written as words over the alphabet $A_k = \{0, 1, \ldots, k-1\}$ and which do not start with 0 (but for the representation of 0 itself). The sequence of the representations of the integers in the binary system is: $\{0, 1, 10, 11, 100, 101, 110, \ldots\}$.

While keeping the notion of position numeration system, the k-ary systems can be generalised by replacing the sequence $(k^n)_{n\geqslant 0}$ with some increasing sequence $U = (U_n)_{n\geqslant 0}$ of integers such that $U_0 = 1$. Using a greedy algorithm, every integer n is then given a representation in the 'base' U, called its U-representation and denoted by $\langle n \rangle_U$. A well-known example is the Fibonacci numeration system based on the sequence $F = (F_n)_{n\geqslant 0}$ of Fibonacci numbers starting with $F_0 = 1$

*LTCI (UMR 5141), Telecom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13, France, angrand@enst.fr — Corresponding author.

†LTCI (UMR 5141), CNRS / Telecom ParisTech, sakarovitch@enst.fr.

and $F_1 = 2$. In this system, every positive integer is given a *canonical* representation which is computed by the greedy algorithm and which is characterised by the fact it does not contain 11 as a factor. The sequence of the representations of the integers in the Fibonacci system is: $\{0, 1, 10, 100, 101, 1000, 1001, 1010, \ldots\}$.

It is possible to look at these two numeration systems, the 2-ary system and the Fibonacci system, independently from the sequences $(2^n)_{n\geqslant 0}$ and $(F_n)_{n\geqslant 0}$ and the greedy algorithm, and by just considering *the set of words* that represent the integers: $1\{0,1\}^* \cup \{0\}$ in the first case, $1\{0,1\}^* \setminus \{0,1\}^*11\{0,1\}^* \cup \{0\}$ in the second case and by *enumerating the element of this set* in the radix order.[1] In both cases, every integer will be given the same representation without reference to the way this representation is computed. It is the language of all representations that matters and this naturally leads to the definition of abstract numeration systems.

Definition 1 ([14])**.** *An* abstract numeration system *(or* ANS *for short) is a triple $\mathcal{S} = (L, A, <)$ where A is an alphabet equipped with a total order $<$ and L is an infinite language of A^*.*

The system $\mathcal{S}$ allows to define a one-to-one correspondence between $\mathbb{N}$ and L by associating every integer n with the $(n+1)$-th word of L in the radix order defined on A^* by $<$. This representation *of n is denoted by $\langle n \rangle_{\mathcal{S}}$ and conversely the corresponding* value *of a word w of L is denoted by $\pi_{\mathcal{S}}(w)$. Of course, the following holds:*

$$\langle \pi_{\mathcal{S}}(w) \rangle_{\mathcal{S}} = w \qquad \textit{and} \qquad \pi_{\mathcal{S}}(\langle n \rangle_{\mathcal{S}}) = n \ .$$

In most cases, the alphabet A and the order $<$ on A are fixed and understood and we speak of the ANS defined by the language L and we use the simpler notations $\langle n \rangle_L$ and $\pi_L(w)$.

If L is a rational language of A^*, we say that the ANS is rational.

Example 1. Let $A = \{a, b\}$, with $a < b$ and let L_1 be the language of words with an even number of b's: $L_1 = \{w \in \{a,b\}^* \mid |w|_b \equiv 0 \mod 2\}$. The sequence of the representations of the integers is: $\{\varepsilon, a, aa, bb, aaa, abb, bab, \ldots\}$ and, for instance,

$$\langle 18 \rangle_{L_1} = a\,a\,b\,a\,b \qquad \text{and} \qquad \pi_{L_1}(b\,b\,a\,b\,b) = 29 \ .$$

Beyond the irrepressible appeal to generalisation and abstraction, a true motivation that supports the definition of ANS is to understand which properties of a numeration system depend upon the whole language of the representations only, and which are more directly related to the way the representation of every number is computed. For instance, we have shown in a previous paper [1] that the successor function in a rational ANS is a piecewise cosequential function, whereas the characterisation of those systems for which this successor function is co-sequential is known in the case of β-numeration systems (*cf.* [9]) but seems to be out of reach for arbitrary rational ANS so far.

[1] The definition of radix order will be given below.

The purpose of this paper is to set up even tighter bonds between rational abstract numeration systems and classical automata theory. We reach this goal *via* the definition of the *enumerating series* of a numeration system and with the use of its *representation* in the case it is rational.

Definition 2. *Let $\mathcal{S} = (L, A, <)$ be an abstract numeration system. The enumerating series of $\mathcal{S}$ is the $\mathbb{N}$-series over A^* denoted by $\mathbf{E}_{\mathcal{S}}$ and defined by:*

$$\mathbf{E}_{\mathcal{S}} = \sum_{w \in L} \left(\pi_{\mathcal{S}}(w) + 1 \right) w \quad .$$

As above, the notation can be simplified as $\mathbf{E}_L = \sum_{w \in L} \left(\pi_L(w) + 1 \right) w$ *.*

Remark 1. The above definition has been taken so that the language L is entirely determined by $\mathbf{E}_L$. Indeed,

$$L = \mathsf{supp}\left(\mathbf{E}_L\right) \quad .$$

One certainly could have taken $\mathbf{E}_L = \sum_{w \in L} \left(\pi_L(w) \right) w$ as a definition for $\mathbf{E}_L$. All the results we are going to describe would have been valid and it may have looked more natural. But we would have lost the information on the first word of L, that is, the representation of 0.

The starting point of our work is a direct proof of the following result (the definition of $\mathbb{N}$-rational series will be recalled below).

Theorem 1 ([7])**.** *The enumerating series of a rational abstract numeration system is an $\mathbb{N}$-rational series.*

In [7], Theorem 1 was a corollary of constructions set up for establishing the rationality or algebraicity of a family of counting problems by means of rational transductions. Theorem 1 was also given another and specialised proof in [16]. Even if both this and the original proofs are effective, the one we give below in Section 3 amounts to compute directly a *representation* of $\mathbf{E}_L$ from a representation of (the characteristic series of) L and also to give an even more compact algorithm for calculating the coefficient of a word w in $\mathbf{E}_L$, that is, the value of w in the system L increased by 1. We then deduce from this latter algorithm the construction of the automaton that recognises the set of representations in the system L of a recognisable set of numbers (Section 4). It is to be noted that the same last construction was also given in [13] (*cf.* Remark 6).

The next result plays the role of a converse of Theorem 1: of course not every $\mathbb{N}$-rational series is the enumerating series of a rational number system, but one can at least know when it is the case.

Theorem 2. *It is decidable whether an $\mathbb{N}$-rational series is the enumerating series of a (rational) abstract numeration system or not.*

We end the paper with some problems that are directly inspired by Theorem 1.

2 Preliminary and notation

This paper makes use of several notions of automata theory such as unambiguous, or deterministic, automata, rational series and languages, with which the reader is supposed to be familiar. Definitions that are not given here are to be found in reference books such as [8, 3, 17]. Our notation are mainly those used in [17].

In the sequel, A is a finite alphabet, A^* the free monoid generated by A, 1_{A^*} the empty word, identity of A^*. The length of a word w in A^* is denoted by $|w|$. Let A be *totally ordered* by $<$. The *radix order* $\prec$ on A^* is defined by:[2]

$$u \prec v \qquad \text{if} \quad \begin{cases} \text{either} & |u| < |v| \ , \\ \text{or} & |u| = |v| \ , \quad u = w\,a\,u' \ , \quad v = w\,b\,v' \quad \text{and} \quad a < b \ . \end{cases}$$

The radix order is a *well order*, that is, every non empty subset of A^* has a smallest element for $\prec$ and can thus be used to *enumerate* any subset of A^*.

Let $\mathbb{K}$ be a semiring; for instance, $\mathbb{N}$, the semiring of non negative integers. A ($\mathbb{K}$-)*series* s (over A^*) is a *map* from A^* to $\mathbb{K}$, and the image of a word w by s is called the *coefficient of* w *in* s and is denoted by $\langle s, w\rangle$. The set of series over A^* with coefficients in $\mathbb{K}$ is denoted by $\mathbb{K}\langle\langle A^*\rangle\rangle$. The *support* of a series s is the language, denoted by $\mathsf{supp}\ s$, which contains those words whose coefficient in s is different from $0_{\mathbb{K}}$. Conversely, the *characteristic series* of a language L of A^* is the $\mathbb{N}$-series, denoted by $\underline{L}$, defined by $\langle \underline{L}, w\rangle = 1$ if w is in L and $\langle \underline{L}, w\rangle = 0$ otherwise.

We call ($\mathbb{K}$-)*representation*, of *dimension* n, a triple (λ, μ, ν) where μ is a morphism $\mu\colon A^* \to \mathbb{K}^{n\times n}$ from A^* to the $n \times n$-matrices with entries in $\mathbb{K}$, and λ and ν are two vectors of dimension n with entries in $\mathbb{K}$, λ a row vector and ν a column vector. A series s in $\mathbb{K}\langle\langle A^*\rangle\rangle$ is ($\mathbb{K}$-)*recognisable* if there exists a representation (λ, μ, ν) such that, for every w in A^*,

$$\langle s, w\rangle = \lambda \cdot \mu(w) \cdot \nu \ .$$

A series is ($\mathbb{K}$-)*rational* if it is the behaviour of a finite ($\mathbb{K}$-)automaton, that is, an automaton with multiplicity in $\mathbb{K}$ (the behaviour of an automaton $\mathcal{A}$ is the series where the coefficient of a word w is the sum of the multiplicities of all computations in $\mathcal{A}$ with label w). Finite $\mathbb{K}$-automata whose transitions are labelled by *letters* and $\mathbb{K}$-representations are two ways to describe the same concept.[3] The illustration given with Example 1 suffices for the definition. As every $\mathbb{K}$-automaton is equivalent to one which is labelled by letters, the families of $\mathbb{K}$-rational and $\mathbb{K}$-recognisable series coincide.

Example 2. The language L_1 of words with an even number of b's is recognised by the automaton $\mathcal{A}_1$ drawn at Figure 1. The representation $(\lambda_1, \mu_1, \nu_1)$ associated with $\mathcal{A}_1$ is

$$\lambda_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \ , \quad \mu_1(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ , \quad \mu_1(b) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad \nu_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ .$$

[2] Notice that $\prec$ is not reflexive and is not the order but the *strict part* of the radix order.

[3] This is true only because A^* is a *free monoid*.

In the sequel, we mostly use $\mathbb{N}$ as the semiring, and we may call *representation* an $\mathbb{N}$-representation.

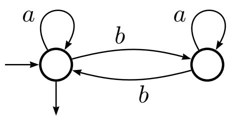

Figure 1: A *DFA* accepting words with an even number of b's

3 Representation of the enumerating series

The proof of Theorem 1, as given in [7] where Theorem 1 is (part of) Corollary 8, is based on the construction of an unambiguous rational transduction that associates to every word u all words v that are greater than u in the radix order. From this, it is easy to derive that the image of the characteristic series of a rational language L is a recognisable series, and equal to $\mathbf{E}_L$ — up to the intersection (or Hadamard product) with L. The advantage of this construction is that it can be applied to unambiguous context-free languages and to various other counting functions as well. The inconvenient is that it does not provide directly the representation of $\mathbf{E}_L$, although it is very similar to the one we develop below.

In [16], Theorem 1 is Proposition 29; its proof is more direct than in [7] in the sense it does not rely on the rational transduction machinery but makes use instead of the characterisation of recognisable series as those which belong to a finitely generated stable submodule of $\mathbb{K}\langle\langle A^*\rangle\rangle$. But this proof yields neither the representation of $\mathbf{E}_L$ nor a simple mean to compute it.

3.1 Preparation

If a is a letter of A, let us denote by A_a the set of letters of A smaller than a:

$$A_a = \{b \in A \mid b < a\} \ .$$

If u be a word of A^*, let us denote by $P(u)$ the set of words of A^* (strictly) smaller than u in the radix order:

$$P(u) = \{v \in A^* \mid v \prec u\} \ .$$

This set $P(u)$ can be defined by induction on the length of u by the following remark. Any word smaller than u followed by *any* letter is smaller than $u\,a$, and so is $u\,b$ for any letter b smaller than a, and the empty word is also smaller than $u\,a$. These three sets are pairwise disjoint and any word smaller than $u\,a$ falls in one of them. Altogether, we have proved the following lemma:[4]

Lemma 3. $\quad \forall u \in A^*\,,\ \forall a \in A \qquad P(u\,a) = 1_{A^*} \cup uA_a \cup P(u)A\,.$

[4] *cf.* Remark 6 below.

Let L be a rational language of A^* and (λ, μ, ν) the $\mathbb{N}$-representation which corresponds to an *unambiguous* finite automaton which recognises L:

$$\forall w \in A^* \qquad \lambda \cdot \mu(w) \cdot \nu = 1 \quad \Longleftrightarrow \quad w \in L \ .$$

We use the following notation: if K is a (finite) subset of A^*, then $\mu(K) = \sum_{w \in K} \mu(w)$. As (λ, μ, ν) corresponds to an *unambiguous* automaton, we have:

$$\forall K \subseteq A^* \qquad \lambda \cdot \mu(K) \cdot \nu = \sum_{w \in K} \lambda \cdot \mu(w) \cdot \nu = \operatorname{card}(K \cap L) \ . \tag{1}$$

3.2 Proof of Theorem 1

Let $\mathcal{S} = (L, A, <)$ be a rational ANS, $\mathcal{A}$ an unambiguous automaton that recognises L, (λ, μ, ν) the corresponding $\mathbb{N}$-representation, and k its dimension. From (1) follows:

$$\forall u \in A^* \qquad \lambda \cdot \mu(P(u)) \cdot \nu = \operatorname{card}(\{v \in A^* \mid v \in L \quad \text{and} \quad v \prec u\}) \ .$$

and thus:

$$\forall w \in L \qquad \lambda \cdot \mu(P(w)) \cdot \nu = \pi_L(w) \ .$$

From Lemma 3 follows:[5]

$$\begin{aligned} &\forall u \in A^* ,\ \forall a \in A \\ &\qquad \lambda \cdot \mu(P(ua)) \cdot \nu = \lambda \cdot \mu(1_{A^*}) \cdot \nu + \lambda \cdot \mu(u) \cdot \mu(A_a) \cdot \nu \\ &\qquad\qquad + \lambda \cdot \mu(P(u)) \cdot \mu(A) \cdot \nu \ . \end{aligned} \tag{2}$$

Let $\sigma = \mu(A)$ and, for every a in A, $\sigma_a = \mu(A_a)$. Thus (2) is rewritten as:

$$\begin{aligned} &\forall u \in A^* ,\ \forall a \in A \\ &\qquad \lambda \cdot \mu(P(ua)) \cdot \nu = \lambda \cdot \nu + \lambda \cdot \mu(u) \cdot \sigma_a \cdot \nu + \lambda \cdot \mu(P(u)) \cdot \sigma \cdot \nu \ . \end{aligned} \tag{3}$$

Let (η, κ, ζ) be the representation of dimension $2k+1$ described by the following $(1, k, k)$-block decomposition:

$$\eta = \begin{pmatrix} 1 & \lambda & 0 \end{pmatrix} , \quad \forall a \in A \quad \kappa(a) = \begin{pmatrix} 1 & 0 & \lambda \\ 0 & \mu(a) & \sigma_a \\ 0 & 0 & \sigma \end{pmatrix} , \quad \zeta = \begin{pmatrix} 0 \\ 0 \\ \nu \end{pmatrix} \ .$$

It is routine to verify, by induction on the length of u, and based on Lemma 3, that $\lambda \cdot \mu(P(u)) \cdot \nu = \eta \cdot \kappa(u) \cdot \zeta$ for every u in A^*.

Let now $\xi = \begin{pmatrix} 1 \\ 0 \\ \nu \end{pmatrix}$ and let s be the series realised by (η, κ, ξ):

$$\forall u \in A^* \qquad \langle s, u \rangle = 1 + \operatorname{card}(\{v \in A^* \mid v \in L \quad \text{and} \quad v \prec u\}) \ .$$

[5] *cf.* Remark 6 below.

In order to get the enumerating series of L, we must retain the words that belong to L only, that is, to make the Hadamard product with the *characteristic series* $\underline{L}$ of L:

$$\mathbf{E}_{\mathcal{S}} = s \odot \underline{L} \ ,$$

and $\mathbf{E}_{\mathcal{S}}$ is $\mathbb{N}$-rational as the Hadamard product of two $\mathbb{N}$-rational series (this is often referred to as (another) Schützenberger Theorem[6]). □

Remark 2. The construction underlying the proof yields for $\mathbf{E}_L$ an $\mathbb{N}$-representation of dimension $2k^2 + k$.

3.3 Computation of the value of a word

The description of a $\mathbb{N}$-rational series s by a $\mathbb{N}$-representation gives a way to compute the coefficient of any word w in s. Theorem 1 thus solves *ipso facto* the problem of computing the value $\pi_L(w)$ of a word w in a rational abstract number system L (which occupies the whole Sect. 2 in [15]).

If s has a representation (χ, ω, ϕ) of dimension n, and if w is of length ℓ, the general algorithm consists in computing $\chi \cdot \omega(w_{i+1}) = (\chi \cdot \omega(w_i)) \cdot \omega(a_{i+1})$ for $i = 0$ to $i = \ell - 1$, where a_i is the i-th letter of w and w_i its prefix of length i. Every step costs $2n^2$ operations, thus in total, roughly $2\ell n^2$ operations.

It would be not such a good idea, however, to apply this general algorithm to the representation of dimension $2k^2+k$ we have obtained in the proof of Theorem 1 above. Its particular form allows, in fact, to compute with vectors and matrix of dimension k only.

Given as above the unambiguous automaton $\mathcal{A}$ of dimension k which recognises L and the corresponding $\mathbb{N}$-representation (λ, μ, ν), we associate a pair $\big(\alpha(w), \gamma(w)\big)$ with every w in A^*, where $\alpha(w)$ and $\gamma(w)$ are two (row) vectors of dimension k, $\alpha(w)$ with entries in $\{0, 1\}$, $\gamma(w)$ with entries in $\mathbb{N}$. The pair $\big(\alpha(w), \gamma(w)\big)$ is computed by induction on the length of w in the following way. Let ℓ be the length of w, let

$$\alpha(1_{A^*}) = \lambda\,, \quad \beta(1_{A^*}) = \lambda\,, \quad \text{and} \quad \gamma(1_{A^*}) = 0 \ ,$$

and, for every $0 \leqslant i < \ell$, let

$$\alpha(w_{i+1}) = \alpha(w_i) \cdot \mu\,(a_{i+1}) \ , \quad \beta(w_{i+1}) = \alpha(w_i) \cdot \sigma_{a_{i+1}} \ , \quad \text{and} \quad \gamma(w_{i+1}) = \lambda + \beta(w_{i+1}) + \gamma(w_i) \cdot \sigma \ .$$

All $\alpha(w)$ have entries in $\{0, 1\}$ since $\mathcal{A}$ is unambiguous. As a simple reformulation of the preceding subsection, we have $\pi_L(w) = \gamma(w) \cdot \nu$ if $\alpha(w) \cdot \nu = 1$, that is, if w is recognised by $\mathcal{A}$ and thus in L, $\pi_L(w)$ undefined otherwise. This algorithm, that is, the computation of $\big(\alpha(w), \gamma(w)\big)$, costs roughly $6\ell k^2$ operations.

[6] *cf.* [3, Th. I.5.3], [8, Prop. VI.7.1] or [17, Cor. III.3.9].

Example 3. Let us consider again the language L_1 and the DFA $\mathcal{A}_1$ of Figure 1. We have thus:

$$\sigma_a = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} , \quad \sigma_b = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad \sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} .$$

The computation of $\pi_{L_1}(bbabb)$ for instance takes the following steps.

i	a_i	α_i	β_i	γ_i
0		$(1,0)$	$(1,0)$	$(0,0)$
1	b	$(0,1)$	$(1,0)$	$(2,0)$
2	b	$(1,0)$	$(0,1)$	$(3,3)$

i	a_i	α_i	β_i	γ_i
3	a	$(1,0)$	$(0,0)$	$(7,6)$
4	b	$(0,1)$	$(1,0)$	$(15,13)$
5	b	$(1,0)$	$(0,1)$	$(29,29)$

And finally $\pi_{L_1}(bbabb) = (29,29) \cdot \nu_1 = 29$.

Remark 3. The computation of $\big(\alpha(w), \gamma(w)\big)$ is very similar to the construction called *product of an automaton by a skew action* in [18, 19].

4 Representation of recognisable subsets of numbers

If s is an $\mathbb{N}$-rational series, that is, a map $s\colon A^* \to \mathbb{N}$, it is well known that for any recognisable set of numbers X, $s^{-1}(X)$ is a rational set of A^* (see [3, Corol. III.2.4], [8, Th. VI.10.1] or [17, Corol. III,4,21], for instance). Theorem 1 thus directly implies the following statement, which has also been proved without reference to it in [14] and in [13].

Corollary 4 ([14])**.** *A recognisable set of numbers is L-recognisable in any rational abstract numeration system L.*

If Corollary 4 requires formally no proof after the characterisation of rational abstract numeration systems given by Theorem 1, it is interesting to further investigate the construction which, given L and a recognisable set of numbers X computes an automaton which recognises the set $\langle X \rangle_L$. The computation method used in the preceding section (which is not the mere application of the general result that yields Corollary 4) allows to establish easily the following statement.

Proposition 5. *Let L be a rational language over A^* recognised by a* deterministic *automaton of dimension k. For any integers p and $r < p$, let $X_{p,r} = p\mathbb{N} + r$ be the set of integers congruent to r modulo p. Then the language $\langle X_{p,r} \rangle_L$ of representations of numbers in $X_{p,r}$ is recognised by a deterministic automaton with at most $k\,p^k$ states.*

Proof. Let $\mathcal{A}$ be an automaton, with set of states Q of cardinal k, which recognises L and (λ, μ, ν) the corresponding $\mathbb{N}$-representation. If $\mathcal{A}$ is deterministic, then λ and μ are row monomial and so are all $\alpha(w)$, for w in A^*, which are thus in 1-1 correspondence with the elements of Q.

Let $\mathcal{C}$ be the automaton whose set of states is

$$R = \{\big(\alpha(w), \delta(w)\big) \mid w \in A^*\} \qquad \text{where} \qquad \delta(w) = \gamma(w) \mod p \ .$$

Thus, $R \subseteq Q \times (\mathbb{Z}/p\mathbb{Z})^k$. The transitions of $\mathcal{C}$ are defined by, for every a in A:

$$\forall w \in A^* \, , \; \forall a \in A \qquad \big(\alpha(w), \delta(w)\big) \xrightarrow[\mathcal{C}]{a} \big(\alpha(wa), \delta(wa)\big) \ .$$

The initial state of $\mathcal{C}$ is $(\lambda, 0)$ and its final states are those $\big(\alpha(w), \delta(w)\big)$ where $\alpha(w)$ is final in $\mathcal{A}$ and $\delta(w) \cdot \nu = r \mod p$. It then follows that the language accepted by $\mathcal{C}$ is $\langle X_{p,r} \rangle_L$. □

Remark 4. If we start from an unambiguous automaton $\mathcal{A}$ of dimension k, the same method yields a deterministic automaton $\mathcal{C}$ with at most $2^k p^k$ states.

Example 4. The automaton built in this way from $\mathcal{A}_1$ and for the recognisable set of numbers $3\mathbb{N} + 1$ is the automaton $\mathcal{C}_1$ shown at Figure 2 (this automaton is not minimal; its minimal quotient has only 8 states).

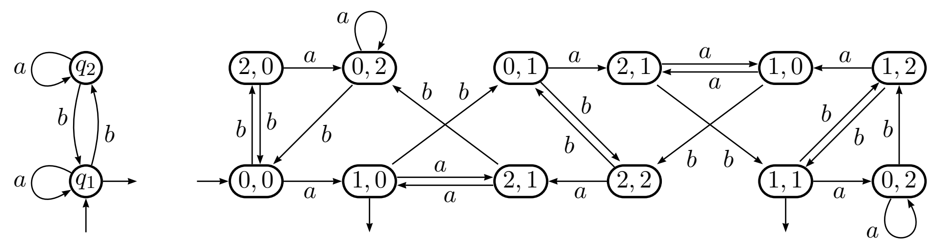

Figure 2: A DFA recognising the set $3\mathbb{N} + 1$ in the ANS L_1.

Remark 5. In [15], another construction has been given for the same purpose. The automaton $\mathcal{D}$ built with this other method and which recognises $\langle X_{p,r} \rangle_L$ is not deterministic, but *codeterministic* and has, roughly, $k\, p^{k+1}$ states. Since $\mathcal{D}$ is codeterministic, its determinisation yields the minimal automaton of $\langle X_{p,r} \rangle_L$ and thus, thanks to Proposition 5, does not produce an exponential blow-up. We do not know of a direct proof of this fact.

Remark 6. After the submitted version was written (and sent), we have learned of the reference [13]. Not only Corollary 4 is established there, but with a method of proof which is very similar to ours. Our Lemma 3 is Lemma 1 in [13]. The term *representation* is not used there but the matrices $\mu(a)$, σ_a and σ are defined (under other notation) and used to give the same proof of Equation (2) (Lemma 2 in [13]).

Afterwards, [13] develops in another direction than this paper: it proves lower bounds for the state complexity of $\langle X_{p,r} \rangle_L$ and shows that the property corresponding to Corollary 4 does not hold for context-free languages.

Remark 7. If the numeration system considered is a positional numeration system (and still a rational one), and under some supplementary hypotheses, then the exact number of states for the minimal automaton of $\langle X_{p,r} \rangle_L$ can be computed (*cf.* [6]).

5 Proof of Theorem 2

The image of a rational language by a rational relation (or transduction) is a rational language; this classical result, due to Nivat and called Evaluation Theorem in [8], extends to rational series, as we state now (*cf.* [17]).

Proposition 6. *Let* $\varphi\colon A^* \to B^*$ *be an unambiguous rational relation and* s *a* $\mathbb{K}$*-rational series over* A^**. Then the series*

$$\underline{\varphi}(s) = \sum_{w \in A^*} <s, w> \underline{\varphi}(w) = \sum_{u \in B^*} <s, \underline{\varphi^{-1}}(u)> u \quad , \tag{4}$$

if it is defined, is a $\mathbb{K}$*-rational series over* B^**.*

It is this result that was used in [7] for the proof of Theorem 1.

Proof of Theorem 2. Let s in $\mathbb{N}\mathrm{Rat}\, A^*$ and $L = \mathsf{supp}\ s$ in $\mathrm{Rat}\, A^*$. The set L is totally ordered by the radix order

$$L = \{w_0 < w_1 < w_2 < \ldots < w_n < \ldots\}$$

and Succ_L is the function from A^* into itself whose domain is L and which maps every w_i to w_{i+1}. It is well-known that Succ_L is a rational function ([4, 10]) and hence unambiguous ([8, 17]). It then follows that the series

$$\underline{\mathsf{Succ}_L}(s) = \sum_{i=0}^{i=\infty} <s, w_i> w_{i+1}$$

is an $\mathbb{N}$-rational series and $t = s - \underline{\mathsf{Succ}_L}(s)$ is a $\mathbb{Z}$-rational series. Now, s is the enumerating series $\mathbf{E}_L$ of the abstract numeration system L if, and only if, for every positive integer i, $<t, w_i> = 1$, that is, if, and only if, $t - \underline{L \setminus \{w_0\}} = 0$, a condition which is known to be decidable as $\mathbb{Z}$ is a sub(semi)ring of a field (*cf.* [8, 17]). □

6 Problems and future work

Looking at abstract number systems as $\mathbb{N}$-rational series naturally leads to two families of questions. The first family consists in questions on $\mathbb{N}$-rational series which ask to which extent the series is related to abstract number systems; the second in questions which generalise to $\mathbb{N}$-rational series questions that are usually considered for (abstract) numeration systems.

An example of questions in the first family is to ask if it is decidable whether a given $\mathbb{N}$-rational series is the enumeration (in a radix ordering) of its (rational) support in a certain, and unknown, abstract numeration system. This seems to be rather a difficult problem. An obvious necessary condition for a series to be a positive instance of this problem is itself a non trivial problem that can be formulated in the following way.

Conjecture 7. *It is decidable whether an $\mathbb{N}$-rational series is a monotone increasing function (for a given order of letters).*

A result due to Honkala [11] provides a kind of converse of Corollary 4 in the case of p-ary numeration systems and states that it decidable whether a p-recognisable set of numbers is recognisable. The generalisation of this result to larger class of numeration systems has been recently studied in [2, 5]. Its generalisation to abstract number systems has been stated as a problem in [12]. It is also a typical example of a question in the second family.

Conjecture 8. *It is decidable whether the set of coefficients of an $\mathbb{N}$-rational series is a recognisable set of numbers.*

7 Summary

In this short paper, we have presented a new idea for the study of abstract number systems, which brings to the subject the whole power of weighted automata theory. In return, the subject of abstract number systems naturally opens new questions for the theory of $\mathbb{N}$-rational series.

Acknowledgements

The authors are grateful to Sylvain Lombardy for the discussions with him on this problem. They also have greatly benefited from having access to the chapter of Lecomte and Rigo on abstract numeration systems in the book *Combinatorics, Automata and Number Theory* to be published by Cambridge University Press.

The authors are particularly indebted to Michel Rigo who, after seeing a first version of this paper, gave them the information that Theorem 1 was already known, the pertinent references, and the encouragements to nevertheless write down a complete version.

References

[1] Angrand, P.-Y. and Sakarovitch J., Radix enumeration of rational languages, *RAIRO Theor. Inf. Appl. 44* (2010) 19–36.

[2] Bell, J., Charlier, E., Fraenkel, A. and Rigo, M., A decision problem for ultimately periodic sets in non-standard numeration systems, *Internat. J.Algebra Comput.* **19** (2009), 809–839.

[3] Berstel, J. and Reutenauer, Ch., Rational Series and their Languages, Springer, 1988.

[4] Berthé, V., Frougny, Ch., Rigo, M., and Sakarovitch, J., On the cost and complexity of the successor function, In *Proc. WORDS 2007* (P. Arnoux, N. Bédaride and J. Cassaigne, eds.), Tech. Rep., Institut de Mathématiques de Luminy (Marseille), (2007) 43–56.

[5] Charlier, E., Abstract numeration systems: recognizability, decidability, multidimensional S-automatic sequences, and real numbers, Thèse de Doctorat, Université de Liège, 2009.

[6] Charlier, E., Rampersad, N., Rigo, M. and Waxweiler, L., State complexity of testing divisibility, In *Proc. DCFS 2010* (I. McQuillan and G. Pighizzini, eds.) Saskatoon Univ. (2010).

[7] Choffrut, Ch. and Goldwurm M., Rational transductions and complexity of counting problems, *Math. Sci. Theory.* **28** (1995) 437–450.

[8] Eilenberg, S., *Automata, Languages and Machines*, vol. A, Academic Press, 1974.

[9] Frougny, Ch., On the sequentiality of the successor function, *Inform. and Comput.* **139** (1997) 17–38.

[10] Frougny, Ch. and Sakarovitch, J., Number representation and finite automata, In *Combinatorics, Automata and Number Theory* (V. Berthé and M. Rigo, eds.) Cambridge University Press, to appear.

[11] Honkala, J., A decision method for the recognizability of sets defined by number systems, *RAIRO Inform. Theor. and Appl.* **20** (1986) 395–403.

[12] Honkala, J. and Rigo, M., A note on decidability questions related to abstract numeration systems, *Discrete Math* **285** (2004) 329–333.

[13] Krieger, D., Miller, A., Rampersad, N., Ravikumar, B. and Shallit, J., Decimation of languages and state complexity, *Theoret. Comp. Sci.* **410** (2009) 2401–2409.

[14] Lecomte, P. and Rigo, M., Numeration systems on a regular language, *Theory Comput. Syst.* **34** (2001) 27–44.

[15] Lecomte, P. and Rigo, M., Abstract numeration systems, In *Combinatorics, Automata and Number Theory* (V. Berthé and M. Rigo, eds.) Cambridge University Press, to appear.

[16] Rigo, M., Numeration systems on a regular language : arithmetic operations, recognizability and formal power series, *Theoret. Comp. Sci.* **269** (2001) 469–498.

[17] Sakarovitch, J., *Eléments de théorie des automates.* Vuibert, 2003. English corrected edition: *Elements of Automata Theory*, Cambridge University Press, 2009.

[18] Sakarovitch, J. and de Souza, R., On the decomposition of k-valued rational relations, In *Proc. STACS 2008* (S. Albers, P. Weil, eds.), (2008) 621–632.

[19] de Souza, R., Etude structurelle des transducteurs de norme bornée, Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications, Paris (2008).