Available online at www.sciencedirect.com
SCIENCE DIRECT®

Theoretical Computer Science 354 (2006) 405–420

Theoretical Computer Science

www.elsevier.com/locate/tcs

Operations preserving regular languages

Jean Berstel[a], Luc Boasson[b], Olivier Carton[b], Bruno Petazzoni[c], Jean-Eric Pin[b,*]

[a]*Institut Gaspard-Monge, Université de Marne-la-Vallée and CNRS, 5, boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2, France*
[b]*LIAFA, Université Paris VII and CNRS, Case 7014, 2 Place Jussieu, F-75251 Paris Cedex 05, France*
[c]*Lycée Marcelin Berthelot, Saint-Maur, France*

Abstract

Given a strictly increasing sequence s of non-negative integers, filtering a word $a_0a_1 \cdots a_n$ by s consists in deleting the letters a_i such that i is not in the set $\{s_0, s_1, \dots\}$. By a natural generalization, denote by $L[s]$, where L is a language, the set of all words of L filtered by s. The filtering problem is to characterize the filters s such that, for every regular language L, $L[s]$ is regular. In this paper, the filtering problem is solved, and a unified approach is provided to solve similar questions, including the removal problem considered by Seiferas and McNaughton. Our approach relies on a detailed study of various *residual notions*, notably residually ultimately periodic sequences and residually rational transductions.

1. Introduction

The original motivation of this paper was to solve an automata-theoretic puzzle, proposed by the fourth author (see also [12]), that we shall refer to as the *filtering problem*. Given a strictly increasing sequence s of non-negative integers, *filtering a word* $a_0a_1 \cdots a_n$ by s consists in deleting the letters a_i such that i is not in the set $\{s_0, s_1, \dots\}$. By a natural generalization, denote by $L[s]$, where L is a language, the set of all words of L filtered by s. The filtering problem is to characterize the filters s such that, for every regular language L, $L[s]$ is regular. The problem is non-trivial since, for instance, it can be shown that the filters n^2 and $n!$ preserve regular languages, while the filter $\binom{2n}{n}$ does not.

The quest for this problem led us to search for analogous questions in the literature. Similar puzzles were already investigated in the seminal paper of Stearns and Hartmanis [19], but the most relevant reference is the paper [18] of Seiferas and McNaughton, in which the so-called *removal problem* was solved: characterize the subsets S of $\mathbb{N}^2$ such that, for each regular language L, the language

$$P(S, L) = \{u \in A^* \mid \text{there exists } v \in A^* \text{ such that } (|u|, |v|) \in S \text{ and } uv \in L\}$$

is regular.

* Corresponding author.
E-mail addresses: Jean.Berstel@univ-mlv.fr (J. Berstel), Luc.Boasson@liafa.jussieu.fr (L. Boasson), Olivier.Carton@liafa.jussieu.fr (O. Carton), Bruno.Petazzoni@ac-creteil.fr (B. Petazzoni), Jean-Eric.Pin@liafa.jussieu.fr (J.-E. Pin).

0304-3975/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2005.11.034

The aim of the present paper is to provide a unified approach to solve at the same time the filtering problem, the removal problem and similar questions. It turns out that these problems are intimately related to the study of *regulators* [6]. A transduction τ from A^* into B^* is a *regulator* if the image under τ of any regular set is regular. It is *continuous* if the inverse image under τ of any regular set is regular. Thus a transduction is continuous if and only if its inverse is a regulator.

Now, the characterization obtained in [18] for the removal problem states that, for any regular subset R of $\mathbb{N}$, the set

$$\{x \in \mathbb{N} \mid \text{ there exists } y \in R \text{ such that } (x, y) \in S\}$$

has to be regular, which exactly means that the relation S is continuous.

Our characterization for the filtering problem is somewhat similar: a filter s preserves regular languages if and only if its differential sequence ∂s (defined by $(\partial s)_n = s_{n+1} - s_n$) is continuous. An equivalent, but more explicit, characterization is the following: for any positive integer t, the two sequences $\partial s \pmod t$ and $\min(\partial s, t)$ have to be ultimately periodic.

The emergence of this differential sequence may appear rather surprising to the reader, but the mystery disappears if, following [13,14], we observe that $L[s] = \tau^{-1}(L)$ where $\tau : A^* \to A^*$ is the transduction defined by

$$\tau(a_0 a_1 \cdots a_n) = A^{s_0} a_0 A^{s_1 - s_0 - 1} a_1 \cdots A^{s_n - s_{n-1} - 1} a_n (1 \cup A)^{s_{n+1} - s_n - 1}.$$

The removal problem can also be interpreted in terms of transductions. It suffices to observe that $P(S, L) = \sigma^{-1}(L)$, where $\sigma : A^* \to A^*$ is the transduction defined by $\sigma(u) = uA^{S(|u|)}$.

Once these problems are interpreted in terms of transductions, the techniques of [13,14] seem to trace an easy road towards their solutions. However, this approach fails, because the above transductions need not be rational or even representable (in the sense of [13,14]).

This failure lead us to a detailed study of transductions by the so-called *residual* approach, which roughly consists in approximating an *infinite* object by a collection of *finite* objects. Profinite techniques (see [1]) and p-adic topology in number theory are good examples of this approach. Another example is the notion of residually ultimately periodic sequence, introduced in [18] as a generalization of a similar notion due to Siefkes [16]. Applying these ideas to transductions, we were lead to the following definitions: a transduction is residually rational if, when it is composed with any morphism onto a finite monoid, the resulting transduction is rational. We analyze in some detail these properties and prove in particular that a transduction is continuous if and only if it is residually rational. This is the key to our problems, since it is now not too difficult to see when our transductions τ and σ are residually rational.

To answer a frequently asked question, we also solve the filtering problem for context-free languages, but the answer is slightly disappointing: only differentially ultimately periodic filters preserve context-free languages.

Our paper is organized as follows. Section 2 introduces the main definitions used in the paper: rational and recognizable sets, relations, transductions, rational transducers, regulators and sequences. The precise formulation of the filtering problem is given in Section 3. Residual properties are studied at length in Section 4 and the properties of differential sequences are analyzed in Section 5. The solutions to the filtering problem and the removal problem are given in Sections 6 and 7. Further properties of residually ultimately periodic sequences are discussed in Section 8 and the filtering problem for context-free languages is solved in Section 9. The paper ends with a short conclusion.

Part of the results of this paper were presented in [3].

2. Preliminaries and background

2.1. Rational and recognizable sets

Given a multiplicative monoid M, the subsets of M form a semiring $\mathcal{P}(M)$ under union as addition and subset multiplication defined by

$$XY = \{xy \mid x \in X \text{ and } y \in Y\}.$$

Recall that the *rational* (or *regular*) subsets of a monoid M form the smallest subset $\mathcal{R}$ of $\mathcal{P}(M)$ containing the finite subsets of M and closed under finite union, product, and star (where X^* is the submonoid generated by X). The set of rational subsets of M is denoted by $\mathrm{Rat}(M)$. It is a subsemiring of $\mathcal{P}(M)$. Rational subsets are closed under rational operations (union, product and star) and under morphisms. This means that if $\varphi : M \to N$ is a monoid morphism, $X \in \mathrm{Rat}(M)$ implies $\varphi(X) \in \mathrm{Rat}(N)$.

Recall that a subset P of a monoid M is *recognizable* if there exists a finite monoid F and a monoid morphism $\varphi : M \to F$ such that $P = \varphi^{-1}(\varphi(P))$. The set of recognizable subsets of M is denoted by $\mathrm{Rec}(M)$. It is also a subsemiring of $\mathcal{P}(M)$. Recognizable subsets are closed under boolean operations, quotients and inverse morphisms.

Let us briefly remind some important results about recognizable and rational sets.

Theorem 2.1 (*Kleene*). *For every finite alphabet A*, $\mathrm{Rec}(A^*) = \mathrm{Rat}(A^*)$.

Theorem 2.2 (*McKnight*). *Let M be a finite monoid. The following conditions are equivalent*:

(1) *M is finitely generated*,
(2) *Every recognizable subset of M is rational*,
(3) *The set M is a rational subset of M*.

Theorem 2.3. *The intersection of a rational set and of a recognizable set is rational.*

Theorem 2.4 (*Mezei*). *Let* $M_1, \ldots, M_n$ *be monoids. A subset of* $M_1 \times \cdots \times M_n$ *is recognizable if and only if it is a finite union of subsets of the form* $R_1 \times \cdots \times R_n$, *where* $R_i \in \mathrm{Rec}(M_i)$.

2.2. Relations

Given two sets E and F, a relation on E and F is a subset of $E \times F$. The *inverse* of a relation S on E and F is the relation S^{-1} on $F \times E$ defined by

$$(y, x) \in S^{-1} \quad \text{if and only if } (x, y) \in S.$$

A relation S on E and F can also be considered as a function from E into $\mathcal{P}(F)$, the set of subsets of F, by setting, for each $x \in E$,

$$S(x) = \{y \in F \mid (x, y) \in S\}.$$

It can also be viewed as a function from $\mathcal{P}(E)$ into $\mathcal{P}(F)$ by setting, for each subset X of E:

$$S(X) = \bigcup_{x \in X} S(x) = \{y \in F \mid \text{there exists } x \in X \text{ such that } (x, y) \in S\}.$$

Dually, S^{-1} can be viewed as a function from $\mathcal{P}(F)$ into $\mathcal{P}(E)$ defined, for each subset Y of F, by

$$S^{-1}(Y) = \{x \in E \mid S(x) \cap Y \neq \emptyset\}.$$

When this *dynamical* point of view is adopted, we say that S is a relation from E into F and we use the notation $S : E \to F$.

2.3. Transductions

Relations between monoids are often called *transductions*. Transductions were intensively studied in connection with context-free languages [2]. In this paper, we shall mainly consider transductions from a finitely generated free monoid A^* into an arbitrary monoid M. A transduction $\tau : A^* \to M$ is *rational* if it is a rational subset of $A^* \times M$.

Let us first recall a standard, but non-trivial property of rational transductions (it is proved for instance right after [2, Proposition III.4.3, p. 67]).

Proposition 2.5. *Let* $\tau : A^* \to M$ *be a rational transduction. If R is a rational subset of* A^*, *then* $\tau(R)$ *is a rational subset of M*.

2.4. Continuous transductions and regulators

A transduction $\tau : A^* \to M$ is called *continuous* [1] if, for each recognizable subset R of M, $\tau^{-1}(R)$ is regular. Continuous transductions were called *recognizability preserving* in [3].

It follows from Proposition 2.5 that every rational transduction is continuous. *Representable transductions*, introduced in [13,14] are other examples of continuous transductions. A characterization of continuous transductions will be given in Section 4.

Following Conway [6], we say that a transduction $\tau : A^* \to B^*$ is a *regulator* if, for each regular language R of A^*, $\tau(R)$ is regular. It follows immediately from the definition that τ is a regulator if and only if its inverse is continuous. In particular, every rational transduction from A^* into B^* is a regulator.

2.5. Rational transducers

Let A be a finite alphabet. The Kleene–Schützenberger theorem [2] states that a transduction $\tau : A^* \to M$ is rational if and only if it can be realized by a *rational transducer*.

Roughly speaking, a rational transducer is a non-deterministic automaton with output in $\mathrm{Rat}(M)$. More precisely, it is a 6-tuple $\mathcal{T} = (Q, A, M, I, F, E)$ where Q is a finite *set of states*, A is the *input alphabet*, M is the *output monoid*, $I = (I_q)_{q\in Q}$ and $F = (F_q)_{q\in Q}$ are arrays of elements of $\mathrm{Rat}(M)$, called respectively, the *initial* and *final outputs*. The set of transitions E is a finite subset of $Q \times A \times \mathrm{Rat}(M) \times Q$. Intuitively, a transition (p, a, R, q) is interpreted as follows: if a is an input letter, the automaton moves from state p to state q and produces the output R.

It is convenient to represent a transition (p, a, R, q) as an edge $p \xrightarrow{a|R} q$. Initial (resp. final) outputs are represented by incoming (resp. outcoming) arrows, which are omitted if the corresponding input (resp. output) is empty. An other standard convention is to simply denote by m the singleton $\{m\}$, for any $m \in M$. The label to the arrow represents the output, but might be omitted if it is equal to the identity of M.

Example 2.1. Let $\mathcal{T} = (Q, A, M, I, E, F)$ be the transducer defined by $Q = \{1, 2\}$, $A = \{a, b\}$, $M = \{a, b\}^*$, $I = (a^*b^*, \emptyset)$, $F = (a^*, b^*)$ and

$$E = \{(1, a, \{1\}, 1), (1, a, \{b\}, 2), (1, b, \{ab\}, 2), (2, a, ba^*, 2), (2, b, \{ba\}, 1)\}.$$

It is represented in Fig. 1

A path is a sequence of consecutive transitions:

$$q_0 \xrightarrow{a_1|R_1} q_1 \xrightarrow{a_2|R_2} q_2 \ \cdots \ q_{n-1} \xrightarrow{a_n|R_n} q_n.$$

The (input) *label* of the path is the word $a_1a_2 \cdots a_n$. Its *output* is the set $I_{q_0}R_1R_2 \cdots R_nF_{q_n}$. The transduction realized by $\mathcal{T}$ maps each word u of A^* onto the union of the outputs of all paths of input label u. For instance, if τ is the transduction realized by the transducer of Example 2.1, there are three paths of input label ab

$$1 \xrightarrow{a|1} 1 \xrightarrow{b|ab} 2 \qquad 1 \xrightarrow{a|b} 2 \xrightarrow{b|ba} 1 \qquad 2 \xrightarrow{a|ba^*} 2 \xrightarrow{b|ba} 1.$$

Since $I_2 = \emptyset$, it follows that $\tau(ab) = (a^*b^*)(1)(ab)(b^*) \cup (a^*b^*)(b)(ba)(a^*)$.

2.6. Sequences

A sequence $(s_n)_{n\geqslant 0}$ of elements of a set is *ultimately periodic* (u.p.) if there exist two integers $m \geqslant 0$ and $r > 0$ such that, for each $n \geqslant m$, $s_n = s_{n+r}$.

The (*first*) *differential sequence* of an integer sequence $(s_n)_{n\geqslant 0}$ is the sequence ∂s defined by

$$(\partial s)_n = s_{n+1} - s_n.$$

[1] We chose this terminology for the following reason: a map from A^* into B^* is continuous in our sense if and only if it is continuous for the *profinite topology* [1] on A^* and B^*.

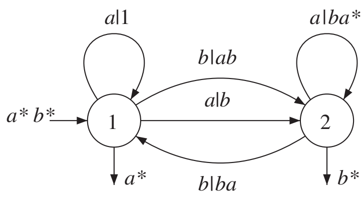

Fig. 1. A transducer.

Note that the integration formula $s_n = s_0 + \sum_{0 \leqslant i \leqslant n-1} (\partial s)_i$ allows one to recover the original sequence from its differential and s_0. A sequence is *syndetic* if its differential sequence is bounded.

If S is an infinite subset of $\mathbb{N}$, the *enumerating sequence* of S is the unique strictly increasing sequence $(s_n)_{n \geqslant 0}$ such that

$$S = \{s_n \mid n \geqslant 0\}.$$

The differential sequence of this sequence is simply called the *differential sequence* of S. A set is *syndetic* if its enumerating sequence is syndetic.

The *characteristic sequence* of a subset S of $\mathbb{N}$ is the sequence c_n defined by

$$c_n = \begin{cases} 1 & \text{if} n \in S, \\ 0 & \text{otherwise.} \end{cases}$$

The following elementary result is folklore.

Proposition 2.6. *Let S be a set of non-negative integers. The following conditions are equivalent*:

(1) *S is a regular subset of* $\mathbb{N}$,
(2) *S is a finite union of arithmetic progressions*,
(3) *the characteristic sequence of S is ultimately periodic*.

If S is infinite, these conditions are also equivalent to the following conditions

(4) *the differential sequence of S is ultimately periodic*.

Example 2.2. Let $S = \{1, 3, 4, 9, 11\} \cup \{7 + 5n \mid n \geqslant 0\} \cup \{8 + 5n \mid n \geqslant 0\} = \{1, 3, 4, 7, 8, 9, 11, 12, 13, 17, 18, 22, 23, 27, 28, \ldots\}$. Then S is a finite union of arithmetic progressions. Its characteristic sequence

$$0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, \ldots$$

and its differential sequence

$$2, 1, 3, 1, 1, 2, 1, 1, 4, 1, 4, 1, 4, \ldots$$

are ultimately periodic.

3. The removal and the filtering problems

A *filter* is a finite or infinite strictly increasing sequence of non-negative integers. If $u = u_0 u_1 u_2 \cdots$ is an infinite word (the u_i are letters), we set

$$u[s] = u_{s_0} u_{s_1} \cdots.$$

Similarly, if $u = u_0 u_1 u_2 \cdots u_n$ is a finite word, we set

$$u[s] = u_{s_0} u_{s_1} \cdots u_{s_k},$$

where k is the largest integer such that $s_k \leqslant n < s_{k+1}$. Thus, for instance, if s is the sequence of squares, *abracadabra* $[s] = abcr$.

By extension, if L is a language (resp. a set of infinite words), we set

$$L[s] = \{u[s] \mid u \in L\}.$$

A filter s *preserves regularity* if, for every regular language L, the language $L[s]$ is regular. The *filtering problem* is to characterize the regularity-preserving filters.

The removal and the filtering problems are instances of a more general question: find out whether a given operator on languages preserves regular languages. The main idea of [13,14] to solve this kind of problem is to write a n-ary operator Ω on languages as the inverse of some transduction $\tau : A^* \to A^* \times \cdots \times A^*$, in such a way that, for all languages $L_1, \ldots, L_n$ of A^*,

$$\Omega(L_1, \ldots, L_n) = \tau^{-1}(L_1 \times \cdots \times L_n)$$

and then to show that τ is a continuous.

Let us try this idea on the removal and the filtering problems. As a first step, we have to express $P(S, L)$ and $L[s]$ as the inverse image of L under a suitable transduction.

We first consider the removal problem. Given a subset S of $\mathbb{N}^2$, we claim that $P(S, L) = \sigma_S^{-1}(L)$, where $\sigma_S : A^* \to A^*$ is the *removal transduction* of S defined by $\sigma_S(u) = uA^{S(|u|)}$. Indeed, we have

$$\begin{aligned}\sigma_S^{-1}(L) &= \{u \in A^* \mid uA^{S(|u|)} \cap L \neq \emptyset\}\\ &= \{u \in A^* \mid \text{there exists } v \in A^* \text{ such that } (|u|, |v|) \in S \text{ and } uv \in L\}\\ &= P(S, L).\end{aligned}$$

Let us now turn to the filtering problem. Let s be a filter. Then $L[s] = \tau_s^{-1}(L)$ where $\tau_s : A^* \to A^*$ is the *filtering transduction* of s defined by

$$\tau_s(a_0a_1 \cdots a_n) = A^{s_0}a_0A^{s_1-s_0-1}a_1 \cdots A^{s_n-s_{n-1}-1}a_n(1 \cup A)^{s_{n+1}-s_n-1}.$$

Observe that $(1 \cup A)^k = 1 \cup A \cup A^2 \cup \cdots \cup A^k$. It remains to find out when σ_S and τ_s are continuous. To show the continuity of a given transduction $\tau : A^* \to M$, a standard technique is to prove that τ is rational or at least representable [13,14].

Unfortunately, except for some special values of S and s, neither σ_S nor τ_s is a rational or even a representable transduction and the methods of [13,14] cannot be applied directly. To overcome this difficulty, we first need to introduce our second major tool, the residual properties.

4. Residual properties

4.1. Residually rational transductions

A transduction $\tau : A^* \to M$ is *residually rational* if, for any morphism $\varphi : M \to F$, where F is a finite monoid, the transduction $\varphi \circ \tau : A^* \to F$ is rational. The next proposition gives a useful characterization of these transductions.

Proposition 4.1. *A transduction $\tau : A^* \to M$ is residually rational if and only if it is continuous.*

Proof. Suppose that τ is residually rational and let $R \in \mathrm{Rec}(M)$. By definition, there exists a morphism φ from M onto a finite monoid F and a subset P of F such that $R = \varphi^{-1}(P)$.

Since τ is residually rational, $\varphi \circ \tau$ is a rational subset of $A^* \times F$. Now F is finite, and thus P is a recognizable subset of F. By Mezei's theorem, $A^* \times P$ is a recognizable subset of $A^* \times F$ and by Theorem 2.3, the set $S = (\varphi \circ \tau) \cap (A^* \times P)$ is a rational subset of $A^* \times F$. Since $S = \bigcup_{x \in P} \tau^{-1}(\varphi^{-1}(x)) \times \{x\}$, the projection of S on A^* is $\tau^{-1}(R)$. Since rational subsets are closed under morphisms, $\tau^{-1}(R)$ is a rational subset of A^*.

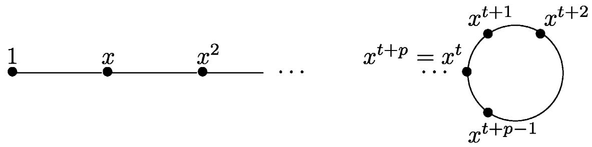

Fig. 2. The monoid $\mathbb{N}_{t,p}$.

Conversely, suppose that, for every $R \in \mathrm{Rec}(M)$, $\tau^{-1}(R) \in \mathrm{Rat}(A^*)$. We claim that τ is residually rational. Let F be a finite monoid and let $\varphi : M \to F$ be a morphism. Then

$$\varphi \circ \tau = \bigcup_{x \in F} \tau^{-1}(\varphi^{-1}(x)) \times \{x\}.$$

Now, for each $x \in F$, $\varphi^{-1}(x)$ is a recognizable subset of M and thus $\tau^{-1}(\varphi^{-1}(x))$ is rational. Since $\{x\}$ is a rational subset of F, $\tau^{-1}(\varphi^{-1}(x)) \times \{x\}$ is a rational subset of $A^* \times F$ and thus $\varphi \circ \tau$ is rational. $\square$

A consequence of Proposition 4.1 is the following.

Corollary 4.2. *Every rational transduction is residually rational.*

Proof. It follows from Propositions 4.1 and 2.5, applied to τ^{-1}. $\square$

The *representable transductions*, introduced in [13,14], are other examples of residually rational transductions.

4.2. Residually ultimately periodic sequences

Let M be a monoid. A sequence $(s_n)_{n \geqslant 0}$ of elements of M is *residually ultimately periodic* (r.u.p.) if, for each monoid morphism φ from M into a finite monoid F, the sequence $\varphi(s_n)$ is ultimately periodic.

We are mainly interested in the case where M is the additive monoid $\mathbb{N}$ of non-negative integers. The following connection with regulators was established in [9,11,18,21].

Proposition 4.3. *A sequence $(s_n)_{n \geqslant 0}$ of non-negative integers is residually ultimately periodic if and only if the function $n \to s_n$ is continuous.*

The finite quotients of $\mathbb{N}$ are the multiplicative cyclic monoids

$$\mathbb{N}_{t,p} = \{1, x, x^2, \ldots, x^{t+p-1}\}$$

presented by the relation $x^{t+p} = x^t$. In other words, $\mathbb{N}_{t,p}$ is the quotient of $\mathbb{N}$ by the monoid congruence $\equiv_{t,p}$ defined as follows:

$$x \equiv_{t,p} y \quad \text{if and only if} \quad \begin{cases} x = y & \text{if } x < t \text{ or } y < t, \\ x \equiv y \pmod{p} & \text{otherwise.} \end{cases}$$

The structure of $\mathbb{N}_{t,p}$ is represented in Fig. 2.

It is well-known that the subsemigroup $\{x^t, \ldots, x^{t+p-1}\}$ is isomorphic to the cyclic group $\mathbb{Z}/p\mathbb{Z}$ and in particular, contains an idempotent.

The two special cases $t = 0$ and $p = 1$ are worth a separate treatment. For $t = 0$, the congruence $\equiv_{t,p}$ is simply the congruence modulo p. For $p = 1$, the congruence $\equiv_{t,1}$, called the *congruence threshold* t, is defined by $x \equiv_{t,1} y$

if and only if $\min(x, t) = \min(y, t)$. Thus threshold counting can be viewed as a formalization of children's counting: zero, one, two, three, ... , many.

A sequence s of non-negative integers is said to be *ultimately periodic modulo p* if, for each monoid morphism $\varphi : \mathbb{N} \to \mathbb{Z}/p\mathbb{Z}$, the sequence $u_n = \varphi(s_n)$ is ultimately periodic. It is equivalent to state that there exist two integers $m \geqslant 0$ and $r > 0$ such that, for each $n \geqslant m$, $u_n \equiv u_{n+r} \pmod p$. A sequence is said to be *cyclically ultimately periodic* (c.u.p.) if it is ultimately periodic modulo p for every $p > 0$. These sequences are called *ultimately periodic reducible* in [18,16].

Example 4.1. The sequences n^2 and $n!$ are both cyclically ultimately periodic. Indeed, for every $p > 0$, and for every $n \geqslant p$, $(n + p)^2 \equiv n^2 \pmod p$ and $n! \equiv 0 \pmod p$.

Example 4.2. It is shown in [16] that the sequence $\lfloor\sqrt{n}\rfloor$ is not cyclically ultimately periodic. Indeed, this sequence is constant on any interval $[n^2, (n + 1)^2[$ and thus cannot be ultimately periodic modulo p (for any p).

Example 4.3. The *Catalan numbers* c_n are defined by $c_n = 1/(n + 1)\binom{2n}{n}$, for $n \geqslant 0$. The sequence of Catalan numbers is not cyclically ultimately periodic. Indeed, let $v_2(m)$ by the highest power of 2 that divides m. Then it is well-known that $v_2(\binom{2n}{n}) = 2^{\beta}(n)$, where $\beta(n)$ is the number of 1's in the binary expansion of n. It follows that $v_2(\binom{2n}{n}) = 2$ if and only if n is a power of 2, and $\binom{2n}{n}$ is divisible by 4 otherwise.

Similarly, a sequence s of non-negative integers is said to be *ultimately periodic threshold t* if, for each monoid morphism $\varphi : \mathbb{N} \to \mathbb{N}_{t,1}$, the sequence $u_n = \varphi(s_n)$ is ultimately periodic. It is equivalent to state that there exist two integers $m \geqslant 0$ and $r > 0$ such that, for each $n \geqslant m$, $\min(u_n, t) = \min(u_{n+r}, t)$.

Example 4.4. For each integer $n \geqslant 0$, denote by $\beta(n)$ the number of 1's in the binary expansion of n. The first values are

n	0	1	2	3	4	5	6	7	8	9	$\cdots$
$\beta(n)$	0	1	1	2	1	2	2	3	1	2	$\cdots$

Of course, $\beta(n) = 1$ if and only if n is a power of 2, and so the sequence $\beta(n)$ is not ultimately periodic with threshold t for any $t > 1$.

Proposition 4.4. *A sequence of non-negative integers is residually ultimately periodic if and only if it is cyclically ultimately periodic and ultimately periodic threshold t for all $t \geqslant 0$.*

Proof. It follows immediately from the definition that a residually ultimately periodic sequence is cyclically ultimately periodic and ultimately periodic threshold t for all $t \geqslant 0$.

Consider now a sequence $(u_n)_{n\geqslant 0}$ which is ultimately periodic modulo p for all $p > 0$ and ultimately periodic threshold t for all $t \geqslant 0$. Let $\varphi : \mathbb{N} \to \mathbb{N}_{t,p}$ be a morphism and let $v_n = \varphi(u_n)$. Denote by e the identity of the cyclic group $G = \{x^t, \dots , x^{t+p-1}\}$. Then the map $\alpha : \mathbb{N}_{t,p} \to G$ defined by $\alpha(s) = se$ is a monoid morphism. Similarly, the map $\beta : \mathbb{N}_{t,p} \to \mathbb{N}_{t,1}$ defined by

$$\beta(x^k) = \begin{cases} x^k & \text{if } k < t, \\ x^t & \text{otherwise,} \end{cases}$$

is a monoid morphism. Note that if x and y are two elements of $\mathbb{N}_{t,p}$ such that $\alpha(x) = \alpha(y)$ and $\beta(x) = \beta(y)$, then $x = y$. Now, by assumption, the sequences $\alpha(v_n)$ and $\beta(v_n)$ are ultimately periodic. That is, there exist integers s, t, p, q such that, for all $n \geqslant s$, $\alpha(v_{n+p}) = \alpha(v_n)$ and, for all $n \geqslant t$, $\beta(v_{n+q}) = \beta(v_n)$. It follows that for all $n \geqslant \max(s, t)$, $\alpha(v_{n+pq}) = \alpha(v_n)$ and $\beta(v_{n+pq}) = \beta(v_n)$ and thus $v_{n+pq} = v_n$. Therefore v_n is ultimately periodic and thus u_n is residually ultimately periodic. $\square$

The next proposition gives a very simple criterion to generate sequences that are ultimately periodic threshold t for all t.

Proposition 4.5. *A sequence $(u_n)_{n \geqslant 0}$ of integers such that $\lim_{n\to\infty} u_n = +\infty$ is ultimately periodic threshold t for all $t \geqslant 0$.*

Proof. Let $t \geqslant 0$. Since $\lim_{n\to\infty} u_n = \infty$, there exists an integer n_0 such that, for all $n \geqslant n_0$, $u_n \geqslant t$. It follows that $\min(u_n, t)$ is ultimately equal to t. □

Example 4.5. The sequences n^2 and $n!$ are residually ultimately periodic. Indeed, we have already seen they are cyclically ultimately periodic. Since they both tend to infinity, Proposition 4.5 shows they are ultimately periodic threshold t for each $t \geqslant 0$ and Proposition 4.4 can be applied.

The sequence $\binom{2n}{n}$ is ultimately periodic threshold t for all t, but is not cyclically ultimately periodic (see Example 4.3).

Let us mention a last example, first given in [5]. Let b_n be a non-ultimately periodic sequence of 0 and 1. The sequence $u_n = (\sum_{0 \leqslant i \leqslant n} b_i)!$ is residually ultimately periodic. It follows that the sequence ∂u is cyclically ultimately periodic. However, it is not residually ultimately periodic since $\min((\partial u)_n, 1) = b_n$.

The class of cyclically ultimately periodic functions has been studied by Siefkes [16], who gave in particular a recursion scheme for producing such functions. The class of residually ultimately periodic sequences was also thoroughly studied [5,9,11,18,21]. Their properties are summarized in the next proposition.

Theorem 4.6 (*Zhang [21], Carton and Thomas [5]*). *Let $(u_n)_{n \geqslant 0}$ and $(v_n)_{n \geqslant 0}$ be r.u.p. sequences. Then the following sequences are also r.u.p.*:

(1) (*composition*) u_{v_n},
(2) (*sum*) $u_n + v_n$,
(3) (*product*) $u_n v_n$,
(4) (*difference*) $u_n - v_n$ *provided that* $u_n \geqslant v_n$ *and* $\lim_{n\to\infty}(u_n - v_n) = +\infty$,
(5) (*exponentiation*) $u_n^{v_n}$,
(6) (*generalized sum*) $\sum_{0 \leqslant i \leqslant v_n} u_i$,
(7) (*generalized product*) $\prod_{0 \leqslant i \leqslant v_n} u_i$.

In particular, the sequences n^k and k^n (for a fixed k), are residually ultimately periodic.

The sequence $2^{2^{2^{\cdot^{\cdot^{2}}}}}$ (exponential stack of 2's of height n) is also considered in [18]. It is also a r.u.p. sequence, according to the following result.

Proposition 4.7. *Let k be a positive integer. Then the sequence u_n defined by $u_0 = 1$ and $u_{n+1} = k^{u_n}$ is r.u.p.*

Proof. Since u_n tends to infinity, it suffices, by Proposition 4.5, to show that u_n is cyclically ultimately periodic. But this follows from the recursion scheme given in [16]. □

The existence of non-recursive, r.u.p. sequences was established in [18]: if $\varphi : \mathbb{N} \to \mathbb{N}$ is a strictly increasing, non-recursive function, then the sequence $u_n = n!\varphi(n)$ is non-recursive but is residually ultimately periodic. The proof is similar to that of Example 4.5.

5. Differential sequences

An integer sequence is called *differentially residually ultimately periodic* (d.r.u.p. in abbreviated form), if its differential sequence is residually ultimately periodic.

What are the connections between d.r.u.p. sequences and r.u.p. sequences? First, the following result holds:

Proposition 5.1 (*Carton and Thomas [5, Corollary 28]*). *Every d.r.u.p. sequence is r.u.p.*

Example 4.5 shows that the two notions are not equivalent. However, if only cyclic counting were used, it would make no difference:

Proposition 5.2. *Let p be a positive number. A sequence is ultimately periodic modulo p if and only if its differential sequence is ultimately periodic modulo p.*

Proof. Let $s = (s_n)_{n \geqslant 0}$ be an integer sequence. If it is ultimately periodic modulo p, then there exist integers t and q such that, for each $n \geqslant t$, $s_{n+q} \equiv s_n \pmod p$. It follows that $s_{n+q+1} - s_{n+q} \equiv s_{n+1} - s_n \pmod p$, showing that the differential sequence of s is ultimately periodic modulo p.

Suppose now that ∂s is ultimately periodic modulo p. Then the proof of [5, Lemma 27] shows that the sequence $s_n = \sum_{0 \leqslant i \leqslant n-1} (\partial s)_i$ is also ultimately periodic modulo p. $\square$

There is a special case for which the notions of r.u.p. and d.r.u.p. sequences are equivalent. Indeed, if the differential sequence is bounded, Proposition 2.6 can be completed as follows.

Lemma 5.3. *If a syndetic sequence is residually ultimately periodic, then its differential sequence is ultimately periodic.*

Proof. Let s be a syndetic sequence and let p be an upper bound for ∂s. If s is r.u.p., Proposition 5.2 shows that ∂s is ultimately periodic modulo p. But since p is an upper bound for ∂s, ∂s is actually ultimately periodic. $\square$

Putting everything together, we obtain

Proposition 5.4. *Let s be a syndetic sequence of non-negative integers. The following conditions are equivalent*:

(1) *s is residually ultimately periodic*,
(2) *∂s is residually ultimately periodic*,
(3) *∂s is ultimately periodic*.

Proof. Proposition 5.1 shows that (2) implies (1). Furthermore (3) implies (2) is trivial. Finally, Lemma 5.3 shows that (1) implies (3). $\square$

Proposition 5.5. *Let S be an infinite syndetic subset of $\mathbb{N}$. The following conditions are equivalent*:

(1) *S is regular*,
(2) *the enumerating sequence of S is residually ultimately periodic*,
(3) *the differential sequence of S is residually ultimately periodic*,
(4) *the differential sequence of S is ultimately periodic*.

Proof. The last three conditions are equivalent by Proposition 5.4 and the equivalence of (1) and (4) follows from Proposition 2.6. $\square$

The class of d.r.u.p. sequences was thoroughly studied in [5].

Theorem 5.6 (*Carton and Thomas [5, Theorem 22]*). *Let $(u_n)_{n \geqslant 0}$ and $(v_n)_{n \geqslant 0}$ be differential residually ultimately periodic sequences. Then the following sequences are also differential residually ultimately periodic*:

(1) (*sum*) $u_n + v_n$,
(2) (*product*) $u_n v_n$,
(3) (*difference*) $u_n - v_n$ *provided that* $u_n \geqslant v_n$ *and* $\lim_{n\to\infty} (\partial u)_n - (\partial v)_n = +\infty$,
(4) (*exponentiation*) $u_n^{v_n}$,
(5) (*generalized sum*) $\sum_{0 \leqslant i \leqslant v_n} u_i$,
(6) (*generalized product*) $\prod_{0 \leqslant i \leqslant v_n} u_i$.

6. A solution to the filtering problem

In this section, we solve completely the filtering problem. Let us start by giving a necessary condition to be a regularity-preserving filter.

Proposition 6.1. *Every regularity-preserving filter is differentially residually ultimately periodic.*

Proof. Let s be a regularity-preserving filter. By Propositions 4.4 and 5.2, it suffices to prove the following properties:

(1) for each $p > 0$, s is ultimately periodic modulo p,
(2) for each $t \geqslant 0$, ∂s is ultimately periodic threshold t.

(1) Let p be a positive integer and let $A = \{0, 1, \dots (p-1)\}$. Let $u = u_0 u_1 \cdots$ be the infinite word whose ith letter u_i is equal to s_i modulo p. At this stage, we shall need two elementary properties of ω-rational sets. The first one states that an infinite word u is ultimately periodic if and only if the ω-language $\{u\}$ is ω-rational. The second one states that, if L is a regular language of A^*, the set of infinite words

$$\overrightarrow{L} = \{v \in A^\omega \mid v \text{ has infinitely many prefixes in } L\}$$

is ω-rational.

We claim that u is ultimately periodic. Define L as the set of prefixes of the infinite word $(0123 \cdots (p-1))^\omega$. Then $L[s]$ is the set of prefixes of u. Since L is regular, $L[s]$ is regular, and thus the set $\overrightarrow{L[s]}$ is ω-rational. But this set reduces to $\{u\}$, which proves the claim. Therefore, the sequence $(s_n)_{n \geqslant 0}$ is ultimately periodic modulo p.

(2) The proof is quite similar to that of (1), but is slightly more technical. Let t be a non-negative integer and let $B = \{0, 1, \dots, t\} \cup \{a\}$, where a is a special symbol. Let $d = d_0 d_1 \cdots$ be the infinite word whose ith letter d_i is equal to $s_{i+1} - s_i - 1$ threshold t. Let us prove that d is ultimately periodic. Consider the regular prefix code

$$P = \{0, 1a, 2a^2, 3a^3, \dots, ta^t, a\}.$$

Then $P^*[s]$ is regular, and so is the language $R = P^*[s] \cap \{0, 1, \dots, t\}^*$. We claim that, for each $n > 0$, the word $p_n = d_0 d_1 \cdots d_{n-1}$ is the maximal word of R of length n in the lexicographic order induced by the natural order $0 < 1 < \cdots < t$. First, $p_n = u[s]$, where $u = a^{s_0} d_0 a^{s_1 - s_0 - 1} d_1 \cdots d_{n-1} a^{s_n - s_{n-1} - 1}$ and thus $p_n \in R$. Next, let $p'_n = d'_0 d'_1 \cdots d'_{n-1}$ be another word of R of length n. Then $p'_n = u'[s]$ for some word $u' \in P^*$. Suppose that p'_n comes after p_n in the lexicographic order. We may assume that, for some index $i \leqslant n-1$, $d_0 = d'_0, d_1 = d'_1, \dots, d_{i-1} = d'_{i-1}$ and $d_i < d'_i$. Since $u' \in P^*$, the letter d'_i, which occurs in position s_i in u', is followed by at least d'_i letters a. Now $d'_i > d_i$, whence $d_i < t$ and $d_i = s_{i+1} - s_i - 1$. It follows in particular that in u', the letter in position s_{i+1} is an a, a contradiction, since $u'[s]$ contains no occurrence of a. This proves the claim.

Let now $\mathcal{A}$ be a finite deterministic trim automaton recognizing R. It follows from the claim that in order to read d in $\mathcal{A}$, starting from the initial state, it suffices to choose, in each state q, the unique transition with maximal label in the lexicographic order. It follows at once that d is ultimately periodic. Therefore, the sequence $(\partial s) - 1$ is ultimately periodic threshold t, and so is (∂s). $\square$

We now show that the converse to Proposition 6.1 is true.

Proposition 6.2. *Let s be a differentially residually ultimately periodic sequence. Then the filtering transduction τ_s is residually rational.*

Proof. Let d be the sequence defined by $d_0 = s_0$ and $d_n = s_n - s_{n-1} - 1$ for $n > 0$. Since s is differentially residually ultimately periodic, d is residually ultimately periodic. Let α be a morphism from A^* into a finite monoid F and $\gamma_s = \alpha \circ \tau_s$. Setting $R = \alpha(A)$, $S = 1 \cup R$ and $\bar{a} = \alpha(a)$ for each $a \in A$, one has

$$\gamma_s(a_0 a_1 \cdots a_n) = R^{d_0} \bar{a}_0 R^{d_1} \bar{a}_1 \cdots R^{d_n} \bar{a}_n S^{d_{n+1}}.$$

Finally, let $\varphi : \mathbb{N} \to \mathcal{P}(F)$ be the monoid morphism defined by $\varphi(n) = R^n$. Since $\mathcal{P}(F)$ is finite and d_n is residually ultimately periodic, the sequence $\varphi(d_n) = R^{d_n}$ is ultimately periodic. Therefore, there exist two integers $t \geqslant 0$ and

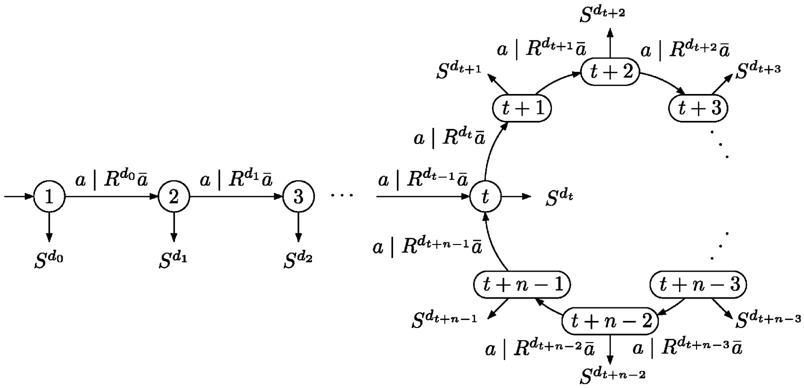

Fig. 3. A transducer realizing γ_s.

$p > 0$ such that, for all $n \geqslant t$, $R^{d_{n+p}} = R^{d_n}$. It follows that the transduction γ_s can be realized by the transducer $\mathcal{T}$ represented in Fig. 3, in which a stands for a generic letter of A.

Formally, $\mathcal{T} = (Q, A, \mathcal{P}(F), I, F, E)$ with $Q = \{1, \ldots, t+n-1\}$, $I_1 = \{1\}$ and $I_q = \emptyset$ for $q \neq 1$, $F_q = S^{q-1}$ for $q \in Q$, and the transitions are of the form $(p, a, R^{p-1}\bar{a}, p+1)$, with $a \in A$ and $p \in Q$ ($p+1$ is of course calculated modulo $\equiv_{t,p}$). Therefore γ_s is rational and thus τ_s is residually rational. $\square$

Putting Propositions 6.1 and 6.2 together, we obtain the characterization announced in the Introduction.

Theorem 6.3. *A filter preserves recognizability if and only if it is differentially residually ultimately periodic.*

7. A solution to the removal problem

A solution to the removal problem was given in [18]. In this section, we only give a proof of the fact that if the relation S is continuous, then the transduction σ_S is also continuous. In view of Proposition 4.1, it is equivalent to prove the following result.

Proposition 7.1. *Let S be a continuous relation on $\mathbb{N}$. The removal transduction σ_S is residually rational.*

Proof. Let α be a morphism from A^* into a finite monoid F. Let $\beta_S = \alpha \circ \sigma_S$ and $R = \alpha(A)$. Since the monoid $\mathcal{P}(F)$ is finite, the sequence $(R^n)_{n \geqslant 0}$ is ultimately periodic. Therefore, there exist two integers $r \geqslant 0$ and $q > 0$ such that, for all $n \geqslant r$, $R^n = R^{n+q}$. Consider the following subsets of $\mathbb{N}$:

$$\begin{aligned}
K_0 &= \{0\} \qquad K_1 = \{1\} \qquad \ldots \qquad K_{r-1} = \{r-1\}\\
K_r &= \{r, r+q, r+2q, \ldots\}\\
K_{r+1} &= \{r+1, r+q+1, r+2q+1, \ldots\}\\
&\vdots\\
K_{r+q-1} &= \{r+q-1, r+2q-1, r+3q-1, \ldots\}.
\end{aligned}$$

The sets K_i, for $i \in \{0, 1, \ldots, r+q-1\}$ are regular and since S is continuous, each set $S^{-1}(K_i)$ is also regular. By Proposition 2.6, there exist two integers $t_i \geqslant 0$ and $p_i > 0$ such that, for all $n \geqslant t_i$, $n \in S^{-1}(K_i)$ if and only if $n + p_i \in S^{-1}(K_i)$. Setting

$$t = \max_{0 \leqslant i \leqslant r+q-1} t_i \quad \text{and} \quad p = \operatorname*{lcm}_{0 \leqslant i \leqslant r+q-1} p_i,$$

we conclude that, for all $n \geqslant t$ and for $0 \leqslant i \leqslant r+q-1$, $n \in S^{-1}(K_i)$ if and only if $n + p \in S^{-1}(K_i)$, or equivalently

$$S(n) \cap K_i \neq \emptyset \iff S(n+p) \cap K_i \neq \emptyset.$$

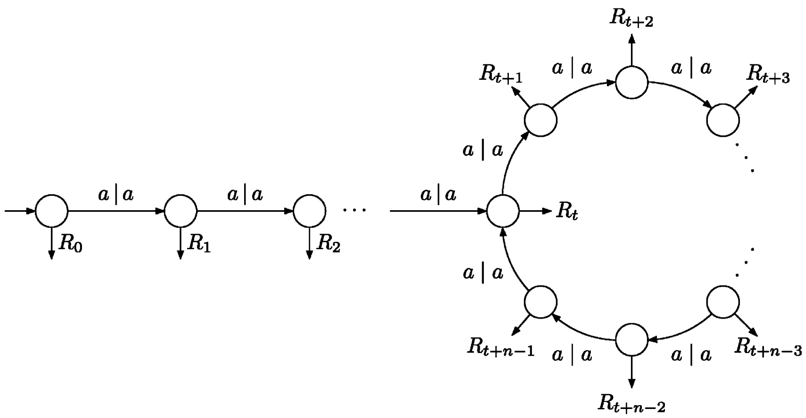

Fig. 4. A transducer realizing β_S.

It follows that the sequence R_n of $\mathcal{P}(F)$ defined by $R_n = R^{S(n)}$ is ultimately periodic of threshold t and period p, that is, $R_n = R_{n+p}$ for all $n \geqslant t$. Consequently, the transduction β_S can be realized by the transducer represented in Fig. 4, in which a stands for a generic letter of A. Therefore β_S is rational and σ_S is residually rational. $\square$

8. Further properties of d.r.u.p. sequences

In this section, we come back to the filtering problem. Filters were defined as strictly increasing sequences, but we could have as well used subsets of $\mathbb{N}$. Indeed, if S is an infinite subset of $\mathbb{N}$, it suffices to set $L[S] = L[s]$ where s is the enumerating sequence of S.

In this setting, the question arises to characterize the filters S such that, for every regular language L, both $L[S]$ and $L[\mathbb{N} \setminus S]$ are regular. By Theorem 6.3, the sequences defined by S and its complement should be d.r.u.p. This implies that S is regular, according to the following slightly more general result.

Proposition 8.1. *Let S and S' be two infinite subsets of $\mathbb{N}$ such that $S \cup S'$ and $S \cap S'$ are regular. If the enumerating sequence of S is d.r.u.p. and if the enumerating sequence of S' is r.u.p., then S and S' are regular.*

Proof. Let s (resp. s') be the enumerating sequence of S (resp. S'). First assume that S' is syndetic. By Proposition 5.5, S' is regular. Now

$$S = \Big((S \cup S') \setminus S'\Big) \cup (S \cap S')$$

and since regular sets are closed under boolean operations, S is regular.

Assume now that S' is not syndetic. Since $S \cup S'$ is an infinite regular subset of $\mathbb{N}$, it contains an arithmetic sequence, say $u_n = a + rn$, for some $a \geqslant 0$ and $r > 0$. Since s is d.r.u.p., the sequence ∂s, counted threshold r, is ultimately periodic. Therefore, there exist n_0 and p such that, for all $n \geqslant n_0$

$$\min((\partial s)_n, r) = \min((\partial s)_{n+p}, r). \tag{1}$$

Since S' is not syndetic, one can find a gap of size p in S'. In other words, there is an interval $I = [b, b + pr]$ such that $I \cap S' = \emptyset$. Without loss of generality, we may assume that $b \geqslant a$ and $b \geqslant s_{n_0}$. Now, at least pr elements of the sequence u_n are in I. These elements belong to $S \cup S'$, and even to S, since I and S' are disjoint. Therefore, $|I \cap S| \geqslant p$. Since S contains all the elements $a + nr$ which are in I, ∂s is bounded by r on I. It follows now from (1) that ∂s is ultimately periodic. It follows by Proposition 5.5 that S is regular. We conclude that S' is regular by the same argument as in the syndetic case, the role of S and S' being swapped. $\square$

The following counter-example shows that the conclusion of Proposition 8.1 no longer holds if S' is only assumed to be residually ultimately periodic. Define a partition $\{S, S'\}$ of $\mathbb{N}$ as follows. Both sets consist of blocks of consecutive integers, obtained by distributing the integers between $n!$ and $(n+1)!$ into n blocks of length $n!$, which are then alternatively allocated to S and S'. Thus we have, with a concise notation,

$$S = \{0, 2, 3, 6\text{–}11, 18\text{–}23, 48\text{–}71, 96\text{–}119, \dots\},$$
$$S' = \{1, 4, 5, 12 - 17, 24 - 47, 72 - 95, 120 - 239, \dots\}.$$

More precisely, given a positive integer m, there is a unique triple of integers (n, k, r) with $n > 0$ and $k > 0$ such that

$$m = kn! + r, \quad 1 \leqslant k \leqslant n \quad \text{and} \quad 0 \leqslant r < n!$$

We use this decomposition of m to define S and S' formally

$$S = \{0\} \cup \{kn! + r | 1 \leqslant k \leqslant n, 0 \leqslant r < n! \text{ and } \lfloor n/2 \rfloor \equiv k \pmod 2\},$$
$$S' = \{kn! + r | 1 \leqslant k \leqslant n, 0 \leqslant r < n! \text{ and } \lfloor n/2 \rfloor \not\equiv k \pmod 2\}.$$

Now, neither S nor S' is ultimately periodic, but the sequences defined by S and S' are both residually ultimately periodic.

We let a last statement as an exercise to the reader.

Proposition 8.2. *Let $S_1, \dots, S_n$ be infinite subsets of $\mathbb{N}$ such that the sets $\bigcup_{1 \leqslant i \leqslant n} S_i$ and $S_i \cap S_j$, for $i \neq j$, are regular. If, for each i, the enumerating sequence of S_i is d.r.u.p., then the sets S_i are all regular.*

9. Filters and context-free languages

We characterized the filters preserving regular languages. What about filters preserving context-free languages? The answer is simple:

Theorem 9.1. *A filter s preserves context-free languages if and only if its differential sequence is ultimately periodic.*

Proof. If the differential sequence of s is ultimately periodic, the filtering transduction τ_s is rational. It follows that the transduction τ_s^{-1} is also rational. Now by a well-known result [2], context-free languages are closed under rational transductions. Since $L[s] = \tau_s^{-1}(L)$, it follows that s preserves context-free languages.

To establish the opposite direction of the theorem, take an infinite filter $s = (s_0, s_1, \dots)$ that preserves context-free languages. Consider the context-free language L over the alphabet $\{a, b, c, d\}$ given by

$$L = \{a^n du \mid n \geqslant 1, u \in \{b, c\}^*, \ |u|_b = n\},$$

and define another language M by $M = L[s] \cap a^+ d\{b, c\}^*$. We claim that

$$M = \{a^n dv \mid n \geqslant 1, v \in \{b, c\}^*, \ 0 \leqslant |v|_b \leqslant s_n - 1\}.$$

Indeed, a word in M has the form $w = a^n \, dv$ for some $n \geqslant 1$ and $v \in \{b, c\}^*$. A word x in L such that $w = x[s]$ has the form

$$x = a^{s_n - 1} \, dy$$

with $y \in \{b, c\}^*$ and $|y|_b = s_n - 1$. It follows that $0 \leqslant |v|_b \leqslant s_n - 1$ and, by choosing the word y in an appropriate way, any value between 0 and $s_n - 1$ can be obtained for $|v|_b$. Consider the projection $\varphi : \{a, b, c, d\}^* \to \{a, b\}^*$. Then

$$N = \varphi(M) = \{a^n b^m \mid 0 \leqslant m \leqslant s_n - 1\}.$$

Since s preserves context-free languages, the language $L[s]$, and consequently also M and N are context-free. Because N is a context-free bounded language over two letters, this is equivalent to the condition that the set

$$H = \{(n, m) \mid 0 \leqslant m \leqslant s_n - 1\}$$

is semilinear or, equivalently, is a rational subset of the free commutative monoid $\mathbb{N}^2$ (see e.g. [7,15]).

Rational subset of $\mathbb{N}^2$ are closed under complementation, so the set $H' = (H + \{(0, 1)\}) \setminus H = \{(n, s_n) \mid n \geqslant 0\}$ is rational. Also, rational subsets of $\mathbb{N}^2$ have unambiguous representations, that is H' is the finite disjoint union of sets of the form $(p_0, q_0) + \sum_{i=1}^{h} (p_i, q_i)\mathbb{N}$, and in our case even with $h = 1$. Indeed, otherwise there are elements $(p_0, q_0) + p_2(p_1, q_1)$ and $(p_0, q_0) + p_1(p_2, q_2)$ in H' and $p_2(p_1, q_1) = p_1(p_2, q_2)$ contradicting the unambiguity.

It follows that H' is a finite disjoint union of sets of the form $(p_0, q_0) + (p, q)\mathbb{N}$. Let P be the lcm of the integers p in these expressions. Then $n \mapsto s_n$ is a linear affine function on each arithmetic progression mod P. $\square$

10. Conclusion

We solved the filtering and the removal problems by using the new concept of residually rational transduction. There are several advantages to this approach.

First, it can be applied to solve most of the automata-theoretic puzzles proposed in the literature [8–11,13,14,17–19]. Next, this approach leads to explicit computations. For instance, given a sequence s and a finite automaton recognizing a language L, one can compute an automaton recognizing $L[s]$. More generally, given an operator on languages Ω, it permits to compute a monoid recognizing $\Omega(L_1, \ldots, L_n)$, given the syntactic monoids of $L_1, \ldots, L_n$. This is a powerful tool for the study of operators on varieties of recognizable languages.

It is easy to create more sophisticated examples, and we do not resist to the temptation to add our own puzzle: show that if L is a recognizable language of A^*, the set

$$\{u \in A^* \mid u^{\underbrace{2^{2^{2^{\cdot^{\cdot^{2}}}}}}_{|u| \text{ times}}} \in L\}$$

is recognizable. The solution follows from the results of this paper.

References

[1] J. Almeida, Finite semigroups and universal algebra, World Scientific Publishing Co. Inc., River Edge, NJ, 1994 (Translated from the 1992 Portuguese original and revised by the author).

[2] J. Berstel, Transductions and Context-free Languages, Teubner, Stuttgart, 1979.

[3] J. Berstel, L. Boasson, O. Carton, B. Petazzoni, J.-E. Pin, Operations preserving recognizable languages, in: A. Lingas, B.J. Nilsson (Eds.), FCT'2003, Lecture Notes in Computer Science, vol. 2751, 2003, pp. 343–354.

[5] O. Carton, W. Thomas, The monadic theory of morphic infinite words and generalizations, Inform. and Comput. 176 (2002) 51–76.

[6] J.H. Conway, Regular Algebra and Finite Machines, Chapman & Hall, London, 1971.

[7] S. Ginsburg, The Mathematical Theory of Context-free Languages, McGraw-Hill Book Co., New York, 1966.

[8] S.R. Kosaraju, Finite state automata with markers, in: Proc. Fourth Annual Princeton Conf. Information Sciences and Systems, Princeton, NJ, 1970, 380.

[9] S.R. Kosaraju, Regularity preserving functions, SIGACT News 6 (2) (1974) 16–17 (Correction to "Regularity preserving functions", SIGACT News 6(3) (1974) 22).

[10] S.R. Kosaraju, Context-free preserving functions, Math. Systems Theory 9 3 (1975) 193–197.

[11] D. Kozen, On regularity-preserving functions, Bull. Europ. Assoc. Theor. Comput. Sc. 58 (1996) 131–138 (Erratum: On regularity-preserving functions, Bull. Europ. Assoc. Theor. Comput. Sc. 59 (1996) 455).

[12] A.B. Matos, Regularity-preserving letter selections. DCC-FCUP Internal Report, Porto, Portugal.

[13] J.-E. Pin, J. Sakarovitch, Operations and transductions that preserve rationality, in: 6th GI Conf., Lecture Notes in Computer Science 145, 1983, pp. 617–628.

[14] J.-E. Pin, J. Sakarovitch, Une application de la représentation matricielle des transductions, Theoret. Comp. Sci. 35 (1985) 271–293.

[15] J. Sakarovitch, Éléments de théorie des automates, Vuibert, Paris, 2003.

[16] D. Siefkes, Decidable extensions of monadic second order successor arithmetic, in: Automatentheorie und formale Sprachen (Tagung, Math. Forschungsinst., Oberwolfach, 1969), Bibliographisches Inst., Mannheim, 1970, pp. 441–472.

[17] J.I. Seiferas, A note on prefixes of regular languages, SIGACT News 6 (1974) 25–29.

[18] J.I. Seiferas, R. McNaughton, Regularity-preserving relations, Theoret. Comput. Sci. 2 (1976) 147–154.

[19] R.E. Stearns, J. Hartmanis, Regularity preserving modifications of regular expressions, Inform. and Control 6 (1963) 55–69.

[21] G.-Q. Zhang, Periodic functions for finite semigroups, Preprint.

Further reading

[4] O. Carton, W. Thomas, The monadic theory of morphic infinite words and generalizations, in: M. Nielsen, B. Rovan (Eds.), MFCS 2000, Lecture Notes in Computer Science, vol. 1893, 2000, 275–284.

[20] G.-Q. Zhang, Automata, Boolean matrices, and ultimate periodicity, Inform. and Comput. 152 (1999) 138–154.