Operations preserving recognizable languages

Jean Berstel[1], Luc Boasson[2], Olivier Carton[2], Bruno Petazzoni[3], and Jean-Éric Pin[2]

[1] Institut Gaspard Monge, Université de Marne-la-Vallée, 5, boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2, berstel@univ-mlv.fr,
[2] LIAFA, Université Paris VII and CNRS, Case 7014, 2 Place Jussieu, F-75251 Paris Cedex 05, FRANCE. Work supported by INTAS project 1224.
Olivier.Carton@liafa.jussieu.fr, Luc.Boasson@liafa.jussieu.fr, Jean-Eric.Pin@liafa.jussieu.fr
[3] Lycée Marcelin Berthelot, Saint-Maur. bpetazzoni@ac-creteil.fr

Abstract. Given a subset S of $\mathbb{N}$, filtering a word $a_0a_1\cdots a_n$ by S consists in deleting the letters a_i such that i is not in S. 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 recognizable language L, $L[S]$ is recognizable. 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. There are two main ingredients on our approach: the first one is the notion of residually ultimately periodic sequences, and the second one is the notion of representable transductions.

1 Introduction

The original motivation of this paper was to solve an automata-theoretic puzzle, proposed by the fourth author (see also [8]), that we shall refer to as the *filtering problem*. Given a subset S of $\mathbb{N}$, *filtering a word* $a_0a_1\cdots a_n$ by S consists in deleting the letters a_i such that i is not in S. 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 recognizable language L, $L[S]$ is recognizable. The problem is non trivial since, for instance, it can be shown that the filter $\{n! \mid n \in \mathbb{N}\}$ preserves recognizable languages.

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 [14], but the most relevant reference is the paper [12] 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 recognizable 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 recognizable.

The aim of this paper is to provide a unified approach to solve at the same time the filtering problem, the removal problem and similar questions. There are two main ingredients in our approach. The first one is the notion of residually ultimately periodic sequences, introduced in [12] as a generalization of a similar notion introduced by Siefkes [13]. The second one is the notion of representable transductions introduced in [9, 10]. Complete proofs will be given in the extended version of this article.

Our paper is organized as follows. Section 2 introduces some basic definitions: rational and recognizable sets, etc. The precise formulation of the filtering problem is given in Section 3. Section 4 is dedicated to transductions. Residually ultimately periodic sequences are studied in Section 5 and the properties of differential sequences are analyzed in Section 6. Section 7 is devoted to residually representable transductions. Our main results are presented in Section 8. Further properties of residually ultimately periodic sequences are discussed in Section 9. The paper ends with a short conclusion.

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 $XY = \{xy \mid x \in X \text{ and } y \in Y\}$. Throughout this paper, we shall use the following convenient notation. If X is a subset of M, and K is a subset of $\mathbb{N}$, we set $X^K = \bigcup_{n \in K} X^n$.

Recall that the *rational* 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 $Rat(M)$. It is a subsemiring of $\mathcal{P}(M)$.

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 $P = \varphi^{-1}(\varphi(P))$. By Kleene's theorem, a subset of a finitely generated free monoid is recognizable if and only if it is rational. Various characterizations of the recognizable subsets of $\mathbb{N}$ are given in Proposition 1 below, but we need first to introduce some definitions.

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

The *(first) differential sequence* of an integer sequence $(s_n)_{n\geq 0}$ is the sequence ∂s defined by $(\partial s)_n = s_{n+1} - s_n$. Note that the integration formula $s_n = s_0 + \sum_{0 \leq i \leq n-1} (\partial s)_i$ 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 $(s_n)_{n\geq 0}$ such that $S = \{s_n \mid n \geq 0\}$. 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 equal to 1 if $n \in S$ and to 0 otherwise. The following elementary result is folklore.

Proposition 1. *Let S be a set of non-negative integers. The following conditions are equivalent:*

(1) *S is recognizable,*

(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 1. Let $S = \{1, 3, 4, 9, 11\} \cup \{7 + 5n \mid n \geq 0\} \cup \{8 + 5n \mid n \geq 0\} = \{1, 3, 4, 7, 8, 9, 11, 12, 13, 17, 18, 22, 23, 27, 28, \ldots\}$. 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.

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 $(y, x) \in S^{-1}$ 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 $S : E \to F$.

A relation $S : \mathbb{N} \to \mathbb{N}$ is *recognizability preserving* if, for each recognizable subset R of $\mathbb{N}$, the set $S^{-1}(R)$ is recognizable.

3 Filtering languages

A *filter* is a finite or infinite increasing sequence s of non-negative integers. If $u = a_0 a_1 a_2 \cdots$ is an infinite word (the a_i are letters), we set $u[s] = a_{s_0} a_{s_1} \cdots$. Similarly, if $u = a_0 a_1 a_2 \cdots a_n$ is a finite word, we set $u[s] = a_{s_0} a_{s_1} \cdots a_{s_k}$, where k is the largest integer such that $s_k \leq n < s_{k+1}$. 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

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

If s is the enumerative sequence of a subset S of $\mathbb{N}$, we also use the notation $L[S]$. If, for every recognizable language L, the set $L[s]$ is recognizable, we say that the filter S *preserves recognizability.* The *filtering problem* is to characterize the recognizability preserving filters.

4 Transductions

In this paper, we consider transductions that are relations from a free monoid A^* into a monoid M. Transductions were intensively studied in connection with context-free languages [1].

Some transductions can be realized by a non-deterministic automaton with output in $\mathcal{P}(M)$, called transducer. More precisely, a *transducer* 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 $\mathcal{P}(M)$, called respectively the *initial* and *final outputs*. The set of transitions E is a finite subset of $Q \times A \times \mathcal{P}(M) \times Q$. 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.

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.

A transduction $\tau : A^* \to M$ is said to be *rational* if τ is a rational subset of the monoid $A^* \times M$. By the Kleene-Schützenberger theorem [1], a transduction $\tau : A^* \to M$ is rational if and only if it can be realized by a *rational transducer*, that is, a transducer with outputs in $Rat(M)$.

A transduction $\tau : A^* \to M$ is said to *preserve recognizability*, if, for each recognizable subset P of M, $\tau^{-1}(P)$ is a recognizable subset of A^*. It is well known that rational transductions preserve recognizability, but this property is also shared by the larger class of *representable transductions*, introduced in [9, 10].

Two types of transduction will play an important role in this paper, the *removal transductions* and the *filtering transductions*. Given a subset S of $\mathbb{N}^2$, considered as a relation on $\mathbb{N}$, the *removal transduction* of S is the transduction $\sigma_S : A^* \to A^*$ defined by $\sigma_S(u) = \bigcup_{(|u|,n) \in S} uA^n$. The *filtering transduction* of a filter s is the transduction $\tau_s : A^* \to A^*$ defined by $\tau_s(a_0a_1 \cdots a_n) = A^{s_0}a_0A^{s_1}a_1 \cdots A^{s_n}a_nA^{\{0,1,\ldots,s_{n+1}\}}$.

The main idea of [9, 10] is to write an n-ary operator Ω on languages as the inverse of some transduction $\tau : A^* \to A^* \times \cdots \times A^*$, that is $\Omega(L_1, \ldots, L_n) = \tau^{-1}(L_1 \times \cdots \times L_n)$. If the transduction τ turns out to be representable, the results of [9, 10] give an explicit construction of a monoid recognizing $\Omega(L_1, \ldots, L_n)$, given monoids recognizing $L_1, \ldots, L_n$, respectively.

In our case, we claim that $P(S, L) = \sigma_S^{-1}(L)$ and $L[s] = \tau_{\partial s-1}^{-1}(L)$. Indeed, we have on the one hand

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

and on the other hand

$$\begin{aligned}
\tau_{\partial s-1}^{-1}(L) = \{&a_0 a_1 \cdots a_n \in A^* \mid \\
&A^{s_0 - 1} a_0 A^{s_1 - s_0 - 1} a_1 \cdots A^{s_n - s_{n-1} - 1} a_n A^{\{0,1,\ldots,s_{n+1}-s_n-1\}} \cap L \neq \emptyset\} \\
= L&[s]
\end{aligned}$$

Unfortunately, the removal transductions and the filtering transductions are not in general representable. We shall see in Section 7 how to overcome this difficulty. But we first need to introduce our second major tool, the residually ultimately periodic sequences.

5 Residually ultimately periodic sequences

Let M be a monoid. A sequence $(s_n)_{n \geq 0}$ 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 connexion with recognizability preserving sequences was established in [5, 7, 12, 16].

Proposition 2. *A sequence $(s_n)_{n \geq 0}$ of non-negative integers is residually ultimately periodic if and only if the function $n \to s_n$ preserves recognizability.*

For each non-negative integer t, define the *congruence threshold* t by setting:

$$x \equiv y \quad (\text{thr } t) \quad \text{if and only if } x = y < t \text{ or } x \geq t \text{ and } y \geq t.$$

Thus threshold counting can be viewed as a formalisation of children counting: zero, one, two, three, ..., many.

A function $s : \mathbb{N} \to \mathbb{N}$ 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 $m \geq 0$ and $r > 0$ such that, for each $n \geq m$, $u_n \equiv u_{n+r} \pmod{p}$. A sequence is said to be *cyclically ultimately periodic* (c.u.p.) if it is ultimately periodic modulo p for every $p > 0$. These functions are called "ultimately periodic reducible" in [12, 13].

Similarly, function $s : \mathbb{N} \to \mathbb{N}$ 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 $m \geq 0$ and $r > 0$ such that, for each $n \geq m$, $u_n \equiv u_{n+r}$ (thr t).

Proposition 3. *A sequence of non-negative integers is residually ultimately periodic if and only if it is ultimately periodic modulo p for all $p > 0$ and ultimately periodic threshold t for all $t \geq 0$.*

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

Proposition 4. *A sequence $(u_n)_{n\geq 0}$ of integers such that $\lim_{n\to\infty} u_n = +\infty$ is ultimately periodic threshold t for all $t \geq 0$.*

Example 2. The sequence $n!$ is residually ultimately periodic. Indeed, let p be a positive integer. Then for each $n \geq p$, $n! \equiv 0 \bmod p$ and thus $n!$ is ultimately periodic modulo p. Furthermore, Proposition 4 shows that, for each $t \geq 0$, $n!$ is ultimately periodic threshold t.

The class of cyclically ultimately periodic functions has been thoroughly studied by Siefkes [13], who gave in particular a recursion scheme for producing such functions. Residually ultimately periodic sequences have been studied in [3, 5, 7, 12, 15, 16]. Their properties are summarized in the next proposition.

Theorem 1. [16, 3] *Let $(u_n)_{n\geq 0}$ and $(v_n)_{n\geq 0}$ 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* $u_n \geq v_n$ *and* $\lim\limits_{n\to\infty}(u_n - v_n) = +\infty$,
(5) *(exponentation)* $u_n^{v_n}$,
(6) *(generalized sum)* $\sum_{0\leq i\leq v_n} u_i$,
(7) *(generalized product)* $\prod_{0\leq i\leq v_n} u_i$.

In particular, the sequences n^k and k^n (for a fixed k), are residually ultimately periodic. However, r.u.p. sequences are not closed under quotients. For instance, let u_n be the sequence equal to 1 if n is prime and to $n! + 1$ otherwise. Then nu_n is r.u.p. but u_n is not r.u.p.. This answers a question left open in [15].

The sequence $2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$ (exponential stack of 2's of height n), considered in [12], is also a r.u.p. sequence, according to the following result.

Proposition 5. *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.*

The existence of non-recursive, r.u.p. sequences was established in [12]: 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 2.

6 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 6. [3, Corollary 28] *Every d.r.u.p. sequence is r.u.p.*

However, the two notions are not equivalent: for instance, it was shown in [3] that if b_n is a non-ultimately periodic sequence of 0 and 1, the sequence $u_n = (\sum_{0 \leq i \leq n} b_i)!$ is r.u.p. but is not d.r.u.p. It suffices to observe that $(\partial u)_n \equiv b_n$ threshold 1.

Note that, if only cyclic counting were used, it would make no difference:

Proposition 7. *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.*

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 1 can be completed as follows.

Lemma 1. *If a syndetic sequence is residually ultimately periodic, then its differential sequence is ultimately periodic.*

Putting everything together, we obtain

Proposition 8. *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 6 shows that (2) implies (1). Furthermore (3) implies (2) is trivial. Finally, Lemma 1 shows that (1) implies (3).

Proposition 9. *Let S be an infinite syndetic subset of $\mathbb{N}$. The following conditions are equivalent:*

(1) *S is recognizable,*
(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 8 and the equivalence of (1) and (4) follows from Proposition 1.

The class of d.r.u.p. sequences was thoroughly studied in [3].

Theorem 2. [3, Theorem 22] *Differential residually ultimately periodic sequences are closed under sum, product, exponentation, generalized sum and generalized product. Furthermore, given two d.r.u.p. sequences $(u_n)_{n \geq 0}$ and $(v_n)_{n \geq 0}$ such that $u_n \geq v_n$ and $\lim_{n \to \infty} (\partial u)_n - (\partial v)_n = +\infty$, the sequence $u_n - v_n$ is d.r.u.p.*

7 Residually representable transductions

Let M be a monoid. A transduction $\tau : A^* \to M$ is *residually rational* (resp. *residually representable*) if, for every monoid morphism α from M into a finite monoid N, the transduction $\alpha \circ \tau : A^* \to N$ is rational (resp. representable).

Since a rational transduction is (linearly) representable, every residually rational transduction is residually representable. Furthermore, every representable transduction is residually representable. We now show that the removal transductions and the filtering transductions are residually rational. We first consider the removal transductions.

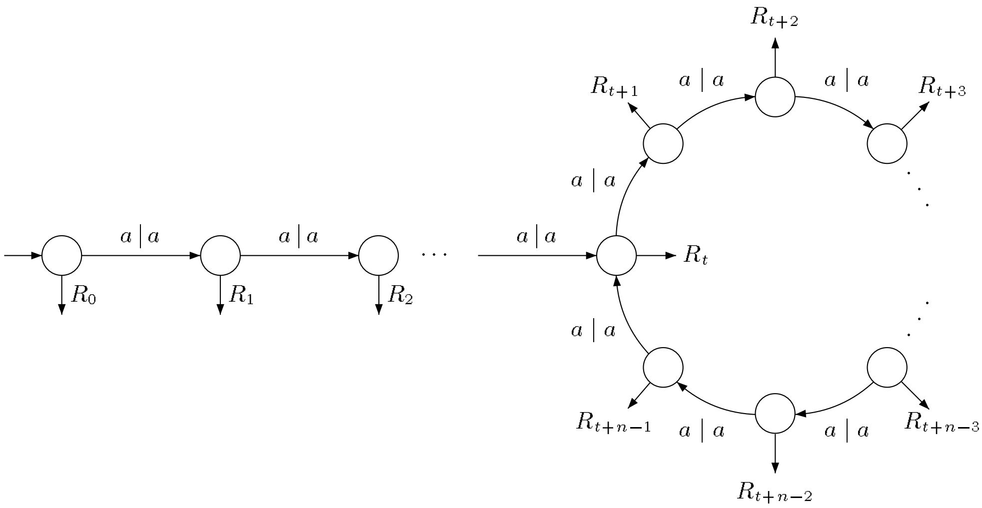

Fig. 1. A transducer realizing β.

Proposition 10. *Let S be a recognizability preserving relation on $\mathbb{N}$. The removal transduction of S is residually rational.*

Proof. Let α be a morphism from A^* into a finite monoid N. Let $\beta = \alpha \circ \tau_s$ and $R = \alpha(A)$. Since the monoid $\mathcal{P}(N)$ is finite, the sequence $(R^n)_{n\geq 0}$ is ultimately periodic. Therefore, there exist two integers $r \geq 0$ and $q > 0$ such that, for all $n \geq r$, $R^n = R^{n+q}$. Consider the following subsets of $\mathbb{N}$: $K_0 = \{0\}$, $K_1 = \{1\}, \ldots,$ $K_{r-1} = \{r-1\}$, $K_r = \{r, r+q, r+2q, \ldots\}$, $K_{r+1} = \{r+1, r+q+1, r+2q+1, \ldots\}$, $\ldots,$ $K_{r+q-1} = \{r+q-1, r+2q-1, r+3q-1, \ldots\}$. The sets K_i, for $i \in \{0, 1, \ldots, r+q-1\}$ are recognizable and since S is recognizability preserving, each set $S^{-1}(K_i)$ is also recognizable. By Proposition 1, there exist two integers $t_i \geq 0$ and $p_i > 0$ such that, for all $n \geq t_i$, $n \in S^{-1}(K_i)$ if and only if $n + p_i \in S^{-1}(K_i)$. Setting $t = \max_{0\leq i\leq r+q-1} t_i$ and $p = \mathrm{lcm}_{0\leq i\leq r+q-1}\, p_i$, we conclude that, for all $n \geq t$ and for $0 \leq i \leq r+q-1$, $n \in S^{-1}(K_i)$ if and only if $n + p \in S^{-1}(K_i)$, or equivalently

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

It follows that the sequence R_n of $\mathcal{P}(N)$ 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 $n \geq t$. Consequently, the transduction β can be realized by the transducer represented in Figure 1, in which a stands for a generic letter of A. Therefore β is rational and τ_s is residually rational.

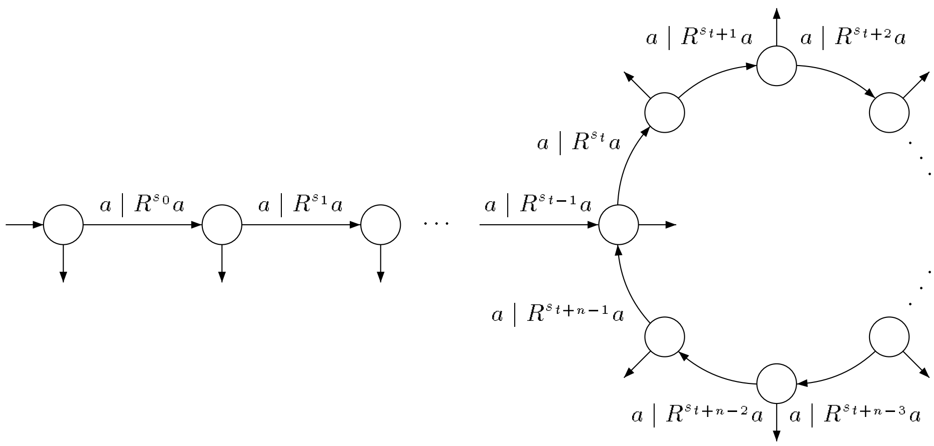

Fig. 2. A transducer realizing γ_s.

Proposition 11. *Let s be a residually ultimately periodic sequence. Then the filtering transduction τ_s is residually rational.*

Proof. Let α be a morphism from A^* into a finite monoid N. Let $\gamma_s = \alpha \circ \tau_s$ and $R = \alpha(A)$. Finally, let $\varphi : \mathbb{N} \to \mathcal{P}(N)$ be the morphism defined by $\varphi(n) = R^n$. Since $\mathcal{P}(N)$ is finite and s_n is residually ultimately periodic, the sequence $\varphi(s_n) = A^{s_n}$ is ultimately periodic. Therefore, there exist two integers $t \geq 0$ and $p > 0$ such that, for all $n \geq t$, $R^{s_{n+p}} = R^{s_n}$. It follows that the transduction γ_s can be realized by the transducer represented in Figure 2, in which a stands for a generic letter of A. Therefore γ_s is rational and thus τ_s is residually rational.

The fact that the two previous transducers preserve recognizability is now a direct consequence of the following general statement.

Theorem 3. *Let M be a monoid. Any residually rational transduction $\tau : A^* \to M$ preserves recognizability.*

Proof. Let P be a recognizable subset of M and let $\alpha : M \to N$ be a morphism recognizing P, where N is a finite monoid. By definition, $\alpha^{-1}(\alpha(P)) = P$. Since τ is residually rational, the transduction $\alpha \circ \tau : A^* \to N$ is rational. Since N is finite, every subset of N is recognizable. In particular, $\alpha(P)$ is recognizable and since τ preserves recognizability, $(\alpha \circ \tau)^{-1}\alpha(P)$ is recognizable. The theorem follows, since $(\alpha \circ \tau)^{-1}\alpha(P) = \tau^{-1}(\alpha^{-1}(\alpha(P))) = \tau^{-1}(P)$.

8 Main results

The aim of this section is to provide a unified solution for the filtering problem and the removal problem.

8.1 The filtering problem

Theorem 4. *A filter preserves recognizability if and only if it is differentially residually ultimately periodic.*

Proposition 11 and Theorem 3 show that if a filter is d.r.u.p., then it preserves recognizability. We now establish the converse property.

Proposition 12. *Every recognizability preserving filter is differentially residually ultimately periodic.*

Proof. Let s be a recognizability preserving filter. By Proposition 3 and 7, it suffices to prove the following properties:

(1) for each $p > 0$, s is ultimately periodic modulo p,

(2) for each $t \geq 0$, ∂s is ultimately periodic threshold t.

(1) Let p be a positive integer and let $A = \{0, 1, ...(p-1)\}$. Let $u = a_0 a_1 \cdots$ be the infinite word whose i-th letter a_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 recognizable language of A^*, then $\overrightarrow{L}$ (the set of infinite words having 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 recognizable, $L[s]$ is recognizable, and thus the set $\overrightarrow{L[s]}$ is ω-rational. But this set reduces to $\{u\}$, which proves the claim. Therefore, the sequence $(s_n)_{n \geq 0}$ is ultimately periodic modulo p.

(2) The proof is quite similar to that of (1), but is sligthly more technical. Let t be a non-negative integer and let $B = \{0, 1, \ldots, t\} \cup \{a\}$, where a is a special symbol. Let $d = d_0 d_1 \cdots$ be the infinite word whose i-th letter d_i is equal to $s_{i+1} - s_i - 1$ threshold t. Let us prove that d is ultimately periodic. Consider the recognizable prefix code $P = \{0, 1a, 2a^2, 3a^3, \ldots, ta^t, a\}$. Then $P^*[s]$ is recognizable, and so is the language $R = P^*[s] \cap \{0, 1, \ldots, t\}^*$. 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 $0 < 1 < \ldots < t$. 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 $i \leq n-1$, $d_0 = d'_0$, $d_1 = d'_1$, $\ldots$, $d_{i-1} = d'_{i-1}$ 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).

8.2 The removal problem

The solution of the removal problem was given in [12].

Theorem 5. *Let S be a subset of $\mathbb{N}^2$. The following conditions are equivalent:*

(1) *for each recognizable language L, the language $P(S, L)$ is recognizable,*
(2) *S is a recognizability preserving relation*

The most difficult part of the proof, (2) implies (1), follows immediately from Proposition 10 and Theorem 3.

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

Coming back to the filtering problem, the question arises to characterize the filters S such that, for every recognizable language L, both $L[S]$ and $L[\mathbb{N} \setminus S]$ are recognizable. By Theorem 4, the sequences defined by S and its complement should be d.r.u.p. This implies that S is recognizable, according to the following slightly more general result.

Proposition 13. *Let S and S' be two infinite subsets of $\mathbb{N}$ such that $S \cup S'$ and $S \cap S'$ are recognizable. 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 recognizable.*

One can show that the conclusion of Proposition 13 no longer holds if S' is only assumed to be residually ultimately periodic.

10 Conclusion

Our solution to the filtering problem was based on the fact that any residually rational transduction preserves recognizability. There are several advantages to our approach.

First, it gives a unified solution to apparently disconnected problems, like the filtering problem and the removal problem. Actually, most of – if not all – the automata-theoretic puzzles proposed in [4–7, 9–12, 14] and [15, Section 5.2], can be solved by using the strongest fact that any residually representable transduction preserves recognizability.

Next, refining the approach of [9, 10], if $\tau : A^* \to A^* \times \cdots \times A^*$ is a residually representable transduction, one could give an explicit construction of a monoid recognizing $\tau^{-1}(L_1 \times \cdots \times L_n)$, given monoids recognizing $L_1, \ldots, L_n$, respectively (the details will be given in the full version of this paper). This information can be used, in turn, to see whether a given operation on languages preserves star-free languages, or other standard classes of rational languages.

Acknowledgements

Special thanks to Michèle Guerlain for her careful reading of a first version of this paper and to the anonymous referees for their suggestions.

References

1. J. Berstel, *Transductions and context-free languages*, Teubner, Stuttgart, (1979).
2. O. Carton and W. Thomas, The monadic theory of morphic infinite words and generalizations, in *MFCS 2000, Lecture Notes in Computer Science* **1893**, M. Nielsen and B. Rovan, eds, (2000), 275–284.
3. O. Carton and W. Thomas, The monadic theory of morphic infinite words and generalizations, Inform. Comput. **176**, (2002), 51–76.
4. S. R. Kosaraju, Finite state automata with markers, in *Proc. Fourth Annual Princeton Conference on Information Sciences and Systems*, Princeton, N. J. (1970), 380.
5. S. R. Kosaraju, Regularity preserving functions, *SIGACT News* **6 (2)**, (1974), 16-17. Correction to "Regularity preserving functions", *SIGACT News* **6 (3)**, (1974), 22.
6. S. R. Kosaraju, Context-free preserving functions, *Math. Systems Theory* **9**, (1975), 193-197.
7. 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.
8. A. B. Matos, Regularity-preserving letter selections, DCC-FCUP Internal Report.
9. J.-É. Pin and J. Sakarovitch, Operations and transductions that preserve rationality, in *6th GI Conference, Lecture Notes in Computer Science* **145**, Springer Verlag, Berlin, (1983), 617–628.
10. J.-É. Pin and J. Sakarovitch, Une application de la représentation matricielle des transductions, *Theoret. Comp. Sci.* **35** (1985), 271–293.
11. J. I. Seiferas, A note on prefixes of regular languages, *SIGACT News* **6**, (1974), 25-29.
12. J. I. Seiferas and R. McNaughton, Regularity-preserving relations, *Theoret. Comp. Sci.* **2**, (1976), 147–154.
13. D. Siefkes, Decidable extensions of monadic second order successor arithmetic, in: *Automatentheorie und formale Sprachen*, (Mannheim, 1970), J. Dörr and G. Hotz, Eds, B.I. Hochschultaschenbücher, 441–472.
14. R. E. Stearns and J. Hartmanis, Regularity preserving modifications of regular expressions, *Information and Control* **6**, (1963), 55–69.
15. Guo-Qiang Zhang, Automata, Boolean matrices, and ultimate periodicity, *Information and Computation*, **152**, (1999), 138–154.
16. Guo-Qiang Zhang, Periodic functions for finite semigroups, preprint.