# Operations preserving recognizable languages*

J. Berstel[1], L. Boasson[2], O. Carton[2], B. Pettazoni[3], J.-E. Pin[2]

[1] IGM, Université de Marne-la-Vallée

[2] LIAFA, Université de Paris VII

[3] Lycée M. Berthelot, Saint-Maur

*Presented at FCT'2003, Malmö.

# Filters

Filter: increasing sequence $(s_n)_{n\geq 0}$ of integers

Example: $s = 0, 1, 4, 9, 16, 25, \ldots$

Image of a word $w = a_0 \cdots a_n$

$$w[s] = a_{s_0} a_{s_1} \cdots a_{s_k} \quad \text{where} \quad s_k \leq n < s_{k+1}$$

Example: $w = abracadabra$

| $w$ | $a$ | $b$ | $r$ | $a$ | $c$ | $a$ | $d$ | $a$ | $b$ | $r$ | $a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $s$ | 0 | 1 | | | 4 | | | | | 9 | |
| $w[s]$ | $a$ | $b$ | | | $c$ | | | | | $r$ | |

$$w[s] = abcr$$

Image of a set $L \subset A^*$ of words : $L[s] = \{w[s] \mid w \in L\}$

# Filtering problem

A filter preserves recognizable sets if, for any recognizable language $L$, the language $L[s]$ recognizable.

**Problem**: characterize filters preserving recognizable sets.

Examples: the following filters preserve recognizable sets:

- $\{2n \mid n \geq 0\}$, (it is a rational transduction)
- $\{n^2 \mid n \geq 0\}$, ( ! it is not a rational transduction)
- $\{2^n \mid n \geq 0\}$, ( !! )
- $\{n! \mid n \geq 0\}$. ( !?! )

But $\{\binom{2n}{n} \mid n \geq 0\}$ **does not** preserve recognizable sets.

# The filter $s = \{n^2 \mid n \geq 0\}$

This filter does not preserve context-free languages, and so is not a rational transduction.

Let $L = \{ca^n ba^{n+1} \mid n \geq 1\}$. The language $M = L[s] \cap ca^+ba^+$ is not context-free.

| 0 | 1 | 4 | 9 | 16 | 25 | 36 | 49 | 64 | 81 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| $c$ | $a$ | $a$ | $b$ | $a$ | | | | | | |
| $c$ | $a$ | $a$ | $a$ | $b$ | $a$ | | | | | |
| $c$ | $a$ | $a$ | $a$ | $a$ | $b$ | $a$ | $a$ | | | |
| $c$ | $a$ | $a$ | $a$ | $a$ | $a$ | $b$ | $a$ | $a$ | | |

General form of words in $M$: $ca_1a_4 \cdots a_{k^2}b_{(k+1)^2}a^\lambda$, where $\lambda$ is the number of squares between $(k+1)^2$ and $2(k+1)^2$.

$\lambda = \lambda_k = \lfloor\sqrt{2}(k+1)\rfloor - (k+1)$ and the set $\{(k, \lambda_k) \mid k \geq 1\}$ is not "semilinear".

# Solution of the filtering problem

Let $r \geq 0$ be a threshold and $q \geq 1$ a period. Two integers $k$ and $k'$ satisfy

$$k \equiv_{r,q} k' \quad \textit{iff} \quad \begin{cases} k = k' & if\ k < r\ or\ k' < r \\ k \equiv k' \mod q & otherwise. \end{cases}$$

$(s_n)_{n\geq 0}$ is residually ultimately periodic, if for any threshold $r$ and period $q \geq 1$, there exist $t \geq 0$ and $p \geq 1$, such that

$$s_n \equiv_{r,q} s_{n+p} \quad \text{for all} \quad n \geq t$$

i.e. the sequence $s_n \mod r, q$ is ultimately periodic.

**Theorem 1** *A filter $(s_n)_{n\geq 0}$ preserves recognizable sets iff the sequence $\partial s_n = s_{n+1} - s_n$ is residually ultimately periodic.*

# Example

Recall

$$k \equiv_{r,q} k' \quad \textit{iff} \quad \begin{cases} k = k' & \textit{when } k < r \textit{ or } k' < r \\ k \equiv k' \mod q & \textit{otherwise}. \end{cases}$$

The representative of $k$ is $k$ itself if $0 \leq k < r$, and is the unique integer $\bar{k} \equiv k \mod q$ and $r \leq \bar{k} < r + q$ otherwise.

For $r = 7, q = 5$, integers greater than $12$ are reduced to one among $7, 8, 9, 10, 11 \mod 5$.

The set of squares has representatives

$$0, 1, 4, 9, 11, 10, 11, 9, 11, 10, 11, \cdots$$

It is ultimately periodic for this $r$ and this $q$.

# Residually ultimately periodic

**Proposition 2** *$s$ is residually ultimately periodic if and only if*

1. *$s$ is ultimately periodic for each $p > 0$,*
2. *$s$ is ultimately periodic with threshold $t$ for each $t \geq 0$*

By definition, $s$ is ultimately periodic with threshold $t$ iff the sequence $(\min(s_n, t))$ is ultimately periodic.

Examples : The sequence of squares.

The sequence

$$010\mathbf{2}010\mathbf{3}010201\mathbf{4}01020103010201\mathbf{5}\cdots$$

# Removal problem

Let $S$ be a relation over $\mathbb{N}$ and $L \subset A^*$. Define

$$L/S = \{u \mid \exists v \ \ (|u|, |v|) \in S \text{ and } uv \in L\}$$

Example : Let $S = \{(n, n) \mid n \in \mathbb{N}\}$. Then $L/S$ is the set of first halves of words in $L$.

A relation $S$ preserves recognizable sets over $\mathbb{N}$ if, for any recognizable $K \subseteq \mathbb{N}$, the set $S(K)$ is recognizable over $\mathbb{N}$ (i.e. a finite union of arithmetic progressions and of a finite set).

**Theorem 3 (Seiferas, McNaughton)**
$L/S$ *is recognizable for any recognizable set* $L$ *iff* $S^{-1}$ *preserves recognizable sets over* $\mathbb{N}$*.*

# Transductions

Transductions are relations from $A^*$ into $B^*$ and later into some monoid $M$.

**Filtering transduction**: Let $s = (s_n)_{n\geq 0}$ be a sequence of integers. Define $\tau_s$

$$\tau_s(a_0 \cdots a_n) = A^{s_0} a_0 A^{s_1 - s_0 - 1} \ldots A^{s_n - s_{n-1} - 1} a_n A^{\leq s_{n+1} - s_n - 1}$$

One has

$$L[s] = \tau_s^{-1}(L).$$

**Removal transduction**: Let $S$ be a relation over $\mathbb{N}$. Define $\tau_S$

$$\tau_S(u) = \bigcup_{(|u|,n)\in S} uA^n.$$

One has

$$L/S = \tau_S^{-1}(L).$$

# Transducers

Let $A$ be an alphabet and $M$ be a monoid.

$$\mathcal{T} = (Q, A \times \mathfrak{P}(M), E, I, F)$$

$Q$: states; $E$: transitions; $I$: initial states; $F$: final states.

Transitions: $q \xrightarrow{a|R} q'$ where $a \in A$ and $R \in \mathfrak{P}(M)$.

Initial and final labels: The entries of the vectors $I, F \in \mathfrak{P}(M)^Q$.

A transducer realizes a transduction $\tau$ from $A^*$ to $M$ defined as follows.

For $w = a_1 \cdots a_n$,

$\tau(w)$ is the union of all products $I_0 R_1 \cdots R_n F_n$ for all paths

$$\xrightarrow{I_0} q_0 \xrightarrow{a_1|R_1} q_1 \xrightarrow{a_2|R_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{a_n|R_n} q_n \xrightarrow{F_n}$$

# Rational transductions

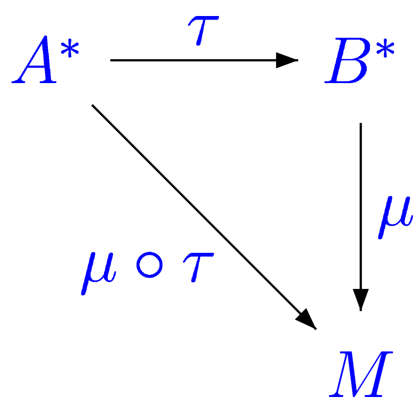
A transduction is rational if it can be realized by a finite transducer with output labels that are rational subsets of $M$.

**Theorem 4** *Let $\tau$ be a rational transduction from $A^*$ to $M$. If $K$ is a recognizable subset of $A^*$, then $\tau(K)$ is rational subset of $M$. If $L$ is a recognizable subset of $M$, then $\tau^{-1}(L)$ is a regular language over $A$.*

However, the filtering transduction and the removal transduction are **not** rational.

# Residually rational transductions

A transduction $\tau$ from $A^*$ to $B^*$ is residually rational if for any morphism $\mu$ from $B^*$ into a **finite** monoid $M$, $\mu \circ \tau$ is rational.



**Theorem 5** *If $\tau$ is residually rational and $L \subset B^*$ is recognizable, then $\tau^{-1}(L)$ is also recognizable.*

**Proof**. Let $\mu : B^* \to M$ be the syntactic morphism of $L$. Then

$$\tau^{-1}(L) = (\mu \circ \tau)^{-1}(P).$$

where $L = \mu^{-1}(P)$.

# Filtering transduction

**Proposition 6** *The filtering transduction is residually rational.*

Recall that $\tau_s : A^* \to A^*$ is

$$\tau_s(a_0 \cdots a_n) = A^{s_0} a_0 A^{d_1} a_1 \cdots a_{n-1} A^{d_n} a_n A^{\leq d_{n+1}}$$

where $d_n = s_{n+1} - s_n - 1$.
Let $R = \mu(A)$ be the image of $A$ in a finite monoid $M$. Since $\mathfrak{P}(M)$ is finite, there $r$ and $q$ such that

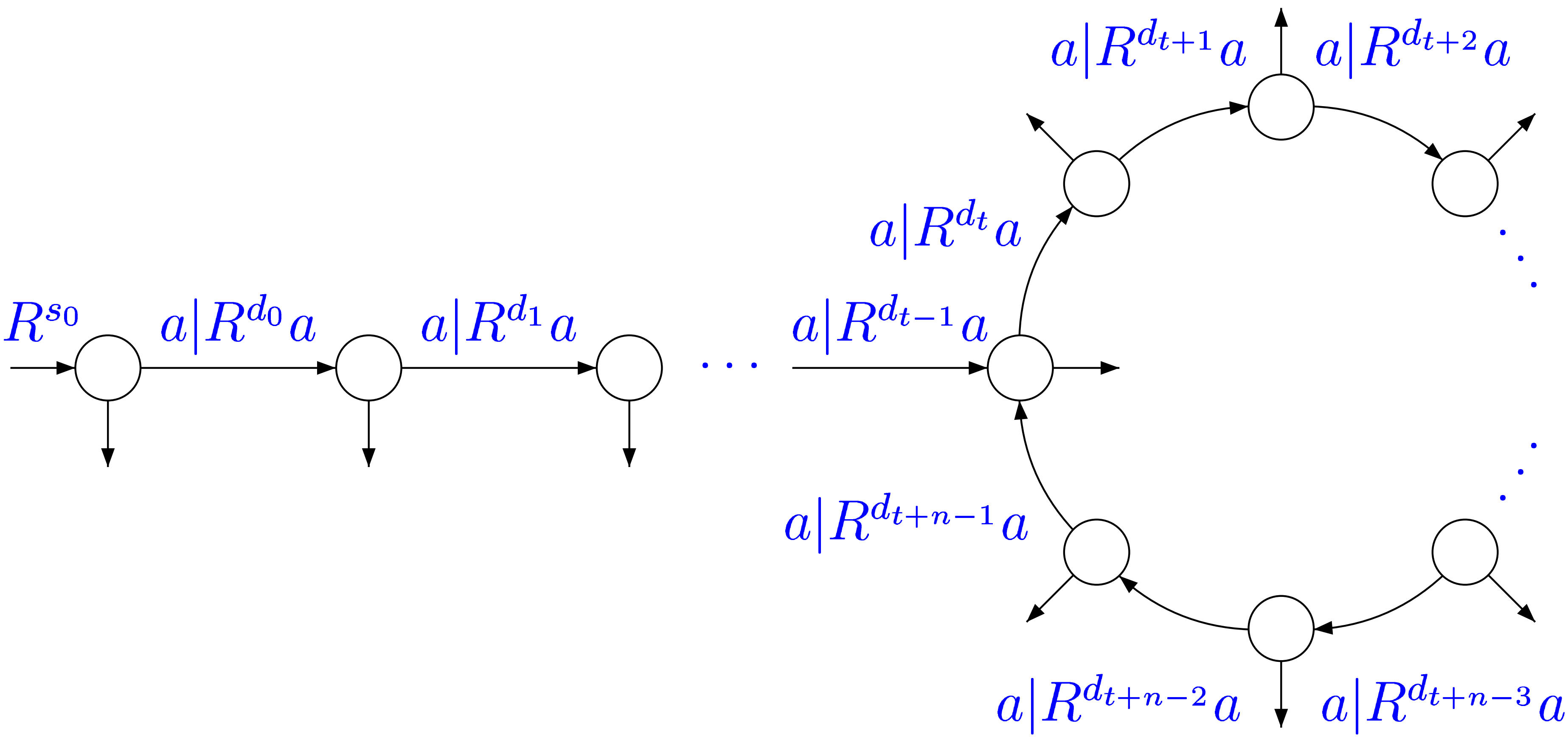
$$R^r = R^{r+q}.$$

Since $(d_n)_{n \geq 0}$ is residually ultimately periodic, there are $t$ and $p$ such that

$$R^{d_n} = R^{d_{n+p}} \quad \text{for every} \quad n \geq t.$$

Thus, $\mu \circ \tau_s$ is realized by the following transducer:

# Filtering transducer



# Removal transduction

**Proposition 7** *The removal transduction is residually rational.*

Recall that the removal transduction is defined by

$$\tau_S(u) = \bigcup_{(|u|,m)\in S} uA^m.$$

Let $R = \mu(A)$ be the image of $A$ in a finite monoid $M$. Since $\mathfrak{P}(M)$ is finite, there $r$ and $q$ such that

$$R^r = R^{r+q}.$$

Define $r + q$ recognizable sets $K_i$ of integers by

$$K_i = \begin{cases} \{i\} & \text{if } 0 \leq i < r \\ \{i + qn \mid n \geq 0\} & \text{if } r \leq i < r + q. \end{cases}$$

# Removal transduction

Since the sets $S^{-1}(K_i)$ are recognizable, there are $t$ and $p$ such that for any $0 \leq i < r + q$ and any $n \geq t$,

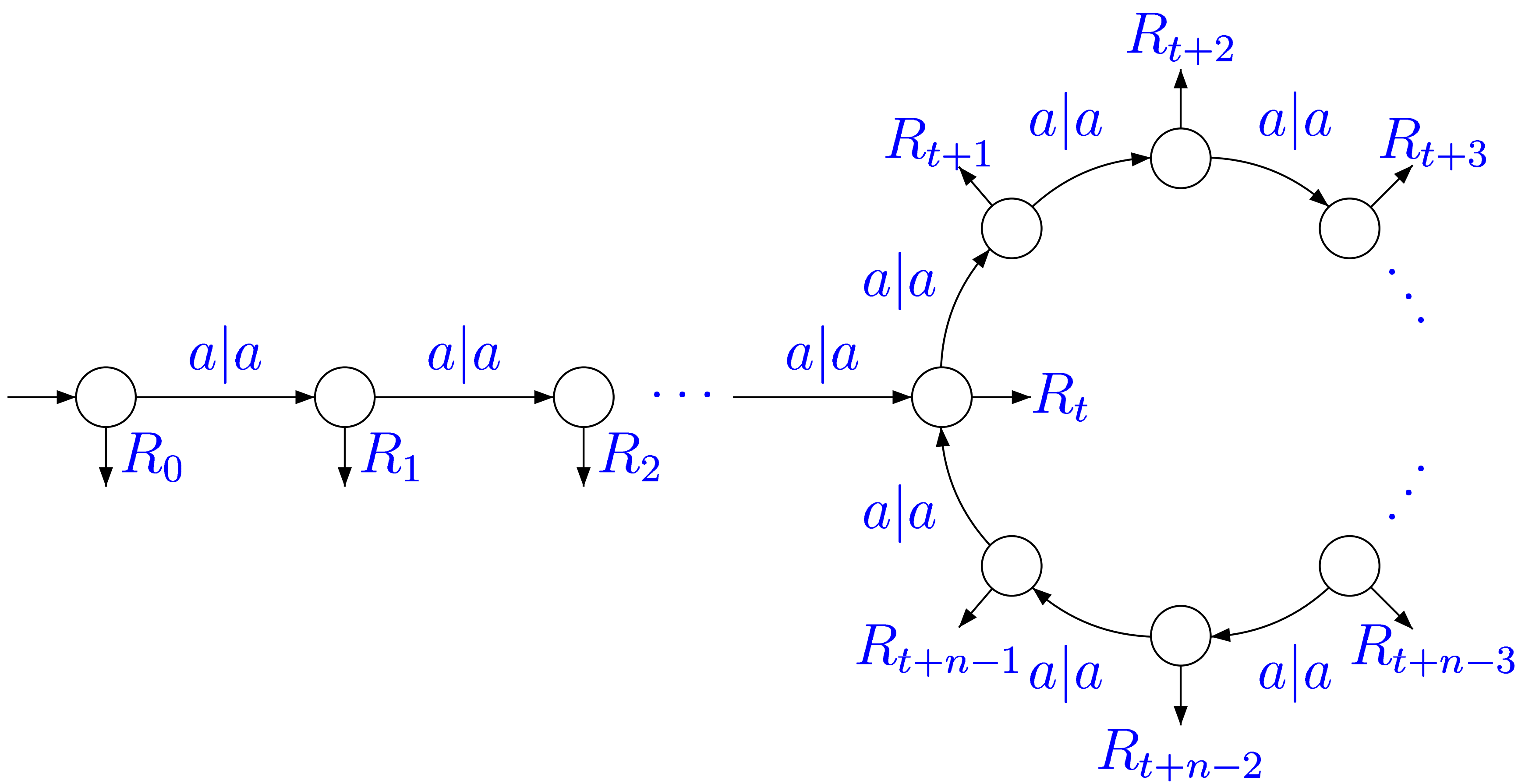
$$n \in S^{-1}(K_i) \iff n + p \in S^{-1}(K_i) \quad n \geq t$$

whence

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

Setting $R_n = R^{S(n)} = \bigcup_{m \in S(n)} R^m$, one gets $R_n = R_{n+p}$ for $n \geq t$.

# Removal transducer



# A filter preserving recognizable sets is drup

**Proposition 8** *A filter preserving recognizable sets is ultimately periodic for each* $p > 0$*.*

Let $A = \{0, 1, \ldots, p-1\}$, and let $u = a_0 a_1 \cdots$ be the infinite word defined by $a_i = s_i \bmod p$:

$$\begin{aligned} s &= s_0 \ s_1 \ s_2 \ \cdots \\ u &= a_0 \ a_1 \ a_2 \ \cdots \end{aligned}$$

Set $v = (01 \cdots (p-1))^\omega$. The letter at position $s_i$ in $v$ is $a_i$.

Let $L$ be the set of prefixes of $v$. Then $L[s]$ is the set of prefixes of $u$.

Since $L[s]$ is regular, the infinite word $u$ is ultimately periodic.

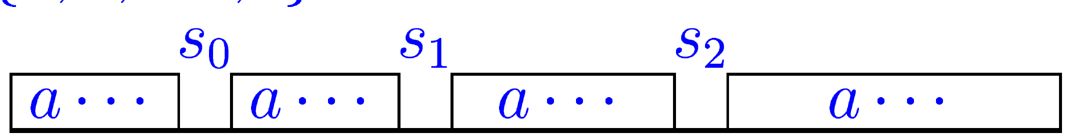
# A filter preserving recognizable sets is drup (2)

**Proposition 9** *If a filter $s$ preserves recognizable sets, then $\partial s$ is ultimately periodic with threshold $t$ for each $t \geq 0$.*

Set $d_i = \min(t, s_{i+1} - s_i - 1)$. We show that the infinite word $d = d_0 d_1 \cdots$ is ultimately periodic.
Set $B = \{0, 1, \ldots, t\} \cup \{a\}$ and define a prefix code

$$P = \{0, 1a, 2a^2, \ldots, ta^t, a\}$$

The language $P^*[s]$ is recognizable, and so is $R = P^*[s] \cap \{0, 1, \ldots, t\}$.




The maximal word (for the order $0 < 1 < \cdots < t$) of length $n$ in $R$ is $d_0 d_1 \cdots d_{n-1}$. The word $d$ is read in a trim automaton recognizing $R$ by taking at each state the edge with maximal label. Thus it is ultimately periodic.

# Rup and drup

A sequence $s$ is differentially residually ultimately periodic if $\partial s = (s_{n+1} - s_n)$ is residually ultimately periodic.
$s$ drup $\Rightarrow$ $s$ rup.

$s$ rup and $\lim \partial s_n = \infty \Rightarrow s$ drup.

$s$ rup and $\partial s$ bounded ($s$ "syndetic") $\Rightarrow \partial s$ ultimately periodic.

The set of residually ultimately periodic sequences is closed under sum, product, exponentiation, composition ($u_{v_n}$) etc.
The set of differentially residually ultimately periodic sequences is closed sum, product, exponentiation, etc.

Sequences that are not rup:
Spectra $\{\lfloor \alpha n \rfloor \mid n \geq 1\}$ for irrational $\alpha$.
The sequence of positions of $1$'s in the Thue-Morse sequence.
The sequence of Catalan numbers.