Regularity preserving functions and transductions, a survey

Jean-Éric Pin

IRIF, CNRS and University Paris

June 23, RAQM 2021

Outline

(1) Some history
(2) Functions from $\mathbb{N}$ to $\mathbb{N}$
(3) Matrix representations
(4) Topological approach
(5) The case of p-group languages

Transductions

Let M and N be monoids. A transduction $\tau\colon M \to N$ is a relation on M and N, viewed as a function from M to $\mathcal{P}(N)$.

One extends τ to a function $\mathcal{P}(M) \to \mathcal{P}(N)$ by setting $\tau(P) = \bigcup_{m \in P} \tau(m)$.

The inverse transduction $\tau^{-1}\colon N \to M$ is defined by

$$\tau^{-1}(Q) = \{m \in M \mid \tau(m) \cap Q \neq \emptyset\}.$$

Regularity-preserving functions and transductions

A function $f : A^* \to B^*$ is regularity-preserving if, for each regular language L of B^*, $f^{-1}(L)$ is also regular.

More generally, let $\mathcal{C}$ be a class of regular languages. A function $f : A^* \to B^*$ is $\mathcal{C}$-preserving if, for each $L \in \mathcal{C}$, $f^{-1}(L)$ is also in $\mathcal{C}$.

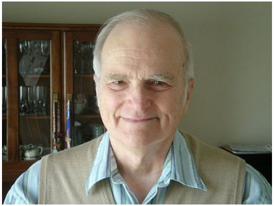
Same definitions for transductions.

Extensions to rational formal power series (Droste and Zhang, 2003) will not be covered in this lecture.

Part I

Some history

Back to Werner's youth...



Stearns and Hartmanis,

Regularity preserving modifications of regular expressions (1963).

Deleting a W-factor

Exercise. Let W be any language. Show that if L is regular [star-free], then so is

$$K = \{u \mid u = xy \text{ and } xwy \in L \text{ for some } w \in W\}$$

Deleting a W-factor, an algebraic proof

Exercise. Let W be any language. Show that if L is regular [star-free], then so is

$$K = \{u \mid u = xy \text{ and } xwy \in L \text{ for some } w \in W\}$$

Proof. Let $h\colon A^* \to M$ be the syntactic morphism of L. Setting

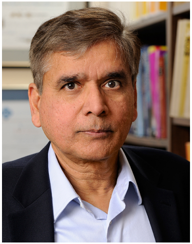
$$T = \{(n, m) \in M \times M \mid nh(W)m \cap h(L) \neq \emptyset\}$$

one gets

$$K = \bigcup_{(n,m)\in T} h^{-1}(n)h^{-1}(m)$$

and the result follows.


Hopcroft and Ullman,
Formal Languages and their relation to Automata (1969).



Kosaraju,
- Finite state automata with markers (1970).
- Regularity preserving functions (1974).

Seiferas,
A note on prefixes
of regular languages (1974)

Seiferas and McNaughton,
Regularity-preserving functions (1976)

Suffix removals

Let $\tau : \mathbb{N} \to \mathbb{N}$ be a transduction and L be a language. Let

$$P(\tau, L) = \{\, p \mid \text{such that } ps \in L \text{ for some } s \\ \text{such that } |s| \in \tau(|p|)\}$$

When does L regular imply $P(\tau, L)$ regular?

Theorem (Seiferas and McNaughton)

This happens iff τ is regularity-preserving.

Subword filtering problem (A. B. Matos)

Let $f : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function. Filtering a word $u = a_0 a_1 \cdots a_n$ through f consists in just keeping the letters a_i such that i is in the range of f.

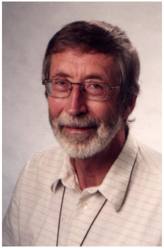
If L is regular, is the set of words of L filtered by f always regular?

Theorem (Berstel, Boasson, Carton, Petazzoni, P. (2006))

This happens iff the function Δf defined by $\Delta f(n) = f(n+1) - f(n)$ is regularity-preserving.

Part II

Functions from $\mathbb{N}$ to $\mathbb{N}$



Siefkes,
Decidable extensions of monadic second order successor arithmetic (1970)

Ultimately periodic functions

A function $f : \mathbb{N} \to \mathbb{N}$ is ultimately periodic if there exists $t \geqslant 0$ and $p > 0$ such that, for all $n \geqslant t$, $f(n+p) = f(n)$. For instance, the sequence

$$1, 4, 0, 2, 8, 1, \underbrace{2, 3, 5}, \underbrace{2, 3, 5}, \underbrace{2, 3, 5}, \underbrace{2, 3, 5}, \ldots$$

is ultimately periodic.

A function $f : \mathbb{N} \to \mathbb{N}$ is ultimately periodic modulo n if the function $f \bmod n$ is ultimately periodic. It is cyclically ultimately periodic if it is ultimately periodic modulo n for all $n > 0$.

Regularity-preserving functions from $\mathbb{N}$ to $\mathbb{N}$

Theorem (Siefkes 1970, Seiferas-McNaughton 1976)

A function $f : \mathbb{N} \to \mathbb{N}$ is ultimately periodic modulo n iff for $0 \leqslant k < n$, the set $f^{-1}(k + n\mathbb{N})$ is regular.

Theorem (Siefkes 1970, Seiferas-McNaughton 1976)

A function $f : \mathbb{N} \to \mathbb{N}$ is regularity-preserving iff it is cyclically ultimately periodic and, for every $k \in \mathbb{N}$, the set $f^{-1}(k)$ is regular.

Regularity-preserving functions from $\mathbb{N}$ to $\mathbb{N}$

[Siefkes 1970]

- Every polynomial function
- $n \to 2^n$
- $n \to n!$
- $n \to 2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$ (exponential stack of 2's of height n)

[Carton-Thomas 02]

- $n \to F_n$ (Fibonacci number)
- $n \to t_n$, where t_n is the prefix of length n of the Prouhet-Thue-Morse sequence.

Counterexamples [Siefkes 1970]

- $n \to \lfloor \sqrt{n} \rfloor$ **is not** cyclically ultimately periodic and hence **not** regularity-preserving.
- $n \to \binom{2n}{n}$ **is not** ultimately periodic modulo 4 and hence **not** regularity-preserving. Indeed

$$\binom{2n}{n} \bmod 4 = \begin{cases} 2 & \text{if } n \text{ is a power of } 2, \\ 0 & \text{otherwise.} \end{cases}$$

Open problem?

- Is the function $n \to p_n$ regularity-preserving? (p_n is the n-th prime number).

Closure properties

Theorem (Siefkes 70, Zhang 98, Carton-Thomas 02)

Let $f, g : \mathbb{N} \to \mathbb{N}$ be cyclically ultimately periodic functions. Then so are the following functions:

(1) $g \circ f$, $f + g$, fg, f^g, *and* $f - g$ *provided that* $f \geqslant g$ *and* $\lim_{n\to\infty} (f - g)(n) = +\infty$,

(2) *(generalised sum)* $n \to \sum_{0 \leqslant i \leqslant g(n)} f(i)$,

(3) *(generalised product)* $n \to \prod_{0 \leqslant i \leqslant g(n)} f(i)$.

Connections with logic

A function $f : \mathbb{N} \to \mathbb{N}$ is effectively regularity-preserving if, for each given regular subset of $\mathbb{N}$, $f^{-1}(R)$ is regular and effectively computable.

Recall that $\Delta f(n) = f(n+1) - f(n)$.

Theorem (Carton-Thomas 02)

Let χ_P be the characteristic function of a predicate $P \subseteq \mathbb{N}$. If $\Delta\chi_P$ is effectively regularity-preserving, then the monadic second order theory $\mathrm{MTh}(\mathbb{N}, <, P)$ is decidable.

Recursivity

Let $f : \mathbb{N} \to \{0, 1\}$ be a non-recursive function. Then the function $n \to (\sum_{0 \leqslant i \leqslant n} f(i))!$ is regularity-preserving but non-recursive.

Open problem. Is it possible to describe all recursive regularity-preserving functions, respectively all recursive cyclically ultimately periodic functions?

One could try to use Siefkes' primitive recursion scheme (1970).

Siefkes' recursion scheme

Theorem

Let $g : \mathbb{N}^k \to \mathbb{N}$ and $h : \mathbb{N}^{k+2} \to \mathbb{N}$ be cyclically ultimately periodic functions satisfying three technical conditions. Then the function f defined from g and h by primitive recursion, i.e.

$$f(0, x_1, \ldots, x_k) = g(x_1, \ldots, x_k),$$
$$f(n+1, x_1, \ldots, x_k) = h(n, x_1, \ldots, x_k, f(n, x_1, \ldots, x_k))$$

is cyclically ultimately periodic.

The three technical conditions

(1) h is cyclically ultimately periodic in x_{k+2} of decreasing period,

(2) g is essentially increasing in x_k,

(3) for all $x \in \mathbb{N}^{k+2}$, $x_{k+2} < h(x_1, \ldots, x_{k+2})$.

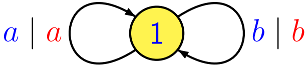
A function f is essentially increasing in x_j iff, for all $z \in \mathbb{N}$, there exists $y \in \mathbb{N}$ such that for all $x \in \mathbb{N}^n$, $y \leqslant x_j$ implies $z \leqslant f(x_1, \ldots, x_n)$.

A function f is c.u.p. of decreasing period in x_j iff, for all p, the period of the function $f \bmod p$ in x_j is $\leqslant p$.

Part III

Matrix representations

Matrix representations



$$\mu(a) = a \quad \mu(b) = b \quad \mu(u) = u$$

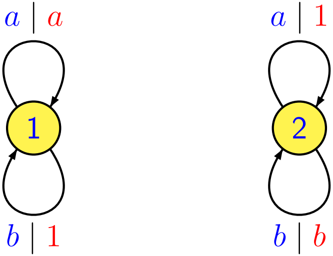
$$f_1(u) = uu \qquad f_1(u) = (\mu(u))^2$$

$$f_2(u) = uau^2 \qquad f_2(u) = \mu(u)a\mu(u)^2$$

$$\tau_1(u) = u^* \qquad \tau_1(u) = \sum_{n \geqslant 0} \mu(u)^n$$

$$\tau_2(u) = \bigcup_{p \text{ prime}} u^p \qquad \tau_2(u) = \sum_{p \text{ prime}} \mu(u)^p$$

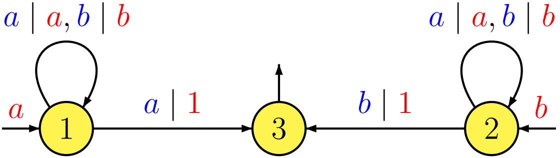
$f(u) = a^{|u|_a} b^{|u|_b}$



$$\mu(a) = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \quad \mu(b) = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix} \quad \mu(u) = \begin{pmatrix} a^{|u|_a} & 0 \\ 0 & b^{|u|_b} \end{pmatrix}$$

$$f(u) = \mu_{1,1}(u)\mu_{2,2}(u)$$

$f(u) = \mathrm{Last}(u)u$



$$\mu(a) = \begin{pmatrix} a & 0 & 1 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mu(b) = \begin{pmatrix} b & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mu(ua) = \begin{pmatrix} ua & 0 & u \\ 0 & ua & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \mu(ub) = \begin{pmatrix} ub & 0 & 0 \\ 0 & ub & u \\ 0 & 0 & 0 \end{pmatrix}$$

$$f(u) = a\mu_{1,3}(u) + b\mu_{2,3}(u)$$

Matrix representations

A transduction $\tau\colon A^* \to M$ admits a matrix representation (S, μ) of degree n if there exist a monoid morphism $\mu\colon A^* \to \mathcal{P}(M)^{n\times n}$ and a possibly infinite union of products S involving arbitrary subsets of M and n^2 variables $X_{1,1}, \ldots, X_{n,n}$, such that, for all $u \in A^*$,

$$\tau(u) = S[\mu_{1,1}(u), \ldots, \mu_{n,n}(u)].$$

Example for $n = 2$: Let $(P_n)_{n\geqslant 0}$ be subsets of M.

$$S = \bigcup_{n\in\mathbb{N}} P_0 X_{1,1}^n P_n X_{2,1} X_{1,1}^n X_{2,2} P_{n!} X_{1,1} P_{2n}$$

Matrix representation of transducers

Theorem (Pin-Sakarovitch 1983)

Let (S, μ) be a matrix representation of degree n of a transduction $\tau\colon A^* \to M$. Let P be a subset of M recognised by a morphism $\eta\colon M \to N$. Then the language $\tau^{-1}(P)$ is recognised by the submonoid $\eta\mu(A^*)$ of the monoid of matrices $\mathcal{P}(N)^{n\times n}$.

Corollary

Every transduction having a matrix representation is regularity-preserving.

Marseilles transductions

aka streaming string transducers, HDTOL

A substitution $\sigma : A^* \to B^*$ is a monoid morphism from A^* to $\mathcal{P}(B^*)$.

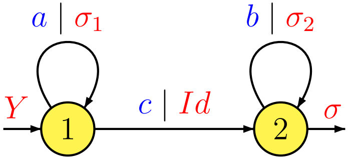
A Marseilles transducer is a sequential transducer whose outputs are substitutions.

Proposition (Pin, Reynier, Villevallois, 2018)

Marseilles transductions are regularity-preserving.

Marseilles transducers

The function $f(a^n c b^p) = a^p b^{pn}$ can be realized by the following Marseilles transducer:

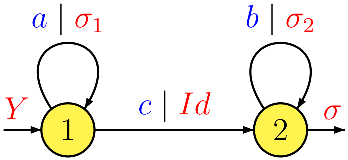


where $A = \{a, b, c\}$, $B = A \cup \{X, Y\}$ and $\sigma, \sigma_1, \sigma_2 : B^* \to B^*$ are substitutions defined by

$$\begin{array}{lll} X\sigma_1 = X & Y\sigma_1 = YX & d\sigma_1 = d \text{ for } d \in A \\ X\sigma_2 = Xb & Y\sigma_2 = Ya & d\sigma_2 = d \text{ for } d \in A \\ X\sigma = 1 & Y\sigma = 1 & d\sigma = d \text{ for } d \in A \end{array}$$

Marseilles transducers at work

The function $f(a^n c b^p) = a^p b^{pn}$ can be realized by the following Marseilles transducer:



$$\tau(a^n c b^p) = Y \sigma_1^n \sigma_2^p \sigma = (Y X^n) \sigma_2^p \sigma = ((Y \sigma_2^p)(X \sigma_2^p)^n) \sigma$$
$$= ((Y a^p)(X b^p)^n) \sigma = a^p b^{pn}$$

$X\sigma_1 = X$	$Y\sigma_1 = YX$	$d\sigma_1 = d$ for $d \in A$
$X\sigma_2 = Xb$	$Y\sigma_2 = Ya$	$d\sigma_2 = d$ for $d \in A$
$X\sigma = 1$	$Y\sigma = 1$	$d\sigma = d$ for $d \in A$

Part IV

Topology

Residually finite monoids

A monoid F separates two elements $x, y \in M$ if there exists a morphism $\varphi : M \to F$ such that $\varphi(x) \neq \varphi(y)$.

A monoid is residually finite if any pair of distinct elements of M can be separated by a finite monoid.

Finite monoids, free monoids, free groups are residually finite. The monoids $A_1^* \times A_2^* \times \cdots \times A_n^*$ are residually finite.

Profinite metric

Let M be a residually finite monoid. The profinite metric d is defined by setting, for $u, v \in M$:

$$r(u, v) = \min\Big\{|F| \;\Big|\; F \text{ separates } u \text{ and } v\Big\}$$
$$d(u, v) = 2^{-r(u,v)}$$

with the conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$. Then

$$d(u, w) \leqslant \max(d(u, v), d(v, w)) \quad \text{(ultrametric)}$$
$$d(uw, vw) \leqslant d(u, v)$$
$$d(wu, wv) \leqslant d(u, v)$$

Recognisable subsets of a monoid

A subset P of a monoid M is recognizable if there exists a finite monoid F, a monoid morphism $\varphi : M \to F$ and a subset Q of F such that $P = \varphi^{-1}(Q)$.

A function $f : M \to N$ is recognizability-preserving if, for each recognizable subset R of N, $f^{-1}(R)$ is recognizable in M.

Same definition for recognizability-preserving transductions.

Recognizability-preserving functions

Let M and N be two finitely generated, residually finite monoids.

Theorem (Pin-Silva 2005)

A function $M \to N$ is recognizability-preserving iff it is uniformly continuous for the profinite metrics.

What about recognizability-preserving transductions?

Example

Proposition (Pin-Silva 2005)

The function $\tau : M \times \mathbb{N} \to M$ defined by $\tau(x, n) = x^n$ is recognizability-preserving.

Corollary. The function $u \to u^{|u|}$ is recognizability-preserving. Indeed it can be decomposed as

$$\begin{aligned} A^* &\to A^* \times \mathbb{N} \\ u &\to (u, |u|) \end{aligned} \qquad \qquad \begin{aligned} A^* \times \mathbb{N} &\to A^* \\ (u, n) &\to u^n \end{aligned}$$

Another example

Let $\tau_n \colon A^* \to (A^*)^n$ be defined by

$$\tau_n(u) = \{(u_1, \ldots, u_n) \mid u_1 \cdots u_n = u\}$$

Then both τ_n and τ_n^{-1} are recognizability-preserving.

Completion

Let M be a finitely generated, residually finite monoid. Let $\widehat{M}$ be the completion of the metric space (M, d).

Proposition

$\widehat{M}$ *is a compact monoid.*

Hausdorff metric

Let (M, d) be a compact metric monoid. Then the set $\mathcal{K}(M)$ of compact subsets of M is also a compact monoid for the Hausdorff metric.

The Hausdorff metric on $\mathcal{K}(M)$ is defined as follows. For $K, K' \in \mathcal{K}(M)$, let

$$\begin{aligned} \delta(K, K') &= \sup_{x \in K} d(x, K') \\ h(K, K') &= \max(\delta(K, K'), \delta(K', K)) \\ &\quad + \text{special definition if } K \text{ or } K' \text{ is empty} \end{aligned}$$

Back to transductions

Let M and N be two finitely generated, residually finite monoids and let $\tau : M \to N$ be a transduction.

Define a map $\widehat{\tau} : M \to \mathcal{K}(\widehat{N})$ by setting, for each $x \in M$, $\widehat{\tau}(x) = \overline{\tau(x)}$.

Theorem (Pin-Silva 2005)

The transduction τ is recognizability preserving iff $\widehat{\tau}$ is uniformly continuous.

Part V

p-group languages

p-group languages

Let p be a prime number. A p-group is a group in which every element has order a power of p.

Target class: $\mathcal{G}_p$, the class of languages recognized by a finite p-group.

Goal. Characterization of $\mathcal{G}_p$-preserving functions.

Separation by p-groups

Let u and v be two words of A^*. A p-group G separates u and v if there is a monoid morphism φ from A^* onto G such that $\varphi(u) \neq \varphi(v)$.

Proposition

Any pair of distinct words can be separated by a finite p-group.

Pro-p metric

Let u and v be two words. Put

$$r_p(u,v) = \min\Big\{|G| \;\Big|\; G \text{ is a } p\text{-group that separates } u \text{ and } v\Big\}$$

$$d_p(u,v) = p^{-r_p(u,v)}$$

with the usual convention $\min\emptyset = -\infty$ and $p^{-\infty} = 0$. Then d_p is an ultrametric:

(1) $d_p(u,v) = 0$ if and only if $u = v$,

(2) $d_p(u,v) = d_p(v,u)$,

(3) $d_p(u,v) \leqslant \max(d_p(u,w), d_p(w,v))$

Binomial coefficients (see Eilenberg or Lothaire)

Let u and $v = a_1 \cdots a_n$ be two words of A^*. Then v is a subword of u if there exist $u_0, \ldots, u_n \in A^*$ such that $u = u_0 a_1 u_1 \cdots u_{n-1} a_n u_n$ (the u_i's might be empty words).

The binomial coefficient $\binom{u}{v}$ is the number of times that v appears as a subword of u.

$$abab, abab, abab. \text{ Thus } \binom{abab}{ab} = 3.$$

If $u = a^n$ and $v = a^m$, then $\binom{u}{v} = \binom{n}{m}$.

An equivalent metric

Let us set

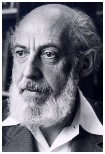
$$r'_p(u,v) = \min \left\{ |x| \;\middle|\; \binom{u}{x} \not\equiv \binom{v}{x} \pmod p \right\}$$

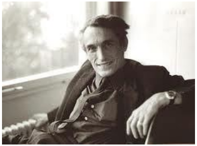
$$d'_p(u,v) = p^{-r'_p(u,v)}$$

Proposition

d'_p *is an ultrametric uniformly equivalent to* d_p.

p-recognisable languages





Theorem (Eilenberg-Schützenberger 1976)

A language is recognized by a finite p-group iff it is a finite Boolean combination of the languages

$$L(x, r, p) = \left\{ u \in A^* \mid \binom{u}{x} \equiv r \bmod p \right\},$$

for $0 \leqslant r < p$ and $x \in A^*$.

The noncommutative difference operator

Let $f : A^* \to F(B)$ be a function. For each letter a, the difference operator $\Delta^a f : A^* \to F(B)$ by

$$(\Delta^a f)(u) = f(u)^{-1} f(ua)$$

The operator $\Delta^w f : A^* \to F(B)$ is defined for each word $w \in A^*$ by setting $\Delta^1 f = f$, and for each letter $a \in A$ and each word $w \in A^*$,

$$\Delta^{aw} f = \Delta^a(\Delta^w f)$$

In fact, for all $v, w \in A^*$, $\Delta^{vw} f = \Delta^v(\Delta^w f)$

Taking $u = 1$

For $w \in A^*$, let $\delta_w f = (\Delta^w f)(1)$. Then

$$
\begin{aligned}
\delta_1 f &= f(1) \\
\delta_a f &= f(1)^{-1} f(a) \\
\delta_{aa} f &= f(a)^{-1} f(1) f(a)^{-1} f(aa) \\
\delta_{baa} f &= f(aa)^{-1} f(a) f(1)^{-1} f(a) f(ba)^{-1} f(b) \\
&\quad f(ba)^{-1} f(baa) \\
\delta_{abaa} f &= f(baa)^{-1} f(ba) f(b)^{-1} f(ba) f(a)^{-1} f(1) \\
&\quad f(a)^{-1} f(aa) f(aaa)^{-1} f(aa) f(a)^{-1} \\
&\quad f(aa) f(aba)^{-1} f(ab) f(aba)^{-1} f(abaa)
\end{aligned}
$$

$\mathcal{G}_p$-preserving functions

Theorem (Pin-Reutenauer 2018)

Let f be a function from A^* to B^*. Are equivalent:

(1) f *is $\mathcal{G}_p$-preserving,*

(2) f *is uniformly continuous for d_p (or d'_p),*

(3) $\lim_{|u|\to\infty} d_p(\delta_u f, 1) = 0$,

Happy birthday Werner!