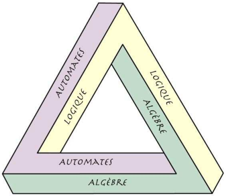
Regularity preserving transductions

Jean-Éric Pin

IRIF, CNRS and University Paris 7



March 2018, Paris

Outline

(1) Motivation
(2) A series of examples
(3) Matrix representation of transducers
(4) Topological characterizations
(5) Subwords and p-group languages

Part I

Motivation

Original motivation

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 $\mathcal{C}$.

Goal. Find a complete description of regularity-preserving [$\mathcal{C}$-preserving] functions.

Same questions for transductions.

A quizz (J.-E. Pin, P.A. Reynier, D. Villevallois)

Which of these functions are regularity-preserving? Star-free preserving?

$$u \to u^2 \qquad u \to \tilde{u}u$$

$$u \to u^{|u|} \qquad u \to a^{|u|_a} b^{|u|_b}$$

$$a^m c b^n \to a^n b^{mn}$$

$$a^n \to a^{2^n} \qquad a^n \to \underbrace{a^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{n \text{ times}}$$

$$u_0 \# u_1 \# u_2 \to u_2 \# u_1 \# u_0 \# u_1 \# u_2$$

$$u \# v \to (v[a \to u])^{|u|}$$

$$u \# baaba \to (buubu)^{|u|}$$

A quizz on transductions (Pin-Sakarovitch 1983)

Which transductions are regularity-preserving?
Star-free preserving?

$$u \to u^*$$

$$u \to \bigcup_{p \text{ prime}} u^p$$

$$u \to A^{|u|} u A^{|u|}$$

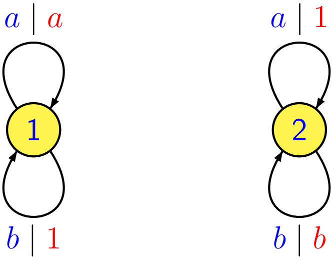
$$u \to \{\tilde{v}v \mid v \text{ is a subword of } u\}$$

Part II

Matrix representations

Matrix representation of transducers [Pin-Sakarovitch]

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



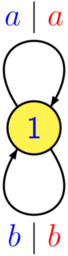
$$\mu(a) = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \quad \mu(b) = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix} \quad \mu(u) = \begin{pmatrix} a^{|u|_a} & 0 \\ 0 & b^{|u|_b} \end{pmatrix}$$

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

Example 2

$$f_1(u) = uu \qquad f_2(u) = \tilde{u}u$$

$$\tau_1(u) = u^* \qquad \tau_2(u) = \bigcup_{p \text{ prime}} u^p$$



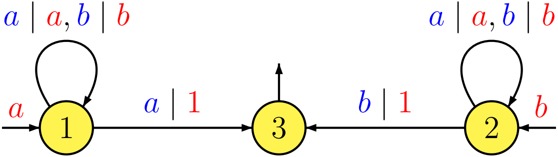
$$\mu(a) = a \quad \mu(b) = b \quad \mu(u) = u$$

$$f_1(u) = (\mu(u))^2 \qquad f_2(u) = \tilde{\mu}(u)\mu(u)$$

$$\tau_1(u) = \sum_{n \geqslant 0} \mu(u) \qquad \tau_2(u) = \sum_{p \text{ prime}} \mu(u)^p$$

Example 3

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

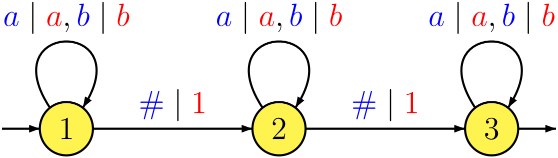


$$\mu(a) = \begin{pmatrix} a & 0 & 1 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mu(b) = \begin{pmatrix} b & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

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

Example 4

$$f(u_0\#u_1\#u_2) = u_2\#u_1\#u_0\#u_1\#u_2$$



$$\mu(\#) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \mu(u_0\#u_1\#u_2) = \begin{pmatrix} u_0 & 0 & 0 \\ 0 & u_1 & 0 \\ 0 & 0 & u_2 \end{pmatrix}$$

$$f(u) = \mu_{3,3}(u)\#\mu_{1,1}(u)\#\mu_{1,1}(u)\#\mu_{2,2}(u)\#\mu_{3,3}(u)$$

Matrix representation of transducers

Theorem (Pin-Sakarovitch 1983)

Every transduction having a matrix representation is regularity-preserving.

Formal definition. Let M be a monoid. A transduction $\tau : A^* \to M$ admits a matrix representation (s, μ) if there exist $n > 0$, a monoid morphism $\mu : A^* \to \mathcal{P}(M)^{n \times n}$, and a series s in n^2 variables $X_{i,j}$ such that for all $u \in A^*$, $\tau(u) = s[X_{i,j} \to \mu_{i,j}(u)]$.

Other results

A transduction $\tau : A^* \to M$ admits a bilateral matrix representation (s, μ) if there exist $n > 0$, a monoid morphism $\mu : A^* \to \mathcal{P}(M)^{n \times n}$, and a series s in $2n^2$ variables $X_{i,j}, \tilde{X}_{i,j}$ such that for all $u \in A^*$, $\tau(u) = s[X_{i,j} \to \mu_{i,j}(u), \tilde{X}_{i,j} \to \tilde{\mu}_{i,j}(u)]$.

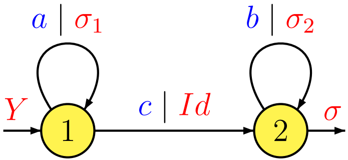
Every transduction having a bilateral matrix representation is regularity-preserving.

Proposition (Pin, Reynier, Villevallois, 2018)

Marseille transducers are regularity-preserving.

Marseille transducers

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

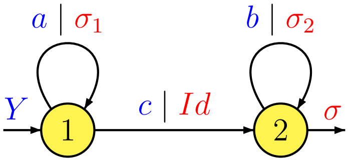


where $A = \{a, b, c\}$, $B = A \cup \{X, Y\}$ and $\sigma, \sigma_1, \sigma_2 : B^* \to B^*$ are substitutions defined by

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

Marseille transducers at work

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



$$\tau(a^n cb^p) = Y\sigma_1^n \sigma_2^p \sigma = (YX^n)\sigma_2^p \sigma = ((Y\sigma_2^p)(X\sigma_2^p)^n)\sigma$$
$$= ((Ya^p)(Xb^p)^n)\sigma = a^p b^{pn}$$

$X\sigma_1 = X$	$Y\sigma_1 = YX$	$d\sigma_1 = d$ 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 III

Topological characterizations

Monoids

A monoid is a set M equipped with an associative binary operation (the product) and an identity 1 for this operation.

A monoid M is finitely generated if there exists a finite subset F of M which generates M.

Examples. The free monoid A^*, with A finite.

Given a monoid M, the set $\mathcal{P}(M)$ of subsets of M is a monoid under the product defined, for $X, Y \subseteq M$, by $XY = \{xy \mid x \in X, y \in Y\}$.

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 $P = \varphi^{-1}(Q)$.

$\mathrm{Rec}(M)$ = set of recognizable subsets of M .

If $M = A^*$, then recognizable = rational = regular [Kleene].

Transductions

Given two monoids M and N, a transduction from M into N is a relation on M and N, usually viewed as a map from M into the monoid $\mathcal{P}(N)$.

If $\tau : M \to N$ is a transduction, then the inverse relation $\tau^{-1} : N \to M$ is also a transduction. If $R \subseteq N$, then

$$\tau^{-1}(R) = \{x \in M \mid \tau(x) \cap R \neq \emptyset\}$$

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

Same definition for recognizability-preserving transductions.

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 $u, v \in M$:

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

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

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

Uniform continuity

Let (X_1, d_1) and (X_2, d_2) be metric spaces.

A function $f : X_1 \to X_2$ is uniformly continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in X_1$,

$$d_1(x, y) < \delta \implies d_2(f(x), f(y)) < \varepsilon.$$

Recognizability-preserving functions

Let M and N be two finitely generated, residually finite monoids. (For instance $M = A^*$ and $N = B^*$).

Theorem (Pin-Silva 2005)

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

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

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 $n > 0$.

Theorem (Siefkes 1970, SeiferasMcNaughton 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. It is regularity-preserving iff it is cyclically ultimately periodic and, for every $k \in \mathbb{N}$, the set $f^{-1}(k)$ is regular.

Two examples

Theorem (Siefkes 1970)

The functions $n \to 2^n$ *and* $n \to \underbrace{2^{\cdot^{\cdot^{\cdot^{2}}}}}_{n \text{ times}}$ *are* *regularity-preserving*.

Two counterexamples

The function $n \to \binom{2n}{n}$ **is not** regularity-preserving. Indeed

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

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 (using some recursion scheme as in Siefkes 1970)?

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

Another result

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

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

Recognizability-preserving transductions

Let M and N be two finitely generated, residually finite monoids. (For instance $M = A^*$ and $N = B^*$).

Theorem

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

What about recognizability-preserving transductions?

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 IV

p-group languages

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

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

p-groups

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

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 metrics

Let u and v be two words. Put

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

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

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

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.

Binomial coefficients (see Eilenberg or Lothaire)

Let $x = a_1 \cdots a_n$ and u be two words of A^*. The binomial coefficient of u and x is

$$\binom{u}{x} = |\{(v_0, \ldots, v_n) \mid v = v_0 a_1 v_1 \ldots a_n v_n\}|$$

If a is a letter, then $\binom{u}{a} = |u|_a$. If $u = a^n$ and $x = a^m$, then $\binom{u}{x} = \binom{m}{n}$.

$$\binom{abab}{a} = 2 \quad \binom{abab}{b} = 2 \quad \binom{abab}{ab} = 3 \quad \binom{abab}{ba} = 1$$

Languages recognized by a p-group

A language recognized by a finite p-group is called a p-group language.

Theorem (Eilenberg-Schützenberger 1976)

A language of A^* is a p-group language iff it is a Boolean combination of the languages

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

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^*. TFCAE:

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

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

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