Context-free sequences & context-free numbers

Marion Le Gonidec, Université de la Réunion

Table of contents

Characterizations of automatic sequences

$x \in A^\omega$ is k-automatic if :

- there exists an finite automaton s.t. x_n is the output of the automaton feeding with $[n]_k$;
- fibers $Fib(x, a) = \{[n]_k,\, x_n = a\}$ are rational languages ;
- x is generated by a k-unifom tag-machine (substitution) ;
- The k-kernel of x is finite ;

Operations

sequence $x = x_0 x_1 \ldots x_n \ldots$ over A

left shift : $S(x) = x_1 \ldots x_n \ldots$

right shift by letter a : $a x_0 x_1 \ldots x_n \ldots$

p-morphism $\sigma : A \longrightarrow A^p$

$$\sigma(x) = \sigma(x_0)\sigma(x_1) \ldots \sigma(x_n) \ldots$$

p-substitution $\sigma : A \longrightarrow 2^{A^p}$

injective : $\sigma(a) \cap \sigma(b) = \emptyset$ for any letters $a \neq b$

applied by inverse :

$$\sigma^{-1}(x) = y_0 y_1 \ldots y_n \ldots \quad \text{with} \quad x_{np} \ldots x_{(n+1)p-1} \in \sigma(y_n)$$

product : $x \times y = (x_0 \times y_0) \ldots (x_n \times y_n) \ldots$

Robustness of the notion

The set of k-automatic sequences is closed under

- finite modifications
- product
- left and right shifts
- uniform morphism
- inverse injective uniform substitution
- Various regular extractions of k-automatic sequences are ultimately periodic

contains ultimately periodic sequences $u\, v\, v \ldots v \ldots$

Real numbers with automatic *b*-adic expansion

Let α and β 2 real number whose *b*-adic expansion is *k*-automatic.

- $\alpha + \beta$ is *k*-automatic
- $r\alpha$ for $r \in \mathbb{Q}$ ins *k*-automatic.

Some other significant results about k-automatic sequences :

- (Cobham 69) Any sequence which is both k- and p-automatic with k and p multiplicatively independent : $k^i \neq p^j$ for any $i, j > 0$, is ultimately periodic
- (Cobham 72) The factor complexity of automatic sequences is either ultimately constant or linear
  The Kolmogorov complexity of their n-prefixes is in $\log n$
- (Christol & Al. 80) If a sequence x over A is k-automatic if there exists a finite field $\mathbb{K}$ with caracteristic k and an injection $\imath : A \to \mathbb{K}$ such that the formal serie $\sum_{n \in \mathbb{N}} \imath(x_n) X^n$ is algebraic over $\mathbb{K}$.
- (Adamczewsky-Bugeaud 07) The k-adic expansion of any algebraic irrational number is not an automatic sequence

Several extensions of the concept

- using ~~k-uniform~~ tag machines/substitutions : substitutive sequences
- using finitely generated k-kernel : regular sequences
- using next step in Chomsky hierarchy : context-free sequences

Pushdown automaton

- states Q,
- stack symbols $P \cup \{\#\}$
- pushdown rule : $X\,p \xrightarrow{a} U\,q$ with

$$X \in P\ ,\ \ U \in P^*\ ,\ \ p,q \in Q\ ,\ \ a \in T$$

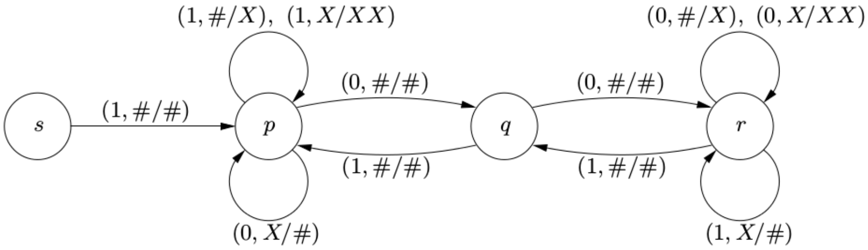
  noted on the graph by (a, X/U)
- initial configuration : $c_0 \in P^*Q$
- final configurations : $F \in Rat(P^*Q)$

Transition : $V\,X\,p \xrightarrow{a} V\,U\,q$ with $V \in P^*$

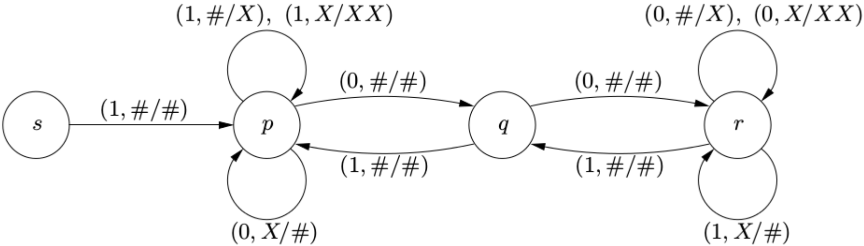
- states $Q = \{s, p, q, r\}$,
- stack symbols $\{X\} \cup \{\#\}$
- pushdown rule : $\mathrm{X\,p} \xrightarrow{\mathbf{a}} \mathrm{U\,q}$ with

$$\mathrm{X} \in \mathrm{P}\ ,\ \ \mathrm{U} \in \mathrm{P}^*\ ,\ \ \mathrm{p,q} \in \mathrm{Q}\ ,\ \ \mathrm{a} \in \mathrm{T}$$

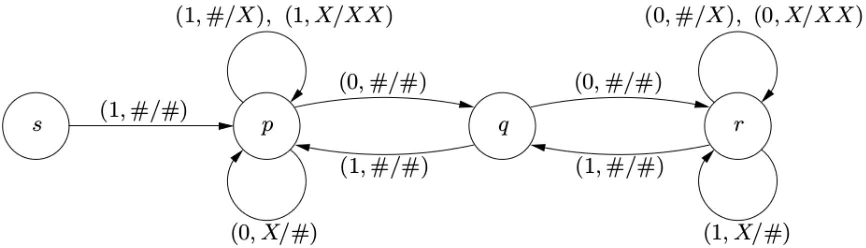
  noted on the graph by (a, X/U)
- initial configuration : $\#$s
- final configurations : F $= \{\#q\}$



Pushdown automata & Context-free languages



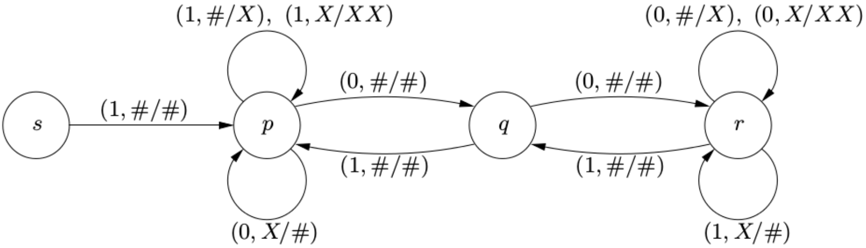
Ex : reading 1111000001 gives the following moves :



Ex : reading 1111000001 gives the following moves :

$\#s$

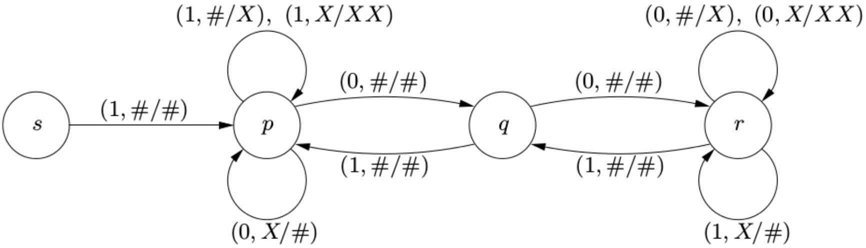
Pushdown automata & Context-free languages



Ex : reading 1111000001 gives the following moves :

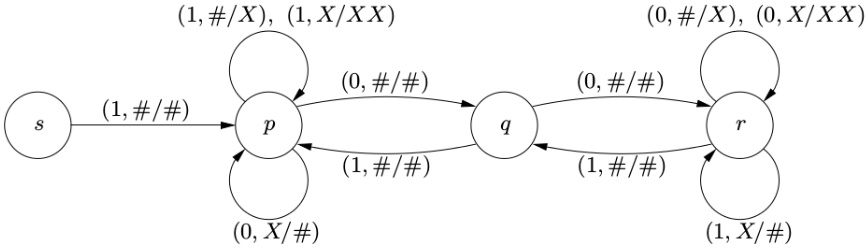
$\#s \xrightarrow{1} \#p$

Pushdown automata & Context-free languages



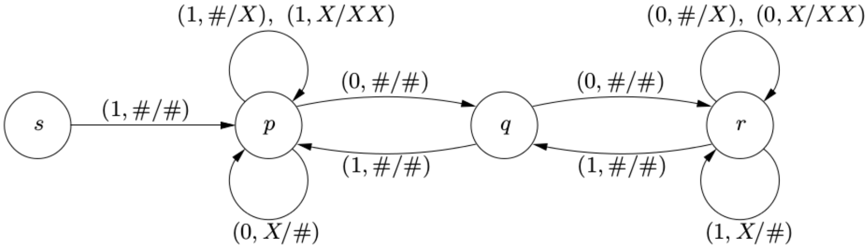
Ex : reading 1111000001 gives the following moves :

$\# s \xrightarrow{1} \# p \xrightarrow{1} Xp$



Ex : reading 1111000001 gives the following moves :

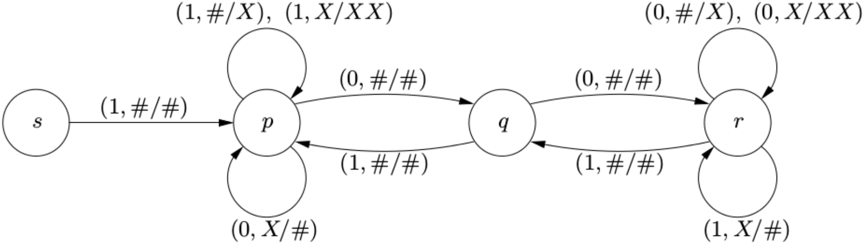
$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp$



Ex : reading 1111000001 gives the following moves :

$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp$

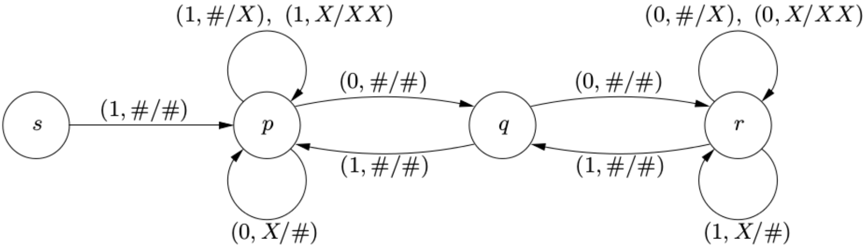
Pushdown automata & Context-free languages



Ex : reading 1111000001 gives the following moves :

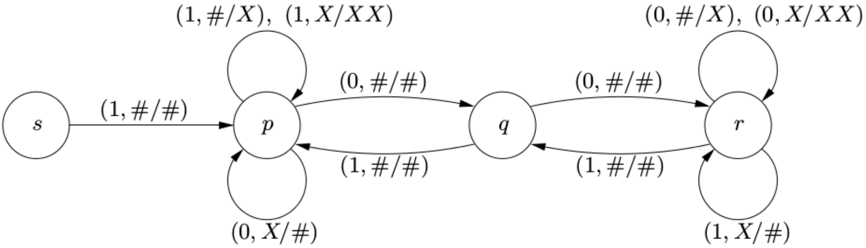
$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp$

Pushdown automata & Context-free languages



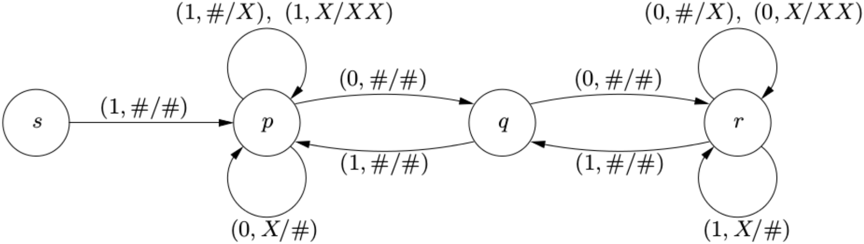
Ex : reading 1111000001 gives the following moves :

$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} X{\color{blue}Xp} \xrightarrow{{\color{blue}0}} X{\color{red}p}$



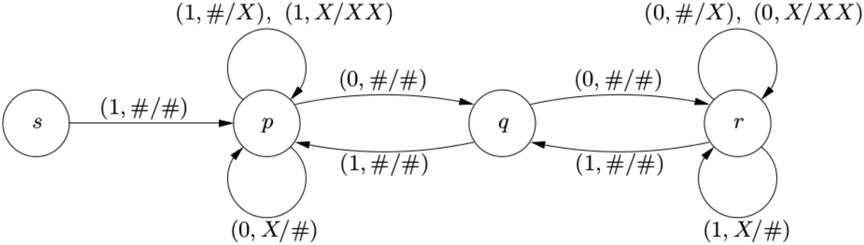
Ex : reading 1111000001 gives the following moves :

$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp \xrightarrow{0} Xp \xrightarrow{0} \#p$



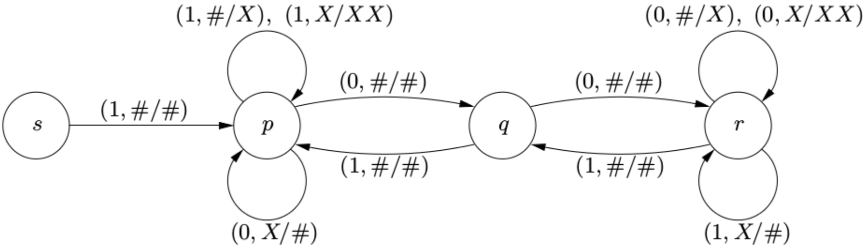
Ex : reading 1111000001 gives the following moves :

$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp \xrightarrow{0} Xp \xrightarrow{0} \#p \xrightarrow{0} \#q$



Ex : reading 1111000001 gives the following moves :

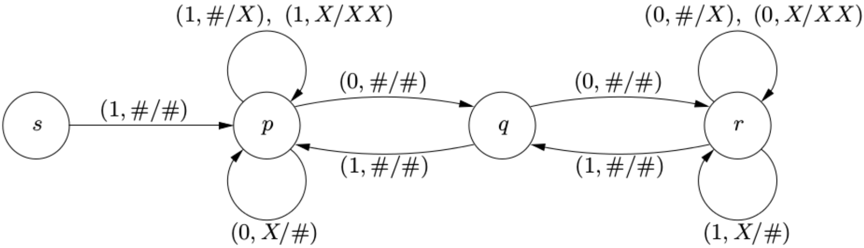
$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp \xrightarrow{0} Xp \xrightarrow{0} \#p \xrightarrow{0} \#q \xrightarrow{0} \#r$



Ex : reading 1111000001 gives the following moves :

$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp \xrightarrow{0} Xp \xrightarrow{0} \#p \xrightarrow{0} \#q \xrightarrow{0}$
$\#r \xrightarrow{1} \#q$

Pushdown automata & Context-free languages



Ex : reading 1111000001 gives the following moves :
$\#s \xrightarrow{1} \#p \xrightarrow{1} Xp \xrightarrow{1} XXp \xrightarrow{1} XXXp \xrightarrow{0} XXp \xrightarrow{0} Xp \xrightarrow{0} \#p \xrightarrow{0} \#q \xrightarrow{0}$
$\#r \xrightarrow{1} \#q$

Context-free languages

Context-free language : $\{ u \in T^* \mid \exists c \in F \, (c_0 \xrightarrow{u} c) \}$

(real-time) deterministic context-free language :

$$(Xp \xrightarrow{a} Uq \ \wedge \ Xp \xrightarrow{a} U'q') \implies Uq = U'q'$$

k-context-free sequences

Definition Hamm 98

A sequence x is k-context-free if for each letter a

$$\mathrm{Fib}_k(x,a) = \{ \mathrm{Rep}_k(n) \mid x_n = a \} \text{ is a context-free language}$$

Example : $A = \{a, b\}$ and $k = 2$

The sequence x defined by

$$\mathrm{Fib}_k(x,a) = \{ 1u \mid u = \widetilde{u} \}$$

$$\mathrm{Fib}_k(x,b) = \{ 1u \mid u \neq \widetilde{u} \} \cup \{\varepsilon\}$$

is context-free

$x =$	b	a	a	a	a	b	b	a	a	b	$\ldots$
	ε	1	10	11	100	101	110	111	1000	1001	

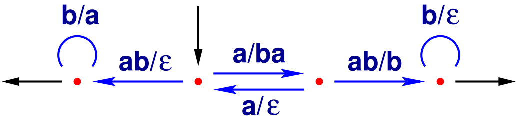
Stability properties

The family $\mathrm{Cf}_k(A^\omega)$ of k-context-free sequences is closed under

- finite modifications
- product with any k-automatic sequence
- left and right shifts
- uniform morphism
- ~~inverse injective k-substitution~~

Transformations realized by finite transducers

Finite transducer



Rational relation R : aabb $\longmapsto$ bab ; aaabbb $\longmapsto$ baaa

$$R(\{ a^n b^n \mid n \geq 0 \}) = R(\{ a^{2n} b^{2n} \mid n > 0 \}) \cup R(\{ a^{2n+1} b^{2n+1} \mid n \geq 0 \})$$
$$= \{ (ba)^n b \mid n > 0 \} \cup \{ (ba)^n a^{2n} \mid n \geq 0 \}$$

Stability property

The context-free languages are preserved by rational relation

corresponds to the preservation by

morphism, inverse morphism, intersection with a regular language

Example

$$L \;=\; 1^*\{\,0^n1^n \mid n>0\,\} \quad ; \quad M \;=\; \{\,1^n0^n \mid n>0\,\}1^*$$

2-context-free sequence x over $\{a,b,c,d\}$ defined by

$$\mathrm{Fib}_2(x,a) \;=\; L.0 \quad ; \quad \mathrm{Fib}_2(x,c) \;=\; \{\varepsilon\} \cup (1\{0,1\}^* - L)0$$

$$\mathrm{Fib}_2(x,b) \;=\; M.1 \quad ; \quad \mathrm{Fib}_2(x,d) \;=\; \{1\} \cup (1\{0,1\}^* - M)1$$

these fibers are (real-time) deterministic context-free languages

injective 2-substitution h defined by

$$h(a) \;=\; ab \quad ; \quad h(b) \;=\; \{a,b,c,d\}^2 - \{ab\}$$

$h^{-1}(x)$ is not a 2-context-free sequence :

$$\mathrm{Fib}_2(h^{-1}(x),a) \;=\; \{\,1^n0^n1^n \mid n>0\,\}$$ not context-free

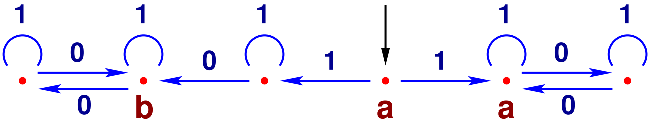
k-graph G over A

- edge labels : $0, \ldots, k-1$
- vertex labels : $a \in A$

  $\mathrm{L}(G, a)$ labels of the initial paths ending to an a-vertex
- $\mathrm{L}(G, a)_{a \in A}$ partition of $\{0, \ldots, k-1\}^*$ not beginning by 0

G recognizes the sequence x defined by $\mathrm{Fib}_k(x, a) = \mathrm{L}(G, a)$

$$x_n = a \iff \mathrm{Rep}_k(n) \in \mathrm{L}(G, a)$$



$$\mathrm{L}(G, a) = 1^+(01^*01^*)^* \cup \{\varepsilon\} \quad ; \quad \mathrm{L}(G, b) = 1^+01^*(01^*01^*)^*$$

k-regular graph

Property

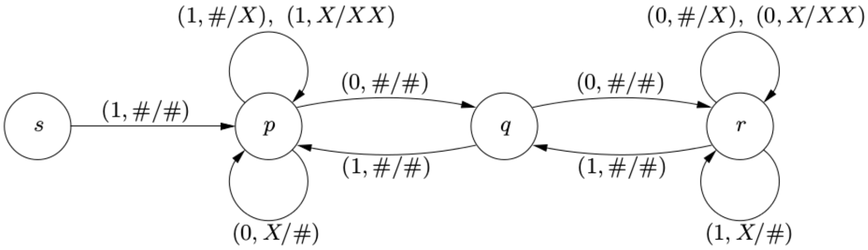
A sequence is context-free if and only if

it is recognized by a k-regular graph of finite degree

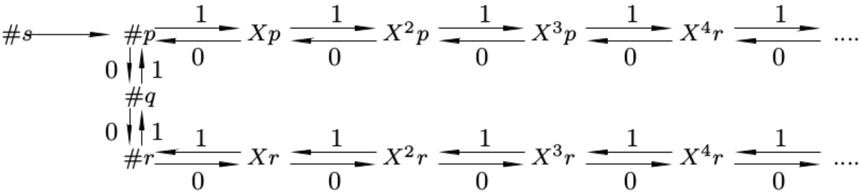
$$
\begin{aligned}
& \text{regular graph of finite degree} \\
= \;& \text{transition graph of a pushdown automaton} \\
= \;& \text{graph finitely decomposable by distance}
\end{aligned}
$$

Pushdown automata and regular graphs

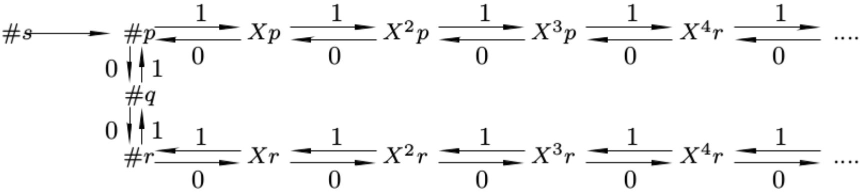
- pushdown rule : $\mathrm{X\,p} \xrightarrow{\mathbf{i}} \mathrm{U\,q}$ with $i \in \{0, \ldots, k-1\}$
- transition : $\mathrm{V\,X\,p} \xrightarrow{\mathbf{i}} \mathrm{V\,U\,q}$
- The graph of transition of the pushdown automaton is :
  - Vertices : set of all the transitions accessible from the set of initial configurations
  - Edges : $V\,X\,p \xrightarrow{\mathbf{i}} V\,U\,q$
  - Initial set of vertices : initial configurations of the PDA.
  - Each vertex is labelled by some $a \in A$, with the condition that for every $a \in A$, the set of configuration labelled by a is a regular Set.

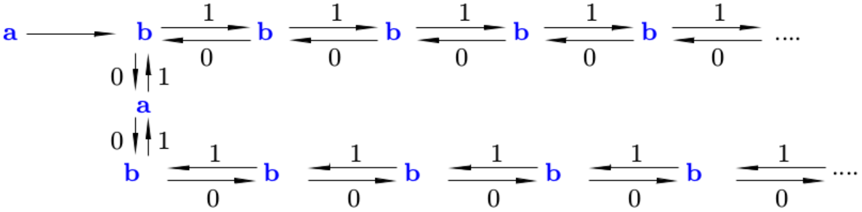


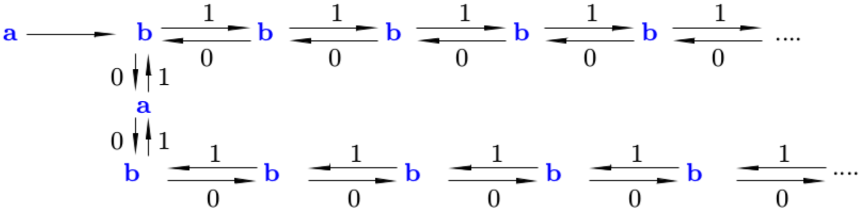
Initial configuration : $\{\#s\}$,
Labelling of configurations : $R_a = \{\#s, \#q\}$, $R_b = C \setminus R_a$



Initial configuration : $\{\#s\}$,
Labelling of configurations : $R_a = \{\#s, \#q\}$, $R_b = C \setminus R_a$





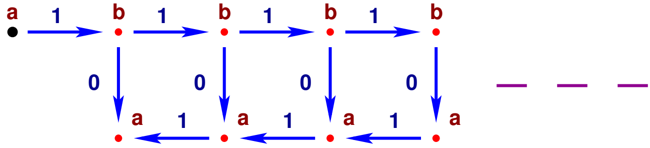


This transition graph acts like a finite automaton with output function :

Feed it with the base 2 expansion of any integer and you get an output *a* or *b*...

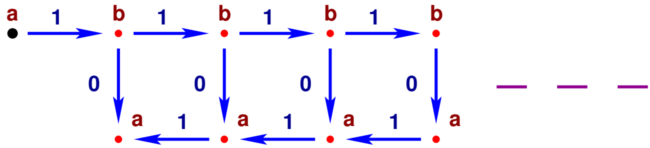
What kind of structure for such graphs ?

Graphs finitely decomposable by distance

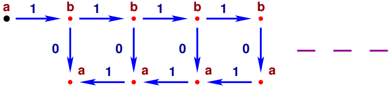


Connected components

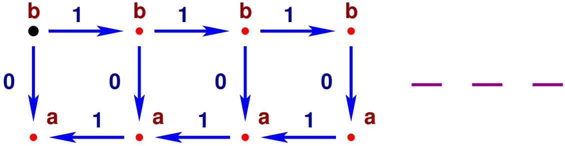
Graphs finitely decomposable by distance



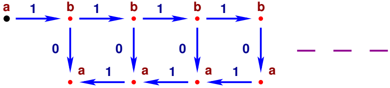
Connected components



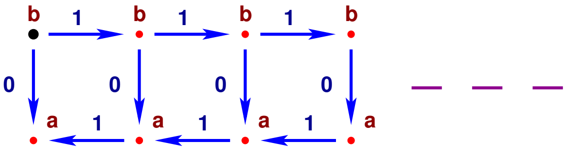
Graphs finitely decomposable by distance



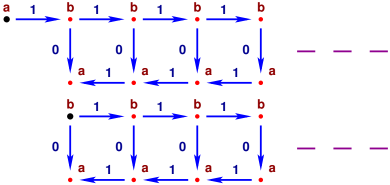
Connected components

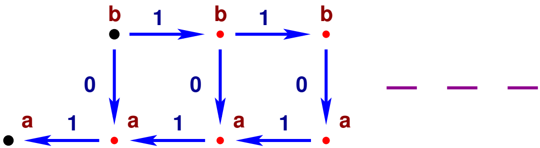


Graphs finitely decomposable by distance

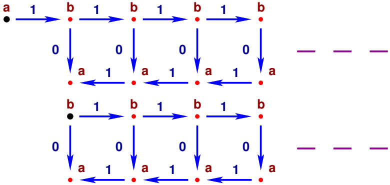


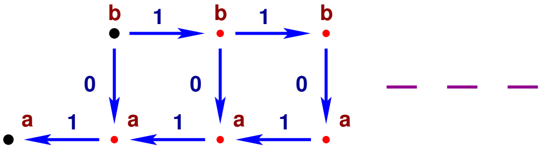
Connected components



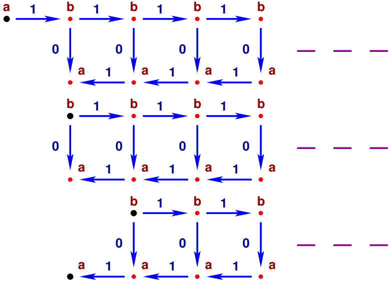


Connected components

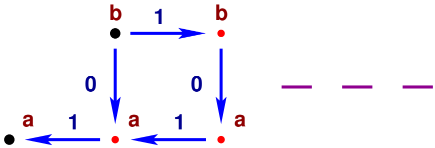




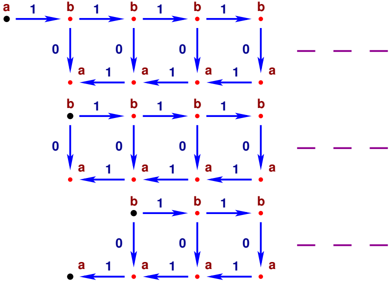
Connected components

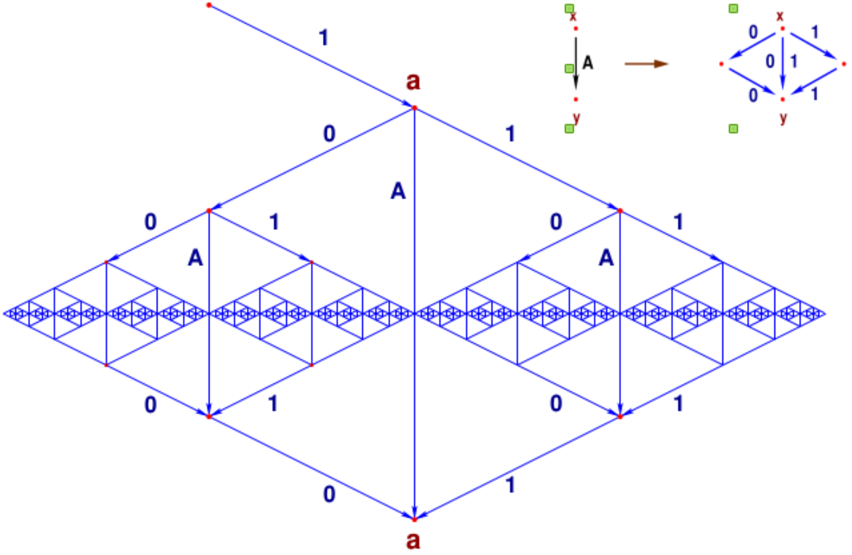


Graphs finitely decomposable by distance

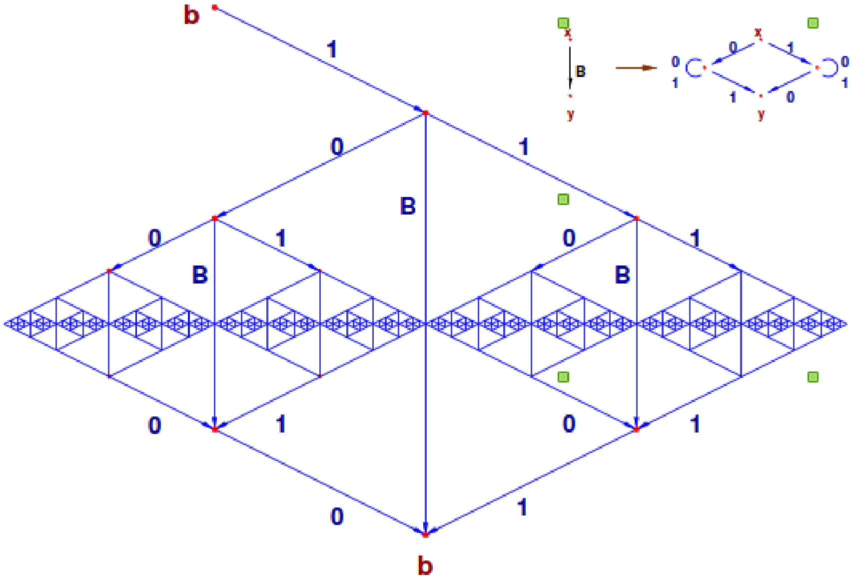


Connected components

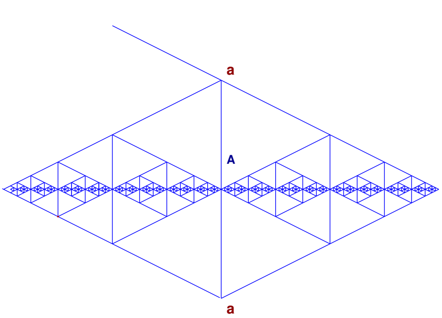




$$\mathrm{Fib}_k(x,a) \;=\; \{\, 1u \mid u = \widetilde{u} \,\}$$



$$\mathrm{Fib}_k(x, b) \;=\; \{\, 1u \mid u \neq \widetilde{u} \,\}$$



2-regular graph recognizing the sequence x defined by

$$\mathrm{Fib}_k(x,a) = \{\, 1u \mid u = \widetilde{u} \,\} \text{ and } \mathrm{Fib}_k(x,b) = \{\, 1u \mid u \neq \widetilde{u} \,\} \cup \{\varepsilon\}$$

Main difficulties to reach an “automatic-style” or “tag-machine”-style efficient construction of this sequences :

- Regular graphs are not always deterministic and/or unambiguous.
- Even if all fibers of a sequences are deterministic context-free, its usually impossible to generate them with a single deterministic regular graph.
- Moreover, the family of (real-time) deterministic k-context-free sequences is closed under
  - ~~product with any k-automatic sequence~~
  - ~~left and right shifts~~
  - ~~uniform morphism~~
  - ~~inverse injective k-substitution~~

k-context-free automatic sequences

Definition

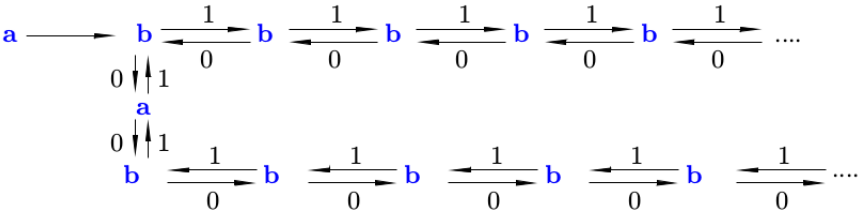
A sequence is (resp. strict) k-context-free automatic if

it is recognized by a deterministic k-regular graph

(resp. and of finite degree)

- deterministic complete coloured k-graph with an initial 0-loop
- sequence by length lexicographic order of the unfolding

An exemple



The sequence x defined by : $\text{Fib}(x, a) = \{u, |u|_1 = |u|_0\}$ and $\text{Fib}(x, b) = \{u, |u|_1 \neq |u|_0\}$ is 2-context-free automatic

Property

x (resp. strict) k-context-free automatic sequence

$\Longrightarrow$ $\mathrm{Fib}_k(x, a)$ deterministic (resp. and real-time) context-free

for each $a \in A$

- converse true for $|A| = 2$ and false in general

Stability properties

The family $\text{DetCf}_k(A^\omega)$ of k-context-free automatic sequences

is closed under

- finite modifications
- product with any k-automatic sequence
- ~~left and right shifts~~
- uniform morphism
- inverse injective k^p-substitution

realized by graph transformations (inverse path functions) which preserve regularity of graphs

In the Cobham's direction...

(Cobham 69) Any sequence which is both k- and p-automatic with k and p multiplicatively independent : $k^i \neq p^j$ for any $i, j > 0$, if and only if ultimately periodic.

Automatic number style : Any number is both k- and p-automatic with k and p multiplicatively independent : $k^i \neq p^j$ for any $i, j > 0$ if and only if is rational.

Base dependence

For any $k, p > 1$,

$$\mathrm{Cf}_k(A^\omega) = \mathrm{Cf}_{k^p}(A^\omega) \quad \text{and} \quad \mathrm{DetCf}_k(A^\omega) \subsetneq \mathrm{DetCf}_{k^p}(A^\omega)$$

$$\mathrm{Cf}_k(A^\omega) = \mathrm{Cf}_p(A^\omega) \iff \exists\, i, j > 0,\ k^i = p^j$$

Algebraicity of context-free automatic numbers

(Adamczewski, Bugeaud and Luca 2004)The k-adic expansion of any algebraic irrational number is not a automatic sequence.

Theorem (Adamczewski,Cassaigne, LG)

The k-adic expansion of any algebraic irrational number is not a context-free automatic sequence.

Relies on the fact that algebraic numbers which are not rational have bad rationnal approximations.

An idea of the proof

Combinatorial criterion for transcendence (Adamczewski, Bugeaud and Luca 2004)

Let x be an infinite sequence over a finite alphabet $A \subset \mathbb{N}$.
If one can find 2 a real number $\epsilon > 0$ and 2 sequences of finite words $(U_n)_{n \in \mathbb{N}}$ and $(V_n)_{n \in \mathbb{N}}$ such that :

- $U_n V_n^{1+\epsilon}$ is a prefix of x,
- the sequence $\left(\frac{|U_n|}{|V_n|}\right)_{n \in \mathbb{N}}$ is bounded,
- the sequence $(|V_n|)_{n \in \mathbb{N}}$ is increasing.

then, for all $b \geq 2$, the real number whose b-expansion is x is either rational or transcendent.

Context-free automatic sequences satisfy the conditions of this result

- By hand, using the underlying structure of pushdown automaton
- ... or using properties of regular graphs.

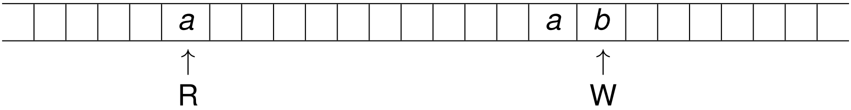
An other consequence of this criterion

Theorem (Adamczewski,Cassaigne, LG)

The k-adic expansion of any algebraic irrational number cannot be generated by a tag-machine (substitutive sequence) with dilation factor larger than 1.

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



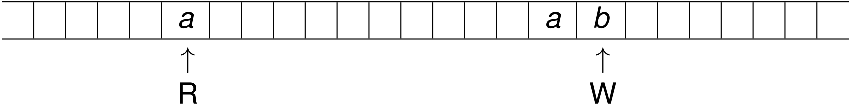
Read and Write start from an initial a

Read a $\longrightarrow$ Write ab and move right

Read b $\longrightarrow$ Write ba and move right

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

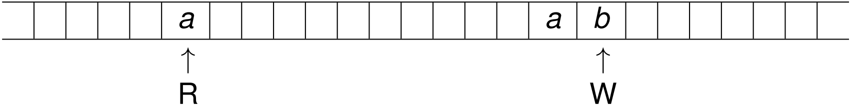
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ a

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

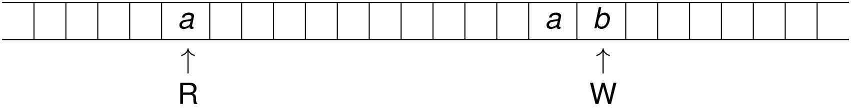
Read a $\longrightarrow$ Write ab and move right

Read b $\longrightarrow$ Write ba and move right

TM = ab

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

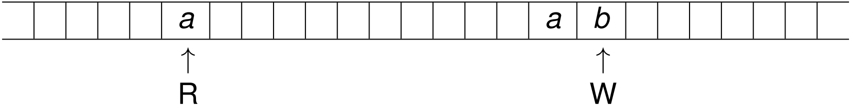
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abba$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abbaba$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:

Read and Write start from an initial a

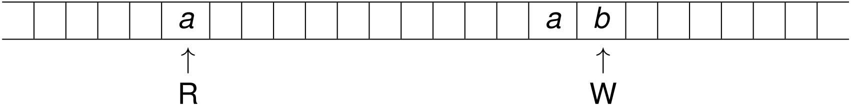
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abbabaab$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

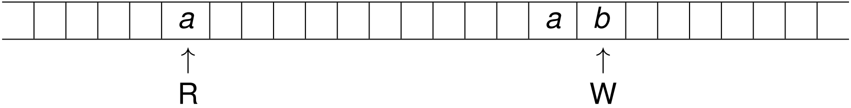
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abbabaabba$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

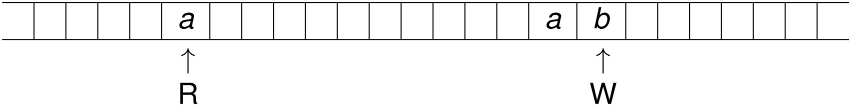
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abbabaabbaab$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

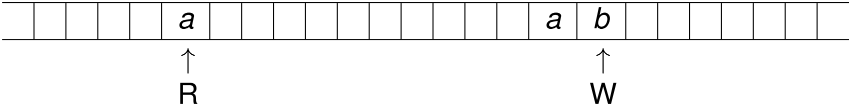
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ $abbabaabbaabab$

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

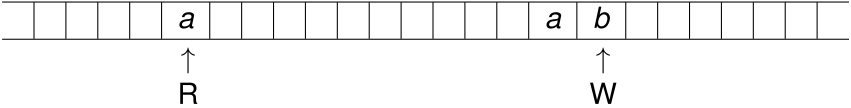
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ *abbabaabbaababba*

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

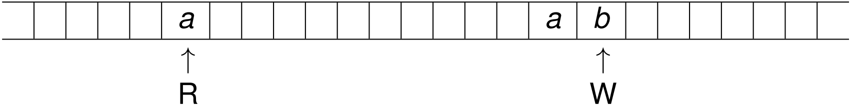
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

TM $=$ *abbabaabbaababbaba*

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial *a*

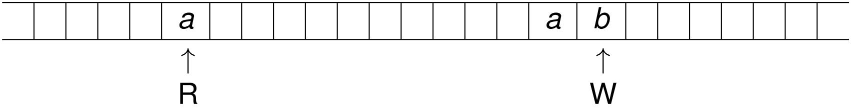
Read *a* $\longrightarrow$ Write *ab* and move right

Read *b* $\longrightarrow$ Write *ba* and move right

TM $=$ *abbabaabbaababbabaab*

The famous Thue-Morse example

A 2-tag machine over $A = \{a, b\}$:



Read and Write start from an initial a

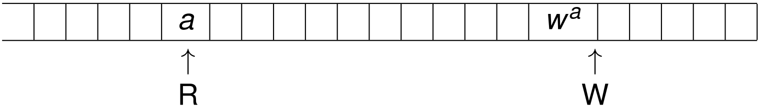
Read $a \longrightarrow$ Write ab and move right

Read $b \longrightarrow$ Write ba and move right

$$\text{TM} = abbabaabbaababbabaababbaabbabaabbaababbaabba\ldots$$

k-automatic sequences

A k-tag machine over A :



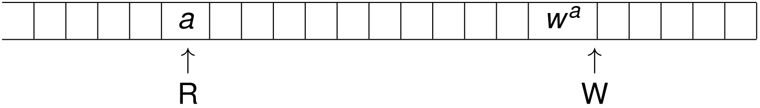
Read and Write start from an initial letter a_0

Read a $\longrightarrow$ Write w^a and move right

where w^a is a word over A

One can also assume that the word w^{a_0} starts with a_0

Dilation factor



$W(n)$: position of the writing head when the reading head R is at position n.

$$Dil(T) = \liminf_{n \to +\infty} \frac{W(n)}{n}$$

If $Dil(T) > 1$, then the number generated by this tag machine is either rational or transcendental.

Complexity

For any strict k-context-free automatic sequence x, its subword complexity is at most polynomial.
More precisely, if S is the number of stack symbol of the underlying pushdown automaton, then :

$$\mathrm{C}_x \in \begin{cases} O(n\log^2 n) & \text{if } S = 1 \\ O(n^{1+2\log_k S}) & \text{if } S > 1. \end{cases}$$

Somes ideas :

- level up in Maslov hierarchy ;
- Cobham's theorem for context-free automatic sequences ;
- Properties of formal series $\sum_{n\in\mathbb{N}} w_n X^n$ where w is context-free automatic sequence ovec $\mathbb{F}_p$;
- Stability of the set of context-free numbers by addition ? multiplication by rational ?...
- Structure of the k-kernels of context-free sequences ?