

How to prove that a language is regular or star-free?

Jean-Éric Pin¹

¹IRIF, CNRS and Université de Paris Cité

CIRM, Mai 2022, Luminy

Table of Contents

- 1 Definitions and early results
- 2 Iteration properties
- 3 Rewriting systems
- 4 Well quasi-orders
- 5 Logic
- 6 Transductions and profinite topology
- 7 Functions from \mathbb{N} to \mathbb{N}

Part I

Definitions and early results

The class of **regular languages** is the smallest class of languages containing the **finite languages** that is closed under finite **union**, **product** and **star**.

The class of **star-free languages** is the smallest class of languages containing the **finite languages** that is closed under finite **union**, **product** and **complement**.



Some examples

The language A^* is **star-free**, since $A^* = \emptyset^c$. More generally, if B is a subset of A , then B^* is **star-free** since

$$B^* = A^* - \bigcup_{a \in A-B} A^* a A^* = \left(\bigcup_{a \in A-B} \emptyset^c a \emptyset^c \right)^c$$

On the alphabet $\{a, b\}$, the language $(ab)^*$ is **star-free** since

$$(ab)^* = (b\emptyset^c \cup \emptyset^c a \cup \emptyset^c a a \emptyset^c \cup \emptyset^c b b \emptyset^c)^c.$$

Every **star-free** language is **regular**, but the language $(aa)^*$ is not **star-free**.

Syntactic monoid

A language L is **recognized** by a monoid M if there is a **monoid morphism** $h : A^* \rightarrow M$ and a subset P of M such that $L = h^{-1}(P)$.

The **syntactic congruence** of L is the equivalence relation \sim_L on A^* defined as follows: $u \sim_L v$ iff, for every $x, y \in A^*$, xuy and xvy are either both in L or both outside of L . The **syntactic monoid** of L is the quotient monoid A^*/\sim_L .

The **syntactic monoid** of a regular language is also the **transition monoid** of its minimal automaton.

Theorem (Kleene 1956, Rabin-Scott 1959)

Let L be a language. Are equivalent:

- (1) L is *regular*,
- (2) L is recognized by a *finite automaton*,
- (3) L is recognized by a *finite monoid*,
- (4) the *syntactic monoid* of L is finite.

Star-free languages

A finite monoid M is **aperiodic** if, for every $x \in M$, there exists $n \geq 0$ such that $x^{n+1} = x^n$.

Theorem (Schützenberger 1965)

For a language L , are equivalent:

- (1) L is **star-free**,
- (2) L is recognised by a **finite aperiodic monoid**,
- (3) the **syntactic monoid** of L is a **finite aperiodic monoid**.

Quotients

The **left quotient** of L by u is $u^{-1}L = \{v \mid uv \in L\}$.

The **right quotient** of L by u is $Lu^{-1} = \{v \mid vu \in L\}$.

The quotients of a **regular** [**star-free**] language are also **regular** [**star-free**].

Theorem (Nerode)

Let L be a language. Are equivalent:

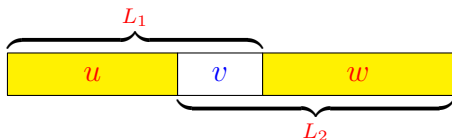
- (1) L is **regular**,
- (2) L has **finitely many left quotients**,
- (3) L has **finitely many right quotients**.

An exercise using quotients

If L_1 and L_2 are regular [star-free], then

$$L = \{uw \mid \exists v \text{ such that } uv \in L_1 \text{ and } vw \in L_2\}$$

is also regular [star-free].



Proof. Note that $u \in L_1 v^{-1}$ and $w \in v^{-1} L_2$. Thus

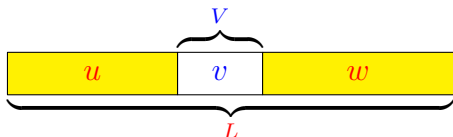
$$L = \bigcup_{v \in A^*} (L_1 v^{-1})(v^{-1} L_2)$$

If L_1 and L_2 are regular, the union is a finite union and L is regular. If L_1 and L_2 are star-free, so is L .

Stearns and Hartmanis's example

Let V be any language. If L is regular [star-free], then so is

$$f_V(L) = \{uw \mid \exists v \in V \text{ such that } uvw \in L\}.$$



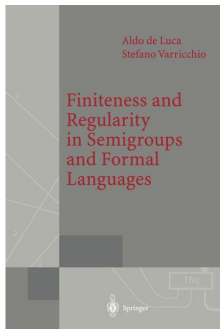
Proof. Let $h: A^* \rightarrow M$ be the syntactic morphism of L and let $P = h(L)$. Setting $R = h(V)$ and $T = \{(m, n) \in M \times M \mid mRn \cap P \neq \emptyset\}$ one gets

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

Part II

Iteration properties

See the book of de Luca and Varricchio (1999).



Finiteness and
Regularity
in Semigroups
and Formal
Languages

Pumping

Jaffe 1978, Stanat-Weiss 1982, Ehrenfeucht, Parikh-Rozenberg 1981 proposed extensions of the **pumping lemma** that **characterise regular languages**. The best versions are due to De Luca and Varricchio.

Proposition (Varricchio 1997)

A language L is **regular** iff there exists m such that, for all words x, u_1, \dots, u_m and y , there exist $i < j$ such that for all $n > 0$,

$$xu_1 \cdots u_{i-1}(u_i \cdots u_j)u_{j+1} \cdots u_my \in L$$

$$\iff xu_1 \cdots u_{i-1}(u_i \cdots u_j)^n u_{j+1} \cdots u_my \in L$$

A slight improvement

Proposition (de Luca, Varricchio)

A language L is *regular* iff there exist m and s such that for all words u_1, \dots, u_m , there exist $i \leq j$, such that for all words x, y and all $n \geq s$,

$$xu_1 \cdots u_{i-1} (u_i \cdots u_j) u_{j+1} \cdots u_m y \in L \iff$$
$$xu_1 \cdots u_{i-1} (u_i \cdots u_j)^n u_{j+1} \cdots u_m y \in L$$

Periodicity and permutation

L is **periodic** if, for all $u \in A^*$, there exist $n, k > 0$ such that $xu^n y \in L \iff xu^{n+k} y \in L$.

It is **permutable** if there exist $n > 0$ such that, for each sequence u_1, \dots, u_n of words, there is a nontrivial **permutation** σ such that $xu_1 \cdots u_n y \in L \iff xu_{\sigma(1)} \cdots u_{\sigma(n)} y \in L$.

Proposition (Restivo and Reutenauer 1984)

*A language is **regular** iff it is **periodic** and **permutable**.*

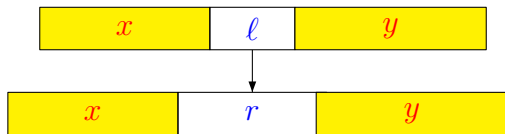
Part III

Rewriting systems

Rewriting systems

A **rewriting system** is a set R of **rewrite rules** $l \rightarrow r$.

Single-step reduction: $u \rightarrow_R v$ if $u = xly$ and $v = xry$ for some rule $l \rightarrow r$.



Reduction relation $\xrightarrow{*}_R$ = reflexive and transitive closure of \rightarrow_R .

R -closure of a language:

$$[L]_{\xrightarrow{*}_R} = \{v \in A^* \mid \exists u \in L \text{ such that } u \xrightarrow{*}_R v\}$$

Theorem

Rewriting systems in which $|r| \leq 1$ for all r preserve regularity.

Proof. Start from the minimal automaton of L .

- for each rule $u \rightarrow 1$ and each path $p \overset{u}{\rightsquigarrow} q$, add a new transition $p \xrightarrow{1} q$;
- for each rule $v \rightarrow a$ with $a \in A$ and each path $p \overset{v}{\rightsquigarrow} q$, add a new transition $p \xrightarrow{a} q$.

The automaton obtained at the end of the iteration process accepts $[L] \xrightarrow{*} \gamma_R$.

Suffix rewriting systems

Suffix rewriting systems: Single-step reductions are **modified** as follows: $u \rightarrow_R v$ if $u = x\ell$ and $v = xr$ for some rule $\ell \rightarrow r$.

Theorem (Büchi 1964)

Suffix [prefix] rewriting systems preserve regularity.

Total regulators

A **rewriting system** R is a **total regulator** if the R -closure of **any** language is **regular**.

Theorem (Bovet-Varricchio 1992)

Let $A = \{a, b\}$. The following rewriting systems are **total regulators**:

- (1) $\{u \rightarrow uu \mid u \in A^*\}$,
- (2) $\{u \rightarrow uu \mid u \in A^*, |u| \leq 2\}$.

Part IV

Well quasi-orders

Let \leq be a quasi-order (or preorder) on A^* . The upper set generated by L is

$$\uparrow L = \{v \in A^* \mid \exists u \in L \text{ such that } u \leq v\}.$$

A quasi-order \leq is a well quasi-order (wqo) if every upper set is generated by some finite language.

A quasi-order is stable (or monotone) if, for all u, v, x, y , the condition $u \leq v$ implies $xuy \leq xvy$.

Well quasi-orders and regular languages

Theorem (Ehrenfeucht, Haussler, Rozenberg 1983)

A language is *regular* iff it is an *upper set* with respect to some *stable well quasi-order* on A^* .

Theorem

A rewriting system whose *reduction relation* is a *well quasi-order* is a *total regulator*.

Some examples

The rewriting system $\{1 \rightarrow a \mid a \in A\}$.

Theorem (Higman)

For all L , the shuffle product $L \sqcup A^*$ is regular.

A set H of words is **unavoidable** if the language $A^* - A^*HA^*$ is finite.

Theorem (Ehrenfeucht, Haussler, Rozenberg 83)

If H is a **unavoidable finite set**, then the reduction relation of the rewriting system $\{1 \rightarrow u \mid u \in H\}$ is a **well quasi-order**.

Ordered semigroups

Let (S, \leq) be a **finite ordered semigroup** and let $\sigma : A^+ \rightarrow S$ be a **onto semigroup morphism**. Let

$$R_\sigma = \{a \rightarrow u \mid a \in A, u \in A^+ \text{ and } \sigma(a) \leq \sigma(u)\}.$$

Theorem (Bucher, Ehrenfeucht, Haussler 1985)

The following conditions are equivalent:

- (1) $\xrightarrow{*} R_\sigma$ is a **well quasi-order**,
- (2) $\{a u a \mid a \in A, u \in A^*, \sigma(a) \leq \sigma(a u a)\}$ is **unavoidable** over A ,
- (3) $\{a u \mid a \in A, u \in A^*, \sigma(a) \leq \sigma(a u)\}$ is **unavoidable** over A .

Congenial ordered semigroups

When is $\xrightarrow{*} R_\sigma$ a well quasi-order?

Proposition (Klima-Kolegar 2022)

This property only depends on (S, \leq) and not on the choice of σ . It is a decidable property.

An ordered semigroup satisfying this property is called congenial (sympatique).

Theorem (Kunc 2005)

A finite semigroup equipped with the equality order is *congenial* iff it is a *chain of simple semigroups*.

Algebraically, it means that, for all $s, t \in S$, either $(st)^\omega s = s$ or $t(st)^\omega s = t$, a property that can be tested in *polynomial time*.

Congenial ordered semigroups

Theorem (Klima-Kolegar 2022)

A finite ordered semigroup S is *congenial* iff for all $n > 0$, for all $s_1, \dots, s_n \in S$, there exists $i \in [1, n]$ such that $s_i \leq s_i(s_{i+1} \cdots s_n s_1 \cdots s_i)^\omega$.

Unfortunately, it is **not possible to bound n** : there are ordered semigroups S_m which satisfy the condition for $n < m$ but not for $n = m$.

Examples

Let $B_2 = \{a, b, ab, ba, 0\}$, with $aba = a$, $bab = b$
and $a^2 = b^2 = 0$.

$*ab$	a
b	$*ba$

$*0$

- (1) With the equality, B_2 is not congenial.
- (2) With the order $x \leq 0$ for all x , it is congenial.
- (3) With the order $0 \leq x$ for all x , it is not congenial.

Systems of inequations (Kunc and Okhotin)

Consider a **finite system of inequations** of the form

$$\bigcup_{j \in J_i} P_{i,j}(X_1, \dots, X_n) \subseteq C_i \quad (1 \leq i \leq k)$$

where each $P_{i,j}$ is a product of **arbitrary constant languages** and **variables**, each C_i is a **constant regular language** and each J_i is possibly **infinite**.

Example. $A = \{a, b, c\}$, $L_0 = \{a^n b^n c^n \mid n \geq 0\}$,
 $L_1 = \{u^{|u|} \mid u \in A^*\}$

$$\bigcup_{p \text{ prime}} X_0 X_1^p L_0 X_1 X_0 L_1 X_0^{p^2} X_1 \subseteq A^* abc A^*$$

Equations and inequalities

$$\bigcup_{j \in J_i} P_{i,j}(X_1, \dots, X_n) \subseteq C_i \quad (1 \leq i \leq k)$$

Theorem (Kunc and Okhotin 2020)

The system has only *finitely many maximal solutions* and every *maximal solution* has *all components regular*.

If all C_i are *star-free*, then the *maximal solutions* are *star-free*.

Furthermore, the result *still holds* if any *inequation* is replaced by an *equation*.

Inequations of the form $XK \subseteq LX$

Theorem (Kunc 2005)

Let K be an *arbitrary* language and let L be a *regular* language. Then the *greatest solution* of the inequation $XK \subseteq LX$ is regular.

The situation is *totally different* for equations of the type $XK = LX$. Kunc (2007) has shown that there exists a *finite* language L such that the *greatest solution* of the equation $XL = LX$ is co-recursively enumerable complete.