# Regular Languages Are Church-Rosser Congruential


**Volker Diekert**
University of Stuttgart

Manfred Kufleitner
University of Stuttgart

Klaus Reinhardt
University of Tübingen

Tobias Walter
University of Stuttgart


*Dedicated to the 60th birthday of Stuart Margolis*


International Conference on Geometric, Combinatorial and Dynamics aspects of Semigroup and Group Theory
Bar Ilan University, June 13th, 2013

# Complexity of formal language classes

Given $L \subseteq A^*$. What is the complexity of its *Word Problem* WP($L$)? **Input:** $w \in A^*$. **Question:** $w \in L$?.

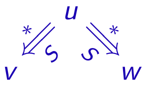
- $L$ regular: WP($L$) real time by reading the input.
- $L$ context-free: WP($L$) is roughly cubic time.
- $L$ context-sensitive: WP($L$) is in PSPACE.
- $L$ recursively enumerable: WP($L$) is recursively enumerable.

If $L$ is deterministic context-free, then WP($L$) is solvable in linear time.

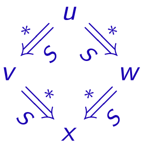
Idea: Go beyond deterministic context-free and keep linear time.

# Semi-Thue systems

- $S \subseteq A^* \times A^*$ *semi-Thue system*, $S$ finite
- elements of $S$ are *rules* $\ell \to r$
- derivation $u \underset{S}{\Longrightarrow} v$ if $u = p\ell q$, $v = prq$, $(\ell \to r) \in S$
- $A^*/S$ are the congruence classes modulo $\overset{*}{\underset{S}{\Longleftrightarrow}}$
- $S$ is *length-reducing* if $|\ell| > |r|$ for all $(\ell \to r) \in S$
- $S$ is *confluent* if




implies




- unique normal forms $\Longrightarrow$ efficient parsing

# Go beyond det.c.f. by confluent string rewriting

**McNaughton, Narendran, and Otto, JACM 1988**

$L \subseteq A^*$ is called a *Church-Rosser congruential language* (CRCL), if there is some finite, confluent and length reducing semi-Thue system $S \subseteq A^* \times A^*$ such that $L$ is a finite union of congruence classes mod $S$.

Algorithm to decide WP:

- **Input:** $w \in A^*$.
- Compute in linear time $w \overset{*}{\underset{S}{\Longrightarrow}} \widehat{w} \in \mathrm{IRR}(S)$.
- Check whether $\widehat{w}$ appears in a finite precomputed table.

**Conjecture 1988-2012**

All regular languages are in CRCL.

# Our contribution: Solution of the 1988 conjecture

Theorem [DKRW2012] Let $L \subseteq A^*$. The following are equivalent:

- $L$ is regular.
- $L$ is strongly Church-Rosser congruential.

$L$ is *strongly Church-Rosser congruential* (sCRCL), if there is some semi-Thue system $S \subseteq A^* \times A^*$ such that

1. $S$ is finite, confluent and length reducing.
2. $S$ is of finite index, i.e., the quotient monoid $A^*/S$ is finite.
3. $L$ is a union of congruence classes mod $S$.

Best known result before 2012 was in an unfinished manuscript by Reinhardt and Thérien (2003): Conjecture is true, if the syntactic monoid is a group.
Idea 2011: Let's use the concept of *Local Divisor*.

# Examples (1)

- $L_1 = \{a^n b^n \mid n \geq 0\}$
  - $S = \{aabb \to ab\}$
  - $L_1 = [ab]_S \cup [\varepsilon]_S$, $L_1$ is *Church-Rosser congruential*
  - $A^*/S$ is infinite, contains $\{[a^n]_S \mid n \geq 1\}$
- $L_2 = \{a^m b^n \mid m \geq n \geq 0\}$
  - not *Church-Rosser congruential* since $a^m$ is irreducible
- $L_3 = \{a, b\}^* \, a \, \{a, b\}^*$
  - $S = \{aa \to a, \; b \to \varepsilon\}$
  - $L_3 = [a]_S$
  - $A^*/S$ is finite, $L_3$ is *strongly Church-Rosser congruential*

# Examples (2)

- $L_4 = (ab)^*$
  - $S = \{aba \to a\}$
  - $L_4 = [ab]_S \cup [\varepsilon]_S$
  - $A^*/S$ is infinite
  - $T = \{aaa \to aa,\ aab \to aa,\ baa \to aa,$
    $bbb \to aa,\ bba \to aa,\ abb \to aa,$
    $aba \to a,\ bab \to b\}$
  - $L_4 = [ab]_T \cup [\varepsilon]_T$
  - $A^*/T$ has 7 elements

# Examples (3)

- $L_5 = \{w \in a^* \mid |w| \equiv 0 \bmod 3\}$
  - $S = \{aaa \to \varepsilon\}$

- $L_6 = \{w \in \{a, b\}^* \mid |w| \equiv 0 \bmod 3\}$
  - $S = \{u \to \varepsilon \mid |u| = 3\}$?
  - NO: $S$ is not confluent: $a \underset{S}{\Longleftarrow} aabb \underset{S}{\Longrightarrow} b$
  - $T = \{aaa \to 1,\ baab \to b,\ (ba)^3 b \to b\}$
    $\cup \{bb\,u\,bb \to b^{|u|+1} \mid 1 \le |u| \le 3\}$
  - $L_6$ is a union of elements in $A^*/T$
  - $A^*/T$ contains 272 elements,
    longest irreducible word has length 16

- $L_7 = \{w \in \{a, b, c\}^* \mid |w| \equiv 0 \bmod 3\}$ ???

# Simple non-cyclic groups

- $\varphi : A^* \to G$ surjective hom., $G$ simple non-cyclic group
- $L_G = \{w \mid \varphi(w) = 1\}$
- Assume $|w| \equiv 0 \bmod n > 1$ for all $w \in L_G$.
- Then $w \mapsto |w| \bmod n$ induces surjective hom. $G \to \mathbb{Z}/n\mathbb{Z}$.
- Contradiction.
- Thus we find $u, v \in L_G$ such that $|u| - |v| = 1$.
- Padding with $u$ and $v$ yields normal forms $v_g \in A^*$ for $g \in G$:
  - $\varphi(v_g) = g$,
  - $|v_g| = |v_h|$ for all $g, h \in G$.
- $S = \left\{w \to v_{\varphi(w)} \;\middle|\; |w| = |v_g| + 1\right\}$,
  works for any language recognized by $\varphi$

# Local divisors

- Let $M$ be a monoid and let $c \in M$.
- Composition $\circ$ on $cM \cap Mc$ defined by $xc \circ cy = xcy$.
- Let $xc = x'c$ and $cy = cy'$. Then

$$xc \circ cy = xcy = x'cy = x'cy' = x'c \circ cy'.$$

- Let $cx = x'c$ and $cy$ be elements in $cM \cap Mc$. Then

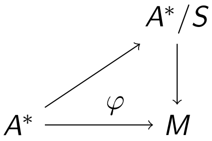
$$cx \circ cy = x'c \circ cy = x'cy = cxy.$$

- It follows
  - $(cM \cap Mc, \circ, c)$ is a monoid.
  - If $c$ is not invertible, then $|cM \cap Mc| < |M|$.

# Weights

- Let $\|\cdot\| : A \to \mathbb{N}$ assign a positive *weight* to each letter.
- $\|a_1 \cdots a_n\| = \|a_1\| + \cdots + \|a_n\|$.

- Theorem: For every weighted alphabet $(A, \|\cdot\|)$ and every homomorphism $\varphi : A^* \to M$ there exists a weight-reducing confluent semi-Thue system $S$ of finite index such that $\varphi$ factorizes through $A^*/S$.



# Proof sketch

- $\varphi : A^* \to M$ homomorphism, $c \in A$ not invertible
- Define $B = A \setminus \{c\}$
- $\varphi_c : B^* \to M$ restriction
- Induction on the alphabet: system $R$ for $\varphi_c$
- $K = \mathrm{IRR}_R(B^*)c$
- $K$ inherits its weights from $A$
- $\psi : K^* \to \varphi(c)M \cap M\varphi(c) : uc \mapsto \varphi(cuc)$ homomorphism
- Induction on the monoid: system $T$ for $\psi$.
- Combine $R$ and $T$ in order to get a system $S$ for $\varphi$.

- Base case: $A = \emptyset$ is trivial.
- Base case: $M$ is a group is highly non-trivial.

# Open problems

- Complexity improvements: size of $S$, size of $A^*/S$
- Lower bounds on size of $S$ and $A^*/S$
- Parikh-reducing systems

# Thank you!