Word problems and formal language theory

Rick Thomas

Department of Informatics

University of Leicester

York Semigroup Seminar

13th February 2019

Generating sets

If Σ is a finite set of symbols we let Σ^* denote the set of all finite words of symbols from Σ (including the empty word ϵ). If we only want to consider non-empty words then we denote the resulting set by Σ^+.

Σ^+ is the *free semigroup* on Σ and Σ^* is the *free monoid* on Σ.

A *language* is a subset of Σ^* for some finite set Σ.

If we have a group G (or a monoid M) with a finite set of generators Σ, then we have a natural homomorphism $\varphi : \Sigma^* \longrightarrow G$ (or $\varphi : \Sigma^* \longrightarrow M$).

For a semigroup S generated by a finite set Σ we have a natural homomorphism $\varphi : \Sigma^+ \longrightarrow S$.

Word problems

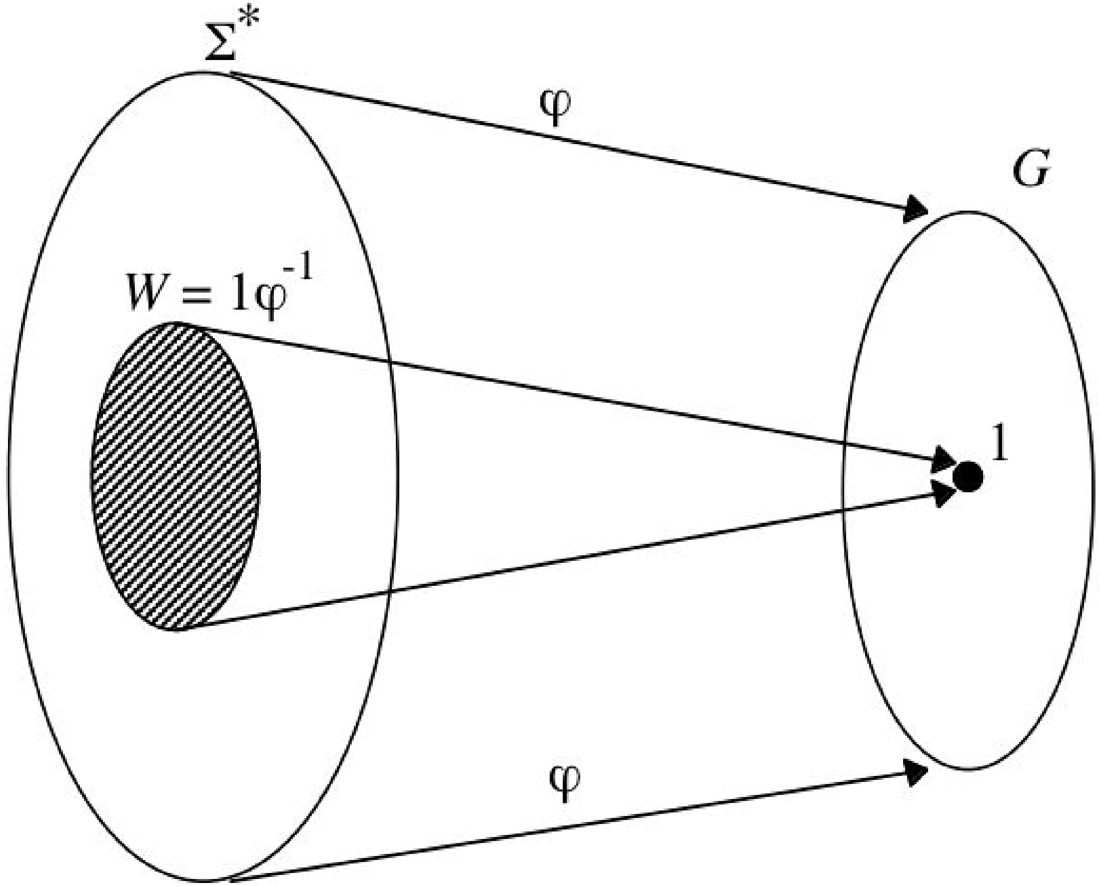
The *word problem* in such a structure is the following question:

Input: Two words α and β in Σ^* (or Σ^+ in the case of a semigroup);

Output: **Yes** if α and β represent the same element of the group (monoid, semigroup);

No otherwise.

In a group, given a word β representing an element g, let γ be a word representing g^{-1}. Now α and β represent the same element of the group if and only if $\alpha\gamma$ represents the identity.


Σ*
φ
G
W = 1φ-1
1
φ

Word problems

Given this, we can define the *word problem* $W = W(\mathrm{G})$ of a finitely generated group G to be the set of all words in Σ^* that represent the identity element of G. (This is not appropriate for monoids and does not make sense in semigroups.)

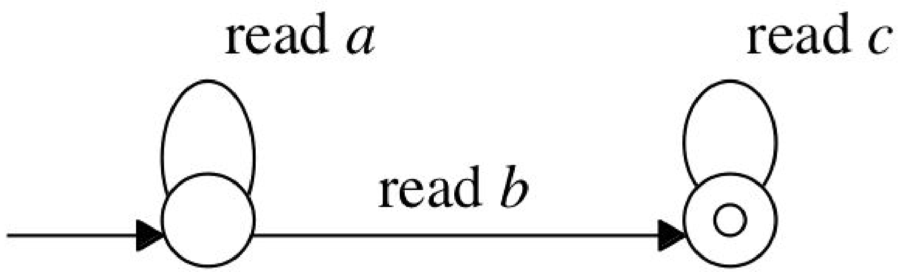
In this way, we can think of the word problem of a group as being a language.

We will focus on some relatively simple families of languages, the *regular languages*, the *one-counter languages*, the *context-free languages* and the *Petri net languages*. Saying that the word problem of a group is regular (or one-counter or context-free or Petri net) does not depend on the choice of finite generating set.

Automata

We can define families of languages using various notions of "automata".

Regular languages are accepted by *finite automata*.

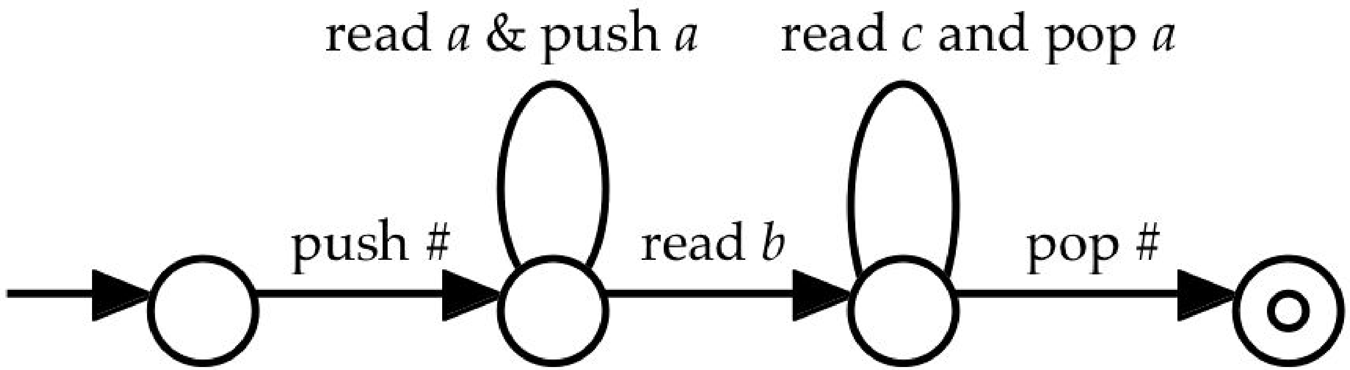


A word is *accepted* if we reach an "accept state" after reading the word.

The *language* $\mathrm{L}(\mathrm{M})$ of M is the set of all words accepted by M.

Pushdown automata

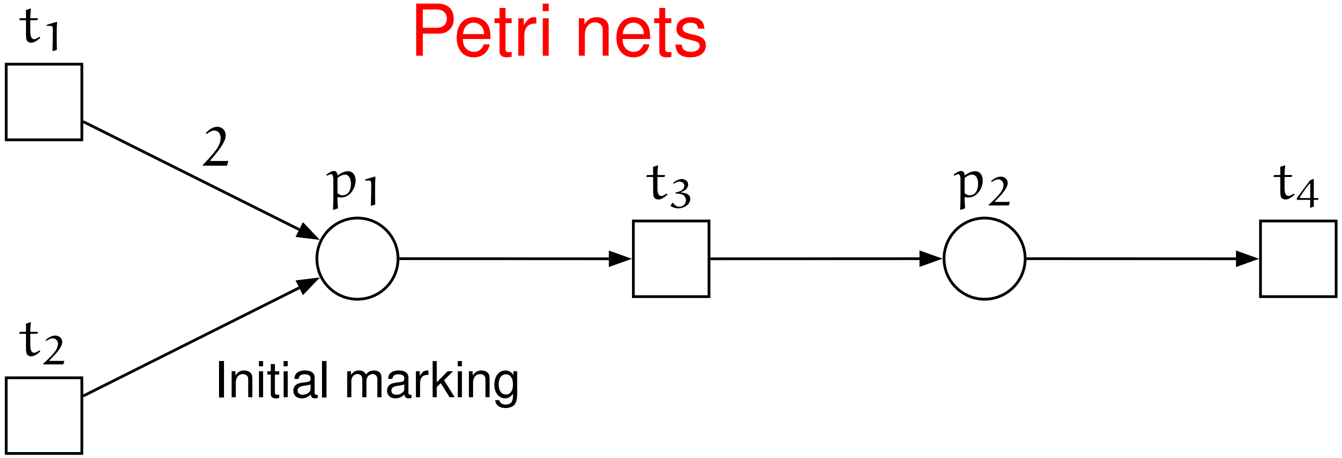
Context-free languages are accepted by *pushdown automata* where we add a “stack” to the machine.



If we restrict to one stack symbol (apart from a fixed bottom marker #) we have a *one-counter language*.

An automaton is said to be *deterministic* if there can never be a possibility of choice as regards to which move to make.

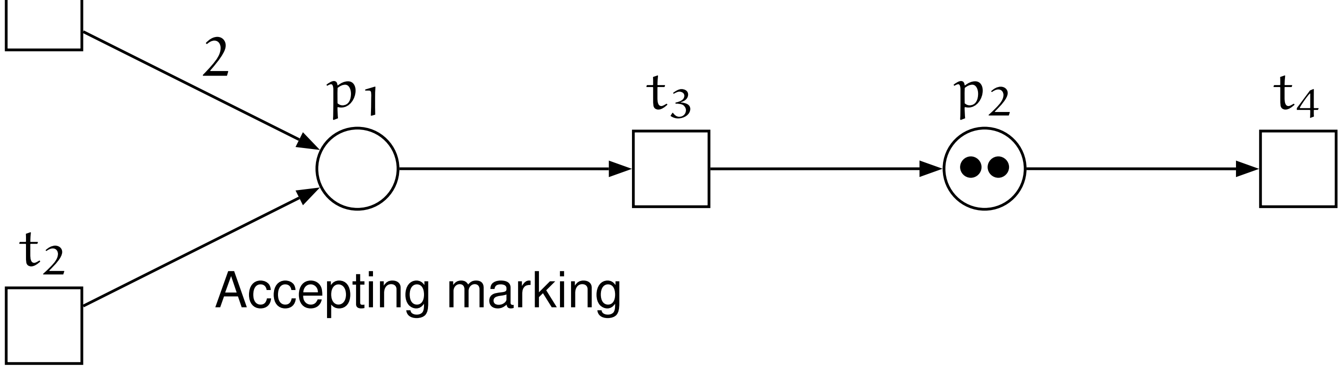
Petri nets



Initial marking

Each transition t_1, t_2, t_3, t_4 has a label.

The set of all accepted words is a *Petri net language*.



Accepting marking

Word problems of groups

If G is a finitely generated group, then $W(G)$ is regular if and only if G is finite. (Anisimov)

If G is a finitely generated group, then $W(G)$ is context-free if and only if G is virtually free. (Muller & Schupp)

As a consequence, if $W(G)$ is context-free, then it is deterministic context-free.

If G is a finitely generated group, then $W(G)$ is a one-counter language if and only if G is virtually cyclic. (Herbst)

Word problems of groups

The following are equivalent for a finitely generated group G and $n \geqslant 1$:

(i) The word problem of G is the intersection of n one-counter languages.

(ii) The word problem of G is the intersection of n deterministic one-counter languages.

(iii) G is virtually abelian of free abelian rank $\leqslant n$. (Holt, Owens & Thomas)

G being virtually abelian is also equivalent to the word problem of G being a Petri net language. (Rino Nesin & Thomas)

Conjecture. The word problem of a finitely generated group G is the intersection of n context-free languages (for some n) if and only if G is virtually a finitely generated subgroup of a direct product of free groups. (Brough)

Word problems of groups

A language L over an alphabet Σ is the word problem of a group with generating set Σ if and only if L satisfies the following two conditions:

(UPP) for all $\alpha \in \Sigma^*$ there exists $\beta \in \Sigma^*$ such that $\alpha\beta \in L$;

(DC) $\alpha \in \Sigma^*, \beta \in L, \gamma \in \Sigma^*, \alpha\beta\gamma \in L \Longrightarrow \alpha\gamma \in L$. (Parkes & Thomas)

As a consequence of the Muller-Schupp clasification we have:

If L is a context-free language satisfying (UPP) and (DC) then L is deterministic context-free.

More properties satisfied by word problems of groups

(USP) for all $\alpha \in \Sigma^*$ there exists $\beta \in \Sigma^*$ such that $\beta\alpha \in L$;

(UFP) for all $\alpha \in \Sigma^*$ there exist $\beta, \gamma \in \Sigma^*$ such that $\beta\alpha\gamma \in L$;

(CRD) $\alpha \in \Sigma^*, \beta \in L, \alpha\beta \in L \Longrightarrow \alpha \in L$;

(CLD) $\alpha \in L, \beta \in \Sigma^*, \alpha\beta \in L \Longrightarrow \beta \in L$;

(IC) $\alpha \in \Sigma^*, \beta \in \Sigma^*, \gamma \in L, \alpha\beta \in L \Longrightarrow \alpha\gamma\beta \in L$;

(CCS) $\alpha \in \Sigma^*, \beta \in \Sigma^*, \alpha\beta \in L \Longrightarrow \beta\alpha \in L$;

(CC) $\alpha \in L, \beta \in L \Longrightarrow \alpha\beta \in L$.

More properties satisfied by word problems of groups

There is a complete characterization as to which sets of properties from

$$S = \{\textbf{(UPP)}, \textbf{(DC)}, \textbf{(USP)}, \textbf{(UFP)}, \textbf{(CRD)}, \textbf{(CLD)}, \textbf{(IC)}, \textbf{(CCS)}, \textbf{(CC)}\}$$

characterize word problems of groups. There are eleven minimal such sets. (Jones & Thomas)

All the properties in S are easily seen to be decidable for the family of regular languages.

However, all of them are undecidable for the family of one-counter languages. (Jones & Thomas)

Some other characterizations

If Σ is an alphabet and $\emptyset \neq L \subseteq \Sigma^*$, then the following are equivalent:

(i) there is a monoid M and a monoid homomorphism $\varphi : \Sigma^* \longrightarrow M$ such that $L = \{1\}\varphi^{-1}$;

(ii) L satisfies **(DC)** and **(IC)**. (Jones & Thomas)

Let Σ be an alphabet, G be a group and $\varphi : \Sigma^* \longrightarrow G$ be a surjective monoid homomorphism. Let $S \subseteq G$ and $L = S\varphi^{-1}$. Then the following are equivalent:

(i) S is a subgroup of G.

(ii) L satisfies **(USP)**, **(CC)** and **(CLD)**. (Jones & Thomas)

Some other characterizations

Let Σ be an alphabet, G be a group and $\varphi : \Sigma^* \to G$ be a surjective monoid homomorphism. Let H be a subgroup of G and then let $L = H\varphi^{-1}$. Then the following are equivalent:

(i) H is a normal subgroup of G.

(ii) L satisfies **(IC)**.

(iii) L satisfies **(DC)**.

(iv) L satisfies **(CCS)**. (Jones & Thomas)

Decidability

There is no algorithm that, given a context-free language $\mathbb{L}$, will decide whether or not $\mathbb{L}$ is the word problem of a group. (Lakin & Thomas)

This can be generalized to the fact that there is no algorithm that, given a one-counter language $\mathbb{L}$, will decide whether or not $\mathbb{L}$ is the word problem of a group. (Jones & Thomas)

However, there is an algorithm that, given a deterministic context-free language $\mathbb{L}$, will decide whether or not $\mathbb{L}$ is the word problem of a group. (Jones & Thomas)

Word problems of semigroups

Duncan and Gilman proposed the following definition of the *word problem for a semigroup* S generated by a finite set A:

$$W(S) = \{\alpha\#\beta^{rev} : \alpha, \beta \in A^+, \alpha =_S \beta\}.$$

This is a natural generalization of the word problem of a group G which was

$$W(G) = \{\alpha\beta^{-1} : \alpha, \beta \in A^*, \alpha =_G \beta\}.$$

In this way, we can consider the word problem of a semigroup as a language.

If S is a finitely generated semigroup, then $W(S)$ is regular if and only if S is finite. (Duncan & Gilman)

One-counter word problems

For a semigroup S with a specified finite generating set A and $n \in \mathbb{N}$, let $\gamma_{S,A}(n)$ be the number of elements of S that are represented by words in A^+ of length at most n. Then $\gamma_{S,A}$ is called the *growth function* of S with respect to A.

Properties such as whether the growth function is linear, polynomial of degree d, or exponential are generally independent of the chosen finite generating set for S.

If a finitely generated semigroup S has word problem a one-counter language, then S has a linear growth function. (Holt, Owens & Thomas)

One-counter word problems

If S is a finitely generated semigroup with a linear growth function then there exist finitely many elements $a_i,\ b_i,\ c_i \in S \cup \{\epsilon\}$ such that every element of S is represented by a word of the form $a_i b_i^n c_i$ for some i and some $n \geqslant 0$. (Holt, Owens & Thomas)

A finitely generated semigroup with linear growth need not have one-counter word problem – in fact its word problem need not even be decidable!

This is in contrast to the situation in groups: a finitely generated group with linear growth is virtually cyclic.

Context-free word problems

For context-free word problems in semigroups there are some partial results (particularly with regards to semigroup constructions) but we are far from a classification.

In contrast to the situation with groups, there is a semigroup S whose word problem is context-free but not deterministic context-free; moreover, this semigroup S is not finitely presented. (Hoffmann, Holt, Owens & Thomas)

Thank you!