SIMPLE LANGUAGE EQUATIONS

MICHAL KUNC *

Abstract. We survey results, both positive and negative, on regularity of maximal solutions of systems of implicit language equations and inequalities. These results concern inequalities with constant right-hand sides, one-sided linear inequalities, inequalities with restrictions on constants, and commutation equations and inequalities. In addition, we present some of these results in a generalized form in order to underline common principles.

1. Introduction

Solving equations is one of the basic topics considered for every algebraic structure. For equations over words a rich theory has been developed and deep results were obtained (see e.g. [21]). But only little is known about equations over languages. Systems of language equations and inequalities were intensively studied especially in connection with context-free languages, since these languages can be elegantly described as components of least solutions of systems of explicit polynomial equations, i.e. equations with the operations of union and concatenation. Much less attention was devoted to implicit language equations and to equations employing other operations than union and concatenation. For standard systems of explicit equations, it makes no significant difference whether one deals with smallest solutions or largest solutions (see [25]). But in the case of implicit equations where the operation of union occurs, it is more natural to study their maximal solutions. For instance, every system of inequalities employing only the operations of union and concatenation with constants possesses a largest solution, although it can have arbitrarily many minimal solutions (just consider the equation $X \cup Y = L$, where X and Y are variables and L is a given language).

This article is intended to be a step towards understanding what the "simple language equations" are, from the point of view of complexity of their maximal solutions. We survey here recent investigations of systems of implicit language equations and inequalities, aiming to discover which types of systems have maximal solutions regular no matter what regular constants they contain and which systems can have maximal solutions non-regular even though all constants are regular languages. There are several large classes of systems whose maximal solutions are known to be regular although they employ non-regular expressions or non-regular constant languages. Here we attempt to find some common factors of different results ensuring regularity of solutions. On the other hand, one can also find equations of a very simple form over regular or even finite languages whose largest solutions are extremely complicated, potentially even non-recursively enumerable. An example of such an equation is the commutation equation. In this article, we restrict our attention to basic monotone language operations, i.e. concatenation,

* Supported by the project MSM 143100009 of the Ministry of Education of the Czech Republic.

union and intersection. Systems of inequalities which employ also the operation of complementation were studied for example by Leiss [20].

Interestingly, for every kind of language equations where maximal solutions are known to be regular, such a result can be directly extended to language inequalities. This is why most of the material presented here actually deals with inequalities rather than equations. Of course, every language inequality $E \subseteq F$, where E and F are expressions over variables and constant languages built using some language-theoretic operations, can be equivalently written as an equation $E \cup F = F$, but the inequality representation appears to be the right one because the properties of expressions E and F required to obtain regularity of solutions are always different; for instance, it is usually sufficient to require that constant languages occurring in these inequalities be regular only on the right-hand sides.

Let us mention that some results related to those presented in this survey were also obtained when language equations were studied from a different perspective, namely, when asking which languages can be defined by systems of implicit equations belonging to a given class. Recent results of Okhotin show that many natural classes of languages can be characterized in this way (see [22]), for example, recursive (recursively enumerable, co-recursively enumerable) languages are exactly languages definable as unique (smallest, largest) solutions of systems of implicit language inequalities using concatenation and all Boolean operations [23], and each such language can be encoded even in a single explicit equation and precisely defined by a system of two explicit equations [24].

The article is organized as follows. Some of the basic notions are recalled in the following section; for a more comprehensive introduction to formal languages, to semigroup theory and to well quasi-orders the reader is referred to [27], [10] and [7], respectively. In Sections 3 and 4 we consider positive results which can be obtained by constructing equivalence relations of finite index on free monoids. Section 5 is devoted to systems of inequalities where regularity of maximal solutions is proved using more general structures on free monoids, namely, well quasi-orders. Finally, the aim of Sections 6 and 7 is to describe some simple systems whose largest solutions are non-regular.

2. Preliminaries

We denote the sets of positive and non-negative integers by $\mathbb{N}$ and $\mathbb{N}_0$, respectively. Throughout the paper we consider a finite alphabet A. As usual, we write A^+ for the set of all non-empty finite words over A, and A^* for the set obtained from A^+ by adding the empty word ε. We use the same symbols A^+ and A^* to denote the free semigroup and the free monoid, respectively, which arise from these sets when we equip them with the operation of concatenation.

If $w = uv$ for some words $u, v, w \in A^*$, then the words u and v are called a *prefix* and a *suffix* of w, respectively, and we write $v = u^{-1}w$. The set of all prefixes of a word w is denoted $\mathrm{Pref}(w)$.

Languages over an alphabet A are arbitrary subsets of A^* and we say that a language $L \subseteq A^*$ is *ε-free* if $\varepsilon \notin L$. For a language $L \subseteq A^*$ and a word $u \in A^*$ we use the usual notation $u^{-1}L = \{v \in A^* \mid uv \in L\}$.

Regular languages are languages accepted by finite automata. An equivalence relation $\sim$ on A^* is a *right congruence* if it satisfies

$$\forall u, v \in A^* \ \forall a \in A \colon u \sim v \implies ua \sim va \,.$$

Right congruences of A^* of finite index directly correspond to finite automata; the state set of the automaton determined by $\sim$ is $A^*/\!\sim$ and the transition function is defined by the rule $\delta(u\!\sim, a) = (ua)\!\sim$ for $u \in A^*$ and $a \in A$. A language is accepted by an automaton if and only if it is a union of certain classes of the corresponding right congruence. The basic tool for proving non-regularity of languages is the well-known pumping lemma (see e.g. [27]).

Let $\mathfrak{S} = (S, *)$ be a semigroup and let $\sigma\colon A^+ \to \mathfrak{S}$ be a homomorphism. We say that an ε-free language $L \subseteq A^+$ is *recognized* by σ if $\sigma^{-1}\sigma(L) = L$, i.e. if there exists a subset $T \subseteq S$ such that $L = \sigma^{-1}(T)$. Analogously to the case of automata and right congruences, homomorphisms from A^+ to finite semigroups correspond to congruences of A^+ of finite index. The *syntactic congruence* $\equiv_L$ of a language $L \subseteq A^+$ is the congruence of the free semigroup A^+ defined by the condition

$$u \equiv_L v \iff (\forall\, x, y \in A^*)(xuy \in L \iff xvy \in L)\,.$$

In other words, the relation $\equiv_L$ is the largest congruence of A^+ such that the corresponding projection homomorphism recognizes L. The factor semigroup $A^+/\equiv_L$ is called the *syntactic semigroup* of L and denoted $\mathcal{S}(L)$; the projection homomorphism $\sigma_L\colon A^+ \to \mathcal{S}(L)$ is referred to as the *syntactic homomorphism* of L. It is well-known that an ε-free language is regular if and only if its syntactic semigroup is finite. Recognizability of arbitrary languages by monoids is defined analogously as recognizability of ε-free languages by semigroups.

Regular operations on languages are union, concatenation given by the rule $K \cdot L = \{uv \mid u \in K,\ v \in L\}$ and Kleene star defined by $L^* = L^+ \cup \{\varepsilon\}$, where $L^+ = \bigcup_{m\in\mathbb{N}} L^m$. Kleene's theorem states that regular languages are exactly those languages which can be obtained from finite languages using regular operations. And a regular language is called *star-free* if it can be constructed from finite languages using the operations of union, complementation and concatenation.

Solutions of language inequalities employing other than regular operations and possibly infinitely many variables will be described using the following notation. Let E be an arbitrary expression built from languages over A (called constants) and variables from an infinite set $\mathcal{V}$ using some symbols for language operations and let $\alpha\colon \mathcal{V} \to \wp(A^*)$ be a mapping assigning to each variable some language over A. Then $\alpha(E)$ denotes the language obtained by replacing each occurrence of every variable $X \in \mathcal{V}$ in E with the language $\alpha(X)$ and evaluating the resulting expression. A language inequality is a formal inequality $E \subseteq F$ of two expressions over constant languages and variables. A solution of the inequality $E \subseteq F$ is any mapping $\alpha\colon \mathcal{V} \to \wp(A^*)$ satisfying $\alpha(E) \subseteq \alpha(F)$. We call such a solution α regular if all the languages $\alpha(X)$, for $X \in \mathcal{V}$, are regular. Solutions of a given system of language inequalities are partially ordered by componentwise inclusion

$$\alpha \leq \beta \iff \forall X \in \mathcal{V}\colon \alpha(X) \subseteq \beta(X)\,,$$

and when $\alpha \leq \beta$ holds, we say that α is contained in β. We are mainly interested in solutions maximal with respect to this ordering (notice that if a variable $X \in \mathcal{V}$ does not occur in the system, then $\alpha(X) = A^*$ for every maximal solution α).

3. Bounded by Constants

The first implicit language inequalities which were studied are inequalities that employ only the operations of union and concatenation and whose right-hand sides are constant languages. It turns out that calculation of maximal solutions of such inequalities can be performed in a monoid recognizing these constant languages and therefore maximal solutions of such inequalities are regular.

Very simple inequalities of this form are in fact considered already when one defines the notion of minimal automaton. It is well-known that states of the minimal automaton of a given regular language L can be identified with languages of the form $w^{-1}L$, where $w \in A^*$ is an arbitrary word, which are precisely largest solutions of inequalities $wX \subseteq L$. The transition function of the automaton is then defined by setting $\delta(w^{-1}L, a) = (wa)^{-1}L$ for every $a \in A$, the initial state is L and terminal states are $w^{-1}L$ for $w \in L$.

Similarly, maximal solutions of the inequality $XY \subseteq L$ in two variables X and Y are precisely states of the so-called universal (non-deterministic) automaton of the language L (see [28]). There is a transition labelled by a letter $a \in A$ from a maximal solution (M, N) to a maximal solution (P, Q) if and only if $MaQ \subseteq L$ holds; a maximal solution (M, N) represents an initial state of this automaton if $\varepsilon \in M$ and it represents a terminal state if $\varepsilon \in N$.

A general form of inequalities whose right-hand sides are constant regular languages was considered by Conway [6]. He observed that inequalities of this form possess only finitely many maximal solutions, all of them are regular and computable. More precisely, every component of a maximal solution of such an inequality $E \subseteq L$, where E is a regular expression and L a regular language, is a union of certain classes of the syntactic congruence of L. In particular, this result leads to an algorithm for calculating best approximations of a given regular language by other given languages. Let us state this result for the case of finite systems of inequalities.

Theorem 1. *Let $n \in \mathbb{N}$ and for $i = 1, \dots, n$ let L_i be a regular language over A and E_i be an expression of the form $\bigcup_{k \in K} e_k$, where K is a (possibly infinite) index set and each expression e_k is a product of arbitrary constant languages and variables from $\mathcal{V}$. Let Σ be the system $\{E_1 \subseteq L_1, \dots, E_n \subseteq L_n\}$ of language inequalities. Then every solution of the system Σ is contained in a regular solution of Σ; in particular, every maximal solution of Σ is regular. If only finitely many variables occur in Σ, then Σ has only finitely many maximal solutions and every solution of Σ is contained in a maximal one.*

Proof. Let $\sigma \colon A^* \to \mathfrak{S}$ be a homomorphism to a finite monoid recognizing all languages L_i for $i = 1, \dots, n$. Consider an arbitrary solution $\alpha \colon \mathcal{V} \to \wp(A^*)$ of Σ. Define a mapping $\beta \colon \mathcal{V} \to \wp(A^*)$ by the rule $\beta(X) = \sigma^{-1}\sigma(\alpha(X))$ for all $X \in \mathcal{V}$, i.e. we add to each component $\alpha(X)$ of the solution α all words which are mapped by σ to the same element of $\mathfrak{S}$ as some word belonging to $\alpha(X)$. Then it is clear that all languages $\beta(X)$ are recognized by σ and because σ is a homomorphism, for every $i \in \{1, \dots, n\}$ we have $\sigma(\beta(E_i)) = \sigma(\alpha(E_i)) \subseteq \sigma(L_i)$ and consequently $\beta(E_i) \subseteq L_i$ since σ recognizes L_i. Therefore every solution of Σ is contained in a regular solution of Σ whose every component is recognized by σ. Then it is easy to see that maximal solutions of Σ are exactly solutions of Σ maximal among solutions whose components are recognized by σ. The second statement of the theorem now immediately follows from the fact that there are only finitely many languages recognizable by σ. $\square$

4. One-Sided Linearity Can Be Managed

It is one of the basic facts about regular languages that they are precisely languages definable by right-linear (or left-linear) grammars. Because languages generated by such grammars are nothing but smallest solutions of systems of explicit right-linear or left-linear equations and it is easy to see that also largest solutions of such systems are always regular, it is natural to ask whether the same can be proved true also for largest solutions of systems of implicit one-sided linear equations.

This question was answered by Baader and Küsters [1], who showed that largest solutions of finite systems of equations of the form

$$K \cup X_1K_1 \cup \cdots \cup X_nK_n = L \cup X_1L_1 \cup \cdots \cup X_nL_n\,,$$

where K, K_1, …, K_n and L, L_1, …, L_n are regular languages, are regular and their computation is an ExpTime-complete problem. Notice that if such a system is solvable, it possesses the largest solution over a given alphabet (this is true because concatenation distributes over infinite union and since no two variables are concatenated in these equations, the union of all solutions is again a solution of this system). We will meet other systems of inequalities having this property in Subsection 5.2 and Sections 6 and 7.

Similarly to other types of inequalities considered in this article, the result of Baader and Küsters can be generalized to systems of implicit left-linear inequalities where constant languages are required to be regular only on right-hand sides of inequalities. For the detailed proof of this result see [18].

Theorem 2. *Let $m, n \in \mathbb{N}$ and let Σ be the system of language inequalities*

$$K_j \cup X_1K_{1j} \cup \cdots \cup X_nK_{nj} \subseteq L_j \cup X_1L_{1j} \cup \cdots \cup X_nL_{nj}\,, \quad j = 1, \ldots, m\,,$$

where L_j, L_{1j}, …, L_{nj} are regular languages and K_j, K_{1j}, …, K_{nj} are arbitrary languages over A. If the system Σ possesses a solution, then the largest solution of Σ is regular.

In order to prove this result one considers a homomorphism σ to a finite monoid recognizing all languages L_j and L_{ij} and uses it to construct a right congruence $\sim$ of finite index on A^* such that every solution of Σ is contained in a regular solution of Σ whose every component is a union of certain classes of this right congruence.

Notice that in order to test whether a given word u belongs to the i-th component of the largest solution of Σ, we have to append to u some word x from one of the languages K_{ij} and then ask whether either the word ux belongs to L_j or we can remove from ux a suffix belonging to a certain language L_{kj} in such a way that the remaining prefix $\bar{u}$ of ux belongs to the k-th component of the solution. In the latter case, the same test can be applied to the resulting word $\bar{u}$. Continuing this way, each test produces a new word $\bar{u}$ which is either a prefix of u or it is constructed from such a prefix v by appending a certain word w obtained by successive addition of elements of some languages $K_{\hat{\imath}\hat{\jmath}}$ and removal of elements of some $L_{\hat{k}\hat{\jmath}}$ without further modifying this prefix. The proof of the theorem is based on the fact that the way in which possible words w are constructed depends only on the component of the solution where v belongs to, and not on the actual word v. And when we are eventually going to remove not only a suffix of w but the whole word w together with a suffix of v, we only need to know for every language $L_{\hat{k}\hat{\jmath}}$ the behaviour with respect to this language of all suffixes of v whose removal produces a prefix of u

belonging to the $\hat{k}$-th component of the solution. In other words, we are interested only in the σ-images of these suffixes of v.

Therefore we consider for every word u all possible choices for membership of prefixes of u in components of a solution of Σ, and for each such guess of a solution classify σ-images of differences between prefixes of u belonging to certain components of the solution. The required right congruence $\sim$ of A^* is defined in such a way that for arbitrary $\sim$-equivalent elements the sets of the results of this classification obtained for all possible guesses of a solution are the same. And for an arbitrary solution of Σ, a larger solution respecting $\sim$ can be constructed as follows: We put a word u into a new solution if we can prescribe membership of prefixes of u to the components of this solution in such a way that the resulting classification is the same as for some word of the original solution.

Let us remark that the right congruence $\sim$ can be calculated algorithmically. Therefore, given right-hand sides of our inequalities, we can algorithmically find all (finitely many) possible components of the largest solution of our system. In particular, if all constant languages occurring in the system are regular, then solvability of the system can be algorithmically decided and the largest solution can be computed.

5. Under Control of Well Quasi-orders

Notice that all types of inequalities considered in the previous sections have some common features. More precisely, right-hand sides of these inequalities define a (one-sided) congruence of the free monoid of finite index, this congruence can be algorithmically calculated from syntactic semigroups of constant languages, and every component of every maximal solution is a union of certain classes of this congruence. This in particular means that once the right-hand sides are fixed, there are only finitely many possible components of maximal solutions. Moreover, arbitrary finite systems of such inequalities can be treated in the same way as single inequalities.

In this section, we relax our requirements on the structure determining possible components of maximal solutions; instead of using equivalence relations of finite index, we consider arbitrary well quasi-orders of the free monoid. There are already many results on regularity of languages based on the concept of well quasi-orders (see [7]). The most important result of this kind is a generalization of Myhill theorem due to Ehrenfeucht et al. [8] stating that a language is regular if and only if it is upward closed with respect to a monotone well quasi-order.

The method of demonstrating that maximal solutions of certain systems of inequalities are regular presented in this section is based on constructing a suitable well quasi-order of the free monoid and showing that every solution of our system is in fact contained in some solution upward closed with respect to this quasi-order. First we apply this method to prove that all maximal solutions of systems of inequalities of a very general form (involving even infinitary union and intersection operations) are regular when only constant languages recognized by finite simple semigroups are allowed (this in particular covers the case of group languages). Then in Subsection 5.2 we explain how well quasi-orders can be used to show that the largest solution of the inequality $XK \subseteq LX$ is regular provided the language L is regular. Results described in this section were proved in [16].

Let us now recall the concept of well quasi-orders. A *quasi-order* $\leq$ on a set S is a reflexive and transitive binary relation. We say that a subset T of S is *upward closed* with respect to $\leq$ if for every $t \in T$ and $s \in S$, the inequality $t \leq s$ implies $s \in T$.

Definition 1. A quasi-order $\leq$ on a set S is called a *well quasi-order* if the following equivalent conditions are satisfied:

(1) There exists neither an infinite strictly descending sequence in S nor an infinite sequence of mutually incomparable elements of S.
(2) For every subset $T \subseteq S$ there exists a finite subset U of T such that for each $t \in T$ there exists some element $s \in U$ satisfying $s \leq t$.
(3) There does not exist an infinite sequence of upward closed subsets of S strictly ascending with respect to inclusion.

A basic example of a well quasi-ordering of words is the (scattered) subword ordering; it is a well quasi-ordering due to Higman's Lemma [9]. Ehrenfeucht et al. [8] proved a generalization of this fact stating that the derivation relation on the set A^* associated with the set of context-free rules $\{\varepsilon \to u_1, \ldots, \varepsilon \to u_n\}$ is a well quasi-order if and only if the language $\{u_1, \ldots, u_n\}$ is unavoidable over A. Another generalization of this result to derivation relations of arbitrary context-free systems due to Bucher et al. [3] is in fact used to prove Theorem 4.

5.1. **When Decompositions Define Well Quasi-orders.** Let us start this subsection by describing systems of inequalities which will be considered here. Let $\mathfrak{L}$ be a finite set of ε-free languages over A. We say that an inequality $E \subseteq F$ is an $\mathfrak{L}$-*inequality* if the expression E is a product of variables and arbitrary constants and the expression F is built from variables and languages belonging to the set $\mathfrak{L} \cup \{\{\varepsilon\}\}$ using symbols for the operations of concatenation, arbitrary (possibly infinite) union and arbitrary (possibly infinite) intersection.

Let $\sigma\colon A^+ \to \mathfrak{S}$ be a homomorphism onto a finite semigroup $\mathfrak{S}$. We define a quasi-order $\leq_\sigma$ on A^* by setting $v \leq_\sigma u$ if and only if $v = a_1 \cdots a_n$, where $a_1, \ldots, a_n \in A$, and $u = u_1 \cdots u_n$, where $u_j \in A^+$ and $\sigma(u_j) = \sigma(a_j)$ for $j = 1, \ldots, n$. This quasi-order is monotone, i.e. from $v_1 \leq_\sigma u_1$ and $v_2 \leq_\sigma u_2$ it follows that $v_1 v_2 \leq_\sigma u_1 u_2$. Notice that if $v \leq_\sigma u$ then either $u = v = \varepsilon$ or $u, v \in A^+$ and $\sigma(v) = \sigma(u)$.

The following theorem states that all maximal solutions of arbitrary systems of $\mathfrak{L}$-inequalities are regular provided there exists a homomorphism σ recognizing all languages from $\mathfrak{L}$ for which the relation $\leq_\sigma$ is a well quasi-order.

Theorem 3. *Let $\mathfrak{L}$ be a finite set of ε-free languages over A and let $\sigma\colon A^+ \to \mathfrak{S}$ be a homomorphism onto a finite semigroup $\mathfrak{S}$ recognizing all languages in $\mathfrak{L}$, and such that A^* is well quasi-ordered by $\leq_\sigma$. Let I be an arbitrary (possibly infinite) set and let $\Sigma = \{E_i \subseteq F_i \mid i \in I\}$ be a system of $\mathfrak{L}$-inequalities. Then every solution of Σ is contained in a regular solution of Σ; in particular, every maximal solution of the system Σ is regular. If only finitely many variables occur in Σ, then every solution of Σ is contained in a maximal one.*

Proof. Let $\alpha\colon \mathcal{V} \to \wp(A^*)$ be a solution of Σ. For every $X \in \mathcal{V}$ define the language

$$\beta(X) = \{u \in A^* \mid \exists v \in \alpha(X)\colon v \leq_\sigma u\}\,.$$

It is clear that $\alpha(X) \subseteq \beta(X)$. We are going to show that β is a regular solution of Σ.

First observe that because the quasi-order $\leq_\sigma$ is monotone, if a word u belongs to the language $\beta(E_i)$, there exists $v \in \alpha(E_i)$ such that $v \leq_\sigma u$. We prove by induction with respect to the structure of the expression F_i that if $v \in \alpha(F_i)$ and $v \leq_\sigma u$ for some words u and v, then $u \in \beta(F_i)$, which is enough to conclude that β is a solution of Σ. So assume a word v belongs to $\alpha(e)$ for some subexpression e of F_i and $v \leq_\sigma u$. If e is a variable, we have $u \in \beta(e)$ by the definition of β. In the case e is a language from $\mathfrak{L}$, one obtains $u \in \beta(e)$ from the fact $\sigma(u) = \sigma(v)$. For $e = \{\varepsilon\}$, the only possibility is $u = v = \varepsilon \in \beta(e)$. If the expression e is of the form $\bigcup_{k\in K} e_k$ or $\bigcap_{k\in K} e_k$ for some set K, then $u \in \beta(e)$ is clear from the induction hypothesis. Finally, consider the case $e = e_1 \cdot e_2$. Then $v = v_1 \cdot v_2$, where $v_1 \in \alpha(e_1)$ and $v_2 \in \alpha(e_2)$. From $v \leq_\sigma u$ we deduce $v_1 = a_1 \cdots a_m$ and $v_2 = a_{m+1} \cdots a_n$, where $0 \leq m \leq n$ and $a_j \in A$, and $u = u_1 \cdots u_n$ for some words $u_1, \ldots, u_n \in A^+$ satisfying $\sigma(u_j) = \sigma(a_j)$ for $j = 1, \ldots, n$. Therefore $v_1 \leq_\sigma u_1 \cdots u_m$ and $v_2 \leq_\sigma u_{m+1} \cdots u_n$ and we can apply the induction hypothesis to these words. Hence $u \in \beta(e)$.

In order to prove that $\beta(X)$ is a regular language, observe that $\beta(X)$ is upward closed with respect to the well quasi-order $\leq_\sigma$, therefore it can be generated by finitely many elements of A^* due to condition 2 of Definition 1, i.e. $\beta(X)$ is a union of finitely many languages of the form $\langle v \rangle = \{u \in A^* \mid v \leq_\sigma u\}$ for a word $v \in A^*$. And it is easy to see that for arbitrary letters $a_1, \ldots, a_n \in A$ we have

$$\langle a_1 \cdots a_n \rangle = (\sigma^{-1}\sigma(a_1)) \cdots (\sigma^{-1}\sigma(a_n)),$$

which shows that each language $\langle v \rangle$ is regular.

We have already proved that every solution of Σ is contained in a regular solution whose every component is a language upward closed with respect to the well quasi-order $\leq_\sigma$. Because there is no infinite strictly ascending sequence of such upward closed sets by condition 3 of Definition 1, this immediately implies that if there are only finitely many variables, every solution is in fact contained in a maximal solution. □

In order to give a precise characterization of those semigroup homomorphisms $\sigma\colon A^+ \to \mathfrak{S}$ for which the relation $\leq_\sigma$ is a well quasi-order on A^* and so Theorem 3 can be applied, we need a few notions from semigroup theory.

An *ideal* of a semigroup $(S, *)$ is a non-empty subset $I \subseteq S$ such that for all $s \in I$ and $t \in S$ we have $s * t \in I$ and $t * s \in I$. A semigroup is called *simple* if it has no proper ideal. For example, every group is a simple semigroup. We say that a semigroup $(S, *)$ is a *chain* if it is commutative and for all elements $s, t \in S$ either $s * t = s$ or $s * t = t$. If a semigroup $\mathfrak{S}$ possesses a congruence relation $\equiv$ such that the factor semigroup $\mathfrak{S}/\equiv$ is a chain, then every congruence class $(s\equiv)$, for $s \in S$, is a subsemigroup of $\mathfrak{S}$ and the semigroup $\mathfrak{S}$ is called a *chain of semigroups* $(s\equiv)$.

Theorem 4. *For an arbitrary homomorphism $\sigma\colon A^+ \twoheadrightarrow \mathfrak{S}$ onto a finite semigroup, the relation $\leq_\sigma$ is a well quasi-order on A^* if and only if $\mathfrak{S}$ is a chain of simple semigroups.*

When we are concerned with regular languages, the most interesting inequalities are those built using regular operations. Systems of such inequalities are in fact a special case of systems considered in this subsection: the star operation is constructed from the operations of concatenation and infinite union and as our systems are allowed to be infinite, we can actually use the operation of infinite union (and consequently the star operation) also on left-hand sides of inequalities. Because all languages from a given set $\mathfrak{L}$ can be recognized by the product of semigroups recognizing individual languages and any product of simple semigroups is again simple, the following result is an immediate consequence of Theorems 3 and 4.

Corollary 1. *Let Σ be a finite system of inequalities of the form $E_i \subseteq F_i$, where E_i are arbitrary regular expressions and F_i are regular expressions over variables, the language $\{\varepsilon\}$ and regular languages recognizable by homomorphisms to finite simple semigroups. Then every solution of Σ is contained in a maximal solution and every maximal solution of Σ is regular.*

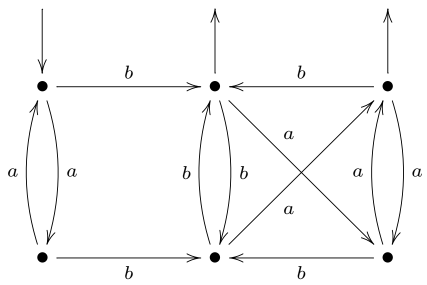
If only one constant language occurs on the right hand sides of inequalities, then in order to apply Theorems 3 and 4 it is sufficient to know that the language is recognized by a chain of finite simple semigroups. Notice that unlike for languages recognized by groups or simple semigroups, recognizability of a regular language by a chain of simple semigroups is independent of the underlying alphabet since additional letters not employed by the language form a zero element in the syntactic semigroup, which becomes the least element of the chain.

Corollary 2. *Let $L \subseteq A^+$ be a regular language recognized by a chain of finite simple semigroups. Let Σ be a system of inequalities of the form $E_i \subseteq F_i$ over finitely many variables, where E_i are arbitrary regular expressions and F_i are regular expressions over variables and the languages L and $\{\varepsilon\}$. Then every solution of Σ is contained in a maximal solution and every maximal solution of Σ is regular.*

Let us now mention how constant languages allowed in the previous results can be characterized by means of their minimal automata. Recall that languages recognizable by finite groups are precisely those regular languages whose minimal automaton is codeterministic, i.e. contains no distinct states p and q such that $\delta(p, a) = \delta(q, a)$ for some $a \in A$. This condition can be transformed into conditions corresponding to the cases of simple semigroups and chains of simple semigroups by considering codeterminism for two-letter (or longer) words instead of single letters. More precisely, a regular ε-free language L over A can be recognized by a finite simple semigroup if and only if its minimal automaton contains no states p and q which for some letters $a, b \in A$ satisfy $\delta(p, a) \neq \delta(q, a)$ and $\delta(p, ab) = \delta(q, ab)$. And it can be recognized by a finite chain of finite simple semigroups if and only if there exists a total ordering $\leq$ of A such that the minimal automaton of L contains no states p and q for which there exist a word $u \in A^*$ and letters $a, b \in A$ satisfying $a \leq b$, $\delta(p, au) \neq \delta(q, au)$ and $\delta(p, aub) = \delta(q, aub)$.

The following example illustrates Theorem 3 on a concrete non-trivial inequality with one constant language recognized by a chain of simple semigroups.

Example 1. Let L be the ε-free language over the alphabet $A = \{a, b\}$ whose minimal automaton is the following:



The language L consists exactly of those words $u \in A^+$ which contain some occurrence of b and where the difference between the length of u and the number of blocks of occurrences of b in u is even. The syntactic semigroup of L is defined by the relations $a^3 = a$, $b^3 = b$, $ab^2 = a^2b$, $ba^2 = b^2a$ and $bab = b^2$; it is a chain of two simple semigroups whose elements are represented by the words a, a^2 and b, b^2, ab, ab^2, ba, b^2a, aba, ab^2a, respectively. The above condition on minimal automaton corresponding to chains of simple semigroups can be verified for L by setting $b < a$.

Let us consider the inequality $aXaXa \subseteq LXL$ with one variable X. It is easy to verify that this inequality possesses a largest solution, namely the regular language $(a^2)^*ab^2a(a^2)^* \cup A^*bA^+bA^*$. In the proof of Theorem 3 we have seen that this solution is upward closed with respect to the well partial order $\leq_{\sigma_L}$. In fact, there are precisely 87 minimal elements in this solution with respect to $\leq_{\sigma_L}$:

$$\{ab^2a\} \cup \big(\{\varepsilon, a, a^2\} \cdot (\{b, b^2\}\{a, a^2\}\{b, b^2\} \cup \{b^3, b^4, babab\}) \cdot \{\varepsilon, a, a^2\}\big) \setminus \\ \setminus (\{\varepsilon, a, a^2\}b^2a^2b^2\{\varepsilon, a, a^2\} \cup \{ab^4a, ab^2ab^2a, aba^2b^2a, ab^2a^2ba\})\,.$$

5.2. **Semi-commutation Inequalities.** This subsection is devoted to another application of well quasi-orders to proving regularity of solutions. We explain here why the largest solution of the inequality $XK \subseteq LX$ is always regular provided L is a regular language. This time, in order to define a suitable well quasi-order on A^*, we have to consider more involved structures than just plain sequences as we did in Subsection 5.1.

The basic idea of the proof is to think of the inequality $XK \subseteq LX$ as a game of two players, the attacker and the defender. The language K determines possible actions of the attacker and the language L determines possible actions of the defender. A position of the game is an arbitrary word w from A^*. At each step of the game, both players successively modify the word according to the following rules. When the game is in a position w, the attacker chooses any element v of K and appends it to w. If no word from L is a prefix of wv, the attacker wins. Otherwise the defender removes any word belonging to L from the beginning of wv. The resulting word is a new position of the game. The defender wins the game if and only if he manages to continue playing forever.

Observe that if the defender has a winning strategy for a given initial position $w \in A^*$, then the set of all positions of the game reachable from w in some scenario corresponding with a chosen winning strategy forms a solution of the inequality $XK \subseteq LX$ containing w. Conversely, given any solution $M \subseteq A^*$ of $XK \subseteq LX$, one can easily construct winning strategies of the defender for all elements of M.

Therefore the largest solution of the inequality $XK \subseteq LX$ is exactly the set of all positions of the game where the defender has a winning strategy. The result described in this subsection can then be reformulated as follows: If the set of possible actions of the defender is regular, then the set of all winning positions of the defender is regular no matter what actions are available to the attacker.

Given an initial position $w \in A^*$, we consider all possible sequences of actions of the defender which can be performed without removing any letters previously added by the attacker. In other words, we deal with all sequences $(w_1, \ldots, w_n)$ of elements of L whose concatenation $w_1 \cdots w_n$ is a prefix of the word w. We arrange these sequences into the form of a tree expressing the order of actions, i.e. the node $(w_1, \ldots, w_n)$ of the tree will be an immediate successor of the node $(w_1, \ldots, w_{n-1})$. Then every move of the defender can be seen as a choice of one of the immediate successors of the node representing the current position of the game. In addition, we have to consider for each node $(w_1, \ldots, w_n)$ the suffix u of w satisfying $w = w_1 \cdots w_n u$. This word u can be removed by the defender in the following turn together with several letters previously added by the attacker. The only information the defender needs to know when removing u is which words can be appended to u in order to get a word from L. This is uniquely determined by the $\equiv_L$-class of u, and therefore it is sufficient to label the node $(w_1, \ldots, w_n)$ with the element $\sigma_L(u) \in \mathcal{S}(L)$.

In this way, we construct a labelled tree for every $w \in A^*$. Then one can introduce a quasi-order on the set of such trees expressing possibility of using winning strategies for one initial position of the game also for another one and prove that the largest solution of the inequality $XK \subseteq LX$ is upward closed with respect to the quasi-order $\leq_L$ induced on A^*. Regularity of the largest solution can then be deduced from the fact that the resulting quasi-order on trees is actually a well quasi-order due to Kruskal's Tree Theorem [15].

Theorem 5. *If $K \subseteq A^*$ is an arbitrary language and $L \subseteq A^*$ is a regular language, then the largest solution of the inequality $XK \subseteq LX$ is regular.*

Although Theorem 5 does not provide us with an algorithm calculating largest solutions of inequalities of the form $XK \subseteq LX$, where K and L are regular languages, it enables us to algorithmically decide whether a given word belongs to the largest solution of such an inequality, because both the largest solution and its complement can be recursively enumerated. First, in [23] it is proved that complements of largest solutions of arbitrary finite systems of language inequalities with Boolean operations are recursively enumerable. Actually, a uniform procedure for all such equations is described there. This in particular means that there exists a procedure which for given regular languages K and L recursively enumerates all words over A which do not belong to the largest solution of the inequality $XK \subseteq LX$. This result is based on the observation that if the defender has no winning strategy for a given position $w \in A^*$, then the attacker has a finite winning strategy for w. The reason why this is true is that every word has only finitely many prefixes which the defender can choose from. On the other hand, there exists also a procedure which for given regular languages K and L recursively enumerates all words from the largest solution of $XK \subseteq LX$; because the largest solution is regular by Theorem 5, it is sufficient to enumerate for every word w all regular languages containing w and for each of them test whether it is a solution. Altogether, we obtain the following corollary:

Corollary 3. *There exists an algorithm which decides for given regular languages K and L and a word $w \in A^*$ whether w belongs to the largest solution of the inequality $XK \subseteq LX$.*

Now we give an example which in particular shows that it is essential to consider the whole tree structure associated with each word, not only the corresponding elements of the syntactic semigroup and the lengths of paths in the tree.

Example 2. Consider the alphabet $A = \{a, b, c, d, e, f, g, h, i\}$ and let $K = \{e, i\}$ and

$$L = \{a, b, bc, f, fg, gh, hc, e, i\} \cup cdKiK \cup dKeK \cup g^*\{b, h\}cdK^3 .$$

One can show that the largest solution of the inequality $XK \subseteq LX$ is the language

$$L^* \cup L^*cdKi \cup L^*dKe \cup L^*g^*\{b, h\}cdK^2 \cup \\ \cup L^+g^*\{b, h\}cdK \cup bcdK \cup LL^+g^*\{b, h\}cd \cup Lbcd .$$

In order to calculate this solution, observe that if u belongs to a solution M of the inequality, then for every $n \in \mathbb{N}$ we have $ue^n \in MK^n \subseteq L^nM$. Because $\varepsilon \notin L$, if we take n sufficiently large, we deduce that the word ue^3 is a prefix of a word from L^*. Since this is possible only if $ue^3 \in L^*$, there remain only few cases to deal with.

We know that the largest solution of the inequality $XK \subseteq LX$ is upward closed with respect to the well quasi-order $\leq_L$, therefore one can find finitely many labelled trees characterizing the solution, i.e. corresponding to minimal elements of the solution. In our case, there are four one-node trees corresponding to the words ε, cdi^2, de^2 and $gbcde^2$, which generate the languages L^*, L^*cdKi, L^*dKe and $L^*g^*\{b, h\}cdK^2$, respectively. Further there are four trees with more than one node corresponding to the words $agbcde$, $bcde$, a^2gbcd and $abcd$, respectively:

$$\begin{tikzcd} 0 \arrow[d] \\ bcde \end{tikzcd} \qquad \begin{tikzcd} & bcde \arrow[dl] \arrow[dr] & \\ cde & & de \end{tikzcd} \qquad \begin{tikzcd} 0 \arrow[d] \\ 0 \arrow[d] \\ bcd \end{tikzcd} \qquad \begin{tikzcd} & 0 \arrow[d] & \\ & bcd \arrow[dl] \arrow[dr] & \\ cd & & d \end{tikzcd}$$

where each node is labelled with the σ_L-image of the corresponding word and by the symbol 0 we mean the zero element of the semigroup $\mathcal{S}(L)$. These four words generate the languages $L^+g^*\{b, h\}cdK$, $L^*\{b, fgh\}cdK$, $LL^+g^*\{b, h\}cd$ and $L^+\{b, fgh\}cd$, respectively.

Finally, let us point out that the word $fghcd$ does not belong to the solution even though the tree corresponding to it is very similar to the one of $abcd$ described above:

$$\begin{tikzcd} & 0 \arrow[dl] \arrow[dr] & \\ bcd \arrow[d] & & bcd \arrow[d] \\ cd & & d \end{tikzcd}$$

This is a consequence of the facts $abcd \equiv_L fghcd$ and $bcd \equiv_L hcd \equiv_L ghcd$. When the inequality is viewed as a game, the difference between these two trees is that for the word $fghcd$ the defender has to make his decision immediately after the first turn of the attacker whereas for the word $abcd$ he can decide according to the attacker's second move.

6. Well Quasi-orders Cannot Be Combined

A common property of the positive results mentioned in Sections 3 and 4 is that they can be formulated not only for single equations, but also for arbitrary finite systems of such equations. The reason for this is that the equivalence relations on A^* constructed for individual inequalities can be easily combined into an equivalence relation recognizing maximal solutions of a system. But when well quasi-orders are used to deal with individual inequalities instead of equivalence relations, it may not be possible to create a well quasi-order which demonstrates regularity of maximal solutions for a given system consisting of two inequalities. The aim of this section is to show that this happens in the case of inequalities of the form $XK \subseteq LX$, where K and L are regular languages. Results presented here are proved in [19].

In Section 5 we have seen that for regular languages K and L the largest solution of the inequality $XK \subseteq LX$ is regular. In this section we will be concerned with results stating that when we either consider a system of two inequalities of this form or add to such an inequality an additional restriction on the solution provided by a regular language, we can obtain a system whose largest solution is not recursively enumerable. Moreover, these results hold even if we consider only finite languages K and star-free languages L. On the other hand, since complements of largest solutions of such systems of inequalities are always recursively enumerable (see [23]), these results are the most negative possible.

Let us first consider systems consisting of one inequality $XK \subseteq LX$ and an additional constraint $X \subseteq M$.

Theorem 6. *There exists a finite language K and star-free languages L, M such that the largest solution of the system*

$$XK \subseteq LX, \; X \subseteq M \tag{1}$$

is not recursively enumerable.

In the following example we demonstrate the idea used to prove this theorem by encoding testing of equality of two counters into the largest solution of a system of the form (1), which is enough to conclude that the solution is non-regular. The general result can be achieved in the same way as this example by encoding an arbitrary Minsky machine.

Example 3. Consider the alphabet $B = \{a, b, c\}$ together with its disjoint copy $\hat{B} = \{\hat{a}, \hat{b}, \hat{c}\}$ and let $A = B \cup \hat{B}$. Then define languages $K = \{a\hat{a}, b\hat{b}, c\hat{c}\}$,

$$\begin{aligned} L = {} & \{a\hat{a}, b\hat{b}a\hat{a}, c\hat{c}a\hat{a}, b\hat{b}c\hat{c}b\} \cup \hat{B}B \\ & \cup c \cdot (A \setminus \{c\})^* \cdot a \cup b \cdot (A \setminus \{b\})^* \cdot a \cup A^+ bA^* b \cup A^+ cA^* c \end{aligned}$$

and $M = (B\hat{B})^+ \cup \hat{B}$ over A and let us denote by S the largest solution of the system (1) obtained this way. We are going to prove that the language S is not regular by means of the pumping lemma.

First we show that for every $n \in \mathbb{N}_0$ the word $b\hat{b}(a\hat{a})^n c\hat{c}(a\hat{a})^n$ belongs to S by constructing a solution N of (1) which contains this word. The required solution is defined as

$$N = \hat{B} \cup \{u_{n,m}, v_{n,m} \mid n \in \mathbb{N}_0,\ 0 \le m \le n\}\,,$$

where $u_{n,m} = (a\hat{a})^m b\hat{b}(a\hat{a})^n c\hat{c}(a\hat{a})^{n-m}$ and $v_{n,m} = (a\hat{a})^m c\hat{c}(a\hat{a})^{n+1} b\hat{b}(a\hat{a})^{n-m}$.

Let us verify that $NK \subseteq LN$. Clearly $\hat{B}K \subseteq (\hat{B}B) \cdot \hat{B} \subseteq LN$. Now let us show that $u_{n,m}K \subseteq LN$. First, we have

$$u_{n,0} \cdot a\hat{a} \in (b \cdot (A \setminus \{b\})^* \cdot a) \cdot \hat{B} \subseteq LN\,,$$
$$u_{n,m+1} \cdot a\hat{a} = a\hat{a} \cdot u_{n,m} \in LN\,.$$

For the word $b\hat{b}$ we distinguish three cases:

$$u_{0,0} \cdot b\hat{b} = b\hat{b}c\hat{c}b \cdot \hat{b} \in LN\,,$$
$$u_{n+1,0} \cdot b\hat{b} = b\hat{b}a\hat{a} \cdot v_{n,n} \in LN\,,$$
$$u_{n,m+1} \cdot b\hat{b} \in (A^+bA^*b) \cdot \hat{B} \subseteq LN\,.$$

And finally, for the word $c\hat{c} \in K$ we directly obtain $u_{n,m} \cdot c\hat{c} \in (A^+cA^*c) \cdot \hat{B} \subseteq LN$.

For every word $v_{n,m} \in N$ the verification is very similar. And because the fact $N \subseteq M$ is trivial, we have actually shown that N is a solution of (1) and therefore the word $b\hat{b}(a\hat{a})^n c\hat{c}(a\hat{a})^n$ belongs to the largest solution S for every $n \in \mathbb{N}_0$.

Now we are going to show that $b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \notin S$ for every $m, n \in \mathbb{N}_0$ satisfying $m < n$. The proof proceeds by induction with respect to m.

First, let $m = 0$ and assume for contradiction that $b\hat{b}c\hat{c}(a\hat{a})^n \in S$. Then we obtain

$$b\hat{b}c\hat{c}(a\hat{a})^n b\hat{b} \in SK \subseteq LS \subseteq LM\,,$$

which is not true because $n \ge 1$. Therefore $b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \notin S$ holds for $m = 0$.

Now consider the case when $m \ge 1$. The induction hypothesis states that $b\hat{b}(a\hat{a})^{m-1} c\hat{c}(a\hat{a})^{n-1} \notin S$ and we prove the required fact $b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \notin S$ by contradiction. From $b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \in S$ it follows that

$$b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \cdot b\hat{b}(a\hat{a})^{m-1} \in SK^m \subseteq L^m S \subseteq L^m M\,.$$

One can see that every prefix of this word belonging to L^m is either $b\hat{b}(a\hat{a})^m$ or an element of $b \cdot (A \setminus \{b\})^* \cdot a \cdot (\hat{B}B)^{m-1}$. Because in the latter case, for none of such prefixes the corresponding suffix belongs to the language M, we deduce that $c\hat{c}(a\hat{a})^n b\hat{b}(a\hat{a})^{m-1} \in S$. Now we repeat the previous argument using the word $c\hat{c}(a\hat{a})^{n-1}$ from K^n instead of $b\hat{b}(a\hat{a})^{m-1}$ and obtain $b\hat{b}(a\hat{a})^{m-1} c\hat{c}(a\hat{a})^{n-1} \in S$, which contradicts the induction hypothesis.

Altogether, we have demonstrated both $b\hat{b}(a\hat{a})^n c\hat{c}(a\hat{a})^n \in S$ for every $n \in \mathbb{N}_0$ and $b\hat{b}(a\hat{a})^m c\hat{c}(a\hat{a})^n \notin S$ for $m < n$, and so the largest solution of (1) is not regular due to the pumping lemma.

Let us now formulate the analogue of Theorem 6 for systems of two inequalities of the form $XK \subseteq LX$.

Theorem 7. *There exist finite languages K, P and star-free languages L, Q such that the largest solution of the system*

$$XK \subseteq LX,\ XP \subseteq QX \tag{2}$$

is not recursively enumerable.

This result can be obtained by encoding system (1) of Theorem 6 into a system of the form (2). Let us demonstrate this by performing the encoding on the system constructed in Example 3.

Example 4. Consider the languages K, L and M over the alphabet A which were defined in Example 3. We enrich the alphabet A with a new letter by setting $\tilde{A} = A \cup \{d\}$ and define the languages $P = \{d\hat{a}\}$ and $Q = Md$ over $\tilde{A}$. We are going to prove that any language containing the word $\hat{a}$ is a solution of system (2) if and only if it is a solution of (1). This in particular means that the largest solutions of these systems are equal, which proves that the largest solution of (2) is not regular.

It is clear that every solution N of (1) satisfying $\hat{a} \in N$ is a solution of (2) too. Conversely, assume that a language N is a solution of (2). First notice that the inequality $XK \subseteq LX$ implies $NK^n \subseteq L^n N$ for arbitrarily large $n \in \mathbb{N}$ and, since $\varepsilon \notin L$, we have $N \subseteq A^*$. Therefore the letter d occurs only once in every word of the language $NP = Nd\hat{a}$ and from the definition of Q one can immediately see that the inclusion $NP \subseteq QN$ implies $N \subseteq M$.

7. Commutation Equations

In this section we deal with another situation where the largest solution of a system of two inequalities, whose largest solutions can be proved regular by means of well quasi-orders, is a non-regular language. Namely, we address the question whether the largest solution of any language equation of the form $XL = LX$ is regular provided L is a regular language. It is known that in several algebraic structures related to algebras of formal languages, two elements commute if and only if they can be generated by the same element. For instance, two words commute if and only if they are powers of the same word (due to the defect theorem) and such a characterization was proved valid also for polynomials and formal series in non-commuting variables over a field (in particular, for multisets of words) by Bergman [2] and Cohn [5], respectively. But in the case of languages the situation is completely different.

The problem of regularity of the largest language commuting with a given regular language was formulated already by Conway [6, p. 55 and 124] in 1971. The problem was recently studied in several articles (e.g. [26, 4, 13]), but affirmative answers were given only for at most ternary sets of words [12] and for regular codes [11]. Moreover, it even remained an open problem whether the largest language commuting with a given finite set of words is recursive. A summary of known results concerning commutation of languages and some examples can be found in the survey by Karhumäki and Petre [14].

Recently the author [17] gave the most negative possible answer to Conway's problem by proving the following result:

Theorem 8. *There exists a finite language L such that the largest language commuting with L is not recursively enumerable.*

This in particular means that the positive result of Subsection 5.2 is not preserved if we combine an inequality $XK \subseteq LX$ with a similar inequality where the order of concatenation is reversed. In addition, Theorem 8 disproves Conway's conjecture [6, p. 129] stating that every maximal solution of a system of so-called semi-linear inequalities is context-free.

Theorem 8 can be proved by encoding the complement of the language computed by an arbitrary Minsky machine into a solution of a commutation equation. We demonstrate this idea on an example which shows that the largest language commuting with a given star-free language need not be regular. Again, we encode into the language L two counters (as powers of a certain letter a) and the operation of simultaneous decrementation of both counters together with testing whether both counters are equal to zero.

Example 5. Let us take the alphabet $A = \{a, b, c, e, \hat{e}, f, \hat{f}, g, \hat{g}\}$. We consider auxiliary languages

$$M = efga^+ba^* \cup ga^*ba^*\hat{g}\hat{f} \cup a^*ba^*\hat{g}\hat{f}\hat{e} \cup fga^*ba^*\hat{g}\,,$$
$$L_0 = (A \setminus \{c\})^*b(A \setminus \{c\})^* \setminus (\{efg, fg, g, \varepsilon\} \cdot a^*ba^* \cdot \{\varepsilon, \hat{g}, \hat{g}\hat{f}, \hat{g}\hat{f}\hat{e}\})$$

and define a star-free language

$$L = \{c, ef, ga, e, fg, \hat{f}\hat{e}, a\hat{g}, \hat{e}, \hat{g}\hat{f}, fgba\hat{g}\} \cup A^*bA^*bA^* \cup L_0 \cup cM \cup Mc\,.$$

Let us verify that the largest solution N of the equation $XL = LX$ is not regular. First observe that $fga^mba^n\hat{g}\hat{f} \notin N$ for every $m, n \in \mathbb{N}_0$. Indeed, assuming the converse, we obtain $c^2fga^mba^n\hat{g}\hat{f} \in L^2N = NL^2$, which is a contradiction because this word has no suffix belonging to L^2.

Now we are going to demonstrate that $efga^mba^n \notin N$ for every $m, n \in \mathbb{N}_0$ satisfying $m < n$. We proceed by induction on m. If $m = 0$ then the converse of this fact implies $cefgba^n \in LN = NL$, which is impossible as no suffix of the word $cefgba^n$ lies in L. Let $m \geq 1$ and suppose we already know that $efga^{m-1}ba^{n-1} \notin N$. By way of contradiction, assume $efga^mba^n \in N$. Then $efga^mba^n\hat{g}\hat{f} \in NL = LN$. Since there are just two prefixes of $efga^mba^n\hat{g}\hat{f}$ which belong to L, we have either $fga^mba^n\hat{g}\hat{f} \in N$ or $ga^mba^n\hat{g}\hat{f} \in N$. Due to our initial observation, the former case is false. From the latter one we deduce $ga^mba^n\hat{g}\hat{f}\hat{e} \in NL = LN$. This immediately gives $a^{m-1}ba^n\hat{g}\hat{f}\hat{e} \in N$ and therefore $fga^{m-1}ba^n\hat{g}\hat{f}\hat{e} \in LN = NL$. The word $fga^{m-1}ba^n\hat{g}\hat{f}\hat{e}$ has exactly two suffixes from L and repeating the previous argument we get $fga^{m-1}ba^n\hat{g} \in N$. Consequently $efga^{m-1}ba^n\hat{g} \in LN = NL$, and since we have $n \geq 2$, this finally leads to $efga^{m-1}ba^{n-1} \in N$, contradicting our assumption.

On the other hand, the language

$$K = L^* \cup \{efga^nba^n, ga^nba^n\hat{g}\hat{f}, a^{n-1}ba^n\hat{g}\hat{f}\hat{e}, fga^{n-1}ba^n\hat{g} \mid n \in \mathbb{N}\}$$

commutes with L because one can easily calculate that both products KL and LK are equal to the language

$$L^+ \cup \{efga^nba^n\hat{g}\hat{f}, ga^nba^n\hat{g}\hat{f}\hat{e}, fga^{n-1}ba^n\hat{g}\hat{f}\hat{e}, efga^{n-1}ba^n\hat{g} \mid n \in \mathbb{N}\}\,;$$

notice that this holds due to the fact $efgba\hat{g} \in L^2$. Therefore the word $efga^nba^n$ belongs to N for all $n \in \mathbb{N}$.

Altogether, we have proved both $efga^nba^n \in N$ for $n \in \mathbb{N}$ and $efga^mba^n \notin N$ for $m < n$, hence the language N cannot be regular due to the pumping lemma.

8. Conclusion

Although there is already a number of both positive and negative results on regularity of solutions of implicit language inequalities, these results are more or less isolated and obtained using several different methods. Therefore the question of what do "simple language equations" look like seems to be far from being understood. But some of the results already indicate a certain borderline between those types of inequalities whose maximal solutions are regular and those where regularity of constants does not ensure regularity of maximal solutions. Moreover, one can find several common features of these results. For instance, regularity of maximal solutions is usually guaranteed by regularity of constant languages occurring on right-hand sides of inequalities, whereas on left-hand sides arbitrary constants can be used and restrictions have to be imposed only on the structure of the expression. Further, it appears that whenever systems of implicit inequalities of some simple form happen to have non-regular maximal solutions, then for some choice of regular constant languages, maximal solutions become even non-recursively enumerable.

There are still many simple equations and inequalities where we know almost nothing about their solutions. In particular, it would be interesting to study inequalities of a very simple form similar to commutation equations, for instance those where multiplication by constant languages from both sides is involved or the intersection operation is used. Additionally, for classes of inequalities where regularity of maximal solutions is proved by means of well quasi-orders, one should consider the question whether these maximal solutions can be computed algorithmically; this question is still open for all such classes mentioned in this survey.

Acknowledgements. I am very grateful to Ondřej Klíma for carefully reading a preliminary version of this paper and for providing me with numerous valuable comments.

References

[1] F. Baader and R. Küsters. Unification in a description logic with transitive closure of roles. In *Proc. 8th Internat. Conf. on Logic for Programming, Artificial Intelligence and Reasoning (LPAR 2001)*, volume 2250 of *Lecture Notes in Comput. Sci.*, pages 217–232. Springer, 2001.

[2] G.M. Bergman. Centralizers in free associative algebras. *Trans. Amer. Math. Soc.*, 137:327–344, 1969.

[3] W. Bucher, A. Ehrenfeucht and D. Haussler. On total regulators generated by derivation relations. *Theoret. Comput. Sci.*, 40:131–148, 1985.

[4] C. Choffrut, J. Karhumäki and N. Ollinger. The commutation of finite sets: A challenging problem. *Theoret. Comput. Sci.*, 273:69–79, 2002.

[5] P.M. Cohn. Factorization in non-commutative power series rings. *Proc. Cambridge Philos. Soc.*, 58:452–464, 1962.

[6] J.H. Conway. *Regular Algebra and Finite Machines.* Chapman and Hall, 1971.

[7] A. de Luca and S. Varricchio. *Finiteness and Regularity in Semigroups and Formal Languages.* Springer, 1999.

[8] A. Ehrenfeucht, D. Haussler and G. Rozenberg. On regularity of context-free languages. *Theoret. Comput. Sci.*, 27:311–332, 1983.

[9] G. Higman. Ordering by divisibility in abstract algebras. *Proc. London Math. Soc. (3)*, 2:326–336, 1952.

[10] J.M. Howie. *Fundamentals of Semigroup Theory.* Clarendon Press, 1995.

[11] J. Karhumäki, M. Latteux and I. Petre. Commutation with codes. *Theoret. Comput. Sci.*, to appear.

[12] J. Karhumäki, M. Latteux and I. Petre. The commutation with ternary sets of words. *Theory Comput. Syst.*, to appear.

[13] J. Karhumäki and I. Petre. Conway's problem for three-word sets. *Theoret. Comput. Sci.*, 289:705–725, 2002.

[14] J. Karhumäki and I. Petre. Two problems on commutation of languages. In *Current Trends in Theoretical Computer Science, The Challenge of the New Century, vol. 2*, pages 477–494. World Scientific, 2004.

[15] J.B. Kruskal. Well-quasi-ordering, the tree theorem, and Vazsonyi's conjecture. *Trans. Amer. Math. Soc.*, 95:210–225, 1960.

[16] M. Kunc. Regular solutions of language inequalities and well quasi-orders. *Theoret. Comput. Sci.*, to appear. Extended abstract in *Proc. 31st Internat. Colloq. on Automata, Languages and Programming (ICALP 2004)*, volume 3142 of *Lecture Notes in Comput. Sci.*, pages 870–881. Springer, 2004.

[17] M. Kunc. The power of commuting with finite sets of words. Manuscript, 2004. Available at `http://www.math.muni.cz/~kunc/`. Extended abstract in *Proc. 22nd Sympos. on Theoretical Aspects of Computer Science (STACS'05)*, volume 3404 of *Lecture Notes in Comput. Sci.*. Springer, 2005.

[18] M. Kunc. Largest solutions of left-linear language inequalities. Manuscript, 2004. Available at `http://www.math.muni.cz/~kunc/`

[19] M. Kunc. On language inequalities $XK \subseteq LX$. Manuscript, 2004. Available at `http://www.math.muni.cz/~kunc/`

[20] E.L. Leiss. *Language Equations*. Springer, 1999.

[21] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[22] A. Okhotin. A characterization of the arithmetical hierarchy by language equations. In *Preproc. Descriptional Complexity of Formal Systems (DCFS 2004)*, 225–237.

[23] A. Okhotin. Decision problems for language equations. Submitted for publication. Available at `http://www.cs.queensu.ca/home/okhotin/`. Preliminary version in *Proc. 30th Internat. Colloq. on Automata, Languages and Programming (ICALP 2003)*, volume 2719 of *Lecture Notes in Comput. Sci.*, pages 239–251. Springer, 2003.

[24] A. Okhotin. On computational universality in language equations. In *Proc. Machines, Computations and Universality (MCU 2004)*, volume 3354 of *Lecture Notes in Comput. Sci.*. Springer, to appear.

[25] A. Okhotin. Greatest solutions of language equations. Submitted for publication. Available at `http://www.cs.queensu.ca/home/okhotin/`

[26] B. Ratoandromanana. Codes et motifs. *RAIRO Inform. Théor. Appl.*, 23:425–444, 1989.

[27] G. Rozenberg and A. Salomaa (Eds.). *Handbook of Formal Languages, Vol. 1*. Springer, 1997.

[28] J. Sakarovitch. *Eléments de théorie des automates*. Vuibert, 2003. English translation *Elements of Automata Theory*. Cambridge University Press, to appear.

Department of Mathematics, Masaryk University,
Janáčkovo nám. 2a, 662 95 Brno, Czech Republic
E-mail address: `kunc@math.muni.cz`
URL: `http://www.math.muni.cz/~kunc/`