Two-variable word equations

Lucian ILIE
Turku Centre for Computer Science TUCS
FIN-20520, Turku, Finland
Wojciech PLANDOWSKI
Turku Centre for Computer Science TUCS
FIN-20520, Turku, Finland

Turku Centre for Computer Science
TUCS Technical Report No 306

ISBN 952-12-0534-2
ISSN 1239-1891

Abstract

We consider languages expressed by word equations in two variables and give a complete characterization for their complexity functions, that is, the functions that give the number of words of the same length. Specifically, we prove that there are only five types of complexities: constant, linear, exponential, and two in between constant and linear. For the latter two, we give precise characterizations in terms of the number of solutions of Diophantine equations of certain types. In particular, we show that the linear upper bound on the non-exponential complexities by Karhumäki et al., cf. [KMP], is tight. There are several consequences of our study. First, we derive that both of the sets of all finite Sturmian words and of all finite Standard words are expressible by word equations. Second, we characterize the languages of non-exponential complexity which are expressible by two-variable word equations as finite unions of several simple parametric formulae and solutions of a two-variable word equation with a finite graph. Third, we find optimal upper bounds on the solutions of (solvable) two-variable word equations, namely, linear bound for one variable and quadratic for the other. From this, we obtain an $\mathcal{O}(n^6)$ algorithm for testing the solvability of two-variable word equations, improving thus very much Charatonik and Pacholski's $\mathcal{O}(n^{100})$ algorithm, cf. [ChPa].

Keywords: word equation, expressible language, complexity function, minimal solution, solvability

TUCS Research Group
Mathematical Structures of Computer Science

1 Introduction

Word equations constitute one of the basic parts of combinatorics on words. The fundamental result in word equations is Makanin's algorithm, cf. [Ma], which decides whether or not a word equation has a solution. The algorithm is one of the most complicated ones existing in the literature. However, the structure of solutions of word equations is not well understood; see [Hm, Raz1, Raz2]. A new light on that topic has been led recently by [KMP] where the languages which are defined by solutions of word equations are studied. These languages possess some pumping-like properties.

The structure of languages which are defined by equations with one variable is very simple. Infinite languages which are defined by one-variable word equations consist of a finite part and an infinite part which is of the form $A^n A'$ for A' a prefix of A. The structure of the finite part is not completely known [OGM]. Our analysis deals with languages which are defined by two-variable word equations. We prove that the complexity of those languages, which is measured by the number of words of a given length, belongs to one of five classes: constant, $\mathcal{D}_1$-type, $\mathcal{D}_2$-type, linear and exponential. The complexities $\mathcal{D}_1$-type and $\mathcal{D}_2$-type are in between linear and constant and they are related to the number of solutions of certain Diophantine equations. As a side effect of our considerations we prove that the linear upper bound given in [KMP] for languages which do not contain a pattern language is tight. Another interesting related result is that the sets of Sturmian and Standard words are expressible by simple word equations. As another consequence of our study, we characterize the languages of non-exponential complexity which are expressible by two-variable word equations as finite unions of several simple parametric formulae and solutions of a two-variable word equation with a finite graph.

Based on our analysis, we find optimal upper bounds on the solutions of (solvable) two-variable word equations, namely, linear bound for one variable and quadratic for the other. From this, we obtain an $\mathcal{O}(n^6)$ algorithm for testing the solvability of two-variable word equations. Recall that there is only one polynomial-time algorithm known for this problem and it works in time $\mathcal{O}(n^{100})$, cf. [ChPa].

2 Expressible languages

In this section we give basic definitions we need later on, as well as some previous results. For an alphabet Σ, we denote by $\mathrm{card}(\Sigma)$ the number of elements of Σ; Σ^* is the set of words over Σ with 1 the empty word. For $w \in \Sigma^*$, $|w|$ is the length of w; for $a \in \Sigma$, $|w|_a$ is the number of occurrences of a in w; for $0 \le k \le |w|$, $\mathrm{pref}_k(w)$ denotes the prefix of length k of w. By $\rho(w)$ we denote the primitive root of w. If $w = uv$, then we denote $u^{-1}w = v$ and $wv^{-1} = u$. For any notions and results of combinatorics on words, we refer to [Lo] and [ChKa].

Consider two disjoint alphabets, of constants, Σ, and of variables, Ξ. A *word* *equation* e is a pair of words $\varphi, \psi \in (\Sigma \cup \Xi)^*$, denoted $e : \varphi = \psi$. The *size* of e, denoted $|e|$, is the sum of the lengths of φ and ψ. The equation e is said to be *reduced* if φ and ψ start with different letters and end with different letters, as words over $\Sigma \cup \Xi$. Throughout the paper, all equations we consider are assumed to be reduced.

A *solution* of e is a morphism $h : (\Sigma \cup \Xi)^* \longrightarrow \Sigma^*$ such that $h(a) = a$, for any $a \in \Sigma$, and $h(\varphi) = h(\psi)$. The set of solutions of e is denoted by $\mathrm{Sol}(e)$.

Notice that a solution can be given also as an ordered tuple of words, each component of the tuple corresponding to a variable of the equation. Therefore, we may take, for a variable $X \in \Xi$, the X-component of all solutions of e, that is,

$$L_X(e) = \{x \in \Sigma^* \mid \text{ there is a solution } h \text{ of } e \text{ such that } h(X) = x\}.$$

The set $L_X(e)$ is called *the language expressed by X in e*. A language $L \subseteq \Sigma^*$ is *expressible* if there is a word equation e and a variable X such that

$$L = L_X(e).$$

Notice that, if X does not appear in e, then $L_X(e) = \Sigma^*$ as soon as $\mathrm{Sol}(e) \neq \emptyset$. Also, if $\mathrm{card}(\Sigma) = 1$, that is, there is only one constant letter, then all expressible languages are trivially regular, as we work here with numbers. Therefore, we shall assume that always $\mathrm{card}(\Sigma) \geq 2$.

The *complexity function* of a language $L \subseteq \Sigma^*$, is the natural function

$$\#_L : \mathbb{N} \longrightarrow \mathbb{N}$$

defined by

$$\#_L(n) = \mathrm{card}\{w \in L \mid |w| = n\}.$$

Example 1. Consider the equation $e : XX = Y$. The complexity of its solutions with respect to Y is

$$\#_{L_Y(e)}(n) = \begin{cases} \mathrm{card}(\Sigma)^{\frac{n}{2}}, & \text{if } n \text{ is even,} \\ 0, & \text{if } n \text{ is odd.} \end{cases}$$

Since the function $\#_L$ can be very unregular, as can be seen from the above example, we use in our considerations a function $\bar{\#}_L$, which is defined by

$$\bar{\#}_L(n) = \max_{1 \leq i \leq n} \#_L(i).$$

We say that a function f is *constant* if $f(n) = \Theta(1)$, is *linear* if $f(n) = \Theta(n)$, and is *exponential* if $f(n) = 2^{\Theta(n)}$.

We make the following conventions concerning notations:

– $a, b, \ldots \in \Sigma$ are constant letters,

– $A, B, \ldots \in \Sigma^*$ are (fixed) constant words,

– $X, Y, \ldots \in \Xi$ are variables,

– $x, y, \ldots \in \Sigma^*$ may denote some arbitrary constant words but may also stand for images of variables by some morphisms from $(\Sigma \cup \Xi)^*$ to Σ^*, that is, $x = h(X), y = h(Y)$, etc.,

– $\varphi, \psi, \ldots \in (\Sigma \cup \Xi)^*$ are mixed words, which *may* (but need not) contain both constants and variables.

We shall use also the following notation (due to Hmelevskii, cf. [Hm]); for $\alpha_i \in (\Sigma \cup \Xi)^*, 1 \leq i \leq n$, we denote

$$[\alpha_i]_{i=1}^n = \alpha_1 \alpha_2 \cdots \alpha_n.$$

We give next two simple examples of languages expressible by word equations. For counterexamples, i.e., examples of languages that are not expressible, see [KMP]. (It is actually quite difficult to give such examples.)

Example 2. For a fixed word $A \in \Sigma^*$, the language

$$L_1 = \{A^n \mid n \geq 0\}$$

is expressed by the variable Y in the two-variable word equation

$$e_1 : XAY = AXX^t$$

where t is such that $A = \rho(A)^t$. Then $\bar{\#}_{L_Y(e)} = \bar{\#}_{L_1}$ is constant.

Example 3. The language of all words containing a sequence of letters a, b, a in this order, that is

$$L_2 = \{xaybuav \mid x, y, u, v \in \Sigma^*\}$$

is expressed by the variable W in the equation

$$e_2 : W = XaYbUaV.$$

We recall the following two results from [KMP]. The first is a lemma which we shall use later.

Lemma 4. (Karhumäki et al. [KMP]) *Assume that e is a word equation over $\Xi = \{X, Y\}$ such that none of the complexity functions $\bar{\#}_{L_X(e)}$ and $\bar{\#}_{L_Y(e)}$ is exponential. Then, for any $n, m \in \mathbb{N}$, there is at most one solution (x, y) of e with $|x| = n$ and $|y| = m$.*

The second is the upper bound on the non-exponential complexities. We shall prove it to be tight.

Theorem 5. (Karhumäki et al. [KMP]) *Any non-exponential complexity function of languages expressible by two-variable word equations is at most linear.*

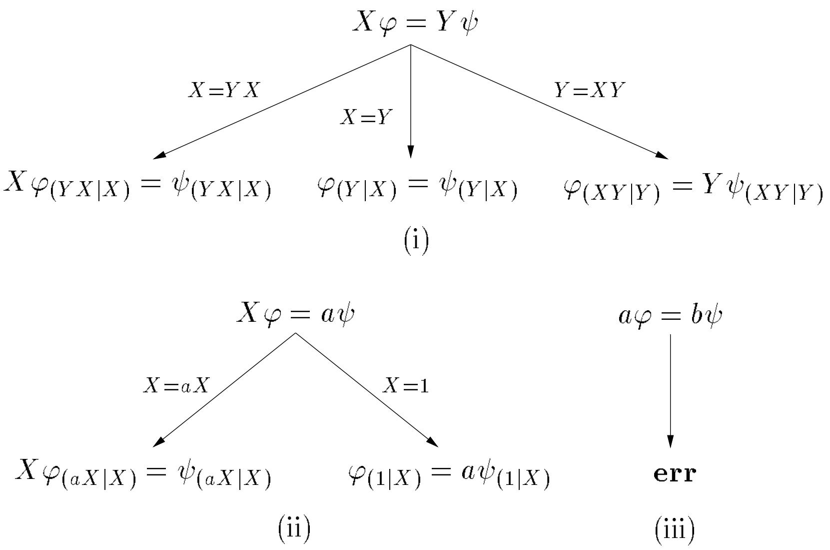

Figure 1: The graph associated with a word equation

We shall need also the *graph* associated with an equation $e : \varphi = \psi$, see [Lo]. It is constructed by applying exhaustively the so-called Levi's lemma which states that if $uv = wt$, for some words u, v, w, t, then either u is a proper prefix of w or $u = w$ or w is a proper prefix of u. The vertices of the graph are different (reduced!) equations (including e) and the directed edges are put as follows. We start with e and draw the graph by considering iteratively the following three cases, depicted in Fig. 1: (i) both sides of e start with variables (which are different since the equation is assumed to be reduced), (ii) one side starts with a constant and the other starts with a variable, and (iii) the two sides start with constants which are different. Clearly, in the last case the equation has no solution, which is marked by an error node. In Fig. 1, we denote by $\varphi_{(\alpha|\beta)}$ the word obtained from φ by replacing all occurrences of β by α.

Thus, we start by processing e and then process all unprocessed vertices. When we find an equation already obtained, we do not create a new vertex but direct the corresponding edge to the old one.

Examples of such graphs are given in the next sections.

We notice that the graph associated with a word equation may be infinite but, if it is finite, then all solutions of the equation are obtained starting from a vertex with no outgoing edges and different from **err** and going in the oposite direction of the edges to the root; at the same time, the corresponding operations on the values of the variables are performed.

We recall that the *Euler's totient function* $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ is defined by

$$\phi(n) = \text{card}\{k | k \text{ is coprime with } n\} = n\Big(1 - \frac{1}{p_1}\Big)\Big(1 - \frac{1}{p_2}\Big)\dots\Big(1 - \frac{1}{p_k}\Big)$$

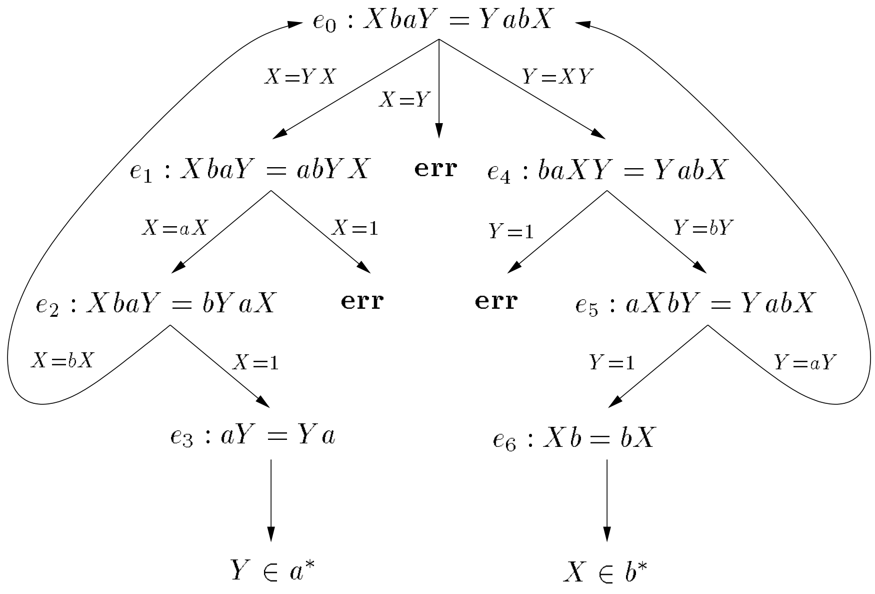

Figure 2: The graph of $\boldsymbol{e_0}$

where $p_1, p_2, \ldots, p_k$ are all the distinct prime factors of n. Clearly, the function $\bar{\phi}(n) = \max_{1 \leq i \leq n} \phi(i)$ is linear.

3 The equation $XbaY = YabX$

The starting point of our analysis is the equation

$$\boldsymbol{e_0} : XbaY = YabX$$

which we study in this section. We show first that there is a very close connection between solutions of $\boldsymbol{e_0}$ and the family of Standard words which we define below. Using then some strong properties of the Standard words, we prove that both functions $\bar{\#}_{L_X(e_0)}$ and $\bar{\#}_{L_Y(e_0)}$ are linear.

Let us consider the set of solutions of our equation $\boldsymbol{e_0}$. For this we draw its associated graph in Fig. 2.

Consider the following two mappings

$$\alpha_1, \alpha_2 : \{a,b\}^* \times \{a,b\}^* \longrightarrow \{a,b\}^* \times \{a,b\}^*$$

defined by

$$\begin{aligned}\alpha_1(u,v) &= (u, ubav),\\ \alpha_2(u,v) &= (vabu, v),\end{aligned}$$

for any $u, v \in \{a, b\}^*$. Using these two mappings, we give the following result which characterizes the set of solutions of $\boldsymbol{e}_0$.

Lemma 6. *The solutions of* $\boldsymbol{e}_0$ *are precisely the pairs of words obtained by:*
(i) starting with a pair (u, v) *of words in the set*

$$\{(a^{n+1}, a^n) \mid n \geq 0\} \cup \{(b^n, b^{n+1}) \mid n \geq 0\},$$

(ii) applying to (u, v) *a finite (possible empty) sequence* $\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_k}$*, for some* $k \geq 0, 1 \leq i_j \leq 2$*, for any* $1 \leq j \leq k$.

Proof. The proof is based on the graph of $\boldsymbol{e}_0$ which is presented in Fig. 2. The solutions (a^{n+1}, a^n) (respectively (b^n, b^{n+1})), for $n \geq 0$, are obtained by composing substitutions which are labels of the path e_0,e_1,e_2,e_3 (respectively e_0,e_4,e_5,e_6). The other solutions are obtained by composing these solutions with substitutions which corresponds to loops e_0, e_4, e_5, e_0, and e_0, e_1, e_2, e_0. These are, respectively, $(X, Y) \to (X, XbaY)$ which is α_1 and $(X, Y) \to (YabX, Y)$ which is α_2. ∎

We define next the Standard words. The set $\mathcal{R}$ of *Standard pairs* (as defined by Rauzy, cf. [Ra]) is the minimal set included in $\{a, b\}^* \times \{a, b\}^*$ such that:
(i) $(a, b) \in \mathcal{R}$ and
(ii) $\mathcal{R}$ is closed under the two mappings

$$\beta_1, \beta_2 : \{a, b\}^* \times \{a, b\}^* \longrightarrow \{a, b\}^* \times \{a, b\}^*$$

defined by

$$\beta_1(u, v) = (u, uv),$$
$$\beta_2(u, v) = (vu, v),$$

for any $u, v \in \{a, b\}^*$. The set $\mathcal{S}$ of *Standard words* is defined by

$$\mathcal{S} = \{u \in \{a, b\}^* \mid \text{ there is } v \in \{a, b\}^* \text{ such that either } (u, v) \in \mathcal{R} \text{ or } (v, u) \in \mathcal{R}\}.$$

We shall use the following strong properties of Standard words, proved by de Luca and Mignosi, cf. [dLMi].

Lemma 7. (de Luca, Mignosi [dLMi]) *The set of Standard words verifies the formula*

$$\mathcal{S} = \{a, b\} \cup \Pi\{ab, ba\}$$

where

$$\Pi = \{w \in \{a, b\}^* \mid w \text{ has two periods } p, q \text{ which are coprime and } |w| = p + q - 2\}.$$

Lemma 8. (de Luca, Mignosi [dLMi]) $\#_\Pi(n) = \phi(n+2)$, *for any* $n \geq 0$, *where* ϕ *is Euler's totien function.*

We now establish a connection between the solutions of $\boldsymbol{e}_0$ and the set of Standard pairs $\mathcal{R}$.

Lemma 9. $\mathrm{Sol}(\boldsymbol{e}_0) = \{(u,v) \mid (uba, vab) \in \mathcal{R}\}$.

Proof. We notice first that, for any $u, v, w, z \in \{a,b\}^*$ and any $i = 1, 2$, we have the equivalence

$$\alpha_i(u,v) = (w,z) \quad \text{iff} \quad \beta_i(uba, vab) = (wba, zab). \tag{1}$$

Indeed, this equivalence is proved by straightforward calculation. Next, we change slightly the definition of the Standard pairs into an equivalent one which is more suitable for our purpose. We claim that all Standard pairs (u,v) such that $|u| \geq 2$ and $|v| \geq 2$ (only these are of interest for us here) are obtained by
(i) starting with a pair (u,v) of words in the set

$$\{(a^{n+1}ba, a^n ab) \mid n \geq 0\} \cup \{(b^n ba, b^{n+1}ab) \mid n \geq 0\},$$

(ii) applying β_i's in any order finitely many times.

To see this, it is enough to emphasize all steps in the forming of the Standard pairs when all components become of length at least two; this is seen below:

$$\begin{array}{lclcl} (a,b) & \xrightarrow{\beta_1^{n+1}} & (a, a^{n+1}b) & \xrightarrow{\beta_2} & (a^{n+1}ba, a^n ab), \\ (a,b) & \xrightarrow{\beta_2^{n+1}} & (b^{n+1}a, b) & \xrightarrow{\beta_1} & (b^n ba, b^{n+1}ab). \end{array}$$

Now, using (1), the equality in the statement is clear by Lemma 6. ∎

Theorem 10. $\#_{L_X(e_0)}(n) = \#_{L_Y(e_0)}(n) = \phi(n)$, *for any* $n \geq 1$.

Proof. We have, by Lemma 9, that

$$\mathcal{S} = L_X(\boldsymbol{e}_0)ba \cup L_Y(\boldsymbol{e}_0)ab \cup \{a,b\}.$$

Thus, by Lemma 7, we get

$$L_X(\boldsymbol{e}_0) = L_Y(\boldsymbol{e}_0) = \Pi \tag{2}$$

and the claim follows from Lemma 8. ∎

Remark. We notice the unexpected equality (2).

As a corollary of Theorem 10 we obtain that the upper bound of Karhumäki et al. in Theorem 5 is optimal.

Corollary 11. *The linear upper bound for the non-exponential complexities of languages expressible by two-variable word equations is tight.*

Furthermore, using the above considerations, we prove that both sets, of Standard and of Sturmian (finite) words are expressible by word equations.

Example 12. Standard words. The set of Standard words $\mathcal{S}$ is expressed by the variable Z in the following system:

$$\begin{cases} XbaY = YabX \\ Z = Xba \text{ or } Z = Yab \text{ or } Z = a \text{ or } Z = b \end{cases}$$

First, by Lemma 9, it is clear that the Z-components of all solutions of the system above are precisely the Standard words. Second, from the above system we can derive a single equation, as well known; see, e.g., [KMP]. Hence, the set of Standard words is expressible.

Example 13. Sturmian words. There are many definitons of the finite Sturmian words (see, e.g., [Be] and references therein). We use here only the fact that the set $\mathcal{S}t$ of finite Sturmian words is the set of factors of Π, cf. [dLMi]. Therefore, the set $\mathcal{S}t$ is expressed by the variable Z in the system

$$\begin{cases} XbaY = YabX \\ X = WZT \end{cases}$$

Again, this can be expressed using a single equation.

4 Linear cases

Our next step in the analysis of the two-variable word equations is to consider equations of a form which is a generalization of $\boldsymbol{e_0}$, namely

$$\boldsymbol{e_1} : XAY = YBX,$$

where $A, B \in \Sigma^*$. The study of this type of equations is not only interesting in itself, but also it will be of essential help later, for other more complicated equations.

Notice first that, if $A = B$, then both functions $\#_{L_X(e_1)}$ and $\#_{L_Y(e_1)}$ are exponential since, in this case, we have

$$\{(w, w) \mid w \in \Sigma^*\} \subseteq \mathrm{Sol}(\boldsymbol{e_1}).$$

Consider the other case, that is, $A \neq B$. As already noticed by Hmelevskii, cf. [Hm], $\boldsymbol{e_1}$ has a non-empty set of solutions if and only if there are $P, Q, R \in \Sigma^*$ such that $A = PQR, B = RQP$. We assume in the sequel that this is the case for some fixed constants A and B and draw the partial graph of $\boldsymbol{e_1}$ in Fig. 3. In the graph, the substitutions $X = wX$, for a fixed word w and a variable X, correspond to a sequence

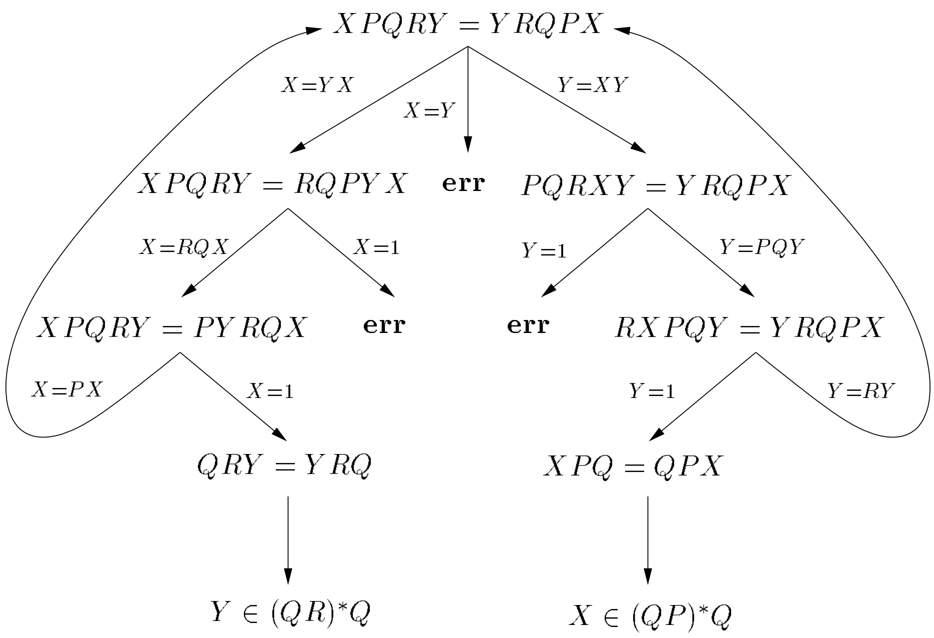

Figure 3: A part of the graph of $\boldsymbol{e_1}$

of substitutions $X = a_1X$, $X = a_2X$, ..., $X = a_nX$, where $w = a_1a_2 \ldots a_n$, and therefore the edge labeled by $X = wX$ corresponds to a path of edges labeled by $X = a_1X$, $X = a_2X$, ..., $X = a_nX$.

As it will be seen, considering only the part in Fig. 3 of the graph of $\boldsymbol{e_1}$ will be enough to conclude the linearity of the functions $\#_{L_X(e_1)}$ and $\#_{L_Y(e_1)}$. (In fact, it is clearly enough to prove this only for one of the two functions; the result for the other will follow by symmetry.) Notice here that the tuple (P, Q, R) need not to be uniquely determined by constant words A and B as it is in the example $A = aabab$, $B = babaa$ where the possible tuples are $(aa, 1, bab)$, (a, a, bab), (a, aba, b), $(1, aa, bab)$ and $(aa, bab, 1)$.

We first establish a connection betwen the solutions of $\boldsymbol{e_1}$ and those of $\boldsymbol{e_0}$. Consider a new letter $\# \notin \{a, b\}$, the mapping

$$^{\#} : \{a, b\}^* \longrightarrow \{a, b, \#\}^*$$

defined by

$$\begin{aligned} &1^{\#} = 1, \\ &(a_1a_2 \cdots a_n)^{\#} = \#a_1\#a_2\# \cdots \#a_n\#, \end{aligned}$$

for any $n \geq 1, a_i \in \{a, b\}, 1 \leq i \leq n$, and the morphism

$$h : \{a, b, \#\}^* \longrightarrow \Sigma^*,$$

given by

$$h(a) = R,$$
$$h(b) = P,$$
$$h(\#) = Q.$$

Lemma 14. *The composition* $^{\#} \circ h$ *is an injective mapping.*

Proof. We argue by contradiction. Assume there are $u_1, u_2, \in \{a, b\}^*$, $u_1 \neq u_2$, such that $h(u_1^{\#}) = h(u_2^{\#})$. Let $u \in \#(\{a, b\}\#)^*$ be the longest common prefix of $u_1^{\#}$ and $u_2^{\#}$. As $^{\#}$ is injective, there are $a_1, a_2 \in \{a, b\}$, $a_1 \neq a_2$, such that

$$u_1^{\#} = ua_1\#u_1',$$
$$u_2^{\#} = ua_2\#u_2',$$

for some $u_1', u_2' \in (\{a, b\}\#)^*$. Without loss of generality, we may assume that $a_1 = a, a_2 = b$. Thus

$$RQh(u_1') = PQh(u_2'),$$

hence either P is a prefix of R or R is a prefix of P. If $|P| = |R|$, then $P = R$ and consequently $A = PQP = B$, a contradiction. So $|P| \neq |R|$ and, again without loss of generality, we may assume that $|P| > |R|$. Denote $PQ = P'$ and $RQ = R'$. Then

$$R'h(u_1')Q = P'h(u_2')Q \tag{3}$$

and $h(u_1')Q$, $h(u_2')Q \in (P' \cup R')^*$. The identity (3) is a nontrivial identity on two words R' and P'. Therefore R', P' commutes, see [Lo]. Hence, $P'R' = PQRQ = R'P' = RQPQ$ and finally $A = PQR = RQP = B$, a contradiction. This completes the proof. ■

We now can give the relation between the solutions of $\boldsymbol{e_0}$ and $\boldsymbol{e_1}$.

Lemma 15. *For any solution* (u, v) *of* $\boldsymbol{e_0}$*, the pair of words* $(h(u^{\#}), h(v^{\#}))$ *is a solution of* $\boldsymbol{e_1}$*.*

Proof. From Fig. 3, it is clear that all solutions of $\boldsymbol{e_1}$ obtained from the partial graph there are also obtained by the steps (i) and (ii) below:
(i) start with a pair of words in the set

$$\{((QR)^{n+1}Q, (QR)^nQ) \mid n \geq 0\} \cup \{((QP)^nQ, (QP)^{n+1}Q) \mid n \geq 0\}$$

(ii) apply a sequence $\gamma_{i_1}\gamma_{i_2}\cdots\gamma_{i_k}$ to this pair of words, for some $k \geq 0, 1 \leq i_j \leq 2$, for any $1 \leq j \leq k$, where

$$\gamma_1(u, v) = (u, uPQRv),$$
$$\gamma_2(u, v) = (vRQPu, v),$$

for any $u, v \in \Sigma^*$. It is then enough to compare this with the procedure of obtaining the solutions of $\boldsymbol{e_0}$ in Lemma 6. ■

Theorem 16. *If $A = PQR, B = RQP$, and $A \neq B$, then the function $\bar{\#}_{L_X(e_1)}$ is linear.*

Proof. According to Lemma 15, for any $u \in L_X(\boldsymbol{e_0})$, we have $h(u^{\#}) \in L_X(\boldsymbol{e_1})$. Also

$$|h(u^{\#})| = (|u|+1)|Q| + |u|_a|R| + |u|_b|P| \leq 3|u| \max(|P|,|Q|,|R|). \tag{4}$$

Consider a fixed $k \geq 1$. Then, by (4), all words in $L_X(\boldsymbol{e_0})$ which are not longer than

$$\frac{k}{3\max(|P|,|Q|,|R|)}$$

have their images through $^{\#} \circ h$ in $L_X(\boldsymbol{e_1})$ and not longer than k. Moreover, by Lemma 14, all these images are distinct. Denote $c = 3\max(|P|,|Q|,|R|)$. Then, by Theorem 10,

$$\sum_{i=0}^{k} \#_{L_X(e_1)}(i) \geq \sum_{i=0}^{[\frac{k}{c}]} \phi(i+2). \tag{5}$$

We use next the following property of the totient function, cf., e.g., [Si]:

$$\lim_{x\to\infty} \frac{\sum_{i=1}^{[x]} \phi(i)}{\frac{3x^2}{\pi^2}} = 1. \tag{6}$$

Thus, by (5) and (6), we get that, for any $k \geq 1$,

$$\bar{\#}_{L_X(e_1)}(k) \geq \frac{1}{k}\sum_{i=0}^{k} \#_{L_X(e_1)}(i) \geq \frac{1}{k}\sum_{i=0}^{[\frac{k}{c}]} \phi(i+2) \geq (\frac{3}{\pi^2 c^2} - \varepsilon)k$$

with $\varepsilon \to 0$ when $k \to \infty$. Since, by Theorem 5, $\bar{\#}_{L_X(e_1)}$ is at most linear, the theorem is proved. ∎

Another type of equation which is very similar to $\boldsymbol{e_1}$ is

$$\boldsymbol{e_2} : AXY = YXB.$$

If the set $L_Y(\boldsymbol{e_2})$ is finite, then certainly $\bar{\#}_{L_X(e_2)}$ is constant. Otherwise, we may restrict the analysis to those words in $L_Y(\boldsymbol{e_2})$ which are longer than $|A| + |B|$ and make the substitution $Y = AYB$. We obtain the equation $XAY = YBX$, which is of type $\boldsymbol{e_1}$. Therefore, we have

Theorem 17. *The functions $\bar{\#}_{L_X(e_2)}$ and $\bar{\#}_{L_Y(e_2)}$ are either constant or linear or else exponential.*

5 Periodic solutions

We study in this section equations in two variables for which all values for one component in the solutions are periodic with a short period; by "short" we mean bounded from above by a fixed constant which depends only on the length of the equation. We show that in this case there are three possible types of complexity for the language expressed by the other component: constant, exponential, and $\mathcal{D}_1$-type. The last type lies in between constant and linear and is defined in terms of the number of solutions of certain Diophantine equations.

Before giving the definition of the $\mathcal{D}_1$-type, we give an example showing how it arises naturally.

Example 18. Consider the equation

$$e : aXXbY = XaYbX.$$

Clearly, the set of solutions of e is

$$\mathrm{Sol}(e) = \{(a^n, (a^n b)^m a^n) \mid n \geq 0\}.$$

Here $\bar{\#}_{L_X(e)}$ is constant but $\#_{L_Y(e)}$ is not. For any $p \geq 0$, $\#_{L_Y(e)}(p)$ is the number of solutions of the Diophantine equation in unknowns n and m

$$(n+1)m + n = p.$$

We now define the $\mathcal{D}_1$-type precisely. A function $f : \mathbb{N} \longrightarrow \mathbb{N}$ is of *divisor-type* if there are some non-negative integers $c_i, 1 \leq i \leq 4$, such that $c_1 \geq c_3, c_2 \geq c_4$ and, $f(k)$ is the number of solutions of the Diophantine equation in unknowns n and m

$$(c_1 n + c_2)m + c_3 n + c_4 = k.$$

A function f is $\mathcal{D}_1$-*type* if there is a divisor-type function g such that $f = \Theta(\bar{g})$ where $\bar{g}(n) = \max_{1 \leq i \leq n} g(i)$.

Before stating the announced result, we need the notion of a *P-factorization* (as defined in [KoPa] and [KMP]). Consider a primitive word $P \in \Sigma^*$. It is well-known that any word $w \in \Sigma^*$ can be uniquely written in the form

$$w = w_1 P^{k_1} w_2 P^{k_2} \cdots w_n P^{k_n} w_{n+1} \tag{7}$$

where $n \geq 0, k_i \geq 0$, for any $1 \leq i \leq n$ and

- w_i does not contain P^2 as a factor, for any $1 \leq i \leq n$,
- P is a proper prefix and a proper suffix of any w_i with $1 \leq i \leq n$,
- P is proper suffix of w_1 or $w_1 = 1$,
- P is proper prefix of w_{n+1} or $w_{n+1} = 1$.

The equation (7) is called the *P-presentation* of w. The *P-factorization* of w is the ordered sequence

$$w_1, P^{k_1}, w_2, P^{k_2}, \ldots, w_n, P^{k_n}, w_{n+1}.$$

We now fix the hypotheses for the result on periodic solutions. Let C, D be two words and e an equation on two variables X and Y. In many parts of our analysis we consider solutions (x, y) such that x is in form $(CD)^n C$. In order to deal with those solutions we define $X_{C,D}(e)$ to be the set of those words of the form $(CD)^n C$ for which $((CD)^n C, y)$ is a solution of e for some y. Similarly, define $Y_{C,D}(e)$ to be the set of words y such that $((CD)^n C, y)$ is a solution of e for some nonnegative n.

Lemma 19. *Assume that $e : \varphi = \psi$ is an equation over $\Xi = \{X, Y\}$ and C, D are two words. Then*
(i) $\bar{\#}_{X_{C,D}(e)}$ is constant,
(ii) $\bar{\#}_{Y_{C,D}(e)}$ is either constant, or exponential, or else of $\mathcal{D}_1$-type.

Proof. Part (i) is clear. Consider part (ii) and assume that $\#_{L_Y(e)}$ is not exponential. Then $\bar{\#}_{Y_{C,D}(e)}$ is not exponential, too. (See Example 21 for an example that it can be also exponential.) Denote $A = CD$ and $A' = C$. Consider now the first variables appearing in φ and ψ. If Y appears first in both, then $\#_{Y_{C,D}(e)}$ is constant. Indeed, we may assume that $\varphi = Y\varphi'$, $\psi = BY\psi'$, for some $B \in \Sigma^+$, $\varphi', \psi' \in (\Sigma \cup \Xi)^*$, and then we have that any $y \in L_Y(e)$ is a prefix of a word in B^*. If X appears first in both φ and ψ, then, for $x = A^n A' \in L_X(e)$ with n large enough (say $n \geq |e|$), the occurrences of x at the beginning of φ and ψ, together with the constants in between can be reduced (if the equation is not contradictory). Therefore, we may assume that the first variable which appears in φ is X whereas for ψ is Y. Hence, the equation is of one of the forms below, where E and F are constant words:

$$\begin{aligned} X \ldots &= EY \ldots, \\ Y \ldots &= FX \ldots, \\ X \ldots &= Y \ldots. \end{aligned}$$

We may assume that Y and X are "long enough" because the solutions with "short" X or Y give, due to Lemma 4, a constant contribution to the complexity functions $\bar{\#}_{L_X(e)}$, $\bar{\#}_{L_Y(e)}$. Then in the first two cases we make the substitution $X = EX'$ and $Y = FY'$, respectively. After making this substitution, the first two cases turn into the third one. We have

$$\begin{aligned} \varphi &= [XB_i]_{i=1}^{l} \varphi', \\ \psi &= Y[C_i X]_{i=1}^{r} C_{r+1} \psi', \end{aligned} \tag{8}$$

where $l \geq 1, r \geq 0, B_i, C_i \in \Sigma^*, \varphi', \psi' \in (\Sigma \cup \Xi)^*$, and ψ' either starts with Y or is empty and φ' either starts with Y or is empty. If φ' is empty then there is a one to one correspondence between the lengths of X and Y (or the length of Y is fixed) and therefore by Lemma 4 both $\#_Y(e)$ and $\#_X(e)$ are constant. Hence, we assume $\varphi = [XB_i]_{i=1}^l Y\varphi'$. Take any $n \geq |e|$ such that $x = A^nA' \in L_X(e)$. There is $y \in \Sigma^*$ such that

$$[A^nA'B_i]_{i=1}^l y\varphi'(x,y) = y[C_iA^nA']_{i=1}^r C_{r+1}\psi'(x,y),$$

where $\varphi'(x,y)$ (resp. $\psi'(x,y)$) is φ' (resp. ψ') where X is replaced by x and Y by y. Depending on the relation between l and r we consider three cases.

Case 1. $l < r$. It follows that

$$[A^nA'B_i]_{i=1}^l y = y\,\mathrm{pref}_{|[A^nA'B_i]_{i=1}^l|}([C_iA^nA']_{i=1}^r C_{r+1})$$

and so, for large enough n, $[A^nA'B_i]_{i=1}^l$ is conjugated with $[C_iA^nA']_{i=1}^{l-1}C_lA^{n'}D$, for some $n' \geq n - |e|, D \in \Sigma^*$, such that either $n' = n$ and $|D| < |e|$ or $n' < n$ and D a proper prefix of A. If $A'B_i \in \rho(A)^*$, for all $1 \leq i \leq l$, then y is a prefix of some power of $\rho(A)$, hence $\bar{\#}_{L_Y(e)}$ is, in this case, constant. Assume that some $A'B_i$ is not a power of $\rho(A)$. Then

$$y = ([A^nA'B_i]_{i=1}^l)^m[A^nA'B_i]_{i=1}^{l'}A^{n+c}D', \tag{9}$$

where $m \geq 0$, $0 \leq l' \leq l-1$, $-|e| \leq c|A| \leq |e|$, and $D' \in \Sigma^*$, $|D'| \leq |e|$. Moreover, assume that there is a solution $(x_0, y_0) \in \mathrm{Sol}(e)$ with $x_0 = A^{n_0}A'$ and y_0 from (9) with $n = n_0, m = m_0$, such that $|A^{n_0}| \geq |e|, m_0 \geq 1$. Then, for any $n \geq n_0, m \geq m_0$, we have that $x = A^nA'$ and y from (9) (with these m and n) constitute a solution of e. To prove this consider the equality $\varphi(x_0,y_0) = \psi(x_0,y_0)$. For any p, the pth occurrence of A^{n_0} in $\varphi(x_0,y_0)$ overlaps longer than $|\rho(A)|$ the pth occurrence of A^{n_0} in $\psi(x_0,y_0)$. Indeed, if this is not the case, then one can easily show that, for any i, $A'B_i$ is a power of $\rho(A)^*$, a contradiction. So, if n_0 is increased by one, the equality is maintained; the new A is introduced inside the overlapping part of the two pth occurrences of A^n in the two sides of the equality. Consider next the case when m_0 is increased. We make the same reasoning as above, just that we consider here the pth occurrences of y_0 or, more precisely, of $([A^{n_0}A'B_i]_{i=1}^l)^{m_0}$.

Now, for different values of n or m we get different words in (9). By possibly decreasing n, we can write y from (9) as

$$y = ([A^nD_i]_{i=1}^l)^m[A^nD_i]_{i=1}^{l'}A^nD_{l'+1},$$

for some $D_i \in \Sigma^*$, $1 \leq i \leq l$. Therefore, the value of the complexity function $\#_{L_Y(e)}(k)$ differs, for large enough k, by at most a constant from the number of solutions of the Diophantine equation in unknowns n and m

$$\left(\sum_{i=1}^{l}|D_i| + l|A|n\right)m + \sum_{i=1}^{l'+1}|D_i| + (l'+1)|A|n = k.$$

Consequently, $\bar{\#}_{Y_{C,D}(e)}$ is of $\mathcal{D}_1$-type.

Case 2. $l = r$. If $|[C_i]_{i=1}^{r+1}| \geq |[B_i]_{i=1}^{l}|$, then the reasoning is similar with the one in Case 1. Otherwise, we replace Y by DY and X by DX, for any $D \in \Sigma^*$ with

$$|D| = |[B_i]_{i=1}^{l}| - |[C_i]_{i=1}^{r+1}|.$$

As $|D| < |e|$, there are finitely many possibilities and each of those is treated as above, since, under the current assumption, ψ' cannot be empty and so, if $\psi' = Y\psi''$, $\psi'' \in (\Sigma \cup \Xi)^*$, then we have

$$[XB_iD]_{i=1}^{l} Y \varphi'(DX, DY) = Y[C_iDX]_{i=1}^{r} C_{r+1} DY \psi''(DX, DY)$$

and $|[B_iD]_{i=1}^{l}| = |[C_iD]_{i=1}^{r+1}|$.

We notice that the case when ψ' is empty is included either in case 1 or in case 2, so we may assume in the sequel that ψ' is not empty.

Case 3. $l > r$. Consider first the solutions (x, y) where

$$|y| \geq |[A^nA'B_i]_{i=1}^{l}| - |[C_iA^nA']_{i=1}^{r} C_{r+1}|.$$

Then y conjugates

$$[A^nA'B_i]_{i=1}^{l}$$

and

$$[C_iA^nA']_{i=1}^{r} C_{r+1} \operatorname{pref}_{|[A^nA'B_i]_{i=1}^{l}| - |[C_iA^nA']_{i=1}^{r} C_{r+1}|}([A^nA'B_i]_{i=1}^{l})$$

since the first letter of ψ is Y. Therefore, the reasoning is similar with the one in case 1.

Consider now the solutions (x, y) where

$$|y| < |[A^nA'B_i]_{i=1}^{l}| - |[C_iA^nA']_{i=1}^{r} C_{r+1}|. \tag{10}$$

Now the solutions where y is a prefix of a word in A^* bring the constant contribution to the complexity function $\#_{L_Y(e)}$. Consider the solutions in which y starts with a prefix which is periodic with a period A and contains a position in which this period is broken. We may assume that y is long enough, i.e., it starts with A^2. Then, due to uniqueness of $\rho(A)$-factorization of the right-hand side of the equation, the second occurrence of y which contains A^2 and a position where this period is broken matches in the left hand side some of l occurrences of A^n in $[A^nA'B_i]_{i=1}^{l}$ and a position where this period is broken. For each n, there are at most l such positions and therefore there are at most l solutions (x, y) such that $x = A^nA'$, y satisfies (10) and y contains a position where the period A is broken. These solutions again bring at most constant contribution to the complexity function $\bar{\#}_{L_Y(e)}$.

Consequently, in all cases, the function $\bar{\#}_{Y_{C,D}(e)}$ is either constant, or $\mathcal{D}_1$-type, or else exponential, as claimed. ■

The next corollary of Lemma 19 shows the possible complexities for equations in two variables in which the variable appearing first in both sides is the same.

Corollary 20. *If $e : \varphi = \psi$ is an equation over $\Xi = \{X, Y\}$ such that the first variable appearing in each of φ and ψ is X, then $\bar{\#}_{L_X(e)}$ is constant and $\bar{\#}_{L_Y(e)}$ is either constant, or $\mathcal{D}_1$-type, or else exponential.*

We now give some examples in order to see that all situations in Corollary 20 (and so in Lemma 19) are indeed possible.

Example 21. (i) Consider first the equation $e_1 : aXaY = XaYa$. Then, clearly, both $\bar{\#}_{L_X(e_1)}$ and $\bar{\#}_{L_Y(e_1)}$ are constant.

(ii) For the equation in Example 18, we have that $\bar{\#}_{L_Y(e)}$ is $\mathcal{D}_1$-type.

(iii) Our last equation is

$$e_2 : aXYXa = XaYaX.$$

Then, clearly,

$$\mathrm{Sol}(e_2) = \{(a^n, w) \mid n \geq 0, w \in \Sigma^*\},$$

hence $\#_{L_Y(e_2)}$ is exponential as soon as $\mathrm{card}(\Sigma) \geq 2$.

We show also, in the next example, that, for almost all divisor-type functions, there is a language expressible by a two-variable word equation such that its complexity function is precisely the given one.

Example 22. As we have defined it, a divisor-type complexity function is precisely determined by the four non-negative integers $c_i, 1 \leq i \leq 4$ such that $c_1 \geq c_3$ and $c_2 \geq c_4$. Consider such integers given and construct the equation

$$e : XaX^{c_3}b^{c_4}X^{c_1-c_3}b^{c_2-c_4}Y = aXYX^{c_1-c_3}b^{c_2-c_4}X^{c_3}b^{c_4}.$$

It is not difficult to see that, if $c_3 \neq c_1 - c_3$ or $c_4 \neq c_2 - c_4$, then

$$L_Y(e) = \{(a^{c_3n}b^{c_4}a^{(c_1-c_3)n}b^{c_2-c_4})^m a^{c_3n}b^{c_4} \mid n, m \geq 0\}.$$

Thus, for any $k \geq 0$, $\#_{L_Y(e)}(k)$ is the number of solutions of the Diophantine equation (in unknowns n and m)

$$(c_1n + c_2)m + c_3n + c_4 = k.$$

We give also another lemma which considers some solutions which behave similarly with those in Lemma 19. Given an equation e in two variables, we consider those solutions of e in which the X component is of the form A^nBA^nC, for fixed words A, B, C. We define X_1 to be the set of the words of the form A^nBA^nC in $L_X(e)$; Y_1 will denote the set of words y such that, for some n, $(A^nBA^nC, y) \in \mathrm{Sol}(e)$.

Lemma 23. *Assume that $e : \varphi = \psi$ is an equation over $\Xi = \{X, Y\}$ and X_1, Y_1 are defined as above. Then*
(i) $\bar{\#}_{X_1}$ is constant,
(ii) $\bar{\#}_{Y_1}$ is either constant, or $\mathcal{D}_1$-type, or else exponential.

Proof. The proof is similar with the proof of Lemma 19. ∎

Even if the proof is similar with the one of Lemma 19, we stated the result in Lemma 23 separately, since it will be applied in some cases where it would be difficult to see how Lemma 19 can be applied directly.

6 Three technical analyses

We study in this section three systems of equations which will be very useful tools in our analysis in the remaining part of the paper.

The first contains two *different* equations of type $\boldsymbol{e_2}$, that is,

$$\boldsymbol{s_1} : \begin{cases} AUV = VUB \\ CUV = VUD \end{cases}$$

where $A, B, C, D \in \Sigma^*, |A| = |B|, |C| = |D|$ and by "different" we mean that the two equations composing $\boldsymbol{s_1}$ are not the same, that is, we assume that either $A \neq C$ or $B \neq D$.

Rather than being interested in the complexity of the sets of solutions of $\boldsymbol{s_1}$, we shall be concerned with the structure and complexity of the set

$$S_1 = \{uvu \mid (u, v) \in \mathrm{Sol}(\boldsymbol{s_1})\}.$$

Let us now start the analysis of $\boldsymbol{s_1}$. If $|A| = |C|$, then $|A| = |B| = |D|$ and either $A \neq C$ or $B \neq D$. Consider the case $A \neq C$ the other being symmetric. Let (u, v) be a solution of $\boldsymbol{s_1}$. Then since vu is a prefix of both A and C we have $|u| + |v| < |A|$, for any $(u, v) \in \mathrm{Sol}(\boldsymbol{s_1})$, so both $L_U(\boldsymbol{s_1})$ and $L_V(\boldsymbol{s_1})$ are finite. Then S_1 is finite, too.

Assume now $|A| \neq |C|$ and, without loss of generality, $|A| < |C|$. Consider a solution (u, v) of $\boldsymbol{s_1}$ and assume that u and v are long enough.

We have then $C = AC', D = D'B$, and $v = v'D = Cv''$, for some $C', D', v', v'' \in \Sigma^*$. Hence, by the first equation, we get $Auv'D'B = vuB$ and, by the second equation, $AC'uv'D = vuD$. Thus $C'uv' = uv'D'$ and so C' and D' are conjugated.

Assume $C' = (PQ)^n, D' = (QP)^n$, for some $n \geq 1$, where $\rho(C') = PQ, \rho(D') = QP$. We have then $uv' = (PQ)^m P$, for some $m \geq 0$.

Similarly, we obtain that $C'v''u = v''uD'$ and so $v''u = (PQ)^m P$, as $|v''| = |v'|$.

Since we may assume that $|u| \geq |PQ|$, we get that $u = (PQ)^r P$ for some $1 \leq r \leq m$ and $v' = (QP)^{m-r}, v'' = (PQ)^{m-r}$. Therefore, $(QP)^{m-r+n}B = A(PQ)^{m-r+n}$. As we may assume that $|v| \geq 2|A| + |PQ|$, we have $A = (QP)^s Q = B$, for some $s \geq 0$, hence $v = Q(PQ)^{m-r+n+s}$. Finally, we obtain that

$$uvu = (PQ)^{m+r+n+s+1}$$

which brings a constant contribution to $\bar{\#}_{S_1}$ since PQ is a fixed word, $PQ = \rho(A^{-1}C)$.

Consider next the case when one of u and v is short. (If both are short, then there are finitely many of them.) If u is short, then the first equation of $\boldsymbol{s_1}$ gives $Au = PQ, uB = QP, v \in (PQ)^* P$, where PQ is short. Thus $AuvuB \in (PQ)^* P$, hence uvu has a short period. When v is short, then, assuming $|u| \geq |PQ|$, we get by the above $uvu = (PQ)^r Pv(PQ)^r P$.

We have thus proved

Lemma 24. *If the two equations in $\boldsymbol{s_1}$ are different, then there is a constant k such that, for any solution (u, v) of $\boldsymbol{s_1}$ with v long enough, uvu has k as period. In particular, $\bar{\#}_{S_1}$ is constant. The words in S_1 such that v is short are of the form $(CD)^n Cv(CD)^n C$.*

The second system which we study here has the form

$$\boldsymbol{s_2} : \begin{cases} AUV = VUB \\ CDU = UDC \end{cases}$$

where $A, B, C, D \in \Sigma^*$, CD primitive. As for $\boldsymbol{s_1}$, we shall be interested in the set

$$S_2 = \{uvu \mid (x, y) \in \mathrm{Sol}(\boldsymbol{s_2})\}.$$

This analysis will be of help in the next section, when defining the $\mathcal{D}_2$-type complexity.

Consider a solution $(u, v) \in \mathrm{Sol}(\boldsymbol{s_2})$. Then

$$u = (CD)^n C,$$

for some $n \geq 0$, and we have that

$$A(CD)^n Cv = v(CD)^n CB.$$

Thus $A(CD)^n C$ and $(CD)^n CB$ are conjugated by v, hence there are $s, t \in \Sigma^*$ such that

$$\begin{array}{rcl} A(CD)^n C & = & st, \\ (CD)^n CB & = & ts. \end{array} \tag{11}$$

We shall assume in the sequel that n is large enough. When n is small, u is short and, by the first equation of s_2, we get a short period for uvu. Indeed, the first equation in s_2 is equivalent to $(UA)UVU = UVU(BU)$ so uA is a short period of uvu. Let i, j, A' be such that $A = (DC)^i A'(CD)^j$ and $(CD)^{j+1}$ is not a suffix of A and $(DC)^{i+1}$ is not a prefix of $A((CD)^j)^{-1}$. Similarly, let k, l, B' be such that $B = (DC)^k B'(CD)^l$ and $(CD)^{l+1}$ is not a suffix of B and $(DC)^{k+1}$ is not a prefix of $B((CD)^l)^{-1}$. Then, assuming that n is large enough, we have, by (11), that $A' = B'$ and $k + l = i + j$. Hence if $A' \neq D$, then

$$\begin{aligned} v &= ((DC)^i A'(CD)^{j+n}C)^m (DC)^i A'(CD)^l, \\ uvu &= (CD)^n C((DC)^i A'(CD)^{j+n}C)^m (DC)^i A'(CD)^l (CD)^n C, \end{aligned}$$

and finally, again for large enough n, we have

$$\begin{aligned} v &= (A(CD)^n C)^m (DC)^i A'(CD)^l, \\ uvu &= ((CD)^n CA)^{m+1} (CD)^{l+n-j} C. \end{aligned} \tag{12}$$

If $A' = B' = D$, then $v = (DC)^t D$ and $uvu = (CD)^{2n+t+1}C$, for some t.

The conclusion of the analysis above, which will be useful in the next section, is summarized in the next lemma.

Lemma 25. *If* $(u, v) \in \mathrm{Sol}(s_2)$, *then* uvu *either has a short period or is of the form in* (12).

We give next an example of a system of type s_2.

Example 26. Let $A = B = a$, $C = 1$, $D = b$. Then we have the following system of equations

$$\left\{ \begin{array}{rcl} aUV &=& VUa \\ bU &=& Ub \end{array} \right.$$

whose solution is $u = b^n$, $v = (ab^n)^m a$, so that $uvu = (b^n a)^{m+1} b^n$.

The third system s_3 we consider here is of the form:

$$s_3 : \left\{ \begin{array}{lcl} AXBYC &=& DYEXF, \\ A'XB'YC' &=& D'YE'XF', \end{array} \right.$$

with long X and Y. Here we are interested in the complexity of the solutions of s_3. We assume that the equations in s_3 are not equivalent, i.e., for either of the equations of s_3, there is a solution of it which is not a solution of the other. Since we are interested only in long solutions, by making substitutions of the form $X = HX'$, $X = X'H$, $Y = HY'$ and $Y = Y'H$, we may turn the second equation into the form $XB'Y = YE'X$. The system is now in the form

$$s_3 : \left\{ \begin{array}{lcl} AXBYC &=& DYEXF, \\ XB'Y &=& YE'X. \end{array} \right.$$

Since the equations are not equivalent we have $A \neq D$ or $C \neq F$ or $B \neq B'$ or $E \neq E'$. Consider solutions (x, y) such that $|y| > |xB'|$ (the case $|x| \leq |y| \leq |xB'|$ leads to periodic solutions and the other cases are symmetric). Then, from the second equation, we have $y = xB'\bar{y} = \bar{y}E'x$, for some $\bar{y}$. We put it to the first equation obtaining

$$AxB\bar{y}E'xC = DxB'\bar{y}ExF.$$

Now, if $A \neq D$ or $C \neq F$ then x has a short period. Otherwise, $B\bar{y}E' = B'\bar{y}E$ and since either $B \neq B'$ or $E \neq E'$ we have $\bar{y} = (PQ)^iP$ for some short primitive word PQ. Thus

$$y = \bar{y}E'x = (PQ)^iPE'x$$

and

$$y = xB'\bar{y} = xB'(PQ)^iP.$$

Let k, j, $\bar{B}$ be such that $B' = (QP)^k\bar{B}(PQ)^j$ and $(PQ)^{j+1}$ is not a suffix of B' and $(QP)^{k+1}$ is not a prefix of $B'((PQ)^j)^{-1}$. Similarly, let m, n, $\bar{E}$ be such that $E' = (QP)^m\bar{E}(PQ)^n$ and $(PQ)^{n+1}$ is not a suffix of E' and $(QP)^{m+1}$ is not a prefix of $E'((PQ)^n)^{-1}$. Then, since x conjugates $B'(PQ)^iP$ and $(PQ)^iPE'$, we have, for large enough i (i.e., large enough $\bar{y}$), $\bar{E} = \bar{B}$. Finally if $\bar{E} \neq Q$ we obtain

$$\begin{array}{rcl} x & = & ((PQ)^iPE')^t(PQ)^{i+j-n}P \\ y & = & ((PQ)^iPE')^{t+1}(PQ)^{i+j-n}P \end{array} \qquad (13)$$

If $\bar{E} = Q$, then $x, y \in (PQ)^*P$.

The above analysis is summarized in

Lemma 27. *If the equations in* $\boldsymbol{s_3}$ *are not trivially equivalent, then either* X *and* Y *are of the form in* (13) *or one of them is periodic with a short period.*

The above result holds with the same proof for similar systems with two or more equations

$$\boldsymbol{s_3'} \ : \ A_iXB_iYC_i = D_iYE_iXF_i, \quad 1 \leq i \leq n.$$

Lemma 28. *If any two equations in* $\boldsymbol{s_3'}$ *are not trivially equivalent, then either* X *and* Y *are of the form in* (13) *or one of them is periodic with a short period.*

7 Equations of the form $A(X)Y = YB(X)$

We study in this section the complexity for the solutions of equations of the form

$$\boldsymbol{e_3} : [XA_i]_{i=1}^k Y = Y[B_iX]_{i=1}^k,$$

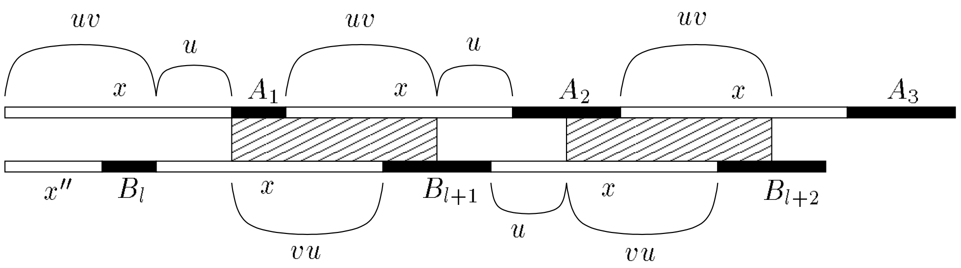

Figure 4: The equation $\boldsymbol{e_3}$ when $|A_i| = |B_{(i+l-2, \mathrm{mod}\, k)+1}|, 1 \le i \le k, |A_1| < |A_2|$.

where $k \geq 2$ and $A_i, B_i \in \Sigma^*, 1 \leq i \leq k$. (For $k = 1$ we have $\boldsymbol{e_1}$.) We shall see also how the $\mathcal{D}_2$-type complexity (defined later in this section) arises naturally. In the following we divide all solutions into several disjoint classes proving that each class brings only constant, $\mathcal{D}_1$-type, $\mathcal{D}_2$-type, linear or exponential contribution to overall complexity.

We denote

$$\begin{aligned} A(X) &= [XA_i]_{i=1}^k, \\ B(X) &= [B_iX]_{i=1}^k. \end{aligned}$$

Consider a solution $(x, y) \in \mathrm{Sol}(\boldsymbol{e_3})$. Then y conjugates $A(x)$ and $B(x)$. If we consider the mapping cycle : $\Sigma^* \longrightarrow \Sigma^*$, defined by $\mathrm{cycle}(aw) = wa$, for any $a \in \Sigma, w \in \Sigma^*$, then

$$A(x) = \mathrm{cycle}^t(B(x)),$$

for some $0 \leq t < |B(x)|$.

Then, there is $l, 2 \leq l \leq k$, such that either $t = |[B_ix]_{i=1}^{l-2}B'|$, where B' is a prefix of B_{l-1}, or $t = |[B_ix]_{i=1}^{l-2}B_{l-1}x'|$, for x' a prefix of x. In the first case x has a period shorter than $|\boldsymbol{e_3}|$ and the contribution of those solutions to $\bar{\#}_{L_X(\boldsymbol{e_3})}$ is constant and, by Lemma 19, the contribution to $\bar{\#}_{L_Y(\boldsymbol{e_3})}$ is either constant, or $\mathcal{D}_1$-type, or else exponential.

Assume now that $t = |[B_ix]_{i=1}^{l-2}B_{l-1}x'|$, for some $2 \leq l \leq k, x' \in \Sigma^*$, such that $x = x'x'', x'' \in \Sigma^*$. Therefore, we have the equality

$$[xA_i]_{i=1}^k = x''[B_ix]_{i=l}^k[B_ix]_{i=1}^{l-2}B_{l-1}x'. \tag{14}$$

the beginning of which is depicted in Fig. 4: above is the beginning of the left-hand side of (14) whereas below is the beginning of the right-hand side.

Let u be the overlapping of the first x above and the first (whole) x (that is, the one after B_l) below. We have that $x = x''B_lu$. Assume that $u \leq \frac{|x|}{2}$. (The case $u \geq \frac{|x|}{2}$ is treated similarly.) If u is short, then x has a short period, $|uA_1|$, and we can apply Lemma 19. Assume that u is long enough.

Since, by our assumption, $|u| \leq |x''B_l|$, we have that u is a prefix of $x''B_l$. Put $x''B_l = uv$. Then $x = uvu$. In Fig. 4, from the overlapping of the second x above and the first x below, we get

$$A_1 uv = vu\,\mathrm{pref}_{|A_1|}(B_{l+1}u), \tag{15}$$

which is of type $\boldsymbol{e_2}$ (in u and v). More precisely, (15) is obtained by considering, in Fig. 4, the prefix uv of the second x above and the suffix vu of the first (whole) x below. If we consider in the same manner the overlapping of the third x above and the second x below, then we obtain another equation of type $\boldsymbol{e_2}$ (in u and v). If this is different from the one in (15) and we consider solutions where v is long then, by Lemma 24, $uvu = x$ has a constant period and hence, by Lemma 19, $\bar{\#}_{L_X(\boldsymbol{e_3})}$ is constant and $\bar{\#}_{L_Y(\boldsymbol{e_3})} \in \{$constant, $\mathcal{D}_1$-type, exponential$\}$. If we consider solutions with short v, then by Lemma 24, $x = uvu$ is of the form $x = (CD)^n Cv(CD)^n C$ and so, by Lemma 23, they bring constant contribution to $\bar{\#}_{L_X(e)}$ and constant or $\mathcal{D}_1$-type contribution to $\bar{\#}_{L_Y(e)}$.

We assume next that all equations of type $\boldsymbol{e_2}$ in u and v we obtain in this way, namely, by considering the prefix uv of the $(k+1)$st x above and the suffix vu of the kth x below in Fig. 4 – see the hatched areas – are identical. Then, necessarily,

$$|A_1| = |B_l|, |A_2| = |B_{l+1}|, \ldots, |A_k| = |B_{l-1}|.$$

We distinguish two cases.

Case 1. $|A_1| = |A_2| = \cdots = |A_k|$. Then, as we assumed that there is only one equation of type $\boldsymbol{e_2}$ in u and v, we obtain that, in fact, $A_1 = A_2 = \cdots = A_k$ and also $B_1 = B_2 = \cdots = B_k$, hence $\boldsymbol{e_3}$ reduces to

$$(XA)^k Y = Y(BX)^k \tag{16}$$

which is equivalent with the equation $XAY = YBX$. If $A = B$, then

$$\{(x, x) \mid x \in \Sigma^*\} \subseteq \mathrm{Sol}(\boldsymbol{e_3}),$$

so both functions $\bar{\#}_{L_X(\boldsymbol{e_3})}$ and $\bar{\#}_{L_Y(\boldsymbol{e_3})}$ are exponential. Otherwise, by Theorem 16, the functions $\bar{\#}_{L_X(\boldsymbol{e_3})}$ and $\bar{\#}_{L_Y(\boldsymbol{e_3})}$ are linear.

Case 2. $|A_i| \neq |A_j|$, for some $1 \leq i, j \leq k$. Without loss of generality, we may assume that $|A_1| < |A_2|$ (this case is shown in Fig. 4). Then u has a short period and we get that (u, v) is a solution of a system of the form $\boldsymbol{s_2}$, say

$$\begin{cases} AUV = VUB \\ CDU = UDC \end{cases} \tag{17}$$

$A, B, C, D \in \Sigma^*, CD$ primitive, where the first equation of (17) is equation (15), that is, $A = A_1, B = \mathrm{pref}_{|A_1|}(B_{l+1}u)$. By our previous assumption, we have a unique such equation.

If u is long enough, then all equations we obtain for u (of the same type as the second equation in (17); given by the segments in between the hatched areas in Fig. 4) are all of the form $(CD)^k U = U(DC)^k$, $k \geq 0$. Therefore, the equation (14) (seen as equation in X) does not only imply (17) but they are even equivalent. We thus have $x = uvu$, for a solution (u, v) of (17) and

$$y = (A(x))^p [B_i x]_{i=1}^{l-2} B_{l-1} x' \tag{18}$$

for $p \geq 0$, where we know that $x' = x(uvB_l^{-1})^{-1} = uA$.

In virtue of (18) and Lemma 25, it appears natural to define the following type of complexity. We say that a natural function $f : \mathbb{N} \longrightarrow \mathbb{N}$ is of *divisor2-type* if there are some integers $c_i, 1 \leq i \leq 8$, such that, $f(k)$ is the number of solutions of the Diophantine equation in unknowns n, m, and p

$$((c_1 n + c_2)m + c_3 n + c_4)p + (c_5 n + c_6)m + c_7 n + c_8 = k.$$

Note here that each divisor-type function is also divisor2-type function We say that a function f is of $\mathcal{D}_2$*-type* complexity if there is divisor2-type function g such that $f = \Theta(\bar{g})$ where $\bar{g}(n) = \max_{1 \leq i \leq n} g(i)$.

The following remark on the complexities will be useful for the general analysis.

Remark 29. According to the definitions, the sum between two $\mathcal{D}_1$-type functions is a function of the same type. Similarly for $\mathcal{D}_2$-type functions. The sum between a $\mathcal{D}_1$-type function and a $\mathcal{D}_2$-type function is a $\mathcal{D}_2$-type function.

As it was intended, $\bar{\#}_{L_Y(e_3)}$ is, in this case, of $\mathcal{D}_2$-type. We have thus proved

Theorem 30. *(i)* $\#_{L_X(e_3)} \in \{$*constant,* $\mathcal{D}_1$*-type, linear, exponential* $\}$,
(ii) $\#_{L_Y(e_3)} \in \{$*constant,* $\mathcal{D}_1$*-type,* $\mathcal{D}_2$*-type, linear, exponential* $\}$.

We give next an example in which the $\mathcal{D}_2$-type complexity is reached.

Example 31. Consider the equation

$$e : XabcXcbabcY = YcbaXcbabcX,$$

which we solve completely in what follows. Consider a solution $(x, y) \in \mathrm{Sol}(e)$. Then $xabcxcbabc$ and $cbaxcbabcx$ are conjugated by y, that is,

$$xabcxcbabc = \mathrm{cycle}^t(cbaxcbabcx),$$

for some $0 \leq t \leq 2|x|+7$. It is not difficult to see that the only possibilities for t are (i) $t = 1$, (ii) $t = |x|+4$, and (iii) $|x|+9 \leq t \leq 2|x|+7$. (The cases $t \in \{0, 2, 3, |x|+3\}$ and $|x|+5 \leq t \leq |x|+8$ are immediately rulled out; for $4 \leq t \leq |x|+2$, we apply the reasoning in the proof of Theorem 30.)

In case (i) we have $x = bab, y = (ba(babc)^3)^n ba(babc)^2 bab, n \geq 0$, and in case (ii) we obtain $x = b, y = (ba(bc)^2 babc)^n babcb, n \geq 0$. Thus, in both cases, we have a constant contribution to either of $\#_{L_X(e)}$ and $\#_{L_Y(e)}$.

The interesting case is (iii). Applying the reasoning in the proof of Theorem 30, we obtain $x = uvu$ where

$$\begin{cases} abcuv = vucba \\ ucb = bcu \end{cases}$$

Therefore, we have

$$\begin{aligned} u &= (bc)^n b, \\ v &= (a(bc)^{n+1}b)^m a, \\ x &= (bc)^n b(a(bc)^{n+1}b)^m a(bc)^n b, \end{aligned}$$

for any $n, m \geq 0$. We then obtain for y the formula

$$y = (xabxaab)^p uv(cba)^{-1}, \tag{19}$$

where u, v, x are given by above and $p \geq 0$. Hence, for any $k \geq 0$, $\#_{L_Y(e)}$ differs by at most 2 from the number of solutions of the Diophantine equation (in unknowns n, m, and p)

$$((4n+8)m + 8n + 11)p + (2n+4)m + 2n - 1 = k.$$

We have used also the fact that, if we denote y in (19) by $y_{m,n,p}$, then $(n_1, m_1, p_1) \neq (n_2, m_2, p_2)$ implies $y_{n_1,m_1,p_1} \neq y_{n_2,m_2,p_2}$. Consequently, $\#_{L_Y(e)}$ is of $\mathcal{D}_2$-type.

8 More technical lemmata

We mainly consider in this section equations in two variables for which all values for one variable are given by some formula involving constant words and some variable integers. We also consider the case when one of the variable is expressed as a fixed formula of the other. These are the last steps before the general analysis in the next section.

Remark 32. Given an equation $e : \varphi = \psi$, if $|\varphi|_X \neq |\psi|_X$ or $|\varphi|_Y \neq |\psi|_Y$, then we obtain a non-trivial relation between lengths of x and y, for any $(x, y) \in \mathrm{Sol}(e)$ and, by Lemma 4, the functions $\bar{\#}_{L_X(e)}$ and $\bar{\#}_{L_Y(e)}$ are constant or exponential.

We may assume also that φ and ψ both start and end with different variables. Indeed, consider the start. (The same reasoning holds for the end.) If both start with constants, then, as the equation is reduced, it has no solution. If one starts with some constants, then there are two possibilities: (i) the same variable appears as the first variable in both φ and ψ and (ii) the first variable appearing in φ and ψ is not the same. In case (i), we apply Lemma 19 and we are done. For (ii), we can eliminate the constants at the begining by a suitable substitution.

The first result is similar with the one in Lemma 23 but a bit more complicated. We consider here those solutions of e in which the X component is of the form $(A^nB)^mA^nC$, for fixed words A, B, C. As above, we define X_2 to be the set of the words of the form $(A^nB)^mA^nC$ in $L_X(e)$; Y_2 will denote the set of words y such that, for some n and m, $((A^nB)^mA^nC, y) \in \mathrm{Sol}(e)$.

Lemma 33. *Assume that $e : \varphi = \psi$ is an equation over $\Xi = \{X, Y\}$ and X_2, Y_2 are defined as above. Then*

(i) $\bar{\#}_{X_2}$ is either constant or $\mathcal{D}_1$-type,

(ii) $\bar{\#}_{Y_2}$ is either constant, or $\mathcal{D}_1$-type, or $\mathcal{D}_2$-type, or else exponential.

Proof. The basic ideas are similar with the ones in the proof of Lemma 19. Consider $e : \varphi = \psi$ and assume $\bar{\#}_{L_Y(e)}$ is not exponential. Then also $\bar{\#}_{Y_2}$ is not exponential.

By Remark 32, we may assume that $|\varphi|_X = |\psi|_X$, $|\varphi|_Y = |\psi|_Y$ and φ and ψ have the forms in (8), where φ' and ψ' end with different variables. Also, if $B \in \rho(A)^*$, then $X_1 \subseteq \rho(A)^*C$ and we can apply Lemma 19. Assume then $B \notin \rho(A)^*$, that is, there is no equality $B\rho(A)^s = \rho(A)^sB$, for some $s \geq 1$.

Take then some $n \geq |e|$, $m \geq |e|$ such that $x = (A^nB)^mA^nC \in L_X(e)$. (If there are no such n or m, then we can apply Lemma 19 or Lemma 23.) Thus, there is $y \in \Sigma^*$ such that

$$[(A^nB)^mA^nCB_i]_{i=1}^l y\varphi'(x, y) = y[C_i(A^nB)^mA^nC]_{i=1}^r C_{r+1}\psi'(x, y).$$

Consider, as in the proof of Lemma 19, three cases, depending on the relation between l and r.

Case 1. $l < r$. We have then

$$y = ([(A^nB)^mA^nCB_i]_{i=1}^l)^p[(A^nB)^mA^nCB_i]_{i=1}^{l'}(A^nB)^{m'}A^nC, \tag{20}$$

where $p \geq 0, 1 \leq l' \leq l-1, 0 \leq m' \leq m$. (For the suffix A^nC we considered the end of the equation.)

Assume first that $m' < m$. Then $CC_i = CB_i = B$, for any $1 \leq i \leq l$. If $B = CB'$, then $C_i = B_i = B'$, for all $1 \leq i \leq l$. The equation becomes

$$e : \quad (XB')^lY\varphi' = Y(B'X)^l\psi'', \quad \text{where } \psi'' = [C_iX]_{i=l+1}^r C_{r+1}\psi'.$$

Therefore, e is equivalent with the system

$$s : \begin{cases} XB'Y = YB'X \\ \varphi' = \psi'' \end{cases}$$

Now, $x = (A^nB)^mA^nC$ implies, by the first equation of s, that $y = (A^nB)^pA^nC$, for some $p \geq 0$. We prove that the contribution to either of $\bar{\#}_{X_2}$ and $\bar{\#}_{Y_2}$ is in this case $\mathcal{D}_1$-type.

First, since the corresponding occurrences of A^n in the two sides of e overlap long enough, n can be increased and the equality is maintained.

Consider the equality $\varphi'(x,y) = \psi''(x,y)$. Since $B \notin \rho(A)^*$, the corresponding occurrences of B in the two sides of the equality must be perfectly matched (i.e., overlap completely). Therefore, $\varphi'(x,y) = \psi''(x,y) = [(A^nB)^{l_i}A^nCD_i]_{i=1}^k, D_i \neq B'$, where each D_i corresponds to a place in between variables, that is, either $D_i = 1$ or $D_i = D$, for some constant $D \neq B'$ of e. Consider each part $(A^nB)^{l_i}A^nC$ separately. There are two possibilities. First, for any i, the number of x's (y's, resp.) in this part is the same in φ' and ψ''. In this case, no matter how m and p are increased, the equality is maintained. Second, for some i, the number of x's or y's is not the same. Then, out of all such i's, we get a unique relation between m and p, of the form $cm = dp$. In this case, we can increase m by $\frac{d}{(c,d)}$ and p by $\frac{c}{(c,d)}$ where (c,d) stands for greatest common divisor of c and d. Consequently, in any case, the contribution to either of $\bar{\#}_{X_2}$ and $\bar{\#}_{Y_2}$ is $\mathcal{D}_1$-type.

Consider next the case $m = m'$. Then y is given by (20) with m instead of m'. Consider the equality $\varphi(x,y) = \psi(x,y)$. The corresponding occurrences of A^n in the left and right must be perfectly matched since otherwise $B \in \rho(A)^*$. Therefore, the corresponding occurrences of $(A^nB)^mA^nC$ are also perfectly matched. Consequently, n and m can be increased. For p, this is true as soon as $p \geq |e|$. Now, if $CB_i = B$, for all $1 \leq i \leq l$, we have that y has the form $y = (A^nB)^pA^nC$ and reason as above. Otherwise, for different triples (n,m,p), we get different y's and x's. Therefore, we have in this case $\mathcal{D}_1$-type contribution to $\bar{\#}_{X_2}$ and $\mathcal{D}_2$-type contribution to $\bar{\#}_{Y_2}$.

Case 2. $l = r$. This reduces to Case 1 as in the proof of Lemma 19.

Case 3. $l > r$. For long y, it reduces to Case 1 as in the proof of Lemma 19. If y is short, that is,

$$|y| < |[(A^nB)^mA^nCB_i]_{i=1}^l| - |[C_i(A^nB)^mA^nC]_{i=1}^r C_{r+1}|,$$

then y has the form $y = (A^nB)^pA^nC$, so we can reason as above. ■

Our next lemma deals with the solutions of the form $(x, \phi(x))$, for some formula $\phi(x) \in (\Sigma^*x)^+\Sigma^*$.

Lemma 34. *Let e be an equation on two variables X and Y. The solutions (x,y) such that $y = \phi(x)$ for a fixed ϕ bring either constant or exponential contribution to the complexity functions $\bar{\#}_{L_Y(e)}$, $\bar{\#}_{L_X(e)}$.*

Proof. We put $\phi(X)$ instead of all occurrences of Y in e obtaining an equation with one variable. If it is trivial (both sides are identical) then all solutions of the form $(x, \phi(x))$ are solutions of the equation and consequently both $\bar{\#}_{L_X(e)}$ and $\bar{\#}_{L_Y(e)}$ are exponential. If the equation is not trivial then its long solutions are of the form A^nA' so that the contribution of them to $\bar{\#}_{L_X(e)}$ is constant and consequently also the contribution of them to $\bar{\#}_{L_Y(e)}$ is constant. ■

9 The general form

We now study two-variable word equations of the general form. We show that one cannot obtain as complexities of their expressed languages anything but the five types we have identified so far, namely constant, $\mathcal{D}_1$-type, $\mathcal{D}_2$-type, linear, and exponential.

Now we describe the basic ideas of the procedure for treating these equations. Depending on the form of the equation we will treat it differently. We also divide the solutions into finite number of smaller parts. For each part we will prove that it brings either constant, $\mathcal{D}_1$-type, $\mathcal{D}_2$-type, linear, or exponential contribution to overall complexity. By Remark 29, this will prove that the overall complexities can be only constant, $\mathcal{D}_1$-type, $\mathcal{D}_2$-type, linear, or exponential. There are several cases when the procedure stops. It stops when it finds a formula for X or Y (then we apply one of the Lemmata 19, 23, and 33 to the starting equation e), or a formula of the form $Y = \phi(X)$ (then we apply Lemma 34). Otherwise, we end up with a set of equations of the form $AXBYC = DYEXF$. If there are at least two of them, then we apply Lemma 28. If there is only one, then, by considerations in Section 4, the solutions of it bring linear or exponential contribution to the complexity function.

Consider an equation $e : \varphi = \psi$ over $\Xi = \{X, Y\}$. We make several observations which help us to restrict a bit the form of e, without loss of generality. First, we may assume that $|\varphi|_X \geq 2$ and $|\varphi|_Y \geq 2$ since the other cases have been already studied. Second, we may assume, by Remark 32, that $|\varphi|_X = |\psi|_X$ and $|\varphi|_Y = |\psi|_Y$ and also that φ and ψ both start and end with different variables.

Consequently, we may assume for e the form

$$e : X\varphi_1(X)Y\varphi(X, Y) = Y\psi_1(X)Y\psi(X, Y),$$

where $\varphi, \psi \in (\Sigma \cup \Xi)^*, \varphi_1, \psi_1 \in (\Sigma \cup \{X\})^*$. Since we investigate both functions $\bar{\#}_{L_X(e)}$ and $\bar{\#}_{L_Y(e)}$, we may take into account only those solutions (x, y) of e with $|x| \leq |y|$.

Case 1. $|\psi_1(X)|_X \geq |X\varphi_1(X)|_X$. Let $\psi_1'(X)$ be the shortest prefix of $\psi_1(X)$ such that $|\psi_1'(X)|_X = |X\varphi_1(X)|_X$. If, additionally, $|\psi_1'(X)| = |X\varphi_1(X)|$, then we can divide the equation e into two parts $X\varphi_1(X)Y = Y\psi_1'(X)$ and $\varphi(X, Y) = \psi_1''(X)Y\psi(X, Y)$, where $\psi_1(X) = \psi_1'(X)\psi_1''(X)$. Using the considerations in Section 7, we conclude that the solutions of the first equation are either the solutions of the equation $XAY = YBX$ for some constants A, B or of the form (12). In the last case we stop with the formula for X. If $XAY = YBX$, then we proceed by processing the equation $e' : \varphi(X, Y) = \psi_1''(X)Y\psi(X, Y)$. Observe that the number of X's and Y's are the same on both sides of e'. Now if e' is in form $A(X)Y = YB(X)$ we either generate another equation of the form $XAY = YBX$ or again a formula for X. In the latter case we stop. In the former case we generate the new equation of the form $XAY = YBX$ and stop. If e' is not of the form $A(X)Y = YB(X)$, then we proceed as we proceeded with e. Notice that e' is strictly shorter.

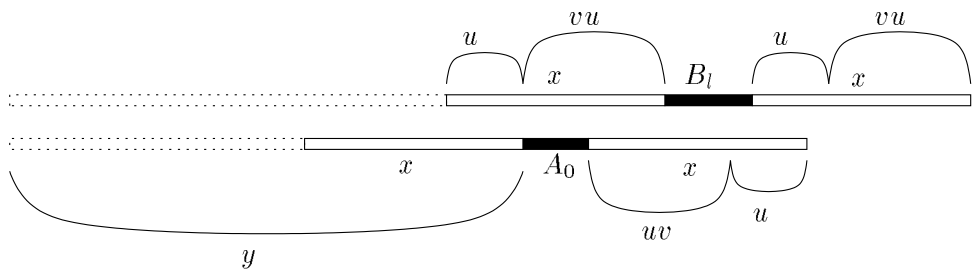

Figure 5: If $|A_0| \neq |B_l|$, then we obtain a system $\boldsymbol{s_1}$ for U and V.

If $|\psi_1'(X)| < |X\varphi_1(X)|$, then we consider all (finitely many) possibilities of completing the word $\psi_1'(X)$ to a word of length $|X\varphi_1(X)|$ by adding a constant word D of length $|X\varphi_1(X)| - |\psi_1'(X)|$. Then, for any such D, we have $X\varphi_1(X)Y = Y\psi_1'(X)D$ and $D\varphi(X,Y) = \psi_1''(X)Y\psi(X,Y)$. We may assume that $\psi_1''(X)$ starts with X or is empty. Now, if the first variable appearing in $\varphi(X,Y)$ and $\psi_1''(X)Y$ is the same, then we apply Corollary 20. Otherwise, for any solution (x,y), we have that x and y are x is a prefix of y. Then $Dx\cdots = x\cdots$ and so x has a short period and we apply Lemma 19.

If $|\psi_1'(X)| > |X\varphi_1(X)|$, then we have a similar reasoning as above but D is such that $\psi_1'(X) = X\varphi_1(X)D$.

Case 2. $|\psi_1(X)|_X < |X\varphi_1(X)|_X$ and consider solutions (x,y) such that $|y| \geq |x\varphi_1(x)| - |\psi_1(x)|$. Let $B'(X)$ be the shortest prefix of $X\varphi_1(X)$ such that $|B'(X)|_X = |X\varphi_1(X)|_X - |\psi_1(X)|_X$. We have then $|X\varphi_1(X)|_X = |\psi_1(X)B'(X)|_X$. If, additionally, $|X\varphi_1(X)| = |\psi_1(X)B'(X)|$, then we may divide the equation e into two parts, $X\varphi_1(X)Y = Y\psi_1(X)B'(X)$ and $B'(X)\varphi(X,Y) = Y\psi(X,Y)$, so that we proceed as in Case 1. If $|X\varphi_1(X)| \neq |\psi_1(X)B'(X)|$, then we show as in Case 1 that x has a short period.

Case 3. $|\psi_1(X)|_X < |X\varphi_1(X)|_X$ and consider solutions (x,y) such that $|y| < |x\varphi_1(x)| - |\psi_1(x)|$. Recall that we consider solutions for which $|y| \geq |x|$. The considerations in this case are similar to those in Section 7. Let $X\varphi_1(X) = [XB_k]_{k=1}^j$ and $\psi_1(X) = A_0[XA_k]_{k=1}^i$. Then $y = [xB_k]_{k=1}^{l-1}x'$ or $y = [xB_k]_{k=1}^{l-1}xB'$ where B' is a prefix of B_l. In the latter case we have a formula of the form $Y = \phi(X)$ so we stop. In the former case, if $|A_0| \neq |B_l|$, see Fig. 5, (we assume here u long and $u \leq \lfloor \frac{x}{2} \rfloor$ – the other case is treated similarly; we assume also that x is a suffix of y) then (u,v) verifies $A_0UV = VUB'$ where $B' = \mathrm{pref}_{|A_0|} B_l x$, and $CDU = UDC$ for some C, D. This is a system of equations $\boldsymbol{s_2}$ from Section 6. The solution of it gives a formula for x, $x = uvu$, so we stop. We obtain a similar formula if $|A_i| \neq |B_{l+i}|$, for some i, so in those cases we stop. Now we consider the case $|A_i| = |B_{l+i}|$. Then $A_iUV = VUB_{l+i}$ and again, if, for some i, j, $A_i \neq A_j$ or $B_{l+i} \neq B_{l+j}$ we have at least two equations forming a system $\boldsymbol{s_2}$ and we stop with the formula for X.

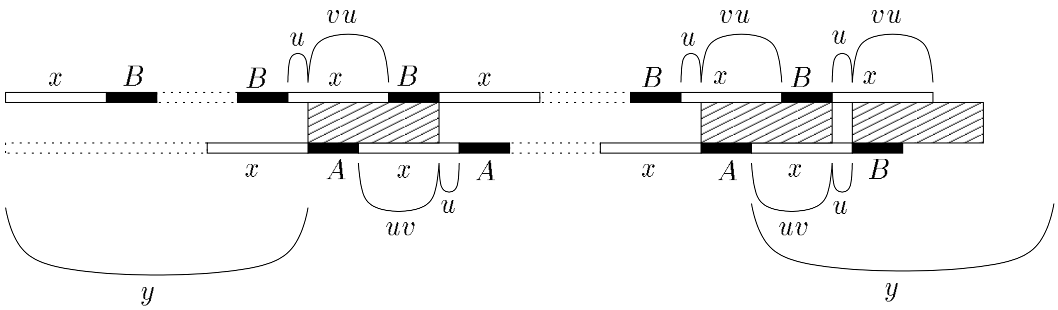

Figure 6: The hatched parts correspond to the word $h = Auv = vuB$.

The remaining case is $A_0 = A_i$ and $B_l = B_{l+i}$, for all i. Now we use the properties of the end of the equation. Since we may assume that one side of the equation ends with X and the other by Y we have $y = x''C_1x \cdots C_sx$. As previously, we either stop with a formula for X or $C_i = A_0$ and $B_i = B_l$, for all i. Denote $B = B_l$, $A = A_0$. We have $X\varphi_1(X) = (XB)^j$, $\psi_1(X) = A(XA)^i$ and $|A| = |B|$. We have the situation depicted in Fig. 6; the hatched areas correspond to the word $h = Auv = vuB$. The second occurrence of y has xB as prefix, so B is a prefix of h. Hence, $A = B$ and the equation is in form $(XA)^iY \cdots = Y(AX)^jAY$ with $j < i$. We have $y = (xA)^ku$, $x = uvu$, and $Auv = vuA$. Then, $xAy = yAx$. We have $(xA)^iy \cdots = y(Ax)^i \cdots = y(Ax)^jAy \cdots$ and finally $(xA)^{i-j-1}x \cdots = y \cdots$. Now we generate $XAY = YAX$ and proceed by considering $(XA)^{i-j-1}X \cdots = Y \cdots$ in the way we considered e. Note here that the last equation is shorter than the starting one, thus the procedure will eventually stop.

10 Complexity and characterization of the expressible languages

Summarizing the above results, we have proved the following theorem which characterizes all possible complexity functions of the languages expressible by two-variable word equations.

Theorem 35. *Let e be an equation with two variables X, Y. Then*

$$\bar{\#}_{L_X(e)}, \bar{\#}_{L_Y(e)} \in \{\textit{constant, } \mathcal{D}_1\textit{-type, } \mathcal{D}_2\textit{-type, linear, exponential}\}.$$

As another consequence of our study, we can give the general forms of the languages expressible by two-variable word equations.

First, we need the notion of pattern language, from [An], cf. also [JSSY]. A *pattern* is a word over the alphabet $\Sigma \cup \Xi$. A *pattern language* generated by a pattern α, denoted $L(\alpha)$ is the set of all morphic images of α under morphisms $h : (\Sigma \cup \Xi)^* \longrightarrow \Sigma^*$ satisfying $h(a) = a$, for any $a \in \Sigma$.

By Theorem 13 in [KMP], we know that, for any language L which is expressible by a two-variable word equation, if $\bar{\#}_L$ is exponential, then there exists a pattern α containing occurrences of one variable only such that $L(\alpha) \subseteq L$.

We have then the following theorem which characterizes the languages expressible by two-variable word equations.

Theorem 36. *For any language L which is expressible by a two-variable word equation, we have*
(i) if $\bar{\#}_L$ is exponential, then L contains a pattern language,
(ii) if $\bar{\#}_L$ is not exponential, then L is a union of
(a) a finite language,
(b) finitely many parametric formulae of the forms

- $A^n B$,
- $(A^n B)^m A^n C$,
- $([A^n B_i]_{i=1}^k)^m \operatorname{pref}([A^n B_i]_{i=1}^k)$,
- $([(A^n B)^m A^n C D_i]_{i=1}^k)^p \operatorname{pref}([(A^n B)^m A^n C D_i]_{i=1}^k)$, *and*

(c) solutions of an equation $XAY = YBX$ (including the case $A = B$); these solutions can be expressed as compositions of finite number of substitutions which can be computed on the basis of the graph for $XAY = YBX$.

11 Minimal solutions and solvability

We consider here the lengths of solutions for two-variable word equations. The reasoning will be again based on the above analysis. We take an equation $e : \varphi = \psi$ over $\Xi = \{X, Y\}$ and assume e is *solvable*, i.e., it has solutions. Then, we show that there is a solution of e for which the length of one component is linear in terms of $|e|$ while the length of the other is quadratic. The following example shows that we cannot hope to improve these bounds by more than a constant.

Example 37. Consider the equation

$$e : \quad aXa^nbX^n = XaXbY.$$

The equation e has a unique solution which is $(x, y) = (a^n, a^{n^2})$. As $|e| = 2n + 8$, we have that $|x| = \mathcal{O}(|e|)$ and $|y| = \mathcal{O}(|e|^2)$.

Before starting the proof of our bounds, we shall give the corresponding result for one-variable word equations, which will be useful.

Lemma 38. *If e is a solvable one-variable word equation, then e has a solution of length at most $|e| - 1$.*

Proof. Put $e : \varphi = \psi$. If $|\varphi|_X \neq |\psi|_X$, then there is only one possible length for the solutions of e and this is at most $|e| - 1$ since it is the solution of a linear equation with coefficients at most $|e| - 1$. Assume $|\varphi|_X = |\psi|_X$ and also that $\varphi = X \cdots$, $\psi = AX \cdots$, A a constant word. Then $x \in \mathrm{Sol}(e)$ implies $x = A^n A'$, for some $n \geq 0$, A' a prefix of A. Since e is solvable, we have also that $|\varphi| = |\psi|$. Thus, if $|A^n A'| \geq |e|$, then, for any $1 \leq k \leq |\varphi|_X$, the kth occurrence of A^n in $\varphi(x)$ overlaps at least $\frac{|A^n|}{2} \geq |A|$ the kth occurrence of A^n in $\psi(x)$. Therefore, $|x| \geq |e|$ implies that n can be decreased such that the equality is maintained. The proof is completed. ∎

We notice that the bound in Lemma 38 is tight as shown by any equation $X = A$, A a constant word. When the equation e is *balanced*, then a better bound can be given, but we shall use the general one above.

The way we are to use the result in Lemma 38 is clear. It is enough to prove that one component of the solution can be made of linear length since this will imply, by way of Lemma 38, a quadratic bound on the other.

Let us now prove the bounds we mentioned. We shall follow the different steps of our analysis, that means, we shall consider first the special forms of solutions and equations and then the general case. The main idea is to consider solutions of length bigger than certain values and then show that they can be shortened such that the respective equation is still satisfied. In order to simplify the arguments, we shall often refer to some of the previous proofs.

First, for the equations of type $\boldsymbol{e_1}$, it is clear from the graph in Fig. 3 that $\boldsymbol{e_1}$, if solvable, has solutions with both components shorter than $|\boldsymbol{e_1}|$. Similarly for $\boldsymbol{e_2}$.

Consider next the periodic solutions in Lemma 19. Asume $x = A^n A'$, $|A| \leq |e|$, A' a prefix of A. Consider first the case when y is also periodic, say $y = B^m B'$, $|B| \leq |e|$, B' a prefix of B. For simplicity, assume A and B primitive. Now, if A and B are not conjugated, then we can decrease n and m independently as soon as $|A^n| \geq |e|, |B| \geq |e|$. Therefore, if there are such solutions, then there are some for which both components are shorter than $|e|$. If A and B are conjugated, then $Y = B'' A^m B'''$ and $\varphi(x, y) = \psi(x, y) = [C_i A^{l_i}]_{i=1}^{k} C_{k+1}$, where $C_i \notin A^*$ corresponds to a place in between two variables in either side of e. Assume $|A^n| \geq |e|, |A^m| \geq |e|$ and consider each part A^{l_i} separately. If, for any i, this part is composed of the same number of x's (y's) in the left and in the right of e, then we can decrease n and m independently. Otherwise, there is a unique relation of the form $cn = dm$ and we can simultaneously decrease n by $\frac{d}{(c,d)}$ and m by $\frac{c}{(c,d)}$. Due to the conditions $|A^n| \geq |e|$ and $|A^m| \geq |e|$, the constants play no important role here but, if we allow, for instance, $|A^n| < |e|$, then we might have that some A^{l_i} is composed in one side of constants only, as it happens for the equation $aXba^n bX^n = XabXbY^n$ (which is a slight modification of the equation in Example 37). In such a case, by Lemma 38, we can decrease m such that $|y| \leq |e|^2$. The worst case is, as in Example 37, when y depends on x only and x cannot be shorthened.

Consider now the general case in Lemma 19, that is, when x is periodic and y is arbitrary. If Y is the first variable occurring in both φ and ψ, then y has a short period and we apply the above reasoning. If X appear first in both sides, then the occurrences of x at the beginning, together with the constants in between can be reduced. We may therefore assume that the first variable appearing in φ is X and in ψ is Y. Assume first e has the form $X \cdots = Y \cdots$. We follow the reasoning in the proof of Lemma 19. Using the same notations, in the case $l < r$ we have that if, for all i, $A'B_i \in \rho(A)^*$, then y has short period and this was studied. Otherwise, n_0 can be decreased such that $|x| \leq |e|$ and we apply Lemma 38. The case $l = r$ is reduced to the previous one and in case $l > r$ we consider only long y which again is reduced to the first case. If e is such that one of the sides starts with constants, then we have a similar reasoning by assuming $|x| \geq 2|e|$.

For the formula in Lemma 23, we have a similar reasoning.

Consider next the case of the formula in Lemma 33. The reasoning is similar. Where we proved that n, m can be increased maintaining the equality, we can decrease them as well.

For the formula in Lemma 34, we replace Y by $\phi(X)$ and if the obtained equation is not trivial, then x has a short period and this was studied. If the equation is trivial, then the length of x can be reduced to zero and so $|y| \leq |e|$..

Finally, for the general form, we use the considerations in Section 9. As shown there the general case reduces to one of the cases we studied above. Notice that in the case when two or more equations of the type $\boldsymbol{e_1}$ are obtained, we have, according to Lemma 28, that either one component has short period, which was studied, or they have the form in (13) which is of the same type as the formula in Lemma 23.

We have therefore proved

Theorem 39. *If e is a solvable two-variable word equation over $\Xi = \{X, Y\}$, then e has a solution (x, y) such that $|x| \leq 2|e|$, $|y| \leq 2|e|^2$.*

Given a two-variable word equation $e : \varphi = \psi$, and two non-negative numbers l_x, l_y, it is clear that we can check in time $|e| + |\varphi\psi|_X(l_x - 1) + |\varphi\psi|_Y(l_y - 1)$ whether e has a solution (x, y) for some x, y with $|x| = l_x$, $|y| = l_y$. Therefore, we get immediately from Theorem 39 the following result.

Theorem 40. *The solvability of two-variable word equations can be tested in time $\mathcal{O}(n^6)$.*

Notice that the only polynomial-time algorithm known for this problem is the one given by Charatonik and Pacholski [ChPa] and it runs in time $\mathcal{O}(n^{100})$. In fact, they intended to prove mainly that the problem can be done in polynomial time.

Another consequence of Theorem 39 concerns the complexity of languages expressible by three-variable word equations. The following result can be proved as Theorem 13 in [KMP].

Theorem 41. *Let L be a language expressible by a three-variable word equation. Then either there is a one-variable pattern α such that $L(\alpha) \subseteq L$ or $\#_L(n) = \mathcal{O}(n^3)$.*

12 Further research

We end our paper by leaving some open problems.

1. What is the precise complexity of $\mathcal{D}_1$-type and $\mathcal{D}_2$-type functions?

2. What are those $\mathcal{D}_1$-type or $\mathcal{D}_2$-type functions that we can obtained as complexities of languages defined by two-variable word equations?

3. How good are the bounds in Theorem 39? By Example 37 they are optimal up to a constant factor, but are all the cases when the bounds are reached pathological?

4. Improve the n^6-algorithm for solvability.

5. Theorem 36 describes precisely the form of the languages of non-exponential complexity which are expressible by two-variable word equations. It remains an open problem to describe the form of those of exponential complexity.

6. Find a polynomial-time algorithm for finding all solutions of an input equation. Observe here that the algorithm which can be derived is not polynomial since in many places, e.g. in the proof of Lemma 19, we split the solutions into exponential number of parts one for each short word D.

Acknowledgement

The authors would like to thank Dr. Filippo Mignosi (University of Palermo, Italy) for useful discussions during the preparation of the paper.

References

[An] Angluin, D., Finding patterns common to a set of strings, *J. Comput. System Sci.* **21**(1) (1980) 46 – 62.

[Be] Berstel, J., Recent results in Sturmian words, in J. Dassow, G. Rozenberg, A. Salomaa, eds., *Developments in Language Theory II*, 13 – 24, World Sci. Publishing, 1996.

[ChPa] Charatonik, W. and Pacholski, L., Word equations with two variables, *Proc. of IWW-ERT'91*, H. Abdulrab, J. P. Pecuchet, eds., 43 – 57, LNCS 667, Springer, Berlin, 1991.

[ChKa] Choffrut, C. and Karhumäki, J., Combinatorics of words, in G. Rozenberg, A. Salomaa, eds., *Handbook of Formal Languages*, 329 – 438, Springer, Berlin, 1997.

[dLMi] de Luca, A. and Mignosi, F., Some combinatorial properties of sturmian words, *Theoret. Comput. Sci.* **136** 361 – 385, 1994.

[OGM] Eyono Obono, S., Goralcik, P., and Maksimenko, M., Efficient solving of the word equations in one variable, in *Proc. of MFCS'94*, 336 – 341, LNCS 841, Springer, Berlin, 1994.

[Hm] Hmelevskii, Yu. I., Equations in free semigroups, *Trudy Mat. Inst. Steklov* **107** 1971. English transl. *Proc Steklov Inst. of Mathematics* **107** (1971), Amer. Math. Soc., 1976.

[JSSY] Jiang, T., Salomaa, A., Salomaa, K., and Yu, S., Decision problems for patterns, *J. Comput. System Sci.* **50**(1) (1995) 53 – 63.

[KMP] Karhumäki, J., Mignosi, F., and Plandowski, W., The expressibility of languages and relations by word equations, in *Proc. of ICALP'97*, 98 – 109, LNCS 1256, Springer, Berlin, 1997.

[KoPa] Koscielski, A. and Pacholski, L., Complexity of Makanin's algorithm, *Journal of the ACM*, **43**(4), 670-684, 1996.

[Lo] Lothaire, M., *Combinatorics on Words*, Addison-Wesley, Reading, Mass., 1983.

[Ma] Makanin, G.S., The problem of solvability of equations in a free semigroup, *Mat. Sb.* **103** (145), 147-233, 1977. English transl. in *Math. U.S.S.R. Sb.* **32**, 1977.

[Ra] Rauzy, G., Mots infinis en arithmetique, in M. Nivat, D. Perrin, eds., *Automata on infinite words*, LNCS 192, Springer, Berlin, 1984.

[Raz1] Razborov, A., On systems of equations in a free group, *Math. USSR Izvestija* **25**(1), 115 – 162, 1985.

[Raz2] Razborov, A., On systems of equations in a free group, Ph.D. Thesis, Moscow State University, 1987.

[Si] Sierpinski, W., *Elementary Theory of Numbers*, Elseviers Science Publishers B.V., Amsterdam, and PWN - Polish Scientific Publishers, Warszawa, 1988.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**