One-Variable Word Equations and Three-Variable Constant-Free Word Equations*

Dirk Nowotka

Department of Computer Science, Kiel University, 24098 Kiel, Germany, dn@informatik.uni-kiel.de

Aleksi Saarela

Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland, amsaar@utu.fi

We prove connections between one-variable word equations and three-variable constant-free word equations, and use them to prove that the number of equations in an independent system of three-variable constant-free equations is at most logarithmic with respect to the length of the shortest equation in the system. We also study two well-known conjectures. The first conjecture claims that there is a constant c such that every one-variable equation has either infinitely many solutions or at most c. The second conjecture claims that there is a constant c such that every independent system of three-variable constant-free equations with a nonperiodic solution is of size at most c. We prove that the first conjecture implies the second one, possibly for a different constant.

1. Introduction

One of the most important open problems in combinatorics on words is the following question: For a given n, what is the maximal size of an independent system of constant-free word equations on n variables? It is known that every system of word equations is equivalent to a finite subsystem and, consequently, every independent system is finite. This is known as *Ehrenfeucht's compactness property*. It was conjectured by Ehrenfeucht in a language theoretic setting, formulated in terms of word equations by Culik and Karhumäki [3], and proved by Albert and Lawrence [1] and independently by Guba [6]. If $n > 2$, no finite upper bound for the size of independent systems is known. The largest known independent systems have size $\Theta(n^4)$ [11]. Some related results and variations of the problem are discussed in [12]. Decreasing chains of word equations, which were first studied by Honkala [9], are a variation of independent systems that is especially important in this article.

The difference between the best known lower and upper bounds is particularly striking in the case of three variables: The largest known independent systems consist of just three equations, but it is not even known whether there exists a constant

*This work has been supported by the DFG Heisenberg grant 590179 (Dirk Nowotka), the DFG research grant 614256 and the Vilho, Yrjö and Kalle Väisälä Foundation (Aleksi Saarela).

c such that every independent system has size c or less. When studying independent systems, it is often additionally required that the system has a nonperiodic solution, and then the largest known examples consist of just two equations. If the above-mentioned constant c exists, then the corresponding constant with this additional requirement is either c or $c - 1$.

There have been some recent advances regarding this topic. The first nontrivial upper bound was proved by Saarela [15]: The size of an independent system on three variables is at most quadratic with respect to the length of the shortest equation in the system. This bound was improved to a linear one by Holub and Žemlička [8]; this is currently the best known result.

Another well-known but less central open problem on word equations is the following question: If a one-variable word equation with constants has only finitely many solutions, then what is the maximal number of solutions it can have? It has been conjectured that the answer is two, but we disprove this conjecture by giving an example with exactly three solutions (see Lemma 9 and Example 10). The best known upper bound, proved by Laine and Plandowski [13], is logarithmic with respect to the number of occurrences of the variable in the equation. Similar but slightly weaker results were proved in [4] and [5].

In this article we establish a connection between three-variable constant-free equations and one-variable equations with constants. This is done by using an old result by Budkina and Markov [2], or a similar result by Spehner [17]. We use this connection to prove two main results.

The first main result is that the size of an independent system of three-variable equations is logarithmic with respect to the length of the shortest equation in the system. This improves the existing linear bound that was mentioned above. This result is based on the logarithmic bound for the number of solutions of one-variable equations. A similar result is proved for decreasing chains of equations.

The second main result is an explicit link between two existing conjectures: If there exists a constant c such that the number of solutions of a one-variable equation is either infinite or at most c, then there exists a constant c' such that the size of an independent system of three-variable constant-free equations with a nonperiodic solution is at most c'. Furthermore, if $c = 3$, then we can let $c' = 17$. The number 17 here is very unlikely to be optimal, and we expect that the result could be improved by a more careful analysis. Again, a similar result is proved for decreasing chains of equations.

2. Preliminaries

Let Ξ be an alphabet of variables and Σ an alphabet of constants. A *constant-free word equation* is a pair $(u, v) \in \Xi^* \times \Xi^*$, and the *solutions* of this equation are the morphisms $h : \Xi^* \to \Sigma^*$ such that $h(u) = h(v)$. A *word equation with constants* is a pair $(u, v) \in (\Xi \cup \Sigma)^* \times (\Xi \cup \Sigma)^*$, and the *solutions* of this equation are the constant-preserving morphisms $h : (\Xi \cup \Sigma)^* \to \Sigma^*$ such that $h(u) = h(v)$. Next we state many definitions that work for both types of equations.

A solution h is *periodic* if $h(pq) = h(qp)$ for all words p, q in the domain of h, and *nonperiodic* otherwise. This definition is interesting mostly for constant-free equations, because if there are two different constant letters, then every constant-preserving morphism is nonperiodic.

A constant-free equation is called *balanced* if every variable appears on the left-hand side exactly the same number of times as on the right-hand side.

Usually we assume that the alphabet of constants is $\Sigma = \{a, b\}$. The case of a unary alphabet is not interesting, and if there are more than two constant letters, they can be encoded using a binary alphabet. For example, if the letters are $a_1, \ldots, a_n$, we can use the encoding $a_i \mapsto a^i b$. This preserves the property of being a solution and the property of not being a solution.

In this article, we are specifically interested in equations with constants on one variable x, and in constant-free equations on three variables x, y, z. We use the notation $[u, v, w]$ for the morphism $h : \{x, y, z\}^* \to \Sigma^*$ defined by $(h(x), h(y), h(z)) = (u, v, w)$, and the notation $[u]$ for the constant-preserving morphism $h : (\{x\} \cup \Sigma)^* \to \Sigma^*$ defined by $h(x) = u$. If U is a set of words, we use the notation $[U] = \{[u] \mid u \in U\}$.

Example 1. *The equation (xab, bax) has infinitely many solutions $[(ab)^i]$, where $i \geq 0$. The equation $(xaxbab, abaxbx)$ has exactly two solutions, $[\varepsilon]$ and $[ab]$. The equation $(xxbaaba, aabaxbx)$ has exactly two solutions, $[a]$ and $[aaba]$. The constant-free equation (xyz, zyx) has solutions $[(pq)^i p, (qp)^j q, (pq)^k p]$, where $p, q \in \Sigma^*$ and $i, j, k \geq 0$. It has no other nonperiodic solutions.*

A set of equations is called a *system of equations*. A system $\{E_1, \ldots, E_N\}$ is often written without the braces as $E_1, \ldots, E_N$. A morphism is a solution of this system if it is a solution of every E_i.

The set of all solutions of an equation E is denoted by $\mathrm{Sol}(E)$ and the set of all solutions of a system of equations $E_1, \ldots, E_N$ by $\mathrm{Sol}(E_1, \ldots, E_N)$. Two equations or systems are *equivalent* if they have exactly the same solutions.

The set of all equations satisfied by a solution h is denoted by $\mathrm{Eq}(h)$. Two solutions h_1 and h_2 are *equivalent* if $\mathrm{Eq}(h_1) = \mathrm{Eq}(h_2)$.

A system of equations $E_1, \ldots, E_N$ is *independent* if it is not equivalent to any of its proper subsystems. Another equivalent definition would be that $E_1, \ldots, E_N$ is independent if there are solutions $h_1, \ldots, h_N$ such that $h_i \in \mathrm{Sol}(E_j)$ if and only if $i \neq j$. The system is *satisfiable* if it has a nonperiodic solution h_{N+1} (systems of constant-free equations always have periodic solutions, so it makes sense to exclude them). If the system is both satisfiable and independent, and if $h_1, \ldots, h_{N+1}$ are as above, then the sequence $(h_1, \ldots, h_{N+1})$ is called its *certificate*. (A system is a set, so the order of the equations is not formally specified, but whenever talking about certificates, it is to be understood that the order of the solutions corresponds to the order in which the equations have been written.)

A sequence of equations $E_1, \ldots, E_N$ is a *decreasing chain* if the systems

$E_1, \ldots, E_{i-1}$ and $E_1, \ldots, E_i$ are nonequivalent for all $i \in \{1, \ldots, N\}$ (the case $i = 1$ means that E_1 cannot be equivalent to the empty system, that is, E_1 cannot be a trivial equation (u, u)). Another equivalent definition would be that $E_1, \ldots, E_N$ is a decreasing chain if there are solutions $h_1, \ldots, h_N$ such that $h_i \in \mathrm{Sol}(E_j)$ whenever $j < i$ but not when $i = j$. The sequence is *satisfiable* if the equations have a common nonperiodic solution h_{N+1}. If the sequence is both satisfiable and a decreasing chain, and if $h_1, \ldots, h_{N+1}$ are as above, then the sequence $(h_1, \ldots, h_{N+1})$ is called its *certificate*.

Note that the equations of an independent system, ordered in any way, form a decreasing chain.

The above definitions can also be stated for infinite systems and sequences. However, by Ehrenfeucht's compactness property, every system of word equations is equivalent to a finite subsystem, and every decreasing chain is finite. We consider only finite systems in this article.

Example 2. *The pair of constant-free equations* $(xyz, zyx), (xyyz, zyyx)$ *is satisfiable and independent. It has a certificate* $([a, b, abba], [a, b, aba], [a, b, a])$.

The sequence of constant-free equations

$$(xyz, zxy), (xyxzyz, zxzyxy), (xz, zx), (z, \varepsilon)$$

is a satisfiable decreasing chain. It has a certificate

$$([a, \varepsilon, b], [a, b, abab], [a, b, ab], [\varepsilon, \varepsilon, a], [a, b, \varepsilon]).$$

The *length* of an equation $E = (u, v)$ is $|uv|$ and it is denoted by $|E|$. If h is a morphism, we use the notation $h(E) = (h(u), h(v))$. The equation E is *reduced* if u and v do not have a common nonempty prefix or suffix. We can always replace an equation with an equivalent reduced equation by canceling the common prefixes and suffixes.

3. Main questions

The following question is one of the biggest open problems on word equations: How large can a satisfiable independent system of constant-free equations on three variables be? The largest known examples are of size two (see Example 2), and it was conjectured by Culik and Karhumäki that these examples are optimal. Even the following weaker conjecture is open:

Conjecture 3. *There exists a number c such that every satisfiable independent system of constant-free equations on three variables is of size c or less.*

We refer to this conjecture as SIND-XYZ, or as SIND-XYZ(c) for a specific value of c. Currently, the best known result is the following.

Theorem 4 (Holub and Žemlička [8]) *Every satisfiable independent system of constant-free equations on three variables is of size $O(n)$, where n is the length of the shortest equation.*

A similar question can be asked for decreasing chains: How long can a satisfiable decreasing chain of constant-free equations on three variables be? The longest known examples are of length four (see Example 2; actually, this example is probably new). The following conjecture is open:

Conjecture 5. *There exists a number c such that every satisfiable decreasing chain of constant-free equations on three variables is of size c or less.*

We refer to this conjecture as SCHA-XYZ, or as SCHA-XYZ(c) for a specific value of c.

Another well-known open problem is the following: How many solutions can a one-variable equation have if it has only finitely many solutions? The best previously known examples have two solutions, and it has been conjectured that these examples are optimal. Even the following weaker conjecture is open:

Conjecture 6. *There exists a number c such that every one-variable equation has either infinitely many solutions or at most c.*

We refer to this conjecture as SOL-XAB, or as SOL-XAB(c) for a specific value of c. We give a counterexample to Conjecture SOL-XAB(2), but Conjecture SOL-XAB(3) remains open. Currently, the best known result is the following.

Theorem 7 (Laine and Plandowski [13]) *If the solution set of a nontrivial one-variable equation is finite, it has size at most $8 \log n + O(1)$, where n is the number of occurrences of the variable. If it is infinite, there are words p, q such that pq is primitive and the solution set is $[(pq)^* p]$.*

As an intermediate question, we can state the following problem: How large can a satisfiable decreasing chain of one-variable equations be? We are not aware of any previous research on this specific problem. We can give the following conjecture:

Conjecture 8. *There exists a number c such that every satisfiable decreasing chain of one-variable equations is of size c or less.*

We refer to this conjecture as SCHA-XAB, or as SCHA-XAB(c) for a specific value of c.

We prove the following implications between the conjectures:

$$\text{SOL-XAB} \implies \text{SCHA-XAB} \iff \text{SCHA-XYZ} \implies \text{SIND-XYZ},$$

or more specifically,

$$\text{SOL-XAB}(c) \implies \text{SCHA-XAB}(c) \begin{cases} \impliedby \text{SCHA-XYZ}(c) \implies \text{SIND-XYZ}(c) \\ \implies \text{SCHA-XYZ}(5c+5) \\ \implies \text{SIND-XYZ}(5c+2). \end{cases}$$

Actually, the implication SCHA-XYZ(c) $\implies$ SIND-XYZ(c) follows directly from the definitions. We also turn Theorem 7 into a result on constant-free equations on three variables.

4. One-variable equations with constants

In this section, we give a family of counterexamples to Conjecture SOL-XAB(2) and prove that Conjectures SCHA-XYZ and SOL-XAB imply Conjecture SCHA-XAB.

Lemma 9. *Let $w \in \{x, a, b\}^*$ be an arbitrary word and h_0 an arbitrary constant-preserving morphism. Let h_1, h_2 be the constant-preserving morphisms defined by*

$$h_1(x) = h_0(xwx), \;\; h_2(x) = h_1(xwx).$$

Then the morphisms h_0, h_1, h_2 are solutions of the equation

$$(xwxh_1(wx)wh_2(x), h_2(x)wh_1(xw)xwx).$$

Moreover, if $h_0(wx)h_1(wx) \neq h_1(wx)h_0(wx)$, then the equation has only finitely many solutions.

Proof. The morphism h_0 is a solution because

$$\begin{aligned} h_0(xwxh_1(wx)wh_2(x)) &= h_0(xwx)h_1(wx)h_0(w)h_2(x) = h_1(xwx)h_0(w)h_1(xwx) \\ &= h_2(x)h_0(w)h_1(xw)h_0(xwx) = h_0(h_2(x)wh_1(xw)xwx). \end{aligned}$$

The morphism h_1 is a solution because

$$\begin{aligned} h_1(xwxh_1(wx)wh_2(x)) &= h_1(xwx)h_1(wx)h_1(w)h_2(x) = h_2(x)h_1(wxw)h_2(x) \\ &= h_2(x)h_1(w)h_1(xw)h_1(xwx) = h_1(h_2(x)wh_1(xw)xwx). \end{aligned}$$

The morphism h_2 is a solution because

$$\begin{aligned} h_2(xwxh_1(wx)wh_2(x)) &= h_2(xwx)h_1(wx)h_2(w)h_2(x) = h_2(xw)h_1(xwxwx)h_2(wx) \\ &= h_2(x)h_2(w)h_1(xw)h_2(xwx) = h_2(h_2(x)wh_1(xw)xwx). \end{aligned}$$

If the equation has infinitely many solutions, then $h_0, h_1, h_2 \in [(pq)^*p]$ for some word pq by Theorem 7. This means that $h_0(wx), h_1(wx) \in (qp)^+$ and thus $h_0(wx)h_1(wx) = h_1(wx)h_0(wx)$. □

Example 10. *Lemma 9 with $w = axb$ and $h_0(x) = \varepsilon$ gives the equation*

$$(xaxbxaabbabaxbabaabbab, abaabbabaxbabaabbxaxbx)$$

with the solutions $[\varepsilon], [ab], [abaabbab]$. It is easy to check that the equation does not have any other solutions. Similarly, letting $w = xb$ and $h_0(x) = a$ would give an equation with three nonempty solutions.

We need the following lemma.

Lemma 11 (Eyono Obono, Goralčík and Maksimenko [5]) *Let E be a one-variable equation and let pq be primitive. The set*

$$\mathrm{Sol}(E) \cap [(pq)^+p]$$

is either $[(pq)^+p]$ or has at most one element.

Lemma 12. *Let $E_1, \dots, E_N$ be a satisfiable decreasing chain of one-variable equations. For all i, E_i has at least $N+1-i$ solutions. If $N \geq 4$, then each one of $E_1, \dots, E_{N-2}$ has only finitely many solutions.*

Proof. If $(h_1, \dots, h_{N+1})$ is a certificate of the chain, then E_i has solutions h_j for all $j > i$. Thus it has at least $N+1-i$ solutions.

Let $N \geq 4$ and let E_i have infinitely many solutions for some $i \leq N-2$. By Theorem 7, $\mathrm{Sol}(E_i) = [(pq)^* p]$ for a primitive word pq. Because $N \geq 4$, there exists j such that $i \neq j \leq N-2$. Then $\mathrm{Sol}(E_i, E_j) = \mathrm{Sol}(E_j) \cap [(pq)^* p]$ contains at least three solutions h_{N-1}, h_N, h_{N+1}, so $\mathrm{Sol}(E_j) \cap [(pq)^+ p]$ contains at least two solutions. By Lemma 11, $\mathrm{Sol}(E_j) \cap [(pq)^+ p] = [(pq)^+ p]$, and by Theorem 7, $\mathrm{Sol}(E_j) = [(pq)^* p]$, so E_i and E_j are equivalent, which is a contradiction. □

Theorem 13. *Every satisfiable decreasing chain of one-variable equations is of size at most $8 \log n + O(1)$, where n is the length of the first equation.*

Proof. Let $E_1, \dots, E_N$ be a satisfiable decreasing chain of one-variable equations and let $N \geq 4$. By Lemma 12, E_1 has at least N solutions but only finitely many. By Theorem 7, $N \leq 8 \log n + O(1)$, where n is the length of the first equation. □

Theorem 14. *Conjecture SOL-XAB(c) implies Conjecture SCHA-XAB(c).*

Proof. We assume that Conjecture SCHA-XAB(c) is false and prove that also Conjecture SOL-XAB(c) is false. We already know that Conjecture SOL-XAB(2) is false, so let $c \geq 3$. There exists a satisfiable decreasing chain of one-variable equations $E_1, \dots, E_{c+1}$. By Lemma 12, E_1 has at least $c+1$ solutions but only finitely many. This is a counterexample to Conjecture SOL-XAB(c). □

Lemma 15. *Let $\Sigma = \{a_1, \dots, a_k\}$ be the alphabet of constants and*

$$\alpha : (\{x\} \cup \Sigma)^* \to \{x, y, z\}^*, \ \alpha(x) = x, \ \alpha(a_i) = y^i z$$

be a morphism. If $E_1, \dots, E_N$ is a satisfiable decreasing chain of equations on $\{x\}$, then $\alpha(E_1), \dots, \alpha(E_N)$ is a satisfiable decreasing chain of constant-free equations on $\{x, y, z\}$.

Proof. Let

$$\beta : \Sigma^* \to \{a, b\}^*, \ \beta(a_i) = a^i b$$

be a morphism. A constant-preserving morphism $h : (\{x\} \cup \Sigma)^* \to \Sigma^*$ is a solution of E_i if and only if the nonperiodic morphism

$$g_h : \{x, y, z\}^* \to \{a, b\}^*, \ g_h(x) = \beta(h(x)), \ g_h(y) = a, \ g_h(z) = b$$

is a solution of $\alpha(E_i)$ (this follows from the fact that $g_h \circ \alpha = \beta \circ h$ and the injectivity of β). So if $(h_1, \dots, h_{N+1})$ is a certificate for $E_1, \dots, E_N$, then $(g_{h_1}, \dots, g_{h_{N+1}})$ is a certificate for $\alpha(E_1), \dots, \alpha(E_N)$. □

Theorem 16. *Conjecture SCHA-XYZ(c) implies Conjecture SCHA-XAB(c).*

Proof. Follows from Lemma 15. □

5. Classification of solutions

We are interested in satisfiable independent systems and satisfiable decreasing chains and their certificates. Recall that morphisms g, h are called equivalent if $\mathrm{Eq}(g) = \mathrm{Eq}(h)$. Every morphism in a certificate can be replaced by an equivalent morphism, so it would be beneficial for us if there were a simple subclass of morphisms containing a representative of every equivalence class. In the three-variable case, this kind of a result follows from a characterization of three-generator subsemigroups of a free semigroup by Budkina and Markov [2], or alternatively from a similar result by Spehner [16, 17]. A comparison of these two results can be found in [7]. The result we present here in Theorem 17 is a simplified version that is perhaps slightly weaker, but it is sufficiently strong for our purposes and easier to work with.

We define classes of morphisms $\{x, y, z\}^* \to \{a, b, c\}^*$:

$$\begin{aligned}
&\mathcal{A} = \{[a, b, c]\},\\
&\mathcal{B} = \{[a^i, a^j, a^k] \mid i, j, k \geq 0\},\\
&\mathcal{C}_{xyz}(i, j) = \{[a, a^i b a^j, w] \mid w \in \{a, b\}^* \wedge (i = 0 \vee w \in b\{a, b\}^*)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \wedge (j = 0 \vee w \in \{a, b\}^* b)\},\\
&\mathcal{C}_{xyz} = \bigcup_{i,j \geq 0} \mathcal{C}_{xyz}(i, j),\\
&\mathcal{D}_{xyz}(i, j, k, l, m, p, q) = \{[a, a^i b (a^m b)^p a^j, a^k b (a^m b)^q a^l]\},\\
&\mathcal{D}_{xyz} = \bigcup \mathcal{D}_{xyz}(i, j, k, l, m, p, q),
\end{aligned}$$

where the last union is taken over all $i, j, k, l, m \geq 0$ and $p, q \geq 1$ such that $ik = jl = 0$ and $\gcd(p + 1, q + 1) = 1$. If (X, Y, Z) is a permutation of (x, y, z), then $\mathcal{C}_{XYZ}(i, j)$, $\mathcal{C}_{XYZ}$, $\mathcal{D}_{XYZ}(i, j, k, l, m, p, q)$ and $\mathcal{D}_{XYZ}$ are defined similarly, with the images of the variables permuted in a corresponding way. For example, in the case of $\mathcal{C}_{XYZ}(i, j)$, X maps to a, Y to $a^i b a^j$, and Z to w. Then we also define

$$\begin{aligned}
&\mathcal{C} = \mathcal{C}_{xyz} \cup \mathcal{C}_{yzx} \cup \mathcal{C}_{zxy} \cup \mathcal{C}_{zyx} \cup \mathcal{C}_{xzy} \cup \mathcal{C}_{yxz},\\
&\mathcal{D} = \mathcal{D}_{xyz} \cup \mathcal{D}_{yzx} \cup \mathcal{D}_{zxy}.
\end{aligned}$$

For $\mathcal{A}$ and $\mathcal{B}$, we do not need to consider different permutations of the variables because the images of the variables are symmetric. For $\mathcal{D}$, we need only three of the six permutations, because the images of the latter two variables are symmetric.

Theorem 17. *Every morphism $\{x, y, z\}^* \to \{a, b, c\}^*$ is equivalent to a morphism in $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D}$.*

Proof. Follows from the characterization of Budkina and Markov [2], or alternatively from the characterization of Spehner [17]. □

Example 18. *The nonperiodic solutions of the equation* (xyz, zyx) *are of the form* $[(pq)^i p, q(pq)^j, (pq)^k p]$. *For example, the equation has the following solutions:*

- $[a, b, (ab)^k a] \in \mathcal{C}_{xyz}(0, 0)$ *and* $[b, a, (ba)^k b] \in \mathcal{C}_{yxz}(0, 0)$ *(these are equivalent),*
- $[(ab)^i a, b, a] \in \mathcal{C}_{zyx}(0, 0)$ *and* $[(ba)^i b, a, b] \in \mathcal{C}_{yzx}(0, 0)$ *(these are equivalent),*
- $[a, b(ab)^j, aba] \in \mathcal{C}_{xzy}(1, 1)$,
- $[aba, b(ab)^j, a] \in \mathcal{C}_{zxy}(1, 1)$,
- $[a, b(ab)^j, (ab)^k a] \in \mathcal{D}_{xyz}(0, 0, 1, 1, 1, j, k-1)$, *where* $j, k-1 \geq 1$ *and* $\gcd(j+1, k) = 1$,
- $[(ab)^i a, b(ab)^j, a] \in \mathcal{D}_{zxy}(1, 1, 0, 0, 1, i-1, j)$, *where* $j, i-1 \geq 1$ *and* $\gcd(j+1, i) = 1$,
- $[(ba)^i b, a, (ba)^k b] \in \mathcal{D}_{yzx}(1, 1, 1, 1, 1, k, i)$, *where* $i, k \geq 1$ *and* $\gcd(i+1, k+1) = 1$.

Lemma 19. *If* E *is a nontrivial constant-free equation on* $\{x, y, z\}$ *and* $i, j \geq 0$, *then* $\mathrm{Sol}(E) \cap \mathcal{C}_{xyz}(i, j) \neq \mathcal{C}_{xyz}(i, j)$.

Proof. Let

$$\alpha : \{x, y, z\}^* \to \{a, b, x\}^*,\ \alpha(x) = a,\ \alpha(y) = a^i b a^j,\ \alpha(z) = x$$

be a morphism. Because α is injective, $\alpha(E)$ is a nontrivial equation on $\{x\}$. By Theorem 7, there exists $u \in b\{a, b\}^* b$ such that $[u]$ is not a solution of $\alpha(E)$. Then $[u] \circ \alpha$ is not a solution of E, but $[u] \circ \alpha \in \mathcal{C}_{xyz}(i, j)$. □

By the following lemma, we can concentrate on solutions in classes $\mathcal{C}$ and $\mathcal{D}$ in our considerations.

Lemma 20. *A satisfiable decreasing chain of* N *constant-free equations on* $\{x, y, z\}$ *has a certificate in* $(\mathcal{C} \cup \mathcal{D})^{N-1} \times (\mathcal{B} \cup \mathcal{C} \cup \mathcal{D}) \times (\mathcal{C} \cup \mathcal{D})$.

A satisfiable independent system of N *constant-free equations on* $\{x, y, z\}$ *has a certificate in* $(\mathcal{C} \cup \mathcal{D})^{N+1}$.

Proof. First, let $E_1, \ldots, E_N$ be a satisfiable decreasing chain and $(h_1, \ldots, h_{N+1})$ its certificate. Every solution in a certificate can be replaced by an equivalent solution, so we can assume that $(h_1, \ldots, h_{N+1}) \in (\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D})^{N+1}$ by Theorem 17.

The morphism in $\mathcal{A}$ is a solution of only the trivial equations (u, u), and these equations cannot be part of any decreasing chain, so none of $h_2, \ldots, h_{N+1}$ can be in $\mathcal{A}$. The only requirement for h_1 is that it must not be a solution of E_1, so we can assume that $h_1 \in \mathcal{C}$ by Lemma 19.

It follows from Proposition 29 in [7] that $E_1, \ldots, E_{N-1}$ are balanced. Every morphism in $\mathcal{B}$ is periodic and thus a solution of every balanced equation, so $h_1, \ldots, h_{N-1} \notin \mathcal{B}$, and $h_{N+1} \notin \mathcal{B}$ by the definition of a certificate.

The case of a satisfiable independent system can be proved in a similar way, except that then also E_N must be balanced and thus $h_N \notin \mathcal{B}$. □

Example 21. *If $(h_1, \ldots, h_5)$ is a certificate of the satisfiable decreasing chain*

$$(xyz, zxy), (xyxzyz, zxzyxy), (xz, zx), (z, \varepsilon)$$

of Example 2, then h_4 must be periodic. This shows that allowing the second-to-last element of a certificate to be in $\mathcal{B}$ is sometimes necessary.

6. Class $\mathcal{C}$

In this section, we study morphisms in class $\mathcal{C}$. This leads to a natural connection between three-variable constant-free equations and one-variable equations with constants.

Lemma 22. *Let E be a nontrivial constant-free equation on $\{x, y, z\}$. There is at most one pair (i, j) such that E has a solution in $\mathcal{C}_{xyz}(i, j)$. For this pair, $i + j \leq |E| - 1$.*

Proof. Let $E = (u, v)$ and $h \in \mathrm{Sol}(E) \cap \mathcal{C}_{xyz}(i, j)$. We can assume that E is reduced. By swapping u and v if necessary, we can assume that one of the following is true:

(1) $v = \varepsilon$.
(2) $u = x^k$, $k \geq 1$, and v begins with y.
(3) u begins with $x^k y$, $k \geq 1$, and v begins with y.
(4) u begins with $x^k z$, $k \geq 1$, and v begins with y.
(5) u begins with x and v begins with z.
(6) u begins with y and v begins with z.

In all of these cases, we get either a contradiction or a single possible value for i as follows:

(1) $u \neq \varepsilon$, so at least one of $h(x), h(y), h(z)$ is ε. The only possibility is $h(z) = \varepsilon$, and then $i = j = 0$.
(2) $h(u) = a^k$ and $h(v)$ contains the letter b, which is a contradiction.
(3) $h(u)$ begins with $a^{k+i}b$ and $h(v)$ begins with $a^i b$, which is a contradiction.
(4) $h(y)$ must begin with a and thus $h(z)$ must begin with b, so $h(u)$ begins with $a^k b$ and $h(v)$ begins with $a^i b$. Thus $i = k$.
(5) $h(z)$ cannot begin with b and thus $h(y)$ must begin with b, so $i = 0$.
(6) It is not possible that $h(y)$ would begin with a and $h(z)$ with b, so $h(y)$ must begin with b and $i = 0$.

By looking at the suffixes of u and v, we similarly see that j is uniquely determined. Moreover, $i + j \leq |E| - 1$. $\square$

Let $E_1, \dots, E_N$ be a satisfiable decreasing chain of constant-free equations on $\{x, y, z\}$ with a certificate $(h_1, \dots, h_{N+1})$. If $h_2, \dots, h_{N+1} \in \mathcal{C}_{xyz}$, then there is a satisfiable decreasing chain $E'_1, \dots, E'_N$ of one-variable equations such that $|E'_n| \leq |E_n|^2$ for all n.

Proof. Let i, j be such that $h_{N+1} \in \mathcal{C}_{xyz}(i, j)$. By Lemma 22, $h_2, \dots, h_N \in \mathcal{C}_{xyz}(i, j)$. By Lemma 19, we can assume that $h_1 \in \mathcal{C}_{xyz}(i, j)$. Let

$$\alpha : \{x, y, z\}^* \to \{a, b, z\}^*, \ \alpha(x) = a, \ \alpha(y) = a^i b a^j, \ \alpha(z) = z$$

be a morphism and let

$$h'_n : \{a, b, z\}^* \to \{a, b\}^*, \ h'_n(z) = h_n(z)$$

be a constant-preserving morphism. For every n, $h_n = h'_n \circ \alpha$ and $\alpha(E_n)$ is a one-variable equation with constants. Then $(h'_1, \dots, h'_{N+1})$ is a certificate of the decreasing chain $\alpha(E_1), \dots, \alpha(E_N)$. The length of $\alpha(E_n)$ is at most $(i + j + 1)|E_n|$, which is at most $|E_n|^2$ by Lemma 22. $\square$

7. Class $\mathcal{D}$

In this section, we study morphisms in class $\mathcal{D}$. This class looks more complicated than class $\mathcal{C}$, but actually there is a lot of structure in the morphisms in $\mathcal{D}$, which allows us to prove stronger results than for $\mathcal{C}$.

Lemma 24. *Let E be a nontrivial constant-free equation on $\{x, y, z\}$. There are i, j, k, l, m, p', q' such that $\mathrm{Sol}(E) \cap \mathcal{D}_{xyz}$ is one of the following sets:*

(1) $\varnothing$,
(2) $\mathcal{D}_{xyz}(i, j, k, l, m, p', q')$,
(3) $\displaystyle\bigcup_{\substack{p,q \geq 1 \\ \gcd(p+1,q+1)=1}} \mathcal{D}_{xyz}(i, j, k, l, m, p, q)$.

Proof. Let $E = (u, v)$. If $u = \varepsilon$ or $v = \varepsilon$, then $\mathrm{Sol}(E) \cap \mathcal{D}_{xyz} = \varnothing$, so let $u \neq \varepsilon \neq v$. We can assume that E is reduced and write it as

$$(x^{a_0} y_1 x^{a_1} \cdots y_r x^{a_r}, x^{b_0} z_1 x^{b_1} \cdots z_s x^{b_s}),$$

where $y_1, \dots, y_r, z_1, \dots, z_s \in \{y, z\}$. We can also assume that $r, s \geq 2$ by replacing (u, v) with the equivalent equation $(uyyu, vyyv)$ if necessary. Let $h \in \mathrm{Sol}(E) \cap \mathcal{D}_{xyz}$ and

$$h(x) = a, \ h(y_t) = a^{i_t} b (a^m b)^{p_t} a^{j_t}, \ h(z_t) = a^{k_t} b (a^m b)^{q_t} a^{l_t},$$

where

$$(i_t, j_t, p_t) = \begin{cases} (i, j, p) & \text{if } y_t = y, \\ (k, l, q) & \text{if } y_t = z, \end{cases} \qquad (k_t, l_t, q_t) = \begin{cases} (i, j, p) & \text{if } z_t = y, \\ (k, l, q) & \text{if } z_t = z. \end{cases}$$

The left-hand side $h(u)$ begins with $a^{a_0+i_1}b$ and the right-hand side $h(v)$ begins with $a^{b_0+k_1}b$, so $a_0 + i_1 = b_0 + k_1$. If $y_1 = z_1$, then $i_1 = k_1$, $a_0 = b_0$, and E is not reduced, a contradiction. Thus $y_1 \neq z_1$ and $i_1 k_1 = ik = 0$. From $a_0 + i_1 = b_0 + k_1$, $i_1 k_1 = 0$, $a_0 b_0 = 0$ it then follows that $k_1 = a_0$ and $i_1 = b_0$. Similarly, by looking at the suffixes of $h(u)$ and $h(v)$ we find out that $y_r \neq z_s$, $l_s = a_r$, and $j_r = b_s$. Thus i, j, k, l are uniquely determined by the equation E.

It must be $\{p_1, q_1\} = \{p, q\}$, and $\gcd(p + 1, q + 1) = 1$, so $p_1 \neq q_1$. If $p_1 < q_1$, then $h(u)$ and $h(v)$ begin with

$$a^{a_0+i_1} b(a^m b)^{p_1} a^{j_1+a_1+i_2} b \qquad \text{and} \qquad a^{b_0+k_1} b(a^m b)^{p_1+1},$$

respectively, so $j_1 + a_1 + i_2 = m$. Similarly, if $p_1 > q_1$, then $l_1 + b_1 + k_2 = m$. Thus

$$m \in \{j_1 + a_1 + i_2, l_1 + b_1 + k_2\}.$$

If $j_1 + a_1 + i_2 = m \neq l_1 + b_1 + k_2$, then there are $n \neq m$, $A \geq 1$, $B \geq 0$ such that $h(u)$ and $h(v)$ begin with

$$a^{a_0+i_1} b(a^m b)^{A(p_1+1)+B(q_1+1)-1} a^n b \qquad \text{and} \qquad a^{b_0+k_1} b(a^m b)^{q_1} a^{l_1+b_1+k_2} b,$$

respectively. It must be $A(p_1 + 1) + B(q_1 + 1) = q_1 + 1$. But then $B > 0$ would be a contradiction, and $B = 0$ would contradict $\gcd(p + 1, q + 1) = 1$. Similarly, $j_1 + a_1 + i_2 \neq m = l_1 + b_1 + k_2$ would lead to a contradiction. Thus it must be $j_1 + a_1 + i_2 = m = l_1 + b_1 + k_2$.

We can write

$$h(u) = a^{c_0} b(a^m b)^{A_1(p+1)+C_1(q+1)-1} a^{c_1} b \cdots b(a^m b)^{A_R(p+1)+C_R(q+1)-1} a^{c_R},$$
$$h(v) = a^{d_0} b(a^m b)^{B_1(p+1)+D_1(q+1)-1} a^{d_1} b \cdots b(a^m b)^{B_S(p+1)+D_S(q+1)-1} a^{d_S},$$

where $c_1, \ldots, c_{R-1}, d_1, \ldots, d_{S-1} \neq m$. It must be $R = S$, $c_t = d_t$, and

$$A_t(p + 1) + C_t(q + 1) = B_t(p + 1) + D_t(q + 1)$$

for all t. Moreover, all values p, q that satisfy these linear relations lead to a solution of the equation. If two of the vectors $(A_t - B_t, C_t - D_t)$ are linearly independent, then there is no solution, if all of them are zero, then all pairs p, q are solutions, and otherwise there is exactly one pair p, q with $\gcd(p + 1, q + 1) = 1$ that is a solution. This concludes the proof. □

The next lemma is a special case of Theorem 5.3 in [15]. Here, the *length type* of a solution h is the vector $(|h(x)|, |h(y)|, |h(z)|)$.

Lemma 25 (Saarela [15]) *Let E_1, E_2 be constant-free equations on three variables. If there exists a nonperiodic solution $h \in \mathrm{Sol}(E_1) \setminus \mathrm{Sol}(E_2)$, then the length types of nonperiodic solutions of the pair E_1, E_2 are covered by a finite union of two-dimensional subspaces of $\mathbb{Q}^3$.*

Let E_1, E_2, E_3 be a satisfiable decreasing chain of constant-free equations on $\{x, y, z\}$ with a certificate (h_1, h_2, h_3, h_4). At most one of h_3, h_4 can be in $\mathcal{D}_{xyz}$.

Proof. Let $h_3, h_4 \in \mathcal{D}_{xyz}$. Then $h_3, h_4 \in \mathrm{Sol}(E_1, E_2) \cap \mathcal{D}_{xyz}$, so the third option of Lemma 24 must be true for this set. We show that the length types of solutions of the pair E_1, E_2 cannot be covered by finitely many two-dimensional spaces, which contradicts Lemma 25.

The length type of $[a, a^i b(a^m b)^p a^j, a^k b(a^m b)^q a^l] \in \mathrm{Sol}(E_1, E_2) \cap \mathcal{D}_{xyz}$ is

$$(1, i + 1 + (m + 1)p + j, k + 1 + (m + 1)q + l).$$

Here i, j, k, l, m are fixed, but p, q can be arbitrary positive integers such that $\gcd(p+1, q+1) = 1$. For every p, there are infinitely many possible values of q, giving infinitely many length types on the line

$$L_p = \{(1, i + 1 + (m + 1)p + j, Z) \mid Z \in \mathbb{Q}\}.$$

The only way to cover these with a finite number of two-dimensional spaces is to have one of them be the unique two-dimensional space containing the whole line. This is true for any p, and different values of p give different two-dimensional spaces, so all length types cannot be covered by finitely many two-dimensional spaces. $\square$

8. Main results

Putting our results together gives the following theorem, which improves the linear bound of Theorem 4 to a logarithmic one.

Theorem 27. *A satisfiable decreasing chain of constant-free equations on $\{x, y, z\}$ has at most $O(\log n)$ equations, where n is the length of the first equation.*

A satisfiable independent system of constant-free equations on $\{x, y, z\}$ has at most $O(\log n)$ equations, where n is the length of the shortest equation.

Proof. By Lemma 20, a satisfiable decreasing chain $E_1, \ldots, E_N$ has a certificate $(h_1, \ldots, h_{N+1})$, where $h_1, \ldots, h_{N-1}, h_{N+1} \in (\mathcal{C} \cup \mathcal{D})$ and $h_N \in (\mathcal{B} \cup \mathcal{C} \cup \mathcal{D})$. By Lemma 26, at most one of the solutions $h_3, \ldots, h_{N+1}$ can be in $\mathcal{D}_{xyz}$, and the same is true for $\mathcal{D}_{yzx}$ and $\mathcal{D}_{zxy}$. Thus at most three of the solutions $h_3, \ldots, h_{N+1}$ can be in $\mathcal{D}$. Let k of $h_2, \ldots, h_{N+1}$ be in $\mathcal{C}_{xyz}$. We get a satisfiable decreasing chain of length k, for which we can use Lemma 23, and then Theorem 13 to conclude that $k = O(\log n)$. Similarly, we can prove that the number of i such that $h_i \in \mathcal{C}_{XYZ}$ is $O(\log n)$ for all permutations (X, Y, Z) of (x, y, z).

The claim about satisfiable independent systems follows because the equations of an independent system, ordered in any way, form a decreasing chain. $\square$

By Theorem 14, Conjecture SCHA-XAB could be replaced by Conjecture SOL-XAB in the next theorem. The constants are probably not optimal.

Theorem 28. *Conjecture SCHA-XAB(c) implies Conjectures SCHA-XYZ($5c+5$) and SIND-XYZ($5c+2$). In particular, if SCHA-XAB(3) is true, then a satisfiable independent system of constant-free equations on $\{x, y, z\}$ has at most 17 equations.*

Proof. First, let $E_1, \ldots, E_N$ be a satisfiable decreasing chain of reduced constant-free equations on $\{x, y, z\}$ with a certificate $(h_1, \ldots, h_{N+1})$. By Lemma 20, we can assume that $h_n \in \mathcal{C} \cup \mathcal{D}$ for all $n \neq N$. If $E_1 = (u, v)$, then at least one of the variables appears both at the beginning of u or v and at the end of u or v. We can assume it is z. Because $\mathcal{C}_{xyz}(0, 0)$ and $\mathcal{C}_{yxz}(0, 0)$ are the same up to swapping a and b, we can assume that $h_n \notin \mathcal{C}_{yxz}(0, 0)$ for all n. It follows from the proof of Lemma 22 that $\mathrm{Sol}(E_1) \cap \mathcal{C}_{yxz} \subseteq \mathcal{C}_{yxz}(0, 0)$, so $h_n \notin \mathcal{C}_{yxz}$ for all $n \geq 2$.

By Lemma 23 and the assumption about Conjecture SCHA-XAB, at most c of the solutions $h_2, \ldots, h_{N+1}$ can be in $\mathcal{C}_{xyz}$, and the same is true for $\mathcal{C}_{yzx}$, $\mathcal{C}_{zxy}$, $\mathcal{C}_{zyx}$, and $\mathcal{C}_{xzy}$. It was shown above that none of the solutions is in $\mathcal{C}_{yxz}$. Thus at most $5c$ of the solutions $h_2, \ldots, h_{N+1}$ can be in $\mathcal{C}$. In addition, h_1 might be in $\mathcal{C}$.

By Lemma 26, at most one of the solutions $h_3, \ldots, h_{N+1}$ can be in $\mathcal{D}_{xyz}$, and the same is true for $\mathcal{D}_{yzx}$ and $\mathcal{D}_{zxy}$. Thus at most three of the solutions $h_3, \ldots, h_{N+1}$ can be in $\mathcal{D}$. In addition, h_1 and h_2 might be in $\mathcal{D}$.

This with the possibility $h_N \in \mathcal{B}$ proves that the total number of the solutions h_n, which is $N + 1$, cannot be more than $5c + 6$.

If $E_1, \ldots, E_N$ is a satisfiable independent system with a certificate $(h_1, \ldots, h_{N+1})$, then we can improve the above proof as follows. By Lemma 20, we can assume that $h_n \in \mathcal{C} \cup \mathcal{D}$ for all n, also for $n = N$. For an equation $E_i = (u, v)$, at least one of the variables appears both at the beginning of u or v and at the end of u or v. We can assume that the same variable z works for two of the equations, say, for E_1 and E_2. Because $\mathcal{C}_{xyz}(0, 0)$ and $\mathcal{C}_{yxz}(0, 0)$ are the same up to swapping a and b, we can assume that $h_n \notin \mathcal{C}_{yxz}(0, 0)$ for all n. It follows from the proof of Lemma 22 that $\mathrm{Sol}(E_1) \cap \mathcal{C}_{yxz} \subseteq \mathcal{C}_{yxz}(0, 0)$ and $\mathrm{Sol}(E_2) \cap \mathcal{C}_{yxz} \subseteq \mathcal{C}_{yxz}(0, 0)$. Every h_n is in $\mathrm{Sol}(E_1)$ or in $\mathrm{Sol}(E_2)$, so $h_n \notin \mathcal{C}_{yxz}$ for all n. Because the equations $E_1, \ldots, E_N$ form a satisfiable decreasing chain in any order, at most $5c$ of the solutions $h_1, \ldots, h_{N+1}$ can be in $\mathcal{C}$, and at most three of the solutions $h_1, \ldots, h_{N+1}$ can be in $\mathcal{D}$. Thus the total number of the solutions h_n cannot be more than $5c + 3$. □

9. Conclusion

We can mention several further research goals. Two obvious ones are improving the constants in Theorem 28, and proving Conjecture SOL-XAB or Conjecture SCHA-XAB (ideally SOL-XAB(3)), and thus also Conjecture SIND-XYZ.

A different topic would be to study the complexity of determining whether a three-variable constant-free equation has a nonperiodic solution. This decision problem is known to be in NP [14]. Based on the connection to one-variable equations that we have proved, a better result could probably be obtained, because one-variable equations can be solved efficiently, even in linear time in the RAM model, as proved by Jeż [10].

Finally, the question about the maximal sizes of independent systems could be studied for more than three variables. This is of course a big question, and our techniques do not help here, because they are specific to the three-variable case.

References

[1] M. H. Albert and J. Lawrence, A proof of Ehrenfeucht's conjecture, *Theoret. Comput. Sci.* **41**(1) (1985) 121–123.

[2] L. G. Budkina and A. A. Markov, F-semigroups with three generators, *Mat. Zametki* **14** (1973) 267–277.

[3] K. Culik, II and J. Karhumäki, Systems of equations over a free monoid and Ehrenfeucht's conjecture, *Discrete Math.* **43**(2–3) (1983) 139–153.

[4] R. Dąbrowski and W. Plandowski, On word equations in one variable, *Algorithmica* **60**(4) (2011) 819–828.

[5] S. Eyono Obono, P. Goralčík and M. Maksimenko, Efficient solving of the word equations in one variable, *Proceedings of the 19th MFCS*, *LNCS* **841**, (Springer, 1994), pp. 336–341.

[6] V. S. Guba, Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems, *Mat. Zametki* **40**(3) (1986) 321–324.

[7] T. Harju and D. Nowotka, On the independence of equations in three variables, *Theoret. Comput. Sci.* **307**(1) (2003) 139–172.

[8] Š. Holub and J. Žemlička, Algebraic properties of word equations, *J. Algebra* **434** (2015) 283–301.

[9] J. Honkala, On chains of word equations and test sets, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **68** (1999) 157–160.

[10] A. Jeż, One-variable word equations in linear time, *Algorithmica* **74**(1) (2016) 1–48.

[11] J. Karhumäki and W. Plandowski, On the defect effect of many identities in free semigroups, *Mathematical aspects of natural and formal languages*, ed. G. Paun (World Scientific, 1994), pp. 225–232.

[12] J. Karhumäki and A. Saarela, On maximal chains of systems of word equations, *Proc. Steklov Inst. Math.* **274** (2011) 116–123.

[13] M. Laine and W. Plandowski, Word equations with one unknown, *Internat. J. Found. Comput. Sci.* **22**(2) (2011) 345–375.

[14] A. Saarela, On the complexity of Hmelevskii's theorem and satisfiability of three unknown equations, *Proceedings of the 13th DLT*, *LNCS* **5583**, (Springer, 2009), pp. 443–453.

[15] A. Saarela, Systems of word equations, polynomials and linear algebra: A new approach, *European J. Combin.* **47** (2015) 1–14.

[16] J.-C. Spehner, Quelques problémes d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre, PhD thesis, Univ. Paris (1976).

[17] J.-C. Spehner, Les systemes entiers d'équations sur un alphabet de 3 variables, *Semigroups*, (1986), pp. 342–357.