An Analysis and a Reproof of Hmelevskii's Theorem *

(Extended Abstract)

Juhani Karhumäki and Aleksi Saarela

Department of Mathematics and Turku Centre for Computer Science TUCS, University of Turku, 20014 Turku, Finland karhumak@utu.fi, amsaar@utu.fi

Abstract. We analyze and reprove the famous theorem of Hmelevskii, which states that the general solutions of constant-free equations on three unknowns are finitely parameterizable, that is expressible by a finite collection of formulas of word and numerical parameters. The proof is written, and simplified, by using modern tools of combinatorics on words. As a new aspect the size of the finite representation is estimated; it is bounded by a double exponential function on the size of the equation.

1 Introduction

Theory of word equations is a fundamental part of combinatorics on words. It plays an essential role in a number of areas of mathematical research, such as in representation results of algebra, theory of algorithms and pattern matching. During the few last decades it has provided several challenging problems as well as fundamental, or even breakthrough, results in discrete mathematics.

Remarkable achievements of the topic are the decidability of the satisfiability problem for word equations, and the compactness result of systems of word equations, see [9] for the first and [1] and [4] for the second. The first result was reproved and sharpened to a PSPACE algorithm in [10]. For the latter one the question of bounding the size of an equivalent finite subset is still a challenge.

In the case of word equations with only three unknowns fundamental results have also been achieved. In one direction Hmelevskii [6] proved already in 1970 that any such constant-free equation is finitely parameterizable, that is the general solution can be expressed as a finite formula on word and numerical parameters. On other direction Spehner [11, 12] classified all sets of relations a given solution, that is a triple of words, can satisfy. A remarkable thing is that both of these results have only very complicated proofs. This, if any, is a splendid example of a challenging nature of word problems. Indeed, even the basic question of finding any upper bound for the maximal size of independent system of word equations on three unknowns is still open, see [5] and [3].

The goal of this paper is to analyze the proof of Hmelevskii's theorem. The result itself is, of course, very well known, see e.g. [8]. However, a compact and

* Supported by the Academy of Finland under grant 8121419

readable presentation of it seems to be lacking. We hope to fill this gap. In other words, we search for a self-contained proof using achievements and tools of combinatorics on words obtained over the last 30 years. The hope will be completed only in the full paper, but we do believe that already this presentation will give the reader justified impression of the proof. In addition, we conclude, for the first time, an upper bound for the size of the formula giving the general solution of a constant-free equation on three unknowns. Our bound is double exponential in terms of the length of the equation – and thus not likely to be even close to the optimal one.

In this extended abstract the proof is outlined in modern terms and tools of combinatorics on words, but many details are left to the final version of the full paper [7].

2 Definitions and Basic Results

In this section we fix the terminology and state the basic auxiliary results needed, for more see [2].

We consider word equations $U = V$, where $U, V \in \Xi^*$ and Ξ is the alphabet of unknowns. A morphism $h : \Xi^* \to \Sigma^*$ is a solution of this equation, if $h(U) = h(V)$. We also consider *one-sided* equations $xU \rightrightarrows yV$. A morphism $h : \Xi^* \to \Sigma^*$ is a solution of this equation, if $h(xU) = h(yV)$ and $|h(x)| \geq |h(y)|$.

A solution h is *periodic*, if there exists such $t \in \Sigma^*$ that every $h(x)$, where $x \in \Xi$, is a power of t. Otherwise h is *nonperiodic*. Periodic solutions are easy to find and represent, so in many cases it is enough to consider nonperiodic ones.

If a word u is a *prefix* of a word v, that is $v = uw$ for some w, the notation $u \leq v$ is used. If also $u \neq v$, then u is a *proper prefix*; this is denoted by $u < v$.

Let $w = a_1 \dots a_n$. Its *reverse* is $w^R = a_n \dots a_1$, and its *length* is $|w| = n$. The number of occurrences of a letter a in w is denoted by $|w|_a$.

If $\Sigma = \{a_1, \dots, a_n\}$, then $U \in \Sigma^*$ can be denoted $U(a_1, \dots, a_n)$, and its image under a morphism h can be denoted $h(U) = U(h(a_1), \dots, h(a_n))$. If $u \in \Sigma^*$, then the morphism $a_1 \mapsto u$ means the morphism, which maps $a_1 \mapsto u$ and $a_i \mapsto a_i$, when $i = 2, \dots, n$.

The following theorems and lemmas are easy to prove by using standard methods for solving equation. They give solutions to some simple equations. These solutions will be the basis of parametric solutions of all equations with three unknowns. We start with the well known lemmata, see [2].

Theorem 2.1. *Let $U, V \in \{x, y\}^*$ and $U \neq V$. Assume that $|U|_x = a$, $|U|_y = b$, $|V|_x = c$ and $|V|_y = d$. The solutions of the equation $U = V$ are $x = t^i$, $y = t^j$, where $t \in \Sigma^*$, $ai + bj = ci + dj$ and $i, j \geq 0$.*

Theorem 2.2. *The solutions of the equation $xz = zy$ are $x = pq$, $y = qp$, $z = p(qp)^i$ or $x = y = 1$, $z = p$, where $p, q \in \Sigma^*$ and $i \geq 0$.*

Lemma 2.3. *The nonperiodic solutions of the equation $xyz = zyx$ are $x = (pq)^i p$, $y = q(pq)^j$, $z = (pq)^k p$, where $p, q \in \Sigma^*$, $i, j, k \geq 0$, $pq \neq qp$ and pq can be assumed to be primitive.*

Lemma 2.4. *The nonperiodic solutions of the equation* $xyz = zxy$ *are* $x = (pq)^i p$, $y = q(pq)^j$, $z = (pq)^k$, *where* $p, q \in \Sigma^*$, $i, j, k \geq 0$ *and* $pq \neq qp$.

Lemma 2.5. *Let* $a \geq 2$. *The nonperiodic solutions of the equation* $xzx = y^a$ *are* $x = (pq)^i p$, $y = (pq)^{i+1} p$, $z = qp((pq)^{i+1}p)^{a-2}pq$, *where* $p, q \in \Sigma^*$, $i \geq 0$ *and* $pq \neq qp$.

Lemma 2.6. *Let* $a \geq 2$. *The nonperiodic solutions of the equation* $xy^a z = zy^a x$ *are* $x = (pq^a)^i p$, $y = q$, $z = (pq^a)^j p$ *or*

$$\begin{cases} x &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^i, \\ y &= (pq)^{k+1}p, \\ z &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^j, \end{cases}$$

where $p, q \in \Sigma^*$, $i, j, k \geq 0$ *and* $pq \neq qp$.

Lemma 2.7. *The nonperiodic solutions of the equation* $xyxz \rightrightarrows zx^2y$ *are* $x = (pq)^i p$, $y = qp((pq)^{i+1}p)^j pq$, $z = pq$, *where* $p, q \in \Sigma^*$, $i, j \geq 0$ *and* $pq \neq qp$.

Lemma 2.8. *Let* $a, b \geq 1$ *and* $U, V \in \Xi^*$. *If* h *is a solution of the equation* $x^a yU = y^b xV$, *then* $h(x)$ *and* $h(y)$ *commute.*

The following corollary of the graph lemma is also useful. A proof can be found in [2]. This result simplifies the original proof of Hmelevskii in several places.

Theorem 2.9. *Let* $A, B, C, D \in \{x, y, z\}^*$. *If* h *is a solution of the pair of equations* $xA = yB$, $xC = zD$, *then* h *is periodic or one of* $h(x), h(y), h(z)$ *equals* 1.

3 Parametric Words

In this section, we define the central notions of this presentation, namely parametric words, parameterizability and parametric solutions.

Fix the alphabet of *word parameters* Δ and the set of *numerical parameters* Λ. Now *parametric words* are defined inductively as follows:

(i) if $a \in \Delta \cup \{1\}$, then (a) is a parametric word,
(ii) if α and β are parametric words, then so is $(\alpha\beta)$,
(iii) if α is a parametric word and $i \in \Lambda$, then (α^i) is a parametric word,

The set of parametric words is denoted by $\mathcal{P}(\Delta, \Lambda)$. The sets of parameters are always denoted by Δ and Λ.

When there is no danger of confusion, unnecessary parenthesis can be omitted and notations like $\alpha^i\alpha^j = \alpha^{i+j}$ and $(\alpha^i)^j = \alpha^{ij}$ can be used. Then parametric words form a monoid, if the product of α and β is defined to be $\alpha\beta$.

If f is a function $\Lambda \to \mathbb{N}_0$, we can abuse the notation and use the same symbol for the function, which maps parametric words by giving values for the

numerical parameters with f: if $a \in \Delta \cup \{1\}$, then $f((a)) = a$; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then $f((\alpha\beta)) = f(\alpha)f(\beta)$; if $\alpha \in \mathcal{P}(\Delta, \Lambda)$ and $i \in \Lambda$, then $f((\alpha^i)) = f(\alpha)^{f(i)}$. A parametric word is thus mapped by f to a word of Δ^*. This can be further mapped by a morphism $h : \Delta^* \to \Sigma^*$ to a word of Σ^*. The mapping $h \circ f$ is a *valuation* of a parametric word into Σ^*, and f is its valuation to the set Δ^*.

We define the *length* of a parametric word: the length of 1 is zero; if $a \in \Delta$, then the length of a is one; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then the length of $\alpha\beta$ is the sum of lengths of α and β; if $\alpha \in \mathcal{P}(\Delta, \Lambda) \setminus \{1\}$ and $i \in \Lambda$, then the length of α^i is the length of α plus one.

Next we define the *height* of a parametric word: if $a \in \Delta \cup \{1\}$, then the height of a is zero; if $\alpha, \beta \in \mathcal{P}(\Delta, \Lambda)$, then the height of $\alpha\beta$ is the maximum of heights of α and β; if $\alpha \in \mathcal{P}(\Delta, \Lambda) \setminus \{1\}$ and $i \in \Lambda$, then the height of α^i is the height of α plus one. Parametric words of height zero can be considered to be words of Δ^*.

A *linear Diophantine relation* R is a disjunction of systems of linear Diophantine equations with lower bounds for the unknowns. For example,

$$((x + y - z = 0) \wedge (x \geq 2)) \vee ((x + y = 3) \wedge (x + z = 4))$$

is a linear Diophantine relation over the unknowns x, y, z. We are only interested in the nonnegative values of the unknowns. If $\Lambda = \{i_1, \ldots, i_k\}$, f is a function $\Lambda \to \mathbb{N}_0$ and $f(i_1), \ldots, f(i_k)$ satisfy R, then the notation $f \in R$ can be used.

Let S be a set of morphisms $\Xi^* \to \Sigma^*$, $\Lambda = \{i_1, \ldots, i_k\}$, h_j a morphism from Ξ^* to parametric words and R_j a linear Diophantine relation, when $j = 1, \ldots, m$. The set $\{(h_j, R_j) : 1 \leq j \leq m\}$ is a *parametric representation* of S, if

$$S = \{h \circ f \circ h_j : 1 \leq j \leq m, f \in R_j\},$$

where $h \circ f$ runs over all valuations to Σ^*. The linear Diophantine relations are not strictly necessary, but they make some proofs easier. A set can be *parameterized*, if it has a parametric representation. The *length* of the parametric representation is the sum of the lengths of all $h_j(x)$, where $j = 1, \ldots, m$ and $x \in \Xi$.

It follows immediately that if two sets can be parameterized, then also their union can be parameterized.

Let $S, S_1, \ldots, S_n$ be sets of morphisms $\Xi^* \to \Sigma^*$. The set S can be *parameterized in terms of the sets* $S_1, \ldots, S_n$, if there exists such morphisms $h_1, \ldots, h_n$ from Ξ^* to $\mathcal{P}(\Xi, \Lambda)$ that

$$S = \{g \circ f \circ h_j : 1 \leq j \leq n, g \in S_j\},$$

where f runs over functions $\Lambda \to \mathbb{N}_0$.

Again it is a direct consequence of the definitions that the parameterizability is preserved in compositions. Namely, if S can be parameterized in terms of the sets $S_1, \ldots, S_n$ and every S_i can be parameterized in terms of the sets $S_{i1}, \ldots, S_{in_i}$, then S can be parameterized in terms of the sets S_{ij}.

We conclude these definitions by saying that solutions of an equation can be *parameterized*, if the set of its all solutions can be parameterized. A parametric representation of this set is a *parametric solution* of the equation.

These definitions can be generalized in an obvious way for systems of equations. Theorems 2.1 and 2.2 and Lemmas 2.3 – 2.7 give parametric solutions for some equations. The following theorem states that the basic tool in solving equations, namely the cancellation of the first variable, preserves the parameterizability of solutions.

Theorem 3.1. *Let $U, V \in \Xi^*$, $x, y \in \Xi$ and $x \neq y$. Let $h : \Xi^* \to \Xi^*$ be the morphism $x \mapsto yx$. If the equation $xh(U) = h(V)$ has a parametric solution, then so does the equation $xU \rightrightarrows yV$.*

Let α and β be parametric words. The pair (α, β) can be viewed as an equation, referred to as an *exponential equation*. The *height* of this equation is the height of $\alpha\beta$. The solutions of this equation are the functions $f : \Lambda \to \mathbb{N}_0$ that satisfy $f(\alpha) = f(\beta)$.

If we know some parametric words, which give all solutions of an equation, but which also give some extra solutions, then often the right solutions can be picked by adding some constraints for the numerical parameters. These constraints can be found by exponential equations, and the following theorems prove that they are in our cases equivalent with linear Diophantine relations.

Theorem 3.2. *Let E be an exponential equation of height one. There exists a linear Diophantine relation R such that a function $f : \Lambda \to \mathbb{N}_0$ is a solution of E if and only if $f \in R$.*

In some cases Theorem 3.2 can be generalized for exponential equations of height two.

Theorem 3.3. *Let $\Lambda = \{i, j\}$ and let $s_0, \ldots, s_m$, $t_1, \ldots, t_m$, $u_0, \ldots, u_n$ and $v_1, \ldots, v_n$ be parametric words of height at most one, with no occurrences of parameter j. Assume that i occurs at least in the words $t_1, \ldots, t_m$ and $v_1, \ldots, v_n$. Let $\alpha = s_0 t_1^j s_1 \ldots t_m^j s_m$ and $\beta = u_0 v_1^j u_1 \ldots v_n^j u_n$. Now there exists a linear Diophantine relation R such that a function $f : \Lambda \to \mathbb{N}_0$ is a solution of the exponential equation $E : \alpha = \beta$ if and only if $f \in R$.*

The parametric words in the next theorem come from Lemma 2.6.

Theorem 3.4. *Let $\Delta = \{p, q\}$, $\Lambda = \{i, j, k\}$ and $a \geq 2$. Let $\alpha = (pq^a)^i p$, $\beta = q$, $\gamma = (pq^a)^j p$, or*

$$\begin{cases} \alpha &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^i, \\ \beta &= (pq)^{k+1}p, \\ \gamma &= qp((pq)^{k+1}p)^{a-2}pq(((pq)^{k+1}p)^{a-1}pq)^j. \end{cases}$$

Let $A, B \in \{x, y, z\}^*$ and let h be the morphism mapping $x \mapsto \alpha, y \mapsto \beta, z \mapsto \gamma$. Now there exists a linear Diophantine relation R such that a function $f : \Lambda \to \mathbb{N}_0$ is a solution of the exponential equation $E : h(A) = h(B)$ if and only if $f \in R$.

4 Basic Equations

From now on we only consider equations with three unknowns. The alphabet of unknowns is $\Xi = \{x, y, z\}$. The left-hand side of an equation can be assumed to begin with x. We can also assume that x occurs on the right-hand side, but not as the first letter.

Periodic solutions and solutions, where some unknown has the value 1, are called *trivial*. These are easy to parameterize by Theorem 2.1.

An equation is a *basic equation*, if it is a trivial equation $U = U$, where $U \in \Xi^*$, if it has only trivial solutions, or if it is of one of the following forms, where $a, b \geq 1$, $c \geq 2$ and $t \in \{x, z\}$:

B1. $x^a y \ldots = y^b x \ldots$
B2. $x^2 \ldots \rightrightarrows y^a x \ldots$
B3. $xyt \ldots \rightrightarrows zxy \ldots$
B4. $xyt \ldots \rightrightarrows zyx \ldots$
B5. $xyz \ldots = zxy \ldots$
B6. $xyz \ldots = zyx \ldots$
B7. $xy^c z \ldots = zy^c x \ldots$
B8. $xyt \ldots \rightrightarrows z^a xy \ldots$
B9. $xyxz \ldots \rightrightarrows zx^2 y \ldots$

The parameterizability of basic equations is easy to prove with the help of previous lemmas and theorems.

Theorem 4.1. *Every basic equation has a parametric solution of bounded length.*

Proof. For equations $U = U$ and for equations with only trivial solutions the claim is clear. We prove it for equations B1 – B9. First we reduce equations to other equations by Theorem 3.1. The equation B2 is reduced by the substitution $x \mapsto yx$ to the equation $xyx \ldots = y^a x \ldots$, which is of the form B1. The equations B3 and B4 are reduced by the substitution $x \mapsto zx$ to the equations $xyz \ldots = zxy \ldots$ and $xyz \ldots = yzx \ldots$, which are of the form B5. The equation B8 is reduced by the substitution $x \mapsto zx$ to the equation $xyzA = z^a xyB$ for some $A, B \in \Xi^*$. By Lemma 2.8, this is equivalent with the equation $xyzxyzA = zxyz^a xyB$, which is of the form B5.

Consider the equations B1, B5, B6, B7 and B9. Their solutions are also solutions of $xy = yx$, $xyz = zxy$, $xyz = zyx$, $xy^c z = zy^c x$ and $xyxz \rightrightarrows zx^2 y$, respectively. For B1 this follows from Lemma 2.8, otherwise by a length argument. By Lemmas 2.1, 2.4, 2.3, 2.6 and 2.7, these latter equations have parametric solutions over word parameters p, q and numerical parameters i, j, k. By substituting the parametric words from these solutions to the original basic equations, we get exponential equations, which are equivalent with linear Diophantine relations by Theorems 3.2, 3.3 and 3.4. The parametric solutions with these linear Diophantine relations, together with parametric representations for the periodic solutions, determine parametric solutions for these equations. □

5 Images and θ-Images

In this section we define images and θ-images of equations and prove some results about them. If h is a solution of the equation $xU \rightrightarrows yV$, then $h(y) \leq h(x)$. This fact was already behind Theorem 3.1. This will be generalized.

Let $t_1, \ldots, t_n \in \{y, z\}$ and $V = t_1 \ldots t_n$. Let $t_{n+1} = t_1$. If a morphism h is a solution of the equation $E : xU \rightrightarrows VxW$, then

$$h(x) = h(V^k t_1 \ldots t_i)u \tag{1}$$

for some numbers k, i and word u satisfying $k \geq 0$, $0 < i \leq n$ and $h(t_{i+1}) \not\leq u$.

On the other hand, a morphism h satisfying (1) is a solution of E iff $uh(U) = h(t_{i+1} \ldots t_n t_1 \ldots t_i)uh(W)$. We can write $h = g \circ f$, where f is the morphism $x \mapsto V^k t_1 \ldots t_i x$ and g is the morphism for which $g(x) = u$, $g(y) = h(y)$ and $g(z) = h(z)$. Now h is a solution of E iff g is a solution of

$$xf(U) \leftleftarrows f(t_{i+1} \ldots t_n t_1 \ldots t_i)xf(W).$$

An *image* of an equation $xU(x, y, z) \rightrightarrows V(y, z)xW(x, y, z)$ under the morphism $x \mapsto V^k Px$, where $k \geq 0$, $V = PQ$ and $Q \neq 1$, is

$$xU(V^k Px, y, z) \leftleftarrows QPxW(V^k Px, y, z).$$

If V contains only one of y, z or if $P = 1$, the image is *degenerated*.

Images are needed in the most important reduction steps used in the proof of parameterizability of equations with three unknowns. The solutions of an equation are easily acquired from the solutions of its images, so it is enough to consider them. There are infinitely many images, but a finite number is enough, if one of them is turned from a one-sided equation to an ordinary equation.

Equation E is *reduced to the equations $E_1, \ldots, E_n$ by an n-tuple of substitutions*, if E is of the form $xU(x, y, z) \rightrightarrows t_1 \ldots t_k xV(x, y, z)$, where $1 \leq n \leq k$ and $t_1, \ldots, t_k \in \{y, z\}$, equation E_i is

$$xU(t_1 \ldots t_i x, y, z) \leftleftarrows t_{i+1} \ldots t_k t_1 \ldots t_i xV(t_1 \ldots t_i x, y, z),$$

when $1 \leq i < n$, and equation E_n is

$$xU(t_1 \ldots t_n x, y, z) = t_{n+1} \ldots t_k t_1 \ldots t_n xV(t_1 \ldots t_n x, y, z).$$

By the above, Theorem 3.1 can be generalized.

Theorem 5.1. *Let E be an equation of length n. If E is reduced to the equations $E_1, \ldots, E_m$ by an m-tuple of substitutions, and if $E_1, \ldots, E_m$ have parametric solutions of length at most c, then E has a parametric solution of length $O(mn)c$.*

Reductions with n-tuples of substitutions are not sufficient. Other ways to restrict the considerations to a finite number of images are needed.

Equation

$$xU(x, y, z) \rightrightarrows V(y, z)xW(x, y, z)$$

is of *type I*, if both unknowns y, z occur in V. Equation

$$xy^bU(x,y,z) \rightrightarrows z^cxV(x,z)yW(x,y,z),$$

where $b, c \geq 1$, is of *type II*, if $b > 1$ or $V \neq 1$.

Theorem 5.2. *The solutions of an equation of type I of length n can be parameterized in terms of the solutions of $O(n^2)$ of its images of length $O(n^3)$.*

Theorem 5.2 can be generalized by defining θ-images.

A sequence of equations $E_0, \ldots, E_n$ is a *chain*, if E_i is an image of E_{i-1} for all i, $1 \leq i \leq n$. Then E_n is an *image of order n* of E_0. If every E_i is a degenerated image, then the chain is degenerated and E_n is a degenerated image of order n.

We define *θ-images* of equations of type I and II. For equations of type I all images are θ-images. For equations of type II the degenerated images of order 2 and nondegenerated images of order 3 are θ-images.

The proofs of the following three lemmas, and also the proof of Theorem 5.2, consist of examining images and their solutions and using exponential equations. Especially the proof of Lemma 5.4 is somewhat complicated. We consider an equation of type II

$$xy^bA(x,y,z) \rightrightarrows z^cxB(x,z)yC(x,y,z), \tag{2}$$

where $b, c \geq 1$ and $b > 1$ or $B \neq 1$. Its images are degenerated and of the form

$$xy^bA(z^ix,y,z) \leftleftarrows z^cxB(z^ix,z)yC(z^ix,y,z).$$

Lemma 5.3. *The solutions h of (2) satisfying $|h(y)| \leq |h(z)|$ can be parameterized in terms of the solutions of $O(n^{17})$ of its θ-images of length $O(n^{18})$.*

Lemma 5.4. *If x occurs in B, then the nonperiodic solutions of (2), and some periodic solutions, can be parameterized in terms of the solutions of $O(n^{17})$ of its θ-images of length $O(n^{18})$.*

Lemma 5.5. *If $B = z^d$, where $d \geq 1$, then the solutions of (2) can be parameterized in terms of the solutions of $O(n^{26})$ of its θ-images of length $O(n^{27})$.*

We define a *complete set of θ-images* of an equation of type I or II. For equations of type I it is the set of Theorem 5.2. For equations of the form (2) it is the set of Lemma 5.3, if $B = 1$, the set of Lemma 5.4, if x occurs in B, and the set of Lemma 5.5, if $B = z^d$, $d \geq 1$. The next theorem follows immediately from this definition.

Theorem 5.6. *Every equation of type I or II of length n has a complete set of θ-images consisting of $O(n^{26})$ equations of length $O(n^{27})$.*

We assume that every complete set of θ-images satisfies the conditions of Theorem 5.6. The next theorem requires only little extra work.

Theorem 5.7. *Let E be a word equation of length n. If $\{E_1, \ldots, E_m\}$ is a complete set of θ-images of E and every E_i has a parametric solution of length at most c, then E has a parametric solution of length $O(mn^{26})c$.*

6 Trees of Equations

The proof of the parameterizability of equations with three unknowns consists mainly of reducing equations to other equations. This forms a tree-like structure. The intention is to make all leaf equations in this tree to be basic equations. The possible reduction steps are given in the definition of a neighborhood.

Lemma 6.1. *Let E_0 be the equation $xy^a zy^p s \ldots \rightrightarrows zy^b xy^q t \ldots$, where $s, t \in \{x, z\}$ and $a+p \neq b+q$. Let $k \geq 8+|p-q|$ be even, E_k be the equation $xP \rightrightarrows zQ$ and $E_0, \ldots, E_k$ be a degenerated chain. Now the solutions of E_k satisfying $y \neq 1$ are also solutions of the equation $xy^a zy^b \rightrightarrows zy^b xy^a$.*

The equations $E_1, \ldots, E_n$ form a *neighborhood* of an equation E, if one of the following conditions holds:

N1. $E_1, \ldots, E_n$ form a complete set of θ-images of E,
N2. E reduces to $E_1, \ldots, E_n$ with an n-tuple of substitutions,
N3. E is the equation $U = V$, U and V begin with different letters, $n = 2$, and E_1 and E_2 are equations $U \rightrightarrows V$ and $V \rightrightarrows U$,
N4. $n = 1$ and E is the equation $U = V$ and E_1 is the equation $U^R = V^R$,
N5. E is the equation $SU = TV$, $|S|_t = |T|_t$ for all $t \in \Xi$, $n = 1$ and E_1 is the equation $US = VT$,
N6. $n = 1$ and E_1 is E reduced from the left or multiplied from the right,
N7. $n = 1$ and, with the assumptions of lemma 6.1, E is the equation $xP \rightrightarrows zQ$ and E_1 the equation $xy^a zy^b xP \rightrightarrows zy^b xy^a zQ$.

Theorem 6.2. *Let E be a word equation of length n and let $E_1, \ldots, E_m$ be its neighborhood. If each E_i has a parametric solution of length at most c, then E has a parametric solution of length $O(mn^{26})c$.*

Proof. For N1 this follows from Theorem 5.7, for N2 from Theorem 5.1 and for N7 from Lemma 6.1. The other cases are clear. □

Directed acyclic graph, whose vertices are equations, is a *tree* of E, if the following conditions hold:

(i) only vertex with no incoming edges is E,
(ii) all other vertices have exactly one incoming edge,
(iii) if there are edges from E_0 to exactly $E_1, \ldots, E_n$, then these equations form a neighborhood of E.

Theorem 6.3. *Let E be a word equation of length n. If E has a tree of height k, then all equations in the tree are of length $O(n)^{27^k}$. If each leaf equation in this tree has a parametric solution of length at most c, then E has a parametric solution of length $O(n)^{52 \cdot 27^k} c$.*

Proof. In the case N1 the first claim follows directly from Theorem 5.6, and for the other cases the bound $O(n)^{27^k}$ is more than enough. Now, by Theorem 6.2, there exists a constant a such that E has a parametric solution of length

$$\begin{aligned} &a(an)^{52} \cdot a((an)^{27})^{52} \cdot a((an)^{27^2})^{52} \cdot \cdots \cdot a((an)^{27^{k-1}})^{52} \cdot c \\ <&a^k (an)^{52 \cdot 27^k} c = O(n)^{52 \cdot 27^k} c. \qquad \square \end{aligned}$$

A tree in which all leaves are basic equations is a *basic tree.*

If every θ-image of an equation of type I or II has a basic tree, then the equation has a basic tree, because it has a complete set of θ-images. The rule N1 is used this way instead of explicitly selecting some complete set of θ-images.

The main theorem is proved by a sequence of lemmas. The lemmas are proved by using the rules of the definition of a neighborhood in various ways.

Lemma 6.4. *The equation $xyz^2A(x,y,z) = yz^2xB(x,y,z)$ has a basic tree.*

Lemma 6.5. *The equation $x^2yz\ldots \rightrightarrows zyxy\ldots$ has a basic tree.*

Lemma 6.6. *Let $s \neq x$ and $t \neq y$. The following equations have basic trees:*

(a) $xy^2z\ldots \rightrightarrows zx^2y\ldots$,
(b) $xyzs\ldots \rightrightarrows zx^2y\ldots$,
(c) $xy^2z\ldots \rightrightarrows zxyt\ldots$,
(d) $xyzt\ldots \rightrightarrows zy^2x\ldots$,
(e) $xyz\ldots \rightrightarrows zy^2x\ldots$.

Let $1 \leq a, b \leq 2$, $d \geq 1$ and $t \neq y$. The equations $x^ay^bt\ldots \rightrightarrows zyx\ldots$, $x^ay^bt\ldots \rightrightarrows zxy\ldots$ and $x^ay^bt\ldots \rightrightarrows z(yz)^dx\ldots$ are *supporting equations.*

Lemma 6.7. *Every supporting equation has a basic tree.*

Lemma 6.8. *The equation $xy^azy^ps\ldots \rightrightarrows zy^bxy^qt\ldots$, where $a > 0$, $a+p = b+q$ and $s, t \neq y$, has a basic tree.*

The next proof contains maybe the most critical part of the construction, because very long chains of images are considered. Similar construction was needed also in the proof of Lemma 6.6.

Lemma 6.9. *The equation $xy^az\ldots \rightrightarrows zy^bx\ldots$, where $a > 0$, has a basic tree.*

Proof. The equation can be written in the form $E_0 : xy^azy^pu\ldots \rightrightarrows zy^bxy^qv\ldots$, where $u, v \neq y$. If $a+p = b+q$, then the claim follows from Lemma 6.8. Assume that $a+p \neq b+q$. Let $l \geq 8+|p-q|$ be even. Form a complete set of θ-images of E_0, a complete set of θ-images of these, and so on l times. These θ-images form chains $E_0, \ldots, E_l$. We show that each chain has an equation with a basic tree; this proves the claim.

First, consider chains of degenerated θ-images. There is a corresponding chain of ordinary images and we can use the rule N7. The equation E_l is replaced by the equation $xy^azy^bxP \rightrightarrows zy^bxy^azQ$, which has a basic tree by Lemma 6.8.

Second, consider nondegenerated chains. Assume that the part $E_0, \dots, E_{j-1}$ of the chain is degenerated and that E_j is a nondegenerated θ-image of E_{j-1}. If $b = 0$, the equation E_0 is of the form $xy^a z \dots \rightrightarrows zx \dots$, and E_{j-1} is of the same form. The equation E_j can be seen to be a supporting equation and thus it has a basic tree. If $b > 0$, then E_0 is of the form $xy^a z \dots \rightrightarrows zy^b x \dots$. Equation E_{j-1} is of the same form. Now E_j is of the form $y^c z y^d x \dots \rightrightarrows xy^a z \dots$, where $c + d = a$ and $c \geq 1$. If $c > 1$, then E_j is basic of the form B2. If $c = 1$, then all θ-images of E_j can be seen to have basic trees by Lemmas 6.6 and 6.7. □

Lemma 6.10. *The equation $xy^a t \dots \rightrightarrows z^c x B(x,z) y \dots$, where $a, c \geq 1$ and $t \neq y$, has a basic tree.*

Lemma 6.11. *The equation $x^n y^m t \dots \rightrightarrows zy A(y,z) x \dots$, where $n, m \geq 1$ and $t \neq y$, has a basic tree.*

The proof of the next theorem finally gathers the previous results together and gives the idea of how the height of the tree can be estimated.

Theorem 6.12. *Every equation of length n with three unknowns has a basic tree of height $O(n)$.*

Proof. The trivial equation $U = U$ is a basic equation. All other equations can be reduced from the left and split into one-sided equations. By multiplication from the right, every one-sided equation can be turned into one of the equations

$$x^2 \dots \rightrightarrows y^c x \dots \tag{3}$$
$$xy \dots \rightrightarrows y^c x \dots \tag{4}$$
$$xz^a t \dots \rightrightarrows y^c x B(x,y) z \dots \tag{5}$$
$$x^a y^b s \dots \rightrightarrows y^c z B(y,z) x \dots \tag{6}$$
$$x^a z^b t \dots \rightrightarrows yz B(y,z) x \dots \tag{7}$$
$$x^a z^b t \dots \rightrightarrows y^d z B(y,z) x \dots, \tag{8}$$

where $a, b, c \geq 1$, $d > 1$, $t \neq z$ and $s \neq y$. We prove that these have basic trees.

Equation (3) is basic of the form B2. Equation (4) is reduced by the substitution $x \mapsto yx$ to the equation $xy \dots = y^c x \dots$, which is basic of the form B1. Equation (5) is the equation of Lemma 6.10. Equation (7) is the equation of Lemma 6.11.

The equation (6) is of type I and its images are of the form $xy \dots \leftrightarrows Dx \dots$, where D is a conjugate of $y^c z B$. If $y^2 \leq D$, then this is of the form (3), if $yz \leq D$, then of the form (5), and if $z \leq D$, then of the form (7). So every image of (6) and thus the equation itself has a basic tree.

The equation (8) is of type I and its images are of the form $x(y \dots)^{a-1} z^b y \dots \leftrightarrows Dx \dots$, where D is a conjugate of $y^d z B$. Again it is of the form (3), (5) or (7). So every image of (6) and thus the equation itself has a basic tree.

The constructions of trees in the lemmas produce trees of bounded height with two exceptions: Lemmas 6.6 and 6.9, where a tree with height of order

$|p-q|$ is constructed for the equation

$$xy^a zy^p \ldots \rightrightarrows zy^b xy^q \ldots . \tag{9}$$

We prove that the powers of y here cannot be more than n, which proves this theorem. In the definition of neighborhood, the rules N1, N2, N5 and N6 can produce higher powers than those in the initial equation. There is no need to use N6 to generate high powers and N5 is only used in 6.4, 6.6 and 6.8, where it does not generate high powers. Consider N1 and N2. Here an equation $xU(x,y,z) \rightrightarrows y^a xV(x,y,z)$ can be turned into $xU(y^i x,y,z) \lhd y^a xV(y^i x,y,z)$ for high values of i. But in order for y to be in the position of (9), the rules N1 or N2 must be used again. Then y is replaced by xuy for some $u \in \{x,z\}^*$ and the powers of y disappear. The claim is proved. □

In the next theorem $\exp^2$ denotes the double exponential function $\exp \circ \exp$.

Theorem 6.13. *Every equation of length n with three unknowns has a parametric solution of length $\exp^2(O(n))$.*

Proof. By Theorem 6.12 every equation has a basic tree of height $O(n)$. By Theorem 4.1 the leaf equations have parametric solutions of bounded length. Now from Theorem 6.3 it follows that E has a parametric solution of length $O(n)^{52 \cdot 27^k}$, where $k = O(n)$, that is of length $\exp^2(O(n))$. □

References

1. M.H. Albert and J. Lawrence: A proof of Ehrenfeucht's Conjecture. Theoret. Comput. Sci. **41** (1985) 121–123
2. C. Choffrut, J. Karhumäki: Combinatorics of words. In: G. Rozenberg, A. Salomaa (eds), Handbook of Formal Languages, Springer (1997)
3. E. Czeizler, J. Karhumäki: On non-periodic solutions of independent systems of word equations over three unknowns. Internat. J. Found. Comput. Sci. **18** (2007) 873–897
4. V. S. Guba: Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems. Mat. Zametki **40** (1986) 321–324
5. T. Harju, J. Karhumäki, W. Plandowski: Independent system of equations. In: M. Lothaire (ed.), Algebraic Combinatorics on Words, Cambridge University Press (2002)
6. Y.I. Hmelevskii: Equations in free semigroups. Proc. Steklov Inst. of Math. **107** (1971); Amer. Math. Soc. Translations (1976)
7. J. Karhumäki, A. Saarela: A Reproof of Hmelevskii's Theorem. Manuscript
8. M. Lothaire: Combinatorics on Words. Addison-Wesley (1983)
9. G. S. Makanin: The problem of solvability of equations in a free semigroup. Mat. Sb. **103** (1977) 147–236; English transl. in Math. USSR Sb. **32** 129–198
10. W. Plandowski: Satisfiability of word equations with constants is in PSPACE. J. ACM **51** (2004) 483–496
11. J.-C. Spehner: Quelques Problemes d'extension, de conjugaison et de presentation des sous-monoides d'un monoide libre. Ph.D. Thesis, Univ. Paris (1976)
12. J.-C. Spehner: Les presentations des sous-monoides de rang 3 d'un monoide libre. Semigroups, Proc. Conf. Math. Res. Inst. (1978) 116–155