On the Complexity of Hmelevskii's Theorem and Satisfiability of Three Unknown Equations

Aleksi Saarela *

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

Abstract. We analyze Hmelevskii's theorem, which states that the general solutions of constant-free equations on three unknowns are expressible by a finite collection of formulas of word and numerical parameters. We prove that the size of the finite representation is bounded by an exponential function on the size of the equation. We also prove that the shortest nontrivial solution of the equation, if it exists, is exponential, and that its existence can be solved in nondeterministic polynomial time.

1 Introduction

This work concerns the theory of word equations, which is a fundamental part of combinatorics on words. It has connections to many other areas including representation results of algebra, theory of algorithms and pattern matching.

Some remarkable results of this topic proved during the last few decades are the decidability of the satisfiability problem for word equations, see [11], and the compactness result of systems of word equations, see [1] and [6]. The first result was improved to a PSPACE algorithm in [12]. The satisfiability problem has been conjectured to be in NP [13].

In the case of constant-free word equations with only three unknowns important results have also been achieved. Hmelevskii [8] proved in 1970 that the general solution of any such equation can be expressed as a finite formula on word and numerical parameters. On the other direction Spehner [15, 16] classified all sets of relations a given solution can satisfy. Both of these results have only very complicated proofs. Another example of a challenging nature of word problems is that the question of finding any upper bound for the maximal size of independent system of word equations on three unknowns is still open, see [7] and [3].

The result of Hmelevskii is well known, see e.g. [10], but the original presentation is very hard to read. A simplified proof using modern tools of combinatorics on words has been given, together with a double exponential upper bound of the size of the formula giving the general solution, see [9]. A complete write-up of the results in [9] is in [14].

In this paper we continue the work of analyzing Hmelevskii's result. Based on [9] and [14] we improve the bound of the size of the parametric solution to single

* Supported by the Academy of Finland under grant 8121419

exponential, as well as prove that the length of the shortest nontrivial solution is also exponential (if such a solution exists). This connects our work to the satisfiability problem mentioned above, because Plandowski and Rytter proved in [13] that there is a nondeterministic algorithm solving the problem in time polynomial in $n \log N$, where n is the length of the equation and N is the length of the shortest solution. From this and our result it follows that the problem of deciding if a constant-free equation on three unknowns has a nontrivial solution is in NP.

2 Definitions

We begin by giving some definitions needed in this paper. A basic reference of the subject is [2].

We assume that all word equations are constant-free unless otherwise stated. Thus 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 a $t \in \Sigma^*$ such 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 solutions.

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* and the notation $u < v$ is used.

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 $i = 2, \dots, n$.

Next we define the central notions of this paper: parametric words and parametric solutions.

We 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 parentheses 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 $\Lambda \to \mathbb{N}_0 = \{0, 1, 2, \dots\}$, 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 the 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 the 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 and z. We are only interested in the nonnegative values of the unknowns. If $\Lambda = \{i_1, \dots, i_k\}$, f is a function $\Lambda \to \mathbb{N}_0$, and $f(i_1), \dots, f(i_k)$ satisfy R, then the notation $f \in R$ can be used.

Let S be a set of morphisms $\Xi^* \to \Sigma^*$, $\Lambda = \{i_1, \dots, i_k\}$, h_j a morphism from the monoid Ξ^* to parametric words and R_j a linear Diophantine relation, when $j = 1, \dots, m$. 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 $j = 1, \dots, m$ and $x \in \Xi$.

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.

Example 2.1. The equation $xz = zy$ has a parametric solution $\{(h_1, R), (h_2, R)\}$, where $\Delta = \{p, q\}$, $\Lambda = \{i\}$, $h_1(x) = pq$, $h_1(y) = qp$, $h_1(z) = p(qp)^i$, $h_2(x) = h_2(y) = 1$, $h_2(z) = p$ and R is the trivial relation satisfied by all functions $f : \Lambda \to \mathbb{N}_0$.

3 Remarks About Parametric Solutions

Next we make some remarks about parametric solutions to increase our understanding of them. The various claims made in this section are not needed in this paper. The proofs can be found in [14].

A parametric solution was defined as a set $\{(h_j, R_j) : 1 \le j \le m\}$. This solution can be written less formally as

$$\begin{aligned} &x = h_1(x),\ y = h_1(y),\ z = h_1(z),\ R_1 \ \text{ or} \\ &\ \vdots \\ &x = h_m(x),\ y = h_m(y),\ z = h_m(z),\ R_m, \end{aligned}$$

if the unknowns are x, y, z. Actually, only one pair (h, R) is needed. For example, if we have a parametric solution

$$x = \alpha_1,\ y = \beta_1,\ z = \gamma_1 \qquad \text{or} \qquad x = \alpha_2,\ y = \beta_2,\ z = \gamma_2,$$

we can replace it with

$$x = \alpha_1^i \alpha_2^j,\ y = \beta_1^i \beta_2^j,\ z = \gamma_1^i \gamma_2^j, \quad i + j = 1,$$

where i and j are new parameters.

On the other hand, the linear Diophantine relations are not necessary either, if we again allow many morphisms. We can get rid of the relations by replacing every pair (h, R) with several morphisms h. This follows from article [4].

Example 3.1. Consider the periodic solutions of the equation $x^n = yz$. They are

$$x = t^i,\ y = t^j,\ z = t^k, \quad ni = j + k.$$

We can replace j with $nj' + b$ and k with $nk' + c$, where $0 \le b, c < n$. Then $i = j' + k' + (b + c)/n$. Only those pairs (b, c) for which $b + c$ is divisible by n are possible. Thus we get a representation

$$\begin{aligned} &x = t^{j'+k'}, && y = t^{nj'}, && z = t^{nk'} && \text{or} \\ &x = t^{j'+k'+1}, && y = t^{nj'+1}, && z = t^{nk'+n-1} && \text{or} \\ &x = t^{j'+k'+1}, && y = t^{nj'+2}, && z = t^{nk'+n-2} && \text{or} \\ &\ \vdots \\ &x = t^{j'+k'+1}, && y = t^{nj'+n-1}, && z = t^{nk'+1}, \end{aligned}$$

where the parameters j', k' can now have any nonnegative values.

The periodic solutions of an equation on three unknowns can be represented with just one morphism and without any Diophantine relations. This does not hold, if instead of periodic solutions we consider all solutions. Indeed, a parametric solution for the equation $xyxzyz = zxzyxy$ consists of at least three

morphisms, if linear Diophantine relations are not allowed. The next example gives the solutions of this equation. We are not aware of any better lower bounds for the maximal required number of morphisms in these kinds of parametric solutions.

Example 3.2. The solutions of the equation $xyxzyz = zxzyxy$ are

$$x = p,\ y = q,\ z = 1 \quad \text{or} \quad x = p,\ y = q,\ z = pq \quad \text{or} \quad x = p^i,\ y = p^j,\ z = p^k,$$

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

4 Basic Equations

Hmelevskii proved that every equation on three unknowns has a parametric solution, and the size of this solution was estimated to be at most double exponential in [9]. Our first goal is to improve this bound to single exponential, and our second goal is to prove that the shortest nontrivial solution is also of exponential size. We need to refer to the theorems and proofs in [14]. Often these theorems claim the existence of some object, while we need to know also something about the size or structure of that object. Typically this information can be obtained simply by examining the old proof, but this is not trivial. In these cases we state the more precise form of the theorem, but do not repeat the proof.

In this section the new information is about the coefficients of some linear Diophantine relations. This is necessary for our second goal.

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 $f(\alpha) = f(\beta)$.

The following three theorems were proved in [14, Theorems 5.1, 5.2, 5.3] except for the upper bounds of the sizes of the coefficients in the relation R. These bounds, however, are easily obtained by examining the proofs. The latter two theorems (especially the last one) are technical variations of the first one.

Theorem 4.1. *Let $E : \alpha = \beta$ 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 $f \in R$. The sizes of the coefficients in R are of the same order as the length of $\alpha\beta$.*

Theorem 4.2. *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 equations $E : \alpha = \beta$ if and only if $f \in R$. The sizes of the coefficients in R are of the same order as the length of $\alpha\beta$.*

Theorem 4.3. *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 $f \in R$. The sizes of the coefficients in R are of the same order as the length of $h(A)h(B)$.

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.

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 quite easy to prove and was done in [14, Theorem 6.2]. The $O(n)$ bound for the coefficients in the linear Diophantine relations follows from the bounds in Theorems 4.1, 4.2 and 4.3.

Theorem 4.4. *Every basic equation has a parametric solution. The solution is of length $O(1)$ and the coefficients in the linear Diophantine relations are of size $O(n)$, where n is the length of the equation.*

5 Length of the Parametric Solution

In this section we prove that the size of the parametric solution is exponential. At the same time we improve some of the theorems in [14] so that they can be used later to prove the existence of a nontrivial solution of exponential size.

First we define images of equations and some other related concepts. These definitions are very important in the proof of Hmelevskii's theorem.

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

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

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

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

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

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

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

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.

The following lemma is the same as [14, Lemma 8.1].

Lemma 5.1. *Let* $u, v, w \in \Sigma^*$, $0 < |w| \leq |u|$ *and* $c \geq 1$. *If*

$$wu^{c+1}v \ldots = u^{c+1}vu \ldots \qquad \text{or} \qquad w(uv)^c u^2 \ldots = (uv)^c u^2 \ldots,$$

then $uv = vu$.

The next lemma is a seemingly minor but essential improvement of [14, Lemma 8.2]: the number k in the lemma can be selected to be logarithmic instead of linear with respect to the number $|p - q|$. This is what ultimately leads to an exponential bound for the length of the parametric solution.

Lemma 5.2. *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 *be an even number such that* $2^{(k-4)/2} \geq 1 + |p - q|$. *Let* 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* $xy^a zy^b \rightrightarrows zy^b xy^a$.

Proof. Assume that E_{i+1} is the image of E_i under the morphism $f_i : x \mapsto (zy^b)^{c_i} x$, when i is even, and under the morphism $f_i : z \mapsto (xy^a)^{c_i} z$, when i is odd. Because $f_0(x)$ and $f_0(z)$ and thus $f_0(s)$ and $f_0(t)$ begin with z, the equation E_k is of the form

$$xy^a zy^p r \ldots \rightrightarrows zy^b xy^q r \ldots, \tag{1}$$

where

$$r = (f_k \circ \cdots \circ f_1)(z) = (f_k \circ \cdots \circ f_4)((((xy^a)^{c_3} zy^b)^{c_2} xy^a)^{c_1} (xy^a)^{c_3}).$$

Let $F_m = f_m \circ \cdots \circ f_4$. The words xy^a and zy^b occur as factors of $F_4(xy^a)$ at least once, and if they occur as factors of $F_m(xy^a)$ at least $2^{(m-4)/2}$ times, they occur as factors of $F_{m+2}(xy^a)$ at least $2^{(m-2)/2}$ times. Thus, by induction, they occur as factors of $F_k(xy^a)$ at least $2^{(k-4)/2}$ times. If h is a solution of E_k, then

$$\begin{aligned} ||h(xy^a zy^p)| - |h(zy^b xy^q)|| &= |a + p - b - q||h(y)| \\ \leq (a + b)|h(y)| + |p - q||h(y)| &\leq (1 + |p - q|)|h(xy^a zy^b)| \\ \leq 2^{(k-4)/2}|h(xy^a zy^b)| &\leq |h(F_k(xy^a))|. \end{aligned}$$

Thus, by (1),

$$w((u^{c_3}v)^{c_2}u)^{c_1}u^{c_3}\ldots = ((u^{c_3}v)^{c_2}u)^{c_1}u^{c_3}\ldots,$$

where $u = h(F_k(xy^a))$, $v = h(F_k(zy^b))$ and $|w| \leq |u|$. If $w = 1$, then $h(xy^azy^p) = h(zy^bxy^q)$, which is not possible by the assumptions $h(y) \neq 1$ and $a+p \neq b+q$. Thus it follows from Lemma 5.1 that $uv = vu$. It can be seen that $u, v \in \{h(xy^a), h(zy^b)\}^*$, u ends with $h(xy^a)$ and v ends with $h(zy^b)$. This means that $h(xy^a)$ and $h(zy^b)$ satisfy a nontrivial relation. It follows that they commute, that is $h(xy^azy^b) = h(zy^bxy^a)$. □

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 (see [14]),
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 5.2, E is the equation $xP \rightrightarrows zQ$ and E_1 the equation $xy^azy^bxP \rightrightarrows zy^bxy^azQ$.

The first paragraph of the next theorem, proved in [14, Theorem 8.3], justifies the definition of a neighborhood. The second paragraph can be deduced by examining the rules in the definition of a neighborhood and, most importantly, the definition of a complete set of θ-images.

Theorem 5.3. *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 $O(mn^{26})c$.*

Compared to the parametric solutions of the equations E_i, the parametric words in the parametric solution of E contain $O(1)$ new numerical parameters, the height of the parametric words can increase by $O(1)$, and the coefficients of the linear Diophantine relations are of the same size.

A 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_0.

The first paragraph of the next theorem is from [14, Theorem 8.4], and the second paragraph follows from the second paragraph of Theorem 5.3 and from Theorem 4.4.

Theorem 5.4. *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 $O(n)^{52\cdot 27^k}c$.*

If the leaf equations are basic equations, the parametric words in the parametric solution of E contain $O(k)$ numerical parameters, their height is $O(k)$, and the coefficients of the linear Diophantine relations are of size $O(n)^{27^k}$.

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

The old version of Lemma 5.2 was used in [14, Lemmas 9.3, 10.2]. By using the improved version and making the corresponding small changes in the proof of [14, Theorem 10.5] gives the following theorem.

Theorem 5.5. *Every equation of length n with three unknowns has a basic tree of height $O(\log n)$.*

Now we can prove one of our main results. We note that it seems unlikely that Hmelevskii's methods would give a sub-exponential bound.

Theorem 5.6. *Every equation of length n with three unknowns has a parametric solution of length $\exp(n^{O(1)})$.*

Proof. By Theorem 5.5, every equation has a basic tree of height $O(\log n)$. By Theorem 4.4, the leaf equations have parametric solutions of bounded length. Now from Theorem 5.4 it follows that E has a parametric solution of length $O(n)^{52\cdot 27^k}$, where $k = O(\log n)$, that is of length $\exp(n^{O(1)})$. □

6 Shortest Nontrivial Solution

Based on Theorem 5.6 we can prove that the shortest nontrivial solution is of exponential length. However, this is not trivial. For example, if we have a parametric word $(p^iq)^j$, then by giving the value 1 for the numerical parameters we get a short word, but the problem is that $i = j = 1$ does not necessarily satisfy the linear Diophantine relation. Thus we need to estimate the size of the minimal solution of the relation. We also need to make sure that the solution of the word equation is indeed nontrivial.

Theorem 6.1. *If an equation of length n with three unknowns has a nontrivial solution, it has a nontrivial solution of length $\exp(n^{O(1)})$.*

Proof. Consider an equation $E : x \ldots = y \ldots$ and its parametric solution

$$\{(h_j, R_j) : 1 \le j \le m\}$$

of length $\exp(n^{O(1)})$. If E has a nontrivial solution, it has a solution where x and y begin with the same letter but z begins with a different letter. Let $h \circ f \circ h_j$ be such a solution, where $h \circ f$ is a valuation. Now also $f \circ h_j$ is such a solution,

and so is $g \circ h_j$, if $g \in R_j$ maps exactly the same numerical parameters to zero as f. Thus $g \circ h_j$ is a nontrivial solution. We must select g so that this solution is sufficiently short.

The lengths of the parametric words $h_j(t)$, where $t \in \{x, y, z\}$, are $\exp(n^{O(1)})$. By Theorems 5.4 and 5.5, every occurrence of a word parameter in $h_j(t)$ appears at most $g(i_1) \dots g(i_k)$ times in $g(h_j(t))$, where $i_1, \dots, i_k$ are numerical parameters and $k = O(\log n)$. Thus the length of $g(h_j(t))$ is $g(i_1) \dots g(i_k) \exp(n^{O(1)})$.

The conditions for g are that it must be in R_j and it must map exactly the same numerical parameters to zero as f. The latter condition can be handled by adding either the equation $i = 0$ (if $f(i) = 0$) or the inequality $i > 0$ (if $f(i) > 0$) to R_j for every $i \in \Lambda$. Inequalities $i > c$ can be replaced with $i = c + 1 + i'$, where i' is a new variable. In this way we get a linear Diophantine relation R'_j, which is a disjunction of linear systems of equations. Because $f \in R'_j$, at least one of these systems has a nonnegative integer solution.

According to [5], if a system of linear equations has a nonnegative integer solution, then it has one of size $O(lM)$, where l is the number of unknowns, M is an upper bound for the $r \times r$ subdeterminants of the augmented matrix of the system, and r is the rank of the system. Now r is at most $l = O(\log n)$. The coefficients in the system are of exponential size by Theorems 5.4 and 5.5, so $M = \exp(n^{O(1)})$. Thus there is a nonnegative integer solution of size $\exp(n^{O(1)})$. This solution gives us a function g such that $g(i_1) \dots g(i_k) \exp(n^{O(1)}) = \exp(n^{O(1)})$. This proves the theorem. □

Now we consider the satisfiability problem. Constant-free equations have always the solution, where every unknown gets the value 1, and usually they have also other periodic solutions. The natural question is thus whether a constant-free equation has a nontrivial solution. This can be easily reduced to the satisfiability problem of equations with constants. In this way we get the result that the above-mentioned question is in NP for equations on three unknowns.

Theorem 6.2. *The existence of a nontrivial solution of a constant-free equation on three unknowns can be decided in nondeterministic polynomial time.*

Proof. The equation $xU = yV$, where $U, V \in \Xi^*$, has a nontrivial solution if and only if it has a solution $x = ax', y = ay', z = bz'$, where a and b are different letters and $x', y', z' \in \Sigma^*$. So we are interested in the existence of a solution for the equation obtained from $xU = yV$ by replacing x with ax', y with ay' and z with bz', where x', y', z' are now new unknowns. The length of this equation is twice the length of the original equation.

There is a nondeterministic algorithm (see [13]) that solves the existence of a solution for the last equation in time polynomial in $n \log N$, where n is the length of the equation and N is the length of the shortest solution. The claim now follows from Theorem 6.1. □

References

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

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. 18:873–897 (2007)
4. S. Eilenberg, M.P. Schützenberger: *Rational sets in commutative monoids.* J. Algebra 13:173-191 (1969)
5. J. von zur Gathen, M. Sieveking: *A bound on solutions of linear integer equalities and inequalities.* Proc. Amer. Math. Soc. 72:155–158 (1978)
6. V. S. Guba: *Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems.* Mat. Zametki 40:321–324 (1986)
7. T. Harju, J. Karhumäki, W. Plandowski: *Independent systems of equations.* In: M. Lothaire (ed), Algebraic Combinatorics on Words, Cambridge University Press (2002)
8. Y.I. Hmelevskii: *Equations in free semigroups.* Proc. Steklov Inst. of Math. 107 (1971); Amer. Math. Soc. Translations (1976)
9. J. Karhumäki, A. Saarela: *An analysis and a reproof of Hmelevskii's theorem.* Proc. of 12th International Conference on Developments in Language Theory, 467–478 (2008)
10. M. Lothaire: *Combinatorics on words.* Addison-Wesley (1983)
11. G. S. Makanin: *The problem of solvability of equations in a free semigroup.* Mat. Sb. 103:147–236 (1977); English transl. in Math. USSR Sb. 32:129–198
12. W. Plandowski: *Satisfiability of word equations with constants is in PSPACE.* J. ACM 51:483–496 (2004)
13. W. Plandowski, W. Rytter: *Application of Lempel-Ziv encodings to the solution of word equations.* Proc. of 25th International Colloquium on Automata, Languages, and Programming 731–742 (1998)
14. A. Saarela: *A new proof of Hmelevskii's theorem.* Licentiate thesis, Univ. Turku (2009) (http://users.utu.fi/amsaar/en/licthesis.pdf)
15. 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)
16. J.-C. Spehner: *Les presentations des sous-monoides de rang* 3 *d'un monoide libre.* Semigroups, Proc. Conf. Math. Res. Inst. 116–155 (1978)