Fundamenta Informaticae 162 (2018) 223–235
DOI 10.3233/FI-2018-1722
IOS Press

Studying Word Equations by a Method of Weighted Frequencies

Aleksi Saarela[*]
Department of Mathematics and Statistics
University of Turku, 20014 Turku, Finland
amsaar@utu.fi

Abstract. We briefly survey some results and open problems on word equations, especially on those equations where the right-hand side is a power of a variable. We discuss a method that was recently used to prove one of the results, and we prove improved versions of some lemmas that are related to the method and can be used as tools when studying word equations. We use the method and the tools to give new, simple proofs for several old results.

1. Introduction

We begin by describing some results (both classical and recent) and open problems related to word equations. In this section, we use the term "word equation" informally to refer to an equality of words. Later, we will give a formal definition of word equations and reformulate some theorems in terms of this formalism. For basics in combinatorics on words, we refer to the book of Lothaire [13]. Throughout this article, let Γ denote an alphabet.

The following classical result was proved by Lyndon and Schützenberger [14] (actually, they proved a stronger result where the free monoid Γ^* is replaced by a free group). Other proofs for the free monoid case have been given, for example the very simple proof by Harju and Nowotka [7].

Theorem 1.1. Let $x, y, z \in \Gamma^*$ and $k, l, m \geq 2$. If $x^k y^l = z^m$, then x, y and z commute.

Theorem 1.1 can also be formulated as follows: If $k, l \geq 2$ and x and y do not commute, then $x^k y^l$ is primitive. But what about the case where $k = 1$ or $l = 1$? Then $x^k y^l$ might be nonprimitive, but only for one pair (k, l). This is formally stated in the following extension of Theorem 1.1.

[*] Address for correspondence: Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland

Theorem 1.2. Let $x, y \in \Gamma^*$ be words that do not commute. The language x^+y^+ contains at most one nonprimitive word, and this word is in $xy^+ \cup x^+y$.

Theorem 1.2 was proved by Shyr and Yu [19], and a simpler proof was given by Dömösi, Horváth and Vuillon [4]. However, it is actually a special case of a result that was proved earlier by Spehner [20], and also by Barbin-Le Rest and Le Rest [2]. This stronger result is given in Theorem 1.3. As an introduction to this theorem, let us consider the following question: By Theorem 1.2, the language x^+y^+ contains at most one nonprimitive word, but what about the larger language $\{x, y\}^+ \smallsetminus \{x, y\}$? Trivially, if $u \in \{x, y\}^+$, then u^m is nonprimitive for all $m \geq 2$. If we exclude these trivial cases, then there is, up to conjugacy, at most one nonprimitive word (if u is nonprimitive, then so is every conjugate of u). This is formally stated in the following extension of Theorem 1.2.

Theorem 1.3. Let $x, y \in \Gamma^*$ be words that do not commute. Let $\{X, Y\}$ be an alphabet and let $h : \{X, Y\}^* \to \Gamma^*$ be the morphism defined by $h(X) = x$ and $h(Y) = y$. Up to conjugacy, there is at most one primitive word $W \in \{X, Y\}^+ \smallsetminus \{X, Y\}$ such that $h(W)$ is nonprimitive, and this word W is in $XY^+ \cup YX^+$ (again, up to conjugacy).

Equations $x^k y^l = z^m$ can be generalized by letting the left-hand side consist of more than two powers. Harju and Nowotka [8] proved the following result.

Theorem 1.4. Let $n \geq 2$, $x_0, \dots, x_n \in \Gamma^*$, $k_0, \dots, k_n \geq 3$ and $k_0 \geq n$. Assume that there does not exists $i \in \{1, \dots, n\}$ such that the primitive roots of x_0 and x_i are conjugate. Then $x_1^{k_1} \cdots x_n^{k_n} \neq x_0^{k_0}$.

Actually, there is a small error in the formulation of this theorem in [8]: The assumption about the primitive roots not being conjugate is replaced by the assumption that $|x_0| \neq |x_i|$ for all $i \geq 1$ and $x_0 x_i \neq x_i x_0$ for at least one $i \geq 1$, but this assumption is too weak, as shown by the counterexample $n = 3$, $k_0 = k_2 = k_3 = 3$, $k_1 = 5$, $x_0 = (a^3b^3)^2$, $x_1 = a^3b^3$, $x_2 = a$, $x_3 = b$.

The equations

$$x_1^k x_2^k \cdots x_{n-1}^k x_n^k = x_0^k \tag{1}$$

have been studied a lot. A simple example would be $(ab)^k a^k (ba)^k = (ababa)^k$, which holds for $k \in \{1, 2\}$, but not for any larger k. The main question about these equations has been that if (1) holds for three different positive values of k, then do the words $x_0, \dots, x_n$ necessarily commute. The first article about this topic was the paper of Hakala and Kortelainen [6], the case where all three values of k are greater than one was solved by Holub [9], and the question was completely settled in an article by Saarela [18], where the following theorem was proved.

Theorem 1.5. Let $n \geq 1$ and $x_0, \dots, x_n \in \Gamma^*$. If $x_1^k \cdots x_n^k = x_0^k$ for three positive integers k, then the words $x_0, \dots, x_n$ commute.

We can generalize (1) by adding “midwords” between the powers. The following result was proved by Holub and Kortelainen [10]. It is an open question whether it would be sufficient to assume that the equation holds for three integers $k \geq 1$.

Theorem 1.6. Let $n \geq 1$ and $s_i, t_i, x_i \in \Gamma^*$ for all i. If

$$s_0 x_1^k s_1 \cdots x_n^k s_n = t_0 x_0^k t_1$$

holds for three consecutive integers $k \geq 2$, then it holds for all k.

We can also allow both sides of the equation to have several powers. The first article about this topic was the paper of Kortelainen [12]. The following theorem was proved by Saarela [17] and it is a slight improvement of a result in [12]. It is an open question whether the bound $m+n$ can be replaced by a constant that does not depend on m and n.

Theorem 1.7. Let $m, n \geq 1$ and $s_i, t_i, x_i, y_i \in \Gamma^*$ for all i. If

$$s_0 x_1^k s_1 \dots x_m^k s_m = t_0 y_1^k t_1 \dots y_n^k t_n. \tag{2}$$

holds for $m + n$ values of k, then it holds for all k.

We will conclude this section by mentioning a very significant open problem on word equations (see the next section for formal definitions): What is the maximal size of an independent system of word equations on n variables? By Ehrenfeucht's compactness property, proved by Albert and Lawrence [1] and by Guba [5], an independent system cannot be infinite. Karhumäki and Plandowski gave examples of independent systems of size $\Theta(n^4)$ [11]. In the case of three variables, which is the simplest nontrivial case, the maximal size of an independent system has been conjectured to be three by Culik and Karhumäki [3], and proved to be at most logarithmic with respect to the size of the shortest equation in the system by Nowotka and Saarela [15].

This general problem about sizes of independent systems is connected to the specific equations we have discussed in several ways: First, as pointed out by Plandowski [16], if Theorem 1.5 had turned out to be false, then the above-mentioned lower bound $\Theta(n^4)$ could have been replaced by a larger lower bound $\Theta(n^5)$, and in principle it might still be possible to use the equations (2) to construct large independent systems. Second, in [17], the same algebraic techniques were used both to prove Theorem 1.7 and to analyze independent systems.

In this article, we will present the ideas behind the proof of Theorem 1.5 (although we will not go through the whole proof), and we will use similar ideas to give new proofs for Theorems 1.1, 1.2 and 1.3.

2. Preliminaries

The empty word is denoted by ε. A nonempty word is *primitive* if it is not a power of a shorter word. Every nonempty word w can be uniquely written in the form p^n where p is primitive. Then p is called the *primitive root* of w.

We say that words $x_0, \dots, x_n$ *commute* if $x_i x_j = x_j x_i$ for all $i, j \in \{0, \dots, n\}$. It is well-known that words $x_0, \dots, x_n \in \Gamma^*$ commute if and only if there exists $p \in \Gamma^*$ such that $x_0, \dots, x_n \in p^*$. If $x_0, \dots, x_n$ are nonempty, then this is equivalent to $x_0, \dots, x_n$ having the same primitive root.

The length of a word w is denoted by $|w|$ and the number of occurrences of a letter a in w is denoted by $|w|_a$. Let $\Gamma = \{a_1, \dots, a_n\}$. If we fix an order relation $<$ for the letters and $a_1 < \dots < a_n$, then the *Parikh vector* of $w \in \Gamma^*$ can be defined as $(|w|_{a_1}, \dots, |w|_{a_n})$.

Let Ξ be an alphabet of variables. A *word equation* is a pair $(U, V) \in \Xi^* \times \Xi^*$, and the *solutions* of this word equation over an alphabet Γ are the morphisms $h : \Xi^* \to \Gamma^*$ such that $h(U) = h(V)$. The set of all solutions of an equation E is denoted by $\mathrm{Sol}_\Gamma(E)$. Usually Γ is clear from context and we can use the notation $\mathrm{Sol}(E)$.

The word equations defined above are *constant-free*. Also word equations with constants could be defined, but in this article, we consider only constant-free equations.

A morphism $h : \Xi^* \to \Gamma^*$ is *periodic* if there exists $p \in \Gamma^*$ such that $h(X) \in p^*$ for all $X \in \Xi$. Periodic solutions of word equations are rather trivial, so usually we are only interested in nonperiodic solutions. It is well-known that if $\Xi = \{X, Y\}$ and $U \neq V$, then the equation (U, V) has only periodic solutions.

Example 2.1. Let $\Xi = \{X, Y, Z\}$. The nonperiodic solutions of the equation (XYZ, ZYX) are exactly the morphisms h defined by

$$h(X) = (pq)^i p,\ h(Y) = (qp)^j q,\ h(Z) = (pq)^k p,$$

where $p, q \in \Gamma^*$, $pq \neq qp$, and $i, j, k \geq 0$. Every periodic morphism is a solution.

We can define boolean combinations of word equations in a natural way. Then we have, for example,

$$\begin{aligned} \mathrm{Sol}(E_1 \wedge E_2) &= \mathrm{Sol}(E_1) \cap \mathrm{Sol}(E_2), \\ \mathrm{Sol}(E_1 \vee E_2) &= \mathrm{Sol}(E_1) \cup \mathrm{Sol}(E_2), \\ \mathrm{Sol}(\neg E) &= \mathrm{Sol}(E)^{\complement}, \end{aligned}$$

where the complement is with respect to the set of all morphisms from Ξ^* to Γ^*. A conjunction $E_1 \wedge \dots \wedge E_n$ can also be denoted by $E_1, \dots, E_n$ and called a *system of equations*, or a *pair of equations* in the case $n = 2$.

The property of being nonperiodic can be encoded as a boolean combination of equations: A morphism $h : \Xi^* \to \Gamma^*$ is nonperiodic if and only if it is a solution of the boolean combination

$$\neg \bigwedge_{X, Y \in \Xi} (XY, YX).$$

A system of equations is *independent*, if it is not equivalent to any of its proper subsystems. A system $E_1, \dots, E_n$ is independent if and only if for every i there exists a morphism h such that $h \in \mathrm{Sol}(E_j)$ for all $j \neq i$ but $h \notin \mathrm{Sol}(E_i)$.

3. Sums of words

We assume that our alphabet Γ is a subset of $\mathbb{R}$ (this is not a restriction; we can assign numerical values to the letters in any way we like, as long as no two letters get the same value). Then we can define the *sum* of a word $w \in \Gamma^*$, denoted by $\Sigma(w)$, to be the sum of its letters, that is, if $w = a_1 \cdots a_n$ and $a_1, \dots, a_n \in \Gamma$, then $\Sigma(w) = a_1 + \dots + a_n$. Words w such that $\Sigma(w) = 0$ are called *zero-sum words*.

If $\Gamma = \{b_1, \dots, b_m\}$, then the sum of $w \in \Gamma^*$ is also the inner product of the vector $(b_1, \dots, b_m)$ and the Parikh vector $(|w|_{b_1}, \dots, |w|_{b_m})$. Zero-sum words are then exactly those words whose Parikh vectors are orthogonal to $(b_1, \dots, b_m)$.

The notation $a_1 \cdots a_n$ of course means the word consisting of the letters $a_1, \dots, a_n$ and not a product of numbers. If $w_1, \dots, w_n$ are words, we can use the notation

$$\prod_{i=1}^{n} w_i = w_1 \cdots w_n$$

for their concatenation.

Let $a_1, \dots, a_n \in \Gamma$. The *prefix sum word* of $w = a_1 \cdots a_n$ is the word $\mathrm{psw}(w) = b_1 \cdots b_n$, where $b_i = \Sigma(a_1 \cdots a_i)$ for all i. Usually, $\mathrm{psw}(w)$ is not a word over Γ, but over some other alphabet. We can give a simple formula for the prefix sum word of a product by using the notation $\mathrm{psw}_r(w) = c_1 \cdots c_n$, where $r \in \mathbb{R}$ and $c_i = b_i + r$ for all i. Then, for $w_1, \dots, w_k \in \Gamma^*$,

$$\mathrm{psw}(w_1 \cdots w_k) = \prod_{i=1}^{k} \mathrm{psw}_{\Sigma(w_1 \cdots w_{i-1})}(w_i).$$

If $w_1, \dots, w_{k-1}$ are zero-sum, then we have the simpler formula

$$\mathrm{psw}(w_1 \cdots w_k) = \prod_{i=1}^{k} \mathrm{psw}(w_i).$$

For the kth power of a word w, we get the formula

$$\mathrm{psw}(w^k) = \prod_{i=1}^{k} \mathrm{psw}_{(i-1)\Sigma(w)}(w).$$

If w is zero-sum, then we have $\mathrm{psw}(w^k) = \mathrm{psw}(w)^k$.

Because letters are real numbers, there is a natural order relation for them. The largest and smallest letters in a word w can be denoted by $\max(w)$ and $\min(w)$, respectively. Usually, max and min are used together with prefix sum words. Note that

$$\max(\mathrm{psw}_r(w)) = \max(\mathrm{psw}(w)) + r, \quad \min(\mathrm{psw}_r(w)) = \min(\mathrm{psw}(w)) + r.$$

The above definitions have the following graphical interpretation: Let $w = a_1 \cdots a_n$. The word w can be represented by a plane curve (more specifically, a polygonal chain) by starting at the origin, moving a_1 steps up and one step to the right, a_2 steps up and one step to the right, and so on. If $\mathrm{psw}(w) = b_1 \cdots b_n$, this curve is also obtained by connecting the points $(0,0), (1, b_1), \dots, (n, b_n)$. The properties of this curve can be studied, leading to a graphical way of analyzing words. See Figure 1 for an example.

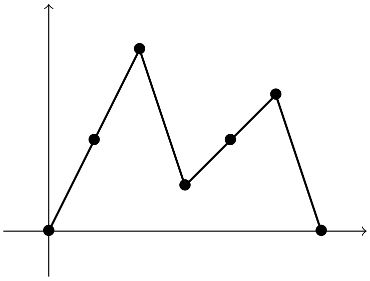

Figure 1. Graphical representation of the word $w = bbcaac$, where $a = 1$, $b = 2$, and $c = -3$. We have $|w| = 6$, $\Sigma(w) = 0$, and $\mathrm{psw}(w) = 241230$.

The last point on the curve of w is $(|w|, \Sigma(w))$. If we start counting from the point $(1, b_1)$ instead of $(0, 0)$, then the biggest y-coordinate is $\max(\mathrm{psw}(w))$ and the smallest y-coordinate is $\min(\mathrm{psw}(w))$. If we use $\mathrm{psw}_r(w)$ instead of $\mathrm{psw}(w)$, then we get a similar curve starting at the point $(0, r)$ instead of $(0, 0)$. The curve of uv consists of the curve of u followed by the curve of v translated in such a way that its starting point matches the endpoint of the curve of u, see Figure 2.

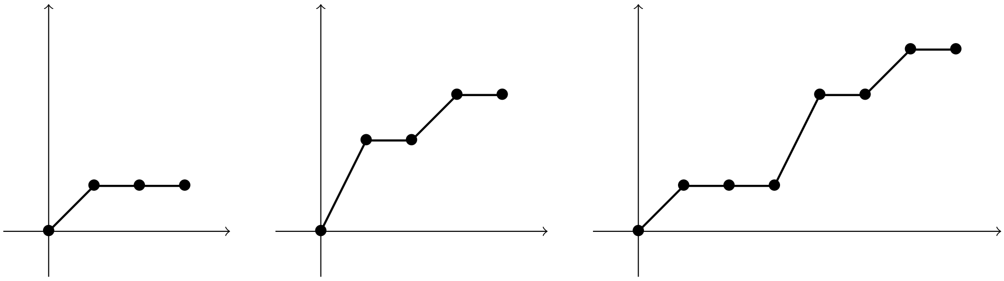

Figure 2. Graphical representation of the words $u = 100$, $v = 2010$ and $uv = 1002010$.

Theorem 1.5 was proved in [18] by using the definitions given in this section. The idea in the proof can be summarized as follows: The alphabet can be normalized so that x_0 becomes a zero-sum word. If all x_i are zero-sum, we can compress them by writing them as products of minimal zero-sum words. After compression and normalization (possibly repeated several times), either the words x_i are unary, which is a trivial case, or x_0 is zero-sum but at least one x_i is not, in which case we can analyze the curves of x_0^k and $x_1^k \cdots x_n^k$. Specifically, we can look at the highest points on the curves, or the points at a certain other height, and count how many times they appear. In this way, we can see that the curves can be equal for at most two values of k. This leads to a proof of Theorem 1.5.

4. Tools

When studying words from a combinatorial point of view, the choice of the alphabet is arbitrary (except for the size of the alphabet). Therefore, we can always change the numerical values of the letters. It was proved in [18] that, given any word u, the alphabet can be normalized so that u becomes a zero-sum word. In Lemma 4.1, we give an improved version of this result.

Before stating and proving the lemma, let us consider the following question: If the alphabet is normalized so that u becomes zero-sum, then which other words become zero-sum? Certainly, if two words have the same primitive root, then either both or neither are zero-sum. More generally, if the Parikh vectors of two nonempty words are linearly dependent, that is, if the vectors are scalar multiples of each other, then either both or neither of the words are zero-sum. If the Parikh vectors are linearly independent, then both words can still be zero-sum in some cases, but this can be avoided if the normalization is done in a suitable way. This is proved in the next lemma. We also prove that we can assume that all letters are integers; usually, this is not important, but in some cases it might be convenient.

Lemma 4.1. Let $n \geq 0$ and $u, v_1, \dots, v_n \in \Gamma^*$. For every $i \in \{1, \dots, n\}$, we assume that the Parikh vector of v_i is not a scalar multiple of the Parikh vector of u. There exists an alphabet $\Delta \subset \mathbb{Z}$ and an isomorphism $g : \Gamma^* \to \Delta^*$ such that $\Sigma(g(u)) = 0$ and $\Sigma(g(v_i)) \neq 0$ for all $i \in \{1, \dots, n\}$.

Proof:
We will give an explicit construction of g. The claim could also be proved by an abstract linear algebraic argument.

The case $u = \varepsilon$ is trivial, so let $u \neq \varepsilon$. Let $\Gamma = \{a_0, \dots, a_m\}$ and let a_m appear in u. Let $B = 1 + |u| \max\{1, |v_1|, \dots, |v_n|\}$. Let

$$b_i = |u|_{a_m} B^i \text{ for } i < m, \qquad b_m = -\sum_{i=0}^{m-1} |u|_{a_i} B^i.$$

Let $\Delta = \{b_0, \dots, b_m\} \subset \mathbb{Z}$ be an alphabet and let $g : \Gamma^* \to \Delta^*$ be the morphism defined by $g(a_i) = b_i$ for all i. Then g is injective and therefore an isomorphism, and $\Sigma(g(u)) = 0$.

To complete the proof, we assume that there exists $j \in \{1, \dots, n\}$ such that $\Sigma(g(v_j)) = 0$ and derive a contradiction. The equality $\Sigma(g(v_j)) = 0$ can be written as

$$\sum_{i=0}^{m-1} |v_j|_{a_i} |u|_{a_m} B^i - |v_j|_{a_m} \sum_{i=0}^{m-1} |u|_{a_i} B^i = 0.$$

from which it follows that

$$\sum_{i=0}^{m-1} |v_j|_{a_i} |u|_{a_m} B^i = \sum_{i=0}^{m-1} |v_j|_{a_m} |u|_{a_i} B^i.$$

The coefficients $|v_j|_{a_i} |u|_{a_m}$ and $|v_j|_{a_m} |u|_{a_i}$ are less than B, so we have B-ary expansions of nonnegative integers on both sides. By the uniqueness of B-ary expansions, $|v_j|_{a_i} |u|_{a_m} = |v_j|_{a_m} |u|_{a_i}$ for all $i \in \{0, \dots, m-1\}$, and trivially also for $i = m$. Therefore $|u|_{a_m}(|v_j|_{a_1}, \dots, |v_j|_{a_m}) = |v_j|_{a_m}(|u|_{a_1}, \dots, |u|_{a_m})$, which contradicts the assumption about the Parikh vectors. □

When using Lemma 4.1, we usually do not care about the words v_i, that is, we can choose $n = 0$.

The next lemma was proved in [18]. For completeness, we repeat the proof here. The lemma claims that every zero-sum word can be written as a product of minimal zero-sum words in a unique way. We can use this to "compress" zero-sum words by replacing these factors by letters.

Lemma 4.2. The set of zero-sum words over Γ is a free monoid.

Proof:
Clearly, zero-sum words form a monoid. This monoid is right unitary, that is, if u and uv are zero-sum, then so is v. It is well-known that a right unitary submonoid of a free monoid is free. □

The next lemma generalizes an idea that was used in [18].

Lemma 4.3. If $h : \Xi^* \to \Gamma^*$ is a length-minimal nonperiodic solution of a boolean combination of word equations over Ξ, then $\Sigma(h(X)) \neq 0$ for at least one variable $X \in \Xi$.

Proof:
Let $h : \Xi^* \to \Gamma^*$ be a nonperiodic solution of the boolean combination of word equations such that $\Sigma(h(X)) = 0$ for all $X \in \Xi$. We will prove that h is not length-minimal.

By Lemma 4.2, we can let Z' be the basis of the free monoid of zero-sum words over Γ. It has a finite subset Z such that $h(X) \in Z^*$ for all $X \in \Xi$. Let Δ be an alphabet and $g : Z^* \to \Delta^*$ an isomorphism. The morphism $g \circ h$ satisfies exactly the same equations as h because g is an isomorphism, and therefore $g \circ h$ is a nonperiodic solution of the boolean combination of word equations.

It remains to be shown that $|(g \circ h)(X)| < |h(X)|$ for at least one $X \in \Xi$. Because h is nonperiodic, there exists a variable X and a nonzero letter a appearing in $h(X)$. There are words $z_1, \dots, z_m \in Z$ such that $h(X) = z_1 \cdots z_m$. Then $g(z_i) \in \Delta$ for all i. The words z_i cannot be empty, and at least one of them contains the nonzero letter a. This means that at least one of them has length at least 2. Thus $|(g \circ h)(X)| = m < |z_1| + \cdots + |z_m| = |h(X)|$. This completes the proof. □

When studying the nonperiodic solutions of a boolean combination of word equations (usually a single equation or a system of equations), we can use Lemmas 4.1 and 4.3 together to assume that there is a nonperiodic solution that maps one freely-chosen variable (or a product of variables) to a zero-sum word, but does not map all variables to zero-sum words. Often we can deduce that the sum of the image of some particular variable X must be nonzero. Then we can also assume that the sum of the image of X is positive, simply by multiplying every letter by -1 if necessary.

5. New proofs

In this section, we give new proofs for Theorems 1.1, 1.2 and 1.3 using prefix sum words and the ideas of Section 3 and the tools of Section 4.

The following theorem is a reformulation of Theorem 1.1 using the formalism of word equations.

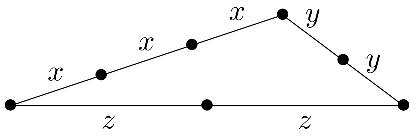

Figure 3. This figure illustrates the proof of Theorem 5.1 in the case $k = 3$, $l = m = 2$, $\Sigma(y) < \Sigma(z) = 0 < \Sigma(x)$. Here we have simplified representations of the curves of x^3y^2 and z^2, where the curves of x, y and z are represented by straight lines, even though they are probably more complicated.

Theorem 5.1. Let X, Y, Z be variables and let $k, l, m \geq 2$. The equation (X^kY^l, Z^m) does not have a nonperiodic solution.

Proof:
We assume that the equation has a nonperiodic solution and derive a contradiction. By Lemmas 4.1 and 4.3, the equation has a solution h such that $\Sigma(h(Z)) = 0$ and at least one of $h(X), h(Y)$ is not zero-sum. Let $h(X) = x$, $h(Y) = y$ and $h(Z) = z$. Because $k\Sigma(x) + l\Sigma(y) = m\Sigma(z) = 0$, one of $\Sigma(x)$ and $\Sigma(y)$ must be positive and the other one negative. We can assume that $\Sigma(y) < 0 < \Sigma(x)$.

Let us first give an informal description of the proof in a geometric way (see Figure 3). The idea is to look at the highest points of the curves of x^ky^l and z^m. The curve of x^ky^l consists of k translated copies of the curve of x, followed by l translated copies of the curve of y. Because $\Sigma(x) > 0$, the last copy of the curve of x is the highest one, and because $\Sigma(y) < 0$, the first copy of the curve of y is the highest one. Therefore, the highest point can only appear within a part of length $|xy|$ in the middle. On the other hand, the curve of z^m consists of m translated copies of the curve of z. Because $\Sigma(z) = 0$, all copies are at the same level. Therefore, the highest point appears within each copy at corresponding positions, so the distance of the first and the last occurrence is at least $|z^{m-1}|$. This means that the curves of x^ky^l and z^m cannot be the same, because $|xy| \leq |z^{m-1}|$.

The proof can be written formally using prefix sum words. We have

$$\prod_{i=0}^{k-1} \mathrm{psw}_{i\Sigma(x)}(x) \prod_{j=0}^{l-1} \mathrm{psw}_{k\Sigma(x)+j\Sigma(y)}(y) = \mathrm{psw}(x^ky^l) = \mathrm{psw}(z^m) = \mathrm{psw}(z)^m. \tag{3}$$

Let $a = \max(\mathrm{psw}(x^ky^l)) = \max(\mathrm{psw}(z^m))$. For all $i < k-1$ and $j > 0$, we have

$$\max(\mathrm{psw}_{i\Sigma(x)}(x)) < \max(\mathrm{psw}_{(k-1)\Sigma(x)}(x)) \leq a,$$
$$\max(\mathrm{psw}_{k\Sigma(x)+j\Sigma(y)}(y)) < \max(\mathrm{psw}_{k\Sigma(x)}(y)) \leq a,$$

so the letter a can appear on the left-hand side of (3) only within the factor

$$\mathrm{psw}_{(k-1)\Sigma(x)}(x)\,\mathrm{psw}_{k\Sigma(x)}(y)$$

of length $|xy| \leq |x^ky^l|/2 = |z^m|/2$. On the other hand, a must appear in $\mathrm{psw}(z)$, so we can write $\mathrm{psw}(z) = paq$ for some words p, q. Then the right-hand side of (3) has a factor $aq\,\mathrm{psw}(z)^{m-2}pa$ of length $|z|^{m-1} + 1 > |z^m|/2$. This is a contradiction. □

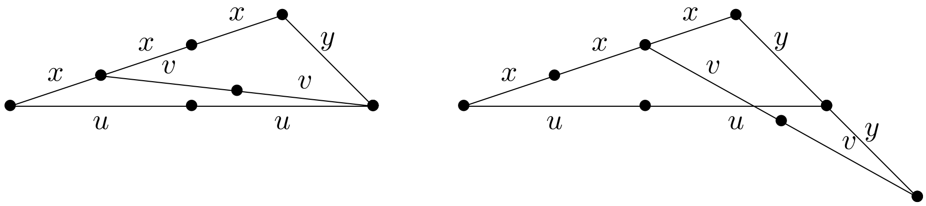

Figure 4. This figure illustrates the proof of Theorem 5.2 in the case where $\Sigma(v), \Sigma(y) < \Sigma(u) = 0 < \Sigma(x)$ and either $k = 3$, $l = l' = 1$, $k' = m = n = 2$ (on the left) or $k = 3$, $l = k' = 1$, $l' = m = n = 2$ (on the right). Here we have simplified representations of the curves of x^3y, xv^2 and u^2 (on the left), and the curves of x^3y^2, x^2v^2 and u^2y (on the right).

The following theorem together with Theorem 5.1 gives a reformulation of Theorem 1.2. The proof is quite similar to the proof of Theorem 5.1.

Theorem 5.2. Let X, Y, U, V be variables. Let $k, l, k', l' \geq 1$, $m, n \geq 2$, and $(k, l) \neq (k', l')$. The pair of equations (X^kY^l, U^m), $(X^{k'}Y^{l'}, V^n)$ does not have a nonperiodic solution.

Proof:
We assume that the pair of equations has a nonperiodic solution and derive a contradiction. By Theorem 5.1, one of k, l is 1 and one of k', l' is 1. Let h be a length-minimal nonperiodic solution and let $h(X) = x$, $h(Y) = y$, $h(U) = u$ and $h(V) = v$. By symmetry, we can assume that $|u^m| \geq |v^n|$ and $l = 1$. Then $l = 1 \leq l'$ and $|x^ky^l| = |u^m| \geq |v^n| = |x^{k'}y^{l'}|$, so it must be $k > k'$. By Lemmas 4.1 and 4.3, we can assume that $\Sigma(u) = 0$ and $\Sigma(y) < 0 < \Sigma(x)$. It follows that $\Sigma(v) < 0$.

We have $u^my^{l'-l} = x^ky^{l'} = x^{k-k'}v^n$. The geometric idea is to conclude that the highest point on the curve of $x^{k-k'}v^n$ can only appear within the first copy of the curve of v, but on the other hand, the highest point of the curve of $u^my^{l'-l}$ must appear within each copy of the curve of u, and this leads to a contradiction (see Figure 4).

Let $a = \max(\mathrm{psw}(x^ky^{l'}))$. We have

$$\mathrm{psw}(x^ky^{l'}) = \prod_{i=0}^{k-1} \mathrm{psw}_{i\Sigma(x)}(x) \prod_{i=0}^{l'-1} \mathrm{psw}_{k\Sigma(x)+i\Sigma(y)}(y)$$

and thus $\max(\mathrm{psw}_{(k-1)\Sigma(x)}(x)) \leq a$ and $\max(\mathrm{psw}_{k\Sigma(x)}(y)) \leq a$.

We have

$$\mathrm{psw}(x^{k-k'}v^n) = \prod_{i=0}^{k-k'-1} \mathrm{psw}_{i\Sigma(x)}(x) \prod_{i=0}^{n-1} \mathrm{psw}_{(k-k')\Sigma(x)+i\Sigma(v)}(v),$$

and a can only appear here within the factor $\mathrm{psw}_{(k-k')\Sigma(x)}(v)$ of length $|v| \leq |v^n|/2 \leq |u^m|/2$.

We have

$$\mathrm{psw}(u^my^{l'-l}) = \mathrm{psw}(u)^m \prod_{i=0}^{l'-l-1} \mathrm{psw}_{i\Sigma(y)}(y), \tag{4}$$

and a can only appear here within the factor $\operatorname{psw}(u)^m$, so we can write $\operatorname{psw}(u) = paq$ for some words p, q. Then (4) has a factor $aq\operatorname{psw}(u)^{m-2}pa$ of length $|u|^{m-1} + 1 > |u^m|/2$. This is a contradiction. □

The following theorem together with Theorem 5.2 gives a reformulation of Theorem 1.3. The proof is a bit more complicated than the proofs of Theorems 5.1 and 5.2.

Theorem 5.3. Let X, Y be variables and let $m \geq 2$. Let $W \in \{X, Y\}^*$ be primitive and let $|W|_X, |W|_Y \geq 2$. The equation (W, Z^m) does not have a nonperiodic solution.

Proof:
We assume that the equation has a nonperiodic solution and derive a contradiction. First, we note that $W \notin (XY)^* \cup (YX)^*$ because it is primitive and contains at least two occurrences of X, and therefore either $W \in (XY)^*X \cup (YX)^*Y$ or W contains XX or YY as a factor. In any case, W has a conjugate that begins and ends with the same variable. By symmetry between X and Y, we can assume that W has a conjugate W' that begins and ends with X. If (W, Z^m) has a nonperiodic solution, then so does (W', Z^m). Let $W' = W_1 \cdots W_n$, where $W_1, \ldots, W_n \in \{X, Y\}$. By Lemmas 4.1 and 4.3, the equation (W', Z^m) has a solution h such that

$$\Sigma(h(Y)) < \Sigma(h(Z)) = 0 < \Sigma(h(X)).$$

Let $h(X) = x$, $h(Y) = y$, $h(Z) = z$ and $h(W_i) = w_i$ for all i.

Let $s_i = \Sigma(w_1 \cdots w_{i-1})$ for $i \in \{1, \ldots, n+1\}$. Let

$$I = \{i \mid s_i = \min\{s_1, \ldots, s_{n+1}\}\} \quad \text{and} \quad J = \{i \mid s_i = \max\{s_1, \ldots, s_{n+1}\}\}.$$

We have

$$s_n = \Sigma(w_1 \cdots w_n) - \Sigma(w_n) = -\Sigma(x) < 0 = s_1 = s_{n+1} < \Sigma(x) = \Sigma(w_1) = s_2,$$

so $1, n+1 \notin I \cup J$. Note that if $i \in J$, then $w_i = y$ (because otherwise $s_{i+1} = s_i + \Sigma(x) > s_i$) and $w_{i-1} = x$ (because otherwise $s_{i-1} = s_i - \Sigma(y) > s_i$).

If $i \in J$, $i' \notin J$ and $W_{i'} = Y$, then $W_i = Y$, $s_{i'} < s_i$, $s_{i'+1} < s_{i+1}$ and $i + 1 \notin I$. This means that if there exists $i \in J$ such that $i + 1 \in I$, then $s_j = s_i$ for every j such that $W_j = Y$, and thus between any two consecutive occurrences of Y in W', there are $-\Sigma(y)/\Sigma(x)$ occurrences of X. This would mean that $W' \in (X^k Y X^l)^*$ for some k, l, which contradicts the primitivity of W'. Therefore, there does not exist $i \in J$ such that $i + 1 \in I$. Similarly, we can show that there does not exist $i \in I$ such that $i + 1 \in J$.

There exists an index j and words p, q such that $z = w_1 \cdots w_{j-1}p$ and $z^{m-1} = qw_{j+1} \cdots w_n$. By the previous paragraph, $j, j+1 \notin J$ or $j, j+1 \notin I$. For the rest of the proof, we assume that $j, j+1 \notin J$. The case $j, j+1 \notin I$ can be proved in a similar way.

Let $a = \max(\operatorname{psw}(w_1 \cdots w_n)) = \max(\operatorname{psw}(z^m))$. Note that if $w_i = y$, then a can appear in $\operatorname{psw}_{s_i}(w_i)$ only if $i \in J$, and if $w_i = x$, then a can appear in $\operatorname{psw}_{s_i}(w_i)$ only if $i + 1 \in J$. We have

$$\operatorname{psw}(w_1 \cdots w_n) = \prod_{i=1}^{n} \operatorname{psw}_{s_i}(w_i),$$

and a can only appear here within the factors

$$\mathrm{psw}_{s_{i-1}}(w_{i-1})\,\mathrm{psw}_{s_i}(w_i) = \mathrm{psw}_{s_{i-1}}(x)\,\mathrm{psw}_{s_i}(y)$$

for $i \in J$. In particular, a does not appear in $\mathrm{psw}_{s_j}(w_j)$. There exists a word u not containing a such that ua is a prefix of $\mathrm{psw}_{s_{i-1}}(x)\,\mathrm{psw}_{s_i}(y)$ for all $i \in J$.

The word $\mathrm{psw}(z) = \mathrm{psw}(w_1 \cdots w_{j-1}p)$ must contain a, so $J \cap \{2, \dots, j-1\} \neq \varnothing$ and we can let $k = \min(J \cap \{2, \dots, j-1\})$. The word $\mathrm{psw}(z^{m-1}) = \mathrm{psw}(qw_{j+1} \cdots w_n)$ must contain a, so $J \cap \{j+2, \dots, n\} \neq \varnothing$ and we can let $l = \min(J \cap \{j+2, \dots, n\})$. Then the shortest prefix of $\mathrm{psw}(z)$ that contains a is $\mathrm{psw}(w_1 \cdots w_{k-2})ua$, and the shortest prefix of $\mathrm{psw}(z^{m-1})$ that contains a is $\mathrm{psw}(qw_{j+1} \cdots w_{l-2})ua$. Therefore, $|w_1 \cdots w_{k-2}| = |qw_{j+1} \cdots w_{l-2}|$ and thus

$$\begin{aligned} |w_{k-1} \cdots w_{l-2}| &= |w_1 \cdots w_{l-2}| - |w_1 \cdots w_{k-2}| \\ &= |w_1 \cdots w_{l-2}| - |qw_{j+1} \cdots w_{l-2}| = |w_1 \cdots w_{j-1}p| = |z|. \end{aligned}$$

This means that there exists a conjugate z' of z such that

$$w_{k-1} \cdots w_{l-2} = z' \quad \text{and} \quad w_{l-1} \cdots w_n w_1 \cdots w_{k-2} = (z')^{m-1}.$$

Let

$$U = W_{k-1} \cdots W_{l-2} \quad \text{and} \quad V = W_{l-1} \cdots W_n W_1 \cdots W_{k-2}.$$

Then $h(U), h(V) \in (z')^*$, so h is a solution of the two-variable equation (UV, VU). If $UV \neq VU$, then this means that x and y commute, which is a contradiction. If $UV = VU$, then U and V are powers of a common word, UV is not primitive, and neither is W because it is a conjugate of UV. This is also a contradiction. □

6. Conclusion

In this article, we have considered word equations. We have mentioned some results and problems that we think are interesting. We have also presented and improved recent methods and tools that we believe will be useful in many situations in the future. Finally, we have given new proofs that we hope are illustrative, both in the sense that they illustrate why the results are true, and in the sense that they illustrate the general proof techniques.

References

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

[2] Barbin-Le Rest E, Le Rest M. Sur la combinatoire des codes à deux mots, *Theoret. Comput. Sci.*, 1985;**41**(1):61–80. URL `https://doi.org/10.1016/0304-3975(85)90060-X`.

[3] Culik IIK, Karhumäki J. Systems of equations over a free monoid and Ehrenfeucht's conjecture, *Discrete Math.*, 1983:**43**(2-3):139–153. doi:10.1016/0012-365X(83)90152-8.

[4] Dömösi P, Horváth G, Vuillon L. On the Shyr-Yu theorem, *Theoret. Comput. Sci.*, 2009;**410**(47-49):4874–4877. URL `https://doi.org/10.1016/j.tcs.2009.06.039`.

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

[6] Hakala I, Kortelainen J. On the system of word equations $x_1^i x_2^i \cdots x_m^i = y_1^i y_2^i \cdots y_n^i$ $(i = 1, 2, \cdots)$ in a free monoid, *Acta Inform.*, 1997;**34**(3):217–230. doi:10.1007/s002360050081.

[7] Harju T, Nowotka D. The equation $x^i = y^j z^k$ in a free semigroup, *Semigroup Forum*, 2004;**68**(3):488–490. doi:10.1007/s00233-003-0028-6.

[8] Harju T, Nowotka D. On the equation $x^k = z_1^{k_1} z_2^{k_2} \cdots z_n^{k_n}$ in a free semigroup, *Theoret. Comput. Sci.*, 2005;**330**(1):117–121. URL `https://doi.org/10.1016/j.tcs.2004.09.012`.

[9] Holub Š. Local and global cyclicity in free semigroups, *Theoret. Comput. Sci.*, 2001;**262**(1–2):25–36. URL `https://doi.org/10.1016/S0304-3975(00)00156-0`.

[10] Holub, Š, Kortelainen J. On systems of word equations with simple loop sets, *Theoret. Comput. Sci.*, 2007;**380**(3):363–372. doi:10.1016/j.tcs.2007.03.026.

[11] Karhumäki J, Plandowski W. On the defect effect of many identities in free semigroups, in: *Mathematical aspects of natural and formal languages* (G. Paun, Ed.), World Scientific, 1994, 225–232. ISBN:9-8102-1914-8.

[12] Kortelainen J. On the system of word equations $x_0 u_1^i x_1 u_2^i x_2 \cdots u_m^i x_m = y_0 v_1^i y_1 v_2^i y_2 \cdots v_n^i y_n$ $(i = 0, 1, 2, \cdots)$ in a free monoid, *J. Autom. Lang. Comb.*, 1998;**3**(1):43–57.

[13] Lothaire M. *Algebraic Combinatorics on Words*, Cambridge University Press, 2002. ISBN: 0521812208.

[14] Lyndon RC, Schützenberger M-P. The equation $a^M = b^N c^P$ in a free group, *Michigan Math. J.*, 1962; **9**(4):289–298.

[15] Nowotka D, Saarela A. One-variable word equations and three-variable constant-free word equations, *Internat. J. Found. Comput. Sci.*, To appear.

[16] Plandowski W. Test sets for large families of languages, *Proceedings of the 7th DLT*, 2710, Springer, 2003. doi:10.1007/3-540-45007-6_6.

[17] Saarela A. Systems of word equations, polynomials and linear algebra: A new approach, *European J. Combin.*, 2015;**47**:1–14. URL `https://doi.org/10.1016/j.ejc.2015.01.005`.

[18] Saarela A. Word equations where a power equals a product of powers, *Proceedings of the 34th STACS*, 66, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

[19] Shyr HJ, Yu S-S. Non-primitive words in the language p^+q^+, *Soochow J. Math.*,1994;**20**(4):535–546.

[20] Spehner J-C. *Quelques problémes d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre*, Ph.D. Thesis, Univ. Paris, 1976.