Word Equations Where a Power Equals a Product of Powers

Aleksi Saarela

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

Abstract

We solve a long-standing open problem on word equations by proving that if the words $x_0, \dots, x_n$ satisfy the equation $x_0^k = x_1^k \cdots x_n^k$ for three positive values of k, then the words commute. One of our methods is to assign numerical values for the letters, and then study the sums of the letters of words and their prefixes. We also give a geometric interpretation of our methods.

1998 ACM Subject Classification G.2.1 Combinatorics, F.4.3 Formal Languages

Keywords and phrases Combinatorics on words, Word equations

Digital Object Identifier 10.4230/LIPIcs.STACS.2017.55

1 Introduction

We say that words $x_0, \dots, x_n$ *commute* if $x_i x_j = x_j x_i$ for all $i, j \in \{0, \dots, n\}$. One of the early results on word equations is the result of Lyndon and Schützenberger [13] that if $x^k = y^m z^n$ for some words x, y, z and numbers $k, m, n \geq 2$, then x, y, z commute (or, equivalently, x, y, z are powers of a common word). Many generalizations have been studied, for example by Lentin [11] and by Shyr and Yu [17]. We are interested in the generalizations where the right-hand side can have more than two powers, but all exponents are equal (here $k \geq 1$):

$$x_0^k = x_1^k \cdots x_n^k. \tag{1}$$

In particular, we are interested in systems of equations where the words $x_0, \dots, x_n$ satisfy (1) for many values of k. An even more general family of equations that could be studied is formed by equations of the form

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

Let us briefly mention some connections and applications of the above equations (1) and (2) (more details can be found in the references given). The first application is in the theory of test sets. A subset K of a language L is called a test set if, for all morphisms f and g, either $f(x) \neq g(x)$ for some $x \in K$ or $f(x) = g(x)$ for all $x \in L$. This means that to check whether f and g agree on L, it is sufficient to check whether they agree on K. Connections between the above equations and test sets are explained, for example, in [7]. As a second application, the equations come up when studying pumping properties of formal languages. For example, in the article [3], pumping the computations of transducers in two places leads to equations of the form (2) with $m = n = 2$. The equations naturally arise in many other settings as well. As a third application, equations (1) are related to the construction of large independent systems of word equations [9, 15]. In fact, a counterexample to Conjecture 1 (stated below and proved in this article) would have improved the best known lower bounds for the size of independent systems. Our last example is that the pair of equations (1) for

34th Symposium on Theoretical Aspects of Computer Science (STACS 2017).
Editors: Heribert Vollmer and Brigitte Vallée; Article No. 55; pp. 55:1–55:9
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$k = 1, 2$ is connected to Sturmian words, and this connection leads to a large family of interesting solutions [14].

The main question about equations (1) is when do they imply that $x_0, \dots, x_n$ commute. Sometimes a single equation is enough: Appel and Djorup [1] proved that if $k = n$ in (1), then the words $x_0, \dots, x_n$ must commute. Their result was generalized by Harju and Nowotka [5] for certain equations which have many different exponents $k_1, k_2, \dots$ instead of just one exponent k. On the other hand, there are many examples of words $x_0, \dots, x_n$ such that $x_i x_j \neq x_j x_i$ for some i, j, but (1) holds for two different values of k. For instance, $(ababa)^k = (ab)^k a^k (ba)^k$ for $k \in \{1, 2\}$. No such examples are known for three different values of k. In fact, the following conjecture is well-known.

▶ **Conjecture 1.** *If $x_0, \dots, x_n$ are words and $k_1, k_2, k_3 \geq 1$ are different numbers such that* (1) *holds for $k \in \{k_1, k_2, k_3\}$, then $x_0, \dots, x_n$ commute.*

In some form, this conjecture has been open for at least about two decades. The case $\{k_1, k_2, k_3\} = \{1, 2, 3\}$ appeared as a question (and was proved for $n \leq 5$) in an article by Hakala and Kortelainen [4], and as an explicit conjecture in an article by Plandowski [15], and a prize for a proof was offered by Holub in 2009[1]. The case where one of k_1, k_2, k_3 is 1 was asked as a question in [6], and the case $k_1, k_2, k_3 \geq 2$ was proved in [7] by Holub. If (1) holds for $k = 1$, then the other equations can be replaced by

$$(x_1 \cdots x_n)^k = x_1^k \cdots x_n^k, \tag{3}$$

so the conjecture often appears in the following form: If (3) holds for two values of $k \geq 2$, then $x_1, \dots, x_n$ commute.

For equations (2), we could ask the following question: For how many values of k does (2) need to hold to guarantee that it holds for all $k \geq 0$? This was first studied by Kortelainen [10]. Currently it is known that $m + n$ different values of k are sufficient [16]. Holub and Kortelainen [8] proved that a constant number of different values is sufficient in some special cases, but it is not known whether this is true in general.

In this article, we will prove Conjecture 1. As mentioned above, it would be sufficient to prove it in the case $k_1 = 1$, but our proof works for all values. This also makes the paper self-contained; to understand the proof, it is only necessary to be familiar with some basics of combinatorics on words, see, e.g., [12]. The possibility of applying the techniques of this paper to the more general equations (2) remains open.

The basic idea behind our proof is to assign numerical values for the letters in a specific way (so that the sum of the letters of x_0 is zero), and then study the sums of the letters of words and their prefixes. We will also give a geometric interpretation of our methods; using geometric intuition was crucial when trying to prove the result.

2 Preliminaries

Let Γ be an alphabet. We can assume that Γ is a subset of $\mathbb{R}$. This allows us to define $\Sigma(w)$ to be the sum of the letters of a word $w \in \Gamma^*$, that is, if $w = a_1 \cdots a_n$ and $a_1, \dots, a_n \in \Gamma$, then $\Sigma(w) = a_1 + \dots + a_n$. The mapping Σ is a morphism from the free monoid Γ^* to the additive monoid $\mathbb{R}$. Words w such that $\Sigma(w) = 0$ are called *zero-sum words.*

[1] http://www.karlin.mff.cuni.cz/~holub/soubory/prizeproblem.pdf

The notation $a_1 \cdots a_n$ of course means the word consisting of the letters $a_1, \ldots, a_n$ and not a product of numbers. When we actually want to compute the product of two numbers, it should be clear from context. If $w_1, \ldots, w_n$ are words, we can also use the notation

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

for their concatenation.

Whenever the symbol Γ appears in this article, it is always used to denote an alphabet. Occasionally we will use other alphabets as well. All of them can be assumed to be subsets of $\mathbb{R}$. Alphabets are also assumed to be finite, unless otherwise specified.

Let $a_1, \ldots, a_k \in \Gamma$. The *prefix sum word* of $w = a_1 \cdots a_k$ is the word $\operatorname{psw}(w) = b_1 \cdots b_k$, where $b_i = \Sigma(a_1 \cdots a_i)$ for all i. Of course, $\operatorname{psw}(w)$ is usually not a word over Γ, but over some other alphabet. The word $\operatorname{psw}(w)$ has the same length as w and the last letter is $\Sigma(w)$.

The mapping psw is injective. It is not a morphism, but we can give a simple formula for the prefix sum word of a product by using the notation $\operatorname{psw}_r(w) = c_1 \cdots c_k$, where $r \in \mathbb{R}$ and $c_i = b_i + r$ for all i. Then, for $w_1, \ldots, w_n \in \Gamma^*$,

$$\operatorname{psw}(w_1 \cdots w_n) = \prod_{i=1}^{n} \operatorname{psw}_{\Sigma(w_1 \cdots w_{i-1})}(w_i).$$

If $w_1, \ldots, w_n$ are zero-sum, then we have the simpler formula

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

so in this case the mapping psw actually does behave like a morphism. For the nth power of a word w, we get the formula

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

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

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. The length of w is denoted by $|w|$, and the number of occurrences of a letter a in w is denoted by $|w|_a$. The size of a set S is denoted by $|S|$.

▶ **Example 2.** Let $w = bbcaac$, where $a = 1$, $b = 2$, and $c = -3$. We have $|w| = 6$, $\max(w) = 2$, and $\min(w) = -3$. Because $\Sigma(w) = 2 + 2 - 3 + 1 + 1 - 3 = 0$, w is a zero-sum word. The prefix sum word of w is $\operatorname{psw}(w) = 241230$, and $\max(\operatorname{psw}(w)) = 4$ and $\min(\operatorname{psw}(w)) = 0$.

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 assign numerical values to the letters in any way we like, as long as no two letters get the same value. The next lemma shows in a formal way that, given any word w, the alphabet can be normalized so that w becomes a zero-sum word.

▶ **Lemma 3.** *Let $w \in \Gamma^*$. There exists an alphabet Δ and an isomorphism $h : \Gamma^* \to \Delta^*$ such that $h(w)$ is zero-sum.*

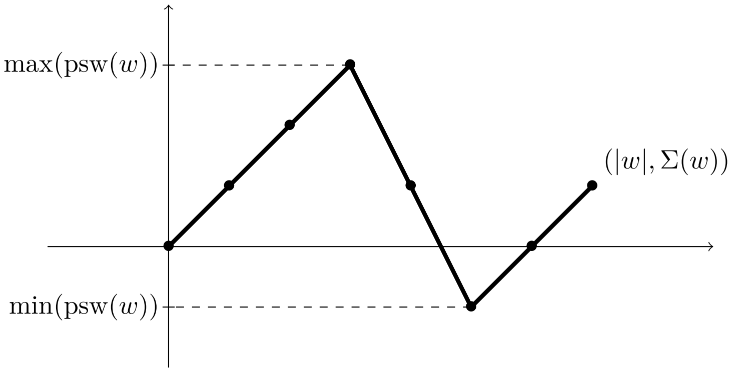

Figure 1 Geometric representation of the word $\text{psw}(w)$, where $w = aaabbaa$, $a = 1$, and $b = -2$. We have $|w| = 7$, $\Sigma(w) = 1$, $\max(\text{psw}(w)) = 3$, and $\min(\text{psw}(w)) = -1$.

Proof. If w is the empty word, it is already zero-sum. Otherwise, let $d = \Sigma(w)/|w|$. We can define an alphabet $\Delta = \{a - d \mid a \in \Gamma\}$ and a morphism $h : \Gamma^* \to \Delta^*$ by $h(a) = a - d$ for all $a \in \Gamma$. Clearly, h is a bijection and therefore an isomorphism, and $\Sigma(h(w)) = \Sigma(w) - d|w| = 0$. ◀

By the next lemma, every zero-sum word can be written as a product of minimal zero-sum words in a unique way. Later we will use this to "compress" zero-sum words by replacing these factors by letters. The free monoid in the lemma can be either trivial (just the empty word) or infinitely generated.

▶ **Lemma 4.** *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 claim could easily be proved directly as well.) ◀

3 Geometric intuition

In this section, we give some geometric intuition, which is not necessary for the proofs, but it might be helpful in understanding them (at least it was helpful in inventing the proofs).

The above definitions have the following geometric interpretation: Let $w = a_1 \cdots a_k$. The word $\text{psw}(w)$ (or the word w depending on the point of view) can be represented by 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 $\text{psw}(w) = b_1 \cdots b_k$, then this curve is also obtained by connecting the points $(0,0), (1,b_1), \dots, (k,b_k)$. The last point is $(|w|, \Sigma(w))$. If we start counting from the point $(1, b_1)$ instead of $(0,0)$, then the biggest y-coordinate is $\max(\text{psw}(w))$ and the smallest y-coordinate is $\min(\text{psw}(w))$. See Figure 1 for an example. The word $\text{psw}_r(w)$ could be represented in a similar way by starting at the point $(0, r)$ instead of $(0,0)$. The curve of $\text{psw}(uv)$ consists of the curve of $\text{psw}(u)$ followed by the curve of $\text{psw}(v)$ translated in such a way that its starting point matches the endpoint of the curve of $\text{psw}(u)$.

The geometric interpretation is similar to the relation between Dyck words and Dyck paths, or the definition of Sturmian words as mechanical words. Representations of words as paths (or paths as words) can also be used in discrete geometry. This can, for example, lead to connections between word equations and tilings of a plane, see [2] for a survey.

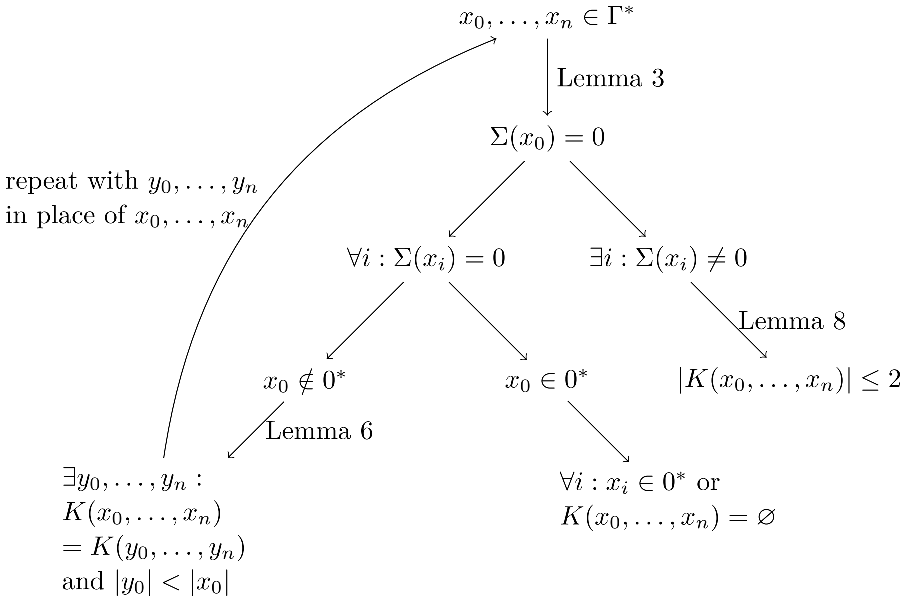
$x_0, \dots, x_n \in \Gamma^*$
Lemma 3
$\Sigma(x_0) = 0$
repeat with $y_0, \dots, y_n$ in place of $x_0, \dots, x_n$
$\forall i : \Sigma(x_i) = 0$
$\exists i : \Sigma(x_i) \neq 0$
Lemma 8
$x_0 \notin 0^*$
$x_0 \in 0^*$
$|K(x_0, \dots, x_n)| \leq 2$
Lemma 6
$\exists y_0, \dots, y_n :$ $K(x_0, \dots, x_n)$ $= K(y_0, \dots, y_n)$ and $|y_0| < |x_0|$
$\forall i : x_i \in 0^*$ or $K(x_0, \dots, x_n) = \varnothing$

Figure 2 Structure of the proof.

4 Idea of the proof

For $x_0, \dots, x_n \in \Gamma^*$, let

$$K(x_0, \dots, x_n) = \{k \in \mathbb{Z}_+ \mid x_0^k = x_1^k \cdots x_n^k\}.$$

We are going to prove that either $x_0, \dots, x_n$ commute or $K(x_0, \dots, x_n)$ contains at most two numbers. This proves Conjecture 1.

Before the formal treatment in Section 5, let us outline the strategy (also illustrated in Figure 2): First we use Lemma 3 to change the alphabet so that x_0 becomes zero-sum. If all x_i are zero-sum, then either x_0 is unary or we can use Lemma 6 to compress them so that the set $K(x_0, \dots, x_n)$ is preserved. The compression can possibly be repeated several times, but only finitely many times. After that we end up either in the unary case, in which case also the original words x_i commute, or in the case where some x_i is not zero-sum. If some x_i is not zero-sum, then we can use Lemma 8 to prove that $K(x_0, \dots, x_n)$ contains at most two numbers; this is the most complicated part of the proof. The idea in Lemma 8 is to compare the number of occurrences of a certain letter in the words $\operatorname{psw}(x_0^k)$ and $\operatorname{psw}(x_1^k \cdots x_n^k)$. If these are different, then $k \notin K(x_0, \dots, x_n)$, because $k \in K(x_0, \dots, x_n)$ is equivalent to $\operatorname{psw}(x_0^k) = \operatorname{psw}(x_1^k \cdots x_n^k)$.

▶ **Example 5.** Consider the case $x_0 = abbaabbaab$, $x_1 = abba$, $x_2 = ab$, and $x_3 = baab$.

Our alphabet is $\{a, b\}$, and we can choose arbitrary numerical values for a and b. We want x_0 to be zero-sum, so let $a = 1$ and $b = -1$. Then also x_1, x_2 and x_3 are zero-sum. We can do the compression, formalized in Lemma 6, by writing the words as products of the zero-sum words ab and ba, and then replacing ab with a letter c and ba with a letter d. We get the words $y_0 = cdcdc$, $y_1 = cd$, $y_2 = c$, and $y_3 = dc$, and $K(x_0, \dots, x_n) = K(y_0, \dots, y_n)$.

Our new alphabet is $\{c, d\}$, and we can choose arbitrary numerical values for c and d. We want y_0 to be zero-sum, so let $c = 2$ and $d = -3$. Then y_1, y_2 and y_3 are not zero-sum. Lemma 8 shows that $|K(y_0, \dots, y_n)| \leq 2$. The idea in Lemma 8 is to compare the number of occurrences of a certain letter in the words $\mathrm{psw}(y_0^k)$ and $\mathrm{psw}(y_1^k \cdots y_n^k)$. In this particular example, it is sufficient to look at the largest letters: We see that $\max(\mathrm{psw}(y_0^k)) = 2$ for all k, but $\max(\mathrm{psw}(y_1^k y_2^k y_3^k)) \geq \Sigma(y_1^k y_2^k) = k$. Thus $y_0^k = y_1^k y_2^k y_3^k$ can hold only for $k \in \{1, 2\}$, and it does hold for these two numbers, so $K(x_0, \ldots, x_n) = \{1, 2\}$.

5 Proof of the conjecture

In this section we will prove our main result, Theorem 9. It is preceded by three lemmas: Lemma 6 and Lemma 8 were already mentioned in Section 4, and Lemma 7 is needed in the proof of Lemma 8.

▶ **Lemma 6.** *Let $x_0, \ldots, x_n \in \Gamma^*$ be zero-sum words. If x_0 is not unary (or equivalently, if x_0 contains a nonzero letter), then there are words $y_0, \ldots, y_n$ such that $|y_0| < |x_0|$, $K(x_0, \ldots, x_n) = K(y_0, \ldots, y_n)$, and $y_0, \ldots, y_n$ commute if and only if $x_0, \ldots, x_n$ commute.*

Proof. By Lemma 4, we can let Z be the basis of the free monoid of zero-sum words over Γ, Δ be an infinite alphabet, and $h : Z^* \to \Delta^*$ be an isomorphism. Let $y_i = h(x_i)$ for all i. Because h is an isomorphism, the words $y_0, \ldots, y_n$ satisfy exactly the same equations as $x_0, \ldots, x_n$. In particular, $y_0^k = y_1^k \cdots y_n^k$ is equivalent to $x_0^k = x_1^k \cdots x_n^k$, and $y_i y_j = y_j y_i$ is equivalent to $x_i x_j = x_j x_i$.

It remains to be shown that $|y_0| < |x_0|$. There are words $z_1, \ldots, z_m \in Z$ such that $x_0 = z_1 \cdots z_m$. Then $h(z_i) \in \Delta$ for all i. The words z_i cannot be empty, and at least one of them contains a nonzero letter, because x_0 is not unary. This means that at least one of them has length at least 2. Thus $|y_0| = m < |z_1| + \cdots + |z_m| = |x_0|$. ◀

In Lemma 8, we will study the words $\mathrm{psw}(x_0^k)$ and $\mathrm{psw}(x_1^k \cdots x_n^k)$. If $s_i = \Sigma(x_1 \cdots x_{i-1})$ for $i \in \{1, \ldots, n\}$, then

$$\mathrm{psw}(x_1^k \cdots x_n^k) = \prod_{i=1}^{n} \mathrm{psw}_{ks_i}(x_i^k),$$

so we will also need to study the words $\mathrm{psw}_{ks_i}(x_i^k)$ that appear in this product. The following technical lemma will be used to analyze these words.

▶ **Lemma 7.** *Let $x \in \Gamma^*$, $s \in \mathbb{R}$, and $\Sigma(x) \neq 0$ or $s \neq 0$. Let $k_1, k_2, k_3 \in \mathbb{Z}_+$ and $k_1 < k_2 < k_3$. Let $w_k = \mathrm{psw}_{ks}(x^k)$. Let $a \geq \max(w_{k_1} w_{k_2} w_{k_3})$. Then $|w_{k_1}|_a \geq |w_{k_2}|_a$.*

Proof. Before the actual proof, let us give the intuitive idea using the geometric concepts in Section 3. The heights of the endpoints of the curve of $\mathrm{psw}_{ks}(x^k)$ are ks and $ks + k\Sigma(x)$. Their positivity or negativity does not depend on k, and at least one of them is nonzero by the assumptions. If one of them is positive, then it becomes larger as k grows, and the highest point on the curve becomes higher as k grows. This means that $\max(w_{k_1} w_{k_2} w_{k_3}) = \max(w_{k_3}) > \max(w_{k_2})$, and thus $|w_{k_2}|_a = 0$. If both of the heights of the endpoints are negative, then they become smaller as k grows, and the highest point on the curve becomes lower as k grows. This means that $\max(w_{k_1} w_{k_2} w_{k_3}) = \max(w_{k_1}) > \max(w_{k_2})$, and thus $|w_{k_2}|_a = 0$. If one of the heights of the endpoints is zero and the other is negative, then the highest point on the curve remains the same as k grows, and also the number of times it occurs remains the same, so $|w_{k_1}|_a = |w_{k_2}|_a$.

Let us now move on to the formal proof. For every $k \in \{k_1, k_2, k_3\}$, we have

$$w_k = \prod_{i=0}^{k-1} \mathrm{psw}_{ks + i\Sigma(x)}(x) \tag{4}$$

and thus

$$\begin{aligned}\max(w_k) &= \max\{ks + i\Sigma(x) \mid 0 \le i < k\} + \max(\operatorname{psw}(x)) \\ &= \begin{cases} ks + \max(\operatorname{psw}(x)) & \text{if } \Sigma(x) \le 0, \\ k(s + \Sigma(x)) - \Sigma(x) + \max(\operatorname{psw}(x)) & \text{if } \Sigma(x) \ge 0. \end{cases}\end{aligned} \tag{5}$$

If $\Sigma(x) \le 0$ and $s \ne 0$ or if $\Sigma(x) \ge 0$ and $s + \Sigma(x) \ne 0$, then (5) is strictly decreasing or strictly increasing with respect to k, and either $\max(w_{k_1}) > \max(w_{k_2})$ or $\max(w_{k_3}) > \max(w_{k_2})$. In this case, $a > \max(w_{k_2})$ and thus $|w_{k_2}|_a = 0$, which proves the claim.

We still need to consider the case $\Sigma(x) < 0$ and $s = 0$, and the case $\Sigma(x) > 0$ and $s + \Sigma(x) = 0$ (if $\Sigma(x) = 0$, then $s = s + \Sigma(x) \ne 0$ by the assumptions). If $\Sigma(x) < 0$ and $s = 0$, then a can appear in the product (4) only in the term corresponding to $i = 0$, which is $\operatorname{psw}(x)$. This does not depend on k, so $|w_{k_1}|_a = |w_{k_2}|_a$. Similarly, if $\Sigma(x) > 0$ and $s + \Sigma(x) = 0$, then a can appear in the product (4) only in the term corresponding to $i = k - 1$, which is $\operatorname{psw}_{-\Sigma(x)}(x)$. This does not depend on k, so $|w_{k_1}|_a = |w_{k_2}|_a$. This completes the proof. ◀

Words u and v are called *abelian equivalent* if $|u|_a = |v|_a$ for every letter a. (We could replace abelian equivalence by equality in the next lemma; then the result would be weaker, but still strong enough for our purposes.)

▶ **Lemma 8.** *Let x_0 be a zero-sum word, and let $x_1, \dots, x_n$ be words not all of which have zero sum. There are at most two positive integers k such that $\operatorname{psw}(x_0^k)$ and $\operatorname{psw}(x_1^k \cdots x_n^k)$ are abelian equivalent.*

Proof. Let $s_1 = 0$ and $s_i = \Sigma(x_1 \cdots x_{i-1})$ for $i \in \{2, \dots, n\}$. Then

$$\operatorname{psw}(x_1^k \cdots x_n^k) = \prod_{i=1}^{n} \operatorname{psw}_{ks_i}(x_i^k). \tag{6}$$

Let

$$I_0 = \{i \in \{1, \dots, n\} \mid s_i = 0 \text{ and } \Sigma(x_i) = 0\} \quad \text{and} \quad I_1 = \{1, \dots, n\} \smallsetminus I_0.$$

The set I_1 is nonempty by the assumptions. We have

$$\operatorname{psw}(x_0^k) = \operatorname{psw}(x_0)^k \quad \text{and} \quad \operatorname{psw}_{ks_i}(x_i^k) = \operatorname{psw}(x_i)^k \text{ for all } i \in I_0. \tag{7}$$

Let $\operatorname{psw}(x_0^k)$ and $\operatorname{psw}(x_1^k \cdots x_n^k)$ be abelian equivalent for $k \in \{k_1, k_2, k_3\}$ and $k_1 < k_2 < k_3$.

Before the actual proof, let us give the intuitive idea by using the geometric concepts in Section 3. To simplify the explanation, we consider only the case $x_0 = x_1 \cdots x_n$ here. Because x_0 is zero-sum, the curve of $\operatorname{psw}(x_0^k)$ consists of k copies of the curve of $\operatorname{psw}(x_0)$, translated horizontally but not vertically. If $i \in I_0$, then the curve of $\operatorname{psw}(x_i)$ appears here k times, also translated horizontally but not vertically, by the definition of the set I_0. Similarly, the curve of $\operatorname{psw}(x_i)$ appears k times in the curve of $\operatorname{psw}(x_1^k \cdots x_n^k)$, again translated horizontally but not vertically. Because we are only interested in abelian equivalence, we can cancel these appearances out (in the formal proof, this corresponds to moving the sum related to I_0 to the left-hand side in (8)). Effectively, this means that we can assume that I_0 is empty. Then, let a be the largest letter appearing in $\operatorname{psw}(x_0^k)$. The letter a does not depend on k, and the number of occurrences grows as k grows. The number of occurrences of a in $\operatorname{psw}(x_1^k \cdots x_n^k)$ must be the same, but this leads to a contradiction with Lemma 7.

Let us now move on to the formal proof. For $k \in \{k_1, k_2, k_3\}$ and every letter a, we have $|\mathrm{psw}(x_0^k)|_a = |\mathrm{psw}(x_1^k \cdots x_n^k)|_a$, so it follows from (6) that

$$|\mathrm{psw}(x_0^k)|_a = \sum_{i=1}^{n} |\mathrm{psw}_{ks_i}(x_i^k)|_a = \sum_{i \in I_0} |\mathrm{psw}_{ks_i}(x_i^k)|_a + \sum_{i \in I_1} |\mathrm{psw}_{ks_i}(x_i^k)|_a.$$

From (7) it then follows that

$$|\mathrm{psw}(x_0)^k|_a = \sum_{i \in I_0} |\mathrm{psw}(x_i)^k|_a + \sum_{i \in I_1} |\mathrm{psw}_{ks_i}(x_i^k)|_a,$$

and, because $|w^k|_a = k|w|_a$ for all words w,

$$k\Big(|\mathrm{psw}(x_0)|_a - \sum_{i \in I_0} |\mathrm{psw}(x_i)|_a\Big) = \sum_{i \in I_1} |\mathrm{psw}_{ks_i}(x_i^k)|_a. \tag{8}$$

Consider the largest letter a which appears in $\mathrm{psw}_{ks_i}(x_i^k)$ for at least one $i \in I_1$ and one $k \in \{k_1, k_2, k_3\}$. Such a letter must exist, because the set I_1 is not empty. The right-hand side of (8) is positive for this a and at least one k, so also the left-hand side must be positive and thus

$$|\mathrm{psw}(x_0)|_a - \sum_{i \in I_0} |\mathrm{psw}(x_i)|_a > 0.$$

But then the left-hand side is positive for all k, and strictly increasing with respect to k. For every $i \in I_1$, we can use Lemma 7 with $x = x_i$ and $s = s_i$. The assumption $\Sigma(x) \neq 0$ or $s \neq 0$ is satisfied because of the definition of I_1, and the assumption $a \geq \max(w_{k_1} w_{k_2} w_{k_3})$ is satisfied by the definition of a. It follows from Lemma 7 that the right-hand side of (8) cannot be larger for k_2 than for k_1. This contradicts the left-hand side being strictly increasing, and this contradiction proves the claim. ◀

Now we can state as a formal theorem and proof what was said at the beginning of Section 4.

▶ **Theorem 9.** *Let $x_0, \ldots, x_n \in \Gamma^*$. If $x_0^k = x_1^k \cdots x_n^k$ for three positive integers k, then the words $x_0, \ldots, x_n$ commute.*

Proof. We assume that $x_i x_j \neq x_j x_i$ for some i, j and prove that $|K(x_0, \ldots, x_n)| \leq 2$. We can assume that x_0 is minimal in the sense that there does not exist words $y_0, \ldots, y_n$ such that $y_i y_j \neq y_j y_i$ for some i, j, $K(y_0, \ldots, y_n) = K(x_0, \ldots, x_n)$, and $|y_0| < |x_0|$. By Lemma 3, we can assume that x_0 is zero-sum.

If $x_0 \in 0^*$, then some $x_i \notin 0^*$ because of the assumption $x_i x_j \neq x_j x_i$, and thus $K(x_0, \ldots, x_n) = \varnothing$. If $x_0 \notin 0^*$ and $x_1, \ldots, x_n$ are zero-sum, then Lemma 6 contradicts the minimality assumption. If at least one of $x_1, \ldots, x_n$ is not zero-sum, then $|K(x_0, \ldots, x_n)| \leq 2$ by Lemma 8. This completes the proof. ◀

6 Conclusion

In this article, we have proved Conjecture 1. One possible direction for further research would be to study equations of the form (2), or some subfamily of these equations (for example, the equations with $m = 1$). Another direction would be to try to apply the methods used in this paper (in particular, Lemma 3, prefix sum words, and the geometric intuition) to some other entirely different problems on word equations. We hope and believe that, in addition to the immediate impact of solving a major open problem, this article will also lead to further advances in the future.

References

1 K. I. Appel and F. M. Djorup. On the equation $z_1{}^n z_2{}^n \cdots z_k{}^n = y^n$ in a free semigroup. *Trans. Amer. Math. Soc.*, 134:461–470, 1968.

2 Srečko Brlek. Interactions between digital geometry and combinatorics on words. In *Proceedings of the 8th WORDS*, volume 63 of *EPTCS*, pages 1–12, 2011. doi:10.4204/EPTCS.63.1.

3 Emmanuel Filiot, Olivier Gauwin, Pierre-Alain Reynier, and Frédéric Servais. Streamability of nested word transductions. In *Proceedings of the 31st FSTTCS*, volume 13 of *LIPIcs*, pages 312–324. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2011.

4 Ismo Hakala and Juha Kortelainen. 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.*, 34(3):217–230, 1997. doi:10.1007/s002360050081.

5 Tero Harju and Dirk Nowotka. 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.*, 330(1):117–121, 2005. doi:10.1016/j.tcs.2004.09.012.

6 Štěpán Holub. A solution of the equation $(x_1^2 \cdots x_n^2)^3 = (x_1^3 \cdots x_n^3)^2$. In *Contributions to general algebra, 11 (Olomouc/Velké Karlovice, 1998)*, pages 105–111. Heyn, 1999.

7 Štěpán Holub. Local and global cyclicity in free semigroups. *Theoret. Comput. Sci.*, 262(1–2):25–36, 2001. doi:10.1016/S0304-3975(00)00156-0.

8 Štěpán Holub and Juha Kortelainen. On systems of word equations with simple loop sets. *Theoret. Comput. Sci.*, 380(3):363–372, 2007. doi:10.1016/j.tcs.2007.03.026.

9 Juhani Karhumäki and Wojciech Plandowski. On the defect effect of many identities in free semigroups. In Gheorghe Paun, editor, *Mathematical aspects of natural and formal languages*, pages 225–232. World Scientific, 1994.

10 Juha Kortelainen. 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.*, 3(1):43–57, 1998.

11 André Lentin. Sur l'équation $a^M = b^N c^P d^Q$ dans un monoïde libre. *C. R. Acad. Sci. Paris*, 260:3242–3244, 1965.

12 M. Lothaire. *Combinatorics on Words.* Addison-Wesley, 1983.

13 Roger C. Lyndon and Marcel-Paul Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.*, 9(4):289–298, 1962. doi:10.1307/mmj/1028998766.

14 Jarkko Peltomäki and Markus Whiteland. A square root map on Sturmian words (extended abstract). In *Proceedings of the 10th WORDS*, volume 9304 of *LNCS*, pages 197–209. Springer, 2015. doi:10.1007/978-3-319-23660-5_17.

15 Wojciech Plandowski. Test sets for large families of languages. In *Proceedings of the 7th DLT*, volume 2710 of *LNCS*, pages 75–94. Springer, 2003. doi:10.1007/3-540-45007-6_6.

16 Aleksi Saarela. Systems of word equations, polynomials and linear algebra: A new approach. *European J. Combin.*, 47:1–14, 2015. doi:10.1016/j.ejc.2015.01.005.

17 Huei Jan Shyr and Shyr-Shen Yu. Non-primitive words in the language p^+q^+. *Soochow J. Math.*, 20(4):535–546, 1994.