International Journal of Foundations of Computer Science
Vol. 22, No. 2 (2011) 345–375

DOI: 10.1142/S0129054111008088

WORD EQUATIONS WITH ONE UNKNOWN

MARKKU LAINE

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

WOJCIECH PLANDOWSKI

Institute of Informatics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
w.plandowski@mimuw.edu.pl

Received 6 April 2009
Accepted 14 July 2010
Communicated by Volker Diekert and Dirk Nowotka

We consider properties of the solution set of a word equation with one unknown. We prove that the solution set of a word equation possessing infinite number of solutions is of the form $(pq)^*p$ where pq is primitive. Next, we prove that a word equation with at most four occurrences of the unknown possesses either infinitely many solutions or at most two solutions. We show that there are equations with at most four occurrences of the unknown possessing exactly two solutions. Finally, we prove that a word equation with at most $2k$ occurrences of the unknown possesses either infinitely many solutions or at most $8\log k + O(1)$ solutions. Hence, if we consider a class $\mathcal{E}_k$ of equations with at most $2k$ occurrences of the unknown, then each equation in this class possesses either infinitely many solutions or $O(\log k)$ number of solutions. Our considerations allow to construct the first alphabet independent linear time algorithm for computing the solution set of an equation in a nontrivial class of equations.

Keywords: Combinatorics on words; word equations; unification theory.

2000 Mathematics Subject Classification: 68R15

1. Introduction

The theory of word equations is a part of combinatorics on words [5, 6] and unification theory [1]. It has applications in graph theory [9], constraint logic programming [8], artificial intelligence [4], and formal languages [2].

As in number theory the problems in this theory are simple to explain to a nonexpert and difficult to solve. As an evidence of the difficulty of the problems in the theory of word equations is that not everything is known on word equations with one unknown. Our paper deals with such equations. The satisfiability problem for such equations can be solved in $O(n \log n)$ time where n is the size of the equation [7]. For finite alphabets of fixed size it can be solved in $O(n + \#_X \log n)$ time where $\#_X$ is the number of occurrences of the unknown X in the equation [3]. If $\#_X = \Theta(n)$, then the algorithm is not linear. It is still not known whether there is a linear time algorithm for the problem.

Our main results deal with the structure of the solution set of an equation with one unknown. The set is either finite or infinite. In [7] it was implicitely proved that in the former case it is of size at most $2 \log n$ where n is the size of the equation and in the latter case it is a union of a finite set of size at most $2 \log n$ and a set of the form $(pq)^+p$ with pq-primitive and q nonempty. We improve the latter case by proving that for such equations the solution set is of the form $(pq)^*p$ with pq-primitive and q nonempty. As an immediate effect of our considerations we obtain a linear time algorithm independent on the size of the alphabet, for finding the solution set of an equation which have infinitely many solutions.

It is quite elementary to find out that equations with at most three occurrences of the unknown can have only zero, one or infinitely many solutions, see Example 2. We show an example of an equation with exactly four occurrences of the unknown which has exactly two solutions. Next we prove that this number is tight. Namely we prove that equations with at most four occurrences of the unknown have either at most two solutions or infinitely many solutions.

At the end we consider equations with at most $2k$ occurrences of the unknown and prove that they have either at most $8 \log k + O(1)$ solutions or infinitely many solutions. If we fix k, then this gives a constant which bounds the number of solutions of an equation containing at most $2k$ occurrences of the unknown in case it has finitely many solutions. Our considerations allow one to construct the $O(n \log k)$ worst case time algorithm where n is the size of the input equation for the satisfiabiity problem and for finding a solution set of an equation containing at most $2k$ occurrences of the unknown. If k is fixed, then it is the first linear time algorithm independent of the size of the alphabet, for these problems for a nontirivial class of equations.

2. Basic Notions

For two numbers p and q, $gcd(p, q)$ denotes the greatest common divisor of p and q. Let Σ be any set. The set Σ is called *alphabet*. The elements of Σ are called *letters*. Sequences of letters are called *words*. The empty sequence is called *the empty word* and is denoted by 1.

Let w be a word. The length of the word w is denoted by $|w|$. A *prefix* of the word w is a word p such that $w = ps$, for some word s. If $s \neq 1$, then the word p is called a *proper prefix*. A *suffix* of the word w is a word s such that $w = ps$, for some word p. If $p \neq 1$, then the word s is called a *proper suffix*. A *subword* of the word w is a word u such that $w = pus$, for some words p and s. Denote by $w[i..j]$, for $1 \leq i \leq j \leq |w|$ the subword of the word w which starts at letter i of the word w and ends at letter j of the word w. The i-th letter of the word w is denoted by

$w[i]$. The prefix of the word w of length k is denoted by $\mathrm{pref}_k(w)$. The suffix of the word w of length k is denoted by $\mathrm{suff}_k(w)$.

For a prefix p of the word w, we denote by $p^{-1}w$ the word s such that $w = ps$. Similarly, for a suffix s of the word w, we denote by ws^{-1} the word p such that $w = ps$. For two words, u, w, by equality $u\cdots = w\cdots$ we mean that there are words u', w' such that $uu' = ww'$. Since it means that either u is a prefix of w or vice versa, we say then that the words u and w are *prefix comparable*. Similarly, by equality $\cdots u = \cdots w$ we mean that there are two words u' and w' such that $u'u = w'w$. Since it means that either u is a suffix of w or vice versa, we say then that the words u and w are *suffix comparable*.

For a word u and a positive integer k, by u^k we mean a word which is a repetition of the word u, k times. If $k = 0$, we define $u^0 = 1$. We say that a word w is *primitive*, if it is not of the form u^k, for any word u and $k \geq 2$. A primitive word u such that $w = u^k$, for some $k \geq 1$, is called *the primitive root* of the word w and denoted by $\mathit{root}(w)$. The function *root* is well-defined for nonempty words, since such a word is unique.

By u^ω we mean a right infinite word which consists of the word u repeated infinitely many times. Similarly, by ${}^\omega u$ we mean a left infinite word which consists of the word u repeated infinitely many times. A *period* of a word w is a prefix u of the word w such that w is a prefix of u^ω. We call a period also the number $|u|$. We say that a period u of a prefix of w *breaks* at position p of w if u is a period of $w[1..p-1]$ and u is not a period of $w[1..p]$. The proof of the next well-known proposition can be found in [5, 6].

Proposition 1 (Fine and Wilf's Lemma) *Let u and v be periods of a word w satisfying $|u|+|v| \leq |w|$. Then the prefix p of w of length $gcd(|u|, |v|)$ is also a period of w.*

We say that two words u and v *commute* if $uv = vu$. We say that words x and y are *conjugates*, if there are words p, q such that $x = pq$ and $y = qp$.

The *reverse* of a word $w = w[1]\cdots w[k]$ is $w[k]\cdots w[1]$.

We will use several well known facts in combinatorics on words. The last one of them is the famous Lyndon-Schützenberger lemma, see the proof in [5].

Fact 1. *(1) Let u and v with $|u| > |v|$ be two periods of a word w. Then the number $|u| - |v|$ is a period of the word $v^{-1}w$.*

(2) If the word p is a period of the word w and the word p is primitive and the word p is a suffix of the word w, then $w = p^l$, for some $l \geq 1$.

(3) Let x and y be two words. Let x be a prefix of the word yx. If $|y| \leq |x|$, then the word y is a period of the word x.

(4) Let x, x' and y be three words. Let the word x be a prefix of the word yx' where the word x' is a prefix of the word x. Then the word y is a period of the word yx'.

(5) Let x and y be two words. Suppose $x^iy\cdots = y^jx\cdots$ with i, $j \geq 1$. Then the words x and y are powers of the same word.

(6) *Let x and y be two words. Suppose the words x and y satisfy a nontrivial identity. Then the words x and y are powers of the same word. In particular, if x and y commute, then they are powers of the same word.*
(7) *Let x, y, z be three words. If $x^n y^m = z^k$ where n, m, $k \geq 2$, then the words x, y and z are powers of the same word.*

A *word equation* e with one unknown X is an equation of the form

$$A_0 X A_1 X \cdots A_k = X B_0 X B_1 X \cdots B_l$$

where A_i, $B_i \in \Sigma^*$ and either A_k or B_l is the empty word. A *solution* of a word equation is a word x such that $A_0 x A_1 x \cdots A_k = x B_0 x B_1 x \cdots B_l$. Observe that always the word x is a prefix of the word A_0^ω. The *size* of the equation e is the number $\sum_{i=0}^{k} |A_i| + \sum_{i=0}^{l} |B_i|$.

Example 2. *A word equation with one occurrence of X is of the form $X = A_0$. It has exactly one solution $x = A_0$.*

A word equation with two occurrences of X is either of the form $XB_0X = A_0$ or of the form $A_0X = XB_0$.

The length argument shows that the only possible length of a solution of the first equation is $\frac{1}{2}(|A_0| - |B_0|)$. Hence, it can have one or zero solutions. The only possible solution is a prefix of appropriate length of A_0.

The second equation is well known in combinatorics on words. It has either zero or infinitely many solutions. In the latter case there are words p, q with pq-primitive and q nonempty such that $A_0 = (pq)^k$, $B_0 = (qp)^k$. Then the solution set is $(pq)^*p$. For any pair of words A_0 and B_0, there is at most one such pair (p, q). Consequently, if pq is primitive and q nonempty, then the solution set of the equation $(pq)^k X = X(qp)^k$ is $(pq)^*p$.

The length argument shows that, if the number of occurrences of X on the left hand side of the equation is different from the number of occurrences of X on the right hand side, then there is at most one solution of such equation. In particular, each equation with odd number of occurrences of X has either zero or one solution. Hence, the only equations we consider next contain the same number of occurrences of X on both sides of the equation.

3. The Family of Equations $u_1Xu_2X = Xv_1Xv_2$

In this section we are interested in equations with finitely many solutions. As we concluded in Section 2 the simplest interesting case of equations with at most four occurrences of X are equations of the form $u_1Xu_2X = Xv_1Xv_2$ or $u_1Xu_2Xu_3 = Xv_1X$. We will prove that such equations have at most two solutions and this boundary is sharp. Since all solutions are prefixes of the same word, then the shortest solution is a prefix of all the other solutions. Let this solution be x. Consider the equation $u_1xXu_2xX = xXv_1xXv_2$ (resp. $u_1xXu_2xXu_3 = xXv_1xX$). It has the same number of solutions as the original equation and, additionally $X = 1$ is a solution of it. This means that we may restrict our considerations to equations such that $X = 1$ is a solution of them. Then, $u_1u_2 = v_1v_2$ (resp. $u_1u_2u_3 = v_1$). If $u_1w = v_1$, then $u_2 = wv_2$, and the equation is of the form $(u_1X)w(v_2X) = (Xu_1)w(Xv_2)$ (resp. $(u_1X)u_2(Xu_3) = (Xu_1)u_2(u_3X)$). These split into simpler equations with two occurrences of X and either one or infinitely many solutions. If $root(u_1) \neq root(v_2)$ (resp. $root(u_1) \neq root(u_3)$), then only $X = 1$ is a solution of the system. Otherwise the solution set of the system is the set $root(u_1)^*$.

If $u_1 = v_1w$, then $wu_2 = v_2$ and the equation is of the form $v_1wXu_2X = Xv_1Xwu_2$. Renaming the constant words we obtain an equation of the form $uvXwX = XuXvw$. This is the family of equations we consider in this section.

Example 3. *Consider the equation $XbXcbbccbc = bcbbccbXcX$. Since all solutions are prefixes of the word $(bcbbccb)^\omega$ it is easy to check that all solutions shorter than 6 are 1 and bc. All solutions of length at least 6 start with $bcbbcc$. Hence, the original equation can be split so that the solutions satisfy $Xbbcbbcc = bcbbccbX$. Hence, such solutions are of the form $x = (bcbbccb)^ibcbbcc$. However, if we look at the original equation we see that bc should be a suffix of x. Hence, there are no such solutions.*

Similarly we prove that the solution set of the equation $XbXcbc = bcbXcX$ is $\{1, bc\}$. Our last example is an equation $XbXcbbbcb = bcbbbcXbX$, which has only the solutions 1 *and bcb. Observe, that if we would replace XbX with X on both sides of the equation, we would get a simple conjugacy equation with the solution set $(bcbbbc)^*b$.*

An equation can have exactly two nonempty solutions. For instance the solution set of the equation $dbdcdbdbdbdcXbX = XbXcdbdbdbdcdbd$ equals $\{d, dbdcdbd\}$. This equation can be obtained from the previous one by adding the letter d between conecutive constant letters of the previous equation. In particular, if we remove letters d from the equation we obtain the previous equation. Observe here, that this trick allows to transform any equation with one unknown possesing exactly k solutions into equation possesing exactly k nonempty solutions.

Lemma 4. *Let p, q, p', q' are four words where q and q' are nonempty. Suppose that the words pq and $p'q'$ are primitive. Let, for some nonnegative integers i, j, n, $m \geq 0$, $(pq)^ip = (p'q')^jp'$ and $(pq)^np = (p'q')^mp'$. If either i, $j \geq 2$, or n, $m \geq 2$, or $n \neq i$ or $m \neq j$, then $p = p'$, $q = q'$, $i = j$ and $n = m$.*

Proof. Case i, $j \geq 2$ or n, $m \geq 2$. We may assume that i, $j \geq 2$, since the other case is symmetric. Let $w = (pq)^ip = (p'q')^jp'$. Then both words pq and $p'q'$ are periods of the word w. Since i, $j \geq 2$, then $|pq| + |p'q'| \leq |w|$. Hence, we may apply Fine and Wilf's Lemma. Consequently, the number $t = gcd(|pq|, |p'q'|)$ is a period of the word w, too. Since the number t divides both the number $|pq|$ and the number $|p'q'|$ and the words pq and $p'q'$ are primitive, we must have $pq = p'q'$. Since the words q and q' are nonempty, then both $|p'q'| = |pq| > |p|$ and $|pq| = |p'q'| > |p'|$. This, with the equation $pq = p'q'$, implies $i = j$, and, consequently, $p' = p$. By equation $pq = p'q'$, we have $q = q'$. Hence, $n = m$.

Case $n \neq i$ **or** $m \neq j$. We may assume that $n \neq i$, since the other case is symmetric. Moreover we may assume that $n > i$, since the case $n < i$ is symmetric. Then, we have

$$(p'q')^m p' = (pq)^n p = (pq)^i p(qp)^{n-i} = (p'q')^j p'(qp)^{n-i}.$$

Hence, $m > j$ and $(q'p')^{m-j} = (qp)^{n-i}$. Since the words pq and $p'q'$ are primitive, then their conjugates qp and $q'p'$ are primitive, too. Hence $q'p' = qp$. Since q and q' are nonempty, then both $|p'q'| = |pq| > |p|$ and $|pq| = |p'q'| > |p'|$. This, with $pq = p'q'$, implies $i = j$, and, consequently, $p' = p$. By equation $pq = p'q'$, we have $q = q'$. Hence, $n = m$. $\square$

In the following, the equation

$$XuXvw = uvXwX \tag{1}$$

is assumed to have a finite solution set $S = \{s_0 = 1, \cdots, s_{k-1}\}$. Since all the words in S are prefixes of the word $(uv)^\omega$ and suffixes of the word $^\omega(vw)$ we may assume that the word s_{i-1} is a proper prefix and a proper suffix of the word s_i, for $i = 1, \cdots, k-1$.

We have

$$s_i u s_i v w = u v s_i w s_i. \tag{2}$$

Hence, the word $s_i u$ is a prefix of the word uvs_i, and, consequently, there is a word v_i such that $s_i u v_i = uvs_i$. By length argument, $|v_i| = |v|$. We have $s_i u s_i v w = uvs_i w s_i = s_i u v_i w s_i$. Hence, we have $s_i v w = v_i w s_i$. Summing this together, we have

$$s_i u v_i = u v s_i \qquad \text{and} \qquad s_i v w = v_i w s_i, \tag{3}$$

Looking at the beginnings of the both sides of the second equality and at the ends of the both sides of the first equality we obtain $v_i = \mathrm{pref}_{|v|}(s_i v) = \mathrm{suff}_{|v|}(v s_i)$, see also Fig. 1.

On the other hand, a common solution s of equations

$$Xuv_i = uvX \qquad \text{and} \qquad Xvw = v_i w X \tag{4}$$

is also a solution of equation (1), since

$$susvw = (su)(svw) = (su)(v_i ws) = (suv_i)(ws) = uvsws.$$

Moreover, looking at the beginnings of both sides of the first equality in equations (3) and at the ends of both sides of the second equality in equations (3), we also get the equalities

$$s_i u = u s_i' \qquad \text{and} \qquad s_i'' w = w s_i, \tag{5}$$

By length argument, we have $|s_i| = |s_i'| = |s_i''|$. Moreover $s_i' = \mathrm{suff}_{|s_i|}(s_i u)$ and $s_i'' = \mathrm{pref}_{|s_i|}(w s_i)$.

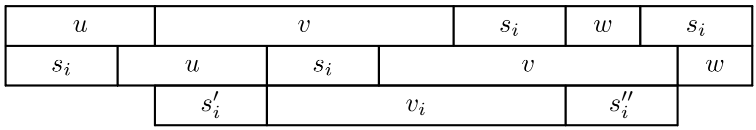

Fig. 1. The dependences between words s_i, s_i', s_i'', v_i and other words. The letters in the same columns are the same. The relative position of boundaries between words in first two rows may be different.

We have $us_i's_ivw = s_ius_ivw = uvs_iws_i$. By cacelling common prefix u of the first and the last sides of equation, we have that s_i' is a prefix of vs_i. Similarly, by cancelling common suffix w, the equation $s_ius_ivw = uvs_iws_i = uvs_is_i''w$ we have, that $s_i'' = \mathrm{suff}_{|s_i|}(s_iv)$. Summing this we have $s_i' = \mathrm{suff}_{|s_i|}(s_iu) = \mathrm{pref}_{|s_i|}(vs_i)$ and $s_i'' = \mathrm{pref}_{|s_i|}(ws_i) = \mathrm{suff}_{|s_i|}(s_iv)$.

We have $us_i'v_iws_i =_{\text{by (3)}} us_i's_ivw = s_ius_ivw = uvs_iws_i$. By cancelling common begining and common end of the first and last sides of this equality, we obtain $s_i'v_i = vs_i$. Similarly, we have $s_ius_ivw = uvs_iws_i = uvs_is_i''w =_{\text{by (3)}}= s_iuv_is_i''w$. By reducing common beginning and end of the first and last sides of the equation we obtain $s_iv = v_is_i''$. Consequently, we have

$$s_i'v_i = vs_i \qquad \text{and} \qquad s_iv = v_is_i''. \tag{6}$$

The dependencies between words are illustrated in Fig. 1.

From here on, by symmetry, we assume, that $|u| \geq |w|$.

Lemma 5. *The length of any solution s_i of equation* (1) *is less than* $2|uv|$.

Proof. Assume, on the contrary, that $|s_i| \geq 2|uv|$. By equalities (3), the word uv is a conjugate of the word uv_i and the word vw is a conjugate of the word v_iw. Again, by equalities (3), there exist words p, q, p' and q', such that the words pq and $p'q'$ are primitive, q and q' are nonempty, $uv \in (pq)^+$, $uv_i \in (qp)^+$, $v_iw \in (p'q')^+$, $vw \in (q'p')^+$, and $s_i = (pq)^ip = (p'q')^jp'$, where $i, j \geq 2$. This implies that $p = p'$ and $q = q'$, by Lemma 4. Therefore, all words in the set $(pq)^*p$ are solutions of equalities (3) and, consequently, of the equation (1), which is a contradiction. □

Lemma 6. *The word s_i is the only solution for the equation pair* (4).

Proof. Let u and v be two solutions for the equation pair (4). Since each equation in the pair is a conjugacy equation, then there are four words p, p', q, q' with nonempty q and q' and primitive pq and $p'q'$, such that $u, v \in (pq)^*p \cap (p'q')^*p'$, $uv_i \in (qp)^+$, $uv \in (pq)^+$, $vw \in (q'p')^+$, $v_iw \in (p'q')^+$. Since u is different of v, then $|(pq)^*p \cap (p'q')^*p'| \geq 2$. By Lemma 4 $p = p'$ and $q = q'$. Hence all words in $(pq)^*p = (p'q')^*p'$ are solutions equation pair (4). Hence, the equation $XuXvw = uvXwX$ has infinitely many solutions. A contridiction. Hence the equation pair (4) has at most one solutions. By equalities (3) the word s_i is a solution of the equation pair (4). This means that the word s_i is the only solution of the equation pair (4). $\square$

Corollary 7. *If $i \neq j$, then $v_i \neq v_j$.*

Corollary 8. *The equation* (1) *has at most one solution of length at least $|v|/2$.*

Proof. Assume, on the contrary, that $|v|/2 \leq |s_i|$, for some $i < k-1$. Then, since $v_i = \mathrm{pref}_{|v|}(s_i v) = \mathrm{suff}_{|v|}(v s_i)$, we have

$$\begin{aligned} v_i &= \mathrm{pref}_{\lfloor |v|/2 \rfloor}(s_i v)\,\mathrm{suff}_{\lceil |v|/2 \rceil}(v s_i) \\ &= \mathrm{pref}_{\lfloor |v|/2 \rfloor}(s_i)\,\mathrm{suff}_{\lceil |v|/2 \rceil}(s_i) \\ &=_{\text{since } s_i \text{ is a prefix and a suffix of } s_{i+1}} \mathrm{pref}_{\lfloor |v|/2 \rfloor}(s_{i+1})\,\mathrm{suff}_{\lceil |v|/2 \rceil}(s_{i+1}) \\ &= \mathrm{pref}_{\lfloor |v|/2 \rfloor}(s_{i+1} v)\,\mathrm{suff}_{\lceil |v|/2 \rceil}(v s_{i+1}) = v_{i+1}. \end{aligned}$$ $\square$

Lemma 9. *For all $i > 0$, $s_i' \neq s_i$ and $s_i \neq s_i''$.*

Proof. We may assume, on the contrary, that $s_i' = s_i$, since the case $s_i'' = s_i$ is symmetric. We get from equalities (6):

$$s_i s_i v = s_i' s_i v = s_i' v_i s_i'' = v s_i s_i''. \tag{7}$$

Hence, the words s_i^2 and $s_i s_i''$ are conjugates. If the word r is the primitive root of the word s_i, then the word s_i^2 is a power of the word r and the word $s_i s_i''$ is a power of a conjugate r' of the word r. Since the word $s_i s_i''$ starts with the word r and with the word r' and the lengths of the words r and r' are equal, then $r = r'$. Consequently, the word s_i'' is a power of the word r. As its length is the same as the length of the word s_i, we have $s_i'' = s_i$. By equation (7) the word v is also a power of the word r. Thus, $s_i, s_i', s_i'', v \in r^+$, for some primitive word r.

By equations (5) $u, w \in r^+$. Hence, for each $k \geq 0$, $r^k u r^k v w = u v r^k w r^k$ and, consequently, all the words in r^+ are solutions of the equation (1), a contradiction. $\square$

Lemma 10. *If $k \geq 3$, then $|s_2| > |u| + |s_1|$.*

Proof. Assume, on the contrary, that $|s_2| \leq |u| + |s_1|$. Assume also, that the word r is the primitive root of the word s_1. Since the word u is a prefix of the word $s_1 u$, the word u is a prefix of the word r^ω. By equalities (3), the word $s_2 u$ is a prefix of the word $u v s_2$, which implies, that the words s_2 and $u v s_1$ are prefix comparable, since the word s_1 is a prefix of the word s_2. Since $|s_2| \leq |u s_1|$, the word s_2 is a prefix of the word $u v s_1$. Moreover, again by equalities (3), the word $s_1 u$ is also a prefix of the word $u v s_1$. Thus, since $|s_2| \leq |s_1 u|$, the word s_2 is a prefix of the word $s_1 u$. The word $s_1 u$ is a prefix of the word r^ω. Hence, the word s_2 is a prefix of the word r^ω. Since the word s_1 is a suffix of the word s_2, the word r is also the primitive root of the word s_2, .

By Corollary 8, we may assume, that the word s_1 is shorter than the word v. Equality $us_1' = s_1u$ implies, that there are integers k, l and n, with $0 < k < l$, and words p and q with pq primitive and q nonempty, such that $s_1 = (pq)^k$, $s_1' = (qp)^k$, $s_2 = (pq)^l$ and $u = (pq)^n p$. By equation (2) (see also Fig. 1), the words s_1us_1 and uvs_1 are prefix comparable. Since the word s_1 is shorter than the word v, then the word s_1us_1 is a prefix of the word uvs_1. Hence, we get that the word $w = (pq)^{k+n}ppq$ is a prefix of the word uvs_1. By equation (2) the words uvs_2 and s_2us_2 are prefix comparable, see Fig. 1. Hence, since the word s_1 is a prefix of the word s_2, the words uvs_1 and s_2u are prefix comparable. Hence, we have that $(pq)^{k+n}pqp$, as the length $|w|$ prefix of the word s_2u, is a prefix of the word uvs_1. Hence $pq = qp$, and, consequently, the word s_1' equals the word s_1, a contradiction. □

We use the following well known fact in combinatorics on words.

Fact 2. *Let x, y, z be three words. Then*

$$x^n z = zy^n, \text{ for some } n \geq 1 \iff xz = zy.$$

Our next lemma gives the necessary and sufficient condition for an equation to have an infinite number of solutions.

Lemma 11. *The solution set S of* (1) *is infinite, iff uv commutes with vw.*

Proof. If the words uv and vw commutes, then $uv, vw \in z^+$. Then any word $s \in z^*$ is a solution, since $susvw = suvws = uvsws$.

Assume then, that the set S is infinite. Let $v' = \text{pref}_{|v|}(uv)$. Since all solutions are prefixes of the infinite word $(uv)^\omega$, each solution of length at least $|v|$ starts with the word v'. Hence it is in form $v'Y$ for some Y which satisfies $v'Yuv'Yvw = uvv'Ywv'Y$. By length argument Y satisfies two equations $v'Yuv' = uvv'Y$ and $Yvw = wv'Y$. Hence, the word $X = v'Y$ satisfies $Xuv' = uvX$ and $Xvw = v'wX$. By Lemma 4, the primitive root pq of the word uv is the primitive root of the word $v'w$ and the primitive root qp of the word uv' is the primitive root of the word vw.

We need to show, that $pq = qp$. Assume, on the contrary, that $pq \neq qp$. As the word u is a prefix of both words uv' and uv, the word u is prefix comparable with both words pq and qp. Hence, the word u has to be shorter than the word pq. Similarly $|w| < |pq|$. If $|uv| \neq |v'w|$, then $||u| - |w|| = ||uv| - |v'w|| \geq |pq|$. Hence, $uv = v'w$. Similarly, $uv' = vw$. By computing $u^2v' = uuv' = uvw = v'ww = v'w^2$, by Fact 2, we get that $uv' = v'w$. Consequently, $uv' = v'w = uv$ and $pq = qp$, a contradiction. □

Lemma 12. *Let x and x' be conjugates such that $x \neq x'$. Then the word xx' is primitive.*

Proof. We have $x = rs$ and $x' = sr$, for some words r, $s \neq 1$. Suppose that the word xx' is not primitive. Then there is a word ρ such that $rssr = \rho^n$ for some $n \geq 2$.

Then the word r^2s^2 is a conjugate of the left hand side. Hence, there is a conjugate ρ_1 of the word ρ such that $r^2s^2 = \rho_1^n$. But this is the Lyndon-Schützenberger equation, see Fact 1. Hence, the words r, s are powers of the same word. Hence, $x = rs = sr = x'$ — a contradiction. $\square$

A period of a word is called *primitive* if it is a primitive word.

Lemma 13. *There is at most one non-empty solution s_i such that $|s_i| \leq |v|/2$.*

Proof. By equations (6) we have

$$s_i's_iv = s_i'v_is_i'' = vs_is_i'', \tag{8}$$

see also Fig. 1. By adding s_i as a prefix and a suffix of the first and the last sides of equality (8) we obtain $(s_is_i')(s_ivs_i) = (s_ivs_i)(s_i''s_i)$. By equations (5) the words s_i, s_i' and s_i'' are conjugates and, by Lemma 9, $s_i \neq s_i'$ and $s_i \neq s_i''$. Hence, by Lemma 12, the word s_is_i' is primitive, and, consequently, the word s_ivs_i has a primitive period of length $2|s_i|$. Suppose that $0 < |s_1| < |s_2| \leq |v|/2$. Since the word s_1 is a prefix and a suffix of the word s_2 the word s_1vs_1 is a central subword of s_2vs_2. The word s_2vs_2 has a primitive period of length $2|s_2|$. Hence, the word s_1vs_1 has two primitive periods of lengths $2|s_1|$ and $2|s_2|$. Also $|s_1vs_1| \geq 2|s_1| + 2|s_2|$, and, by Proposition 1, $|s_1| = |s_2|$ raising a contradiction. $\square$

Summing Lemma 13 and Corollary 8, we have at most one non-empty solution of length at most $|v|/2$ and at most one nonempty solution of length at least $|v|/2$. Together with the empty solution our equation has at most three solutions.

We proceed by assuming that the solution set of our equation is $\{s_0 = 1, s_1, s_2\}$ where $|s_1| < |v|/2 \leq |s_2|$. Next we will gradually close the window of possible lengths for s_2.

We already know, by Lemma 5, that $|s_1| < |v|/2 \leq |s_2| < 2|uv|$. By equations (6) and by $|s_1| < |v| = |v_1|$, the word s_1 is a prefix of the word v_1. Still by equations (6) and by $|s_1's_1| = 2|s_1| < |v|$ the word $s_1's_1$ is a proper prefix of the word v. Let t_1 be such that $v = s_1's_1t_1$. Again by equations (6), $s_1'v_1 = vs_1 = s_1's_1t_1s_1$. By reducing common prefix s_1', we obtain $v_1 = s_1t_1s_1$, Once again by equations (6), $s_1v = v_1s_1'' = s_1t_1s_1s_1''$. By reducing common prefix s_1 in the last equation, we obtain $v = t_1s_1s_1''$. Summing everything together we obtain $v = s_1's_1t_1 = t_1s_1s_1''$ and $v_1 = s_1t_1s_1$, see Fig. 2. Since the word s_1 is a prefix and a suffix of the word s_2 and $|s_1| \leq |v| = |v_2|$, then, by equations (6), the word s_1 is a prefix and a suffix of the word v_2. Hence, by $2|s_1| \leq |v| = |v_2|$, $v_2 = s_1t_2s_1$, for some non-empty word t_2 satisfying $|t_1| = |t_2|$. Hence, we have equations:

$$v = s_1's_1t_1 = t_1s_1s_1'',\ v_1 = s_1t_1s_1,\ v_2 = s_1t_2s_1, \tag{9}$$

for some nonempty words t_1 and t_2, such that $|t_1| = |t_2|$.

By Lemma 9, Lemma 11 and Corollary 7 we know, that none of the three conditions

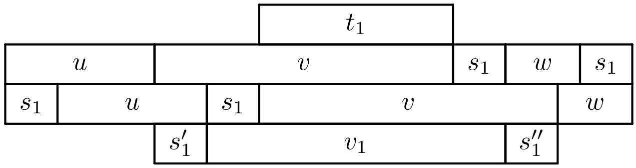

Fig. 2. Illustration for dependencies between the word t_1 and words defined earlier.

(1) $s_1' = s_1$ ($s_1'' = s_1$);
(2) uv commutes with vw; and
(3) $t_1 = t_2$

is possible. The proofs of each one of the following lemmas start from a counter assumption and then one of these three contradictions is deduced.

Lemma 14. $|s_2| > 2\,|s_1|$

Proof. Assume, on the contrary, that $|s_2| \leq 2\,|s_1|$. Hence, $|s_1| < |s_2| \leq |s_1 t_2 s_1| = |v_2| = |v|$. Since the word s_1 is a prefix and a suffix of the word s_2 and since $|s_2| \leq 2|s_1|$, we have $s_2 = \alpha s_1 = s_1 \beta$, for some words α, β satisfying $|\alpha| = |\beta| \leq |s_1|$. Moreover, for some words β' and α'', $s_2' = s_1'\beta'$ and $s_2'' = \alpha'' s_1''$, since, by equality (6) and by $|s_1|$, $|s_2| < |v|$, both words s_1' and s_2' are prefixes and both words s_1'' and s_2'' are suffixes of the word v. Observe, that $|\alpha| = |\beta| = |\alpha''| = |\beta'| \leq |s_1|$. By equality (6) and by equality (9), both words $s_1' s_1$ and $s_2' = s_1'\beta'$ are prefixes of the word v. Hence, the word β' is a prefix of the word s_1. By $\alpha s_1 = s_1 \beta$, the word α is also a prefix of the word s_1 of length $|\alpha| = |\beta'|$. Hence, $\beta' = \alpha$. Similarly, we obtain $\alpha'' = \beta$. We have the following implications that start from equations (6) and apply equations (9) and equations already deduced in this proof:

$$\begin{cases} v s_2 = s_2' v_2 \\ v_2 s_2'' = s_2 v \end{cases} \Rightarrow \begin{cases} s_1' s_1 t_1 \alpha s_1 = s_1' \alpha s_1 t_2 s_1 \\ s_1 t_2 s_1 \beta s_1'' = s_1 \beta t_1 s_1 s_1'' \end{cases} \Rightarrow \begin{cases} s_1 t_1 \alpha = \alpha s_1 t_2 \\ t_2 s_1 \beta = \beta t_1 s_1 \end{cases}$$

$$\Rightarrow \begin{cases} s_1 t_1 \alpha = s_1 \beta t_2 \\ t_2 \alpha s_1 = \beta t_1 s_1 \end{cases} \Rightarrow \begin{cases} t_1 \alpha = \beta t_2 \\ t_2 \alpha = \beta t_1 \end{cases} \Rightarrow \begin{cases} t_1 \alpha = \beta t_2 \\ t_2 \alpha^2 = t_2 \alpha\alpha = \beta t_1 \alpha = \beta\beta t_2 = \beta^2 t_2 \end{cases}$$

$$\Rightarrow\ t_1 \alpha = \beta t_2 = t_2 \alpha.$$

The last implication comes from Fact 2. We come to the contradiction $t_1 = t_2$. □

Since $|s_2| > 2|s_1|$ and the word s_1 is a prefix and a suffix of the word s_2, we may write $s_2 = s_1 \gamma s_1$, for some nonempty word γ. By equations (5), we have $us_2' = s_2 u$. Hence, since the word s_1 is a suffix of the word s_2, we have that the word $us_1' = s_1 u$ is a suffix of the word $s_2 u = us_2'$. By Lemma 10, the word s_2' is of length at least $|s_1 u|$. Hence, the word us_1' is a suffix of the word s_2'. In particular the word s_1' is a suffix of the word s_2'. By equalities (6), the words v and s_2' are

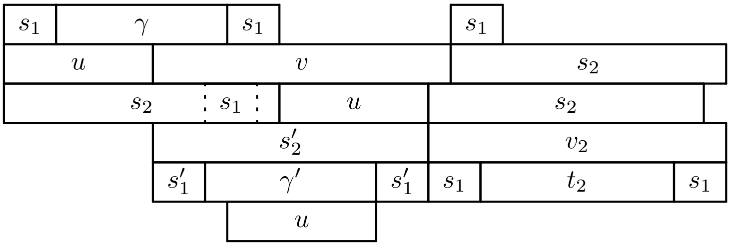

Fig. 3. Dependencies on words in the proof of Lemma 15.

prefix comparable. By equations (9), the word s_1' is a prefix of the word v and, consequently a prefix of the word s_2'. Consequently, $s_2' = s_1'\gamma's_1'$, for some word γ'. Symmetrically, $s_2'' = s_1''\gamma''s_1''$. Observe also, that since $v = s_1's_1t_1$ (resp. $v = t_1s_1s_1''$) - by equations (9) - and the words v and s_2' are prefix comparable (resp. the words v and s_2'' are suffix comparable) - by equations (6)-, then the word s_1 is a prefix of the word $\gamma's_1'$ (resp. a suffix of the word $s_1''\gamma''$). Summing the coclusions of this paragraph we have

$$s_2 = s_1\gamma s_1,\ s_2' = s_1'\gamma's_1',\ s_2'' = s_1''\gamma''s_1'', \tag{10}$$

$$s_1 \text{ is a prefix of } \gamma's_1',\ s_1 \text{ is a suffix of } s_1''\gamma'' \tag{11}$$

Lemma 15. $|\gamma| > |t_1|$. *Hence,* $|s_2| > |v|$.

Proof. Assume, on the contrary, that $|\gamma| \le |t_1| = |t_2|$ and, consequently, $|s_2| = 2|s_1| + |\gamma| \le 2|s_1| + |t_2| = |v_2| = |v|$, or, equivalently, $|s_2u| \le |uv|$, see Fig. 3.

We show first, that $\gamma' \in (s_1s_1')^*s_1$. The word $s_1's_1$ is a period of the word vs_1, since, by equations (6), $(vs_1)s_1''s_1 = s_1'v_1s_1''s_1 = (s_1's_1)(vs_1)$. By Lemma 12, the word $s_1's_1$ is primitive. Hence, all occurrences of the word $s_1's_1$ in the word vs_1 are at positions which are multiples of $|s_1's_1|$.

Since the words $us_2's_2 = s_2us_2$ and uvs_2 are prefix comparable, then the words $s_2's_2$ and vs_2 are prefix comparable, and, since the word s_1 is a prefix of the word s_2 (by equations (10)), the words $s_2's_1$ and vs_1 are prefix comparable. By $|s_2| \le |v|$, $|s_2's_1| \le |vs_1|$ and, consequently, the word $s_2's_1$ is a prefix of the word vs_1. By equations (10), $s_1'\gamma's_1's_1 = s_2's_1$. Hence, the word $s_1'\gamma'(s_1's_1)$ is a prefix of the word vs_1. In particular, the occurrence of the suffix $s_1's_1$ of the word $s_1'\gamma's_1's_1$ in the word vs_1 is at position being a multiple of $|s_1's_1|$. Consequently, since the word $s_1's_1$ is a period of the word vs_1, $s_1'\gamma' \in (s_1's_1)^+$, and, finally $\gamma' \in (s_1s_1')^*s_1$.

Symmetrically we can prove that $\gamma'' \in (s_1s_1'')^*s_1$.

Let $k \ge 0$ be such that $\gamma' = (s_1s_1')^ks_1$. By equalities (10) and $|\gamma'| = |\gamma''|$ and $|s_1s_1'| = |s_1s_1''|$, we have $\gamma'' = (s_1s_1'')^ks_1$. Using the equalities (5) and (10) we compute

$$wu(s_1's_1)^{k+1}s_1' = wus_2' = ws_2u = s_2''wu = (s_1''s_1)^{k+1}s_1''wu = \\ = (s_1''s_1)^{k+1}ws_1u = (s_1''s_1)^{k+1}wus_1'. \tag{12}$$

Hence, $wu(s_1's_1)^{k+1} = (s_1''s_1)^{k+1}wu$. Hence, by Fact 2, $wus_1's_1 = s_1''s_1wu$. Thus, by equalities (5), $s_1''wus_1 = ws_1us_1 = wus_1's_1 = s_1''s_1wu$. Reducing common prefix s_1'' in the first and last side of the obtained equation, we get that the word wu commutes with the word s_1.

Case $|wu| \geq |s_1|$. Since the words wu and s_1 commutes, then the word s_1 is a suffix of the word wu. By equalities (12) the word s_1' is also a suffix of the word wu. Hence, the words s_1' and s_1 are equal, since both of them are suffixes of the word wu of the same length. We get a contradiction.

Case $|wu| < |s_1|$**.** Since the words s_1 and wu commutes, they are powers of the same word. In particular, the word wu is a prefix and a suffix of the word s_1. By equalities (5), the word u is a prefix of the word s_1 and the word w is a suffix of the word s_1. Hence, the word u is a prefix of the word wu and the word w is a suffix of the word wu. But $|wu| = |uw|$. Hence, $wu = uw$ and, consequently, u, $w \in \rho^+$, for some primitive word ρ. Since the words wu and s_1 commutes, $s_1 \in \rho^+$. This implies, that the word s_1 commutes with the word u. Hence, and by equalities (5), $us_1 = s_1u = us_1'$. Reducing common prefix u of the first and the last side of the obtained equation we get $s_1 = s_1'$. A contradiction. □

Lemma 16. $|\gamma| > |t_1| + |s_1|$. *Hence,* $|s_2| > |v| + |s_1|$.

Proof. Assume, on the contrary, that $|\gamma| \leq |t_1| + |s_1|$. By Lemma 15, we know, that $|t_2| = |t_1| < |\gamma|$. By equalities (6), (9), (10), we have $vs_1\gamma s_1 = vs_2 = s_2'v_2 = s_2's_1t_2s_1$. Hence, the word $s_1t_2s_1$ is a suffix of the word $s_1\gamma s_1$. Again by equalities (6), (9), (10), we have

$$s_1\gamma s_1v = s_2v = v_2s_2'' = s_1t_2s_1s_2''.$$

Hence, the word $s_1t_2s_1$ is a prefix of the word $s_1\gamma s_1$. Summing the conclusions of last two sentences we get that the word $s_1t_2s_1$ is a prefix and a suffix of the word $s_1\gamma s_1$. Thus, there are words p and q and a positive integer k, such that the word pq is primitive and q nonempty, and

$$(pq)^k s_1t_2s_1 = s_1\gamma s_1 = s_1t_2s_1(qp)^k. \tag{13}$$

We have

$$|pq| \leq |(pq)^k| = |s_1\gamma s_1| - |s_1t_2s_1| = |\gamma| - |t_2| \leq |s_1|.$$

Hence, the word qp is a sufix of the word s_1. Since the word pq is primitive, then its conjugate qp is primitive, too. Since $s_1\gamma s_1 \in (pq)^+p$, the word qp occurs in the word $s_1\gamma s_1$ only at positions which are of the form $|p| + i|pq|$, for some nonnegative integer i. Hence, $s_1 \in (pq)^+p$ and, consequently, by equation (13), t_2, $\gamma \in q(pq)^*$.

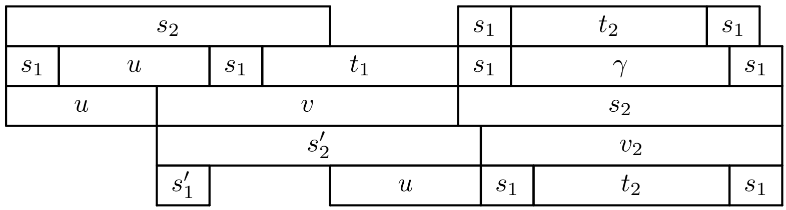

Fig. 4. The dependences between words occurring in the proof of Lemma 16.

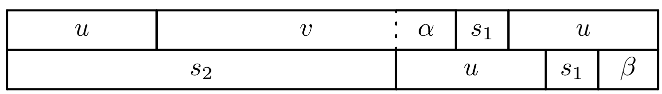

Fig. 5. Illustration for the proof of Lemma 17.

Since $|\gamma| > |t_2|$, we have $\gamma \in q(pq)^+$. By equations (10) and (5) the word us_1' is a prefix of the word $us_2' = s_2u$. Hence, by Lemma 10 and by equalities (5), the word $s_1u = us_1'$ is a prefix of the word s_2. By equalities (10), $s_2 = s_1\gamma s_1 \in (pq)^+p$. Therefore, the word u is a prefix of the word $(qp)^\omega$.

By equalities (6), $us_2' = s_2u$. Since $s_2 \in (pq)^+p$ and the word u is a prefix of $(qp)^\omega$ the word us_2' is a prefix of the word $(pq)^\omega$. Hence, the number $|pq|$ is a period of the word us_2'. By equalities (13), (10), (5), (9), we get equalities $uv(pq)^k s_1t_2s_1 = uvs_1\gamma s_1 = uvs_2 = us_2'v_2 = us_2's_1t_2s_1$. Reducing the common suffix $s_1t_2s_1$ from the first and last sides of the obtained equality we get $uv(pq)^k = us_2'$. Hence, by primitiveness of pq which is a period of us_2', we deduce $uv \in (pq)^+$. Hence, the word t_1, as a suffix of the word v (by (9)), is a suffix of the word ${}^\omega(pq)$. As the word $t_2 \in q(pq)^+$ is also is a suffix of the word ${}^\omega(pq)$ of same length, we get the contradiction $t_1 = t_2$. □

Lemma 17. $|s_2| > |uv|$.

Proof. Assume, on the contrary, that $|s_2| \leq |uv|$. By equalities (3), the word s_2 is a prefix of the word uv. Again by equalities (3), the word s_2u is a prefix of the word uvs_2. Since the word uvs_1 is a prefix of the word uvs_2, the words uvs_1 and s_2u are prefix comparable. By Lemma 16, we know, that the word uvs_1 is a prefix of the word s_2u. By equalities (3), the word s_1u is a prefix of the word uvs_1. Hence, the word s_1u is a prefix of the word s_2u. Hence, the word uvs_1u is a prefix of the word uvs_2u. By equalities (3) and (9), $uvs_2u = s_2uv_2u = s_2us_1\cdots$. Hence, the words uvs_1u and s_2us_1 are prefix comparable. Since the word s_2us_1 is shorter than the word uvs_1u, then it is one of its prefixes. Hence, there are words α and β, $|\alpha| = |\beta|$, such that $\alpha s_1u = us_1\beta$, see Fig. 5. By Lemma 16, we have $|\alpha| = |\beta| = |uv| - |s_2| < |u| - |s_1|$.

By equalities (6), we have $us_1' = s_1u$. Hence, there are words p and q, such that $s_1 = pq$, $s_1' = qp$ and $u \in (pq)^*p$. Since, by Lemma 16 and by our assumption, $|v|+|s_1| < |s_2| \le |u|+|v|$, then $|u| > |s_1|$ and, consequently, $u \in (pq)^+p$. Therefore, the word $qppq$ is a suffix of the word us_1. Since $|us_1|-|\alpha| > |us_1|-(|u|-|s_1|) = 2|s_1|$, we have $|\alpha^{-1}us_1| > 2|s_1| \ge 2|pq|$. Hence the word $qppq$ is a suffix of the word $\alpha^{-1}us_1$ which is a subword of the word s_1u. Thus, the word $qppq$ is a subword of the word $s_1u \in (pq)^+p$. This implies, that the word $s_1's_1 = qppq$ is not primitive, which is only possible, by Lemma 12, if $s_1' = s_1$. A contradiction. □

The next Lemma finally shows, that the assumption of three solutions is not possible. It also improves Lemma 5.

Lemma 18. *The length of any solution s_i of equation* (1) *is at most* $|uv|$.

Proof. Assume, on the contrary, that $|s_i| > |uv|$. By equations (3), there are words p, q, p', q' such that q and q' are non-empty and $s_i \in (pq)^*p$, $s_i \in (p'q')^*p'$, $uv = pq$, $uv_i = qp$, $v_iw = p'q'$ and $vw = q'p'$. By Corollary 7, $v_i \ne v_0 = v$, we have $uv \ne uv_i$ and $v_iw \ne vw$ and, consequently, p and p' are non-empty. At the beginning of this section we assumed that $|v_iw| = |vw| \le |uv|$, since the other case is symmetric. Since $|s_i| > |uv| = |pq| \ge |vw| = |p'q'|$, there are integers $k \ge j \ge 0$ such that $s_i = (pq)^{j+1}p = (p'q')^{k+1}p'$. By Lemma 5, $j = 0$. We have $s_i = pqp = (p'q')^{k+1}p'$. Observe that the words p and $(p'q')^kp'$ are prefix and suffix comparable.

Assume first, that $|uv| = |vw|$. Then, $|pq| = |uv| = |vw| = |p'q'|$. Since $s_i \in (pq)^+p \cap (p'q')^+p'$, then $pq = p'q'$ and, consequently, $p = p'$ and $q = q'$. Then any word $s \in (pq)^*p$ is a solution of equations $Xuv_i = uvX$ and $Xvw = v_iwX$, and hence also a solution of equation (1). Therefore, $|uv| \ne |vw|$.

Now

$$\begin{aligned} 0 < |uv| - |vw| &= |pq| - |q'p'| = (|pq| - |s_i|) + (|s_i| - |q'p'|) \\ &= (|pq| - |pqp|) + (|(p'q')^{k+1}p'| - |q'p'|) = |(p'q')^kp'| - |p|\,. \end{aligned}$$

Thus, the word p is both a proper prefix and a proper suffix of the word $(p'q')^kp'$, and there are words α and β and a non-negative integer h, such that $p = (\alpha\beta)^h\alpha$ and $(p'q')^kp' = (\alpha\beta)^{h+1}\alpha$. We have $|u| \ge |u|-|w| = |uv|-|vw| = |(p'q')^kp'|-|p| = |\alpha\beta|$. Since $(p'q')^{k+1}p' = s_i = pqp = puv_i$, then the words $(p'q')^kp'$ and pu are prefix comparable. Hence the word u starts by the word $\beta\alpha$. Moreover, by equations (3), the word u is a prefix of the word s_i, so it starts from the same prefix as the word $(p'q')^kp'$, that is $\alpha\beta$. Hence, $\alpha\beta = \beta\alpha$. Consequently, α, $\beta \in r^*$, for some primitive word r. Hence, $p = r^m$, and $(p'q')^kp' = r^n$, for some integers m, n satisfying $n > m > 0$ (recall that the word p is nonempty).

Since the words u and pu are prefixes of the word $s_i = pqp$, then the word u is a prefix of the word pu. Hence, the word u is a prefix of the word r^ω. Similarly , the word w is a suffix of the word ${}^\omega r$. Since $puv_i = pqp = (p'q')^{k+1}p' = v_iw(p'q')^kp'$, we get $r^{\ell_1}r'v_i = v_ir''r^{\ell_2}$, for some integers $\ell_1,\ \ell_2 \ge 0$, where the word r' is a proper

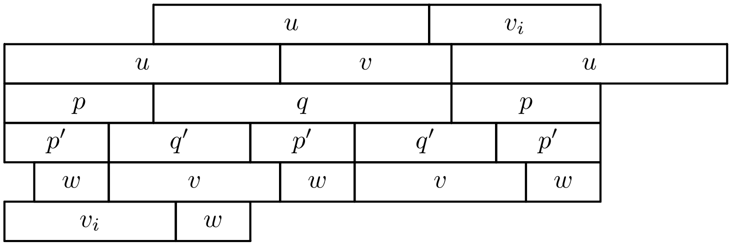

Fig. 6. Illustration for the proof of Lemma 18.

prefix of the word r, the word r'' is a proper suffix of the word r. Comparing the lengths of sides of the last equation we obtain $\ell_1 = \ell_2$ and $|r'| = |r''|$. Consequently, $r^\ell r' v_i = v_i r'' r^\ell$, for some $\ell \geq 0$, and $|r'| = |r''|$. Observe here the equalities $r'' r^\ell = w(p'q')^k p'$ and $pu = r^\ell r'$.

Since $|u| \geq |\alpha\beta| \geq |r|$ and $|p| \geq |r|$, then $|pu| \geq |r^2|$, and, consequently, $\ell \geq 2$. Assume first that $r' = r'' = \varepsilon$. Then $v_i \in r^+$ and, consequently, $|v| = |v_i| \geq |r|$. Otherwise, $r' \neq \varepsilon$ and $r'' \neq \varepsilon$.

Let σ be a word such that $r = r'\sigma$. Observe here that since the word r is primitive and the words r' and r'' are, respectively, nonempty proper prefix and nonempty proper suffix of the word r, then both words $r^\ell r'$ and $r'' r^\ell$ are primitive. We have, that $r^\ell r' = \varphi\psi$, $r'' r^\ell = \psi\varphi$ and, by primitiveness of $\varphi\psi$, $v_i \in (\varphi\psi)^* \varphi$, for some words φ and ψ. Then the word $r^\ell r' = \varphi\psi$ is a subword of the word $r'' r^\ell r'' r^\ell = \psi(\varphi\psi)\varphi$. Denote $x = r'' r^\ell r'' r^\ell$. Since the word r is a period of the word $r^\ell r'$ and the word r is primitive, then all occurrences of the word r in the word $r^\ell r'$ are positions which are multiplies of $|r|$. Since $\ell \geq 2$, then $\ell - 1$ of ℓ occurrences of the word r from the word $r^\ell r'$ occurs in the first and in second half of the word $r'' r^\ell r'' r^\ell$. Since the word r is a period of the word $r'' r^\ell$, then these occurrences has to match one of ℓ rs in the first or second half of the word $r'' r^\ell r'' r^\ell$. There cannot be two words rs from the word $r^\ell r'$ one which occur in the first half of the word x and the second one in the second half of the word x, since, by $r' \neq \varepsilon$, the distances between occurrences of the word r in first half and second half of the word x are of the form $|r''| + i|r|$, for some nonegative integer i, and, consequently, they are not multiplies of $|r|$. Hence, these $\ell - 1$ occurrences of the word r are in one of two halves of the word x. Hence, either these are first or last $\ell - 1$ occurrences of the word r in the word $r^\ell r'$.

In the first case $r' = r''$ and the occcurrence is in bold in $x = r' \mathbf{r}^{\ell} \mathbf{r}' r^\ell$. Consequently, $\varphi = r^\ell$ and $\psi = r' = r''$.

In the second case $r'\sigma = r = \sigma r''$ and the occurrence is in bold in $x = r'' r^{\ell-1} r' \sigma \mathbf{r}'' \mathbf{r}^{\ell-1} \mathbf{r}' \sigma r''$. Consequently, $\varphi = \sigma$ and $\psi = r'' r^{\ell-1} r'$.

Summing the two cases we have, $v_i \in (\varphi\psi)^* \varphi \subseteq (r^\ell r')^* \{r^\ell, \sigma\}$. This leads to two cases: either $|v| = |v_i| > |r|$, or $v_i = \sigma$. Together with the case $|v| = |v_i| \geq |r|$ which came from the case $r' = r'' = \varepsilon$, it is enough to consider two cases: either $|v| = |v_i| \geq |r|$, or $v_i = \sigma$.

Case $|v| \geq |r|$. Since the word vp is a suffix of the word $uvp = pqp = s_i = (p'q')^{k+1}p'$, then the words vp and $(p'q')^k p'$ are suffix comparable. Then, the word r is a suffix of the word v, since $vp \in vr^+$ and $(p'q')^k p' \in r^+$ are suffix comparable and $|p| = |r^m| < |p|+|r| \leq |r^n| = |(p'q')^k p'|$. Since $vw = q'p'$ and the word w is a suffix of the word $v_i w = p'q'$, then the words vw and $w(p'q')^k p'$ are suffix comparable. Since the word r is a suffix of the word v, then the words rw and $w(p'q')^k p'$ are sufffix comparable. Since the word rw is not loneger than the word $w(p'q')^k p' \in wr^+$, then the word rw is a suffix of the word $w(p'q')^k p' \in wr^+$. Hence, by primitiveness of r, $w \in r^*$. Consequently, since $r''r^\ell = w(p'q')^k p'$, then $r'' = 1$. Since $|r'| = |r''| = 0$, then $r' = 1$. Hence, $r^\ell v_i = v_i r^\ell$, and, consequently, $v_i \in r^*$. Since $pu = r^\ell r' = r^\ell$, and $p \in r^*$, we have $u \in r^*$. Finally, $v \in r^*$, since $puv_i = pqp = uvp$. We have reached the contradiction, that uv and vw commute.

Case $v_i = \sigma$. Since $r''r^\ell = w(p'q')^k p' \in wr^+$, then $w \in r''r^*$. Since $(p'q')^k p'q'p' = v_i w(p'q')^k p' \in \sigma r'' r^+ \subseteq r^+$ and $(p'q')^k p' \in r^+$, then $q'p' \in r^+$. Hence, $pqp = ((p'q')^k p')(q'p') \in r^+$ and finally, since $p \in r^+$, then $pq \in r^+$. We have again the same contradiction, that $uv = pq$ and $vw = q'p'$ commute. □

As an immediate consequence of our considerations we obtain.

Theorem 19. *Each equation with at most four occurrences of X possesses either at most two or infinitely many solutions. There are equations with at most four occurrences of X which possesses exactly two solutions.*

4. Infinite Solution Set

Let e be a nontrivial equation to solve. We assume that equations in this section contain infinitely many solutions. Then e is of the form $A_0XA_1 \cdots A_kX = XB_0XB_1X \cdots B_k$ with $k \geq 0$ and $A_0 \neq 1$ or of the form $A_0XA_1 \cdots XA_k = XB_0X \cdots B_{k-1}X$ with $k \geq 1$ and $A_0 \neq 1$. An equation in the first form is called *an equation of type (1)*. An equation in the second form is called *an equation of type (2)*.

An equation of type (1) can have a solution only if $\sum_{i=0}^{k} |A_i| = \sum_{i=0}^{k} |B_i|$. An equation of type (2) can have a solution only if $\sum_{i=0}^{k} |A_i| = \sum_{i=0}^{k-1} |B_i|$. The equation e can be split into two equations which are equivalent to e if for some $0 \leq l < k$: $\sum_{i=0}^{l} |A_i| \leq \sum_{i=0}^{l} |B_i|$. Let l be the minimal one. Then the equation of type (1) splits into two ones. The first one is $A_0XA_1 \cdots XA_lX = XB_0XB_1 \cdots XB_{l-1}XB'$ where the word B' is a prefix of the word B_l of length

$$\sum_{i=0}^{l} |A_i| - \sum_{i=0}^{l-1} |B_i|.$$

The second equation is $A_{l+1}XA_{l+2}X \cdots A_kX = B''XB_{l+1} \cdots XB_k$ where $B_l = B'B''$. The equation of type (2) splits into two ones. The first one is

$$A_0XA_1 \cdots XA_lX = XB_0XB_1 \cdots XB_{l-1}XB'$$

where the word B' is a prefix of the word B_l of length $\sum_{i=0}^{l}|A_i| - \sum_{i=0}^{l-1}|B_i|$. The second equation is $A_{l+1}XA_{l+2}X\cdots XA_k = B''XB_{l+1}\cdots B_{k-1}X$ where $B_l = B'B''$. Observe here, that, since the first equation is of type (1), the equation of type (2) can always be splitted. Additionally, for both types of equations, the first equation of two cannot be splitted.

Observation 1. *Any equation e is equivalent to a set of equations of type (1) each of which cannot be splitted.*

Proof. We use the splitting techinque we have described and obtain two equations: an equation of type (1) and another one e'. We reduce in e' a common prefix of both sides of the equation obtaining an equation e''. Next, we continue splitting starting from the equation e''. ☐

Each of equations obtained in the set of equations in Observation 1 is of type (1) that is it is of the form $A_0XA_1\cdots A_kX = XB_0XB_1X\cdots B_k$ with $k \geq 0$ and cannot be splitted so it satisfies additionally $\sum_{i=0}^{l}|A_i| > \sum_{i=0}^{l}|B_i|$, for $0 \leq l < k$. Such an equation is called *irreducible*.

We want to prove that, if e has infinitely many solutions, then the solution set is of the form r^*, for some primitive word r, or of the form $(pq)^*p$, for some primitive word pq and nonempty words p and q. Since each solution of e is a solution of all equations in the system, each irreducible equation in the system has infinitely many solutions. If we prove the theorem for irreducible equations, we prove the theorem for all equations. Indeed, if $|(p_1q_1)^*p_1 \cap (p_2q_2)^*p_2| = \infty$, for some words p_i, q_i, where the words q_i are nonempty and the word p_iq_i is primitive, then $(p_1q_1)^*p_1 = (p_2q_2)^*p_2$.

If $k = 0$, then there is nothing to prove. Assume then that $k \geq 1$.

We define now *elementary* equations. An equation e is *elementary* if

$$\sum_{i=0}^{l}|A_i| > \sum_{i=0}^{l-1}|B_i|, \text{ for each } 0 \leq l \leq k.$$

Observe here that each irreducible equation is elementary but not vice versa. Indeed, the equation $aXbX = XaXb$ is elementary and not irreducible. It can be split into two equations $aX = Xa$ and $bX = Xb$.

Lemma 20. *Assume that an elementary equation has infinitely many solutions of the form r^i where r is a primitive word. Then all words of the form r^i, for $i \geq 0$, are solutions of the equation. Moreover $A_0 = r^m$, for some $m \geq 1$ and $B_k = r^t$, for some $t \geq 1$. Additionally, r is a prefix of B_lr, for $0 \leq l \leq k$, and, for $1 \leq l \leq k$, $B_{l-1}r^j = r^sA_l$, for some j, $s > 0$.*

Proof. Consider the words $w_{i,l} = A_0r^iA_1r^i\cdots A_lr^i$, for $i \geq 0$ and $0 \leq l \leq k$, and the words $u_{i,l} = r^iB_0r^i\cdots r^iB_{l-1}r^i$, for $i \geq 0$ and $0 \leq l \leq k$. We have

$$|w_{i,l}| - |u_{i,l}| = \sum_{s=0}^{l}|A_s| - \sum_{s=0}^{l-1}|B_s|.$$

Observe here that this difference does not depend on i. Since the equation is elementary the word $w_{i,l}$ is longer than the word $u_{i,l}$, for all l.

Claim 1. *For each $i \geq 0$, the word $u_{i,l}$ is a prefix of the word $w_{i,l}$ and the word $u_{i,l}^{-1}w_{i,l}$ is of the form r^s, for $s \geq 1$ and s does not depend on i. Moreover $A_0 = r^m$, for some $m \geq 1$.*

Proof of the claim. We have already proved that s does not depend on i. We prove the claim by induction on l.

First we prove our statement for $l = 0$. We have $w_{i,0} = A_0 r^i$ and $u_{i,0} = r^i$. Since our equation has infinitely many solutions, then, for arbitrary large j, $u_{j,0} = r^j$ is a prefix of $w_{j,0} = A_0 r^j$. Hence $A_0 = r^m$, for some $m \geq 1$. Then, for each $i \geq 1$, $u_{i,0}^{-1}w_{i,0} = r^m$. Hence, the statement for $l = 0$ holds.

Suppose that the induction hypothesis is true, for $l - 1 \geq 0$. Then, for each i, the word $u_{i,l-1}$ is a prefix of the word $w_{i,l-1}$. By induction hypothesis the word $u_{i,l-1}^{-1}w_{i,l-1}$ is of the form r^s for some s and s does not depend on i. Since our equation has infinitely many solutions of the form r^i, there is arbitrarily large number j such that the word $u_{j,l}$ is a proper prefix of the word $w_{j,l}$. Since $w_{j,l} = w_{j,l-1}A_l r^j$ and $u_{j,l} = u_{j,l-1}B_{l-1}r^j$ we have that the word $B_{l-1}r^j$ is a prefix of the word $r^s A_l r^j$. Since r is primitive and the number j is arbitrarily large, the word $(B_{l-1}r^j)^{-1}(r^s A_l r^j)$ is of the form r^n for some $n > 0$. Hence, for each i, $(B_{l-1}r^i)^{-1}(r^s A_l r^i)$ is equal to r^n. Hence, the result. $\square$

Proof of Lemma 20 (continued). Let $l > 0$. By the claim there is $s > 0$ such that, for all i, $w_{i,l-1} = u_{i,l-1}r^s$. As $w_{i,l} = w_{i,l-1}A_l r^j$ and $u_{i,l} = u_{i,l-1}B_{l-1}r^j$, we have that the word $B_{l-1}r^j$ is a prefix of the word $r^s A_l r^j$, and then the word r is a prefix of the word $B_{l-1}r$. Since, for each i, the word $u_{i,l}^{-1}w_{i,l} = (B_{l-1}r^i)^{-1}(r^s A_l r^i)$ is equal to the word r^n with $n > 0$, then $B_{l-1}r^j = r^s A_l$, for some $j > 0$.

We have $w_{i,k} = u_{i,k}r^t$, for some t and all i. Since, for some i, the word r^i is a solution to our equation, then $B_k = r^t$. Hence, for all i, the word r^i is a solution of the equation. Since the word $u_{i,k}$ is a proper prefix of the word $w_{i,k}$, then $t > 0$. $\square$

Lemma 21. *Assume that an irreducible equation has infinitely many solutions of the form $(pq)^i p$ for nonempty words p, q such that the word pq is primitive. Then all words of the form $(pq)^i p$, for $i \geq 0$, are solutions of the equation. Moreover $A_0 = (pq)^m$, for some $m \geq 1$ and $B_k = (qp)^t$, for some $t \geq 1$. Additionally, for each $0 \leq i \leq k$, the word qp is a prefix of $B_i pq$, and the word pq is a suffix of qpA_i.*

Proof. Let e be the equation. Make a substitution which replaces the variable X by the word pX. Since the equation e is irreducible, then the new equation e' is elementary. It has infinitely many solutions of the form $(qp)^i$ where the word qp is primitive. By Lemma 20 all words $(qp)^i$, for $i \geq 0$, are solutions to the equation e'. Hence, all words $(pq)^i p$, for $i \geq 0$, are solutions of e.

The equation e' has infinitely many solutions of the form $(qp)^i$. By Lemma 20 $B_k = (qp)^t$, for some $t \geq 1$. Since the equation e is irreducible, the equation e'

after reduction of constants from the front of its sides is of type (1). Denote its coefficients by A'_i and B'_i. Then $A'_0 = p^{-1}A_0p = (qp)^m$, for some integer $m \geq 1$. Hence, $A_0 = (pq)^m$.

Moreover, by Lemma 20 the word qp is a prefix of the word $B'_iqp = B_ipqp$. Hence, the word qp is a prefix of the word B_ipq. Clearly, the word pq is a suffix of the word qpA_0. Additionally $B'_{i-1}(qp)^j = (qp)^sA'_i$, for $1 \leq i \leq k$ and some j, $s > 0$. Hence, $B_{i-1}p(qp)^j = (qp)^sA_ip$. Consequently, the word pq is a suffix of the word qpA_i. □

Let e be an irreducible equation. Then $|B_0| < |A_0|$. Let B' be the prefix of the word A_0 of length $|A_0| - |B_0|$. We distinguish two types of equations: equations where $A_0 = B_0B'$ and equations where $A_0 \neq B_0B'$. The equations of the first type are called *of type* r and the equations of the second type are called *of type* (p, q) where r, p, q are nonempty words which will be determined later by A_0 and B_0B'.

We divide solutions of the equation e into two groups which are called *long solutions* and *short solutions*. *Long solutions* are those which are not shorter than $|B'|$. *Short solutions* are those which are shorter than $|B'|$. Since all solutions are prefixes of the word A_0^ω, then long solutions start by the word B'. Hence they all satisfy $A_0X = XB_0B'$ where B' comes from the prefix of the X which stays in equation just to the right of the word B_0. If our equation has a long solution then A_0 and B_0B' are conjugates. Hence, either $B_0B' = A_0 = r^m$, for some primitive word r and $m \geq 1$, or $A_0 = (pq)^m$, $B_0B' = (qp)^m$, for $m \geq 1$ and some nonempty words p, q such that the word pq is primitive. In the former case the equation is called *of type* r. In the latter case the equation is called *of type* (p, q).

As a result of our previous considerations we have

Lemma 22. *Let e be an irreducible equation. Assume that the equation e has infinitely many solutions. If the equation is of type r, then all words in r^* are solutions for it. If the equation is of type (p, q), then all words in $(pq)^*p$ are solutions for it. In both cases there are no other long solutions.*

By Lemma 22 it is enough to prove that there are no short solutions except the ones mentioned in the lemma.

In the following subsections the meaning of B', m, t, A_i, B_i are the same as in this section. We recall it now. Considered irreducible equations are in form $A_0XA_1X \ldots A_kX = XB_0XB_1 \ldots XB_k$. The word B' is the prefix of the word A_0 of length $|A_0| - |B_0|$. m is a positive integer such that $A_0 = r^m$, if the eqaution is of type r and $A_0 = (pq)^m$, if the equation is of type (p, q). t is a positive integer such that $B_k = r^t$, if the equation is of type r and $B_k = (qp)^t$, if the equation is of type (p, q).

4.1. *Equations of type r*

Theorem 23. *Let e be an irreducible equation. Let e be an equation of type r with infinitely many solutions. Then the solution set of e is r^*.*

Proof. Observe first that, if a word x is a short solution, then the word xB_0 is a proper prefix of the word A_0. Since the equation is of type r, we have $r^m = A_0 = B_0B'$ where the word B' is a prefix of the word $A_0 = r^m$.

If $|B_0| \geq |r|$, then since the word xB_0 is a prefix of the word $A_0 = r^m$ and r is a primitive word, we have $x = r^n$, for some n. In this case we do not obtain new solutions. If $|B'| \geq |r|$, then the word B' starts with the word r. Assume additionally that $B_0 \neq 1$. Since the word B' is a suffix of the word r^m, $B_0 = r^l$, for some $l \geq 1$. Hence, we returned to considered case $|B_0| \geq |r|$. It remains to consider two cases: either $B_0 = 1$, or $0 < |B_0| < |r|$ and $|B'| < |r|$.

In the latter case $|r^m| = |B_0B'| < 2|r|$. Hence, $m = 1$. We have $r = A_0 = B_0B'$ and the word B' is a prefix of the word r. Hence, the word B_0 is a period of the word r. Let $B_0 = \rho^s$ for some primitive word ρ. Since the word xB_0 is a prefix of the word $A_0 = r$, $x = \rho^k$ for some $k \geq 0$. If $k = 0$, then $x = 1 \in r^*$. Otherwise, $k \geq 1$. We have that the word x is a suffix of the word $B_k = r^t$. Hence, the word ρ is a suffix of the word r. Moreover, the word ρ is also a prefix of the word r and its period. This contradicts primitiveness of r.

The last case is $B_0 = 1$. Let l be the minimal index such that $B_l \neq 1$. Such an index exists, since $|B_k| = |r^t| > 0$. If $l < k$, then, by Lemma 20, $B_lr^j = r^sA_{l+1}$ with $s \geq 1$. If $l = k$ then $B_l = r^t$. Hence, either the word r is a prefix of the word B_l or the word B_l is a period of the word r.

Let x be a nonempty short solution. Then the words A_0x and $x^{l+1}B_l$ with $l \geq 1$ are prefix comparable. Since the word x is a short solution we have $|x| \leq |A_0|$. Then the word $xB_0x = xx$ is a prefix of the word $A_0x = r^mx$. Since the word x is a prefix of the word $A_0 = r^m$, the word r is a period of the word $A_0x = r^mx$. Hence, if $|x| \geq |r|$, then the word xx has two periods: x and r. We have $|x| + |r| \leq |xx|$. Hence, by Proposition 1 and primitiveness of r, $x = r^n$, for some $n \geq 1$. We may thus assume that $|x| < |r|$. Let $x = \rho^d$, for some primitive word ρ. If the word x is a period of the word r, then the word ρ is a period of the word r. Then the word ρ is a prefix of the word r and since the word x is a suffix of the word $B_k = r^t$, the word ρ is also a suffix of the word r. This contradicts primitiveness of the word r. Hence, the word x is not a period of the word r.

We have that the words A_0x and $x^{l+1}B_l$ are prefix comparable and $|x| < |r|$. Hence the words r and $x^{l+1}B_l$ are prefix comparable. There are three cases to consider: the word r is a prefix of the word x^{l+1}, or $|x^{l+1}| < |r| \leq |x^{l+1}B_l|$, or the word $x^{l+1}B_l$ is a prefix of the word r. In each case we will prove that the word x is a period of the word r which raises a contradiction.

Case the word r is a prefix of the word x^{l+1}. Then the word x is a period of the word r.

Case $|x^{l+1}| < |r| \leq |x^{l+1}B_l|$. Recall that either the word B_l starts with the word r or the word B_l is a period of the word r. Then, if the word B_l starts with the word r, then the word x is a period of the word r. If the word B_l is a period of the word r, then it is a prefix of the word r and the word x is a period of the word $x^{l+1}B_l$ and, consequently, the word x is a period of the word r. In all cases the word x is a period of the word r.

Case the word $x^{l+1}B_l$ is a proper prefix of the word r. Hence, $|B_l| < |r|$ and, consequently, as we concluded earlier, the word B_l is a period of the word r. Let $B_l = \rho^t$ for some primitive word ρ. Then $x^{l+1} = \rho^j$, for some $j \geq 1$. Since the word ρ is primitive $x = \rho^i$, for some i. Hence, again the word x is a period of the word r. □

4.2. *Equations of type* (p, q)

We start by proving a corollary of Theorem 23.

Corollary 24. *Let e be an irreducible equation of type (p, q) with infinitely many solutions. Then all solutions of e which are not shorter than p form the set $(pq)^*p$.*

Proof. Let e' be an equation which can be obtained from e after replacing the unknown X by the word Xp. Then e' is splitted into a set of irreducible equations of type (p, q) with infinitely many solutions. Hence, the solution set of e' is $(pq)^*$. Since each solution of e which is not shorter than p is of the form xp, such solutions form the set $(pq)^*p$. □

By Corollary 24 it is enough to prove that there are no short solutions which are shorter than p. To prove this we need a technical lemma.

Lemma 25. *Let e be an irreducible equation of type (p, q) with infinitely many solutions. Then the empty word is not a solution for the equation e.*

We use Lemma 25 to prove our main result of this section.

Theorem 26. *Let e be an irreducible equation of type (p, q) with infinitely many solutions. Then all solutions form the set $(pq)^*p$.*

Proof. We use notation from the end of Section 4. By Corollary 24 it is enough to prove that the equation e has no solutions shorter than p. Suppose on the contrary, that a prefix of p, x with $|x| < |p|$, is a solution of e. Replace in equation e each occurrence of the unknown X by the word xX obtaining an equation e'. The equation e' has infinitely many solutions and the empty word is a solution of it. Now apply splitting to the equation e'. We obtain an irreducible equation e'' and another equation.

The equation e'' has infinitely many solutions and the empty word is a solution of it. Let A'_i, B'_i, for $0 \leq i \leq l$, be the constants of the equation e''. Then $A'_0 =$

$x^{-1}(pq)^m x$. Additionally, if $|A_0| > |B'_0|$, then $B'_0 = B_0 x$. Otherwise, the word B'_0 is a prefix of the word $B_0 x$ of length $|A_0|$ and, consequently, $B'_0 = B_0 B' = (qp)^m$.

In the former case, let B'' be the prefix of A'_0 of length $|A'_0| - |B'_0|$. Then $B'' = x^{-1}B'$. Hence, $B'_0 B'' = B_0 B' = (qp)^m$. Since $|x| < |p|$ and the word pq is primitive, then $A'_0 = x^{-1}(pq)^m x \neq (qp)^m = B'_0 B''$. Consequently, the equation e'' is an equation of type (p', q'). By Lemma 25 it is impossible.

In the latter case $B'_0 = (qp)^m$ and the equation e'' is of the form $A'_0 X = X B'_0$. Again, since $|x| < |p|$ and the word pq is primitive, the empty word is not a solution of e''. A contradiction. □

4.3. *The proof of Lemma 25*

We use notation from the end of Section 4. We divide the proof into three cases. The first case is $|B_0| \geq |q|$ and $|A_k| \geq |q|$. The second one is $0 < |B_0| < |q|$ or $0 < |A_k| < |q|$. The third one is $B_0 = 1$ or $A_k = 1$.

First we eliminate the third case. Suppose that $B_0 = 1$, the case $A_k = 1$ being symmetric. Then $(qp)^m = B_0 B' = B'$. Since B' is a prefix of $A_0 = (pq)^m$, then $A_0 = B'$ and, consequently, $qp = pq$. Hence, pq is not primitive raising a contradiction.

Now, observe that the subcase of the second case: $0 < |B_0| < |q|$ and the subcase: $0 < |A_k| < |q|$ are symmetric since we may consider the equation in which both sides are reversed sides of the original equation. Hence, it is enough to consider the subcase $0 < |B_0| < |q|$.

4.3.1. *Case* $|B_0| \geq |q|$ *and* $|A_k| \geq |q|$

Recall that, by Lemma 21, for each $0 \leq i \leq k$, the word pq is a suffix of the word qpA_i and the word qp is a prefix of the word $B_i pq$.

Lemma 27. *Let e be an irreducible equation of type (p, q) possessing infinitely many solutions. If $|B_0| \geq |q|$ and $|A_k| \geq |q|$, then the empty word is not a solution of e.*

Proof. Assume on the contrary, that the empty word is a solution of e. Let r be the smallest index $k \geq r \geq 0$ such that $B_r \neq q$. Such an index exists, since $B_k = (qp)^t$ with $t \geq 1$ is longer than q. Observe here that if $r = k$, then $|B_r| = |(qp)^t| \geq |pq|$. Hence, if $|B_r| < |pq|$, then $r < k$. Observe also that, since $B_r \neq q$ and the word qp is a prefix of the word $B_r pq$, we may assume that $|B_r| \neq |q|$.

Since the empty word is a solution of e, we have $\cdots A_{k-1}A_k = \cdots B_{k-1}B_k$. Since the word pq is a suffix of the word qpA_k and A_k is longer than q and $B_k = (qp)^t$, then the words p and q are suffix comparable.

Case $|B_r| \geq |pq|$. Since the word qp is a prefix of the word $B_r pq$, then $B_r = qpw$, for some word w. Since the empty word is a solution of e, we have $A_0 A_1 \cdots = B_0 B_1 \cdots$. Hence, $(pq)^m \cdots = q^r qpw \cdots$ with $m \geq 1$ and $r \geq 0$. Consequently, by

Fact 1, the words p and q are powers of the same word and the word pq is not primitive raising a contradiction..

Case $|q| < |B_r| < |pq|$ **and** $|q| \geq |p|$. As we concluded earlier, we have $r < k$. We have $pq \cdots = A_0 A_1 \cdots = B_0 B_1 \cdots = q^r q \cdots$. Since, additionally, $|q| \geq |p|$, by Fact 1, the word p is a period of the word q. Since the words p and q are suffix comparable and $|q| \geq |p|$, the word p is a suffix of the word q. Since the word p is a period of the word q and a suffix of the word q, then, by Fact 1, the words p and q are powers of the same word. Hence, the word pq is not primitive raising a contradiction.

Case $|B_r| < |q|$ **and** $|q| \geq |p|$. Again we have $r < k$. Since $|B_0| \geq |q|$, we have $r \geq 1$. Hence, $B_0 = q$ and, consequently, $pq \cdots = A_0 A_1 \cdots = B_0 B_1 \cdots = q \cdots$. We continue as in the previous case and raise a contradiction.

Case $|q| < |B_r| < |qp|$ **and** $|q| < |p|$. Again we have $r < k$. Since the word qp is a prefix of the word $B_r pq$, we have $B_r = qp'$, for some nonempty proper prefix p' of the word p. Again since the word qp is a prefix of the word $B_r pq$, the word p' is a period of the word p. We have $pq \cdots = A_0 A_1 \cdots = B_0 B_1 \cdots = q^r qp' \cdots$.

Either $|q^r qp'| \geq |p|$ or $|q^r qp'| < |p|$. In the former case, since $|q| < |p|$ and the word p' is a prefix of the word p, then, by Fact 1, the word q is a period of the word p. Since the words q and p are suffix comparable, the word q is a suffix of the word p. Hence, the words p and q are powers of the same word and, consequently, the word pq is not primitive raising a contradiction.

In the latter case, since the word p' is a period of the word p and the occurrence of the word p' in the word $q^r qp'$ is a subword of the word p, we have $root(p') = root(q)$. Since the word p' is a period of the word p, this means that the word q is a period of the word p. Since the words q and p are suffix comparable and $|q| < |p|$, the word q is a suffix of the word p. Hence, the words p and q are powers of the same word and, consequently, the word pq is not primitive raising a contradiction.

Case $|B_r| < |q|$ **and** $|q| < |p|$. Since the words p and q are suffix comparable, the word q is a suffix of the word p. As we concluded we have also $r < k$. Since the word qp is a prefix of the word $B_r pq$, $q = B_r p'$ and the word p' is a nonempty proper prefix of the word p. Since the word $qp = B_r p' p$ is a prefix of the word $B_r pq$, the word p' is a period of the word p. Since the word B_r is shorter than the word q, then $r > 0$. Since the empty word is a solution of e, then $pq \cdots = A_0 A_1 \cdots = B_0 B_1 \cdots = q^r B_r \cdots$ with $r > 0$. Since $r > 0$, we have $pq \cdots = q \cdots = B_r p' \cdots$. Since $|q| < |p|$, the word $q = B_r p'$ is a prefix of the word p. Since the word p' is a period of the word p and the occurrence of the word p' in the word q is a subword of the word p, then $root(p') = root(q)$. Hence, the word q is a period of the word p. Since, additionally the word q is a suffix of the word p, then the words q and p are powers of the same word. Consequently, the word pq is not primitive raising a contradiction. $\square$

4.3.2. *Case* $0 < |B_0| < |q|$

Denote by $u \wedge v$ the longest common prefix of the words u and v. Let x and y be two nonempty words. Let $w_1 \in x(x \cup y)^\omega$ and $w_2 \in y(x \cup y)^\omega$ be two infinite words. Since $w_1 \wedge w_2 = x^\omega \wedge y^\omega$ we have

Fact 3. *Let x and y be two nonempty words. Let $w_1 \in x(x\cup y)^*$ and $w_2 \in y(x\cup y)^*$. Then the word x is a period of each longer than $|x|$ common prefix of w_1 and w_2.*

Lemma 28. *Let e be an equation of type (p, q) possessing infinitely many solutions. If $0 < |B_0| < |q|$, then the empty word is not a solution to e.*

Proof. Assume on the contrary that the empty word is a solution of e. Let r be the smallest index such that $B_r \neq B_0$. Such an index exists since $|B_k| = |(qp)^t| \geq |qp| > |q| > |B_0|$. The word qp is a prefix of both words B_0pq and B_rpq. Since the empty word is a solution of e, then $pq \cdots = A_0A_1 \cdots = B_0B_1 \cdots = B_0^rB_r \cdots$ with $r \geq 1$. Hence, the word qp is prefix comparable with the words $w_1 = B_0B_0^rB_r$ and $w_2 = B_rB_0^rB_r$.

If the word qp is shorter than both words w_1 and w_2, then, by Fact 3, the word B_0 is a period of the word qp.

If the word qp is longer than the word w_1 or the word w_2, then the words w_1 and w_2 are prefix comparable. Hence, by Fact 1, $root(B_0) = root(B_r)$. Denote $s = root(B_0)$. Then $B_0pq = s^ipq$ and $B_rpq = s^jpq$, for some $i \neq j$. Suppose that $i > j$, the other case being symmetric. Then the word $B_rpq = s^jpq$ is a suffix of the word $B_0pq = s^ipq$. Since the word qp is a prefix of both words B_rpq and B_0pq, then the word qp is a prefix of the word $s^{i-j}qp$. Hence, the word s^{i-j} is a period of the word qp. Hence, the word B_0 is a period of the word qp.

In both cases the word B_0 is a period of the word qp. Since the word qp is a prefix of the word B_0pq, then the word B_0p which is shorter than the word qp, is also a period of the word qp. Hence, by Fact 1, the word p is a period of the word $w = B_0^{-1}qp$. Let $d = root(p)$. Then $p = d^j$ for some $j \geq 1$ and, consequently, the word d is a period of the word $w = B_0^{-1}qd^j$. Since the word d is primitive and the word d is a suffix of the word w, we have $w = d^l$. Hence, $q = B_0d^i$, for some $i \geq 1$. Observe here that, since $B_r \neq B_0$ and the word qp is a prefix of the words B_0pq and B_rpq, and $|B_0| < |q|$, we may assume that $|B_0| \neq |B_r|$. We consider four cases.

Case $|B_r| \geq |q|$. Since the word qp is a prefix of the word B_rpq, then $B_r = q \cdots = B_0s^i \cdots$. Hence, the word s^jB_0 is a prefix of the word pq and the word $B_0^{r+1}s$ is a prefix of the word $B_0^{r+1}s^i = B_0^rB_r$. Since the empty word is a solution of the equation e, the words pq and $B_0^rB_r$ are prefix comparable. Hence, the words $B_0^{r+1}s$ and s^jB_0 are prefix comparable. Consequently, by Fact 1, the words B_0 and s are powers of the same word. Since the word s is primitive, $B_0 = s^o$. Consequently, $q = s^{o+i}$ and finally the words p, q are powers of s. Hence, the word pq is not primitive raising a contradiction.

Case $0 < |B_r| < |B_0| < |q|$. We may apply the same reasoning as we did for B_0 to prove that $q = B_r d^e$, for some $e > i$. Hence, $B_0 = B_r d^{e-i}$. We have that the word $d^j B_r$ is a prefix of the word pq and the word $B_0 = B_r d^{e-i}$ is a prefix of the word $B_0^r B_r$. Since the empty word is a solution of e, then the words pq and $B_0^r B_r$ are prefix comparable, and, consequently, the words $d^j B_r$ and $B_r d^{e-i}$ are prefix comparable. By primitivity of d and by Fact 1, the word B_r is a power of the word d. Consequently, the word q is a power of the word d contradicting the primitiveness of pq.

Case $0 < |B_0| < |B_r| < |q|$. We may apply the same reasoning as we did for B_0 to prove that $q = B_r d^e$, for some $e < i$. Hence, $B_r = B_0 d^{i-e}$. We have that the word $d^j B_0$ is a prefix of the word pq and the word $B_0^{r+1} d^{i-e}$ is a prefix of the word $B_0^r B_r$. Since the empty word is a solution of e, then the words pq and $B_0^r B_r$ are prefix comparable, and, consequently, the words $d^j B_0$ and $B_0^{r+1} d^{i-e}$ are prefix comparable. By primitivity of d and by Fact 1, the word B_0 is a power of the word d. Consequently, the word q is a power of the word d contradicting the primitiveness of pq.

Case $B_r = 1$. Then the word qp is a prefix of the word $B_r pq = pq$. Hence, $pq = qp$, contradicting the primitiveness of pq. □

5. A Family of Equations with Constant Number of Occurrences of *X*

Let $\mathcal{E}_k$ be the family of equations containing at most $2k$ occurrences of the unknown X and possessing finitely many solutions. In this section we prove the following theorem, which is a consequence of Lemma 33 and Lemma 34.

Theorem 29. *If $e \in \mathcal{E}_k$, then it possesses at most $8 \log k + O(1)$ solutions. A solution set of an equation in this class can be found in linear time.*

Consider $e \in \mathcal{E}_k$. If it is not irreducible, then it splits into irreducible equations, each of which contains at most $2k$ occurrences of X. Hence, it is enough to consider irreducible equations of the form

$$A_0 X A_1 \cdots A_{k-1} X = X B_0 X B_1 \cdots X B_{k-1}.$$

Assume additionally that $|B_{k-1}| \geq |A_0|$, the other case being symmetric.

Denote by $B(x)$ the word $xB_0xB_1x \cdots xB_{k-2}x$. Similarly, denote by $A(x)$ the word $xA_1xA_2 \cdots xA_{k-1}x$. We divide all solutions of e into two sets: the solutions x such that $|B(x)| \leq |A_0|$ and the solutions x such that $|B(x)| > |A_0|$. Denote by t the number $|B(1)| = |B_0 B_1 \cdots B_{k-2}|$.

Our proof is constructive in the sense that we are able to find a solution set of the equation in linear time.

The following proposition was proved in [7].

Proposition 30. *Let p and q be two words. Assume that the word pq is primitive and $p \neq 1$. Let e be an equation with one unknown. Let $Sol(e)$ be the solution set* *of the equation e. Let $T = Sol(e) \cap (pq)^*p$. Then either $T = (pq)^*p$ or $T = (pq)^+p$ or $T \subseteq \{p, (pq)^i p\}$ for some $i \geq 1$. Moreover, the set T can be found in linear time.*

Remark. We proved that the case $T = (pq)^+p$ in Proposition 30 is not possible.

In [7], the following proposition was implicitely proved.

Proposition 31. *Let $e : sX... = XtX...$ be an equation with one unknown X and let m be an integer. Then there are at most two solutions $X = x$ such that $m/2 < |tx| \leq m$.*

In [3], the following proposition was proved.

Proposition 32. *$O(\log n)$ candidates for short solutions of an equation with one unknown can be found in linear time independently of the size of the alphabet. The candidates satisfy the constraint in Proposition 31.*

5.1. *The solutions x such that $|B(x)| \leq |A_0|$*

Lemma 33. *Let S_1 be the set of solutions x of e such that $|B(x)| \leq |A_0|$. Then $|S_1| \leq 2\log k + O(1)$. Moreover, the set S_1 can be found in linear time.*

Proof. If the set S_1 is nonempty, then $t = |B(1)| \leq |A_0|$. Denote by s the suffix of the word A_0 of length $|A_0| - t$. If $x \in S_1$, then the word $B(x)$ is a prefix of the word A_0 and there is an overlap between the words s and B_{k-1} of length $|A_0| - |B(x)| = |s| - k|x|$. Consequently, $|x| \leq \frac{|s|}{k}$.

All overlaps between the words s and B_{k-1} can be found in linear time using failure table from Knuth-Morris-Pratt string-matching algorithm for the word $B_{k-1}\$s$ where \$ is a new symbol which does not occur in the alphabet. Let i_1, i_2, ..., i_l be start positions in s of overlaps between B_{k-1} and s.

Suppose first, that there is at most one index j such that $i_j \leq \frac{|s|}{2}$. Then, for $j \geq 2$, $i_j > \frac{|s|}{2}$ and, consequently, if the word x is not the shortest solution in the set S_1, then $|A_0| - |B(x)| = |s| - k|x| = |s| - i_j$, for some $j \geq 2$, and, consequently, $|x| > \frac{|s|}{2k}$. By Proposition 31, there are at most two solutions in the set S_1 such that longer than $\frac{|s|}{2k}$ and shorter than $\frac{|s|}{k}$. By Proposition 32 they can be chosen in linear time from the candidates by checking candidates of appropriate lengths.

Suppose now, that there are at least two indices j such that $i_j \leq \frac{|s|}{2}$. Suppose these are $i_1, \dots i_n$. Then, for $j < n$, the number $i_{j+1} - i_j$ is a period of the second half of the word s, see Fig. 7. We have $(i_{j+2} - i_{j+1}) + (i_{j+1} - i_j) = i_{j+2} - i_j \leq i_{j+2} \leq \frac{|s|}{2}$, for $j < n-1$. Hence, we may apply Proposition 1 to periods $i_{j+1} - i_j$ and $i_{j+2} - i_{j+1}$ of the second half of the word s, obtaining $i_{j+2} - i_{j+1} = i_{j+1} - i_j$, for $j < n - 1$. Denote $d = i_{j+1} - i_j$, for $j < n$. Each $i_j + t$, for $j \leq n$, is a candidate for $|B(x)| + 1$. Hence, each i_j, for $j \leq n$, uniquely determines a candidate for $|x|$, and hence it determines a prefix of the word A_0^ω equal to x. Having such a candidate we have to check whether it is in the set S_1. Then we say that i_j *corresponds to* $x \in S_1$.

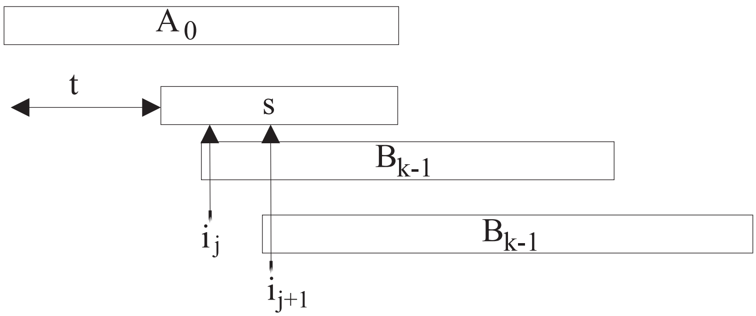

Fig. 7. $i_{j+1} - i_j$ is a period of second half of s.

We check first whether i_1 corresponds to $x \in S_1$. Clearly, it can be done in linear time.

If i_j, for some $j \geq 2$, corresponds to $x \in S_1$, then $|B(x)| = t + k|x| \geq t + d$. Hence, $|x| \geq \frac{d}{k}$ and, consequently, by Proposition 31, there are at most $2 \log k + 1$ solutions in the set S_1 which are shorter than d. By Proposition 32, they can be selected from the set of candidates in linear time by checking each candidate longer than $\frac{d}{k}$ and shorter than d.

Consider any solution $x \in S_1$ which is of length at least d. Let u and v be, respectively, the prefix of length d and the suffix of length d of x. Since the word x is a prefix of the word A_0 and a suffix of the word B_{k-1}, then the words u and v do not depend on the choice of x.

If the number d is a period of the word x, then the words u and v are conjugates and, consequently, there is at most one pair of words p, q with $q \neq 1$ and pq primitive such that $u = (pq)^m$ and $v = (qp)^m$. Then, since the word x is longer than the word $u = (pq)^m$, $x = (pq)^i p$, for some $i \geq 1$. By Proposition 30 there is at most one i such that the word x is a solution of e. Hence, there is at most one solution in the set S_1 such that d is its period and it can be found in linear time.

If the number d is not a period of the word x, then the period d of a suffix of the word x breaks somewhere in x. Since the word x is shorter than the word A_0, it is a suffix of the word B_{k-1}. Hence, the distance p between the position of break in x and the end of x does not depend on the chosen x. The word $B(x)$ ends in the word A_0 to the right of the position $i_2 = i_1 + d$. Hence, the suffix v of the last x in $B(x)$ is inside s. Hence, the periods of length d of s and a suffix of v has to match. We may find the place in A_0 from the right end of A_0 where this period is broken. The positions of breaks of periods in last x of the word $B(x)$ and in A_0 have to be the same. Hence, the word $B(x)$ has to end exactly p positions to the right of this place. This position does not depend on x. Hence, we have only one possible candidate for the number $|B(x)|$. Hence, the number of x in this case is at most one and a candidate for such x can be found in linear time. □

5.2. *The solutions x such that $|B(x)| > |A_0|$*

Lemma 34. *Let S_2 be the set of solutions x of e such that $|B(x)| > |A_0|$. Then $|S_2| \leq 6\log k + O(1)$. Moreover, the set S_2 can be found in linear time.*

Proof. Let $x \in S_2$. Assume that $|x| \geq |A_0|$. Since the equation e is irreducible, then $|A_0| > |B_0|$. Let B' be a prefix of the word A_0 of length $|A_0| - |B_0|$. Then $A_0x = xB_0B'$. Hence, there is unique pair of words p, q with $q \neq 1$ and pq primitive such that $A_0 = (pq)^m$, $B_0B' = (qp)^m$ and $x = (pq)^ip$, for some $i \geq 1$. By Proposition 30, there is at most one i such that x of the form $(pq)^ip$ is a solution of e and it can be found in linear time.

Assume that $\frac{|A_0|}{(2k+1)k} \leq |x| < |A_0|$. By Proposition 31 there are at most $4\log k+2$ solutions x in this case. By Proposition 32 they can chosen from the set of candidates in linear time.

It remains to prove that the number of $x \in S_2$ satisfying $|x| < \frac{|A_0|}{(2k+1)k}$, is at most $2\log k + O(1)$ and that they can be chosen from the set of candidates in linear time. Denote the set of such x by S_3. We have $|B(x)| = t + k|x| > |A_0|$. Hence, $\frac{|A_0|-t}{k} < |x| < \frac{|A_0|}{(2k+1)k}$. Let $c_1 = \frac{1}{2k+1}$. Then $|x| < c_1\frac{|A_0|}{k}$.

We have $\frac{|A_0|-t}{k} < c_1\frac{|A_0|}{k}$. Hence, $t > (1-c_1)|A_0|$. Since $t = |B_0 \cdots B_{k-2}|$, there is an index i such that $|B_i| > (1-c_1)\frac{|A_0|}{k}$. Let l be the minimal such index. Clearly, l can be found in linear time. Let B'_l be length $\lceil (1-c_1)\frac{|A_0|}{k}\rceil$ prefix of the word B_l.

Now we present the claim the proof of which is left to the reader.

Claim 2. *(i) Let t be a real number and k a positive integer. Then*

$$(k-1)\lfloor \frac{t}{k} \rfloor + \lceil \frac{t}{k} \rceil \leq \lceil t \rceil$$

(ii) Let q be an integer and t be a real number. Then

$$\lceil q - t \rceil + \lfloor t \rfloor = q$$

By point (i) of Claim 2, we have $|B_0 \cdots B_{l-1}B'_l| \leq (k-1)\lfloor (1-c_1)\frac{|A_0|}{k} \rfloor + \lceil (1-c_1)\frac{|A_0|}{k} \rceil \leq \lceil (1-c_1)|A_0| \rceil$. By point (ii) of Claim 2, we have

$$|xB_0xB_1x\cdots xB_{l-1}xB'_l| \leq |B_0 \cdots B_{l-1}B'_l| + k|x| \leq \lceil (1-c_1)|A_0| \rceil + \lfloor c_1|A_0| \rfloor = |A_0|.$$

Denote $B'(u) = uB_0uB_1u\cdots uB_{l-1}uB'_l$. We have just proved that $|B'(x)| \leq |A_0|$. Hence, the word $B'(x)$ is a prefix of the word A_0.

Let $d = |B'(x)| - |B'(1)|$. Then

$$d = l|x| \leq k|x| \leq c_1|A_0|$$

Since $c_1|A_0| = \frac{1}{2}\frac{(1-c_1)|A_0|}{k}$, we have $d \leq \frac{1}{2}|B'_l|$. Since the word $B'(x)$ is a prefix of the word A_0 and the word B'_l is a suffix of the word $B'(x)$, the word B'_l occurs in the word A_0 at position i such that $|B'(1)| - |B'_l| + 1 \leq i \leq |B'(1)| - |B'_l| + 1 + \frac{1}{2}|B'_l|$. All occurrences of the word B'_l between these positions can be found in linear time by any linear time string matching algorithm that searches the word B'_l in the word

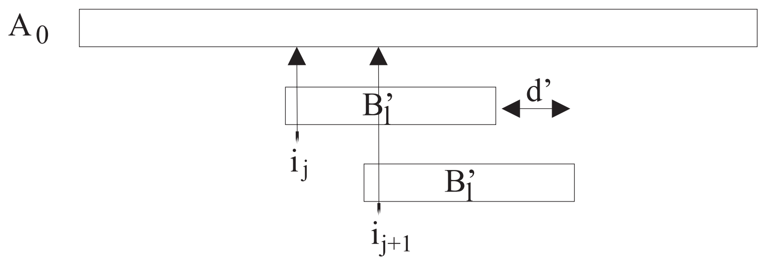

Fig. 8. The occurrences of B'_l at positions i_{j+1} and i_j overlap.

A_0. Let $i_1 < \cdots < i_r$ be such positions. Since $i_r - i_1 \le \frac{1}{2}|B'_l|$ each $i_{j+1} - i_j$ is a period of the word B'_l of length at most half of B'_l, see Fig. 8. By Proposition 1 such a period is a multiple of the shortest period d' of B'_l. Hence, $i_{j+1} - i_j = d'$, for all j and, consequently, $i_j = i_1 + (j-1)d'$.

Each of i_j, for $1 \le j \le r$, is a candidate for $|B'(x)| - |B'_l| + 1$, for $x \in S_3$, and hence uniquely determines $|x|$ and, consequently, the prefix of A_0^ω which is equal to x. If $x \in S_3$, then we say that i_j *corresponds* to $x \in S_3$.

First we verify whether i_1 corresponds to an element $x \in S_3$. It can be done in linear time.

Assume that i_j, for some $j \ge 2$, corresponds to an element x of S_3. Then $|B(1)| - |B'_l| + 1 + l|x| = |B'(x)| - |B'_l| + 1 = i_j$. Hence, $k|x| > l|x| \ge i_j - (|B(1)| - |B'_l| + 1) \ge d'$. Consequently, $|x| > \frac{d'}{k}$. Hence, by Proposition 31 the number of solutions of length at most d' is at most $2\log k + 1$. By Propositions 31 and 32, candidates for solutions of length at most d' can be verified in linear time.

Consider any solution $x \in S_3$ which is of length at least d'. Let u and v be, respectively, a prefix and a suffix of the word x of length d'. Since the word x is a prefix of the word A_0 and a suffix of the word B_{k-1}, the words u and v do not depend on the choice of x.

If the number d' is a period of the word x, then the words u and v are conjugates and, consequently, there is at most one pair of words p, q with $q \ne 1$ and pq primitive such that $u = (pq)^m$ and $v = (qp)^m$. Then, since the word x is longer than the word $u = (pq)^m$, $x = (pq)^i p$, for some $i \ge 1$. By Proposition 30 there is at most one i such that x is a solution of e. Hence, there is at most one solution in S_3 such that d' is its period. By Proposition 30 it can be found in linear time.

If the number d' is not a period of the word x, then the period d' of a suffix of the word x breaks somewhere in the word x. Since the word x is shorter than the word A_0, it is a suffix of the word B_{k-1}. Hence, the distance p between the position of break in the word x and the end of the word x does not depend on the chosen x. The word $B'(x)(B'_l)^{-1}$ ends in A_0 between positions $i_2 = i_1 + d'$ and i_r. Hence, the suffix v of the last x in $B'(x)$ is in the periodic part of A_0 between positions i_2 and i_r. We find a position p' to the left of i_2 where the period d' breaks. The position of break of the period d' in last x of $B'(x)(B'_l)^{-1}$ and the position p' in A_0

has to be the same. Hence, $B'(x)(B'_l)^{-1}$ has to end exactly p positions to the right of this place. This position does not depend on x. Hence, we have only one possible candidate for the number $|B'(x)|$. Hence, the number of x in this case is at most one and such a solution can be found in linear time.

We have just proved that $|S_3| \leq 2\log k + O(1)$. □

References

[1] F. Baader and W. Snyder. Unification theory. In J.A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning Vol. 1*, pages 447–533. Elsevier Science Publishers, 2001.

[2] S. Bala. *Decision Problems on Regular Expressions*. PhD thesis, Universytet Wroclawski, Wroclaw, Poland, 2005.

[3] R. Dabrowski and W. Plandowski. On word equations in one variable. In *Proceedings of Mathematical Foundations of Computer Science MFCS'02, Lecture Notes in Computer Science 2420*, pages 212–221, 2002.

[4] A. Gleibman. Knowledge representation via verbal description generalization: alternative programming in sampletalk language. In *Proceedings of Workshop on Inference for Textual Question Answering, Pittsburgh, Pennsylvania. AAAI-05 - the Twentieth National Conference on Artificial Inteligence.*, pages 59–68, 2005. See also `www.sampletalk.com`.

[5] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, 1997.

[6] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2002.

[7] S. Eyono Obono, P. Goralcik, and M. N. Maksimenko. Efficient solving of the word equations in one variable. In *Mathematical Foundations of Computer Science 1994, 19th International Symposium, MFCS'94, Lecture Notes in Computer Science 841*, pages 336–341. Springer, 1994.

[8] A. Rajasekar. Applications in constraint logic programming with strings. In *Proceedings of the Principles and Practice of Constraint Programming, Second International Workshop, PPCP'94, Lecture Notes in Computer Science 874*, pages 109–122. Springer, 1994.

[9] M. Schaefer, E. Sedgwick, and D. Stefankovic. Recognizing string graphs is in NP. *Journal of Computer and System Sciences*, 67(2):365–380, 2003.