Systems of Word Equations and Polynomials: a New Approach [*]

Aleksi Saarela

Turku Centre for Computer Science TUCS and Department of Mathematics
University of Turku, FI-20014 Turku, FINLAND
`amsaar@utu.fi`

We develop new polynomial methods for studying systems of word equations. We use them to improve some earlier results and to analyze how sizes of systems of word equations satisfying certain independence properties depend on the lengths of the equations. These methods give the first nontrivial upper bounds for the sizes of the systems.

1 Introduction

Word equations are a fundamental part of combinatorics on words, see e.g. [20] or [2] for a general reference on these subjects. One of the basic results in the theory of word equations is that a nontrivial equation causes a defect effect. In other words, if n words satisfy a nontrivial relation, then they can be represented as products of $n-1$ words. Not much is known about the additional restrictions caused by several independent relations [9].

In fact, even the following simple question, formulated already in [3], is still unanswered: how large can an independent system of word equations on three unknowns be? The largest known examples consist of three equations. The only known upper bound comes from the Ehrenfeucht Compactness Property, proved in [1] and independently in [8]: an independent system cannot be infinite. This question can be obviously asked also in the case of $n > 3$ unknowns. Then there are independent systems of size $\Theta(n^4)$ [16]. Some results concerning independent systems on three unknowns can be found in [11], [5] and [6], but the open problem seems to be very difficult to approach with current techniques.

There are many variations of the above question: we may study it in the free semigroup, i.e. require that $h(x) \neq \varepsilon$ for every solution h and unknown x, or examine only the systems having a solution of rank $n-1$, or study chains of solution sets instead of independent systems. See e.g. [10], [9], [4] and [17].

In this article we will try to use polynomials to study some questions related to systems of word equations. Algebraic techniques have been used before, most notably in the proof of Ehrenfeucht's conjecture, which is based on Hilbert's Basis Theorem. However, the way in which we use polynomials is quite different and allows us to apply linear algebra to the problems.

One of the main contributions of this article is the development of new methods for attacking problems on word equations. This is done in Sections 3 and 5. Other contributions include simplified proofs and generalizations for old results in Sect. 4 and in the end of Sect. 5, and studying maximal sizes of independent systems of equations in Sect. 6. Thus the connection between word equations and linear algebra is not only theoretically interesting, but is also shown to be very useful at establishing simple-looking results that have been previously unknown, or that have had only very complicated proofs. In addition to the results of the paper, we believe that the techniques may be useful in further analysis of word equations.

Now we give a brief overview of the paper. First, in Sect. 2 we define a way to transform words into polynomials and prove some basic results using these polynomials.

In Sect. 3 we prove that if the lengths of the unknowns are fixed, then there is a connection between the ranks of solutions of a system of equations and the rank of a certain polynomial matrix. This theorem is very important for all the later results.

Section 4 contains small generalizations of two earlier results. These are nice examples of the methods developed in Sect. 3 and have independent interest, but they are not important for the later sections.

In Sect. 5 we analyze the results of Sect. 3, when the lengths of the unknowns are not fixed. For every solution these lengths form an n-dimensional vector, called the *length type* of the solution. We prove that the length types of all solutions of rank $n-1$ of a pair of equations are covered by a finite union of $(n-1)$-dimensional subspaces, if the equations are not equivalent on solutions of rank $n-1$. This means that the solution sets of pairs of equations are in some sense more structured than the solution sets of single equations. This theorem is the key to proving the remaining results. We conclude Sect. 5 by proving a theorem about unbalanced equations. This gives a considerably simpler reproof and a generalization of a result in [11]

Finally, in Sect. 6 we return to the question about sizes of independent systems. There is a trivial bound for the size of a system depending on the length of the longest equation, because there are only exponentially many equations of a fixed length. We prove that if the system is independent even when considering only solutions of rank $n-1$, then there is an upper bound for the size of the system depending quadratically on the length of the shortest equation. Even though it does not give a fixed bound even in the case of three unknowns, it is a first result of its type – hence opening, we hope, a new avenue for future research.

2   Basic Theorems

Let $|w|$ be the length of a word w and $|w|_a$ be the number of occurrences of a letter a in w. We use the notation $u \leq v$, if u is a prefix of v. We denote the set of nonnegative integers by $\mathbb{N}_0$ and the set of positive integers by $\mathbb{N}_1$. The empty word is denoted by ε.

In this section we give proofs for some well-known results. These serve as examples of the polynomial methods used. Even though the standard proofs of these are simple, we hope that the proofs given here illustrate how properties of words can be formulated and proved in terms of polynomials.

Let $\Sigma \subset \mathbb{N}_1$ be an alphabet of numbers. For a word $w = a_0 \ldots a_{n-1} \in \Sigma^n$ we define a polynomial

$$P_w = a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1}.$$

Now $w \mapsto P_w$ is an injective mapping from words to polynomials (here we need the assumption $0 \notin \Sigma$). If $w_1, \ldots, w_m \in \Sigma^*$, then

$$P_{w_1 \ldots w_m} = P_{w_1} + P_{w_2} X^{|w_1|} + \cdots + P_{w_m} X^{|w_1 \ldots w_{m-1}|}. \tag{1}$$

If $w \in \Sigma^+$ and $k \in \mathbb{N}_0$, then

$$P_{w^k} = P_w \frac{X^{k|w|} - 1}{X^{|w|} - 1}$$

The polynomial P_w can be viewed as a characteristic polynomial of the word w. We could also replace X with a suitable number b and get a number whose reverse b-ary representation is w. Or we could let the coefficients of P_w be from some other commutative ring than $\mathbb{Z}$. Similar ideas have been used to analyze words in many places, see e.g. [19], [23] and [15].

Example 2.1. If $w = 1212$, then $P_w = 1 + 2X + X^2 + 2X^3$.

A word w is *primitive*, if it is not of the form u^k for any $k > 1$. If $w = u^k$ and u is primitive, then u is a *primitive root* of w.

Lemma 2.2. *If w is primitive, then P_w is not divisible by any polynomial of the form $(X^{|w|} - 1)/(X^n - 1)$, where $n < |w|$ is a divisor of $|w|$.*

Proof. If P_w is divisible by $(X^{|w|} - 1)/(X^n - 1)$, then there are numbers $a_0, \ldots, a_{n-1}$ such that

$$P_w = (a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1}) \frac{X^{|w|} - 1}{X^n - 1} = (a_0 + a_1 X^1 + \cdots + a_{n-1} X^{n-1})(1 + X^n + \cdots + X^{|w|-n}),$$

so $w = (a_0 \ldots a_{n-1})^{|w|/n}$. $\qquad \square$

The next two theorems are among the most basic and well-known results in combinatorics on words (except for item (4) of Theorem 2.4).

Theorem 2.3. *Every nonempty word has a unique primitive root.*

Proof. Let $u^m = v^n$, where u and v are primitive. We need to show that $u = v$. We have

$$P_u \frac{X^{m|u|} - 1}{X^{|u|} - 1} = P_{u^m} = P_{v^n} = P_v \frac{X^{n|v|} - 1}{X^{|v|} - 1}.$$

Because $m|u| = n|v|$, we get $P_u(X^{|v|} - 1) = P_v(X^{|u|} - 1)$. If $d = \gcd(|u|, |v|)$, then $\gcd(X^{|u|} - 1, X^{|v|} - 1) = X^d - 1$. Thus P_u must be divisible by $(X^{|u|} - 1)/(X^d - 1)$ and P_v must be divisible by $(X^{|v|} - 1)/(X^d - 1)$. By Lemma 2.2, both u and v can be primitive only if $|u| = d = |v|$. $\qquad \square$

The primitive root of a word $w \in \Sigma^+$ is denoted by $\rho(w)$.

Theorem 2.4. *For $u, v \in \Sigma^+$, the following are equivalent:*

1. *$\rho(u) = \rho(v)$,*

2. *if $U, V \in \{u, v\}^*$ and $|U| = |V|$, then $U = V$,*

3. *u and v satisfy a nontrivial relation,*

4. *$P_u/(X^{|u|} - 1) = P_v/(X^{|v|} - 1)$.*

Proof. (1) $\Rightarrow$ (2): $U = \rho(u)^{|U|/|\rho(u)|} = \rho(u)^{|V|/|\rho(u)|} = V$.

(2) $\Rightarrow$ (3): Clear.

(3) $\Rightarrow$ (4): Let $u_1 \ldots u_m = v_1 \ldots v_n$, where $u_i, v_j \in \{u, v\}$. Now

$$0 = P_{u_1 \ldots u_m} - P_{v_1 \ldots v_n} = \frac{P_u}{X^{|u|} - 1} p - \frac{P_v}{X^{|v|} - 1} p$$

for some polynomial p. If $m \neq n$ or $u_i \neq v_i$ for some i, then $p \neq 0$, and thus $P_u/(X^{|u|} - 1) = P_v/(X^{|v|} - 1)$.

(4) $\Rightarrow$ (1): We have $P_{u^{|v|}} = P_{v^{|u|}}$, so $u^{|v|} = v^{|u|}$ and $\rho(u) = \rho(u^{|v|}) = \rho(v^{|u|}) = \rho(v)$. $\qquad \square$

Similarly, polynomials can be used to give a simple proof for the theorem of Fine and Wilf. In fact, one of the original proofs in [7] uses power series. Algebraic techniques have also been used to prove variations of this theorem [21].

Theorem 2.5 (Fine and Wilf). *If u^i and v^j have a common prefix of length $|u| + |v| - \gcd(|u|, |v|)$, then $\rho(u) = \rho(v)$.*

3   Solutions of Fixed Length

In this section we apply polynomial techniques to word equations. From now on, we will assume that the unknowns are ordered as $x_1, \ldots, x_n$ and that Ξ is the set of these unknowns.

A (coefficient-free) *word equation* $u = v$ on n unknowns consists of two words $u, v \in \Xi^*$. A *solution* of this equation is any morphism $h : \Xi^* \to \Sigma^*$ such that $h(u) = h(v)$. The equation is *trivial*, if u and v are the same word.

The (combinatorial) *rank* of a morphism h is the smallest number r for which there is a set A of r words such that $h(x) \in A^*$ for every unknown x. A morphism of rank at most one is *periodic*.

Let $h : \Xi^* \to \Sigma^*$ be a morphism. The *length type* of h is the vector

$$L = (|h(x_1)|, \ldots, |h(x_n)|) \in \mathbb{N}_0^n.$$

This length type L determines a morphism $\mathrm{len}_L : \Xi^* \to \mathbb{N}_0, \mathrm{len}_L(w) = |h(w)|$.

For a word equation $E : y_1 \ldots y_k = z_1 \ldots z_l$, where $y_i, z_i \in \Xi$, a variable $x \in \Xi$ and a length type L, let

$$Q_{E,x,L} = \sum_{y_i = x} X^{\mathrm{len}_L(y_1 \ldots y_{i-1})} - \sum_{z_i = x} X^{\mathrm{len}_L(z_1 \ldots z_{i-1})}.$$

Theorem 3.1. *A morphism $h : \Xi^* \to \Sigma^*$ of length type L is a solution of an equation $E : u = v$ if and only if*

$$\sum_{x \in \Xi} Q_{E,x,L} P_{h(x)} = 0.$$

Proof. Now $h(u) = h(v)$ if and only if $P_{h(u)} = P_{h(v)}$, and the polynomial $P_{h(u)} - P_{h(v)}$ can be written as $\sum_{x \in \Xi} Q_{E,x,L} P_{h(x)}$ by (1). $\qquad\qquad\square$

Example 3.2. Let $\Xi = \{x, y, z\}$, $E : xyz = zxy$ and $L = (1, 1, 2)$. Now

$$Q_{E,x,L} = 1 - X^2, \qquad Q_{E,y,L} = X - X^3, \qquad Q_{E,z,L} = X^2 - 1.$$

If h is the morphism defined by $h(x) = 1$, $h(y) = 2$ and $h(z) = 12$, then h is a solution of E and

$$Q_{E,x,L} P_{h(x)} + Q_{E,y,L} P_{h(y)} + Q_{E,z,L} P_{h(z)} = (1 - X^2) \cdot 1 + (X - X^3) \cdot 2 + (X^2 - 1)(1 + 2X) = 0.$$

A morphism $\phi : \Xi^* \to \Xi^*$ is an *elementary transformation*, if there are $x, y \in \Xi$ so that $\phi(y) \in \{xy, x\}$ and $\phi(z) = z$ for $z \in \Xi \setminus \{y\}$. If $\phi(y) = xy$, then ϕ is *regular*, and if $\phi(y) = x$, then ϕ is *singular*. The next lemma follows immediately from results in [20].

Lemma 3.3. *Every solution h of an equation E has a factorization $h = \theta \circ \phi \circ \alpha$, where $\alpha(x) \in \{x, \varepsilon\}$ for all $x \in \Xi$, $\phi = \phi_m \circ \cdots \circ \phi_1$, every ϕ_i is an elementary transformation and $\phi \circ \alpha$ is a solution of E. If $\alpha(x) = \varepsilon$ for s unknowns x and t of the ϕ_i are singular, then the rank of $\phi \circ \alpha$ is $n - s - t$.*

Lemma 3.4. *Let $E : u = v$ be an equation on n unknowns. Let $h : \Xi^* \to \Sigma^*$ be a solution of length type L that has rank r. There is an r-dimensional subspace V of $\mathbb{Q}^n$ such that $L \in V$ but those length types of the solutions of E of rank r that are in V are not covered by any finite union of $(r-1)$-dimensional spaces.*

Proof. Let $h = \theta \circ \phi_m \circ \cdots \circ \phi_1 \circ \alpha$ as in Lemma 3.3. Let $f_k = \phi_k \circ \cdots \circ \phi_1 \circ \alpha$. Now $g \circ f_m$ is a solution of E for every morphism $g : \Xi^* \to \Sigma^*$. The length type of $g \circ f_m$ is

$$\sum_{i=1}^{n} |g(x_i)| \cdot (|f_m(x_1)|_{x_i}, \ldots, |f_m(x_n)|_{x_i}) \tag{2}$$

To prove the theorem, we need to show that at least r of the vectors in this sum are linearly independent.

Let A_k be the $n \times n$ matrix $(|f_k(x_i)|_{x_j})$. If there are s unknowns x such that $\alpha(x) = \varepsilon$, then the rank of A_0 is $n - s$. If ϕ_k is regular, then the matrix A_k is obtained from A_{k-1} by adding one of the columns to another column, so the ranks of these matrices are equal. If ϕ_k is singular, then A_k is obtained from A_{k-1} by adding one of the columns to another column and setting some column to zero, so the rank of the matrix is decreased by at most one. If t of the ϕ_i are singular, then the rank of A_m is at least $n - s - t$. The rank of f_m is $n - s - t$, so $r \leq n - s - t$ and at least r of the columns of A_m are linearly independent. $\qquad \square$

Lemma 3.5. *Let $E : u = v$ be an equation and $h : \Xi^* \to \Sigma^*$ be a solution of length type L that has rank r. There are morphisms $f_m : \Xi^* \to \Xi^*$ and $g_m : \Xi^* \to \Sigma^*$ and polynomials p_{ij} such that the following conditions hold:*

1. *$h = g_m \circ f_m$,*

2. *f_m is a solution of E,*

3. *$P_{(g \circ f_m)(x_i)} = \sum p_{ij} P_{g(x_j)}$ for all i, j, if $g : \Xi^* \to \Sigma^*$ is a morphism of the same length type as g_m,*

4. *r of the vectors $(p_{1j}, \ldots, p_{nj})$, where $j = 1, \ldots, n$, are linearly independent.*

Proof. Let f_k be as in the proof of Lemma 3.4 and let g_k be such that $h = g_k \circ f_k$. For every k, there are polynomials p_{ijk} so that $P_{h(x_i)} = \sum_{j=1}^{n} p_{ijk} P_{g_k(x_j)}$ for all $i \in \{1, \ldots, n\}$ (p_{ijk} "encodes" the positions of the word $g_k(x_j)$ in $h(x_i)$). Let B_k be the $n \times n$ matrix (p_{ijk}). The matrix B_{k+1} is obtained from B_k by adding one of the columns to another column, and multiplying some column with a polynomial. Like in Lemma 3.4, we conclude that at least $n - s - t$ of the columns of B_m are linearly independent and $r \leq n - s - t$. If we let $p_{ij} = p_{ijm}$, then the four conditions hold. $\qquad \square$

With the help of these lemmas, we are going to analyze solutions of some fixed length type. Fundamental solutions (which were implicitly present in the previous lemmas, see [20]) have been used in connection with fixed lengths also in [13] and [12].

Theorem 3.6. *Let $E_1, \ldots, E_m$ be a system of equations on n unknowns and let $L \in \mathbb{N}_0^n$. Let $q_{ij} = Q_{E_i, x_j, L}$. If the system has a solution of length type L that has rank r, then the rank of the $m \times n$ matrix (q_{ij}) is at most $n - r$. If the rank of the matrix is 1, at most one component of L is zero and the equations are nontrivial, then they have the same solutions of length type L.*

Proof. Let h be a solution of length type L that has rank r. If $r = 1$, the first claim follows from Theorem 3.1, so assume that $r > 1$. Let E be an equation that has the same nonperiodic solutions as the system. We will use Lemma 3.5 for this equation. Fix k and let $g : \Xi^* \to \Sigma^*$ be the morphism determined by $g(x_k) = 10^{|g_m(x_k)|-1}$ and $g(x_i) = 0^{|g_m(x_i)|}$ for all $i \neq k$ (we assumed earlier that $0 \notin \Sigma$, but it does not matter here). Then $g \circ f_m$ is a solution of every E_l, $P_{(g \circ f_m)(x_i)} = \sum_{j=1}^{n} p_{ij} P_{g(x_j)}$ and

$$0 = \sum_{i=1}^{n} Q_{E_l, x_i, L} \sum_{j=1}^{n} p_{ij} P_{g(x_j)} = \sum_{i=1}^{n} Q_{E_l, x_i, L} p_{ik}$$

for all l by Theorem 3.1. Thus the vectors $(p_{1j}, \ldots, p_{nj})$ are solutions of the linear system of equations determined by the matrix (q_{ij}). Because at least r of these vectors are linearly independent, the rank of the matrix is at most $n - r$.

If at most one component of L is zero and the equations are nontrivial, then all rows of the matrix are nonzero. If also the rank of the matrix is 1, then all rows are multiples of each other and the second claim follows by Theorem 3.1. $\qquad \square$

4 Applications

The *graph* of a system of word equations is the graph, where Ξ is the set of vertices and there is an edge between x and y, if one of the equations in the system is of the form $x \cdots = y \cdots$. The following well-known theorem can be proved with the help of Theorem 3.6.

Theorem 4.1 (Graph Lemma). *Consider a system of equations whose graph has r connected components. If h is a solution of this system and $h(x_i) \neq \varepsilon$ for all i, then h has rank at most r.*

Proof. We can assume that the connected components are

$$\{x_1, \ldots, x_{i_2-1}\}, \{x_{i_2}, \ldots, x_{i_3-1}\}, \ldots, \{x_{i_r}, \ldots, x_n\}$$

and the equations are

$$x_j \cdots = x_{k_j} \cdots,$$

where $j \in \{1, \ldots, n\} \smallsetminus \{1, i_2, \ldots, i_r\}$ and $k_j < j$. Let q_{ij} be as in Theorem 3.6. If we remove the columns $1, i_2, \ldots, i_r$ from the $(n-r) \times n$ matrix (q_{ij}), we obtain a square matrix M, where the diagonal elements are not divisible by X, but all elements above the diagonal are divisible by X. This means that $\det(M)$ is not divisible by X, so $\det(M) \neq 0$. Thus the rank of the matrix (q_{ij}) is $n-r$ and h has rank at most r by Theorem 3.6. $\square$

The next theorem generalizes a result from [5] for more than three unknowns.

Theorem 4.2. *If a pair of nontrivial equations on n unknowns has a solution h of rank $n-1$, where no two of the unknowns commute, then there is a number $k \geq 1$ such that the equations are of the form $x_1 \cdots = x_2^k x_3 \cdots$.*

Proof. By Theorem 4.1, the equations must be of the form $x_1 \cdots = x_2 \cdots$. Let them be

$$x_1 u y \cdots = x_2 v z \cdots \qquad \text{and} \qquad x_1 u' y' \cdots = x_2 v' z' \cdots,$$

where $u, v, u', v' \in \{x_1, x_2\}^*$ and $y, z, y', z' \in \{x_3, \ldots, x_n\}$. We can assume that $z = x_3$ and $|h(x_2 v)| \leq |h(x_1 u)|, |h(x_1 u')|, |h(x_2 v')|$. If it would be $|h(x_1 u)| = |h(x_2 v)|$, then $h(x_1)$ and $h(x_2)$ would commute, so $|h(x_1 u)| > |h(x_2 v)|$. If v would contain x_1, then $h(x_1)$ and $h(x_2)$ would commute by Theorem 2.5, so $v = x_2^{k-1}$ for some $k \geq 1$.

Let L be the length type of h and let q_{ij} be as in Theorem 3.6. By Theorem 3.6, the rank of the matrix (q_{ij}) must be 1 and thus $q_{12} q_{23} - q_{13} q_{22} = 0$. The term of $q_{13} q_{22}$ of the lowest degree is $X^{|h(x_2^k)|}$. The same must hold for $q_{12} q_{23}$, and thus the term of q_{23} of the lowest degree must be $-X^{|h(x_2^k)|}$. This means that $|h(x_2 v')| = |h(x_2^k)| \leq |h(x_1 u')|$ and $z' = x_3$. As above, we conclude that $|h(x_2 v')| < |h(x_1 u')|$, v' cannot contain x_1 and $v' = x_2^{k-1}$. $\square$

It was proved in [18] that if

$$s_0 u_1^i s_1 \ldots u_m^i s_m = t_0 v_1^i t_1 \ldots v_n^i t_n$$

holds for $m+n+3$ consecutive values of i, then it holds for all i. By using similar ideas as in Theorem 3.6, we improve this bound to $m+n$ and prove that the values do not need to be consecutive. In [18] it was also stated that the arithmetization and matrix techniques in [24] would give a simpler proof of a weaker result. Similar questions have been studied in [14] and there are relations to independent systems [22].

Theorem 4.3. *Let $m,n \geq 1$, $s_j, t_j \in \Sigma^*$ and $u_j, v_j \in \Sigma^+$. Let $U_i = s_0 u_1^i s_1 \ldots u_m^i s_m$ and $V_i = t_0 v_1^i t_1 \ldots v_n^i t_n$. If $U_i = V_i$ holds for $m+n$ values of i, then it holds for all i.*

Proof. The equation $U_i = V_i$ is equivalent with $P_{U_i} - P_{V_i} = 0$. This equation can be written as

$$\sum_{j=0}^{m} y_j X^{i|u_1 \ldots u_j|} + \sum_{k \in K} z_k X^{i|v_1 \ldots v_k|} = 0, \tag{3}$$

where y_j, z_k are some polynomials, which do not depend on i, and K is the set of those $k \in \{0, \ldots n\}$ for which $|v_1 \ldots v_k|$ is not any of the numbers $|u_1 \ldots u_j|$ ($j = 0, \ldots, m$). If $U_{i_1} = V_{i_1}$ and $U_{i_2} = V_{i_2}$, then

$$(i_1 - i_2)|u_1 \ldots u_m| = |U_{i_1}| - |U_{i_2}| = |V_{i_1}| - |V_{i_2}| = (i_1 - i_2)|v_1 \ldots v_n|.$$

Thus $|u_1 \ldots u_m| = |v_1 \ldots v_n|$ and the size of K is at most $n - 1$. If (3) holds for $m + 1 + \#K \leq m + n$ values of i, it can be viewed as a system of equations, where y_j, z_k are unknowns. The coefficients of this system form a generalized Vandermonde matrix, whose determinant is nonzero, so the system has a unique solution $y_j = z_k = 0$ for all j, k, (3) holds for all i and $U_i = V_i$ for all i. □

5 Sets of Solutions

Now we analyze how the polynomials $Q_{E,x,L}$ behave when L is not fixed. Let

$$\mathcal{M} = \{a_1 X_1 + \cdots + a_n X_n \mid a_1, \ldots, a_n \in \mathbb{N}_0\} \subset \mathbb{Z}[X_1, \ldots, X_n]$$

be the additive monoid of linear homogeneous polynomials with nonnegative integer coefficients on the variables $X_1, \ldots, X_n$. The *monoid ring* of $\mathcal{M}$ over $\mathbb{Z}$ is the ring formed by expressions of the form

$$a_1 X^{p_1} + \cdots + a_k X^{p_k},$$

where $a_i \in \mathbb{Z}$ and $p_i \in \mathcal{M}$, and the addition and multiplication of these generalized polynomials is defined in a natural way. This ring is denoted by $\mathbb{Z}[X; \mathcal{M}]$. If $L \in \mathbb{Z}^n$, then the value of a polynomial $p \in \mathcal{M}$ at the point $(X_1, \ldots X_n) = L$ is denoted by $p(L)$, and the polynomial we get by making this substitution in $s \in \mathbb{Z}[X; \mathcal{M}]$ is denoted by $s(L)$.

The ring $\mathbb{Z}[X; \mathcal{M}]$ is isomorphic to the ring $\mathbb{Z}[Y_1, \ldots, Y_n]$ of polynomials on n variables. The isomorphism is given by $X^{X_i} \mapsto Y_i$. However, the generalized polynomials, where the exponents are in $\mathcal{M}$, are suitable for our purposes.

If $a_i \leq b_i$ for $i = 1, \ldots, n$, then we use the notation

$$a_1 X_1 + \cdots + a_n X_n \preceq b_1 X_1 + \cdots + b_n X_n.$$

If $p, q \in \mathcal{M}$ and $p \preceq q$, then $p(L) \leq q(L)$ for all $L \in \mathbb{N}_0^n$.

For an equation $E : x_{i_1} \ldots x_{i_r} = x_{j_1} \ldots x_{j_s}$ we define

$$S_{E,x} = \sum_{x_{i_k}=x} X^{X_{i_1}+\cdots+X_{i_{k-1}}} - \sum_{x_{j_k}=x} X^{X_{j_1}+\cdots+X_{j_{k-1}}} \in \mathbb{Z}[X; \mathcal{M}].$$

Now $S_{E,x}(L) = Q_{E,x,L}$. Theorem 3.1 can be formulated in terms of the generalized polynomials $S_{E,x}$.

Theorem 5.1. *A morphism $h : \Xi^* \to \Sigma^*$ of length type L is a solution of an equation E if and only if*

$$\sum_{x \in \Xi} S_{E,x}(L) P_{h(x)} = 0.$$

Example 5.2. Let $E : x_1 x_2 x_3 = x_3 x_1 x_2$. Now

$$S_{E,x_1} = 1 - X^{X_3}, \qquad S_{E,x_2} = X^{X_1} - X^{X_1 + X_3}, \qquad S_{E,x_3} = X^{X_1 + X_2} - 1.$$

The *length* of an equation $E : u = v$ is $|E| = |uv|$.

Theorem 5.3. *Let E_1, E_2 be a pair of nontrivial equations on n unknowns that don't have the same sets of solutions of rank $n - 1$. The length types of solutions of the pair of rank $n - 1$ are covered by a union of $|E_1|^2$ $(n - 1)$-dimensional subspaces of $\mathbb{Q}^n$. If $V_1, \ldots, V_m$ is a minimal such cover and $L \in V_i$ for some i, then E_1 and E_2 have the same solutions of length type L and rank $n - 1$.*

Proof. Let $s_{ij} = S_{E_i,x_j}$ for $i = 1, 2$ and $j = 1, \ldots, n$. If all 2×2 minors of the $2 \times n$ matrix (s_{ij}) are zero, then for all length types L of solutions of rank $n - 1$ the rank of the matrix (q_{ij}) in Theorem 3.6 is 1 and E_1 and E_2 are equivalent, which is a contradiction. Thus there are k, l such that $t_{kl} = s_{1k} s_{2l} - s_{1l} s_{2k} \neq 0$. The generalized polynomial t_{kl} can be written as

$$t_{kl} = \sum_{i=1}^{M} X^{p_i} - \sum_{i=1}^{N} X^{q_i},$$

where $p_i, q_i \in \mathcal{M}$ and $p_i \neq q_j$ for all i, j. If L is a length type of a solution of rank $n - 1$, then $M = N$ and L must be a solution of the system of equations

$$p_i = q_{\sigma(i)} \qquad (i = 1, \ldots, M) \tag{4}$$

for some permutation σ. For every σ the equations determine an at most $(n - 1)$-dimensional space. Let

$$s_{1k} = \sum_i X^{a_i} - \sum_i X^{a_i'}, \quad s_{2l} = \sum_i X^{b_i} - \sum_i X^{b_i'}, \quad s_{1l} = \sum_i X^{c_i} - \sum_i X^{c_i'}, \quad s_{2k} = \sum_i X^{d_i} - \sum_i X^{d_i'},$$

where $a_i \preceq a_{i+1}$, $a_i' \preceq a_{i+1}'$, and so on. The polynomials p_i form a subset of the polynomials $a_i + b_j$, $a_i' + b_j'$, $c_i + d_j'$ and $c_i' + d_j$ (the reason that they form just a subset is that we assumed $p_i \neq q_j$ for all i, j). For any i, let j_i be the smallest index j such that $a_i + b_j = p_m$ for some m. Now for every i, j, m such that $a_i + b_j = p_m$ we have $a_i + b_{j_i} \preceq p_m$. We can do a similar thing for the polynomials a_i', b_i' and c_i, d_i' and c_i', d_i. In this way we obtain at most $|E_1|$ polynomials p_i such that for any L the value of one of these polynomials is minimal among the values $p_i(L)$. Similarly we obtain at most $|E_1|$ "minimal" polynomials q_i. It is sufficient to consider only those systems (4), where one of the equations is formed by these "minimal" polynomials p_i, q_i. There are at most $|E_1|^2$ possible pairs of such polynomials, and each of them determines an $(n - 1)$-dimensional space.

Consider the second claim. Because the cover is minimal, there is a solution of rank $n - 1$ whose length type is in V_i, but not in any other V_j. By Lemma 3.4, the length types of solutions of rank $n - 1$ in this space cannot be covered by a finite union of $(n - 2)$-dimensional spaces. Thus one of the systems (4) must determine the space V_i. The same holds for systems coming from all other nonzero 2×2 minors of the matrix (s_{ij}), so E_1 and E_2 have the same solutions of rank $n - 1$ and length type L for all $L \in V_i$ by Theorem 3.6. $\qquad\square$

The following example illustrates the proof of Theorem 5.3. It gives a pair of equations on three unknowns, where the required number of subspaces is two. We do not know any example, where more spaces would be necessary.

Example 5.4. Consider the equations $E_1 : x_1x_2x_3 = x_3x_1x_2$ and $E_2 : x_1x_2x_1x_3x_2x_3 = x_3x_1x_3x_2x_1x_2$ and the generalized polynomial

$$
\begin{aligned}
s =& S_{E_1,x_1} S_{E_2,x_3} - S_{E_1,x_3} S_{E_2,x_1} \\
=& X^{2X_1+X_2} + X^{2X_1+2X_2+X_3} + X^{X_1+2X_3} + X^{X_1+X_2+X_3} - X^{2X_1+X_2+X_3} - X^{X_1+X_3} - X^{2X_1+2X_2} - X^{X_1+X_2+2X_3}.
\end{aligned}
$$

If L is a length type of a nontrivial solution of the pair E_1, E_2, then $s(L) = 0$. If $s(L) = 0$, then L must satisfy an equation $p = q$, where $p \in \{2X_1 + X_2, X_1 + 2X_3, X_1 + X_2 + X_3\}$ and $q \in \{X_1 + X_3, 2X_1 + 2X_2\}$. The possible relations are

$$
X_3 = 0, \qquad X_1 + X_2 = X_3, \qquad X_2 = 0, \qquad X_1 + 2X_2 = 2X_3.
$$

If L satisfies one of the first three, then $s(L) = 0$. If L satisfies the last one, then $s(L) \neq 0$, except if $L = 0$. So if h is a nonperiodic solution, then

$$
|h(x_3)| = 0 \quad \text{or} \quad |h(x_1x_2)| = |h(x_3)| \quad \text{or} \quad |h(x_2)| = 0.
$$

There are no nonperiodic solutions with $h(x_2) = \varepsilon$, but every h with $h(x_3) = \varepsilon$ or $h(x_1x_2) = h(x_3)$ is a solution.

An equation $u = v$ is *balanced*, if $|u|_x = |v|_x$ for every unknown x. In [11] it was proved that if an independent pair of equations on three unknowns has a nonperiodic solution, then the equations must be balanced. With the help of Theorem 5.3 we get a significantly simpler proof and a generalization for this result.

Theorem 5.5. *Let E_1, E_2 be a pair of equations on n unknowns having a solution of rank $n - 1$. If E_1 is not balanced, then every solution of E_1 of rank $n - 1$ is a solution of E_2.*

Proof. The length types of solutions of E_1 are covered by a single $(n - 1)$-dimensional space V. Because the pair E_1, E_2 has a solution of rank $n - 1$, V is a minimal cover for the length types of the solutions of the pair of rank $n - 1$. By Theorem 5.3, E_1 and E_2 have the same solutions of length type L and rank $n - 1$ for all $L \in V$. $\qquad\square$

Another way to think of this result is that if E_1 is not balanced but has a solution of rank $n - 1$ that is not a solution of E_2, then the pair E_1, E_2 causes a larger than minimal defect effect.

6 Independent Systems

A system of word equations $E_1, \ldots, E_m$ is *independent*, if for every i there is a morphism that is not a solution of E_i, but is a solution of all the other equations.

A sequence of equations $E_1, \ldots, E_m$ is a *chain*, if for every i there is a morphism that is not a solution of E_i, but is a solution of all the preceding equations.

The question of the maximal size of an independent system is open. Only things that are known are that independent systems cannot be infinite and there are systems of size $\Theta(n^4)$, where n is the number of unknowns. For a survey on these topics, see [17].

We study the following variation of the above question: how long can a sequence of equations $E_1,\ldots,E_m$ be, if for every i there is a morphism of rank $n-1$ that is not a solution of E_i, but is a solution of all the preceding equation? We prove an upper bound depending quadratically on the length of the first equation. For three unknowns we get a similar bound for the size of independent systems and chains.

Theorem 6.1. *Let $E_1,\ldots,E_m$ be nontrivial equations on n unknowns having a common solution of rank $n-1$. For every $i \in \{1,\ldots,m-1\}$, assume that there is a solution of the system $E_1,\ldots,E_i$ of rank $n-1$ that is not a solution of E_{i+1}. If the length types of solutions of the pair E_1,E_2 of rank $n-1$ are covered by a union of N $(n-1)$-dimensional subspaces, then $m \le N+1$. In general, $m \le |E_1|^2 +1$.*

Proof. We can assume that E_i is equivalent with the system $E_1,\ldots,E_i$ for all $i \in \{1,\ldots,m\}$. Let the length types of solutions of E_2 of rank $n-1$ be covered by the $(n-1)$-dimensional spaces $V_1,\ldots,V_N$. Some subset of these spaces forms a minimal cover for the length types of solutions of E_3 of rank $n-1$. If this minimal cover would be the whole set, then E_2 and E_3 would have the same solutions of rank $n-1$ by the second part of Theorem 5.3. Thus the length types of solutions of E_3 of rank $n-1$ are covered by some $N-1$ of these spaces. We conclude inductively that the length types of solutions of E_i of rank $n-1$ are covered by some $N-i+2$ of these spaces for all $i \in \{2,\ldots,m\}$. It must be $N-m+2 \ge 1$, so $m \le N+1$. By the first part of Theorem 5.3, $N \le |E_1|^2$. $\square$

In Theorem 6.1 it is not enough to assume that the equations are independent and have a common solution of rank $n-1$. If the number of unknowns is not fixed, then there are arbitrarily large such systems, where the length of every equation is 10, see e.g. [10].

In the case of three unknowns, Theorem 6.1 gives an upper bound depending on the length of the shortest equation for the size of an independent system of equations, or an upper bound depending on the length of the first equation for the size of a chain of equations. A better bound in Theorem 5.3 would immediately give a better bound in the following corollary.

Corollary 6.2. *If $E_1,\ldots,E_m$ is an independent system on three unknowns having a nonperiodic solution, then $m \le |E_1|^2 +1$. If $E_1,\ldots,E_m$ is a chain of equations on three unknowns, then $m \le |E_1|^2 +5$.*

References

[1] M. H. Albert & J. Lawrence (1985): *A proof of Ehrenfeucht's conjecture*. Theoret. Comput. Sci. 41(1), pp. 121–123, doi:10.1016/0304-3975(85)90066-0.

[2] Christian Choffrut & Juhani Karhumäki (1997): *Combinatorics of Words*. In Grzegorz Rozenberg & Arto Salomaa, editors: *Handbook of Formal Languages*, 1, Springer-Verlag, pp. 329–438.

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

[4] Elena Czeizler (2008): *Multiple constraints on three and four words*. Theoret. Comput. Sci. 391(1-2), pp. 14–19, doi:10.1016/j.tcs.2007.10.026.

[5] Elena Czeizler & Juhani Karhumäki (2007): *On non-periodic solutions of independent systems of word equations over three unknowns*. Internat. J. Found. Comput. Sci. 18(4), pp. 873–897, doi:10.1142/S0129054107005030.

[6] Elena Czeizler & Wojciech Plandowski (2009): *On systems of word equations over three unknowns with at most six occurrences of one of the unknowns*. Theoret. Comput. Sci. 410(30-32), pp. 2889–2909, doi:10.1016/j.tcs.2009.01.023.

[7] N. J. Fine & H. S. Wilf (1965): *Uniqueness theorems for periodic functions*. *Proc. Amer. Math. Soc.* 16, pp. 109–114, doi:10.1090/S0002-9939-1965-0174934-9.

[8] V. S. Guba (1986): *Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems*. *Mat. Zametki* 40(3), pp. 321–324, doi:10.1007/BF01142470.

[9] Tero Harju & Juhani Karhumäki (2004): *Many aspects of defect theorems*. *Theoret. Comput. Sci.* 324(1), pp. 35–54, doi:10.1016/j.tcs.2004.03.051.

[10] Tero Harju, Juhani Karhumäki & Wojciech Plandowski (2002): *Independent systems of equations*. In M. Lothaire, editor: *Algebraic Combinatorics on Words*, Cambridge University Press, pp. 443–472.

[11] Tero Harju & Dirk Nowotka (2003): *On the independence of equations in three variables*. *Theoret. Comput. Sci.* 307(1), pp. 139–172, doi:10.1016/S0304-3975(03)00098-7.

[12] Štěpán Holub (2000): *In search of a word with special combinatorial properties*. In: *Computational and geometric aspects of modern algebra*, *London Math. Soc. Lecture Note Ser.* 275, Cambridge Univ. Press, pp. 120–127, doi:10.1017/CBO9780511600609.011.

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

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

[15] Štěpán Holub & Juha Kortelainen (2009): *On partitions separating two words*. In: *Proceedings of the 7th International Conference on Words*.

[16] Juhani Karhumäki & Wojciech Plandowski (1996): *On the size of independent systems of equations in semigroups*. *Theoret. Comput. Sci.* 168(1), pp. 105–119, doi:10.1016/S0304-3975(96)00064-3.

[17] Juhani Karhumäki & Aleksi Saarela: *On maximal chains of systems of word equations*. *Proc. Steklov Inst. Math.* To appear.

[18] Juha Kortelainen (1998): *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), pp. 43–57.

[19] Werner Kuich (1997): *Semirings and formal power series*. In Grzegorz Rozenberg & Arto Salomaa, editors: *Handbook of Formal Languages*, 1, Springer-Verlag, pp. 609–677.

[20] M. Lothaire (1983): *Combinatorics on Words*. Addison-Wesley.

[21] Filippo Mignosi, Jeffrey Shallit & Ming-wei Wang (2001): *Variations on a theorem of Fine & Wilf*. In: *Proceedings of the 26th International Symposium on Mathematical Foundations of Computer Science*, pp. 512–523, doi:10.1007/3-540-44683-4_45.

[22] Wojciech Plandowski (2003): *Test sets for large families of languages*. In: *Developments in Language Theory*, pp. 75–94, doi:10.1007/3-540-45007-6_6.

[23] Arto Salomaa (1985): *The Ehrenfeucht conjecture: a proof for language theorists*. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* 27, pp. 71–82.

[24] Paavo Turakainen (1987): *The equivalence of deterministic gsm replications on Q-rational languages is decidable*. *Math. Systems Theory* 20(4), pp. 273–282, doi:10.1007/BF01692070.