Quadratic Word Equations

Volker Diekert[1] and John Michael Robson[2]*

[1] Institut für Informatik, Universität Stuttgart
Breitwiesenstr. 20-22, D-70565 Stuttgart, Germany
[2] LaBRI, Université Bordeaux I
351, cours de la Libération, F-33405 Talence Cedex, France

Summary. We consider word equations where each variable occurs at most twice (quadratic systems). The satisfiability problem is NP-hard (even for a single equation), but once the lengths of a possible solution are fixed, then there is a deterministic linear time algorithm to decide whether there is a corresponding solution. If the lengths of a minimal solution were at most exponential, then the satisfiability problem of quadratic systems would be NP-complete.

In the second part we address the problem with regular constraints: The uniform version is PSPACE-complete. Fixing the lengths of a possible solution doesn't make the problem much easier. The non-uniform version remains NP-hard (in contrast to the linear time result above). The uniform version remains PSPACE-complete.

In the third part we show that for quadratic systems the exponent of periodicity is at most linear in the denotational length.

1 Introduction

The existential theory of equations over free monoids is decidable. This major result was obtained by Makanin [10], who showed that the satisfiability of word equations with constants is decidable. For the background we refer to [14], to the corresponding chapter in the Handbook of Formal Languages (Eds. G. Rozenberg and A. Salomaa), [2], or to the forthcoming [3]. In 1990 Schulz [19] showed an important generalization: Makanin's result remains true when adding regular constraints. Thus, we may specify for each word variable x a regular language L_x and we are only looking for solutions where the value of each variable x is in L_x. This was used e.g. when generalizing Makanin's result to free partially commutative monoids, see [4,12].

A lower bound of NP-hardness on the complexity of this problem is not hard to prove. For an upper bound Gutiérrez [5] showed that the problem is in EXPSPACE, and Plandowski stated a better NEXPTIME result in [15]. Both results imply that the satisfiability problem is in 2-DEXPTIME. Plandowski's method is based on another recent result due to Rytter and Plandowski [18] showing that the minimal solution of a word equation is highly compressible in terms of Lempel-Ziv encodings. It is conjectured that the length of a minimal solution is at most exponential in the denotational length of the equation.

* This work was partially supported by the French German project PROCOPE

In this contribution we only deal with quadratic systems, i.e., systems of word equations where each variable occurs at most twice. These systems are easier to handle. In combinatorial group theory these systems have been introduced by [8], see also [9]. They play an important rôle in the classification of closed surfaces and basic ideas of how to handle quadratic equations go back to [13]. The explicit statement of an algorithm for the solution of quadratic systems of word equations appears in [11].

The satisfiability problem of quadratic systems is still NP-hard (even for a single equation), but once the lengths of a possible solution are fixed, there is a deterministic linear time algorithm to decide whether there is a corresponding solution. A corollary is that if the lengths of a minimal solution of solvable quadratic systems were at most exponential, then the satisfiability problem would be NP-complete. In fact this is strongly conjectured. The conclusion of containment in NP follows also from [18], but the direct method yields a much simpler approach to the special situation of quadratic systems.

In the second part we address the problem with regular constraints. The uniform version is PSPACE-complete. We also show that fixing the lengths of a possible solution doesn't make the problem much easier. The non-uniform version remains NP-hard (in contrast to the linear time result above). The uniform version remains PSPACE-complete.

In the third part we have a closer look at exponent of periodicity for quadratic systems. This is the maximum power of a primitive word which can appear in a minimal solution. We show the exponent of periodicity is at most linear in the denotational length of quadratic systems whereas it can be exponential in general.

The first two parts of the paper are based on the extended abstract [17]. We have however decided to be complementary in the sense that omitted proofs there are given here and full proofs given there are omitted here. The third part on the exponent of periodicity is new. It uses ideas due to Anca Muscholl.

2 Quadratic equations

Let A be an alphabet of constants and let Ω be a set of variables. As usual, $(A \cup \Omega)^*$ means the free monoid over the set $A \cup \Omega$. A *word equation* $L = R$ is a pair $(L, R) \in (A \cup \Omega)^* \times (A \cup \Omega)^*$, and a *system* of word equations is a set of equations $\{L_1 = R_1, \dots, L_k = R_k\}$; its denotational length is defined as $|L_1 R_1 \cdots L_k R_k|$. A *solution* is a homomorphism $\sigma\colon (A \cup \Omega)^* \to A^*$ leaving the letters of A invariant such that $\sigma(L_i) = \sigma(R_i)$ for all $1 \leq i \leq k$. A solution $\sigma\colon \Omega \to A^*$ is called *minimal*, if the sum $\sum_{x \in \Omega} |\sigma(x)|$ is minimal. A system of word equations is called *quadratic*, if each variable occurs at most twice. In the present paper we consider only quadratic systems. The satisfiability problem for word equations is to decide whether a given system of word equations has a solution. Often, this problem is stated only for a single equation, since a system of word equations can be transformed into a single word equation. However this transformation multiplies the number of occurrences of variables. So, since we are dealing with quadratic systems, we cannot use this trick here.

The best known bound on the length of a minimal solution of a quadratic system is doubly exponential [11]. This seems to be quite an overestimation. We have the following conjecture.

Conjecture The length of a minimal solution of a solvable quadratic system of word equations is at most polynomial in the input size.

The value of Theorem 2 would already increase, if only the following much weaker conjecture were true.

Conjecture (weak form) The length of a minimal solution of a solvable quadratic system of word equations is at most exponential in the input size.

The conjectures above are supported by the fact that the exponent of periodicity is at most linear in the denotational length of the system. This result is stated in Theorem 5 below.

3 Complexity results

The first result states that the satisfiability problem of word equations remains NP-hard, even in the restricted case of a single quadratic equation.

Theorem 1. *Let $|A| \geq 2$. The following problem is NP-hard.*
INSTANCE: A quadratic word equation.
QUESTION: Is there a solution $\sigma\colon \Omega \longrightarrow A^*$?

Proof. We give a reduction from 3-SAT. Let $F = C_0 \wedge \cdots \wedge C_{m-1}$ be a propositional formula in 3-CNF over a set of variables Ξ. Each clause has the form

$$C_i = (\tilde{X}_{3i} \vee \tilde{X}_{3i+1} \vee \tilde{X}_{3i+2})$$

where the $\tilde{X}_j$ are literals. We can assume that every variable has both positive and negative occurrences.

First we construct a quadratic system of word equations using word variables

$$\begin{aligned} c_i,\ d_i,\ 0 \leq i \leq m-1, \\ x_j,\ 0 \leq j \leq 3m-1, \\ y_X,\ z_X,\ \text{for each } X \in \Xi. \end{aligned}$$

We use the constants $a, b, a \neq b$. For each clause C_i we have two equations:

$$c_i x_{3i} x_{3i+1} x_{3i+2} = a^{3m} \quad \text{and} \quad c_i d_i = a^{3m-1}.$$

Now let $X \in \Xi$. Consider the set of positions $\{i_1, \ldots, i_k\}$ where $X = \tilde{X}_{i_1} = \cdots = \tilde{X}_{i_k}$ and the set of positions $\{j_1, \ldots, j_n\}$ where $\bar{X} = \tilde{X}_{j_1} = \cdots = \tilde{X}_{j_n}$. We deal with the case $k \leq n$; the case $n \leq k$ is symmetric. With each X we define two more equations:

$$y_X z_X = b \text{ and } x_{i_1} \cdots x_{i_k} y_X a^n b x_{j_1} \cdots x_{j_n} z_X = a^n b a^n b.$$

It is easy to see that the formula is satisfiable if and only if the quadratic system has a solution.

Next, a system of k word equations $L_1 = R_1, \ldots, L_k = R_k$, $k \geq 1$ with $R_1 \cdots R_k \in \{a, b\}^*$ is equivalent to a single equation temporarily using a third constant c: $L_1 c \cdots L_{k-1} c L_k = R_1 c \cdots R_{k-1} c R_k$. Finally, we can eliminate the use of the third letter c without increasing the number of occurrences of any variable by the well known technique of coding the three letters as aba, $abba$ and $abbba$.

The following theorem is the main result of [17]. In a slightly different form it appeared first in an unpublished manuscript of the second author [16].

Theorem 2. *There is a linear time algorithm to solve the following problem (on a unit cost RAM).*

INSTANCE: A quadratic system of word equations with a list of natural numbers $b_x \in \mathbb{N}$, $x \in \Omega$, written in binary.

QUESTION: Is there a solution $\sigma \colon \Omega \longrightarrow A^*$ such that $|\sigma(x)| = b_x$ for all $x \in \Omega$?

Proof. In a linear time preprocessing we can split the system into equations each containing a maximum of three variable occurrences: to see this let $x_1 \cdots x_g = x_{g+1} \cdots x_d$ be a word equation of the system with $1 \leq g < d$, $x_i \in A \cup \Omega$ for $1 \leq i \leq d$. Then the equation is equivalent to:

$$\begin{aligned} x_1 &= y_1, & x_{g+1} &= y_{g+1}, \\ y_1 x_2 &= y_2, & y_{g+1} x_{g+2} &= y_{g+2}, \\ &\vdots & &\vdots \\ y_{g-1} x_g &= y_g, & y_{d-1} x_d &= y_d, \end{aligned}$$
$$y_g = y_d.$$

Here $y_1, \ldots, y_d$ denote new variables, each of them occurring exactly twice. After the obvious simplification of equations with only one variable or constant on each side, we obtain a system where each equation has the form $z = xy$, $x, y, z \in A \cup \Omega$. In fact, using (for the first time) that the lengths b_x are given, we may assume that $b_x \neq 0$ for all variables $x \in \Omega$ and that each equation has the form $z = xy$, where z is a variable. If m denotes the number of equations, then we can define the input size of a problem instance E as:

$$d(E) = m + \sum_{x \in \Omega} \log_2(b_x).$$

For $x \in A \cup \Omega$ let $|x| = b_x$ if $x \in \Omega$, and $|x| = 1$ if $x \in A$. We are looking for a solution σ such that $|\sigma(x)| = |x|$ for all $x \in A \cup \Omega$. A variable $z \in \Omega$ is called *doubly defined*, if E contains two equations $z = xy$ and $z = uv$. Let $dd(E)$ be the number of doubly defined variables. Define $c = 0.55$ and $k = 3/\ln(1/c)$ (≈ 5.01). Finally, we define the weight of the instance E as follows

$$W(E) = |\Omega| + dd(E) + k \sum_{x \in \Omega} \ln |x|.$$

We show in [17] a process which, in $\mathcal{O}(1)$ operations finds a simplification of the system reducing its weight by at least 1.

Remark 1. The method above yields a most general solution in the following sense. Let E be an instance to the problem of Theorem 2 and assume that E is solvable. Then we produce in linear time a quadratic system over a set of variables Γ (but without doubly defined variables) such that the set of solutions satisfying the length constraints is in a canonical one-to-one correspondence with the set of mappings $\psi : \Gamma \longrightarrow A^*$ where $|\psi(x)| = |b_x|$ for $x \in \Gamma$.

Corollary 1. *If the conjecture (weak form) above is true, then the satisfiability problem for quadratic systems of word equations is NP-complete.*

Remark 2. Given Theorem 1, Corollary 1 follows also from the work of Rytter and Plandowski [18]. They have shown that if the lengths $b_x, x \in \Omega$, are given in binary as part of the input together with a word equation (not necessarily quadratic), then there is a deterministic polynomial time algorithm for the satisfiability problem. Their method is based on Lempel-Ziv encodings and technically involved. Our contribution shows that the situation becomes much simpler for quadratic systems. In particular, we can reduce polynomial time to linear time; and our method is fairly straightforward using variable splitting. In view of the conjectures above it is not clear that the use of Lempel-Ziv encodings can improve the running time for deciding the satisfiability of quadratic systems. The most difficult part is apparently to get an idea of the lengths b_x for $x \in \Omega$. Once these lengths are known (or fixed), the corresponding satisfiability problem for quadratic systems of word equation becomes extremely simple.

4 Regular constraints

There is an interesting generalization of Makanin's result due to Schulz [19]. It says that if a word equation is given with a list of regular languages $L_x \subseteq A^*, x \in \Omega$, then one can decide whether there is a solution $\sigma : \Omega \longrightarrow A^*$ such that $\sigma(x) \in L_x$ for all $x \in \Omega$. In the following we shall assume that regular languages are specified by non-deterministic finite automata (NFA). In the uniform version the NFA are part of the input. In the non-uniform version the NFA are restricted so that each is allowed to have at most k states, where k is a fixed constant not part of the input. Using a recent result of Gutiérrez [5] one can show that the uniform satisfiability problem of word equations with regular constraints can be solved in EXPSPACE (more precisely in DSPACE($2^{\mathcal{O}(d^3)}$), if d denotes the input size), see [3]. So, from the general case it is not really clear whether adding regular constraints makes the satisfiability problem of word equations harder. We give here however some evidence that, indeed, it does. Restricted to quadratic systems the uniform satisfiability problem with regular constraints becomes PSPACE complete.

The non-uniform version is NP-hard and it remains NP-hard, even if the lengths $b_x, x \in \Omega$, are given in unary as part of the input. This is in sharp contrast to Theorem 2. Having regular constraints it is also easy to find examples where the length of a minimal solution increases exponentially.

Theorem 3. *The following problem is PSPACE-complete.*

INSTANCE: A quadratic system of word equations with a list of regular constraints $L_x \subseteq A^*$, $x \in \Omega$.

QUESTION: Is there a solution $\sigma \colon \Omega \longrightarrow A^*$ such that $\sigma(x) \in L_x$ for all $x \in \Omega$?

Moreover, the problem remains PSPACE-complete, if the input is given together with a list of numbers b_x, $x \in \Omega$ (a number $b \in \mathbb{N}$ resp.), written in binary, and if we ask for a solution satisfying in addition the requirement $|\sigma(x)| = b_x$ ($|\sigma(x)| = b$ resp.) for all $x \in \Omega$?

Proof. The PSPACE-hardness follows directly from a well-known result on regular sets. Let $L_1, \ldots, L_n$ be regular languages specified by NFA. Then the emptiness problem $L_1 \cap \cdots \cap L_n = \emptyset$ is PSPACE-complete, [7]. If the intersection is not empty, then there is a witness of at most exponential length. Let b be this upper bound on the length of a witness. Using a new letter c such that $c \notin A$, we can ask whether the intersection

$$L_1 c^* \cap \cdots \cap L_n c^*$$

contains a word of length b. (Instead of using a new letter we may also use some coding provided $|A| \geq 2$.) The quadratic system is given by n variables $x_1, \ldots, x_n$ and regular constraints $L_{x_i} = L_i c^*$ for $1 \leq i \leq n$. The equations are trivial:

$$x_1 = x_2, \; x_2 = x_3, \; \ldots \; , \; x_{n-1} = x_n.$$

The PSPACE algorithm for the uniform satisfiability problem is a modification of the proof of Theorem 2. Let r_x be the number of states of the NFA specifying the language L_x. Let $r = \sum_{x \in \Omega} r_x$. Then there are a homomorphism $\varphi : A^* \longrightarrow \mathbb{B}^{r \times r}$, and vectors $I_x, F_x \in \mathbb{B}^r$ such that

$$w \in L_x \Longleftrightarrow I_x^T \cdot \varphi(w) \cdot F_x = 1$$

We run the non-deterministic decision algorithm for satisfiability as usual. A variable $x \in \Omega$ is however represented as a pair (x, B_x) where $B_x \in \mathbb{B}^{r \times r}$.

When we use a variable for the first time, we check in PSPACE that $B_x \in \varphi(A^*)$. A constant $a \in A$ is represented by the pair $(a, \varphi(a))$. Whenever we meet an equation $z = xy$, we check (in polynomial time) that $B_z = B_x \cdot B_y$.

Theorem 4. *Let $r \geq 4$ be a fixed constant, which is not part of the input. The following problem is NP-complete.*

INSTANCE: A quadratic system of word equations with a list of natural numbers $b_x \in \mathbb{N}$ written in binary, a list of regular constraints $L_x \subseteq A^*$, $x \in \Omega$, such that each language can be specified by some NFA of at most r states, and $|A| \geq 2$.

QUESTION: Is there a solution $\sigma\colon \Omega \longrightarrow A^*$ such that $|\sigma(x)| = b_x$ and $\sigma(x) \in L_x$ for all $x \in \Omega$?

Moreover, the problem remains NP-hard, if the numbers b_x, $x \in \Omega$, are written in unary, $|A| = 2$, and the system is a single equation.

Proof. The non-uniform problem is in NP, since a test whether a Boolean matrix B belongs to the image $\varphi(A^*)$ can now be performed in polynomial time (in fact, in constant time). Therefore the algorithm mentioned in the proof of Theorem 3 runs in non-deterministic polynomial time.

We have to show that the problem is NP-hard, even if the numbers $b_x, x \in \Omega$, are written in unary. First we use a reduction to a system of equations similar to that from 3-SAT in the proof of Theorem 1. Hence, we start with $F = C_0 \wedge \cdots \wedge C_{m-1}$ and each clause C_i has the form

$$C_i = (\tilde{X}_{3i} \vee \tilde{X}_{3i+1} \vee \tilde{X}_{3i+2})$$

We use word variables:

$$\begin{aligned} c_i &, \; 0 \leq i \leq m-1, \\ x_j &, \; 0 \leq j \leq 3m-1, \\ y_X, z_X &, \text{ for each } X \in \Xi. \end{aligned}$$

As above for $X \in \Xi$ let $\{i_1, \ldots, i_k\}$ be the set of positions where $X = \tilde{X}_{i_1} = \cdots = \tilde{X}_{i_k}$ and let $\{j_1, \ldots, j_n\}$ be the set of positions where $\bar{X} = \tilde{X}_{j_1} = \cdots = \tilde{X}_{j_n}$. We define the following equations:

$$\begin{aligned} c_i &= x_{3i} x_{3i+1} x_{3i+2} \;, \quad 0 \leq i \leq m-1 \\ y_X &= x_{i_1} \cdots x_{i_k} z_X x_{j_1} \cdots x_{j_n} \;, \quad X \in \Xi. \end{aligned}$$

The lengths are given by $|c_i| = 3$, $0 \leq i \leq m-1$, $|x_j| = 1$, $0 \leq j \leq 3m-1$, $|z_X| = 2$ and $|y_X| = 2 + k + n$ for every $X \in \Xi$. The regular constraints are from the following list: $A^*, A^*aA^*, ab \cup ba$, and $a^*b^* \cup b^*a^*$. Allowing several initial states we need at most 4 states for each language. That is why we have chosen $r \geq 4$. Finally, the regular constraints are:

$$c_i \in A^*aA^* \;, \quad 0 \leq i \leq m-1,$$

$$\begin{aligned} x_j &\in A^* , \quad 0 \le j \le 3m-1, \\ z_X &\in ab \cup ba , \\ y_X &\in a^*b^* \cup b^*a^* . \end{aligned}$$

Again it is straightforward to see that the word equations have a solution satisfying the constraints if and only if the expression is satisfiable.

The final reduction from a system of equations to a single equation over a binary alphabet is simpler using regular constraints than the one sketched in the proof of Theorem 1. We omit the details.

5 On the exponent of periodicity

Let $E = \{L_1 = R_1, \cdots, L_k = R_k\}$ be a system of word equations. The *exponent of periodicity* of E is defined by

$$\sup\{\alpha \in \mathbb{N} \mid \sigma(x) = up^\alpha v, \sigma \text{ is a minimal solution}, x \in \Omega, p \neq \epsilon\}.$$

Example 1. Consider the following quadratic system

$$(x_i = ax_{i-1})_{2 \le i \le n}.$$

Its exponent of periodicity is $n-1$.

It is well-known that the exponent of periodicity is at most singly exponential in the denotational length of the system, see [6]. For quadratic systems we will show a linear bound. A polynomial bound has been proposed by Anca Muscholl and is stated in [1]. Our approach is slightly different and yields the following result.

Theorem 5. *Let E be a quadratic system of word equations. Then the exponent of periodicity is at most linear in the denotational length.*

Proof. The first part is a transformation of a system of word equations to a linear Diophantine system. It relies on the notion of a p-stable normal form and was already used in the original paper by Makanin [10]. This is standard and also explained in detail in [3]. Therefore we only sketch this part:

We may assume that all equations are of the form $z = xy$ with $x,\ y,\ z \in (A \cup \Omega)$. Consider a minimal solution σ and a primitive word $p \in A^+$ such that $\sigma(x) = up^\alpha v$ for some $x \in \Omega$, $u, v \in A^*$ and where α is the exponent of periodicity of the system E. Every word $u \in A^*$ can be written in its p-stable normal form. This is a factorization

$$u = u_0 p^{\alpha_1} u_1 \cdots p^{\alpha_k} u_k$$

such that k is minimal and, if p^2 is a factor of u, then $k > 0$, $u_0 \in A^*p \setminus A^*p^2A^*$, $u_1, \ldots, u_{k-1} \in (A^*p \cap pA^*) \setminus A^*p^2A^*$ and $u_k \in pA^* \setminus A^*p^2A^*$. As the name suggests, the p-stable normal form is unique. Now, an equation $z = xy$ yields an equation $\sigma(z) = \sigma(x)\sigma(y)$. We obtain an equation on the p-stable normal forms:

$$w_0 p^{\gamma_1} w_1 \cdots p^{\gamma_k} w_k p^{\gamma_{k+1}} \cdots p^{\gamma_m} w_m = u_0 p^{\alpha_1} u_1 \cdots p^{\alpha_k} u_k v_0 p^{\beta_1} v_1 \cdots p^{\beta_l} v_l.$$

Reading the exponents α_i, β_i, γ_i as integer variables, we obtain a number of trivial equations $\alpha_1 = \gamma_1$, $\alpha_2 = \gamma_2$, $\cdots$ and at most 2 non-trivial equations. For instance we might have $\gamma_k = \alpha_k + 1$, $\gamma_{k+1} = \beta_1 + 1$. Another possibility is $\gamma_k = \alpha_k + \beta_1 + 3$. In fact, as the reader can easily verify there are nine cases and which one appears depends only on the p-stable normal form of $u_k v_0$. The two cases just mentioned are so to speak *worst cases*.

Putting together the equations derived from each word equation, we obtain a (perhaps huge) linear Diophantine system where every non-zero coefficient is 1 and every variable occurs at most twice. There are $\mathcal{O}(d)$ non-trivial equations and the constant in each equation has absolute value at most 3. Moreover σ provides us with a positive solution and this solution is minimal among all positive solutions (because σ is a minimal solution of E). The trivial equations such as $\alpha_i = \gamma_i$ can be eliminated leaving us with a linear Diophantine system with a rather special structure: there are at most m equations and n variables $X = \{x_1, \cdots, x_n\}$ where $m,\ n \in \mathcal{O}(d)$. We can write each equation as $\sum a_{i,j} x_j = b_i$ where $|b_i| \leq 3$ and $\sum_i |a_{i,j}| \leq 2$ for each j. This is just a system of 2-type as introduced in the next section. The exponent of periodicity is bounded by the maximum value occurring in all minimal solutions of this system. Therefore Theorem 5 is a direct consequence of Theorem 6 which will be stated and shown in the next section.

Remark 3. The statement of Theorem 5 can be generalized in order to include regular constraints by the same techniques as in [3]. This is left to the interested reader.

6 Linear Diophantine Equations of 2-type

Let $X = \{x_1, \cdots, x_n\}$ be a set of integer variables. A system of linear Diophantine equations is specified by an integer matrix $A = (a_{i,j}) \in \mathbb{Z}^{m \times n}$ and a vector $b \in \mathbb{Z}^m$. Without restriction we have $m = n$. A solution (resp. positive solution) is a vector $x \in \mathbb{Z}^n$ (resp. $x \in \mathbb{N}^n \setminus \{0\}$) such that $Ax = b$.

A positive solution $x \in \mathbb{N}^n \setminus \{0\}$ is minimal if, for all other solutions $x' \in \mathbb{N}^n \setminus \{0\}$, there is at least one component j such that $x_j < x'_j$. The system is called here of *2-type* if for all $1 \leq j \leq k$ we have $\sum_i |a_{i,j}| \leq 2$. We consider only systems of 2-type. These systems correspond to quadratic systems of word equations. The aim of this section is to give an upper bound on the size of all minimal positive solutions. The size of a vector $z \in \mathbb{Z}^n$ is its *norm* defined by

$$||z|| = \sum_{1 \leq i \leq n} |z_i|.$$

Lemma 1. *Every system of 2-type $Ax = b$ with a non-zero integer solution has a non-zero integer solution where $|x_\ell| \leq \max\{2,\ ||b||\}$ for all $1 \leq \ell \leq n$.*

Proof. We give an algorithm to find such a solution. The algorithm works in two phases, *elimination* and *solution.* The elimination phase produces a *core* system of equations in the uneliminated variables such that no variable occurs in two equations and a *defining* equation for each eliminated variable.

Then the solution phase finds a solution to the core system and extends it to the eliminated variables by iterated substitution in the defining equations. As a preliminary step, we replace each $b_i \neq 0$ by a new variable x_{n+i}.

The elimination phase proceeds in stages, each of which eliminates one variable. Initially all equations are in the core, no variables are eliminated, and all variables occur in some equation. If no variable occurs in two equations of the core, then the phase is complete; otherwise suppose that x_i occurs in equations j and j'. It follows that $i \leq n$, and we must have $|\alpha_{i,j}| = |\alpha_{i,j'}| = 1$ so we can rewrite these two equations in the form $x_i = \sum_{k\neq i} \alpha_{j,k} x_k$ and $x_i = \sum_{k\neq i} \alpha_{j',k} x_k$. We replace these two equations in the core by $\sum_{k\neq i}(\alpha_{j,k} - \alpha_{j',k})x_k = 0$, eliminate x_i and give x_i the defining equation $x_i = \sum_{k\neq i} \alpha_{j,k} x_k$.

The elimination phase produces a set of equations (core and defining) equivalent to the initial system. Now the solution phase starts by replacing the new variables x_{n+i} by the original constants b_i in the core and in the defining equations. The final core equations now have the form

$$\sum_{k\in I} \alpha_k x_k + \sum_{\ell\in J} \beta_\ell x_\ell = b'$$

where $|\alpha_k| = 1$, $|\beta_\ell| \in \{0,\ 2\}$, $|b'| \leq ||b||$, and I and J are subsets of uneliminated variables of $\{1, \ldots, n\}$. (Note that we might have $b' = 0$ although $||b|| \neq 0$, but not vice versa.) The final core equations can be solved independently. Since the system has a solution, each equation can be solved in such a way that $\sum_{k\in I} |x_k| + \sum_{\ell\in J} 2|x_\ell| \leq |b'| \leq ||b||$. Moreover, if $||b|| = 0$, then, since the system has a non-zero solution, at least one core equation has a non-zero solution; for any such equation we have a solution with $\sum_{k\in I} |x_k| + \sum_{\ell\in J} 2|x_\ell| \leq 4$. Hence in all cases, there is a non-zero solution satisfying

$$\sum_{k\in I} |x_k| + \sum_{\ell\in J} 2|x_\ell| + ||b|| \leq \max\{4,\ 2||b||\}.$$

We define for each uneliminated variable x_k and each core equation i a weight $w_{i,k}$. In the initial system $w_{i,k} = |\alpha_{i,k}|$. For the replacement equation obtained in an elimination stage (to have a name let's call the replacement equation r) we define $w_{r,k}$ as $w_{j,k} + w_{j',k}$. Thus we always have $|\alpha_{i,k}| \leq w_{i,k} \leq 2$ and $w_{i,k} = 1$, if $|\alpha_{i,k}| = 1$, for every equation and uneliminated variable.

Claim. For every solution $x \in \mathbb{Z}^n$ and at every stage of the elimination phase, for every ℓ, there is a core equation j such that $2|x_\ell| \leq \sum_k w_{j,k}|x_k|$.

The claim is clearly true for all uneliminated variables since a positive $w_{j,\ell}$ occurs in at least one core equation and either $w_{j,\ell} = 2$ (and the claim becomes trivial) or $w_{j,\ell} = 1$ and then we have

$$|x_\ell| = |\sum_{k\neq\ell} \alpha_{j,k} x_k| \leq \sum_{k\neq\ell} |\alpha_{j,k}||x_k| \leq \sum_{k\neq\ell} w_{j,k}|x_k|.$$

In particular the claim is true for the initial system. Now consider a stage eliminating x_i and replacing equations j and j' by the new core equation r. As can be seen by the defining equation for x_i we have $2|x_i| \leq \sum_{k\neq i} w_{r,k}|x_k|$. So the claim is trivial for i. Thus, the elimination stage can affect the truth of the claim, only if it removes the core equation j satisfying the claim for x_ℓ and $\ell \neq i$. We can make the following calculation:

$$\begin{array}{lll}
2|x_\ell| & \leq \sum_k w_{j,k}|x_k| & \text{before the elimination} \\
 & \leq (\sum_{k\neq i} w_{j,k}|x_k|) + |x_i| & \text{because } w_{j,i} = 1 \\
 & \leq (\sum_{k\neq i} w_{j,k}|x_k|) + (\sum_{k\neq i} \alpha_{j',k}|x_k|) & \\
 & = \sum_{k\neq i}(w_{j,k} + \alpha_{j',k})|x_k| & \\
 & \leq \sum_{k\neq i}(w_{j,k} + w_{j',k})|x_k| & \\
 & = \sum_k w_{r,k}|x_k| & \text{after the elimination.}
\end{array}$$

This completes the proof of the claim. Finally the assertion of the lemma follows since for the final core equations we have constructed a solution (yielding a non-zero solution of the initial system) such that

$$2|x_\ell| \leq \sum_k w_{r,k}|x_k| \leq \sum_{k\in I} |x_k| + \sum_{\ell\in J} 2|x_\ell| + ||b|| \leq \max\{4, 2||b||\}.$$

Lemma 2. *Let* $\det(A) \neq 0$. *If* $x \in \mathbb{Z}^n$ *solves* $Ax = b$, *then for all* $1 \leq \ell \leq n$ *we have* $|x_\ell| \leq \max\{2,\ ||b||\}$. *In particular, if* $||b|| \in \mathcal{O}(n)$, *then* $|x_\ell| \in \mathcal{O}(n)$.

Proof. Immediate by Lemma 1 since the solution is unique.

Lemma 3. *Let* $\det(A) = 0$. *Then the homogeneous system* $Ax = 0$ *has a non-zero solution* $x \in \mathbb{Z}^n$ *with* $|x_\ell| \leq 2$ *for all* $1 \leq \ell \leq n$.

Proof. Since the determinant is zero, there is a non-zero integer solution and we use Lemma 1.

Theorem 6. *Let* $x \in \mathbb{N}^n \setminus \{0\}$ *be a minimal solution of* $Ax = b$, $1 \leq \ell \leq n$ *and* $||b|| \in \mathcal{O}(n)$. *Then* $|x_\ell| \in \mathcal{O}(n)$.

Proof. For $\det(A) \neq 0$ this is clear by Lemma 2. Hence we assume that $\det(A) = 0$. Consider now first the case that $Ax = 0$ has a positive solution with all components at most 2. Since the given solution is minimal, we cannot subtract this positive solution from it and remain in $\mathbb{N}^n \setminus \{0\}$; hence it has some component x_j with value v_j less than or equal to 2. We replace x_j

by v_j to obtain a new system $A'x = b'$ for which $(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots x_n)$ is still minimal and $\|b'\| \leq \|b\| + 4$. Repeating as long as $A'x = 0$ has a positive solution with all components at most 2, we eventually are done or we reach a system $A'x = b'$ with $\|b'\| \in \mathcal{O}(n)$ for which all positive solutions of $A'x = 0$ have a component at least 3. Again, if $\det(A') \neq 0$, Lemma 2 gives the result. So we assume $\det(A') = 0$. By Lemma 3 there is still a non-zero solution z of $A'z = 0$ where the absolute value of all components is at most 2. If necessary replacing z by $-z$, we have $z_\ell \geq 0$. There is a maximum non-negative integer λ such that $x' = x + \lambda z$ is a positive solution of $A'x = b'$. The solution x' has some component (say x'_j) with value v_j at most 2 and $x'_\ell \geq x_\ell$. (It might happen that x' is not minimal anymore.) Again we replace x_j by v_j to get a new system where no positive solution of $A'x = 0$ has all components at most 2. (Because otherwise the old system had such a solution with $v_j = 0$.) Repeating this process, we eventually reach a system $A''x = b''$ with $\|b''\| \in \mathcal{O}(n)$ and $\det(A'') \neq 0$ and having a solution x'' with $x''_\ell \geq x_\ell$. Thus using Lemma 2 we have a $\mathcal{O}(n)$ bound on x''_ℓ and so on x_ℓ completing the proof.

The following example is the Diophantine version of Example 1 showing that our bounds are essentially optimal.

Example 2. Consider the following linear Diophantine system of 2-type

$$(x_i - x_{i-1} = 1)_{2 \leq i \leq n}.$$

Its minimal positive solution is $x = (0, 1, \ldots, n-1)$.

References

1. Thierry Arnoux. Untersuchungen zum Makaninschen Algorithmus. Diplomarbeit 1613, Universität Stuttgart, 1998.
2. Christian Choffrut and Juhani Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
3. Volker Diekert. Makanin's Algorithm. In M. Lothaire: *Algebraic Combinatorics on Words*. Cambridge University Press. A preliminary version is on the web: http://www-igm.univ-mlv.fr/ berstel/Lothaire/index.html.
4. Volker Diekert, Yuri Matiyasevich, and Anca Muscholl. Solving trace equations using lexicographical normal forms. In P. Degano et al., editors, *Proc. 24th ICALP, Bologna (Italy) 1997*, number 1256 in Lect. Notes Comp. Sci., pages 336–347. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
5. Claudio Gutiérrez. Satisfiability of word equations with constants is in exponential space. In *Proc. 39th FOCS*, pages 112–119, Los Alamitos, California (USA), 1998. IEEE Computer Society Press.
6. Antoni Kościelski and Leszek Pacholski. Complexity of Makanin's algorithm. *J. Assoc. Comput. Mach.*, 43(4):670–684, 1996. Preliminary version in *Proc. 31st FOCS*, Los Alamitos, California (USA), 1990.

7. Dexter Kozen. Lower bounds for natural proof systems. In *Proc. 18th FOCS, Providence, Rhode Island (USA)*, pages 254–266. IEEE Computer Society Press, 1977.
8. Roger C. Lyndon. Equations in free groups. *Transactions of the American Mathematical Society*, 96, 1960.
9. Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, Berlin, Heidelberg, New York, 1977.
10. Gennadií S. Makanin. The problem of solvability of equations in a free semigroup. *Mat. Sb.*, 103(2):147–236, 1977. In Russian; English translation in: *Math. USSR Sbornik, 32*, 129–198, 1977.
11. Yuri Matiyasevich. A connection between systems of word and length equations and Hilbert's Tenth Problem. *Sem. Mat. V. A. Steklov Math. Inst. Leningrad*, 8:132–144, 1968. In Russian; English translation in: *Seminars in Mathematics, V. A. Steklov Mathematical Institute, 8*, 61–67, 1970.
12. Yuri Matiyasevich. Some decision problems for traces. In S. Adian et al., editors, *Proc. 4th LFCS, Yaroslavl (Russia) 1997*, number 1234 in Lect. Notes Comp. Sci., pages 248–257. Springer-Verlag, Berlin, Heidelberg, New York, 1997. Invited lecture.
13. Jakob Nielsen. Die Isomorphismen der allgemeinen, unendlichen Gruppe mit zwei Erzeugenden. *Mathematische Annalen*, 78, 1918.
14. Dominique Perrin. Equations in words. In H. Ait-Kaci et al., editors, *Resolution of equations in algebraic structures*, volume 2, pages 275–298. Academic Press, 1989.
15. Wojciech Plandowski. Satisfiability of word equations with constants is in NEXPTIME. In *Proc. 31st STOC, Atlanta, Georgia (USA)*, 1999. To appear.
16. John Michael Robson. Word equations with at most 2 occurrences of each variable. Preprint, LaBRI, Université de Bordeaux I, 1998.
17. John Michael Robson and Volker Diekert. On quadratic word equations. In Ch. Meinel et al., editors, *Proc. 16th STACS, Trier (Germany)*, Lect. Notes Comp. Sci. Springer-Verlag, Berlin, Heidelberg, New York, 1999. To appear.
18. Wojciech Rytter and Wojciech Plandowski. Application of Lempel-Ziv encodings to the solution of word equations. In K. G. Larsen et al., editors, *Proc. 25th ICALP, Aalborg (Denmark)*, number 1443 in Lect. Notes Comp. Sci., pages 731–742. Springer-Verlag, Berlin, Heidelberg, New York, 1998.
19. Klaus U. Schulz. Makanin's algorithm for word equations: Two improvements and a generalization. In K.-U. Schulz, editor, *Proc. IWWERT'90, Tübingen (Germany)*, number 572 in Lect. Notes Comp. Sci., pages 85–150. Springer-Verlag, Berlin, Heidelberg, New York, 1992.

Index