Upper Bounds on the Length of Minimal Solutions to Certain Quadratic Word Equations

Joel D. Day[1]
Loughborough University, UK
Kiel University, Germany
J.Day@lboro.ac.uk

Florin Manea
Kiel University, Germany
flm@informatik.uni-kiel.de

Dirk Nowotka
Kiel University, Germany
dn@informatik.uni-kiel.de

Abstract

It is a long standing conjecture that the problem of deciding whether a quadratic word equation has a solution is in NP. It has also been conjectured that the length of a minimal solution to a quadratic equation is at most exponential in the length of the equation, with the latter conjecture implying the former. We show that both conjectures hold for some natural subclasses of quadratic equations, namely the classes of regular-reversed, k-ordered, and variable-sparse quadratic equations. We also discuss a connection of our techniques to the topic of unavoidable patterns, and the possibility of exploiting this connection to produce further similar results.

2012 ACM Subject Classification Mathematics of computing → Combinatorics on words; Theory of computation → Problems, reductions and completeness

Keywords and phrases Quadratic Word Equations, Length Upper Bounds, NP, Unavoidable Patterns

Digital Object Identifier 10.4230/LIPIcs.MFCS.2019.44

Funding Florin Manea's work was supported by the DFG grant MA 5725/2-1.

1 Introduction

A *word equation* is an equation $\alpha = \beta$ in which the two sides, α and β, are words consisting of letters from a *terminal alphabet* $\Sigma = \{\mathtt{a}, \mathtt{b}, \ldots\}$ and *variables* from a set $X = \{x, y, z, \ldots\}$. It has a solution if the variables may be substituted for words over Σ in such a way that the two sides become identical. For example, the equation $\mathtt{ab}x\mathtt{b}y = x\mathtt{b}yx\mathtt{b}$ has a solution where x is substituted by $\mathtt{a}$, and y is substituted by $\mathtt{ab}$. Usually such a substitution is represented by a morphism $h : (X \cup \Sigma)^* \to \Sigma^*$ which preserves the terminal symbols (i.e. $h(\mathtt{a}) = \mathtt{a}$ for all $\mathtt{a} \in \Sigma$). Situated in the intersection between computer science and algebra, word equations are an important tool for describing structural relations between words, and as such are of interest in a variety of areas, ranging from combinatorial group and monoid theory [20, 19, 7], to unification [25, 12, 14], database theory [11, 10], model checking, verification and security, where there has been much interest recently in developing so-called string solvers capable of dealing with word equations such as HAMPI [17], CVC4 [3], Stranger [28], ABC [2], Norn [1], S3P [27] and Z3str3 [4]. Of course, not all equations have solutions (consider the trivial example $\mathtt{a}x = \mathtt{b}x$, or less trivial ones such as $\mathtt{abc}x = x\mathtt{cba}$), and the problem of deciding whether a given equation has a solution – the satisfiability problem – has been at the

[1] Corresponding author

44th International Symposium on Mathematical Foundations of Computer Science (MFCS 2019).
Editors: Peter Rossmanith, Pinar Heggernes, and Joost-Pieter Katoen; Article No. 44; pp. 44:1–44:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

centre of research on word equations since their inception. It is not difficult to see that the satisfiability problem contains an inherent degree of computational complexity. In particular, there are several simple reductions from NP-complete problems such as the membership problem for pattern languages [9, 8] and linear arithmetic. On the other hand, bounding the complexity from above has proven considerably more challenging. After considerable effort, Makanin [21] famously showed that the satisfiability problem for word equations is algorithmically decidable. This result was later improved by Plandowski [23] who gave an algorithm which requires only polynomial space, and more recently this has been refined further to linear space by Jeż [13, 15] using the elegant method of recompression. Nevertheless, determining the precise complexity, and in particular whether the problem is contained in NP, remains one of the outstanding open problems in the area.

One method for obtaining upper bounds on the complexity is to consider the lengths of minimal solutions – those for which no shorter solution exists to the same equation. Since it may easily be checked in polynomial time (in the length of the substitution) whether a given substitution satisfies a given word equation: simply apply the substitution to each side and compare the resulting words, we have a clear relation between the lengths of minimal solutions and the complexity of the satisfiability problem. If the minimal solutions are guaranteed to be short enough (e.g. polynomial in the length of the equation), we have a non-deterministic polynomial time algorithm which simply guesses a solution and then checks it. In fact, Plandowski and Rytter [24] showed that minimal solutions may be compressed substantially in such a way that their compressed versions may still be checked efficiently, so a similar approach works for equations whose minimal solutions are at most exponentially long in the length of the equation. Unfortunately, the best known upper bound on the length of minimal solutions to word equations in general is double exponential. It is worth pointing out that it is not difficult to construct examples of equations for which the lengths of minimal solutions are (single) exponential in the length of the equation.

In the absence of matching upper and lower bounds on the complexity for the whole class of word equations, it makes sense to first consider subclasses. For example, equations with at most two distinct variables (which may each occur multiple times) can be solved in polynomial time (see, e.g. [6]), and thus do not exhibit the same intractability as more general classes, while another class which is generally well understood is the class of equations in which variables may only occur on one side (i.e. either $\alpha \in \Sigma^*$ or $\beta \in \Sigma^*$). In this case, the satisfiability problem is simply the (NP-complete) problem of pattern matching with variables, for which the computational complexity has been studied extensively. Aside from these examples, however, there seem to be surprisingly few classes of equations for which the satisfiability problem is known to be contained in NP, especially amongst those with corresponding hardness results.

Quadratic word equations are word equations in which each variable may occur at most twice – although the number of variables is unrestricted. There are many reasons that the quadratic equations form a particularly interesting subclass of equations. On one hand, NP-hardness remains (see [8]) but also even for the very restricted case in which the variables must occur in exactly the same order on both sides and only the terminal symbols may vary (see [5]). On the other hand, unlike the general case, there is a straightforward proof that the satisfiability problem for quadratic word equations is decidable, using so-called Nielsen transformations (see [18]). Moreover, while examples of equations with three occurrences of each variable are known for which the minimal solutions are exponentially long in the length of the equation, no such examples are known for quadratic equations. However, while these results seem to indicate that quadratic equations may not be as complex as the general case, understanding quadratic equations, and in particular determining whether the satisfiability of quadratic equations is in NP, has proven exceptionally difficult and as with the general case, remains a long-standing open problem.

Our Contribution. In the current paper, we further develop a method for showing upper bounds on the length of solutions to quadratic word equations first introduced in [5], and use it to obtain such bounds for several subclasses of quadratic equations. The method extends the existing technique of *filling the positions* (see [16, 24]), and relies on arranging the individual positions of a solution, as referenced by their origin (e.g. the third letter in the second occurrence of the variable x) into chains, which may be represented as words (chain-words). This chains-representation of solutions is discussed in detail in Section 3.

We show firstly that quadratic equations with a high concentration of terminal symbols and variables occurring only once – with few variables occurring the maximal two times – have minimal solutions at length at most $n2^{2^{O(V^4)}}$ where V is the number of variables occurring twice and n is the length of the equation. Using the previously mentioned algorithm of Plandowski and Rytter, it follows that for the class of equations for which $V < \log(n)$, which we shall refer to as *variable-sparse* quadratic equations, the satisfiability problem is in NP. Moreover, by observing that variables occurring only once do not have a dramatic impact on the length of minimal solutions (Proposition 2), we also obtain that if V is bounded by a constant, the satisfiability problem may be solved in polynomial time.

As a straightforward consequence, we are also able to show that equations which may be obtained by concatenating many "small" quadratic equations (over disjoint, constant-size sets of variables) also have short solutions, thus may be solved in non-deterministic polynomial time. Such equations may be arbitrarily disordered at a local level – we do not restrict the structure of each individual equation before concatenating – but possess a global order in which the sets of variables must occur from left to right in each side of the equation. Since these equations, which we shall call k-ordered equations, generalise the regular-ordered equations considered in [5], we also get the corresponding NP-hardness lower bound in this case. Thus, the remaining "interesting", cases must possess some global disorder among the variables.

Our main result (Theorem 21) establishes upper bounds of $n2^{3n}$ on the length of minimal solutions to a subclass of quadratic equations for which this necessary separation of the variables is maximal, namely the class of regular-reversed equations, in which the variables occur in the opposite order on the LHS as on the RHS. We also show (Proposition 14) that this bound is close to the best that we may expect using our approach. The proof of Theorem 21, although technical, revolves around two simple operations on the chains-representation of a solution. The first is a compression of certain subchains, which produces a new, shorter solution to an equation of the same length. In a sense, this kind of compression generalises the Nielsen transformations mentioned previously. The second operation, given by Lemma 22 and taking advantage of the structure induced by the carefully chosen compression, is a simple deletion of a variable from the equation (and solution) such that the property of being chain-square-free is preserved.

The final part to our contribution focusses on potential forward steps in the general case, and as such begins to explore a connection to the topic of avoidability of patterns within the field of combinatorics on words. Our main result in this direction is a characterisation of possible chain words in the case of regular equations (equations in which each variable occurs at most once on each side) which is considerably simpler than the more abstract definition given in Section 3. This leaves open a particular interesting problem to determine whether there exist "long" non-repetitive (square-free) words satisfying this characterisation, where a negative answer, for an appropriate definition of "long", would, due to results presented in Section 3, yield that the satisfiability problem for this large subclass of quadratic word equations is in NP. We also discuss, with Lemma 27 as an example, how looking at various structures other than direct repetitions might also be sufficient to obtain the same result.

The rest of the paper is organised as follows. In Section 2, we establish the basic notations and definitions we will need, along with some preliminary results. In Section 3, we shall introduce the main framework used to establish our results. This section recaps the main construction (the chains representation of solutions) and lemma (Lemma 11) originally presented in [5] which form the basis of the framework, and mentions some new additional useful results such as Lemma 12 and Proposition 14. Our main results concerning upper bounds on the length of minimal solutions to equations are presented in Sections 4 and 5. Finally, in Section 6, we present our characterisation of chain-words and discuss the connection to the topic of avoidability of patterns and the resulting possibilities for exploiting Lemma 11 in Section 3 further.

2 Preliminaries

Let Σ be an alphabet. We denote by Σ^* the set of all words over Σ. The empty word is denoted ε. A word u is a *prefix* of a word w if there exists v such that $w = uv$. Similarly, u is a *suffix* of w if there exists v such that $w = vu$, and u is a *factor* of w if there exist v, v' such that $w = vuv'$. A *square* is a word ww for some $w \in \Sigma^*\backslash\{\varepsilon\}$. The length of a word w is denoted $|w|$, and the number of occurrences of a letter $\mathtt{a}$ in a word w is denoted $|w|_{\mathtt{a}}$. The i^{th} letter of w, as counted from the left, is denoted $w[i]$. The *reversal* of a word w is the word $w^R = w[|w|]w[|w|-1]\dots w[2]w[1]$. For two alphabets Σ_1, Σ_2, a morphism $h : {\Sigma_1}^* \to {\Sigma_2}^*$ is a mapping such that, for all $u, v \in {\Sigma_1}^*$, $h(u)h(v) = h(uv)$. Thus, a morphism is uniquely defined by its image on each letter in Σ_1. In the rest of the paper, we shall distinguish between two alphabets: an (infinite) set X of *variables*, and a *terminal alphabet* Σ. We shall generally assume that the terminal alphabet contains at least two letters, but otherwise its cardinality will not be important for our purposes. For a word $\alpha \in (X \cup \Sigma)^*$, we shall denote the set $\{x \in X \mid |\alpha|_x \geq 1\}$ by $\text{var}(\alpha)$ and the set $\{\mathtt{a} \in \Sigma \mid |\alpha|_{\mathtt{a}} \geq 1\}$ by $\text{alph}(\alpha)$. We shall say a morphism $h : (X \cup \Sigma)^* \to \Sigma^*$ is a *substitution* if $h(\mathtt{a}) = \mathtt{a}$ for all $\mathtt{a} \in \Sigma$.

A *word equation* is a tuple (α, β) (usually written as $\alpha = \beta$) such that $\alpha, \beta \in (X \cup \Sigma)^*$. α and β are called the left hand side (LHS) and right hand side (RHS) respectively. Solutions are substitutions $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ such that $h(\alpha) = h(\beta)$. The length of a word equation $\alpha = \beta$ is $|\alpha\beta|$. The length of a solution h to the equation is $|h(\alpha)|$. A solution is *minimal* if no shorter solution exists. The *Satisfiability Problem* is the decision problem of determining whether, for a given word equation, there exists a solution. The following result of Plandowski and Rytter establishes a relationship between the length of minimal solutions (when they exist) and the computational complexity of the satisfiability problem.

▶ **Theorem 1** ([24]). *Suppose that for a given class of word equations, there exists a polynomial P such that any equation in the class which has a solution, has one whose length is at most $2^{P(n)}$ where n is the length of the equation. Then the satisfiability problem for that class is in* NP*.*

A word equation $\alpha = \beta$ is *quadratic* if $|\alpha|_x + |\beta|_x \leq 2$ for every $x \in X$. It is *regular* if $|\alpha|_x \leq 1$ and $|\beta|_x \leq 1$ for all $x \in X$. It is regular-ordered if it is regular, and the variables occur in the same order in both sides of the equation (i.e. there do not exist $x, y \in X$ and $\alpha', \beta' \in (X \cup \Sigma)^*$ such that $x\alpha'y$ is a factor of α and $y\beta'x$ is a factor of β). It was shown in [5] that the satisfiability for regular-ordered word equations is NP-hard.

Finally, we remark that, when considering asymptotic bounds on the length of minimal solutions to quadratic equations, it is not necessary to consider equations in which a variable occurs only once. Hence we shall, unless otherwise stated, consider classes of equations in which the variables occur exactly twice. All of our results providing upper bounds on the length of minimal solutions (and thus containment in NP) can easily be adapted for classes permitting variables which occur only once.

▶ **Proposition 2.** *Let $\alpha = \beta$ be a quadratic word equation and let $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a minimal solution. Let X_1 be the set of variables occurring exactly once in $\alpha\beta$ and let X_2 be the set of variables occurring twice. Let $g : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to (\mathrm{var}(\alpha\beta) \cup \Sigma)^*$ be the morphism such that $g(x) = h(x)$ if $x \in X_1$ and $g(x) = x$ otherwise. Then the equation $g(\alpha) = g(\beta)$ has length less than $2|\alpha\beta|$, and has a minimal solution $h' : (X_2 \cup \Sigma)^* \to \Sigma^*$ such that $|h'(g(\alpha))| = |h(\alpha)|$.*

3 A Method for Upper Bounds

Despite Theorem 1, showing inclusion in NP often remains a challenge, particularly among those classes for which the corresponding NP-hardness lower bound exists. In the present section, we outline a framework, first presented in [5] for reasoning about the (non)-minimality of solutions which we shall rely on in the proofs of most of our results. The main idea centers around the simple principle that a solution is not minimal if we can remove some parts of it to obtain another, shorter solution. While this is in itself a trivial statement, the consequences of our approach will lead, as we shall see in Section 6, to an entirely non-trivial method of reasoning generally about the lengths of minimal solutions to quadratic equations and a surprising link to the topic of avoidability of patterns, a central theme within combinatorics on words.

Positions in a Solution. Our approach extends the method of “filling the positions”, used to determine whether a word equation has a solution when the lengths of the substitutions for the variables are given explicitly (see [16, 24] for an overview). In this respect, we must first specify precisely what we mean by a position. For our reasoning, it will be convenient to be able to reference parts of the substitution/solution both purely from their location in the full image, and relative to the part of the equation – an occurrence of a variable or terminal symbol – from which they originate. Accordingly, for an equation E given by $\alpha = \beta$ and a substitution $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ such that $|h(\alpha)| = |h(\beta)|$, we define the set of *absolute positions* of h w.r.t. E as $\mathcal{AP}_E^h = \{i \mid 1 \leq i \leq |h(\alpha)|\}$, and the set of *relative positions* (or just positions) of h w.r.t. E as $\mathcal{RP}_E^h = \{(x, i, d) \mid x \in \mathrm{var}(\alpha\beta) \cup \mathrm{alph}(\alpha\beta), 1 \leq i \leq |\alpha\beta|_x, \text{ and } 1 \leq d \leq |h(x)|\}$.

Intuitively, the relative position (x, i, d) corresponds to “the d^{th} letter in the image of the i^{th} occurrence of x”, where occurrences are counted from left to right in $\alpha\beta$. As such, we have exactly two relative positions corresponding to each absolute position (one for the LHS and one for the RHS). Let $f_E^h : \mathcal{RP}_E^h \to \mathcal{AP}_E^h$ be the (non-injective) function mapping a relative position to the corresponding absolute position. More formally, for a relative position $r = (x, i, d)$, let $f_E^h(r)$ be the unique integer j such that $j - d = |h(\gamma)| \mod |h(\alpha)|$ where γ is the prefix of $\alpha\beta$ up to, but not including, the i^{th} occurrence of x (i.e. the longest prefix of $\alpha\beta$ such that $|\gamma|_x < i$). We shall say that a relative position $(x, i, d) \in \mathcal{RP}_E^h$ is a *terminal position* if $x \in \Sigma$ (in which case it is guaranteed that $d = 1$), and otherwise we shall say that the position is *non-terminal* and *belongs* to the variable x.

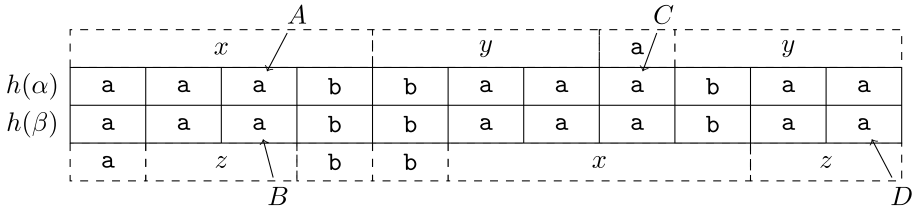

Figure 1 The solution h given by $h(x) = \mathtt{aaab}$, $h(y) = \mathtt{baa}$ and $h(z) = \mathtt{aa}$ to the equation E given by $xy\mathtt{a}y = \mathtt{a}z\mathtt{bb}xz$. Each individual rectangle is a position in $\mathcal{RP}_E^h$. For example, A is the position $(x, 1, 3)$, while D is the position $(z, 2, 2)$, and C is the (terminal) position $(\mathtt{a}, 1, 1)$. The positions A and B are neighbours. B and D are siblings, meaning that A is the successor of D.

Neighbours and Siblings. We shall now introduce the two relations on the set of relative positions used in the method of filling the positions. Firstly, we shall say that two relative positions r_1, r_2 are *siblings* if there exist x, d, i, i' with $i \neq i'$ such that $r_1 = (x, i, d)$ and $r_2 = (x, i', d)$. Secondly, we shall say that two relative positions r_1 and r_2 are *neighbours* if $f_E^h(r_1) = f_E^h(r_2)$ and $r_1 \neq r_2$. The following remarks are immediate:

▶ Remark 3. Each relative position $r = (x, i, d)$ has exactly one neighbour. It has exactly $|\alpha\beta|_x - 1$ siblings. In particular, if $\alpha = \beta$ is quadratic, then either r is a terminal position or r has at most one sibling.

While we will not give a full description of the method of filling the positions in the present paper, it essentially consists of constructing, for a given equation E and solution h, the equivalence relation $\mathcal{R}_E^h$ obtained by the reflexive and transitive closure of the union of the neighbour and sibling relations. The following remarks are standard facts regarding the method of filling the positions and can easily be verified.

▶ Remark 4. A substitution $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ is a solution to the equation E given by $\alpha = \beta$ if and only if $|h(\alpha)| = |h(\beta)|$ and, for any two positions $r_1 = (x, i, d), r_2 = (x', i', d')$ such that $(r_1, r_2) \in \mathcal{R}_E^h$, we have that $h(x)[d] = h(x')[d']$.

▶ Remark 5. If $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ is a solution to the equation E given by $\alpha = \beta$, and there exists an equivalence class of the relation $\mathcal{R}_E^h$ which does not contain a terminal position, then the positions belonging to this class can be substituted by any word without affecting the fact that h is a solution. In particular, they can be substituted by the empty word (i.e. removed altogether), so the solution is not minimal.

The Chains Representation and Chain-words. Suppose now that we have a quadratic equation E given by $\alpha = \beta$ and a solution $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ to E. From now on, we shall only consider quadratic equations. Then by Remark 3, we can organise the relative positions into sequences, or *chains* using the sibling and neighbour relations as follows.

▶ **Definition 6** (Chains Representation of a Solution). *A sequence $r_1, r_2, \ldots, r_n$ of positions from $\mathcal{RP}_E^h$ is a chain-sequence of h with respect to E if*

- *$r_1 = (x_1, i_1, d_1)$ where either x_1 is a terminal symbol from Σ or $|\alpha\beta|_{x_1} = 1$,*
- *r_2 is the neighbour of r_1,*
- *$r_n = (x_n, i_n, d_n)$ where either x_n is a terminal symbol from Σ or $|\alpha\beta|_{x_n} = 1$,*
- *for all j, $2 < j \leq n$, r_j is the neighbour of the sibling of r_{j-1}.*

The set of all chain-sequences is the chains-representation *of h with respect to E. For a position r in a chain, the next position r' in the chain, if it exists, is called the* successor *of r, while r is the* predecessor *of r'. For the sake of brevity, we shall usually refer to chain-sequences simply as "chains".*

▶ Remark 7. Throughout the rest of the paper, we shall only consider solutions for which all equivalence classes of the relation $\mathcal{R}_E^h$ contain a terminal position (otherwise, we can just erase the appropriate parts to obtain a shorter solution, see Remark 5). Under this assumption, all positions $r \in \mathcal{RP}_E^h$ which are not terminal and have a sibling r' will occur (exactly once) in exactly one chain, while terminal positions and positions without a sibling will occur twice (once at the start of a chain-sequence, and once at the end). Thus the length of the solution will be at most half the sum of the lengths of all chains in the chains representation.

▶ Remark 8. For every chain $r_1, r_2, \ldots, r_n$ in the chains representation, there will also exist a dual chain $r_n, \overline{r_{n-1}}, \ldots, \overline{r_2}, r_1$ where, for $2 \leq i \leq n-1$, $\overline{r_i}$ denotes the sibling of r_i. Similarly, for every subchain $r_i, r_{i+1}, \ldots, r_j$ such that r_i and r_j are not terminal positions, there exists a dual subchain $\overline{r_j}, \overline{r_{j-1}}, \ldots, \overline{r_i}$, obtained by reversing and swapping each position for its neighbour. Each equivalence class of the relation $\mathcal{R}_E^h$ corresponds exactly to the positions contained within one chain and its dual.

▶ Remark 9. In the particular case of regular equations, the sibling of a position will always occur on the opposite side of the equation. Since the same is always true by definition for the neighbour of a position, it follows that the successor in the chains representation of a position will belong to the same side. In particular, every position (x, i, d) in the chain which belongs to a variable, with the exception of the first, will have the same value i.

▶ **Definition 10** (Chain-Words and Similarity). *Let Γ_E^h be the set $\{(x, i) \mid x \in \mathrm{var}(\alpha\beta)$ and $i \in \{1, 2\}\}$, and let $\rho : \mathcal{RP}_E^h \to \Gamma_E^h$ be the projection of $\mathcal{RP}_E^h$ onto the first two elements (so that $\rho((x, i, d)) = (x, i)$). For each chain-sequence $C = r_1, r_2, \ldots, r_n$ of h with respect to E, the chain-word induced by C is the word $w = \rho(r_2)\rho(r_3)\ldots\rho(r_{n-1})$ (over the alphabet Γ_E^h). The set of all chain-words induced by chain-sequences of h w.r.t. E is denoted $\Delta_{h,E}$. We shall say that two (sub)chains are* similar *if they induce the same chain-words.*

The following lemma is the main motivation for extending the framework of filling the positions to take into account the additional order expressed in the chains representation. It shall provide the basis for our reasoning later that long solutions to (certain) equations cannot be minimal due to the fact that their chains must contain some repetitive structure.

▶ **Lemma 11** ([5]). *Let E be a quadratic word equation given by $\alpha = \beta$ and let $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a solution to E. If there exists a chain-word $w \in \Delta_{h,E}$ such that w contains a square, then h is not minimal.*

If, for a solution $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ to an equation $\alpha = \beta$, there does not exist a chain-word which contains a square, we shall say that the chains-representation is *square-free*, or that the solution h is *chain-square-free*. The simplest examples of solutions which are not chain-square-free are when two occurrences of the same variable overlap (i.e. the parts of the solution word corresponding to the occurrences of these variables intersect), leading to a square of length one in a chain-word. Generalising this slightly to squares of length two, the following lemma gives another simple example of when a solution is not chain-square-free, which is used in the proofs of our later results.

▶ **Lemma 12.** *Let $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a solution to a quadratic equation E given by $\alpha = \beta$. Let $i, i', j, j' \in \{1, 2\}$ with $i \neq i'$ and $j \neq j'$, and let $x, y \in \mathrm{var}(\alpha\beta)$. If $f_E^h((x, i, 1)) \leq f_E^h((y, j, 1)) \leq f_E^h((x, i, |h(x)|)) \leq f_E^h((y, j, |h(y)|))$ and $f_E^h((x, i', 1)) \leq f_E^h((y, j', 1)) \leq f_E^h((x, i', |h(x)|)) \leq f_E^h((y, j', |h(y)|))$, then h is not chain-square-free.*

The following gives a full example of a solution to a quadratic equation along with its chains representation. The solution is not chains-square-free, and thus a shorter one exists.

▶ **Example 13.** Consider the equation E given by $xy\mathtt{a}y = \mathtt{a}z\mathtt{bb}xz$ over variables $x, y, z \in X$ and terminal symbols $\mathtt{a}, \mathtt{b} \in \Sigma$. Consider the solution $h : ((x, y, z) \cup \Sigma)^* \to \Sigma^*$ such that $h(x) = \mathtt{aaab}$, $h(y) = \mathtt{baa}$ and $h(z) = \mathtt{aa}$. Then the set of relative positions is:

$$\begin{aligned}\mathcal{RP}^h_E = \{&(x,1,1),(x,1,2),(x,1,3),(x,1,4),(x,2,1),(x,2,2),(x,2,3),(x,2,4),\\ &(y,1,1),y(1,2),(y,1,3),(y,2,1),(y,2,2),(y,2,3),\\ &(z,1,1),(z,1,2),(z,2,1),(z,2,2),(\mathtt{a},1,1),(\mathtt{a},2,1),(\mathtt{b},1,1),(\mathtt{b},2,1)\},\end{aligned}$$

the chains-representation of h consists of the four chains:

$$\begin{aligned}C_1 &= (\mathtt{a},1,1),(x,2,3),(z,1,2),(y,2,3),(x,2,2),(z,1,1),(y,2,2),(x,2,1),(\mathtt{a},2,1)\\ C_2 &= (\mathtt{b},1,1),(x,1,4),(y,2,1),(\mathtt{b},2,1)\\ C_3 &= (\mathtt{a},2,1),(x,1,1),(y,1,2),(z,2,1),(x,1,2),(y,1,3),(z,2,2),(x,1,3),(\mathtt{a},1,1)\\ C_4 &= (\mathtt{b},2,1),(y,1,1),(x,2,4),(\mathtt{b},1,1)\end{aligned}$$

where C_1 and C_3 are dual, as are C_2 and C_4. The set of chain-words is given as follows, where for ease of reading, the elements $(x, 1)$,$(x, 2)$, $(y, 1)$, $(y, 2)$, $(z, 1)$ and $(z, 2)$ of Γ^h_E are denoted A, B, C, D, E and F respectively.

$$\Delta_{h,E} = \{BEDBEDB, AD, ACFACFA, CB\}.$$

The fact that one of the chain-words contains a square (e.g. $BEDBED$ in the first one) means that, due to Lemma 11, a shorter solution exists. In this case, we have the solution h' given by $h'(x) = \mathtt{aab}$, $h'(y) = \mathtt{ba}$ and $h'(z) = \mathtt{a}$.

Finally, we point out that this general approach of analysing the chain-words for squares can, at best, give exponential upper bounds on the length of minimal solutions to quadratic word equations, and additionally that the property of being chain-square-free, while necessary, is not sufficient for being minimal.

▶ **Proposition 14.** *Let $n \in \mathbb{N}$. Then the equation $x_n x_{n-1} \dots x_1 \mathtt{a} = \mathtt{a} x_1 \dots x_{n-1} x_n$ has a solution $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ which is not minimal but is chain-square-free such that $|h(x_n x_{n-1} \dots x_1 \mathtt{a})| = 2^n$.*

4 Variable-Sparse and k-Ordered Quadratic Equations

We begin in the current section by considering quadratic equations without many repeating variables – those equations $\alpha = \beta$ for which the set $\{x \mid |\alpha\beta|_x = 2\}$ is "small". Let the variable-sparse equations be defined as follows.

▶ **Definition 15.** *A word equation $\alpha = \beta$ is* variable-sparse *if $|\{x \mid |\alpha\beta|_x \geq 2\}| \leq \log(|\alpha\beta|)$.*

While there are practical reasons to care about variable-sparse equations – it seems reasonable to expect that equations encountered in practice may often have this form – we can also take advantage of the insights gained by considering this class to settle the complexity of another, more general class, namely the k-ordered equations, which complements, and thus motivates, the class of regular-reversed equations considered in the next section.

As the next proposition shows, the lengths of possible chains in the chains-representation of a minimal solution is bounded by a double-exponential function in number of repeating variables, and consequently so is the length of minimal solutions. The double-exponential bound is derived from the upper bound on lengths of minimal solutions to the whole class of word equations (see [22]), and thus we do not expect it to be tight. However, it is sufficient to show that in the case of variable-sparse quadratic equations, minimal solutions are at most single exponential in the length of the equation, and thus that the satisfiability problem for this class is contained in NP.

▶ **Proposition 16.** *Let $\alpha = \beta$ be a quadratic equation and let $V = |\{x \mid |\alpha\beta|_x = 2\}|$. If there exists a solution to $\alpha = \beta$, then there exists a solution $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ to $\alpha = \beta$ such that each chain in the chains representation of h w.r.t. $\alpha = \beta$ has length at most $2^{2^{O(V^4)}}$. Consequently, if h is minimal, then $|h(\alpha)| \leq |\alpha\beta| 2^{2^{O(V^4)}}$. Moreover, it follows that the satisfiability problem for the class of variable-sparse quadratic equations is in* NP.

If we limit the number of variables further, bounded by a constant instead of $\log(|\alpha\beta|)$, then the bound on the length of minimal solutions becomes polynomial in the length of the equation, and we are able to infer the following.

▶ **Proposition 17.** *Let $V \in \mathbb{N}$ be a constant. Then the satisfiability problem for the class of quadratic equations with at most V variables can be solved in polynomial time.*

The k-ordered equations generalise the regular-ordered equations considered in [5], and essentially rely on the same idea of an order in which the variables must occur from left to right in both sides of the equation. However, this order is only enforced on small subsets of variables as a whole, and does not dictate the local order in which variables in a single subset may occur. For example, $x_1x_2x_3x_4x_5x_6 = x_3x_2x_1x_6x_4x_5$ is 3-ordered, but not 2-ordered.

▶ **Definition 18.** *Let $k \in \mathbb{N}$ and let $\alpha = \beta$ be a quadratic word equation. Then $\alpha = \beta$ is* k-ordered *if there exist pairwise disjoint sets of variables $X_1, X_2, \ldots, X_\ell$ with $|X_i| \leq k$ for $1 \leq i \leq \ell$, and $\alpha_i, \beta_i \in (X_i \cup \Sigma)^*$ for $1 \leq i \leq \ell$ such that $\alpha = \alpha_1\alpha_2 \ldots \alpha_\ell$ and $\beta = \beta_1\beta_2 \ldots \beta_\ell$.*

Length bounds for minimal solutions to k-ordered equations can be derived in a relatively straightforward manner from Proposition 17: after removing variables occurring only once (see Proposition 2), a basic length argument can be used to (non-deterministically) divide a k-ordered equation into linearly many individual equations over pairwise disjoint sets X_i of at most k variables X_i. Since these equations do not share any variables, solutions to the original equation can be obtained by simply combining solutions to the individual equations, which if they exist, due to Proposition 16, may be chosen to be short.

▶ **Proposition 19.** *Let $k \in \mathbb{N}$ be a constant and let $\alpha = \beta$ be a k-ordered quadratic word equation. Let $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a minimal solution to $\alpha = \beta$. Then $|h(\alpha)| \leq |\alpha\beta| 2^{2^{O(k^4)}}$. Thus, the satisfiability problem for k-ordered quadratic equations is in* NP.

5 Regular-Reversed Quadratic Equations

Proposition 19 settles the complexity of the satisfiability problem for a large class of quadratic equations, in which the variables occur according to some global order on both sides of the equation. In the present section, we present our main result, concerning a class of equations which are, in a sense, opposite to k-ordered equations, namely the regular-reversed equations, in which the orders in which variables occur on each side are reversed. More formally, we define regular-reversed equations as follows:

▶ **Definition 20.** *Let $\pi : (X \cup \Sigma)^* \to X^*$ be the morphism such that $\pi(x) = x$ if $x \in X$ and $\pi(x) = \varepsilon$ otherwise. A quadratic word equation $\alpha = \beta$ is* regular-reversed *if $\pi(\alpha) = \pi(\beta)^R$.*

Since the equations described in Proposition 14 are in fact regular-reversed, we cannot hope to achieve polynomial bounds on the length of minimal solutions by relying on Lemma 11; the proposition tells us that in fact, exponentially long chain-square-free solutions exist. Nevertheless, we are able to exploit Lemma 11 to obtain exponential upper bounds, which due to Theorem 1, is sufficient to show that the satisfiability problem is NP for this class.

▶ **Theorem 21.** *Let $\alpha = \beta$ be a regular-reversed equation with n distinct variables. Let $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a chain-square-free solution to $\alpha = \beta$. Then $|h(\alpha)| \leq 2^{3n}|\alpha\beta|$. Consequently, the satisfiability of regular-reversed equations is in* NP.

The proof in full is rather technical and too long to include in the main exposition. However, in order to give a flavour of the reasoning, we include the following crucial lemma. Essentially, the lemma tells us that if a regular-reversed equation has a chain-square-free solution fulfilling certain combinatorial conditions, then we may erase a variable from both the equation and solution, to obtain a shorter equation over fewer variables and a new, shorter solution which is at least half as long as the original, but which remains chain-square-free. This reduction in the number of variables allows us to apply a straightforward induction, which, due to the linear reduction in the size of the solution in each step, yields exponential bounds on the length of chain-square-free (and hence also on minimal) solutions. The majority of the remaining effort in the proof is focused on reaching a point at which these combinatorial conditions are met. It is a straightforward observation that if the conditions are met, then the solution is not minimal, highlighting the usefulness of using chain-square-free solutions in place of minimal ones.

▶ **Lemma 22.** *Let $\alpha = \beta$ be a regular-reversed equation given by $u_0x_1u_1x_2 \ldots x_nu_n = v_nx_nv_{n-1}x_{n-1} \ldots v_1x_1v_0$ where $u_i, v_i \in \Sigma^*$ for $0 \leq i \leq n$, and $x_i \in X$ for $1 \leq i \leq n$. Suppose there exists a chain-square-free solution $h : (\text{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ such that for some $1 \leq p < q \leq n$:*

(1) *$h(u_0x_1 \ldots u_{p-1}x_p) = h(v_nx_n \ldots v_p)$ and $h(u_px_{p+1} \ldots u_{n-1}x_nu_n) = h(x_pv_{p-1} \ldots x_1v_0)$, and*

(2) *$|h(u_px_{p+1} \ldots x_q)| < |h(x_p)|$ and $|h(x_qv_{q-1} \ldots v_p)| \leq |h(x_p)|$, and*

(3) *$|h(u_px_{p+1} \ldots x_qu_q)| + |h(v_qx_qv_{q-1} \ldots v_p)| \geq |h(x_p)|$.*

Then the substitution $g : ((\text{var}(\alpha\beta)\backslash\{x_p\}) \cup \Sigma)^* \to \Sigma^*$ given by $g(x) = h(x)$ for all $x \in \text{var}(\alpha\beta)\backslash\{x_p\}$ is a chain-square-free solution to the equation $\alpha' = \beta'$ obtained by erasing x_p from α and β, and moreover $|g(\alpha')| \geq \frac{|h(\alpha)|}{2}$.

Proof. To see that g is a solution to $\alpha' = \beta'$, let $w = h(u_0x_1 \ldots u_{p-1})$ and let $w' = h(v_{p-1} \ldots x_1v_0)$. Note that $g(v_nx_n \ldots v_p) = wh(x_p)$ and $g(u_px_{p+1} \ldots u_{n-1}x_nu_n) = h(x_p)w'$. It follows that $g(\alpha') = wh(x_p)w' = g(\beta')$, while $h(\alpha) = wh(x_p)h(x_p)w' = h(\beta)$. It follows immediately that $|h(x_p)| \leq \frac{|h(\alpha)|}{2}$, and thus that $|g(\alpha')| = |h(\alpha)| - |h(x_p)| \geq \frac{|h(\alpha)|}{2}$. It remains to see that g is chain-square-free, for which we must understand the chains representation of g compared to that of h. In particular, we shall show (via Claims 23 and 24) that the chain(-words) of g are obtained by simply removing positions belonging to the variable x_p from the chain(-words) of h. Before we prove this statement, we observe some basic facts about the sibling and neighbour relations for each solution. Note firstly that $\mathcal{RP}^g_{\alpha'=\beta'} = \mathcal{RP}^h_{\alpha=\beta}\backslash\{(x_p, i, d) \mid i \in \{1, 2\}, d \in [1..|h(x_p)|]\}$. Moreover, note that two positions $r_1, r_2 \in \mathcal{RP}^g_{\alpha'=\beta'}$ are siblings in the chains representation of g if and only if they are siblings in the chains representation of h. Finally, for each $r \in \mathcal{RP}^g_{\alpha'=\beta'}$, we have $f^g_{\alpha'=\beta'}(r) = f^h_{\alpha=\beta}(r) - \mu$ where $\mu = 0$ if $f^h_{\alpha=\beta}(r) \leq |wh(x_p)|$ and $\mu = |h(x_p)|$ otherwise. Consequently, for any position $r \in \mathcal{RP}^g_{\alpha'=\beta'}$ such that $f^h_{\alpha=\beta}(r) \leq |w|$ or $f^h_{\alpha=\beta}(r) > |h(\alpha)| - |w'|$, the neighbour of r in the chains representation of g is the same as the neighbour of r in the chains representation of h.

The following two claims describe the successor relation (and hence the chains) for g w.r.t. $\alpha' = \beta'$ in terms of the successor relation for h w.r.t. $\alpha = \beta$.

▷ Claim 23. Let r, r' be a subchain of some chain in the chains representation of h w.r.t. $\alpha = \beta$. Suppose that neither r nor r' belongs to x_p. Then r, r' is also a subchain of some chain in the chains representation of g w.r.t. $\alpha' = \beta'$.

Proof (Claim 23). Let $\overline{r} = r$ if r belongs to a terminal symbol, and let $\overline{r}$ be the sibling of r otherwise. Note that since r does not belong to x_p, the definition of $\overline{r}$ is the same for both chains-representations. Moreover, note that since no variables occur only once in the equations $\alpha = \beta$ and $\alpha' = \beta'$, the successor of r is the neighbour of $\overline{r}$ in both chains representations. If $f^h_{\alpha=\beta}(\overline{r}) \leq |w|$ or $f^h_{\alpha=\beta}(\overline{r}) > |h(\alpha)| - |w'|$, then as previously mentioned, the neighbour of $\overline{r}$, and hence successor of r, is the same in both chains representations and the claim follows. If instead $|w| < f^h_{\alpha=\beta}(\overline{r}) \leq |h(\alpha)| - |w'|$, then either $\overline{r}$ belongs to x_p, or the neighbour of $\overline{r}$ in the chains representation of h w.r.t $\alpha = \beta$ belongs to x_p. If $\overline{r}$ belongs to x_p, then r belongs to x_p which is a contradiction. Similarly, since the neighbour of $\overline{r}$ is the successor of r, namely r', if it belongs to x_p, then we again get a contradiction, so neither case is possible under the assumptions of the claim. ◁

▷ Claim 24. Let r, r', r'' be a subchain of some chain in the chains representation of h w.r.t. $\alpha = \beta$ such that r' belongs to x_p. Then neither r nor r' belongs to x_p, and moreover, r, r'' is also a subchain of some chain in the chains representation of g w.r.t. $\alpha' = \beta'$.

Proof (Claim 24). The fact that r and r'' do not belong to x_p follows immediately from the fact that h is chain-square-free along with Remark 9. As before, let $\overline{r} = r$ if r belongs to a terminal symbol and let $\overline{r}$ be the sibling of r otherwise. Again, note that since r does not belong to x_p, the definition of $\overline{r}$ is the same for both chains-representations. Let $\overline{r'}$ be the sibling of r' in the chains representation of h w.r.t $\alpha = \beta$ (recall that x_p occurs twice in $\alpha = \beta$, so the sibling exists). Then r' is the neighbour of $\overline{r}$ and r'' is the neighbour of $\overline{r'}$ in the chains representation of h w.r.t $\alpha = \beta$. We shall proceed by distinguishing two cases based on r'. Suppose firstly that $r' = (x_p, 2, d)$ for some $d \in [1..|h(x_p)|]$ (so r' belongs to the occurrence of x_p on the RHS). Then $|wh(x_p)| < f^h_{\alpha=\beta}(r') = f^h_{\alpha=\beta}(\overline{r}) \leq |h(\alpha)| - |w'|$ and $f^h_{\alpha=\beta}(\overline{r'}) + |h(x_p)| = f^h_{\alpha=\beta}(r')$. Hence $f^g_{\alpha'=\beta'}(\overline{r}) = f^h_{\alpha=\beta}(\overline{r}) - |h(x_p)|$, and

$$f^h_{\alpha=\beta}(r'') = f^h_{\alpha=\beta}(\overline{r'}) = f^h_{\alpha=\beta}(r') - |h(x_p)| = f^h_{\alpha=\beta}(\overline{r}) - |h(x_p)| \leq |h(\alpha)| - |w'| - |h(x_p)|.$$

Since $|h(\alpha)| - |w'| - |h(x_p)| = |wh(x_p)|$, this implies that $f^g_{\alpha'=\beta'}(r'') = f^h_{\alpha=\beta}(r'')$. Thus:

$$f^g_{\alpha'=\beta'}(\overline{r}) = f^h_{\alpha=\beta}(\overline{r}) - |h(x_p)| = f^h_{\alpha=\beta}(r') - |h(x_p)| = f^h_{\alpha=\beta}(\overline{r'}) = f^h_{\alpha=\beta}(r'') = f^g_{\alpha'=\beta'}(r'').$$

Symmetrically, if instead $r' = (x_p, 1, d)$ for some $d \in [1..|h(x_p)|]$ (so r' belongs to the occurrence of x_p on the LHS), then in the same manner, we can derive:

$$f^g_{\alpha'=\beta'}(\overline{r}) = f^h_{\alpha=\beta}(\overline{r}) = f^h_{\alpha=\beta}(r') = f^h_{\alpha=\beta}(\overline{r'}) - |h(x_p)| = f^h_{\alpha=\beta}(r'') - |h(x_p)| = f^g_{\alpha'=\beta'}(r'').$$

In both cases we get that $f^g_{\alpha'=\beta'}(\overline{r}) = f^g_{\alpha'=\beta'}(r'')$, so r'' and $\overline{r}$ are neighbours in the chains representation of g w.r.t $\alpha' = \beta'$ and hence r'' is the successor of r in the chains representation of g w.r.t $\alpha' = \beta'$ and the statement of the claim follows. ◁

It follows easily from Claims 23 and 24 that each chain in the chains representation of g is obtained by removing all positions belonging to x_p from some chain in the chains-representation of h. It remains to check that the act of removing the positions belonging to x_p does not introduce any squares in the resulting chain-words.

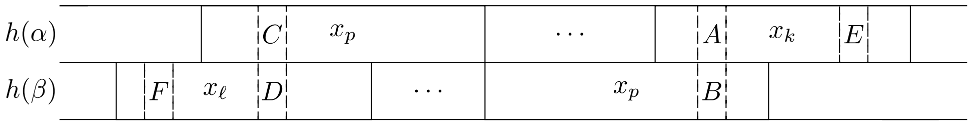

Figure 2 By Condition (1), the two occurrences of x_p are adjacent. Consequently, if (x_k, i', d_1) and (x_p, i, d_2) are neighbours, and (x_ℓ, i, d_3) and (x_p, i', d_2) are neighbours, then it cannot be the case that (x_k, i', d_4) and (x_ℓ, i, d_5) are neighbours. The case $i = 2$ is shown. The case $i = 1$ is symmetric and may be obtained by swapping the locations of x_k and x_ℓ. The positions (x_k, i', d_1), (x_p, i, d_2), (x_p, i', d_2), (x_ℓ, i, d_3), (x_k, i', d_4) and (x_ℓ, i, d_5) are marked as A, B, C, D, E and F respectively.

Suppose for contradiction that a new square is introduced. That is, there exists a subchain C in the chains representation of h for which the induced chain-word is not a square, but for which the chain-word induced by the new subchain $\hat{C}$ obtained by removing all positions belonging to x_p from C is a square. There are two ways in which this may happen. The first is that the original subchain C has the form $C', (x_p, i, d_1), C''$ where C' and C'' are subchains of length at least one and are similar (thus inducing a square in the chain-word once the central (x_p, i, d) is removed). Note that in this case there might be further positions belonging to x_p in C' and C'', but the act of removing them will not alter their similarity so we do not need to keep track of them explicitly.

By Condition (3), every position belonging to x_p is either the successor, or predecessor of a position which is either terminal or belongs to a variable x_j with $p < j \leq q$. Since the subchains C' and C'' do not contain terminal positions (this would contradict the assumption that they are similar), they must both either start with a position belonging to some x_j with $p < j \leq q$, or both end with a position to belonging to some x_j with $p < j \leq q$. It follows from Condition (2) that the neighbour of any position belonging to x_j must belong to x_p. Thus the successor and predecessor of a position belonging to x_j must also belong to x_p. However, this implies that the original subchain C occurs directly before, or after a position (x_p, i, d_2) (by Remark 9, it will have the same index i as the position (x_p, i, d_1)). However, this results in a subchain of the form $(x_p, i, d_2), C', (x_p, i, d_1), C''$ or of the form $C', (x_p, i, d_1), C'', (x_p, i, d_2)$, which in either case induces chain-word containing a square, a contradiction to the fact that h is chain-square-free.

The second possibility is that C may be divided into two further subchains C' and C'' (so $C = C', C''$), where C' and C'' are not similar, but become similar once positions belonging to x_p are removed. This implies the existence of (not necessarily consecutive) subchains $(x_k, i, d_1), (x_p, i, d_2), (x_\ell, i, d_3)$ and $(x_k, i, d_4), (x_\ell, i, d_5)$ of C (one occurring in C' and the other in C''). Clearly, since h is chain-square-free, we must have that $k \neq \ell \neq p$. Let $i' = i + 1 \mod 2$. Then (x_k, i', d_1) and (x_p, i, d_2) are neighbours, (x_p, i', d_2) and (x_ℓ, i, d_3) are neighbours and (x_k, i', d_4) and (x_ℓ, i, d_5) are neighbours. However this is only possible if the two occurrences of x_p in the solution $h(x)$ are not adjacent (see Fig. 2), or more precisely, it implies that $f^h_{\alpha=\beta}((x_p, 1, |h(x_p)|)) < f^h_{\alpha=\beta}((x_p, 2, 1)) - 1$. This clearly contradicts Condition (1). In all cases we get a contradiction, so g must be chain-square-free as claimed. ◀

6 Avoiding Squares and other Patterns

Lemma 11 invites an obvious question: do there exist long solutions for which the chain-words (which must then also be long) are square-free? This question is particularly interesting as a negative answer for the appropriate meaning of long would be sufficient to show that the satisfiability of quadratic word equations is in NP. For regular equations with two variables, the chain-words will be over an alphabet of size two, meaning we get an immediate answer: a quick exhaustive search reveals that any word over two letters of length at least 4 contains a square. Thus, for a regular equation with two variables, the chain-words of a minimal solution can have length at most 3, and thus, any minimal solution must have length at most $3n$ where n is the number of terminal symbols in the equation.

Unfortunately, a famous result of Thue [26] reveals that there exist infinitely long words over three letters which do not contain squares, meaning such a simple proof will not work for equations with more variables. This does not mean, however, that Lemma 11 is of no use in more the more general case. It is easily established that not all words may occur as chain-words of some solution to an equation (note, e.g. that the number of possible different factors of length two in a chain-word is $2n-1$ where n is the number of variables, while in general there are n^2 such factors). The next theorem gives a characterisation of when a word w is a chain-word of some solution to a regular word equation.

▶ **Theorem 25.** *Let w be a word and let Γ be the alphabet of letters occurring in w. There exists a regular word equation E with solution h such that w is a chain-word in $\Delta_{h,E}$ if and only if, there exist letters $\$, \# \notin \Gamma$ and linear orders $<_1$, $<_2$ on the sets $\Gamma \cup \{\#\}$ and $\Gamma \cup \{\$\}$ respectively such that for every $u \in \Gamma^*$ and $A, B, C, D \in \Gamma \cup \{\$, \#\}$ with $A \neq B$ and $C \neq D$, if AuC and BuD are both factors of $\#w\$$, then either that $A <_2 B$ and $C <_1 D$ or that $B <_2 A$ and $D <_1 C$.*

▶ **Corollary 26.** *Let E be a regular word equation and let h be a solution to E. Let $w \in \Delta_{h,E}$. Let A, B, C, D be letters from w such that $A \neq B$ and $C \neq D$ Then for any word u, at least one of AuC, BuC, AuD, BuD is not a factor of w.*

While this characterisation appears not to reveal immediately whether "long" square-free chain-words exist, we can make use of it to derive further conditions which may be more useful. As an example. we show in the following lemma that chain-words avoiding squares must also avoid other types of pattern (which are not necessarily avoidable in general).

▶ **Lemma 27.** *Let E be a quadratic word equation given by $\alpha = \beta$ and let $h : (\mathrm{var}(\alpha\beta) \cup \Sigma)^* \to \Sigma^*$ be a solution to E. If there exists a chain-word $w \in \Delta_{E,h}$ which contains factor of the form $x_1x_2x_3x_4x_2x_1x_3$ such that x_3 is not a prefix of x_1 or x_2, then there exists a (possibly distinct from w) chain-word $w' \in \Delta_{E,h}$ which contains a square.*

Unlike squares, it follows from the famous Zimin algorithm [18] that all words which are "long enough" will encounter a factor of the form $x_1x_2x_3x_4x_2x_1x_3$. In other words, $x_1x_2x_3x_4x_2x_1x_3$ is an *unavoidable pattern*. Unfortunately, however, this does not guarantee the additional condition that x_3 is not a prefix of x_1 or x_2, so Lemma 27 does not immediately provide a bound on the length of chain-square-free solutions. Nevertheless, a quick exhaustive search again reveals that any word of length at least 8 over three letters contains such a factor of the form $x_1x_2x_3x_4x_2x_1x_3$ or its reversal where x_1, x_2 and x_3 are all distinct letters (and so satisfying the prefix/suffix conditions given in Lemma 27). Moreover, it is not difficult to adapt the proof of Lemma 27 to produce other "forbidden factors" which must be avoided in chain-words of chain-square-free solutions. Thus we expect it to be a promising direction to try to obtain upper bounds on the lengths of (subclasses of) quadratic word equations through the lens of unavoidable patterns: do there exist combinations of patterns which are unavoidable, at least when considered over some over-approximation of all possible chain-words which ultimately guarantee the existence of squares in the chain-words of the same solutions?

References

1 P. A. Abdulla, M. F. Atig, Y. Chen, L. Holík, A. Rezine, P. Rümmer, and J. Stenman. Norn: An SMT Solver for String Constraints. In *Proc. CAV 2015*, volume 9206 of *LNCS*, pages 462–469, 2015.

2 A. Aydin, L. Bang, and T. Bultan. Automata-Based Model Counting for String Constraints. In *Proc. CAV 2015*, volume 9206 of *LNCS*, pages 255–272, 2015.

3 C. Barrett, C. L. Conway, M. Deters, L. Hadarean, D. Jovanović, T. King, A. Reynolds, and C. Tinelli. CVC4. In *Proc. CAV 2011*, volume 6806 of *LNCS*, pages 171–177, 2011.

4 M. Berzish, V. Ganesh, and Y. Zheng. Z3str3: A string solver with theory-aware heuristics. In *Proc. FMCAD 2017*, pages 55–59. IEEE, 2017.

5 J. D. Day, F. Manea, and D. Nowotka. The Hardness of Solving Simple Word Equations. In *Proc. MFCS 2017*, volume 83 of *LIPIcs*, pages 18:1–18:14, 2017.

6 R. Dąbrowski and W. Plandowski. Solving two-variable word equations. In *Proc. 31th International Colloquium on Automata, Languages and Programming, ICALP 2004*, volume 3142 of *Lecture Notes in Computer Science*, pages 408–419, 2004.

7 V. Diekert, A. Jez, and M. Kufleitner. Solutions of Word Equations Over Partially Commutative Structures. In *Proc. 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 127:1–127:14, 2016.

8 V. Diekert and J. M. Robson. On Quadratic Word Equations. In *Proc. 16th Annual Symposium on Theoretical Aspects of Computer Science, STACS 1999*, volume 1563 of *Lecture Notes in Computer Science*, pages 217–226, 1999.

9 A. Ehrenfeucht and G. Rozenberg. Finding a Homomorphism Between Two Words is NP-Complete. *Information Processing Letters*, 9:86–88, 1979.

10 D. D. Freydenberger. A Logic for Document Spanners. In *Proc. 20th International Conference on Database Theory, ICDT 2017*, Leibniz International Proceedings in Informatics (LIPIcs), 2017. To appear.

11 D. D. Freydenberger and M. Holldack. Document Spanners: From Expressive Power to Decision Problems. In *Proc. 19th International Conference on Database Theory, ICDT 2016*, volume 48 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:17, 2016.

12 J. Jaffar. Minimal and Complete Word Unification. *Journal of the ACM*, 37(1):47–85, 1990.

13 A. Jeż. Recompression: a simple and powerful technique for word equations. In *Proc. STACS 2013*, volume 20 of *LIPIcs*, pages 233–244, 2013.

14 A. Jez. Context Unification is in PSPACE. In *Proc. 41st International Colloquium on Automata, Languages, and Programming, ICALP 2014*, volume 8573 of *Lecture Notes in Computer Science*, pages 244–255. Springer, 2014.

15 A. Jeż. Word Equations in Nondeterministic Linear Space. In *Proc. ICALP 2017*, volume 80 of *LIPIcs*, pages 95:1–95:13, 2017.

16 J. Karhumäki, F. Mignosi, and W. Plandowski. The expressibility of languages and relations by word equations. *Journal of the ACM (JACM)*, 47(3):483–505, 2000.

17 A. Kiezun, V. Ganesh, P. J. Guo, P. Hooimeijer, and M. D. Ernst. HAMPI: a solver for string constraints. In *Proc. ISSTA 2009*, pages 105–116. ACM, 2009.

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

19 R. C. Lyndon. Equations in free groups. *Transactions of the American Mathematical Society*, 96:445–457, 1960.

20 R. C. Lyndon and P. E. Schupp. *Combinatorial Group Theory.* Springer, 1977.

21 G. S. Makanin. The problem of solvability of equations in a free semigroup. *Sbornik: Mathematics*, 32(2):129–198, 1977.

22 W. Plandowski. Satisfiability of Word Equations with Constants is in NEXPTIME. In *Proc. STOC 1999*, pages 721–725. ACM, 1999.

23 W. Plandowski. Satisfiability of word equations with constants is in PSPACE. In *Proc. FOCS 1999*, pages 495–500. IEEE, 1999.

24 W. Plandowski and W. Rytter. Application of Lempel-Ziv Encodings to the Solution of Words Equations. In *Proc. ICALP 1998*, volume 1443 of *LNCS*, pages 731–742, 1998.

25 K. U. Schulz. Word Unification and Transformation of Generalized Equations. *Journal of Automated Reasoning*, 11:149–184, 1995.

26 A. Thue. Über unendliche Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I Mat. Nat. Kl.*, 7, 1906.

27 M. Trinh, D. Chu, and J. Jaffar. Progressive Reasoning over Recursively-Defined Strings. In *Proc. CAV 2016*, volume 9779 of *LNCS*, pages 218–240, 2016.

28 F. Yu, M. Alkhalaf, and T. Bultan. STRANGER: An Automata-based String Analysis Tool for PHP. In *Proc. TACAS 2010*, volume 6015 of *LNCS*, 2010.