Universität Stuttgart

Institut für Formale Methoden der Informatik
Abteilung Theoretische Informatik

Universitätsstraße 38
70569 Stuttgart

Bachelorarbeit

Quadratic Equations in Free Groups and Free Monoids

Silas Natterer

Studiengang: B.Sc. Informatik

1. Prüfer: Prof. Dr. Volker Diekert

2. Prüfer: -

Betreuer: Prof. Dr. Volker Diekert

begonnen am: 17.08.2023

beendet am: 17.02.2024

Quadratic Equations in Free Groups and Free Monoids

Silas Natterer

February 12, 2024

Abstract

We consider word equations over a free monoid or group where every variable occurs at most twice, also called quadratic equations. First, we recount some previously established facts about quadratic equations. We then present the classic solution algorithm for quadratic equations over free monoids based on Nielsen transformations. Next, we prove a theorem that is in some ways analogous to the Pumping Lemma for regular languages: If a quadratic equation permits infinitely many solutions, this already implies the existence of solutions with arbitrarily high exponent of periodicity; this statement is proven both for free monoids and free groups. Finally, we present an algorithm which computes the exponent of periodicity for general equations over free monoids.

1 Introduction

Equations over free monoids and free groups have been studied extensively. The most prominent results have been achieved by Makanin, who in 1977, showed that the satisfiability problem for word equations over free monoids is decidable [8]. He later generalized this to free groups [9] and Schulz [12] proved that the result remains true when adding regular constraints. Makanin's algorithm is, in its essence, the construction of a finite search graph. The algorithm is often considered to possess one of the most complicated termination proofs in computer science literature. The best known upper bound for its complexity is Expspace [3]. A modern description of the algorithm and its termination proof can, for example, be found in Lothaire [6]. In 1999, using an entirely different method, Plandowski [11] was able to construct a Pspace algorithm. This brings the upper bound closer to the lower bound of NP-hardness, which follows directly from the NP-hardness of integer linear programming.

This paper will focus on equations where every variable occurs at most twice, also called quadratic equations. Even though the satisfiablity problem is known to still be NP-hard for this subclass [2], solution algorithms are considerably less complicated. A simple Pspace algorithm, orignally due to Matiyasevich [10], is based on Nielsen transformations and presented in section 3. Diekert and Robson [2] were even able to construct an algorithm that, given the length of each variable in binary, runs in linear time. However, the exact complexity for the satisfiablity problem of quadratic equations over free monoids is still unknown. Quadratic equations over free groups are intimately connected to the classification of closed surfaces and have been a subject in combinatorial group theory for a long time. In contrast to free monoids, the satisfiability problem of quadratic equations over the free group is known to be NP-complete [4]. In fact, the length of a minimal solution is at most polynomial in the length of the equation [7].

The main goal of this paper is to show, that the set of solutions to a quadratic word equation satisfies a property somewhat similar to the Pumping Lemma for regular languages: If it contains infinitely many solutions, this necessarily implies the existence of solutions with arbitrarily high exponent of periodicity. This was first demonstrated by Bastien Laboureix but remained unpublished [5]. Section 3 establishes the property for free monoids and section 4 deals with the analogous statement in free groups.

Finally, section 5 provides an algorithm for computing the exponent of periodicity, and we show that this is not much harder than solving the respective equation. The results in this section are originally due to Volker Diekert.

2 Notation

Let F_Σ denote the free group and $\Sigma^* \subseteq F_\Sigma$ the free monoid over a given alphabet Σ; $1 \in \Sigma^*$ is used to denote the empty word and $\Sigma^+ = \Sigma^* \setminus \{1\}$. For a word $w \in F_\Sigma$ let $|w|_M \in \mathbb{N}$ denote the number of occurrences of letters from $M \subseteq \Sigma$ in the reduced word corresponding to w and let $|w| = |w|_\Sigma$ be its length.

The exponent of periodicity $\exp(w) \in \mathbb{N}$ of $w \in \Sigma^*$ is defined by

$$\exp(w) = \max\{n \in \mathbb{N} | \exists u, v \in \Sigma^*, p \in \Sigma^+ : w = up^n v\}$$

In order to define $\exp(w)$ for $w \in F_\Sigma$, consider w as a freely reduced word in $(\Sigma \cup \Sigma^{-1})^*$, where Σ^{-1} is the alphabet of inverses to elements in Σ.

Let M be monoid and $F \in \{F_\Sigma, \Sigma^*\}$ depending on context. Given $x \in \Sigma$ and $w \in M$ let $\phi = (x \mapsto w)$ denote the homomorphism $\phi : F \to M$ that fulfills $\phi|_{\Sigma \setminus \{x\}} = \mathrm{id}|_{\Sigma \setminus \{x\}}$ and $\phi(x) = w$; it is uniquely defined due to the fundamental property of free monoids and groups.

For the remainder of the paper, let Σ be a nonempty alphabet of constants and Ω a nonempty set of variables.

3 Quadratic Equations in Free Monoids

We start by defining the concept of an equation over the free monoid and its solution:

Definition 3.1. An equation $L = R$ over the free monoid Σ^* is a pair of words $(L, R) \in (\Sigma \cup \Omega)^* \times (\Sigma \cup \Omega)^*$, such that L and R do not share any common prefix or suffix. It is called trivial, if $L, R \in \Sigma^*$. It is called quadratic, if $|L|_{\mathbf{x}} + |R|_{\mathbf{x}} \leq 2$ for each $\mathbf{x} \in \Omega$. Let $|L{=}R| = |L| + |R|$ denote the length of $L = R$ and $\operatorname{var}(L{=}R) \subseteq \Omega$ the set of variables occurring in $L = R$.

Definition 3.2. Let $L = R$ be an equation over Σ^*. A homomorphism

$$\sigma : (\Sigma \cup \Omega)^* \to \Sigma^*$$

with $\sigma|_\Sigma = \mathrm{id}_\Sigma$ is called a solution of $L = R$, if $\sigma(L) = \sigma(R)$. The length of σ is defined as $|\sigma| = \sum_{\mathbf{x} \in \Omega} |\sigma(\mathbf{x})|$.

Remark. One may also allow equations where L and R share a common prefix or suffix; they are then identified with the corresponding reduced equation. Equations which differ by swapping L and R are also identified. This is well defined, since identified equations obviously share the same set of solutions.

Also note that if $\operatorname{var}(L{=}R) \subset \Omega$ and σ is a solution, then the values $\sigma|_{\Omega \setminus \operatorname{var}(L=R)}$ of variables absent in $L = R$ can be changed arbitrarily to obtain new solutions.

Example 3.1. Consider the equation $au = bv$ with $a, b \in \Sigma$ and $u, v \in (\Sigma \cup \Omega)^*$. This equation has a solution, if and only if the equation $u = v$ has a solution and $a = b$. Indeed, if σ is a solution to $u = v$ and $a = b$, then also $\sigma(au) = a\sigma(u) = b\sigma(v) = \sigma(bv)$. Conversely, σ being a solution to $au = bv$ implies $a\sigma(u) = \sigma(av) = \sigma(bu) = b\sigma(v)$ and hence $\sigma(u) = \sigma(v)$ and $a = b$. In particular, if $L = R$ is a solvable equation and not equivalent to the trivial equation $1 = 1$, then we can iteratively reduce any prefixes until at least one side begins with a variable; the same holds for suffixes.

We want to solve equations by factoring them into Nielsen transformations. It is, however, sufficient to only consider a subset of all possible Nielsen transformations, so called related transformations:

Definition 3.3. Let $L = R$ be an equation over Σ^*. An endomorphism

$$\tau : (\Sigma \cup \Omega)^* \to (\Sigma \cup \Omega)^*$$

is called a transformation related to $L = R$, if

- $\tau = (\mathbf{x} \mapsto 1)$ and $L = \mathbf{x}w$ for some variable $\mathbf{x} \in \Omega$ and $w \in (\Sigma \cup \Omega)^*$
- $\tau = (\mathbf{x} \mapsto \alpha\mathbf{x})$ with $L = \mathbf{x}u$ and $R = \alpha v$ for some variable $\mathbf{x} \in \Omega$, some $\alpha \in \Sigma \cup \Omega \setminus \{\mathbf{x}\}$ and $u, v \in (\Sigma \cup \Omega)^*$

Remark. One may additionally allow permutations $\tau \in \operatorname{Sym}\Omega$ of the variables or substitutions where the variable occurs at the end of either L or R; they arise when considering the reversed equation $L^{\mathrm{R}} = R^{\mathrm{R}}$. However, the two transformations presented above are sufficient to establish lemma 3.1.

Lemma 3.1. *Let $L = R$ be a nontrivial equation over Σ^* and σ be a corresponding solution. Then there exists a transformation τ related to $L = R$ such that $\sigma = \sigma' \circ \tau$ and $|\operatorname{var}(\tau(L){=}\tau(R))| < |\operatorname{var}(L{=}R)|$ or $|\operatorname{var}(\tau(L){=}\tau(R))| \leq |\operatorname{var}(L{=}R)|$ and $|\sigma'| < |\sigma|$.*

Proof. Since $L = R$ is nontrivial and solvable, one may assume $L = \mathbf{x}u$ for some $\mathbf{x} \in \operatorname{var}(L{=}R)$ and $u \in (\Sigma \cup \Omega)^*$ (see example 3.1). If $\sigma(\mathbf{x}) = 1$ choose $\tau = (\mathbf{x} \mapsto 1)$ and $\sigma' = \sigma$, and observe that $\sigma = \sigma' \circ \tau$ and

$$|\operatorname{var}(\tau(L){=}\tau(R))| \leq |\operatorname{var}(L{=}R) \setminus \{\mathbf{x}\}| < |\operatorname{var}(L{=}R)|$$

Next $\sigma(\mathbf{x}) \neq 1$ implies $\sigma(R) = \sigma(L) = \sigma(\mathbf{x})\sigma(u) \neq 1$; hence $R \neq 1$ and $R = \alpha v$ for some $\alpha \in \Sigma \cup \Omega \setminus \{\mathbf{x}\}$ and $v \in (\Sigma \cup \Omega)^*$. If $\sigma(\alpha) = 1$, then $\alpha \in \Omega$ must be a variable and we reduce to the previous case. Therefore $\sigma(\alpha) \neq 1$ as well, and by potentially swapping sides we get

$$0 < |\sigma(\alpha)| \leq |\sigma(\mathbf{x})|$$

Since σ is a solution, we have $\sigma(\mathbf{x})\sigma(u) = \sigma(\alpha)\sigma(v)$, and so $\sigma(\alpha)$ must be a prefix of $\sigma(\mathbf{x})$. We get $\sigma(\mathbf{x}) = \sigma(\alpha)w$ for some $w \in \Sigma^*$ and choose

$$\tau = (\mathbf{x} \mapsto \alpha\mathbf{x}) \qquad \sigma' = (\mathbf{x} \mapsto w) \circ \sigma|_{\Omega\setminus\{\mathbf{x}\}}$$

as the desired factorization with $|\operatorname{var}(\tau(L){=}\tau(R))| \leq |\operatorname{var}(L{=}R)|$ and

$$|\sigma'| = \sum_{\mathbf{y}\in\Omega} |\sigma'(\mathbf{y})| = |w| - |\sigma(\mathbf{x})| + \sum_{\mathbf{y}\in\Omega} |\sigma(\mathbf{y})| = |\sigma| - |\sigma(\alpha)| < |\sigma|$$

□

As long as the given equation is nontrivial, lemma 3.1 allows us to always find a related transformation that can be factored out of the given solution. Thus, repeated application of the lemma and induction over $(|\operatorname{var}(L{=}R)|, |\sigma|)$ with lexicographic order immediately yields

Corollary 3.2. *Let $L = R$ be an equation over Σ^* and σ be a corresponding solution. Then there exists $\sigma_0 : (\Sigma \cup \Omega)^* \to \Sigma^*$ and transformations $(\tau_i)_{1\leq i\leq n}$ such that*

$$\sigma = \sigma_0 \circ \psi = \sigma_0 \circ \tau_n \circ \cdots \circ \tau_1$$

where each τ_i is related to the respective previous equation

$$(\tau_{i-1} \circ \cdots \circ \tau_1)(L) = (\tau_{i-1} \circ \cdots \circ \tau_1)(R)$$

and $\psi(L) = \psi(R)$ is the trivial equation.

This allows us to factor solutions of any word equation in terms of related transformations. The final homomorphism σ_0 is necessary, since there might be variables which do not occur in $L = R$, or which get cancelled somewhere along the transformation process as seen in example 3.5.

Now this factorization implies, that every solution can be found on a graph, where the vertices are equations $L = R$, and the outgoing edges are the transformations τ related to $L = R$. Solving word equations can then be reduced to finding paths on this graph.

Definition 3.4. Let $L = R$ be an equation over Σ^*. Let $V_0 = \{L{=}R\}$ and

$$V_n = \{\tau(U){=}\tau(V) : (U{=}V) \in V_{n-1} \text{ and } \tau \text{ is related to } U = V\}$$
$$E_n = \{U{=}V \xrightarrow{\tau} \tau(U){=}\tau(V) : (U{=}V) \in V_{n-1} \text{ and } \tau \text{ is related to } U = V\}$$

The directed graph $G_{L=R} = (\bigcup_{n=0}^{\infty} V_n, \bigcup_{n=1}^{\infty} E_n)$ is then called the solution graph of the equation $L = R$.

Theorem 3.3. *Let $L = R$ be an equation over Σ^* such that its solution graph $G_{L=R}$ is finite. Then the satisfiability problem for $L = R$ is decidable.*

Proof. Starting at the node $L = R$ perform a (nondeterministic) depth first search for the trivial equation $1 = 1$. This search terminates since $G_{L=R}$ is finite. If it finds a path

$$L = R \xrightarrow{\tau_1} \tau_1(L) = \tau_1(R) \xrightarrow{\tau_2} \cdots \xrightarrow{\tau_n} 1 = 1$$

then we use $\sigma_0 = (\mathbf{x} \mapsto 1)_{\mathbf{x}\in\Omega}$ and set $\sigma = \sigma_0 \circ \tau_n \circ \cdots \circ \tau_1$. This is a solution since the equation $\sigma(L) = \sigma(R)$ is equivalent to $(\tau_n \circ \cdots \circ \tau_1)(L) = (\tau_n \circ \cdots \circ \tau_1)(R)$, which is equivalent to the trivial equation $1 = 1$.

Conversely, if $L = R$ permits a solution σ, then it possesses a factorization

$$\sigma = \sigma_0 \circ \psi = \sigma_0 \circ \tau_n \circ \cdots \circ \tau_1$$

according to corollary 3.2. Since all the τ_i are related transformations and $\psi(L) = \psi(R)$ is the trivial equation $1 = 1$, this again corresponds to a path from $L = R$ to $1 = 1$ in $G_{L=R}$, and hence a solution $\sigma' = (\mathbf{x} \mapsto 1)_{\mathbf{x}\in\Omega} \circ \psi$ that will be discovered by the algorithm. □

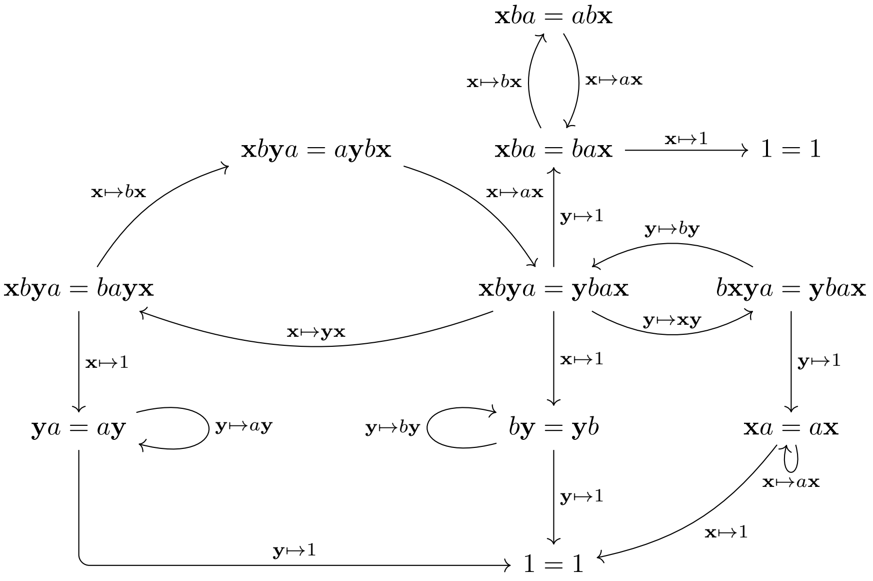
This algorithm is not restricted to quadratic equations (see example 3.6) but it fails in general (see example 3.7). For quadratic equations, the finiteness of $G_{L=R}$ follows because related transformations do not increase their length. This is obvious for transformations of the form $\mathbf{x} \mapsto 1$ for $\mathbf{x} \in \Omega$. To see that it also holds for transformations of the form $\tau = (\mathbf{x} \mapsto \alpha\mathbf{x})$, consider $L = \mathbf{x}u$ and $R = \alpha v$ for some $\alpha \in \Sigma \cup \Omega$ and $u, v \in (\Sigma \cup \Omega)^*$. Then $\tau(L) = \alpha\mathbf{x}\tau(u)$ and $\tau(R) = \alpha\tau(v)$ and $|\tau(u)| + |\tau(v)| \leq |u| + |v| + 1$, since $\mathbf{x}$ can occur at most one more time in u or v. Since $\tau(L)$ and $\tau(R)$ share the common prefix α we get

$$|\tau(L){=}\tau(R)| = |\mathbf{x}\tau(u){=}\tau(v)| = 1 + |\tau(u)| + |\tau(v)| \leq |u| + |v| + 2 = |L{=}R|$$

The nondeterministic algorithm presented in theorem 3.3 thus has linear space complexity when applied to quadratic equations, which proves

Corollary 3.4. *There is a* Pspace *algorithm for the satisfiability problem of quadratic equations over the free monoid Σ^*.*

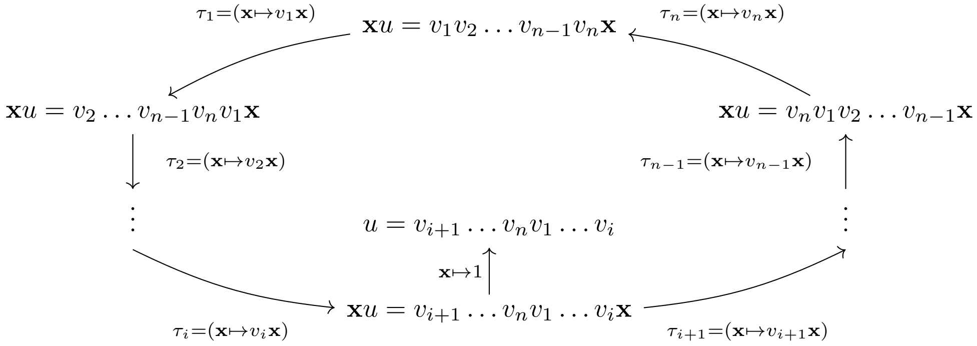
Example 3.2. Let $\Omega = \{\mathbf{x}, \mathbf{y}\}$ and consider the equation $\mathbf{x}b\mathbf{y}a = \mathbf{y}ba\mathbf{x}$ for some $a, b \in \Sigma$. Its solution graph looks like this:



$\mathbf{x}ba = ab\mathbf{x}$
$\mathbf{x}\mapsto b\mathbf{x}$
$\mathbf{x}\mapsto a\mathbf{x}$
$\mathbf{x}b\mathbf{y}a = a\mathbf{y}b\mathbf{x}$
$\mathbf{x}ba = ba\mathbf{x}$
$\mathbf{x}\mapsto 1$
$1 = 1$
$\mathbf{x}\mapsto b\mathbf{x}$
$\mathbf{x}\mapsto a\mathbf{x}$
$\mathbf{y}\mapsto 1$
$\mathbf{y}\mapsto b\mathbf{y}$
$\mathbf{x}b\mathbf{y}a = ba\mathbf{y}\mathbf{x}$
$\mathbf{x}b\mathbf{y}a = \mathbf{y}ba\mathbf{x}$
$b\mathbf{x}\mathbf{y}a = \mathbf{y}ba\mathbf{x}$
$\mathbf{x}\mapsto\mathbf{y}\mathbf{x}$
$\mathbf{y}\mapsto\mathbf{x}\mathbf{y}$
$\mathbf{x}\mapsto 1$
$\mathbf{x}\mapsto 1$
$\mathbf{y}\mapsto 1$
$\mathbf{y}a = a\mathbf{y}$
$\mathbf{y}\mapsto a\mathbf{y}$
$\mathbf{y}\mapsto b\mathbf{y}$
$b\mathbf{y} = \mathbf{y}b$
$\mathbf{x}a = a\mathbf{x}$
$\mathbf{x}\mapsto a\mathbf{x}$
$\mathbf{y}\mapsto 1$
$\mathbf{x}\mapsto 1$
$\mathbf{y}\mapsto 1$
$1 = 1$

Example 3.3. Let $\Omega = \{\mathbf{x}\}$. Consider a nontrivial quadratic equation of the form $\mathbf{x}u = v\mathbf{x}$ with $u, v \in \Sigma^*$. It is a fundamental theorem due to Lyndon and Schützenberger, that this equation has a solution, if and only if $u = sr$ and $v = rs$ for some $r, s \in \Sigma^*$; then $\mathbf{x} \mapsto (rs)^k r$ is a solution for any $k \in \mathbb{N}$.

Instead of the conventional proof [1], we utilize the solution graph $G_{\mathbf{x}u=v\mathbf{x}}$ to reestablish this fact:



$\tau_1{=}(\mathbf{x}\mapsto v_1\mathbf{x})$
$\mathbf{x}u = v_1 v_2 \dots v_{n-1} v_n \mathbf{x}$
$\tau_n{=}(\mathbf{x}\mapsto v_n\mathbf{x})$
$\mathbf{x}u = v_2 \dots v_{n-1} v_n v_1 \mathbf{x}$
$\mathbf{x}u = v_n v_1 v_2 \dots v_{n-1} \mathbf{x}$
$\tau_2{=}(\mathbf{x}\mapsto v_2\mathbf{x})$
$\tau_{n-1}{=}(\mathbf{x}\mapsto v_{n-1}\mathbf{x})$
$\vdots$
$u = v_{i+1} \dots v_n v_1 \dots v_i$
$\vdots$
$\mathbf{x}\mapsto 1$
$\mathbf{x}u = v_{i+1} \dots v_n v_1 \dots v_i \mathbf{x}$
$\tau_i{=}(\mathbf{x}\mapsto v_i\mathbf{x})$
$\tau_{i+1}{=}(\mathbf{x}\mapsto v_{i+1}\mathbf{x})$

Now each path σ in $G_{\mathbf{x}u=v\mathbf{x}}$ can be written in the form

$$\sigma = (\mathbf{x} \mapsto 1) \circ \tau_i \circ \dots \circ \tau_1 \circ (\tau_n \circ \dots \circ \tau_1)^k$$

for some $0 \leq i < n$ and $k \in \mathbb{N}$. It is a solution, if and only if it ends at the vertex $1 = 1$, which is the case, if and only if $u = v_{i+1} \dots v_n v_1 \dots v_i$. We choose $r = v_1 \dots v_i$ and $s = v_{i+1} \dots v_n$ and since $G_{\mathbf{x}u=v\mathbf{x}}$ contains all solutions, the theorem follows.

Example 3.4. Let $\Omega = \{\mathbf{x}\}$ and consider a nontrivial quadratic equation of the form $\mathbf{x}u\mathbf{x} = v$ for some $u, v \in \Sigma^*$ and its solution graph $G_{\mathbf{x}u\mathbf{x}=v}$:

$$\begin{tikzcd}
\mathbf{x}u\mathbf{x} = v_1v_2\dots v_{n-1}v_n \arrow[r, "\mathbf{x}\mapsto v_1\mathbf{x}"] & \mathbf{x}uv_1\mathbf{x} = v_2\dots v_{n-1}v_n \arrow[d, "\mathbf{x}\mapsto v_2\mathbf{x}"] \\
 & \vdots \arrow[d, "\mathbf{x}\mapsto v_i\mathbf{x}"] \\
uv_1\dots v_i = v_{i+1}\dots v_n & \mathbf{x}uv_1\dots v_i\mathbf{x} = v_{i+1}\dots v_n \arrow[l, "\mathbf{x}\mapsto 1"'] \arrow[d, "\mathbf{x}\mapsto v_{i+1}\mathbf{x}"] \\
 & \vdots \arrow[d, "\mathbf{x}\mapsto v_{n-1}\mathbf{x}"] \\
\mathbf{x}uv_1v_2\dots v_{n-1}v_n\mathbf{x} = 1 & \mathbf{x}uv_1v_2\dots v_{n-1}\mathbf{x} = v_n \arrow[l, "\mathbf{x}\mapsto v_n\mathbf{x}"']
\end{tikzcd}$$

Therefore, the equation has the unique solution $\mathbf{x} \mapsto v_1 \dots v_i$ for $i \in \mathbb{N}$, if and only if $uv_1 \dots v_i = v_{i+1} \dots v_n$.

We also observe that if $\mathbf{x} \in \Omega$ is a variable which appears twice on the same side, then repeated application of transformations $\mathbf{x} \mapsto \alpha\mathbf{x}$ with $\alpha \in \Sigma \cup \Omega$ always ends in a vertex of the form $L = 1$, at which point $\mathbf{x} \mapsto 1$ is the only available related transformation.

Example 3.5. Let $\Omega = \{\mathbf{x}, \mathbf{y}\}$. We consider the quadratic equation $\mathbf{xy} = \mathbf{yx}$. We will investigate its solutions using the graph $G_{\mathbf{xy}=\mathbf{yx}}$:

$$\begin{tikzcd}
1 = 1 & \mathbf{xy} = \mathbf{yx} \arrow[l, "\mathbf{x}\mapsto 1"'] \arrow[r, "\mathbf{y}\mapsto 1"] \arrow[loop above, "\mathbf{x}\mapsto\mathbf{yx}"] \arrow[loop below, "\mathbf{y}\mapsto\mathbf{xy}"] & 1 = 1
\end{tikzcd}$$

The graph tells us, that every solution has the form

$$\sigma = \sigma_0 \circ \epsilon \circ \phi \text{ where } \epsilon \in \{\mathbf{x} \mapsto 1, \mathbf{y} \mapsto 1\} \text{ and } \phi \in \{\mathbf{x} \mapsto \mathbf{yx}, \mathbf{y} \mapsto \mathbf{xy}\}^*$$

Now obviously $\phi(\mathbf{x}), \phi(\mathbf{y}) \in \{\mathbf{x}, \mathbf{y}\}^*$, and thus $\epsilon(\phi(\mathbf{x})) = \mathbf{z}^n$ and $\epsilon(\phi(\mathbf{y})) = \mathbf{z}^m$ with $\mathbf{z} \in \{\mathbf{x}, \mathbf{y}\}$ being the remaining variable after the elimination ϵ. Finally, some $\sigma_0(\mathbf{z}) = w$ with $w \in \Sigma^*$ must be chosen to retrieve the full solution σ, which then satisfies $\sigma(\mathbf{x}) = w^n$ and $\sigma(\mathbf{y}) = w^m$. In particular, this shows that if two words commute, they must both be powers of some common third word.

Example 3.6. Let $L = w$ be an equation with $w \in \Sigma^*$. Even if it is not quadratic, its solution graph $G_{L=w}$ is still finite: The length of the right side $|w|$ or the number of variables reduces in each transformation step, implying a linear bound on the length of paths in $G_{L=w}$.

Example 3.7. Consider the equation $\mathbf{x}^2 = u\mathbf{x}v$ for some $u, v \in \Sigma^*$. This is the minimal example for an equation, which does not have a finite solution graph:

$$\mathbf{x}^2 = u\mathbf{x}v \xrightarrow{\mathbf{x} \mapsto u\mathbf{x}} \mathbf{x}u\mathbf{x} = u\mathbf{x}v \xrightarrow{\mathbf{x} \mapsto u\mathbf{x}} \cdots \xrightarrow{\mathbf{x} \mapsto u\mathbf{x}} \mathbf{x}u^n\mathbf{x} = u\mathbf{x}v \xrightarrow{\mathbf{x} \mapsto u\mathbf{x}} \cdots$$

This is true despite the fact that $\mathbf{x}^2 = u\mathbf{x}v$ can have at most one solution.

The exponent of periodicity is a useful tool when studying word equations and plays a crucial role in the termination proof of Makanin's algorithm. We start with the definition of the exponent of periodicity of an equation, before investigating its finiteness in the quadratic case.

Definition 3.5. Let $L = R$ be an equation over Σ^*. The exponent of periodicity of $L = R$ is defined as

$$\exp(L{=}R) = \sup\{\exp(\sigma) : \sigma \text{ solves } L = R\} \in \mathbb{N} \cup \{\infty\}$$

where $\exp(\sigma) = \max\{\exp(\mathbf{x}) : \mathbf{x} \in \Omega\} \in \mathbb{N}$.

Now an infinite exponent of periodicity obviously implies the existence of an infinite number of solutions. We show, that the reverse holds for quadratic equations:

Theorem 3.5. *A quadratic equation $L = R$ over Σ^* permits infinitely many solutions, if and only if $\exp(L{=}R) = \infty$.*

Proof. Consider the solution graph $G_{L=R}$ of $L = R$. It is finite since $L = R$ is quadratic. By corollary 3.2, every solution can be factored through a path on $G_{L=R}$. Since $L = R$ has infinitely many solutions and $G_{L=R}$ is finite, there must either exist a cycle in $G_{L=R}$, or there are infinitely many solutions corresponding to the same path. The second case implies that the corresponding factorizations

$$\sigma = \sigma_0 \circ \psi = \sigma_0 \circ \tau_k \circ \cdots \circ \tau_1$$

differ in σ_0 for at least one $\mathbf{x} \in \Omega$ and are equal otherwise. For every $n \in \mathbb{N}$ and some $a \in \Sigma$, we hence get the solution $\sigma_n = (\mathbf{x} \mapsto a^n) \circ \psi$, which implies

$$\exp(L{=}R) \geq \lim_{n\to\infty} \exp(\sigma_n) \geq \lim_{n\to\infty} \exp(\sigma_n(\mathbf{x})) \geq \lim_{n\to\infty} \exp(a^n) = \infty$$

Next we assume the existence of a cycle in $G_{L=R}$; then the situation is illustrated by the following diagram:

$$\begin{tikzcd}
 & & & L_i = R_i \arrow[dr, bend left, "\tau_i"] & & & \\
 & & \vdots \arrow[ur, bend left, "\tau_{i-1}"] & & \vdots \arrow[dl, "\tau_t"] & & \\
L = R \arrow[r, "\tau_1"] & \cdots \arrow[rr, "\tau_{s-1}"] & & L_s = R_s \arrow[ul, "\tau_s"] \arrow[rr, "\tau_{t+1}"] & & \cdots \arrow[r, "\tau_k"] & 1 = 1
\end{tikzcd}$$

We will examine, which transformations can occur in the cycle $\phi = \tau_t \circ \cdots \circ \tau_s$, hence let $s \leq i < t$.

First, notice that any transformation $\tau_i = (\mathbf{x} \mapsto 1)$ with $\mathbf{x} \in \operatorname{var}(L{=}R)$ obviously reduces the length of the equation. The same holds for any related transformation $\tau_i = (\mathbf{x} \mapsto \alpha\mathbf{x})$ where $\mathbf{x} \in \Omega$ occurs only once in the equation. To see this, consider $L_i = \mathbf{x}u$ and $R_i = \alpha v$ for some $\alpha \in \Sigma \cup \Omega$ and $u, v \in (\Sigma \cup \Omega)^*$. Then

$$\tau_i(L_i) = \alpha\mathbf{x}\tau_i(u) = \alpha\mathbf{x}u \qquad \tau_i(R_i) = \alpha\tau_i(v) = \alpha v$$

since $\mathbf{x}$ does not occur in u nor v. Now $\tau_i(L_i)$ and $\tau_i(R_i)$ share the common prefix α and thus

$$|L_{i+1}{=}R_{i+1}| = |\mathbf{x}u{=}v| = |u| + |v| + 1 < |u| + |v| + 2 = |L_i{=}R_i|$$

Since applying further related transformations can never increase the length of a quadratic equation again, these two types cannot occur in our cycle ϕ.

Next, we partition the set of doubly occurring variables into two sets, based on whether or not they occur on the same side of the equation $L_i = R_i$:

$$\begin{aligned} \Omega_i^+ &= \{\mathbf{x} \in \Omega : |L_i|_{\mathbf{x}} = 1 \text{ and } |R_i|_{\mathbf{x}} = 1\} \\ \Omega_i^- &= \{\mathbf{x} \in \Omega : |L_i|_{\mathbf{x}} = 2 \ \text{ or } \ |R_i|_{\mathbf{x}} = 2\} \end{aligned}$$

Due to our previous considerations, we know that each transformation τ_i is of the form $\tau_i = (\mathbf{x} \mapsto \alpha\mathbf{x})$ with $L_i = \mathbf{x}u$ and $R_i = \alpha v$ for some $\mathbf{x} \in \Omega_i^+ \cup \Omega_i^-$, some $\alpha \in \Sigma \cup \Omega$ and $u, v \in (\Omega \cup \Sigma)^*$.

If there is a τ_i with $\mathbf{x} \in \Omega_i^+$ or $\alpha \in \Omega_i^+$, then by possibly swapping sides we can assume $\mathbf{x} \in \Omega_i^+$ and $R_i = v_1\mathbf{x}v_2$ for some $v_1 = \alpha v_1', v_2 \in (\Sigma \cup \Omega)^*$. This then allows us to insert a new cycle

$$\begin{tikzcd} \cdots \arrow[r, "\tau_{i-1}"] & \mathbf{x}u = v_1\mathbf{x}v_2 \arrow[loop above, "\delta=(\mathbf{x}\mapsto v_1\mathbf{x})"] \arrow[r, "\tau_{i+1}"] & \cdots \end{tikzcd}$$

starting and ending at $L_i = R_i$. For each $n \in \mathbb{N}$, we thus get the solution

$$\sigma_n = \sigma_0 \circ \tau_k \circ \cdots \circ \tau_i \circ \delta^n \circ \tau_{i-1} \circ \cdots \circ \tau_1$$

Now $\delta^n(\mathbf{x}) = v_1^n\mathbf{x}$ since $\mathbf{x}$ does not occur in v_1 and hence $\delta(v_1) = v_1$. Since all the transformations are related, we additionally get $(\tau_{i-1} \circ \cdots \circ \tau_1)(\mathbf{x}) = w\mathbf{x}$ for some $w \in (\Sigma \cup \Omega)^*$. With $\varphi = \sigma_0 \circ \tau_k \circ \cdots \circ \tau_i$ we then obtain

$$\sigma_n(\mathbf{x}) = \varphi(\delta^n(w\mathbf{x})) = \varphi(\delta^n(w)v_1^n\mathbf{x})) = \varphi(\delta^n(w))\varphi(v_1)^n\varphi(\mathbf{x})$$

and since $\exp(\sigma_n(\mathbf{x})) = \exp(\varphi(\delta^n(w))\varphi(v_1)^n\varphi(\mathbf{x})) \geq n$ we get

$$\exp(L{=}R) \geq \lim_{n\to\infty} \exp(\sigma_n) \geq \lim_{n\to\infty} \exp(\sigma_n(\mathbf{x})) \geq \lim_{n\to\infty} n = \infty$$

Now assume by contradiction, that there is no τ_i with $\mathbf{x} \in \Omega_i^+$ or $\alpha \in \Omega_i^+$. Then $\mathbf{x} \in \Omega_i^-$ and $\alpha \in \Sigma \cup \Omega_i^-$ for every τ_i. We can therefore write $L_i = \mathbf{x}u_1\mathbf{x}u_2$ for some $u_1, u_2 \in (\Sigma\cup\Omega)^*$. Now consider how $N_i = |\Omega_i^-|$ changes, after applying τ_i to $L_i = R_i$. Since $\mathbf{x}$ does not occur in u_1, u_2 or v, we get

$$\tau_i(L_i) = \alpha\mathbf{x}\tau(u_1)\alpha\mathbf{x}\tau(u_2) = \alpha\mathbf{x}u_1\alpha\mathbf{x}u_2 \qquad \tau_i(R_i) = \alpha\tau_i(v) = \alpha v$$

The preceding α cancels once again and we obtain $L_{i+1} = \mathbf{x}u_1\alpha\mathbf{x}u_2$ and $R_{i+1} = v$; hence only α switches sides and we get

$$N_{i+1} = \begin{cases} N_i & \text{if } \alpha \in \Sigma \\ N_i - 1 & \text{if } \alpha \in \Omega_i^- \end{cases}$$

Observe that N_i never increases as long as $\mathbf{x} \in \Omega_i^-$ and $\alpha \in \Sigma \cup \Omega_i^-$. Since ϕ is a cycle, we conclude that it can never decrease either, and hence $\mathbf{x} \in \Omega_i^-$ and $\alpha \in \Sigma$ for all τ_i; consulting example 3.4 shows that this is also impossible. This contradiction concludes the proof. □

The following examples are meant to clear up any misunderstandings about the proof of theorem 3.5.

Example 3.8. One may mistakenly assume, that there cannot exist a cycle $\phi = \tau_1 \cdots \circ \ldots \tau_n$ in $G_{L=R}$ that consist only of transformations of the form $\tau_i = (\mathbf{x} \mapsto \alpha\mathbf{x})$ with $\mathbf{x} \in \Omega_i^-$. This is true for $|\Omega| < 7$. However, for

$$\Omega = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}, \mathbf{g}\}$$

we find the cycle ϕ of length 21 given by

$$\begin{tikzcd}
\mathbf{abca} = \mathbf{efcgdfdbge} \arrow[d, "\mathbf{a}\mapsto\mathbf{efcga}"] & & \mathbf{acga} = \mathbf{ebgfdbdcfe} \arrow[ll, "\varrho"'] \\
\mathbf{abcefcga} = \mathbf{dfdbge} \arrow[d, "\mathbf{d}\mapsto\mathbf{abd}"] & \begin{array}{c} \mathbf{b}\mapsto\mathbf{f} \\ \mathbf{f}\mapsto\mathbf{g} \\ \mathbf{g}\mapsto\mathbf{c} \\ \mathbf{c}\mapsto\mathbf{b} \end{array} & \mathbf{fdbdcfacga} = \mathbf{ebge} \arrow[u, "\mathbf{e}\mapsto\mathbf{fdbdcfe}"] \\
\mathbf{cefcga} = \mathbf{dfabdbge} \arrow[dr, bend right, "\mathbf{c}\mapsto\mathbf{dfac}"'] & & \mathbf{fdfacga} = \mathbf{bdcebge} \arrow[u, "\mathbf{f}\mapsto\mathbf{bdcf}"] \\
& \mathbf{cefdfacga} = \mathbf{bdbge} \arrow[ur, bend right, "\mathbf{b}\mapsto\mathbf{ceb}"'] &
\end{tikzcd}$$

If one wants to avoid edges $\varrho \in \operatorname{Sym}\Omega$, the cycle ϕ^4 of length 80 can be chosen alternatively. Since $\operatorname{ord}\varrho = 4$, this cycle can be represented using only related transformations defined in the strict sense.

In fact, for $|L{=}R| = 7$ there already exist strictly different cycles, meaning that they are not the same up to a permutation of variables or choosing a different starting point.

Now the crucial observation is, that even though all $\mathbf{x}$ are in Ω_i^- not all α are: For instance, the third substitution after ϱ is $\mathbf{a} \mapsto \mathbf{ca}$ where $\mathbf{a} \in \Omega_i^-$ but $\mathbf{c} \in \Omega_i^+$. However, it is unclear whether or not ϕ itself still induces solutions with infinite exponent of periodicity.

Example 3.9. An involution on some monoid M is a function $M \to M$ that satisfies $\overline{uv} = \bar{v}\bar{u}$ for all $u, v \in M$. One can generalize the notion of an equation over Σ^* by considering $L, R \in (\Sigma \cup \Omega \cup \overline{\Omega})^*$. A solution σ must then additionally satisfy $\sigma(\bar{\mathbf{x}}) = \overline{\sigma(\mathbf{x})}$. The definition of related transformations and the proof of corollary 3.2 also generalize in a straightforward manner.

Now let $\Omega = \{\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}\}$ and consider the cycle ϕ of length 4 given by

$$\begin{tikzcd}
\bar{\mathbf{x}}\mathbf{x}\mathbf{y}\bar{\mathbf{z}} = \mathbf{z}\mathbf{w}\bar{\mathbf{w}}\bar{\mathbf{y}} \arrow[d, "\mathbf{z}\mapsto\bar{\mathbf{x}}\mathbf{z}"] & & \mathbf{y}\bar{\mathbf{z}}\mathbf{z}\mathbf{x} = \mathbf{w}\bar{\mathbf{w}}\bar{\mathbf{x}}\bar{\mathbf{y}} \arrow[ll, bend right, "\varrho"', "\substack{\mathbf{x}\mapsto\bar{\mathbf{y}} \\ \mathbf{y}\mapsto\mathbf{z} \\ \mathbf{z}\mapsto\bar{\mathbf{w}} \\ \mathbf{w}\mapsto\bar{\mathbf{x}}}"] \\
\mathbf{x}\mathbf{y}\bar{\mathbf{z}}\mathbf{x} = \mathbf{z}\mathbf{w}\bar{\mathbf{w}}\bar{\mathbf{y}} \arrow[rr, "\mathbf{x}\mapsto\mathbf{x}\mathbf{z}"] & & \mathbf{x}\mathbf{y}\bar{\mathbf{z}}\mathbf{z}\mathbf{x} = \mathbf{w}\bar{\mathbf{w}}\bar{\mathbf{y}} \arrow[u, "\mathbf{w}\mapsto\mathbf{x}\mathbf{w}"']
\end{tikzcd}$$

Again since ord $\varrho = 8$, one can alternatively use the cycle ϕ^8 of length 24. Looking closely, we see that at no point in ϕ a variable occurs on both sides and with the same involution state. Therefore, we cannot insert the desired simple cycle of infinite exponent into ϕ. This demonstrates, that the proof of theorem 3.5 cannot be generalized to include involutions.

However, it is unclear if ϕ itself induces solutions with infinite exponent of periodicity. If it does not, it may be possible to construct a counterexample to theorem 3.5 for equations with involutions by inserting constants at the correct positions in the equation.

Example 3.10. Let $\Omega = \{\mathbf{x}, \mathbf{y}\}$ and $w_0, \ldots, w_4 \in (\Omega \cup \Sigma)^*$. When considering quadratic systems of equations one finds the cycle

$$\begin{tikzcd}
\begin{matrix} w_0\mathbf{x}w_1 = \mathbf{y}w_2 \\ \mathbf{y}w_3 = \mathbf{x}w_4 \end{matrix} \arrow[r, bend right, "\mathbf{y}\mapsto w_0\mathbf{y}"'] & \begin{matrix} \mathbf{x}w_1 = \mathbf{y}w_2 \\ w_0\mathbf{y}w_3 = \mathbf{x}w_4 \end{matrix} \arrow[l, bend right, "\mathbf{x}\mapsto w_0\mathbf{x}"']
\end{tikzcd}$$

Again this shows that the proof of theorem 3.5 does not work for systems of equations. However, in this case it is obvious that the above cycle induces an infinite exponent of periodicity in the solutions of $\mathbf{x}$ and $\mathbf{y}$.

4 Quadratic Equations in Free Groups

We again start by defining the concept of equations and their solutions, this time over free groups:

Definition 4.1. An equation E over the free group F_Σ is a cyclically reduced word $E \in F_{\Sigma\cup\Omega}$. It is called trivial, if $E \in F_\Sigma$. It is called quadratic, if $|E|_{\mathbf{x}} \leq 2$ for each $\mathbf{x} \in \Omega$. Let $\operatorname{var} E \subseteq \Omega$ denote the set of variables occurring in E.

Definition 4.2. Let E be an equation over F_Σ. A homomorphism

$$\sigma : F_{\Sigma\cup\Omega} \to F_\Sigma$$

with $\sigma|_\Sigma = \mathrm{id}_\Sigma$ is called a solution of E, if $\sigma(E) = 1$.

Remark. Let E be an equation over F_Σ and $w \in F_{\Sigma\cup\Omega}$. Now every solution σ of E is also a solution of E^{-1} and of wEw^{-1}. Therefore, equations which can be transformed into each other by inversion or cyclic permutation will be identified; one may thus also allow equations even though they are not cyclically reduced. This identification coincides with the identification for equations $L = R$ over Σ^* by setting $E = LR^{-1}$ to be the corresponding equation over F_Σ.

Again note, that if $\operatorname{var} E \subset \Omega$, then the values $\sigma|_{\Omega\setminus\operatorname{var} E}$ of a given solution σ can be changed arbitrarily since they do not occur in E.

Example 4.1. Let $E = \mathbf{x}w$ for some $w \in F_{\Sigma\cup\Omega\setminus\{\mathbf{x}\}}$ and $\mathbf{x} \in \Omega$. If $w \in F_\Sigma$ and $\Omega = \{\mathbf{x}\}$ then, by the properties of the free group, we get the unique solution $\mathbf{x} \mapsto w^{-1}$. Otherwise, we can construct arbitrary solutions as follows: For every $\mathbf{y} \in \Omega \setminus \{\mathbf{x}\}$ choose some $w_{\mathbf{y}} \in F_\Sigma$. Now set $\sigma(\mathbf{y}) = w_{\mathbf{y}}$ and $\sigma(\mathbf{x}) = \sigma(w)^{-1}$. This is possible, since $w \in F_{\Sigma\cup\Omega\setminus\{\mathbf{x}\}}$ and a solution, since $\sigma(E) = \sigma(\mathbf{x})\sigma(w) = 1$.

Example 4.2. Let $\Omega = \{\mathbf{x}\}$ and $E = \mathbf{x}u\mathbf{x}^{-1}v$ for some $u, v \in F_\Sigma$. Since $\mathbf{x}u\mathbf{x}^{-1}v = 1$, if and only if $\mathbf{x}u\mathbf{x}^{-1} = v^{-1}$, this equation has a solution, if and only if u and v^{-1} are conjugate. If that is the case, then for every $h \in F_\Sigma$ with $v^{-1} = huh^{-1}$ and every $k \in \mathbb{Z}$ the homomorphism $\mathbf{x} \mapsto hu^k$ is a solution.

Example 4.3. Let $\Omega = \{\mathbf{x}\}$ and $E = \mathbf{x}u\mathbf{x}v$ for some $u, v \in F_\Sigma$. This equation has a solution, if and only if $v^{-1}u$ is a square:

$$\mathbf{x}u\mathbf{x}v = 1 \xrightarrow{\mathbf{x}\mapsto\mathbf{x}u^{-1}} \mathbf{x}\mathbf{x}u^{-1}v = 1 \Longleftrightarrow \mathbf{x}^2 = v^{-1}u$$

If that is the case, then $v^{-1}u = s^2$ for exactly one $s \in F_\Sigma$, and $\mathbf{x} \mapsto su^{-1}$ is the unique solution to E. In particular, the number of solutions of E is always finite.

Example 4.4. Let $\Omega = \{\mathbf{x}, \mathbf{y}\}$ and $E = \mathbf{x}\mathbf{y}\mathbf{x}^{-1}\mathbf{y}^{-1}$. Now $E = [\mathbf{x}, \mathbf{y}]$ is the commutator of $\mathbf{x}$ and $\mathbf{y}$; hence σ is a solution, if and only if $\sigma(\mathbf{x})$ and $\sigma(\mathbf{y})$ commute. Since F_Σ is free, this is the case if and only if both of them are powers of some common $w \in F_\Sigma$. One then gets solutions $\sigma = (\mathbf{x} \mapsto w^n) \circ (\mathbf{y} \mapsto w^m)$ for all choices of $n, m \in \mathbb{Z}$.

We now turn to the exponent of periodicity again. The additional structure of free groups lets us prove an even stronger statement than for free monoids.

Definition 4.3. Let E be an equation over F_Σ and σ be a corresponding solution. The exponent of periodicity of σ is defined as

$$\exp(E) = \sup\{\exp(\sigma) : \sigma \text{ solves } E\} \in \mathbb{N} \cup \{\infty\}$$

where $\exp(\sigma) = \max\{\exp(\mathbf{x}) : \mathbf{x} \in \Omega\} \in \mathbb{N}$.

Theorem 4.1. *Let E be a solvable quadratic equation over F_Σ. Then the following three statements are equivalent:*

(i) $\Omega = \{\boldsymbol{x}\}$ *and* $E \in \mathfrak{F} = \{\boldsymbol{x}w, \boldsymbol{x}u\boldsymbol{x}v : u, v, w \in F_\Sigma\}$*.*

(ii) E *permits only finitely many solutions.*

(iii) $\exp(E) < \infty$*.*

Proof. For the implication *(i)* ⇒ *(ii)* see example 4.1 and 4.3. The implication *(ii)* ⇒ *(iii)* is trivial.

To prove *(iii)* ⇒ *(i)* consider the contraposition: Let E be a quadratic equation and σ be a corresponding solution. Since $E \notin \mathfrak{F}$, several cases must be distinguished. For every case solutions $(\sigma_n)_{n\in\mathbb{N}}$ are constructed, such that $\exp(E) \geq \lim_{n\to\infty} \exp(\sigma_n) = \infty$.

We might first assume that $\operatorname{var} E = \Omega$, since $\operatorname{var} E \subset \Omega$ implies that $\sigma_n|_{\Omega\setminus\operatorname{var} E}$ can be chosen arbitrarily. Moreover, if some variable $\mathbf{x} \in \Omega$ occurs only once in E, then $\Omega \neq \{\mathbf{x}\}$ since $E \notin \mathfrak{F}$; again we can choose $\sigma_n|_{\Omega\setminus\{\mathbf{x}\}}$ as we wish (see example 4.1). In both cases, we can for example choose $\sigma_n(\mathbf{y}) = a^n$ for the respective variables, where $a \in \Sigma$. This obviously satisfies

$$\lim_{n\to\infty} \exp(\sigma_n) \geq \lim_{n\to\infty} \exp(\sigma_n(\mathbf{y})) = \lim_{n\to\infty} \exp(a^n) = \infty$$

For the remainder of the proof, we can now assume that all variables occur twice in E and that $\operatorname{var} E = \Omega$.

In case some $\mathbf{x} \in \Omega$ occurs in E twice with different exponents, we get $E = \mathbf{x}u\mathbf{x}^{-1}v$ for some $u, v \in F_{\Sigma\cup\Omega}$. If $\sigma(u) = 1$, then also

$$\sigma(v) = \sigma(\mathbf{x})\sigma(u)\sigma(\mathbf{x})^{-1}\sigma(v) = \sigma(E) = 1$$

Similarly, $\sigma(v) = 1$ implies $\sigma(u) = 1$. Since $\mathbf{x} \notin \operatorname{var} \sigma|_{\Omega\setminus\{\mathbf{x}\}}(E)$ its solution can be chosen arbitrarily; take for example $\sigma_n = (\mathbf{x} \mapsto a^n) \circ \sigma|_{\Omega\setminus\{\mathbf{x}\}}$ for some $a \in \Sigma$ again.

Now proceed with $\sigma(u) \neq 1$ and set $\sigma_n = \sigma \circ (\mathbf{x} \mapsto \mathbf{x}u^n)$. This is a solution since

$$E = \mathbf{x}u\mathbf{x}^{-1}v \xrightarrow{\mathbf{x}\mapsto\mathbf{x}u^n} \mathbf{x}u\mathbf{x}^{-1}v \xrightarrow{\sigma} \sigma(\mathbf{x}u\mathbf{x}^{-1}v) = \sigma(E) = 1$$

Now some cancellation may occur in $\sigma_n(\mathbf{x}) = \sigma(\mathbf{x})\sigma(u)^n$ reducing its exponent of periodicity. Then $\sigma(\mathbf{x}) = vp^{-1}\sigma(u)^{-k}$ for some $k \in \mathbb{N}, v \in F_\Sigma$ and some factorization $\sigma(u) = pq$ with $p, q \in F_\Sigma$; hence $\sigma_n(\mathbf{x}) = vq\sigma(u)^{n-k-1}$ and again

$$\lim_{n\to\infty} \exp(\sigma_n) \geq \lim_{n\to\infty} \exp(\sigma_n(\mathbf{x})) \geq \lim_{n\to\infty} (n - k - 1) = \infty$$

Finally, if every variable occurs in E twice with the same exponent, then E must contain at least two variables since $E \notin \mathfrak{F}$. Up to equivalence of equations and permutation of variables and their inverses, there are exactly two ways an equation can be arranged in this case:

$$E = \mathbf{x}w_1\mathbf{y}w_2\mathbf{x}w_3\mathbf{y}w_4 \xrightarrow{\tau=(\mathbf{x}\mapsto\mathbf{x}\mathbf{y}^{-1}w_1^{-1})} \mathbf{x}w_2\mathbf{x}\mathbf{y}^{-1}w_1^{-1}w_3\mathbf{y}w_4 = E'$$

$$E = \mathbf{x}w_1\mathbf{x}w_2\mathbf{y}w_3\mathbf{y}w_4 \xrightarrow{\tau=(\mathbf{x}\mapsto\mathbf{x}\mathbf{y}^{-1}w_2^{-1})} \mathbf{x}\mathbf{y}^{-1}w_2^{-1}w_1\mathbf{x}w_3\mathbf{y}w_4 = E'$$

for some $w_1, \ldots .w_4 \in F_{\Sigma\cup\Omega}$ and $\mathbf{x}, \mathbf{y} \in \Omega$. Now E' is also solvable, since $\sigma' = \sigma \circ \tau^{-1}$ is a solution. In both cases, E' contains $\mathbf{y}$ with two different exponents; hence we reduce to the previous case and find solutions σ'_n of E' with $\lim_{n\to\infty} \exp(\sigma'_n(\mathbf{y})) = \infty$. At last, this yields solutions $\sigma_n = \sigma'_n \circ \tau$ of E with $\lim_{n\to\infty} \exp(\sigma_n) \geq \lim_{n\to\infty} \exp(\sigma_n(\mathbf{y})) = \lim_{n\to\infty}(\sigma'_n(\mathbf{y})) = \infty$. □

Theorem 4.1 basically classifies all quadratic equations over F_Σ based on the finiteness of their exponent of periodicity: Solvable equations in $\mathfrak{F}$ with $|\Omega| = 1$ are the only ones for which $0 < \exp(E) < \infty$. Every other equation is either unsolvable with $\exp(E) = 0$ or it has an infinite number of solutions with $\exp(E) = \infty$. In particular, this characterization implies

Corollary 4.2. *A quadratic equation E over F_Σ permits infinitely many solutions, if and only if* $\exp(E) = \infty$.

5 Computing the Exponent of Periodicity

This final section is concerned with computing the exponent of periodicity for general word equations over free monoids. At first glance this seems much harder than solving the respective equation. However, the following lemma will show that these two problems can, in fact, be reduced to each other.

Lemma 5.1. *Let $L = R$ be an equation over Σ^* and $k \in \mathbb{N}$ encoded in binary. Then the problem of deciding whether* $\exp(L{=}R) \geq k$ *can be reduced to the satisfiability problem of some word equation in polynomial time.*

Proof. We start by guessing a variable $\mathbf{x} \in \Omega$ and some constant $a \in \Sigma$. Next we consider the binary expansion

$$k = \sum_{i=0}^{l} b_i 2^i$$

of k and the following system of equations:

$$\begin{array}{cc} L = R & \mathbf{q}_1 = \mathbf{q}_0^2 \\ \mathbf{x} = \mathbf{u}\mathbf{q}_0^{b_0} \cdots \mathbf{q}_l^{b_l}\mathbf{v} & \vdots \\ \mathbf{q}_0 = a\mathbf{p} & \mathbf{q}_l = \mathbf{q}_{l-1}^2 \end{array}$$

where $\mathbf{u}, \mathbf{v}, \mathbf{p}, \mathbf{q}_0, \ldots, \mathbf{q}_l \in \hat{\Omega}$ are new variables and $\Omega \subset \hat{\Omega}$. We can solve this system using Makanin's algorithm or some other algorithm of our choice; we can even transform it into a single equation [6]. Note that the size of this system is linear in the size of the original equation and the length of the binary encoding of k. It is not hard to show that $\exp(L{=}R) \geq k$, if and only if the above system has a solution:

First, let σ be a solution of the system; then σ must also be a solution of $L = R$ by construction, and one easily proves by induction that $\sigma(\mathbf{q}_i) = \sigma(\mathbf{q}_0)^{2^i}$. We therefore find

$$\sigma(\mathbf{x}) = \sigma(\mathbf{u})\sigma(\mathbf{q}_0)^{b_0} \cdots \sigma(\mathbf{q}_l)^{b_l}\sigma(\mathbf{v}) = \sigma(\mathbf{u})\sigma(\mathbf{q}_0)^k\sigma(\mathbf{v})$$

and $\exp(L{=}R) \geq \exp(\sigma(\mathbf{x})) \geq k$ since $\sigma(\mathbf{q}_0) = a\sigma(\mathbf{p})$ is nonempty.

Conversely, if $\exp(L{=}R) \geq k$, then there must be some variable $\mathbf{x} \in \Omega$ and a solution σ such that $\exp(\sigma(\mathbf{x})) \geq k$. We can thus find a nonempty word $q = ap \in \Sigma^+$ for some $a \in \Sigma$ and words $u, v \in \Sigma^*$ such that $\sigma(\mathbf{x}) = uq^kv$. This obviously yields a solution $\hat{\sigma}$ of the above system by choosing

$$\hat{\sigma}|_\Omega = \sigma, \hat{\sigma}(\mathbf{u}) = u, \hat{\sigma}(\mathbf{v}) = v, \hat{\sigma}(\mathbf{p}) = p \text{ and } \hat{\sigma}(\mathbf{q}_i) = q^{2^i}$$

□

Remark. We can also perform a trivial reduction in the opposite direction: A word equation $L = R$ has a solution, if and only if $\exp(L{=}R) \geq 1$. This shows that the two problems are in fact equivalent, and in particular, computing the exponent of periodicity must be NP-hard too.

With this lemma in hand, we are almost ready to compute the exponent of periodicity. However, we would also like to ascertain whenever the exponent of periodicity is infinite. Here a central lemma in the termination proof of Makanin's algorithm proves useful.

Lemma 5.2. *Let $L = R$ be an equation over Σ^* with $|L{=}R| = n$ and let σ be a corresponding solution. By considering the p-stable normal form of σ we can find a linear system of equations, where each of its nonnegative diophantine solutions induces another solution of $L = R$.*

In particular, there is a function $e(n) \in 2^{\Theta(n)}$ and a minimal solution σ_0 with $\exp(\sigma_0) < e(n)$; if $\exp(\sigma) \geq e(n)$, then there must be infinitely many solutions with $\exp(L{=}R) = \infty$.

For the details of this construction and the proof of the lemma we refer to Lothaire's chapter on Makanin's algorithm [6]. We also remark that in the special case of quadratic equations, the exponent of periodicity can be further bounded by a linear function, as shown by Diekert and Robson [2]. After these preparations, we can easily prove the final theorem of this section:

Theorem 5.3. *Let $L = R$ be an equation of length $|L{=}R| = n$ over Σ^*. There is a* PSPACE *algorithm which computes* $\exp(L{=}R)$.

Proof. Perform a binary search over the range $k = 1, \ldots, e(n)$; for each k use lemma 5.1 to determine whether or not $\exp(L{=}R) \geq k$ and use the result to continue the search. If the search terminates with $k = e(n)$ then we know $\exp(L{=}R) \geq e(n)$ and hence $\exp(L{=}R) = \infty$ by lemma 5.2. Otherwise, we find $1 \leq k < e(n)$ such that $\exp(L{=}R) \geq k$ but $\exp(L{=}R) < k + 1$; hence $\exp(L{=}R) = k$.

The complexity of this binary search is at most linear, since $e(n) \in 2^{\Theta(n)}$ by lemma 5.2. The cost of each invokation of lemma 5.1 depends on the algorithm used to solve the system of word equations; using Plandowski's PSPACE algorithm [11], the entire search can be done in PSPACE as well. Also note, that any upper bound on $e(n)$ suffices and that we do not actually have to compute the exact value of $e(n)$. □

6 Conclusion

We were able to prove, that a quadratic equation over some free group or monoid has infinitely many solutions, if and only if its exponent of periodicity is infinte. Future work in this area may generalize this result in several different directions. It may be interesting to consider quadratic systems of equations or quadratic equations with involution. Alternatively, one may also consider regular constraints. Finally, the case of general equations remains open. Since their solution graph is in general infinte, it seems likely that entirely different methods need to be utilized here.

Additionally, we demonstrated how to compute the exponent of periodicity. This algorithm yields an upper bound of containment in PSPACE for the problem, which is essentially optimal unless NP-completeness for the satisfiability problem of word equations is established.

References

[1] Volker Diekert, Manfred Kufleitner, Gerhard Rosenberger, and Ulrich Hertrampf. *Discrete algebraic methods: Arithmetic, cryptography, automata and groups.* Walter de Gruyter GmbH & Co KG, 2016.

[2] Volker Diekert and John Michael Robson. Quadratic word equations. *Jewels are Forever: Contributions on Theoretical Computer Science in Honor of Arto Salomaa*, pages 314–326, 1999.

[3] Claudio Gutiérrez. Satisfiability of word equations with constants is in exponential space. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, pages 112–119. IEEE, 1998.

[4] Olga Kharlampovich, Igor G Lysënok, Alexei G Myasnikov, and Nicholas WM Touikan. Quadratic equations over free groups are np-complete. *arXiv preprint arXiv:0802.3839*, 2008.

[5] Bastien Laboureix. Rapport de stage. On word equations over free monoids and free groups (french, unpublished).

[6] M. Lothaire. *Algebraic combinatorics on words*, volume 90. Cambridge university press, 2002.

[7] Igor G Lysenok and Alexei G Myasnikov. A polynomial bound on solutions of quadratic equations in free groups. *Proceedings of the Steklov Institute of Mathematics*, 274:136–173, 2011.

[8] Gennadiy Semenovich Makanin. The problem of solvability of equations in a free semigroup. *Matematicheskii Sbornik*, 145(2):147–236, 1977.

[9] Gennadiy Semenovich Makanin. Equations in a free group. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 46(6):1199–1273, 1982.

[10] Yuri Matiyasevitch. A connection between systems of word and length equations and hilbert's tenth problem. *Seminar in Mathematics, V.A. Steklov Mathematic Institute, 8*, pages 61–67, 1970.

[11] Wojciech Plandowski. Satisfiability of word equations with constants is in pspace. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 495–500. IEEE, 1999.

[12] Klaus U Schulz. Makanin's algorithm for word equations-two improvements and a generalization. In *International Workshop on Word Equations and Related Topics*, pages 85–150. Springer, 1990.

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Datum und Unterschrift

Declaration

I hereby declare that the work presented in this thesis is entirely my own. I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted hard copies.

Date and Signature

Danksagung

An dieser Stelle möchte ich mich noch herzlichst bei meinem Betreuer und Prüfer Prof. Dr. Volker Diekert für seine Unterstützung sowie für die anregenden Gespräche und den kleinen Einblick in die mathematische Forschung bedanken.