The existential theory of equations with rational constraints in free groups is PSPACE–complete

Volker Diekert[1], Claudio Gutiérrez[2], Christian Hagenah[1]

[1] Inst. für Informatik, Universität Stuttgart,
Breitwiesenstr. 20-22, D-70565 Stuttgart
diekert@informatik.uni-stuttgart.de christian@hagenah.de
[2] Depto. de Ciencias de la Computación, Universidad de Chile,
Blanco Encalada 2120, Santiago, Chile
cgutierr@dcc.uchile.cl

ACM Classification F.2. Analysis of Algorithms and Problem Complexity, F.2.2. Computation on Discrete Structures, F.4. Mathematical Logic and Formal Languages.
Subject Descriptor Equations in Free Groups.

Abstract

It is known that the existential theory of equations in free groups is decidable. This is a famous result of Makanin. On the other hand it has been shown that the scheme of his algorithm is not primitive recursive. In this paper we present an algorithm that works in polynomial space, even in the more general setting where each variable has a rational constraint, that is, the solution has to respect a specification given by a regular word language. Our main result states that the existential theory of equations in free groups with rational constraints is PSPACE–complete. We obtain this result as a corollary of the corresponding statement about free monoids with involution.

1 Introduction

Around the 1980's a great progress was achieved on the algorithmic decidability of elementary theories of free monoids and groups. In 1977 Makanin [17] proved that the existential theory of equations in free monoids is decidable by presenting an algorithm which solves the satisfiability problem for a single word equation with constants. In 1983 he extended his result to the more complicated framework in free groups [18]. In fact, using a result by Merzlyakov [22] he also showed that the positive theory of equations in free groups is decidable [19], and Razborov was able to give a description of the whole solution set [28]. The algorithms of Makanin are very complex: For word equations the running time was first estimated by several towers of exponentials and it took more than 20 years to lower it down to the best known bound for Makanin's original algorithm, which is to date EXPSPACE [9]. For solving equations in free groups Kościelski and Pacholski [14] have shown that the scheme proposed by Makanin is not primitive recursive.
In 1999 Plandowski invented another method for solving word equations and he showed that the satisfiability problem for word equations is in PSPACE, [26]. One ingredient of his work is to use data compression to reduce the exponential space to polynomial space. The importance of data compression was first recognized by Rytter and Plandowski when applying Lempel-Ziv encodings to the minimal solution of a word equation [27]. Another important notion is the definition of an ℓ-factorization of the solution being explained below. Gutiérrez extended Plandowski's method to the case of free groups, [10]. Thus, a non-primitive recursive scheme for solving equations in free groups has been replaced by a polynomial space bounded algorithm. Hagenah and Diekert worked independently in the same direction and using some ideas of Gutiérrez they obtained a result which includes the presence of rational constraints. This appeared as extended abstract in [4] and also as a part of the PhD-thesis of Hagenah [11].
The present paper is a journal version of [4, 10]. It shows that the existential theory of equations in free groups with rational constraints is PSPACE–complete. Rational constraints mean that a possible solution has to respect a specification which is given by a regular word language. The idea to consider regular constraints for word equations goes back to Schulz [29] who also pointed out the importance of this concept, see also [6, 8]. The PSPACE–completeness for the case of word equations with regular constraints has been stated by Rytter already, as cited in [26, Thm. 1].

Our proof reduces the case of equations with rational constraints in free groups to the case equations with regular constraints in free monoids with involution, which turns out to be the central object. (Makanin uses the notion of "paired alphabet", but a main difference is that he considered "non contractible" solutions only, whereas we deal with general solutions and, in addition, we have constraints.) During our work we extend the method of [26] such that it copes with the involution and the method of [10] such that it copes with rational constraints. The first step is a reduction to the satisfiability problem of a single equation with regular constraints in a free monoid with involution. In order to avoid an exponential blow-up, we do not use a reduction as in [19], but a simpler one. In particular, we can handle negations simply by positive rational constraints. In the second step we show that the satisfiability problem of a single equation with regular constraints in a free monoid with involution is still in PSPACE. This part is rather technical and we introduce several new notions like base-change, projection, partial solution, and free interval. The careful handling of free intervals is necessary because of the constraints. In some sense this is the only additional difficulty which we will meet when dealing with constraints. After these preparations we can follow Plandowski's method. Throughout we shall use many of the deep ideas which were presented in [26], and apply them in a different setting. Hence, as we cannot use Plandowski's result as a black box, we have to go through the whole construction again. As a result our paper is (involuntarily) self-contained, up to standard knowledge in combinatorics on words and linear Diophantine equations.

2 Free Groups and their Rational Subsets

Let Σ be a finite alphabet. By $F(\Sigma)$ we denote the free group over Σ. Elements of $F(\Sigma)$ can be represented by words in $(\Sigma \cup \overline{\Sigma})^*$, where $\overline{\Sigma} = \{ \overline{a} \mid a \in \Sigma \}$. We read $\overline{a}$ as a^{-1} in $F(\Sigma)$ and we use the convention that $\overline{\overline{a}} = a$. Hence the set $\Gamma = \Sigma \cup \overline{\Sigma}$ is equipped with an involution $\overline{} : \Gamma \to \Gamma$; the involution is extended to Γ^* by $\overline{a_1 \cdots a_n} = \overline{a_n} \cdots \overline{a_1}$ for $n \geq 0$ and $a_i \in \Gamma$, $1 \leq i \leq n$. The empty word as well as the unit element in other monoids is denoted by 1. By $\psi : \Gamma^* \to F(\Sigma)$ we denote the canonical homomorphism. A word $w \in \Gamma^*$ is *freely reduced*, if it contains no factor of the form $a\overline{a}$ with $a \in \Gamma$. The reduction of a word $w \in \Gamma^*$ can be computed by using the Noetherian and confluent rewriting system $\{ a\overline{a} \to 1 \mid a \in \Gamma \}$. For $w \in \Gamma^*$ we

denote by $\widehat{w}$ the freely reduced word which denotes the same group element in $F(\Sigma)$ as w. Hence, $\psi(u) = \psi(v)$ if and only if $\widehat{u} = \widehat{v}$ in Γ^*.
The class of *rational languages* in $F(\Sigma)$ is inductively defined as follows: Every finite subset of $F(\Sigma)$ is rational. If $P_1, P_2 \subseteq F(\Sigma)$ are rational, then $P_1 \cup P_2$, $P_1 \cdot P_2$, and P_1^* are rational. Hence, $P \subseteq F(\Sigma)$ is rational if and only if $P = \psi(P')$ for some regular language $P' \subseteq \Gamma^*$.[1] In particular, we can use a non-deterministic finite automata over Γ for specifying rational group languages over $F(\Sigma)$.
The following proposition is due to M. Benois [1], see also [2, Sect. III. 2].

Proposition 1 *Let $P' \subseteq \Gamma^*$ be a regular language and $P = \psi(P') \subseteq F(\Sigma)$. Then we effectively find a regular language $\widetilde{P}' \subseteq \Gamma^*$ such that $\widetilde{P}' = \{\, \widehat{w} \in \Gamma^* \mid \psi(w) \notin P \,\}$. Hence, the complement of P is the rational group language $\psi(\widetilde{P}')$ and the family of rational group languages is an effective Boolean algebra.*

Proof. (Sketch) Using the same state set (and some additional transitions which are labeled with the empty word) we can construct (in polynomial time) a finite automaton which accepts the following language

$$P'' = \{\, v \in \Gamma^* \mid \exists u \in P' : u \xrightarrow{*} v \,\}$$

where $u \xrightarrow{*} v$ means that v is a descendant of u by the convergent rewriting system $\{\, a\overline{a} \to 1 \mid a \in \Gamma \,\}$. Then we complement P'' with respect to Γ^*; and we build the intersection with the regular set of freely reduced words. □

3 The Existential Theory

In the following Ω denotes a finite set of variables (or unknowns) and we let $^{-} : \Omega \to \Omega$ be an involution without fixed points. Clearly, if $X \in \Omega$ has an interpretation in $F(\Sigma)$, then we read $\overline{X}$ as $X^{-1} \in F(\Sigma)$.
The *existential theory of equations with rational constraints in free groups* is inductively defined as follows. Atomic formulae are either of the form $W = 1$, where $W \in (\Gamma \cup \Omega)^*$ or of the form $X \in P$, where X is in Ω and $P \subseteq F(\Sigma)$ is

[1] We follow the usual convention to call a rational subset of a free monoid *regular*. This convention is due to Kleene's Theorem stating that regular, rational, and recognizable have the same meaning in free monoids. But in free groups these notions are different and we have to be more precise.

a rational language. A propositional formula is build up by atomic formulae using negations, conjunctions and disjunctions. The existential theory refers to closed existentially quantified propositional formulae which evaluate to *true* over $F(\Sigma)$.

Theorem 2 *The following problem is* PSPACE*–complete.*
INPUT: A closed existentially quantified propositional formula with rational constraints in the free group $F(\Sigma)$ for some finite alphabet Σ.
QUESTION: Does the formula evaluate to true *over $F(\Sigma)$?*

The PSPACE–hardness follows from a result of Kozen [15], since (due to the constraints) the *empty intersection* problem of regular sets can easily be encoded in the problem above. The same argument applies to Theorems 4 and 5 below and therefore the PSPACE–hardness is not discussed further in the sequel: We have to show the inclusion in PSPACE, only.
The PSPACE algorithm for solving Theorem 2 will be described by a (highly) non-deterministic procedure. We will make sure that if the input evaluates to true, then at least one possible output is true. If it evaluates to false, then no (positive) output is possible. By standard methods (Savitch's Theorem) such a procedure can be transformed into a polynomial space bounded deterministic decision procedure, see any textbook on complexity theory, e.g. [12, 23].
We start the procedure as follows. Using the rules of DeMorgan we may assume that there are no negations at all, but the atomic formulae are now of the either form: $W = 1$, $W \neq 1$, $X \in P$, $X \notin P$ with $W \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $P \subseteq F(\Sigma)$ rational.[2]
The next step is to replace every formula $W \neq 1$ by

$$\exists X : WX = 1 \wedge X \notin \{1\},$$

where X is a fresh variable, hence we can put $\exists X$ to the front. Now we eliminate all disjunctions. More precisely, every subformula of type $A \vee B$ is non-deterministically replaced either by A or by B. At this stage the propositional formula has become a conjunction of formulae of type $W = 1$, $X \in P$, $X \notin P$ with $W \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $P \subseteq F(\Sigma)$ rational.
We may assume that $|W| \geq 3$, since if $1 \leq |W| < 3$, then we may replace $W = 1$ by $Wa\overline{a} = 1$ for some $a \in \Gamma$. For the following it is convenient to

[2]The reason that we keep $X \notin P$ instead of $X \in \widetilde{P}$ where $\widetilde{P} = F(\Sigma) \setminus P$ is that the complementation may involve an exponential blow-up of the state space; this has to be avoided.

assume that $|W| = 3$ for all subformulae $W = 1$. This is also easy to achieve. As long as there is a subformula $x_1 \cdots x_k = 1$, $x_i \in \Gamma \cup \Omega$ for $1 \leq i \leq k$ and $k \geq 4$, we replace it by the conjunction

$$\exists Y : x_1 x_2 Y = 1 \wedge \overline{Y} x_3 \cdots x_k = 1,$$

where Y is a fresh variable and $\exists Y$ is put to the front, and then proceed recursively. This finishes the first phase. The output of this phase is a system of atomic formulae of type $W = 1$, $X \in P$, $X \notin P$ with $W \in (\Gamma \cup \Omega)^3$, $X \in \Omega$, and $P \subseteq F(\Sigma)$ rational.

At this point we switch to the existential theory of equations with regular constraints in free monoids where these monoids have an involution. Recall that $X \in P$ (resp. $X \notin P$) means in fact $X \in \psi(P')$ (resp. $X \notin \psi(P')$) where $P' \subseteq \Gamma^*$ is a regular word language specified by some finite non-deterministic automaton. Using ψ-symbols we obtain an interpretation over $(\Gamma^*, \overline{})$ without changing the truth value by replacing syntactically each subformula $X \in P$ (resp. $X \notin P$) by $\psi(X) \in \psi(P')$ (resp. $\psi(X) \notin \psi(P')$) and by replacing each subformula $W = 1$ by $\psi(W) = 1$.

We keep the interpretation over words, but we eliminate now all occurrences of ψ again. We begin with the occurrences of ψ in the constraints. Let $P' \subseteq \Gamma^*$ be regular being accepted by some finite automaton with state set Q. As stated in the first part of the proof of Proposition 1, we construct a finite automaton, using the same state set, which accepts the following language

$$P'' = \{\, v \in \Gamma^* \mid \exists u \in P' : u \overset{*}{\rightarrow} v \,\}.$$

In particular, $\psi(P') = \psi(P'')$ and $\widehat{P} \subseteq P''$ where $\widehat{P} = \{\, \widehat{u} \in \Gamma^* \mid u \in P' \,\}$.

We replace all positive atomic subformulae of the form $\psi(X) \in \psi(P')$ by $X \in P''$. A simple reflection shows that the truth value has not changed since we can think of X of being a freely reduced word. For a negative formulae $\psi(X) \notin \psi(P')$ we have to be a little more careful. Let $N \subseteq \Gamma^*$ be the regular set of all freely reduced words. The language N is accepted by a deterministic finite automaton with $|\Gamma| + 1$ states. We replace $\psi(X) \notin \psi(P')$ by

$$X \notin P'' \wedge X \in N,$$

where P'' is as above. Again the truth value did not change.

We now have to deal with the formulae $\psi(xyz) = 1$ where $x, y, z \in \Gamma \cup \Omega$. Observe that the underlying propositional formula is satisfiable over Γ^* if and only if it is satisfiable in freely reduced words. The following lemma is well-known. Its easy proof is left to the reader.

Lemma 3 *Let $u, v, w \in \Gamma^*$ be freely reduced words. Then we have $\psi(uvw) = 1$ (i.e. $uvw = 1$ in $F(\Sigma)$) if and only if there are words $P, Q, R \in \Gamma^*$ such that $u = PQ$, $v = \overline{Q}R$, and $w = \overline{R}\,\overline{P}$ holds in Γ^*.*

Based on this lemma we replace each atomic subformulae $\psi(xyz) = 1$ with $x, y, z \in \Gamma \cup \Omega$ by a conjunction

$$\exists P \exists Q \exists R : x = PQ \wedge y = \overline{Q}R \wedge z = \overline{R}\,\overline{P},$$

where P, Q, R are fresh variables and the existential block is put to the front. The new existential formula has no occurrence of ψ anymore. The atomic subformulae are of the form $x = yz$, $X \in P$, $X \notin P$, where $x, y, z \in \Gamma \cup \Omega$ and $P \subseteq \Gamma^*$ is regular. The size of the formula is linear in the size of the original formula. Therefore Theorem 2 is a consequence of Theorem 4.

4 Free Monoids with Involution

As above, let Γ be an alphabet of constants and Ω be an alphabet of variables. There are involutions $^- : \Gamma \to \Gamma$ and $^- : \Omega \to \Omega$. The involution on Ω is without fixed points, but we explicitly allow fixed points for the involution on Γ. [3] The involution is extended to $(\Gamma \cup \Omega)^*$ by $\overline{x_1 \cdots x_n} = \overline{x_n} \cdots \overline{x_1}$ for $n \geq 0$ and $x_i \in \Gamma \cup \Omega$, $1 \leq i \leq n$.
From now on, all monoids M under consideration are equipped with an involution $^- : M \to M$, i.e. we have $\overline{1} = 1$ for the unit element, $\overline{\overline{x}} = x$, and $\overline{xy} = \overline{y}\,\overline{x}$ for all $x, y \in M$. A homomorphism between monoids M and M' is therefore a mapping $h : M \to M'$ such that $h(1) = 1$, $h(xy) = h(x)h(y)$, and $h(\overline{x}) = \overline{h(x)}$ for all $x, y \in M$. The pair $(\Gamma^*, ^-)$ is called a *free monoid with involution.* [4]
The existential theory of equations with regular constraints in free monoids with involution is based on atomic formulae of type $U = V$ where $U, V \in (\Gamma \cup \Omega)^*$ and of type $X \in P$ where $X \in \Omega$ and $P \subseteq \Gamma^*$ is a regular language

[3] Fixed points for the involution on constants are needed in the proof later anyhow and this more general setting leads to further applications, [5]

[4] Note that $(\Gamma^*, ^-)$ is a free monoid which has an involution, but it is not a free object in the category of monoids with involution, as soon as the involution has fixed points.

specified by some non-deterministic finite automaton. Again, a propositional formula is build up by atomic formulae using negations, conjunctions and disjunctions. The existential theory refers to closed existentially quantified propositional formulae which evaluate to *true* over $(\Gamma^*, \bar{\ })$.
The following statement is the main result of the paper.

Theorem 4 *The following problem is* PSPACE*–complete.*
INPUT: A closed existentially quantified propositional formula with regular constraints in a free monoid with involution over $(\Gamma, \bar{\ })$.
QUESTION: Does the formula evaluate to true *over* $(\Gamma^*, \bar{\ })$*?*

The proof of Theorem 4 is in a first step (next section) a reduction to Theorem 5. The proof of Theorem 5 will be the essential technical contribution.

5 From Regular Constraints to Boolean Matrices and a Single Equation

The first part of the proof is very similar to what we have done above. By DeMorgan we have no negations and all subformulae are of type $U = V$, $U \neq V$, $X \in P$, $X \notin P$, where $U, V \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $P \subseteq \Gamma^*$ is regular.
Since we work over a free monoid Γ^* it is easy to handle inequalities $U \neq V$ where $U, V \in (\Gamma \cup \Omega)^*$. We recall it under the assumption $|\Gamma| \geq 2$: A subformulae $U \neq V$ is replaced by

$$\exists X \exists Y \exists Z : \bigvee_{a \neq b} (U = VaX \vee V = UaX \vee (U = XaY \wedge V = XbZ)).$$

Making guesses we can eliminate all disjunctions and we obtain a propositional formula which is a single conjunction over subformulae of type $U = V$, $X \in P$, and $X \notin P$ where $U, V \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $P \subseteq \Gamma^*$ is regular.
By another standard procedure we can replace a conjunction of word equations over $(\Gamma \cup \Omega)^*$ by a single word equation $L = R$ with $L, R \in (\Gamma \cup \Omega)^+$. For example, we may choose a new letter a and then we can replace a system $L_1 = R_1$, $L_2 = R_2, \ldots, L_k = R_k$ by $L_1 a L_2 a \cdots a L_k = R_1 a R_2 a \cdots a R_k$ and a list $X \in \Gamma^*$ for all $X \in \Omega$; this works since $a \notin \Gamma$.
Therefore we may assume that our input is given by a single equation $L = R$ with $L, R \in (\Gamma \cup \Omega)^+$ and by two lists $(X_j \in P_j, 1 \leq j \leq m)$ and $(X_j \notin$

$P_j, m < j \leq k)$ where $X_j \in \Omega$ and each regular language $P_j \subseteq \Gamma^*$ is specified by some non-deterministic automaton $\mathcal{A}_j = (Q_j, \Gamma, \delta_j, I_j, F_j)$ where Q_j is the set of states, $\delta_j \subseteq Q_j \times \Gamma \times Q_j$ is the transition relation, $I_j \subseteq Q_j$ is the subset of initial states, and $F_j \subseteq Q_j$ is the subset of final states, $1 \leq j \leq k$. Of course, a variable X may occur several times in the list with different constraints, therefore we might have k greater than $|\Omega|$. The question is whether there is a solution.

A *solution* is a mapping $\sigma : \Omega \to \Gamma^*$ being extended to a homomorphism $\sigma : (\Gamma \cup \Omega)^* \to \Gamma^*$ by leaving the letters from Γ invariant such that the following conditions are satisfied:

$$\begin{array}{rcll} \sigma(L) & = & \sigma(R), & \\ \sigma(\overline{X}) & = & \overline{\sigma(X)} & \text{for } X \in \Omega, \\ \sigma(X_j) & \in & P_j & \text{for } 1 \leq j \leq m, \\ \sigma(X_j) & \notin & P_j & \text{for } m < j \leq k. \end{array}$$

For the next steps it turns out to be more convenient to work within the framework of Boolean matrices instead of finite automata: Let Q be the disjoint union of the state spaces Q_j, $1 \leq j \leq k$. We may assume that $Q = \{1, \ldots, n\}$. Let $\delta = \bigcup_{1 \leq j \leq k} \delta_j$, then $\delta \subseteq Q \times \Gamma \times Q$ and with each $a \in \Gamma$ we can associate a Boolean $n \times n$ matrix $g(a) \in \mathbb{B}^{n \times n}$ such that $g(a)_{i,j} =$ "$(i, a, j) \in \delta$" for $1 \leq i, j \leq n$. Since our monoids should have an involution, we shall in fact work with $2n \times 2n$ matrices. Henceforth $M \subseteq \mathbb{B}^{2n \times 2n}$ denotes the following monoid with involution:

$$M = \{ \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \mid A, B \in \mathbb{B}^{n \times n} \},$$

where

$$\overline{\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}} = \begin{pmatrix} B^T & 0 \\ 0 & A^T \end{pmatrix}$$

and the operator ^{T} denotes the transposition. We define a homomorphism $h : \Gamma^* \to M$ by

$$h(a) = \begin{pmatrix} g(a) & 0 \\ 0 & g(\overline{a})^T \end{pmatrix} \text{ for } a \in \Gamma,$$

where the mapping $g : \Gamma \to \mathbb{B}^{n \times n}$ is defined as above. The homomorphism h can be computed in polynomial time and it respects the involution. Now, for each regular language P_j, $1 \leq j \leq k$ we compute vectors $I_j, F_j \in \mathbb{B}^{2n}$ such that for all $w \in \Gamma^*$ and $1 \leq j \leq k$ we have the equivalence:

$$w \in P_j \Leftrightarrow I_j^T h(w) F_j = 1.$$

Having done these computations we make a non-deterministic guess $\rho(X) \in M$ for each variable $X \in \Omega$. We verify $\rho(\overline{X}) = \overline{\rho(X)}$ for all $X \in \Omega$ and whenever there is a constraint of type $X \in P_j$ for some $1 \leq j \leq m$ (or $X \notin P_j$ for some $m < j \leq k$), then we verify $I_j^T \rho(X) F_j = 1$, if $1 \leq j \leq m$ (or $I_j^T \rho(X) F_j = 0$, if $m < j \leq k$).
After these preliminaries, we introduce the formal definition of an *equation E with constraints*: Let $d, n \in \mathbb{N}$ and let $M \subseteq \mathbb{B}^{2n \times 2n}$ be the monoid with involution defined above. We consider an equation of length d over some Γ and Ω with constraints in M being specified by a list E containing the following items:

- The alphabet $(\Gamma, \overline{\ })$ with involution.
- The homomorphism $h : \Gamma^* \to M$ which is specified by a mapping $h : \Gamma \to M$ such that $h(\overline{a}) = \overline{h(a)}$ for all $a \in \Gamma$.
- The alphabet $(\Omega, \overline{\ })$ with involution without fixed points.
- A mapping $\rho : \Omega \to M$ such that $\rho(\overline{X}) = \overline{\rho(X)}$ for all $X \in \Omega$.
- The equation $L = R$ where $L, R \in (\Gamma \cup \Omega)^+$ and $|LR| = d$.

We will denote this list simply by

$$E = (\Gamma, h, \Omega, \rho; L = R).$$

A convenient definition for the input size is given by $n + d + \log_2(|\Gamma| + |\Omega|)$. This definition takes into account that there might be constants or variables with constraints which are not present in the equation. Recall that n refers to the dimension of the boolean matrices, and this parameter is part of the input.
A *solution* of E is a mapping $\sigma : \Omega \to \Gamma^*$ (being extended to a homomorphism $\sigma : (\Gamma \cup \Omega)^* \to \Gamma^*$ by leaving the letters from Γ invariant) such that the following three conditions are satisfied:

$$\sigma(L) = \sigma(R),$$

$$\begin{aligned}\sigma(\overline{X}) &= \overline{\sigma(X)} \text{ for all } X \in \Omega,\\ h\sigma(X) &= \rho(X) \text{ for all } X \in \Omega.\end{aligned}$$

By the reduction above, Theorem 4 is a consequence of the next statement which says that the satisfiability problem of equations with constraints can be solved in polynomial space.

Theorem 5 *The following problem is* PSPACE*–complete.*
INPUT: An equation E_0 with constraints $E_0 = (\Gamma_0, h_0, \Omega_0, \rho_0; L_0 = R_0)$.
QUESTION: Is there a solution $\sigma : \Omega_0 \to \Gamma_0^*$?

For the proof we need an explicit space bound. Therefore we fix some polynomial p and and we allow working space $p(n + d + \log_2(|\Gamma| + |\Omega|)$. An appropriate choice of the polynomial p can be calculated from the presentation below. What is important is that the notions of *admissibility* being used in the next sections always refer to some fixed polynomials. The following lemma states that some basic operations, which we have to perform several times can be done in PSPACE.

Lemma 6 *The following two problems can be solved in polynomial space with respect to the input size $n + \log(|\Gamma|)$.*
INPUT: A matrix $A \in M$ and a mapping $h : \Gamma \to M$.
QUESTION: Is there some $w \in \Gamma^*$ such that $h(w) = A$?
INPUT: A matrix $A \in M$ and a mapping $h : \Gamma \to M$.
QUESTION: Is there some $w \in \Gamma^*$ such that $h(w) = A$ and $w = \overline{w}$?

Proof. The first question can be solved by guessing a word w letter by letter and calculating $h(w)$. The second question can be solved since $w = \overline{w}$ implies $w = ua\overline{u}$ for some $u \in \Gamma^*$ and $a \in \Gamma \cup \{1\}$ with $a = \overline{a}$. Hence we can guess u and a. During the guess we compute $B = h(u)$ and then we verify $A = Bh(a)\overline{B}$. □

Here is a first application of Lemma 6: Assume that an equation with constraints $E = (\Gamma, h, \Omega, \rho; L = R)$ contains in the specification some variable X which does not occur in $LR\overline{LR}$, then the equation might be unsolvable, simply because $\rho(X) \notin h(\Gamma^*)$. However, by the lemma above we can test this in PSPACE. If $\rho(X) \in h(\Gamma^*)$, then we can safely cancel X and $\overline{X}$. Thus, we put this test in the preprocessing, and in the following we shall assume that all variables occur somewhere in $LR\overline{LR}$. In particular, we may assume $|\Omega| \leq 2|LR|$.

6 The Exponent of Periodicity

A key step in proving Theorem 5 is to find a bound on the exponent of periodicity in a minimal solution. This idea is used in all known algorithms for solving word equations in general, c.f., [17, 26].
Let $w \in \Gamma^*$ be a word. The exponent of periodicity $\exp(w)$ is defined by

$$\exp(w) = \sup\{\, \alpha \in \mathbb{N} \mid \exists u, v, p \in \Gamma^*, p \neq 1 : w = up^\alpha v \,\}.$$

We have $\exp(w) > 0$ if and only if w is not the empty word. Let $E = (\Gamma, h, \Omega, \rho, L = R)$ be an equation with constraints. The exponent of periodicity of E is also denoted by $\exp(E)$. It is defined by

$$\exp(E) = \inf\{\{\, \exp(\sigma(L)) \mid \sigma \,\text{is a solution of}\, E \,\} \cup \{\infty\}\,\}.$$

By definitions we have $\exp(E) < \infty$ if and only if E is solvable. Here we show that the well-known result from word equations [13] transfers to the situation here. The exponent of periodicity of a solvable equation can be bounded by a singly exponential function. Thus, in the following sections we shall assume that if E_0 is solvable, then $\exp(E_0) \in 2^{\mathcal{O}(d+n\log n)}$. This is the content of the next proposition.

Proposition 7 *Let $E = (\Gamma, h, \Omega, \rho; L = R)$ be an equation with constraints and let $\sigma : \Omega \to \Gamma^*$ be a solution. Then we find effectively a solution $\sigma' : \Omega \to \Gamma^*$ such that* $\exp(\sigma'(L)) \in 2^{\mathcal{O}(d+n\log n)}$.

The rest of this section is devoted to prove Proposition 7. Since it follows standard lines, the proof can be skipped in a first reading.
Proof. Let $p \in A^+$ be a primitive word. In our setting the definition of the *p-stable normal form* of a word $w \in A^*$ depends on the property whether or not $\overline{p}$ is a factor of p^2. So we distinguish two cases and in the following we also write p^{-1} for denoting $\overline{p}$. Then, for example, p^{-3} means the same as $\overline{p}^3$.
First case: We assume that $\overline{p}$ is not a factor of p^2. The idea is to replace each maximal factor of the form p^α with $\alpha \geq 2$ by a sequence $p, \alpha - 2, p$ and each maximal factor of the form $\overline{p}^\alpha$ with $\alpha \geq 2$ by a sequence $\overline{p}, -(\alpha-2), \overline{p}$.
This leads to the following notion:
The *p-stable normal form* (first kind) of $w \in A^*$ is a shortest sequence (k is minimal)

$$(u_0, \varepsilon_1\alpha_1, u_1, \ldots, \varepsilon_k\alpha_k, u_k)$$

such that $k \geq 0$, $u_0, u_i \in A^*$, $\varepsilon_i \in \{+1, -1\}$, $\alpha_i \geq 0$ for $1 \leq i \leq k$, and the following conditions are satisfied:

- $w = u_0 p^{\varepsilon_1 \alpha_1} u_1 \cdots p^{\varepsilon_k \alpha_k} u_k$.
- $k = 0$ if and only if neither p^2 nor $\overline{p}^2$ is a factor of w.
- If $k \geq 1$, then:

$$\begin{array}{rcl} u_0 & \in & A^* p^{\varepsilon_1} \setminus A^* p^{\pm 2} A^*, \\ u_i & \in & (A^* p^{\varepsilon_{i+1}} \cap p^{\varepsilon_i} A^*) \setminus A^* p^{\pm 2} A^* \text{ for } 1 \leq i < k, \\ u_k & \in & p^{\varepsilon_k} A^* \setminus A^* p^{\pm 2} A^*. \end{array}$$

The p-stable normal form of $\overline{w}$ becomes

$$(\overline{u_k}, -\varepsilon_k \alpha_k, u_{k-1}, \ldots, -\varepsilon_1 \alpha_1, \overline{u_0}).$$

Example 8 *Let $p = a\overline{a}ba\overline{a}$ with $b \neq \overline{b}$ and $w = p^4 \overline{b}a\overline{a}p^{-1}a\overline{a}\overline{b}p^{-2}$. Then the p-stable normal form of w is:*

$$(\overline{a}a\overline{a}b, 2, a\overline{a}ba\overline{a}\overline{b}a\overline{a}, -1, a\overline{a}\overline{b}a\overline{a}\overline{b}a\overline{a}, 0, a\overline{a}\overline{b}a\overline{a}).$$

Second case: We assume that $\overline{p}$ is a factor of p^2. Then we can write $p = rs$ with $\overline{p} = sr$ and $r = \overline{r}$, $s = \overline{s}$. We allow $r = 1$, hence the second case includes the case $p = \overline{p}$. In fact, if $r = 1$, then below we obtain the usual definition of p-stable normal form. Moreover, by switching to some conjugated word of p we could always assume that $r \in \{1, a\}$ for some letter a being fixed by the involution, $a = \overline{a}$, but this switch is not made here. The idea is to replace each maximal factor of the form $(rs)^\alpha r$ with $\alpha \geq 2$ by a sequence $rs, \alpha - 2, sr$. In this notation $\alpha - 2$ is representing the factor $(rs)^{\alpha-2} r = p^{\alpha-2} r = r\overline{p}^{\alpha-2}$. The *p-stable normal form* (second kind) of $w \in A^*$ is now a shortest sequence (k is minimal)

$$(u_0, \alpha_1, u_1, \ldots, \alpha_k, u_k)$$

such that $k \geq 0$, $u_0, u_i \in A^*$, $\alpha_i \geq 0$ for $1 \leq i \leq k$, and the following conditions are satisfied:

- $w = u_0 p^{\alpha_1} r u_1 \cdots p^{\alpha_k} r u_k$.
- $k = 0$ if and only if $p^2 r$ is not a factor of w.

- If $k \geq 1$, then:

$$\begin{array}{rcl} u_0 & \in & A^*rs \setminus (A^*p^2rA^* \cup A^*rsrs), \\ u_i & \in & (A^*rs \cap srA^*) \setminus (srsrA^* \cup A^*p^2rA^* \cup A^*rsrs) \text{ for } 1 \leq i < k, \\ u_k & \in & srA^* \setminus (A^*p^2rA^* \cup srsrA^*). \end{array}$$

Since $\overline{rs} = sr$, the p-stable normal form of $\overline{w}$ becomes

$$(\overline{u_k}, \alpha_k, u_1, \ldots, \alpha_1, \overline{u_0}).$$

So, for the second kind no negative integers interfere.

Example 9 *Let $p = a\overline{a}b$ with $b = \overline{b}$. Then $r = a\overline{a}$ and $s = b$. Let $w = \overline{a}p^4ap^3a$ Then the p-stable normal form of w is:*

$$(\overline{a}ba\overline{a}b, 2, ba\overline{a}baa\overline{a}b, 0, ba\overline{a}ba).$$

In both cases we can write the p-stable normal form of w as a sequence

$$(u_0, \alpha_1, u_1, \ldots, \alpha_k, u_k)$$

where u_i are words and α_i are integers.
For every finite semigroup S there is a number $c(S)$ such that for all $s \in S$ the element $s^{c(S)}$ is idempotent, i.e., $s^{c(S)} = s^{2c(S)}$. It is clear that the number $c(M)$ for our monoid $M \subseteq \mathbb{B}^{2n \times 2n}$ is the same as the number $c(\mathbb{B}^{n \times n})$. It is well-known [21] that we can take $c(\mathbb{B}^{n \times n}) = n!$ (it is however more convenient to define $c(M) = 3$ for $n = 1$). Hence in the following $c(M) = \max\{3, n!\}$. For specific situations this might be an overestimation, but this choice guarantees $h(uv^{c(M)}w) = h(uv^{2c(M)}w)$ for all $u, v, w \in \Gamma^*$ and all $h : \Gamma^* \to M$.
Now, let $w, w' \in \Gamma^*$ be words such that the p-stable normal forms are identical up to one position where for w appears an integer α_i and for w' appears an integer α'_i. We know $h(w) = h(w')$ whenever the following conditions are satisfied: $\alpha_i \cdot \alpha'_i > 0$, $|\alpha_i| \geq c(M)$, $|\alpha'_i| \geq c(M)$, and $\alpha_i \equiv \alpha'_i \pmod{c(M)}$. Then we have $h(w) = h(w')$. This is the reason to change the syntax of the p-stable normal form. Each non-zero integer α' is written as $\alpha' = \varepsilon(q + \alpha c(M))$ where ε, q, α are uniquely defined by $\varepsilon \in \{+1, -1\}$, $0 \leq q < c(M)$, and $\alpha \geq 0$. For $\alpha' = 0$ we may choose $\varepsilon = q = \alpha = 0$. We shall read α as a variable ranging over non-negative integers, but ε, q, and $c(M)$ are viewed

as constants. In fact, if $|\alpha'| < c(M)$, then we best view α also as a constant in order to avoid problems with the constraints.

Let u, v, and w be words such that $uv = w$ holds. Write these words in their p-stable normal forms:

$$\begin{array}{ll} u: & (u_0, \varepsilon_1(q_1 + \alpha_1 c(S)), u_1, \ldots, \varepsilon_k(q_k + \alpha_k c(S)), u_k), \\ v: & (v_0, \varepsilon_1'(s_1 + \beta_1 c(S)), v_1, \ldots, \varepsilon_\ell'(s_\ell + \beta_\ell c(S)), v_\ell), \\ w: & (w_0, \varepsilon_1''(t_1 + \gamma_1 c(S)), w_1, \ldots, \varepsilon_m''(t_m + \gamma_m c(S)), w_m). \end{array}$$

Since $uv = w$ there are many identities. For example, for $k, \ell \geq 2$ we have $u_0 = w_0$, $v_l = w_m$, $q_1 = t_1$, $\alpha_1 = \gamma_1$, etc. What exactly happens depends only on the p-stable normal form of the product $u_k v_0$. There are several cases, which easily can be listed. We treat only one of them, which is in some sense the worst case in order to produce a large exponent of periodicity. This is the case where $p = rs$ with $r = \overline{r}$ and $s = \overline{s}$. Then it might be that $u_k = srsr_1$ and $v_0 = r_2 srs$ with $r_1 r_2 = r$ (and $r_1 \neq 1 \neq r_2$). Hence we have $u_k v_0 = sp^3$ and $k + \ell = m + 1$. It follows $\alpha_1 = \gamma_1, \ldots, \alpha_{k-1} = \gamma_{k-1}$, $\beta_2 = \gamma_{k+1}, \ldots, \beta_\ell = \gamma_m$, and there is only one non-trivial identity:

$$q_k + s_1 + 4 + (\alpha_k + \beta_1)c(S) = t_k + \gamma_k c(S).$$

Since by assumption $c(S) \geq 3$, the case $u_k v_0 = sp^3$ leads to the identity:

$$\gamma_k = \alpha_k + \beta_1 + c \text{ with } c \in \{0, 1, 2\}.$$

Assume now that $\alpha_k \geq 1$ and $\beta_1 \geq 1$. If we replace α_k, β_1, and γ_k by some $\alpha_k' \geq 1$, $\beta_1' \geq 1$, and $\gamma_k' \geq 1$ such that still $\gamma_k' = \alpha_k' + \beta_1' + c$, then we obtain new words u', v', and w' with the same images under h in M and still the identity $u'v' = w'$.

What follows then is completely analogous to what has been done in detail in [13, 10, 11, 3]. Using the p-stable normal form we can associate with an equation $L = R$ of denotational length d together with its solution $\sigma : \Omega \to \Gamma^*$ some linear Diophantine system of d equations in at most $3d$ variables. The variables range over natural numbers since zeros are substituted. (In fact the number of variables can be reduced to be at most $2|\Omega|$). The parameters of this system are such that maximal size of a minimal solution (with respect to the component wise partial order of $\mathbb{N}^d$) is in $\mathcal{O}(2^{1.6d})$ with the same approach as in [13]. This tight bound is based in turn on the work of [30]; a more moderate bound $2^{\mathcal{O}(d)}$ (which is enough for our purposes) is easier to obtain, see e.g. [3]. The maximal size of a minimal solution of the linear Diophantine system has a backward translation to a bound on the exponent of periodicity. For this translation we have to multiply with the factor $c(M) \in 2^{\mathcal{O}(n \log n)}$ and to add $c(M)+1$. Putting everything together we obtain the claim of the proposition. $\square$

7 Exponential Expressions

During the procedure which solves Theorem 5 various other equations with constraints are considered but the monoid M will not change.
There will be not enough space to write down the equation $L = R$ in plain form, in general. In fact, there is a provable exponential lower bound for the length $|LR|$ in the worst case which we can meet during the procedure. In order to overcome this difficulty Plandowski's method uses data compression for words in $(\Gamma \cup \Omega)^*$ in terms of exponential expressions.
Exponential expressions (their evaluation and their size) are inductively defined:

- Every word $w \in \Gamma^*$ denotes an exponential expression. The evaluation $\mathrm{eval}(w)$ is equal to w, its size $\|w\|$ is equal to the length $|w|$.
- Let e, e' be exponential expressions. Then ee' is an exponential expression. Its evaluation is the concatenation $\mathrm{eval}(ee') = \mathrm{eval}(e)\mathrm{eval}(e')$, its size is $\|ee'\| = \|e\| + \|e'\|$.
- Let e be an exponential expression and $k \in \mathbb{N}$. Then $(e)^k$ is an exponential expression. Its evaluation is $\mathrm{eval}((e)^k) = (\mathrm{eval}(e))^k$, its size is $\|(e)^k\| = \log(k) + \|e\|$ where $\log(k) = \max\{1, \lceil \log_2(k) \rceil\}$.

It is not difficult to show that the length of $\mathrm{eval}(e)$ is at most exponential in the size of e, a fact which is, strictly speaking, not needed for the proof of Theorem 5. What we need however is the next lemma. Its proof can be done easily by structural induction and it is omitted.

Lemma 10 *Let $u \in \Gamma^*$ be a factor of a word $w \in \Gamma^*$. Assume that w can be represented by some exponential expression of size p. Then we find an exponential expression of size at most p^2 that represents u.*

We say that an exponential expression e is *admissible*, if its size $\|e\|$ is bounded by some fixed polynomial in the input size of E_0. The lemma above states that if e is admissible, then we find admissible exponential expressions for all factors of $\mathrm{eval}(e)$. But now the admissibility is defined with respect to some polynomial which is the square of the original polynomial, so, in a nested way, we can apply this procedure a constant number of times, only. In our application the nested depth does not go beyond two.
The next lemma is straightforward since we allow a polynomial space bound without any time restriction. Again, the proof is left to the reader.

Lemma 11 *The following two problems can be solved in* PSPACE.
INPUT: Exponential expressions e and e'.
QUESTION: Do we have $\mathrm{eval}(e) = \mathrm{eval}(e')$*?*
INPUT: A mapping $h : \Gamma \to M$ and an exponential expression e.
OUTPUT: The matrix $h(\mathrm{eval}(e)) \in M$.

Remark 12 *The computation above can actually be performed in polynomial time, but this is not evident for the first question, see [24] for details.*

Henceforth we allow that the part $L = R$ of an equation with constraints may also be given by a pair of exponential expressions (e_L, e_R) with $\mathrm{eval}(e_L) = L$ and $\mathrm{eval}(e_R) = R$. We say that $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ is *admissible*, if $e_L e_R$ is admissible, $|\Gamma \setminus \Gamma_0|$ has polynomial size, $\Omega \subseteq \Omega_0$, and $h(a) = h_0(a)$ for $a \in \Gamma \cap \Gamma_0$.
For two admissible equations with constraints $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ and $E' = (\Gamma, h, \Omega, \rho; e'_L = e'_R)$ we write $E \equiv E'$, if $\mathrm{eval}(e_L) = \mathrm{eval}(e'_L)$ and $\mathrm{eval}(e_R) = \mathrm{eval}(e'_R)$ as strings in $(\Gamma \cup \Omega)^*$. This means that they represent exactly the same equations.

8 Base Changes

In this section we fix a mapping $h : \Gamma \to M$ which respects the involution. Let $(\Gamma', {}^-)$ be an alphabet with involution and let $\beta : \Gamma' \to \Gamma^*$ be some mapping β such that $\beta(\overline{a}) = \overline{\beta(a)}$ for all $a \in \Gamma'$. We define $h' : \Gamma' \to M$ such that $h' = h\beta$. We also extend to a homomorphism $\beta : (\Gamma' \cup \Omega)^* \to (\Gamma \cup \Omega)^*$ by leaving the variables invariant.

Let $E' = (\Gamma', h', \Omega, \rho; L' = R')$. be an equation with constraints. The *base change* $\beta_*(E')$ is defined by

$$\beta_*(E') = (\Gamma, h, \Omega, \rho; \beta(L') = \beta(R')).$$

We also refer to $\beta : \Gamma' \to \Gamma^*$ as a base change and we say that β is *admissible*, if $|\Gamma'|$ has polynomial size and if $\beta(a)$ can be represented by some admissible exponential expression for all $a \in \Gamma'$.

Remark 13 *If $\beta : \Gamma' \to \Gamma^*$ is an admissible base change and if $L' = R'$ is given by a pair of admissible exponential expressions, then we can represent $\beta_*(E')$ by some admissible equation with constraints. A representation of $\beta_*(E')$ is computable in polynomial time.*

Lemma 14 *Let E' be an equation with constraints and $\beta : \Gamma' \to \Gamma^*$ be a base change. If σ' is a solution of E', then $\sigma = \beta\sigma'$ is a solution of $\beta_*(E')$.*

Proof. Clearly $\sigma(\overline{X}) = \overline{\sigma(X)}$ and $h\sigma(X) = h\beta\sigma'(X) = h'\sigma'(X) = \rho(X)$ for all $X \in \Omega$. Next by definition $\sigma(a) = a$ for $a \in \Gamma$ and $\beta(X) = X$ for $X \in \Omega$. Hence $\sigma\beta(a) = \beta\sigma'(a)$ for $a \in \Gamma'$ and therefore $\sigma\beta = \beta\sigma' : (\Gamma' \cup \Omega)^* \to \Gamma^*$. This means $\sigma\beta(L) = \beta\sigma'(L) = \beta\sigma'(R) = \sigma\beta(R)$ since $\sigma'(L) = \sigma'(R)$. □
The lemma above leads to the first rule.

Rule 1 *If E is of the form $\beta_*(E')$ and if we are looking for a solution of E, then it is enough to find a solution for E'. Hence, during a non-deterministic search we may replace E by E'.*

Example 15 *Consider the following equation E with constraints over $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$:*

$$X\overline{X} = Y\bar{b}\bar{c}\bar{b}\bar{a}\bar{b}\bar{c}\bar{b}YZabcb\overline{Y}.$$

Let there be the constraints for X and Z saying $X \in \Gamma^{300}\Gamma^*$ and $Z \in \bar{b}\bar{c}\bar{b}\bar{a}\Gamma^*$. Define $\Gamma' = \{a, b, \bar{a}, \bar{b}\}$ and a base change $\beta : \Gamma' \to \Gamma^*$ by $\beta(a) = abcb$ and $\beta(b) = bcb$. Then the equation E is of the form $\beta_*(E')$ where E' is given by

$$X\overline{X} = Y\bar{a}\bar{b}YZa\overline{Y}$$

and the new (and sharper) constraint for Z is simply $Z \in \bar{a}\Gamma'^*$, for X we may sharpen the constraint to $X \in \Gamma'^{100}\Gamma'^*$ According to Rule 1 it is enough *to solve E'. The effect of the base change β is that the equation E' is shorter and the alphabet of constants becomes smaller, since the letter c is not used anymore. Note also that the length restriction on X became smaller, too. However this has a prize; in general, $E = \beta_*(E')$ might have a solution, whereas E' is unsolvable. As we will see later, our guess has been correct in the sense that E' still has a solution.*

9 Projections

Let $(\Gamma, \overline{\ })$ and $(\Gamma', \overline{\ })$ be alphabets with involution such that $(\Gamma, \overline{\ }) \subseteq (\Gamma', \overline{\ })$. A *projection* is a homomorphism $\pi : \Gamma'^* \to \Gamma^*$ such that both $\pi(a) = a$ for $a \in \Gamma$ and $\pi(\overline{a}) = \overline{\pi(a)}$ for all $a \in \Gamma'$. If $h : \Gamma \to M$ is given, then a projection π defines also $h' : \Gamma' \to M$ by $h' = h\pi$.
Let E be an equation with constraints $E = (\Gamma, h, \Omega, \rho; L = R)$. Then we can define an equation with constraints $\pi^*(E)$ by

$$\pi^*(E) = (\Gamma', h\pi, \Omega, \rho; L = R).$$

The difference between E and $\pi^*(E)$ is only in the alphabets of constants and in the mappings h and $h' = h\pi$. Note that every projection $\pi : \Gamma'^* \to \Gamma^*$ defines a base change π_* such that $\pi_*\pi^*(E) = E$.

Lemma 16 *Let $E = (\Gamma, h, \Omega, \rho; L = R)$ and $E' = (\Gamma', h', \Omega, \rho; L = R)$ be equations with constraints. Then the following two statements hold.*

i) There is a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $\pi^*(E) = E'$, if and only if both $h'(\Gamma') \subseteq h(\Gamma^*)$ and for all $a \in \Gamma'$ with $a = \overline{a}$ there is some $w \in \Gamma^*$ with $w = \overline{w}$ such that $h'(a) = h(w)$.

ii) If we have $\pi^*(E) = E'$ and if $\sigma' : \Omega \to \Gamma'^*$ is a solution of E', then we effectively find a solution σ for E such that $|\sigma(L)| \leq 2|M||\sigma'(L)|$.

Proof. i) Clearly, the only-if condition is satisfied by the definition of a projection since then $h' = h\pi$. For the converse, assume that $h'(\Gamma') \subseteq h(\Gamma^*)$ and that $a = \overline{a}$ implies $h'(a) \in h(\{w \in \Gamma^* \mid w = \overline{w}\})$. Then for each $a \in \Gamma' \setminus \Gamma$ we can choose a word $w_a \in \Gamma^*$ such that $h'(a) = h(w_a)$. We can make the choice such that $w_{\overline{a}} = \overline{w_a}$ for all $a \in \Gamma' \setminus \Gamma$. If $a \neq \overline{a}$, then we can find w_a such that $|w_a| < |M|$, since we can take the shortest word $w_a \in \Gamma^*$ such that $h(w_a) = h'(a) \in M$. For $a = \overline{a}$ we know that there is some word $w_a \in \Gamma^*$ with

$h'(a) = h(w_a)$ and $w_a = \overline{w_a}$. Hence we can write $w_a = vb\overline{v}$ with $b \in \Gamma \cup \{1\}$ and $b = \overline{b}$. For $b \neq 1$ we can demand $|w_a| \leq 2|M| - 1$. For $b = 1$ we can demand $|w_a| \leq 2|M| - 2$. Thus, we find a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $\pi^*(E) = E'$ and moreover, $|\pi(a)| < 2|M|$ for all $a \in \Gamma'$.
ii) Using the reasoning in the proof of i) we may assume that $\pi : \Gamma'^* \to \Gamma^*$ satisfies $|\pi(a)| < 2|M|$ for all $a \in \Gamma'$. Since π defines a base change with $\pi_*(E') = E$, we know by Lemma 14 that $\sigma = \pi\sigma'$ is a solution of E. Clearly, $|\sigma(L)| = |\pi\sigma'(L)| \leq 2|M||\sigma'(L)|$. □

Remark 17 *In the following we will meet the problem to decide whether there is a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $\pi^*(E) = E'$. We actually need not too much space for this test. It is not necessary to write down π. We can use the criterion in the lemma above and Lemma 6. Then we have to store in the working space only some Boolean matrices of $\mathbb{B}^{2n \times 2n}$. In particular, if n is a constant (or logarithmically bounded in the input size), then the test $\exists \pi : \pi^*(E) = E'$ can be done in polynomial time. However, if n becomes a substantial part of the input size, then the test might be difficult in the sense that we might need the full power of* PSPACE.

The lemma above leads now to the second rule.

Rule 2 *If π is a projection and if we are looking for a solution of E, then it is enough to find a solution for $\pi^*(E)$. Hence, during a non-deterministic search we may replace E by $\pi^*(E)$.*

Example 18 *Let us continue with the equation which has been obtained by the transformation in Example 15. In order to simplify notations, we will call E the equation $X\overline{X} = Y\bar{a}\bar{b}YZa\overline{Y}$, and $\Gamma = \{a, b, \bar{a}, \bar{b}\}$.*
Remember that the constraint on X demanded a rather long solution. Therefore we may reintroduce a letter c and put $\Gamma' = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. Then we may define a projection $\pi : \Gamma' \to \Gamma^*$ by, say, $\pi(c) = b^{100}$. The equation $E' = \pi^*(E)$ looks as above, but in E' we may change the constraint for X. We may sharpen the new constraint for X to be $X \in \Gamma^* c \Gamma^*$. Thus, the solution for X might be very short now.

10 Partial Solutions

Let $\Omega' \subseteq \Omega$ be a subset of the variables which is closed under involution. We assume that there is a mapping $\rho' : \Omega' \to M$ with $\rho'(\overline{x}) = \overline{\rho'(x)}$, but we do not require that ρ' is the restriction of $\rho : \Omega \to M$. Consider an equation with constraints $E = (\Gamma, h, \Omega, \rho; L = R)$. A *partial solution* is a mapping $\delta : \Omega \to \Gamma^*\Omega'\Gamma^* \cup \Gamma^*$ such that the following conditions are satisfied:

i) $\delta(X) \in \Gamma^* X \Gamma^*$ for all $X \in \Omega'$,

ii) $\delta(X) \in \Gamma^*$ for all $X \in \Omega \setminus \Omega'$,

iii) $\delta(\overline{X}) = \overline{\delta(X)}$ for all $X \in \Omega$.

The mapping δ is extended to a homomorphism $\delta : (\Gamma \cup \Omega)^* \to (\Gamma \cup \Omega')^*$ by leaving the elements of Γ invariant. Let $E' = (\Gamma, h, \Omega', \rho'; L' = R')$ be another equation with constraints (using the same Γ and h). We write $E' = \delta_*(E)$, if there exists some partial solution $\delta : \Omega \to \Gamma^*\Omega\Gamma^* \cup \Gamma^*$ such that the following conditions hold: $L' = \delta(L)$, $R' = \delta(R)$, $\rho(X) = h(u)\rho'(X)h(v)$ for $\delta(X) = uXv$, and $\rho(X) = h(w)$ for $\delta(X) = w \in \Gamma^*$.

Lemma 19 *In the notation of above, let $E' = \delta_*(E)$ for some partial solution $\delta : \Omega \to \Gamma^*\Omega\Gamma^* \cup \Gamma^*$. If σ' is a solution of E', then $\sigma = \sigma'\delta$ is a solution of E. Moreover, we have $\sigma(L) = \sigma'(L')$ and $\sigma(R) = \sigma'(R')$.*

Proof. By definition, δ and σ' are extended to homomorphisms $\delta : (\Gamma \cup \Omega)^* \to (\Gamma \cup \Omega')^*$ and $\sigma' : (\Gamma \cup \Omega')^* \to \Gamma^*$ leaving the letters of Γ invariant. Since $E' = \delta_*(E)$ we have $\delta(L) = L'$ and $\delta(R) = R'$. Since σ' is a solution, we have $\sigma(L) = \sigma'\delta(L) = \sigma'(L') = \sigma'(R') = \sigma'\delta(R) = \sigma(R)$ and σ leaves the letters of Γ invariant. The solution σ' satisfies $h\sigma'(X) = \rho'(X)$ for all $X \in \Omega'$. Hence, if $\delta(X) = uXv$, then $\rho(X) = h(u)\rho'(X)h(v) = h(u\sigma'(X)v) = h\sigma'(uXv) = h\sigma'\delta(X) = h\sigma(X)$. If $\delta(X) = w \in \Gamma^*$, then $\sigma(X) = \sigma'\delta(X) = w$ and $\rho(X) = h(w)$, again by the definition of a partial solution. □

Lemma 20 *The following problem can be solved in* PSPACE.

INPUT: Two equations with constraints $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ and $E' = (\Gamma, h, \Omega', \rho'; e_{L'} = e_{R'})$.

QUESTION: Is there some partial solution δ such that $\delta_*(E) \equiv E'$?

Moreover, if $\delta_*(E) \equiv E'$ is true, then there are admissible exponential expressions e_u, e_v for each $X \in \Omega'$ and an admissible exponential expression e_w for each $X \in \Omega \setminus \Omega'$ such that

$$\begin{array}{rcll} \delta(X) & = & \mathrm{eval}(e_u) X \mathrm{eval}(e_v) & \text{for } X \in \Omega', \\ \delta(X) & = & \mathrm{eval}(e_w) & \text{for } X \in \Omega \setminus \Omega'. \end{array}$$

Proof. Let $L = \mathrm{eval}(e_L)$, $R = \mathrm{eval}(e_R)$, $L' = \mathrm{eval}(e_{L'})$, and $R' = \mathrm{eval}(e_{R'})$. The non-deterministic algorithm works as follows:
For each $X \in \Omega'$ we guess admissible exponential expressions e_u and e_v with $\mathrm{eval}(e_u), \mathrm{eval}(e_v) \in \Gamma^*$. We define an exponential expressions $e_X = e_u X e_v$ and $\delta(X) = \mathrm{eval}(e_X)$. For each $X \in \Omega \setminus \Omega'$ we guess an admissible exponential e_X with $\mathrm{eval}(e_X) \in \Gamma^*$ and $\delta(X) = \mathrm{eval}(e_X)$.
Next we verify whether or not $\delta_*(E) \equiv E'$. During this test we have to create an exponential expression f_L (and f_R, resp.) by replacing X in e_L (and e_R, resp.) with the expression e_X. This increases the size in the worst case by a factor of $\max\{||e_X|| \mid X \in \Omega\}$. The other tests whether $\rho(X) = h(u)\rho'(X)h(v)$ for $\delta(X) = uXv$ and $\rho(X) = h(w)$ for $\delta(X) = w \in \Gamma^*$ involve admissible exponential expressions over Boolean matrices and can be done in polynomial time.
The correctness of the algorithm follows from our general assumption that all $X \in \Omega$ appear in $LR\overline{LR}$. Therefore, if we have $\delta_*(E) \equiv E'$, then $\delta(X)$ (or $\delta(\overline{X})$) appears necessarily as a factor in $L'R' = \delta(LR)$. Hence $\delta(X)$ has an exponential expression of polynomial size by Lemma 10. Therefore guesses of e_u, e_v, and e_w as above are possible without running out of space. □

Remark 21 *Actually, the test for $\delta_*(E) \equiv E'$ can be performed in non-deterministic polynomial time by Remark 12.*

The lemma above leads to the third and last rule.

Rule 3 *If δ is a partial solution and if we are looking for a solution of E, then it is enough to find a solution for $\delta_*(E)$. Hence, during a non-deterministic search we may replace E by $\delta_*(E)$.*

Remark 22 *We can think of a partial solution $\delta : \Omega \to \Gamma^*\Omega'\Gamma^* \cup \Gamma^*$ in the following sense. Assume we have an idea about $\sigma(X)$ for some $X \in \Omega$. Then we might guess $\sigma(X)$ entirely. In this case we can define $\delta(X) = \sigma(X)$*

and we have $X \notin \Omega'$. For some other X we might guess only some prefix u and some suffix v of $\sigma(X)$. Then we define $\delta(X) = uXv$ and we have to guess some $\rho'(X) \in M$ such that $\rho(x) : h(u)\rho'(X)h(v)$. If our guess was correct, then such a matrix $\rho'(X) \in M$ must exist. We have partially specified the solution and applying Rule 3, we continue this process by replacing the equation $L = R$ by the new equation $\delta(L) = \delta(R)$.

Example 23 *We continue with our running example. After renaming, the equation E is given by*

$$X\overline{X} = Y\bar{a}\bar{b}YZa\overline{Y},$$

and the alphabet of constant is given by $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. The constraints are $X \in \Gamma^* c \Gamma^*$ and $Z \in \bar{a}\{a, b, \bar{a}, \bar{b}\}^*$.
We may guess the partial solution as follows: $\delta(X) = aX$, $\delta(Y) = Y$, and $\delta(Z) = \bar{a}b$. The new equation $\delta_*(E)$ is

$$aX\overline{X}\bar{a} = Y\bar{a}\bar{b}Y\bar{a}ba\overline{Y}.$$

The remaining constraint is that the solution for X has to use the letter c. The process can continue, for example, we can apply Rule 1 again by defining another base change $\beta(b) = ba$ to get the equation

$$aX\overline{X}\bar{a} = Y\bar{b}Y\bar{a}b\overline{Y}$$

over $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. Since the last equation has a solution (e.g., given by $\sigma(X) = bc\bar{c}\bar{b}\bar{b}abc$ and $\sigma(Y) = abc\bar{c}\bar{b}$), the first equation with constraints in Example 15 has a solution too.

11 The Search Graph and Plandowski's Algorithm

In the following we show that there is some fixed polynomial (which can be calculated from the presentation below) such that the high-level description of Plandowski's algorithm is as follows: On input E_0 compute the maximal space bound, given by the polynomial, to be used by the procedure. Then apply non-deterministically Rules 1, 2, and 3 until an equation with a trivial solution is found.

From the description above it follows that the specification of the algorithm just uses Rules 1, 2, 3. The algorithm is simple but it demands a good heuristics to explore the search graph. The hard part is to prove that this schema is correct; for this we have to be more precise.
The *search graph* is a directed graph: The nodes are admissible equations with constraints. For two nodes E, E', we define an arc $E \to E'$, if there are an admissible base change β, a projection π, and a partial solution δ such that $\delta_*(\pi^*(E)) \equiv \beta_*(E')$.

Lemma 24 *The following problem can be decided in* PSPACE.
INPUT: Admissible equations with constraints E and E'.
QUESTION: Is there an arc $E \to E'$ in the search graph?

Proof. We first guess some alphabet $(\Gamma'', \bar{\ })$ of polynomial size together with $h'' : \Gamma'' \to M$. Then we guess some admissible base change $\beta : \Gamma' \to \Gamma''^*$ such that $h' = h''\beta$ and we compute $\beta_*(E')$.
Next we guess some admissible equation with constraints E'' which uses Γ'' and Ω. We check using Lemma 20 that there is some partial solution $\delta : \Omega \to \Gamma''^*\Omega'\Gamma''^* \cup \Gamma''^*$ such that $\delta_*(E'') \equiv \beta_*(E')$. (Note that every equation with constraints E'' satisfying $\delta_*(E'') \equiv \beta_*(E')$ for some δ is admissible by Lemma 10.) Finally we check using Remark 22 and that there is some projection $\pi : \Gamma'' \to \Gamma$ such that $\pi^*(E) \equiv E''$. We obtain $\delta_*(\pi^*(E)) \equiv \beta_*(E')$.
□

Remark 25 *Following Remarks 12 and 21 the problem in Lemma 24 can be decided in non-deterministic polynomial time, if the monoid M is not part of the input and viewed as a constant. If, as in our setting, M is part of the input, then* PSPACE *is the best we can prove, because the test for the projection becomes difficult.*

Plandowski's algorithm works as follows:

begin
    $E := E_0$
    while $\Omega \neq \emptyset$ **do**
        Guess an admissible equation E' with constraints
        Verify that $E \to E'$ is an arc in the search graph
        $E := E'$
    endwhile
    return "eval(e_L) = eval(e_R)"
end

By Rules 1–3 (Lemmata 14, 16 *ii*), and 19), if $E \to E'$ is an arc in the search graph and E' is solvable, then E is solvable, too. Thus, if the algorithm returns *true*, then E_0 is solvable. The proof of Theorem 5 is therefore reduced to the statement that if E_0 is solvable, then the search graph contains a path to some node without variables and the exponential expressions defining the equation evaluate to the same word. This existence proof is the hard part, it covers the rest of the paper.

Remark 26 *If $E \to E'$ is due to some $\pi : \Gamma''^* \to \Gamma^*$, $\delta : \Omega \to \Gamma''^*\Omega'\Gamma''^* \cup \Gamma''^*$, and $\beta : \Gamma'^* \to \Gamma''^*$, then a solution $\sigma' : \Omega' \to \Gamma'^*$ of E' yields the solution $\sigma = \pi(\beta\sigma')\delta$. Hence we may assume that the length of a solution has increased by at most an exponential factor by Lemma 16 ii). Since we are going to perform the search in a graph of at most exponential size, we get automatically a doubly exponential upper bound for the length of a minimal solution by backwards computation on such a path. This is still the best known upper bound (although an singly exponential bound is conjectured), see [25].*

12 Free Intervals

In this section we introduce the notion of *free interval* in order to cope with long factors in the solution which are not related to any cut. If there were no constraints, then these factors would not appear in a minimal solution. In our setting we cannot avoid these factors.

For a word $w \in \Gamma^*$ we let $\{0, \ldots, |w|\}$ be the set of its *positions*. The interpretation is that factors of w are between positions. To be more specific let $w = a_1 \cdots a_m$, $a_i \in \Gamma$ for $1 \leq i \leq m$. Then $[\alpha, \beta]$ with $0 \leq \alpha < \beta \leq m$ is called a *positive interval* and the factor $w[\alpha, \beta]$ is defined by the word $w[\alpha, \beta] = a_{\alpha+1} \cdots a_\beta$.

It is convenient to have an involution on the set of intervals. Therefore $[\beta, \alpha]$ is also called an interval (but it is never positive), and we define $w[\beta, \alpha] = \overline{w[\alpha, \beta]}$. We allow also $\alpha = \beta$ and we define $w[\alpha, \alpha]$ to be the empty word. For all $0 \leq \alpha, \beta \leq m$ we let $\overline{[\alpha, \beta]} = [\beta, \alpha]$, then always $\overline{w[\alpha, \beta]} = w\overline{[\alpha, \beta]}$.

Let us focus on the word $w_0 \in \Gamma_0^*$ which in our notation is the solution $w_0 = \sigma(L_0) = \sigma(R_0)$, where $L_0 = x_1 \cdots x_g$ and $R_0 = x_{g+1} \cdots x_d$, $x_i \in (\Gamma_0 \cup \Omega_0)$ for $1 \leq i \leq d$. We are going to define an equivalence relation $\approx$ on the set of intervals of w_0. For this we have to fix some few more notations. We let $m_0 = |w_0|$ and for $i \in \{1, \ldots, d\}$ we define positions $\mathrm{l}(i) \in \{0, \ldots, m_0 - 1\}$

and $\mathrm{r}(i) \in \{1, \ldots, m_0\}$ by the congruences

$$\begin{aligned}
\mathrm{l}(i) &\equiv |\sigma(x_1 \cdots x_{i-1})| \mod m_0, \\
\mathrm{r}(i) &\equiv |\sigma(x_{i+1} \cdots x_d)| \mod m_0.
\end{aligned}$$

This means, the factor $\sigma(x_i)$ starts in w_0 at the left position $\mathrm{l}(i)$ and it ends at the right position $\mathrm{r}(i)$. In particular, we have $\mathrm{l}(1) = \mathrm{l}(g+1) = 0$ and $\mathrm{r}(g) = \mathrm{r}(d) = m_0$. The set of l and r positions is called the set of *cuts*. Thus, the set of cuts is $\{\mathrm{l}(i), \mathrm{r}(i) \mid 1 \leq i \leq d\}$. There are at most d cuts. These positions cut the word w_0 in at most $d-1$ factors. For convenience we henceforth assume $2 \leq g < d < m_0$ whenever necessary. We make also the assumption that $\sigma(x_i) \neq 1$ for all $1 \leq i \leq d$. This assumption can be realized e.g. by a first step in Plandowski's algorithm using a partial solution δ which sends a variable X to the empty word, if $\sigma(X) = 1$ and sends X to itself otherwise. Another choice to realize this assumption is by a guess in some preprocessing.
We have $\sigma(x_i) = w_0[\mathrm{l}(i), \mathrm{r}(i)]$ and $\sigma(\overline{x_i}) = w_0[\mathrm{r}(i), \mathrm{l}(i)]$ for $1 \leq i \leq d$. By our assumption, the interval $[\mathrm{l}(i), \mathrm{r}(i)]$ is positive. Let us consider a pair (i, j) such that $i, j \in 1, \ldots, d$ and $x_i = x_j$ or $x_i = \overline{x_j}$. For $\mu, \nu \in \{0, \ldots, \mathrm{r}(i) - \mathrm{l}(i)\}$ we define a relation $\sim$ by:

$$\begin{aligned}
[\mathrm{l}(i)+\mu, \mathrm{l}(i)+\nu] &\sim [\mathrm{l}(j)+\mu, \mathrm{l}(j)+\nu], \text{ if } x_i = x_j, \\
[\mathrm{l}(i)+\mu, \mathrm{l}(i)+\nu] &\sim [\mathrm{r}(j)-\mu, \mathrm{r}(j)-\nu], \text{ if } x_i = \overline{x_j}.
\end{aligned}$$

Note that $\sim$ is a symmetric relation. Moreover, $[\alpha, \beta] \sim [\alpha', \beta']$ implies both $[\beta, \alpha] \sim [\beta', \alpha']$ and $w_0[\alpha, \beta] = w_0[\alpha', \beta']$. By $\approx$ we denote the reflexive and transitive closure of $\sim$. Then $\approx$ is an equivalence relation and again, $[\alpha, \beta] \approx [\alpha', \beta']$ implies both $[\beta, \alpha] \approx [\beta', \alpha']$ and $w_0[\alpha, \beta] = w_0[\alpha', \beta']$.
Next we define the notion of *free interval*. An interval $[\alpha, \beta]$ is called *free*, if whenever $[\alpha, \beta] \approx [\alpha', \beta']$, then there is no cut γ' with $\min\{\alpha', \beta'\} < \gamma' < \max\{\alpha', \beta'\}$. Clearly, the set of free intervals is closed under involution, i.e., if $[\alpha, \beta]$ is free, then $[\beta, \alpha]$ is free, too. It is also clear that $[\alpha, \beta]$ is free whenever $|\beta - \alpha| \leq 1$.

Example 27 *The last equation in Example 23, namely*

$$aX\overline{X}\bar{a} = Y\bar{b}Y\bar{a}b\overline{Y},$$

has a solution which yields the word

$$w_0 = \overset{0}{|}\ a\ \overset{1}{|}\ bc\bar{c}\bar{b}\ \overset{5}{|}\ \bar{b}\ \overset{6}{|}\ abc\ \overset{9}{|}\ \bar{c}\bar{b}\ \overset{11}{|}\ \bar{a}\ \overset{12}{|}\ b\ \overset{13}{|}\ bc\bar{c}\bar{b}\ \overset{17}{|}\ \bar{a}\ \overset{18}{|}\ .$$

The set of cuts is shown by the bars. The intervals $[1, 5]$, $[13, 17]$, *and* $[6, 9]$ *are not free, since* $[1, 5] \approx [17, 13] \approx [7, 11]$ *and* $[6, 9] \approx [0, 3]$ *and* $[7, 11]$, $[0, 3]$ *contain cuts. There is only one equivalence class of free intervals of length longer than 1 (up to involution), which is given by* $[1, 3] \sim [17, 15] \sim [7, 9] \sim [11, 9] \sim [5, 3] \sim [13, 15]$.

The next lemma says that subintervals of free intervals are free again.

Lemma 28 *Let* $[\alpha, \beta]$ *be a free interval and* μ, ν *such that* $\min\{\alpha, \beta\} \leq \mu, \nu \leq \max\{\alpha, \beta\}$. *Then the interval* $[\mu, \nu]$ *is also free.*

Proof. We may assume that $\alpha \leq \mu < \nu \leq \beta$. By contradiction assume that $[\mu, \nu]$ is not free. Then there is some $k \geq 0$ and some cut γ' such that

$$[\mu, \nu] = [\mu_0, \nu_0] \sim [\mu_1, \nu_1] \sim \cdots \sim [\mu_k, \nu_k]$$

with $\min\{\mu_k, \nu_k\} < \gamma' < \max\{\mu_k, \nu_k\}$. If $k = 0$, then we have a immediate contradiction. For $k \geq 1$ the relation $[\mu, \nu] \sim [\mu_1, \nu_1]$ is due to some pair x_i, x_j with $x_i = x_j$ or $x_i = \overline{x_j}$. Since $[\alpha, \beta]$ contains no cut, we can use the same pair to find an interval $[\alpha_1, \beta_1]$ such that $[\alpha, \beta] \sim [\alpha_1, \beta_1]$ and $\mu_1, \nu_1 \in \{\min\{\alpha_1, \beta_1\}, \ldots, \max\{\alpha_1, \beta_1\}\}$. Using induction on k we see that $[\alpha_1, \beta_1]$ cannot be free. A contradiction, because then $[\alpha, \beta]$ is not free. $\square$

Next we introduce the notion of *implicit cut* for non-free intervals. For our purpose it is enough to define it for positive intervals. So, let $0 \leq \alpha < \beta \leq m_0$ such that $[\alpha, \beta]$ is not free. A position γ with $\alpha < \gamma < \beta$ is called an *implicit cut* of $[\alpha, \beta]$, if we meet the following situation. There is a cut γ' and an interval $[\alpha', \beta']$ such that

$$\begin{aligned} \min\{\alpha', \beta'\} &< \gamma' < \max\{\alpha', \beta'\}, \\ [\alpha, \beta] &\approx [\alpha', \beta'], \\ \gamma - \alpha &= |\gamma' - \alpha'|. \end{aligned}$$

The following observation will be used throughout. If we have $\alpha \leq \mu < \gamma < \nu \leq \beta$ and γ is an implicit cut of $[\alpha, \beta]$, then γ is also an implicit cut of $[\mu, \nu]$. In particular, neither $[\mu, \nu]$ nor $[\nu, \mu]$ is a free interval.[5]

[5]However, if γ is an implicit cut of $[\mu, \nu]$, then it might happen that γ is no implicit cut of $[\alpha, \beta]$, although $[\alpha, \beta]$ is certainly not free.

Lemma 29 *Let* $0 \leq \alpha \leq \alpha' < \beta \leq \beta' \leq m_0$ *such that* $[\alpha, \beta]$ *and* $[\alpha', \beta']$ *are free intervals. Then the interval* $[\alpha, \beta']$ *is free, too.*

Proof. Assume by contradiction that $[\alpha, \beta']$ is not free. Then it contains an implicit cut γ with $\alpha < \gamma < \beta'$. By the observation above: If $\gamma < \beta$, then γ is an implicit cut of $[\alpha, \beta]$ and $[\alpha, \beta]$ is not free. Otherwise, $\alpha' < \gamma$ and α', β' is not free. □

We now consider the maximal elements. A free interval $[\alpha, \beta]$ is called *maximal free*, if there is no free interval $[\alpha', \beta']$ such that both $\alpha' \leq \min\{\alpha, \beta\} \leq \max\{\alpha, \beta\} \leq \beta'$ and $\beta' - \alpha' > |\beta - \alpha|$. With this notion Lemma 29 states that maximal free intervals do not overlap.

Lemma 30 *Let* $[\alpha, \beta]$ *be a maximal free interval. Then there are intervals* $[\gamma, \delta]$ *and* $[\gamma', \delta']$ *such that* $[\alpha, \beta] \approx [\gamma, \delta] \approx [\gamma', \delta']$ *and* γ *and* δ' *are cuts.*

Proof. We may assume that $[\alpha, \beta]$ is a positive interval, i.e., $\alpha < \beta$. We show the existence of $[\gamma, \delta]$ where $[\alpha, \beta] \approx [\gamma, \delta]$ and γ is a cut. The existence of $[\gamma', \delta']$ where $[\alpha, \beta] \approx [\gamma', \delta']$ and δ' is a cut follows by a symmetric argument. If $\alpha = 0$, then α itself is a cut and we can choose $\delta = \beta$. Hence let $1 \leq \alpha$ and consider the positive interval $[\alpha - 1, \beta]$. This interval is not free, but the only possible position for an implicit cut is α. Thus for some cut γ we have $[\alpha - 1, \beta] \approx [\alpha', \beta']$ with $\min\{\alpha', \beta'\} < \gamma < \max\{\alpha', \beta'\}$ and $|\gamma - \alpha'| = 1$. A simple reflection shows that we have $[\alpha - 1, \alpha] \approx [\alpha', \gamma]$ and $[\alpha, \beta] \approx [\gamma, \beta']$. Hence we can choose $\delta = \beta'$. □

Proposition 31 *Let* Γ *be the set of words* $w \in \Gamma_0^*$ *such that there is a maximal free interval* $[\alpha, \beta]$ *with* $w = w_0[\alpha, \beta]$. *Then* Γ *is a subset of* Γ_0^+ *of size at most* $2d - 2$. *The set* Γ *is closed under involution.*

Proof. Let $[\alpha, \beta]$ be maximal free. Then $|\beta - \alpha| \geq 1$ and $[\beta, \alpha]$ is maximal free, too. Hence $\Gamma \subseteq \Gamma_0^+$ and Γ is closed under involution. By Lemma 30 we may assume that α is a cut. Say $\alpha < \beta$. Then $\alpha \neq m_0$ and there is no other maximal free interval $[\alpha, \beta']$ with $\alpha < \beta'$ because of Lemma 29. Hence there are at most $d - 1$ such intervals $[\alpha, \beta]$. Symmetrically, there are at most $d - 1$ maximal free intervals $[\alpha, \beta]$ where $\beta < \alpha$ and α is a cut. □

For a moment let $\Gamma_0' = \Gamma_0 \cup \Gamma$ where $\Gamma \subseteq \Gamma_0^+$ is the set defined in Proposition 31. The inclusion $\Gamma_0' \subseteq \Gamma_0^+$ defines a natural projection $\pi : \Gamma_0' \to \Gamma_0^*$ and a mapping $h_0' : \Gamma_0' \to M$ by $h_0' = h_0 \pi$. Consider the equation with constraints $\pi^*(E)$, this is a node in the search graph, because the size of Γ is linear in d.

The reason to switch from Γ_0 to Γ'_0 is that, due to the constraints, the word w_0 may have long free intervals, even in a minimal solution. Over Γ'_0 long free intervals can be avoided. Formally, we replace w_0 by a solution w'_0 where $w'_0 \in \Gamma^*$. The definition of w'_0 is based on a factorization of w_0 in maximal free intervals. There is a unique sequence $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k = m_0$ such that $[\alpha_{i-1}, \alpha_i]$ is a maximal free interval for each $1 \leq i \leq k$ and

$$w_0 = w_0[\alpha_0, \alpha_i] \cdots w_0[\alpha_{k-1}, \alpha_k].$$

Note that all cuts occur as some α_p, therefore we can think of the factors $w_0[\alpha_{i-1}, \alpha_i]$ as letters in Γ for $1 \leq i \leq k$. Moreover, all constants which appear in $L_0 R_0$ are elements of Γ. We replace w_0 by the word $w'_0 \in \Gamma^*$. Then we can define $\sigma : \Omega \to \Gamma^*$ such that both $\sigma(L_0) = \sigma(R_0) = w'_0$ and $\rho_0 = h'_0\sigma$. In other terms, σ is a solution of $\pi^*(E_0)$. We have $w_0 = \pi(w'_0)$ and $\exp(w'_0) \leq \exp(w_0)$. The crucial point is that w'_0 has no long free intervals anymore. With respect to w'_0 and Γ'_0 all maximal free intervals have length exactly one.

In the next step we show that we can reduce the alphabet of constants to be Γ. The inclusion of Γ in Γ'_0 defines an admissible base change $\beta : \Gamma \to \Gamma'_0$. Consider $E'_0 = (\Gamma, h, \Omega_0, \rho_0; L_0 = R_0)$ where h is the restriction of the mapping h'_0. Then we have $\pi^*(E_0) = \beta_*(E'_0)$. The search graph contains an arc from E_0 to E'_0, since we may choose δ to be the identity. The equation with constraints E'_0 has a solution σ with $\sigma(L_0) = w'_0$ and $\exp(w'_0) \leq \exp(w_0)$. In order to avoid too many notations we identify E_0 and E'_0, hence we also assume $w_0 = w'_0$. However, as a reminder that we have changed the alphabet of constants (recall that some words became letters), we prefer to use the notation Γ rather than Γ_0. Thus, in what follows we shall make the following assumptions:

$$\begin{aligned}
E_0 &= (\Gamma, h, \Omega_0, \rho_0; L_0 = R_0), \\
L_0 &= x_1 \cdots x_g \text{ and } g \geq 2, \\
R_0 &= x_{g+1} \cdots x_d \text{ and } d > g, \\
|\Gamma| &\leq 2d - 2, \\
|\Omega_0| &\leq 2d, \\
M &\subseteq \mathbb{B}^{2n \times 2n}.
\end{aligned}$$

Moreover: All variables X occur in $L_0 R_0 \overline{L_0 R_0}$. There is a solution σ such that $w_0 = \sigma(L_0) = \sigma(R_0)$ with $\sigma(X_i) \neq 1$ for $1 \leq i \leq d$ and $\rho_0 = h\sigma = h_0\sigma$.

We have $|w_0| = m_0$ and $\exp(w_0) \in 2^{\mathcal{O}(d+n\log n)}$. All maximal free intervals have length exactly one, i.e., every positive interval $[\alpha, \beta]$ with $\beta - \alpha > 1$ contains an implicit cut.
It is because of the last sentence that we have worked out the details about free intervals. This difficulty is due to the constraints. Without them the reasoning would have been much simpler. But the good news are that from now on, the presence of constraints will not interfere very much.

Example 32 *Following Example 27, we use the same equation $aX\overline{X}\bar{a} = Y\bar{b}Y\bar{a}b\overline{Y}$ and we consider the solution w_0.*
The new solution is defined by replacing in w_0 each factor bc by a new letter d which represents a maximal free interval. The new w_0 has the form

$$w_0 = \overset{0}{|}\ a\ \overset{1}{|}\ d\bar{d}\ \overset{3}{|}\ \bar{b}\ \overset{4}{|}\ ad\ \overset{6}{|}\ \bar{d}\ \overset{7}{|}\ \bar{a}\ \overset{8}{|}\ b\ \overset{9}{|}\ d\bar{d}\ \overset{11}{|}\ \bar{a}\ \overset{12}{|}\ .$$

Now all maximal intervals have length one.

13 Critical Words and Blocks

In the following ℓ denotes an integer which varies between 1 and m_0. For each ℓ we define the set of critical words C_ℓ by

$$C_\ell = \{\, w_0[\gamma - \ell, \gamma + \ell], w_0[\gamma + \ell, \gamma - \ell] \mid \gamma \,\text{is a cut and}\, \ell \leq \gamma \leq m_0 - \ell \,\}.$$

We have $1 \leq |C_\ell| \leq 2d - 4$ and C_ℓ is closed under involution. Each word $u \in C_\ell$ has length 2ℓ, it can be written in the form $u = u_1 u_2$ with $|u_1| = |u_2| = \ell$. Then u_1 (resp. $\overline{u_2}$) appears as a suffix, left of some cut and u_2 (resp. $\overline{u_1}$) appears as a prefix, right of the same cut.
A triple $(u, w, v) \in (\{1\} \cup \Gamma^\ell) \times \Gamma^+ \times (\{1\} \cup \Gamma^\ell)$ is called a *block* if first, up to a possible prefix or suffix no other factor of the word uwv is a critical word, second, $u \neq 1$ if and only if a prefix of uwv of length 2ℓ belongs to C_ℓ, and third, $v \neq 1$ if and only if a suffix of uwv of length 2ℓ belongs to C_ℓ. The set of blocks is denoted by B_ℓ. It is viewed (as a possibly infinite) alphabet where the involution is defined by $\overline{(u, w, v)} = (\overline{v}, \overline{w}, \overline{u})$. We can define a homomorphism $\pi_\ell : B_\ell^* \to \Gamma^*$ by $\pi_\ell(u, w, v) = w \in \Gamma^+$ being extended to a projection $\pi_\ell : (B_\ell \cup \Gamma)^* \to \Gamma^*$ by leaving Γ invariant. We define $h_\ell : (B_\ell \cup \Gamma) \to M$ by $h_\ell = h\pi_\ell$. In the following we shall consider finite subsets $\Gamma_\ell \subseteq B_\ell \cup \Gamma$ which are closed under involution. Then

by $\pi_\ell : \Gamma_\ell^* \to \Gamma^*$ and $h_\ell : \Gamma_\ell^* \to M$ we understand the restrictions of the respective homomorphisms.

For every non-empty word $w \in \Gamma^+$ we define its *ℓ-factorization* as follows. We write

$$F_\ell(w) = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k) \in B_\ell^+$$

such that $w = w_1 \cdots w_k$ and for $1 \leq i \leq k$ the following conditions are satisfied:

- v_i is a prefix of $w_{i+1} \cdots w_k$,
- $v_i = 1$ if and only if $i = k$,
- u_i is a suffix of $w_1 \cdots w_{i-1}$,
- $u_i = 1$ if and only if $i = 1$.

Note that the ℓ-factorization of a word w is unique. For $k \geq 2$ we have $|w_1| \geq \ell$ and $|w_k| \geq \ell$, but all other w_i may be short. If no critical word appears as a factor of w, then $F_\ell(w) = (1, w, 1)$. In particular, this is the case for $|w| < 2\ell$. If we have $w = puvq$ with $|u| = |v| = \ell$ and $uv \in C_\ell$, then there is a unique $i \in \{1, \ldots, k-1\}$ such that $u = u_{i+1}$, $v = v_i$, and $pu = w_1 \cdots w_i$, $vq = w_{i+1} \cdots w_k$. Thus, $F_\ell(w)$ contains a factor $(u_i, w_i, v)(u, w_{i+1}, v_{i+1})$ where v is a prefix of $w_{i+1}v_{i+1}$ and u is a suffix of $u_i w_i$. For example, the ℓ-factorization of $uv \in C_\ell$ with $|u| = |v| = \ell$ is

$$F_\ell(uv) = (1, u, v)(u, v, 1).$$

We define the head, body, and tail of a word w based on its ℓ-factorization

$$F_\ell(w) = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k)$$

in B_ℓ^* and Γ^* as follows:

$$\begin{aligned}
\mathrm{Head}_\ell(w) &= (u_1, w_1, v_1) \in B_\ell, \\
\mathrm{head}_\ell(w) &= w_1 \in \Gamma^+, \\
\mathrm{Body}_\ell(w) &= (u_2, w_2, v_2) \cdots (u_{k-1}, w_{k-1}, v_{k-1}) \in B_\ell^*, \\
\mathrm{body}_\ell(w) &= w_2 \cdots w_{k-1} \in \Gamma^*, \\
\mathrm{Tail}_\ell(w) &= (u_k, w_k, v_k) \in B_\ell, \\
\mathrm{tail}_\ell(w) &= w_k \in \Gamma^+.
\end{aligned}$$

For $k \geq 2$ (in particular, if $\mathrm{body}_\ell(w) \neq 1$) we have

$$\begin{aligned} F_\ell(w) &= \mathrm{Head}_\ell(w)\mathrm{Body}_\ell(w)\mathrm{Tail}_\ell(w), \\ w &= \mathrm{head}_\ell(w)\mathrm{body}_\ell(w)\mathrm{tail}_\ell(w). \end{aligned}$$

Moreover, u_2 is a suffix of w_1 and v_{k-1} is a prefix of w_k.
Assume $\mathrm{body}_\ell(w) \neq 1$ and let $u, v \in \Gamma^*$ be any words. Then we can view w in the context uwv and $\mathrm{Body}_\ell(w)$ appears as a proper factor in the ℓ-factorization of uwv. More precisely, let

$$F_\ell(uwv) = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k).$$

Then there are unique $1 \leq p < q \leq k$ such that:

$$\begin{aligned} F_\ell(uwv) &= (u_1, w_1, v_1) \cdots (u_p, w_p, v_p)\mathrm{Body}_\ell(w)(u_q, w_q, v_q) \cdots (u_k, w_k, v_k), \\ w_1 \cdots w_p &= u\,\mathrm{head}_\ell(w), \\ w_q \cdots w_k &= \mathrm{tail}_\ell(w)v \end{aligned}$$

Finally, we note that the above definitions are compatible with the involution. We have $F_\ell(\overline{w}) = \overline{F_\ell(w)}$, $\mathrm{Head}_\ell(\overline{w}) = \overline{\mathrm{Tail}_\ell(w)}$, and $\mathrm{Body}_\ell(\overline{w}) = \overline{\mathrm{Body}_\ell(w)}$.

14 The ℓ-Transformation

Our equation with constraints is $E_0 = (\Gamma, h, \Omega_0, \rho_0; x_1 \cdots x_g = x_{g+1} \cdots x_d)$. We start with the ℓ-factorization of $w_0 = \sigma(x_1 \cdots x_g) = \sigma(x_{g+1} \cdots x_d)$. Let

$$F_\ell(w_0) = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k).$$

A sequence $S = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ with $1 \leq p \leq q \leq k$ is called an *ℓ-factor*. We say that S is a *cover* of a positive interval $[\alpha, \beta]$, if both $|w_1 \cdots w_{p-1}| \leq \alpha$ and $|w_{q+1} \cdots w_k| \leq m_0 - \beta$. Thus, $w_0[\alpha, \beta]$ becomes a factor of $w_p \cdots w_q$. It is a *minimal cover*, if neither $(u_{p+1}, w_{p+1}, v_{p+1}) \cdots (u_q, w_q, v_q)$ nor $(u_p, w_p, v_p) \cdots (u_{q-1}, w_{q-1}, v_{q-1})$ is a cover of $[\alpha, \beta]$. The minimal cover exists and it is unique.
We let $\Omega_\ell = \{ X \in \Omega_0 \mid \mathrm{body}_\ell(\sigma(X)) \neq 1 \}$, and we are going to define a new left-hand side $L_\ell \in (B_\ell \cup \Omega_\ell)^*$ and a new right-hand side $R_\ell \in (B_\ell \cup \Omega_\ell)^*$. For L_ℓ we consider those $1 \leq i \leq g$ where $\mathrm{body}_\ell(\sigma(x_i)) \neq 1$. Note that this implies $x_i \in \Omega_\ell$ since $\ell \geq 1$ and then the body of a constant is always empty.

Recall the definition of $\mathrm{l}(i)$ and $\mathrm{r}(i)$, and define $\alpha = \mathrm{l}(i) + |\mathrm{head}_\ell(\sigma(x_i))|$ and $\beta = \mathrm{r}(i) - |\mathrm{tail}_\ell(\sigma(x_i))|$. Then we have $w_0[\alpha, \beta] = \mathrm{body}_\ell(\sigma(x_i))$. Next consider the ℓ-factor $S_i = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ which is the minimal cover of $[\alpha, \beta]$. Then we have $1 < p \leq q < k$ and $w_p \cdots w_q = w_0[\alpha, \beta] = \mathrm{body}_\ell(\sigma(x_i))$. The definition of S_i depends only on x_i, but not on the choice of the index i.
We replace the ℓ-factor S_i in $F_\ell(w_0)$ by the variable x_i. Having done this for all $1 \leq i \leq g$ with $\mathrm{body}_\ell(\sigma(x_i)) \neq 1$ we obtain the left-hand side $L_\ell \in (B_\ell \cup \Omega_\ell)^*$ of the ℓ-transformation E_ℓ. For R_ℓ we proceed analogously by replacing those ℓ-factors S_i where $\mathrm{body}_\ell(\sigma(x_i)) \neq 1$ and $g + 1 \leq i \leq d$.
For E_ℓ we cannot use the alphabet B_ℓ, because it might be too large or even infinite. Therefore we let $\Gamma_{\ell'}$ be the smallest subset of B_ℓ which is closed under involution and which satisfies $L_\ell R_\ell \in (\Gamma_{\ell'} \cup \Omega_\ell)^*$. We let $\Gamma_\ell = \Gamma_{\ell'} \cup \Gamma$. The projection $\pi_\ell : \Gamma_\ell^* \to \Gamma^*$ and the mapping $h_\ell : \Gamma_\ell \to M$ are defined by the restriction of $\pi_\ell : B_\ell \to \Gamma^*$, $\pi_\ell(u, w, v) = w$ and $h_\ell(u, w, v) = h(w) \in M$ and by $\pi_\ell(a) = a$ and $h_\ell(a) = h(a)$ for $a \in \Gamma$.
Finally, we define the mapping $\rho_\ell : \Omega_\ell \to M$ by $\rho_\ell(X) = h(\mathrm{body}_\ell(\sigma(X)))$. This completes the definition of the ℓ-transformation:

$$E_\ell = (\Gamma_\ell, h_\ell, \Omega_\ell, \rho_\ell; L_\ell = R_\ell).$$

Remark 33 *One can verify that* $\sigma_\ell : \Omega_\ell \to \Gamma_\ell^*$, $\sigma_\ell(X) = \varphi_\ell(\mathrm{Body}_\ell(\sigma(X)))$ *defines a solution of* E_ℓ*, where* φ_ℓ *is the identity on* Γ_ℓ *and* π_ℓ *on* $B_\ell \setminus \Gamma_{\ell'}$*. Although, up to the trivial case* $\ell = m_0$*, we make no explicit use of this fact.*

Example 34 *We continue with our example* $aX\overline{X}\bar{a} = Y\bar{b}Y\bar{a}b\overline{Y}$ *and the solution* σ *which has been given by*

$$w_0 = |\ a \mid d\bar{d} \mid \bar{b} \mid ad \mid \bar{d} \mid \bar{a} \mid b \mid d\bar{d} \mid \bar{a}\ |,$$

where the bars show the cuts.
Up to involution, the set C_1 *is given by* $\{ad, bd, \bar{a}b, d\bar{d}\}$ *and* C_2 *is given by* $\{d\bar{d}\bar{b}a, \bar{d}\bar{b}ad, ad\bar{d}\bar{a}, d\bar{d}\bar{a}b\}$*. The 1-factorization of* w_0 *can be obtained letter by letter. The 2-factorization of* w_0 *is given by the following sequence:*

$$(1, ad\bar{d}, \bar{b}a)(d\bar{d}, \bar{b}, ad)(\bar{d}\bar{b}, ad, \bar{d}\bar{a})(ad, \bar{d}, \bar{a}b)(d\bar{d}, \bar{a}, bd)(\bar{d}\bar{a}, b, d\bar{d})(\bar{a}b, d\bar{d}a, 1).$$

Recall $\sigma(X) = d\bar{d}\bar{b}ad$ *and* $\sigma(Y) = ad\bar{d}$*. Hence their 2-factorizations are* $(1, ad\bar{d}, \bar{b}a)(d\bar{d}, \bar{b}, ad)(\bar{d}\bar{b}, ad, 1)$ *and* $(1, ad\bar{d}, 1)$*, respectively.*

By renaming letters, the 2-factorization of w_0 *becomes* $a\bar{b}cdeb\bar{a}$ *and the equation* E *reduces to* $E_2 : aXcdeX\bar{a} = a\bar{b}cdeb\bar{a}$ *since the body of* $\sigma(Y)$ *is empty. The reader can check that the 3-factorization of* w_0 *after renaming is the very same word as the 2-factorization, but the 3-factorization of* $\sigma(X)$ *is now one letter,* $(1, d\bar{d}\bar{b}ad, 1)$*, so* E_3 *becomes a trivial equation. Plandowski's algorithm will return* true *at this stage.*

Remark 35 *i) In the extreme case* $\ell = m_0$*, the* ℓ*-transformation becomes trivial. Let* $a = (1, w_0, 1)$*. Then* $\overline{a} = (1, \overline{w_0}, 1)$ *and* $\Gamma_{m_0} = \{a, \overline{a}\} \cup \Gamma$*. Moreover, we have* $L_{m_0} = R_{m_0} = a$*, and* $h_{m_0}(a) = h(w_0) \in M$*. Since* $\Omega_{m_0} = \emptyset$*, the equation with constraints* E_{m_0} *has trivially a solution. It is clear that* E_{m_0} *is a node in the search graph, and if we reach* E_{m_0}*, then the algorithm will return* true*.*
ii) The other extreme case is $\ell = 1$*. The situation again is simple, but the precise definition is technically more involved. Consider a block* (u, w, v) *which appears in* $F_1(w_0)$*. Then* $w = w_0[\alpha, \beta]$ *for some* $\beta - \alpha \geq 1$*. We cannot have* $\beta - \alpha \geq 2$*, because then* $[\alpha, \beta]$ *would have an implicit cut* γ*, but* $w_0[\gamma - 1, \gamma + 1] \in C_1$ *and no critical word is a factor of* w*. An immediate consequence is* $|\Gamma_1| \leq (|\Gamma| + 1)^3 \in \mathcal{O}(d^3)$*. Let* $X \in \Omega_0$*. Then* $\mathrm{Body}_1(\sigma(X)) \neq 1$ *if and only if* $|\sigma(X)| \geq 3$*. Thus, for* $X \in \Omega_1$ *we have* $\sigma(X) = bcu = vde$ *with* $b, c, d, e \in \Gamma$ *and* $u, v \in \Gamma^+$*. It follows:*

$$F_1(\sigma(X)) = (1, b, c)(b, c, v_2) \cdots (u_{|v|+1}, d, e)(d, e, 1).$$

For example, for $|v| = 1$ *this means* $b = u_{|v|+1}$*,* $c = d$*, and* $v_2 = e$*.*
We can describe $L_1 \in \Gamma_1^*$ *as follows:*
For $1 \leq i \leq g$ *let* $w_i = \sigma(x_i)$ *and* a_i *the last letter of* $\sigma(x_{i-1})$ *if* $i > 1$ *and* $a_1 = 1$*. Let* f_i *the first letter of* $\sigma(x_{i+1})$ *if* $i < g$ *and* $f_g = 1$*. Let* b_i *the first letter of* w_i *and* e_i *the last letter of* w_i*.*
For $|w_i| = 1$ *we replace* x_i *by the* 1*-factor* (a_i, b_i, f_i)*.*
For $|w_i| = 2$ *we replace* x_i *by the* 1*-factor* $(a_i, b_i, e_i)(b_i, e_i, f_i)$*.*
For $|w_i| \geq 3$ *we let* c_i *be the second letter of* w_i *and* d_i *its second last. In this case we replace* x_i *by* $(a_i, b_i, c_i)x_i(d_i, e_i, f_i)$*.*
The definition of R_1 *is analogous. Thus, we obtain* $|L_1R_1| \leq 3|L_0R_0| = 3d$*, and* E_1 *is admissible. We also see that there was an overestimation of the size of* $|\Gamma_1|$*. For each* x_i *we need at most two constants together with their involutions. Since* Γ_1 *contains also* Γ*, we obtain* $|\Gamma_1| \leq 6d$*.*

By the remark above, E_1 and E_{m_0} are admissible and hence nodes of the search graph. The goal is to reach E_{m_0} via E_1 when starting with E_0. For the moment it is even not clear that the ℓ-transformations with $1 < \ell < m_0$ belongs to the search graph. We prove this statement in the next section.

15 The ℓ-transformation E_ℓ is admissible

Proposition 36 *There is a polynomial p (of degree at most 4) such that each E_ℓ is admissible for all $\ell \geq 1$.*

Proof. It is enough to show that L_ℓ and R_ℓ can be represented by exponential expressions of size $\mathcal{O}(d^2(d+n\log n))$. Then Γ_ℓ can have size at most $\mathcal{O}(d^2(d+n\log n))$ and the assertion follows. We will estimate the size of an exponential expression for L_ℓ, only.
We start again with the ℓ-transformation of

$$F_\ell(w_0) = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k).$$

If k is small there is nothing to do since $|L_\ell| \leq |F_\ell(w_0)|$. An easy reflection shows that $|L_\ell|$ can become large, only if there is some $1 \leq i \leq g$ such that $\mathrm{head}_\ell(\sigma(x_i))$ or $\mathrm{tail}_\ell(\sigma(x_i))$ is long. By symmetry we treat the case $\mathrm{head}_\ell(\sigma(x_i))$ only and we fix some notation. We let $1 \leq i \leq g$, $\alpha = \mathrm{l}(i)$, and $\beta = \alpha + |\mathrm{head}_\ell(\sigma(x_i))|$. Let

$$(u_{p-1}, w_{p-1}, v_{p-1}) \cdots (u_{q+1}, w_{q+1}, v_{q+1})$$

be a minimal cover of $[\alpha, \beta]$. We may assume that $q-p$ is large. It is enough to find an exponential expression for the ℓ-factor

$$(u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$$

having size in $\mathcal{O}(d(d+n\log n))$, because we want the whole expression to have size in $\mathcal{O}(d^2(d+n\log n))$.
Note that $w_p \cdots w_q$ is a proper factor of $\mathrm{head}_\ell(\sigma(x_i))$. Hence no critical word of C_ℓ can appear as a factor inside $w_p \cdots w_q$. This means there is some $p \leq s \leq q$ such that both $|w_p \cdots w_{s-1}| < \ell$ and $|w_{s+1} \cdots w_q| < \ell$. Indeed, if $|w_p \cdots w_{q-1}| < \ell$, then we choose $s = q$. Otherwise we let $p \leq s \leq q$ be minimal such that $|w_p \cdots w_s| \geq \ell$. Then $|w_{s+1} \cdots w_q| \geq \ell$ is impossible because $u_{s+1}v_s \in C_\ell$ would appear as a factor in $w_p \cdots w_q$. We can write

$$(u_p, w_p, v_p) \cdots (u_q, w_q, v_q) = S_1(u_s, w_s, v_s)S_2;$$

and since $(u_s, w_s, v_s) \in \Gamma_\ell$ is a letter, it is enough to find exponential expressions for S_i, $i = 1, 2$, of size $\mathcal{O}(d(d + n \log n))$ each. As a conclusion it is enough to prove the following lemma. □

The statement of the next lemma is slightly more general as we need it above. There we need the lemma for $c = 1$, but later we will apply the lemma with values $c \le 32d$.

Lemma 37 *Let $c > 0$ be a number and*

$$S = (u_1, w_1, v_1) \cdots (u_k, w_k, v_k) \in B_\ell^*$$

be a sequence which appears as some ℓ-factor in $F_\ell(w_0)$. If we have $k \le 3$ or $|w_2 \cdots w_{k-1}| \le c\ell$, then we can represent the sequence by some exponential expression of size $\mathcal{O}(cd(d + n \log n))$.

Proof. We show that there is an exponential expression of size $\mathcal{O}(d(d + n \log n))$ under the assumption $|w_1 \cdots w_k| < \ell$. This is enough, because we always can write S as $a_0 S_1 a_1 \cdots S_{c'} a_{c'}$, where $c' \le c$, the a_i are letters, and each S_i satisfies the assumption. Note that the assumption implies $u_1 \ne 1 \ne v_k$ and we may define u_{k+1} as the suffix of length ℓ of $u_1 w_1 \cdots w_k$. For $1 \le i \le k$ let $z_i = u_{i+1} v_i$. Then $z_i \in C_\ell$ is a critical word which appears as a factor in $z = u_1 w_1 w_2 \cdots w_k v_k$. If the words z_i, $1 \le i < k$ are pairwise different, then $k - 1 \le |C_\ell| \in \mathcal{O}(d)$ and we are done. Hence we may assume that there are repetitions. Let j be the smallest index such that a critical word is seen for the second time and let $i < j$ be the first appearance of z_j. This means for $1 \le i < j$ the words $z_1, \cdots, z_{j-1}$ are pairwise different and $z_i = z_j$. Now, $|w_1 \cdots w_k| < \ell$ and $|z_i| = 2\ell$, hence z_i and z_j overlap in z. We can choose r maximal such that $u_1 w_1 \cdots w_i (w_{i+1} \cdots w_j)^r v_j$ is a prefix of the word z. (Note that the last factor v_j insures that the prefix ends with z_j). For some index $s > j$ we can write

$$z = u_1 w_1 \cdots w_i (w_{i+1} \cdots w_j)^r w_s \cdots w_k v_k.$$

We claim that $z_i \notin \{z_s, \ldots, z_k\}$. Indeed, let t be maximal such that $z_i = z_t$ and assume that $j \ne t$. Then both $|w_{i+1} \cdots w_j|$ and $|w_{j+1} \cdots w_t|$ are periods of z_i, but $|w_{i+1} \cdots w_t| \le |z|$. Hence by Fine and Wilf's Theorem [16] we obtain that the greatest common divisor of $|w_{i+1} \cdots w_j|$ and $|w_{j+1} \cdots w_t|$ is a period, too. Due to the definition of an ℓ-factorization (z_j was the first repetition) the length $|w_{j+1} \cdots w_t|$ is therefore a multiple of $|w_{i+1} \cdots w_j|$ and we must have $t = s - 1$. This shows the claim. Moreover, we have

$$\begin{aligned} & (u_1, w_1, v_1) \cdots (u_k, w_k, v_k) \\ = \ & (u_1, w_1, v_1) \cdots (u_i, w_i, v_i)[(u_{i+1}, w_{i+1}, v_{i+1}) \cdots (u_j, w_j, v_j)]^r \, S' \end{aligned}$$

where $S' = (u_s, w_s, v_s) \cdots (u_k, w_k, v_k)$ for $s = i + 1 + r(j - i)$. We have $r \leq \exp(w_0)$, hence $r \in 2^{\mathcal{O}(d+n \log n)}$. It follows that

$$(u_1, w_1, v_1) \cdots (u_i, w_i, v_i)[(u_{i+1}, w_{i+1}, v_{i+1}) \cdots (u_j, w_j, v_j)]^r$$

is an exponential expression of size $j + \log(r) \in \mathcal{O}(d + n \log n)$. More precisely, for some suitable constant $\widetilde{c}$ its size is at most $\widetilde{c}(d + n \log n)$. The constant $\widetilde{c}$ depends only on the constant which is hidden when writing $\exp(w_0) \in 2^{\mathcal{O}(d+n \log n)}$. By induction on the size of the set $\{z_1, \ldots, z_k\}$ we may assume that $S' = (u_s, w_s, v_s) \cdots (u_k, w_k, v_k)$ has an exponential expression of size at most $|\{z_s, \ldots, z_k\}|\widetilde{c}(d + n)$. Hence the exponential expression for S has size at most

$$\widetilde{c}(d + n \log n) + |\{z_s, \ldots, z_k\}|\widetilde{c}(d + n \log n) \leq |\{z_1, \ldots, z_k\}|\widetilde{c}(d + n \log n).$$

Thus, the size is in $\mathcal{O}(d(d + n \log n))$. □

At this stage we know that all ℓ-transformations are admissible (with respect to some suitable polynomial of degree 4). Thus $E_1, \ldots, E_{m_0}$ are nodes of the search graph. Next we show that the search graph contains arcs $E_0 \to E_1$ and $E_\ell \to E_{\ell'}$ for $1 \leq \ell < \ell' \leq 2\ell$. Hence the graph contains a path (of logarithmic length in m_0) from E_0 to E_{m_0}. The non-deterministic procedure is able to find this path and on input E_0 Plandowski's algorithm gives the correct answer.

In order to establish the existence of arcs from E_ℓ to $E_{\ell'}$ for $0 \leq \ell < \ell' \leq \max\{1, 2\ell\}$ we shall define intermediate equations $E_{\ell,\ell'}$ such that there is an admissible base change β, a projection π, and a partial solution δ with

$$\delta_*(\pi^*(E_\ell)) \equiv E_{\ell,\ell'} \equiv \beta_*(E_{\ell'}).$$

16 The arc from E_0 to E_1

Recall the definition of $E_1 = (\Gamma_1, h_1, \Omega_1, \rho_1; L_1 = R_1)$. The letters of Γ_1 can be written either as (a, b, c) or as b with $a, c \in \Gamma \cup \{1\}$ and $b \in \Gamma$. We define

a projection which is used here as a base change $\beta : \Gamma_1 \to \Gamma$ by $\beta(a, b, c) = b$ and leaving the letters of Γ invariant. Clearly, $h_1 = h\beta$, and β defines an admissible base change. Define $E_{0,1} = \beta_*(E_1)$. Then we have $L_{0,1} = \beta(L_1)$ and $R_{0,1} = \beta(R_1)$ where $\beta : (\Gamma_1 \cup \Omega_1)^* \to (\Gamma \cup \Omega_1)^*$ is the extension with $\beta(X) = X$ for all $X \in \Omega_1$. We have $\Gamma_{0,1} = \Gamma$
It is now obvious how to define the partial solution $\delta : \Omega_0 \to \Gamma\Omega_1\Gamma \cup \Gamma^*$ such that $\delta_*(E_0) = E_{0,1}$. If $|\sigma(X)| \leq 2$, then we let $\delta(X) = \sigma(X)$. For $|\sigma(X)| \geq 3$ we write $\sigma(X) = aub$ with $a, b \in \Gamma$ and $u \in \Gamma^+$. Then we have $X \in \Omega_1 = \Omega_{0,1}$ and we define $\delta(X) = aXb$ and $\rho_{0,1}(X) = h(u)$. For $X \in \Omega_1$ we have $\rho_1(X) = h(\mathrm{body}_1(\sigma(X)))$, hence $\rho_{0,1} = \rho_1$, too. This shows that, indeed, $\delta_*(E_0) = \beta_*(E_1)$. Formally, we can write this as $\delta_*(\pi^*(E_0)) = \beta_*(E_1)$, where π is the identity. Hence there is an arc from E_0 to E_1.

17 The equations $E_{\ell,\ell'}$ for $1 \leq \ell < \ell' \leq 2\ell$

In this section we define for each $1 \leq \ell < \ell' \leq 2\ell$ an intermediate equation with constraints

$$\beta_*(E_{\ell'}) = E_{\ell,\ell'} = (\Gamma_{\ell,\ell'}, h_{\ell,\ell'}, \Omega_{\ell'}, \rho_{\ell'}; L_{\ell,\ell'} = R_{\ell,\ell'})$$

by some base change $\beta : \Gamma_{\ell'} \to (B_\ell \cup \Gamma)^*$, then we show that β is admissible. Recall $\Gamma \subseteq \Gamma_{\ell'} \subseteq B_{\ell'} \cup \Gamma$. The base change β leaves the letters of Γ invariant. Consider some $(u, w, v) \in \Gamma_{\ell'} \setminus \Gamma$. It is enough to define $\beta(u, w, v)$ or $\beta(\overline{v}, \overline{w}, \overline{u})$. Hence we may assume that (u, w, v) appears as a letter in the ℓ'-factorization $F_{\ell'}(w_0)$. Therefore we find a positive interval $[\alpha, \beta]$ such that $w = w_0[\alpha, \beta]$ and such that the following two conditions are satisfied:
1) We have $u = 1$ and $\alpha = 0$ or $|u| = \ell'$, $\alpha \geq \ell'$, and $u = w_0[\alpha - \ell', \alpha]$.
2) We have $v = 1$ and $\beta = m_0$ or $|v| = \ell'$, $\beta \leq m_0 - \ell'$, and $v = w_0[\beta, \beta + \ell']$.
Let $(u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ be the ℓ-factor which is the minimal cover of $[\alpha, \beta]$ with respect to the ℓ-factorization $F_\ell(w_0)$. Since $\ell \leq \ell'$ we have $w_p \cdots w_q = w$. Moreover, the word u_p is a suffix of u and v_q is a prefix of v. We define

$$\beta(u, w, v) = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q) \in B_\ell^+.$$

The definition does not depend on the choice of $[\alpha, \beta]$ as long as $0 \leq \alpha < \beta \leq m_0$ and 1) and 2) are satisfied. We have $\overline{\beta(u, w, v)} = \beta(\overline{v}, \overline{w}, \overline{u})$ and $h_\ell\beta = h_{\ell'}$. Now let $\Gamma_{\ell,\ell'} \subseteq B_\ell \cup \Gamma$ be the smallest subset such that $\beta(\Gamma_{\ell'}) \subseteq \Gamma^*_{\ell,\ell'}$. Then

$\Gamma_{\ell,\ell'}$ contains Γ and it is closed under involution (since $\Gamma_{\ell'}$ has this property). A crucial, but easy reflection shows that $\Gamma_\ell \subseteq \Gamma_{\ell,\ell'}$. This will become essential later.
We view β as a homomorphism $\beta : \Gamma^*_{\ell'} \to \Gamma^*_{\ell,\ell'}$ and define $E_{\ell,\ell'} = \beta_*(E_{\ell'})$. Let us show that β defines an admissible base change. Since $E_{\ell'}$ is already known to be admissible with respect to some polynomial of degree 4, it is enough to find some admissible exponential expression (again with respect to some polynomial of degree 4) for the ℓ-factor

$$\beta(u,w,v) = (u_p,w_p,v_p)\cdots(u_q,w_q,v_q)$$

where $(u,w,v) \in \Gamma_{\ell'} \setminus \Gamma$. We use the same notations as above. Thus, for some positive interval $[\alpha,\beta]$ we have $w_p \cdots w_q = w_0[\alpha,\beta]$, the word u is a suffix of $w_0[0,\alpha]$, and v is a prefix of $w_0[\beta,m_0]$. If $q-p$ is small, there is nothing to do. By Lemma 37 we may also assume that $\beta - \alpha > 32d\ell$. We are to define inductively a sequence of positions

$$\alpha = \alpha_0 < \alpha_1 < \cdots < \alpha_i < \cdots < \beta_i < \cdots < \beta_1 < \beta_0 = \beta.$$

Each time we let $W_i = w_0[\alpha_i,\beta_i]$. Thus, $W_0 = w_p \cdots w_q$. Assume that $W_i = w_0[\alpha_i,\beta_i]$ is already defined such that $\beta_i - \alpha_i \geq 2$. The interval $[\alpha_i,\beta_i]$ is not free. Hence, there is some implicit cut γ_i with $\alpha_i < \gamma_i < \beta_i$. The word W_i is a factor of w, hence no factor of W_i belongs to the set of critical words $C_{\ell'}$. This implies $\beta_i - \gamma_i < \ell'$ or $\gamma_i - \alpha_i < \ell'$. If we have $\beta_i - \gamma_i < \ell'$ then we let $\alpha_{i+1} = \alpha_i$ and $\beta_{i+1} = \gamma_i$. In the other case we let $\alpha_{i+1} = \gamma_i$ and $\beta_{i+1} = \beta_i$. Thus W_{i+1} is defined such that W_{i+1} is a proper factor of W_i with $|W_i| - |W_{i+1}| < \ell'$.
We need some additional book keeping. We define $r_i \in \{\mathrm{l},\mathrm{r}\}$ by $r_i = \mathrm{r}$ if $\beta_i = \beta_{i+1}$ and $r_i = \mathrm{l}$ otherwise (i.e., $\alpha_i = \alpha_{i+1}$). Furthermore the implicit cut γ_i corresponds to some real cut γ'_i and $\alpha'_i < \gamma'_i < \beta'_i$ such that $W_i = w_0[\alpha'_i,\beta'_i]$ or $W_i = w_0[\beta'_i,\alpha'_i]$. We define $s_i \in \{+,-\}$ by $s_i = +$ if $W_i = w_0[\alpha'_i,\beta'_i]$ and $s_i = -$ otherwise (in particular, $s_i = -$ implies $\overline{W_i} = w_0[\alpha'_i,\beta'_i]$). The triple (γ'_i, r_i, s_i) is denoted by $\gamma(i)$. There are at most $4(d-2)$ such triples and $\gamma(i)$ is defined whenever W_{i+1} is defined. We stop the induction procedure after the first repetition. Thus we find $0 \leq i < j < 4d$ such that $\gamma(i) = \gamma(j)$. We obtain a sequence $W_0, W_1, \ldots, W_i, \ldots, W_j$ where each word is a proper factor of the preceding one. We have $|W_0| - |W_j| < 4d\ell' \leq 8d\ell$ and due to $|W_0| > 32d\ell$ the sequence above really exists, moreover $|W_j| > 8d\ell$.
Next, we show that W_j has a non-trivial overlap with itself. We treat the case $\gamma(i) = \gamma(j) = (\gamma,\mathrm{r},+)$ only. The other three cases $(\gamma,\mathrm{r},-)$, $(\gamma,\mathrm{l},+)$,

and $(\gamma, \mathrm{l}, -)$ can be treated analogously. For some $\alpha' < \gamma < \beta'$ we have $W_i = w_0[\alpha', \beta']$ and $W_{i+1} = w_0[\gamma, \beta']$. Thus, for some $\gamma \leq \mu < \nu \leq \beta'$ we have $W_j = w_0[\mu, \nu]$ and we can assume that $\mu - \gamma < (j-i)\ell' \leq 4d\ell' - \ell' \leq 8d\ell - \ell'$. On the other hand we have $\gamma(j) = (\gamma, \mathrm{r}, +)$, too. Hence for some $\mu' < \gamma < \nu'$ with $\gamma - \mu' < \ell'$ we have $W_j = w_0[\mu', \nu']$, too. Therefore $0 < \mu - \mu' < 8d\ell$ and W_j has some non-trivial overlap. We can write $W_j = W^e W'$ such that $1 \leq |W| < 8d\ell$ and W' is a prefix of W.

Putting everything together, we arrive in all cases at a factorization $W_0 = UW^eV$ with $e \leq \exp(w_0)$, $1 \leq |W| < 8d\ell$, and $|U| + |V| < 16d\ell$. However, we have not finished yet. Recall that we are looking for an admissible exponential expression for

$$\beta(u, w, v) = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q).$$

Due to $|W_0| > \ell$ we can choose r minimal, $p < r \leq q+1$, and s maximal $p - 1 \leq s < q$ such that $|w_p \cdots w_{r-1}| > |U| + \ell$ and $|w_{s+1} \cdots w_q| > |V| + \ell$. By Lemma 37 we may assume $r < s$ and it is enough to find an exponential expression for

$$S = (u_r, w_r, v_r) \cdots (u_s, w_s, v_s).$$

Note that the word $u_r w_r w_{r+1} \cdots w_s v_s$ is a factor of W^e. Again, we may assume that $w_r w_{r+1} \cdots w_s > 32d\ell$. By switching to some conjugated word W' if necessary, we may assume that $u_r w_r w_{r+1} \cdots w_s v_s$ is a prefix of W^e. Moreover, by symmetry we may choose a positive interval $[\alpha, \beta]$ such that $w_0[\alpha, \beta] = u_r w_r w_{r+1} \cdots w_s v_s$. Clearly, we have $w_0[i, j] = w_0[i + |W|, j + |W|]$ for all $\alpha \leq i < j \leq \beta - |W|$. In particular, the critical word $w_0[\alpha, \alpha + 2\ell]$ appears as $w_0[\alpha + |W|, \alpha + |W| + 2\ell]$ again. This means that there is some $r \leq t < s$ such that $|w_r \cdots w_t| = |W|$. More precisely, we can choose $r \leq t < t' \leq s$ and a maximal $e' \leq e$ such that

$$S = \big((u_r, w_r, v_r) \cdots (u_t, w_t, v_t)\big)^{e'} (u_{t'}, w_{t'}, v_{t'}) \cdots (u_s, w_s, v_s).$$

Since it holds $e' \leq \exp(w_0)$, $|w_r \cdots w_t| = |W|$, and $|w_{t'} \cdots w_s| \leq |W|$, the existence of an admissible exponential expression for $\beta(u, w, v)$ follows. Hence β is an admissible base change.

18 Passing from E_ℓ to $E_{\ell,\ell'}$ for $1 \leq \ell < \ell' \leq 2\ell$

In the final step we have to show that there exists some projection $\pi : \Gamma^*_{\ell,\ell'} \to \Gamma^*_\ell$ and some partial solution $\delta : \Omega_\ell \to \Gamma^*_{\ell,\ell'} \Omega_{\ell'} \Gamma^*_{\ell,\ell'} \cup \Gamma^*_{\ell,\ell'}$ such that $\delta_*(\pi^*(E_\ell)) \equiv E_{\ell,\ell'}$. We don't have to care about admissibility anymore.

For the projection we have to consider a letter in $\Gamma_{\ell,\ell'} \setminus \Gamma_\ell$. Such a letter has the form $(u, w, v) \in B_\ell$ and we may define $\pi(u, w, v) = w$ since $\Gamma \subseteq \Gamma_\ell$. Clearly $\pi(\overline{(u, w, v)}) = \overline{\pi(u, w, v)}$ and $h_{\ell,\ell'}(u, w, v) = h_{\ell'}(u, w, v) = h(w) = h_\ell(\pi(u, w, v))$ are verified. Thus $\pi : \Gamma^*_{\ell,\ell'} \to \Gamma^*_\ell$ defines a projection such that

$$\pi^*(E_\ell) = (\Gamma_{\ell,\ell'}, h_{\ell,\ell'}, \Omega_\ell, \rho_\ell; L_\ell = R_\ell).$$

We have to define a partial solution $\delta : \Omega_\ell \to \Gamma^*_{\ell,\ell'}\Omega_{\ell'}\Gamma^*_{\ell,\ell'} \cup \Gamma^*_{\ell,\ell'}$ such that $\delta(L_\ell) = \beta(L_{\ell'})$ and $\delta(R_\ell) = \beta(R_{\ell'})$. For this, we have to consider a variable $X \in \Omega$ with $\mathrm{body}_\ell(\sigma(X)) \neq 1$. By symmetry, we may assume that $X = x_i$ for some $1 \leq i \leq g$. Hence $\sigma(X) = w_0[\mathrm{l}(i), \mathrm{r}(i)]$.
Let $\alpha = \mathrm{l}(i) + |\mathrm{head}_\ell(\sigma(X))|$ and $\beta = \mathrm{r}(i) - |\mathrm{tail}_\ell(\sigma(X))|$. Then $\mathrm{l}(i) + \ell \leq \alpha < \beta \leq \mathrm{r}(i) - \ell$. Let $(u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ be the minimal cover of $[\alpha, \beta]$ with respect to the ℓ-factorization. We have $w_p \cdots w_q = \mathrm{body}_\ell(\sigma(X))$.
For $\mathrm{body}_{\ell'}(X) = 1$ we have $X \in \Omega_\ell \setminus \Omega_{\ell'}$ and we define

$$\delta(X) = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q).$$

Then $\delta(X) \in B^*_\ell$ and $h_\ell\delta(X) = \rho_\ell(X)$ since $\rho_\ell(X) = h(\mathrm{body}_\ell(\sigma(X)))$. It is also clear that the definition does not depend on the choice of i, and we have $\delta(\overline{X}) = \overline{\delta(X)}$.
Recall the definition of $L_{\ell'}$. Since $\mathrm{body}_{\ell'}(\sigma(X)) = 1$, there is a factor $f_1 \cdots f_r$ of $L_{\ell'}$ which belongs to $\Gamma^*_{\ell'}$ and $f_1 \cdots f_r$ covers $[\alpha, \beta]$ with respect to the ℓ'-factorization $F_{\ell'}(w_0)$. It follows that $\delta(X)$ is a factor of $\beta(f_1 \cdots f_r)$, hence $\delta(X) \in \Gamma^*_{\ell,\ell'}$ by definition of $\Gamma_{\ell,\ell'}$.
For $\mathrm{body}_{\ell'}(X) \neq 1$ we have $X \in \Omega_{\ell'}$ and we find positions $\mu < \nu$ such that $\mu = \mathrm{l}(i) + |\mathrm{head}_{\ell'}(\sigma(X))|$ and $\nu = \mathrm{r}(i) - |\mathrm{tail}_{\ell'}(\sigma(X))|$.
For some $p \leq r \leq s \leq q$ we have $w_0[\alpha, \mu] = w_p \cdots w_{r-1}$, $w_0[\nu, \beta] = w_{s+1} \cdots w_q$, and $\mathrm{body}_{\ell'}(\sigma(X)) = w_r \cdots w_s$. We define

$$\delta(X) = (u_p, w_p, v_p) \cdots (u_{r-1}, w_{r-1}, v_{r-1}) X (u_{s+1}, w_{s+1}, v_{s+1}) \cdots (u_q, w_q, v_q).$$

As above, we can verify that $\delta(X) = UXV$ with $U, V \in \Gamma^*_{\ell,\ell'}$ such that $\delta(\overline{X}) = \overline{V}\,\overline{X}\,\overline{U}$ and $\rho_\ell(X) = h_{\ell,\ell'}(U)\rho_{\ell'}(X)h_{\ell,\ell'}(V)$. Finally, $\delta(L_\ell) = L_{\ell'}$ and $\delta(R_\ell) = R_{\ell'}$. Hence $\delta_*(\pi^*(E_\ell)) = \beta_*(E_{\ell'})$. This proves Theorem 5.

19 Conclusion

In this paper we were dealing with the existential theory, only. For free groups it is also known that the positive theory without constraints is decidable, see [19]. Thus, one can allow also a mixture of existential and universal quantifiers, if there are no negations at all. Since a negation can be replaced with the help of an extra variable and some positive rational constraint, one might be tempted to prove that the positive theory of equations with rational constraints in free groups is decidable. But such a program must fail: Indeed, by [20] and [7] it is known that the positive $\forall\exists^3$-theory of word equations is unsolvable. Since Σ^* is a rational subset of the free group $F(\Sigma)$, this theory can be encoded in the positive theory of equations with rational constraints in free groups, and the later is undecidable, too. On the other hand, a negation leads to a positive constraint of a very restricted type, so the interesting question remains under which type of constraints the positive theory becomes decidable.

Acknowledgments The research was partly supported by the German Research Foundation *Deutsche Forschungsgemeinschaft, DFG.* In addition, C. Gutiérrez thanks Centro de Modelamiento Matemático, FONDAP Matemáticas Discretas, for financial support.

References

[1] Michèle Benois. Parties rationelles du groupe libre. *C. R. Acad. Sci. Paris, Sér. A*, 269:1188–1190, 1969.

[2] Jean Berstel. *Transductions and context-free languages.* Teubner Studienbücher, Stuttgart, 1979.

[3] Volker Diekert. Makanin's Algorithm. In M. Lothaire, editor, *Algebraic Combinatorics on Words.* Cambridge University Press, 2001. To appear. A preliminary version is on the web:
http: //www-igm.univ-mlv.fr/ berstel/Lothaire/index.html.

[4] Volker Diekert, Claudio Gutiérrez, and Christian Hagenah. The existential theory of equations with rational constraints in free groups is PSPACE-complete. In A. Ferreira and H. Reichel, editors, *Proc. of the 18th STACS (STACS'01), Dresden*, Lecture Notes in Computer Science 2010, pages 170–182. Springer, 2001.

[5] Volker Diekert and Markus Lohrey. A note on the existential theory in plain groups. *International Journal of Algebra and Computation*, 2001. To appear.

[6] Volker Diekert, Yuri Matiyasevich, and Anca Muscholl. Solving word equations modulo partial commutations. *Theoretical Computer Science*, 224:215–235, 1999. Special issue of LFCS'97.

[7] Valery G. Durnev. Undecidability of the positive $\forall\exists^3$-theory of a free semi-group. *Sib. Mat. Zh.*, 36(5):1067–1080, 1995. In Russian; English translation: *Sib. Math. J., 36*(5), 917–929, 1995.

[8] Yuri Gurevich and Andrei Voronkov. Monadic simultaneous rigid E-unification and related problems. In P. Degano, R. Gorrieri, and A. Marchetti-Spaccamela, editors, *Proc. 24th ICALP (ICALP'97), Bologna (Italy) 1997*, number 1256 in Lecture Notes in Computer Science, pages 154–165. Springer, 1997.

[9] Claudio Gutiérrez. Satisfiability of word equations with constants is in exponential space. In *Proc. of the 39th Ann. Symp. on Foundations of Computer Science, FOCS'98*, pages 112–119, Los Alamitos, California, 1998. IEEE Computer Society Press.

[10] Claudio Gutiérrez. Satisfiability of equations in free groups is in PSPACE. In *32nd Ann. ACM Symp. on Theory of Computing (STOC'2000)*, pages 21–27. ACM Press, 2000.

[11] Christian Hagenah. *Gleichungen mit regulären Randbedingungen über freien Gruppen.* Ph.D.-thesis, Institut für Informatik, Universität Stuttgart, 2000.

[12] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

[13] Antoni Kościelski and Leszek Pacholski. Complexity of Makanin's algorithm. *Journal of the Association for Computing Machinery*, 43(4):670–684, 1996. Preliminary version in *Proc. of the 31st Annual IEEE Symposium on Foundations of Computer Science*, Los Alamitos, 1990, 824–829.

[14] Antoni Kościelski and Leszek Pacholski. Makanin's algorithm is not primitive recursive. *Theoretical Computer Science*, 191:145–156, 1998.

[15] Dexter Kozen. Lower bounds for natural proof systems. In *Proc. of the 18th Ann. Symp. on Foundations of Computer Science, FOCS 77*, pages 254–266, Providence, Rhode Island, 1977. IEEE Computer Society Press.

[16] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopaedia of Mathematics and its Applications*. Addison Wesley, 1983. Reprinted by *Cambridge University Press*, 1997.

[17] Gennadií Semyonovich Makanin. The problem of solvability of equations in a free semigroup. *Math. Sbornik*, 103:147–236, 1977. English transl. in Math. USSR Sbornik 32 (1977).

[18] Gennadií Semyonovich Makanin. Equations in a free group. *Izv. Akad. Nauk SSR*, Ser. Math. 46:1199–1273, 1983. English transl. in Math. USSR Izv. 21 (1983).

[19] Gennadií Semyonovich Makanin. Decidability of the universal and positive theories of a free group. *Izv. Akad. Nauk SSSR*, Ser. Mat. 48:735–749, 1984. In Russian; English translation in: *Math. USSR Izvestija, 25*, 75–88, 1985.

[20] S. S. Marchenkov. Unsolvability of positive $\forall\exists$-theory of a free semigroup. *Sib. Mat. Zh.*, 23(1):196–198, 1982. In Russian.

[21] George Markowsky. Bounds on the index and period of a binary relation on a finite set. *Semigroup Forum*, 13:253–259, 1977.

[22] Yuri I. Merzlyakov. Positive formulae over free groups. *ALgebra i Logika*, 5(4):25–42, 1966. English translation.

[23] Christos H. Papadimitriou. *Computatational Complexity*. Addison Wesley, 1994.

[24] Wojciech Plandowski. Testing equivalence of morphisms on context-free languages. In Jan van Leeuwen, editor, *Algorithms—ESA '94, Second Annual European Symposium*, volume 855 of *Lecture Notes in Computer Science*, pages 460–470, Utrecht, The Netherlands, 1994. Springer.

[25] Wojciech Plandowski. Satisfiability of word equations with constants is in NEXPTIME. In *Proceedings 31st Annual ACM Symposium on Theory of Computing, STOC'99*, pages 721–725. ACM Press, 1999.

[26] Wojciech Plandowski. Satisfiability of word equations with constants is in PSPACE. In *Proc. of the 40th Ann. Symp. on Foundations of Computer Science, FOCS 99*, pages 495–500. IEEE Computer Society Press, 1999.

[27] Wojciech Plandowski and Wojciech Rytter. Application of Lempel-Ziv encodings to the solution of word equations. In Kim G. Larsen et al., editors, *Proc. 25th ICALP (ICALP'98), Aalborg (Denmark) 1998*, number 1443 in Lecture Notes in Computer Science, pages 731–742. Springer, 1998.

[28] Alexander A. Razborov. On systems of equations in a free group. *Izv. Akad. Nauk SSSR*, Ser. Mat. 48:779–832, 1984. In Russian; English translation in: *Math. USSR Izvestija, 25*, 115–162, 1985.

[29] Klaus U. Schulz. Makanin's algorithm for word equations — Two improvements and a generalization. In Klaus U. Schulz, editor, *Word Equations and Related Topics*, number 572 in Lecture Notes in Computer Science, pages 85–150. Springer, 1991.

[30] Joachim von zur Gathen and Malte Sieveking. A bound on solutions of linear integer equalities and inequalities. *Proc. Amer. Math. Soc.*, 72(1):155–158, october 1978.