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

Volker Diekert [a,*], Claudio Gutierrez [b], Christian Hagenah [a]

[a] *Institut für Formale Methoden der Informatik (FMI), Universität Stuttgart,*
*Universitätsstr. 38 D-70569 Stuttgart, Germany*

[b] *Depto. de Ciencias de la Computación, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile*

## Abstract

It is well-known that the existential theory of equations in free groups is decidable. This is a celebrated result of Makanin which was published 1982. Makanin did not discuss complexity issues, but later it was shown that the scheme of his algorithm is not primitive recursive. In this paper we present an algorithm that works in polynomial space. This improvement is based upon an extension of Plandowski's techniques for solving word equations. We present a PSPACE-algorithm in a 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. We obtain our main result about the existential theory in free groups as a corollary of the corresponding statement in free monoids with involution.

*Keywords:* Equations in free groups

[*] Corresponding author. Fax: +49 711 78 16 310.

*E-mail addresses:* diekert@fmi.uni-stuttgart.de (V. Diekert), cgutierr@dcc.uchile.cl (C. Gutierrez), christian@hagenah.de (C. Hagenah).

## 1. Introduction

Around 1980 great progress was achieved on the algorithmic decidability of elementary theories of free monoids and groups. In 1977 Makanin [24], 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 1982 he extended his result to the more complicated situation in free groups [25]. Using a result by Merzlyakov [31] Makanin also showed that the positive theory of equations in free groups is decidable [26]. In [37] Razborov obtained a description of the general solution of given bounded periodicity exponent of an arbitrary system of equations in a free group.

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 this bound for Makanin's original algorithm to ExpSpace [14]. For solving equations in free groups Kościelski and Pacholski showed that the scheme proposed by Makanin is not primitive recursive [21].

In 1999 Plandowski used another method for solving word equations and he showed that the satisfiability problem for word equations is in Pspace [34,35]. One ingredient of his work is to use data compression to reduce the 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 [36]. Another important definition is the $\ell$-factorization of a solution to a word equation. The roots of the notion of $\ell$-factorization are in the notion of synchronizing factorization from [19].

Gutierrez extended Plandowski's method to the case of free groups [16]. 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 Gutierrez [15] they obtained a result which includes the presence of rational constraints [4,17].

The present paper is the journal version of [16,4]. 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 [38] who also pointed out the importance of this concept, see also [7,13]. The Pspace-completeness for the case of word equations with regular constraints has already been stated by Rytter according to [34, Theorem 1].

Our proof reduces the case of equations with rational constraints in free groups to the case of equations with regular constraints in free monoids with involution, which turn out to be central objects. (Makanin uses the notion of "paired alphabet;" one of the differences is that he considered "non-contractible" solutions only, whereas we deal with general solutions.) Our work extends the method of [34,35] so that it copes with involutions, and it extends the method of [16] so 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. To avoid an exponential blow-up, we do not use a reduction as in [26], 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 in Pspace. This part is technical and first we introduce several notions like base-change, projection, partial solution, and free interval. After these preparations we can follow Plandowski's method. Throughout we shall use many of the deep ideas which were presented in [34,35], but we 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. On the positive side, we obtain a self-contained presentation.

## 2. Basic notions and statements of theorems

### 2.1. Preliminaries

An *involution* on a set is a bijection $\overline{\phantom{x}}$ such that $\overline{\overline{x}} = x$ for all elements $x$. If $M$ is a monoid, then an involution $\overline{\phantom{x}} : M \to M$ means that we also require $\overline{1} = 1$ for the unit element 1 and $\overline{xy} = \overline{y}\,\overline{x}$ for all $x, y \in M$.

Let $\Sigma$ be a finite alphabet. By $\Sigma^*$ we denote the free monoid over $\Sigma$. Elements of $\Sigma^*$ are called *words*. The length of a word $w$ is denoted by $|w|$. A *factor* of a word $w$ is a word $v$ such that $w = w_1 v w_2$; it is called *proper* if $1 \neq v \neq w$. By $F(\Sigma)$ we denote the free group over $\Sigma$. Elements of $F(\Sigma)$ can be represented by words over $\Gamma = \Sigma \cup \overline{\Sigma}$, where $\overline{\Sigma} = \{\, \overline{a} \mid a \in \Sigma \,\}$ is a disjoint copy of $\Sigma$. We let $\overline{\overline{a}} = a$, this defines an involution $\overline{\phantom{x}} : \Gamma \to \Gamma$; and the involution is extended to $\Gamma^*$ by $\overline{a_1 \cdots a_n} = \overline{a_n} \cdots \overline{a_1}$. The meaning of $\overline{w}$ is the inverse $w^{-1}$ in $F(\Sigma)$. A word $w \in \Gamma^*$ is *freely reduced* , if it contains no factor of the form $a\overline{a}$ with $a \in \Gamma$. For $w \in \Gamma^*$ we denote by $\widehat{w}$ the freely reduced word which denotes the same group element in $F(\Sigma)$. Hence, $\widehat{u} = \widehat{v}$ if and only if $\psi(u) = \psi(v)$, where $\psi : \Gamma^* \to F(\Sigma)$ denotes the canonical homomorphism.

The classes of *rational* and *recognizable* subsets are defined for every monoid $M$ [10]. Rational sets (or languages) are defined inductively as follows. All finite subsets of $M$ are rational. If $C_1, C_2 \subseteq M$ are rational, then the union $C_1 \cup C_2$, the concatenation $C_1 \cdot C_2$, and the generated submonoid $C_1^*$ are rational. A subset $C \subseteq M$ is recognizable, if and only if there is a homomorphism $h$ to some finite monoid $M'$ such that $C = h^{-1}h(C)$. Kleene's Theorem states that in finitely generated free monoids both classes coincide, and we follow the usual convention to call a rational (or recognizable) subset of a free monoid *regular*.

The empty word is the unit element of a free monoid, it is denoted by 1 as the unit element in other monoids. The singleton set $\{1\}$ is rational in $F(\Sigma)$, but not recognizable if $\Sigma \neq \emptyset$. A subset $C \subseteq F(\Sigma)$ is rational if and only if $C = \psi(C')$ for some regular language $C' \subseteq \Gamma^*$. In particular, we can use a non-deterministic finite automata over $\Gamma$ for specifying rational group languages over $F(\Sigma)$.

The *existential theory of equations with rational constraints* in a monoid $M$ with a generating set $\Gamma$ is defined as follows. Let $\Omega$ be a set of variables (or unknowns). Atomic formulae are either of the form $L = R$, where $L, R \in (\Gamma \cup \Omega)^*$ or of the form $X \in C$, where $X$ is in $\Omega$ and $C \subseteq M$ is a rational language. An *existentially quantified* formula is a block of existentially quantified variables followed by a Boolean combination of atomic formulae. It is closed, if there are no free variables. The existential theory of equations with rational constraints in $M$ is the set of all closed existentially quantified formulae which are *true* in $M$.

### 2.2. Free groups

The next proposition is due to Benois [1], see also [2, Section III.2].

**Proposition 1.** *The family of rational languages over the free group $F(\Sigma)$ forms an effective Boolean algebra.*

**Proof.** (*Sketch.*) It is enough to show that the family of rational languages is closed under complementation. Let $C' \subseteq \Gamma^*$ be a regular language and $C = \psi(C')$ the corresponding rational group language in $F(\Sigma)$. Assume that $C'$ is given by some non-deterministic finite automaton. Using the same state set we can construct (in polynomial time) a finite automaton which accepts the following language

$$C'' = \{ v \in \Gamma^* \mid \exists u \in C' : u \overset{*}{\to} v \}$$

where $u \overset{*}{\to} v$ means that $v$ is a descendant of $u$ by the rewriting system $\{ a\overline{a} \to 1 \mid a \in \Gamma \}$. Then we complement $C''$ with respect to $\Gamma^*$; and intersect $\Gamma^* \setminus C''$ with the regular set of freely reduced words. We obtain a regular set $\widetilde{C}'$. Hence, the complement of $C$ in $F(\Sigma)$ is the rational group language $\psi(\widetilde{C}')$. □

**Problem 2.** By EFG we denote the following decision problem:

INPUT: A finite alphabet $\Sigma$ and a closed existentially quantified formula with rational constraints in the free group $F(\Sigma)$.

QUESTION: Is the formula *true* in $F(\Sigma)$?

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

The difficult part is to show that EFG is in PSPACE. For this we prove a more general statement about the existential theory of equations with regular constraints in free monoids with involution.

### 2.3. Free monoids with involution

In the following let $\Gamma$ be a finite alphabet of constants and $\Omega$ be an alphabet of variables together with involutions $^- : \Gamma \to \Gamma$ and $^- : \Omega \to \Omega$. The involution on $\Omega$ is without fixed points, but we allow fixed points for the involution on $\Gamma$. The involution is extended to $(\Gamma \cup \Omega)^*$ by $\overline{x_1 \cdots x_n} = \overline{x_n} \cdots \overline{x_1}$ for $n \geqslant 0$ and $x_i \in \Gamma \cup \Omega$, $1 \leqslant i \leqslant n$. Clearly, $\overline{\overline{u}} = u$ for all $u \in (\Gamma \cup \Omega)^*$.

From now on, almost all monoids $M$ under consideration are equipped with an involution $^- : M \to M$. A *morphism* between monoids with involution $M$ and $M'$ is henceforth 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 $x, y \in M$. Thus, a morphism is a homomorphism of monoids which respects the involution. The pair $(\Gamma^*, ^-)$ is called a *free monoid with involution.* A morphism $h : \Gamma^* \to M$ is specified by a list $(h(a); a \in \Gamma)$ such that $h(\overline{a}) = \overline{h(a)}$ for all $a \in \Gamma$.

**Problem 4.** By EFMI we denote the following decision problem:

INPUT: A closed existentially quantified formula with regular constraints in a free monoid with involution $(\Gamma^*, ^-)$.

QUESTION: Is the formula *true* in $(\Gamma^*, ^-)$?

The proof of the following statement is the main technical contribution of the paper.

**Theorem 5.** *The problem EFMI is* PSPACE-*complete.*

### 2.4. Equations with constraints

In the following it is more suitable to work with Boolean matrices instead of finite automata. Let $n \geqslant 1$. Henceforth, $M_{2n} \subseteq \mathbb{B}^{2n \times 2n}$ denotes the following monoid with involution:

$$M_{2n} = \left\{ \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \mid A, B \in \mathbb{B}^{n \times n} \right\},$$

where

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

The operator $^{\mathrm{T}}$ denotes transposition and $\mathbb{B}^{n \times n}$ is the monoid of Boolean $n \times n$–matrices.

**Definition 6.** An *equation E with constraints* is a list

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

containing the following items:

- The alphabet $\Gamma = (\Gamma, \overline{\ })$ with involution.
- The morphism $h : \Gamma^* \to M_{2n}$ which is specified by a mapping $h : \Gamma \to M_{2n}$ such that $h(\overline{a}) = \overline{h(a)}$ for all $a \in \Gamma$.
- The alphabet $\Omega = (\Omega, \overline{\ })$ with involution without fixed points.
- A mapping $\rho : \Omega \to M_{2n}$ such that $\rho(\overline{X}) = \overline{\rho(X)}$ for all $X \in \Omega$.
- The word equation $L = R$ where $L, R \in (\Gamma \cup \Omega)^+$.

A *solution* of $E$ is a mapping $\sigma : \Omega \to \Gamma^*$, which is extended to a morphism $\sigma : (\Gamma \cup \Omega)^* \to \Gamma^*$ by leaving the letters from $\Gamma$ invariant such that the following three conditions are satisfied:

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

Let $d = |LR|$ be the *denotational length* of the word equation $L = R$. The *input size* of $E$ is given by

$$\|E\| = d + n + \log_2(|\Gamma| + |\Omega|).$$

The definition of input size takes into account that there might be constants or variables with constraints which are not present in the equation. Due to this definition we assume that the input to Problem 7 is kept on a separate read-only storage.

**Problem 7.** By EWC we denote the following decision problem:
INPUT: An equation with constraints, $E = (\Gamma, h, \Omega, \rho; L = R)$.
QUESTION: Is there a solution $\sigma : \Omega \to \Gamma^*$?

**Theorem 8.** *The problem EWC is* PSPACE-*complete.*

We now turn to the proofs of Theorems 3, 5, and 8. The PSPACE-hardness of the problems EFMI, EFG, and EWC follows directly from a result of Kozen [22], since the *empty intersection* problem of regular sets can easily be encoded in the problems above. Therefore the PSPACE-hardness is not discussed further in the sequel.

The difficult part is to show that the problems EFG, EFMI, and EWC can be solved in polynomial space. We proceed as follows. Section 3.1 yields a (polynomial time) reduction from the problem EFG to the problem EFMI. Section 3.2 yields a reduction from EFMI to EWC, but this reduction involves non-deterministic steps. It can be performed however in non-deterministic polynomial time. Section 4 is the core of the paper. It shows that the problem EWC can be solved by some non-deterministic PSPACE algorithm. By Savitch's theorem such a procedure can be transformed into a polynomially space bounded deterministic decision procedure, see e.g. [18]. This concludes the proof of Theorems 3, 5, and 8.

**Remark 9.** Problem EWC is *NP*–hard for $n = 1$ already, since then we are in the framework of word equations (without constraints); and linear integer programming can easily be reduced to word equations, see e.g. [3]. We conjecture that the problem is in fact *NP*-complete, if $n$ is bounded by some constant which is not part of the input, see also [36].

## 3. Reductions

### *3.1. Reduction of problem EFG to EFMI*

The next technical lemma follows directly from the well known fact that the Cayley graph of a free group is a tree. The proof of Lemma 10 is therefore omitted. As above, let $\psi : \Gamma^* \to F(\Sigma)$ be the canonical morphism.

**Lemma 10.** *Let $u, v, w \in \Gamma^*$ be freely reduced words. Then we have $uvw = 1$ in $F(\Sigma)$ (i.e. $\psi(uvw) = 1$) if and only if there are words $P, Q, R \in \Gamma^*$ such that $u = P\overline{Q}$, $v = Q\overline{R}$, and $w = R\,\overline{P}$ in $\Gamma^*$.*

**Proposition 11.** *There is a polynomial time reduction of problem EFG to EFMI.*

**Proof.** The reduction follows standard lines. The input to the problem EFG is a closed existentially quantified formula with rational constraints in the free group $F(\Sigma)$. Using De Morgan's law we may assume that there are no negations at all. Since we are in a group, the atomic formulae are now of the either form: $W = 1$, $W \neq 1$, $X \in C$ or $X \notin C$ where $W \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $C \subseteq F(\Sigma)$ is rational. The reason that we keep $X \notin C$ instead of $X \in \widetilde{C}$ where $\widetilde{C} = F(\Sigma) \setminus C$ is that the complementation may involve an exponential blow-up.

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.

We may assume that $|W| \geqslant 3$, since if $1 \leqslant |W| < 3$, then we may replace $W = 1$ by $Wa\overline{a} = 1$ for some $a \in \Gamma$. For the present reduction it is convenient to assume that $|W| = 3$ for all subformulae $W = 1$. This is easy to achieve. As long as there is a subformula $x_1 \cdots x_k = 1, x_i \in \Gamma \cup \Omega$ for $1 \leqslant i \leqslant k$ and $k \geqslant 4$, 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.

Now, there are no negations and all atomic formulae are of type $W = 1, X \in C$ or $X \notin C$, where $W \in (\Gamma \cup \Omega)^+$, $|W| = 3, X \in \Omega$, and $C \subseteq F(\Sigma)$ is rational.

At this point we switch to free monoids with involution. Recall that $\psi : \Gamma^* \to F(\Sigma)$ denotes the canonical morphism and that $X \in C$ (respectively, $X \notin C$) means in fact $X \in \psi(C')$ (respectively, $X \notin \psi(C')$), where $C' \subseteq \Gamma^*$ is a regular language specified by some finite non-deterministic automaton over the alphabet $\Gamma^*$. Using $\psi$-symbols we obtain an interpretation over $(\Gamma^*, \bar{\ })$ without changing the truth value of the input formula: We replace each subformula $X \in C$ (respectively, $X \notin C$) syntactically by $\psi(X) \in \psi(C')$ (respectively, $\psi(X) \notin \psi(C')$) and we replace each subformula $W = 1$ by $\psi(W) = 1$.

We keep the interpretation over words, but we now eliminate all occurrences of $\psi$ again. We begin with the occurrences of $\psi$ in the constraints. Let $C' \subseteq \Gamma^*$ be regular. According to the proof of Proposition 1 we construct a finite automaton, which accepts the following language

$$C'' = \{ v \in \Gamma^* \mid \exists u \in C' : u \overset{*}{\to} v \}.$$

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

We replace all positive atomic subformulae of the form $\psi(X) \in \psi(C')$ by $X \in C''$. A simple reflection shows that the truth value has not changed since we can think of $X$ as being a freely reduced word. For a negative formula $\psi(X) \notin \psi(C')$ 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 some deterministic finite automaton with $|\Gamma| + 2$ states. We replace $\psi(X) \notin \psi(C')$ by

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

where $C''$ 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 quantifier free formula is satisfiable over $\Gamma^*$ if and only if it is satisfiable in freely reduced words.

Based on Lemma 10 we replace each atomic subformulae $\psi(xyz) = 1$ with $x, y, z \in \Gamma \cup \Omega$ by a conjunction

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

where $P$, $Q$, $R$ are fresh variables and the existential block is put to the front. The new existential formula has no occurrence of $\psi$ anymore. The atomic subformulae are of the form $x = yz, X \in C$ or $X \notin C$, where $x, y, z \in \Gamma \cup \Omega$ and $C \subseteq \Gamma^*$ is regular. The size of the new formula is polynomial in the size of the original formula. This finishes the reduction from the problem EFG to EFMI. □

### *3.2. Reduction of Problem EFMI to EWC*

**Proposition 12.** *There is a non-deterministic polynomial time reduction of problem EFMI to EWC.*

**Proof.** The input to problem EFMI is a closed existentially quantified formula $\Phi$ with regular constraints over a free monoid with involution. We define a procedure which transforms the input $\Phi$

into an equation with constraints $E_\Phi$. If $\Phi$ is true, then at least one possible output $E_\Phi$ has a solution. If the output $E_\Phi$ has a solution, then $\Phi$ is true. The procedure will work in non-deterministic polynomial time.

We may assume that the formula $\Phi$ contains no negations and all atomic subformulae are of type $U = V$, $U \neq V$, $X \in C$ or $X \notin C$, where $U, V \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $C \subseteq \Gamma^*$ is regular.

Since we work over a free monoid $\Gamma^*$ it is easy to handle inequalities $U \neq V$ where $U, V \in (\Gamma \cup \Omega)^*$. If two words $u$, $v$ in $\Gamma^*$ are different, then there are three cases: $u$ is a proper prefix of $v$ or $v$ is a proper prefix of $u$ or there is some word $x$ such that $xa$ is a prefix of $u$, $xb$ is a prefix of $v$, and $a \neq b$. Therefore, a subformula $U \neq V$ can be replaced by

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

Making guesses we can eliminate all disjunctions to obtain an existentially quantified formula which consists of a block of existentially quantified variables followed by a single conjunction over atomic subformulae of type $U = V$, $X \in C$ or $X \notin C$, where $U, V \in (\Gamma \cup \Omega)^*$, $X \in \Omega$, and $C \subseteq \Gamma^*$ is regular.

By a standard procedure we can replace a conjunction of word equations over $(\Gamma \cup \Omega)^*$ by a single word equation $L = R$ where neither $L$ nor $R$ is empty. For example, we may choose a new letter $a$ ($a \notin \Gamma$) 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 we add for all variables $X$ the constraint $X \in \Gamma^*$.

Therefore, we may assume that our input is now given by three items: a single word equation $L = R$ with $L, R \in (\Gamma \cup \Omega)^+$ and two lists: $(X_j \in C_j, 1 \leqslant j \leqslant m)$ and $(X_j \notin C_j, m < j \leqslant k)$. Each regular language $C_j \subseteq \Gamma^*$ is specified by some non-deterministic automaton $\mathscr{A}_j = (Q_j, \Gamma, \delta_j, I_j, F_j)$ 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, $1 \leqslant j \leqslant k$. Of course, a variable $X$ may occur several times in the list with different constraints, therefore we might have $k$ greater than $|\Omega|$.

For the reduction to the problem EWC we have to consider Boolean matrices instead of finite automata. This allows us to store all constraints concerning a variable in a single Boolean matrix. Let $Q$ be the disjoint union of the state spaces $Q_j$, $1 \leqslant j \leqslant k$. We may assume that $Q = \{1, \ldots, n\}$. Let $\delta = \bigcup_{1 \leqslant j \leqslant k} \delta_j$, 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 $g(a)_{i,j} = 1$, if $(i, a, j) \in \delta$ and $g(a)_{i,j} = 0$ otherwise. We define a morphism $h : \Gamma^* \to M_{2n}$ by

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

The list of matrices $(h(a); a \in \Gamma)$ can be computed in polynomial time and we have $h(\bar{a}) = \overline{h(a)}$. Now, for each regular language $C_j$, $1 \leqslant j \leqslant k$ we compute vectors $I_j, F_j \in \mathbb{B}^{2n}$ (corresponding to initial and final states) such that for all $w \in \Gamma^*$ and $1 \leqslant j \leqslant k$ we have the equivalence:

$$w \in C_j \Leftrightarrow I_j^{\mathrm{T}} h(w) F_j = 1.$$

Having done these computations we make a non-deterministic guess $\rho(X) \in M_{2n}$ 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 $X \in C_j$

for some $1 \leqslant j \leqslant m$ (or $X \notin C_j$ for some $m < j \leqslant k$), then we verify $I_j^{\mathrm{T}} \rho(X) F_j = 1$, if $1 \leqslant j \leqslant m$ (or $I_j^{\mathrm{T}} \rho(X) F_j = 0$, if $m < j \leqslant k$).

This finishes the reduction of problem EFMI to EWC. □

## 4. Problem EWC is in PSPACE

### 4.1. Road-map

The proof of Theorem 8 is based on three transformation rules for equations with constraints. Each transformation preserves unsolvability; and it can be applied as long as the computation respects a given polynomial space bound. (The notion of *admissibility* given in Definition 31 formalizes the notion that the size of some object is bounded polynomially in the input size.)

No transformation rule introduces any new variable, but it may happen that the number of variables decreases. So, the global strategy is to apply the rules until all variables have been eliminated; the final step is then a direct evaluation of an equation without variables.

If the final output is *yes*, then the input equation is solvable, too. The main difficulty in the proof is the converse. We have to show that we can perform all these transformations within polynomial space such that for a solvable equation with constraints at least one computation path leads to the output *yes*.

To overcome this difficulty various notions and concepts are developed. We follow the approach of Plandowski [34,35], but we have two sources for additional complications. We have to cope with the involution and we have constraints. It is fairly standard to handle regular constraints. It may look rather technical if a reader sees it for the first time, but there is no surprise and the real additional difficulty is condensed in one section.

There are in fact three Sections 4.2, 4.6, and 4.9, where regular constraints play a crucial role. In Section 4.2, we show why an explicit specification of the constants is necessary. On an algebraic level we have to solve a membership problem in a submonoid of Boolean matrices. The submonoid is given by a list of matrices and we ask whether some other matrix $A$ is a product of matrices from the list. Clearly, the answer may be *no*, but if we add $A$ to the list, then it becomes trivially *yes*. In our language this means that it may happen that an equation with constraints becomes solvable by enlarging the alphabet of constants. This effect is not possible without constraints: If a word equation $L = R$ without constraints has a solution, then it has a solution over the alphabet of constants which appear in the string $LR$.

The presence of constraints makes it necessary to formalize the notion of *projection* in Section 4.6. A projection is a controlled way of introducing new constants such that unsolvable equations remain unsolvable. The use of new constants is inherent in Plandowski's method. If during the transformation the underlying word equations becomes too long, long subsequences of constants (factors) are coded as a single new letter. So, the alphabet of constants changes all the time: We remove constants in order to keep the alphabet size polynomially bounded, and introduce them in order to keep the length of the underlying word equation polynomially bounded. The technical preparation for this is done in Section 4.9. It introduces the notion of *free interval* and it is there where our presentation becomes more involved due to constraints.

Dealing with involutions is the main source for new difficulties. For example, we cannot directly apply the usual method for bounding the exponent of periodicity. We need a new concept of $p$-stable normal form in Section 4.3. The result of this section is however as expected: If a $w_0$ represents a solution of minimal length, then the number of repetitions inside $w_0$ is bounded singly exponential in the size of the equation. Thus, if $w_0 = uv^k w$, then in binary notation $k$ uses polynomially many bits only.

This leads directly to Section 4.4. Word equations are not stored in plain form, but Plandowski's method uses data compression to keep them within polynomial size. More specifically, we allow regular expressions with exponents in binary notation.

The following three sections explain the transformation rules in detail: In Section 4.5, we formalize the way to remove constants and Section 4.6 deals with the controlled way of introducing them. In Section 4.7, we formalize guessing a *partial solution*.

The transformation rules lead to the formal description of a *search graph* in 4.8. The difficulty of proving Theorem 8 is reduced to showing that the search graph contains a path from a solvable input equation to some trivial equation. This part is very complex, but the basic ideas can be traced to [36] where Lempel-Ziv encodings of minimal solutions of a word equation are investigated. Key notions are *critical word* and $\ell$-factorization. The technical part is developed in sections 4.10 to 4.15.

### *4.2. A PSPACE-complete subproblem*

The next proposition states that two basic operations, which are used several times as subroutines, can be performed in PSPACE.

**Proposition 13.** *The following problems are* PSPACE-*complete with respect to the input size* $n + \log |\Gamma|$.

*INPUT: A matrix* $B \in \mathbb{B}^{n \times n}$ *and a homomorphism* $g : \Gamma^* \to \mathbb{B}^{n \times n}$ *given as a list of matrices* $(B_1, \ldots, B_{|\Gamma|})$.

*QUESTION: Is there some* $u \in \Sigma^*$ *such that* $g(u) = B$?

*INPUT: A matrix* $A \in M_{2n}$ *and a morphism* $h : \Gamma \to M_{2n}$ *given as a list of matrices* $(A_1, \ldots, A_{|\Gamma|})$ *with* $\overline{A_{a_i}} = A_{\overline{a_i}}$ *for all* $a_i \in \Gamma$.

*QUESTION: Is there some* $w \in \Gamma^*$ *such that* $h(w) = A$ *and* $w = \overline{w}$?

**Proof.** The first problem is closely related to the intersection problem of regular languages and its PSPACE-hardness is again due to Kozen [22], see also [11, MS5]. The PSPACE-algorithm starts with the unit matrix. Then it guesses a word $u$ letter by letter and, simultaneously, calculates $g(u)$: If we guess the letter $a_i$, then we move to the $i$th matrix in the list $(B_1, \ldots, B_{|\Gamma|})$ describing $g$, and we multiply $B_i$ on the right to the current value held in the work space. We terminate if and only if $g(u) = B$.

The second problem can be solved since $w = \overline{w}$ implies $w = ub\bar{u}$ for some $u \in \Gamma^*$ and $b \in \Gamma \cup \{1\}$ with $b = \overline{b}$. Hence we can guess some $B$ and $b$ and we verify $A = Bh(b)\overline{B}$ and $b = \overline{b}$. Then using the first part, we check that $B = h(u)$ for some $u \in \Gamma^*$.

Since there is no reference which shows the PSPACE-hardness of the second problem, we sketch a reduction from the first to the second one: Consider a mapping $g : \Sigma \to \mathbb{B}^{n \times n}$ and $B \in \mathbb{B}^{n \times n}$, the pair $(B, g)$ is an instance of the first problem. Let $\Gamma$ be the disjoint union $\Sigma \cup \overline{\Sigma}$ and let $g(\bar{a}) = 1$, where $1 \in \mathbb{B}^{n \times n}$ is the identity matrix. In the notations of above we have $h(a) = \begin{pmatrix} g(a) & 0 \\ 0 & 1 \end{pmatrix}$ for $a \in \Gamma$.

Let $A = \begin{pmatrix} B & 0 \\ 0 & B^{\mathrm{T}} \end{pmatrix}$, then the pair $(A, h)$ becomes an instance of the second problem. Clearly, if $g(u) = B$ for some $u \in \Sigma^*$, then $h(u\bar{u}) = \begin{pmatrix} B & 0 \\ 0 & B^{\mathrm{T}} \end{pmatrix}$. For the converse note that the matrices $h(a)$ and $h(\bar{b})$ commute for all $a, b \in \Sigma$. If there is some $w = \overline{w} \in \Gamma^*$ with $h(w) = \begin{pmatrix} B & 0 \\ 0 & B^{\mathrm{T}} \end{pmatrix}$, then we can write $w = w_1\overline{w_1}$ and we may assume that $w_1 = u_1\overline{u_2}$ with $u_1, u_2 \in \Sigma^*$. It follows that $g(u_1u_2) = B$. □

Assume that an equation $E = (\Gamma, h, \Omega, \rho; L = R)$ contains a variable $X$ in the specification which does not occur in $LR\overline{LR}$. In this case the equation might be unsolvable, simply because $\rho(X) \notin h(\Gamma^*)$. However, by Proposition 13 we can test this in PSPACE. Therefore, if $X$ does not appear in $LR\overline{LR}$ and $\rho(X) \in h(\Gamma^*)$, then we can remove $X$ and $\overline{X}$ from the specification. This yields the following remark.

**Remark 14.** Henceforth, if $E = (\Gamma, h, \Omega, \rho; L = R)$ is an equation with constraints, then we assume that all variables occur somewhere in $LR\overline{LR}$. As a consequence, we may assume $|\Omega| \leqslant 2|LR|$.

### *4.3. The exponent of periodicity*

A key step in proving Theorem 8 is to find an effective bound on the exponent of periodicity in a solution of minimal length. This idea is used in all known algorithms for solving word equations, c.f. [24,34,35]. It turns out that the well-known result on word equations [20] transfers to the situation here: The exponent of periodicity can be bounded by a singly exponential function.

Let $w \in \Gamma^*$ be a word. The *exponent of periodicity* $\exp(w)$ is defined by

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

**Proposition 15.** *Let* $E = (\Gamma, h, \Omega, \rho; L = R)$ *be a solvable equation with constraints. Then there is a solution* $\sigma : \Omega \to \Gamma^*$ *such that* $\exp(\sigma(L)) \in 2^{\mathcal{O}(d+n\log n)}$.

The proof of Proposition 15 is independent of the rest of the paper. Therefore, we postpone it to the appendix, Section 6.

### *4.4. Exponential expressions*

To keep computations in polynomial space Plandowski's method uses *exponential expressions*. We give inductive definitions for an exponential expression, its evaluation, and its size.

**Definition 16.**

- Every word $w \in \Gamma^*$ is 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 $\|(e)^k\| = \|e\| + \max\{1,\ \lceil \log_2(k) \rceil\}$.

**Lemma 17.** *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 can find an exponential expression of size at most $p^2$ that represents $u$.*

**Proof.** The proof is an easy argument by structural induction. □

Lemma 17 will be applied to exponential expressions where the size $\|e\|$ is bounded by some value which is polynomial in the input size of the equation $E_0$. Since the size of the exponential expressions for factors can be the square of the original polynomial, we can apply this subroutine in nested way 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. The proof of Lemma 18 is omitted.

**Lemma 18.** *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 $h : \Gamma \to M_{2n}$ and an exponential expression $e$.*
*OUTPUT*: *The matrix* $h(\mathrm{eval}(e)) \in M_{2n}$.

**Remark 19.** The computation above can actually be performed in polynomial time, but this is not evident for the first question, see [32] for details.

Henceforth, we allow that the part $L = R$ of an equation with constraints may be given by a pair of exponential expressions $(e_L, e_R)$ with $\mathrm{eval}(e_L) = L$ and $\mathrm{eval}(e_R) = R$.

**Definition 20.** Let $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ and $E' = (\Gamma, h, \Omega, \rho; e'_L = e'_R)$ be equations with constraints. 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 $(\Gamma \cup \Omega)^*$.

The meaning of $E \equiv E'$ is that $E$ and $E'$ represent exactly the same equation if they were written out explicitly. By Lemma 18 we can decide $E \equiv E'$ in polynomial space; moreover, Remark 19 says that this decision is actually possible in polynomial time.

### 4.5. Base changes

In this section, we describe the first transformation rule. Let $h : \Gamma^* \to M_{2n}$ be a morphism. Let $(\Gamma', \overline{\phantom{a}})$ be another alphabet with involution and let $\beta : \Gamma' \to \Gamma^*$ be some mapping such that $\beta(\overline{a}) = \overline{\beta(a)}$ for all $a \in \Gamma'$. We define $h' : \Gamma' \to M_{2n}$ by $h' = h\beta$. We extend $\beta$ to a morphism $\beta : (\Gamma' \cup \Omega)^* \to (\Gamma \cup \Omega)^*$ by leaving the variables invariant and we call the morphism $\beta$ a *base change*.

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

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

**Lemma 21.** *Let $E'$ be an equation with constraints and $\beta : \Gamma' \to \Gamma^*$ be a base change. If $\sigma'$ 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 $\sigma'(L) = \sigma'(R)$. □

Lemma 21 leads to the first rule.

**Rule 1.** *If we have $E \equiv \beta_*(E')$ and 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 $E'$.*

For readability of the following examples all constraints are defined by membership in a regular language rather than by a mapping $\rho$. We also strengthen constraints in the examples (thereby having fewer solutions) in order to avoid lengthy regular expressions.

**Example 22.** Let $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. Consider the following equation $E$:

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

with constraints $X \in \Gamma^{300}\Gamma^*$ and $Z \in \bar{b}\bar{c}\bar{b}\bar{a}\Gamma^*$. Let $\Gamma' = \{a, b, \bar{a}, \bar{b}\}$ and define 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

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

We may strengthen the constraint to $X \in \Gamma'^{100}\Gamma'^*$ and $Z \in \bar{a}\Gamma'^*$. According to Rule 1 it is enough to solve $E'$. The effect of the base change $\beta$ is that both the equation $E'$ and the alphabet of constants are smaller. (The letter $c$ is not used anymore.) Note also that the length restriction on $X$ switched from $|X| \geqslant 300$ to $|X| \geqslant 100$. However, base changes may have a prize: It might be that $E = \beta_*(E')$ has a solution, whereas $E'$ is unsolvable. As we will see later, in our example the guess has been correct in the sense that $E'$ has a solution.

### 4.6. Projections

Let $(\Gamma, \overline{\ })$ and $(\Gamma', \overline{\ })$ be alphabets with involution such that $\Gamma \subseteq \Gamma'$. A *projection* is a morphism $\pi : \Gamma'^* \to \Gamma^*$ such that both, $\pi(a) = a$ for $a \in \Gamma$ and $\pi(\overline{a}) = \overline{\pi(a)}$ for all $a \in \Gamma'$.

Let $E$ be an equation with constraints $E = (\Gamma, h, \Omega, \rho; L = R)$. Then we define an equation with constraints $\pi^*(E)$ by

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

The equation $\pi^*(E)$ uses a larger alphabet of constants than $E$ does, but the word equation $L = R$ is exactly the same. Therefore $\pi^*(E)$ uses constants which do not appear in $L = R$. These constants may help to find (short) solutions which satisfy regular constraints. Note that every projection $\pi : \Gamma'^* \to \Gamma^*$ defines a base change $\pi$ such that $\pi_*\pi^*(E) = E$. Let $E' = \pi^*(E)$. By Rule 1 we may replace $\pi_*(E')$ by $\pi^*(E)$. We formulate this special case a second rule.

**Rule 2.** *Let $\pi$ be a projection. 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 $\pi^*(E)$.*

**Remark 23.** The reason to introduce Rule 2 will become clear only later. In Section 4.8, we define the formal notion of a search graph. We restrict the use of Rule 1 to so-called *admissible* base changes (cf. Definition 31), whereas there is no such restriction for the projection $\pi$ when we apply Rule 2.

**Lemma 24.** *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*) *Let $\pi^*(E) = E'$ for some projection $\pi$ and let $\sigma' : \Omega \to \Gamma'^*$ be a solution of $E'$. Then there is a solution $\sigma$ for $E$ such that $|\sigma(L)| \leqslant 2|M_{2n}||\sigma'(L)|$.*

**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 $|w_a| < |M_{2n}|$, since we can take the shortest word $w_a \in \Gamma^*$ such that $h(w_a) = h'(a) \in M_{2n}$. 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 $|w_a| \leqslant 2|M_{2n}| - 1$. For $b = 1$ we can demand $|w_a| \leqslant 2|M_{2n}| - 2$. Thus, we find a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $\pi^*(E) = E'$ and moreover, $|\pi(a)| < 2|M_{2n}|$ for all $a \in \Gamma'$.

(ii) According to the proof of (i) we may assume that $\pi : \Gamma'^* \to \Gamma^*$ satisfies $|\pi(a)| < 2|M_{2n}|$ for all $a \in \Gamma'$. Since $\pi$ defines a base change with $\pi_*(E') = E$, we know by Lemma 21 that $\sigma = \pi\sigma'$ is a solution of $E$. Clearly, $|\sigma(L)| = |\pi\sigma'(L)| \leqslant 2|M_{2n}||\sigma'(L)|$. □

**Example 25.** Let us continue with the equation which has been obtained by the transformation in Example 22. To simplify notations, we let $E$ be 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 $X$ has changed to $|X| \geqslant 100$. Let us reintroduce a letter $c$ and put $\Gamma' = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. We may define a projection $\pi : \Gamma' \to \Gamma^*$ by $\pi(c) = b^{100}$. The equation $E' = \pi^*(E)$ looks as above, but in $E'$ the constraint for $X$ has changed. The new constraint is $|X| \geqslant 100 \vee X \in \Gamma^* c \Gamma^*$. Thus, a solution for $X$ might be very short now.

During the procedure we occasionally have to decide whether there is a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $\pi^*(E) = E'$. This decision is possible according to the next proposition.

**Proposition 26.** *The following problem* PSPACE*-complete.*
*INPUT: Alphabets $(\Gamma, \overline{\phantom{x}}) \subseteq (\Gamma', \overline{\phantom{x}})$ and mappings $h, h'$, where $h$ is the restriction of $h'$.*
*QUESTION: Is there a projection $\pi : \Gamma'^* \to \Gamma^*$ such that $h' = h\pi$?*

**Proof.** This follows from Lemma 24 *(i)* and Proposition 13. □

### *4.7. Partial solutions*

Let $\Omega' \subseteq \Omega$ be a subset of the variables which is closed under involution. We assume that there is a mapping $\rho' : \Omega' \to M_{2n}$ with $\rho'(\overline{X}) = \overline{\rho'(X)}$, but we do not require that $\rho'$ is the restriction of

$\rho : \Omega \to M_{2n}$. Consider an equation with constraints $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$. 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 $\delta$ is extended to a morphism $\delta : (\Gamma \cup \Omega)^* \to (\Gamma \cup \Omega')^*$ by leaving the elements of $\Gamma$ invariant. Let $E' = (\Gamma, h, \Omega', \rho'; e_{L'} = e_{R'})$ be another equation with constraints (using the same $\Gamma$ and $h$). By abuse of language, we write $E' \equiv \delta_*(E)$, 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 27.** *In the notation of above, let $E' \equiv \delta_*(E)$ for some partial solution $\delta : \Omega \to \Gamma^*\Omega'\Gamma^* \cup \Gamma^*$. If $\sigma'$ 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, $\delta$ and $\sigma'$ are extended to morphisms $\delta : (\Gamma \cup \Omega)^* \to (\Gamma \cup \Omega')^*$ and $\sigma' : (\Gamma \cup \Omega')^* \to \Gamma^*$ leaving the letters of $\Gamma$ invariant. Since $E' = \delta_*(E)$ we have $\delta(L) = L'$ and $\delta(R) = R'$. Since $\sigma'$ is a solution, we have $\sigma(L) = \sigma'\delta(L) = \sigma'(L') = \sigma'(R') = \sigma'\delta(R) = \sigma(R)$ and $\sigma$ leaves the letters of $\Gamma$ invariant. The solution $\sigma'$ 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 28.** *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 $\delta$ such that $\delta_*(E) \equiv E'$?*

*If $\delta_*(E) \equiv E'$ is true, then there are exponential expressions of polynomial size $e_u, e_v$ for each $X \in \Omega'$ and $e_w$ for each $X \in \Omega \setminus \Omega'$ such that*

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

**Proof.** Let $L = \text{eval}(e_L)$, $R = \text{eval}(e_R)$, $L' = \text{eval}(e_{L'})$, and $R' = \text{eval}(e_{R'})$. The non-deterministic PSPACE algorithm works as follows:

For each variable $X \in \Omega'$ we guess exponential expressions $e_u$ and $e_v$ with $\text{eval}(e_u), \text{eval}(e_v) \in \Gamma^*$. We define exponential expressions $e_X = e_u X e_v$ and we define $\delta(X) = \text{eval}(e_X)$. For each $X \in \Omega \setminus \Omega'$ we guess an exponential expression $e_X$ with $\text{eval}(e_X) \in \Gamma^*$ and we define $\delta(X) = \text{eval}(e_X)$.

Next we verify whether or not $\delta_*(E) \equiv E'$. During this test we have to create an exponential expression $f_L$ (and $f_R$, respectively) by replacing $X$ in $e_L$ (and $e_R$, respectively) 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 exponential expressions over Boolean matrices and can be done in polynomial time.

The correctness of the algorithm follows from our general assumption (Remark 14) that all $X \in \Omega$ appear in $LR\overline{LR}$. Therefore, if we have $\delta_*(E) \equiv E'$, then every factor of $\delta(X)$ (or of $\delta(\overline{X})$) appears necessarily as a factor in $L'R' = \delta(LR)$. Hence every factor of $\delta(X)$ has an exponential expression of polynomial size by Lemma 17. □

**Remark 29.** Actually, the test for $\delta_*(E) \equiv E'$ can be performed in non-deterministic polynomial time by Remark 19.

Lemma 27 leads to the third and last rule.

**Rule 3.** *If $\delta$ 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 $\delta_*(E)$.*

**Example 30.** 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 $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. We strengthen constraints such that $X \in \Gamma^* c \Gamma^*$ 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 22 has a solution too.

### *4.8. The search graph and Plandowski's algorithm*

The input of Problem EWC is an equation with constraints. In order to fix notations we call it $E_0 = (\Gamma_0, h_0, \Omega_0, \rho_0; L_0 = R_0)$ and we let $d = |L_0R_0|$. According to Remark 14 we assume $|\Omega_0| \leqslant 2d$.

**Definition 31.** Let $p_0$ be a polynomial. The notion of *admissibility* is defined with respect to $p_0(\|E_0\|)$.

- An exponential expression $e$ is *admissible*, if $\|e\| \leqslant p_0(\|E_0\|)$.
- A base change $\beta : \Gamma' \to \Gamma^*$ is *admissible*, if $|\Gamma'| \leqslant p_0(\|E_0\|)$ and for all $a \in \Gamma'$ there is an admissible exponential expression for $\beta(a)$.
- An equation with constraints $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ is *admissible*, if $|\Gamma \setminus \Gamma_0| \leqslant p_0(\|E_0\|)$, $h(a) = h_0(a)$ for $a \in \Gamma_0$, and $e_Le_R$ is admissible.

In the following we assume that a polynomial $p_0$ (of large enough degree) has been fixed whenever we speak about admissibility. We do not calculate $p_0$ explicitly, but it will become clear from the context what *large enough* actually means.

**Definition 32.** The *search graph* of $E_0$ is a directed graph where nodes are admissible equations with constraints. For two nodes $E, E'$ there is an arc $E \to E'$, if there are an admissible base change $\beta$, a projection $\pi$, and a partial solution $\delta$ such that $\delta_*(\pi^*(E)) \equiv \beta_*(E')$.

**Lemma 33.** *Let $p_0$ be a polynomial of degree at least 1. The following problem is* PSPACE*-complete.*

*INPUT: Equations with constraints $E_0, E$, and $E'$ such that $E$ and $E'$ are admissible with respect to $p_0(\|E_0\|)$.*

*QUESTION: Is there an arc $E \to E'$ in the search graph of $E_0$?*

**Proof.** The arc from $E$ to $E'$ is established by applying Rules 2, 1, and 3 (in this order) to $E$. More precisely, we let $E_0 = (\Gamma_0, h_0, \Omega_0, \rho_0; L_0 = R_0)$, $E = (\Gamma,, \Omega, \rho; e_L = e_R)$, and $E' = (\Gamma', h', \Omega', \rho'; e_{L'} = e_{R'})$. We first guess some alphabet $(\Gamma'', \bar{\ })$ of polynomial size together with $h'' : \Gamma'' \to M_{2n}$ and we guess some admissible base change $\beta : \Gamma' \to \Gamma''^*$ such that $h' = h''\beta$. We compute $\beta_*(E')$. Next, we guess some equation with constraints $E''$ which uses $\Gamma''$ and $\Omega$. We check using Lemma 28 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 $\delta$ can be represented in polynomial space by Lemma 17.) Finally, we check using Proposition 26 that there is some projection $\pi : \Gamma'' \to \Gamma$ such that $\pi^*(E) \equiv E''$. We obtain $\delta_*(\pi^*(E)) \equiv \beta_*(E')$.

The PSPACE-hardness follows by Proposition 13 which shows that the problem is PSPACE–hard on instances of the following type: The equation for $E$ and $E'$ is $X = X$, we have $\rho(X) = \rho'(X) = A \in M_{2n}$, and $\Gamma' \setminus \Gamma = \{a, \bar{a}\}$ with $h'(a) = A$. □

**Remark 34.** Following Remarks 19 and 29, the problem presented in Lemma 33 can be decided in non-deterministic polynomial time, if the monoid $M_{2n}$ is not part of the input and the parameter $n$ is viewed as a constant.

On a high-level description, Plandowski's algorithm applies Rules 1 to 3 in a non-deterministic way until a trivial equation is found. An actual implementation of the algorithm depends on the chosen polynomial $p_0$ and it has the following structure.

```
begin
    $E := E_0$
    while $\Omega \neq \emptyset$ do
        Guess an equation with constraints $E'$, which is admissible with respect to $p_0(|E_0|)$
        Verify that $E \to E'$ is an arc in the search graph of $E_0$
        $E := E'$
    endwhile
    return "eval($e_L$) = eval($e_R$)"
end
```

Lemmata 21, 24 (*ii*), and 27 say that the algorithm returns *true* only if $E_0$ is solvable. The proof of Theorem 8 is therefore reduced to the statement that there is a polynomial $p_0$ such that for all $E_0$ we have, if $E_0$ is solvable, then the search graph contains a path to some solvable equation without variables.

**Remark 35.** If the arc $E \to E'$ 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 24 (*ii*). Since we are going to perform the search in a graph of at most exponential size, we automatically get 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 a singly exponential bound is conjectured), see [33].

### 4.9. Free intervals

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

It is convenient to have an involution on the set of intervals. If $[\alpha, \beta]$ is a positive interval, then $[\beta, \alpha]$ is also called a (non-positive) interval, and we define $w[\beta, \alpha] = \overline{w[\alpha, \beta]}$. Moreover, we define $w[\alpha, \alpha]$ to be the empty word. For all $0 \leqslant \alpha, \beta \leqslant m$ we let $\overline{[\alpha, \beta]} = [\beta, \alpha]$; therefore, $\overline{w[\alpha, \beta]} = w\overline{[\alpha, \beta]}$.

In the following we assume that the input equation $E_0$ has a solution $\sigma$ with $w_0 = \sigma(L_0) = \sigma(R_0)$ and $m_0 = |w_0|$. We have $w_0 \in \Gamma_0^*$, but in this section the alphabet $\Gamma_0$ is replaced by some other alphabet $\Gamma$, which turns out to be a set of non-empty words over $\Gamma_0$. We let $d$ be the denotational length of the equation and $L_0 = x_1 \cdots x_g$ and $R_0 = x_{g+1} \cdots x_d$, $x_i \in (\Gamma_0 \cup \Omega_0)$ for $1 \leqslant i \leqslant d$. We assume $2 \leqslant g < d < m_0$ whenever necessary. We also make the assumption that $\sigma(x_i) \neq 1$ for all $1 \leqslant i \leqslant d$. This assumption can be realized, e.g., in the preprocessing.

We are going to define an equivalence relation $\approx$ on the set of intervals of $w_0$. For this we need some preparation. For $i \in \{1, \ldots, d\}$ we define positions $\mathrm{l}(i)$ and $\mathrm{r}(i)$ such that $\sigma(x_i)$ starts in $w_0$ at the left position $\mathrm{l}(i)$ and it ends at the right position $\mathrm{r}(i)$. Formally, we define $\mathrm{l}(i) \in \{0, \ldots, m_0 - 1\}$ and $\mathrm{r}(i) \in \{1, \ldots, m_0\}$ by the congruences:

$$\mathrm{l}(i) \equiv |\sigma(x_1 \cdots x_{i-1})| \bmod m_0$$
$$\mathrm{r}(i) \equiv |\sigma(x_1 \cdots x_i)| \bmod m_0$$

We have $\mathrm{l}(1) = \mathrm{l}(g+1) = 0$ and $\mathrm{r}(g) = \mathrm{r}(d) = m_0$ since the range for the congruences are different for left- and right positions. 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 $1 \leqslant i \leqslant d$. The interval $[\mathrm{l}(i), \mathrm{r}(i)]$ is positive, because $\sigma(x_i) \neq 1$.

The set of l- and r-positions is called the set of *cuts*. Thus, the set of cuts is $\{\, \mathrm{l}(i), \mathrm{r}(i) \mid 1 \leqslant i \leqslant d \,\}$. The positions 0 and $m_0$ are cuts and there are at most $d$ cuts. These positions split the word $w_0$ into at most $d - 1$ factors.

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:

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

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* using this equivalence and cuts.

**Definition 36.** An interval $[\alpha, \beta]$ is *free*, if, whenever $[\alpha, \beta] \approx [\alpha', \beta']$, then there is no cut $\gamma'$ with $\min\{\alpha', \beta'\} < \gamma' < \max\{\alpha', \beta'\}$.

Clearly, the set of free intervals is closed under involution, i.e., $[\alpha,\beta]$ is free if and only if $[\beta,\alpha]$ is free. It is also clear that $[\alpha,\beta]$ is free if $|\beta-\alpha|\leqslant 1$.

Free intervals correspond to 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 solution where $m_0$ is minimal. In our setting we cannot avoid these factors.

**Example 37.** The last equation in Example 30, namely

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

has a solution which yields the word

$$w_0 = \overset{0}{|}\,\mathrlap{\underbrace{\phantom{a\,|\,bc\bar c}}_{Y}}a\,\overset{1}{|}\,\mathrlap{\overbrace{\phantom{bc\bar c\bar b\,|\,\bar b\,|\,abc}}^{X}}bc\bar c\bar b\,\overset{5}{|}\,\bar b\,\overset{6}{|}\,\mathrlap{\underbrace{\phantom{abc\,|\,\bar c}}_{Y}}abc\,\overset{9}{|}\,\mathrlap{\overbrace{\phantom{\bar c\bar b\,|\,\bar a\,|\,b\,|\,bc\bar c\bar b}}^{\overline{X}}}\bar c\bar b\,\overset{11}{|}\,\bar a\,\overset{12}{|}\,b\,\overset{13}{|}\,b\mathrlap{\underbrace{\phantom{c\bar c\bar b\,|\,\bar a}}_{\overline{Y}}}c\bar c\bar b\,\overset{17}{|}\,\bar a\,\overset{18}{|}\;.$$

The set of cuts is shown by the vertical 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 38.** *Let* $[\alpha,\beta]$ *be a free interval and* $\mu,\nu$ *having the property* $\min\{\alpha,\beta\}\leqslant\mu,\nu\leqslant\max\{\alpha,\beta\}$. *Then the interval* $[\mu,\nu]$ *is also free.*

**Proof.** We may assume that $\alpha\leqslant\mu<\nu\leqslant\beta$. By contradiction assume that $[\mu,\nu]$ is not free. Then there is some $k\geqslant 0$ and some cut $\gamma'$ 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 $k\geqslant 1$ 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, the same pair $x_i, x_j$ defines an interval $[\alpha_1,\beta_1]$ such that $[\alpha,\beta]\sim[\alpha_1,\beta_1]$ and $\min\{\alpha_1,\beta_1\}\leqslant\mu_1,\nu_1\leqslant\max\{\alpha_1,\beta_1\}$. Using induction on $k$ we see that $[\alpha_1,\beta_1]$ is not free. But then $[\alpha,\beta]$ is not free, and this is a contradiction. □

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 $0\leqslant\alpha<\beta\leqslant m_0$ such that $[\alpha,\beta]$ is not free. A position $\gamma$ with $\alpha<\gamma<\beta$ is called an *implicit cut* of $[\alpha,\beta]$, if there is a cut $\gamma'$ 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 $\alpha \leqslant \mu < \gamma < \nu \leqslant \beta$ and $\gamma$ is an implicit cut of $[\alpha,\beta]$, then $\gamma$ is also an implicit cut of $[\mu,\nu]$. In particular, neither $[\mu,\nu]$ nor $[\nu,\mu]$ is a free interval.

**Definition 39.** A free interval $[\alpha,\beta]$ is called *maximal free*, if there is no free interval $[\alpha',\beta']$ such that both, $\alpha' \leqslant \min\{\alpha,\beta\} \leqslant \max\{\alpha',\beta'\} \leqslant \beta'$ and $|\beta-\alpha| < \beta' - \alpha'$.

Lemma 40 states that maximal free intervals do not overlap.

**Lemma 40.** *Let* $0 \leqslant \alpha \leqslant \alpha' < \beta \leqslant \beta' \leqslant m_0$ *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 $\gamma$ with $\alpha < \gamma < \beta'$. By the observation above: If $\gamma < \beta$, then $\gamma$ is an implicit cut of $[\alpha,\beta]$ and $[\alpha,\beta]$ is not free. Otherwise, $\alpha' < \gamma$ and $[\alpha',\beta']$ is not free. □

Lemma 41 states the main observation of this section.

**Lemma 41.** *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* $\gamma$ *and* $\delta'$ *are cuts.*

**Proof.** We may assume that $\alpha < \beta$. We show the existence of $[\gamma,\delta]$ where $[\alpha,\beta] \approx [\gamma,\delta]$ and $\gamma$ is a cut. (The existence of $[\gamma',\delta']$ where $[\alpha,\beta] \approx [\gamma',\delta']$ and $\delta'$ is a cut follows by a symmetric argument.)

If $\alpha = 0$, then $\alpha$ is a cut and we can choose $\delta = \beta$. Hence let $1 \leqslant \alpha$ and consider the positive interval $[\alpha-1,\beta]$. This interval is not free and the only possible position for an implicit cut is $\alpha$. Thus, for some cut $\gamma$ 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 $\delta = \beta'$. □

In the following proposition the symbol $\Gamma$ refers to some set of factors of the word $w_0$. (Recall that $w_0 = \sigma(L_0) = \sigma(R_0)$ and $\sigma$ is a solution of the input equation $E_0$.) The set $\Gamma$ becomes the basic alphabet later.

**Proposition 42.** *Let* $\Gamma$ *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* $\Gamma$ *is a subset of* $\Gamma_0^+$ *of size at most* $2d-2$. *The set* $\Gamma$ *is closed under involution.*

**Proof.** Let $[\alpha,\beta]$ be maximal free. Then $|\beta-\alpha| \geqslant 1$ and $[\beta,\alpha]$ is also maximal free by definition. Hence $\Gamma \subseteq \Gamma_0^+$ and $\Gamma$ is closed under involution. By Lemma 41 we may assume that $\alpha$ 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 40. 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 $\alpha$ is a cut. □

For the moment let $\Gamma_0' = \Gamma_0 \cup \Gamma$ where $\Gamma \subseteq \Gamma_0^+$ is the set defined in Proposition 42. The inclusion $\Gamma_0' \subseteq \Gamma_0^+$ defines a natural projection $\pi : \Gamma_0' \to \Gamma_0^*$ and a mapping $h_0' : \Gamma_0' \to M_{2n}$ by $h_0' = h_0\pi$. Consider the equation with constraints $\pi^*(E_0)$, this is a node in the search graph, because the size of $\Gamma$ is linear in $d$.

The reason to switch from $\Gamma_0$ to $\Gamma_0'$ is that, due to the constraints, the word $w_0$ may have long free intervals, even in a minimal solution. Over $\Gamma_0'$ 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 $w_0$ into 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 all $1 \leqslant i \leqslant k$ and

$$w_0 = w_0[\alpha_0, \alpha_1] \cdots w_0[\alpha_{k-1}, \alpha_k].$$

Note that all cuts occur as some $\alpha_p$, therefore we can think of the factors $w_0[\alpha_{i-1}, \alpha_i]$ as letters in $\Gamma$ for $1 \leqslant i \leqslant k$. Moreover, all constants which appear in $L_0 R_0$ are elements of $\Gamma$, too. We replace $w_0$ by the word $w'_0 \in \Gamma^*$. Then we can define $\sigma' : \Omega_0 \to \Gamma^*$ such that both, $\sigma'(L_0) = \sigma'(R_0) = w'_0$ and $\rho_0 = h'_0 \sigma'$. In other words, $\sigma'$ is a solution of $\pi^*(E_0)$. We have $w_0 = \pi(w'_0)$ and $\exp(w'_0) \leqslant \exp(w_0)$. The crucial point is that $w'_0$ has no long free intervals anymore. With respect to $w'_0$ and $\Gamma'_0$ all maximal free intervals have length exactly one.

**Example 43.** Following Example 37, 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 free intervals have length one.

In the next step we show that we can reduce the alphabet of constants to be $\Gamma$. The inclusion of $\Gamma$ into $\Gamma'_0$ 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 $\delta$ to be the identity. The equation with constraints $E'_0$ has a solution $\sigma'$ with $\sigma'(L_0) = w'_0$ and $\exp(w'_0) \leqslant \exp(w_0)$.

In order to avoid an excess of notation we identify $E_0$ and $E'_0$, hence we also assume $\sigma = \sigma'$ and $w_0 = w'_0$. However, as a reminder that we have changed the alphabet of constants (recall that some words became letters), we prefer to use the symbol $\Gamma$ rather than $\Gamma_0$. Thus, in what follows we use the following.

**Assumption 44.** The input equation $E_0$ satisfies the following conditions:

$$\begin{aligned}
E_0 &= (\Gamma, h, \Omega_0, \rho_0; L_0 = R_0),\\
L_0 &= x_1 \cdots x_g \text{ and } g \geqslant 2,\\
R_0 &= x_{g+1} \cdots x_d \text{ and } d > g,\\
|\Gamma| &\leqslant 2d - 2,\\
|\Omega_0| &\leqslant 2d.
\end{aligned}$$

Moreover, all variables $X \in \Omega_0$ occur in $L_0 R_0 \overline{L_0 R_0}$. There is a solution $\sigma$ and a word $w_0$ with $|w_0| = m_0$ and $\exp(w_0) \in 2^{\mathcal{O}(d + n \log n)}$ such that $w_0 = \sigma(L_0) = \sigma(R_0)$ with $\sigma(X_i) \neq 1$ for $1 \leqslant i \leqslant d$ and $\rho_0 = h\sigma : \Omega_0 \to M_{2n} \subseteq \mathbb{B}^{2n \times 2n}$. All maximal free intervals have length exactly one, i.e., every positive interval $[\alpha, \beta]$ with $\beta - \alpha > 1$ contains an implicit cut.

### 4.10. Critical words

For each $1 \leqslant \ell \leqslant m_0$ we define the set of *critical words* $C_\ell$ by

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

We have $1 \leqslant |C_\ell| \leqslant 2d - 4$ and $C_\ell$ is closed under involution. Each word $u \in C_\ell$ has length $2\ell$, it can be written in the form $u = u_1 u_2$ with $|u_1| = |u_2| = \ell$. The word $u_1$ (respectively, $\overline{u_2}$) appears as a suffix, to the left of some cut and $u_2$ (respectively, $\overline{u_1}$) appears as a prefix, to the right of the same cut.

By $B_\ell$ we denote the set of triples $(u, w, v) \in (\{1\} \cup \Gamma^\ell) \times \Gamma^+ \times (\{1\} \cup \Gamma^\ell)$ which satisfy the following four conditions:

(1) No factor of the word $w$ belongs to $C_\ell$.
(2) If a factor of the word $uwv$ belongs to $C_\ell$, then this factor is a prefix or a suffix of $uwv$.
(3) If $u \neq 1$, then a prefix of $uwv$ of length $2\ell$ belongs to $C_\ell$,
(4) If $v \neq 1$, then a suffix of $uwv$ of length $2\ell$ belongs to $C_\ell$.

The set $B_\ell$ 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 morphism $\pi_\ell : B_\ell^* \to \Gamma^*$ by $\pi_\ell(u, w, v) = w \in \Gamma^+$. It is extended to a projection $\pi_\ell : (B_\ell \cup \Gamma)^* \to \Gamma^*$ by leaving $\Gamma$ invariant. We define $h_\ell : (B_\ell \cup \Gamma)^* \to M_{2n}$ by $h_\ell = h\pi_\ell$, i.e., $h_\ell(a) = h(a)$ for $a \in \Gamma$ and $h_\ell(u, w, v) = h(w)$ for $(u, w, v) \in B_\ell$. The symbols $\pi_\ell$ and $h_\ell$ are also used for restrictions of the morphisms $\pi_\ell$ and $h_\ell$.

Later we consider finite sets $\Gamma_\ell, \Gamma_{\ell,\ell'}$ such that $\Gamma \subseteq \Gamma_\ell \subseteq \Gamma_{\ell,\ell'} \subseteq B_\ell \cup \Gamma$. Then $\pi_{\ell,\ell'} : \Gamma_{\ell,\ell'}^* \to \Gamma_\ell^*$ denotes the projection given by $\pi_{\ell,\ell'}(u, w, v) = w \in \Gamma^*$ for $(u, w, v) \in \Gamma_{\ell,\ell'} \setminus \Gamma_\ell$ and $\pi_{\ell,\ell'}(u, w, v) = (u, w, v)$ for $(u, w, v) \in \Gamma_\ell$. By $h_\ell : \Gamma_\ell^* \to M_{2n}$ and $h_{\ell,\ell'} : \Gamma_{\ell,\ell'}^* \to M_{2n}$ we denote the restrictions of $h_\ell : (B_\ell \cup \Gamma)^* \to M_{2n}$. We have $h_{\ell,\ell'} = h_\ell \pi_{\ell,\ell'}$.

For every non-empty word $w \in \Gamma^+$ we define its *$\ell$-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 $1 \leqslant i \leqslant k$ the following conditions are satisfied (see Fig. 1):

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

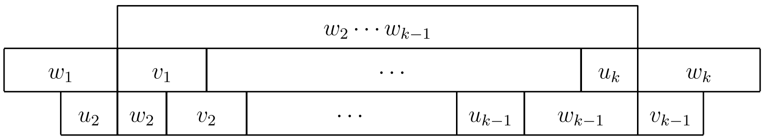

Fig. 1. An $\ell$-factorization.

Note that the $\ell$-factorization of a word $w$ is unique. For $k \geqslant 2$ we have $|w_1| \geqslant \ell$ and $|w_k| \geqslant \ell$, 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 $\ell$-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 $\ell$-factorization

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

in $B_\ell^*$ and $\Gamma^*$ 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 $k \geqslant 2$ (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 $\ell$-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 $1 \leqslant p < q \leqslant k$ such that:

$$F_\ell(uwv) = (u_1, w_1, v_1) \cdots (u_p, w_p, v_p)\mathrm{Body}_\ell(w)(u_q, w_q, v_q) \cdots (u_k, w_k, v_k),$$

$$w_1 \cdots w_p = u\,\mathrm{head}_\ell(w), \quad \text{and} \quad w_q \cdots w_k = \mathrm{tail}_\ell(w)v.$$

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 $\mathrm{Body}_\ell(\overline{w}) = \overline{\mathrm{Body}_\ell(w)}$.

### *4.11. The $\ell$-transformation*

By Assumption 44, $E_0 = (\Gamma, h, \Omega_0, \rho_0; x_1 \cdots x_g = x_{g+1} \cdots x_d)$, and the equation has a solution $\sigma$ where $w_0 = \sigma(x_1 \cdots x_g) = \sigma(x_{g+1} \cdots x_d)$ and $m_0 = |w_0|$. We let $1 \leqslant \ell \leqslant m_0$ and we consider the $\ell$-factorization of the word $w_0$:

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

A sequence $S = (u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ with $1 \leqslant p \leqslant q \leqslant k$ is called an *$\ell$-factor* . We say that $S$ is a *cover* of a positive interval $[\alpha, \beta]$, if both, $|w_1 \cdots w_{p-1}| \leqslant \alpha$ and $|w_{q+1} \cdots w_k| \leqslant m_0 - \beta$. That is, $w_0[\alpha, \beta]$ is a factor of $w_p \cdots w_q$. It is a *minimal cover* , if neither the sequence $(u_{p+1}, w_{p+1}, v_{p+1}) \cdots (u_q, w_q, v_q)$ 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 $L_\ell$ we consider those $1 \leqslant i \leqslant g$ where $\mathrm{body}_\ell(\sigma(x_i)) \neq 1$. Note that this implies $x_i \in \Omega_\ell$ since $\ell \geqslant 1$ and 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 $\beta = \mathrm{r}(i) - |\mathrm{tail}_\ell(\sigma(x_i))|$. We have $w_0[\alpha, \beta] = \mathrm{body}_\ell(\sigma(x_i))$. Next consider the $\ell$-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 $1 < p \leqslant q < k$ and $w_p \cdots w_q = w_0[\alpha, \beta] = \mathrm{body}_\ell(\sigma(x_i))$. The value of $S_i$ depends only on $x_i$, but not on the choice of the index $i$. This means $S_i = S_j$ whenever $x_i = x_j$.

We replace the $\ell$-factor $S_i$ in $F_\ell(w_0)$ by the variable $x_i$. Having done this for all $1 \leqslant i \leqslant g$ 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 $\ell$-transformation $E_\ell$. For $R_\ell$ we proceed analogously by replacing those $\ell$-factors $S_i$ where $\mathrm{body}_\ell(\sigma(x_i)) \neq 1$ and $g + 1 \leqslant i \leqslant d$.

For $E_\ell$ we cannot use the alphabet $B_\ell$, because it might be too large (even infinite). Therefore we let $\Gamma'_\ell$ be the smallest subset of $B_\ell$ which is closed under involution and which satisfies $L_\ell R_\ell \in (\Gamma'_\ell \cup \Omega_\ell)^*$.

We let $\Gamma_\ell = \Gamma'_\ell \cup \Gamma$. (We allow $\Gamma$ because the constants of $\Gamma$ make it easy to cope with the constraints.) Recall that $h_\ell(u, w, v) = h(w)$ for $(u, w, v) \in \Gamma_\ell \setminus \Gamma$ and $h_\ell(a) = h(a)$ for $a \in \Gamma$. Finally, we define the mapping $\rho_\ell : \Omega_\ell \to M_{2n}$ by $\rho_\ell(X) = h(\mathrm{body}_\ell(\sigma(X)))$. The reason is that we know $\rho(X) = h(\sigma(X))$. We can write $\sigma(X) = u\,\mathrm{body}_\ell(\sigma(X))v$, hence $h(\sigma(X)) = h(u)\rho_\ell(X)h(v)$.

The steps above define the $\ell$-transformation and yield the following equation:

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

**Example 45.** We continue with our example $a\overline{X}\overline{X}\bar{a} = Y\bar{b}Y\bar{a}b\overline{Y}$ and the solution $\sigma$ which has been given by

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

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:

$$\begin{aligned} &(1, ad\bar{d}, \bar{b}a)(d\bar{d}, \bar{b}, ad)\ (\bar{d}\bar{b}, ad, \bar{d}\bar{a}) \\ &\qquad\qquad (ad, \bar{d}, \bar{a}b)\ \ (d\bar{d}, \bar{a}, bd)(\bar{d}\bar{a}, b, d\bar{d})(\bar{a}b, d\bar{d}\bar{a}, 1). \end{aligned}$$

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

Let us rename the letters:

$$a = (1, ad\bar{d}, \bar{b}a)$$
$$b = (\bar{d}\bar{a}, b, d\bar{d})$$
$$c = (\bar{d}\bar{b}, ad, \bar{d}\bar{a})$$
$$d = (ad, \bar{d}, \bar{a}b)$$
$$e = (d\bar{d}, \bar{a}, bd)$$

After this renaming the 2-factorization of $w_0$ becomes $a\bar{b}cdeb\bar{a}$ and the equation $E$ reduces to $E_2 : aXcde\overline{X}\bar{a} = a\bar{b}cdeb\bar{a}$ 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 46.** (i) In the extreme case $\ell = m_0$, the $\ell$-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 $h_{m_0}(a) = h(w_0) \in M_{2n}$. Since $\Omega_{m_0} = \emptyset$, the equation with constraints $E_{m_0}$ trivially has 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 $\ell = 1$. We develop the technical details as an example. Consider a triple $(u, w, v) \in \Gamma_1$ which appears in $F_1(w_0)$. Then $w = w_0[\alpha, \beta]$ for some $\beta - \alpha \geqslant 1$. All maximal free intervals have length 1 (by Assumption 44). Assume $\beta - \alpha \geqslant 2$, then $[\alpha, \beta]$ would contain an implicit cut $\gamma$ and $w_0[\gamma - 1, \gamma + 1] \in C_1$. But no critical word is a factor of $w$, $\beta - \alpha = 1$. An immediate consequence is $|\Gamma_1| \leqslant (|\Gamma| + 1)^3 \in \mathcal{O}(d^3)$, since $|\Gamma| \leqslant 2d - 2$. (More precisely, we could bound $|\Gamma_1|$ by $6d$, but $|\Gamma_1| \in \mathcal{O}(d^3)$ is good enough for our purpose.) Let $X \in \Omega_0$. Then $\mathrm{Body}_1(\sigma(X)) \neq 1$ if and only if $|\sigma(X)| \geqslant 3$. 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 $1 \leqslant i \leqslant g$ 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 $|w_i| \geqslant 3$ 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 $|L_1R_1| \leqslant 3|L_0R_0| = 3d$, and $E_1$ is admissible.

By Remark 46 the equations $E_1$ and $E_{m_0}$ are admissible and hence nodes of the search graph of $E_0$. The goal is to reach $E_{m_0}$, but it is not clear yet, neither that the $\ell$-transformations with $1 < \ell < m_0$ belong to the search graph nor that there are arcs from $E_0$ to $E_1$ or from $E_1$ to $E_2$ and so on. We prove these statements in the next sections.

### *4.12. The $\ell$-transformation $E_\ell$ is admissible*

**Proposition 47.** *There is a polynomial $p_0$ (of degree 4) such that each $E_\ell$ is admissible with respect to $p_0$ for all $\ell \geqslant 1$.*

**Proof.** The input size is $d + n + \log_2(|\Gamma| + |\Omega_0|)$. We have $|\Gamma| + |\Omega_0| \leqslant 4d - 2$ and $E_0 = (\Gamma, h, \Omega_0, \rho_0; x_1 \cdots x_g = x_{g+1} \cdots x_d)$. The constraints are Boolean $n \times n$-matrices and $d$ is the length of the equation. It is enough to show that $L_\ell$ and $R_\ell$ can be represented by exponential expressions of size $\mathcal{O}(d^2(d + n \log n))$. Then $\Gamma_\ell$ 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 $L_\ell$, only.

We start again with the $\ell$-transformation

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

If $k$ is small there is nothing to do since $|L_\ell| \leqslant |F_\ell(w_0)|$. An easy reflection shows that $|L_\ell|$ can become large, only if there is some $1 \leqslant i \leqslant g$ 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 $1 \leqslant i \leqslant g$, $\alpha = \mathrm{l}(i)$, 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 $[\alpha, \beta]$. (The definition of a minimal cover has been given at the beginning of Section 4.11.) We may assume that $q - p$ is large. It is enough to show that the $\ell$-factor

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

has an exponential expression of 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 $C_\ell$ can appear as a factor inside $w_p \cdots w_q$. This means there is some $p \leqslant s \leqslant q$ 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 $p \leqslant s \leqslant q$ be minimal such that $|w_p \cdots w_s| \geqslant \ell$. Then $|w_{s+1} \cdots w_q| \geqslant \ell$ 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

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

Since $(u_s, w_s, v_s) \in \Gamma_\ell$ is a letter, it is enough to show that there are exponential expressions for $S_i$ of size $\mathcal{O}(d(d + n \log n))$ for $i = 1, 2$. This follows from Lemma 48 with $c = 1$. □

The statement of Lemma 48 is more general than needed for the proof of Proposition 47, but later it is used for other values of $c$. In fact, it will be used for $c \leqslant 32d$.

**Lemma 48.** *Let $c > 0$ be a value which might depend on $d$ (and $n$) and let*

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

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

**Proof.** Clearly, we may assume $k > 3$. We show that there is an exponential expression of size $\mathcal{O}(d(d + n \log n))$ under the assumption $|w_1 \cdots w_k| < \ell$. (Note that $c$ has been removed from the

$\mathcal{O}$–term.) This is enough, because we can write $S$ as $a_0S_1a_1\cdots S_{c'}a_{c'}$, where $c' \leqslant c$, the $a_i$ are letters, and each $S_i$ satisfies the assumption. Due to the factorization we may also assume $u_1 \neq 1 \neq v_k$ and therefore we may define $u_{k+1}$ as the suffix of length $\ell$ of $u_1w_1\cdots w_k$. For $1 \leqslant i \leqslant k$ let $z_i = u_{i+1}v_i$. Then $z_i \in C_\ell$ is a critical word which appears as a factor in $z = u_1w_1w_2\cdots w_kv_k$. If the words $z_i$, $1 \leqslant i < k$ are pairwise different, then $k-1 \leqslant |C_\ell| \in \mathcal{O}(d)$ 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 $1 \leqslant i < j$ the words $z_1,\ldots,z_{j-1}$ 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 $u_1w_1\cdots w_i(w_{i+1}\cdots w_j)^rv_j$ 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

$$z = u_1w_1\cdots w_i(w_{i+1}\cdots w_j)^r w_s\cdots w_kv_k.$$

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 $j \neq t$. Then both, $|w_{i+1}\cdots w_j|$ and $|w_{j+1}\cdots w_t|$ are periods of $z_i$, but $|w_{i+1}\cdots w_t| \leqslant |z_i|$. Hence by Fine and Wilf's Theorem [23] 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 $\ell$-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

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

where $S' = (u_s,w_s,v_s)\cdots(u_k,w_k,v_k)$ for $s = i+1+r(j-i)$ and $r \geqslant 1$. We have $r \leqslant \exp(w_0)$, 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 $j + \lceil\log_2(r)\rceil \in \mathcal{O}(d+n\log n)$. More precisely, we can effectively calculate some constant $\widetilde{c}$ such that $j + \lceil\log_2(r)\rceil \leqslant \widetilde{c}(d+n\log_2 n)$.

We have $|\{z_s,\ldots,z_k\}| < |\{z_1,\ldots,z_k\}|$. Therefore by induction we may assume that the sequence $S' = (u_s,w_s,v_s)\cdots(u_k,w_k,v_k)$ has an exponential expression of size at most $|\{z_s,\ldots,z_k\}|\widetilde{c}(d+n)\log_2 n$. Hence the exponential expression for $S$ has size at most

$$\begin{aligned}&\widetilde{c}(d+n\log_2 n) + |\{z_s,\ldots,z_k\}|\widetilde{c}(d+n\log_2 n)\\&\qquad\qquad \leq |\{z_1,\ldots,z_k\}|\widetilde{c}(d+n\log_2 n).\end{aligned}$$

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

**Remark 49.** At this stage we know that all $\ell$-transformations are admissible with respect to some suitable polynomial $p_0$ of degree 4. Next we show that we can modify the polynomial $p_0$ such that the search graph also contains arcs $E_0 \to E_1$ and $E_\ell \to E_{\ell'}$ for $1 \leqslant \ell < \ell' \leqslant 2\ell$. For this reason we use the notion of admissibility with respect to the 4-th power $p_0^4$ of $p_0$. Thus, admissibility is meant with respect to a polynomial of degree 16.

### 4.13. The arc from $E_0$ to $E_1$

We present the formal construction of the arc from $E_0$ to $E_1$. We give all technical details since this arc is the model for the more complicated way the other arcs are constructed in the search graph.

An explicit description of $E_1 = (\Gamma_1, h_1, \Omega_1, \rho_1; L_1 = R_1)$ has been given in Remark 46. The letters of $\Gamma_1$ can be written either as $(a, b, c)$ or as $b$ with $a, c \in \Gamma \cup \{1\}$ and $b \in \Gamma$. We define an admissible base change $\beta : \Gamma_1 \to \Gamma$ by $\beta(a, b, c) = b$ and $\beta(b) = b$ for $b \in \Gamma$. Trivially, $h_1 = h\beta$. 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 $|\sigma(X)| \leqslant 2$, then we let $\delta(X) = \sigma(X)$. For $|\sigma(X)| \geqslant 3$ 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, by definition, $\rho_1(X) = h(\mathrm{body}_1(\sigma(X)))$, 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 $\pi$ is the identity. This yields the arc from $E_0$ to $E_1$.

### 4.14. The equation $E_{\ell,\ell'}$ for $1 \leqslant \ell < \ell' \leqslant 2\ell$

To establish the arcs from $E_\ell$ to $E_{\ell'}$ for all $1 \leqslant \ell < \ell' \leqslant 2\ell$ we use an intermediate equation $E_{\ell,\ell'}$ such that there is an admissible base change $\beta$, a projection $\pi$, and a partial solution $\delta$ with

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

The way we move from $E_\ell$ to $E_{\ell'}$ is visualized in Fig. 2.

We begin with the definition of the base change $\beta$. Recall $\Gamma \subseteq \Gamma_{\ell'} \subseteq B_{\ell'} \cup \Gamma$. As expected, we define $\beta(a) = a$ for $a \in \Gamma$. 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 in the $\ell'$-factorization $F_{\ell'}(w_0)$. Therefore we find a positive interval $[\alpha_0, \beta_0]$ such that $w = w_0[\alpha_0, \beta_0]$ and such that the following two conditions are satisfied:

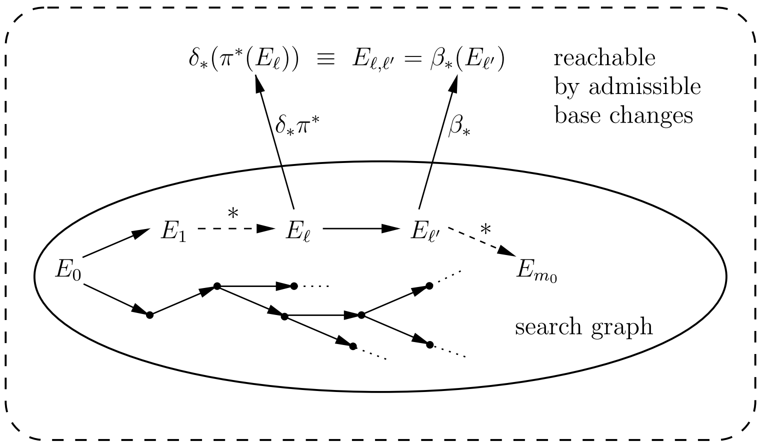

Fig. 2. The search graph and its neighborhood.

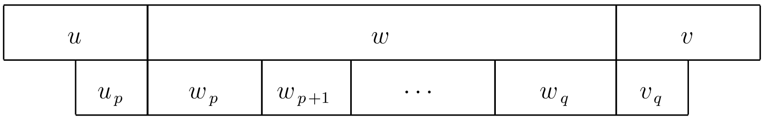

Fig. 3. An $\ell$-factorization of $uwv$.

(1) We have $u = 1$ and $\alpha_0 = 0$ or $|u| = \ell'$, $\alpha_0 \geqslant \ell'$, and $u = w_0[\alpha_0 - \ell', \alpha_0]$.
(2) We have $v = 1$ and $\beta_0 = m_0$ or $|v| = \ell'$, $\beta_0 \leqslant m_0 - \ell'$, and $v = w_0[\beta_0, \beta_0 + \ell']$.

Let $(u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ be the $\ell$-factor which is the minimal cover of $[\alpha_0, \beta_0]$ with respect to the $\ell$-factorization $F_\ell(w_0)$. Since $\ell \leqslant \ell'$ 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

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

We have the picture shown in Fig. 3.

The definition does not depend on the choice of $[\alpha_0, \beta_0]$ as long as $0 \leqslant \alpha_0 < \beta_0 \leqslant m_0$ 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 $\Gamma$ and it is closed under involution (since $\Gamma_{\ell'}$ has this property). An easy reflection shows that $\Gamma_\ell \subseteq \Gamma_{\ell,\ell'}$. This will become essential in Section 4.15.

We view $\beta$ as a morphism $\beta : \Gamma_{\ell'}^* \to \Gamma_{\ell,\ell'}^*$ and we have $h_{\ell,\ell'} \beta = h_\ell$. Define $E_{\ell,\ell'} = \beta_*(E_{\ell'})$. Then

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

Let us show that $\beta$ is admissible. 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 $\ell$-factor above

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

Using the same terminology as above there is some positive interval $[\alpha_0, \beta_0]$ such that $w_p \cdots w_q = w_0[\alpha_0, \beta_0]$, the word $u$ is a suffix of $w_0[0, \alpha_0]$, and $v$ is a prefix of $w_0[\beta_0, m_0]$. If $q - p$ is small, there is nothing to do. By Lemma 48 we may also assume that $\beta_0 - \alpha_0 > 32d\ell$. We inductively define a sequence of positions

$$\alpha_0 \leqslant \alpha_1 \leqslant \cdots \leqslant \alpha_i \leqslant \cdots \leqslant \beta_i \leqslant \cdots \leqslant \beta_1 \leqslant \beta_0.$$

In each step 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 $\beta_i - \alpha_i \geqslant 2$. The interval $[\alpha_i, \beta_i]$ is not free. Hence, there is some implicit cut $\gamma_i$ 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 $\gamma_i$ corresponds to some real cut $\gamma_i'$ 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 $(\gamma'_i, r_i, s_i)$ 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 of some $\gamma(i)$. Thus we find $0 \leqslant i < j < 4d$ 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 $|W_0| - |W_j| < 4d\ell' \leqslant 8d\ell$ and due to $|W_0| > 32d\ell$ the sequence above really exists, moreover $|W_j| > 24d\ell$.

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 $\gamma \leqslant \mu < \nu \leqslant \beta'$ we have $W_j = w_0[\mu, \nu]$ and we can assume that $\mu - \gamma < (j-i)\ell' \leqslant 4d\ell' - \ell' \leqslant 8d\ell - \ell'$. 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 may choose $W = w_0[\mu', \mu]$ and it follows that we can write $W_j = W^e W'$ such that $1 \leqslant |W| < 8d\ell$ and $W'$ is a prefix of $W$.

Putting everything together, we arrive in all cases at a factorization $W_0 = UW^eV$ with $e \leqslant \exp(w_0)$, $1 \leqslant |W| < 8d\ell$, and $|U| + |V| < 16d\ell$.

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, $p < r \leqslant q+1$, and $s$ maximal $p - 1 \leqslant s < q$ such that $|w_p \cdots w_{r-1}| > |U| + \ell$ and $|w_{s+1} \cdots w_q| > |V| + \ell$. By Lemma 48 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 $W^e$. Hence we may factorize $W = W'W''$ in such a way that after replacing $W$ by $W''W'$, we may assume that $u_r w_r w_{r+1} \cdots w_s v_s$ is in fact a prefix of $W^e$. Furthermore, we may assume that $w_r w_{r+1} \cdots w_s > 32d\ell$ and by symmetry we may choose some positive interval $[\alpha_0, \beta_0]$ such that $w_0[\alpha_0, \beta_0] = u_r w_r w_{r+1} \cdots w_s v_s$. Clearly, we have $w_0[i, j] = w_0[i + |W|, j + |W|]$ for all $\alpha_0 \leqslant i < j \leqslant \beta_0 - |W|$. In particular, the critical word $w_0[\alpha_0, \alpha_0 + 2\ell]$ appears as $w_0[\alpha_0 + |W|, \alpha_0 + 2\ell + |W|]$ again. This means that there is some $r \leqslant t < s$ such that $|w_r \cdots w_t| = |W|$. More precisely, we can choose $r \leqslant t < t' \leqslant s$ and a maximal $e' \leqslant e$ 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 $e' \leqslant \exp(w_0)$, $|w_r \cdots w_t| = |W|$, and $|w_{t'} \cdots w_s| \leqslant |W|$, the existence of an admissible exponential expression for $\beta(u, w, v)$ follows. Hence $\beta$ is an admissible base change.

### *4.15. Passing from $E_\ell$ to $E_{\ell,\ell'}$ for $1 \leqslant \ell < \ell' \leqslant 2\ell$*

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 worry about admissibility anymore. Once $\delta_*(\pi^*(E_\ell)) \equiv E_{\ell,\ell'}$ is established, Theorem 8 is proved.

For the definition of the projection $\pi$ consider a letter in $\Gamma_{\ell,\ell'} \setminus \Gamma_\ell$. Such a letter has the form $(u, w, v) \in B_\ell$ with $w \in \Gamma^+$. There is no length bound on $w$ known (or needed). We define $\pi(u, w, v) = w$ and this is possible since $\Gamma \subseteq \Gamma_\ell$.

Clearly $\pi(\overline{(u, w, v)}) = \overline{\pi(u, w, v)}$ and $h_{\ell,\ell'}(u, w, v) = h(w) = h_\ell(\pi(u, w, v))$. 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 $1 \leqslant i \leqslant g$. Hence $\sigma(X) = w_0[\mathrm{l}(i), \mathrm{r}(i)]$.

Let $\alpha_X = \mathrm{l}(i) + |\mathrm{head}_\ell(\sigma(X))|$ and $\beta_X = \mathrm{r}(i) - |\mathrm{tail}_\ell(\sigma(X))|$. Then $\mathrm{l}(i) + \ell \leqslant \alpha_X < \beta_X \leqslant \mathrm{r}(i) - \ell$. Let $(u_p, w_p, v_p) \cdots (u_q, w_q, v_q)$ be the minimal cover of $[\alpha_X, \beta_X]$ with respect to the $\ell$-factorization. We have $w_p \cdots w_q = \mathrm{body}_\ell(\sigma(X))$.

For $\mathrm{body}_{\ell'}(\sigma(X)) = 1$ we have $X \in \Omega_\ell \setminus \Omega_{\ell'}$ and we define

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

It follows that $\delta(X) \in B^*_\ell$ 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 $[\alpha_X, \beta_X]$ with respect to the $\ell'$-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 $\mathrm{body}_{\ell'}(\sigma(X)) \neq 1$ 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 $p \leqslant r \leqslant s \leqslant q$ we have $w_0[\alpha_X, \mu] = w_p \cdots w_{r-1}$, $w_0[\nu, \beta_X] = w_{s+1} \cdots w_q$, 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, $\delta(L_\ell) = \beta(L_{\ell'})$ and $\delta(R_\ell) = \beta(R_{\ell'})$. Hence $\delta_*(\pi^*(E_\ell)) \equiv \beta_*(E_{\ell'})$. The final step in proving Theorem 8 is completed.

## 5. Concluding remarks

The PSPACE-hardness stated in Theorems 3, 5, and 8 is due to rational constraints, but this is a side effect and a nice coincidence, only. The reason to include constraints has been motivated by possible applications to free partially commutative groups (graph groups). When Matiyasevich showed in 1996 that the existential theory of equations in free partially commutative monoids (trace monoids) is decidable [29,30,7], it became clear by his method that regular constraints are a powerful tool in order to extend decidability results to other algebraic structures and, in particular, they are necessary for extending Makanin's result from free groups to graph groups. At that time the result of Schulz [38] on word equations with regular constraints had been available but no such analogue for equations in free groups was known. So, the idea was to look for such an analogue first. Inspired by Gutierrez [15] we were finally led to investigate free monoids with involution and regular constrains. This approach turned out to be fruitful. Based on Theorem 5 of this paper (the results date back to the year 2000) it is shown in [8] that the existential theory of equations in graph groups is decidable. This result is a common generalization of Matiyasevich's decidability result on trace monoids and Makanin's result on free groups. In a continuation of this work on graph groups we obtained various other decidability results about the existential and positive theories in graph products, see [5,6].

Makanin has also shown that the positive theory in free groups is decidable [26]. It remains decidable with recognizable constraints [6]. In contrast, the positive theory of equations with rational constraints is undecidable in free groups, because the positive $\forall\exists^3$-theory of word equations is undecidable [27,9] and $\Sigma^*$ is a rational subset of the free group $F(\Sigma)$. So the question remains under which restricted type of constraints the positive theory of equations in free groups remains decidable.

## 6. Appendix. Proof of Proposition 15

We first repeat the statement of Proposition 15:

*Let* $E = (\Gamma, h, \Omega, \rho; L = R)$ *be a solvable equation with constraints. Then there is a solution* $\sigma : \Omega \to \Gamma^*$ *such that* $\exp(\sigma(L)) \in 2^{\mathcal{O}(d+n\log n)}$.

**Proof.** Let $p \in \Gamma^+$ be a primitive word. This means that $p \neq r^k$ for all $k > 1$ and $r \in \Gamma^*$. In the following we also write $p^{-1}$ instead of $\overline{p}$. Then, $p^{-3}$ for example means the same as $\overline{p}^3$. The definition of the *p-stable normal form* depends on whether or not $\overline{p}$ is a factor of $p^2$. So we distinguish two cases.

**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 $p^\alpha$ with $\alpha \geqslant 2$ by a sequence $p, \alpha - 2, p$ and each maximal factor of the form $\overline{p}^\alpha$ with $\alpha \geqslant 2$ by a sequence $\overline{p}, -(\alpha - 2), \overline{p}$.

The *p-stable normal form* (first kind) of $w \in \Gamma^*$ is a shortest sequence ($k$ is minimal)

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

such that $k \geqslant 0$, $u_0, u_i \in \Gamma^*$, $\varepsilon_i \in \{+1, -1\}$, $\alpha_i \geqslant 0$ for $1 \leqslant i \leqslant k$, 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 $k \geqslant 1$, then:

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

If $(u_0, \varepsilon_1\alpha_1, u_1, \ldots, \varepsilon_k\alpha_k, u_k)$ is the $p$-stable normal form of the word $w$, then the $p$-stable normal form of the word $\overline{w}$ becomes $(\overline{u_k}, -\varepsilon_k\alpha_k, \overline{u_{k-1}}, \ldots, -\varepsilon_1\alpha_1, \overline{u_0})$.

**Example 50.** 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:

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

**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, compare e.g. with [3].

The idea is to replace each maximal factor of the form $(rs)^\alpha r$ with $\alpha \geqslant 2$ by a sequence $rs, \alpha - 2, sr$. In this notation $\alpha - 2$ is representing the factor $(rs)^{\alpha-2}r = p^{\alpha-2}r = r\overline{p}^{\alpha-2} = rp^{2-\alpha}$.

The *$p$-stable normal form* (second kind) of $w \in \Gamma^*$ is the shortest sequence ($k$ is minimal)

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

such that $k \geqslant 0$, $u_0, u_i \in \Gamma^*$, $\alpha_i \geqslant 0$ for $1 \leqslant i \leqslant k$, 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 $k \geqslant 1$, then:
  $u_0 \in \Gamma^* rs \setminus (\Gamma^* p^2 r \Gamma^* \cup \Gamma^* rsrs)$,
  $u_i \in (\Gamma^* rs \cap sr\Gamma^*) \setminus (srsr\Gamma^* \cup \Gamma^* p^2 r \Gamma^* \cup \Gamma^* rsrs)$ for $1 \leqslant i < k$,
  $u_k \in sr\Gamma^* \setminus (\Gamma^* p^2 r \Gamma^* \cup srsr\Gamma^*)$.

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

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

So, for the second kind no negative integers interfere.

**Example 51.** Let $p = a\overline{a}b$ with $b = \overline{b}$. Then $r = a\overline{a}$ and $s = b$. Let $w = \overline{a}bp^4ap^3a$. Then the $p$-stable normal form of $w$ is:

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

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 $\alpha_i$ are integers.

It is well-known [28] that for Boolean matrices $A \in \mathbb{B}^{n \times n}$ we have $A^{n!} = A^{n!}A^{n!}$. Hence the matrix $A^{n!}$ is idempotent. For the following we define and fix $c(M_{2n}) = \max\{4, n!\}$. This choice guarantees $h(uv^{c(M_{2n})}w) = h(uv^{2c(M_{2n})}w)$ for all $u, v, w \in \Gamma^*$ and all $h : \Gamma^* \to M_{2n}$ and, of course, $c(M_{2n}) \geqslant 3$. The fact $c(M_{2n}) \geqslant 3$ is used at some point below.

Now, let $w, w' \in \Gamma^*$ be two words whose $p$-stable normal forms are identical up to the position of the $i$th integer. Assume that in the $p$-stable normal $w$ at this position there is the integer $\alpha_i$ and that for $w'$ at this position there is $\alpha'_i$. We know $h(w) = h(w')$ as soon as the following conditions are satisfied: $\alpha_i \cdot \alpha'_i > 0$, $|\alpha_i| \geqslant c(M_{2n})$, $|\alpha'_i| \geqslant c(M_{2n})$, and $\alpha_i \equiv \alpha'_i \pmod{c(M_{2n})}$. It is therefore convenient to change the syntax of the $p$-stable normal form. Each non-zero integer $\alpha'$ is written as $\alpha' = \varepsilon(q + \alpha c(M_{2n}))$ where $\varepsilon, q, \alpha$ are uniquely defined by $\varepsilon \in \{+1, -1\}$, $0 \leqslant q < c(M_{2n})$, and $\alpha \geqslant 0$. For $\alpha' = 0$ we may choose $\varepsilon = q = \alpha = 0$. The values $\varepsilon$, $q$, and $c(M_{2n})$ are viewed as constants, if $\alpha = 0$, then it is viewed as a constant, too. Otherwise, if $\alpha \geqslant 1$, then we view $\alpha$ as a variable ranging over positive integers.

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

$$\begin{aligned}
&u\colon\ (u_0, \varepsilon_1(q_1 + \alpha_1 c(M_{2n})), u_1, \ldots, \varepsilon_k(q_k + \alpha_k c(M_{2n})), u_k),\\
&v\colon\ (v_0, \varepsilon'_1(s_1 + \beta_1 c(M_{2n})), v_1, \ldots, \varepsilon'_\ell(s_\ell + \beta_\ell c(M_{2n})), v_\ell),\\
&w\colon\ (w_0, \varepsilon''_1(t_1 + \gamma_1 c(M_{2n})), w_1, \ldots, \varepsilon''_m(t_m + \gamma_m c(M_{2n})), w_m).
\end{aligned}$$

Since $uv = w$ there are many identities. For example, for $k, \ell \geqslant 2$ 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 can be listed easily. We treat only one of them, which is in some sense the most difficult one: We treat the case $p = rs$ with $r = \overline{r}$ and $s = \overline{s}$. It may lead to a large exponent of periodicity. It might be that $u_k = srsr_1$ and $v_0 = r_2srs$ with $r_1r_2 = r$ (and $r_1 \neq 1 \neq r_2$). Hence, we have $u_k v_0 = sp^3$ and $k + \ell = m + 1$. It follows that $\alpha_1 = \gamma_1, \ldots, \alpha_{k-1} = \gamma_{k-1}$, $\beta_2 = \gamma_{k+1}, \ldots, \beta_\ell = \gamma_m$, and there is only one non-trivial identity:

$$q_k + s_1 + 4 + (\alpha_k + \beta_1)c(M_{2n}) = t_k + \gamma_k c(M_{2n}).$$

Since by assumption $c(M_{2n}) \geqslant 3$, 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 $\alpha_k \geqslant 1$ and $\beta_1 \geqslant 1$. If we replace $\alpha_k$, $\beta_1$, and $\gamma_k$ by some $\alpha'_k \geqslant 1$, $\beta'_1 \geqslant 1$, and $\gamma'_k \geqslant 1$ such that we still have $\gamma'_k = \alpha'_k + \beta'_1 + c$, then we obtain new words $u'$, $v'$, and $w'$ with the same images under $h$ in $M_{2n}$ and the identity $u'v' = w'$ remains true.

The following step is completely analogous to what has been done in detail in [20,16,17,3]. 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 positive natural numbers.

The parameters of this system are such that the 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 [20]. This tight bound is based in turn on the work of [12]; 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 translates back into a bound on the exponent of periodicity. For this translation we have to multiply the bound using the factor $c(M_{2n})$ and to add $c(M_{2n}) + 1$. Putting everything together we obtain the claim of the proposition since $c(M_{2n}) \in 2^{\mathcal{O}(n \log n)}$. □

## Acknowledgments

The authors thank the anonymous referees and Géraud Sénizergues for detailed comments which helped to improve the presentation of the paper. The research has been supported partly by the German Research Foundation, *Deutsche Forschungsgemeinschaft, DFG* within the project em GWSS. In addition, Claudio Gutierrez thanks the Centro de Modelamiento Matemático, FONDAP Matemáticas Discretas, for financial support.

## References

[1] M. Benois, Parties rationelles du groupe libre, C.R. Acad. Sci. Paris, Sér. A 269 (1969) 1188–1190.

[2] J. Berstel, Transductions and Context-free Languages, Teubner Studienbücher, Stuttgart, 1979.

[3] V. Diekert, Makanin's Algorithm, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, volume 90 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2002, pp. 387–442, Chapter 12.

[4] V. Diekert, C. Gutierrez, C. Hagenah, The existential theory of equations with rational constraints in free groups is PSPACE-complete, in: A. Ferreira, H. Reichel (Eds.), Proceedings of the 18th Annual Symposium on Theoretical Aspects of Computer Science (STACS'01), Dresden (Germany), 2001, Lecture Notes in Computer Science, vol. 2010, Springer-Verlag, Berlin, 2001, pp. 170–182.

[5] V. Diekert, M. Lohrey. Word equations over graph products, in: P.K. Pandya, J. Radhakrishnan (Eds.), Proceedings of the 23rd Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2003), Mumbai (India), Lecture Notes in Computer Science, vol. 2914, Springer-Verlag, Berlin, 2003, pp. 156–167.

[6] V. Diekert, M. Lohrey, Existential and positive theories of equations in graph products, Theory Comput. Syst. 37 (2004) 133–156.

[7] V. Diekert, Yu. Matiyasevich, A. Muscholl, Solving word equations modulo partial commutations, Theoretical Comput. Sci. 224 (1999) 215–235, Special issue of LFCS'97.

[8] V. Diekert, A. Muscholl. Solvability of equations in free partially commutative groups is decidable, in: F. Orejas, P.G. Spirakis, J. van Leeuwen (Eds.), Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP'01), Lecture Notes in Computer Science, vol. 2076, Springer-Verlag, Berlin Heidelberg, 2001, pp. 543–554.

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

[10] S. Eilenberg, Automata, Languages, and Machines, volume A. Academic Press, New York and London, 1974.

[11] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, San Francisco, 1979.

[12] J. von zur Gathen, M. Sieveking, A bound on solutions of linear integer equalities and inequalities, Proc. Am. Math. Soc. 72 (1) (1978) 155–158.

[13] Y. Gurevich, A. Voronkov, Monadic simultaneous rigid E-unification and related problems, in: P. Degano, R. Gorrieri, A. Marchetti-Spaccamela (Eds.), Proceedings of the 24th International Colloquium on Automata, Languages and Programming (ICALP'97), Bologna, Lecture Notes in Computer Science, vol. 1256, Springer-Verlag, Berlin Heidelberg, 1997, pp. 154–165.

[14] C. Gutierrez, Satisfiability of word equations with constants is in exponential space, in: Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS'98), Los Alamitos (California), EEE Computer Society Press, 1998, pp. 112–119.

[15] C. Gutierrez, Equations in free semigroups with anti-involution and their relation to equations in free groups, in: G.H. Gonnet, D. Panario, A. Viola (Eds.), Proceedings Latin American Theoretical INformatics, LATIN'2000, Lecture Notes in Computer Science, vol. 1776, Springer-Verlag, Berlin, 2000, pp. 387–396.

[16] C. Gutierrez, Satisfiability of equations in free groups is in PSPACE, in: Proceedings 32nd Annual ACM Symposium on Theory of Computing, STOC'2000, ACM Press, 2000, pp. 21–27.

[17] C. Hagenah, Gleichungen mit regulären Randbedingungen über freien Gruppen, Ph.D. thesis, Institut für Informatik, Universität Stuttgart, 2000.

[18] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison-Wesley, Reading, MA, 1979.

[19] J. Karhumäki, F. Mignosi, W. Plandowski, The expressibility of languages and relations by word equations, J. Assoc. Comput. Machinery 47 (2000) 483–505.

[20] A. Kościelski, L. Pacholski, Complexity of Makanin's algorithm, J. Assoc. Comput. Machinery 43 (4) (1996) 670–684.

[21] A. Kościelski, L. Pacholski, Makanin's algorithm is not primitive recursive, Theoretical Comput. Sci. 191 (1998) 145–156.

[22] D. Kozen, Lower bounds for natural proof systems, in: Proceedings of the 18th Annual Symposium on Foundations of Computer Science, FOCS'77, Providence, Rhode Island, IEEE Computer Society Press, 1977, pp. 254–266.

[23] M. Lothaire, Combinatorics on words, in: Encyclopedia of Mathematics and its Applications, vol. 17 Addison-Wesley, Reading, MA, 1983, Reprinted by Cambridge University Press, 1997.

[24] G.S. Makanin, The problem of solvability of equations in a free semigroup. Math. Sbornik, 103 (1977) 147–236. English transl. in Math. USSR Sbornik 32 (1977).

[25] G.S. Makanin, Equations in a free group. Izv. Akad. Nauk SSR, Ser. Math. 46 (1982) 1199–1273. English transl. in Math. USSR Izv. 21 (1983).

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

[27] S.S. Marchenkov, Unsolvability of positive ∀∃-theory of a free semi-group, Sibirsky Matematicheskie Jurnal 23 (1) (1982) 196–198, In Russian.

[28] G. Markowsky, Bounds on the index and period of a binary relation on a finite set, Semigroup Forum 13 (1977) 253–259.

[29] Yu. Matiyasevich, Reduction of trace equations to word equations. Talk given at the "Colloquium on Computability, Complexity, and Logic", Institut für Informatik, Universität Stuttgart, Germany, Dec. 5–6, 1996.

[30] Yu. Matiyasevich, Some decision problems for traces, in: S. Adian, A. Nerode (Eds.), Proceedings of the 4th International Symposium on Logical Foundations of Computer Science (LFCS'97), Yaroslavl, Russia, July 6–12, 1997, Lecture Notes in Computer Science, vol. 1234, Springer-Verlag, Berlin Heidelberg, 1997, pp. 248–257. Invited lecture.

[31] Yu.I. Merzlyakov, Positive formulae over free groups, Algebra i Logika 5 (4) (1966) 25–42, In Russian.

[32] W. Plandowski, Testing equivalence of morphisms on context-free languages, in: J. van Leeuwen (Ed.), Algorithms—ESA'94, Second Annual European Symposium, Lecture Notes in Computer Science, vol. 855, Utrecht, The Netherlands, Springer, 1994, pp. 460–470.

[33] W. Plandowski, Satisfiability of word equations with constants is in NEXPTIME, in: Proceedings of the 31st Annual ACM Symposium on Theory of Computing, STOC'99, ACM Press, 1999, pp. 721–725.

[34] W. Plandowski, Satisfiability of word equations with constants is in PSPACE, in: Proceedings of the 40th Annual Symposium on Foundations of Computer Science, FOCS'99, IEEE Computer Society Press, 1999, pp. 495–500.

[35] W. Plandowski, Satisfiability of word equations with constants is in PSPACE, J. Assoc. Comput. Machinery 51 (2004) 483–496.

[36] W. Plandowski, W. Rytter, Application of Lempel-Ziv encodings to the solution of word equations, in: K.G. Larsen et al., (Eds.), Proceedings of the 25th International Colloquium on Automata, Languages and Programming (ICALP'98), Aalborg (Denmark), 1998, Lecture Notes in Computer Science, vol. 1443, Springer-Verlag, Berlin Heidelberg, 1998, pp. 731–742.

[37] A.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 (1985) 115–162.

[38] K.U. Schulz, Makanin's algorithm for word equations—Two improvements and a generalization, in: K.U. Schulz (Ed.), Word Equations and Related Topics, Lecture Notes in Computer Science, vol. 572, Springer-Verlag, Berlin Heidelberg, 1991, pp. 85–150.