Solvability of Equations in Free Partially Commutative Groups is Decidable

Volker Diekert[1] and Anca Muscholl[2]

[1] Institut für Informatik, Universität Stuttgart,
Breitwiesenstr. 20-22, D-70565 Stuttgart
diekert@informatik.uni-stuttgart.de
[2] LIAFA, Université Paris VII,
2, place Jussieu, case 7014, F-75251 Paris Cedex 05
muscholl@liafa.jussieu.fr

Abstract. Trace monoids are well-studied objects in computer science where they serve as a basic algebraic tool for analyzing concurrent systems. The question whether the existential theory of trace equations is decidable has been solved positively in 1996. Free partially commutative groups (graph groups) generalize trace monoids in the sense that every element has an inverse. In this paper we show that the existential theory of equations over graph groups is decidable, too. Our decision procedure is non-elementary, but if a certain graph theoretical parameter is viewed as a constant, then we obtain a PSPACE-completeness result. Restricting ourselves to trace monoids we still obtain a better complexity result, as it was known previously.

1 Introduction

Free partially commutative monoids (or *trace monoids*) serve as a basic algebraic tool for investigating concurrent systems, [15, 16]. Atomic actions are represented by letters and independency of actions is reflected by an independence relation on the alphabet. The independency defines a partial commutation. If each atomic action a has an inverse $\overline{a}$, then, on the algebraic level, we switch from monoids to groups. This means that we work with free partially commutative groups, which in algebra are also known as *graph groups*, see e.g. [7].

We show that the existential theory of equations over graph groups is decidable. Our decision procedure is non-elementary, but if a certain graph theoretical parameter is viewed as a constant, then we can establish a PSPACE-completeness result. In fact, we generalize a recent result of [4], which is also the starting point for us.

In the simplest setting we ask whether a single word equation in unknowns and constants is solvable. This problem is well-known to be NP-hard. It becomes PSPACE-hard, as soon as we add regular constraints for the unknowns, simply because the intersection problem for regular languages is PSPACE-complete, [10]. Makanin proved the decidability of word equations [11] and Schulz extended this decidability result in order to include regular constraints [20]. Standard methods imply that the existential theory of word equations is decidable, once the problem of solving a single equation is settled. The same holds for equations over free groups and again Makanin proved decidability [12, 13]. The scheme of Makanin in the case of free groups is however known to be non primitive-recursive, see [9]. Only when Plandowski invented a new method for solving word equations by some polynomial space bounded algorithm [19], the corresponding problem for free groups was reconsidered and Gutiérrez [8] succeeded in extending Plandowski's polynomial space algorithm to free groups. In fact, it has been possible to prove decidability (and PSPACE-completeness) of equations with rational constraints in free groups, see [4].

The situation in trace monoids turned out to be quite different and decidability for trace equations was shown only in 1996 by Matiyasevich, see [14, 5]. It remained open whether the solvability of graph group equations is decidable, but it was known that rational constraints are too powerful: They lead to undecidability (although this is not treated here). The good notion turned out to be *normalized rational constraint* (this concept is introduced here). Moreover, the main result of [4] is not about free groups, but about *free monoids with involution*. This is another key: It is not necessary to reduce to the situation in free groups. We introduce the notion of trace monoid with involution and reduce stepwise to the case of free monoids with involution by removing independency. This reduction is the main technical contribution of our work. We are strictly more general than in the case of trace monoids, and when applying our method in this special case we still obtain a better complexity result as known previously.

2 Preliminaries

2.1 Free partially commutative monoids with involution

Let us recall some basic concepts of trace theory, for details we refer to [6]. By Γ we mean a finite alphabet which is equipped with an involution $\overline{} : \Gamma \to \Gamma$. An involution is a mapping such that $\overline{\overline{a}} = a$ for all $a \in \Gamma$. By $I \subseteq \Gamma \times \Gamma$ we denote an *independence relation*, its complement is $D = (\Gamma \times \Gamma) \setminus I$. It is called a *dependence* relation and we demand both that D is reflexive and symmetric and that $(a, b) \in D$ implies $(\overline{a}, b) \in D$ for all $a, b \in \Gamma$. In particular, we have $(a, \overline{a}) \in D$ for all $a \in \Gamma$. The *free partially commutative* monoid $M(\Gamma, I)$ is defined by the quotient monoid $\Gamma^*/\{ab = ba \mid (a, b) \in I\}$. According to Mazurkiewicz [15] it is also called a *trace monoid* and its elements are *traces*. If the reference to (Γ, I) is clear, we also write $\mathbb{M}$ instead of $M(\Gamma, I)$. The length $|x|$ of a trace x is the length of any representing word. A letter $a \in \Gamma$ is called minimal (maximal resp.) in x, if we can write $x = ay$ ($x = ya$ resp.) for some $y \in \mathbb{M}$. The set of minimal (maximal resp.) elements consists of pairwise independent letters. The set of letters occurring in $x \in \Gamma^*$ or in $x \in \mathbb{M}$ is denoted $\mathrm{alph}(x)$. For $a \in \Gamma$ let $I(a) = \{b \in \Gamma \mid (a, b) \in I\}$. By 1 we denote the empty word, the empty trace, and the unit element in a group.

We shall use node-labeled directed acyclic graphs $[V, E, \lambda]$ in order to represent traces. We assume that V is finite and that $\lambda : V \to \Gamma$ is a labeling such that $(\lambda(v), \lambda(v')) \in D$ implies either $(v, v') \in E^*$ or $(v', v) \in E^*$, so all dependent vertices are ordered. Then $[V, E, \lambda]$ defines a unique trace $x = [V, E, \lambda] \in \mathbb{M}$ in a canonical way. A trace $x \in \mathbb{M}$ may have many different representations. We say that $[V, E, \lambda]$ is a *dependence graph* of a trace $x \in \mathbb{M}$, if $[V, E, \lambda]$ represents x and in addition $(\lambda(v), \lambda(v')) \in D$ is equivalent to $(v, v') \in \mathrm{id}_\Gamma \cup E \cup E^{-1}$. Up to isomorphism, the dependence graph of x is unique, and so is its induced labeled partial order (pomset) $[V, E^*, \lambda]$ which is also denoted by $[V, \leq, \lambda]$.

The involution $^{-} : \Gamma \to \Gamma$ is extended to the free monoid Γ^* by $\overline{a_1 \cdots a_n} = \overline{a_n} \cdots \overline{a_1}$. Since the independence relation I is supposed to be compatible with the involution, the same definition transfers to traces. Hence $(\mathbb{M}, ^{-})$ is a trace monoid with involution. Note that we have $\overline{[V, E, \lambda]} = [V, E^{-1}, \overline{\lambda}]$ and $\overline{[V, \leq, \lambda]} = [V, \geq, \overline{\lambda}]$ respectively, where $\overline{\lambda}(v) = \overline{\lambda(v)}$ for all $v \in V$.

2.2 Recognizable and rational subsets, factor traces

For a moment let M be any finitely generated monoid and $\psi : \Gamma^* \to M$ be a surjective homomorphism. A subset $L \subseteq M$ is called *recognizable*, if $\psi^{-1}(L)$ is a regular word language. Languages of the form $\psi(L)$ with $L \subseteq \Gamma^*$ being regular are called *rational*.

Henceforth by ψ we mean the canonical homomorphism from words to traces, $\psi : \Gamma^* \to \mathbb{M}$. Two words representing the same trace have the same length. Fixing some linear order on Γ we can choose for a given trace $x \in \mathbb{M}$ the lexicographical first word representing x, which is denoted by $\mu(x) \in \Gamma^*$. Clearly $\psi(\mu(x)) = x$. It is well-known [1] that $\mu(\mathbb{M}) \subseteq \Gamma^*$ is a regular word language. If $L \subseteq \mathbb{M}$ is recognizable, then $\psi^{-1}(L) \cap \mu(\mathbb{M})$ is regular, too; and L is a homomorphic image of some regular subset of $\mu(\mathbb{M})$. Ochmański's Theorem says that this correspondence is one-to-one between recognizable subsets of $\mathbb{M}$ and regular subsets of $\mu(\mathbb{M})$, see [17], [18, Thm. 6.3.12].

Let $x = [V, \leq, \lambda] \in \mathbb{M}$ be a trace where $\leq$ is the partial order induced by the dependence graph. A *factor* or *factor trace* is a trace $f \in \mathbb{M}$ such that we can write $x = pfq$. Given a factorization $x = pfq$, we find some $F \subseteq V$ such that the induced pomset of F in $[V, \leq, \lambda]$ represents f. Moreover, F has the property that whenever $v \leq v' \leq v''$ with $v, v'' \in F$, then $v' \in F$, too. Conversely, let $F \subseteq V$ be a subset with this property: $v \leq v' \leq v''$ with $v, v'' \in F$ implies $v' \in F$. Then we can factorize $x = pfq$ such that F represents the factor f, but due to partial commutation the factorization is not unique (even if $F \neq \emptyset$). There is another warning. Assume that $V = F \cup G$ is a disjoint union where F represents f and G represents g. This does not mean that x is a product of f and g, in general. Indeed, let $\{a, b\} \times \{c, d\} \subseteq I$ and $x = abcd$. Then we have $abcd = cadb$, so both $f = bc$ and $g = ad$ are factors. But x is not a product of f and g as soon as both $ab \neq ba$ and $cd \neq dc$.

2.3 Graph groups and their normalized rational subsets

We can switch from monoids with involution to groups by interpreting $\overline{a}$ as a^{-1}. Formally, we define the quotient monoid $G(\Gamma, I, ^{-})$ by $M(\Gamma, I)/\{a\overline{a} = 1 \mid a \in \Gamma\}$, and the canonical homomorphism is denoted by $\varphi : M(\Gamma, I) \to G(\Gamma, I, \overline{})$. Since $\overline{\overline{a}} = a$, this is a group. It is called the *graph group* over $(\Gamma, I, \overline{})$. This is a slight abuse of language, since the involution may have fixed points, i.e., we allow $a = \overline{a}$ for some $a \in \Gamma$. This freedom is very useful for various reasons. Here it means that $G(\Gamma, I, \overline{})$ may have elements of order two. Torsion free groups arise if the involution has no fixed points. This is the usual setting of graph groups as investigated e.g. by Droms in [7]. If the reference to $(\Gamma, I, \overline{})$ is clear we sometimes write $\mathbb{G}$ instead of $G(\Gamma, I, \overline{})$. We can describe the group $\mathbb{G}$ by means of a confluent and Noetherian trace rewriting system $S = \{a\overline{a} \to 1 \mid a \in \Gamma\}$. Termination and confluence of S as a rewriting system over $\mathbb{M}$ can be easily verified. For a general treatment or trace rewriting systems the interested reader is referred to [3]. A trace $x \in \mathbb{M}$ is called *reduced*, if it is irreducible with respect to the system S, i.e., the trace x contains no factor of the form $a\overline{a}$ with $a \in \Gamma$. Then every group element $x \in \mathbb{G}$ has a unique reduced representation. This means there is a unique reduced trace $\rho(x) = \widetilde{x} \in \mathbb{M}$ such that $\varphi(\widetilde{x}) = x$.

The set of all reduced traces $\widetilde{G} \subseteq \mathbb{M}$ is therefore in one-to-one correspondence with $\mathbb{G}$ via some injective mapping ρ. Since $\widetilde{G}$ is defined by some finite set of forbidden factors, it is a recognizable subset of $\mathbb{M}$. Hence $\mu(\widetilde{G})$ is a regular language of words. The mapping $\nu : G(\Gamma, I, \overline{}) \to \Gamma^*$ which maps a group element x to $\nu(x) = \mu(\widetilde{x}) = \mu(\rho(x))$ is a normal-form mapping, i.e., $\varphi(\psi(\nu(x))) = x$. The word $\nu(x)$ is the lexicographical first word among the shortest words which represent x.

We obtain the following commutative diagram:

$$\begin{tikzcd} \widetilde{G} \arrow[r, hook, "\text{can.}"] & \mathbb{M} \arrow[r, hook, shift left, "\mu"] \arrow[d, two heads, "\varphi"] & \Gamma^* \arrow[l, two heads, shift left, "\psi"] \\ & \mathbb{G} \arrow[ul, "\sim", "\rho"'] \arrow[ur, "\nu"'] & \end{tikzcd}$$

A group language $L \subseteq \mathbb{G}$ is called *normalized regular*, if the set of normal forms $\nu(L) \subseteq \Gamma^*$ is a regular word language. Since $\nu(\mathbb{G} \setminus L) = \nu(G) \setminus \nu(L)$, the class of normalized rational languages is an effective Boolean algebra. If $\mathbb{G}$ is a free group, then every rational language is normalized regular. This is non-trivial; it follows from [2]. In general, the class of normalized regular languages is strictly contained in the class of rational subsets.

3 Equations over graph groups and trace monoids with involution

In the following Ω denotes a set of variables (or unknowns) and we let $\overline{} : \Omega \to \Omega$ act as an involution on unknowns without fixed points. A *propositional formula* over Γ and Ω is inductively defined as follows. Atomic formulas are either equations of the form $\alpha = \beta$, where $\alpha, \beta \in (\Gamma \cup \Omega)^*$, or constraints. A constraint is of the form $X \in L$, where $X \in \Omega$, and L is either recognizable (if we deal with monoids) or normalized regular (if we deal with graph groups). For the specification of L we may use some NFA (non-deterministic finite automaton) where the accepted language is either in $\mu(\mathbb{M})$ (if we deal with monoids) or in $\nu(\mathbb{G})$ (if we deal with graph groups). Propositional formulas are built from atomic formulas using the Boolean connectives. A formula in the existential theory is an existentially quantified sentence. An *assignment* for a propositional formula is a mapping $\sigma : \Omega \to \Gamma^*$ such that $\sigma(\overline{X}) = \overline{\sigma(X)}$ for all $X \in \Omega$. As usual, we extend σ to a homomorphism $\sigma : (\Gamma \cup \Omega)^* \to \Gamma^*$ by letting $\sigma(a) = a$ for $a \in \Gamma$. By extending σ further with the homomorphism $\Gamma^* \to \mathbb{M}$ (if we consider trace monoids) and with the homomorphism $\Gamma^* \to \mathbb{G}$ respectively (if we consider graph groups), we read σ also as $\sigma : (\Gamma \cup \Omega)^* \to \mathbb{M}$ and $\sigma : (\Gamma \cup \Omega)^* \to \mathbb{G}$. There will be no risk of confusion. The evaluation of a propositional formula under an assignment $\sigma : \Omega \to \Gamma^*$ is defined in a canonical way: The evaluation under σ of a constraint $X \in L$ is the truth value of $\mu(\sigma(x)) \in \mu(L)$ and of $\nu(\sigma(x)) \in \nu(L)$, respectively. The evaluation of a formula depends on whether we work with a trace monoid $\mathbb{M}$ or with a graph group $\mathbb{G}$. An existentially quantified sentence is true, if and only if the underlying propositional formula evaluates to *true* by some assignment.

Our main theorem states that the existential theory of trace monoids with involution is decidable. The group case becomes a corollary. This is due to the following lemma and proposition which are close to similar statements in [4].

Lemma 1. *Let $\widetilde{x}, \widetilde{y}, \widetilde{z} \in \mathbb{M}$ be reduced traces representing the group elements $x, y, z \in \mathbb{G}$. Then we have $xy = z$ in $\mathbb{G}$ if and only if there are traces $p, q, r \in \mathbb{M}$ such that $\widetilde{x} = pq, \widetilde{y} = \overline{q}r$, and $\widetilde{z} = pr$.*

Proposition 1. *The existential theory of equations with normalized regular constraints over $G(\Gamma, I, \overline{})$ is* PSPACE*-hard. There is a polynomial time reduction of this theory to the existential theory of equations with recognizable constraints over trace monoids with involution $M(\Gamma, I)$.*

4 The Main Theorem

In order to have a convenient complexity bound below we introduce the graph theoretical concept of a complete clan. A *complete clan* in (Γ, D) is a maximal subset $A \subseteq \Gamma$ such that $(a, c) \in D \Leftrightarrow (b, c) \in D$ for all $a, b \in A$, $c \in \Gamma$. Note that a complete clan is a complete subgraph of (Γ, D), since D is reflexive. The set of complete clans is a partition of Γ and we denote by $c(\Gamma, D)$ the number of complete clans. A complete clan A is called *thin*, if there are $a \in A$, $b \in \Gamma \setminus A$ such that $(a, b) \in I$. The following facts are easily verified: The number of thin clans is either $c(\Gamma, D)$ or $c(\Gamma, D) - 1$, it is never 1. If $\mathbb{M}$ is a direct product of d free monoids, then the number of thin clans is d for $d > 1$, and it is 0 for $d = 1$.

Recognizable constraints are a natural concept in the framework of trace equations: They are needed for negations and more importantly for expressing commutation during the reduction. For this we introduce two macros: The formula

$$\bigvee_{a\in\Gamma}\big((X \in a\mathbb{M} \mathbin{\&} Y \notin a\mathbb{M}) \vee (X \notin a\mathbb{M} \mathbin{\&} Y \in a\mathbb{M})\big)$$

represents $\min(X) \neq \min(Y)$. With the help of this macro we can replace an inequality $\alpha \neq \beta$ by the equivalent formula $\exists Z \exists X \exists Y : \alpha = ZX \mathbin{\&} \beta = ZY \mathbin{\&} \min(X) \neq \min(Y)$.

A formula $\mathrm{alph}(X) \times \mathrm{alph}(Y) \subseteq I$ stands for

$$\bigwedge_{a\in\Gamma}(X \notin \Gamma^* a \Gamma^* \ \vee \ Y \in I(a)^*).$$

So, if $I = \emptyset$, then the macro means nothing but $X \in \{1\} \ \vee \ Y \in \{1\}$. A formula as above is called a *commutation constraint.*

Theorem 1. *Let $\tau \geq 0$. The following problem is* PSPACE-*complete.*

Input: *An existentially quantified sentence Φ of equations with recognizable constraints over a trace monoid with involution $(M(\Gamma, I), \overline{})$ such that the number of complete thin clans is at most τ.*

Question: *Is the sentence Φ true?*

Corollary 1. *Let $\tau \geq 0$. The following problem is* PSPACE-*complete.*

Input: *An existentially quantified sentence Φ of equations with normalized rational constraints over a graph group $G(\Gamma, I, \overline{})$ such that the number of complete thin clans is at most τ.*

Question: *Is the sentence Φ true?*

For $\tau = 0$ Thm. 1 follows by [4], since $\tau = 0$ is the situation of free monoids. We assume $\tau \geq 1$ and since τ is a constant, it is enough to give reduction to the case $\tau - 1$. This is done stepwise. First, we deal with negations: Using the first macro above (and the fact that the class of recognizable languages is closed under complementation) we may assume without restriction that Φ is a positive sentence. More precisely, for the polynomial time reduction we will assume that Φ may contain negations only on the constraints. We still call it a *positive sentence.*

Second, we make a reduction to the case where the set of fixed points $\Delta = \{a \in \Gamma \mid a = \overline{a}\}$ is empty. This reduction is omitted for lack of space. Hence, in the following we shall assume that the involution has no fixed points, i.e., $a \neq \overline{a}$ for all $a \in \Gamma$. Moreover, we shall assume that all equations of Φ are in triangulated form, $X = X_1 X_2$.

We fix for the rest of the paper some thin complete clan in (Γ, D) which we write in the form $A \cup \overline{A}$ with $A \cap \overline{A} = \emptyset$. We define $\widehat{D} = D \cup \Gamma \times (A \cup \overline{A}) \cup (A \cup \overline{A}) \times \Gamma$, so the clan is not thin anymore with respect to $\widehat{D}$. Let $\widehat{I} = \Gamma \times \Gamma \setminus \widehat{D}$, then $\widehat{\mathbb{M}} = M(\Gamma, \widehat{I})$ is a trace monoid with involution, but the number of thin complete clans in $(\Gamma, \widehat{D})$ is at most $\tau - 1$.

In the remaining of the section $\hat{\psi}$ means the canonical homomorphism $\hat{\psi} : \widehat{\mathbb{M}} \to \mathbb{M}$. By induction, it is enough to transform Φ into some equivalent sentence

$\widehat{\Phi}$ over $\widehat{\mathbb{M}}$. The handling of recognizable constraints is trivial: A constraint $X \in L$ ($X \notin L$ resp.) is replaced by $X \in \hat{\psi}^{-1}(L)$ ($X \notin \hat{\psi}^{-1}(L)$ resp.). For a replacement of an equation $X = X_1X_2$ let $d = 3\tau^2 + 8\tau + 7$ and choose d new, existentially quantified variables $Y_1, \ldots, Y_d$. The replacement is a disjunction over all permutations $\pi \in \mathrm{Perm}(d)$ and all c with $0 \leq c \leq d$. Each clause in the disjunction is a conjunction of the following equations and commutation constraints:

$$X_1 = Y_1 \cdots Y_c \;\;\&\;\; X_2 = Y_{c+1} \cdots Y_d \;\;\&\;\; X = Y_{\pi(1)} \cdots Y_{\pi(d)}$$
$$\& \bigwedge_{(i-j)(\pi(i)-\pi(j))<0} \mathrm{alph}(Y_i) \times \mathrm{alph}(Y_j) \subseteq I$$

We obtain a system of equations with a larger set of variables and a new sentence $\widehat{\Phi}$. An easy reflection shows that if $\widehat{\Phi}$ evaluates to true over $\widehat{\mathbb{M}}$, then Φ evaluates to true over $\mathbb{M}$. Since τ is a constant, the size of $\widehat{\Phi}$ is polynomial in the size of Φ when using the commutation macro. In any case the transformation from Φ to $\widehat{\Phi}$ can be performed by some polynomially space bounded Turing machine. The heart of the paper is the proof of the converse: If Φ evaluates to true, then $\widehat{\Phi}$ evaluates to true as well.

5 Normal Forms

The aim of this section is to define a normal form mapping $\mathrm{nf} : \mathbb{M} \to \widehat{\mathbb{M}}$ which associates with each trace $x \in \mathbb{M}$ a unique trace $\mathrm{nf}(x) \in \widehat{\mathbb{M}}$ and which is compatible with the involution, i.e., we demand $\hat{\psi}(\mathrm{nf}(x)) = x$ and $\mathrm{nf}(\overline{x}) = \overline{\mathrm{nf}(x)}$. As we will see below the existence of the normal form relies on the following:

Lemma 2. *Let $a \in \Gamma$ such that $a \neq \overline{a}$ and let $w \in \{a, \overline{a}\}^*$ be any word. Then there exists a unique $k \geq 0$ such that $w \in a^*(\overline{a}a^*)^k(\overline{a}^*a)^k\overline{a}^*$.*
For the same k we also have $\overline{w} \in a^*(\overline{a}a^*)^k(\overline{a}^*a)^k\overline{a}^*$.

We view a trace x as a labeled pomset $[V, \leq, \lambda]$, $\lambda : V \to \Gamma$. We write $v \in x$ for a vertex $v \in V$ and $u \parallel v$ whenever u, v are incomparable with respect to $\leq$. We also write $(u, v) \in I$ if $(\lambda(u), \lambda(v)) \in I$, and analogously for $D = (\Gamma \times \Gamma) \setminus I$. Let $a_1 < \cdots < a_q$ be the linearly ordered subset of $(V, \leq)$ containing all vertices with label in the clan $A \cup \overline{A}$. We might have $q = 0$ meaning that there are no vertices with label in $A \cup \overline{A}$. We read $a_1 < \cdots < a_q$ as a word $a_1 \cdots a_q$ in the free monoid $(A \cup \overline{A})^*$. With each vertex $v \in V$ we shall associate the maximal factor of $a_1 \cdots a_q$ consisting of the vertices w with label in $A \cup \overline{A}$ which are incomparable with v, i.e., $w \parallel v$. For this, we introduce the *source* $s(v)$ and the *target point* $t(v)$ of v:

$$s(v) = \sup\{i \mid a_i \leq v\}, \quad t(v) = \inf\{i \mid v \leq a_i\},$$

where by convention, $\sup \emptyset = 0$ and $\inf \emptyset = q + 1$. Thus, $0 \leq s(v) \leq q$, $1 \leq t(v) \leq q + 1$ and $s(v) \leq t(v)$ for all $v \in V$. Note that we have $s(v) = t(v)$ if and only if the label of v belongs to $A \cup \overline{A}$.

For $0 \leq s \leq t \leq q+1$ we define the *median position* $m(s,t)$. For $s = t$ let $m(s,t) = s$. For $s < t$ we choose by Lem. 2 the unique c with $s \leq c < t$ and $k \geq 0$ such that $a_{s+1} \cdots a_c \in A^*(\overline{A}A^*)^k$ and $a_{c+1} \cdots a_{t-1} \in (\overline{A}^* A)^k \overline{A}^*$. Then we define $m(s,t) = c + \frac{1}{2}$ and we call $m(s,t)$ the *median position*. The median position $m(s(v), t(v))$ is called the *global position* of a vertex $v \in V$, it is denoted by $g(v)$. Hence $g(v) = m(s(v), t(v))$.

Lemma 3. *Let $x = [V, \leq, \lambda]$ and $v, w \in V$ be vertices such that $v \leq w$. Then we have $s(v) \leq s(w)$, $t(v) \leq t(w)$, and $g(v) \leq g(w)$.*

We define the normal form $\mathrm{nf}(x) \in \widehat{\mathbb{M}}$ by introducing *new arcs* into the dependence graph $[V, E, \lambda]$ of x. Let $v, w \in V$ such that $\lambda(w) \in A \cup \overline{A}$ and $v \parallel w$. (In particular, $\lambda(v) \notin A \cup \overline{A}$ and $g(v) \neq g(w)$.) We define a new arc from v to w, if $g(v) < g(w)$, otherwise we define a new arc from w to v. The arcs being already present in the dependence graph of x are called *old arcs*. The union $\hat{E}$ of old and new arcs defines a labeled directed graph $[V, \hat{E}, \lambda]$. From the lemma below it follows that $[V, \hat{E}, \lambda]$ is an acyclic graph:

Lemma 4. *Let $x \in \mathbb{M}$ and consider vertices u, v in x. For $u \leq v$ in $\mathrm{nf}(x)$ we have $g(u) \leq g(v)$. Moreover, we have $g(u) < g(v)$ whenever there is a path from u to v in $\mathrm{nf}(x)$ using at least one new arc.*

Thus, $[V, \hat{E}, \lambda]$ defines a unique trace $\mathrm{nf}(x) = [V, \hat{E}, \lambda]$ of $\widehat{\mathbb{M}}$. The important property of the normal form is $\mathrm{nf}(\overline{x}) = \overline{\mathrm{nf}(x)}$ for all $x \in \mathbb{M}$, which can be derived from Lem. 2.

Example 1. Let $x = a\,a\,\overline{a}\,\overline{a}\,a\,\overline{a}\,a\,\overline{a}\,\overline{a}\,a\,a\,b$ and $(a,b) \in I$. Then $k = 3$, $g(b) = 6\frac{1}{2}$ and $\mathrm{nf}(x) = a\,a\,\overline{a}\,\overline{a}\,a\,\overline{a}\,b\,a\,\overline{a}\,\overline{a}\,a\,a$.

Remark 1. A normal form satisfying $\mathrm{nf}(\overline{x}) = \overline{\mathrm{nf}(x)}$ for all $x \in \mathbb{M}$ cannot exist, in general, if the involution has fixed points. Indeed, assume we were in the situation $a, b \in A$, $a = \overline{a}$, $b = \overline{b}$, and $(a,b) \in I$. Then $\overline{ab} = ab$, but $(a,b) \in \widehat{D}$, so necessarily $\mathrm{nf}(\overline{ab}) \neq \overline{\mathrm{nf}(ab)}$ in $\widehat{\mathbb{M}}$.

After these preparations Thm. 1 becomes a consequence of the next one.

Theorem 2. *Let $d = 3\tau^2 + 8\tau + 7$ and let $x, x_1, x_2 \in \mathbb{M}$ such that $x = x_1 x_2$. Then there are $y_1, \ldots, y_d \in \widehat{\mathbb{M}}$, an index $0 \leq c \leq d$, and a permutation $\pi \in \mathrm{Perm}(d)$ with the following properties:*

$$\begin{aligned}
\mathrm{nf}(x) &= y_{\pi(1)} \cdots y_{\pi(d)}, \\
\mathrm{nf}(x_1) &= y_1 \cdots y_c, \\
\mathrm{nf}(x_2) &= y_{c+1} \cdots y_d, \\
\mathrm{alph}(y_i) \times \mathrm{alph}(y_j) &\subseteq I \textit{ for all } i, j, \textit{ where } (i-j)(\pi(i) - \pi(j)) < 0.
\end{aligned}$$

The proof of Thm. 2 covers the rest of the paper. First note that if the theorem is true for some $d \leq 3\tau^2 + 8\tau + 7$, then it is true for $d = 3\tau^2 + 8\tau + 7$. We let $x, x_1, x_2 \in \mathbb{M}$ such that $x = x_1 x_2$. We present x by its pomset $[V, \leq, \lambda]$

and $a_1 < \cdots < a_q$ is the linearly ordered subset of all vertices in V which have a label in $A \cup \overline{A}$. We define p to be the index such that $a_1, \ldots, a_p \in x_1$ and $a_{p+1}, \ldots, a_q \in x_2$. We have $0 \leq p \leq q$.

The case $q = 0$ is trivial (it could be excluded), the case $p = 0$ or $p = q$ is not trivial in the sense that we might have $\mathrm{nf}(x) \neq \mathrm{nf}(x_1)\,\mathrm{nf}(x_2)$, c.f. Ex. 1 above with $x_2 = b$.

In order to determine how $\mathrm{nf}(x)$ can be obtained from $\mathrm{nf}(x_1), \mathrm{nf}(x_2)$, we introduce the notion of *local position*. The local position $\ell(v)$ is the global position of v in x_1, if v belongs to x_1. If v belongs to x_2, then $\ell(v)$ is the global position of v in x_2 plus p, since we define $\ell(v)$ in x. Suppose that v is in x_1. With respect to x_1, the target point $t'(v)$ of v is $\min\{p+1, t(v)\}$, hence we define:

$$\ell(v) = m(s(v), \min\{p+1, t(v)\})\,.$$

Similarly, if v belongs to x_2 then the source point $s'(v)$ of v with respect to x_2 is $s'(v) = \max\{s(v), p\}$ and we define:

$$\ell(v) = m(\max\{s(v), p\}, t(v))\,.$$

The next proposition is a crucial technical result. It shows that among the vertices v with $s(v) \leq p < t(v)$, i.e., among the vertices where the local and global positions may differ, there is a constant number of different source points.

Proposition 2. *Let $x \in \mathbb{M}$ and define $S = \{s(v) \mid s(v) \leq p < t(v), v \in x\}$. Then we have $|S| \leq \tau + 1$.*

Proof. We may assume that $|S| \geq 2$. Choose a sequence $b_1, \ldots, b_k, b_{k+1} \in x$ with k minimal such that $S = \{s(b_i) \mid 1 \leq i \leq k+1\}$. We may assume that $0 \leq s(b_1) < \cdots < s(b_k) < p$ and $p < t(b_i)$ for $1 \leq i \leq k$. Hence we have $b_i \parallel a_p$ for all $1 \leq i \leq k$. We will show that $k \leq \tau$. For each $1 < i \leq k$ we choose a path from $a_{s(b_i)}$ to b_i in the dependence graph of x. On this path we pick a last vertex c_i with $c_i \leq a_p$. This vertex cannot be b_i. Hence there is a next vertex d_i with $(c_i, d_i) \in D$, $c_i < d_i \leq b_i$, and $d_i \parallel a_p$ for $1 < i \leq k$. Let $d_1 = b_1$. We claim that $\{d_i\} \times \{c_{i+1}, \ldots, c_k\} \subseteq I$ for $1 \leq i \leq k$. Indeed assume by contradiction that $(d_i, c_j) \in D$ for some $i < j \leq k$. Then $d_i \leq c_j$ since $s(b_i) < s(b_j)$. But we have $c_j \leq a_p$, hence $d_i \leq c_j$ implies $d_i \leq a_p$ in contradiction to $d_i \parallel a_p$.

Next, $(d_i, c_i) \in D$ and $\{d_i\} \times \{c_{i+1}, \ldots, c_k\} \subseteq I$ imply that c_i and c_j are in different thin clans for all $2 \leq i < j \leq k$. Moreover, there is another thin clan containing d_1. □

Analogously to Prop. 2 it holds that $|T| \leq \tau + 1$, where T is the set of target points: $T = \{t(v) \mid s(v) \leq p < t(v), v \in x\}$.

The positions we are interested in will be called in the following *cutting points*. The set of cutting points C is the union $C = C_g \cup C_\ell \cup \{0, p + \frac{1}{2}, q+1\}$ where

$$C_g = \{g(v) \mid s(v) \leq p < t(v), v \in x\},$$
$$C_\ell = \{\ell(v) \mid s(v) \leq p < t(v), v \in x\}.$$

Proposition 3. *The number of cutting points is bounded by* $\tau^2 + 4\tau + 6$.

The cutting points split the real interval $[0, q+1]$ into open intervals of the form (i, j) where $i, j \in C, i < j$, and $(i, j) \cap C = \emptyset$. There are $|C| - 1$ such intervals, so the number is at most $\tau^2 + 4\tau + 5$. With each interval (i, j) we associate a factor trace of x. The factor trace is denoted either $x[0; i, j]$ or $x[3; i, j]$; we call $x[m; i, j]$, $0 \leq m \leq 3$, a *segment.* (Later on we will define segments with index 1 and 2.) Since $p + \frac{1}{2}$ is a cutting point we have either $j \leq p + \frac{1}{2}$ or $p + \frac{1}{2} \leq i$. For $j \leq p + \frac{1}{2}$ we define

$$x[0; i, j] = \{v \in x \mid i < g(v) \leq j \text{ and } t(v) \leq p\}.$$

For $p + \frac{1}{2} \leq i$ we define

$$x[3; i, j] = \{v \in x \mid i \leq g(v) < j \text{ and } p < s(v)\}.$$

Note that we have $\ell(v) = g(v)$ for all $v \in x[0; i, j]$ or $v \in x[3; i, j]$. A segment as defined above is just a set of vertices of x. However, it is easy to see that a segment defines a factor trace of x and a factor trace of either x_1 or x_2. In the lemma below, we show that this property is still true for the normal forms $\mathrm{nf}(x), \mathrm{nf}(x_1), \mathrm{nf}(x_2)$.

Lemma 5. *Each segment* $x[m; i, j], m = 0, 3$, *defines a factor trace* y *of* $\mathrm{nf}(x)$. *If* $j \leq p + \frac{1}{2}$, *then* $x[0; i, j]$ *also defines a factor trace* y_1 *of* $\mathrm{nf}(x_1)$ *and we have* $y = y_1$ *in* $\widehat{\mathbb{M}}$. *If* $p + \frac{1}{2} \leq i$, *then* $x[3; i, j]$ *also defines a factor trace* y_2 *of* $\mathrm{nf}(x_2)$ *and we have* $y = y_2$ *in* $\widehat{\mathbb{M}}$.

The segments $x[0; i, j], x[3; i, j]$ are pairwise disjoint subsets of x, but they do not cover x, in general. The missing points are those vertices v where $s(v) \leq p < t(v)$. Therefore for each $m = 1, 2$ and $i \leq p < j$ we define an m-segment by:

$$x[m; i, j] = \{v \in x_m \mid s(v) = i \text{ and } t(v) = j\}, \quad m = 1, 2.$$

Note that the indices i, j are not required to be cutting points. However, all vertices $v \in x[m; i, j]$ have the same local position $\ell(v)$ and the same global position $g(v)$, which are both cutting points. The number of non-empty 1 and 2 segments is at most $2(\tau+1)^2 = 2\tau^2 + 4\tau + 2$ by Prop. 2.

Lemma 6. *Each* m*-segment* $x[m; i, j]$, $m = 1, 2$ *defines a factor trace* y *of* $\mathrm{nf}(x)$ *and a factor trace* y_m *of* $\mathrm{nf}(x_m)$. *We have* $y = y_m$ *in* $\widehat{\mathbb{M}}$.

Let us summarize the notations we have introduced up to this point. For each $m = 0, 1, 2, 3$ we have defined m-segments of the form $x[m; i, j]$ which are pairwise disjoint subsets of x and cover x. Hence $F = \{x[m; i, j] \mid x[m; i, j] \neq \emptyset\}$ is a partition of the set x. Each $x[m; i, j]$ with $m \leq 1$ is a factor trace of both $\mathrm{nf}(x)$ and $\mathrm{nf}(x_1)$ (respectively, each $x[m; i, j]$ with $m > 1$ is a factor trace of both $\mathrm{nf}(x)$ and $\mathrm{nf}(x_2)$). Using the counting above (Props. 2 and 3) we have $|F| \leq 3\tau^2 + 8\tau + 7$.

By Lemma 5 and 6 we denote by $y_f \in \widehat{\mathbb{M}}$ the factor trace of $\mathrm{nf}(x)$ associated with $f \in F$. Since the segments in F cover all of $\mathrm{nf}(x), \mathrm{nf}(x_1)$ and $\mathrm{nf}(x_2)$ it remains to show how to write $\mathrm{nf}(x), \mathrm{nf}(x_1)$ and $\mathrm{nf}(x_2)$ as products of y_f, where f is in F. For this we have to compare segments. Every segment is associated with either a pair of consecutive cutting points or a single cutting point. For a non-empty m-segment $f = x[m; i, j]$ we define a *global weight* $\omega_g(f)$ and a *local weight* $\omega_\ell(f)$ as follows. For $m \in \{0, 3\}$ let $\omega_g(f) = \omega_\ell(f) = \frac{i+j}{2}$, which is the center of the half-open interval associated with f. For $m \in \{1, 2\}$ let $\omega_g(f) = g(v)$ and $\omega_\ell(f) = \ell(v)$ for some $v \in f$. Recall that all $v \in f$ in this case have the same global position and the same local position.

We endow the set F of m-segments, $m \in \{0, 1, 2, 3\}$ with a global total order $\sqsubseteq_g$ and a local total order $\sqsubseteq_\ell$. In the following let $h \in \{g, \ell\}$, so h refers either to the global or to the local situation.

For $f = x[m; i, j]$, $f' = x[m'; i', j']$ we write $f \sqsubseteq_h f'$ if one of the following conditions holds:

1. $\omega_h(f) < \omega_h(f')$.
2. $\omega_h(f) = \omega_h(f')$, and $m < m'$.
3. $\omega_h(f) = \omega_h(f')$, $m = m'$, and $i < i'$.
4. $\omega_h(f) = \omega_h(f')$, $m = m'$, $i = i'$, and $j \leq j'$.

It is clear that $\sqsubseteq_g$ and $\sqsubseteq_\ell$ are both total orders. The next proposition has several important consequences. For example, it implies that $\sqsubseteq_g$ and $\sqsubseteq_\ell$ are both linearizations of the partial order $\leq$ of x.

Proposition 4. *Let $v, v' \in x$ be vertices such that $v \in f = x[m; i, j]$ and $v' \in f' = x[m'; i', j']$, where $f, f' \in F$. Then we have:*

- *If $v \leq v'$ or $g(v) < g(v')$ holds, then $f \sqsubseteq_g f'$.*
- *If $v \leq v'$ or $\ell(v) < \ell(v')$ holds, then $f \sqsubseteq_\ell f'$.*

Corollary 2. *Let $F = \{f_1, \ldots, f_d\}$ be sorted such that $f_i \sqsubseteq_g f_{i+1}$ for all i and let $y_i = y_{f_i}$ be the associated factors of $\mathrm{nf}(x)$. Then we have $\mathrm{nf}(x) = y_1 \cdots y_d$.*

Corollary 3. *Let $F = \{f_1, \ldots, f_d\}$ be sorted such that $f_i \sqsubseteq_\ell f_{i+1}$ for all i and let $y_i = y_{f_i}$ be the associated factors of $\mathrm{nf}(x)$. Then there exists some c with $0 \leq c \leq d$ satisfying*

$$\mathrm{nf}(x_1) = y_1 \cdots y_c \quad \text{and} \quad \mathrm{nf}(x_2) = y_{c+1} \cdots y_d\,.$$

Corollary 4. *Let $f, f' \in F$ be segments such that $f \sqsubseteq_\ell f'$. Then $f' \sqsubseteq_g f$ and $f \neq f'$ imply $\mathrm{alph}(y_f) \times \mathrm{alph}(y_{f'}) \subseteq I$.*

For the final step we consider the normal forms of x, x_1, x_2 and express them as products of factor traces y_f associated with segments $f \in F$. We sort F such that $f_1 \sqsubseteq_\ell \cdots \sqsubseteq_\ell f_d$. Let $\pi \in \mathrm{Perm}(d)$ be the permutation such that $f_{\pi(1)} \sqsubseteq_g \cdots \sqsubseteq_g f_{\pi(d)}$ is F sorted with respect to $\sqsubseteq_g$. By Cor. 4 we have $\mathrm{alph}(y_i) \times \mathrm{alph}(y_j) \subseteq I$ whenever $(i - j)(\pi(i) - \pi(j)) < 0$. (As above, y_i is the factor trace associated with $f_i \in F$.) This completes the proof of Thms. 2 and 1.

Acknowledgment We thank Yuri Matiyasevich for various contributions which were at the beginning of this work.

References

1. A. V. Anisimov and D. E. Knuth. Inhomogeneous sorting. *International Journal of Computer and Information Sciences*, 8:255–260, 1979.
2. M. Benois. Parties rationelles du groupe libre. *C. R. Acad. Sci. Paris, Sér. A*, 269:1188–1190, 1969.
3. V. Diekert. *Combinatorics on Traces.* LNCS 454. Springer, 1990.
4. V. Diekert, C. Gutiérrez, and C. Hagenah. The existential theory of equations with rational constraints in free groups is PSPACE-complete. In *Proc. 18th Ann. Symp. on Theor. Aspects of Comp. Sci. (STACS'01)*, LNCS 2010:170–182, Springer, 2001.
5. V. Diekert, Yu. Matiyasevich, and A. Muscholl. Solving word equations modulo partial commutations. *Theoretical Computer Science*, 224:215–235, 1999. Special issue of LFCS'97.
6. V. Diekert and G. Rozenberg, editors. *The Book of Traces.* World Scientific, Singapore, 1995.
7. C. Droms. Isomorphisms of graph groups. *Proc. American Mathematical Society*, 100:407–408, 1987.
8. C. Gutiérrez. Satisfiability of equations in free groups is in PSPACE. In *Proc. 32nd Ann. ACM Symp. on Theory of Computing, STOC'2000*, pp. 21–27. ACM Press 2000.
9. A. Kościelski and L. Pacholski. Makanin's algorithm is not primitive recursive. *Theoretical Computer Science*, 191:145–156, 1998.
10. D. Kozen. Lower bounds for natural proof systems. In *Proc. 18th Ann. Symp. on Found. of Comp. Sci., FOCS'77*, pp. 254–266, IEEE Computer Society Press 1977.
11. G. S. Makanin. The problem of solvability of equations in a free semigroup. *Math. Sbornik*, 103:147–236, 1977. English transl. in Math. USSR Sbornik 32 (1977).
12. G. S. Makanin. Equations in a free group. *Izv. Akad. Nauk SSR*, Ser. Math. 46:1199–1273, 1983. English transl. in Math. USSR Izv. 21 (1983).
13. G. S. Makanin. Decidability of the universal and positive theories of a free group. *Izv. Akad. Nauk SSSR*, Ser. Mat. 48:735–749, 1984. In Russian; English translation in: *Math. USSR Izvestija, 25*, 75–88, 1985.
14. Yu. Matiyasevich. Some decision problems for traces. In *Proc. 4th Int. Symp. on Log. Found. of Comp. Sci. (LFCS'97)*, LNCS 1234: 248–257, Springer, 1997. Invited lecture.
15. A. Mazurkiewicz. Concurrent program schemes and their interpretations. DAIMI Rep. PB 78, Aarhus University, Aarhus, 1977.
16. A. Mazurkiewicz. Trace theory. In *Petri Nets, Applications and Relationship to other Models of Concurrency*, LNCS 255: 279–324, Springer, 1987.
17. E. Ochmański. Regular behaviour of concurrent systems. *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 27:56–67, 1985.
18. E. Ochmański. Recognizable trace languages. In *The Book of Traces*, Chapter 6: 167–204. World Scientific, Singapore, 1995.
19. W. Plandowski. Satisfiability of word equations with constants is in PSPACE. In *Proc. 40th Ann. Symp. on Found. of Comp. Sci., FOCS'99*, pages 495–500. IEEE Computer Society Press 1999.
20. K. U. Schulz. Makanin's algorithm for word equations — Two improvements and a generalization. In *Word Equations and Related Topics*, LNCS 572: 85–150, Springer 1991.