Solving equations with rational constraints: Plandowski's method

Volker Diekert [1]

Alagna, December 2008

[1] Joint work with Claudio Gutierrez and Christian Hagenah

Background

The existential theory of equations in free monoids is decidable. The existential and positive theories of equations in free groups are decidable. These are celebrated result of Makanin published 1977, 1982 and 1984 . Makanin did not discuss complexity issues, but later it was shown that **his** algorithm for free groups is not primitive recursive.

The best known bound to date is PSPACE by an extension of Plandowski's techniques for solving word equations.

We deal with also with rational constraints, that is, the solution has to respect a specification given by a regular word language.

Notations

An *involution* on a set is a bijection $\overline{}$ such that $\overline{\overline{x}} = x$.
An *involution* on a monoid. In addition:

$$\overline{1} = 1, \quad \overline{xy} = \overline{y}\,\overline{x}.$$

A *factor* of a word $w \in \Sigma^*$ is a word v such that $w = w_1 v w_2$.
Elements of $F(\Sigma)$ are represented by the regular (!) set of freely reduced words over $\Gamma = \Sigma \cup \overline{\Sigma}$.
The involution is extended to Γ^* by $\overline{a_1 \cdots a_n} = \overline{a_n} \cdots \overline{a_1}$.

Rational and *recognizable* subsets

Let M be a monoid.
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*.

Rational subsets in groups

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 Γ for specifying rational group languages over $F(\Sigma)$.

Benois result

Proposition (Michele Benois)

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

The *existential theory of equations with rational constraints*

Let Ω be a set of variables (or unknowns).
Atomic formulae are either $L = R$, where $L, R \in (\Gamma \cup \Omega)^*$ or $X \in C$, where $X \in \Omega$ and $C \subseteq M$ is rational.
The existential theory of equations with rational constraints in M is the set of all closed existentially quantified formulae which are *true* in M.

The main result

Theorem

The following problem is PSPACE*–complete.*
INPUT: A finite alphabet Σ *and a closed existentially quantified formula with rational constraints in the free group* $F(\Sigma)$*.*
QUESTION: Is the formula true *in* $F(\Sigma)$*?*

No negations for the existential theory

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.
$X \notin \{1\} \iff X \in F(\Sigma) \setminus \{1\}$ is a rational constraint!

Reduction to Free Monoids with Involution

Theorem

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

Reduction

Proposition

There is a polynomial time reduction of problem over free groups to free monoids with involution.

Lemma

Let $u, v, w \in \Gamma^*$ be freely reduced words. Then:

$$uvw = 1 \in F(\Sigma)$$

if and only if

$$\begin{aligned} \exists P, Q, R \in \Gamma^* : \\ u &= P\overline{Q} \\ w &= R\,\overline{P} \\ v &= Q\overline{R} \end{aligned}$$

Boolean matrices

It is better to work with Boolean matrices instead of finite automata.
We have a natural involution:

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

where

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

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

Equation with constraints

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 given by a mapping $\sigma : \Omega \to \Gamma^*$ such that the following three conditions are satisfied:

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

Yet another formulation

Theorem

The following problem is $\textsc{Pspace}$*–complete.*
INPUT: An equation with constraints, $E = (\Gamma, h, \Omega, \rho; L = R)$.
QUESTION: Is there a solution $\sigma : \Omega \to \Gamma^*$*?*

The New Look

Our input is given by three items: a single word equation $L = R$ with $L, R \in (\Gamma \cup \Omega)^+$ and two lists: $(X_j \in C_j, 1 \leq j \leq m)$ and $(X_j \notin C_j, m < j \leq k)$. Each regular language $C_j \subseteq \Gamma^*$ is specified by some non-deterministic automaton $\mathcal{A}_j = (Q_j, \Gamma, \delta_j, I_j, F_j)$ where Q_j is the set of states, $\delta_j \subseteq Q_j \times \Gamma \times Q_j$ is the transition relation, $I_j \subseteq Q_j$ is the subset of initial states, and $F_j \subseteq Q_j$ is the subset of final states, $1 \leq j \leq k$.

Road-Map

The proof of the result 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.
- 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 a direct evaluation of an equation without variables.

Why PSPACE-hardness?

Proposition

The following problems are PSPACE*–complete.*

INPUT: A matrix $B \in \mathbb{B}^{n \times n}$ *and a homomorphism* $g : \Gamma^* \longrightarrow \mathbb{B}^{n \times n}$ *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* $h : \Gamma \rightarrow M_{2n}$ *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* $w = \overline{w}$*?*

The Exponent of Periodicity

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

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* $\exp(\sigma(L)) \in 2^{\mathcal{O}(d + n \log n)}$.

Exponential Expressions

Definition

- Every word $w \in \Gamma^*$ is an exponential expression.
- Let e, e' be exponential expressions. Then ee' is an exponential expression.
- Let e be an exponential expression and $k \in \mathbb{N}$. Then $(e)^k$ is an exponential expression.
  Its size is $\|(e)^k\| = \|e\| + \log_2(k)$.

Lemma

Let w be represented by some exponential expression of size p. Then we can find for any factor u an exponential expression of size at most p^2.

Base Changes

The first transformation rule. Replace words by letters.
Let $h : \Gamma^* \to M_{2n}$ be a morphism and $\beta : \Gamma' \to \Gamma^*$ be some mapping such that $\beta(\overline{a}) = \overline{\beta(a)}$.
We call the morphism β a *base change*.
Define:

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

The idea is to move from E to E'.

Lemma

If σ' is a solution of E', then $\sigma = \beta\sigma'$ is a solution of $\beta_*(E')$.

Proof.

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

Basechange

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'*.*

Example

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'.

Projections

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

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

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.

Rule 2 *Let π 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 $\pi^*(E)$.*

Example

$X\overline{X} = Y\bar{a}\bar{b}YZa\overline{Y}$, and $\Gamma = \{a, b, \bar{a}, \bar{b}\}$.
Constraint: $|X| \geq 100$. 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 the new constraint is $|X| \geq 100 \vee X \in \Gamma^* c \Gamma^*$.
Thus, a solution for X might be very short now.

Partial Solutions

A *partial solution* is a mapping $\delta : \Omega \to \Gamma^*\Omega'\Gamma^* \cup \Gamma^*$ such that the following conditions are satisfied:

1. $\delta(X) \in \Gamma^* X \Gamma^*$ for all $X \in \Omega'$,
2. $\delta(X) \in \Gamma^*$ for all $X \in \Omega \setminus \Omega'$,
3. $\delta(\overline{X}) = \overline{\delta(X)}$ for all $X \in \Omega$.

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:

1. $L' = \delta(L)$, $R' = \delta(R)$,
2. $\rho(X) = h(u)\rho'(X)h(v)$ for $\delta(X) = uXv$,
3. $\rho(X) = h(w)$ for $\delta(X) = w \in \Gamma^*$.

Lemma

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

Lemma

The following problem can be solved in PSPACE.

INPUT: Two equations with constraints $E = (\Gamma, h, \Omega, \rho; e_L = e_R)$ *and* $E' = (\Gamma, h, \Omega', \rho'; e_{L'} = e_{R'})$.
QUESTION: Is there some partial solution δ *such that* $\delta_*(E) \equiv E'$*?*

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*

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

Proof.

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

Next we verify whether or not $\delta_*(E) \equiv E'$. During this test we have to create an exponential expression f_L (and f_R, resp.) by replacing X in e_L (and e_R, resp.) with the expression e_X. This increases the size in the worst case by a factor of $\max\{||e_X|| \mid X \in \Omega\}$.

The correctness of the algorithm follows from our assumption 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. □

Guessing partial solutions

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

Example

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

and $\Gamma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. Constraints: $X \in \Gamma^* c \Gamma^*$ 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.

Example

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 equation with constraints in the first example has a solution too.

Admissibility

The input 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 $d = |L_0 R_0|$. We may assume $|\Omega_0| \leq 2d$.

Definition

Let p_0 be a polynomial. The notion of *admissibility* is defined with respect to $p_0(\|E_0\|)$ (which is fixed and can be calculated.)

- An exponential expression e is *admissible*, if $\|e\| \leq p_0(\|E_0\|)$.
- A base change $\beta : \Gamma' \to \Gamma^*$ is *admissible*, if $|\Gamma'| \leq p_0(\|E_0\|)$ 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 $|\Gamma \setminus \Gamma_0| \leq p_0(\|E_0\|)$, $h(a) = h_0(a)$ for $a \in \Gamma_0$, and $e_L e_R$ is admissible.

Search graph

Definition

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 β, a projection π, and a partial solution δ such that $\delta_*(\pi^*(E)) \equiv \beta_*(E')$.

Lemma

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?

Plandowski's algorithm

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 "$\mathrm{eval}(e_L) = \mathrm{eval}(e_R)$"
end

The algorithm returns *true* only if E_0 is solvable.
The challenge is to show that we find a fixed polynomial p_0 such that if E_0 is solvable, then the search graph contains a path to some solvable equation without variables.

Length of a shortest solultion

Remark

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. 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).

START

So far we have done nothing but preparation.
The work starts now.

The set-up from yesterday

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 given by a mapping $\sigma : \Omega \to \Gamma^*$ such that the following three conditions are satisfied:

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

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 $0 \leq \alpha < \beta \leq m$ 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 $0 \leq \alpha, \beta \leq m$ we let $\overline{[\alpha, \beta]} = [\beta, \alpha]$; therefore, $\overline{w[\alpha, \beta]} = w\overline{[\alpha, \beta]}$.

Cuts

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 $\mathrm{r}(i)$.
We have $\mathrm{l}(1) = \mathrm{l}(g+1) = 0$ and $\mathrm{r}(g) = \mathrm{r}(d) = m_0$.
We have $\sigma(x_i) = w_0[\mathrm{l}(i), \mathrm{r}(i)]$ and $\sigma(\overline{x_i}) = w_0[\mathrm{r}(i), \mathrm{l}(i)]$ for $1 \leq i \leq d$.
The interval $[\mathrm{l}(i), \mathrm{r}(i)]$ is positive, because $\sigma(x_i) \neq 1$.
The set of l– and r–positions is the set of *cuts*. Thus, the set of cuts is $\{\, \mathrm{l}(i), \mathrm{r}(i) \mid 1 \leq i \leq d \,\}$. 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.

Equivalent intervals

Let us consider a pair (i, j) such that $i, j \in \{1, \ldots, d\}$ and $x_i = x_j$ or $x_i = \overline{x_j}$. For $\mu, \nu \in \{0, \ldots, \mathrm{r}(i) - \mathrm{l}(i)\}$ we define a relation $\sim$ by:

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

Note that $\sim$ is a symmetric relation.
By $\approx$ we denote the reflexive and transitive closure of $\sim$. Then $\approx$ is an equivalence relation; and $[\alpha, \beta] \approx [\alpha', \beta']$ implies:

1. $[\beta, \alpha] \approx [\beta', \alpha']$.
2. $w_0[\alpha, \beta] = w_0[\alpha', \beta']$.

Free intervals

Definition

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

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

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

has a solution:

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

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
$[1,3] \sim [17,15] \sim [7,9] \sim [11,9] \sim [5,3] \sim [13,15]$.

Maximal free

Definition

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

Maximal free intervals do not overlap.

Lemma

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

Main observation on free intervals

Lemma

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

Proposition

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

Example

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.

Thus, we can assume that the alphabet of constants is Γ.

Critical words

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

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

Each word $u \in C_\ell$ has length 2ℓ, it can be written in the form $u = u_1 u_2$ with $|u_1| = |u_2| = \ell$.

The ℓ-factorization

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

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

such that $w = w_1 \cdots w_k$ and:

- 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 $i = k$.
- $v_i u_{i+1} \in C_\ell$ and these are all of them.

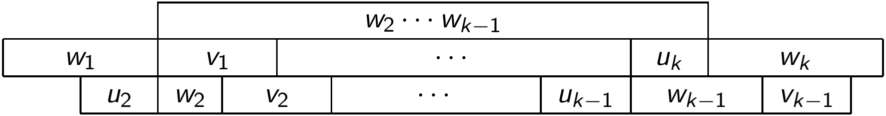

Figure: An ℓ-factorization

If no critical word appears as a factor of w, then $F_\ell(w) = (1, w, 1)$.
The ℓ-factorization of $uv \in C_\ell$ with $|u| = |v| = \ell$ is

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

Body and bodies

Let

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

Define:

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

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

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

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

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

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

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

The ℓ-Transformation

We consider the ℓ-factorization of the solution 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 $1 \leq p \leq q \leq k$ is called an ℓ-*factor*.
$w_0[\alpha, \beta]$ is a factor of $w_p \cdots w_q$.

The ℓ-Transformation

New variables:

$$\Omega_\ell = \{ X \in \Omega_0 \mid \mathrm{body}_\ell(\sigma(X)) \neq 1 \}$$

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 each $X \in \Omega_\ell$ find the subsequences in

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

corresponding to $\mathrm{body}_\ell(\sigma(X))$ replaces these subsequences by X.
The steps above define the ℓ-transformation and yield the following equation:

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

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

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

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:

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

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:

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

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

i) In the extreme case $\ell = m_0$, the ℓ-transformation becomes trivial. Let $a = (1, w_0, 1)$. Then $\overline{a} = (1, \overline{w_0}, 1)$ and $\Gamma_{m_0} = \{a, \overline{a}\} \cup \Gamma$. Moreover, we have $L_{m_0} = R_{m_0} = a$, and $h_{m_0}(a) = h(w_0) \in M_{2n}$. 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 can describe $L_1 \in \Gamma_1^*$ as follows:

For $1 \leq i \leq g$ let $w_i = \sigma(x_i)$ and a_i the last letter of $\sigma(x_{i-1})$ if $i > 1$ and $a_1 = 1$. Let f_i the first letter of $\sigma(x_{i+1})$ if $i < g$ and $f_g = 1$. Let b_i the first letter of w_i and e_i the last letter of w_i.

For $|w_i| = 1$ we replace x_i by the 1-factor (a_i, b_i, f_i).

For $|w_i| = 2$ we replace x_i by the 1-factor $(a_i, b_i, e_i)(b_i, e_i, f_i)$.

For $|w_i| \geq 3$ we let c_i be the second letter of w_i and d_i its second last. In this case we replace x_i by $(a_i, b_i, c_i)x_i(d_i, e_i, f_i)$.

The definition of R_1 is analogous. Thus, we obtain $|L_1 R_1| \leq 3|L_0 R_0| = 3d$, and E_1 is admissible.

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 ℓ-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.

This involves combinatorics on words and many technical details which can be found in the paper:

Volker Diekert, Claudio Gutiérrez, and Christian Hagenah.
The existential theory of equations with rational constraints in free groups is PSPACE-complete.
Information and Computation, 105–140, **202** (2005)