J. Group Theory **12** (2009), 317–330
DOI 10.1515/JGT.2008.080

Journal of Group Theory

Solving one-variable equations in free groups

Dimitri Bormotov, Robert Gilman and Alexei Myasnikov

(Communicated by A. V. Borovik)

Abstract. Equations in free groups have become prominent recently in connection with the solution to the well-known Tarski conjecture. Results of Makanin and Rasborov show that solvability of systems of equations is decidable and there is a method for writing down in principle all solutions. However, no practical method is known; the best estimate for the complexity of the decision procedure is P-space.

The special case of one-variable equations in free groups has been open for a number of years, although it is known that the solution sets admit simple descriptions. We use cancellation arguments to give a short and direct proof of this result and also to give a practical polynomial-time algorithm for finding solution sets. One-variable equations are the only general subclass of equations in free groups for which such results are known.

We improve on previous attempts to use cancellation arguments by employing a new method of reduction motivated by techniques from formal language theory. Our paper is self-contained; we assume only knowedge of basic facts about free groups.

1 Introduction

A one-variable equation $E(x) = 1$ in a finitely generated free group F is an expression of the form

$$u_0 x^{\varepsilon_0} u_1 x^{\varepsilon_1} \dots u_{d-1} x^{\varepsilon_{d-1}} = 1 \tag{1}$$

composed of elements $u_i \in F$, integers $\varepsilon_i = \pm 1$ and a symbol x not in F. A solution to (1) is an element $g \in F$ such that substitution of g for x yields 1 in F.

Lyndon [18] was the first to study equations of this sort. He characterized solution sets in terms of parametric words. The parametric words involved were simplified by Lorents [20], [21] and Appel [1]. However, Lorents announced his results without proof, and Appel's published proof has a gap (see [6]). A complete proof has been provided recently by Chiswell and Remeslennikov [6].

The novel analysis of Chiswell and Remeslennikov involves algebraic geometry ([2], [23]). First they describe the isomorphism types of the coordinate groups of irreducible one-variable equations over F (i.e., equations with solutions sets irreducible in the Zariski topology), and then they deduce the structure of the solution sets. The latter part is easy, but the former requires sophisticated techniques involving ultrapowers and Lyndon length functions. The key point is that coordinate groups of irreducible equations over F are subgroups of the ultrapower F^I/D of F over a countable set I with a non-principal ultrafilter D.

One can view the group F as a subgroup of F^I/D under the canonical diagonal embedding. From this point of view the coordinate groups are precisely the finitely generated subgroups of F^I/D containing F, i.e., the so-called F-subgroups. In particular, up to isomorphism the coordinate groups of irreducible one-variable equations over F are the subgroups of F^I/D of the form $\langle F, g \rangle$ with $g \in F^I/D$.

Investigation of such F-subgroups of F^I/D is not easy and involves a careful analysis of Lyndon functions. (It might be interesting to see whether it is easier to use free actions on Λ-trees.) The computations can be simplified by employing a result from [11] which states that the coordinate groups of irreducible varieties are precisely the finitely generated F-subgroups of the free exponential Lyndon group $F^{\mathbb{Z}[t]}$. As this group is the union of an infinite ascending chain of extensions of centralizers of F (see [24]), one can use Bass–Serre theory to study F-subgroups of $F^{\mathbb{Z}[t]}$.

The method of Chiswell and Remeslennikov is very powerful and potentially useful for more than just free groups. However, it does have the disadvantage of not giving an algorithm for explicitly describing the set of solutions.

This paper is a refinement and extension of [9] where results from formal language theory are used to describe solution sets of one-variable equations in free groups. It seems likely that these arguments can be extended to other groups admitting suitable (not necessarily Lyndon) length functions. The main advantage of this method is that it is short and yields a polynomial-time algorithm for producing a description of all solutions. This algorithm has been implemented by the first author [4].

Theorem 1. *The solution set for a one-variable equation in a free group F is either the whole group F, or the empty set, or a finite union of sets $\{uv^iw \mid i \in \mathbb{Z}\}$, where $u, v, w \in F$. There is a polynomial-time algorithm for finding these sets.*

Since $uv^iw = uw(w^{-1}vw)^i$, the solution set of a one-variable equation in F can be described as a finite (perhaps empty) union of finitely many cosets of centralizers in F.

Let Σ be a set of free generators for F together with their inverses, and let Σ^* be the free monoid over Σ. We consider Equation (1) in terms of words in Σ^*. Each coefficient u_i is represented by a freely reduced word (also denoted by u_i) in Σ^*. From this point of view $E(x) = u_0x^{\varepsilon_1}u_2x^{\varepsilon_2}\dots u_{d-1}x^{\varepsilon_d}u_d$ is a word in the free monoid over $\Sigma \cup \{x, x^{-1}\}$, and a solution to $E(x) = 1$ is a word $s \in \Sigma^*$ such that $E(s)$ is freely equal to the empty word. We assume without loss of generality that $E(x)$ is freely reduced, and call d the degree of $E(x)$. If $d = 0$, then $E(x) = u_0$. In this case the solution set is empty if $E(x) \neq 1$ and all of Σ^* if $E(x) = 1$. If the equation has degree one, it is easy to find its unique solution. From now on we consider only equations of degree at least two.

The first assertion of Theorem 1 is equivalent to saying that for some finite union of sets of words $uv^{\mathbb{Z}}w = \{uv^iw \mid i \in \mathbb{Z}\}$ the solution set consists of all words freely equal to elements of the finite union. A set $uv^{\mathbb{Z}}w$ can be viewed as the set of values of a *parametric* word uv^tw when a new variable (parameter) t is specialized into an arbitrary integer. In terms of algebraic geometry over F (see [2]) the sets $uv^{\mathbb{Z}}w$ are precisely the irreducible components of the solution set of $E(x) = 1$ in F and the parametric words uv^tw are the *generic points* of these varieties [15]. In [18] Lyndon described parametric words as elements of a free operator group $F^{\mathbb{Z}[T]}$ with operators from a polynomial ring $\mathbb{Z}[T]$ in a finite set of indeterminates $T = \{t_1, \dots, t_m\}$. In particular, one can multiply parametric words and take their exponents in $\mathbb{Z}[T]$.

We begin with some lemmas on cancellation, after which we find a finite number of parametric words $uv^{t_i}w$ and $uv^{t_i}wr^{t_j}s$, values of which contain all solutions to $E(x) = 1$ up to free equivalence. Next we show that two parameters are not required and that uv^iw is a solution either for all integers i or for an effectively determined finite subset. At the end we present the algorithm and estimate its time complexity.

To explain our argument in more detail we require a few definitions. For any (word) $g \in F$ we say that the ith occurrence of g *cancels out* in $E(g)$ if there exists a way to freely reduce $E(g)$ such that all letters from g^{ε_i} cancel out during this reduction process.

We say that g is a *pseudo-solution* of $E(x) = 1$ if some occurrence of g cancels out in $E(g)$. Obviously every solution of $E(x) = 1$ is also a pseudo-solution of $E(x) = 1$. However, unlike solutions, pseudo-solutions admit a nice reduction theory.

Our key idea is to study pseudo-solutions of equations instead of solutions. The first result in this direction (stated in [9] in a slightly different form) reduces the situation to cubic equations. Namely, Lemmas 13 and 14 show that if g is a pseudo-solution of $E(x) = 1$ in F then g is a pseudo-solution of a cubic equation of the type

$$x^{\varepsilon_{j-1}} u_j x^{\varepsilon_j} u_{j+1} x^{\varepsilon_{j+1}},$$

where $0 < j < d$ and indices are read modulo d (so $u_d = u_0$). Next in Lemma 15 we show that pseudo-solutions of cubic equations are in fact pseudo-solutions of some particular quadratic equations which one can find effectively. Finally, Lemmas 7 and 8 give a precise description of pseudo-solutions of quadratic one-variable equations over F in terms of parametric words. Combining all these results we obtain description of all pseudo-solutions of $E(x) = 1$ in terms of parametric words in two parameters.

The rest of our proof explains precisely how to use only one parameter to describe solutions of $E(x) = 1$. The method of big powers (see [3]) is the key tool in the second part. This means that the argument is rather general: it works in many other groups that satisfy the big powers condition (see [17]), for example torsion-free hyperbolic groups.

One-variable equations are the only general class of equations in free groups for which a good description of solution sets as well as a practical (polynomial-time) algorithm are known. In his seminal paper [22] Makanin proved decidability of the Diophantine problem in free groups F (whether or not a given equation has a solution in F); however, his original algorithm is very inefficient—not even primitive recursive (see [16]). In the fundamental paper [26] Razborov gave a description of solution sets of arbitrary equations in F. Though this description is extremely complicated, it has been useful in the solution of several deep problems in group theory (e.g. in [5], [12], [13]) including Tarski's problems [14]. In another paper [27] Razborov showed that, in general, there is no easy description of solution sets of equations in F. Later, Plandowski [25] gave a much improved P-space version of the decision algorithm for equations in free monoids, and Gutierrez [10] devised a P-space algorithm for the decision problem for equations in free groups. Recent results due to Diekert, Gutierrez, and Hagenah [7] indicate that the decision problem for equations in free groups might be P-space-complete, though nothing definite has been proven so far. These results on the complexity of the decision problem for equations in free groups and for their solution sets make the existence of subclasses of equations admitting polynomial decision algorithms and descriptions of solutions sets in closed form, all the more remarkable.

2 Cancellation lemmas

As above, Σ is a set of free generators and their inverses for a free group F, and Σ^* is the free monoid over Σ. Let p, q, r, s, t, u, v, w be words in Σ^*. We write $u \sim v$ if u is freely equal to v, and $u \to v$ if u can be reduced to v by cancellation of subwords aa^{-1}, $a \in \Sigma$. In particular $u \to u$. The empty word is denoted by 1, and the length of u is $|u|$. Recall that for any word u there is a unique freely reduced word v such that $u \to v$, and further $u \sim w$ if and only if $w \to v$.

We introduce some additional notation.

Definition 1. Let w be any word.

(1) w' stands for an arbitrary prefix of w and w'' for an arbitrary suffix.

(2) $|w|_c$ is the length of a cyclicly reduced word conjugate to w.

Lemma 1. *If $v \to u$ and $u = u_1u_2 \dots u_m$, then $v = v_1v_2 \dots v_m$ with $v_i \to u_i$.*

Proof. We use induction on n, the number of cancellations necessary to reduce v to u. If $n = 0$, then $u = v$ and there is nothing to prove. Otherwise let the first reduction be $v \to w$. By induction $w = w_1w_2 \dots w_m$ with $w_i \to u_i$. As v is obtained from w by inserting a subword aa^{-1} into some w_i or appending it to the beginning or end of some w_i, v has the desired factorization. □

Lemma 2. *Consider a fixed sequence of cancellations which reduces u to v. If two particular letters of u cancel at some point in the sequence, then either they are adjacent in u or the subword between them has been reduced to 1 by previous cancellations.*

Proof. We use induction on the length of the cancellation sequence. □

Now we slightly generalize the definition of a pseudo-solution of equation to the following situation.

Definition 2. A subword s of w is a pseudo-solution if there is a sequence of cancellations in w which consumes all letters in s.

We are dealing with words over Σ, not group elements. For example $s = ab^{-1}$ is a pseudo-solution of $asba^{-1}a$ but not of asb.

Lemma 3. *Let* $W = u_0v_0u_1v_1 \dots u_{d-1}v_{d-1} \sim 1$. *Then one of the following holds*:

(1) v_0 *is a pseudo-solution of* $u_0v_0u_1v_1$;

(2) v_j *is a pseudo-solution of* $v_{j-1}u_jv_ju_{j+1}v_{j+1}$ *for some* j *strictly between* 0 *and* $d-1$;

(3) v_{d-1} *is a pseudo-solution of* $v_{d-2}u_{d-1}v_{d-1}$.

Proof. Fix a sequence of cancellations which reduces W to 1, and let v_j be the first of the subwords v_i to be consumed. If there is a tie, pick either subword. Observe that the letters in v_j must cancel with nearby letters in W. If a letter in v_j cancelled to the right of v_{j+1}, then by Lemma 2, v_{j+1} would disappear before v_j. Likewise no letter of v_j cancels to the left of v_{j-1}. This proves the lemma. □

The next two lemmas can be proved by straightforward induction on the length of an appropriate cancellation sequence.

Lemma 4. *If* s *is a pseudo-solution of* $w = usv$, *then* $s = s_1s_2$ *with* s_1 *a pseudo-solution of* us_1 *and* s_2 *a pseudo-solution of* s_2v.

Lemma 5. *Let* s *be a pseudo-solution of* w, *and fix a cancellation sequence. The smallest subword of* w *which contains* s *and all letters in* w *cancelling with letters of* s *is freely equal to* 1.

Lemma 6. *A subword* s *of* w *is a pseudo-solution if and only if there is a word* t *such that* s *is a subword of* t, t *is a subword of* w, *and* $t \sim 1$.

Proof. If t exists, then $t \sim 1$ implies $t \to 1$ whence t and all of its subwords are pseudo-solutions of w. For the converse apply Lemma 5. □

Lemma 7. *If* us *and* sw *are freely reduced and if either occurrence of* s *is a pseudo-solution of* $usvsw$, *then* $s \sim v_3^{-1}v_1^{-1}$ *for some factorization* $v = v_1v_2v_3$.

Proof. We argue by induction on n, the length of a cancellation sequence. If $n = 0$, then $s = 1$ in which case we take $v_1 = v_3 = 1$ and $v_2 = v$. If $v = 1$, then $usvsw = ussw$. As us and sw are freely reduced, the only reduction possible involves cancellation at the boundary between us and sw. It follows that $ss \sim 1$, whence $s \sim 1$.

Assume that $n > 0$ and $v \neq 1$. If the first reduction is within v, then $v \to p$ and by induction $s \sim p_3^{-1}p_1^{-1}$ for some factorization $p = p_1p_2p_3$. Lemma 1 implies $v = v_1v_2v_3$ with $v_i \sim p_i$ and $s \sim v_3^{-1}v_1^{-1}$.

The remaining possibilities are cancellation at the boundary between s and v or the boundary between v and s. Consider the first case; the second is similar. We have $s = ta^{-1}$ and $v = ap$ for some letter a and words t and p. The induction hypothesis applied to $utpt(a^{-1}w)$ yields $p = p_1p_2p_3$ and $t \sim p_3^{-1}p_1^{-1}$. But then $v = ap = (ap_1)p_2p_3$ and $s = ta^{-1} \sim p_3^{-1}(ap_1)^{-1}$ as desired. □

Lemma 8. *Suppose that $v \not\sim 1$. If us and $s^{-1}w$ are freely reduced and s or s^{-1} is a pseudo-solution of $usvs^{-1}w$, then $s \sim v''v^k$ for some integer k.* (*See Definition* 1.) *Likewise if us^{-1} and sw are freely reduced and s or s^{-1} is a pseudo-solution of $us^{-1}vsw$, then $s \sim v^kv'$.*

Proof. Consider the first part; as before use induction on n, the number of cancellations. If $n = 0$, then $s = 1$. Take $v'' = 1$ and $k = 0$. Otherwise the first reduction is either within v or at one end or the other of v. In the first case $v \to v_1$, and the induction hypothesis applied to $usv_1s^{-1}w$ yields the desired result.

Suppose then that there is a reduction at the left end of v; the other case is similar. We have $s = ta^{-1}$, $v = ap$, and application of the induction hypothesis to $ut(pa)t^{-1}w$ yields $pa = p_1p_2$ and $t \sim p_2(pa)^k$. It follows that

$$s \sim p_2(pa)^k a^{-1} \sim p_2a^{-1}a(pa)^k a^{-1} \sim p_2a^{-1}v^k.$$

If $p_2 \neq 1$, then $p_2 = v_2a$ for some suffix v_2 of v whence $s \sim v_2v^k$. If $p_2 = 1$, then

$$s \sim a^{-1}v^k \sim a^{-1}vv^{k-1} \sim pv^{k-1}.$$

As p is a suffix of v, the first assertion holds. The second assertion follows from the first upon replacement of s by s^{-1}. □

Lemma 9. *If s is a pseudo-solution of tus, t is a pseudo-solution of tvs, and st is freely reduced, then $s \sim (v^{-1}u)^i(v^{-1}u)'$ and $t \sim (vu^{-1})''(vu^{-1})^j$ for some integers i, j.*

Proof. Application of Lemma 5 to tus implies either $u = u_1u_2$ with $u_2s \sim 1$ or $t = t_1t_2$ with $t_2us \sim 1$. Consider the first case. We have $s \sim u_2^{-1} \sim (v^{-1}u)^{-1}(v^{-1}u_1)$ as required. Further $u_2^{-1} \to s$ implies that t is a pseudo-solution of tvu_2^{-1} and hence of tvu^{-1}. Thus either $v = v_1v_2$ with $t \sim v_1^{-1}$ or $u = u_3u_4$ with $t \sim (vu_4^{-1})^{-1}$. But then $t \sim (v_2u^{-1})(vu^{-1})^{-1}$ or $t \sim (u_3^{-1})(vu^{-1})^{-1}$, and we see that t has the right form. A similar analysis starting with tvs also works.

It remains to consider the case $t = t_1t_2$ with $t_2us \sim 1$ and $s = s_1s_2$ with $tvs_1 \sim 1$. Suppose that $u \sim v$. We have $t = t_1t_2$ with $t_2us \sim 1$ and $s = s_1s_2$ with $tus_1 \sim 1$. If $t_1 = 1$, then $tus \sim 1$ implies $st \sim u^{-1}$. As st is freely reduced, Lemma 1 yields $s \sim (u^{-1})'$, $s \sim (v^{-1})'$, and $t \sim (u^{-1})''$ which is included in $i = j = 0$. If $t_1 \neq 1$, it follows from $t_2us_1s_2 \sim 1 \sim t_1t_2us_1$ that $t_1 \sim s_2$. As st is freely reduced, t_1 and s_2 are too. Thus $t_1 = s_2 \neq 1$. Hence $s_1s_2t_2 = s_1t_1t_2$ is freely reduced. But then $s_1s_2t_2 \sim u^{-1} \sim s_1t_1t_2$ implies $s \sim (v^{-1})'$, $t \sim (u^{-1})''$ as before.

Finally suppose that $t = t_1t_2$ with $t_2us \sim 1$, $s = s_1s_2$ with $tvs_1 \sim 1$, and $u \not\sim v$. From $t_2us \sim 1$ we deduce $u^{-1}t_2^{-1} \to s$. Hence t is a pseudo-solution of $tvu^{-1}t_2^{-1}$ and all the more of $tvu^{-1}t_2^{-1}t_1^{-1} = tvu^{-1}t^{-1}$. Likewise s is a pseudo-solution of $s^{-1}v^{-1}us$. We are done by Lemma 8. □

Lemma 10. *Let st be freely reduced. If the right-hand occurrence of s is a pseudo-solution in $stus$ but not in tus, then $st \sim u_3^{-1}u_1^{-1}$ for some factorization $u = u_1u_2u_3$. Likewise if the left-hand occurrence of t is a pseudo-solution in $tvst$ but not in tvs, then $st \sim v_3^{-1}v_1^{-1}$ for some factorization $v = v_1v_2v_3$.*

Proof. Consider the first part; the second is treated similarly. We have $s = pq$ with $q \neq 1$ and $qtus \sim 1$ (by Lemma 5). Since st is freely reduced, so is qt. It follows that t is a pseudo-solution of tus. If t is not a pseudo-solution of tu, then $s = ef$ with $tue \sim 1$ (by Lemma 5). But then $qf \sim 1$ forces f to be a pseudo-solution of $pqf = sf = eff$, and Lemma 7 implies $f \sim 1$. Consequently $tuef = tus \sim 1$ contrary to our hypothesis that s is not a pseudo-solution of tus.

It remains to deal with the possibility that t is a pseudo-solution of tu. In this case $u = w_1w_2$ with $t \sim w_1^{-1}$ (by Lemma 5). It follows that $qw_2s \sim 1$ whence the right-hand occurrence of s is a pseudo-solution in sw_2s. An application of Lemma 7 completes the proof. □

Lemma 11. *Suppose that $p^iuq^j \sim v$, $i, j \geqslant 0$, and*

$$i|p|_c + j|q|_c \geqslant 2|p| + 2|q| + |u| + |v|.$$

(*Recall Definition* 1.) *Further assume that p and q are not freely equal to proper powers. Under these conditions $uqu^{-1} \sim p^{-1}$.*

Proof. Assume that the lemma holds when both p and q are cyclicly reduced, and consider the case that they are not. Free reduction of p and q yields reduced words $p_1p_2p_1^{-1} \sim p$, $q_1q_2q_1^{-1} \sim q$ with both p_2, q_2 cyclicly reduced. Hence $p_2^i(p_1^{-1}uq_1)q_2^j \sim p_1^{-1}vq_1$. Rewriting $p_1^{-1}uq_1$ as u_2 and $p_1^{-1}vq_1$ as v_2 we obtain $p_2^iu_2q_2^j \sim v_2$. As

$$i|p_2|_c + j|q_2|_c = i|p|_c + j|q|_c \geqslant 2|p| + 2|q| + |u| + |v| \geqslant 2|p_2| + 2|q_2| + |u_2| + |v_2|,$$

we have $u_2q_2u_2^{-1} \sim p_2^{-1}$. Hence

$$uqu^{-1} \sim (p_1u_2q_1^{-1})(q_1q_2q_1^{-1})(q_1u_2^{-1}p_1^{-1}) \sim p_1u_2q_2u_2^{-1}p_1^{-1} \sim p_1p_2^{-1}p_1^{-1} \sim p^{-1}.$$

It remains to deal with the case that p and q are cyclicly reduced. Without loss of generality assume that u and v are freely reduced and $i, j \geqslant 0$. Thus there is a sequence of

$$\tfrac{1}{2}(|p^iuq^j| - |v|) = \tfrac{1}{2}(i|p| + j|q| + |u| - |v|) \geqslant |p| + |q| + |u|$$

cancellations which reduces p^iuq^j to v.

Since cancellation can occur only at either end of u, the first $|u|$ cancellations must consume u. In other words u cancels with a suffix of p^i and a prefix of q^j. For some factorizations $p = p_1 p_2$ and $q = q_1 q_2$ we have $u = (p_2 p^{i_2})^{-1}(q^{j_1} q_1)^{-1}$ for some $i_1, i_2, j_1, j_2 \geqslant 0$ with $i = i_1 + 1 + i_2$ and $j = j_1 + 1 + j_2$. Consequently $p^{i_1} p_1 q_2 q^{j_2}$ admits at least $|p| + |q|$ cancellations. Thus the infinite sequences $q_2 q_1 q_2 q_1 \ldots$ and $p_1^{-1} p_2^{-1} p_1^{-1} p_2^{-1} \ldots$ have the same prefix of length $|p| + |q|$. As these sequences have periods $|q|$ and $|p|$ respectively, they are identical by [8, Theorem 1]. But then the fact that $(p_1^{-1} p_2^{-1})^{|q|}$ and $(q_2 q_1)^{|p|}$ have the same length implies that they are equal. Since p and q are not proper powers, neither are $(p_1^{-1} p_2^{-1})$ and $q_2 q_1$. It follows that $p_1^{-1} p_2^{-1} = q_2 q_1$, and this equation implies in a straightforward way that $uqu^{-1} \sim p^{-1}$. □

Lemma 12. *Suppose that q^j is a pseudo-solution of $p^i u q^j v r^k$, that*

$$|j|\,|q|_c \geqslant 7(|p| + |u| + |q| + |v| + |r|),$$

and that p, q, r are not proper powers. Then either $q \sim 1$ or $|i| \geqslant 1$ and $u^{-1} p u \sim q^{\pm 1}$ or $|k| \geqslant 1$ and $vrv^{-1} \sim q^{\pm 1}$.

Proof. Without loss of generality assume that $i, j, k \geqslant 0$. By Lemma 4, q factors as $q_1 q_2$ and q^j factors as $(q^{j_1} q_1)(q_2 q^{j_2})$ in such a way that $q^{j_1} q_1$ is a pseudo-solution of $p^i u q^{j_1} q_1$, and $q_2 q^{j_2}$ is a pseudo-solution of $q_2 q^{j_2} v r^k$. Clearly one of j_1, j_2 is no smaller than $(j-1)/2$. Assume that it is j_1; the argument is similar in the other case.

By Lemma 5, $q^{j_1} q_1$ extends to a suffix of $p^i u q^{j_1} q_1$ which is freely equal to 1. If that suffix is contained in $u q^{j_1} q_1$, then $u = u_1 u_2$ with $q^{j_1} \sim u_2^{-1} q_1^{-1}$. Hence q^{j_1} freely reduces to a word w with $|w| \leqslant |u| + |q|$. On the other hand

$$|w| \geqslant j_1 |q|_c \geqslant .5(j-1)|q|_c \geqslant 3.5(|p| + |u| + |q|) - .5|q| \geqslant 3(|p| + |q| + |u|).$$

But then $|p| = |q| = |u| = 0$, which implies $q \sim 1$.

It remains to consider the case that the suffix is not contained in $u q^{j_1} q_1$. In particular $i \geqslant 1$. For some factorization $p = p_1 p_2$ and $m \leqslant i$ we have $p_2 p^m u q^{j_1} q_1 \sim 1$. Thus $p^m u q^{j_1} \sim p_2^{-1} q_1^{-1}$. As above

$$j_1 |q|_c \geqslant 3(|p| + |q| + |u|) \geqslant 2|p| + 2|q| + |u| + |p_2^{-1} q_1|.$$

Lemma 11 applies and yields $uqu^{-1} \sim p^{\pm 1}$. □

3 Parametric words

In this section we show how to find a finite set of words and parametric words $uv^{t_1} w r^{t_2} s$, values of which together contain all solutions to the equation $E(x) = 1$, where

$$E(x) = u_0 x^{\varepsilon_0} u_1 x^{\varepsilon_1} \ldots u_{d-1} x^{\varepsilon_{d-1}}.$$

Let s be any fixed freely reduced word which is a solution to $E(x) = 1$.

Lemma 13. *One of the following holds*:

(1) s^{ε_0} *is a pseudo-solution of* $u_0 s^{\varepsilon_0} u_1 s^{\varepsilon_1}$;

(2) s^{ε_j} *is a pseudo-solution of* $s^{\varepsilon_{j-1}} u_j s^{\varepsilon_j} u_{j+1} s^{\varepsilon_{j+1}}$ *for some* j *strictly between* 0 *and* $d-1$;

(3) $s^{\varepsilon_{d-1}}$ *is a pseudo-solution of* $s^{\varepsilon_{d-2}} u_{d-1} s^{\varepsilon_{d-1}}$.

Proof. This follows from Lemma 3 for $W = E(s)$. □

It is convenient to use the following immediate consequence of Lemma 13.

Lemma 14. s^{ε_j} *is a pseudo-solution of* $s^{\varepsilon_{j-1}} u_j s^{\varepsilon_j} u_{j+1} s^{\varepsilon_{j+1}}$ *for some* j *between* 0 *and* $d-1$. *Here indices are read modulo* d; *e.g.*, $u_d = u_0$.

Remark 1. Lemmas 13 and 14 hold also (by the same argument) in the case when ε_i are arbitrary integers (not only numbers 1 or -1).

It follows from Lemma 14 that application of the following lemma to all successive pairs of coefficients $u = u_i$, $v = u_{i+1}$ (with indices read modulo d) yields a set of words and parametric words containing s or s^{-1} for every solution s to $E(x) = 1$.

Lemma 15. *If* $\alpha, \beta = \pm 1$ *and* s *is a freely reduced pseudo-solution to* $s^{\alpha} u s v s^{\beta}$, *then one of the following holds.* (*Recall Definition* 1.)

(1) $s \sim (v^{-1}u)^i (v^{-1}u)'(vu^{-1})''(vu^{-1})^j$.

(2) $s \sim (u^{-1})'(u^{-1})''$ *or* $(v^{-1})'(v^{-1})''$.

(3) $s \sim (u^{-1})' v^i v' v'' v^j$ *or* $u^i u' u'' u^j (v^{-1})''$.

(4) $s \sim u^i u' v'' v^j$.

Proof. By Lemma 4 we have $s = s_1 s_2$ with s_1 a pseudo-solution of $s^{\alpha} u s_1$ and s_2 a pseudo-solution of $s_2 v s^{\beta}$. There are four cases. First if $\alpha = -1$, $\beta = -1$, Lemma 8 applied to $s_2^{-1} s_1^{-1} u \underline{s_1}$ and $\underline{s_2} v s_2^{-1} s_1^{-1}$ yields (4).

If $\alpha = \beta = 1$, we have $s_1 s_2 u \underline{s_1}$ and $\underline{s_2} v s_1 s_2$ where the pseudo-solutions are underlined. It may happen that s_1 is pseudo-solution of $s_2 u s_1$ and s_2 is a pseudo-solution of $s_2 v s_1$. In this case Lemma 9 applies and (1) holds. Otherwise either s_1 is not a pseudo-solution of $s_2 u s_1$ or s_2 is not a pseudo-solution of $s_2 v s_1$. In both cases Lemma 10 implies (2).

Suppose that $\alpha = 1$, $\beta = -1$. In this case $s_1 s_2 u \underline{s_1}$ and $\underline{s_2} v s_2^{-1} s_1^{-1}$. By Lemma 10, either s is included in (2) or $s_2 u \underline{s_1}$ whence s_1 is freely equal to the inverse of a suffix of $s_2 u$. Equivalently s_1 is freely equal to a prefix of $(s_2 u)^{-1}$. But Lemma 8 implies $s_2 \sim v_2 v^j$ for some integer j and factorization $v = v_1 v_2$. It follows from Lemma 1 that s_1 is freely equal to a prefix of $(v_2 v^j u)^{-1}$. Consideration of the possible cases yields (3).

A similar argument works when $\alpha = -1$, $\beta = 1$ and shows that (2) or (3) holds. □

4 Solutions

In order to find all solutions to $E(x) = 1$ we need to test the possibilities given by Lemma 15. It is straightforward to test the single words; the parametric words require more work. They have the form rp^isq^jt or rp^is. Without loss of generality we assume that p and q are not proper powers and $p \not\sim 1 \not\sim q$.

Consider rp^is. Substitute rys for x in $E(x) = 1$ to obtain an equation $E'(y) = v_0y^{\varepsilon_0} \dots v_{d-1}y^{\varepsilon_{d-1}}$ in the indeterminate y with coefficients v_j of the form su_jr, su_js^{-1} etc. Note that rp^is is a solution of $E(x)$ if and only if p^i is a solution of $E'(y)$. Also the sum of the lengths of the coefficients of $E'(y)$ is $|v_0 \dots v_{d-1}| = |u_0 \dots u_{d-1}| + d|rs|$. Denote this number by K_1.

If a coefficient v_j commutes with p, i.e. $v_jp \sim pv_j$, then the subword $y^{\varepsilon_{j-1}}v_jy^{\varepsilon_j}$ of $E'(y)$ may be replaced by $v_jy^{\varepsilon_{j-1}+\varepsilon_j}$ without affecting the set of indices i for which p^i is a solution. This is true even if indices are read modulo d. The coefficients in $E'(y)$ will change, but $E'(1) = v_0 \dots v_{d-1}$ remains constant. In particular the sum of the length of the coefficients is still K_1.

Continue replacements of this sort until reaching an equation of the form $E''(y) = w_0y^{k_0} \dots w_my^{k_m}$ with m minimal. It may be that $m = 0$ and $E''(y) = w_0$. In this case p^i is a solution for all i if $w_0 = v_0 \dots v_{d-1} \sim 1$ and for no i otherwise. Similarly if $E''(y) = w_0y^{k_0}$, then p^i is a solution if and only if $w_0 \sim p^{-ik_0}$. In this case the free reduction of p^{ik_0} is a word of length at least $|ik_0|\,|p|_c$ and at most $|w_0| = K_1$. Consequently $|i|\,|p|_c \leqslant |ik_0|\,|p|_c \leqslant K_1$.

The remaining possibility is that $E''(y) = w_0y^{k_0} \dots w_my^{k_m}$ with $m \geqslant 2$, all $k_j \neq 0$ and no w_j commuting with p. No w_j conjugates p to p^{-1} either, as p and p^{-1} are not conjugate in the free group F. If p^i is a solution, then by Lemma 14 (with $E''(y)$ in place of $E(x)$) some p^{ik_j} must be a pseudo-solution of $p^{ik_{j-1}}w_jp^{ik_j}w_{j+1}p^{ik_{j+1}}$. Lemma 12 now implies that $|i|\,|p|_c < 7(|p| + |w_j| + |p| + |w_{j+1}| + |p|) \leqslant 21|p| + 7K_1$. We have proved the following lemma.

Lemma 16. *If rp^is is a solution to $E(x) = 1$ for some i with*

$$|i|\,|p|_c > 21|p| + 7(|u_0 \dots u_{d-1}| + d|rs|),$$

then rp^is is a solution for all i.

Consider a solution rp^isq^jt to $E(x) = 1$. Define

$$K_2 = 21 \max\{|p|, |q|\} + 7(|u_0 \dots u_{d-1}| + d|rst|).$$

We will show that either $|i|\,|p|_c$ or $|j|\,|q|_c$ is no larger than K_2d. Thus each parametric word rp^isq^jt from Lemma 13 with two parameters may be replaced by a collection of parametric words with just one parameter, namely $rp^{i_0}sq^jt$, $rp^isq^{j_0}t$ with $|i_0|\,|p|_c \leqslant K_2d$ and $|j_0|\,|q|_c \leqslant C_2d$.

Without loss of generality suppose that $i, j \geqslant 0$, and p and q are not proper powers. In particular the centralizers in the free group F of p and q are the cyclic subgroups generated by p and q respectively.

Lemma 17. *Suppose that p is not conjugate to q or q^{-1} and rp^isq^jt is a solution to $E(x) = 1$. Then either $|i|\,|p|_c$ or $|j|\,|q|_c$ is no larger than K_2, where*

$$K_2 = 21\max\{|p|,|q|\} + 7(|u_0 \dots u_{d-1}| + d|rst|).$$

Proof. First suppose that $s \sim 1$ and take the solution to be rp^iq^jt. Write $E(rp^iq^jt)$ in the form $E(rp^iq^jt) = v_0(p^iq^j)^{\varepsilon_1}v_1 \dots v_{d-1}(p^iq^j)^{\varepsilon_d}v_d$. Rewrite now each $(p^iq^j)^{\varepsilon_k}$ as $(p^i)^{\varepsilon_k}(q^j)^{\varepsilon_k}$ for $\varepsilon_k = 1$ and each $(p^iq^j)^{\varepsilon_k}$ as $(q^j)^{\varepsilon_k}(p^i)^{\varepsilon_k}$ for $\varepsilon_k = -1$. Denote the resulting product by W. Call the elements $(p^i)^{\varepsilon_k}$ and $(q^j)^{\varepsilon_k}$ powers, and the elements v_k coefficients. Consider how a coefficient v_k might conjugate the power on one side of itself in W to the power on the other side. As p is not conjugate to q or q^{-1}, v_k would either lie in a subword $v_{k-1}p^iq^jv_kq^{-j}p^{-i}v_{k+1}$ ($\varepsilon_{k-1} = 1, \varepsilon_k = -1$) and centralize q or in a subword $v_{k-1}q^{-j}p^{-i}v_kp^iq^jv_{k+1}$ ($\varepsilon_{k-1} = -1, \varepsilon_k = 1$) and centralize p. Notice that in the former case $v_k = tu_kt^{-1}$, and in the latter one $v_k = r^{-1}u_kr$. Since $E(x)$ is freely reduced $u_k \neq 1$, hence $v_k \neq 1$. Consequently v_k is freely equal to a non-trivial power of q in the first case and a non-trivial power of p in the second. In this event W is freely equal to the word obtained by deleting the powers on each side of v_k.

Let W' be the word obtained from W by performing all of the deletions discussed in the previous paragraph. Notice that the first and last powers of W survive and that the new coefficients are either old coefficients which do not conjugate their adjacent powers into each other or products $v_kv_{k+1} \dots v_{k+m}$ of successive coefficients whose adjacent powers in W have been deleted. In the latter case the coefficient is an alternating product of non-trivial powers of p and q.

Since $W' \sim 1$, Lemma 3 implies that some power is a pseudo-solution in a subword of W' consisting of up to three powers and the coefficients between them. The sum of the length of the coefficients of W' is the same as that of W, namely $\sum|u_k| + d|r| + d|s|$. If $|i|\,|p|_c$ and $|j|\,|q|_c$ exceed the bound given above, then Lemma 12 applies and (as p is not conjugate to q or q^{-1}) implies that some coefficient conjugates one adjacent power to the other. But this is impossible either because the coefficient is inherited from W or because the coefficient is an alternating product of non-trivial powers of p and q, and the conjugation would be a non-trivial relation satisfied by p and q, which generate a free group of rank two.

It remains to reduce to the case $s \sim 1$. Assume that $s \not\sim 1$, and rewrite the solution as $rp^i(sqs^{-1})^j(st)$. One of $|i|\,|p|_c$ or $|j|\,|sqs^{-1}|_c = |j|\,|q|_c$ is at most $21\max\{|p|,|q|\} + 7(|u_0 \dots u_{d-1}| + d|rst|)$. $\square$

Finally, consider a solution rp^isq^jt to $E(x) = 1$ with p conjugate to q or q^{-1}. With appropriate changes to r, s, t and j, rp^isq^jt may be rewritten as rp^isp^jt where p is cyclicly reduced. If s commutes with p then $rp^isp^jt = rp^{i+j}st$ and Lemma 16 applies. Now we may assume that s does not commute with p.

Define $W = E(rp^isp^jt) = v_0(p^isp^j)^{\varepsilon_1}v_1 \dots v_{d-1}(p^isp^j)^{\varepsilon_d}v_d$, and argue as before. The coefficients now include the subwords $s^{\pm 1}$ as well as the coefficients v_k. Consider how a v_k might conjugate the power on one side of itself to the power on the other side. Since all powers are powers of p, v_k would commute with p and hence would itself be freely equal to a power of p. If $\varepsilon_k\varepsilon_{k+1} = -1$, then $v_k \not\sim 1$ and the powers on each side cancel. However, if $\varepsilon_k \varepsilon_{k+1} = 1$, then the powers do not necessarily cancel but combine to form a power $p^{\pm(i+j)}$.

Let W' be the word obtained from W by performing all of the deletions and combinations of powers discussed in the previous paragraph. Notice that the first and last powers of W survive and that the new subwords between powers surviving from W are either coefficients from W which do not conjugate their adjacent powers into each other or alternating products $s^{\pm 1} p^{k_m} v_m s^{\pm 1} p^{k_{m+1}} v_{m+1} \dots s^{\pm 1} p^{k_n} v_n s^{\pm 1} s$ where the elements v_j that occur are freely equal to powers of p. Further if p^{k_j} occurs between s and s^{-1}, then $k_j = 0$ and v_j is freely equal to a non-trivial power of p, while if p^{k_j} occurs between s and s or s^{-1} and s^{-1}, then $k_j = \pm(i+j)$.

There are two possibilities. First, if $|(i+j)|\,|p|_c \geqslant K_2$, then we may consider the subwords $p^{\pm(i+j)}$ to be powers like the terms $p^{\pm i}$ and $p^{\pm j}$ surviving from W and the subwords between the powers as coefficients. Again, by Lemma 3 some power is a pseudo-solution in a subword of W' consisting of up to three powers and the coefficients between them. Then Lemma 12 applies and implies that some coefficient conjugates one adjacent power to the other and hence is a power of p. But this is impossible either because the coefficient is inherited from W or because the coefficient is an alternating product of non-trivial powers of p and s, and the conjugation would be a non-trivial relation satisfied by the subgroup generated by p and s, which is free of rank two.

Second, if $|(i+j)|\,|p|_c < K_2$, we take just the elements $p^{\pm i}$ and $p^{\pm j}$ from W to be powers. The coefficients are either inherited from W or alternating products

$$s^{\pm 1} p^{k_m} v_m s^{\pm 1} p^{k_{m+1}} v_{m+1} \dots s^{\pm 1} p^{k_n} v_n s^{\pm 1} s$$

as above. In this case $|(i+j)|\,|p| = |(i+j)|\,|p_c| \leqslant K_2$, and the total length of the coefficients increases to at most $K_2 + (d-1)|(i+j)|\,|p| \leqslant dK_2$. Lemma 12 applies and yields the following lemma.

Lemma 18. *Suppose that p is conjugate to q or q^{-1} and $rp^i sq^j t$ is a solution to $E(x) = 1$. Then $|i|\,|p|_c$ or $|j|\,|q|_c$ is no larger than dK_2.*

5 The algorithm

The algorithm implicit in the preceding analysis may be described as follows.

(1) The input is an equation $u_0 x^{\varepsilon_0} u_1 x^{\varepsilon_1} \dots u_{d-1} x^{\varepsilon_{d-1}} = 1$ of degree $d \geqslant 2$ and with freely reduced coefficients from a free monoid Σ^* over a set Σ of generators and their inverses for a free group F.

(2) Let L be the list of words and parametric words and their inverses from Lemma 15. Rewrite the parametric words so that they are either ordinary words or have one of the forms $rp^i s$ or $rp^i sq^j t$ with $p \not\sim 1 \not\sim q$, p, q not proper powers, and in the latter case $sqs^{-1} \not\sim p^{\pm 1}$.

(3) For each ordinary word $w \in L$ test $E(w) \sim 1$ and $E(w^{-1}) \sim 1$. Remove w from L.

(4) Replace each parametric word $rp^i sq^j t$ with words $rp^i sq^{j_0} t$ and $rp^{i_0} sq^j t$ for all i_0, j_0 with $|i_0|\,|p|_c \leqslant dK_2$ and $|j_0|\,|q|_c \leqslant dK_2$ where K_2 is as in Lemma 17.

(5) For each word of the form $w = rp^iq$ in L, if $E(rp^{i_0}s) \sim 1$ where i_0 is the least integer greater than $(1/|p|_c)(21|p|) + 7(|u_0 \dots u_{d-1}| + d|rs|)$, then $x = rp^is$ is a solution for all i, otherwise test $E(rp^{i_1}s) \sim 1$ for all $|i_1| < i_0$.

We leave it to the reader to check that our preceding analysis implies the correctness of the above algorithm. To bound the time complexity let $|L|$ be the length of the list from Step 2 and M the maximum of $|rpsqt|$ for each entry rp^isq^jt. Note that M is also an upper bound for the length of the coefficients of $E(x)$ and that the constant K_2 from Lemma 17 is $O(dM)$.

Steps 2 and 3 are accomplished in time $O(M|L|)$, and Step 4 in time $O(dK_2M|L|)$. Let L' be the augmented list from Step 4. Then $|L'| = O(dK_2|L|) = O(d^2M|L|)$, and each entry in L' has the form rp^is with $|rps| = O(dMK_2) = O(d^2M^2)$.

For each entry there are $O(M + dM + d^2MK_2) = O(d^3M^2)$ tests performed in Step 5. The time to test the entry rp^is is linear in the length of $E(rp^is)$, which is

$$O(d|rp^is| + |u_1 \dots u_{d-1}|) = O(id^3M^2) = O(dK_2d^3M^2) = O(d^5M^3).$$

Thus the total time for Step 5 is

$$O(|L'| \cdot (d^3M^2) \cdot (d^5M^3)) = O((d^2M|L|) \cdot (d^3M^2) \cdot (d^5M^3)) = O(d^{10}M^6|L|).$$

Clearly this estimate bounds the time of the complete algorithm.

Finally let m be the maximum size of a coefficient in $E(x)$. If follows from Lemma 15 that $M = O(m)$ and that $|L| = O(dm^3)$. Thus the time complexity of our algorithm is $O(d^{11}m^9)$.

References

[1] K. Appel. One-variable equations in free groups. *Proc. Amer. Math. Soc.* **19** (1968), 912–918.

[2] G. Baumslag, A. Myasnikov and V. Remeslennikov. Algebraic geometry over groups I. Algebraic sets and ideal theory. *J. Algebra* **219** (1999), 16–79.

[3] G. Baumslag, A. Myasnikov and V. Remeslennikov. Discriminating completions of hyperbolic groups. *Geom. Dedicata* **92** (2003), 115–143.

[4] D. Bormotov. A package of algorithms for solving hard problems in group theory. Dissertation thesis. The Graduate School of CUNY, New York (2005).

[5] I. Bumagin, O. Kharlampovich and A. Myasnikov. Isomorphism problem for finitely generated fully residually free groups. *J. Pure Appl. Algebra* **208** (2007), 961–977.

[6] I. Chiswell and V. N. Remeslennikov. Equations in free groups with one variable I. *J. Group Theory* **3** (2000), 445–466.

[7] V. Diekert, C. Gutierrez and C. Hagenah. The existential theory of equations with rational constraints in free groups is PSPACE-complete. *Inform. Comput.* **202** (2005), 105–140.

[8] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965), 109–114.

[9] R. Gilman and A. G. Myasnikov. One variable equations in free groups via context free languages. In *Computational and experimental group theory*, Contemp. Math. 349 (American Mathematical Society, 2004), pp. 83–88.

[10] C. Gutierrez. Satisfiability of equations in free groups is in PSPACE. In *32nd annual ACM symposium on theory of computing* (*STOC'2000*) (ACM Press, 2000), pp. 21–27.
[11] O. Kharlampovich and A. Myasnikov. Irreducible affine varieties over a free group. I. Irreducibility of quadratic equations and Nullstellensatz. *J. Algebra* **200** (1998), 472–516.
[12] O. Kharlampovich and A. Myasnikov. Effective JSJ decompositions. In *Algorithms*, *languages*, *logic*, Contemp. Math. 378 (American Mathematical Society, 2005), pp. 87–212.
[13] O. Kharlampovich and A. Myasnikov. Implicit function theorems over free groups. *J. Algebra* **290** (2005), 1–203.
[14] O. Kharlampovich and A. Myasnikov. Elementary theory of free nonabelian groups. *J. Algebra* **302** (2006), 451–552.
[15] O. Kharlampovich and A. Myasnikov. Algebraic geometry over free groups: lifting solutions into generic points. In *Algorithms*, *languages*, *logic*, Contemp. Math. 378 (American Mathematical Society, 2005), pp. 213–318.
[16] A. Kościelski and L. Pacholski. Makanin's algorithm is not primitive recursive. *Theoret. Comp. Sci.* **191** (1998), 145–156.
[17] A. Kvaschuk and A. Myasnikov. Big powers and free constructions. *Internat. J. Algebra Comput.*, to appear.
[18] R. C. Lyndon. Equations in free groups. *Trans. Amer. Math. Soc.* **96** (1960), 445–457.
[19] R. C. Lyndon. Groups with parametric exponents. *Trans. Amer. Math. Soc.* **96** (1960), 518–533.
[20] A. A. Lorents. The solution of systems of equations in one unknown in free groups. *Soviet Math. Dokl.* **4** (1963), 262–265.
[21] A. A. Lorents. Representations of sets of solutions of systems of equations with one unknown in a free group. *Soviet Math. Dokl.* **9** (1968), 81–84.
[22] G. S. Makanin. Equations in a free group. *Math USSR-Izv.* **21** (1983), 546–582.
[23] A. Myasnikov and V. Remeslennikov. Algebraic geometry. 2. Logical foundations. *J. Algebra* **234** (2000), 225–276.
[24] A. Myasnikov and V. Remeslennikov. Exponential groups. 2. Extension of centralizers and tensor completion of CSA-groups. *Internat. J. Algebra Comput.* **6** (1996), 687–711.
[25] W. Plandowski. Satisfiability of word equations with constants is in PSPACE. *J. ACM* **51** (2004), 483–496.
[26] A. A. Razborov. On systems of equations in a free group. *Math. USSR-Izv.* **25** (1985), 115–162.
[27] A. A. Razborov. An equation in a free group whose set of solutions does not allow a representation as a superposition of a finite number of parametric functions. In *Proc. 9th all-union sympos. group theory* (Moscow, 1984), p. 54.

Received 22 June, 2007

Dimitri Bormotov, Intrinsyc Software International, Inc., 700 W. Pender St., Vancouver, BC, V6C 1G8, Canada
E-mail: dbormotov@intrinsyc.com

Robert Gilman, Stevens Institute of Technology, Hoboken NJ 07030-5991, U.S.A.
E-mail: Robert.Gilman@stevens.edu

Alexei Myasnikov, McGill University, Montreal, QC, H3A 2K6, Canada
E-mail: amiasnikov@math.mcgill.ca