Automorphic orbits in free groups: words versus subgroups *

Pedro V. Silva

Centro de Matemática, Faculdade de Ciências, Universidade do Porto,
R. Campo Alegre 687, 4169-007 Porto, Portugal
e-mail: pvsilva@fc.up.pt

and Pascal Weil

LaBRI, CNRS, 351 cours de la Libération, 33405 Talence Cedex, France
and *Department CSE, IIT Delhi, Hauz Khas, New Delhi 110016, India*
e-mail: pascal.weil@labri.fr

2000 Mathematics Subject Classification: 20E05

ABSTRACT

We show that the following problems are decidable in rank 2 free groups: does a given finitely generated subgroup H contain primitive elements? and does H meet the automorphic orbit of a given word u? In higher rank, we show the decidability of the following weaker problem: given a finitely generated subgroup H, a word u and an integer k, does H contain the image of u by some k-almost bounded automorphism? An automorphism is k-almost bounded if at most one of the letters has an image of length greater than k.

1 Introduction

Orbit problems in general concern the orbit of an element u or a subgroup H of a group F, under the action of a subgroup G of $\mathrm{Aut}\, F$. Conjugacy problems are a special instance of such problems, where G consists of the inner automorphisms of F. In this paper, we restrict our attention to the case where F is the free group F_A with finite basis A.

In this context, orbit problems were maybe first considered by Whitehead [21], who proved that membership in the orbit of u under the action of $\mathrm{Aut}\, F_A$ is decidable. The analogous result regarding the orbit of a finitely generated subgroup H was established by Gersten [6]. Much literature has been devoted as well to the case where $G = \langle\varphi\rangle$ is a cyclic subgroup of $\mathrm{Aut}\, F_A$, e.g. Myasnikov and Shpilrain's work [12] on finite orbits of the form $\langle\varphi\rangle \cdot u$ and Brinkmann's recent proof [2] of the decidability of membership in $\langle\varphi\rangle \cdot u$.

*The first author acknowledges support from Project ASA (PTDC/MAT/65481/2006) and C.M.U.P., financed by F.C.T. (Portugal) through the programmes POCTI and POSI, with national and European Community structural funds. Both authors acknowledge support from the ESF project AutoMathA.

The orbit problem considered in this paper is the following: given an element $u \in F_A$ and a finitely generated subgroup H of F_A, does H meet the orbit of u under $\mathrm{Aut}\, F_A$, that is, does H contain $\varphi(u)$ for some automorphism $\varphi \in \mathrm{Aut}\, F_A$? A particular instance of this problem is the question whether H contains a primitive element, since the set of primitive elements of F_A is the automorphic orbit of each letter $a \in A$.

Our main result states that both these problems are decidable in the 2-generated free group F_2. In free groups with larger rank, we are only able to decide a weaker problem. Say that an automorphism φ of F_A is k-almost bounded if $|\varphi(a)| > k$ for at most one letter $a \in A$. We show that given $k > 0$, $u \in F_A$ and H a finitely generated subgroup of F_A, one can decide whether there exists a k-almost bounded automorphism μ such that $\mu(u) \in H$.

In the rank 2 case, we use a particular factorization of the automorphism group $\mathrm{Aut}\, F_2$ (Theorem 3.8) and a detailed combinatorial analysis of the effect of certain simple automorphisms on the graphical representation of the subgroup H (the representation by means of so-called Stallings foldings [19, 8], see Section 2.2).

The proof of the result on almost bounded automorphisms in arbitrary ranks relies ultimately on Diekert et al.'s result that the existential theory of equations with rational constraints in free groups is decidable [5]. Interesting intermediary results state that the set of primitive elements in F_2 is a context-sensitive language (Proposition 3.5) and that if $|A| = m$ and $v_1, \ldots, v_{m-1} \in F_A$, then the set of elements x such that $v_1, \ldots, v_{m-1}, x$ form a basis of F_A is a constructible rational set (Proposition 5.3).

2 Preliminaries

2.1 Free groups

Let A denote a finite alphabet and let A^{-1} denote a set of formal inverses of A. The *free group on* A is the quotient

$$F_A = (A \cup A^{-1})^*/\eta,$$

where $(A \cup A^{-1})^*$ is the free monoid over $A \cup A^{-1}$ and η denotes the congruence on $(A \cup A^{-1})^*$ generated by the relation

$$\{(aa^{-1}, 1) \mid a \in A \cup A^{-1}\}.$$

We denote the canonical projection $(A \cup A^{-1})^* \to F_A$ by π. (The ${}^{-1}$ notation is extended to $(A \cup A^{-1})^*$ as usual.)

Let

$$R_A = (A \cup A^{-1})^* \setminus \Big(\bigcup_{a \in A \cup A^{-1}} (A \cup A^{-1})^* aa^{-1} (A \cup A^{-1})^* \Big)$$

denote the set of all *reduced* words in $(A \cup A^{-1})^*$ and let $\iota : (A \cup A^{-1})^* \to R_A$ denote the *reduction map*. Since $\eta = \mathrm{Ker}\,\iota$, we abuse notation and denote also by ι the induced bijection $F_A \to R_A$. The *length* of $g \in F_A$ is defined by $|g| = |\iota(g)|$. To simplify notation, we shall usually write $\overline{u} = \iota(u)$.

Let $u \in R_A$. We say that u is *cyclically reduced* if $uu \in R_A$. For every word $u \in R_A$, there exist unique words $v, w \in R_A$ such that $u = vwv^{-1}$ and w is cyclically reduced. We say that w is the *cyclically reduced core* of u.

Given $X \subseteq F_A$, we denote by $\langle X \rangle$ the subgroup of F_A generated by X.

Let $\mathrm{Aut}\, F_A$ denote the group of all automorphisms of F_A. If $\varphi \in \mathrm{Aut}\, F_A$ and no confusion arises, we shall denote also by φ the corresponding bijection of R_A.

Given $B \subseteq F_A$, we say that B is a *basis* of F_A if the homomorphism $F_B \to F_A$ induced by the inclusion map $B \to F_A$ is an isomorphism. Equivalently, B is a basis of F_A if and only if $B = \varphi(A)$ for some $\varphi \in \mathrm{Aut}\, F_A$.

2.2 Automata

An A-*language* is a subset of A^*. Following the standard language theory convention, we usually omit brackets in the representation of singleton sets.

We say that $\mathcal{A} = (Q, q_0, T, E)$ is a (finite) A-automaton if:

- Q is a (finite) set;
- $q_0 \in Q$ and $T \subseteq Q$;
- $E \subseteq Q \times A \times Q$.

A *nontrivial path* in $\mathcal{A}$ is a sequence

$$p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} \ldots \xrightarrow{a_n} p_n$$

with $(p_{i-1}, a_i, p_i) \in E$ for $i = 1, \ldots, n$. Its *label* is the word $a_1 \ldots a_n \in A^*$. It is said to be a *successful* path if $p_0 = q_0$ and $p_n \in T$. We consider also the *trivial path* $p \xrightarrow{1} p$ for each $p \in Q$. It is successful if $p = q_0 \in T$. The *language* $L(\mathcal{A})$ *recognized by* $\mathcal{A}$ is the set of all labels of successful paths in $\mathcal{A}$.

The automaton $\mathcal{A} = (Q, q_0, T, E)$ is said to be *deterministic* if, for all $p \in Q$ and $a \in A$, there is at most one edge of the form (p, a, q). We write then $q = p \cdot a$. We say that $\mathcal{A}$ is *trim* if every $q \in Q$ lies in some successful path.

The *star* operator on A-languages is defined by

$$L^* = \bigcup_{n \geq 0} L^n,$$

where $L^0 = \{1\}$. An A-language L is said to be *rational* if L can be obtained from finite A-languages using finitely many times the operators union, product and star. Alternatively, L is rational if and only if it is recognized by a finite (deterministic) A-automaton $\mathcal{A} = (Q, q_0, T, E)$. The set of all rational A-languages is denoted by Rat A.

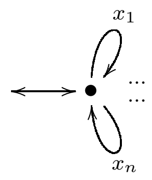
In the context of a particular result or claim, we say that a rational language L is *effectively constructible* if there exists an algorithm to produce a finite automaton recognizing L from the concrete structures containing the input.

If $\mathcal{A} = (Q, q_0, T, E)$ is an $(A \cup A^{-1})$-automaton, the *dual* of an edge (p, a, q) is (q, a^{-1}, p). Then $\mathcal{A}$ is said to be *dual* if E contains the duals of all edges. It is said to be *inverse* if it is dual, deterministic, trim and $|T| = 1$.

Given a finitely generated subgroup H of F_A, we denote by $\mathcal{A}(H)$ the finite automaton associated to H by the construction often referred to as *Stallings foldings*. This construction, that can be traced back to the early part of the twentieth century [16, Chap. 11], was made explicit by Serre [17] and Stallings [19] (see also [8]).

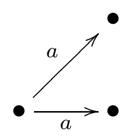
We can describe it briefly as follows.

1. We take a finite generating set $X = \{x_1, \ldots, x_n\}$ for H in reduced form.
2. We build the *flower automaton*



where the petals are paths labelled by the generators and their dual edges.

3. We successively identify ("fold") pairs of edges of the form



($a \in A \cup A^{-1}$) until no further folding applies.

The following proposition summarizes some of the relevant properties of $\mathcal{A}(H)$ (see [8]):

Proposition 2.1 *Let $H \leq_{f.g.} F_A$. Then:*

(i) *$\mathcal{A}(H)$ is a finite inverse automaton;*

(ii) *if $p \xrightarrow{u} q$ is a path in $\mathcal{A}(H)$, so is $p \xrightarrow{\overline{u}} q$;*

(iii) *$\mathcal{A}(H)$ does not depend on the finite reduced generating set chosen;*

(iv) *for every $u \in R_A$, $u \in L(\mathcal{A}(H))$ if and only if $\pi(u) \in H$;*

(v) *$L(\mathcal{A}(H)) \subseteq \pi^{-1}(H)$;*

(vi) *for every cyclically reduced $u \in F_A$, $wuw^{-1} \in H$ for some $w \in F_A$ if and only if u labels some loop in $\mathcal{A}(H)$.*

We conclude with the fundamental Benois Theorem [1]:

Theorem 2.2 *If $L \in \mathrm{Rat}\,(\mathrm{A} \cup \mathrm{A}^{-1})$, then $\overline{L} \in \mathrm{Rat}\,(\mathrm{A} \cup \mathrm{A}^{-1})$ and is effectively constructible.*

2.3 Automorphisms of F_2

In most of the paper, we shall be discussing the free group on 2 generators. We shall fix the alphabet $A_2 = \{a, b\}$ and use the notation $F_2 = F_A$, $R_2 = R_A$.

Given a basis $\{u, v\}$ of F_2, we denote by $\varphi_{u,v}$ the automorphism defined by $\varphi_{u,v}(a) = u$ and $\varphi_{u,v}(b) = v$. For every $w \in F_2$, let $\lambda_w = \varphi_{waw^{-1}, wbw^{-1}}$ be the inner automorphism defined by w.

We introduce the notation

$$\Sigma = \{\varphi_{a,ba}, \varphi_{b^{-1},a^{-1}}\};$$

$\Phi = \{\varphi_{a,ba}, \varphi_{ab,b}, \varphi_{a,ab}, \varphi_{ba,b}\}$;

$\Psi = \{\varphi \in \mathrm{Aut}\, F_2 : |\varphi(a)| = |\varphi(b)| = 1\}$;

$\Lambda = \{\lambda_w;\ w \in R_A\}$;

$\Delta = \{\varphi_{a,a^m b^\varepsilon a^n};\ m, n \in \mathbb{Z},\ \varepsilon \in \{1, -1\}\}$;

It is immediate that these sets consist of automorphisms of F_2. The following lemma summarizes some of their properties. The proof (by straightforward verification) is omitted.

Lemma 2.3 *Let $w \in R_2$, $\theta \in Aut F_2$, $m, n \in \mathbb{Z}$ and $\varepsilon \in \{1, -1\}$. Then*

(i) $\theta\lambda_w = \lambda_{\theta(w)}\theta$;

(ii) $\varphi_{a,ab} = \lambda_a \varphi_{a,ba}$;

(iii) $\varphi_{ab,b} = \varphi_{b,a}\varphi_{a,ba}\varphi_{b,a}$;

(iv) $\varphi_{ba,b} = \lambda_b \varphi_{ab,b} = \lambda_b \varphi_{b,a}\varphi_{a,ba}\varphi_{b,a}$;

(v) $\varphi^{-1}_{a,ba} = \varphi_{a^{-1},b}\varphi_{a,ba}\varphi_{a^{-1},b}$;

(vi) $\varphi_{b,a} = \varphi_{a^{-1},b}\varphi_{b^{-1},a^{-1}}\varphi_{a^{-1},b}$;

(vii) $\varphi_{a,a^m b^\varepsilon a^n} = \lambda_{a^m}\varphi_{a,b^\varepsilon a^{m+n}}$;

(viii) $\varphi_{a,b^{-1}a^n} = \lambda_{a^{-n}}\varphi_{a,ba^{-n}}\varphi_{a,b^{-1}}$;

(ix) $\varphi_{a,ba^n} = \varphi^n_{a,ba}$.

From now on, we apply the language formalism and conventions to automorphisms.

Proposition 2.4 *(i) $X\Lambda = \Lambda X$ for every $X \subseteq Aut F_2$;*

(ii) $\Lambda\Psi\Phi^* \subseteq \Lambda\Psi(\Sigma^{-1})^*\varphi_{a^{-1},b}$;

(iii) $\Delta \subseteq \Lambda(\varphi^*_{a,ba} \cup \varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b})(1 \cup \varphi_{a,b^{-1}})$.

Proof. (i) By Lemma 2.3(i).

(ii) By Lemma 2.3(i)-(vi), we get

$$\begin{aligned}\Lambda\Psi\Phi^* &\subseteq \Lambda\Psi(\varphi_{a,ba} \cup \varphi_{b,a})^* = \Lambda\Psi(\varphi^{-1}_{a^{-1},b}\varphi^{-1}_{a,ba}\varphi^{-1}_{a^{-1},b} \cup \varphi_{a^{-1},b}\varphi_{b^{-1},a^{-1}}\varphi_{a^{-1},b})^* \\ &= \Lambda\Psi(\varphi_{a^{-1},b}\varphi^{-1}_{a,ba}\varphi_{a^{-1},b} \cup \varphi_{a^{-1},b}\varphi^{-1}_{b^{-1},a^{-1}}\varphi_{a^{-1},b})^* = \Lambda\Psi\varphi_{a^{-1},b}(\Sigma^{-1})^*\varphi_{a^{-1},b} \\ &= \Lambda\Psi(\Sigma^{-1})^*\varphi_{a^{-1},b}.\end{aligned}$$

(iii) By Lemma 2.3(v)-(ix), we have

$$\varphi_{a,a^m ba^n} = \lambda_{a^m}\varphi_{a,ba^{m+n}} \in \Lambda(\varphi^*_{a,ba} \cup (\varphi^{-1}_{a,ba})^*) = \Lambda(\varphi^*_{a,ba} \cup \varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b}),$$

$$\begin{aligned}\varphi_{a,a^m b^{-1}a^n} &= \lambda_{a^m}\varphi_{a,b^{-1}a^{m+n}} = \lambda_{a^{-n}}\varphi_{a,ba^{-(m+n)}}\varphi_{a,b^{-1}} \\ &\in \Lambda(\varphi^*_{a,ba} \cup \varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b})\varphi_{a,b^{-1}}.\end{aligned}$$

□

3 Primitive words

Let us first consider a particular automorphic orbit in F_A, namely the set P_A of primitive words. Recall that a word is *primitive* if it belongs to some basis of F_A. In particular, P_A is the automorphic orbit of each letter from A. We shall often view P_A as a subset of R_A. We denote by P_2 the set of all primitive words in F_2. We establish certain language-theoretic properties of P_2 and we use combinatorial properties of the words in P_2 to derive a technical factorization of the group $\mathrm{Aut}\, F_2$ of automorphisms of F_2, that will be used in Section 4.

Let us first recall three known results from the literature. The first is due to Nielsen [13] (see also [4, 2.2] and [14]) and the second is due to Wen and Wen [20]. An interesting perspective on either is offered in [9, Chapter 2] and [3, Chapter I-5].

Proposition 3.1 *(i) Up to conjugation, every primitive word $u \in P_2$ is either a letter, or of the form $u = a^{n_1}b^{m_1}...a^{n_k}b^{m_k}$ where*

- *either $n_1 = ... = n_k \in \{1, -1\}$ and $\{m_1, ..., m_k\} \subseteq \{n, n+1\}$ for some integer n,*
- *or $m_1 = ... = m_k \in \{1, -1\}$ and $\{n_1, ..., n_k\} \subseteq \{n, n+1\}$ for some integer n.*

(ii) The set of positive primitive words $P_2 \cap \{a, b\}^+$ is equal to $\Phi^*(\{a, b\}) = b \cup \Phi^*(a)$.

Corollary 3.2 *$P_2 = \Lambda\Psi\Phi^*(a)$.*

Proof. By Proposition 3.1(i), a primitive word contains at most two letters from $A \cup A^{-1}$. Moreover, Proposition 3.1 implies that the set of all cyclically reduced primitive words is precisely

$$\Psi(P_2 \cap \{a, b\}^+) = \Psi(b) \cup \Psi\Phi^*(a).$$

Since $\Psi(a) = A \cup A^{-1} = \Psi(b)$, we conjugate to get

$$P_2 = \Lambda\Psi\Phi^*(a).$$

□

3.1 The language P_2

Recall that a *context-sensitive A-grammar* is a triple $\mathcal{G} = (V, P, S)$ such that

- V is a finite set containing $A \,\dot\cup\, \{S\}$;
- P is the *set of rules* of the grammar, a finite subset of $(V^+ \setminus A^+) \times V^+$ satisfying

$$(x, y) \in P \Rightarrow |x| \leq |y|.$$

For all $x, y \in V^+$, we write $x \Rightarrow y$ if there exist $r, s \in V^*$ and $(p, q) \in P$ such that $x = rps$ and $y = rqs$. We denote by $\overset{*}{\Rightarrow}$ the transitive and reflexive closure of $\Rightarrow$. The language generated by $\mathcal{G}$ is

$$L(\mathcal{G}) = \{w \in A^+ \mid S \overset{*}{\Rightarrow} w\}.$$

A language $L \subseteq A^+$ is said to be *context-sensitive* if it is generated by some context-sensitive A-grammar. As usual, a language $L \subseteq A^*$ is called *context-sensitive* if $L \cap A^+$ is context-sensitive.

Lemma 3.3 *The class of context-sensitive languages is closed under union, intersection, right and left quotient by a word, ε-free substitutions, inverse morphisms and non-erasing morphisms (that is, homomorphisms in which every letter is mapped to a non-empty word).*

Proof. Closure under union, intersection, ε-free substitutions, inverse homomorphisms and non-erasing morphisms is well-known [7, Exercise 9.10]. In particular, the family of context-sensitive languages forms a *trio* [7, Section 11.1] and as such, it is closed under *limited erasing* [7, Lemma 11.2]. By definition, this means that if $k \geq 1$, L is context-sensitive and φ is a morphism such that $\varphi(v) \neq 1$ for each $u \in L$ and each factor v of u of length greater than k, then $\varphi(L)$ is context-sensitive as well.

Now let $L \subseteq A^*$, $a \in A$ and $\$ \notin A$. Let σ be the substitution that maps a to $\sigma(a) = \{a, \$\}$ and which fixes every other letter of A. Let also $\varphi\colon (A \cup \{\$\})^* \to A^*$ be the morphism which fixes every letter of A and erases \$. Then $a^{-1}L = \varphi(\sigma(L) \cap \$A^*)$ and $La^{-1} = \varphi(\sigma(L) \cap A^*\$)$. Since the σ-images of the letters are finite, and hence context-sensitive, the languages $\sigma(L) \cap \$A^*$ and $\sigma(L) \cap A^*\$$ are context-sensitive; moreover φ exhibits limited erasing on these languages, so $a^{-1}L$ and La^{-1} are context-sensitive as well. $\Box$

Lemma 3.4 *Let A be a finite alphabet and let Γ be a finite set of endomorphisms of A^+. For every $u \in A^+$, $\Gamma^*(u)$ is a context-sensitive language.*

Proof. Take $b \notin A$. We define a context-sensitive $(A \cup \{b\})$-grammar $\mathcal{G} = (V, P, S)$ by $V = A \cup \{R, S, T\} \cup \{F_\varphi \mid \varphi \in \Gamma\}$ and

$$P = \{\, S \to bF_\varphi uR,\ S \to bub^2,\ F_\varphi a \to \varphi(a)F_\varphi,\ F_\varphi R \to TR,\\ F_\varphi R \to b^2,\ aT \to Ta,\ bT \to bF_\varphi;\ a \in A,\ \varphi \in \Gamma\}.$$

We show that $L(\mathcal{G}) = b\Gamma^*(u)b^2$.

Clearly, $F_\varphi v \overset{*}{\Rightarrow} \varphi(v)F_\varphi$ for all $\varphi \in \Gamma$ and $v \in A^*$ and so

$$bvTR \overset{*}{\Rightarrow} bTvR \Rightarrow bF_\varphi vR \overset{*}{\Rightarrow} b\varphi(v)F_\varphi R \Rightarrow b\varphi(v)TR.$$

Since $S \Rightarrow bF_\varphi uR \overset{*}{\Rightarrow} b\varphi(u)F_\varphi R \Rightarrow b\varphi(u)TR$ for every $\varphi \in \Gamma$, it follows that $S \overset{*}{\Rightarrow} b\theta(u)F_\varphi R \Rightarrow b\theta(u)b^2$ for every $\theta \in \Gamma^+$. Together with $S \Rightarrow bub^2$, this yields $b\Gamma^*(u)b^2 \subseteq L(\mathcal{G})$.

To prove the opposite inclusion, let

$$Z = \{S\} \cup \{bxyb^2,\ bxTyR,\ b\varphi(x)F_\varphi yR;\ xy \in \Gamma^*(u)\}.$$

Then

$$(X \in Z \ \wedge\ X \Rightarrow Y) \Rightarrow Y \in Z.$$

Since $S \in Z$, it follows that $L(\mathcal{G}) \subseteq Z \cap A^* = b\Gamma^*(u)b^2$ and so $L(\mathcal{G}) = b\Gamma^*(u)b^2$. Thus $b\Gamma^*(u)b^2$ is context-sensitive and by Lemma 3.3, $\Gamma^*(u) = b^{-1}(b\Gamma^*(u)b^2)(b^2)^{-1}$ is context-sensitive as well. $\Box$

Theorem 3.5 *P_2 is a context-sensitive language.*

Proof. Since the class of context-sensitive languages is closed under union (Lemma 3.3), it follows from Proposition 3.1(ii) and Lemma 3.4 that $P_2 \cap \{a,b\}^+$ is context-sensitive. Let $\mathcal{G} = (V, P, S)$ be a context-sensitive A-grammar generating $P_2 \cap \{a,b\}^+$. We build a context-sensitive A-grammar $\mathcal{G}' = (V', P', S')$ by letting $V' = \{S'\} \cup V$ and

$$P' = P \cup \{S' \to S\} \cup \{S' \to cS'c^{-1};\ c \in A_2 \cup A_2^{-1}\}.$$

It is immediate that $L(\mathcal{G}') \cap R_2$ is the set of all reduced words having their cyclically reduced core in $P_2 \cap \{a,b\}^+$. It follows from Proposition 3.1(i) or Corollary 3.2 that $P_2 = P_2 \cap R_2 = \Psi(L(\mathcal{G}') \cap R_2)$, and in view of the closure properties in Lemma 3.3, P_2 is context-sensitive. □

This result cannot be improved to the next level of Chomsky's hierarchy:

Proposition 3.6 *P_2 is not a context-free language.*

Proof. We show that $P_2 \cap ab^+ab^+ab^+$ is not a context-free language. Since the class of context-free languages is closed under intersection with rational languages, it shows that P_2 is not context-free either.

It follows easily from Proposition 3.1(i) that

$$P_2 \cap ab^*ab^*ab^* = \{ab^m ab^n ab^k;\ m,n,k \in \mathbb{N},\ \max\{m,n,k\} = \min\{m,n,k\} + 1\}. \quad (1)$$

It is now a classical exercise to show that $P_2 \cap ab^+ab^+ab^+$ is not context-free since it fails the Pumping Lemma for context-free languages [7, Section 6.1]. □

3.2 A factorization of Aut F_2

The following result constitutes a simple application of Proposition 2.1:

Lemma 3.7 *Let $A = \{a_1, \ldots, a_m\}$ and $u \in R_A$. Then $\{a_1, \ldots, a_{m-1}, u\}$ is a basis of F_A if and only if $u = va_m^{\varepsilon}w$ for some $v, w \in R_{\{a_1,\ldots,a_{m-1}\}}$ and $\varepsilon \in \{1,-1\}$.*

Proof. It is immediate that if $u = va_m^{\epsilon}w$ with $v, w \in R_{\{a_1,\ldots,a_{m-1}\}}$, then $\{a_1, ..., a_{m-1}, u\}$ generates F_A, and by the Hopfian property of free groups (see [10]), $\{a_1, ..., a_{m-1}, u\}$ is a basis of F_A.

Conversely, let us assume that $u \in R_A$ contains several occurrences of a_m or a_m^{-1}, and let $u = vzw$ with $v, w \in R_{a_1,\ldots,a_{m-1}}$ of maximal length. It is immediate that if $H = \langle a_1, ..., a_{m-1}, u\rangle$, then $H = \langle a_1, ..., a_{m-1}, z\rangle$ and $\mathcal{A}(H)$ is equal to $\mathcal{A}(\langle z\rangle)$ with loops labelled $a_1, ..., a_{m-1}$ attached at the origin. Thus, if $\{a_1, ..., a_{m-1}, u\}$ is a basis of F_A, then $\mathcal{A}(\langle z\rangle)$ must consist of a single loop labeled a_m, and hence z must be equal to a_m or a_m^{-1}. □

Theorem 3.8 *$Aut\, F_2 = \Lambda\Psi\Phi^*\Delta = \Psi(\Sigma^{-1})^*\Lambda\varphi_{a,ba}^*(\varphi_{a^{-1},b} \cup \varphi_{a^{-1},b^{-1}})$.*

Proof. We start by establishing the first equality. Let $\theta \in \mathrm{Aut}\, F_2$. Then $\theta(a) \in P_2$ and so $\theta(a) = \sigma(a)$ for some $\sigma \in \Lambda\Psi\Phi^*$ by Corollary 3.2. Now $\sigma^{-1}\theta \in \mathrm{Aut}\, F_2$, hence $\{a = \sigma^{-1}\theta(a), \sigma^{-1}\theta(b)\}$ is a basis of F_2. By Lemma 3.7, it follows that $\sigma^{-1}\theta(b) = a^m b^{\varepsilon} a^n$ for some $m, n \in \mathbb{Z}$ and $\varepsilon \in \{1,-1\}$. Thus $\theta(b) = \sigma(a^m b^{\varepsilon} a^n)$ and we get

$$\theta = \sigma\varphi_{a,a^m b^{\varepsilon} a^n} \in \Lambda\Psi\Phi^*\Delta.$$

The opposite inclusion is trivial since $\mathrm{Aut}\, F_2$ is closed under composition. Therefore $\mathrm{Aut}\, F_2 = \Lambda\Psi\Phi^*\Delta$.

Now Proposition 2.4 yields

$$\begin{aligned}\mathrm{Aut}\, F_2 &= \Lambda\Psi\Phi^*\Delta \\ &\subseteq \Lambda\Psi\Phi^*(\varphi^*_{a,ba} \cup \varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b})(1 \cup \varphi_{a,b^{-1}}) \\ &= \Lambda\Psi\Phi^*(1 \cup \varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b})(1 \cup \varphi_{a,b^{-1}}) \\ &= \Lambda\Psi\Phi^*\varphi_{a^{-1},b}\varphi^*_{a,ba}\varphi_{a^{-1},b}(1 \cup \varphi_{a,b^{-1}}) \\ &\subseteq \Lambda\Psi(\Sigma^{-1})^*\varphi^*_{a,ba}\varphi_{a^{-1},b}(1 \cup \varphi_{a,b^{-1}}) \\ &= \Psi(\Sigma^{-1})^*\Lambda\varphi^*_{a,ba}(\varphi_{a^{-1},b} \cup \varphi_{a^{-1},b^{-1}}).\end{aligned}$$

The converse inclusion is of course trivial. □

4 The orbit problem in F_2

The aim of this section is to prove

Theorem 4.1 *Given $u \in F_2$ and $H \leq_{f.g.} F_2$, it is decidable whether or not $\mu(u) \in H$ for some $\mu \in Aut F_2$.*

In view of Theorem 3.8, we will pay detailed attention to the action of the automorphisms of Σ, namely $\varphi_{b^{-1},a^{-1}}$ and $\varphi_{a,ba}$, on the automata of the form $\mathcal{A}(H)$.

Let us also note the interesting corollary below.

Corollary 4.2 *Given $H \leq_{f.g.} F_2$, it is decidable whether H contains a primitive element of F_2.*

4.1 Singularities, bridges and automorphisms in Σ

Given $H \leq_{f.g.} F_2$, we say that a state q of $\mathcal{A}(H)$ is a

- *source* if $q \cdot a, q \cdot b \neq \emptyset$, $\qquad \xleftarrow{a} q \xrightarrow{b}$
- *sink* if $q \cdot a^{-1}, q \cdot b^{-1} \neq \emptyset$. $\qquad \xrightarrow{a} q \xleftarrow{b}$

We use the general term *singularities* to refer to both sources and sinks.

We denote by $\mathrm{Sing}(H)$ the set of all singularities of $\mathcal{A}(H)$ plus the origin. If we emphasize the vertices of $\mathrm{Sing}(H)$ in $\mathcal{A}(H)$, it is immediate that $\mathcal{A}(H)$ can be described as the union of *positive paths*, i.e. paths with label in $(a \cup b)^+$, between the vertices of $\mathrm{Sing}(H)$, and these positive paths do not intersect each other except at $\mathrm{Sing}(H)$. We call such paths *bridges*. Note that every positive path whose internal states are not singularities can be extended into a uniquely determined bridge.

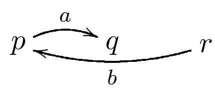
We discuss now the evolution of the Stallings automata under the influence of Σ. The next result follows immediately.

Fact 4.3 *The automaton $\mathcal{A}(\varphi_{b^{-1},a^{-1}}(H))$ has the same vertex set as $\mathcal{A}(H)$, edges are reverted and labels changed. In particular, sources and sinks are exchanged. If β is a bridge in $\mathcal{A}(H)$, $\beta = p \xrightarrow{w} q$, then there is a bridge $q \longrightarrow p$ labeled $\varphi_{b^{-1},a^{-1}}(w^{-1})$, of equal length, which we denote by $\varphi_{b^{-1},a^{-1}}(\beta)$.*

Fact 4.4 *The automaton $\mathcal{A}(\varphi_{a,ba}(H))$ is obtained from $\mathcal{A}(H)$ by the following 3 steps:*

(S1) If $p\xrightarrow{b}q$ is an edge of $\mathcal{A}(H)$ and q is not a sink, we replace that edge by a path $p\xrightarrow{b}\bullet\xrightarrow{a}q$, adding a new intermediate vertex for each such edge.

(S2) If $p\xrightarrow{a}q\xleftarrow{b}r$ is a sink in $\mathcal{A}(H)$, we replace this configuration by



(S3) We successively remove all the vertices of degree 1 different from the origin.

Proof. Following [15, Subsection 1.2], the automaton $\mathcal{A}(\varphi(H))$ may be obtained from $\mathcal{A}(H)$ in three steps:

(1) We replace each edge labelled by b by a path labelled ba (introducing a new intermediate vertex for each such edge), producing a dual automaton $\mathcal{B}$.

(2) We execute the complete folding of $\mathcal{B}$.

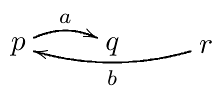
(3) We successively remove all the vertices of degree 1 different from the origin.

How much folding is involved in the process? Let us consider the first level of folding, i.e. those pairs of edges that can be immediately identified in $\mathcal{B}$.

- There are no b-edges involved in the first level of folding: indeed, the b-edges keep their origin when we go from $\mathcal{A}(H)$ to $\mathcal{B}$, and their target is always a new vertex where folding cannot take place.
- If we have a sink $p\xrightarrow{a}q\xleftarrow{b}r$ in $\mathcal{A}(H)$, we get

  $$p\xrightarrow{a}q\xleftarrow{a}\bullet\xleftarrow{b}r$$

  in $\mathcal{B}$ and therefore an instance of first level folding, yielding



- These are the only instances of first level folding: we cannot fold two "new" a-edges $\xrightarrow{a}q\xleftarrow{a}$ in $\mathcal{B}$ since that would imply the existence of two b-edges $\xrightarrow{b}q\xleftarrow{b}$ in $\mathcal{A}(H)$.

Let $\mathcal{C}$ denote the automaton obtained by performing all the instances of first level folding in $\mathcal{B}$. It follows from the above remarks that $\mathcal{C}$ can be obtained from $\mathcal{A}(H)$ by application of (S1) and (S2).

We actually need no second level of folding because $\mathcal{C}$ is already deterministic. Indeed, it is clear from (S1) and (S2) that configurations such as $\xleftarrow{a}q\xrightarrow{a}$ or $\xleftarrow{b}q\xrightarrow{b}$ cannot occur in $\mathcal{C}$.

Suppose that $\xrightarrow{b}q\xleftarrow{b}$ does occur. Then both edges must have been obtained through (S2) which is impossible since $p\cdot a$ is uniquely determined in $\mathcal{A}(H)$.

Finally, suppose that $\xrightarrow{a}q\xleftarrow{a}$ does occur. At least one of these edges must have been obtained through (S1), but not both, otherwise we would have a configuration $\xrightarrow{b}q\xleftarrow{b}$ in $\mathcal{A}(H)$. But then we would have a configuration $\xrightarrow{a}q\xleftarrow{b}$ in $\mathcal{A}(H)$ and q would be a sink, contradicting the application of (S1). Thus $\mathcal{C}$ is deterministic and so $\mathcal{A}(\varphi(H))$ is obtained from $\mathcal{A}(H)$ by successive application of (S1), (S2) and (S3). □

Fact 4.5 *(i) When applying $\varphi_{a,ba}$, a state of $\mathcal{A}(H)$ is trimmed in step (S3) if and only if it is a sink of $\mathcal{A}(H)$ without outgoing edges. Moreover, no consecutive states can be trimmed.*

(ii) The sources of $\mathcal{A}(\varphi_{a,ba}(H))$ are precisely the sources p of $\mathcal{A}(H)$ such that $p \cdot a$ is not a sink or has outgoing edges in $\mathcal{A}(H)$.

(iii) The sinks of $\mathcal{A}(\varphi_{a,ba}(H))$ are precisely the states p of $\mathcal{A}(H)$ with incoming edges such that $p \cdot a$ is a sink of $\mathcal{A}(H)$.

Proof. (i) The origin cannot be trimmed and the number of outgoing edges never decreases, so the only possible candidates to (S3) are the states that are decreasing the number of incoming edges, which are precisely the sinks of $\mathcal{A}(H)$. Clearly, their fate will then depend on the previous existence of some outgoing edge. Note that $\mathcal{A}(H)$ cannot possess two consecutive sinks with no outgoing edges, hence the trimming of a vertex will not be followed by the trimming of any of its neighbours.

(ii) Since outgoing edges can be at most redirected through (S1) and (S2), it is clear that every source p of $\mathcal{A}(\varphi_{a,ba}(H))$ must be a source of $\mathcal{A}(H)$. Thus everything will depend on $p \cdot a$ being trimmed or not, and part (i) yields the claim.

(iii) Clearly, no new intermediate vertex obtained through (S1) can become a sink, and any sink of $\mathcal{A}(H)$ will not remain such after application of (S2). Thus the only remaining candidates are the non sinks of $\mathcal{A}(H)$ that are increasing the number of incoming edges, which are precisely those of the form $q \cdot a^{-1}$, where q is a sink of $\mathcal{A}(H)$. Clearly, to have two distinct incoming edges in $\mathcal{A}(\varphi_{a,ba}(H))$, $p = q \cdot a^{-1}$ must have at least one incoming edge in $\mathcal{A}(H)$. In such a case, it is easy to check that after (S1)/(S2), p has indeed become a sink of $\mathcal{A}(\varphi_{a,ba}(H))$. We remark also that the subsequent trimming by (S3) does not affect the presence of singularities. $\square$

Fact 4.6 *Let β be a bridge in $\mathcal{A}(H)$ of length at least 2, say $\beta = p \xrightarrow{w} q$, and let $w = w'cd$ where $c, d \in A$.*

(i) $\mathcal{A}(\varphi_{a,ba}(H))$ has a positive path $p \xrightarrow{\varphi(w'c)} s$, which extends to a uniquely determined bridge, denoted by $\varphi_{a,ba}(\beta)$.

(ii) $|\varphi_{a,ba}(\beta)| \geq |\beta| - 1$, and we have $|\varphi_{a,ba}(\beta)| = |\beta| - 1$ exactly if $w \in a^+$, p is a source or the origin in $\mathcal{A}(H)$, and q is a sink in $\mathcal{A}(H)$.

Proof. Write $\beta = p \xrightarrow{w'} r \xrightarrow{c} s \xrightarrow{d} q$.

(i) By Fact 4.5, no state of the path $p \xrightarrow{\varphi(w'c)} s$ risks trimming. Hence it suffices to check that no intermediate vertex of this path can become a singularity (let alone the origin). This follows easily from Fact 4.5(ii) and (iii).

(ii) The inequality $|\varphi_{a,ba}(\beta)| \geq |\beta| - 1$ follows at once from part (i). It follows also that $|\varphi_{a,ba}(\beta)| = |\beta| - 1$ if and only if $w'c \in a^+$ (otherwise $|\varphi_{a,ba}(\beta)| \geq |\varphi_{a,ba}(w'c)| > |w'c| = |\beta| - 1$) and $p, s \in \mathrm{Sing}(\varphi_{a,ba}(H))$. Thus we assume that $w'c \in a^+$.

Clearly, if p is the origin, it must remain so. If p is a source, it follows from Fact 4.5(ii) that p remains a source (since $p \cdot a$ is not a sink in $\mathcal{A}(H)$). Finally, if p is a sink, it will no longer be a singularity in $\mathcal{A}(\varphi_{a,ba}(H))$ by Fact 4.5(iii). Therefore $p \in \mathrm{Sing}(\varphi_{a,ba}(H))$ if and only if it is a source or the origin in $\mathcal{A}(H)$.

Clearly, q can never become the origin or a source. Since q has incoming edges in $\mathcal{A}(H)$, it follows from Fact 4.5(iii) that s becomes a sink in $\mathcal{A}(\varphi_{a,ba}(H))$ if and only if $s \cdot a$ is a sink in $\mathcal{A}(H)$. Since the unique outgoing edge of s in $\mathcal{A}(H)$ has label d, then $s \in \mathrm{Sing}(\varphi_{a,ba}(H))$ if and only if $d = a$ and q is a sink in $\mathcal{A}(H)$. □

Let $\sigma(H)$ be the *number of singularities* of $\mathcal{A}(H)$, i.e. the number of sources plus the number of sinks. Note that a vertex may be a source and a sink, and is then counted twice. Facts 4.3 and 4.5 yield:

Lemma 4.7 *Let $H \leq_{f.g.} F_2$ and $\varphi \in \Sigma$. Then $\sigma(\varphi(H)) \leq \sigma(H)$.*

We say that a path $p \xrightarrow{w} r$ is *homogeneous* if $w \in R_a \cup R_b$. Given $H \leq_{f.g.} F_2$, we define

$$\delta_0(H) = \max\{\, \sigma(H), \{|\kappa| \mid \kappa \text{ is a homogeneous cycle in } \mathcal{A}(H)\}\},$$
$$\delta(H) = \max\{\ \delta_0(H), \{|\kappa| \mid \kappa \text{ is a homogeneous cycle-free path in } \mathcal{A}(H)\}\}.$$

Lemma 4.8 *Let $H \leq_{f.g.} F_2$ and $\varphi \in \Sigma$. Then $\delta_0(\varphi(H)) \leq \delta_0(H)$.*

Proof. We may assume that $\varphi = \varphi_{a,ba}$. In view of Lemma 4.7, we only need to show that $\mathcal{A}(\varphi(H))$ has no homogeneous cycle of length greater than $\delta_0(H)$.

Assume that $q \circlearrowleft^{w}$ is a homogeneous cycle in $\mathcal{A}(\varphi(H))$. Assume first that $w = a^n$. Since no a-edge obtained through (S1) can be part of an a-cycle, it follows that the cycle existed already in $\mathcal{A}(H)$ and so $n \leq \delta_0(H)$.

Assume now that $w = b^n$. Once again, no b-edge obtained through (S1) can be part of a b-cycle, hence all edges in the cycle must have been obtained through (S2). But producing a b-edge through (S2) requires a sink, and any such sink produces a unique b-edge. Thus n cannot exceed the number of sinks in $\mathcal{A}(H)$ and so $n \leq \delta_0(H)$ as required. □

Given $H \leq_{f.g.} F_2$, we consider the *geodesic metric* d defined on the vertex set of $\mathcal{A}(H)$ by taking $d(u, v)$ to be the length of the shortest path connecting u and v. Since $\mathcal{A}(H)$ is inverse, it is irrelevant to consider directed or undirected paths. As usual, we have

$$d(u, \mathrm{Sing}(H)) = \min\{d(u, v);\ v \in \mathrm{Sing}(H)\}.$$

Given $t > 0$, we denote by $\mathcal{A}_t(H)$ the automaton obtained by removing from $\mathcal{A}(H)$ all vertices u such that $d(u, \mathrm{Sing}(H)) > t$ and their adjacent edges. We say that $\mathcal{A}(H)$ is the *t-truncation* of $\mathcal{A}(H)$.

By Fact 4.6, we know that, for every bridge β in $\mathcal{A}(H)$ of length at least 2, and $\varphi \in \Sigma$, we have $|\varphi(\beta)| \geq |\beta| - 1$. The next lemma will provide sufficient conditions to ensure $|\varphi(\beta)| \geq |\beta|$.

Lemma 4.9 *Let $\varphi \in \Sigma$, $H \leq_{f.g.} F_2$ and $K \in \Sigma^*(H)$. If β is a bridge in $\mathcal{A}(K)$ and $|\beta| > 2\delta(H)$, then $|\varphi(\beta)| \geq |\beta|$.*

Proof. We may assume that H is nontrivial, i.e., $\delta(H) > 0$.

The result is easily verified if $\varphi = \varphi_{b^{-1},a^{-1}}$ in view of Fact 4.3, since that automorphism preserves state set, singularities and distances. So we now assume that $\varphi = \varphi_{a,ba}$.

Since $\varphi_{b^{-1},a^{-1}}$ has order 2, we may assume that $\varphi_{b^{-1},a^{-1}}\varphi_{b^{-1},a^{-1}}$ is not a factor of μ as a word on Σ, i.e., we may replace Σ^* by

$$L = \Sigma^* \setminus (\Sigma^* \varphi_{b^{-1},a^{-1}} \varphi_{b^{-1},a^{-1}} \Sigma^*) \qquad (2)$$

at our convenience. Hence we may write $\mu = \varphi_{a,ba}^{j}\psi$ with $j \geq 0$ and $\psi \in \{1, \varphi_{b^{-1},a^{-1}}\} \cup \varphi_{b^{-1},a^{-1}}\varphi_{a,ba}L$.

We first observe that by Lemma 4.8,

$$\text{If } x \xrightarrow{a^n} x \text{ is a cycle in } \mathcal{A}(\psi(H)), \text{ then } n \leq \delta(H). \tag{3}$$

Next we show that

$$\text{If } x \xrightarrow{a^n} y \text{ is a cycle-free path in } \mathcal{A}(\psi(H)), \text{ then } n \leq \delta(H). \tag{4}$$

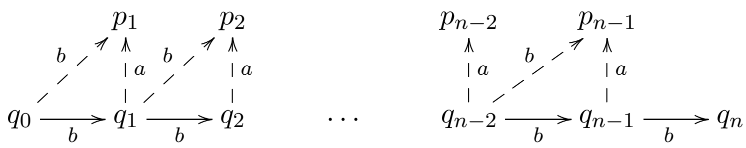
The result is trivial if ψ is trivial or equal to $\varphi_{b^{-1},a^{-1}}$ (since in that case $\delta(\psi(H)) = \delta(H)$). Let us now assume that $\psi \notin \{1, \varphi_{b^{-1},a^{-1}}\}$, so that $\psi = \varphi_{b^{-1},a^{-1}}\varphi_{a,ba}\chi$ for some $\chi \in L$.

Let $p \xrightarrow{a^n} q$ be a cycle-free path in $\mathcal{A}(\psi(H))$ with $n > \delta(H)$. Then $q \xrightarrow{b^n} p$ is a cycle-free path in $\mathcal{A}(\varphi_{a,ba}\chi(H))$. Since $\delta(H) > 0$, we have $n \geq 2$. Observe that the application of $\varphi_{a,ba}$ shatters to pieces any b-path existing in $\mathcal{A}(\chi(H))$, hence transformations of type (S2) must be involved in the genesis of $q \xrightarrow{b^n} p$.

Let

$$q = q_0 \xrightarrow{b} q_1 \xrightarrow{b} \dots \xrightarrow{b} q_n = p$$

be our path in $\mathcal{A}(\varphi_{a,ba}\chi(H))$. Since any b-edge obtained through (S1) must be followed only by an a-edge, only $q_{n-1} \xrightarrow{b} q_n$ can be obtained through (S1). Thus there exist edges in $\mathcal{A}(\chi(H))$ (represented through discontinuous lines) of the form



Clearly, the vertices $p_1, \dots, p_{n-1}$ are distinct sinks in $\mathcal{A}(\chi(H))$. If $q_{n-1} \xrightarrow{b} q_n$ is also obtained through (S2), we get an nth sink in $\mathcal{A}(\chi(H))$. If $q_{n-1} \xrightarrow{b} q_n$ is obtained through (S1), there is an edge $q_{n-1} \xrightarrow{b} z$ in $\mathcal{A}(\chi(H))$ and since $n \geq 2$, it follows that q_{n-1} must be a source in $\mathcal{A}(\chi(H))$. In any case, we obtain n singularities in $\mathcal{A}(\chi(H))$ and so

$$\delta_0(\chi(H)) \geq n > \delta(H) \geq \delta_0(H),$$

contradicting Lemma 4.8. Therefore (4) holds.

Let us finally consider a bridge β in $\mathcal{A}(K)$ such that $|\beta| > 2\delta(H)$. By Fact 4.6, if $|\varphi(\beta)| < |\beta|$, then $\beta = p \xrightarrow{a^m} q$, where $m > 2\delta(H)$, p is a source or the origin in $\mathcal{A}(K)$, and q is a sink of $\mathcal{A}(K)$. Since the action of $\varphi_{a,ba}$ does not increase the length of a-cycles and since all a-cycles in $\mathcal{A}(\psi(H))$ have length at most $\delta(H)$ (see (3)), we see that β is not part of an a-cycle. Now, the action of $\varphi_{a,ba}$ can only increase the length of a cycle-free a-path by one unit, so (4) shows that $j > 2\delta(H) - \delta(H) = \delta(H)$.

By Fact 4.6(ii), p is either the origin or a source in $\mathcal{A}(K)$. By Fact 4.5(ii), p is still the origin or a source in $\mathcal{A}(\psi(H))$. Moreover, successive application of Fact 4.5(iii) yields that $q \cdot a^j$ exists in $\mathcal{A}(\psi(H))$ and is a sink in that automaton (note that, since p is either the origin or a source, all the a-edges in the corresponding path of $\mathcal{A}(\varphi_{a,ba}^{i}\psi(H))$ must already exist in $\mathcal{A}(\varphi_{a,ba}^{i-1}\psi(H))$ for $i = 1, \dots, j$). In particular, a^j labels a cycle-free path in $\mathcal{A}(\psi(H))$ (if the path were not cycle-free, β would be part of an a-cycle, a contradiction). Since $j > \delta(H)$, this contradicts (4), and hence concludes the proof. □

Theorem 4.10 *Let $\varphi \in \Sigma$, $H \leq_{f.g.} F_2$, $t \geq \delta(H)$ and $K, K' \in \Sigma^*(H)$. Then*

$$\mathcal{A}_t(K) = \mathcal{A}_t(K') \quad \Rightarrow \quad \mathcal{A}_t(\varphi(K)) = \mathcal{A}_t(\varphi(K')).$$

Proof. As in the proof of Lemma 4.9, we may assume that $\delta(H) > 0$ and $\varphi = \varphi_{a,ba}$.

By Lemma 4.7, we know that the number of singularities does not increase by application of automorphisms from Σ. By Lemma 4.9, we also know that, once the length of a bridge reaches the threshold $2\delta(H) + 1$, it can only get longer. As it turns out from the definition, truncation affects only bridges of length at least $2\delta(H) + 1$. We must therefore discuss the truncation mechanism for such long bridges.

Assume that $\beta : p \xrightarrow{w} q$ is a bridge in $\mathcal{A}(\mu(H))$ ($\mu \in \Sigma^*$) with $|w| \geq 2t+1$. Then we may write $w = uzv$ with $|u| = |v| = t$. By Lemma 4.9, the label of $\varphi(\beta)$ is of the form $u'z'v'$ with $|u'| = |v'| = t$ and $|z'| \geq |z|$. We only need to prove that u' and v' depend only on $\mathcal{A}_t(\mu(H))$ and are therefore independent from z.

In view of Fact 4.5, it is clear that u' depends only on $\mathcal{A}_t(\mu(H))$ (remember that $w = uzv$ is a positive word and singularities cannot *move forward* along a positive path). The nontrivial case is of course the case of q being a sink in $\mathcal{A}(\mu(H))$, since by Fact 4.5(iii) a sink can actually be transferred to the preceding state along a positive path. We claim that even in this case v' is independent from z.

Indeed, assume first that b occurs in v. Then $|\varphi(v)| \; > \; |v|$ provides enough compensation for the sink moving backwards one position. Hence we may assume that $v = a^t$. We claim that $v' = a^t$ as well, independently from z. Suppose not. Since we are assuming that the sink has moved from q to its predecessor, and $\varphi(a^{t-1}) = a^{t-1}$, it follows that $v' = ba^{t-1}$. Hence b occurs in w. Write $w = xba^m$. Since $\varphi(ba^m) = ba^{m+1}$, and taking into account the mobile sink, we obtain by comparison $ba^m = ba^{t-1}$ and so $m = t-1$, a contradiction, since a^t is a suffix of w. Therefore $v' = a^t$ and so is independent from z as required. □

Corollary 4.11 *Let $H \leq_{f.g.} F_2$ and $t \geq \delta(H)$. Then the set*

$$\mathcal{X}(H) = \{\mathcal{A}_t(K) \mid K \in \Sigma^*(H)\}$$

is finite and effectively constructible.

Proof. By Lemma 4.7, $\mathcal{X}(H)$ is finite. The proof of Theorem 4.10 provides a straighforward algorithm to compute all its elements. Indeed, all we need is to compute the finite sets

$$\mathcal{X}_n(H) = \{\mathcal{A}_t(K) \mid K \in \Sigma^n(H)\}$$

until reaching

$$\mathcal{X}_{n+1}(H) \subseteq \bigcup_{i=0}^{n} \mathcal{X}_i(H), \tag{5}$$

which must occur eventually since $\mathcal{X}(H) = \cup_{i \geq 0} \mathcal{X}_i(H)$ is finite. Why does (5) imply $\mathcal{X}(H) = \cup_{i \geq 0}^{n} \mathcal{X}_i(H)$? Suppose that $\mathcal{B} \in \mathcal{X}_m(H) \setminus (\cup_{i \geq 0}^{n} \mathcal{X}_i(H))$ for $m > n$ minimal, say $\mathcal{B} = \mathcal{A}_t(\varphi(K))$ with $K \in \Sigma^{m-1}(H)$ and $\varphi \in \Sigma$. By minimality of m, we have $\mathcal{A}_t(K) \in \cup_{i \geq 0}^{n} \mathcal{X}_i(H)$. Thus $\mathcal{A}_t(K) = \mathcal{A}_t(K')$ for some $K' \in \cup_{i=0}^{n} \Sigma^i(H)$. Now Theorem 4.10 yields

$$\mathcal{B} = \mathcal{A}_t(\varphi(K)) = \mathcal{A}_t(\varphi(K')) \in \bigcup_{i=0}^{n+1} \mathcal{X}_i(H) = \bigcup_{i=0}^{n} \mathcal{X}_i(H),$$

a contradiction. Therefore $\mathcal{X}(H) = \cup_{i \geq 0}^{n} \mathcal{X}_i(H)$ as claimed. □

4.2 Proof of Theorem 4.1

Let $u \in F_2$ and $H \leq_{f.g.} F_2$. We want to show that it is decidable whether $\mu(u) \in H$ for some $\mu \in \mathrm{Aut}\, F_2$. By Theorem 3.8, and since $\Psi^{-1} = \Psi$, it suffices to decide whether there exist $w \in F_2$ and $n \geq 0$ such that one of the following conditions hold:

- $\lambda_w \varphi_{a,ba}^n \varphi_{a^{-1},b}(u) \in \Sigma^* \Psi(H)$;
- $\lambda_w \varphi_{a,ba}^n \varphi_{a^{-1},b^{-1}}(u) \in \Sigma^* \Psi(H)$.

Since Ψ is finite, it suffices to be able to decide whether

$$\text{there exist } w \in F_2 \text{ and } n \geq 0 \text{ such that } \lambda_w \varphi_{a,ba}^n(u) \in \mu(H) \text{ for some } \mu \in \Sigma^*. \tag{6}$$

As noted before, we may use L instead of Σ^*. In view of Proposition 2.4(i), we may also replace $\lambda_w \varphi_{a,ba}^n$ by $\varphi_{a,ba}^n \lambda_w$.

We start by considering the case $n = 0$. Again by Proposition 2.4(i), we may assume that u is cyclically reduced, and by Proposition 2.1(vi), our problem further reduces to asking if one can decide whether

$$u \text{ labels a loop in } \mathcal{A}(\mu(H)) \text{ for some } \mu \in L. \tag{7}$$

We note that every loop contains either the origin or a singularity: if it does not contain the origin, then there is a path from the origin to a state in the loop, and the first contact between that path and the loop is a source or a sink. In particular, every loop labelled by u in $\mathcal{A}(\mu(H))$ is also in $\mathcal{A}_t(\mu(H)$ if $t > |u|/2$. Let us then fix $t > \max\{|u|/2, \delta(H)\}$. Then for every $\mu \in L$, u labels a loop in $\mathcal{A}(\mu(H))$ if and only if u labels a loop in $\mathcal{A}_t(\mu(H))$. By Corollary 4.11 we can effectively compute the finite set

$$\mathcal{X}(H) = \{\mathcal{A}_t(K) \mid K \in \Sigma^*(H)\}.$$

Thus (7) is decidable, and hence (6) is decidable for $n = 0$. It is also decidable for any fixed n (applying the case $n = 0$ to $\varphi_{a,ba}^n(u)$ instead of u).

We now consider (6) in its full generality. If $u \in R_a$, then we are reduced to the case $n = 0$ since $\varphi_{a,ba}(u) = u$. So we will assume that b or b^{-1} occurs in u, and by conjugation again, we may assume that u starts with b or ends with b^{-1} (and not both since u is cyclically reduced).

Let M be the least common multiple of $1, 2, \ldots, \delta_0(H)$. In order to prove (6), it suffices to show that if there exist $w \in F_2$ and $n \geq 0$ such that $\lambda_w \varphi_{a,ba}^n(u) \in \mu(H)$ for some $\mu \in \Sigma^*$, then $\lambda_w \varphi_{a,ba}^n(u) \in \mu(H)$ for some $w \in F_2$, $n < |u| + \max\{|u|, M + \delta(H)\}$ and $\mu \in \{1, \varphi_{b^{-1},a^{-1}}\} \cup \varphi_{b^{-1},a^{-1}} \varphi_{a,ba} L$. Since we have proved (6) for bounded n, the latter property is decidable, and hence (6) is decidable in general.

We now proceed to proving this reduction, assuming that $\lambda_w \varphi_{a,ba}^n(u) \in \mu(H)$ for some $n \geq 0$, $w \in F_2$ and $\mu \in \Sigma^*$. We consider such a triple (w, n, μ) with n minimal and we want to show that $n < |u| + \max\{|u|, M + \delta(H)\}$. So let us assume that $n \geq |u| + \max\{|u|, M + \delta(H)\}$. As already observed, $\varphi_{a,ba}^n \lambda_{w'}(u) \in \mu(H)$ for some $w' \in F_2$, and if μ starts with $\varphi_{a,ba}$ (as a word on Σ), we may cancel $\varphi_{a,ba}$ on both sides, and hence reduce n. Thus, by minimality, we may assume that μ does not start with $\varphi_{a,ba}$, that is, $\mu \in \{1, \varphi_{b^{-1},a^{-1}}\} \cup \varphi_{b^{-1},a^{-1}} \varphi_{a,ba} L$.

Write $u = a^{i_0}b^{\varepsilon_1}a^{i_1}\dots b^{\varepsilon_k}a^{i_k}$ with $\varepsilon_\ell = \pm 1$ for every ℓ. For every $m \geq |u|$, we have $m > |i_\ell|$ for every ℓ and it follows easily that

$$\varphi^m_{a,ba}(u) = \varphi_{a,ba^m}(u) = a^{j_0}b^{\varepsilon_1}a^{j_1}\dots b^{\varepsilon_k}a^{j_k}$$

with

$$j_\ell = \begin{cases} i_\ell + m & \text{if } \epsilon_\ell = \epsilon_{\ell+1} = 1, \text{ or } \ell = k \text{ and } \epsilon_k = 1 \\ i_\ell - m & \text{if } \epsilon_\ell = \epsilon_{\ell+1} = -1, \text{ or } \ell = 0 \text{ and } \epsilon_1 = -1 \\ i_\ell & \text{in all other cases} \end{cases}$$

Since we assumed that u is cyclically reduced and u starts with b or ends with b^{-1}, it is easily verified that $\varphi^m_{a,ba}(u)$ is cyclically reduced as well.

If $n \geq |u| + \max\{|u|, M + \delta(H)\}$, then we have $n, n-M \geq |u|$, so if

$$\varphi^n_{a,ba}(u) = a^{r_0}b^{\varepsilon_1}a^{r_1}\dots b^{\varepsilon_k}a^{r_k}, \quad \varphi^{n-M}_{a,ba}(u) = a^{s_0}b^{\varepsilon_1}a^{s_1}\dots b^{\varepsilon_k}a^{s_k},$$

we may write

$$s_\ell = \begin{cases} r_\ell & \text{if } |r_\ell| < |u| \\ r_\ell + M & \text{if } r_l \leq -M - |u| \\ r_\ell - M & \text{if } r_l \geq M + |u| \end{cases}$$

for $l = 0, \dots, k$. Indeed, if $|r_\ell| < |u|$, then $r_\ell = i_\ell$ since otherwise $|r_\ell| \geq n - |i_\ell| \geq |u|$. On the other hand, if $r_\ell \leq -M - |u|$, then $r_\ell = i_\ell - n$ and so $r_\ell = s_\ell - M$. The case $r_l \geq M + |u|$ is similar.

By Proposition 2.1(vi), $\varphi^n_{a,ba}(u)$ labels a loop in $\mathcal{A}(\mu(H))$. For every ℓ, $|r_\ell| < |u|$ or $|r_\ell| > n - |u| \geq M + \delta(H)$. But every cycle-free a-path in $\mathcal{A}(\mu(H))$ has length at most $\delta(H)$ (by (4) in Subsection 4.1). So every factor a^{r_ℓ} of u such that $|r_\ell| \geq |u|$ must be read in a cycle of $\mathcal{A}(\mu(H))$ (in an inverse automaton, if a homogeneous path contains a cycle, then it reads entirely along that cycle). Note that, by definition, M is a multiple of the length c_ℓ of that cycle. Now compare $\varphi^{n-M}_{a,ba}(u)$ and $\varphi^n_{a,ba}(u)$: the difference in each a-path is either a^M, or a^{-M}, or non-existent. In any case, it consists of a whole number of passages around the length c_ℓ cycle, and hence $\varphi^{n-M}_{a,ba}(u)$ labels a path in $\mathcal{A}(\mu(H))$ as well. This contradicts the minimality of n and completes the proof. □

5 Beyond rank 2

We do not know how to extend Theorem 4.1 to arbitrary finite alphabets, but we can get decidability for weakened versions of the problem. The first such result involves a restriction on the subgroups considered.

Theorem 5.1 *Let $u \in F_A$ and let $H \leq_{f.g.} F_A$. If H is cyclic or a free factor of F_A, it is decidable whether or not $\mu(u) \in H$ for some $\mu \in Aut F_A$.*

Proof. Let us first assume that H is a free factor of F_A, with rank k. It is easily verified that $\mu(u) \in H$ for some automorphism μ if and only if u sits in some rank k free factor of F_A. We conclude using the result from [11], which shows that one can effectively compute the least free factor of F_A containing u (the algebraic closure of the subgroup $\langle u \rangle$).

Let us now assume that $H = \langle v \rangle$. Without loss of generality, we may assume that u and v are cyclically reduced. Say that a word x is root-free if it is not equal to a non-trivial power of a shorter word. Then $u = x^k$ for some uniquely determined integer $k \geq 1$ and root-free word x, and similarly, $v = y^\ell$ for some uniquely determined $\ell \geq 1$ and root-free y. It is an elementary verification that the image of a cyclically reduced root-free word by an automorphism is also cyclically reduced and root-free. Thus, an automorphism maps u into H if and only if and only it maps x to y or y^{-1}, and k is a multiple of ℓ. Decidability follows from the fact that we can decide whether two given words are in each other's automorphic orbit, using Whitehead's algorithm [10]. $\square$

The second result on a weakened version of our orbit problem involves *almost bounded automorphisms*. Given a finite alphabet A and $k \in \mathbb{N}$, we say that an automorphism φ of F_A is *k-almost bounded* if $|\varphi(a)| > k$ for at most one letter $a \in A$. We let $\mathrm{AlmB}_k F_A$ denote the set of k-almost bounded automorphisms of F_A.

Theorem 5.2 *Given $u \in F_A$, $L \subseteq R_A$ rational and $k \in \mathbb{N}$, it is decidable whether or not $\overline{\mu(u)} \in L$ for some $\mu \in AlmB_k F_A$.*

The proof of this theorem relies on Diekert et al.'s result on the decidability of the existential theory of equations with rational constraints in free groups [5]. It also requires the following result, which generalizes Lemma 3.7.

Proposition 5.3 *Let $m = |A|$ and $v_1, \ldots, v_{m-1} \in R_A$. Then*

$$X = \{x \in R_A \mid (v_1, \ldots, v_{m-1}, x) \text{ is a basis of } F_A\}$$

is rational and effectively constructible.

Proof. First note that X is nonempty if and only if $(v_1, ..., v_{m-1})$ is a basis of a free factor of F_A. This is decidable. In fact, it is verified in [18] that if $K = \langle v_1, ..., v_{m-1} \rangle$, then K is a free factor of F_A if and only if there are vertices p and q of $\mathcal{A}(K)$ whose identification leads (via foldings) to the bouquet of circles $\mathcal{A}(F_A)$. In addition, if u_p and u_q are the labels of geodesic paths of $\mathcal{A}(K)$ from the origin to p and q, and if $z = \overline{u_p u_q^{-1}}$, then $z \in X$. Thus it is decidable whether $X = \emptyset$, and if it is not, then we can effectively construct an element z of X.

Let $\varphi \in \mathrm{Aut}\, F_A$ be defined by $\varphi(a_i) = v_i$ $(i = 1, \ldots, m-1)$ and $\varphi(a_m) = z$. Then $x \in X$ if and only if $(a_1, \ldots, a_{m-1}, \varphi^{-1}(x))$ is a basis of F_A. Write $R = R_{\{a_1,\ldots,a_{m-1}\}}$. By Lemma 3.7, this is equivalent to say that $\varphi^{-1}(x) \in R(a_m \cup a_m^{-1})R$ and therefore

$$X = \overline{\varphi(R(a_m \cup a_m^{-1})R)} = \overline{V(z \cup z^{-1})V}$$

for $V = \{v_1, \ldots, v_{m-1}, v_1^{-1}, \ldots, v_{m-1}^{-1}\}^*$.

Since $V(z \cup z^{-1})V$ is a rational subset of $(A \cup A^{-1})^*$, we conclude that X is rational by Theorem 2.2. Moreover, the formula $X = \overline{V(z \cup z^{-1})V}$ provides an effective construction of X. $\square$

Proof of Theorem 5.2. Write $A = \{a_1, \ldots, a_m\}$. Without loss of generality, we may restrict ourselves to the case $|\mu(a_i)| \leq k$ for $i = 1, \ldots, m-1$. Since there are only finitely many choices for these $\mu(a_i)$, we may as well assume them to be fixed, say $\mu(a_i) = v_i$ for $i = 1, \ldots, m-1$.

Write $u = u_0 a_m^{\varepsilon_1} u_1 \ldots a_m^{\varepsilon_n} u_n$ with $n \geq 0$, $u_i \in F_{\{a_1,\ldots,a_{m-1}\}}$ and $\varepsilon_i = \pm 1$ for every i. Then we must decide if there exists some

$$y \in X = \{x \in R_A \mid (v_1, \ldots, v_{m-1}, x) \text{ is a basis of } F_A\}$$

such that

$$\overline{u'_0 y^{\varepsilon_1} u'_1 \ldots y^{\varepsilon_n} u'_n} \in L,$$

where u'_i is the word obtained by replacing each a_j by v_j in u_i. Note that X is rational by Proposition 5.3. This is equivalent to deciding whether or not the equation

$$u'_0 y^{\varepsilon_1} u'_1 \ldots y^{\varepsilon_n} u'_n = z$$

on the variables y, z has some solution in F_A with the rational constraints $y \in X$ and $z \in L$. By [5], this is decidable. □

Corollary 5.4 *Given $u \in F_A$, $H \leq_{f.g.} F_A$ and $k \in \mathbb{N}$, it is decidable whether or not $\mu(u) \in H$ for some $\mu \in AlmB_kF_A$.*

Proof. In view of Theorem 2.2, the reduced words of H constitute a rational language and so we may apply Theorem 5.2. □

References

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

[2] P. Brinkmann, Detecting automorphic orbits in free groups, arXiv:0806.2889v1.

[3] J. Berstel, A. Lauve, C. Reutenauer and F. Saliola. *Combinatorics on words: Christoffel words and repetitions in words*, AMS Monographs, to appear.

[4] M. Cohen, W. Metzler and A. Zimmermann. What does a basis of $F(a, b)$ look like?, *Math. Ann.* 257 (1981), 435–445.

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

[6] S. Gersten, On Whitehead's algorithm, *Bull. Am. Math. Soc.* 10 (1984) 281–284.

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

[8] I. Kapovich and A. Myasnikov, Stallings foldings and subgroups of free groups, *J. Algebra* 248 (2002), 608–668.

[9] M. Lothaire. *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, 2002.

[10] R. C. Lyndon and P. E. Schupp, *Combinatorial Group Theory*, Springer-Verlag 1977.

[11] A. Miasnikov, E. Ventura and P. Weil, Algebraic extensions in free groups, in *Algebra and Geometry in Geneva and Barcelona* (G.N. Arzhantseva, L. Bartholdi, J. Burillo and E. Ventura eds.), Trends in Mathematics, Birkhaüser (2007), pp. 225–253.

[12] A.G. Myasnikov and V. Shpilrain, Automorphic orbits in free groups, *J. Algebra* 269 (2003), 18–27.

[13] J. Nielsen. Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden, *Math. Ann.* 78 (1918), 385–397.

[14] R. P. Osborne and H. Zieschang. Primitives in the free group on two generators, *Invent. Math.* 63 (1981), 17–24.

[15] A. Roig, E. Ventura and P. Weil, On the complexity of the Whitehead minimization problem, *Int. J. Alg. Comput.* 17 (2007), 1611-1634.

[16] J. Rotman. *An introduction to the theory of groups*, 4th edition, Springer, 1995.

[17] J.-P. Serre. *Arbres, amalgames, SL_2*, Astérisque 46, Soc. Math. France, 1977. English translation: *Trees*, Springer Monographs in Mathematics, Springer, 2003.

[18] P. V. Silva and P. Weil. On an algorithm to decide whether a free group is a free factor of another, *RAIRO Theoretical Informatics and Applications* 42 (2008), 395–414.

[19] J. Stallings. Topology of finite graphs, *Invent. Math.* 71 (1983), 551–565.

[20] Z. X. Wen and Z. Y. Wen. Local isomorphisms of invertible substitutions, *C. R. Acad. Sci. Paris Sér. I Math.* 318 (1994), 299–304.

[21] J.H.C. Whitehead, On equivalent sets of elements in a free group, *Annals of Mathematics* 37 (1936) 782–800.