

## Subgroups of free groups and primitive elements

A. Clifford and R. Z. Goldstein

(Communicated by J. Howie)

**Abstract.** An algorithm is developed to determine whether a subgroup of a free group contains a primitive element. This answers question 39b in the list of problems on free groups posted on the World of Groups website, [www.grouptheory.info](http://www.grouptheory.info).

### 1 Introduction

The World of Groups website [12] maintains a list of open problems in group theory. Question F39 has two parts. The first part asks whether there is an algorithm which, when given a subgroup  $H$  of a free group  $F$  and an element  $w \in F$ , can determine whether there is an automorphism  $\phi : F \rightarrow F$  so that  $\phi(w) \in H$ . In this paper, we address the second part by developing an algorithm which, given a subgroup  $H$  of a free group  $F$ , determines whether  $H$  contains an element which is part of some basis for  $F$ . Such an element is called *primitive*. Question F39a still remains unanswered in general. However, in a recent preprint [9], Silva and Weil have given an algorithm in the case where the ambient free group has rank two.

### 2 Automorphisms

Throughout,  $F$  will denote a free group on  $n$  generators. We will take as a basis for  $F$  the set  $X = \{x_1, x_2, \dots, x_n\}$ . We begin by discussing various automorphisms of  $F$ . Let  $\sigma$  be a permutation of the set  $\{1, 2, \dots, n\}$  and for each  $i$ , let  $\hat{x}_i$  be a choice of either  $x_i$  or  $x_i^{-1}$ . Then we define the automorphism  $\phi_{\sigma, \wedge} : F \rightarrow F$  on the basis by  $\phi_{\sigma, \wedge}(x_i) = \hat{x}_{\sigma(i)}$ . This is a *type I automorphism*. We note that the set of type I automorphisms is closed under composition.

Let  $k$  and  $l$  be two integers satisfying  $1 \leq k \leq l \leq n$ . Then a *type II automorphism* is of the form  $\phi_{k,l} : F \rightarrow F$  defined on the basis  $X = \{x_1, x_2, \dots, x_n\}$  by:

$$\begin{aligned}\phi_{k,l}(x_1) &= x_1, \\ \phi_{k,l}(x_i) &= x_1 x_i \text{ for } 2 \leq i \leq k, \\ \phi_{k,l}(x_i) &= x_1 x_i x_1^{-1} \text{ for } k < i \leq l, \\ \phi_{k,l}(x_i) &= x_i \text{ fixed for } l < i \leq n.\end{aligned}$$

To ensure that  $\phi_{k,l}$  is not the identity, we make the further assumption that  $l \geq 2$ . One can easily obtain all Nielsen automorphisms as compositions of Type I and Type II automorphisms. It is well known (cf. [5]) that every automorphism of  $F$  can be written as a composition of Nielsen automorphisms. Therefore, the collection of Type I and Type II automorphisms generates the automorphism group of  $F$ .

### 3 Graphs

All graphs considered will be finite, directed and labeled with the elements of  $X = \{x_1, x_2, \dots, x_n\}$ . All graph maps between such graphs will be assumed to preserve the direction and labeling of the edges. Given a graph  $K$ , we shall use  $E_K$  to denote the edge set of  $K$  and  $|K|$  to denote the cardinality of  $E_K$ .

Let  $K$  be such a graph. A path in  $K$  corresponds to an element of  $F$  in the usual way. We pick a vertex  $v_0$  in  $K$ , and define the subgroup  $H = \pi K$  as the set of elements obtained from closed paths at  $v_0$ . If we choose a different vertex, we get a subgroup  $H_1$  conjugate to  $H$ . Since  $H$  contains a primitive element if and only if any conjugate of it does, we will not make reference to base points and consider  $K$  to represent any subgroup in the conjugacy class of  $\pi K$ .

The graph  $K$  is reduced if there is no vertex  $v$  adjacent to two edges both labeled with some  $x_i$ , either both directed toward  $v$ , or both directed away from  $v$ . If a graph  $K$  is not reduced, we can perform a sequence of Stallings foldings, as described in [7], to obtain a reduced graph  $K^*$  so that  $\pi K = \pi K^*$ . Moreover, there is an induced graph map  $p_K : K \rightarrow K^*$  which we call a folding map. Throughout, we will omit subscripts when they become too burdensome. Note that if  $v$  and  $w$  are vertices in  $K$ , then  $p_K(v) = p_K(w)$  if and only if there is a path in  $K$  from  $v$  to  $w$  whose word freely reduces to 1 in  $F$ . Let  $f : L \rightarrow K$  be a graph map between the graphs  $L$  and  $K$ . There is an induced map  $f^* : L^* \rightarrow K^*$  so that  $p_K \circ f = f^* \circ p_L$ .

For any finitely generated subgroup  $H$  of  $F$ , there is a finite reduced graph  $K$  so that  $H = \pi K$ . To see this, let  $L$  be the covering of the wedge of  $n$  circles corresponding to  $H$ . Then we can take  $K$  to be any subgraph of  $L$ , appropriately directed and labeled, so that the inclusion induced  $i_* : \pi_1(K) \rightarrow \pi_1(L)$  is an isomorphism. Conversely, if  $\pi K = H$ , we can recognize it as a subgraph of  $L$ . We call such a graph a *core* for  $H$ . We note that typically core graphs are precluded from having degree one vertices. However, for technical reasons, we shall find it convenient to allow core graphs to have degree one vertices as may result from our definition.

Let  $K$  be a core for  $H$ , a subgroup of  $F$ . For a given automorphism  $\phi$  of  $F$ , we wish to describe a core for  $\phi[H]$ . If  $\phi$  is a type I automorphism, clearly by relabeling some edges and by redirecting some edges, we can obtain a graph  $K'$  so that  $\phi_{\sigma, \wedge}(\pi K) = \pi K'$ .

We wish to define a similar process for the type II automorphism  $\phi_{k,l}$ . Let  $K$  be a core graph for the subgroup  $H$  as described above. For each edge  $e$  labeled  $x_i$  for some  $i \in \{2, \dots, k\}$ , we introduce a vertex,  $v_e$  and two edges, one from  $\iota(e)$  to  $v_e$  labeled  $x_1$ , and the other from  $v_e$  to  $\tau(e)$  labeled  $x_i$ . We then remove the original edge  $e$ . Similarly, for each edge  $e$  labeled  $x_i$  for some  $i \in \{k+1, \dots, l\}$ , we introduce two vertices,  $u_e$  and  $w_e$  and three edges, two labeled  $x_1$ , one from  $\iota(e)$  to  $u_e$  and one from  $\tau(e)$

to  $w_e$ , and one from  $u_e$  to  $w_e$  labeled  $x_i$ . We then remove the original edge  $e$ . In this process, the edge that we remove has the same label as exactly one of the two or three edges that we added. We call the added edge with the same label as the removed edge a *replacement* edge. The other added edges are all labeled  $x_1$ . We call these edges *new*  $x_1$  edges. The edges of  $K$  labeled  $x_1$  are *old*  $x_1$  edges. We call this new graph  $\hat{K}$ . Now, there is probably not a graph map from  $K$  to  $\hat{K}$ , but there is an injection  $\rho_K : E_K \rightarrow E_{\hat{K}}$  where  $\rho_K(e)$  is the replacement edge of  $e$  if  $e$  is labeled  $x_i$  for some  $i$  with  $2 \leq i \leq l$ , and  $\rho_K(e) = e$  otherwise. The proof of the following lemma will be left to the reader.

**Lemma 1.** *Let  $f : L \rightarrow K$  be a graph map. Let  $\hat{L}$  and  $\hat{K}$  be obtained from  $L$  and  $K$  via the automorphism  $\phi_{k,l}$ . Then there is an induced graph map  $\hat{f} : \hat{L} \rightarrow \hat{K}$  so that for any edge  $e$  of  $L$ ,  $\hat{f}[\rho_L(e)] = \rho_K(f[e])$ .*

The graph  $\hat{K}$  may not be a core subgraph as it may not be reduced. We perform a sequence of Stallings foldings to obtain a core graph  $K' = (\hat{K})^*$  which may contain degree one vertices. We will find the following lemmas useful. The proof of the first is left to the reader.

**Lemma 2.** *Let  $F[x_1, x_2, \dots, x_n] = F$  be the free group on  $n$  generators. Let  $H$  be a subgroup of  $F$  with core graph  $K$ . Let  $\phi : F \rightarrow F$  be an automorphism which acts as a type II basis change. Let  $K' = (\hat{K})^*$  be the core graph obtained from  $K$  via  $\phi$  as described above. Then  $K'$  is a core graph of  $\phi(H)$ .*

**Lemma 3.** *Let  $K$  be a core graph. Let  $\hat{K}$  be obtained from  $K$  via the type II automorphism  $\phi_{k,l}$  and let  $p_{\hat{K}} : \hat{K} \rightarrow K'$  be a folding map. If  $e_1$  and  $e_2$  are two edges so that  $p_{\hat{K}}(e_1) = p_{\hat{K}}(e_2)$ , then both edges are labeled  $x_1$  and at least one is a new  $x_1$  edge.*

*Proof.* Let  $e_1$  and  $e_2$  be two edges of  $\hat{K}$  adjacent to  $v_1$  and  $v_2$  respectively so that  $p_{\hat{K}}(e_1) = p_{\hat{K}}(e_2)$  and  $p_{\hat{K}}(v_1) = p_{\hat{K}}(v_2)$ . Clearly, this implies that  $e_1$  and  $e_2$  have the same label. There is a path in  $\hat{K}$  from  $v_1$  to  $v_2$  whose label is a word freely equal to 1. If both  $e_1$  and  $e_2$  correspond to edges of  $K$ , then there is a path between the corresponding vertices in  $K$  whose label is a word freely equal to 1. This is impossible as  $K$  is reduced. The result follows.  $\square$

**Lemma 4.** *Let  $f : L \rightarrow K$  be a map between reduced graphs. Let  $L'$  and  $K'$  be the reduced graphs obtained from  $L$  and  $K$  via  $\phi_{k,l}$  and let  $f' : L' \rightarrow K'$  be the induced graph map. If  $e_1$  and  $e_2$  are distinct edges of  $L$  which are identified by  $f$ , then the corresponding edges  $e'_1$  and  $e'_2$  are distinct and identified by  $f'$ .*

*Proof.* That  $e'_1$  and  $e'_2$  are distinct is a consequence of Lemma 3. It is a consequence of Lemma 1 and the fact that the folding maps commute with induced maps that  $f'$  identifies  $e'_1$  and  $e'_2$ .  $\square$

#### 4 Manifolds

In [10] and [11], Whitehead developed techniques for studying automorphisms of free groups using  $M = (S^1 \times S^2) \# (S^1 \times S^2) \# \cdots \# (S^1 \times S^2)$ , the 3-manifold which is the connected sum of  $n$  copies of  $S^1 \times S^2$ . We will provide a brief review of these techniques, but refer the interested reader to [3], [4] and [8] for more details. A sphere basis for  $M$  is an ordered  $n$ -tuple of spheres  $\Sigma = (S_1, S_2, \dots, S_n)$  embedded in  $M$  so that  $M - \bigcup_{i=1}^n S_i$  is connected. If we split each  $S_i$ , we get a *pre-manifold*  $M_1$  which is homeomorphic to the 3-sphere  $S^3$  with  $2n$  copies of the open 2-ball  $B^2$  removed. So each  $S_i$  of  $M$  corresponds to two boundary components of  $M_1$  which we call  $S_i^-$  and  $S_i^+$ . Now,  $M$  is obtained from  $M_1$  by identifying each  $S_i^-$  to  $S_i^+$  by reversing orientation. For clerical expedience, we let  $\psi : M_1 \rightarrow M$  be that identification map. Now, let  $U$  be a collar on  $S_i$  in  $M$ . Then  $U - S_i$  has two components  $U_1$  and  $U_2$ . If the closure of  $\psi^{-1}[U_j]$  contains  $S^+$ , we call  $U_j$  the *positive side* of  $S_i$ . Otherwise, the closure of  $\psi^{-1}[U_j]$  contains  $S^-$  and we call  $U_j$  the *negative side* of  $S_i$ .

Clearly,  $\pi_1(M) = F$  is isomorphic to a free group on  $n$  generators. Since we are looking for primitive elements of  $F$ , we will only be interested in elements of  $F$  up to conjugacy. For this reason, we will not pay attention to the base point of  $M$ . Each choice of sphere basis for  $M$  corresponds to a choice of generators of  $F$  as follows: Let  $C$  be a loop in  $M$  which meets exactly one element  $S_i$  of the sphere basis. We can pull  $C$  back to  $M_1$  where it appears as an arc between  $S_i^+$  and  $S_i^-$ . If the arc is directed from  $S_i^+$  to  $S_i^-$ , then the curve  $C$  represents the primitive element  $x_i$  of  $F$ ; if it is directed the other way, it represents  $x_i^{-1}$ . In this way we can determine the conjugacy class of elements of  $F$  represented by a curve embedded in  $M$ .

Again, let  $\Sigma = (S_1, S_2, \dots, S_m)$  be a sphere basis for  $M$ . Reordering the elements of this  $n$ -tuple, or exchanging the roles of  $S_j^-$  and  $S_j^+$  for some  $j$ , is tantamount to choosing a different basis of  $\pi_1(M) = F$ . Such a change of basis corresponds to a type I automorphism.

Next, we discuss Whitehead moves on the sphere basis for  $M$ . These moves correspond to type II automorphisms.

Let  $S_*$  be a sphere properly embedded in  $M$  disjoint from the sphere basis. We will find it easier to refer to  $M_1$ , in which we denote  $\psi^{-1}(S_*)$  also as  $S_*$ . We note that  $S_*$  disconnects  $M_1$  into two submanifolds, and assume that there is some fixed  $i_0$  so that  $S_*$  separates  $S_{i_0}^-$  from  $S_{i_0}^+$ . We call the submanifold containing  $S_{i_0}^-$  the *inside* of  $S_*$  and the one containing  $S_{i_0}^+$  the *outside* of  $S_*$ . By a type I basis change, we can assume that  $i_0 = 1$  and that there are  $k$  and  $l$  so that the following assertions hold: for all  $i \in \{1, \dots, k\}$ ,  $S_i^-$  is in the inside of  $S_*$  and  $S_i^+$  is in the outside of  $S_*$ ; for all  $i \in \{k+1, \dots, l\}$ , both  $S_i^-$  and  $S_i^+$  are in the inside of  $S_*$ ; and for all  $i \in \{l+1, \dots, n\}$ , both  $S_i^-$  and  $S_i^+$  are in the outside of  $S_*$ .

We may use  $S_*$  to make a type II basis change by replacing  $S_1$  with  $S_*$ . We begin by splitting  $M_1$  along  $S_*$  to obtain a disconnected manifold  $D$  with  $2n+2$  boundary components. This manifold has two components, one which contains  $S_1^-$  and a copy of  $S_*$  which we call  $S_*^+$ ; and one which contains  $S_1^+$  and a copy of  $S_*$  which we call  $S_*^-$ . We identify  $S_1^-$  to  $S_1^+$  using  $\psi$  to obtain  $M_2$ . We point out that by identifying  $S_*^-$  to  $S_*^+$  in the obvious way and identifying  $S_i^-$  to  $S_i^+$  using  $\psi$ , we regain the manifold

$M$ , now with sphere basis  $\Sigma' = (S_*, S_2, S_3, \dots, S_n)$ . With this sphere basis, a curve that only meets  $S_*$  corresponds either to  $x_1$  or  $x_1^{-1}$ , depending on how it is directed. So this process induces the automorphism  $\phi_{k,l}$  on  $F$ .

## 5 Curves in $M$

Let  $\alpha : S^1 \rightarrow M$  be an embedding of a simple closed curve  $C$  in  $M$  which is transverse to the elements of the sphere basis  $\Sigma = (S_1, S_2, S_3, \dots, S_n)$  representing the cyclically reduced element  $w = y_1 y_2 \dots y_m$  of  $F$  written in the basis  $X = \{x_1, x_2, \dots, x_n\}$ , where  $y_j \in X \cup X^{-1}$ . We will impose a graph structure on  $C$  using  $\Sigma$ . We will call the resulting graph  $C_\Sigma$ . An *arc of  $C$  relative to  $\Sigma$*  is the closure of a component of  $C - \Sigma$ . In terms of the identification map  $\psi : M_1 \rightarrow M$ , an arc is a component of  $\psi^{-1}[C]$ . The ordering on  $S^1$  gives us an ordering of the arcs  $C_1, C_2, \dots, C_m$  of  $C$ , and directs each arc. We choose a point  $q_j$  on  $C_j$ . These points will be the vertices of  $C_\Sigma$ .

We call the components of  $C - \{q_1, q_2, \dots, q_m\}$  *intervals*. The intervals of  $C$  correspond to the edges of  $C_\Sigma$  in the following way. The interval  $I_j$  of  $C$  is that part of  $C$  bounded by  $q_j$  and  $q_{j+1}$ . It meets exactly one element  $S_i$  from the sphere basis where  $y_j = x_i^{\pm 1}$ . The corresponding edge of  $C_\Sigma$  is labeled  $x_i$  and directed from the vertex in the arc adjacent to  $S_i^-$  to the vertex in the arc adjacent to  $S_i^+$ . We note that the graph  $C_\Sigma$  is a circle graph.

Let  $K$  be a core for the subgroup  $H$  and assume that  $w \in H$ . Then there is a graph map  $\gamma : C_\Sigma \rightarrow K$  which reads  $w$ . The subgraph  $\gamma[C_\Sigma]$  of  $K$  is called *the support of  $C_\Sigma$  in  $K$* . The number of edges in the support of  $C_\Sigma$  in  $K$  is the *width of  $C_\Sigma$  in  $K$* , denoted  $w_K(C)$ . We define the *deficiency of  $C_\Sigma$  in  $K$*  to be  $\delta(C_\Sigma) = |C_\Sigma| - w_K(C)$ . The deficiency measures the number of repeated edges obtained when reading the word  $w$  in  $K$ . In particular, if  $\delta(C_\Sigma) = 0$ , then the word  $w$  has no repeated edges in  $K$ .

Clearly the structure that we have put on the graph  $C_\Sigma$  depends on our choice of sphere basis  $\Sigma = (S_1, S_2, S_3, \dots, S_n)$ . Next we will describe how this structure changes when we choose a new sphere basis via a Whitehead move. To this end, let  $S_*$  be a sphere embedded in  $M$  disjoint from the elements of  $\Sigma$ , and let  $\Sigma' = (S_*, S_2, S_3, \dots, S_n)$  be a sphere basis so that the change of basis from  $\Sigma$  to  $\Sigma'$  corresponds to the automorphism  $\phi_{k,l}$ . We assume that the arcs  $C_j$  are transverse to  $S_*$  and that no arc intersects  $S_*$  more than once. While this latter assumption is quite strong, it will be the case in our eventual application. It ensures that the word described by  $C$  is reduced with respect to the new sphere basis  $\Sigma' = (S_*, S_2, \dots, S_n)$ . We also will make the inconsequential assumption that no vertex of an arc lies on  $S_*$ .

The next lemma follows directly from Lemma 3.

**Lemma 5.** *Let  $C_\Sigma$  and  $C_{\Sigma'}$  be the circle graphs corresponding to the curve  $C$  in  $M$  with sphere bases  $\Sigma = (S_1, S_2, \dots, S_n)$  and  $\Sigma' = (S_*, S_2, \dots, S_n)$  respectively where  $\Sigma'$  is obtained from  $\Sigma$  by a Whitehead move corresponding to the automorphism  $\phi_{k,l}$  of  $F$ . Let  $(C_\Sigma)' = (\hat{C}_\Sigma)^*$  be the graph obtained by applying  $\phi_{k,l}$  to  $C_\Sigma$  as described in Section 3. Then  $(C_\Sigma)'$  is precisely the circle graph  $C_{\Sigma'}$  together with some edges labeled  $x_1$  each adjacent to a vertex of degree one. In this way, we can consider  $C_{\Sigma'}$  as a subgraph of  $(C_\Sigma)'$ .*

Given a sphere basis  $\Sigma = (S_1, S_2, \dots, S_n)$  and a curve  $C_\Sigma$ , we say the sphere  $S_*$  is *stingy* with respect to  $C_\Sigma$  if every arc of  $C_\Sigma$  that meets  $S_*$  also meets  $S_1^-$ . For the rest of this paper, we will assume that our sphere  $S_*$  is stingy.

Next, assume that  $C$  represents an element of some subgroup  $H$  of  $F$  and let  $K$  be a core for  $H$ . In this situation, we have two graph maps  $\gamma: C_\Sigma \rightarrow K$  and  $\gamma': C_{\Sigma'} \rightarrow K'$ . We wish to compare the width of  $C$  in  $K$  to its width in  $K'$ .

**Theorem 1.** *Let  $C$  be a curve embedded in  $M$  and let  $C_\Sigma$  be the circle graph corresponding to  $C$  with respect to the sphere basis  $\Sigma = (S_1, S_2, \dots, S_n)$ . Let  $S_*$  be a stingy sphere embedded in  $M$  so that the change of basis from  $\Sigma$  to  $\Sigma' = (S_*, S_2, \dots, S_n)$  induces the automorphism  $\phi_{k,1}$ . Further, assume that there is a subgroup  $H$  with core graph  $K$  so that  $C_\Sigma$  represents an element  $w \in H$ . Let  $K'$  be the core graph obtained from  $K$  via  $\phi_{k,1}$ , and let  $C_{\Sigma'}$  be the circle graph corresponding to the curve  $C$  with respect to the basis  $\Sigma'$ . Then  $w_{K'}(C) \leq w_K(C)$ .*

*Proof.* From the above lemma, we know that  $(C_\Sigma)'$  is precisely the circle graph  $C_{\Sigma'}$  together with edges labeled  $x_1$  attached to some of the vertices of  $C_{\Sigma'}$ . Since  $S_*$  is stingy, it follows that in  $C_{\Sigma'}$  every new  $x_1$  edge is adjacent to either another new  $x_1$  edge or an old  $x_1$  edge with the opposite orientation. So every new  $x_1$  edge gets folded. Therefore,  $|C_{\Sigma'}| \leq |C_\Sigma|$ . Now from Lemma 4, we know that if  $\gamma$  identifies  $e_1$  and  $e_2$  for some  $e_1, e_2 \in E_{C_\Sigma}$  then  $\gamma'$  identifies  $e'_1$  and  $e'_2$ . The result follows.  $\square$

## 6 Primitive curves in $M$

A curve  $C$  in  $M$  is *primitive* if it represents a primitive element of  $F$ . The following theorem will help us detect primitive curves in  $M$ . A slightly incorrect version of this statement was given in [1]. We include a corrected version here.

**Theorem 2.** *Let  $w$  be a reduced element of  $F$ . Then  $w$  is primitive if and only if there is a primitive curve  $C$  in  $M$  representing  $w$  with respect to the basis  $\Sigma$ , and an embedding of a sphere  $T$  so that  $C \cap T$  consists of a single point.*

*Proof.* If  $w$  is primitive then there exists an automorphism  $\phi: F \rightarrow F$  such that  $\phi(x_1) = w$ . Moreover, there is a homeomorphism  $f: M \rightarrow M$  that induces  $\phi$ . Whitehead in [11] has shown that we can take  $f$  so that  $T = f[S_1]$  is normal to  $\Sigma$ . This means that in  $M_1$ , no component of  $\psi^{-1}[T - \Sigma]$  has a closure that meets a boundary sphere in more than one circle. Let  $C_1$  be a simple closed curve representing  $x_1$  which intersects the element  $S_1$  of the sphere basis  $\Sigma$  exactly once and does not meet any other element of  $\Sigma$ . We note that  $f[C_1]$  meets  $T$  exactly once and the word which  $f[C_1]$  represents is equivalent to  $w$  in  $F$  although it may not be reduced. Now choose a pair  $C$  and  $T$  satisfying the following conditions: the word which  $C$  represents is equal to  $w$  in  $F$ ;  $T$  is normal to  $\Sigma$ ;  $C \cap T$  has exactly one point; and  $C$  has the minimal number of intersections with the sphere basis  $\Sigma$  among such pairs. We claim that the word that  $C$  represents is reduced.

To see this, assume that it is not reduced. Then there is a subarc  $\gamma$  of  $C$  with end points  $p$  and  $q$  so that  $p$  and  $q$  are on some  $S_i$ ; no other point of  $\gamma$  is on an element of  $\Sigma$ ; and in  $M_1$  the closure of  $\psi^{-1}[\gamma]$  has its two boundary points on the same boundary sphere of  $M_1$ , either  $S_i^-$  or  $S_i^+$ . Without loss of generality, we will assume that it is the latter.

Now, the sphere  $T$  meets  $S_i$  in a collection of circles  $\{D_1, D_2, \dots, D_r\}$ . We will assume that neither  $p$  nor  $q$  is on any of these circles. If  $p$  and  $q$  are on the same component of  $S_i - \bigcup D_j$ , then  $\gamma$  does not meet  $T$  because  $\gamma \cup \gamma'$ , being trivial, must meet  $T$  in an even number of points. So we can obtain a primitive curve  $C'$  representing  $w$  meeting  $T$  in one point with fewer intersections with  $\Sigma$  as follows: let  $\gamma'$  be a path from  $p$  to  $q$  on  $S_i$  which is disjoint from  $T$ . Now, let  $C' = (C - \gamma) \cup \gamma'$  and alter  $C'$  by a homotopy to push  $\gamma'$  into the negative side of  $S_i$ . This contradicts the minimality of  $C$ .

Next, assume there is some  $D_j$  separating  $p$  and  $q$  on  $S_i$ . In this case, since  $T$  is normal to  $\Sigma$ , there is a component  $\Delta$  of  $T - \Sigma$  whose closure contains  $D_j$  so that  $\gamma$  meets  $\Delta$  in a point  $y$ . Clearly,  $y$  must be the unique point in  $T \cap C$ . We see that the normality of  $T$  also implies that if there is such a  $D_j$ , there is only one. Here, let  $\gamma'$  be a path on  $S_i$  from  $p$  to  $q$  that meets  $D_j$  in one point and no other points of  $T$ . As above, let  $C' = (C - \gamma) \cup \gamma'$  and alter  $C'$  by a homotopy to push  $\gamma'$  into the negative side of  $S_i$ . Again we see that  $C'$  meets  $T$  in one point and represents the word  $w$  in  $F$ . Moreover,  $C'$  has fewer points of intersection with  $\Sigma$  than does  $C$ .

Now, to prove the converse, suppose that there is a curve  $C$  which represents  $w$  with respect to  $\Sigma$  and a 2-sphere  $T$  such that  $C$  intersects  $T$  in precisely one point. According to [3],  $T$  is a primitive sphere, since the complement of  $T$  in  $M$  is connected. Thus we can find 2-spheres  $\Sigma_2, \dots, \Sigma_n$  so that  $\Sigma' = \{T, \Sigma_2, \dots, \Sigma_n\}$  is a sphere basis. Thus the sphere bases  $\Sigma$  and  $\Sigma'$  represent an automorphism  $\phi$  of  $F$ . Then label the sphere  $T$  as  $x_1$  and label the other spheres accordingly. Now  $\phi(w)$  has exactly one occurrence of  $x_1$  or exactly one occurrence of  $x_1^{-1}$  but not both. Thus  $w$  is primitive.  $\square$

Such a sphere  $T$  as described in this theorem is called a *detection sphere* for  $C$ . Let  $\Sigma = (S_1, S_2, \dots, S_n)$  be a sphere basis of  $M$ . We assume that any detection sphere is transverse to  $\Sigma$ . It is not true that every primitive curve has a detection sphere. If the primitive curve  $C$  has a detection sphere, then we say that  $C$  is a *proper primitive curve*. If  $C$  is a proper primitive curve, the *complexity of  $C$  relative to  $\Sigma$* , denoted  $\chi_\Sigma(C)$ , is the minimal number of components of  $T \cap \bigcup S_i$  among detection spheres  $T$  for  $C$ .

**Theorem 3.** *Let  $K$  be a core graph of some subgroup  $H$ . Let  $C$  be a proper primitive curve in  $M$  representing the primitive element  $w$  of  $H$ , and let  $\Sigma = (S_1, S_2, \dots, S_n)$  be a sphere basis. Assume that  $\chi_\Sigma(C) > 0$ . Then there is another sphere basis for  $M$ ,  $\Sigma'$  obtained from  $\Sigma$  by a composition  $\phi$  of a type I automorphism and a type II automorphism so that  $\chi_{\Sigma'}(C) < \chi_\Sigma(C)$  and  $w_{K'}(C) \leq w_K(C)$ , where  $K'$  is the graph obtained from  $K$  via  $\phi$ .*

*Proof.* Let  $T$  be the detection sphere realizing the complexity of  $C$ . Then, since  $\chi_\Sigma(C) > 0$ ,  $T$  must meet the sphere basis  $\Sigma$ . An endcap  $E$  is a subset of  $T$  which is homeomorphic to a closed disk so that the boundary of  $E$  lies on some  $S_i$  and no point on the interior of  $E$  lies on any element of  $\Sigma$ . Clearly,  $T$  has at least two endcaps, at least one of which does not meet  $C$ . Let us denote this endcap as  $E$ .

We consider  $E$  as a subset of  $M_1$ . Then the boundary of  $E$  meets either  $S_i^+$  or  $S_i^-$ . Without loss of generality, we assume it is the latter. Moreover,  $E$  separates some elements of  $\Sigma^\pm$  from the others, since otherwise  $E$  would be a trivial endcap as described in [10], in which case we could alter  $T$  by an isotopy to reduce  $\chi_\Sigma(C)$ .

Let  $j: S^2 \times I \rightarrow M_1$  be an embedding describing a collar of  $S_i^-$  so that  $j[S^2 \times \{0\}] = S_i^-$ . We let  $J$  be the image of  $j$  and let  $J_1$  be  $j[S^2 \times \{1\}]$ . We take this collar small enough so that the following conditions hold:  $C$  and  $T$  are transverse to  $J_1$ ;  $E \cap J_1$  has one component;  $T \cap J$  is a collection of annuli each of which has one boundary component on each of  $S_i^-$  and  $J_1$ ; and  $C \cap J$  consists of a collection of subarcs of  $C$  each of which meet  $S_i^-$ .

Now  $E$  meets  $J_1$  in a circle. We let  $\Delta$  be the disk on  $J_1$  bounded by this circle so that  $E \cup \Delta$  separates  $S_i^+$  from  $S_i^-$ . We let  $S_*$  be the sphere  $(E - J) \cup \Delta$ . Clearly,  $S_*$  is not transverse to  $T$ , but we can push  $S_*$  slightly so that  $S_*$  is disjoint from  $E$ . Now  $S_*$  is a sphere embedded in  $M_1$  and every arc of  $C$  that meets  $S_*$  also meets  $S_i^-$ .

Using a type I automorphism, we can assume that  $i = 1$  and that there are  $k$  and  $l$  so that the following assertions hold: for all  $i \in \{1, \dots, k\}$ ,  $S_i^-$  is in the inside of  $S_*$  and  $S_i^+$  is in the outside of  $S_*$ ; for all  $i \in \{k+1, \dots, l\}$ , both  $S_i^-$  and  $S_i^+$  are in the inside of  $S_*$ ; and for all  $i \in \{l+1, \dots, n\}$ , both  $S_i^-$  and  $S_i^+$  are in the outside of  $S_*$ . We note that  $S_*$  is stingy.

Now we do a Whitehead move corresponding to  $\phi_{k,l}$  to obtain the basis  $\Sigma' = (S_*, S_2, \dots, S_n)$  and the graph  $K'$ . Since  $S_*$  does not meet  $E$ , we have  $\chi_{\Sigma'}(C) < \chi_\Sigma(C)$ . Moreover, by Theorem 1, we know that  $w_{K'}(C) \leq w_K(C)$ .  $\square$

**Theorem 4.** *Let  $K$  be a core graph of some subgroup  $H$ . Let  $C$  be a proper primitive curve in  $M$  representing the primitive element  $w$  of  $H$ , and let  $\Sigma = (S_1, S_2, \dots, S_n)$  be a sphere basis. Assume that  $\chi_\Sigma(C) = 0$ . Then there is another proper primitive curve  $C_1$  so that  $\chi_\Sigma(C_1) = 0$  and the deficiency  $\delta_K(C_1) = 0$ .*

*Proof.* We choose a proper primitive curve  $C_1$  so that  $\chi_\Sigma(C_1) = 0$  and  $\delta_K(C_1)$  is as small as possible. By way of contradiction, assume  $\delta_K(C_1) > 0$ . Then there are two intervals  $I_1$  and  $I_2$  of  $C_1$  so that  $\gamma(I_1) = \gamma(I_2)$ . We assume that this edge of  $K$  is labeled  $x_i$  so that there are points  $p_1 \in S_i \cap I_1$  and  $p_2 \in S_i \cap I_2$ . These points separate  $C_1$  into two intervals  $D_1$  and  $D_2$ .

Since  $C_1$  is a proper primitive curve and  $\chi_\Sigma(C_1) = 0$ , there is some detection sphere  $T$  so that  $T$  is disjoint from each element of  $\Sigma$  and  $T$  meets  $C_1$  in exactly one point, which, without loss of generality, we assume is on  $D_1$ . Since  $T$  does not meet  $S_i$ , there is a small interval  $D_3$  on  $S_i$  from  $p_1$  to  $p_2$  which is disjoint from  $T$ . We define the curve  $C_2 = D_1 \cup D_3$  and alter  $C_2$  by a homotopy near  $D_3$  so that it is transverse to  $S_i$ . Now  $C_2$  meets  $T$  in exactly one point, so it is a proper primitive curve. Moreover, the construction ensures that there is a corresponding map  $\gamma_2: C_2 \rightarrow K$ . Lastly,



we note that  $\delta_K(C_2) < \delta_K(C_1)$ . This contradicts the minimality of  $C_1$ . Therefore,  $\delta_K(C_1) = 0$ .  $\square$

## 7 The algorithm

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a basis for the free group  $F_n$ . In [10], Whitehead gave an algorithm which determines whether or not a word  $w$  in the alphabet  $X$  is a primitive element of  $F_n$ . Significant improvements in the complexity of this algorithm have been given in [6].

Let  $K$  be a core graph of the subgroup  $H$  of  $F$ . Determining whether  $H$  contains a primitive element is equivalent to determining whether there is a path in  $K$  that represents a primitive element. Such a path will be called a *primitive path*. A path is *simple* if it uses no edge more than once. Clearly, there is an algorithm to determine all the simple paths in  $K$ . Since there is an algorithm which determines whether a given path is a primitive path, it is possible to determine whether  $K$  has a simple primitive path.

Given the graph  $K$  with  $N$  edges, a graph  $K'$  is *one step away from  $K$*  if there is a graph  $K''$  so that the following conditions hold:  $K'$  is subgraph of  $K''$ ;  $K'$  has no more than  $N$  edges; and  $K''$  is obtained from  $K$  by performing a type I and a type II automorphism on  $K$ . We see that if  $K'$  is one step away from  $K$  and  $K'$  contains a primitive path, then so does  $K$ . (We point out that it is possible that  $K$  contains a primitive path while  $K'$  does not.) Clearly, given  $K$  there is an algorithm which produces all graphs which are one step away from  $K$ . We say that  $K'$  is  $M$  steps away from  $K$  if there is a sequence of graphs  $K = K_0, K_1, \dots, K_M = K'$  so that each  $K_j$  is one step away from  $K_{j-1}$  and  $M$  is minimal in this regard.

Given  $K$ , let

$$\Omega_K = \{K' \mid \text{there is some } M \text{ so that } K' \text{ is } M \text{ steps away from } K\}.$$

Since each  $K' \in \Omega_K$  has no more than  $N$  edges,  $\Omega_K$  is a finite class of graphs. Also, if some  $K' \in \Omega_K$  contains a primitive path, then so does  $K$ . Moreover,  $K' \in \Omega_K$  if and only if  $\Omega_{K'} \subset \Omega_K$ .

Now there is an algorithm which produces all of the elements of  $\Omega_K$ . So there is an algorithm which can determine whether there is an element of  $\Omega_K$  that has a simple primitive path. This is the algorithm that determines whether  $K$  has a primitive path as is made clear in the next theorem.

**Theorem 5.**  *$K$  has a primitive path if and only if there is some  $K'$  in  $\Omega_K$  that has a simple primitive path.*

*Proof.* Clearly, if there is some  $K'$  in  $\Omega_K$  that has a simple primitive path, then  $K$  has a primitive path.

Conversely, assume that  $K$  has a primitive path  $\alpha$ . Let  $C$  be a proper primitive curve representing  $\alpha$  embedded in  $M$  with sphere basis  $\Sigma$  and detection sphere  $T$  so

that  $\chi_\Sigma(C)$  is as small as possible. We proceed by induction on  $\chi_\Sigma(C)$ . If  $\chi_\Sigma(C) = 0$ , then by Theorem 3,  $K$  contains a simple primitive path.

If  $\chi_\Sigma(C) > 0$ , then by Theorem 2, there is another sphere basis  $\Sigma'$  for  $M$ , obtained from  $\Sigma$  by a composition  $\phi$  of a type I automorphism and a type II automorphism so that  $\chi_{\Sigma'}(C) < \chi_\Sigma(C)$  and  $w_{K''}(C) \leq w_K(C)$ , where  $K''$  is the graph obtained from  $K$  via  $\phi$ . We let  $K'''$  be the support of  $C$  in  $K''$ . Clearly  $K''' \in \Omega_K$ . Since  $\chi_{\Sigma'}(C) < \chi_\Sigma(C)$ , there is some  $K' \in \Omega_{K'''}$  that contains a simple primitive path. But  $\Omega_{K'''} \subset \Omega_K$ , so  $K' \in \Omega_K$ .  $\square$

We end this note with a brief discussion of the complexity of this algorithm. Let  $K$  be a graph with  $n$  edges whose rank is  $r$ . If  $\alpha$  is a simple path in  $K$ , clearly  $\alpha$  has no more than  $n$  edges. In [6], it is shown that it can be determined whether or not  $\alpha$  represents a primitive element in linear time in  $n$ . Unfortunately, the number of simple paths in  $K$  is of the order of  $(2r)^r$ . Lastly, we note that for the algorithm to conclude that  $K$  does not have a primitive path, we need to check every simple path not only in  $K$ , but in each graph in  $\Omega_K$ , which although finite, can be quite large. In conclusion, we see that the algorithm described here is very time consuming. While some efficiencies in this algorithm may be obtainable, we think that if ever a reasonably practical algorithm is found, it will be considerably different.

**Acknowledgement.** The authors would like to thank the referee for numerous suggestions which led to significant improvements to this paper. In particular, the referee informed us about the existence of the very interesting preprint [9].

## References

- [1] A. Clark and R. Goldstein. On a method of J. H. C. Whitehead. *Comm. Algebra* **34** (2006), 1615–1623.
- [2] S. M. Gersten. Reducible diagrams and equations over groups. In *Essays in group theory*, Math. Sci. Res. Inst. Publ. 8 (Springer, 1987), pp. 15–73.
- [3] R. Z. Goldstein. The length and thickness of words in a free group. *Proc. Amer. Math. Soc.* **127** (1999), 2857–2863.
- [4] R. Z. Goldstein and E. C. Turner. Monomorphisms of finitely generated free groups have finitely generated equalizers. *Invent. Math.* **82** (1985), 283–289.
- [5] R. C. Lyndon and P. E. Schupp. *Combinatorial group theory* (Springer-Verlag, 1977).
- [6] A. Roig, E. Ventura and P. Weil. On the complexity of the Whitehead minimization problem. *Internat. J. Algebra Comput.* **17** (2007), 1611–1634.
- [7] J. R. Stallings. Topology of finite graphs. *Invent. Math.* **71** (1983), 551–565.
- [8] J. R. Stallings. Whitehead graphs on handlebodies. In *Geometric group theory down under (Canberra, 1996)* (de Gruyter, 1999), pp. 317–330.
- [9] P. V. Silva and P. Weil. Automorphic orbits in free groups: words versus subgroups. Preprint (2009). <http://arxiv.org/abs/0809.4386>.
- [10] J. H. C. Whitehead. On certain sets of elements in a free group. *Proc. London Math. Soc.* (2) **41** (1936), 48–56.
- [11] J. H. C. Whitehead. On equivalent sets of elements in a free group. *Ann. of Math.* (2) **37** (1936), 782–800.
- [12] World of Groups website. <http://www.grouptheory.info/>

Received 3 February, 2009; revised 21 August, 2009

A. Clifford, Department of Mathematics and Statistics, The College of New Jersey, P.O. Box 7718, Ewing, NJ 08628-0718, U.S.A.  
E-mail: cliffan@tcnj.edu

R. Z. Goldstein, Department of Mathematics, 1400 Washington Avenue, Albany, NY 12222, U.S.A.  
E-mail: richardg@csc.albany.edu