FUNDAMENTA
MATHEMATICAE
199 (2008)

Reidemeister conjugacy for finitely generated free fundamental groups

by

Evelyn L. Hart (Hamilton, NY)

*Dedicated to the memory of mathematical friends
Janet Andersen and Laura Sanchis*

Abstract. Let X be a space with the homotopy type of a bouquet of k circles, and let $f : X \to X$ be a map. In certain cases, algebraic techniques can be used to calculate $N(f)$, the Nielsen number of f, which is a homotopy invariant lower bound on the number of fixed points for maps homotopic to f. Given two fixed points of f, x and y, and their corresponding group elements, W_x and W_y, the fixed points are Nielsen equivalent if and only if there is a solution $z \in \pi_1(X)$ to the equation $z = W_y^{-1} f_\sharp(z) W_x$. The Nielsen number is the number of equivalence classes that have nonzero fixed point index.

A variety of methods for determining the Nielsen classes, each with their own restrictions on the map f, have been developed by Wagner, Kim, and (when the fundamental group is free on two generators) by Kim and Yi. In order to describe many of these methods with a common terminology, we introduce new definitions that describe the types of bounds on $|z|$ that can occur. The best directions for future research become clear when this new nomenclature is used.

To illustrate the new concepts, we extend Wagner's ideas, regarding W-characteristic maps and maps with remnant, to two new classes of maps that have only partial remnant. We prove that for these classes of maps Wagner's algorithm will find almost all Nielsen equivalences, and the algorithm is extended to find all Nielsen equivalences. The proof that our algorithm does find the Nielsen number is complex even though these two classes of maps are restrictive.

For our classes of maps (MRN maps and 2C3 maps), the number of possible solutions z is at most 11 for MRN maps and 14 for 2C3 maps. In addition, the length of any solution is at most three for MRN maps and four for 2C3 maps. This makes a computer search reasonable. Many examples are included.

1. Introduction. Let X be a space that has the homotopy type of a bouquet of circles, and let $f : X \to X$ be a map. Our goal is to estimate

2000 *Mathematics Subject Classification*: Primary 55M20.
Key words and phrases: fixed point, Nielsen number, free group, remnant.
Supported in part by a grant from the Colgate University Research Council.

$\min[f] = \min\{|\mathrm{Fix}(g)| : g \sim f\}$. For some spaces that have a free fundamental group on two generators, there are algorithms for calculating $\min[f]$. For most situations we seek to estimate $\min[f]$ with the Nielsen number, $N[f]$. The Nielsen number of f is a homotopy invariant lower bound for $\min[f]$. For a map that is homotopic to a homeomorphism, $N[f] = \min[f]$. Otherwise, it is possible for $\min[f] - N[f]$ to be arbitrarily large. A summary of results regarding Wecken properties (the sharpness of the lower bound $N[f]$) is presented in [11]. Standard references for an introduction to Nielsen fixed point theory are [1] and [9].

For maps that are homotopic to a homeomorphism on surfaces with negative Euler characteristic, Kelly [12] uses geometric results of Bestvina and Handel [2] to produce an algorithm for calculating $N[f]$. Recently in [3] Bogopolski et al. proved that free-by-cyclic groups have solvable conjugacy problems, but it is not clear how to apply this result to our situation.

We concentrate here on results that apply to free fundamental groups of any finite rank and algorithms that allow us to find easily the Nielsen number of a self-map on the associated space. There are significant results for the special case in which the fundamental group is free with rank two. For the disc with two holes, Kelly [10] presents a geometric algorithm for computing $\min[f]$. In the same paper he proves that $\min[f] - N[f]$ can be arbitrarily large on this space. Wagner produces examples of this in [17]. Llibre and Nunes [15] provide an algorithm for finding $\min[f]$ when the space is the figure eight. Recently, Kim [14] extended Yi's work from [18] so that there is now an algebraic algorithm for finding $N[f]$ on any space for which $\pi_1(X)$ is free of rank two. They use mutants to replace f with a map homotopic to f that has remnant. Then Wagner's algorithm can be used to find $N[f]$. (We discuss remnant and Wagner's algorithm below.) However, all of these results are algorithmic only on free fundamental groups of rank two.

In the study of Nielsen periodic point theory, there is interest in spaces that have the homotopy type of a wedge of circles (see [7]). In order to calculate the Nielsen periodic point numbers, information about the Nielsen classes of the fixed points for each iterate of the original map is needed.

To calculate the Nielsen number of a map f, we must find the number of orbits of a group action that have nonzero fixed point index. The group action is the Reidemeister action of $\pi_1(X)$ on $\pi_1(X)$. The element z acts on the element α to produce

$$z \circ \alpha = f_\sharp(z)\alpha z^{-1}.$$

Two words that are in the same orbit are said to be Reidemeister equivalent. The difficulty in finding $N[f]$ arises when we have two group elements α and β, each representing a fixed point of f, and we must determine whether they are in the same Reidemeister orbit. That is, we must search for a word $z \in \pi_1(X)$ so that z acts on α to produce β. We study here cases in which there are upper bounds on the length of the z's that must be considered. When such bounds exist and are sufficiently small, a computer search can be used to determine Reidemeister equivalence and the Nielsen number.

Suppose that $\pi_1(X)$ is a free group of rank k. We begin, as did Wagner in [16], by replacing X by the wedge of k circles and by replacing f with a new map of a particular form that induces the same homomorphism on $\pi_1(X)$. We call the new space and map X and f. For this new map, the fixed points correspond to certain initial subwords of the $f(a_i)$ for the generators a_i of $\pi_1(X)$. These subwords can be determined, along with the index of the corresponding fixed point, by using the Fox calculus as in [4]. We call the resulting element of $\mathbb{Z}[\pi_1(X)]$ the Fox trace for f. Once the terms of the Fox trace are collected into Reidemeister orbits, the resulting element of the free $\mathbb{Z}$-module generated by the Reidemeister orbits is equal to the Reidemeister trace (or the generalized Lefschetz number). See, for example, [5]. The Nielsen number is the number of terms with nonzero coefficient in the Reidemeister trace. In [16], Wagner describes an equivalent method for finding the Fox trace. For each $x \in \mathrm{Fix}(f)$, we use W_x to denote the element of $\pi_1(X)$ that represents x in the Fox trace.

In [16] Wagner makes a breakthrough when she concentrates on algebraic techniques for determining Reidemeister equivalence only between terms of the Fox trace (instead of between any two elements of $\pi_1(X)$). She provides a method for finding many Reidemeister equivalences between terms in the Fox trace. When this method finds all such Reidemeister equivalences for a homomorphism $f_\sharp$, the homomorphism is called W-characteristic, and the method becomes Wagner's algorithm. But which maps are W-characteristic? This is difficult to determine in general.

Wagner defines a property for maps called having remnant, which we repeat in Definition 5.1. Roughly, a map has remnant if it has limitations on the amount of cancellation in products of the form $f_\sharp(a_i)f_\sharp(a_j)$, where a_i and a_j are generators of $\pi_1(X)$. The first proof of the following theorem is in [16], and simplified proofs appear in [6] and [13].

THEOREM 1.1 (Wagner's Theorem). *Any map with remnant is W-characteristic.*

It is the proof of this theorem that has been the inspiration for the work of Kim in [13] and the results in this paper. Wagner includes in [16] an example of a map that is W-characteristic and does not have remnant.

In [17], Wagner defines a simple map and provides a formula for $N(f)$ for any map on a hyperbolic surface that is both simple and W-characteristic. This formula will become more useful as more classes of W-characteristic maps are identified. For now, if a map does not have remnant then the only way to determine whether it is W-characteristic is to calculate its Nielsen number by some other means and compare the result with Wagner's method. Potentially more useful is Corollary 2.3 from [17], in which Wagner presents an upper bound for $N(f)$ when f is simple but might not be W-characteristic.

Let $x, y \in \mathrm{Fix}(f)$ be represented, respectively, by the words W_x and W_y in the Fox trace. If W_x and W_y are Reidemeister equivalent (and thus the points x and y are Nielsen equivalent), we define the algebraic distance between them, $d(x,y) = d(W_x, W_y)$, to be the minimum length of all words z for which $z \circ W_x = W_y$. We define $\mathrm{far}(f_\sharp)$ to be the maximum such distance over all pairs of Nielsen equivalent fixed points.

The work of Kim in [13] on maps with bounded solution length can be expressed using these new terms, as we discuss at the end of Section 3.

As a result of Wagner's work we conclude in Theorem 3.3 that many maps with remnant have $\mathrm{far}(f_\sharp) \leq 2$. For these maps we have an even stronger result. Given any two terms of the Fox trace, one need only consider four possible solutions z to determine whether the words are Reidemeister equivalent. In general, a map f with remnant has $\mathrm{far}(f_\sharp) \leq |\mathrm{Fix}(f)| + 1$.

A class of maps called MRN maps, in the restricted setting of a fundamental group of rank two, is defined in [6]. These maps have partial remnant and have many restrictions on the cancellation of words in the image of $f_\sharp$. It is announced in [6] that MRN maps have $\mathrm{far}(f_\sharp) \leq 3$. As with Theorem 3.3, it turns out that one need only check 11 words rather than the 53 words of length at most three. See Theorem 5.9. The original plan was to include the proof of this result in the present paper.

Instead, we introduce a larger class of maps defined for any rank fundamental group, maps of type 2C3, and prove that these maps have $\mathrm{far}(f_\sharp) \leq 4$. In fact, for a given pair of terms from the Fox trace, there are only 14 words z to check in order to determine whether the terms are Reidemeister equivalent. See Theorems 5.12 and 5.13. The techniques needed for the proof about MRN maps are the same as the techniques used in the proofs we provide. We are able to describe precisely the situations in which Wagner's method will fail to find an existing Reidemeister equivalence for an MRN map or a 2C3 map.

There are certainly many maps that do not fit into the classes of maps we have studied. Consider Example 6.2, where for any $n \in \mathbb{N}$ there is a map with no remnant for which $\mathrm{far}(f_\sharp) \geq n$.

Having an upper bound for $\mathrm{far}(f_\sharp)$ is useful, but Wagner's results are deeper than this. In fact, for any $n \in \mathbb{N}$ there is a map with remnant for which $\mathrm{far}(f_\sharp) \geq n$. See Example 6.1. Yet Wagner's algorithm can find all Reidemeister equivalences easily for maps with remnant. If we study only

$\mathrm{far}(f_\sharp)$, we miss the opportunity to extend Wagner's techniques in all its generality.

We define (as we did in [6]) a map to be n-characteristic if for any two Nielsen equivalent fixed points x and y there is a sequence of fixed points $x_1, \ldots, x_t$ such that $x = x_1$, $y = x_t$, and for each $i = 1, \ldots, t-1$ we have x_i Nielsen equivalent to x_{i+1} and $d(x_i, x_{i+1}) \leq n$.

If a map is n-characteristic, then to determine the Nielsen classes one need only act on each word that represents a fixed point by each z of length at most n and keep track of equivalences as they are found.

In Theorems 4.2, 5.9, and 5.12, respectively, we note that any W-characteristic map is 2-characteristic, that any MRN map is 2-characteristic, and that any map of type 2C3 is 3-characteristic. As mentioned above for $\mathrm{far}(f_\sharp)$, we prove in the first two theorems that we need only check a limited number of z's rather than all that have length at most 2.

We have determined some properties that affect whether a map can be proven to be n-characteristic, and we give these properties names in Section 5.1. The minimum of the length of the remnant of $f_\sharp(a_i)$ is crucial. Cyclical cancellation is restrictive but makes proofs much easier to manage. A map has second order stable remnant if the remnant as defined by Wagner remains uncancelled in triples $f_\sharp(a_i)f_\sharp(a_j)f_\sharp(a_k)$. This property is crucial when $f_\sharp(a_j)$ does not have remnant.

While this paper was in preparation, this author and Kim [8] defined maps with k-remnant and proved that such maps are $(k+1)$-characteristic. A map has k-remnant if, roughly, there is limited cancellation in the product of the images of two words with length k.

In the future, there will be many more results involving maps that do or do not have these properties. Perhaps significant results will result from a combination of Kim's bounded solution length arguments and the techniques used here to prove that maps of type 2C3 are 3-characteristic.

We prove in Example 6.3 that there is a map of type 2C3 that has $\mathrm{far}(f_\sharp) = 4$, and in Example 6.4 we prove that there is a map of type 2C3 that is not 2-characteristic. Thus the bounds in Theorem 5.12 are sharp. We cannot do better. Example 6.4 is also an example of a map that has type 2C3 and is not W-characteristic.

The paper is organized as follows. Section 2 contains an overview of the necessary background. In Section 3 we provide the new definitions of the algebraic distance between fixed points and $\mathrm{far}(f_\sharp)$. The concept of an n-characteristic homomorphism is defined in Section 4. Section 5.1 contains new definitions of minimum remnant length, cyclical cancellation, and second order stable remnant. The rest of Section 5 contains definitions of MRN maps and 2C3 maps and the statements of the theorems regarding their properties as described above. The examples are worked out in Section 6. In Section 7 we prove Theorems 5.12 and 5.13.

Thanks to Bob Brown, Seung Won Kim, and Ed Keppelmann for helpful conversations. The referee deserves thanks for careful reading and many useful comments.

2. Preliminaries. A detailed description of the necessary background can be found in [5] and [6], and the reader is encouraged to refer to these sources for details that are omitted here.

2.1. *The standard form of the self-map.* We begin with a self-map on a surface with boundary. Let φ be the induced endomorphism on the fundamental group, which is a free group generated by $a_1, \ldots, a_k$. As in [16], we replace the surface with a bouquet of k circles, X, and we replace the map with a map $f : X \to X$ such that $f_\sharp = \varphi$. We use Wagner's precise definition of this new map f (see [16] or [6] for details). Thus the base point x_0 (the wedge point) is fixed by f, and all fixed points of f are isolated. In addition, there is a bijection between $\mathrm{Fix}(f)\backslash\{x_0\}$ and the occurrences of a_i and a_i^{-1} in the word $f_\sharp(a_i)$ for each i.

2.2. *The Fox trace and the Reidemeister action.* The *Fox trace*, previously called the unreduced Reidemeister trace, is the element of $\mathbb{Z}[\pi_1(X)]$ given by

$$\mathrm{FT}(f_\sharp) = 1 - \sum_j \frac{\partial f_\sharp(a_j)}{\partial a_j},$$

where the a_j are the generators of $\pi_1(X)$ and the derivatives use the Fox calculus. See [4]. We do not even cancel identical terms that have opposite coefficients. The terms are labeled W_i in order, and these are called the terms of the Fox trace. We set $\mathcal{F}$ to be the sequence of these terms. Note that we may have two terms being the same group element. The terms are exactly the W's that Wagner defines in [16], which can be verified by using the so called product rule from the Fox calculus.

The map f has exactly one fixed point for each term in $\mathcal{F}$. We name the fixed points x_0 (the wedge point), $x_1, \ldots, x_m$ following each loop in order, and we see that x_i is represented by W_i in the Fox trace. See Example 6.1. When convenient, we use W_x to denote the term in the Fox trace that represents the fixed point x.

The *Reidemeister action* of $\pi_1(X)$ on $\pi_1(X)$ is defined as follows. For any $z, \alpha \in \pi_1(X)$, z acts on α by

$$z \circ \alpha = f_\sharp(z)\alpha z^{-1}.$$

The orbits of the action are *Reidemeister classes*, and two group elements are *Reidemeister equivalent* if they are in the same Reidemeister class. We are interested in Reidemeister equivalences only between the terms of the Fox trace, so in some sense we are interested in the action of $\pi_1(X)$ on $\mathcal{F}$.

Two fixed points $x, y \in \mathrm{Fix}(f)$ are *Nielsen equivalent* (in the same Nielsen class) if and only if the corresponding terms, W_x and W_y, are Reidemeister equivalent. That is, $x \sim y$ if and only if there is a solution $z \in \pi_1(X)$ to the equation

$$z = W_y^{-1} f_\sharp(z)\, W_x. \tag{1}$$

The *Nielsen number* $N[f]$ is the number of Nielsen classes with nonzero index.

3. Algebraic distance and $\mathrm{far}(f_\sharp)$

DEFINITION 3.1. Let $x, y \in \mathrm{Fix}(f)$ be Nielsen equivalent fixed points. Then the *algebraic distance* between x and y is

$$d(x,y) = \min\{|z| : z \in \pi_1(X) \text{ and } z = W_y^{-1} f_\sharp(z) W_x\}.$$

When convenient, we extend this definition to terms in the Fox trace so that $d(W_x, W_y) = d(x, y)$.

When the algebraic distances between related fixed points are known to be bounded by n, then all Nielsen equivalences can be discovered by acting on each W_x repeatedly using all words $z \in \pi_1(X)$ with $|z| \le n$. The next definition makes this idea precise.

DEFINITION 3.2 (The algebraic distance to the farthest fixed point). Let $x \in \mathrm{Fix}(f)$. We define $\mathrm{far}(x)$ and $\mathrm{far}(f_\sharp)$ by

$$\mathrm{far}(x) = \max\{d(x,y) : y \in \mathrm{Fix}(f)\}, \qquad \mathrm{far}(f_\sharp) = \max\{\mathrm{far}(x) : x \in \mathrm{Fix}(f)\}.$$

The Nielsen class containing x equals the set of fixed points y for which $W_y \in \{f_\sharp(z^{-1}) W_x z : z \in \pi_1(X) \text{ and } |z| \le \mathrm{far}(x)\}$. Thus if $\mathrm{far}(x) \le n$ the entire Nielsen class containing x can be determined by repeatedly acting on W_x using all $z \in \pi_1(X)$ for which $|z| \le n$ and comparing the results with the terms in $\mathcal{F}$. As we mentioned in the introduction, Examples 6.1 and 6.2 both show that maps can have arbitrarily large values of $\mathrm{far}(f_\sharp)$.

We can now restate some of the known results for Reidemeister equivalence using $\mathrm{far}(f_\sharp)$. See Section 5 for the definition of remnant.

THEOREM 3.3. *Let f be a map.*

1. *Suppose that f has remnant such that for each generator a_i the length of the remnant of $f_\sharp(a_i)$ is at least two. Then* $\mathrm{far}(f_\sharp) \le 2$.
2. *In fact, if f has remnant and for each a_i the remnant of $f_\sharp(a_i)$ is not equal to a_i, then* $\mathrm{far}(f_\sharp) \le 2$.

3. *Suppose that* $p, q \in \mathrm{Fix}(f)$. *Even if* f *does not have remnant, if Wagner's method finds a Reidemeister equivalence between* W_p *and* W_q, *then* $d(p,q) \leq |\mathrm{Fix}(f)| + 1$.
4. *Thus, if* f *has remnant we can conclude that* $\mathrm{far}(f_\sharp) \leq |\mathrm{Fix}(f)| + 1$.

Outline of proof of Theorem 3.3. A careful reading of the proof of Theorem 1.1 (see [6], [13], or [16]) will be sufficient to prove the first two parts of the theorem.

Wagner's method finds a Reidemeister equivalence when there is, for some k, a sequence of fixed points $x_1, \ldots, x_k$ connecting W_p and W_q as described, for example, in [16] and [6]. The solution z that results has length at most $k+1$. The length $k+1$ can occur only when $\overline{W}_1 = \overline{W}_2$ and $W_i = \overline{W}_{i+1}$ for $i = 2, \ldots, k-1$. But these occur only if $d(x_1, x_2) = 2$ and $d(x_i, x_{i+1}) = 1$ for each i. These short solutions when concatenated become the z that we need to connect p and q. Any other chain of equivalences results in a shorter total length of z. ■

3.1. *Bounded solution length.* In [13], Kim extends the ideas of Wagner in a different way. Given a map f that does not have remnant, he first uses Wagner's method to find as many Reidemeister equivalences as possible with that method. Then, for any two fixed points x and y that Wagner's method does not find to be equivalent, he considers the possible lengths of potential solutions to equation (1).

Definition 3.4 (Kim). Let $x, y \in \mathrm{Fix}(f)$ be such that Wagner's method does not find an equivalence between them. Then x and y have *bounded solution length* if there is an $n \in \mathbb{Z}$ such that for every $z \in \pi_1(X)$ for which $|z| > n$ we have $|W_y^{-1} f_\sharp(z) W_x| > |z|$. When x and y have bounded solution length, the smallest such n is the *solution bound* for x and y, written $n_{x,y}$.

If for any two fixed points of f either Wagner's method finds an equivalence between them or else they have bounded solution length, then f *has bounded solution length* with *solution bound*, SB, given by the maximum of the solution bounds for the pairs of fixed points.

Combining Theorem 3.3 with this definition results in the following.

Theorem 3.5. *Any map* f *that has bounded solution length with* $\mathrm{SB} = s$ *has* $\mathrm{far}(f_\sharp) \leq \max\{|\mathrm{Fix}(f)| + 1, s\}$.

Kim goes on to prove that several classes of maps have SB ≤ 16 and announces that (by a tedious proof) the upper bound can be reduced to 10. These upper bounds are too large to be useful in computer calculations in general, but the ideas in the proofs will surely contribute to better bounds in the future. A map f that has $\mathrm{far}(f_\sharp)$ bounded does not necessarily have bounded solution length because the existence of a short solution z for a pair of fixed points does not preclude the existence of arbitrarily long solutions to the same equation. Kim relates his results to ours for MRN maps, as we will mention in Section 5.2.

4. n-characteristic homomorphisms. In Example 6.1, we present for each $n \in \mathbb{Z}$ an endomorphism $f_\sharp$ that is W-characteristic but has $\mathrm{far}(f_\sharp) > n$. Yet Wagner's algorithm can quickly determine all of the Reidemeister equivalences and the Nielsen number. Thus the qualities of W-characteristic endomorphisms that we would like to extend to other settings cannot be described in terms of bounds on $\mathrm{far}(f_\sharp)$. For this reason we present another concept, first introduced in [6].

Definition 4.1. The endomorphism $f_\sharp$ is *n-characteristic* if for any two Nielsen related fixed points x and y there is a sequence $(q_1, \ldots, q_r)$ of Nielsen related fixed points such that $x = q_1$, $y = q_r$ and for $i = 1, \ldots, r-1$ we have $d(q_i, q_{i+1}) \leq n$.

If $f_\sharp$ is n-characteristic then the Nielsen classes can be determined in a finite number of steps. Suppose that $\pi_1(X)$ is generated by $a_1, \ldots, a_k$ and that $f_\sharp$ is n-characteristic. For a given fixed point $x = q_1$ on the a_j loop, we would like to find all sequences of fixed points as in the definition above. If we act on W_x repeatedly using each of the many elements $z \in \pi_1(X)$ with $|z| \leq n$, we will find all possible elements q_2.

The following theorem, which is Remark 4.9 from [6], is the key to extending Wagner's results. It states that the number of calculations needed to find the Nielsen classes and Reidemeister structure for W-characteristic maps is much smaller than the number of calculations needed for an arbitrary 2-characteristic map. We seek other classes of maps for which the number of calculations needed is known to be relatively small.

Theorem 4.2. *Any W-characteristic homomorphism is also 2-characteristic. Given fixed points x on the a_i loop and y on the a_j loop, $d(x, y) \leq 2$ if and only if the set $\{1, a_i^{-1}, a_j, a_j a_i^{-1}\}$ contains a solution z to* (1). *On the other hand, Example* 6 *of* [6] *proves that a 2-characteristic map need not be W-characteristic.*

Proof. Given an endomorphism $f_\sharp$ that is W-characteristic, suppose that W_i and W_j are terms from the Fox trace that come from partial derivatives involving generators a_i and a_j, respectively. Then (using ideas introduced by Wagner) we have three cases. If $W_i = W_j$, then W_i and W_k are equivalent using $z = 1$. If $W_i = \overline{W}_k$, then $W_i = \overline{W}_k = f_\sharp(a_k^{-1}) W_k\, a_k$. Thus $z = a_k^{-1}$ is a solution to (1). If $\overline{W}_i = \overline{W}_k$, then

$$f_\sharp(a_i^{-1})W_i\, a_i = f_\sharp(a_k^{-1})W_k a_k \quad \text{and} \quad W_i = f_\sharp(a_i a_k^{-1})W_k\, a_k a_i^{-1}.$$

Thus in this case $z = a_i a_k^{-1}$ is a solution to (1). ∎

In Example 4.6 of [13], Kim proves that a W-characteristic map need not have bounded solution length. But, by combining Theorem 4.2 with the definition of bounded solution length, Kim proved the following theorem.

THEOREM 4.3 (Kim, Thm. 4.5 from [13]). *If f has bounded solution length with* SB $= s$, *then f is n-characteristic with* $n = \max\{2, s\}$.

5. New concepts and two classes of maps. Let $\pi_1(X) = \langle a_1, \ldots, a_k\rangle$. Let $f_\sharp : \pi_1(X) \to \pi_1(X)$ be any homomorphism, and let $\mathcal{G}^\pm$ be $\{a_1, \ldots, a_k\} \cup \{a_1^{-1}, \ldots, a_k^{-1}\}$.

We use the notation of [6]. Then for each word $f_\sharp(a_i)$ in $\pi_1(X)$ we consider the initial subwords of $f_\sharp(a_i)$ that cancel when we reduce each of the products in the set $\{f_\sharp(a_j)f_\sharp(a_i) : j = 1, \ldots, k\} \cup \{f_\sharp(a_j^{-1})f_\sharp(a_i) : j \neq i\}$. We let U_i denote the longest of all these initial segments. Similarly, we consider the terminal subwords of $f_\sharp(a_i)$ that cancel when we reduce each of the products in the set $\{f_\sharp(a_i)f_\sharp(a_j) : j = 1, \ldots, k\} \cup \{f_\sharp(a_i)f_\sharp(a_j^{-1}) : j \neq i\}$. The longest of these terminal segments is denoted by V_i.

DEFINITION 5.1 (Wagner). If $|U_i| + |V_i| < |f_\sharp(a_i)|$, then $f_\sharp(a_i)$ *has remnant*, and *the remnant of* $f_\sharp(a_i)$ is $\overline{X}_i = U_i^{-1} f_\sharp(a_i) V_i^{-1}$. When each word $f_\sharp(a_i)$ has remnant, then the homomorphism $f_\sharp$ *has remnant*.

Note that if $f_\sharp(a_i)$ does not have remnant, then it is possible to have $|U_i| + |V_i| > |f_\sharp(a_i)|$, and in this case U_i and V_i overlap in the word $f_\sharp(a_i)$.

Let $z = z_1 \cdots z_r$ with $z_i \in \mathcal{G}^\pm$. Then $z_i = a_j$ or else $z_i = a_j^{-1}$ for some j. If $z_i = a_j$ (respectively, $z_i = a_j^{-1}$), we say that $Q_i = U_j$ ($Q_i = V_j^{-1}$) and $S_i = V_j$ ($S_i = U_j^{-1}$). Also, if $f_\sharp(a_j)$ has remnant, then we write $T_i = \overline{X}_j$ ($T_i = \overline{X}_j^{-1}$).

EXAMPLE 5.2. Let $\pi_1(X) = \langle a_1, a_2, a_3, a_4, a_5\rangle$, and let $f_\sharp$ be of the form

$$\begin{aligned}
f_\sharp(a_1) &= a_1 \cdot a_4 = U_1 V_1,\\
f_\sharp(a_2) &= a_4^{-1} \cdot \underline{a_2 a_1^{-1} a_2} \cdot a_5 = U_2 \overline{X}_2 V_2 = V_1^{-1}\overline{X}_2 V_2,\\
f_\sharp(a_3) &= a_1 \cdot \underline{a_1 a_1 a_1} \cdot a_5 = U_3 \overline{X}_3 V_3 = U_1 \overline{X}_3 V_2,\\
f_\sharp(a_4) &= a_1^{-1} \cdot \underline{a_5 a_2^{-1} a_5} \cdot a_3 = U_4 \overline{X}_4 V_4 = U_4 \overline{X}_4 V_4,\\
f_\sharp(a_5) &= a_1^{-1} \cdot \underline{a_2 a_3 a_2} \cdot a_3 = U_5 \overline{X}_5 V_5 = U_4 \overline{X}_5 V_4,\\
f_\sharp(a_6) &= a_6^{-1} \cdot \underline{a_1 a_5^{-1} a_1} \cdot a_6 = U_6 \overline{X}_6 V_6 = U_6 \overline{X}_6 U_6.
\end{aligned}$$

Here we have underlined the remnant of $f_\sharp(a_i)$ for each i. Suppose that $z = a_5 a_3^{-1} a_1$. Then we say that $Q_2 T_2 S_2 = f_\sharp(z_2) = f_\sharp(a_3^{-1}) = V_3^{-1} \overline{X}_3^{-1} U_3^{-1}$ with $Q_2 = V_3^{-1}$, etc.

5.1. *New concepts.* For a map in which at least one of the $f_\sharp(a_i)$ does not have remnant, the lengths of the remnants in the other image words are crucial. The longer the minimum of these lengths, the easier it is to prove results concerning Reidemeister equivalence for the map.

Definition 5.3. We define the *minimum remnant length* for $f_\sharp$ to be $\min\{|\overline{X}_i| : f_\sharp(a_i) \text{ has remnant}\}$.

If a homomorphism $f_\sharp$ has remnant, then for any word $w = w_1 \cdots w_q \in \pi_1(X)$ we know that in the product $f_\sharp(w_1) \cdots f_\sharp(w_q)$ none of the remnants will cancel at all. However, when we study homomorphisms with only partial remnant, we can have problems. Note that if we were to change the above example so that $f_\sharp(a_2)$ were $a_4^{-1} \cdot \underline{a_1 a_1 a_2} \cdot a_5$, then in the product $f_\sharp(a_3^{-1}) f_\sharp(a_1) f_\sharp(a_2)$ we would have cancellation occurring between $\overline{X}_3^{-1}$ and $\overline{X}_2$. The next definition describes the additional requirement that we need to avoid this situation. (Compare with the definition of efficient cancellation in [13].)

Definition 5.4. Let $f_\sharp$ be an endomorphism on $\pi_1(X)$ such that $f_\sharp(a_i)$ has remnant. Then $f_\sharp(a_i)$ has *second order stable remnant* if in any product of the form $f_\sharp(a_i) f_\sharp(g_1) f_\sharp(g_2)$ or $f_\sharp(g_2) f_\sharp(g_1) f_\sharp(a_i)$, with $g_1, g_2 \in \mathcal{G}^\pm$, there is cancellation involving part or all of the subword $\overline{X}_i$ only if $g_1 = a_i^{-1}$.

We say that $f_\sharp$ has *second order stable remnant* if for each i such that $f_\sharp(a_i)$ has remnant, $f_\sharp(a_i)$ also has second order stable remnant.

Definition 5.5. A homomorphism $f_\sharp$ on $\pi_1(X)$ has *cyclical cancellation* if for each i the following hold:

1. We have $U_i \neq 1$ and $V_i \neq 1$.
2. There is a unique j such that either $j \neq i$ and $U_j^{-1} U_i = 1$ or else $V_j U_i = 1$.
3. There is a unique j such that either $j \neq i$ and $V_i V_j^{-1} = 1$ or else $V_i U_j = 1$.
4. For any j, each product of the form $U_j^{-1} U_i$, $V_j U_i$, $V_i V_j^{-1}$, and $V_i V_j$ is either equal to 1 or has no cancellation at all.

Definition 5.6 (The functions λ and ϱ). Let $f_\sharp$ have cyclical cancellation. The uniqueness of the cancellations ensures that there are functions $\lambda : \mathcal{G} \to \mathcal{G}^\pm$ (respectively, $\varrho : \mathcal{G} \to \mathcal{G}^\pm$) such that $\lambda(a_i)$ (resp. $\varrho(a_i)$) is the element g of $\mathcal{G}^\pm$ for which $f_\sharp(g)$ cancels with the left (resp. right side) of $f_\sharp(a_i)$ as in item 2 (resp. 3) above. Refer to the example below. These functions can be extended to all of $\mathcal{G}^{\pm}$ by setting $\lambda(a_i^{-1}) = (\varrho(a_i))^{-1}$ and $\varrho(a_i^{-1}) = (\lambda(a_i))^{-1}$.

Example 5.2, Part 2. First note that in this example $f_\sharp$ has minimum remnant length three and does have second order stable remnant, even though it does not have remnant.

In addition, $f_\sharp$ has cyclical cancellation. The orbits of the action of ϱ on $\mathcal{G}^{\pm}$ are given by

$$a_1 \to a_2 \to a_3^{-1} \to a_1, \quad a_1^{-1} \to a_3 \to a_2^{-1} \to a_1^{-1}, \quad a_4 \to a_5^{-1} \to a_4,$$
$$a_4^{-1} \to a_5 \to a_4^{-1}, \quad a_6 \to a_6, \quad a_6^{-1} \to a_6^{-1}.$$

Example 5.7. Next we show what can happen when a map does not have cyclical cancellation. Suppose that $\pi_1(X) = \langle a, b, c \rangle$ and that $f_\sharp$ is given by

$$f_\sharp(a) = ac, \quad f_\sharp(b) = c^{-1} \cdot \underline{bab} \cdot c, \quad f_\sharp(c) = a \cdot \underline{c^{-1}b^2} \cdot a^{-1}.$$

As usual, the remnant is underlined. Note that $\varrho(a)$ is not well defined. It could be b or b^{-1}. There are many possible words w for which $f_\sharp(w)$ has significant cancellation. For example, $f_\sharp(aba^{-1}ca) = f_\sharp(a)f_\sharp(b)f_\sharp(a^{-1})f_\sharp(c)f_\sharp(a)$, and there is cancellation between each of the consecutive factors. The advantage to us in requiring that a map have cyclical cancellation is that there is a well understood form to words w for which $f_\sharp(w)$ has cancellation.

5.2. *The class of MRN homomorphisms.* This class of maps was first defined in [6]. We repeat the definition here using the new terminology from Section 5.1. This class is similar to but more restrictive than the class of 2C3 maps that will be introduced in Section 5.3. For one thing, MRN maps are always defined on spaces of the homotopy type of a wedge of only two circles.

Definition 5.8. Suppose that $\pi_1(X) = \langle a_1, a_2 \rangle$. A map f has *Property MRN* if the following hold:

1. $f_\sharp(a_1)$ does not have remnant, and $|U_1| + |U_2| = f_\sharp(a_1)$.
2. $f_\sharp(a_2)$ does have remnant with minimum remnant length at least two.
3. $f_\sharp$ has cyclical cancellation.
4. Either $f_\sharp$ has second order stable remnant or else $\overline{X}_2 = sts^{-1}$ with tt reduced (so t is cyclically reduced) and $|t| \geq 2$.

Any map that is an MRN map is of the form $f_\sharp(a_1) = U_1V_1$ and either $f_\sharp(a_2) = U_1\overline{X}_2V_1$ (type MRN1) or $f_\sharp(a_2) = V_1^{-1}\overline{X}_2U_1^{-1}$ (type MRN2). In [6] we said that $f_\sharp(a_2) = mrn$, with r the remnant. Hence the name. Kim proves, in Theorem 4.12 of [13], that an MRN map with $|t| \geq 3$ has bounded solution length with SB ≤ 16, but our results are more useful in this case.

The following theorem was stated in [6], and a proof was promised in that paper. Instead, we include here proofs of Theorems 5.12 and 5.13, concerning maps of type 2C3. The proofs for MRN maps use the same techniques but are longer and more tedious.

THEOREM 5.9. *Maps of type MRN have* $\mathrm{far}(f_\sharp) \leq 3$ *and are* 2*-characteristic. For maps of type MRN1, fixed points* x *and* y *are Nielsen equivalent if and only if the set*

$$\{1, a_1, a_2, a_1a_2^{-1}, a_1^{-1}, a_2^{-1}, a_2a_1^{-1}, a_1^2a_2^{-1}, a_2a_1^{-2}, a_2a_1^{-1}a_2^{-1}, a_2a_1a_2^{-1}\}$$

contains a solution z *to* (1). *The corresponding set for maps of type MRN2 is*

$$\{1, a_1, a_2, a_2a_1, a_1^{-1}, a_2^{-1}, a_1^{-1}a_2^{-1}, a_1^{-2}a_2^{-1}, a_2a_1^2, a_2a_1a_2^{-1}, a_2a_1^{-1}a_2^{-1}\}.$$

Note that there are 53 elements of the fundamental group that have length at most two, and here we need only check 11 possible solutions for (1).

5.3. *The class of 2C3 homomorphisms*

DEFINITION 5.10. Assume that $\pi_1(X)$ is the free group on k generators $\langle a_1, \ldots, a_k\rangle$. Then any endomorphism $f_\sharp$ on $\pi_1(X)$ is *of type 2C3* if the following hold:

1. $f_\sharp$ has cyclical cancellation.
2. $|U_1| + |V_1| = |f_\sharp(a_1)|$. Thus $f_\sharp(a_1) = U_1V_1$ is reduced and $f_\sharp(a_1)$ does not have remnant.
3. For $j \geq 2$, $f_\sharp(a_j)$ does have remnant and has second order stable remnant. Thus $f_\sharp$ has second order stable remnant. Also, the minimum remnant length for $f_\sharp$ is three.

The name 2C3 comes from second order stable remnant, cyclical cancellation, and minimum remnant length three.

REMARK 5.11. An MRN map has type 2C3 if it has second order stable remnant and has minimum remnant length at least 3.

We prove the following theorem in Section 7.

THEOREM 5.12. *Any homomorphism* $f_\sharp$ *of type 2C3 is* 3*-characteristic and has* $\mathrm{far}(f_\sharp) \leq 4$.

Example 6 in [6] demonstrates that an MRN map need not be W-characteristic. But this example also has type 2C3. Thus a map of type 2C3 need not be W-characteristic. We provide, in Example 6.4, another homomorphism of type 2C3 that is not W-characteristic. This homomorphism also has the property that it is 3-characteristic but not 2-characteristic. Thus our result that any 2C3 map is 3-characteristic is as strong as possible. Example 6.3 is a map of type 2C3 that has $\mathrm{far}(f_\sharp) = 4$. Again, our result is as strong as possible.

THEOREM 5.13. *For 2C3 maps, with W_γ and W_τ terms of the Fox trace that correspond to fixed points on loops represented by generators a_γ and a_τ, our goal is to determine whether there is a word $z \in \pi_1(X)$ that satisfies $z = W_\gamma^{-1} f_\sharp(z) W_\tau$. The following statements hold:*

1. *Only 14 of the many words of length at most four must be checked as possible solutions to determine whether W_γ and W_τ are Reidemeister equivalent. (When $\pi_1(X)$ has rank k, the number of words of length at most four is more than $8k^4$.) These 14 words are presented in the following table. If none of the 14 conditions in the right column apply, then W_τ and W_γ are not Reidemeister equivalent.*
2. *In fact, Wagner's method will successfully determine whether W_γ and W_τ are equivalent whenever a_γ and a_τ are not equal to $\varrho(a_1)$, $\lambda(a_1)$, and their inverses. In this case, there are only four words that could be solutions, as in Theorem 4.2.*

The following table contains the crucial information needed to conclude that the theorem is true. The proof of both theorems is in Section 7.

	z	This z can occur as a solution only if
$\lvert z\rvert = 0$	1	$W_\gamma = W_\tau$
$\lvert z\rvert = 1$	a_γ	$W_\tau = \overline{W}_\gamma$ (Wagner's notation)
	a_τ^{-1}	$W_\gamma = \overline{W}_\tau$
	a_1	$\lambda(a_1) = a_\gamma^{-1}$ or $\varrho(a_1) = a_\tau$
	a_1^{-1}	$\lambda(a_1) = a_\gamma$ or $\varrho(a_1^{-1}) = a_\tau$
$\lvert z\rvert = 2$	$a_\gamma a_\tau^{-1}$	$\overline{W}_\gamma = \overline{W}_\tau$
	$a_\gamma a_1^{-1}$	$\lambda(a_1^{-1}) = a_\gamma$
	$a_\gamma a_1$	$\lambda(a_1) = a_\gamma$
	$a_1 a_\tau^{-1}$	$\varrho(a_1) = a_\tau^{-1}$
	$a_1^{-1} a_\tau^{-1}$	$\varrho(a_1^{-1}) = a_\tau^{-1}$
$\lvert z\rvert = 3$	$a_\gamma a_1 a_\tau^{-1}$	$\lambda(a_1) = a_\gamma$ or $\varrho(a_1) = a_\tau^{-1}$
	$a_\gamma a_1^{-1} a_\tau^{-1}$	$\lambda(a_1^{-1}) = a_\gamma$ or $\varrho(a_1^{-1}) = a_\tau^{-1}$
$\lvert z\rvert = 4$	$a_\gamma a_1^2 a_\tau^{-1}$	$\lambda(a_1) = a_\gamma$ and $\varrho(a_1) = a_\tau^{-1}$
	$a_\gamma a_1^{-2} a_\tau^{-1}$	$\lambda(a_1^{-1}) = a_\gamma$ and $\varrho(a_1^{-1}) = a_\tau^{-1}$

6. Examples. Example 6.1 provides for each $n \in \mathbb{N}$ a W-characteristic (and hence 2-characteristic) map with $\mathrm{far}(f_\sharp) \geq n$.

EXAMPLE 6.1. Let $n \in \mathbb{N}$. Let $\pi_1(X) = \langle a_1, \dots, a_{n-1}\rangle$. We define $f_\sharp$ by

$$f_\sharp(a_1) = a_1^4, \qquad f_\sharp(a_2) = a_2^2 a_1^3,$$
$$f_\sharp(a_k) = a_k^2 a_{k-1}^{-1} \quad \text{for } k = 3, \dots, n-1.$$

We have fixed points x_4 corresponding to the last a_1 in $f_\sharp(a_1)$ and y corresponding to the second occurrence of a_k in $f_\sharp(a_k)$. It turns out that $d(x_4, y) = n$ and thus $\mathrm{far}(f_\sharp) \geq n$. To simplify the notation, we will prove this for $n = 5$. The proof for other values of n is similar.

Let $\pi_1(X) = \langle a, b, c, d \rangle$. We use upper case letters to represent inverses of generators.

Let $f_\sharp$ be given by

$$a \mapsto a^4 = \underline{a}aaa, \qquad b \mapsto b^2a^3 = b\underline{b}aaa,$$
$$c \mapsto c^2B = c\underline{c}B, \quad d \mapsto d^2C = d\underline{d}C.$$

This map has remnant, and the remnant of each image word is underlined above. Thus the map is W-characteristic. By Theorem 4.2, the map is 2-characteristic.

Consider the following table, which for the fixed point x_i on loop a_j provides the index of the fixed point as well as W_{x_i} and $\overline{W}_{x_i}$. See [6, Section 4.1] for explanations.

i	index	loop $= a_j$	W_i	$\overline{W}_i = f_\sharp(a_j^{-1})\, W_i\, a_j$
0	$+1$	$--$	1	1 (by definition)
1	-1	a	1	A^3
2	-1	a	a	A^2
3	-1	a	a^2	A
4	-1	a	a^3	1
5	-1	b	1	A^3B
6	-1	b	b	A^3B^2
7	-1	c	1	bC
8	-1	c	c	b
9	-1	d	1	cD
10	-1	d	d	cD

Note that x_4 and $y = x_{10}$ are Nielsen equivalent because (using Wagner's algorithm) we have a chain of Reidemeister equivalences from W_4 to W_{10} given by

$$W_4 \sim \overline{W}_4 = W_1 \sim \overline{W}_1 = \overline{W}_6 \sim W_6 = \overline{W}_8 \sim W_8 = \overline{W}_{10} \sim W_{10}.$$

The values of z used for the Reidemeister equivalences above are $z = a, aB, C$, and D, respectively. Thus we can conclude (and easily check) that $z = a^2BCD$ is a solution to the equation $z = W_4^{-1} f_\sharp(z) W_{10}$.

Is there a shorter word that is also a solution? Here we use a new version of abelianization to prove that the answer is no, which proves that $\mathrm{far}(f_\sharp) \geq 5$.

In the free abelian group G_{ab} generated by a, b, c, and d, any element has a unique expression of the form $a^ib^jc^kd^l$ for $i, j, k, l \in \mathbb{Z}$. We use here the symbol $f_\sharp$ to represent the induced endomorphism on G_{ab}. We seek information about any solution z to the equation in G_{ab} given by $z = W_4^{-1} f_\sharp(z) W_{10}$, which can be rewritten as $W_4 W_{10}^{-1} = z^{-1} f_\sharp(z)$.

Let $z = a^i b^j c^k d^l \in G_{ab}$. Then the equation can be written as $a^3 D = a^{-i} b^{-j} c^{-k} d^{-l} f_\sharp(a^i b^j c^k d^l) = a^{3i+3j} b^{j-k} c^{k-l} d^l$. Thus any solution $z \in G_{ab}$ must have $i = 2$, $j = -1$, $k = -1$, and $l = -1$. This implies that in the fundamental group we must have $|z| \geq |i| + |j| + |k| + |l| = 5$. ∎

Even when there is no remnant in any of the image words for $f_\sharp$, we can have $\operatorname{far}(f_\sharp)$ arbitrarily large, as we demonstrate in the next example.

Example 6.2. Let $n \in \mathbb{N}$, $\pi_1(X) = \langle a, b, c \rangle$, and let $f_\sharp$ be given by

$$f_\sharp(a) = (bc)^{2n} a (bc)^{2n}, \quad f_\sharp(b) = c^{-1}, \quad f_\sharp(c) = b^{-1}.$$

Note that this homomorphism has no remnant in each image word. There is a fixed point x with $W_x = (bc)^{2n}$. Using $z = (bc)^n$ in (1), we find that $z = W_{x_0}^{-1} f_\sharp(z) W_x$. Thus $d(x_0, x) \leq 2n$.

Using abelianization, we will now prove that $d(x_0, x) = 2n$. Suppose that z is any solution to $z = W_{x_0}^{-1} f_\sharp(z) W_x$. We will prove that $|z| \geq 2n$.

For this paragraph, we work in the abelianization G_{ab} of $\pi_1(X)$, the free abelian group generated by a, b, and c. Then we can let $z = a^i b^j c^k$ for some $i, j, k \in \mathbb{Z}$. We see that $f_\sharp(z) z^{-1} = W_{x_0} W_x^{-1} = (bc)^{-2n}$. But $f_\sharp(z) z^{-1} = a^{-i+i} b^{4ni-j-k} c^{4ni-j-k}$. Thus we must have $j + k = 4ni + 2n = 2n(2i + 1)$.

Returning now to $\pi_1(X)$, we have $|z| \geq |i| + |j| + |k| \geq |j + k| = 2n|2i + 1|$. But $|2i + 1| \geq 1$ because i is an integer. Thus any solution to $z = W_{x_0}^{-1} f_\sharp(z) W_x$ must have length at least $2n$.

We conclude that $\operatorname{far}(f_\sharp) \geq 2n$. ∎

In the proof of Theorem 5.12 we show that there can be a type 2C3 map with $\operatorname{far}(f_\sharp) = 4$ that is 3-characteristic. Here is an illustration of that case.

Example 6.3. Let $\pi_1(X) = \langle a, b, c \rangle$ with $f_\sharp$ given by

$$f_\sharp(a) = aa, \quad f_\sharp(b) = cb^3 a^{-1} = c \cdot \underline{b^3} \cdot a^{-1}, \quad f_\sharp(c) = c^4 a = c \cdot \underline{c^3} \cdot a.$$

Here the remnant is underlined, and we have two cycles for the function ϱ: $a \to c^{-1} \to b \to a$ and also $a^{-1} \to b^{-1} \to c \to a$. (Reverse the arrows to find the orbits for λ.) The remaining requirements for a type 2C3 map are also met.

Consider fixed points x_γ on the b loop that has $W_\gamma = cb^2$ and x_τ on the c loop that has $W_\tau = c^3$. Using $z = ba^2 c^{-1}$, we find that these fixed points are Nielsen equivalent and $d(W_\gamma, W_\tau) \leq 4$.

By abelianization, as in Example 6.2, there is no shorter z. Thus $d(W_\gamma, W_\tau) = 4$ and $\operatorname{far}(f_\sharp) \geq 4$. But because $f_\sharp$ is of type 2C3, $\operatorname{far}(f_\sharp) = 4$.

On the other hand, there are intermediate fixed points as follows:

List the 10 fixed points as usual with x_0 the base point. Then $x_\gamma = x_5$ and $x_\tau = x_9$. We have $d(x_5, x_2) = 1$, using $z = b$, and $d(x_2, x_1) = 1$, using $z = a$, and finally $d(x_1, x_9) = 2$, using $z = ac^{-1}$. Note that the three solutions here multiply to become the solution for connecting x_5 and x_9 in one step. ∎

Note that in Example 6 of [6] we present a homomorphism that is an MRN map and is not W-characteristic. Thus Wagner's method does not find all Nielsen equivalences. That example is also of type 2C3.

EXAMPLE 6.4. Here we present another homomorphism that is of type 2C3 and is not W-characteristic. In addition, this example has two fixed points x_2 and x_3 for which $d(x_2, x_3) = 3$ and no intermediate fixed point exists. Thus we can conclude that 2C3 maps are not in general 2-characteristic.

Let $\pi_1(X) = \langle a, b, c\rangle$ and let $f_\sharp(a) = a^{-1}b$, $f_\sharp(b) = a \cdot \underline{aba} \cdot a$, and $f_\sharp(c) = a \cdot \underline{b^2c} \cdot b$. This is a homomorphism of type 2C3. The fixed points are described in the table below.

i	index	loop $= a_j$	W_i	$\overline{W}_i = f_\sharp(a_j^{-1})\, W_i\, a_j$
0	$+1$	$--$	1	1 (by definition)
1	$+1$	a	a^{-1}	$b^{-1}a$
2	-1	b	a^2	a^{-2}
3	-1	c	ab^2	b^{-1}

Wagner's method finds no Reidemeister equivalences, but $z = bac^{-1}$ works to prove that W_2 and W_3 are Reidemeister equivalent. Thus this homomorphism is not W-characteristic.

Could it be that this map is 2-characteristic? We answer this question by seeking Reidemeister equivalences for W_2 for which $|z| \le 2$. Any such equivalence must also hold in the abelianization of $\pi_1(X)$. If the abelianization of z is $a^i b^j c^k$ for integers i, j, and k, then we must have $|i| + |j| + |k| \le 2$. It is straightforward to check that in the abelianized setting there is no equivalence between W_2 and W_0 nor between W_2 and W_1 using z satisfying $|i| + |j| + |k| \le 2$. Thus in $\pi_1(X)$ itself there can be no such equivalences with $|z| \le 2$. Thus the map is not 2-characteristic, and our statement that 2C3 maps are 3-characteristic is a sharp result. ∎

7. The proof of Theorems 5.12 and 5.13. Let $f_\sharp$ be a homomorphism of type 2C3. Assume $z \in \pi_1(X)$ is a solution to the equation

$$z = W_\gamma^{-1} f_\sharp(z) W_\tau, \tag{2}$$

where W_γ and W_τ are terms of the Fox trace.

Let x_γ and x_τ be the corresponding Nielsen related fixed points, and let a_γ (respectively a_τ) be the generator of $\pi_1(X)$ that corresponds to the loop containing x_γ (resp. x_τ). We assume that $|z| \ge 4$. In most cases this leads to a contradiction. When it is possible for $|z|$ to be equal to four, we prove that there is a sequence of intermediate fixed points connecting x_γ and x_τ in such a way as to prove that $f_\sharp$ is 3-characteristic.

Remark 7.1. Let $z \in \pi_1(X)$. If z occurs as a solution to (2), then z^{-1} also occurs as a solution to (2) using the W's in the opposite order. This means that if we have proven that z cannot occur as a solution for any W_γ and W_τ, then we also know that z^{-1} cannot occur as a solution for any W_γ and W_τ.

7.1. *Classifying segments of a word in* $\pi_1(X)$. We require a method of dividing z into subwords w_i for which we know the structure of $f_\sharp(w_i)$ and for which there is no cancellation between images of adjacent subwords. See Remark 7.3 for an example.

Definition 7.2. Let $z = z_1 \cdots z_q$ be reduced with each $z_j \in \mathcal{G}^\pm$.

A subword $w = z_s \cdots z_t$ of z is an *RP* (a reducing piece) if $|w| \geq 2$, $\varrho(z_j) = z_{j+1}$ for $j = s, \ldots, t-1$, and w is as long as possible with this property (so that $\varrho(z_{j-1}) \neq z_j$ for example).

A subword $w = z_s \cdots z_t$ of z is an *NRP* (a nonreducing piece) if $\varrho(z_j) \neq z_{j+1}$ for $j = s, \ldots, t-1$, and w is as long as possible with this property.

Remark 7.3. To demonstrate, consider again the orbits for Example 5.2. Suppose that z is the (arbitrary) word given below, with the RPs underlined. We have

$$z = a_5 \cdot \underline{a_3^{-1} a_1} \cdot \underline{a_2^{-1} a_1^{-1} a_3 a_2^{-1}} \cdot a_2^{-1} a_4 \cdot \underline{a_6 a_6} \cdot a_1 \cdot \underline{a_4 a_5^{-1}}.$$

Thus the image is (with the images of the RPs underlined)

$$\begin{aligned} f_\sharp(z) = {} & U_5 \overline{X}_5 V_5 \cdot \underline{V_3^{-1} \overline{X}_3^{-1} V_1} \cdot \underline{V_2^{-1} \overline{X}_2^{-1} \overline{X}_3 \overline{X}_2^{-1} U_2^{-1}} \\ & \cdot V_2^{-1} \overline{X}_2^{-1} U_2^{-1} U_4 \overline{X}_4 V_4 \cdot \underline{U_6 \overline{X}_6 \overline{X}_6 V_6} \cdot U_1 V_1 \cdot \underline{U_4 \overline{X}_4 \overline{X}_5^{-1} U_5^{-1}}. \end{aligned}$$

Note that this product is reduced as written. In addition, the remnant of each image word remains intact in the reduced form of $f_\sharp(z)$.

Lemma 7.4. *Between any two NRPs there is at least one RP. Because $f_\sharp$ has cyclical cancellation and because $f_\sharp(a_1) = U_1 V_1$ is reduced, we have $\lambda(a_1) \neq a_1$ and $\varrho(a_1) \neq a_1$. Also, no RP contains both a_1 and a_1^{-1} because they are in different cycles.*

We now introduce more notation. Let $z = a_\gamma z_2 \cdots z_q$ with each $z_j \in \mathcal{G}^\pm$. For some i, z_j is equal to either a_i or a_i^{-1}. We write $f_\sharp(z_j) = Q_j T_j S_j$. If $z_j = a_i$, then $Q_j = U_i$, $T_b = \overline{X}_i$, and $S_j = V_i$. If $z_j = a_i^{-1}$, then $Q_j = V_i^{-1}$, $T_b = \overline{X}_i^{-1}$, and $S_j = U_i^{-1}$.

If z is an RP, then $f_\sharp(z) = Q_1T_1T_2\cdots T_{q-1}T_qS_q$ is reduced except that each T_i might be trivial. If z is an NRP, then

$$f_\sharp(z) = Q_1T_1S_1Q_2T_2S_2\cdots Q_{q-1}T_{q-1}S_{q-1}Q_qT_qS_q$$

is reduced except that each T_i might be trivial.

7.2. *Bounds on the length of* $W_\gamma^{-1}f_\sharp(z)W_\tau$. Let $z = a_\gamma z_2\cdots z_q$ with each $z_j \in \mathcal{G}^\pm$ and $q \geq 4$. We now analyze the possible solutions to (2), which is $z = W_\gamma^{-1}f_\sharp(z)W_\tau$.

First consider $f_\sharp(z)$, which before reduction is

$$f_\sharp(z) = Q_1T_1S_1Q_2T_2S_2\cdots Q_{q-1}T_{q-1}S_{q-1}Q_qT_qS_q.$$

Let $\widehat{B} = Q_1T_1S_1Q_2$, $M = T_2S_2\cdots Q_{q-1}T_{q-1}$, and $\widehat{E} = S_{q-1}Q_qT_qS_q$. Then we have chopped $f_\sharp(z)$ into a beginning, middle, and end. Note that, for any z with length at least four, there is no cancellation at the dots in the product

$$f_\sharp(z) = \widehat{B}\cdot M\cdot \widehat{E}.$$

Next we define $B = W_\gamma^{-1}\widehat{B}$ and $E = \widehat{E}W_\tau$.

Remark 7.5. For example, consider the element z given in Remark 7.3. We have

$$\widehat{B} = U_5\overline{X}_5V_5V_3^{-1} = f_\sharp(a_5)V_3^{-1} = a_1^{-1}a_2a_3a_2a_3a_5^{-1}.$$

Similarly,

$$\widehat{E} = V_4V_5^{-1}\overline{X}_5^{-1}U_5^{-1} = a_3a_3^{-1}a_2^{-1}a_3^{-1}a_2^{-1}a_1 = a_2^{-1}a_3^{-1}a_2^{-1}a_1.$$

Remark 7.6. Recall that W_γ (resp. W_τ) is an initial segment of $f_\sharp(a_\gamma)$ (resp. $f_\sharp(a_\tau)$). Because of this, W_γ^{-1} is never long enough to cancel with M in $W_\gamma^{-1}f_\sharp(z)W_\tau$ and a similar statement is true about W_τ. Then

$$|z| = |W_\gamma^{-1}f_\sharp(z)W_\tau| = |B| + |M| + |E|$$

because there can be no cancellation at the dots in the product $W_\gamma^{-1}f_\sharp(z)W_\tau = B\cdot M\cdot E$.

We first put bounds on $|B| + |E|$ and then on $|M|$. This provides the facts necessary for the proof of Theorem 5.12.

Let Δ be the number of times that z begins or ends in an NRP. Thus if z is all one NRP, $\Delta = 2$.

Lemma 7.7. *Assume that* $|z| \geq 4$. *Then*

1. $|B| + |E| \geq 2\Delta$.
2. *If* $B = 1$, *then* z *begins with an RP and for some* $\varepsilon \in \{-1,1\}$ *we have* $z_1 = a_1^\varepsilon = \varrho(a_\gamma^{-1})$.
3. *If* $E = 1$, *then* z *ends in an RP and for some* $\varepsilon \in \{-1,1\}$ *we have* $z_q = a_1^\varepsilon = \lambda(a_\tau)$.

Proof. First suppose that z begins in an NRP. Then $Q_1T_1S_1Q_2$ is reduced. Thus

$$|B| = |W_\gamma^{-1}Q_1T_1S_1| + |Q_2| \geq |W_\gamma^{-1}Q_1T_1S_1| + 1.$$

If $|B| = 1$, then $W_\gamma^{-1}Q_1T_1S_1 = 1$, which forces $z_1 = a_\gamma$ and $W_\gamma = f_\sharp(a_\gamma)$. But this happens only if $f_\sharp(a_\gamma)$ ends in a_γ^{-1}. Thus Q_2 does not begin in a_γ because z begins in an NRP. But z begins with Q_2, and Q_2 never cancels from the right. Thus $z_1 \neq a_\gamma$. This is a contradiction, so $|B| \geq 2$.

Similarly, if z ends in an NRP, then $|E| \geq 2$. Hence $|B| + |E| \geq 2\Delta$.

Next, suppose that $B = 1$. By the first part of this lemma, z begins in an RP, and thus $S_1 = Q_2^{-1}$. In addition, we must have $W_\gamma = Q_1T_1$.

Suppose that $z_1 \notin \{a_1, a_1^{-1}\}$. Then we must have $z_1 = a_\gamma$ or else T_1 does not cancel, and thus $W_\gamma = Q_1T_1 = U_\gamma\overline{X}_\gamma$.

This means that

$$z = T_2S_2 \cdots Q_qT_qS_qW_\tau,$$

and T_2 does not cancel.

There are two ways that W_γ can equal $U_\gamma\overline{X}_\gamma$. Either $\overline{X}_\gamma$ ends in a_γ^{-1} or V_γ begins with a_γ.

Suppose that $\overline{X}_\gamma$ ends in a_γ^{-1}. Then $T_2S_2 \cdots Q_qT_qS_qW_\tau$ cannot begin with a_γ. But $z_1 = a_\gamma$. This is a contradiction.

Next suppose that V_γ begins with a_γ. Because z begins in an RP we infer that Q_2 ends in a_γ^{-1}. As above, $T_2S_2 \cdots Q_qT_qS_qW_\tau$ cannot begin with a_γ, and we have a contradiction.

We have proven that if $B = 1$ then z must begin with an RP, with either a_1 or a_1^{-1}, and in either case $z_1 = \varrho(a_\gamma^{-1})$.

The third item is proven similarly. ∎

7.3. *Tedious definitions for the bound on* $|M|$. Recall that

$$M = T_2S_2Q_3T_3 \cdots T_{q-2}S_{q-2}Q_{q-1}T_{q-1}.$$

First we let $y = z_2 \cdots z_{q-1}$ and we seek a lower bound for $|f_\sharp(y)|$. Then we consider the fact that $f_\sharp(y) = Q_2MS_{q-1}$ and find a lower bound for $|M|$.

We consider the contributions of each NRP and RP in y to the length of $f_\sharp(y)$.

First we consider the NRPs in y.

Each z_j in an NRP of y that is not a_1 nor a_1^{-1} contributes at least 5 to $|f_\sharp(y)|$ because $|f_\sharp(z_j)| = |Q_j| + |T_j| + |S_j| \geq 1 + 3 + 1$, and none of this cancels in $f_\sharp(y)$. Let n be the number of such z_j.

Each z_j that is equal to a_1 or a_1^{-1} contributes at least 2 to $|f_\sharp(y)|$ because $|f_\sharp(z_j)| = |Q_j| + |S_j| \geq 1 + 1$, and none of this cancels in $f_\sharp(y)$. Let m be the number of such z_j.

Then the total length of all the NRPs in y is $n+m$, and these NRPs contribute at least $5n+2m$ to $|f_\sharp(y)|$.

Let ℓ_1 be the number of RPs in y that contain neither a_1 nor a_1^{-1}. Similarly, let ℓ_2 be the number of RPs in y that do contain one of a_1 or a_1^{-1}.

Next we consider the RPs in y. The RPs that include neither a_1 nor a_1^{-1} are the easiest to analyze. Recall that ℓ_1 is the number of such RPs in y. Let r be the total number of z_j's in all the RPs of y that do not contain a_1 nor a_1^{-1}. Each of these letters z_j contributes $|T_j| \geq 3$ to $|f_\sharp(y)|$. But note that the first letter of each such RP also contributes $|Q_j| \geq 1$ and the last letter of the RP also contributes $|S_j| \geq 1$.

Thus the RPs that contain neither a_1 nor a_1^{-1} contribute r to $|y|$ and contribute $3r+2\ell_1$ to $|f_\sharp(y)|$.

Finally, we consider the RPs of y that do contain either a_1 or a_1^{-1}. We have that ℓ_2 is the number of such RPs in y. In such an RP, each z_j that is equal to neither a_1 nor a_1^{-1} contributes $|T_j| \geq 3$ to $|f_\sharp(y)|$, as above. Let s be the number of such z_j. The z_j that are either a_1 or a_1^{-1} contribute nothing to $|y|$ unless they are the first or last letter of the RP. Let α be the number of those z_j that are a_1 or a_1^{-1} and are at the beginning or the end of an RP. As above, the first (resp. last) letter of each such RP contributes $|Q_j| \geq 1$ ($|S_j| \geq 1$) to $|f_\sharp(y)|$.

Thus the RPs that do contain a_1 or a_1^{-1} contribute $s+\alpha$ to $|y|$ and $3s+2\ell_2$ to $f_\sharp(y)$.

So far we know that

$$|y| = n+m+r+s+\alpha, \qquad |f_\sharp(y)| \geq 5n+2m+3r+2\ell_1+3s+2\ell_2.$$

Recall that $f_\sharp(y) = Q_2 M S_{q-1}$ and that Q_2 and S_{q-1} each contributed 1 to the lower bound for $|f_\sharp(y)|$. Thus

$$|z| = |y|+2 = n+m+r+s+\alpha+2, \qquad |M| \geq 5n+2m+3r+2\ell_1+3s+2\ell_2-2,$$

and hence, using the first part of Lemma 7.7, we have

$$\begin{aligned} n+m+r+s+\alpha+2 &= |z| = |W_\gamma^{-1} f_\sharp(z) W_\tau| \\ &= |B|+|M|+|E| \geq 5n+2m+3r+2\ell_1+3s+2\ell_2-2+2\varDelta. \end{aligned}$$

In proving Theorem 5.12 we will repeatedly use the following:

$$4+\alpha \geq 4n+m+2r+2\ell_1+2s+2\ell_2+2\varDelta. \tag{3}$$

Remark 7.8. Consider again the word z given in Remark 7.3. We have

$$y = \underline{a_3^{-1}a_1} \cdot \underline{a_2^{-1}a_1^{-1}a_3a_2^{-1}} \cdot a_2^{-1}a_4 \cdot \underline{a_6a_6} \cdot a_1a_4.$$

Note that z ends in an RP ($\underline{a_4a_5^{-1}}$), but if the last letter is removed from z to form y then what was part of this RP (a_4) is now part of an NRP in y.

The NRP in y are $a_2^{-1}a_4$ and a_1a_4. We have $n = 3$ and $m = 1$.

The RPs in y are $a_3^{-1}a_1$, followed by $a_2^{-1}a_1^{-1}a_3a_2^{-1}$, and a_6a_6. Thus $\ell_1 = 1$, $r = 2$, $\ell_2 = 2$, $s = 4$, and $\alpha = 1$.

Thus $|y| = n + m + r + s + \alpha = 12$. Also,

$$f_\sharp(y) = \underline{V_3^{-1}\overline{X}_3^{-1}V_1} \cdot \underline{V_2^{-1}\overline{X}_2^{-1}\overline{X}_3\overline{X}_2^{-1}U_2^{-1}} \\ \cdot V_2^{-1}\overline{X}_2^{-1}U_2^{-1}U_4\overline{X}_4V_4 \cdot \underline{U_6\overline{X}_6\overline{X}_6V_6} \cdot U_1V_1U_4\overline{X}_4V_4,$$

and $|f_\sharp(y)|$ is bounded below by the sum of the number of U's and V's plus three times the number of X's. Thus $|f_\sharp(y)| \ge 14 + 9 \cdot 3 = 41$. Our equation gives us the lower bound $|f_\sharp(y)| \ge 5n + 2m + 3r + 2\ell_1 + 3s + 2\ell_2 = 41$.

Next we note that $|M| = |y| - |V_3^{-1}| - |V_4| \ge 39$ because V_3^{-1} and V_4 are each counted as contributing 1 to the lower bound for $|y|$.

LEMMA 7.9. *In addition*, $\alpha \le s + \ell_2$. *Thus* (3) *becomes*

$$4 \ge 4n + m + 2r + 2\ell_1 + s + \ell_2 + 2\Delta. \tag{4}$$

Proof. For a given RP that contains a_1, the largest number of occurrences of a_1 occurs when the length of the cycle of a_1 is two. That is, to maximize α, we need to consider the case in which $\varrho^2(a_1) = a_1$. In this case, we can have the RP equal to $(a_1\varrho(a_1))^t a_1$. If this is the form of each RP in $f_\sharp(z)$, then $\alpha = s + \ell_2$. The same holds if we have an RP containing a_1^{-1}. An RP of any other form will contribute less to the count for α. ■

7.4. *The proof of Theorem 5.12 at last.* We begin the proof by assuming that $|z| \ge 4$. We prove that this forces z to equal one of a short list of words that all have length 4. For each of these words, we will show that there is an intermediate fixed point for x_γ and x_τ so that there are shorter solutions z that can be used to connect x_γ and x_τ.

In addition to (4), repeated here,

$$4 \ge 4n + m + 2r + 2\ell_1 + s + \ell_2 + 2\Delta,$$

we have

$$4 \le |z| = n + m + r + s + \alpha \le n + m + r + 2s + \ell_2.$$

Note that these two inequalities imply that $n = 0$ and that Δ is 1 or 0. Therefore without loss of generality we may assume that z begins in an RP.

CASE 1: $s = 0$. In this case there are no RPs in y, and the only NRPs in y are made up entirely of powers of a_1. Thus $y = a_1^m$ or $y = a_1^{-m}$ and $m = |z| - 2 = q - 2$.

By Remark 7.1, we may assume that $z = z_1a_1^{q-2}z_q$. Because z begins in an RP, we know that $z_1 = \lambda(a_1)$. (This seems contradictory, but remember that $\ell_1 + \ell_2$ is the number of RPs in the subword y. Here z_2 is part of an NRP in y and yet part of an RP in z.)

In addition, because $\lambda(a_1) \neq a_1^{\pm 1}$, Lemma 7.7 guarantees that $|B| \neq 0$. The same is true about $|E|$ either because z ends in an NRP (so $|E| \geq 2$) or because z ends in an RP and $z_q = \varrho(a_1) \neq a_1^{\pm 1}$. Thus $|B| + |E| \geq 2$.

(4) becomes

$$4 \geq m + |B| + |E| \geq q - 2 + 2 = q = |z|.$$

This forces $z = \lambda(a_1)a_1a_1\varrho(a_1) = (W_\gamma^{-1}Q_1T_1) \cdot (V_1U_1) \cdot (T_4S_4W_\tau)$, with B, M, and E indicated by parentheses on the right side.

We now prove that there are intermediate fixed points connecting x_γ and x_τ. We have $B = W_\gamma^{-1}Q_1T_1 = \lambda(a_1) = z_1$. Thus $z_1 \neq a_1^{\pm 1}$ and $|T_1| \geq 3$. In order for most of T_1 to cancel when B is reduced, we must have $a_\gamma = z_1 = \lambda(a_1)$. Thus $f_\sharp(a_\gamma) = U_\gamma \overline{X}_\gamma U_1^{-1}$.

Similarly, $a_\tau^{-1} = z_q = \varrho(a_1)$. Thus $f_\sharp(a_\tau) = U_\tau \overline{X}_\tau V_1$.

We also have $M = V_1U_1 = a_1^2$, so $f_\sharp(a_1) = a_1^2$. To summarize:

$$f_\sharp(a_1) = a_1^2, \quad f_\sharp(a_\gamma) = U_\gamma \overline{X}_\gamma a_1^{-1} = W_\gamma a_\gamma a_1^{-1},$$
$$f_\sharp(a_\tau) = U_\tau \overline{X}_\tau a_1 = W_\tau a_\tau a_1.$$

Let x_1 and x_2 be the two fixed points on the a_1 loop. Using Wagner's notation, we have the following information:

	W_i	$\overline{W}_i = f_\sharp(a_j^{-1})\, W_i\, a_j$
x_0	1	1
x_1	1	a_1^{-1}
x_2	a_1	1
x_γ	W_γ	a_1
x_τ	W_τ	a_1^{-1}

Using the ideas in the proof of Theorem 4.2, we find that $d(x_\gamma, x_2) \leq 1$ (via $z = a_\gamma$); $d(x_2, x_1) \leq 1$ (using $z = a_1$); and $d(x_1, x_\tau) \leq 2$ (using $z = a_1a_\tau$). Note that combining these three shorter z's gives us the original $z = a_\gamma a_1 a_1 a_\tau^{-1}$. Thus $x_\tau \sim x_2 \sim x_1 \sim x_\tau$ with each equivalence requiring a solution of length at most 2.

Case 2: $s \geq 1$. Thus $\ell_2 \geq 1$ and $1 \leq \alpha$. This forces $r = 0$, $\ell_1 = 0$ (because $r \geq \ell_1$) and $m + 2\Delta \leq 2$. Thus $\Delta = 0$ or $\Delta = 1$.

Case 2A: $s \geq 1$ and $\Delta = 1$ (so that one end of z is an RP and the other is an NRP). Without loss of generality, we can assume that z_1z_2 is an RP and z_q is part of an NRP in z. Thus by the proof of Lemma 7.7, $|E| \geq 2$. Returning to (3), replacing 2Δ with $|B| + |E|$, we see that

$$2 \geq m + s + \ell_2 + |B|.$$

This means that $s = 1 = \ell_2$ and $m = 0 = |B|$. Using (3) and then (4), we find that $4 \leq |z| \leq s + \alpha + 2 \leq 2s + \ell_2 + 2 = 4$, and $4 = |z| \geq 3s + 2\ell_2 - 2 + |B| + |E| \geq 5$. Thus we have a contradiction.

Case 2B: $s \geq 1$ and $\Delta = 0$ (so that z begins and ends in RPs). We have $2 \geq m + |B| + |E|$.

Subcase 2B(i): $|B|+|E| = 2$. We have $m = 0$ and $4 \geq s+\ell_2+|B|+|E| \geq 4$. This forces $s = \ell_2 = 1$, and $1 \leq \alpha \leq 2$.

Then $|z| \geq 3s = 2\ell_2 - 2 + |B| + |E| = 5$.

Because y has no NRPs and z begins and ends in RPs, it follows that z is composed entirely of one RP. Thus $|z| = 5$ and $\alpha = 2$. Then z must have the form (for some $\varepsilon \in \{1, -1\}$)

$$z = \lambda(a_1^\varepsilon)a_1^\varepsilon\varrho(a_1^\varepsilon)a_1^\varepsilon\varrho(a_1^\varepsilon).$$

This can happen only if $a_1^\varepsilon = \varrho^2(a_1^\varepsilon)$ and thus $\lambda(a_1^\varepsilon) = \varrho(a_1^\varepsilon)$.

Note that we now have

$$z_1a_1^\varepsilon z_1a_1^\varepsilon z_1 = z = (W_\tau^{-1}Q_1T_1) \cdot (T_1) \cdot (T_1S_1W_\tau).$$

In addition, $z_1 = \varrho(a_1^\varepsilon) \neq a_1^\varepsilon$ (because $f_\sharp$ is a 2C3 homomorphism). Thus $|B| \neq 0$ and $|E| \neq 0$. This forces $B = z_1 = E$. Because $M = T_1$ with length at least three, and because $|B|+|E| = 2$, we must have $T_1 = a_1^\varepsilon z_1a_1^\varepsilon$. But this contradicts the fact that T_1 must end in z_1 because $W_\tau^{-1}Q_1T_1 = B = z_1$. Therefore this subcase cannot occur.

Subcase 2B(ii): $|B| + |E| = 1$. In this subcase we have $m \leq 1$, and without loss of generality we may assume that $|B| = 0$ and $|E| = 1$. Then by Lemma 7.7 (for some $\varepsilon \in \{1, -1\}$)

$$z_1z_2 = a_1^\varepsilon\varrho(a_1^\varepsilon).$$

Recall that because $n = 0$, any NRP in y must be a power of a_1. Because $\varrho(a_1^\varepsilon) \neq a_1^\varepsilon$, z_2 cannot be part of an NRP in y. Thus z_2 must be part of an RP in y. The fact that $z_2 \neq a_1^\varepsilon$ implies that α cannot reach its maximum value of $s + \ell_2$. We now have $\alpha \leq s + \ell_2 - 1$.

Thus our inequalities combine to become

$$3 \geq m + s + \ell_2 + |E| \geq m + 1 + s + \ell_2,$$

which implies that $m = 0$, $s = 1$, $\ell_2 = 1$, $\alpha = 1$, and $|z| = 4$.

We assume that $\varepsilon = 1$ because the proof for $\varepsilon = -1$ is very similar. Using arguments similar to those in the previous subcase, we find that

$$z = a_1\varrho(a_1)a_1\varrho(a_1) = (W_\gamma^{-1}U_1) \cdot (T_2) \cdot (T_2S_2W_\tau).$$

Also $E = z_4 = \varrho(a_1)$ is the first letter of T_2, and $M = T_2 = a_1\varrho(a_1)a_1$. But $a_1 \neq \varrho(a_1)$. This is a contradiction.

Subcase 2B(iii): $|B|+|E| = 0$. Here, for some $\varepsilon_1, \varepsilon_2 \in \{1, -1\}$, we have $z_1z_2 = a_1^{\varepsilon_1}\varrho(a_1^{\varepsilon_1})$ and $z_{q-1}z_q = \lambda(a_1^{\varepsilon_2})a_1^{\varepsilon_2}$.

As in the subcases above, $z_1z_2z_3$ and $z_{q-2}z_{q-1}z_q$ must be parts of RPs in z. Thus $s \geq 2$ and $\alpha \leq s + \ell_2 - 2$ because the number of a_1's and a_1^{-1}'s in y is limited by the fact that y begins and ends in letters other than these.

We have

$$m + 2s + \ell_2 \geq m + s + \alpha + 2 = |z| \geq 2m + 3s + 2\ell_2 - 2,$$

and thus $2 \geq m + s + \ell_2 \geq 3$. Thus this subcase cannot occur.

Thus both theorems are proven. ■

References

[1] R. F. Brown, *The Lefschetz Fixed Point Theorem*, Scott, Foresman and Co., 1971.

[2] M. Bestvina and M. Handel, *Train-tracks for surface homeomorphisms*, Topology 34 (1995), 109–140.

[3] O. Bogopolski, A. Martino, O. Maslakova and E. Ventura, *The conjugacy problem is solvable in free-by-cyclic groups*, Bull. London Math. Soc. 38 (2006), 787–794.

[4] E. Fadell and S. Husseini, *The Nielsen number on surfaces*, in: Contemp. Math. 21, Amer. Math. Soc., 1983, 59–98.

[5] E. Hart, *The Reidemeister trace and the calculation of the Nielsen number*, in: Nielsen Theory and Reidemeister Torsion, Banach Center Publ. 49, Inst. Math., Polish Acad. Sci., 1999, 151–157.

[6] —, *Algebraic techniques for calculating the Nielsen number on hyperbolic surfaces*, in: Handbook of Topological Fixed Point Theory, Kluwer, 2005, 463–487.

[7] E. Hart, P. Heath and E. Keppelmann, *An algorithm for Nielsen type periodic numbers of maps with remnant on surfaces with boundary and on bouquets of circles*, submitted.

[8] E. Hart and S. W. Kim, *The Nielsen number for free fundamental groups and maps without remnant*, J. Fixed Point Theory Appl. 2 (2007), 261–275.

[9] B. Jiang, *Lectures on Nielsen Fixed Point Theory*, Contemp. Math. 14, Amer. Math. Soc., Providence, RI, 1983.

[10] M. Kelly, *Minimizing the number of fixed points for self-maps of compact surfaces*, Pacific J. Math. 126 (1987), 81–123.

[11] —, *Nielsen fixed point theory on surfaces*, in: Handbook of Topological Fixed Point Theory, Kluwer, 2005, 647–658.

[12] —, *Computing Nielsen numbers of surface homeomorphisms*, Topology 35 (1996), 13–25.

[13] S. W. Kim, *Computation of Nielsen numbers for maps of compact surfaces with boundary*, J. Pure Appl. Algebra 208 (2007), 467–479.

[14] —, *Nielsen numbers of maps of polyhedra with fundamental group free on two generators*, submitted.

[15] J. Llibre and A. Nunes, *Minimum number of fixed points for maps of the figure eight space*, Int. J. Bifur. Chaos Appl. Sci. Engrg. 9 (1999) 1795–1802.

[16] J. Wagner, *An algorithm for calculating the Nielsen number on surfaces with boundary*, Trans. Amer. Math. Soc. 351 (1999), 41–62.

[17] —, *Classes of Wecken maps of surfaces with boundary*, Topology Appl. 76 (1997), 27–46.

[18] P. Yi, *An algorithm for computing the Nielsen number of maps on the pants surface*, Ph.D. Thesis, UCLA, 2003.

Department of Mathematics
Colgate University
Hamilton, NY 13346-1398, U.S.A.
E-mail: ehart@mail.colgate.edu

Received 17 May 2005;
in revised form 2 August 2006 and 16 October 2007