Pergamon

Topology Vol. 35, No. 1, pp. 13–25, 1996
Elsevier Science Ltd
Printed in Great Britain

0040–9383(95)00011–9

COMPUTING NIELSEN NUMBERS OF SURFACE HOMEOMORPHISMS

MICHAEL R. KELLY

(*Received* 26 *April* 1994; *in revised form* 6 *February* 1995)

1. INTRODUCTION

LET X be a compact ANR and let $f: X \to X$ be a continuous self-mapping. Then X admits a local fixed point index and one defines the *Nielsen number*, $N(f)$, of f as follows: first, an equivalence relation is defined on the set of fixed points of f, denoted $Fix(f)$, by $x \sim y$ if and only if there exists a path α in X going from x to y such that $f(\alpha)$ is homotopic to α rel endpoints. An equivalence class under this relation is often referred to as a *Nielsen class*. Each component of $Fix(f)$ has an integer-valued index, and the index of a Nielsen class is the sum of the indices of the components which belong to that class. Then $N(f)$ is defined to be the number of Nielsen classes which have a nonzero index. The *fixed point data* of f consists of the Nielsen classes together with their indices. An excellent treatment of the basics of Nielsen theory can be found in [2]. An alternative approach, using covering space theory, can be found in [7]. In this context, the fixed point data naturally appears as an element in an appropriate 0th-Hochschild homology group [5]. It is classically known as the *Reidemeister trace* of the map.

Clearly, $N(f)$ gives a lower bound for the number of fixed points of f, and by the homotopy invariance of the Nielsen number, a lower bound among all maps homotopic to f. In a number of cases this lower bound is realizable; for example, when X is a topological manifold and f is a homeomorphism. In such cases, we may think of the Nielsen number as giving the fixed point count of an "optimal" representative for a given homotopy class of maps.

This paper is concerned with the problem of computing the Nielsen number. With input data being the topological space X and a homotopy class of self-mappings of X, denoted by f, we are interested in an effective method which outputs the value of $N(f)$. This should depend on the topology of X and on the given homotopy class of self-mappings. Once a base point is chosen, for many spaces the latter is equivalent to being given the induced action on the fundamental group of X. Some results on the computation of $N(f)$ can be found in [12] where computational algorithms are given when X is either a polyhedron with finite fundamental group, a Jiang space, or an Infrasolvmanifold. As the main result of this paper we present an algorithm which allows one to compute $N(f)$ in the case that X is a compact surface and the homotopy class of self-mappings is a *mapping class*; that is, a class that contains a homeomorphism. This allows us to take advantage of the Nielsen–Thurston theory of surface automorphisms and the recent algorithm due to Bestvina and Handel [1].

The basic idea of this algorithm can be described as follows. By a theorem of Jiang and Guo [8], the Nielsen number is the same as the minimal number of fixed points occurring among all maps homotopic to the given mapping f. This minimal number is denoted by $MF[f]$. The Jiang–Gao result allows us to use methods developed by this author [9, 10] for studying the fixed point behavior of self-mappings of surfaces. In Section 2 we present these methods in the context needed for our algorithm. For self-mappings in general it is not known whether or not these methods are algorithmic, but in the special case of a mapping class an improvement is given in Section 3. This improvement reduces the computation of $MF[f]$ to the problem of finding a maximal reducing family of simple closed curves for the mapping class. The Bestvina–Handel algorithm is now applied to find such a reducing family, and hence, the algorithm for $N(f)$. Section 4 gives the details of the algorithm for computing Nielsen numbers and in Section 5 some illustrative examples are given.

We would like to point out here that the Bestvina–Handel algorithm can also be applied to the computation of Nielsen numbers in the case of an orientation-preserving homeomorphism of an orientable surface. In practice this turns out to be much more complicated than the algorithm presented here. The reason is that one needs to find train-tracks for both the stable and unstable laminations as well as have a good estimation of the induced action on these laminations so as to determine precisely which domains are invariant under the action. Our algorithm approaches the computation of $N(f)$ directly. It starts by considering a certain representative mapping for the given mapping class and then decides when, if ever, any of its fixed points are in the same Nielsen class. Thus, in addition to the Nielsen number, the algorithm gives the Reidemeister trace of the mapping class.

2. A METHOD FOR FINDING THE LEAST NUMBER OF FIXED POINTS

Let F be a compact surface with nonempty boundary, ∂F. Throughout this section we assume that F is equipped with a handle structure consisting of 0-handles $D_1, \ldots, D_n$ together with 1-handles $H_1, \ldots, H_k$ attached to the boundary of $\bigcup D_i$. Let A denote the union of the $2k$ attaching arcs $A_1, \ldots, A_{2k}$. By a *1-dimensional submanifold of F* we mean a finite collection of properly embedded curves Γ in F which (i) are pairwise disjoint, (ii) are transverse to A, (iii) contains no one-sided simple closed curves, (iv) contains no inessential simple closed curves or any arcs which are homotopic (rel. endpoints) to an arc in the boundary. The collection of curves Γ is said to be *taut* if the cardinality of $\Gamma \cap A$ is minimal among all collections Γ' isotopic rel endpoints to Γ. If we allow isotopies which move endpoints in the above, we say that Γ has *minimal geometric intersection with A.*

Let Γ be a 1-dimensional submanifold of F and suppose that $\mu : \Gamma \to \{1, \ldots, 2k\}$ is a continuous function. The pair (Γ, μ) is called a *combinatorial mapping* if there exists a self-mapping f of F such that (i) $f^{-1}(A) = \Gamma$, (ii) $\mu(C) = l$ if and only if $f(C) \subset A_l$ for each component C of Γ, (iii) f has no fixed points on Γ, and (iv) for each open set $\mathcal{O}$ meeting Γ, $f(\mathcal{O})$ meets at least two components of $F \backslash A$. The mapping f is referred to as a *topological representative* for Γ.

Given $f : F \to F$ such that $f^{-1}(A)$ is a 1-dimensional submanifold of F there is a natural induced combinatorial mapping, denoted Γ_f with $\Gamma_f = f^{-1}(A)$. (The obvious function μ is suppressed in the notation.) More generally, given a self-mapping f and a combinatorial mapping (Γ, μ) we say that (Γ, μ) *represents* f if (Γ, μ) has a topological representative which is homotopic to f. The use of homotopy in this definition does not create any ambiguity due to the following lemma.

LEMMA 2.1. *If f and g are self-mappings of F such that $\Gamma_f = \Gamma_g$, then f is homotopic to g.*

Proof. The given handle structure for F together with the function μ determine the induced map $f_{\#}$ of $\pi_1(F)$ up to an inner automorphism of π_1. Since surfaces are $K(\pi, 1)$'s the result follows. □

Given the pair (Γ, μ) and a topological representative f for the pair, assign to each component X of $F \backslash (A \cup \Gamma)$ an integer, denoted $index(X, f)$, which is the topological fixed point index of f on the domain X. The value of this index depends on the positioning of X inside the surface F, and on the action of f on the points of $\Gamma \cap A$ which lie on the frontier of X. An algorithm for the computation of $index(X, f)$ is given in Section 2 of [9]. This algorithm is part of the algorithm given in Proposition 2.2 below. For a given topological representative f let $M(\Gamma, f)$ denote the number of components which have nonzero index. This gives a rough estimate for the number of fixed points of f, and provides an upper bound for estimating $MF[f]$.

In order to eliminate the dependence on the choice of topological representative in the estimate $M(\Gamma, f)$ we consider the following. Let C be an oriented curve in F which meets both A and Γ transversely. The *A-itinerary* of C is a finite sequence, taken from the set $\{1, \ldots, 2k\}$, which measures the curve's points of intersection (in order and without repitition) with the components of A. For example, an arc which traverses A_1, then A_2 twice followed by A_4 5-times and then A_3 once would have $(1, 2, 4, 3)$ as its A-itinerary. For closed curves this is well defined only up to a cyclic permutation. We say that 1-dimensional submanifolds Γ_1 and Γ_2 are *A-equivalent* if there is a choice of orientations and an isotopy of F taking Γ_1 to Γ_2 which is A-itinerary preserving. The *Γ-itinerary* of the curve C is the finite sequence whose ith entry is the μ-value of the ith point of intersection of C with Γ. This allows for repetition. For example, the Γ-itinerary $(1, 2, 2, 1)$ indicates an oriented curve intersecting Γ in components $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ such that $\mu(\gamma_1) = \mu(\gamma_4) = 1$ and $\mu(\gamma_2) = \mu(\gamma_3) = 2$.

Let $FPN(\Gamma, A)$ denote the minimum among all possible values of $M(\Omega, g)$ where Ω is A-equivalent to Γ and g is a topological representative for Ω.

PROPOSITION 2.2. *Given the combinatorial mapping (Γ, μ) there is an algorithm for computing $FPN(\Gamma, A)$.*

Proof. For a fixed choice of Ω and g let R_Ω denote the set of components of $F \backslash (A \cup \Omega)$ whose frontier does not consist of a single subarc of Ω together with a single subarc of A. Define a graph G in the surface F as follows: first choose one vertex in each $X \in R_\Omega$ for which X and $g(X)$ lie in the same component of $F \backslash A$. For a vertex v, let X_v denote the component corresponding to v. Join vertices v and w with an edge if and only if there is a point x common to the frontiers of X_v and X_w with the property that x is in $A_i \cap \mu^{-1}(i)$ for some i. Observe that for any two A-equivalent combinatorial mappings their associated graphs are naturally isomorphic. In fact, there is an isotopy of F taking one to the other for which the path of each vertex under the isotopy is contained in a single handle. Thus, the first step in the algorithm is to produce the graph G in F.

The second and main step involves computing indices. Assign to each vertex v the value $i(v) = index(X_v, g)$. We now need to optimize the number of vertices with $i(v) = 0$. This turns out to be a linear optimization problem (using the algorithm for computing $i(v)$). First, for each isolated vertex v of G, the value of $i(v)$ is independent of the choice of topological representative. Next, on each maximal connected subgraph of G whose vertices are all of valence at most two (in G), optimize the number having $i($[redacted]$) = 0$. Finally, each such subgraph can be done in such a way as to optimize the number of remaining vertices having $i(\blacksquare) = 0$. □

The following lemma gives two invariance properties of FPN. The proof follows directly from the definition.

LEMMA 2.3. *(a) Let* (Γ, μ) *be a combinatorial mapping and* $g: F \to F$ *a homeomorphism. Then* $g(\Gamma)$ *is a combinatorial mapping (with respect to* $g(A)$*) and* $FPN(g(\Gamma), g(A)) = FPN(\Gamma, A)$. *(b) If* Γ *and* Γ' *are A-equivalent combinatorial mappings, then* $FPN(\Gamma, A) = FPN(\Gamma', A)$.

Our approach to finding $MF[f]$ is to note that this minimal number for maps is achieved by some mapping g for which Γ_g is a combinatorial mapping. Hence, $MF[f] = FPN(\Gamma_g, A)$. In the light of Proposition 2.2 it remains to find an algorithm for producing this Γ_g given the input homotopy class $[f]$ and the handle structure which determines A.

A *stable arc* for (Γ, μ) is an arc λ in the interior of F with the properties: (1) its endpoints are in distinct components of the complement of $\Gamma \cup A$; (2) it has minimal geometric intersection (rel. endpoints) with A; and (3) its A- and Γ-itineraries are the same. Two stable arcs are said to be equivalent if their respective endpoints lie in the same components and they have the same A-itineraries. A stable arc is *minimal* if any subarc which is also a stable arc is A-equivalent to the entire arc. Thus, minimal stable arcs are those that cannot be split into two proper arcs each of which is stable. A stable arc is said to be *trivial* if it is A-equivalent to a stable arc which intersects Γ only in points of $A \cap \Gamma$. The graph G constructed in the proof of Proposition 2.2 contains all possible equivalence classes of trivial stable arcs for the given combinatorial mapping. Hence, trivial stable arcs are easy to find. Theorem 2.4 below indicates the need for finding nontrivial minimal stable arcs.

The following construction defines an action of stable arcs on combinatorial mappings. Let Γ be a combinatorial mapping and let α be a stable arc for Γ with $\alpha \cap \Gamma \cap A = \emptyset$. There is a natural pairing between the finite sets $\alpha \cap A$ and $\alpha \cap \Gamma$. Given a point p in the latter, let p' be its paired point in the former, and let α_p be the subarc of α joining the two. Let C_p be a disk which contains α_p and meets both of A and Γ in a number of arcs, one for each point in $\alpha_p \cap (A \cup \Gamma)$. Let γ_p be the arc in $C_p \cap \Gamma$ containing p, let B_p be the component of $C_p \backslash \gamma_p$ which contains p', and let ζ_p denote the arc $\partial C_p \cap B_p$. Replace γ_p by ζ_p to obtain a new curve. Loosely, a "finger-push" of γ_p along α_p has been performed to obtain this new curve. If q is another point in $\alpha \cap \Gamma$ the disk C_q is chosen as above but with the additional requirement that the arc ζ_q is kept disjoint from ζ_p. Repeating this for each point in $\alpha \cap \Gamma$ we obtain a new collection of curves $\alpha(\Gamma)$, and a new combinatorial mapping by keeping the assignment μ unchanged. An ordered sequence Λ of minimal, nontrivial stable arcs which are pairwise disjoint is called a *merging sequence*. The action described above applied in the same manner to the sequence Λ yields a combinatorial mapping $(\Lambda(\Gamma), \mu)$.

In our present notation, Theorem 3.1 of [10] generalizes naturally to the following.

THEOREM 2.4. *Given a self-mapping* f *of* F, *there is a taut combinatorial mapping* (Γ, μ) *representing* f *and a merging sequence* Λ *such that* $FPN(\Lambda(\Gamma), A) = MF[f]$. *Moreover, if* h *is a given fixed point minimizing map we may assume that* Γ *is isotopic to* $h^{-1}(A)$.

Proof. The proof of [10; Theorem 3.1] goes through unchanged. Merging arcs are replaced here by nontrivial, minimal stable arcs and the map $f_{\min}$ is replaced by the computation of $FPN(\blacksquare)$. □

3. SURFACE AUTOMORPHISMS AND FIXED POINTS

In order to use Theorem 2.4 to get an algorithm for $MF[f]$ one needs to produce all possible taut combinatorial mappings representing f, and for each one determine all possible merging sequences. This would be a difficult task in general, but in the special case that f is a homeomorphism we give an improvement below in Theorem 3.1 which will lead to an algorithm.

Suppose that F is a compact surface with nonempty boundary and negative Euler characteristic. Let $h: F \to F$ be a homeomorphism and suppose that C_h is a maximal family of h-invariant simple closed curves in F. In particular, $h(C_h)$ is isotopic to C_h, no two components of C_h are isotopic. We assume that C_h is contained in the interior of F, has minimal geometric intersection with A, and no component borders two distinct periodic components of $F \backslash C_h$ such that their union is also periodic. The collection C_h is often referred to as a *reducing family* for h.

For each component A_i of A, we let P_i denote the disk formed by taking the union of all handles which meet A in exactly two arcs, both of which are isotopic to A_i. Suppose Σ is a collection of taut proper arcs in F each isotopic to A_i. We say that Σ is *A-admissible* if (1) neither endpoint lies in P_i, (2) any two components have the same A-itinerary, and (3) if the A-itinerary is empty, then Σ is contained in a single component of $F \backslash A$. In general, Γ is A-admissible if a taut collection of curves which is isotopic to Γ rel. endpoints has the property that for each i, the subcollection of arcs isotopic to A_i is A-admissible.

THEOREM 3.1. *Given h and C_h as described above, there exist an A-admissible combinatorial mapping Γ representing h, which has minimal geometric intersection with A and is isotopic to $h^{-1}(A)$, and a merging sequence Λ such that $N(h) = FPN(\Lambda(\Gamma), A)$ and each stable arc in Λ is contained in C_h.*

Remark. If no component of $h^{-1}(A)$ is isotopic to a component of A, then Γ is unique up to A-equivalence. If the subcollection of curves in Γ which are isotopic to A_i is nonempty, then it follows from the definition of admissible that there are exactly two choices for the placement of these curves.

Proof of Theorem 3.1. Without loss we assume that $h(C_h) = C_h$. First consider the case when C_h only contains simple closed curves which are parallel to a boundary component of F. Then by the Nielsen–Thurston classification theorem [13], h is isotopic to either a periodic or pseudo-Anosov homeomorphism.

Suppose h is isotopic to a pseudo-Anosov mapping. Let B_0 be a collection of proper geodesic arcs in F such that each A_i is isotopic to a unique member of B_0. Then there is a pseudo-Anosov mapping ρ isotopic to h and a train-track τ which carries the stable lamination for ρ and further has the property that τ meets B_0 efficiently [6, Theorem 2.5.1]. As a result $\rho^{-1}(B_0)$ is a taut (relative to B_0) collection of curves. By choosing a number of arcs sufficiently close to those in B_0 we obtain a collection B and an associated handle structure on F such that B is isotopic to A and the collection of curves $\rho^{-1}(B)$ is taut (relative to B) and is also B-admissible. Let $I: F \to F$ denote the end of an isotopy with $I(A) = B$.

Now the mapping ρ may have more than $N(h)$ fixed points, but for a pseudo-Anosov any excess fixed points occur on invariant boundary components and have zero index. Thus by an arbitrarily small deformation we can replace ρ by an embedding which, by abuse of notation, we will also denote by ρ. This embedding has exactly $N(h)$ fixed points and is chosen so that $\rho^{-1}(B)$ is taut and so that no fixed points occur on B. It follows directly from the definition that the induced combinatorial mapping Γ_ρ is such that $FPN(\Gamma_\rho, B) = N(h)$.

In order to get the desired combinatorial mapping and merging sequence in the theorem the following construction is used to produce a new combinatorial mapping representing h which also realizes $N(h)$: let $\hat{F}$ denote the surface obtained by attaching an annulus to each component of ∂F. Let K denote the closure of $\hat{F} \backslash F$. Extend each arc component of A (resp. B) vertically outwards in the product structure of the annuli to obtain the collection of arcs $\hat{A}$ (resp. $\hat{B}$) in $\hat{F}$. The isotopy taking A to B extends to one of $\hat{F}$ taking $\hat{A}$ to $\hat{B}$. Let $\hat{I}$ denote the end of this isotopy. Since C_h has minimal geometric intersection with A there is a homeomorphism $J: \hat{F} \to F$ such that $J(\hat{A}_i) = A_i$ and $J(\partial F) = C_h$.

We now extend Γ_ρ to a combinatorial mapping of $\hat{F}$. Let N be an arbitrarily small regular neighborhood of ∂F in $\hat{F}$. Let ϕ be a component of Γ_ρ. Since ϕ is B-admissible, we can attach two taut proper arcs in K to ϕ so that the union is a $\hat{B}$-admissible proper arc in $\hat{F}$ which, up to $\hat{B}$-equivalence, has the following form: $u_1 \cdot u_2 \cdot u_3 \cdot u_4 \cdot u_5$ where u_1 and u_5 are contained in $K \backslash \hat{B}$; u_2 and u_4 are contained in N, with each having a $\hat{B}$-itinerary of the form $(Y, x, r(Y))$ where x is an integer and $r(Y)$ denotes the reverse of the sequence Y; and $u_3 \subset F \backslash int(N)$ is such that if u_2 and u_4 are replaced by arcs in $N \backslash \hat{B}$, then the resulting curve has minimal geometric intersection with $\hat{B}$. Repeating this construction for each component of Γ_ρ produces a $\hat{B}$-admissible combinatorial mapping denoted by $\hat{\Gamma}_\rho$. Since any topological representative for Γ_ρ extends to a self-map g of $\hat{F}$ such that $g^{-1}(\hat{B}) = \hat{\Gamma}_\rho$ and $g(K) \subset F$, it follows that $FPN(\hat{\Gamma}_\rho, \hat{B}) = FPN(\Gamma_\rho, B)$.

Now, proceed as in the proof of Theorem 3.1 in [10], but only making alterations to $\hat{\Gamma}_\rho$ inside the neighborhood N defined above. Here, the only move required involves reducing curves with $\hat{B}$-itinerary of the form $(Y, x, r(Y))$ by cutting along an arc parallel to a subarc of $\hat{B}$ (this is [9, Lemma 3.5]). After removing inessential simple closed curves the resulting collection of curves is isotopic (rel. endpoints) to the original $\hat{\Gamma}_\rho$. Also, it follows that after this reduction process, all nontrivial stable arcs can be chosen to lie on ∂F. As a result, we obtain a $\hat{B}$-admissible combinatorial mapping Ψ which has minimal geometric intersection with $\hat{B}$, and a merging sequence Λ contained in ∂F with $FPN(\Lambda(\Psi), \hat{B}) = FPN(\hat{\Gamma}_\rho, \hat{B}) = N(h)$. Finally, Lemma 2.3 ensures that $J \circ \hat{I}^{-1}(\Psi)$ is the desired combinatorial mapping and that $J \circ \hat{I}^{-1}(\Lambda)$ is the desired merging sequence. This completes the proof in the pseudo-Anosov case.

If h is isotopic to the identity, one can construct by hand an A-admissible combinatorial mapping Γ such that $\Gamma \cap A = \emptyset$ and $FPN(\Gamma, A) = 1$.

If h is isotopic to a nontrivial periodic mapping, choose B_0 as in the pseudo-Anosov case and then fix a metric on F so that each component of B_0 is a proper geodesic arc and h is isotopic to an isometry ρ with this metric. If B_0 contains an arc τ such that $\rho(\tau) = \tau$, replace τ by an arc parallel to τ. This ensures that $\rho^{-1}(B_0)$ is taut relative to B_0. Since each fixed point class of ρ is connected a small deformation will produce an embedding having exactly $N(h)$ fixed points. Now proceed as in the pseudo-Anosov case to get the desired combinatorial mapping.

In the reducible case first divide C_h into two classes: those parallel to the boundary C_∂ and the rest, C_{int}. For each $C \in C_{int}$ let N_C be a regular neighborhood of C. Since C_h has minimal geometric intersection with A we can assume the N_C are chosen so that the original handle structure on F induces a handle structure on each component of $F \backslash \bigcup N_C$, as well as on each N_C.

On each h-invariant component Y of $F \backslash \bigcup N_C$ get an embedding ρ as in the pseudo-Anosov or periodic case. Thus, $\rho^{-1}(A \cap Y)$ is taut and ρ has $N(h|_Y)$ fixed points. On pieces that are permuted by h define ρ so that $\rho^{-1}(A \cap Y)$ is taut. For each $C \in C_{int}$ which is

h-invariant extend ρ to N_C as follows: first, set $\rho(C) = C$ so that $\rho^{-1}(A \cap C) \cap A = \emptyset$ and $\mathrm{Fix}(\rho|_C)$ consists of two points if orientation-reversing and empty otherwise. Second, extend to $N_C \backslash C$ so that each component of $\rho^{-1}(A) \cap N_C \backslash C$ is taut and no fixed points occur in $N_C \backslash C$. These two conditions are easy to guarantee. (Since ρ maps ∂N_C away from C, we may assume that ρ maps circles parallel to C away from C. In doing so any violation of tautness can be removed by pulling across A without introducing any fixed points.) Finally, if C is permuted by h, then extend ρ so that tautness is obtained just as above.

It follows from the results in [8] that ρ has exactly $N(h)$ fixed points. Now follow the proof given in the pseudo-Anosov case—starting with the construction of $\hat{F}$ and reducing curves as needed in $\bigcup N_C$ as well as to a regular neighborhood of C_∂. □

This section is concluded with two variations on the previous theorem. The first is a relative version of the theorem while the second is used to cover the case when the surface has no boundary. The idea behind the second variation is that other collections besides A can be used to locate fixed points.

For a relative version we simply observe that each combinatorial mapping with Γ isotopic to $h^{-1}(A)$ corresponds to a homeomorphism of F which is isotopic to h. Thus, following through the proof of Theorem 3.1 with $N(h)$ replaced by the relative Nielsen number $N_\partial(h)$ [8] we obtain Theorem 3.2 below. In this setting the computation of FPN is slightly different in that indices are computed assuming that boundary is mapped to boundary and interior to interior. Here the notation FPN_∂ will be used.

THEOREM 3.2. *Given h and C_h as described above, there exist an A-admissible combinatorial mapping Γ representing h, which has minimal geometric intersection with A and is isotopic to $h^{-1}(A)$, and a merging sequence Λ such that $N_\partial(h) = FPN_\partial(\Lambda(\Gamma), A)$ and each stable arc in Λ is contained in C_h.*

For the following let F be a compact orientable surface, possibly with empty boundary. Suppose that K is a collection of pairwise disjoint simple closed curves in F which give a pants decomposition for F. That is, each component of $F\backslash K$ is homeomorphic to a three times punctured sphere. If F is closed, a *transversal* to K is a simple closed curve T which has minimal geometric intersection with K and is such that each component of $F\backslash(K \cup T)$ is simply connected. If $\partial F \neq \emptyset$, a transversal is a finite collection of proper arcs satisfying the same property. In place of A we will use the pair (K, T) which we denote by $\mathscr{K}$. In this setting, a combinatorial mapping means the collection $h^{-1}(\mathscr{K})$, where h is a homeomorphism transverse to $\mathscr{K}$ with no fixed points on $K \cup T$. Alternatively, a combinatorial mapping Γ can be thought of as a pair (Γ_K, Γ_T) together with a pair of assignments (μ_K, μ_T). It is said to be taut if each component taken from the pair has minimal geometric intersection with each member of $K \cup T$.

As before, a combinatorial mapping Γ determines a fixed point count which we now denote by $FPN(\Gamma, \mathscr{K})$. The analog of $M(\Gamma, f)$ defined in Section 2 is obtained by finding the index of each component of $F\backslash(\Gamma_K \cup \Gamma_T \cup K \cup T)$. The proof of Theorem 3.1 applies directly using $\mathscr{K}$ in place of A to obtain the following.

THEOREM 3.3. *Let F and $\mathscr{K}$ be as above and let $h: F \to F$ be a homeomorphism with maximal reducing family C_h. Then there exist a taut combinatorial mapping Γ representing h, and a merging sequence Λ, such that $N(h) = FPN(\Lambda(\Gamma), \mathscr{K})$ and each stable arc in Λ is contained in C_h.*

4. THE ALGORITHM

Let F denote a compact surface which we assume has negative Euler characteristic. The computation of the Nielsen number for the surfaces with nonnegative Euler characteristic is straightforward. Let h be a self-homeomorphism of F which for our purposes we regard as being given either as a composition of Dehn twists and periodic homeomorphisms, or by its induced action on the fundamental group (but knowing *a priori* that this action comes from a homeomorphism). Before describing the algorithm we first give some preliminary results needed for its implementation. The most important is the algorithm due to Bestvina and Handel [1]. We state in the following Proposition 4.1 the only part of that algorithm needed for the algorithm given in this section. Other authors have also developed algorithms which find maximal reducing families for orientation-preserving homeomorphisms. See for instance [4] or [11].

PROPOSITION 4.1. *Let F be an orientable surface and h an orientation-preserving self-homeomorphism of F. Then there is an algorithm for finding a maximal reducing family C_h.*

A homeomorphism $h:F \to F$ is said to be *fixed point reduced* if $\mathrm{Fix}(h)$ has exactly $N(h)$ components and each component is either a point or a surface with negative Euler characteristic. Let n_h denote the number of nonorientable surfaces fixed by h and let

$$\omega_h = \max_g\{n_g | g \text{ is isotopic to } h \text{ and is fixed point reduced}\}.$$

LEMMA 4.2. *Suppose F is nonorientable and let $\tilde{F}$ be the orientable surface which double covers F. Let p be a nontrivial involution of $\tilde{F}$ corresponding to the covering. Given any homeomorphism $h:F \to F$ and a lift $\tilde{h}:\tilde{F} \to \tilde{F}$*

$$N(h) = 1/2(N(\tilde{h}) + N(p \circ \tilde{h}) + \omega_h).$$

Proof. We follow the proof of [3; Theorem 2.5] when restricted to self-mappings of F. Given $h:F \to F$ let h_0 be a homeomorphism isotopic to h and in standard form as in [8]. Perturb h_0 slightly to a homeomorphism g which is fixed point reduced and such that $n_g = \omega_h$ [8; Corollary 3.5]. Each of the nonorientable surface components of $\mathrm{Fix}(g)$ is covered by exactly one component from $\mathrm{Fix}(\tilde{g}) \cup \mathrm{Fix}(p \circ \tilde{g})$, where $\tilde{g}$ is a lift of g. This gives exactly one essential Nielsen class and a contribution of one towards ω_h. All other components of $\mathrm{Fix}(g)$ are covered either by two components of $\mathrm{Fix}(\tilde{g})$ or by two components of $\mathrm{Fix}(p \circ \tilde{g})$. By construction, these two components must be in distinct Nielsen classes. As each is essential and the contribution towards ω_h is zero, the result follows. □

LEMMA 4.3. *Let $\mathscr{C}$ be a maximal reducing family for h^2, then $\mathscr{C}$ is also a maximal reducing* family for h.

Proof. Without loss, we assume that $\mathscr{C}$ and h are chosen so that $\mathscr{C}$ meets $h(\mathscr{C})$ transversely and each pair of components has minimal geometric intersection. Let $C \in \mathscr{C}$ such that $h(C) \cap C \neq \emptyset$. If either domain of $F \backslash \mathscr{C}$ bordered by C is a pseudo-Anosov piece, then the sequence $\mathit{length}(h^n(h(C))) \to \infty$ as $n \to \infty$. Similarly, if each of the domains is periodic and h acts by a nontrivial Dehn twist along C. But this is a contradiction as $\mathit{length}(h^n(h(C))) = \mathit{length}(h^{n+1}(C)) \leq K$ where K is the maximal length of a member of $\mathscr{C} \cup h(\mathscr{C})$. It follows that $\mathscr{C} \cup h(\mathscr{C}) = \emptyset$ and thus, by the maximality of $\mathscr{C}$, that $h(\mathscr{C})$ is isotopic to $\mathscr{C}$. The result now follows. □

We are now in a position to give the algorithm for computing $N(h)$.

ALGORITHM 4.4 ($N(h)$ for homeomorphisms of bounded surfaces). *Given a compact surface F with nonempty boundary and a homeomorphism $h: F \to F$ the integer $N(h)$ can be computed using the following procedure*:

Step 1. If F is nonorientable, construct the double covering by an orientable surface and use Lemma 4.2 *to compute $N(h)$. For the remainder of the algorithm we assume that F is orientable.*

Step 2. Appeal to Proposition 4.1 *and Lemma* 4.3 *to get a maximal reducing family C_h for h.*

Step 3. Introduce a handle structure on F to obtain A. List all of the combinatorial mappings, up to A-equivalence, which satisfy the hypothesis of Theorem 3.1. *If a handle structure with only one 0-handle is used, then they are in one-to-one correspondence with the set of components of A which are isotopic to some component of $h^{-1}(A)$.*

Step 4. For each Γ from Step 3, it follows from Theorem 3.1 *that any stable arc for Γ must lie on a collection $\hat{C}_h$ which is A-equivalent to C_h. Thus it is possible to list all of the relevant merging sequences for Γ. For each one, compute the value of FPN.*

Step 5. Repeat Step 4 for each combinatorial mapping from Step 3. By Theorem 3.1 *the minimal value among all the FPNs is equal to $N(h)$.*

For closed surfaces the above algorithm does not apply. Instead we use the variation leading to Theorem 3.3. This variation could also be used for surfaces with boundary but it turns out that Algorithm 4.4 is much easier to implement than Algorithm 4.5.

ALGORITHM 4.5. ($N(h)$ for homeomorphisms of closed surfaces). *Given a closed surface F and a homeomorphism $h: F \to F$ the integer $N(h)$ can be computed using the following procedure*:

Step 1. Same as in Algorithm 4.4.

Step 2. Same as in Algorithm 4.4.

Step 3. Choose a pants decomposition K and a transversal T as defined in Section 3. To facilitate computations one can make choices so that $C_h \subset K$. List all of the combinatorial mappings which satisfy the hypothesis of Theorem 3.3.

Step 4. As in Algorithm 4.4 *list all of the relevant merging sequences for a given combinatorial mapping Γ. Note that if $C_h \subset K$, then the merging sequences do not intersect Γ_K. Compute $FPN(\Lambda(\Gamma), \mathscr{K})$ for each such Λ.*

Step 5. Same as in Algorithm 4.4.

Remark. If $C_h \neq \emptyset$ in the case of a closed surface, one could use Algorithm 4.4 applied to each invariant piece, together with the results of [8], to obtain $N(h)$.

5. SOME EXAMPLES

In this section we present two examples which illustrate some of the features of Algorithm 4.4. For the sake of brevity the calculation involved in finding maximal reducing families is omitted. Also, the homeomorphisms in the examples are defined in terms of combinatorial mappings. If a homeomorphism h is given as a composition of Dehn twists and periodics, then a combinatorial mapping satisfying the hypothesis of Theorem 3.1 can be found by computing $h^{-1}(A)$ and then isotoping to get minimal geometric intersection with A. On the other hand, if the induced action on the fundamental group is given, first find some Γ that represents the homotopy class. Now, if Γ has more components than A, the fact that we are dealing with a mapping class ensures that two components with the same μ-value can be joined by an arc whose interior is disjoint from Γ. By fusing these components together along the arc, and removing arcs parallel to the boundary, one reduces the number of components. Repeat this process until the desired combinatorial mapping is reached.

Example 5.1. Let F denote the twice punctured torus. It has a handle structure consisting of three 1-handles attached to a disk D. Let $A_1, \ldots, A_6$ be the attaching arcs for the handles as indicated by the dashed segments in Fig. 1(a). For each $1 \le k \le 3$, let H_k denote the 1-handle which meets A_{2k-1} and A_{2k}. The three proper arcs $\delta_1, \delta_2, \delta_3$ appearing in Fig. 1(a) determine an orientation-preserving homeomorphism h of F as follows: thicken δ_i to a disk Δ_i which meets ∂F in two arcs. Define h by first mapping the pair $(\Delta_i, \Delta_i \cap \partial F)$ homeomorphically onto $(H_i, H_i \cap \partial F)$. Up to isotopy there are four ways to do this. Choose the one that preserves the orientation on F and also the direction of the arrows given in the figure. One can check that $h|_{\cup \Delta_i}$ extends to all of F by mapping the remaining domain onto D.

An application of the Bestvina–Handel algorithm shows that h is irreducible and, in fact, gives a pseudo-Anosov mapping class. As defined in Section 2, the combinatorial mapping Γ_h consists of, for each i, two arcs running parallel to δ_i. It satisfies the hypothesis of Theorem 3.1 and is one of two that must be considered in order to use the theorem to compute $N(h)$. The other A-admissible combinatorial mapping, Γ', is obtained from the same construction, but with δ_3 appearing on the other side of the handle H_3.

We first compute $FPN(\Gamma_h, A)$. The graph G constructed in the proof of Proposition 2.2 is indicated in Fig. 1(b). The vertices are labelled $x_0, \ldots, x_7$ with x_7 lying in the region between two curves in Γ_h parallel to δ_1. In general, each edge of G must cross A exactly once. Thus, one of the vertices must be x_7. The orientation of δ_1 determines x_2x_7 and x_7x_5 as the edges of G. A direct index calculation on the isolated vertices yields

$$i(x_0) = i(x_6) = 0$$

$$i(x_1) = i(x_3) = i(x_4) = -1.$$

(a)

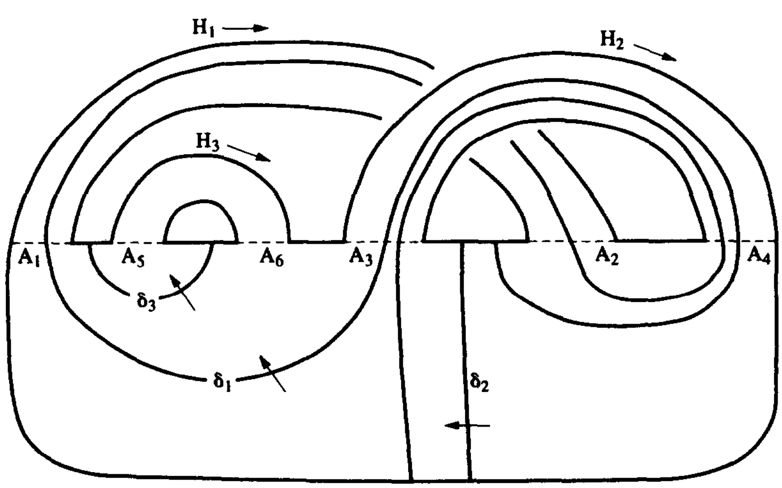

Fig. 1.

To optimize on the remaining 1-dimensional component of G simply arrange that h is defined so that it maps the point $A_1 \cap h^{-1}(A_1)$ to the right (in the figure) along A_1. This makes $i(x_2) = 0$. Similarly, map $A_2 \cap h^{-1}(A_2)$ to the left. Then $i(x_7) = 0$ and $i(x_5) = -1$ and so $FPN(\Gamma_h, A) = 4$.

According to Theorem 3.1 we need only find those minimal stable arcs which are parallel to ∂F. Inspection shows that there are three: one from x_3 to x_1 with A-itinerary $(4, 3)$; one from x_3 to x_6 with A-itinerary $(1, 2)$; the third from x_4 to x_6 with A-itinerary $(3, 4)$. Let λ denote the one from x_3 to x_1. The combinatorial mapping $\lambda(\Gamma_h)$ is depicted in Fig. 1(c). Except for the part obtained by pushing Γ_h along λ, each curve in the figure corresponds to two parallel arcs in $\lambda(\Gamma_h)$. Also in the figure are the vertices $y_0, \ldots, y_9$ of the corresponding graph needed for computing FPN. Note that $y_0, \ldots, y_7$ correspond naturally to $x_0, \ldots, x_7$ in Fig. 1(b). The 1-dimensional components for this graph are $y_2 y_7 y_5$ and

(b)

(c)

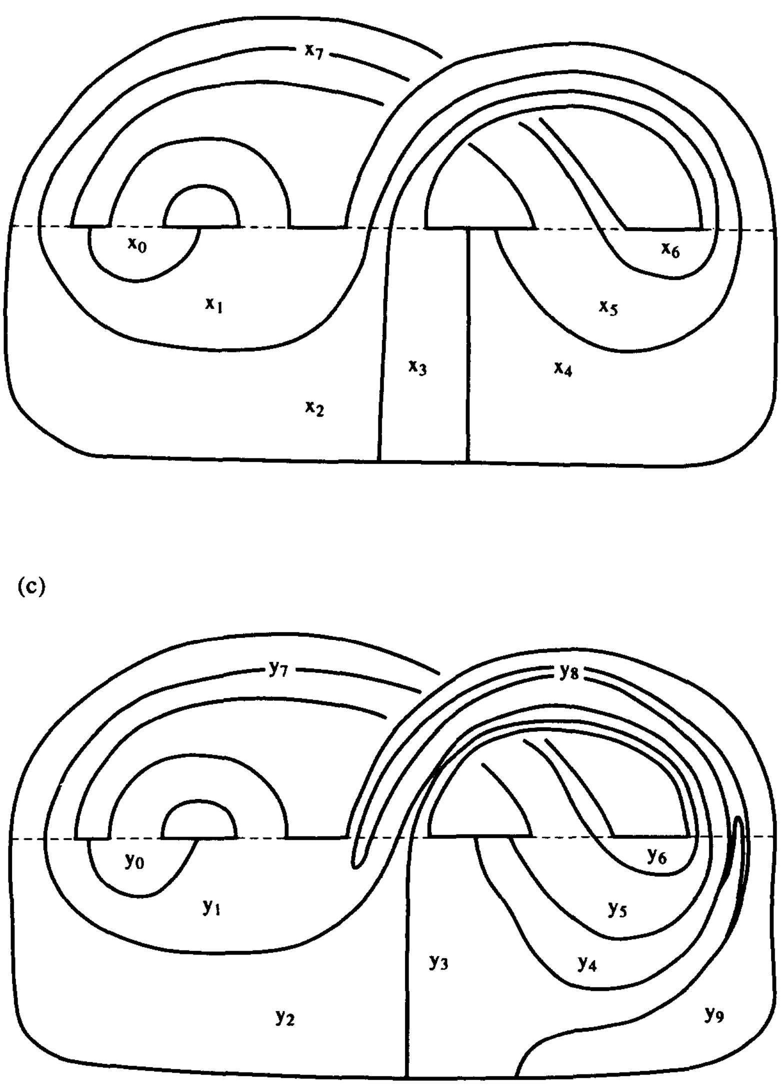

Fig. 1(c) (*continued*)

$y_3y_8y_1$. Similar to those for the x's, the index calculations for this combinatorial mapping are

$$i(y_0) = i(y_6) = i(y_9) = i(y_2) = i(y_7) = 0$$

$$i(y_4) = i(y_5) = -1.$$

For the new component $y_3y_8y_1$ it is possible to arrange that

$$i(y_3) = i(y_8) = 0 \quad \text{and} \quad i(y_1) = -2.$$

In the language of Nielsen theory, the stable arc λ had the effect of joining the Nielsen equivalent fixed points x_1 and x_3 and forming a single fixed point of index -2.

To finish the computation of $N(h)$ we remark that neither of the other two stable arcs will have the effect of reducing the value of FPN. Also, due to the symmetry, the same calculations occur when Γ' is used in place of Γ_h. Thus, $N(h) = FPN(\lambda(\Gamma_h), A) = 3$.

Example 5.2. For this example F denotes the three times punctured torus. A handle structure is determined by $A_1, \ldots, A_8$ as indicated in Fig. 2. Let H_k denote the 1-handle which meets A_{2k-1} and A_{2k}. Similar to Example 5.1, a homeomorphism g is determined by the oriented arcs $\zeta_1, \ldots, \zeta_4$ given in Fig. 2. This time choices are made so that g is orientation-reversing. The Bestvina–Handel algorithm applied to g^2 shows that the simple closed curve with A-itinerary $(1, 2, 4, 3, 2, 1, 3, 4)$ is a reducing curve which reduces F into two components. The mapping class g^2 reduces to the identity on the "outside" (this is a 4 times punctured sphere) and is pseudo-Anosov on the punctured torus "inside". By Lemma 4.3, this same curve acts as the reducing family for g.

To compute $FPN(\Gamma_g, A)$ let $r_1, \ldots, r_8$ denote the vertices of G which lie in the 0-handle (Fig. 2). There is one other vertex r_9, which lies in H_2, and two resulting edges r_4r_9 and r_9r_7. After adjusting g appropriately a calculation yields

$$i(r_1) = i(r_2) = i(r_7) = 1$$

with all others being equal to zero. Thus, $FPN(\Gamma_g, A) = 3$.

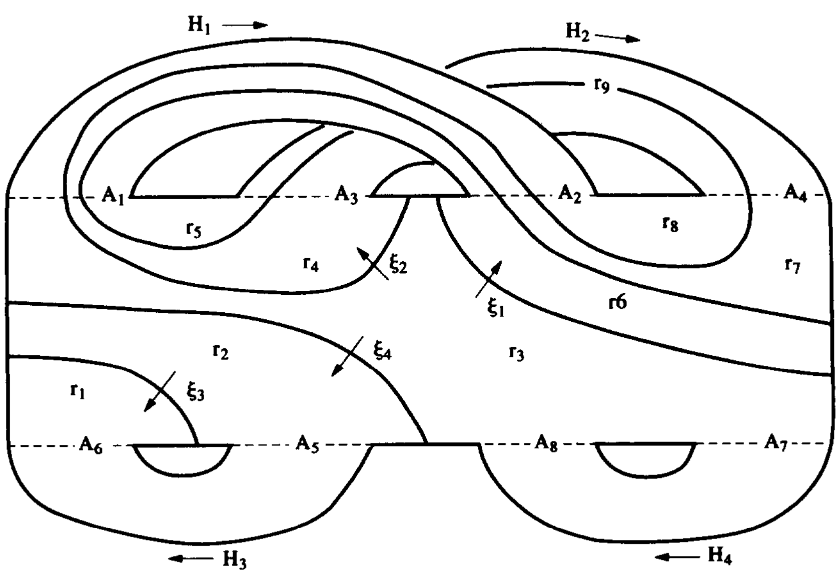

Fig. 2.

The combinatorial mapping Γ_g has exactly two nontrivial stable arcs which satisfy the hypothesis of Theorem 3.1. One goes from r_3 to r_8 with A-itinerary $(3, 4)$; the other goes from r_6 to r_5 with A-itinerary $(4, 3)$. But as neither have an endpoint in common, and neither joins two vertices having nonzero index, the value of FPN cannot be reduced by a merging sequence taken from the two. Finally, as in the previous example, there is one other combinatorial mapping satisfying the hypothesis of Theorem 3.1. It is obtained by isotoping ζ_3 to the other side of H_3. It can be checked that all of the above computations for Γ_g are identical in this case and hence, $N(g) = FPN(\Gamma_g, A) = 3$.

REFERENCES

1. M. BESTVINA and M. HANDEL: Train tracks for surface homeomorphisms, to appear in *Topology*.
2. R. F. BROWN: *The Lefschetz fixed point theorem*, Scott-Foresman, Chicago, (1971).
3. R. DOBRENKO and J. JEZIERSKI: The coincidence Nielsen number on non-orientable manifolds, *Rocky Mountain Journal of Math.* **23** (1993), 67–85.
4. J. FRANKS and M. MISIUREWICZ: Cycles for disk homeomorphisms and thick trees, *Contemp. Math.* **152** (1993), 69–139.
5. R. GEOGHEGAN and A. NICAS: Parametrized Lefschetz–Nielsen fixed point theory and Hochschild homology traces, *Amer. J. Math.* **116** (1994), 397–446.
6. J. HARER and R. PENNER: *The combinatorics of train tracks*, Annals of Math. Studies, No. **125**, Princeton University Press (1992).
7. B. J. JIANG: *Lectures on Nielsen fixed point theory*, Contemporary Mathematics **14**, Amer. Math. Soc., Providence, RI (1983).
8. B. JIANG and J. GUO: Fixed points of surface diffeomorphisms, *Pacific Journal of Math.* **160** (1993), 67–89.
9. M. R. KELLY: Minimizing the number of fixed points for self-maps of compact surfaces, *Pacific Journal of Math.* **126** (1987), 81–123.
10. M. R. KELLY: Minimizing the cardinality of the fixed point set for self-maps of surfaces with boundary, *Michigan Math. J.* **39** (1992), 201–217.
11. J. E. LOS: Pseudo-Anosov maps and invariant train tracks in the disc: a finite algorithm, *Proc. London Math. Soc.* **66** (1993), 400–430.
12. C. K. MCCORD: Computing Nielsen numbers, *Contemp. Math.* **152** (1993), 249–267.
13. W. P. THURSTON: On the geometry and dynamics of diffeomorphisms of surfaces, *Bull. AMS* **19** (1988), 417–431.

Department of Mathematics and Computer Science
Loyola University
6363 St Charles Avenue
New Orleans
LA 70118
USA