Ergod. Th. & Dynam. Sys. (1997), **17**, 1393–1408
Printed in the United Kingdom © *1997 Cambridge University Press*

A bound on the fixed-point index for surface mappings

MICHAEL R. KELLY

Department of Mathematics and Computer Science, Loyola University, New Orleans, Louisiana, 70118, USA

(*Received* 31 *May* 1996 *and accepted in revised form* 1 *October* 1996)

Abstract. In this paper we consider the class of surface mappings consisting of those maps which have the least number of fixed points possible among all maps in their homotopy class. When the surface has non-empty boundary, we show that for mappings in this class the index of a fixed point is bounded above by 1 and below by $2\chi - 1$. This generalizes a well known result for pseudo-Anosov homeomorphisms. A proof of a Jiang–Guo type inequality is also given.

1. *Introduction*

One aspect of topological dynamics is to be able to make qualitative statements about the periodic points of some class of continuous mappings defined on some restricted class of spaces. This paper is concerned with local information which can be obtained as a result of the restricted setting. In particular, information about the index of an isolated fixed point of the mapping. See [**B**] for the definition of the fixed-point index.

The simplest such case might be that of an expanding map on a compact manifold, where the index of each fixed point is ± 1 depending only on the dimension of the manifold. An analogous result for hyperbolic linear maps, or more generally, Anosov mappings of the n-torus $S^1 \times \cdots \times S^1$, (see, for example, [**KH**]) is that each of the fixed points has the same index, $I = \pm 1$. In this setting the number of fixed points of a map f is easily computed using the formula

$$NF(f) = |L(f)|$$

where $L(f)$ denotes the Lefschetz number of the mapping, and $NF(f)$ the number of fixed points of f. Generalizations of this idea can be found in the work of a number of authors. See [**M**] for a survey of some of these results. Another setting is that of the area-preserving homeomorphisms of a compact 2-manifold. When the surface in question is topologically an annulus and the homeomorphism is smooth, Simon [**S**] showed that the index of a fixed point is bounded above by one. Pelikan and Slaminka [**PS**] removed the smoothness hypothesis and generalized the result to all orientation- and area-preserving homeomorphisms on orientable surfaces.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

Another direction is provided by the class of pseudo-Anosov homeomorphisms on a compact surface defined by Thurston [**T**]. Here, given $h : F \to F$, it follows from the classical Euler–Poincaré formula for singular foliations of surfaces (see [**FLP**, Exposé 5]) that for each fixed point x of h

$$2\chi(F) - 1 \leq \text{index}(x) \leq 1,$$

where $\chi(F)$ denotes the Euler characteristic of F. This result was extended by Jiang and Guo [**JG**] to a broader class of surface homeomorphisms. Their class has the feature that any given homeomorphism is isotopic to some member. In addition, each member has exactly the number of fixed points predicted by the Nielsen number. See [**B**] or [**J**] for the definition and properties of the Nielsen number.

Remark. Actually, more is true for the surface homeomorphisms cited above. The lower bound of $2\chi(F) - 1$ can only occur for a pseudo-Anosov homeomorphism on a closed surface, and all additional fixed points have index ± 1. In general, the sum of the indices of all fixed points which have index less than -1 is bounded below by $4\chi(F)$. Also, the lower bounds can be improved when the surface has boundary.

The purpose of this paper is to generalize the above result about homeomorphisms of surfaces to a fairly natural class of self-mappings of surfaces. This generalization for surfaces is optimal in some sense and also points out the essential topological feature needed for such a result.

Let F be a surface; a compact two-dimensional manifold with or without boundary. The Euler characteristic of F is denoted by $\chi(F)$. Given a continuous self-mapping $f : F \to F$ we let $NF(f)$ denote the cardinality of the set of fixed points of f. We say that f is *fixed-point minimal* (or simply minimal) if $NF(f)$ is equal to the least number of fixed points possible among all maps homotopic to f.

THEOREM A. *Let F be a surface with non-empty boundary and let f be a fixed-point minimal self-mapping. For each fixed point x of f,*

$$2\chi(F) - 1 \leq \text{index}(x) \leq 1.$$

As remarked earlier, the lower bound of $2\chi(F) - 1$ does occur for homeomorphisms, but only when the surface is closed. The following simple example illustrates that, for self-mappings, this lower bound is the best possible whenever the surface has non-empty boundary.

Example 1. Let F be a surface with non-empty boundary. Let W denote a wedge of n circles $\omega_1, \ldots, \omega_n$ embedded in F and based at a point p. We assume that n and W are chosen so that the inclusion $i : W \to F$ is a homotopy equivalence. Define a map $f : W \to W$ by $\omega_i \mapsto \omega_i\omega_i$ for each i, and such that $\text{Fix}(f) = p$. Extend to a map $f : F \to W$. Define $g : F \to F$ by $g(x) = i(f(x))$. Clearly, p is the only fixed point of g so g is fixed-point minimal. Since $L(g) = 2\chi(F) - 1$, it follows that $\text{index}(p) = 2\chi(F) - 1$.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

The inequality given in Theorem A has a variation which is more algebraic in nature. Namely, that the index of a Nielsen fixed-point class is bounded below by $2\chi - 1$ and above by 1. These two formulations are the same for homeomorphisms [**JG**] but not necessarily for self-mappings. When the surface in question has non-empty boundary Wagner [**W**] defines a special class of mappings called W-characteristic. The author then shows that this class is quite large (in a probabilistic sense) and that these bounds for the index of a Nielsen class hold for all maps in this class.

Given a surface F with non-positive Euler characteristic and a homeomorphism $h : F \to F$, Jiang and Guo [**JG**] use the estimates on the indices for their class of surface homeomorphisms to prove the inequality

$$|L(h) - \chi(F)| \leq N(h) - \chi(F),$$

where $N(h)$ denotes the Nielsen number of h. Thus, the Lefschetz number serves as a lower bound for the number of fixed points of a homeomorphism. In [**JG**] the authors raise the question as to whether or not this same inequality holds for all self-mappings of surfaces. For surfaces with non-empty boundary, Wagner [**W**] gives evidence for an affirmative answer to this question by showing that the inequality holds for the same class of self-mappings mentioned above. In this paper we prove that a weak version of the Jiang–Guo inequality holds in the case when the surface has non-empty boundary; namely, the following.

THEOREM B. *Let F be a surface with non-empty boundary. Then for any self-mapping $f : F \to F$,*

$$|L(f) - \chi(F)| \leq NF(f) - \chi(F).$$

§2–5 of this paper are devoted to the proofs of Theorems A and B. The idea behind these proofs is to define a fairly technical class of maps so that each homotopy class contains a member which is fixed-point minimal and, at the same time, index calculations are feasible. §2 gives some notation needed for the proof. In §3 we present a method for computing indices of fixed points. In §4 this special class of mappings, called index reduced, is defined and relevant properties are established. Finally, in §5 the proofs of Theorems A and B are given.

The method of proof relies quite heavily on the fact that the surface can be described with a handle structure consisting of 0-handles and 1-handles, but no 2-handles. It is well known that all surfaces with boundary have such a handle decomposition, while no closed surfaces admit such. On the other hand, the results corresponding to Theorems A and B are known to be true for the torus and Klein bottle, and as all maps on surfaces with negative Euler characteristic have |degree| ≤ 1, it seems quite reasonable that these results hold for closed surfaces (other than S^2 and RP^2). Evidence for this is given below in Theorem E, where a very natural (but at this time unknown) topological result concerning surface mappings is shown to be sufficient to prove these results. The proof of Theorem E is given in §6.

CONJECTURE C. *The statements of Theorems A and B hold when F is a closed surface with non-positive Euler characteristic.*

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

CONJECTURE D. *Let F be a closed surface with negative Euler characteristic and let $f : F \to F$ be a self-mapping which is not homotopic to a homeomorphism. Then there is a fixed-point minimal map g, homotopic to f, such that $g(F) \neq F$.*

THEOREM E. *Conjecture D $\Longrightarrow$ Conjecture C.*

2. *Definitions and notation*

Let D be a two-dimensional disk and let F denote the surface obtained by attaching the 1-handles $H_1, \ldots, H_k$ to D:

$$F = D \cup H_1 \cup H_2 \cup \cdots \cup H_k.$$

Let $\{A_{2i-1}, A_{2i}\}$, $1 \leq i \leq k$, denote the attaching region for H_i. For our purposes each A_j is viewed as a proper arc in F. Let $A = \bigcup_{j=1}^{2k} A_j$.

Let $f : F \to F$ be a continuous mapping. Throughout this paper we will assume that $\mathrm{Fix}(f)$ is a finite set disjoint from A. A further assumption will be that $f^{-1}(A)$ is a proper one-dimensional submanifold of F, transverse to A, and contains no simple closed curves disjoint from A. The following is some terminology regarding f that was introduced in [**K1**] and will be used throughout this paper. The closure of a component, U, of $F \setminus (f^{-1}(A) \cup A)$ is a *critical region* for f if both $f(U)$ and U are contained in the same component of $F \setminus A$. The set of critical regions for f will be denoted by $CR(f)$. Clearly, the union of their interiors contains $\mathrm{Fix}(f)$. The absence of simple closed curves disjoint from A in $f^{-1}(A)$ ensures that critical regions are just two-dimensional disks. The *index* of a critical region, C, is the topological fixed point index of f on the interior of C (see [**B**] or [**J**]). A critical region C is said to be *taut* if no component of $f^{-1}(A) \cap C$ has its endpoints contained in a single component of A.

If α is an arc in $f^{-1}(A)$ let $\partial\alpha$ denote its endpoints. If α is a single point, set $\partial\alpha = \alpha$. For $1 \leq i \leq 2k$, let ∂_i denote the union of the two components of $(\partial F \setminus A) \cup \partial A_i$ which intersect A_i and let T_i denote the collection of connected subsets α of $f^{-1}(A)$ that satisfy (i) $\partial\alpha \subset A_i \cup \partial_i$, and (ii) $\alpha \cap A_j = \emptyset$ when $i \neq j$.

Definitions. Let $i, j \in \{1, \ldots, 2k\}$ and suppose that α is a component of $\cup T_i$ with $f(\alpha) \subset A_j$. The arc α is said to be an *ij-turn* if $\partial\alpha \cap \partial F = \emptyset$ and $\alpha \cap A_i$ consists of an even number of points; α is an *ij-crossing* if $\partial\alpha \cap \partial F = \emptyset$ and $\alpha \cap A_i$ consists of an odd number of points; and α is an *ij-end* if $\partial\alpha \cap \partial F \neq \emptyset$. The terms turn, crossing, and end will be used when the particular indices are not important for the discussion.

The following construction will be used to create a new map $\hat{f}$ which is homotopic to the given map f. Let α denote either a turn, a crossing or an end. If α is an entire component of $f^{-1}(A)$ let $\alpha^+ = \alpha$ and let α_0 be a curve isotopic to α, which is near α and disjoint from A. Otherwise, define α^+ and α_0 as follows. When α is either a turn or a crossing let α^+ be a regular neighborhood of α in $f^{-1}(A)$ with $\alpha^+ \cap A = \alpha \cap A$. Let α_0 be an arc in F such that $\partial\alpha_0 = \partial\alpha^+$, the interior of α_0 is disjoint from $f^{-1}(A) \setminus \alpha^+$ and $\alpha_0 \cap A$ consists of at most one point. When α is an end choose α_0 so that one of its endpoints is in ∂F and so that $\alpha_0 \cap A = \emptyset$. If $\{\alpha_n\}$ is a collection of turns, crossings and ends, this construction can be applied to each member with all choices being made so that the resulting α_0's are pairwise disjoint. Let N denote

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

the set consisting of all turns, crossings and ends for f. Let N^+ (respectively N_0) denote the set containing all the α^+ (respectively α_0). Associate to f any map $\hat{f} : F \to F$ that satisfies $\hat{f}^{-1}(A) = (f^{-1}(A) \setminus N^+) \cup N_0$.

Also associated to a given mapping f, we consider a graph G_f defined in the following manner. For each critical region of f we create one vertex. Vertices u and v are joined by a single edge if and only if the respective critical regions have common to their boundary an endpoint of either an ii-turn, an ii-crossing or an ii-end for some i. The graph G_f embeds naturally in the surface F by placing a vertex in the interior of each critical region and placing each edge so that it passes through the corresponding turn, crossing or end in exactly one point.

Example 2. This example is given to illustrate the definitions, notations and constructions described above. We consider the surface F obtained by removing five open disks from the 2-sphere. A map $f : F \to F$ is indicated in Figure 1, where $f^{-1}(A_i)$ is the curve(s) labelled with the letter i. The strip between the curves labelled $2i-1$ and $2i$ is mapped to the 1-handle H_i while the rest is mapped into D. The critical regions for f are labelled by $a, \ldots, o$ in Figure 1. There are three turns and two ends for this map; an 11-turn, 22-turn, 44-turn, 33-end, and 34-end. Among the many crossings the 55- and 66-crossings are the only ones which contribute to G_f. The graph G_f can be embedded in F as follows. A vertex at each of $a, \ldots, o$ with edges bc, cd, ef, fg, gh, hi and fn, each edge crossing A in one point. By definition there is no edge at the vertex m. In Figure 2, the

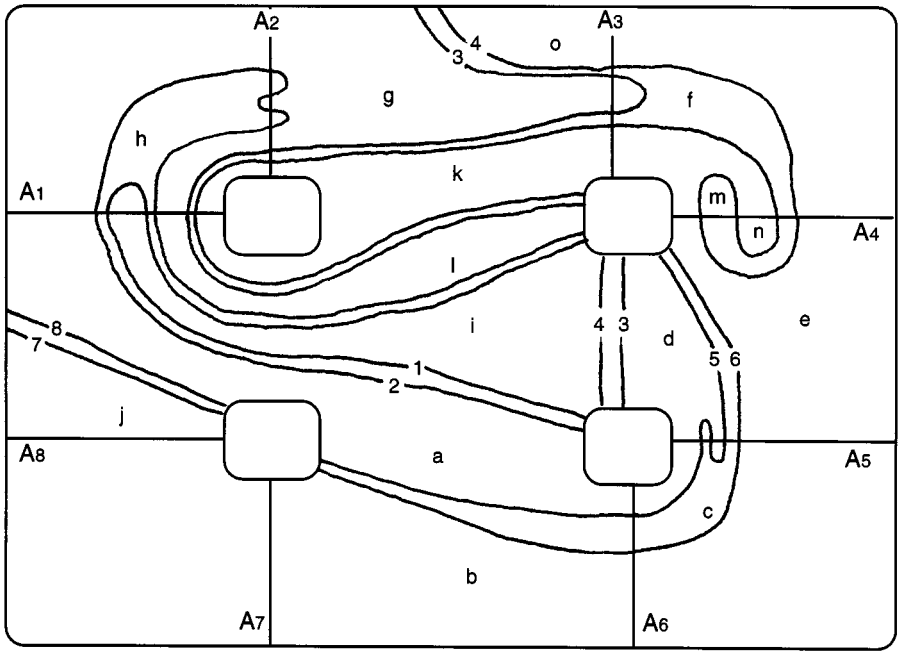


FIGURE 1.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

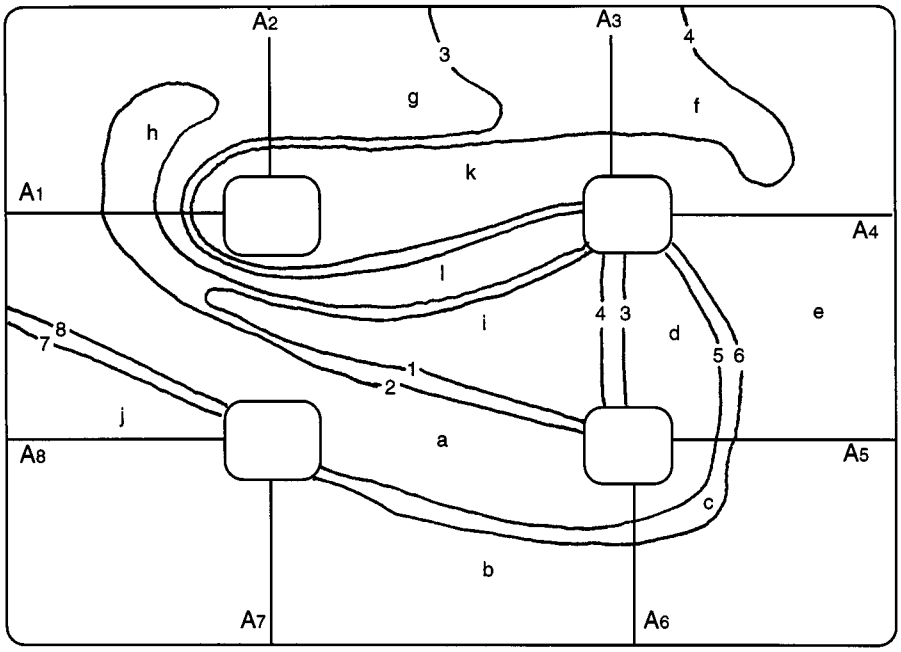


FIGURE 2.

curves $\hat{f}^{-1}(A)$ for an associated map $\hat{f}$ are depicted. Note that the construction removes the critical regions labelled m, n, o and there is a natural correspondence between the remaining regions. This feature is developed further in §4.

Clearly, a map f has at least as many fixed points as the number of critical regions with non-zero index. In fact, there is a map f_0 homotopic to f such that $f_0^{-1}(A) = f^{-1}(A)$ and f_0 has the same number of fixed points as the number of critical regions of f with non-zero index [**K1**; Lemma 2.1]. The graph G_f gives more information about the size of $\mathrm{Fix}(f)$ as indicated in Proposition 2.1 below. The index of a component of G_f is defined to be the sum of the indices of the critical regions corresponding to its vertices. It follows from the definition that any two fixed points taken from critical regions belonging to the same component of G_f are in the same Nielsen fixed-point class. The following proposition gives a minimum type theorem in fixed-point theory.

PROPOSITION 2.1. *Given $f : F \to F$, let M denote the number of path components of G_f with non-zero index. There is a map $h : F \to F$ homotopic to f which has exactly M fixed points. Moreover, G_h is isomorphic to G_f.*

Proof. Let C be a component of G_f and let T be a spanning tree in C. Let V denote the set of vertices, and E the set of edges of T. Fix a vertex $z \in V$. Since T is a tree, there is a unique bijection $\eta : E \to (V \setminus \{z\})$ which has the property that $\eta(e)$ is an endpoint

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

of e for each $e \in E$. Apply Lemma 2.2 below to each $e \in E$ so that $\text{index}(X_e) = 0$, where X_e is the natural critical region corresponding to $\eta(e)$.

Repeat the above for each component of G_f. This gives a map g with G_g isomorphic to G_f and such that each component of G_g yields at most one critical region with non-zero index. The desired map h agrees with g away from the interiors of the critical regions and is extended, as in the proof of [**K1**; Lemma 2.1] so that in each critical region with zero index there are no fixed points, and one fixed point in all other critical regions. □

LEMMA 2.2. *Let e be an edge in the graph G_f with endpoints v, w. Let X_v and X_w denote the critical regions for f corresponding to v and w respectively. Let R be a disk in F which contains the turn, crossing, or end corresponding to e, meets each of A and $f^{-1}(A)$ in an arc, and only meets the critical regions X_v and X_w. Then there is a map g which agrees with f on $F \setminus R$ such that* $\text{index}(Y_v) = 0$, *where Y_v is the critical region for g which naturally corresponds to X_v.*

Proof. The map g is defined by first setting $g^{-1}(A) = (f^{-1}(A) \setminus R) \cup L$, where L is some arc in R meeting A in a finite set of points. The region Y_v is the component of $F \setminus (g^{-1}(A) \cup A)$ which meets ∂R in the same set as did X_v. It follows from Proposition 3.1 in the next section that the influence of $g|_R$ on the value of $\text{index}(Y_v)$ is determined by the action of g on $A \cap L$. In fact, if the cardinality of $A \cap L$ is suitably large, one can easily define g on $A \cap L$ so that $\text{index}(Y_v) = 0$. The proof is finished by noting that g extends to all of R so that $L \mapsto A$ and it agrees with f on ∂R. □

3. *Index computations for critical regions*

The purpose of this section is give a combinatorial formula for the index of a critical region based on the description of a map given previously. A special case of the formula, Corollary 3.2, will be used to derive the inequalities in Theorems A and B.

Let Y be a critical region for f and let E denote the handle in the given handle structure that contains Y. Suppose that there exists a subarc ζ of $\partial F \cap E$ such that $Y \cap \zeta = \emptyset$. Fix a homeomorphism $h : E \to [0, 1] \times [0, 1]$ such that $h(\zeta) = (\{1\} \times [0, 1]) \cup ([0, 1] \times \{0, 1\})$. Observe that if p is an endpoint of a component of $f^{-1}(A) \cap \partial Y$, then $h(p)$ and $h(f(p))$ are disjoint and contained in $\{0\} \times [0, 1]$.

Definition. Let γ be a component of $f^{-1}(A) \cap \partial Y$ with $\partial\gamma = \{p, q\}$. Set $h(p) = (0, x_p)$, $h(q) = (0, x_q)$, $h(f(p)) = (0, y_p)$ and $h(f(q)) = (0, y_q)$ and assume without loss that $x_p < x_q$. The arc γ is *contracting* if $y_p > x_p$ and $y_q < x_q$; *expanding* if $y_p < x_p$ and $y_q > x_q$; and *neutral* otherwise.

With the notation given above we now compute the fixed-point index of a critical region. Let Y be a critical region for f with the curve ζ as described above. There is a unique component γ_ζ of $f^{-1}(A) \cap \partial Y$ with the property that γ_ζ separates $Y \setminus \gamma_\zeta$ from ζ in E.

PROPOSITION 3.1. *Suppose that Y is a critical region for f and ζ is a subarc of $\partial F \cap E$ as described above. Suppose further that Y is homeomorphic to a disk. Let h be a*

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

homeomorphism as defined above. Set $\iota_0 = 1$ *if* γ_ζ *is contracting,* -1 *if expanding and* 0 *if neutral. Let* ι_1 *denote the number of contracting components of* $(f^{-1}(A) \cap \partial Y) \setminus \gamma_\zeta$ *and let* ι_2 *denote the number of expanding components. Then*

$$\text{index}(Y) = \iota_0 - \iota_1 + \iota_2.$$

Proof. The definition of index(Y), given in terms of h, can be expressed as follows. Fix an orientation on E. The frontier of Y, denoted by ∂Y inherits an induced orientation. The homeomorphism h induces an orientation on $[0,1] \times [0,1] \subset \mathbb{R}^2$ and hence on the unit sphere S^1 in $\mathbb{R}^2$. Define the unit vector field $\phi : \partial Y \to S^1$ on ∂Y by

$$\phi(x) = \frac{h(x) - h(f(x))}{|h(x) - h(f(x))|}.$$

Since ∂Y is a simple closed curve the degree of ϕ is well defined and is equal to index(Y).

As $(0,1) \in S^1$ is a regular value, the degree of ϕ can be computed by considering the action of ϕ on a neighborhood of each point in $\phi^{-1}((0,1))$. To do so, first observe that $\phi^{-1}((0,\pm 1))$ consists of all of the endpoints of the components of $\partial Y \cap f^{-1}(A)$. If a component is either contracting or expanding, $\phi^{-1}((0,1))$ contains exactly one of its endpoints. If neutral, then either none or both of the endpoints belong to $\phi^{-1}((0,1))$. Now if γ is a component other than γ_ζ, then one can check that ϕ is locally orientation reversing at the point $\gamma \cap \phi^{-1}((0,1))$ when γ is contracting, and locally orientation preserving when γ is expanding. If γ is neutral and two points occur, then ϕ reverses orientation at one and preserves orientation at the other. For the component γ_ζ the local action is the opposite; orientation preserving for contracting, reversing for expanding. This yields a total count of $\iota_0 - \iota_1 + \iota_2$ for the degree of ϕ. □

Our use of the above proposition will be in the restricted case when the critical region Y is taut. We also include the case when the arc ζ does not exist and so the proposition cannot be applied.

COROLLARY 3.2. *Let* Y *be a taut critical region. If* Y *does not contain all of* $\partial F \cap E$*, then* index$(Y) = \iota_0 - \iota_1$*, where* ι_0 *and* ι_1 *are as defined above. If* Y *contains all of* $\partial F \cap E$*, then* index$(Y) = 1$.

Proof. Directly from the definition we see that if γ is expanding then both of its endpoints must be contained in a single component of A. Hence, when Y does not contain all of $\partial F \cap E$, each component of $f^{-1}(A) \cap \partial Y$ is either contracting or neutral. If Y is taut and contains all of $\partial F \cap E$, then $Y = E$ and so $f(Y) \subset Y$. □

The following example illustrates the computation of the fixed-point index of a critical region. It also points out a feature of the map $\hat{f}$ that will be used later: a relative ease of computation for the index.

Example 3. We continue with Example 2 and focus on the critical region labelled i in Figure 1. Choose ζ to be the segment of the outer boundary curve from A_4 to A_5. Then γ_ζ is the curve labelled 4. Thus, $f(\gamma_\zeta) \subset A_4$ and so γ_ζ is neutral. For the other two components of $f^{-1}(A) \cap i$, the computation depends on how f maps the two points of

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

contact with A_1. One can check that each component is either contracting or neutral, and so from Proposition 3.1, we can only conclude from Figure 1 that $|\text{index}(i) + 1| \leq 1$.

On the other hand, the calculation for the critical region i in Figure 2 is more exact. Here γ_ζ is the same and the other component of $\hat{f}^{-1}(A) \cap i$ is contracting. Hence, $\text{index}(i) = -1$.

4. *Index reduced maps*

In this section we define a special class of self-mappings and prove various results about these mappings. Let F and A be as defined in §2.

Definition. The map $f : F \to F$ is said to be *index reduced* if it satisfies properties (a1)–(a5) below.

(a1) $f^{-1}(A)$ is a one-dimensional proper submanifold of F transverse to A, with no fixed points occurring on A.
(a2) Each simple closed curve in $f^{-1}(A)$ intersects A.
(a3) For each open set $\mathcal{U}$ meeting $f^{-1}(A)$, $f(\mathcal{U})$ meets at least two components of $F \setminus A$.
(a4) If α is an ij-turn or end, then $i = j$. If α is an ij-crossing with $i \neq j$, then α is a single point.
(a5) If a critical region meets exactly one component of A, then its frontier must contain an entire component of $f^{-1}(A)$.

Example 4. The map f given in Example 2 is not index reduced, but almost so. Each of (a1)–(a3), holds whereas the 34-end violates (a4) and the three critical regions m, n, o each violate (a5). It is index reduced everywhere else, including the area around i. This allowed for the computation of $\text{index}(i)$ as seen in Example 3. Without increasing the number of fixed points, f can be deformed into an index reduced map by moving the 34-end off A_3 and by simplifying the 44-turn. More generally, the following theorem holds.

THEOREM 4.1. *Given $g : F \to F$ there exists a map f homotopic to g which is index reduced and fixed-point minimal.*

In order to prove Theorem 4.1 we need some of the results from [**K1**]. These are summarized in the next two lemmas using the language and notation of this paper. The first lemma gives some fairly elementary results taken from [**K1**, Lemmas 3.2 and 3.6] while the second is part of the main technical lemma [**K1**, Lemma 3.5].

LEMMA 4.2. *Given a map f there exists a map g homotopic to f such that: $\#\text{Fix}(g) \leq \#\text{Fix}(f)$; $g^{-1}(A)$ is a one-dimensional proper submanifold of F transverse to A; g has no fixed points on A; for each open set $\mathcal{U}$ meeting $g^{-1}(A)$, $g(\mathcal{U})$ meets at least two components of $F \setminus A$; and each simple closed curve in $g^{-1}(A)$ meets A.*

Let α be a subarc of $f^{-1}(A)$ with one endpoint x lying on some A_i, the other endpoint y on $\partial F \cup A_i$ and such that $\alpha \cap (A \cup \partial F) = \{x, y\}$. There is a unique disk D_α such that ∂D_α consists of α together with a subarc J_α from $\partial F \cup A_i$. Let X_α denote the closure of the component of $F \setminus (f^{-1}(A) \cup A)$ that contains α and lies in D_α. We assume that X_α is a disk.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

LEMMA 4.3. *Let f be a map with an arc $\alpha \subset f^{-1}(A)$ as above. Suppose that either $f(\alpha) \not\subset A_i$, or X_α is a critical region, or $X_\alpha \neq D_\alpha$. Then there exist a non-empty subset S_α of $J_\alpha \cap f^{-1}(A)$, and a map g homotopic to f such that g satisfies the conclusion of Lemma 4.2 and $g^{-1}(A) \cap A = (f^{-1}(A) \cap A) \setminus S_\alpha$.*

Remark. Only the case $y \in A_i$ is proved in [**K1**]. The other case, when $y \in \partial F$, can be proved using the same argument.

Proof of Theorem 4.1. Without loss of generality we may assume that g is fixed-point minimal. Properties (a1)–(a3) are contained in Lemma 4.2. If either of (a4) or (a5) fails for g, it is easy to see that the hypotheses of Lemma 4.3 are satisfied. Apply the lemma to get a new minimal map satisfying (a1)–(a3), and strictly reducing the cardinality of $g^{-1}(A) \cap A$. After a finite number of applications of Lemma 4.3 we obtain the desired index reduced map f. □

We now give some features of indexed reduced maps which will be used in the proofs of the main results. Recall the construction of the associated mapping $\hat{f}$ and the graph G_f given in §2. If f is index reduced, then there is a natural bijection ϕ between the set of edges of G_f and the set of arcs α_0 given in the construction of $\hat{f}$. This is exploited to promote G_f to a directed graph as follows. Let e be an edge of G_f with endpoints u and v corresponding, respectively, to critical regions U and V. The oriented edge will be $\vec{uv}$ where u and v are labelled according to the conventions below. If e corresponds to an end or a turn, then $\phi(e)$ intersects exactly one of U or V. We assume that u and v are labelled so that $\phi(e)$ intersects U. If e corresponds to a crossing, let p denote the point of intersection of $\phi(e)$ and A_i, and let C denote the component of $A_i \setminus p$ that contains $\hat{f}(p)$. The map $\hat{f}$ has two critical regions which contain the point p. Let Y denote the one that meets C in an arc. Label u and v so that V and Y lie in the same component of $F \setminus A$.

Suppose that X is a critical region for f corresponding to a vertex x of G_f. Let $t_o(X)$ denote the number of edges going out of x that correspond to a turn, and let $t_i(X)$ be the number of edges coming into x that correspond to a turn. Analogously, define $c_o(X), c_i(X), e_o(X)$ and $e_i(X)$ corresponding to crossings and ends.

Another useful consequence of property (a5) is that if X is a critical region for f, then there is a unique critical region Y for $\hat{f}$ such that $X \cap Y \neq \emptyset$. Define a map

$$\psi : CR(f) \to CR(\hat{f})$$

by $\psi(X) = Y$.

LEMMA 4.4. *If $f : F \to F$ is an index reduced map, then the map ψ defined above is a bijection. Furthermore, for each $X \in CR(f)$, $\psi(X)$ is homeomorphic to a disk.*

Proof. The fact that ψ is onto follows directly from the construction of $\hat{f}$. To see that ψ is one-to-one, suppose there exists distinct critical regions X_1 and X_2 such that $\psi(X_1) = \psi(X_2)$. Then there is an arc σ contained in A_i for some i, such that one endpoint is in X_1, the other is in X_2 and $\partial\sigma = \partial\tau$ for some turn τ. By (a4), τ is an

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

ii-turn, and by (a3), the disks bounded by σ and τ which lie on the opposite side of A_i from X_1 must be critical regions. But this contradicts (a5).

To prove the second statement suppose that $\psi(X)$ is not a disk. Then $f^{-1}(A)$ contains a simple closed curve bounding a disk B such that $B \setminus A$ has at most one component whose closure intersects more than one component of A. By (a4) no such component can exist, and from (a5) it follows that $B \cap A = \emptyset$. This contradicts (a2). □

LEMMA 4.5. *Let $\mathcal{X}$ denote the set of critical regions corresponding to the vertices of a component of G_f. Then*

$$\sum_{X \in \mathcal{X}} \text{index}(X) = \sum_{X \in \mathcal{X}} \text{index}(\psi(X)).$$

Proof. There exists a connected open set $\mathcal{O}$ which contains $\mathcal{X} \cup \psi(\mathcal{X})$ and is disjoint from all other critical regions for f. Moreover, there is a map f' which agrees with f on $F \setminus \mathcal{O}$ and with $\hat{f}$ on $\mathcal{O}$. This map is homotopic to f by a homotopy with support on $F \setminus \mathcal{O}$. □

5. *Proofs of Theorems A and B*

LEMMA 5.1. *Let $f : F \to F$ be an index reduced map and let X be a critical region for f. Then* $\text{index}(\psi(X)) \leq 1 - t_o(X) - c_o(X) - e_o(X)$.

Proof. If $\psi(X)$ is not taut, its frontier meets $\hat{f}^{-1}(A)$ in a single arc which is either an ij-turn or ij-end with $i \neq j$. A direct calculation shows that $\text{index}(\psi(X)) = 0$. For an index reduced map such a $\psi(X)$ can only occur when $t_o(X) + e_o(X) = 1$ and $c_o(X) = 0$.

The rest of the proof concerns the case when $\psi(X)$ is taut, where by Lemma 4.4, the methods of §3 can be applied to compute $\text{index}(\psi(X))$. In the exceptional case when $\psi(X)$ contains all of $\partial F \cap E$, where E is as defined in §3, the result follows as $t_o(X) = c_o(X) = e_o(X) = 0$. We now assume that $\psi(X)$ does not contain all of $\partial F \cap E$.

First suppose that $t_o(X) > 0$. Choose a component A_i of A that contains one of the turns going out and choose ζ in $\partial F \cap E$ with one endpoint on A_i and disjoint from X. It follows that $\psi(X) \cap \zeta = \emptyset$, so we have the homeomorphism h as defined in §3 (using $Y = \psi(X)$). From our choice of ζ it follows that γ_ζ is neutral and so by Corollary 3.2, $\text{index}(\psi(X)) = -\iota_1$. Now let γ be a component of $(\hat{f}^{-1}(A) \cap \partial\psi(X))$ other than γ_ζ, and suppose that γ corresponds to an edge going out of the vertex corresponding to X. Then γ corresponds to a jj-turn, end or crossing for f, for some j. If γ corresponds to a turn, then by property (a5) γ must separate A_j and γ_ζ in the handle containing $\psi(X)$. It then follows that γ is contracting. If γ corresponds to an end, then again γ must separate A_j and γ_ζ so γ is contracting. If γ corresponds to a crossing, the convention for orienting edges ensures that γ is contracting. Except for γ_ζ all edges going out are contracting so

$$\iota_1 \geq c_o(X) + t_o(X) + e_o(X) - 1,$$

and the result follows.

The argument in the case that $e_o(X) > 0$ is similar. For the last case we assume that $t_o(X) + e_o(X) = 0$. Choose some ζ. If γ_ζ is contracting, then it does not correspond to a crossing going out. So $\iota_0 = 1$ and $\iota_1 \geq c_o(X)$. If γ_ζ is neutral, then $\iota_1 \geq c_o(X) - 1$. In either case $\text{index}(\psi(X)) \leq 1 - c_o(X)$. □

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

Let X denote the closure of a component of $F \setminus (A \cup f^{-1}(A))$. We assign to X two non-negative quantities: the number of sides and the number of borders. The number of *sides* of X, denoted by $s(X)$, is defined to be the number of components of $f^{-1}(A) \cap X$ which meet two distinct components of $(A \cup \partial F) \setminus (A \cap \partial F)$ with the property that the closures of these two components are disjoint. To find the number of *borders*, $b(X)$, let $n_1(X)$ denote the number of components of $f^{-1}(A) \cap X$ which meet two distinct components of $(A \cup \partial F) \setminus (A \cap \partial F)$ with the property that the closures of these two components intersect. Let $n_2(X)$ denote the number of points of $A \cap \partial F$ in X. Set $b(X) = \frac{1}{2}(n_1(X) + n_2(X))$. Immediate from these definitions we have the following.

LEMMA 5.2. *If $s(X) > 0$, then $s(X) + b(X) \geq 2$.*

Let $\mathcal{Y}$ denote the set of components of $F \setminus (A \cup f^{-1}(A))$ for which the number of sides is greater than zero. Let m denote the cardinality of this set.

LEMMA 5.3. $\sum_{Y \in \mathcal{Y}} (b(Y) + s(Y)) = 2m - 2\chi(F)$.

Proof. Let S be a component of $F \setminus A$ and let $k(S)$ be the total number of sides counted in $\mathcal{Y} \cap S$. Embed a graph $G(S)$ in S as follows. Put one vertex in each member of $\mathcal{Y} \cap S$ and join two vertices by an edge if and only if the corresponding regions have a side in common. Let $e(S)$ denote the number of edges and $v(S)$ the number of vertices. As $G(S)$ contains no cycles it follows that $e(S) + 1 = v(S)$. Also, since $k(S) = 2e(S)$, we have that $k(S) = 2v(S) - 2$. Summing over all $2 - \chi$ components of $F \setminus A$ we get

$$\sum_{Y \in \mathcal{Y}} s(Y) = \sum_{S} k(S) = 2m - 2(2 - \chi).$$

Finally, noting that the total number of borders for regions in $\mathcal{Y}$ is the same as the number of points in $A \cap \partial F$ we get

$$\sum b(Y) = 2(2 - 2\chi). \qquad \square$$

LEMMA 5.4. *Let $f : F \to F$ be an index reduced map and suppose that X is a critical region for f such that $\psi(X)$ is taut. Then*

$$\operatorname{index}(\psi(X)) \geq 1 + c_i(X) - s(\psi(X)).$$

Proof. If no arc ζ (as in §3) exists the result follows immediately from Corollary 3.2. Otherwise, choose some ζ and a homeomorphism h as in §3. First we consider the case when γ_ζ is neutral. In this case $\iota_0 = 0$ and also γ_ζ does not correspond to a crossing coming in. Among the remaining sides of $\psi(X)$ each crossing coming in is neutral and so

$$\iota_1 \leq s(\psi(X)) - 1 - c_i(X).$$

Now suppose that γ_ζ is contracting. Then $\iota_0 = 1$. If γ_ζ corresponds to a crossing coming in, then we have

$$\iota_1 \leq (s(\psi(X)) - 1) - (c_i(X) - 1).$$

Otherwise we have

$$\iota_1 \leq (s(\psi(X)) - 1) - c_i(X).$$

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

In all three cases above we see that

$$\iota_0 - \iota_1 \geq 1 + c_i(X) - s(\psi(X)).$$

The result now follows from Corollary 3.2. □

LEMMA 5.5. *Let X be a critical region for the index reduced map $f : F \to F$. Then*

$$s(X) \geq 1 + c_i(X) + t_o(X) - \text{index}(\psi(X)).$$

Proof. If $\psi(X)$ is taut, then this follows immediately from Lemma 5.4 and the fact that $s(X) - s(\psi(X)) = t_o(X)$. If $\psi(X)$ is not taut, then for an index reduced map $s(X) = 1 + t_o(X)$, $c_i(X) = 0$ and $\text{index}(\psi(X)) = 0$ which gives the desired inequality. □

LEMMA 5.6. *Let $\mathcal{X}$ denote the set of critical regions for f corresponding to the vertices of a component of G_f. Then*

$$\sum_{X \in \mathcal{X}} b(X) \geq \sum_{X \in \mathcal{X}} e_o(X).$$

Proof. For each end-edge of a component of G_f let U and V be the two critical regions corresponding to the endpoints. The end contributes one towards each of $n_1(U) + n_1(V)$ and $n_2(U) + n_2(V)$. □

Proof of Theorem B. First observe that proving Theorem B is equivalent to showing that $L(f) \leq NF(f)$ and $L(f) + NF(f) \geq 2\chi(F)$. Since the theorem is clearly true for the disk we assume that $\chi(F) \leq 0$. Also, without loss of generality, we may assume that f is fixed-point minimal. Choose a handle structure for F as in §2. From Theorem 4.1 we may assume that f is also index reduced.

Let C be a component of G_f and let $\mathcal{X}$ denote the set of critical regions for f corresponding to the vertices of C. Let $v(C)$ and $e(C)$ denote, respectively, the number of vertices and edges of C. Then clearly we have

$$\sum_{X \in \mathcal{X}} (1 - t_o(X) - c_o(X) - e_o(X)) = v(C) - e(C).$$

Since C is connected, $v(C) - e(C) \leq 1$. Thus by Lemmas 4.5 and 5.1,

$$\text{index}(\mathcal{X}) = \sum_{X \in \mathcal{X}} \text{index}(X) \leq 1. \tag{1}$$

Now sum (1) over all $\mathcal{X}$ such that $\text{index}(\mathcal{X}) \neq 0$ to get

$$\sum \text{index}(\mathcal{X}) \leq NF(f),$$

and since the left-hand side of this inequality contains all of the fixed-point index for f we have the desired inequality

$$L(f) \leq NF(f).$$

To verify the other inequality for Theorem B first apply Lemma 5.5 to get

$$\sum_{X \in \mathcal{X}} s(X) \geq v(C) + \sum_{X \in \mathcal{X}} (c_i(X) + t_o(X)) - \sum_{X \in \mathcal{X}} \text{index}(\psi(X)).$$

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

Hence, by Lemmas 4.5 and 5.6,

$$\sum_{X\in\mathcal{X}}(s(X)+b(X)) \geq v(C)+\sum_{X\in\mathcal{X}}(c_i(X)+t_o(X)+e_o(X))-\sum_{X\in\mathcal{X}}\text{index}(X), \quad (2)$$

and since the middle summation counts the number of edges in C,

$$\sum_{X\in\mathcal{X}}(s(X)+b(X)) \geq v(C)+e(C)-\text{index}(\mathcal{X}) \geq 2v(C)-1-\text{index}(\mathcal{X}).$$

Thus,

$$1+\text{index}(\mathcal{X}) \geq 2v(C)-\sum_{X\in\mathcal{X}}(s(X)+b(X)). \quad (3)$$

Summing over each such $\mathcal{X}$ with $\text{index}(\mathcal{X})\neq 0$ we get

$$NF(f)+L(f) \geq 2\sum v(C)-\sum(s(X)+b(X)). \quad (4)$$

The summations on the-right hand side of (4) are assumed to be over the correct regions.

Now let $\mathcal{Z}$ denote the set of components X of $F\setminus(A\cup f^{-1}(A))$ which satisfy either of the following: (i) X is not a critical region but has at least one side; or (ii) X is a critical region where the corresponding $\mathcal{X}$ has index zero. Note that when f is minimal each critical region in $\mathcal{Z}$ has index zero. Also, by Lemma 5.2, for each $X\in\mathcal{Z}$ we have

$$0 \geq 2-(s(X)+b(X)). \quad (5)$$

Finally, sum the inequality (5) over all regions in $\mathcal{Z}$ and combine with (4) to obtain the inequality

$$NF(f)+L(f) \geq 2\sum_{C} v(C)+\sum_{X\in\mathcal{Z}}2-\sum_{X\in\mathcal{Y}}(s(X)+b(X)).$$

Because every critical region has at least one side, the first two summations on the right-hand side of this inequality reduces to twice the cardinality of $\mathcal{Y}$. Thus by Lemma 5.3, the right-hand side is just $2m-(2m-2\chi)$, which completes the proof of Theorem B. □

Proof of Theorem A. Let f be fixed-point minimal and let $x\in\text{Fix}(f)$. Let $\mathcal{X}$ denote the set of critical regions which corresponds to a component of G_f and contains x. By Proposition 2.1, x is the only fixed point in $\mathcal{X}$ and so, by (1), the right-hand inequality in Theorem A is proved.

For the left-hand inequality use (3) only for this particular $\mathcal{X}$, otherwise use (5) for all other regions in $\mathcal{Y}$ to get

$$1+\text{index}(x) \geq \sum_{X\in\mathcal{Y}}2-\sum_{X\in\mathcal{Y}}(s(X)+b(X)).$$

As in the proof of Theorem B, the right-hand side reduces to $2\chi(F)$. □

6. *The closed surface case*

To prove Theorem E we first need the following minor variation of Theorem B resulting from a restricted hypothesis.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

PROPOSITION 6.1. *Let F be a surface with negative Euler characteristic and exactly one boundary component. Suppose that $f : F \to F$ is such that $f(\partial F)$ is null homotopic. Then $NF(f) + L(f) \geq 2\chi(F) + 2$.*

Proof. Fix a handle structure for F. As in the proof of Theorem B we assume that f is an index-reduced minimal mapping. Without loss we assume that $f(\partial F)$ is a single point disjoint from A. Thus $f^{-1}(A)$ consists only of simple closed curves. If any of these curves is isotopic to ∂F, [**K2**, Lemma 2.2] can be used to convert an outermost such curve into two arcs each being isotopic (rel endpoints) to an arc in ∂F. These arcs are removed as in [**K2**, Lemma 2.1]. As a result we can assume that $f^{-1}(A)$ does not contain any simple closed curves isotopic to ∂F. This in turn ensures that there exists a component X of $F \setminus (A \cap f^{-1}(A))$ which is contained in D, and whose closure contains at least four points of $A \cap \partial F$.

Now either X is a critical region with index$(\mathcal{X}) \neq 0$ or $X \in \mathcal{Z}$, where $\mathcal{X}$ and $\mathcal{Z}$ are as given in §5. In the former case we see that

$$\sum_{X \in \mathcal{X}} b(X) \geq 2 \quad \text{while} \quad \sum_{X \in \mathcal{X}} e_o(X) = 0.$$

This creates a difference of at least two in (2). This difference carries through the rest of the argument giving the desired result.

In the case when $X \in \mathcal{Z}$ the argument depends on the number of sides of X. If $s(X) \geq 2$, then we get

$$-2 \geq 2 - (s(X) + b(X))$$

in (4) which, as before, carries a difference of at least two through the argument. Finally, if $s(X) = 1$, some A_j must be contained in the closure of X. Let $A_{j'}$ be such that A_j and $A_{j'}$ form the attaching region for one of the 1-handles, and let Y denote the component of $F \setminus (A \cap f^{-1}(A))$ whose closure contains $A_{j'}$. Note that Y is also contained in D. As Y also contains at least four points of $A \cap \partial F$, the arguments given above allow us to assume that $Y \in \mathcal{Z}$ and $s(Y) = 1$. If $X \neq Y$, then we have that

$$-1 \geq 2 - (s(X) + b(X)) \quad \text{and} \quad -1 \geq 2 - (s(Y) + b(Y)),$$

which leads to the desired result. On the other hand, if $X = Y$, then this region contains at least six points of $A \cap \partial F$ and so

$$-2 \geq 2 - (s(X) + b(X)).$$ □

The proof given above immediately gives us the following variation of Theorem A.

PROPOSITION 6.2. *Let F and f be as in Proposition 6.1 and suppose that f is minimal. For each fixed point x of f, $2\chi(F) + 1 \leq \text{index}(x) \leq 1$.*

Proof of Theorem E. Conjecture C is known to be true when F is either the torus or the Klein bottle, so we assume that $\chi(F) < 0$. As the result is known in the case when f is homotopic to a homeomorphism [**JG**], we assume otherwise. According to Conjecture D, there is a minimal map f_0 which is homotopic to f and such that $f_0(F) \cap E = \emptyset$ for some disk E in F. As in §5, it suffices to prove the result for the map f_0.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press

Let $G = F \setminus \mathrm{int}(E)$ and let $g = f_0|_G$. By Proposition 6.1,

$$NF(g) + L(g) \geq 2\chi(G) + 2.$$

Since g and f_0 have the same fixed point set

$$NF(f_0) + L(f_0) \geq 2(\chi(F) - 1) + 2 = 2\chi(F).$$

From Theorem B we have $NF(g) \leq L(g)$, and so

$$NF(f_0) \leq L(f_0).$$

It follows from Proposition 6.2 that

$$2\chi(F) - 1 \leq \mathrm{index}(x) \leq 1$$

for each fixed point x of f. □

REFERENCES

[B] R. F. Brown. *The Lefschetz Fixed Point Theorem*. Scott-Foresman, Chicago, 1971.

[FLP] A. Fathi, F. Laudenbach and V. Poénaru. Travaux de Thurston sur les surfaces. *Asterisque* **66–67** (1979).

[J] B. Jiang. *Lectures on Nielsen Fixed Point Theory (Contemporary Mathematics 14)*. American Mathematical Society, Providence, RI, 1983.

[JG] B. Jiang and J. Guo. Fixed points of surface diffeomorphisms. *Pacific J. Math.* **160** (1993), 67–89.

[KH] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications 54)*. Cambridge University Press, Cambridge, 1995.

[K1] M. R. Kelly. Minimizing the number of fixed points for self-maps of compact surfaces. *Pacific J. Math.* **126** (1987), 81–123.

[K2] M. R. Kelly. The relative Nielsen number and boundary-preserving surface maps. *Pacific J. Math.* **161** (1993), 139–153.

[M] C. K. McCord. Computing Nielsen numbers. *Nielsen Theory and Dynamical Systems (Contemporary Mathematics 152)*. Ed. C. K. McCord. American Mathematical Society, Providence, RI, 1993, pp. 249–267.

[PS] S. Pelikan and E. E. Slaminka. A bound for the fixed point index of area-preserving homeomorphisms of two-manifolds. *Ergod. Th. & Dynam. Sys.* **7** (1987), 463–479.

[S] C. Simon. A bound for the fixed-point index of an area-preserving map with applications to mechanics. *Inventiones Math.* **26** (1974), 187–200.

[T] W. P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* **19** (1988), 417–431.

[W] J. Wagner. An algorithm for calculating the Nielsen number on surfaces with boundary. *Preprint*, UCLA, 1995.

https://doi.org/10.1017/S014338579709768X Published online by Cambridge University Press