

## FIXED POINTS OF SURFACE DIFFEOMORPHISMS

BOJU JIANG AND JIANHAN GUO

We give a complete proof of the following theorem which was conjectured by Jakob Nielsen for closed oriented surfaces.

**THEOREM.** *Let  $f: M \rightarrow M$  be a homeomorphism of a compact surface. When  $M$  is closed, then  $f$  is isotopic to a diffeomorphism with  $N(f)$  fixed points, where  $N(f)$  is its Nielsen number. When  $M$  has boundary,  $N(f)$  should be replaced by the relative Nielsen number  $N(f; M, \partial M)$  defined by Schirmer.*

Another result is the inequality  $|L(f) - \chi(M)| \leq N(f) - \chi(M)$  when  $\chi(M) < 0$ , where  $L(f)$  is the Lefschetz number and  $\chi(M)$  is the Euler characteristic.

**Introduction.** For a self-map  $f$  of a compact polyhedron  $X$ , the Nielsen number  $N(f)$  is defined to be the number of essential fixed point classes. (See [J3] for an introduction to the Nielsen fixed point theory.) It is a classical theorem of Wecken [W] that  $N(f)$  is a lower bound of the number of fixed points for all maps homotopic to  $f$ , and that if  $X$  is a manifold of dimension  $\geq 3$ , this lower bound is always realizable (see also [Br], [K]). It is now known [J4] that when  $X$  is a surface with negative Euler characteristic, there exists a map  $f: X \rightarrow X$  such that every map homotopic to  $f$  has more than  $N(f)$  fixed points. The purpose of this paper is to show that for homeomorphisms of surfaces the Nielsen number is indeed the least number of fixed points in the isotopy class, as Nielsen himself conjectured (cf. [N2, §31]) in his study of oriented closed surfaces.

**MAIN THEOREM.** *Let  $M$  be a compact surface, closed or with boundary. Let  $f: M \rightarrow M$  be a homeomorphism. Then  $f$  is isotopic to a smooth embedding which has  $N(f)$  fixed points. If, in addition, no boundary component of  $M$  is mapped onto itself by  $f$  in an orientation-reversing manner, then  $f$  is isotopic to a diffeomorphism having  $N(f)$  fixed points.*

This theorem was announced in [J2], here strengthened with smoothness considerations. An example in [J1] shows it is necessary

to allow embeddings in order to get as few as  $N(f)$  fixed points when there are orientation-reversing invariant boundary components. If we insist on diffeomorphisms, the least number of fixed points in the isotopy class turns out to be the relative Nielsen number  $N(f; M, \partial M)$  defined by Schirmer [S]. See Theorem 5.1 below. Another result is an interesting inequality (Theorem 4.1) relating the Nielsen number with the Lefschetz number.

All the recent progress on Nielsen's fixed point conjecture originated in Thurston's theory of surface diffeomorphisms [T]. Thurston himself solved the important case of orientation-preserving pseudo-Anosov maps in Theorem 6 of [T]. See [BK] for the first published proof. The easy periodic case was treated in [J1]. The work [I] considered orientation-preserving maps of closed orientable surfaces and, besides pseudo-Anosov and periodic cases, hinted at the kind of analysis needed for reducible maps. In the present paper, we shall consider general diffeomorphisms of compact surfaces, and give complete proofs.

There are seven connected compact surfaces with positive or zero Euler characteristic. For them the truth of the Main Theorem can be checked case by case. The 2-sphere, the 2-disk and the annulus have respectively 2, 2 and 4 isotopy classes of self-homeomorphisms. The real projective plane has only one isotopy class, the Möbius band has two (see [E, Theorems 5.5 and 5.8]). All these isotopy classes have obvious simple representatives satisfying the requirements of the Main Theorem. The torus and the Klein bottle are more interesting. They are studied in Nielsen's classical paper [N1]. Both have the Euclidean plane  $E^2$  as the universal covering space with isometric covering translations. Each isotopy class contains a linear representative, i.e. one that lifts to a linear map of  $E^2$ . Such a representative, slightly perturbed if necessary to remove inessential fixed point classes, minimizes the number of fixed points.

Henceforth we assume that  $M$  is a compact surface such that each of its connected components has negative Euler characteristic. Our proof of the Main Theorem is based on Thurston's classification of surface homeomorphisms.

**THURSTON THEOREM ([T]).** *Every homeomorphism  $f : M \rightarrow M$  is isotopic to a "diffeomorphism"  $\varphi$  such that either*

(1)  *$\varphi$  is an isometry with respect to some hyperbolic metric on  $M$ , or equivalently,  $\varphi$  is a periodic map, i.e.  $\varphi^m = \text{id}$ ; or*

(2)  $\varphi$  is a pseudo-Anosov map, i.e. there is a number  $\lambda > 1$  and a pair of transverse measured foliations  $(\mathfrak{F}^s, \mu^s)$  and  $(\mathfrak{F}^u, \mu^u)$  such that  $\varphi(\mathfrak{F}^s, \mu^s) = (\mathfrak{F}^s, \frac{1}{\lambda}\mu^s)$  and  $\varphi(\mathfrak{F}^u, \mu^u) = (\mathfrak{F}^u, \lambda\mu^u)$ ; or

(3)  $\varphi$  is a reducible map, i.e. there is a system of disjoint simple closed curves  $\Gamma = \{\Gamma_1, \dots, \Gamma_n\}$  in  $\text{int}M$  such that  $\Gamma$  is invariant by  $\varphi$  (but the  $\Gamma_i$ 's may be permuted) and  $\Gamma$  has a  $\varphi$ -invariant tubular neighborhood  $\mathcal{N}(\Gamma)$  such that each component of  $M - \mathcal{N}(\Gamma)$  has negative Euler characteristic and on each (not necessarily connected)  $\varphi$ -component of  $M - \mathcal{N}(\Gamma)$ ,  $\varphi$  satisfies (1) or (2).

**REMARK.** The pseudo-Anosov map in [T] is not an honest diffeomorphism in that it is not even  $C^1$  at the singularities of  $\mathfrak{F}^s$  and  $\mathfrak{F}^u$ . In the statement given in [T],  $\Gamma_i$  are two-sided simple closed curves on  $M$ , so some components of  $M - \mathcal{N}(\Gamma)$  may be Möbius bands. We prefer to allow one-sided  $\Gamma_i$  in order to guarantee that every component of  $M - \mathcal{N}(\Gamma)$  has negative Euler characteristic.

The proof of this theorem can be found in [B], [FLP], [HT] for the oriented surfaces, and in [Wu] for non-orientable surfaces.

The structure of the paper is as follows. In §§1–2 we develop standard forms for  $\varphi$  on the periodic pieces and the pseudo-Anosov pieces respectively. These are not meant to be representatives of the isotopy classes that have the minimal number of fixed points, rather they are designed to be building blocks that are ready to be glued together. In §3 these models are assembled into a standard form for a general  $\varphi$ , where we shall regard the case (3) in Thurston Theorem as the general case by regarding the first two cases as “reducible” with empty reducing curves ( $\Gamma = \emptyset$ ). The rest of §3 is devoted to the detailed analysis of the fixed point classes of this standard form. This is a technical stepping-stone of the paper. The Main Theorem is then proved in §4 by shrinking these fixed point classes to single points via isotopy. The inequality relating the Nielsen number and the Lefschetz number is also proved there. A discussion on the relative Nielsen numbers is given in §5.

There is another obvious division of the problem into three cases:

- (1) orientation preserving homeomorphisms on oriented surfaces,
- (2) orientation reversing homeomorphisms on oriented surfaces,
- (3) homeomorphisms on nonorientable surfaces.

Although we have chosen Thurston’s structural trichotomy as the organizing principle of the paper, the above division is certainly the

natural order of understanding. The orientation preserving and reversing cases involve complementary types of fixed point classes, while the nonorientable case combines both. So we suggest the reader to first focus on the orientation preserving case which contains the main ideas with only half of the technicalities, then the orientation reversing case and finally the nonorientable case. For the convenience of the reader, we systematically use the superscripts  $+/-$  in the labels to indicate the relevance to orientation preserving/reversing cases.

The following notation and terminology will be used throughout the paper.

**NOTATION.** Let  $f: X \rightarrow X$  be a map. Then  $\text{Fix } f$  denotes the fixed point set  $\{x \in X \mid x = f(x)\}$ . When  $X$  is a polyhedron and  $A \subset X$  is such that  $A \cap \text{Fix } f$  is both open and closed in  $\text{Fix } f$ , then  $\text{index}(f, A)$  denotes the fixed point index of  $A \cap \text{Fix } f$ .

**DEFINITION.** Let  $f: X \rightarrow X$  be a map. A subset  $A \subset X$  is said to be  $f$ -invariant if  $f(A) \subset A$ . Two path-connected  $f$ -invariant subsets  $A_0, A_1$  are said to be  $f$ -related if there is a path  $c: I \rightarrow X$  such that  $c \simeq f \circ c: I, 0, 1 \rightarrow X, A_0, A_1$ .

In general this is not an equivalence relation among  $f$ -invariant subsets. When both  $A_0$  and  $A_1$  are single points, it reduces to the Nielsen equivalence relation between fixed points of  $f$ . Namely, two fixed points  $x$  and  $y$  are in the same fixed point class if there is a path  $c: I \rightarrow X$  connecting them such that  $c \simeq f \circ c$  rel endpoints.

**1. The periodic case.** Suppose  $\varphi: M \rightarrow M$  is periodic, i.e.  $\varphi^m = \text{id}$  for some natural number  $m$ . It is well known that such a  $\varphi$  is an isometry with respect to some hyperbolic metric on  $M$  (of constant Gaussian curvature  $-1$  and with totally geodesic boundary).

In this section, the connected components of  $\text{Fix } \varphi$  are explicitly described, and the fixed point classes of  $\varphi$  are identified with these components.

**LEMMA 1.1.** *Let  $A$  be a component of  $\text{Fix } \varphi$ . Then one of the following is true:*

- (1)<sup>+</sup>  $A$  is a component of  $M$  and  $\text{index}(\varphi, A) = \chi(A) < 0$ .
- (2)<sup>-</sup>  $A$  is a closed geodesic, with a neighborhood diffeomorphic to  $S^1 \times (0, 1)$  where  $\varphi$  acts as the reflection  $(z, t) \mapsto (z, 1 - t)$ ;  $\text{index}(\varphi, A) = 0$ .
- (3)<sup>-</sup>  $A$  is a geodesic arc orthogonally connecting two (not necessarily distinct) components of  $\partial M$ , with a neighborhood diffeomor-

phic to  $I \times (0, 1)$  where  $\varphi$  acts as the reflection  $(s, t) \mapsto (s, 1 - t)$ ;  $\text{index}(\varphi, A) = 1$ .

(4)<sup>+</sup>  $A$  is a point, with a neighborhood diffeomorphic to the open unit disc  $\text{int}D = \{z \in \mathbb{C} \mid |z| < 1\}$  where  $\varphi$  acts as a rotation;  $\text{index}(\varphi, A) = 1$ .

*Proof.* It is clear that each component of the fixed point set of an isometry is a properly embedded totally geodesic submanifold. Hence the four possibilities. The index is easily calculated from the local description of  $\varphi$ . □

**LEMMA 1.2.** *Let  $A_0, A_1$  be either a fixed point of  $\varphi$  or a  $\varphi$ -invariant component of  $\partial M$ . Suppose  $A_0$  and  $A_1$  are  $\varphi$ -related via a path  $c: I, 0, 1 \rightarrow M, A_0, A_1$ . Then there is a path  $\gamma$  in  $\text{Fix } \varphi$  such that  $\gamma \simeq c: I, 0, 1 \rightarrow M, A_0, A_1$ .*

*Proof.* Hyperbolic geometry guarantees a unique shortest geodesic in every homotopy class of paths  $I, 0, 1 \rightarrow M, A_0, A_1$ . (Recall that  $\partial M$  is totally geodesic.) Let  $\gamma$  be the shortest geodesic homotopic to  $c$ . Since  $\varphi$  is isometric,  $\varphi \circ \gamma$  is the shortest geodesic homotopic to  $\varphi \circ c$ . But  $c \simeq \varphi \circ c$  so  $\gamma = \varphi \circ \gamma$ . This means  $\gamma$  is in  $\text{Fix } \varphi$ . □

**COROLLARY 1.3.** *Fixed point classes of  $\varphi$  are connected.* □

**2. The pseudo-Anosov case.** Suppose  $\varphi: M \rightarrow M$  is a pseudo-Anosov map with stable and unstable measured foliations  $(\mathfrak{F}^s, \mu^s)$  and  $(\mathfrak{F}^u, \mu^u)$  respectively, and with expansion constant  $\lambda > 1$ .

The classical model for  $\varphi$  is described in §2.1. It is isotoped in §2.2 to a standard form  $\bar{\varphi}$  which is smooth and moreover, in view of our later need in §3 of gluing periodic pieces and pseudo-Anosov pieces together, is required to be periodic on  $\partial M$ . In §2.3, the fixed point classes of  $\bar{\varphi}$  are identified as the connected components of  $\text{Fix } \bar{\varphi}$ .

2.1. *The classical model.* The following model is adapted from the description of Thurston’s pseudo-Anosov map  $\varphi$  for closed surfaces (cf. [GK, pp. 176, 182]). The strategy to describe  $\varphi$  near  $\partial M$  is as follows: Collapsing each component of  $\partial M$  into a single point (called a puncture) we obtain a generalized pseudo-Anosov map  $\hat{\varphi}: \widehat{M} \rightarrow \widehat{M}$  of a closed surface  $\widehat{M}$  (cf. [FLP, pp.217,243]). The model for pseudo-Anosov maps works for generalized pseudo-Anosov maps as well by allowing the prong number  $p \geq 1$ . We then recover  $M$  and  $\varphi$  by “blowing up” the punctures of  $\widehat{M}$ .

We first introduce some notations. Let us write  $\ell = \ln \lambda > 0$ . Let  $p$  be a natural number.

**DEFINITION.** On the complex plane  $\mathbb{C} = \{s = s_1 + is_2 \mid s_1, s_2 \in \mathbb{R}\}$ , define the linear map

$$F: \mathbb{C} \rightarrow \mathbb{C}, \quad s_1 + is_2 \mapsto \lambda s_1 + is_2/\lambda.$$

It is the time-one map for the vector field  $V$  defined by  $\dot{s}_1 = \lambda s_1$ ,  $\dot{s}_2 = -\lambda s_2$ , or

$$V(s) = \lambda \bar{s}.$$

**DEFINITION.** On the complex plane  $\mathbb{C} = \{z = \rho e^{i\theta} \mid \rho \geq 0, \theta \in \mathbb{R}\}$ , consider the (multi-valued) map

$$\Phi = \Phi_p: \mathbb{C} \rightarrow \mathbb{C}, \quad z \mapsto z^{p/2}$$

and its inverse

$$\Phi^{-1} = \Phi_p^{-1}: \mathbb{C} \rightarrow \mathbb{C}, \quad s \mapsto s^{2/p}.$$

Consider the map  $\Psi: \mathbb{C} - \text{int } D, S^1 \rightarrow \mathbb{C}, 0$  defined by  $\Psi(z) = z - z/|z|$ . It restricts to a diffeomorphism  $\Psi: \mathbb{C} - D \rightarrow \mathbb{C} - \{0\}$  with inverse

$$\Psi^{-1}: \mathbb{C} - \{0\} \rightarrow \mathbb{C} - D, \quad z \mapsto z + \frac{z}{|z|}.$$

**DEFINITION.** Let  $v = v_p$  be the vector field  $\Phi_*^{-1}V$ . A simple calculation gives

$$v(z) = \frac{2\ell}{p} z \left( \frac{z}{|z|} \right)^{-p} \quad \text{or} \quad v(\rho e^{i\theta}) = \frac{2\ell}{p} \rho e^{i(1-p)\theta}.$$

Let  $f = f_p$  be the time-one map for  $v = v_p$ .

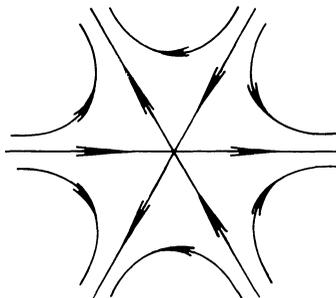


FIGURE 1. The flow of  $v$  and  $\bar{v}$  ( $p = 3$ )

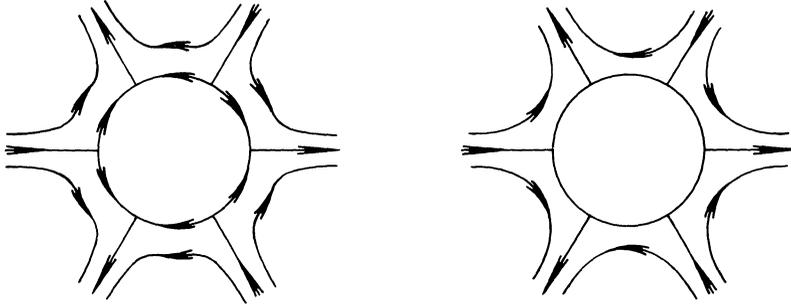


FIGURE 2. The flow of  $v'$  and  $\bar{v}'$  ( $p = 3$ )

Define the vector field  $v'$  on  $\mathbb{C} - D$  by  $v' = \Psi_*^{-1}v$ . The formula for  $v'$  is easily obtained:

$$v'(\rho e^{i\theta}) = \frac{2l}{p} \{ (\rho - 1)e^{i(1-p)\theta} + e^{i(\theta - \pi/2)} \sin p\theta \},$$

so  $v'$  is smoothly extendable to  $\mathbb{C} - \text{int } D$  with  $v'|_{\partial D}$  tangent to  $\partial D$ . Let  $f'$  be the time-one map for  $v'$ . We have a commutative diagram

$$\begin{array}{ccccc} \mathbb{C} & \xleftarrow{\Phi} & \mathbb{C} & \xleftarrow{\Psi} & \mathbb{C} - \text{int } D \\ F \downarrow & & \downarrow f & & \downarrow f' \\ \mathbb{C} & \xleftarrow{\Phi} & \mathbb{C} & \xleftarrow{\Psi} & \mathbb{C} - \text{int } D \end{array}$$

DEFINITION. Let  $r = r_{(p,k)\pm} : \mathbb{C} \rightarrow \mathbb{C}$  be the map

$$r(\rho e^{i\theta}) = \rho e^{\pm i(\theta + 2k\pi/p)}$$

where  $k$  is an integer,  $0 \leq k < p$ . It is simply a rotation or a reflection according as the plus sign or the minus sign is chosen.

*The atlas.* A chart at a point  $x \in \text{int } M$  is an open neighborhood  $U_x$  and an embedding  $u_x : U_x, x \rightarrow \mathbb{C}, 0$ . A chart at a component  $A$  of  $\partial M$  is an open neighborhood  $U_A$  and an embedding  $u_A : U_A, A \rightarrow \mathbb{C} - \text{int } D, S^1$ .

There is a finite smooth atlas  $\mathcal{U}$  of  $M$ , consisting of one chart for each interior singularity, one chart for each boundary component, and some other charts at interior regular points, such that:

(1) The measures  $\mu^s$  and  $\mu^u$  on  $U_x$  are mapped by  $u_x$  to the measures  $|\text{Re } dz^{p/2}| = |\text{Re } d\Phi_p(z)|$  and  $|\text{Im } dz^{p/2}| = |\text{Im } d\Phi_p(z)|$  on  $\mathbb{C}$  respectively, where  $p = p_x \geq 3$  is the number of prongs of  $\mathfrak{F}$  at

a singular point  $x$ , or  $p_x = 2$  if  $x$  is a regular point; the leaves of  $\mathfrak{F}^s$  and  $\mathfrak{F}^u$  get mapped to the lines  $\{\operatorname{Re} \Phi_p(z) = \text{constant}\}$  and  $\{\operatorname{Im} \Phi_p(z) = \text{constant}\}$  respectively. (The prongs of  $\mathfrak{F}^s$  are  $\{z = \rho e^{i\theta} \mid \rho \geq 0, \theta = \frac{\text{odd}}{p}\pi\}$  and the prongs of  $\mathfrak{F}^u$  are  $\{z = \rho e^{i\theta} \mid \rho \geq 0, \theta = \frac{\text{even}}{p}\pi\}$ .)

(2) The measures  $\mu^s$  and  $\mu^u$  on  $U_A$  are mapped by  $u_A$  to the measures  $|\operatorname{Re} d\Phi_p \circ \Psi(z)|$  and  $|\operatorname{Im} d\Phi_p \circ \Psi(z)|$  on  $\mathbb{C} - \text{int } D$  respectively, where  $p = p_A \geq 1$  is the number of prongs of  $\mathfrak{F}$  at  $A$ ; the leaves of  $\mathfrak{F}^s$  and  $\mathfrak{F}^u$  get mapped to the lines  $\{\operatorname{Re} \Phi_p \circ \Psi(z) = \text{constant}\}$  and  $\{\operatorname{Im} \Phi_p \circ \Psi(z) = \text{constant}\}$  respectively.

We may further assume (change scale if necessary) that, in each chart at an interior singular point  $x$ , the closed unit disk  $D$  is contained in  $u_x(U_x)$  and is disjoint from chart overlaps; similarly for the closed annulus  $\{z \mid 1 \leq |z| \leq 2\}$  in each chart  $U_A$ .

*The model for  $\varphi$ .* In such an atlas  $\mathcal{U}$ , denoting  $U'_x = U_x \cap \varphi^{-1}U_{\varphi(x)}$  and  $U'_A = U_A \cap \varphi^{-1}U_{\varphi(A)}$ , we have commutative diagrams

$$\begin{array}{ccc} U'_x, x & \xrightarrow{\varphi} & U_{\varphi(x)}, \varphi(x) & & U'_A, A & \xrightarrow{\varphi} & U_{\varphi(A)}, \varphi(A) \\ u_x \downarrow & & \downarrow u_{\varphi(x)} & & u_A \downarrow & & \downarrow u_{\varphi(A)} \\ \mathbb{C}, 0 & \xrightarrow{r \circ f} & \mathbb{C}, 0 & & \mathbb{C} - \text{int } D, S^1 & \xrightarrow{r \circ f'} & \mathbb{C} - \text{int } D, S^1 \end{array}$$

in charts at interior singularities and boundary components, where  $f = f_p$ ,  $f = f'_p$  and  $r = r_{(p,k)^\pm}$  for suitable  $p, k$ .

If such a singular point  $x$  is a fixed point of  $\varphi$ , it will be called a fixed point of type  $(p, k)^\pm$ . Similarly, if this  $A$  is  $\varphi$ -invariant, it is called a  $\varphi$ -invariant boundary component of type  $(p, k)^\pm$ . Here the sign  $+/-$  indicates  $\varphi|_A : A \rightarrow A$  is orientation preserving or reversing.

Suppose  $x \in \text{int } M$  is a regular point and  $\varphi(x) = x$ . Then there exist local charts at  $x$  and  $\varphi(x)$  (not necessarily in the atlas  $\mathcal{U}$ ) with the above commutative diagram where  $p$  is taken to be 2. In this sense a regular point can be regarded as a “2-prong singularity”, so that our discussion of singularities applies to regular points as well.

**REMARK.** In view of the rotation symmetry of  $v, v', f$  and  $f'$ , for odd  $p$  every type  $(p, k)^-$  is conjugate to, hence regarded the same as, the type  $(p, 0)^-$ ; for even  $p$ , among the types  $(p, k)^-$  there are two essentially different types:  $(p, 0)^-$  and  $(p, 1)^-$ .

Note that in contrast to the non-smoothness of  $\varphi$  at an interior  $p$ -prong singularity ( $p > 2$ ), this model of  $\varphi$  can be smoothly extended to the boundary  $\partial M$ . However, it fails to meet our later need in §3 that  $\varphi$  be periodic on  $\partial M$ .

2.2. *The standard form.* We shall modify the previous model to achieve two goals: to make  $\varphi$  smooth and to make  $\varphi|_{\partial M}$  periodic.

DEFINITION. Let  $\alpha: \mathbb{R} \rightarrow I$  be a non-decreasing smooth function such that  $\alpha(0) = 0$ ,  $\alpha(1/\lambda^2) = 1$ , and  $\alpha(t) \geq 0$  if  $t \geq 0$ . Let  $\bar{V}$  be the smooth vector field on  $\mathbb{C}$  obtained by “slowing down” the vector field  $V$ ,  $\bar{V}(s) = \alpha(|s|)V(s)$ . Then  $\bar{V}$  and its differentials of any order vanish at  $0 \in \mathbb{C}$ . Let  $\bar{F}: \mathbb{C} \rightarrow \mathbb{C}$  be the time-one map for  $\bar{V}$ . It is clear that  $\bar{F}$  is a diffeomorphism and  $\bar{F}(s) = F(s)$  when  $|s| \geq 1$ .

DEFINITION. Let  $\bar{f}: \mathbb{C} - \{0\} \rightarrow \mathbb{C} - \{0\}$  be the time-one map for the vector field  $\bar{v} := \Phi_*^{-1}\bar{V}$  on  $\mathbb{C} - \{0\}$ . Let  $\bar{f}': \mathbb{C} - D \rightarrow \mathbb{C} - D$  be the time-one map for the vector field  $\bar{v}' := \Psi_*^{-1}\bar{v}$  on  $\mathbb{C} - D$ . It follows from the smoothness of  $\bar{V}$  and its flatness at 0 that we can make  $\bar{v}$  smooth on  $\mathbb{C}$  by defining  $\bar{v}(0) = 0$ , and extend  $\bar{v}'$  smoothly to  $S^1 = \partial D$  by defining  $\bar{v}'(z) = 0$  there. The diffeomorphisms  $\bar{f}$  and  $\bar{f}'$  are then extended accordingly. We have a commutative diagram

$$\begin{array}{ccccc} \mathbb{C} & \xleftarrow{\Phi} & \mathbb{C} & \xleftarrow{\Psi} & \mathbb{C} - \text{int } D \\ \bar{F} \downarrow & & \downarrow \bar{f} & & \downarrow \bar{f}' \\ \mathbb{C} & \xleftarrow{\Phi} & \mathbb{C} & \xleftarrow{\Psi} & \mathbb{C} - \text{int } D \end{array}$$

It is clear that  $\bar{f} = f$  when  $|z| \geq 1$ ,  $\bar{f}' = f'$  when  $|z| \geq 2$ , and  $\bar{f}'$  has the same jets (of any order) at  $S^1$  as the identity map.

DEFINITION. In the atlas  $\mathcal{U}$ , let  $\bar{\varphi}: M \rightarrow M$  be defined by the commutativity of the diagrams

$$\begin{array}{ccc} U'_x, x & \xrightarrow{\varphi} & U_{\varphi(x)}, \varphi(x) & & U'_A, A & \xrightarrow{\varphi} & U_{\varphi(A)}, \varphi(A) \\ u_x \downarrow & & \downarrow u_{\varphi(x)} & & u_A \downarrow & & \downarrow u_{\varphi(A)} \\ \mathbb{C}, 0 & \xrightarrow{r \circ \bar{f}} & \mathbb{C}, 0 & & \mathbb{C} - \text{int } D, S^1 & \xrightarrow{r \circ \bar{f}'} & \mathbb{C} - \text{int } D, S^1 \end{array}$$

in the charts at interior singularities and boundary components, and be the same as  $\varphi$  elsewhere. Clearly, it is well-defined and smooth. Since  $\bar{v}$  and  $\bar{v}'$  are obtained from  $v$  and  $v'$  by “slowing down”,  $\bar{\varphi}$  is isotopic to  $\varphi$ . This isotopy does not change the interior fixed

points and the invariant boundary components, but  $\bar{\varphi}|_{\partial M}$  becomes a periodic map. We shall call  $\bar{\varphi}$  the *standard form* for the pseudo-Anosov map  $\varphi$ .

**REMARK 1.** We do not claim that  $\bar{\varphi}$  is topologically conjugate to  $\varphi$  on  $\text{int}M$ . Our concern here is only the isotopy, not the more difficult problem of conjugacy treated in [GK].

For the standard form we have:

**LEMMA 2.1.** *Every interior fixed point of  $\bar{\varphi}$  is isolated. Its index depends on its type:*

<i>Type of <math>x</math></i>	$(p, 0)^+$	$(p, k)^+, p \nmid k$	$(p_{\text{even}}, 0)^-$	$(p_{\text{even}}, 1)^-$	$(p_{\text{odd}}, 0)^-$
$\text{index}(\bar{\varphi}, x)$	$1 - p$	$1$	$-1$	$1$	$0$

*Fixed points on an invariant boundary component  $A$ :*

<i>Type of <math>A</math></i>	$(p, 0)^+$	$(p, k)^+, p \nmid k$	$(p_{\text{even}}, 0)^-$	$(p_{\text{even}}, 1)^-$	$(p_{\text{odd}}, 0)^-$
$\text{Fix } \bar{\varphi} \cap A$	$A$	$\emptyset$	$2 \text{ points}$	$2 \text{ points}$	$2 \text{ points}$
$\text{index}(\bar{\varphi}, A)$	$-p$	$0$	$0 + 0$	$1 + 1$	$1 + 0$

**REMARK 2.** Isolated fixed points of zero index are removable via local perturbation. There are two cases:

For an interior fixed point of type  $(p, 0)^-$ ,  $p$  odd, locally  $\bar{\varphi}$  switches the two sides of the reflection axis and moves every point of the axis (except the fixed point) in the same direction (cf. Fig. 1). This fixed point can be removed by composing  $\bar{\varphi}$  with a slight push along the axis.

For a boundary fixed point on an unstable prong of a type  $(p, 0)^-$  boundary component, locally  $\varphi$  switches the two sides of the unstable prong and moves every point of that prong (except the fixed point) inward (cf. Fig. 2). This fixed point is removed by pushing the unstable prong into  $\text{int}M$ .

### 2.3. Fixed point classes.

**LEMMA 2.2.** *Let  $A_0, A_1$  be either a fixed point of  $\varphi$  or a  $\varphi$ -invariant component of  $\partial M$ . Suppose  $A_0$  and  $A_1$  are  $\varphi$ -related via a path  $c: I, 0, 1 \rightarrow M, A_0, A_1$ . Then there is a path  $\gamma \simeq c: I, 0, 1 \rightarrow M, A_0, A_1$  such that either*

- (1)  $\gamma$  is in  $\text{Fix } \varphi$ , or
- (2)  $\gamma$  is in a component of  $\partial M$  of type  $(p, k)^+$ .

*The same is true with  $\varphi$  replaced by  $\bar{\varphi}$ .*

*Proof.* By assumption we have  $c \simeq \varphi \circ c: I, 0, 1 \rightarrow M, A_0, A_1$ . Deform  $c$  to an immersion  $\gamma$  quasi-transverse to  $\mathfrak{F}^s$ . Then  $\gamma \simeq \varphi \circ \gamma: I, 0, 1 \rightarrow M, A_0, A_1$  and  $\varphi \circ \gamma$  is also an immersion quasi-transverse to  $\mathfrak{F}^s$ . (Cf. [FLP, p.76] for the definition of quasi-transversals. The Propositions II.3 and II.6 of [FLP, Exposé 5] can be naturally generalized to homotopy classes of paths  $I, 0, 1 \rightarrow M, A_0, A_1$ .) Thus

$$\mu^s(\gamma) = \inf\{\mu^s(c') \mid c' \simeq \gamma\} = \inf\{\mu^s(c') \mid c' \simeq \varphi \circ \gamma\} = \mu^s(\varphi \circ \gamma).$$

But  $\mu^s(\varphi \circ \gamma) = \lambda \mu^s(\gamma)$  and  $\lambda > 1$ , hence  $\mu^s(\gamma) = 0$ . This means  $\gamma$  runs along the leaves of  $\mathfrak{F}^s$ . Thus  $\gamma$  is quasi-transverse to  $\mathfrak{F}^u$ . Then a similar argument shows that  $\gamma$  also runs along the leaves of  $\mathfrak{F}^u$ . This can occur only if  $\gamma$  is a constant path in  $\text{int } M$  or  $\gamma$  is in  $\partial M$ . Hence the conclusion for  $\varphi$ .

The isotopy from  $\varphi$  to  $\bar{\varphi}$ , obtained by “gradually slowing down” the vector fields  $v$  and  $v'$  to  $\bar{v}$  and  $\bar{v}'$ , does not change the fixed point set except on invariant boundary components of type  $(p, 0)^+$ . So the fixed point classes of  $\varphi$  and  $\bar{\varphi}$  correspond in an obvious way. Hence the conclusion remains valid for  $\bar{\varphi}$ .  $\square$

**COROLLARY 2.3.** *For the standard form  $\bar{\varphi}$ , every fixed point class is connected.*  $\square$

**3. The general case.** In this section, a standard form for a general  $\varphi$  is introduced. Its restriction on the periodic pieces and the pseudo-Anosov pieces having been specified in the previous sections, it remains to specify its behavior in the neighborhood of the reducing curves. This is done in §3.1. We then concentrate on the fixed point classes of the standard form. Using a book-keeping scheme introduced in §3.2, we in §3.3 identify the fixed point classes with the connected components of the fixed point set. This enables us (in §3.4) to compile a complete list of possible types of fixed point classes.

Our standard form has some other useful features. For example, iterates of a standard  $\varphi$  are still standard. The standard form is also “equivariant” with respect to finite group actions on  $M$ . These will not be discussed in this paper.

Suppose  $\varphi: M \rightarrow M$  is a homeomorphism of a compact surface  $M$  and  $\Gamma = \Gamma_1 \cup \cdots \cup \Gamma_n$  ( $n \geq 0$ ) is a disjoint union of smooth simple closed curves  $\Gamma_i \subset \text{int } M$  such that  $\Gamma$  is  $\varphi$ -invariant. Let  $\mathcal{N}(\Gamma)$  be a  $\varphi$ -invariant tubular neighborhood of  $\Gamma$ . The components  $\{N_i\}$  of  $\overline{\mathcal{N}(\Gamma)}$  are annuli or Möbius bands. The components  $\{M_i\}$

of  $M - \mathcal{N}(\Gamma)$  are subsurfaces with negative Euler characteristic. On these (not necessarily connected)  $\varphi$ -components,  $\varphi$  is either periodic or pseudo-Anosov.

Note that according to the Thurston Theorem this is the general case: If  $M$  has only one  $\varphi$ -component, the case  $\Gamma = \emptyset$  ( $n = 0$ ) is either periodic or pseudo-Anosov, while the case  $\Gamma \neq \emptyset$  ( $n > 0$ ) is reducible.

Suppose  $\varphi$  is given as in the Thurston Theorem, and on pseudo-Anosov pieces  $\varphi$  has already been isotoped into the standard form. Thus  $\varphi|_{\partial\mathcal{N}(\Gamma)}$  is a periodic map. We shall isotope  $\varphi|_{\overline{\mathcal{N}(\Gamma)}}$  rel  $\partial\mathcal{N}(\Gamma)$  into a standard form.

3.1. *Standard form for  $\varphi$  on  $\mathcal{N}(\Gamma)$ .* We now consider the standard form of  $\varphi|_{\overline{\mathcal{N}(\Gamma)}}$  under isotopies relative to  $\partial\mathcal{N}(\Gamma)$ . A component of  $\overline{\mathcal{N}(\Gamma)}$  is either an annulus or a Möbius band. Let  $N$  be a  $\varphi$ -invariant component of  $\overline{\mathcal{N}(\Gamma)}$ . Then  $\varphi|_N: N \rightarrow N$  is a diffeomorphism and  $\varphi|_{\partial N}$  is periodic.

**LEMMA 3.1.** *Let  $N$  be an annulus or a Möbius band,  $\varphi: N \rightarrow N$  be a diffeomorphism such that  $\varphi|_{\partial N}$  is periodic. Then  $\varphi$  is isotopic rel  $\partial N$  to a diffeomorphism  $\bar{\varphi}$  which is either periodic or a twist. More precisely:*

(A) *The annular case:  $N$  is the annulus  $S^1 \times I$ . Then  $\varphi$  is isotopic rel  $S^1 \times \partial I$  to a diffeomorphism  $\bar{\varphi}$  which is conjugate to one of the following standard maps  $\psi: S^1 \times I \rightarrow S^1 \times I$ .*

(1)<sup>+</sup>  $\psi(z, t) = (ze^{2(a+bt)\pi i}, t)$ , where  $a, b$  are rational numbers. If  $b = 0$ ,  $\text{Fix } \psi$  is either  $S^1 \times I$  or empty; if  $b \neq 0$ , such a  $\psi$  will be called a twist.  $\text{Fix } \psi$  is finitely many parallel circles  $S^1 \times \{t\}$ .

(2)<sup>-</sup>  $\psi(z, t) = (ze^{2a\pi i}, 1-t)$  where  $a$  is a rational number;  $\text{Fix } \psi$  is either  $S^1 \times \{\frac{1}{2}\}$  or empty.

(3)<sup>-</sup>  $\psi(z, t) = (\bar{z}, t)$ ;  $\text{Fix } \psi$  is two arcs  $\{1\} \times I$  and  $\{-1\} \times I$ .

(4)<sup>+</sup>  $\psi(z, t) = (\bar{z}e^{a(1-2t)\pi i}, 1-t)$ , where  $a$  is rational. If  $a \neq 0$ , such a  $\psi$  will be called a flip-twist.  $\text{Fix } \psi$  is two points  $(1, \frac{1}{2})$  and  $(-1, \frac{1}{2})$ .

(B) *The Möbius band case:  $N$  is the Möbius band represented as  $S^1 \times I$  modulo the identification  $(z, t) \sim (-z, 1-t)$ . Then  $\varphi$  is isotopic rel  $\partial N$  to a diffeomorphism  $\bar{\varphi}$  which is conjugate to one of the following standard maps  $\psi: N \rightarrow N$ .*

(1)  $\psi(z, t) = (ze^{2a\pi i}, t) \sim (-ze^{2a\pi i}, 1-t)$ , where  $a$  is a rational number;  $\text{Fix } \psi$  is either  $N$ , or empty, or the central circle represented by  $S^1 \times \{\frac{1}{2}\}$ .

(2)  $\psi(z, t) = (\bar{z}, t) \sim (-\bar{z}, 1 - t)$ ;  $\text{Fix } \psi$  consists of an arc  $\{1\} \times I \sim \{-1\} \times I$  and another point  $(1, \frac{1}{2}) \sim (-1, \frac{1}{2})$ .

*Proof.* The analysis is based on the following classical facts.

- Every periodic map of  $S^1$  is either conjugate to a rotation  $z \mapsto ze^{2a\pi i}$ , where the rational number  $a \pmod{1}$  is the Poincaré rotation number, or conjugate to a reflection  $z \mapsto \bar{z}$ .
- A homeomorphism of  $S^1 \times I$  onto itself which is the identity on  $S^1 \times \partial I$  is isotopic rel  $S^1 \times \partial I$  to a Dehn twist  $(z, t) \mapsto (ze^{2kt\pi i}, t)$ , where  $k$  is an integer.
- A homeomorphism of the Möbius band onto itself which is the identity on the boundary is isotopic rel boundary to the identity (cf. [E, Theorem 3.4]).

These standard maps fall into two major types: periodic maps and twists. The characteristic property of the twists is that  $S^1 \times \partial I$  can never be in the same fixed point class of any iterate of  $\varphi$ . We omit the details of the elementary but somewhat tedious arguments.  $\square$

**REMARK 1.** Strictly speaking, in order to guarantee that  $\varphi|_{\overline{\mathcal{N}(\Gamma)}}$  matches smoothly along  $\partial\mathcal{N}(\Gamma)$  with the standard form of  $\varphi|M - \mathcal{N}(\Gamma)$ , the standard formula for the twist should be  $\psi(z, t) = (ze^{2(a+b\delta(t))\pi i}, t)$ , where  $\delta: I \rightarrow I$  is a smooth increasing function,  $\delta(0) = 0$ ,  $\delta(1) = 1$ , and all the derivatives vanish at 0 and 1. Similarly for the flip-twist. This modification would not change the fixed point behavior that concerns us.

**REMARK 2.** Interior fixed circles can be removed via isotopy rel boundary. There are two cases:

Interior fixed circles of type (A1)<sup>+</sup> can be removed by composing  $\psi$  with a diffeomorphism  $S^1 \times I \rightarrow S^1 \times I$ ,  $(z, t) \mapsto (z, \beta(t))$ , where  $\beta: \mathbb{R} \rightarrow \mathbb{R}$  is a diffeomorphism such that  $\beta(t) \neq t$  iff  $0 < t < 1$ .

Fixed circles of types (A2)<sup>-</sup> and (B1) are removed by rotating the central circle  $S^1 \times \{\frac{1}{2}\}$ ; e.g. composing  $\psi$  with a map  $S^1 \times I \rightarrow S^1 \times I$ ,  $(z, t) \mapsto (ze^{i\alpha(t)}, t)$  where  $\alpha: \mathbb{R} \rightarrow I$  is a smooth function with  $\alpha(0) = 0$ ,  $\alpha(\frac{1}{2}) = 1$  and  $\alpha(1 - t) = \alpha(t)$ .

**DEFINITION.** A diffeomorphism  $\varphi$  is said to be in *standard form*, if its restriction to every periodic  $\varphi$ -component is periodic, its restriction to every pseudo-Anosov  $\varphi$ -component is in the standard form  $\bar{\varphi}$  of §2.2, and its restriction to every  $\varphi$ -component of  $\overline{\mathcal{N}(\Gamma)}$  is in the

standard form  $\bar{\varphi}$  of Lemma 3.1. (For simplicity, we shall omit the bar in the notation for the standard forms of §2.2 and Lemma 3.1.)

**COROLLARY 3.2.** *Every fixed point class of  $\varphi|_{\overline{\mathcal{N}(\Gamma)}}$  is connected. Moreover, if two fixed points  $x, y$  are joined by a path  $c$  such that  $c \simeq \varphi \circ c$ , then there is a path  $\gamma$  in  $\text{Fix } \varphi$  such that  $\gamma \simeq c$  rel end-points.  $\square$*

**3.2. Book-keeping in the universal cover.** Consider the following general setting which clearly applies to the situation of §3.3.

Suppose  $M$  is a connected smooth  $m$ -manifold,  $S \subset M$  is a (not necessarily connected) proper  $(m-1)$ -submanifold (proper in the sense that  $S \cap \partial M = \partial S$  and  $S$  has compact intersection with any compact subset of  $M$ ). Let  $p: \widetilde{M} \rightarrow M$  be the universal covering of  $M$ .

**DEFINITION.** The book-keeping graph  $G(S)$  is defined as follows. Each connected component  $\widetilde{U}_i$  of  $\widetilde{M} - p^{-1}(S)$  gives rise to a vertex  $v_i$ . Each connected component  $\widetilde{S}_j$  of  $p^{-1}(S)$  gives rise to an edge  $e_j$ . A vertex  $v_i$  is incident to an edge  $e_j$  if and only if  $\widetilde{S}_j$  is contained in the closure of  $\widetilde{U}_i$ .

**DEFINITION.** Suppose  $\tilde{c}$  is a path in  $\widetilde{M}$  from  $\tilde{x}$  to  $\tilde{y}$  transverse to  $p^{-1}(S)$ . Let  $\tilde{z}_1, \dots, \tilde{z}_k$  be the successive points where  $\tilde{c}$  crosses  $p^{-1}(S)$ . Let  $\tilde{c}_0, \tilde{c}_1, \dots, \tilde{c}_k$  be the successive segments of  $\tilde{c}$  cut by  $p^{-1}(S)$ . Define the *book-keeping path*  $\beta(\tilde{c})$  for  $\tilde{c}$  to be the edge-path  $v_0 e_1 v_1 \cdots v_{k-1} e_k v_k$  on  $G(S)$ , where  $e_h$  corresponds to the connected component  $\widetilde{S}_h$  of  $p^{-1}(S)$  containing  $\tilde{z}_h$ , and  $v_h$  corresponds to the connected component  $\widetilde{U}_h$  of  $\widetilde{M} - p^{-1}(S)$  containing  $\tilde{c}_h$ .

**LEMMA 3.3.** *The book-keeping graph  $G(S)$  is a tree. If every connected component  $U$  of  $M - S$  is  $\pi_1$ -injective in the sense that  $\pi_1(U)$  injects into  $\pi_1(M)$ , and if  $\tilde{c}$  has minimal intersection with  $p^{-1}(S)$  among paths from  $\tilde{x}$  to  $\tilde{y}$ , then the book-keeping path  $\beta(\tilde{c})$  is an arc (possibly degenerate to a single vertex).*

*Proof.* Since  $\widetilde{M}$  is a simply connected  $m$ -manifold and  $\widetilde{S}_j \subset \widetilde{M}$  is a connected proper  $(m-1)$ -submanifold, every  $\widetilde{S}_j$  must separate  $\widetilde{M}$  by homological reasons. Thus every edge  $e_j$  separates  $G(S)$ , hence the graph  $G(S)$  is a tree.

For the second part, it suffices to show there is no spur in the edge-path  $\beta(\tilde{c}) = v_0 e_1 v_1 \cdots v_{k-1} e_k v_k$ . If otherwise  $e_h = e_{h+1}$  for some  $h$ ,

the path  $\tilde{c}_h$  in  $\tilde{U}_h$  would have both ends in the same  $\tilde{S}_h = \tilde{S}_{h+1}$ . Now  $\tilde{U}_h$  is simply connected by the  $\pi_1$ -injectivity assumption. So  $\tilde{c}_h$  could be deformed into  $\tilde{S}_h = \tilde{S}_{h+1}$ , thus the product path  $\tilde{c}_{h-1}\tilde{c}_h\tilde{c}_{h+1}$  could be further deformed into  $\tilde{U}_{h-1} = \tilde{U}_{h+1}$ , contradicting the minimality of  $\tilde{c}$ .  $\square$

**3.3. Connectedness of fixed point classes.** Suppose the diffeomorphism  $\varphi$  is in the standard form defined in §3.1.

**LEMMA 3.4.** *Let  $A_0, A_1$  be either a fixed point of  $\varphi$  or a  $\varphi$ -invariant component of  $\partial M$ . Suppose  $A_0$  and  $A_1$  are  $\varphi$ -related via a path  $c: I, 0, 1 \rightarrow M, A_0, A_1$ . Then there is a path  $\gamma \simeq c: I, 0, 1 \rightarrow M, A_0, A_1$  such that either*

- (1)  $\gamma$  is in  $\text{Fix } \varphi$ , or
- (2)  $\gamma$  is in  $A_0 = A_1$  which is a boundary component containing no fixed point.

*Proof.* (1) Without loss of generality we may assume the path  $c$  has the minimal number of intersections with  $\partial \mathcal{N}(\Gamma)$  in its homotopy class  $I, 0, 1 \rightarrow M, A_0, A_1$ . Note that if  $c$  does not cross  $\partial \mathcal{N}(\Gamma)$ , the truth of the conclusion is already guaranteed by Lemmas 1.2, 2.2 and 3.2.

(2) We shall work on the universal cover, using another form of the notion of  $\varphi$ -relation defined at the end of the Introduction.

*Alternative Definition.* Path-connected  $\varphi$ -invariant subsets  $A_0, A_1 \subset M$  are  $\varphi$ -related (via a path  $c$ ) if and only if: there is a lifting  $\tilde{\varphi}: \tilde{M} \rightarrow \tilde{M}$  of  $\varphi$  on the universal cover  $\tilde{M}$  of  $M$ , such that some connected component  $\tilde{A}_0$  of  $p^{-1}(A_0)$  and some connected component  $\tilde{A}_1$  of  $p^{-1}(A_1)$  (joined by a lifting  $\tilde{c}$  of  $c$ ) are  $\tilde{\varphi}$ -invariant.

(The “if” part is trivial. For the “only if” part, choose a lifting  $\tilde{\varphi}$  of  $\varphi$  and a lifting  $\tilde{c}$  of  $c$  such that the given homotopy  $c \simeq \varphi \circ c$  lifts to a homotopy  $\tilde{c} \simeq \tilde{\varphi} \circ \tilde{c}$ . See [J3, Theorem I.1.10] for the case when both  $A_0, A_1$  are single points.)

(3) Apply the book-keeping scheme of §3.2 to  $M$  with  $m = 2$  and  $S := \partial \mathcal{N}(\Gamma)$ , and use the notation in the definition of  $\beta(\tilde{c})$ . The  $\pi_1$ -injectivity condition in Lemma 3.3 is clearly satisfied in our setting, and the minimality assumption on  $c$  guarantees the minimality of  $\tilde{c}$  with respect to  $p^{-1}(S)$ . So  $\beta(\tilde{c}) = v_0 e_1 v_1 \cdots v_{k-1} e_k v_k$  is an arc in  $G(S)$ .

The diffeomorphism  $\tilde{\varphi}$  leaves  $p^{-1}(S)$  invariant, so the book-keeping path  $\beta(\tilde{\varphi} \circ \tilde{c}) = v'_0 e'_1 v'_1 \cdots v'_{k-1} e'_k v'_k$  is also an arc.

Since  $\tilde{A}_0$  and  $\tilde{A}_1$  are  $\tilde{\varphi}$ -invariant, we have  $v_0 = v'_0$  and  $v_k = v'_k$ . By the uniqueness of joining arcs in the tree  $G(S)$ , we conclude that  $v_h = v'_h$  and  $e_h = e'_h$  for all  $h$ . Thus every  $\tilde{U}_h$  and  $\tilde{S}_h$  is  $\tilde{\varphi}$ -invariant.

(4) Apply Definition (2) to the universal cover  $p : \tilde{U}_h \rightarrow \overline{U}_h$ , we see  $S_h := p(\tilde{S}_h)$  and  $S_{h+1} = p(\tilde{S}_{h+1})$  are  $\varphi$ -related on the subsurface  $\overline{U}_h$ , via  $c_h := p \circ \tilde{c}_h$ , the  $h$ -th segment of  $c$ .

(5) The subsurfaces  $\overline{U}_0, \dots, \overline{U}_k$  are alternately of two different kinds: components of  $M - \mathcal{N}(\Gamma)$  (the  $M_j$ 's with  $\chi < 0$ ), and components of  $\overline{\mathcal{N}(\Gamma)}$  (the  $N_i$ 's with  $\chi = 0$ ).

(6) For every  $h$  of the first kind, Lemmas 1.2 and 2.2 say there is a path  $\gamma_h$  in  $\text{Fix } \varphi$  and paths  $\tau_h, \tau_{h+1}$  in  $S_h, S_{h+1}$  such that  $c_h \simeq \tau_h \gamma_h \tau_{h+1}$  rel endpoints. Replacing the segment  $c_h$  with  $\tau_h \gamma_h \tau_{h+1}$  and then slightly pushing the parts  $\tau_h, \tau_{h+1}$  into the neighboring regions  $U_{h-1}, U_{h+1}$  (of the second kind) respectively, we deform  $c$  into a new  $c$  with  $\gamma_h$  as the new  $c_h$ . This deformation does not affect the minimality of  $c$  with respect to  $S$ . Hence we may *assume from now on* that  $c_h$  is in  $\text{Fix } \varphi$  for every  $h$  of the first kind.

(7) Continue with  $h$  of the first kind. Now that  $\varphi \circ c_h = c_h$ , there exists a covering translation  $\alpha_h$  such that  $\tilde{\varphi} \circ \tilde{c}_h = \alpha_h \circ \tilde{c}_h$ . We claim that  $\alpha_h = 1$ . Indeed, when  $A_0$  or  $A_1$  is a point, this is true for  $h = 0$  or  $k$  because  $\tilde{A}_0$  or  $\tilde{A}_1$  is a point. In all other cases, apply the following geometric observation to the universal cover  $p : \tilde{U}_h \rightarrow \overline{U}_h$ .

*Observation.* Suppose  $p : \tilde{M} \rightarrow M$  is the universal cover of a connected compact surface  $M$  with boundary,  $\chi(M) < 0$ . Suppose  $\alpha$  is a covering translation. If there is a path  $\tilde{c}$  in  $\tilde{M}$  such that  $\tilde{c}$  and  $\alpha \circ \tilde{c}$  join the same pair of different connected components of  $\partial \tilde{M}$ , then  $\alpha$  must be the identity.

(Proof of this observation: Think of  $M$  as a hyperbolic surface with totally geodesic boundary, so that  $\partial \tilde{M}$  consists of hyperbolic straight lines. Via homotopy we may replace  $\tilde{c}$  with the unique shortest geodesic joining that pair of components of  $\partial \tilde{M}$ . Since  $\alpha$  is an isometry,  $\alpha \circ \tilde{c}$  is the same shortest geodesic. So  $\alpha = 1$ .)

(8) We have shown that for  $h$  of the first kind,  $\tilde{c}_h$  is indeed a path in  $\text{Fix } \tilde{\varphi}$ . In particular,  $\tilde{z}_h, \tilde{z}_{h+1}$  are fixed points of  $\tilde{\varphi}$ . But every other  $h$  is of the first kind, so that *all* cut points  $\tilde{z}_1, \dots, \tilde{z}_k$  of  $\tilde{c}$  are fixed points of  $\tilde{\varphi}$ .

(9) Now we turn to those  $h$  of the second kind,  $\overline{U}_h \subset \overline{\mathcal{N}(\Gamma)}$ . By Definition (2),  $z_h, z_{h+1}$  are in the same fixed point class of  $\varphi|_{\overline{U}_h}$ .

Corollary 3.2 tells us there is a path homotopy  $c_h \simeq \gamma_h$  rel endpoints such that  $\gamma_h$  is in  $\text{Fix } \varphi$ . Replacing every such segment  $c_h$  with  $\gamma_h$ , we obtain the desired path  $\gamma$  lying entirely in  $\text{Fix } \varphi$ .  $\square$

**COROLLARY 3.5.** *Every fixed point class of  $\varphi$  is connected.*  $\square$

**3.4. Types of fixed point classes.** The last corollary enables us to identify the fixed point classes of  $\varphi$ . Each is a component of  $\text{Fix } \varphi$ , hence a union of the fixed point components on the standard pieces described in Lemmas 1.1, 2.1 and 3.1. Putting together all the information there and paying attention to the consistency along  $\partial \mathcal{N}(\Gamma)$ , we can get

**LEMMA 3.6.** *The possible types of fixed point classes of  $\varphi$  are listed below, with a description of their local behavior.*

(1) $^\pm$  *Isolated fixed point  $x$ :*

(a) $^+$   $x \in \text{int } M$ ,  $\varphi$  is conjugate to a rotation in a neighborhood of  $x$ ;  $\text{index}(\varphi, x) = 1$ .

(b) $^+$   $x \in \text{int } M$  is a fixed point of an annular flip-twist;  $\text{index}(\varphi, x) = 1$ .

(c) $^+$   $x \in \text{int } M$  is a type  $(p, k)^+$  interior fixed point of a pseudo-Anosov piece;  $\text{index}(\varphi, x) = 1 - p$  or  $1$ .

(d) $^-$   $x \in \text{int } M$  is a type  $(p, k)^-$  interior fixed point of a pseudo-Anosov piece;  $\text{index}(\varphi, x) = 1, -1$  or  $0$ .

(e) $^-$   $x \in \partial M$  and  $x$  is in a type  $(p, k)^-$  invariant boundary component of some pseudo-Anosov piece;  $\text{index}(\varphi, x) = 1$  or  $0$ .

(2) $^\pm$  *Fixed circle  $C$ :*

(a) $^+$   $C \subset \text{int } M$  is a fixed circle of an annular twist;  $\text{index}(\varphi, C) = 0$ .

(b) $^-$   $C \subset \text{int } M$  and in a neighborhood of  $C$ ,  $\varphi$  is conjugate to the reflection  $(z, t) \mapsto (z, 1 - t)$  on the annulus  $S^1 \times I$  or the Möbius band  $S^1 \times I / \sim$ ;  $\text{index}(\varphi, C) = 0$ .

(c) $^+$   $C \subset \text{int } M$ ; on one side  $C$  is a type  $(p, 0)^+$  boundary component of some pseudo-Anosov piece, on the other side  $C$  is a boundary component of an annular twist;  $\text{index}(\varphi, C) = -p$ .

(d) $^+$   $C \subset \partial M$ , and  $C$  is a type  $(p, 0)^+$  boundary component of some pseudo-Anosov piece;  $\text{index}(\varphi, C) = -p$ .

(3) $^-$  *Fixed arc  $A$ , contained in some subsurface  $B$  of  $M$  on which  $\varphi$  acts as an involution. Every endpoint  $x$  of  $A$  is either*

(a)  $x \in \text{int } M$ , on the outside of  $B$   $x$  is in a type  $(p, k)^-$  invariant boundary component of a pseudo-Anosov piece, or

(b)  $x \in \partial M$ .

The possible values of  $\text{index}(\varphi, A)$  are 1,  $-1$  or 0.

(4)<sup>+</sup> Fixed subsurface  $B$  of  $M$  with  $\chi(B) \leq 0$ . The possible forms for a component  $C$  of  $\partial B$ :

- (a)  $C \subset \text{int } M$ , on the outside of  $B$   $C$  is a type  $(p_C, 0)^+$  invariant boundary component of some pseudo-Anosov piece;
- (b)  $C \subset \text{int } M$ , on the outside of  $B$   $C$  is a boundary component of an annular twist;
- (c)  $C \subset \partial M$ .

We have  $\text{index}(\varphi, B) = \chi(B) - \sum p_C < 0$  where the summation is over the components  $C$  of  $\partial B$  of type (a).

*Proof.* These are the only possible combinations of the fixed point sets of the standard models. The calculation of the index can be done using the well-known proposition below.  $\square$

**PROPOSITION 3.7.** *Let  $f: X \rightarrow X$  be a self-map of a compact polyhedron. Suppose  $X_0, X_1, X_2$  are subpolyhedra of  $X$  such that  $X = X_1 \cup X_2$ ,  $X_0 = X_1 \cap X_2$ . We suppose  $f(X_i) \subset X_i$  and write  $f_i: X_i \rightarrow X_i$  for the restriction of  $f$ , for  $i = 0, 1, 2$ . Let  $A \subset \text{Fix } f$  be both open and closed in  $\text{Fix } f$ , and let  $A_i = A \cap X_i$ ,  $i = 0, 1, 2$ . Then*

$$\text{index}(f, A) = \text{index}(f_1, A_1) + \text{index}(f_2, A_2) - \text{index}(f_0, A_0). \quad \square$$

**4. Proof of the Main Theorem.** In this section we first prove the Main Theorem. Then we prove an inequality relating the Lefschetz number and the Nielsen number.

*Proof of the Main Theorem.* We are supposed to show that every essential fixed point class of  $\varphi$  (in the standard form) is shrinkable to a point, and every inessential one is removable, via a smooth isotopy (through diffeomorphisms or through embeddings). In Steps 1–4 below, we examine successively the various types of fixed point classes listed in Lemma 3.6.

Step 1<sup>-</sup>. Isolated fixed points of zero index.

This can occur in types (1d)<sup>-</sup> and (1e)<sup>-</sup>. They can be removed according to Remark 2 of §2.2.

Step 2<sup>±</sup>. Fixed circles.

Circles of types (2a)<sup>+</sup> and (2b)<sup>-</sup> can be removed according to Remark 2 of §3.1. Circles of types (2c)<sup>+</sup> and (2d)<sup>+</sup> will be treated later in Step 4.

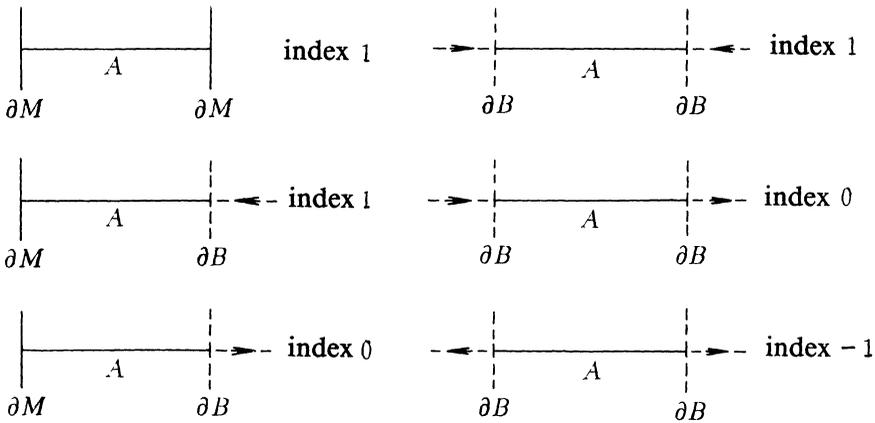


FIGURE 3

Step  $3^-$ . Fixed arcs.

Arcs of type  $(3)^-$  have six possible forms shown in Fig. 3, in which  $\varphi$  switches the two sides of the horizontal axis and moves points on the extension of  $A$  in the direction of the arrows. In the two forms with zero index, the fixed point set  $A$  can be removed by composing  $\varphi$  with a slight push along the axis in the direction of the arrow. In the forms with index  $\pm 1$  the fixed point class can be reduced to a point by composing  $\varphi$  with a slight contraction or stretch along the axis.

Step  $4^+$ . Fixed point classes  $F$  of types  $(2c)^+$ ,  $(2d)^+$  and  $(4)^+$  have some common features and will be given a unified treatment. They all have negative index so the task is to reduce  $F$  to a point via an isotopy. Let us fix a hyperbolic metric on  $M$  and let  $\delta > 0$  be the minimal length of closed geodesics of  $M$ . Then every simple closed curve of length less than  $\delta$  must bound a disk.

For all these three types, it is clear that there exists a neighborhood  $W$  of  $F$  and a smooth vector field  $w$  on  $W$  such that  $w = 0$  on  $F$ ,  $|w| < \delta/2$  on  $W$ ,  $\varphi$  coincides with the time-one map for  $w$  on a smaller neighborhood  $V$  of  $F$ . Perturb  $w$  to get a smooth vector field  $v$  on  $W$  so that  $v(x) \neq w(x)$  only if  $x$  is in a sufficiently small neighborhood  $U$  of  $F$ ,  $|v| < \delta/2$  on  $W$ ,  $v$  is tangent to  $\partial M$  on  $\partial M \cap W$ , and  $v$  has a unique singularity  $x_0$  on  $W$ ,  $x_0$  being any pre-assigned point of  $F$ . The index of the singularity must equal to  $\text{index}(\varphi, F) < 0$ . Let  $\varphi'$  be the time-one map for  $v$ . When  $U$  is small enough,  $\varphi'$  is well defined on  $V$  and coincides with  $\varphi$  near the boundary of  $V$ . So, letting  $\varphi' = \varphi$  on  $M - V$ , we extend  $\varphi'$

to a diffeomorphism  $M \rightarrow M$ . This  $\varphi'$  is clearly isotopic to  $\varphi$ . It remains to show  $\varphi'$  has no fixed points other than  $x_0$ .

The vector field  $v$  on  $W$  cannot have closed orbits shorter than  $\delta$ , because otherwise on the disk bounded by this closed orbit the sum of indices of the singularities of  $v$  must add up to 1, contradicting the fact that  $v$  has only one singularity with negative index. But  $|v| < \delta/2$ , so the time-one map can have no fixed point other than the singularity  $x_0$ .

Thus all types have been treated and our goal is achieved.

Note that isotopies through embeddings are needed only in Step 1, type (1e)<sup>-</sup> and in Step 3, all involving orientation-reversing  $\varphi$ -invariant components of  $\partial M$ .  $\square$

As a by-product of the analysis leading to the Main Theorem, we have

**THEOREM 4.1.** *Let  $M$  be a compact surface with  $\chi(M) < 0$ , and  $f: M \rightarrow M$  be a homeomorphism. Let  $L(f)$  be the Lefschetz number of  $f$ . Then  $N(f) \geq L(f) \geq -N(f) + 2\chi(M)$ , or equivalently,  $|L(f) - \chi(M)| \leq N(f) - \chi(M)$ .*

*Proof.* Since both  $L(f)$  and  $N(f)$  are homotopy invariants of  $f$ , we may assume  $f$  is in the standard form  $\varphi$ . The fixed point classes of  $\varphi$  are as described in Lemma 3.6.  $L(\varphi) \leq N(\varphi)$  simply because every fixed point class of  $\varphi$  has index  $\leq 1$ . The other inequality needs a closer analysis. Clearly

$$L(\varphi) + N(\varphi) \geq \sum_A (\text{index}(\varphi, A) + 1)$$

where the summation is taken over all fixed point classes  $A$  with  $\text{index}(\varphi, A) < -1$ . A 0-dimensional fixed point class  $A$  in the last summation must be a point  $x$  which is an interior  $p_x$ -prong singularity of some pseudo-Anosov piece  $M_j$  and  $\text{index}(\varphi, A) + 1 = 2 - p_x$ . A 1-dimensional fixed point class in this sum is a circle  $C$  which is a  $p_C$ -prong boundary component of some pseudo-Anosov piece of  $\varphi$ ,  $\text{index}(\varphi, A) = -p_C$ . For a 2-dimensional fixed point class  $A$ ,  $\text{index}(\varphi, A) = \chi(A) - \sum p_C$  where the summation is over components  $C$  of  $\partial A$  which is at the same time a  $p_C$ -prong boundary component of a pseudo-Anosov piece  $M_j$ . Hence

$$L(\varphi) + N(\varphi) \geq \sum_1 \left\{ \sum_x (2 - p_x) + \sum_C (-p_C) \right\} + \sum_2 \chi(M_j)$$

where  $\sum_1$  sums over the pseudo-Anosov pieces  $M_j$  of  $\varphi$  and  $\sum_2$  sums over the periodic pieces  $M_j$  of  $\varphi$ , in the braces  $x$  runs over all interior singularities of  $\varphi|_{M_j}$  and  $C$  runs over the boundary components. By the Euler-Poincaré formula of [FLP, p.75] (applied to the stable foliation of  $\varphi|_{M_j}$ ), the sum in the braces equals  $2\chi(M_j)$ . Thus

$$L(\varphi) + N(\varphi) \geq \sum_1 2\chi(M_j) + \sum_2 \chi(M_j) \geq 2\chi(M)$$

since every  $\chi(M_j) \leq 0$ . □

**REMARK.** From the proof we see that the equality  $L(f) + N(f) = 2\chi(M)$  can occur only if  $f$  is isotopic to a pseudo-Anosov map of a closed surface.

*Question.* Is Theorem 4.1 true even for self-maps of  $M$ ? Note that for compact surfaces with  $\chi = 0$ , it is known that  $|L(f)| \leq N(f)$  for every self-map  $f$ .

**5. The relative Nielsen numbers.** What is the best lower bound for the number of fixed points in an isotopy class of diffeomorphisms? The Main Theorem provides the answer only for the isotopy classes without orientation-reversing invariant boundary components. A candidate for a general answer is the relative Nielsen number introduced by Schirmer [S].

Let  $(X, A)$  be a pair of compact polyhedra,  $f: X, A \rightarrow X, A$  be a self-map of the pair. Let  $f_A: A \rightarrow A$  be the restriction of  $f$ . Denote by  $N(f, f_A)$  the number of essential fixed point classes of  $f$  that contain some essential fixed point class of  $f_A$ . Define

$$N(f; X, A) := N(f) + N(f_A) - N(f, f_A).$$

It is shown in [S] that for any map  $g \simeq f: X, A \rightarrow X, A$ , we always have  $N(f; X, A) = N(g; X, A)$ , therefore such a map  $g$  has at least  $N(f; X, A)$  fixed points on  $X$ .

Another relative Nielsen number is introduced by Zhao [Z]. Define  $N(f; X - A)$  to be the number of essential fixed point classes of  $f$  that are not  $f$ -related to the  $f$ -invariant set  $A$ . For any map  $g \simeq f: X, A \rightarrow X, A$ , we always have  $N(f; X - A) = N(g; X - A)$ , hence such a map  $g$  has at least  $N(f; X - A)$  fixed points on the complement  $X - A$ .

Note that a map  $f: X, A \rightarrow X, A$  has exactly  $N(f; X, A)$  fixed points on  $X$  and exactly  $N(f; X - A)$  fixed points on  $X - A$  if and

only if the following three conditions are satisfied:

- every fixed point class of  $f_A$  is a single point;
- every inessential fixed point class of  $f$  consists of essential fixed points of  $f_A$ ;
- an essential fixed point class of  $f$  can have more than one point only if it consists of essential fixed points of  $f_A$ .

For homeomorphisms of surfaces we have:

**THEOREM 5.1.** *Let  $M$  be a compact surface and let  $f: M \rightarrow M$  be a homeomorphism. Then  $f$  is isotopic to a diffeomorphism which has exactly  $N(f; M, \partial M)$  fixed points on  $M$  and exactly  $N(f; M - \partial M)$  fixed points in  $\text{int } M$ .*

*Proof.* For the three connected compact surfaces  $M$  with  $\chi(M) \geq 0$ , the conclusion is obvious. Hence we assume that every component of  $M$  has negative Euler characteristic. By Thurston Theorem we may replace  $f$  with the diffeomorphism  $\varphi$  in the standard form, as at the beginning of §4. Now we isotope  $\varphi$  as in the proof of the Main Theorem, with the following modifications: In Step 1 and Step 3, when we push along the axis we leave the intersection of the axis with  $\partial M$  fixed. In Step 4, we choose the singularity  $x_0$  of the vector field  $v$  to be in  $F \cap \partial M$  when  $F \cap \partial M \neq \emptyset$ . The result  $\varphi'$  is then a diffeomorphism and satisfies the three conditions listed above. Hence  $\varphi'$  has exactly  $N(\varphi'; M, \partial M)$  fixed points on  $M$  and exactly  $N(\varphi'; M - \partial M)$  fixed points on  $\text{int } M$ .  $\square$

**Acknowledgment.** We thank the referee for comments that led to improvement of exposition and to simplification of the analysis in §§3.2–3.

#### REFERENCES

- [B] L. Bers, *An extremal problem for quasiconformal mapping and a theorem by Thurston*, Acta Math., **141** (1978), 73–98.
- [BK] J. S. Birman and M. Kidwell, *Fixed points of pseudo-Anosov diffeomorphisms of surfaces*, Adv. in Math., **46** (1982), 217–220.
- [Br] R. F. Brown, *The Lefschetz Fixed Point Theorem*, Scott-Foresman, Chicago, 1971.
- [E] D. B. A. Epstein, *Curves on 2-manifolds and isotopies*, Acta Math., **115** (1966), 83–107.
- [FLP] A. Fathi, F. Laudenbach, and V. Poénaru, *Travaux de Thurston sur les surfaces*, Séminaire Orsay, Astérisque vol. 66–67, Soc. Math. France, Paris, 1979.
- [GK] M. Gerber and A. Katok, *Smooth models of Thurston's pseudo-Anosov maps*, Ann. Sci. Éc. Norm. Sup., **15** (1982), 173–204.

- [HT] M. Handel and W. P. Thurston, *New proofs of some results of Nielsen*, Adv. in Math., **56** (1985), 173–191.
- [I] N. V. Ivanov, *Nielsen numbers of self-maps of surfaces*, Studies in Topology IV, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI), vol.122, Nauka, Leningrad, 1982, pp. 56–65, 163–164 (Russian); English transl., J. Sov. Math., **26** (1984), 1636–1641.
- [J1] B. Jiang, *Fixed point classes from a differential viewpoint*, Fixed Point Theory (E. Fadell, G. Fournier, eds.), Lecture Notes in Math., vol. 886, Springer-Verlag, Berlin, Heidelberg, New York, 1981, pp. 163–170.
- [J2] ———, *Fixed points of surface homeomorphisms*, Bull. Amer. Math. Soc., **5** (1981), 176–178.
- [J3] ———, *Lectures on Nielsen Fixed Point Theory*, Contemp. Math., vol.14, Amer. Math. Soc., Providence, RI, 1983.
- [J4] ———, *Fixed points and braids*, Invent. Math., **75** (1984), 69–74; II, Math. Ann., **272** (1985), 249–256.
- [K] T.-H. Kiang, *The Theory of Fixed Point Classes*, Science Press, Beijing, 1979, 1986 (Chinese); English edition, 1989, Springer-Verlag, Berlin, Heidelberg, New York.
- [N1] J. Nielsen, *Über die Minimalzahl der Fixpunkte bei den Abbildungstypen der Ringflächen*, Math. Ann., **82** (1921), 83–93.
- [N2] ———, *Untersuchungen zur Topologie des geschlossenen zweiseitigen Flächen*, I, Acta Math., **50** (1927), 189–358; English transl., *Investigations in the topology of closed orientable surfaces*, I, Jakob Nielsen: Collected Mathematical Papers (V.L. Hansen, ed.), vol. 1, Birkhäuser, Boston, 1986, pp. 223–341.
- [S] H. Schirmer, *A relative Nielsen number*, Pacific J. Math., **122** (1986), 459–473.
- [T] W. P. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc., **19** (1988), 417–431.
- [W] F. Wecken, *Fixpunktklassen*, I, Math. Ann., **117** (1941), 659–671; II, Math. Ann., **118** (1942), 216–234; III, Math. Ann., **118** (1942), 544–577.
- [Wu] Y.-Q. Wu, *Canonical reducing curves of surface homeomorphisms*, Acta Math. Sinica, New Series, **3** (1987), 305–313.
- [Z] X.-Z. Zhao, *A relative Nielsen number for the complement*, Topological Fixed Point Theory and Applications (B. Jiang, ed.), Lecture Notes in Math., vol. 1411, Springer-Verlag, Berlin, Heidelberg, New York, 1989, pp. 189–199.

Received October 17, 1990 and in revised form March 3, 1992. Both authors partially supported by NSFC grants.

PEKING UNIVERSITY  
BEIJING 100871, CHINA

AND

HANGZHOU UNIVERSITY  
HANGZHOU, ZHEJIANG 310028, CHINA

