

Bounds for fixed points and fixed subgroups on surfaces and graphs

BOJU JIANG
SHIDA WANG
QIANG ZHANG

We consider selfmaps of hyperbolic surfaces and graphs, and give some bounds involving the rank and the index of fixed point classes. One consequence is a rank bound for fixed subgroups of surface group endomorphisms, similar to the Bestvina–Handel bound (originally known as the Scott conjecture) for free group automorphisms.

When the selfmap is homotopic to a homeomorphism, we rely on Thurston’s classification of surface automorphisms. When the surface has boundary, we work with its spine, and Bestvina–Handel’s theory of train track maps on graphs plays an essential role.

It turns out that the control of empty fixed point classes (for surface automorphisms) presents a special challenge. For this purpose, an alternative definition of fixed point class is introduced, which avoids covering spaces hence is more convenient for geometric discussions.

55M20, 57M07; 20F34, 57M15, 57N05

1 Introduction

Fixed point theory studies fixed points of a selfmap f of a space X . (A *selfmap* is a map from a space to itself.) Nielsen fixed point theory, in particular, is concerned with the properties of the fixed point set $\text{Fix } f := \{x \in X \mid x = f(x)\}$ that are invariant under homotopy of the map f (see Jiang [4] for an introduction).

The fixed point set $\text{Fix } f$ splits into a disjoint union of *fixed point classes*. Two fixed points are in the same class if and only if they can be joined by a *Nielsen path* which is a path homotopic (relative to endpoints) to its own f -image. Each fixed point class \mathbf{F} is an isolated subset of $\text{Fix } f$ hence its index $\text{ind}(f, \mathbf{F}) \in \mathbb{Z}$ is defined.

For an endomorphism $\phi: G \rightarrow G$ of a group, its *fixed subgroup* refers to the subgroup $\text{Fix}(\phi) := \{g \in G \mid g = \phi(g)\} \subset G$. The *stabilizer* of a fixed point $x \in \text{Fix } f$ is

the subgroup $\text{Stab}(f, x) := \text{Fix}(f_\pi) \subset \pi_1(X, x)$ where $f_\pi: \pi_1(X, x) \rightarrow \pi_1(X, x)$ is induced by f . (See the Remark in Section 2.2.) Since fixed points in the same class have isomorphic stabilizers, the *stabilizer* of a fixed point class \mathbf{F} is defined as $\text{Stab}(f, \mathbf{F}) = \text{Stab}(f, x)$ for any $x \in \mathbf{F}$. The *rank* of \mathbf{F} is defined as

$$\text{rank}(f, \mathbf{F}) := \text{rank Stab}(f, \mathbf{F}),$$

where the rank of a group means the minimal number of generators.

We are primarily interested in selfmaps of surfaces and graphs. In this setting, we find it convenient to introduce another term. The *characteristic* of a fixed point class \mathbf{F} is defined as

$$\text{chr}(f, \mathbf{F}) := 1 - \text{rank}(f, \mathbf{F}),$$

with the only exception that $\text{chr}(f, \mathbf{F}) := \chi(X)$ when X is a closed surface and $\text{Stab}(f, \mathbf{F}) = \pi_1(X)$. This number restricts the ‘potential’ shape of \mathbf{F} in the sense that, when f is homotoped to $f': X \rightarrow X$ and \mathbf{F} corresponds to a fixed point class \mathbf{F}' of f' , \mathbf{F}' cannot contain a connected π_1 -injective subset whose Euler characteristic is less than $\text{chr}(f, \mathbf{F})$.

For brevity, we will write $\text{Stab}(\mathbf{F})$, $\text{rank}(\mathbf{F})$ and $\text{chr}(\mathbf{F})$ if no confusion is possible for the selfmap f in the context.

In Nielsen theory, a fixed point class \mathbf{F} is allowed to be empty. In that case the above definition of $\text{rank}(\mathbf{F})$ (and $\text{chr}(\mathbf{F})$) does not make sense. An alternative approach, that works for empty classes as well but avoids covering spaces, will be given in Section 2.4.

Our main result is

Theorem 1.1 *Suppose X is either a connected finite graph or a connected compact hyperbolic surface, and $f: X \rightarrow X$ is a selfmap. Then*

- (A) $\text{ind}(\mathbf{F}) \leq \text{chr}(\mathbf{F})$ for every fixed point class \mathbf{F} of f ;
- (B) when X is not a tree,

$$\sum_{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq 2\chi(X),$$

where the sum is taken over all fixed point classes \mathbf{F} with $\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0$.

Both the index and the rank appear in the inequalities. Weaker bounds that involve only one of them were known before. Replacing $\text{ind}(\mathbf{F})$ with $\text{chr}(\mathbf{F})$ in (B) we get the rank bound $\sum_{\text{chr}(\mathbf{F}) < 0} \{\text{chr}(\mathbf{F})\} \geq \chi(X)$ of Bestvina and Handel [1] which solves the Scott conjecture for free group automorphisms. On the other hand, replacing $\text{chr}(\mathbf{F})$ with 1 in

(A) and (B) gives the index bounds $\text{ind}(\mathbf{F}) \leq 1$ and $\sum_{\text{ind}(\mathbf{F})+1 < 0} \{\text{ind}(\mathbf{F}) + 1\} \geq 2\chi(X)$ for surface maps (see Jiang [5]). So our result can be seen as a unification and strengthening of both.

A consequence of Theorem 1.1 is a rank bound for fixed subgroups of surface group endomorphisms:

Theorem 1.2 *Suppose G is the fundamental group of a closed hyperbolic surface. Then for any endomorphism $\phi: G \rightarrow G$, we have*

- (1) $\text{rank Fix}(\phi) \leq \text{rank}(G)$ if ϕ is epimorphic, with equality if and only if $\phi = \text{id}$;
- (2) $\text{rank Fix}(\phi) \leq \frac{1}{2} \text{rank}(G)$ if ϕ is not epimorphic.

The paper is organized as follows. Section 2 gives a very brief account of the basic notions and facts of Nielsen fixed point theory. Section 3 motivates our quest for Thurston’s theory of surface automorphisms. Section 4 deals with graph maps based on Bestvina–Handel train track theory, and completes the proofs of Theorem 1.1 and Theorem 1.2. Some examples will be given in Section 5.

Acknowledgements We would like to thank the referee for helping to clarify an argument in a proof.

This work is partially supported by a NSFC grant and a MOEC grant.

2 Background

Let X be a connected compact polyhedron, and $f: X \rightarrow X$ a selfmap.

2.1 Fixed point class

Let $p: \tilde{X} \rightarrow \tilde{X}$ be the universal covering of X , with group π of covering translations which we identify with the fundamental group $\pi_1(X)$.

For any lifting $\tilde{f}: \tilde{X} \rightarrow \tilde{X}$ of f , the projection of its fixed point set is called a *fixed point class* of f , written $\mathbf{F} = p(\text{Fix } \tilde{f})$. Strictly speaking, we say two liftings \tilde{f} and \tilde{f}' of f are *conjugate* if there exists $\gamma \in \pi$ such that $\tilde{f}' = \gamma^{-1} \circ \tilde{f} \circ \gamma$. Then $\mathbf{F} = p(\text{Fix } \tilde{f})$ is said to be the fixed point class of f *labeled* by the conjugacy class of \tilde{f} . Thus, a fixed point class always carries a label which is a conjugacy class of liftings. The fixed point set $\text{Fix } f$ decomposes into a disjoint union of fixed point classes. However, *empty* fixed point classes may have different labels hence be regarded as different. We would better think of them as hidden rather than nonexistent.

2.2 Index and stabilizer

The *index* of a fixed point class \mathbf{F} is defined to be the fixed point index $\text{ind}(f, \mathbf{F})$. Empty fixed point classes clearly have index 0. It is well known that the total index of all fixed point classes equals the Lefschetz number, that is, $\sum_{\mathbf{F}} \text{ind}(f, \mathbf{F}) = L(f)$.

Each lifting \tilde{f} induces an endomorphism $\tilde{f}_\pi: \pi \rightarrow \pi$ defined by

$$\tilde{f} \circ \gamma = \tilde{f}_\pi(\gamma) \circ \tilde{f}, \quad \text{for all } \gamma \in \pi.$$

The *stabilizer* of a fixed point class $\mathbf{F} = p(\text{Fix } \tilde{f})$, is defined as the subgroup

$$\text{Stab}(f, \mathbf{F}) := \{\gamma \in \pi \mid \gamma^{-1} \circ \tilde{f} \circ \gamma = \tilde{f}\},$$

which is identical to the fixed subgroup of the endomorphism \tilde{f}_π ,

$$\text{Fix}(\tilde{f}_\pi) = \{\gamma \in \pi \mid \gamma = \tilde{f}_\pi(\gamma)\}.$$

Up to group isomorphism, it is independent of the choice of \tilde{f} in its conjugacy class.

Remark This terminology (and notation) stems from the π -action by conjugation, $\gamma: \tilde{f} \mapsto \gamma^{-1} \circ \tilde{f} \circ \gamma$, on the set of liftings of f . The *orbit* (of a lifting \tilde{f}) of this action is the conjugacy class of liftings labeling \mathbf{F} , and the *stabilizer* under this action is the $\text{Stab}(f, \mathbf{F})$ defined above. For the use of stabilizer in a different setting, see Jiang [4, page 36].

For nonempty fixed point classes this definition reduces to the simpler one given in Section 1.

2.3 Invariance

The following facts are proved in [4]. The fixed point class correspondences involved will be given in our alternative approach of Section 2.4.

Fact (Homotopy invariance) *A homotopy $H = \{h_t\}: f_0 \simeq f_1: X \rightarrow X$ gives rise to a natural bijective correspondence $H: \mathbf{F}_0 \mapsto \mathbf{F}_1$ from f_0 -fixed point classes to f_1 -fixed point classes, and*

$$\text{ind}(f_0, \mathbf{F}_0) = \text{ind}(f_1, \mathbf{F}_1), \quad \text{Stab}(f_0, \mathbf{F}_0) \cong \text{Stab}(f_1, \mathbf{F}_1).$$

Remark A homotopy may create non-empty fixed point classes, or remove fixed point classes. The above correspondence is bijective only when empty fixed point classes are taken into account.

Fact (Morphism) A morphism from a selfmap $X \xrightarrow{f} X$ to a selfmap $Y \xrightarrow{g} Y$ means a map $h: X \rightarrow Y$ such that $h \circ f = g \circ h$. It induces a natural function $\mathbf{F}_f \mapsto \mathbf{F}_g$ from f -fixed point classes to g -fixed point classes, such that $h(\mathbf{F}_f) \subset \mathbf{F}_g$ and $h_\pi \text{Stab}(f, \mathbf{F}_f) \subset \text{Stab}(g, \mathbf{F}_g)$.

Fact (Commutation invariance) Suppose $\phi: X \rightarrow Y$ and $\psi: Y \rightarrow X$ are maps. Then the selfmaps $X \xrightarrow{\psi \circ \phi} X$ and $Y \xrightarrow{\phi \circ \psi} Y$ are said to differ by a commutation. The morphism ϕ between them sets up a natural bijective correspondence $\mathbf{F}_{\psi \circ \phi} \mapsto \mathbf{F}_{\phi \circ \psi}$ from $(\psi \circ \phi)$ -fixed point classes to $(\phi \circ \psi)$ -fixed point classes, and

$$\text{ind}(\psi \circ \phi, \mathbf{F}_{\psi \circ \phi}) = \text{ind}(\phi \circ \psi, \mathbf{F}_{\phi \circ \psi}), \quad \text{Stab}(\psi \circ \phi, \mathbf{F}_{\psi \circ \phi}) \cong \text{Stab}(\phi \circ \psi, \mathbf{F}_{\phi \circ \psi}).$$

2.4 Alternative definitions

The above definitions of fixed point class and stabilizer involve covering spaces. An alternative approach using paths is sometimes more convenient.

Suppose $f: X \rightarrow X$ is a self-map of a connected compact polyhedron.

Definition 2.1 By an f -route we shall mean a homotopy class (rel. endpoints) of paths $w: I \rightarrow X$ from a point $x \in X$ to $f(x)$. For brevity we shall often say the path w (in place of the path class $\langle w \rangle$) is an f -route at $x = w(0)$. An f -route w gives rise to an endomorphism $f_w: \pi_1(X, x) \rightarrow \pi_1(X, x)$, defined as $\langle a \rangle \mapsto \langle w(f \circ a)\bar{w} \rangle$, where a is any loop based at x , and \bar{w} denotes the reverse of w .

Two f -routes $\langle w \rangle, \langle w' \rangle$ are *conjugate* if there is path $q: I \rightarrow X$ from $x = w(0)$ to $x' = w'(0)$ such that $\langle w' \rangle = \langle \bar{q}w(f \circ q) \rangle$, that is, w' and $\bar{q}w(f \circ q)$ are homotopic rel. endpoints. We also say that the (possibly tightened) f -route $\bar{q}w(f \circ q)$ is obtained from w by an f -route move along the path q .

Note that a constant f -route w corresponds to a fixed point $x = w(0) = w(1)$ of f , and the endomorphism f_w becomes the usual $f_\pi: \pi_1(X, x) \rightarrow \pi_1(X, x)$. Two constant f -routes are conjugate if and only if the corresponding fixed points can be joined by a Nielsen path. This motivates the following definition.

Definition 2.2 With an f -route w (more precisely, with its conjugacy class) we associate a *fixed point class* \mathbf{F}_w of f , which consists of the fixed points that correspond to constant f -routes conjugate to w . Thus fixed point classes are associated bijectively with conjugacy classes of f -routes. A fixed point class \mathbf{F}_w can be empty if there is no constant f -route conjugate to w . Empty fixed point classes are distinguished by their associated route classes.

Remark This definition is equivalent to the traditional one in Section 2.1 because an f -route specifies a lifting \tilde{f} .

Definition 2.3 The *fixed subgroup* of the endomorphism f_w is the subgroup

$$\text{Fix}(f_w) := \{\gamma \in \pi_1(X, w(0)) \mid \gamma = f_w(\gamma)\}.$$

The *stabilizer* of the fixed point class \mathbf{F}_w is defined to be $\text{Stab}(f, \mathbf{F}_w) := \text{Fix}(f_w)$, it is well defined up to isomorphism because conjugate f -routes have isomorphic stabilizers.

Remark Let $x \in X$ be any given point. The set of f -routes at x is in bijective correspondence with $\pi_1(X, x)$. The group $\pi_1(X, x)$ acts on the set of f -routes at x , with a loop class $\langle a \rangle$ acting as $\langle w \rangle \mapsto \langle \bar{a}w(f \circ a) \rangle$. The orbits of this action represent all conjugacy classes of f -routes. The stabilizer of $\langle w \rangle$ under this action coincides with the fixed subgroup $\text{Fix}(f_w)$.

The correspondences between fixed point classes mentioned in Section 2.3 can now be explicitly defined.

Definition 2.4 Under a homotopy $H = \{h_t\}_{t \in I}: X \rightarrow X$, each h_0 -route w_0 gives rise to an h_1 -route

$$w_1 = w_0 \cdot H(w_0(0)), \quad \text{where } H(w_0(0)) \text{ is the path } \{h_t(w_0(0))\}_{t \in I}.$$

Clearly w_0 and w_1 share the same starting point, and $(h_0)_{w_0} = (h_1)_{w_1}$. The function $w_0 \mapsto w_1$ defines the fixed point class function $\mathbf{F}_{w_0} \mapsto \mathbf{F}_{w_1}$ induced by the homotopy.

Definition 2.5 For a morphism $h: X \rightarrow Y$ from a selfmap $X \xrightarrow{f} X$ to a selfmap $Y \xrightarrow{g} Y$, h sends an f -route w to a g -route $h \circ w$. The function $w \mapsto h \circ w$ defines the fixed point class function $\mathbf{F}_w \mapsto \mathbf{F}_{h \circ w}$ induced by the morphism.

3 Surface automorphisms

The aim of this section is to prove Theorem 3.1. The special case for empty fixed point classes, stated as Theorem 3.2, is interesting in itself.

Theorem 3.1 *Suppose M is a compact connected surface with negative Euler characteristic, and $f: M \rightarrow M$ is a homeomorphism. Then*

- (A) $\text{ind}(\mathbf{F}) \leq \text{chr}(\mathbf{F})$ for every fixed point class \mathbf{F} of f ;

$$(B) \quad \sum_{\text{ind}(\mathbf{F})+\text{chr}(\mathbf{F})<0} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq 2\chi(M).$$

The proof is based on Thurston’s isotopy classification of surface automorphisms. The following theorem summarizes the Thurston canonical map from [9], supplemented by the standard form developed by Jiang and Guo [6] with fine-tuned local behavior.

Theorem T *Suppose M is a compact connected surface with negative Euler characteristic. Every homeomorphism $f: M \rightarrow M$ is isotopic to a diffeomorphism φ such that either*

- (1) φ is a periodic map, that is, $\varphi^m = \text{id}$ for some $m \geq 1$, or equivalently, φ is an isometry with respect to some hyperbolic metric on M ; or
- (2) φ is a pseudo-Anosov map, that is, there is a number $\lambda > 1$ and a pair of transverse measured foliations (\mathfrak{F}^s, μ^s) and (\mathfrak{F}^u, μ^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) φ is a reducible map, that is, there is a system of disjoint simple closed curves $\Gamma = \Gamma_1 \cup \dots \cup \Gamma_n$ in the interior of M with the properties below.
 - (a) Γ is invariant by φ (but the Γ_i ’s may be permuted), and each component of $M \setminus \Gamma$ has negative Euler characteristic.
 - (b) Γ has a φ -invariant tubular neighborhood $\mathcal{N}(\Gamma)$ such that on each φ -component (= either a single φ -invariant component, or a disjoint union of components which are cyclically permuted by φ) of $M \setminus \mathcal{N}(\Gamma)$, φ is either periodic or pseudo-Anosov. (Note that the definitions in (1) and (2) apply to a φ -component as well.)
 - (c) Γ is minimal among all systems satisfying (a) and (b).
 - (d) φ is in the standard form as defined in [6, page 79].

The φ above will be called a *standard* map isotopic to f . Components of $\overline{\mathcal{N}(\Gamma)}$ will be called *strips*, and components of $M \setminus \mathcal{N}(\Gamma)$ will be called *pieces*. A strip is either an annulus or a Möbius strip. A piece is either *periodic* or *pseudo-Anosov* according as the type of the φ -component of that piece. A periodic piece on which φ is the identity map will be called an *identity* piece.

Remark The minimality condition (c) is equivalent to the condition of maximal periodic pieces, that is, no periodic piece can be combined with a neighboring one by eliminating the reducing curve between them.

Since $\text{chr}(\mathbf{F})$ and $\text{ind}(\mathbf{F})$ are homotopy invariants, in Theorem 3.1 the general automorphism f can be replaced by a standard map φ . For nonempty fixed point classes it is only a matter of observation:

Corollary T Suppose φ is a standard map in Theorem T. Then

(A') $\text{ind}(\mathbf{F}) \leq \chi(\mathbf{F}) = \text{chr}(\mathbf{F})$ for every nonempty fixed point class \mathbf{F} of φ ;

(B')
$$\sum_{\substack{\text{ind}(\mathbf{F})+\text{chr}(\mathbf{F})<0 \\ \mathbf{F} \text{ nonempty}}} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq 2\chi(M).$$

Remark It is the equality $\chi(\mathbf{F}) = \text{chr}(\mathbf{F})$ here that motivates the definition of $\text{chr}(\mathbf{F})$ given in Section 1.

Proof of Corollary T It follows from [6, Lemmas 1.2, 2.2 and 3.4] that each Nielsen path of φ can be deformed (rel. endpoints) into $\text{Fix } \varphi$. Hence every nonempty fixed point class \mathbf{F} is connected, and its stabilizer $\text{Stab}(\varphi, \mathbf{F})$ can be represented in \mathbf{F} itself. Therefore $\text{chr}(\mathbf{F}) = \chi(\mathbf{F})$.

A complete list of possible types of fixed point classes of φ is given in [6, Lemma 3.6]. They are submanifolds of M . Conclusion (A') can be read off from that list. The argument below for (B') parallels the proof of [6, Theorem 4.1].

Let us examine the list and focus on fixed point classes \mathbf{F} that contribute to the summation in (B'). A 0-dimensional \mathbf{F} must be a fixed point x which is an interior p_x -prong singularity of a pseudo-Anosov piece M_j , and $\text{ind}(\varphi, \mathbf{F}) + \text{chr}(\varphi, \mathbf{F}) = 2 - p_x$. A 1-dimensional \mathbf{F} is a fixed circle C which is a p_C -prong boundary component of a pseudo-Anosov piece, and $\text{ind}(\varphi, \mathbf{F}) + \text{chr}(\varphi, \mathbf{F}) = -p_C$. A 2-dimensional \mathbf{F} is either an identity piece, or a strip, or an identity piece combined with some neighboring strips, and moreover, $\text{ind}(\varphi, \mathbf{F}) + \text{chr}(\varphi, \mathbf{F}) = 2\chi(\mathbf{F}) - \sum p_C$ where the summation is over components C of $\partial\mathbf{F}$ which is at the same time a p_C -prong boundary component of a pseudo-Anosov piece M_j . Hence

$$\sum_{\substack{\text{ind}(\mathbf{F})+\text{chr}(\mathbf{F})<0 \\ \mathbf{F} \text{ nonempty}}} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq \sum_1 \left\{ \sum_x (2 - p_x) + \sum_C (-p_C) \right\} + \sum_2 2\chi(M_j)$$

where \sum_1 sums over the pseudo-Anosov pieces M_j of φ , inside the braces x runs over all interior singularities of $\varphi|M_j$ and C runs over all boundary components, and \sum_2 sums over the periodic pieces M_j of φ . Note that non-identity periodic pieces contribute nothing to the left and negative to the right. By the Euler–Poincaré formula of Fathi, Laudenbach and Poénaru [3, page 75] (applied to the stable foliation of $\varphi|M_j$), the sum in the braces equals $2\chi(M_j)$. Thus the inequality (B') follows:

$$\sum_{\substack{\text{ind}(\mathbf{F})+\text{chr}(\mathbf{F})<0 \\ \mathbf{F} \text{ nonempty}}} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq \sum_1 2\chi(M_j) + \sum_2 2\chi(M_j) = 2\chi(M). \quad \square$$

The remainder of this section is devoted to proving the following result

Theorem 3.2 *Suppose φ is a standard map in Theorem T. Then $\text{rank}(\varphi, \mathbf{F}) \leq 1$ for any empty fixed point class \mathbf{F} .*

This Theorem says that empty fixed point classes satisfy Theorem 3.1 (A) and do not contribute to the summation in Theorem 3.1 (B). It supplements Corollary T to complete the proof of Theorem 3.1.

Proof of Theorem 3.2 for periodic φ First assume M is a closed surface. Then φ is an isometry with respect to a hyperbolic metric on M . The universal cover of M is the hyperbolic plane H^2 . Let $\tilde{\varphi}: H^2 \rightarrow H^2$ be a lifting of φ labeling an empty fixed point class \mathbf{F} . Then $\tilde{\varphi}$ is a hyperbolic isometry without fixed points, so it has an axis L .

By definition, any nontrivial deck translation $\gamma \in \text{Stab}(\mathbf{F})$ commutes with $\tilde{\varphi}$, so the axis of γ coincides with L . Thus the elements of the group $\text{Stab}(\mathbf{F})$ share a common axis, hence it is commutative. But nontrivial commutative subgroups of $\pi_1(M)$ are cyclic. So if $\text{Stab}(\mathbf{F})$ is nontrivial then $\text{rank}(\mathbf{F}) = 1$. Thus Theorem 3.2 is valid for closed M .

Next assume M has boundary. Then φ is an isometry with respect to a hyperbolic metric on M (with totally geodesic boundary). Let M' be the double of M . The metric on M extends to a hyperbolic metric on M' , the isometry $\varphi: M \rightarrow M$ extends to an isometry $\varphi': M' \rightarrow M'$, and the φ -fixed point class \mathbf{F} is included in a φ' -fixed point class \mathbf{F}' . The inclusion i and projection p give a commutative diagram of maps (so they are morphisms between selfmaps)

$$\begin{array}{ccccc} M & \xrightarrow{i} & M' & \xrightarrow{p} & M \\ \varphi \downarrow & & \varphi' \downarrow & & \varphi \downarrow \\ M & \xrightarrow{i} & M' & \xrightarrow{p} & M \end{array}$$

and $p \circ i = \text{id}_M$. Hence $p(\mathbf{F}') = \mathbf{F}$ and $p_\pi(\text{Stab}(\varphi', \mathbf{F}')) = \text{Stab}(\varphi, \mathbf{F})$.

Now \mathbf{F} is empty implies \mathbf{F}' is empty. Since M' is a closed surface, it follows that $\text{rank}(\varphi, \mathbf{F}) \leq \text{rank}(\varphi', \mathbf{F}') \leq 1$. □

Proof of Theorem 3.2 for pseudo-Anosov φ Suppose φ is pseudo-Anosov with stable and unstable measured foliations (\mathfrak{F}^s, μ^s) and (\mathfrak{F}^u, μ^u) , respectively, and with expansion constant $\lambda > 1$. Suppose w is a φ -route. Let $g_1, g_2 \in \text{Fix}(\varphi_w)$ be nontrivial elements represented by loops c_1, c_2 , respectively.

Deform c_1 to a loop $\gamma_1: S^1 \rightarrow M$ quasi-transverse to \mathfrak{F}^s , see Fathi, Laudenbach and Poénaru [3, page 76]. (Use \simeq to denote free homotopy.) Then $\gamma_1 \simeq \varphi \circ \gamma_1: S^1 \rightarrow M$ and $\varphi \circ \gamma_1$ is also quasi-transverse to \mathfrak{F}^s . Thus

$$\mu^s(\gamma_1) = \inf\{\mu^s(c'_1) \mid c'_1 \simeq \gamma_1\} = \inf\{\mu^s(c'_1) \mid c'_1 \simeq \varphi \circ \gamma_1\} = \mu^s(\varphi \circ \gamma_1).$$

But $\mu^s(\varphi \circ \gamma_1) = \lambda \mu^s(\gamma_1)$ and $\lambda > 1$, hence $\mu^s(\gamma_1) = 0$. This means γ_1 runs along the leaves of \mathfrak{F}^s . Thus γ_1 is quasi-transverse to \mathfrak{F}^u . Then a similar argument shows that γ_1 also runs along the leaves of \mathfrak{F}^u . This can occur only if γ_1 is in a component Δ of ∂M . So g_1 can be represented by a loop of the form $q\gamma_1q^{-1}$, where q is a path ending in Δ . By a φ -route move along q , we can assume the base point $b = w(0)$ is in Δ , and $c_1 = \gamma_1$. Since $\langle c_1 \rangle = \langle w(\varphi \circ c_1)\bar{w} \rangle$, we can deform w to make it lie in Δ .

The same argument as the previous paragraph, but using based homotopy instead of free homotopy, shows c_2 is homotopic to a loop γ_2 in Δ . Hence g_1, g_2 lie in a cyclic group $\pi_1(\Delta)$.

What we have actually shown is that for pseudo-Anosov φ , any fixed point class \mathbf{F} has $\text{rank}(\mathbf{F}) \leq 1$, and those with nontrivial stabilizer can only live in ∂M . □

We need some preparation for the reducible φ .

Lemma 3.3 *Suppose $\ell: S^1 \rightarrow M$ is a loop that crosses $\partial\mathcal{N}(\Gamma)$ but is freely homotopic to one that does not. Then ℓ must have at least one “loose segment”.*

Here by a “segment” we mean an arc $A \subset S^1$ obtained by cutting S^1 along $\ell^{-1}\partial\mathcal{N}(\Gamma)$; it is “loose” if $\ell|_A$ can be deformed into $\partial\mathcal{N}(\Gamma)$ rel. endpoints.

Proof Suppose $H: S^1 \times I \rightarrow M$ is the homotopy of ℓ . Assume that it is smooth and transverse to $\partial\mathcal{N}(\Gamma)$. By an innermost disk argument, we can deform H so that $H^{-1}\partial\mathcal{N}(\Gamma)$ contains no disk-bounding circles. Then some arc $A \times \{0\} \subset S^1 \times \{0\}$ will combine with an arc in $H^{-1}\partial\mathcal{N}(\Gamma)$ to bound a disk in $S^1 \times I$. This A is a loose segment. □

Proposition 3.4 *Suppose $c: S^1 \rightarrow M$ is a homotopically nontrivial loop freely homotopic to its φ -image. Then c can be freely deformed to lie either in a strip $\mathcal{N}(\Gamma_i)$ or in a piece M_j .*

Proof We can assume that c minimizes (in its free homotopy class) the number of intersections with $\partial\mathcal{N}(\Gamma)$. Take a free homotopy $H: S^1 \times I \rightarrow M$ from c to $\varphi \circ c$ which is smooth and transverse to $\partial\mathcal{N}(\Gamma)$. Then $H^{-1}\partial\mathcal{N}(\Gamma)$ consists of disjoint arcs

and simple closed curves in $S^1 \times I$. There can be no arc in $H^{-1}\partial\mathcal{N}(\Gamma)$ joining two points of $S^1 \times \{0\}$, otherwise c can be deformed to decrease the number of intersections with $\partial\mathcal{N}(\Gamma)$, contradicting its minimality. Recall that c and $\varphi \circ c$ have the same number of intersections with $\partial\mathcal{N}(\Gamma)$ because $\varphi^{-1}\partial\mathcal{N}(\Gamma) = \partial\mathcal{N}(\Gamma)$.

By an innermost disk argument, the map H can be deformed rel. $S^1 \times \partial I$ so that one of the following occurs:

- (1) $H^{-1}\partial\mathcal{N}(\Gamma)$ is empty; or
- (2) $H^{-1}\partial\mathcal{N}(\Gamma)$ consists of disjoint simple closed curves parallel to $S^1 \times \{0\}$; or
- (3) $H^{-1}\partial\mathcal{N}(\Gamma)$ consists of disjoint arcs from $S^1 \times \{0\}$ to $S^1 \times \{1\}$.

In case (1), the loop c is already in a component of $M \setminus \partial\mathcal{N}(\Gamma)$ which is either in a strip or in a piece. So the conclusion holds.

In case (2), the loop c deforms into $\partial\mathcal{N}(\Gamma)$ through the annulus between $S^1 \times \{0\}$ and the component of $H^{-1}\partial\mathcal{N}(\Gamma)$ nearest to $S^1 \times \{0\}$, hence deforms further into Γ . So the conclusion also holds.

We are going to rule out case (3). Note that since an iterate of a standard map is again a standard map, in the argument below we have the freedom to replace φ by an iterate φ^ℓ , and to stack up ℓ copies of H to get a homotopy $H^{(\ell)}$ from c to $\varphi^\ell \circ c$ (but still denoting them as φ and H).

By replacing φ with an iterate of φ , we can assume that all periodic pieces of φ are identity pieces; and since Γ is assumed to be a minimal reduction system, φ is a nontrivial twist on each strip sandwiched between identity pieces.

Suppose $H^{-1}\partial\mathcal{N}(\Gamma)$ consists of arcs a_1, \dots, a_{2m} (in cyclic order), a_k being from $(z_k, 0)$ to $(z'_k, 1)$. By looking at a further iterate of φ if necessary, we can assume $z_k = z'_k$.

Then, $H^{-1}\partial\mathcal{N}(\Gamma)$ cuts $S^1 \times I$ into curvilinear rectangles R_1, R_2, \dots, R_{2m} (in cyclic order) which are, alternatingly, sent by H into pieces and strips.

If R_k is sent into a piece M_{j_k} , the restriction $H|R_k$ defines a φ -relation (see Jiang and Guo [6, page 70] for the definition) between two boundary components of M_{j_k} . On a pseudo-Anosov piece, by [6, Lemma 2.2] the bottom side of R_k would give a segment of c that is deformable into $\partial\mathcal{N}(\Gamma)$, contradicting the minimality assumption on c . So M_{j_k} must be an identity piece, and the top and bottom sides of R_k match. Since the bottom side of R_k is not deformable into $\partial\mathcal{N}(\Gamma)$ (again by the minimality assumption of c), the left and right sides of R_k must be contractible loops. Thus every arc a_k is mapped to a contractible loop in $\partial\mathcal{N}(\Gamma)$.

Now look at a rectangle R_k sent into a strip $\overline{\mathcal{N}(\Gamma_{i_k})}$. Since both its left and right sides are mapped by H to contractible loops in $\partial\mathcal{N}(\Gamma)$, the top and bottom sides must contribute the same degree in the Γ_{i_k} direction. This contradicts the assumption that φ is a nontrivial twist on such a strip. Thus the case (3) is ruled out.

This concludes the proof of the Proposition. \square

Proposition 3.5 *Suppose w is a φ -route, and $g_1, g_2 \in \text{Fix}(\varphi_w)$ are nontrivial elements.*

Then, up to φ -route move, g_1, g_2 can both be represented either in the same strip or in the same piece.

Proof Without loss we assume that the base point $w(0)$ is not in $\partial\mathcal{N}(\Gamma)$. By Proposition 3.4, for $k = 1, 2$, g_k is represented by a loop of the form $u_k c_k \bar{u}_k$, where c_k is a loop in $M \setminus \partial\mathcal{N}(\Gamma)$ and u_k is a path transverse to $\partial\mathcal{N}(\Gamma)$ which crosses the latter m_k times. We assume that the total crossing number $m_1 + m_2$ is minimal among all such representations, up to φ -route moves. We want to show $m_1 = m_2 = 0$.

When $m_k > 0$, let v_k denote the initial subpath of u_k up to its first crossing with $\partial\mathcal{N}(\Gamma)$. When $m_k = 0$, take u_k to be the constant path.

The product $g_1 g_2$ is represented by the loop $c = u_1 c_1 \bar{u}_1 u_2 c_2 \bar{u}_2$. By Proposition 3.4, since $\langle c \rangle$ is in $\text{Fix}(\varphi_w)$, c is freely homotopic to a loop in $M \setminus \partial\mathcal{N}(\Gamma)$.

By Lemma 3.3, c has loose segments. Clearly the part $u_k c_k \bar{u}_k$ cannot contain any loose segment, otherwise m_k can be decreased by tightening.

Case 1 Both m_k 's are positive. The only possible loose segments are $\bar{v}_1 v_2$ and $\bar{v}_2 v_1$, inverse to each other. Suppose $\bar{v}_1 v_2$ deforms into a component C_1 of $\partial\mathcal{N}(\Gamma)$. Define q to be the path v_1 slightly extended across C_1 , and move the φ -route w along q . After tightening, both m_1 and m_2 decrease by 1. This contradicts the minimality of $m_1 + m_2$. So Case 1 is ruled out.

Case 2 One m_k is zero (say, $m_2 = 0$). Then the loose segment must be $\bar{v}_1 c_2 v_1$. Move the φ -route w along q as in Case 1. Then m_1 decreases by 1 as before. Now g_2 is represented no longer by c_2 , but by the loop $\bar{q} c_2 q$ which tightens to a loop c'_2 in $M \setminus \partial\mathcal{N}(\Gamma)$. In other words, the representative of g_2 is changed from c_2 to this c'_2 on the other side of C_1 . Hence $m_2 = 0$ is kept. So Case 2 is also ruled out.

Case 3 Both m_k 's are zero. This is what we want. \square

Proof of Theorem 3.2 for reducible φ Suppose \mathbf{F} is associated with a φ -route w and $\text{rank Fix}(\varphi_w) > 1$. Let $g_1, g_2 \in \text{Fix}(\varphi_w)$ be two elements generating a non-cyclic subgroup. By Proposition 3.5, up to a φ -route move, both can be simultaneously represented either in a strip or in a piece. But the rank condition rules out the strips, so they can be represented in some piece M_j .

Clearly, M_j is invariant under φ . Let $\varphi_j: M_j \rightarrow M_j$ denote the restriction of φ . The φ -route w , now in M_j , associates with a φ_j -fixed point class \mathbf{F}_j . It is clear that $\mathbf{F}_j \subset \mathbf{F}$ and $g_1, g_2 \in \text{Fix}((\varphi_j)_w)$ so $\text{Fix}((\varphi_j)_w) > 1$. From the proof of Theorem 3.2 for periodic and pseudo-Anosov maps, it follows that \mathbf{F}_j is nonempty. Hence \mathbf{F} is nonempty. \square

4 Graph selfmaps

Our discussion of graph maps is based on Bestvina and Handel’s theory of train track maps [1]. We follow their terminology.

A *graph* X is a 1-dimensional (or possibly 0-dimensional) finite cellular complex. The 0-cells and (open) 1-cells are called *vertices* and *edges* respectively. A *graph map* $\alpha: X \rightarrow Y$ is a cellular map, that is, it maps vertices to vertices. Up to homotopy there is no loss to assume that the restriction of α to every edge e of X is either locally injective or equal to a constant map. A graph map $\alpha: X \rightarrow Y$ is π_1 -*injective* if it induces an injective homomorphism of the fundamental group on each component of X . It is an *immersion* if it sends edges to edges and it is locally injective at vertices. Clearly immersions are always π_1 -injective.

A *path* p in a graph X is a map $p: [0, 1] \rightarrow X$ that is either locally injective or equal to a constant map; in the latter case we say that p is a trivial path. For a nontrivial path p in X , its *initial tip* is the maximal initial open subpath that lies in an edge of X . The *terminal tip* is defined similarly.

A graph map $\alpha: X \rightarrow Y$ induces a function $D\alpha$ on the set of oriented edges of X . It sends an oriented edge e to the first oriented edge of $\alpha(e)$; if $\alpha(e)$ is trivial we say $D\alpha(e) = 0$.

For a graph selfmap $\beta: Z \rightarrow Z$, a β -*Nielsen path* is a nontrivial path p in Z joining two fixed points of β such that $\beta(p) \simeq p$ rel. endpoints; it is *indivisible* if it cannot be written as a concatenation $p = p_1 \cdot p_2$, where p_1 and p_2 are subpaths of p that are β -Nielsen paths.

The following theorem summarizes the results of Bestvina and Handel [1] that we need. See Jiang [5, Theorem BH].

Theorem BH Let X be a connected graph but not a tree, and let $f: X \rightarrow X$ be a π_1 -injective map. Then f has the same homotopy type as a graph selfmap $\beta: Z \rightarrow Z$, where Z is a connected graph without vertices of valence 1 and all fixed points of β are vertices, and there is a β -invariant proper subgraph Z_0 , containing all vertices of Z . The map $\beta: (Z, Z_0) \rightarrow (Z, Z_0)$ of the pair is of one of the following types.

Type 1: β sends Z into Z_0 .

Type 2: β cyclically permutes the edges in $Z \setminus Z_0$.

Type 3: β expands edges of $Z \setminus Z_0$ by a factor $\lambda > 1$ with respect to a suitable non-negative metric L supported on $Z \setminus Z_0$, and has the properties (a)–(c) below.

- (a) For every oriented edge e in $Z \setminus Z_0$, $D\beta(e)$ lies in $Z \setminus Z_0$.
- (b) There is at most one indivisible β -Nielsen path that intersects $Z \setminus Z_0$.
- (c) If p is an indivisible β -Nielsen path that intersects $Z \setminus Z_0$, then the tips of p are in $Z \setminus Z_0$ and invariant under β , and exactly one turn of p in $Z \setminus Z_0$ (at a vertex v_p of valence ≥ 3 in Z) degenerates under $D\beta$.

Proof Compared to [5, Theorem BH], we only added the requirement that Z_0 contains all fixed points of β . This is achieved in two easy steps. Firstly, on each β -invariant oriented edge we can adjust β to get rid of interior fixed points, so that β has only finitely many fixed points. Secondly, we can subdivide Z at the non-vertex fixed points and include them into Z_0 . \square

Information on fixed point classes follows.

Corollary BH Denote $\beta_0 := \beta|_{Z_0}: Z_0 \rightarrow Z_0$. If there exists an indivisible β -Nielsen path p that intersects $Z \setminus Z_0$, it is unique. (In Type 2 when $Z \setminus Z_0$ is a single edge e , take $p := e$.) There are three possible cases.

- (i) No such path p exists (as always in Type 1, in Type 2 when $Z \setminus Z_0$ has more than one edge, and possibly in Type 3). Then the β -fixed point classes are the same as the β_0 -fixed point classes. Their characteristics are also carried over:

$$\text{chr}(\beta, \mathbf{F}_0) = \text{chr}(\beta_0, \mathbf{F}_0)$$

for all β_0 -fixed point classes \mathbf{F}_0 .

- (ii) The path p connects two different β_0 -fixed point classes \mathbf{F}'_1 and \mathbf{F}'_2 . Then the β -fixed point classes are the same as the β_0 -fixed point classes, except that \mathbf{F}'_1 and \mathbf{F}'_2 combine into a single β -fixed point class $\mathbf{F}' = \mathbf{F}'_1 \cup \mathbf{F}'_2$. Their characteristics are also carried over, except that

$$\text{chr}(\beta, \mathbf{F}') = \text{chr}(\beta_0, \mathbf{F}'_1) + \text{chr}(\beta_0, \mathbf{F}'_2) - 1.$$

- (iii) The path p has both ends in a β_0 -fixed point class \mathbf{F}'_0 . Then the β -fixed point classes are the same as the β_0 -fixed point classes. Their characteristics are also carried over, except that

$$\text{chr}(\beta, \mathbf{F}'_0) = \text{chr}(\beta_0, \mathbf{F}'_0) - 1.$$

Proof Case (i) The conclusion is immediate, since every Nielsen path is a product of indivisible Nielsen paths, hence lies in Z_0 .

Case (ii) Suppose the path p goes from $a \in \mathbf{F}'_1$ to $b \in \mathbf{F}'_2$. For the chr equation it suffices to show that

$$\text{Stab}(\beta, a) \cong \text{Stab}(\beta_0, a) * \text{Stab}(\beta_0, b).$$

Define the natural homomorphism η from the right hand side to the left hand side by $\eta(\langle u \rangle) = \langle u \rangle$ and $\eta(\langle v \rangle) = \langle pv\bar{p} \rangle$, where u and v are Nielsen paths in Z_0 at a and b , respectively, and $\langle \cdot \rangle$ denotes loop class.

First observe that η is injective. In fact, the η -image of any nontrivial element of the right hand side is represented by a product $w = u_1pv_1\bar{p} \dots u_\ell pv_\ell \bar{p}u_{\ell+1}$, where u_i and v_i are Nielsen paths in Z_0 at a and b , respectively. (We allow that u_1 and $u_{\ell+1}$ be trivial, but assume other subpaths are nontrivial.) By Property (c) in Theorem BH, both tips of p are in $Z \setminus Z_0$, so w is an immersed Nielsen path and represents a nontrivial element in the left hand side.

On the other hand, any nontrivial element of the left hand side is represented by a concatenation of β -Nielsen paths, hence by a product like the w above, so η is surjective. Thus the desired isomorphism is established.

Case (iii) Suppose the path p goes from a to b , both in \mathbf{F}'_0 . Pick a β_0 -Nielsen path q in Z_0 from b back to a . It suffices to show that

$$\text{Stab}(\beta, a) \cong \text{Stab}(\beta_0, a) * J,$$

where J is the infinite cyclic group generated by the loop class represented by the loop $t := pq$ when $a \neq b$, but taking $t := p$ when $a = b$. The argument for establishing the isomorphism is similar to case (ii). □

Theorem 4.1 Let X be a connected graph and $f: X \rightarrow X$ be a graph map. Then

- (A) $\text{ind}(f, \mathbf{F}) \leq \text{chr}(f, \mathbf{F})$ for every fixed point class \mathbf{F} of f ; and
- (B) when X is not a tree,

$$\sum_{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq 2\chi(X).$$

Proof It follows from Bestvina and Handel [1, Lemma 2.1] and Dicks and Ventura [2, Lemma I.5.4] that empty fixed point classes have index equal to zero and non-negative characteristic, so they satisfy (A) and do not contribute to (B). Hence we only need to consider nonempty fixed point classes.

By [5, Lemma A], without loss of generality we may assume that f is π_1 -injective. Then it suffices to prove the bounds (A) and (B) for the graph selfmap $\beta: (Z, Z_0) \rightarrow (Z, Z_0)$ in Theorem BH. We can assume $\chi(Z) < 0$ since Theorem 4.1 is trivial if $\chi(Z) \geq 0$.

Let $Z_i, i = 1, \dots, n$ be the connected components of Z_0 . Suppose the β -invariant ones are $i = 1, \dots, k$. Denote $\beta_0 := \beta|_{Z_0}: Z_0 \rightarrow Z_0$ and $\beta_i := \beta|_{Z_i}: Z_i \rightarrow Z_i$ for $1 \leq i \leq k$.

Since Z is a connected graph without vertices of valence 1, and Z_0 is a proper subgraph, it is easy to see that all $\chi(Z_i) > \chi(Z)$. So, working inductively, we may assume that Theorem 4.1 is true for every β_i .

The characteristics of β -fixed point classes are already analyzed in Corollary BH. We will refer to the cases (i)–(iii) there.

In order to discuss indices, we introduce some notation. For a vertex v of Z , let $\omega(v)$ be the number of oriented edges e in $Z \setminus Z_0$ starting at v ; and let $\delta(v)$ be the number of such oriented edges e with the additional requirement that e gets initially expanded along itself by β . Recall that $\text{ind}(\beta, v) = \text{ind}(\beta_0, v) - \delta(v)$. For a β -invariant subgraph $Y \subset Z$, let $\omega(Y)$ and $\delta(Y)$ denote the sum of $\omega(v)$ and $\delta(v)$, respectively, over all $v \in Y$. So, for every β -fixed point class \mathbf{F} we have $\text{ind}(\beta, \mathbf{F}) = \text{ind}(\beta_0, \mathbf{F}) - \delta(\mathbf{F})$.

The inequality (A) is obvious except for the combined fixed point class \mathbf{F}' in case (ii) of Type 3. For this \mathbf{F}' the index is $\text{ind}(\beta, \mathbf{F}'_1) + \text{ind}(\beta, \mathbf{F}'_2) \leq \text{ind}(\beta_0, \mathbf{F}'_1) + \text{ind}(\beta_0, \mathbf{F}'_2) - 2$ since both $\delta(\mathbf{F}'_1) \geq 1$ and $\delta(\mathbf{F}'_2) \geq 1$ by Property (c). So (A) follows from the inductive hypothesis.

To prove (B), we use the function $\mu(x) := \min\{x, 0\} = \frac{1}{2}(x - |x|)$. It is superadditive in the sense that $\mu(x + y) \geq \mu(x) + \mu(y)$. Denote

$$S_i := \sum_{\mathbf{F}_i} \mu(\text{ind}(\beta_i, \mathbf{F}_i) + \text{chr}(\beta_i, \mathbf{F}_i) - \delta(\mathbf{F}_i)),$$

for $1 \leq i \leq k$, where the summation is taken over all β_i -fixed point classes \mathbf{F}_i ; and let $S_i := 0$ for $k < i \leq n$. Then we observe that

$$\sum_{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq \begin{cases} \sum_{i=1}^n S_i, & \text{in case (i);} \\ \sum_{i=1}^n S_i - 1, & \text{in cases (ii) and (iii).} \end{cases}$$

We shall show that the right hand side $\geq 2\chi(Z)$ by two claims.

Claim 1 $S_i \geq 2\chi(Z_i) - \omega(Z_i)$ for $1 \leq i \leq n$.

In fact, the inequality is trivial for $i > k$. (It is non-obvious only when $\chi(Z_i) = 1$. But $\chi(Z_i) = 1$ implies $\omega(Z_i) \geq 2$ because Z has no vertices of valence 1.) So we now assume $i \leq k$.

If $\chi(Z_i) \leq 0$, by the inductive hypothesis we have

$$\begin{aligned} S_i &= \sum_{\mathbf{F}_i} \mu(\text{ind}(\beta_i, \mathbf{F}_i) + \text{chr}(\beta_i, \mathbf{F}_i) - \delta(\mathbf{F}_i)) \\ &\geq \sum_{\mathbf{F}_i} \mu(\text{ind}(\beta_i, \mathbf{F}_i) + \text{chr}(\beta_i, \mathbf{F}_i)) - \sum_{\mathbf{F}_i} \delta(\mathbf{F}_i) \\ &\geq 2\chi(Z_i) - \delta(Z_i) \\ &\geq 2\chi(Z_i) - \omega(Z_i). \end{aligned}$$

If $\chi(Z_i) = 1$, then Z_i is a tree. The unique β_i -fixed point class \mathbf{F}_i has $\text{ind}(\beta_i, \mathbf{F}_i) = \text{chr}(\beta_i, \mathbf{F}_i) = 1$, so $S_i = \mu(2 - \delta(\mathbf{F}_i)) \geq \mu(2 - \omega(Z_i)) = 2\chi(Z_i) - \omega(Z_i)$. Thus Claim 1 is proved.

Claim 2 In cases (ii) and (iii), there exists $1 \leq h \leq n$ with $S_h > 2\chi(Z_h) - \omega(Z_h)$.

In fact, when β is of Type 3, the argument of Jiang [5, page 475, lines 6–19] repeated verbatim.

When β is of Type 2, $Z \setminus Z_0$ is a single edge e , and no component of Z_0 can be a tree. On the other hand, only one end of e is expanded by β , so $\delta(Z_0) = 1 < 2 = \omega(Z_0)$. Hence $\delta(Z_h) < \omega(Z_h)$ for some component Z_h . Then by the proof of Claim 1 we see $S_h \geq 2\chi(Z_h) - \delta(Z_h) > 2\chi(Z_h) - \omega(Z_h)$. Thus Claim 2 is proved.

It follows from Claims 1–2 that in all cases (i)–(iii),

$$\sum_{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0} (\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})) \geq \sum_{i=1}^n (2\chi(Z_i) - \omega(Z_i)) = 2\chi(Z).$$

This is the inequality (B) for β .

The inductive proof of the Theorem is now complete. □

Proof of Theorem 1.1 We have already taken care of graph selfmaps and surface automorphisms (Theorems 4.1 and 3.1). Surfaces with boundary have the same homotopy type as graphs. So it easily reduces to graph selfmaps. It remains to consider

the case that X is a closed hyperbolic surface and f is not homotopic to a self-homeomorphism. Then by [5, Lemma B], f is homotopic to a non-surjective selfmap. Hence the endomorphism $f_\pi: \pi_1(X) \rightarrow \pi_1(X)$ factors through a free group.

For a group G , let $\text{Ir}(G)$ denote its *inner rank* defined as the maximal rank of free homomorphic images of G . It is known that if G is the fundamental group of a closed surface, then $\text{Ir}(G) = \lfloor \text{rank}(G)/2 \rfloor$. (See Lyndon and Schupp [7, page 52] where it is attributed to Zieschang [10].)

So, the endomorphism $f_\pi: \pi_1(X) \rightarrow \pi_1(X)$ factors through the free group F_r of rank $r = \lfloor \text{rank}(\pi_1(X))/2 \rfloor$, which is isomorphic to $\pi_1(Y)$ for a connected graph Y . Since X and Y are aspherical, the algebraic factorization can be realized topologically, that is, f is homotopic to a composition $X \xrightarrow{\psi \circ \phi} X$.

By the homotopy and commutation invariance, we can replace f with the composition $Y \xrightarrow{\phi \circ \psi} Y$ which is a graph selfmap. □

The above argument actually proves the next Proposition with a bound stronger than that of Theorem 1.1.

Proposition 4.2 *Suppose X is a closed hyperbolic surface, and $f: X \rightarrow X$ is a selfmap homotopic to a non-surjective one. Then*

$$\sum_{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F}) < 0} \{\text{ind}(\mathbf{F}) + \text{chr}(\mathbf{F})\} \geq \begin{cases} \chi(X), & \text{if } \chi(X) \text{ is even,} \\ \chi(X) + 1, & \text{if } \chi(X) \text{ is odd.} \end{cases}$$

Proof of Theorem 1.2 Regard $\phi: G \rightarrow G$ as induced by a selfmap $f: X \rightarrow X$ of a closed hyperbolic surface. Then $\text{rank Fix}(\phi) = \text{rank}(\mathbf{F})$ for some fixed point class \mathbf{F} of f .

When ϕ is epimorphic, it is an automorphism because the group G is Hopfian. So, up to homotopy, we can assume f to be a homeomorphism, even to be a standard map of Theorem T. Replacing $\text{ind}(\mathbf{F})$ with $\text{chr}(\mathbf{F})$ in Corollary T, we see $\text{chr}(\mathbf{F}) \geq \chi(X) = 2 - \text{rank}(G)$. If $\mathbf{F} = X$ then $\phi = \text{id}$, while if not, we have $\text{chr}(\mathbf{F}) = 1 - \text{rank}(\mathbf{F})$, hence $\text{rank Fix}(\phi) < \text{rank}(G)$.

When ϕ is not epimorphic, f is homotopic to a non-surjective selfmap. Replacing $\text{ind}(\mathbf{F})$ with $\text{chr}(\mathbf{F})$ in Proposition 4.2, we have $1 - \text{rank}(\mathbf{F}) = \text{chr}(\mathbf{F}) \geq \frac{1}{2}\chi(X) = 1 - \frac{1}{2} \text{rank}(G)$, so that $\text{rank Fix}(\phi) \leq \frac{1}{2} \text{rank}(G)$. □

5 Examples

Theorem 1.1 places restrictions on the possible (chr, ind) pairs. In Table 1, \boxtimes stands for impossible values, \square for values controlled by (B), \blacksquare for uncontrolled ones, and \bigcirc for situations where empty fixed point classes persist.

| chr \ ind | 1 | 0 | -1 | -2 | ... | χ | ... | 2χ +2 | 2χ +1 | 2χ | 2χ -1 |
|-----------|----------------|-------------|----------------|-------------|-------------|-----------|-------------|---------------|---------------|-------------|---------------|
| 1 | \blacksquare | \bigcirc | \blacksquare | \square | ... | \square | ... | \square | \square | \square | \square |
| 0 | \boxtimes | \bigcirc | \square | \square | ... | \square | ... | \square | \square | \square | \boxtimes |
| -1 | \boxtimes | \boxtimes | \square | \square | ... | \square | ... | \square | \square | \boxtimes | \boxtimes |
| -2 | \boxtimes | \boxtimes | \boxtimes | \square | ... | \square | ... | \square | \boxtimes | \boxtimes | \boxtimes |
| \vdots | \boxtimes | \boxtimes | \boxtimes | \boxtimes | \ddots | \square | \ddots | \boxtimes | \boxtimes | \boxtimes | \boxtimes |
| χ | \boxtimes | \boxtimes | \boxtimes | \boxtimes | \boxtimes | \square | \boxtimes | \boxtimes | \boxtimes | \boxtimes | \boxtimes |

Table 1: Possible values of ind and chr

A natural question is whether all pairs marked with \square are realizable. We shall see below that the answer is “yes” for graphs (hence also for surfaces with boundary). But for closed orientable surfaces, there are exceptions.

Example 5.1 Let X be the wedge sum of $n > 1$ circles. Take the common point of these circles to be the base point. Suppose r and d are integers with $0 \leq r \leq n$ and $0 \leq d \leq 2n - 2r$. We shall construct a selfmap $f: X \rightarrow X$ which has a fixed point class \mathbf{F} such that $\text{chr}(\mathbf{F}) = 1 - r$ and $\text{ind}(\mathbf{F}) = \text{chr}(\mathbf{F}) - d$. This means all pairs (chr, ind) permitted in Table 1 are realizable.

The construction follows. Let γ_i be the simple loop along the i th circle. Define f by

$$f \circ \gamma_i = \begin{cases} \gamma_i & \text{if } i \leq r, \\ \gamma_i^2 & \text{if } r < i \leq r + \frac{d}{2}, \\ \gamma_i \gamma_1 & \text{if } i = r + \frac{d+1}{2} > 1, \\ \gamma_i \gamma_n & \text{if } i = r + \frac{d+1}{2} = 1, \\ 1 & \text{if } r + \frac{d+1}{2} < i \leq n. \end{cases}$$

We can perturb f slightly on the first r circles so that the base point is the only fixed point. Clearly it has the desired characteristic and index.

Now consider a hyperbolic closed surface X . We will try to construct a standard map (see Section 3) $f: X \rightarrow X$ with a fixed point class \mathbf{F} such that $c = \text{chr}(\mathbf{F}) \leq 1$ and $d = \text{chr}(\mathbf{F}) - \text{ind}(\mathbf{F}) \geq 0$. The area marked with \square in Table 1 is characterized by the inequality $\max\{2c + 1, 0\} \leq d \leq 2c - 2\chi(X)$.

We need some direct consequences of Masur and Smillie [8, Theorem 2] (where 2-prong punctures are not taken into account). Let T_g denote the orientable closed surface of genus $g \geq 0$, and let $T_{g,k}$ denote the k -times ($k \geq 0$) punctured T_g . They represent all compact orientable surfaces.

- Lemma 5.2** (1) *There exists a pseudo-Anosov map on $T_{g,k}$ ($k \geq 0$) with a p -prong interior singularity if and only if $3 \leq p \leq 4g - 2 + k = -2\chi(T_{g,k}) - k + 2$ and $(\chi(T_{g,k}), k, p) \neq (-1, 1, 3), (-2, 0, 5)$.*
- (2) *There exists a pseudo-Anosov map on $T_{g,k}$ ($k \geq 1$) such that the total prong number of the punctures is p if and only if $k \leq p \leq 4g - 4 + 2k = -2\chi(T_{g,k})$ and $(\chi(T_{g,k}), k, p) \neq (-1, 1, 1), (-2, 2, 3), (-3, 1, 5)$.*

In the following constructions, the pseudo-Anosov maps are assumed to preserve every prong, and to keep the boundary fixed pointwise.

Example 5.3 (for $\text{chr}(\mathbf{F}) = 1$ on orientable X) Suppose X is a hyperbolic orientable closed surface X . Let $3 \leq d \leq -2\chi(X) + 2$. By Lemma 5.2(1), except when $(\chi(X), d) = (-2, 5)$, there exists a pseudo-Anosov map f on X such that one of its singularities is d -pronged. This singularity forms a fixed point class \mathbf{F} with $\text{chr}(\mathbf{F}) = 1$ and $\text{chr}(\mathbf{F}) - \text{ind}(\mathbf{F}) = d$.

So, every entry (chr, ind) in the first row ($\text{chr} = 1$) of Table 1 is realizable by a pseudo-Anosov map, except when $(\chi; \text{chr}, \text{ind}) = (-2; 1, -4)$. It is easy to see that this one fails also for periodic and reducible maps, so it is actually unrealizable.

Example 5.4 (for $\text{chr}(\mathbf{F}) = 1$ on non-orientable X) A hyperbolic non-orientable closed surface X can be cut open along a circle Γ into an *orientable* surface X' with k punctures, where $k = 1$ or 2 according as $\chi(X)$ is odd or even, respectively. Clearly $\chi(X') = \chi(X)$.

Let f' be a pseudo-Anosov map on X' which has an interior d -prong singularity. Let $f: X \rightarrow X$ be the map induced by f' . The d -prong singularity forms a fixed point class \mathbf{F} of f with $\text{chr}(\mathbf{F}) = 1$ and $\text{ind}(\mathbf{F}) = 1 - d$.

By Lemma 5.2(1), this construction works if and only if $3 \leq d \leq -2\chi(X) - k + 2$ and $(\chi(X), k, d) \neq (-1, 1, 3)$. Hence, in the first row of Table 1, an entry (chr, ind) is realizable if

- for $\chi(X) = -1$: $\text{ind} \neq -2$ and -3 ;
- for odd $\chi(X) \leq -3$: $\text{ind} \neq 2\chi(X) - 1$;
- for even $\chi(X)$: $\text{ind} \neq 2\chi(X)$ and $2\chi(X) - 1$.

We are unable to decide the realizability of these exceptional entries.

Example 5.5 (for $\text{chr}(\mathbf{F}) \leq 0$) Suppose a closed hyperbolic surface X is cut along a system of circles $\Gamma = \Gamma_1 \cup \dots \cup \Gamma_k$, $k \geq 1$, into two pieces $X_1 \sqcup X_2$, both are k -times punctured surfaces other than the disk, and X_2 is orientable. We observe that when $\chi(X_1) = c$ is given, $\chi(X) \leq c \leq 0$, the choice of k is limited to the range $1 \leq k \leq 2 - \max\{c, \chi(X) - c\}$ and $k \equiv \chi(X) - c \pmod{2}$.

Let f_2 be a pseudo-Anosov map on X_2 such that the total prong number of the punctures is d when $d \geq k$ or k when $0 \leq d < k$. If such an f_2 does not exist, then we take f_2 to be the identity map on X_2 . Define f to be the standard map with $f|_{X_1} = \text{id}$, $f|_{X_2} = f_2$, and with no twist on Γ if $d \geq k$, but with Dehn twists on $k - d$ components of Γ if $0 \leq d < k$. Then X_1 forms a fixed point class \mathbf{F} with $\text{chr}(\mathbf{F}) = c$ and $\text{chr}(\mathbf{F}) - \text{ind}(\mathbf{F}) = d$ (see Jiang and Guo [6, Lemma 3.6]). By Lemma 5.2(2), this construction works if and only if $0 \leq d \leq 2c - 2\chi(X)$ and $(\chi(X) - c, k, d) \neq (-1, 1, 1), (-1, 3, 1), (-1, 3, 2), (-2, 2, 3), (-3, 1, 5)$.

On a given X , we can try the construction for every k in the range stated above. Then we see that in the ‘ $\text{chr} < 1$ ’ rows of Table 1, a permitted entry (chr, ind) is realizable on X if $(\chi; \text{chr}, \text{ind}) \neq (\chi(X); \chi(X) + 1, \chi(X)), (-2; 0, -3), (-3; 0, -5), (-3; -1, -4)$. For example, $(-4; -1, -6)$ can be realized for $k = 3$ and $X_2 = T_{1,3}$, $(-4; -2, -5)$ can be realized for $k = 4$ and $X_2 = T_{0,4}$.

When X is orientable, for a fixed point class \mathbf{F} of a standard map f with $(\chi; \text{chr}, \text{ind}) = (\chi(X); \chi(X) + 1, \chi(X))$, \mathbf{F} must be an identity piece next to a pseudo-Anosov piece of Euler characteristic -1 . Such a decomposition is ruled out by Lemma 5.2(2). Hence the entry $(\chi; \text{chr}, \text{ind}) = (\chi(X); \chi(X) + 1, \chi(X))$ is indeed unrealizable.

For the entry $(\chi; \text{chr}, \text{ind}) = (-2; 0, -3)$, X is the double torus, \mathbf{F} is a strip joining pseudo-Anosov pieces. It cuts X either into a 2-punctured torus, or into two 1-punctured tori. The former case is ruled out by Lemma 5.2(2). The latter case is ruled out because the puncture on either component must have 2 prongs, but to make $\text{ind}(\mathbf{F}) = -3$ the prong numbers should add up to 3. Hence the entry $(\chi; \text{chr}, \text{ind}) = (-2; 0, -3)$ is also unrealizable.

When X is non-orientable, we are unable to decide the realizability of the exceptional entries.

References

- [1] **M Bestvina, M Handel**, *Train tracks and automorphisms of free groups*, Ann. of Math. (2) 135 (1992) 1–51 MR1147956
- [2] **W Dicks, E Ventura**, *The group fixed by a family of injective endomorphisms of a free group*, Contemporary Mathematics 195, American Mathematical Society, Providence, RI (1996) MR1385923
- [3] **A Fathi, F Laudenbach, V Poénaru, editors**, *Travaux de Thurston sur les surfaces*, Astérisque 66, Société Mathématique de France, Paris (1979) MR568308 With an English summary
- [4] **B J Jiang**, *Lectures on Nielsen fixed point theory*, Contemporary Mathematics 14, American Mathematical Society, Providence, R.I. (1983) MR685755
- [5] **B Jiang**, *Bounds for fixed points on surfaces*, Math. Ann. 311 (1998) 467–479 MR1637903
- [6] **B J Jiang, J H Guo**, *Fixed points of surface diffeomorphisms*, Pacific J. Math. 160 (1993) 67–89 MR1227504
- [7] **R C Lyndon, P E Schupp**, *Combinatorial group theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete 89, Springer, Berlin (1977) MR0577064
- [8] **H Masur, J Smillie**, *Quadratic differentials with prescribed singularities and pseudo-Anosov diffeomorphisms*, Comment. Math. Helv. 68 (1993) 289–307 MR1214233
- [9] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. (N.S.) 19 (1988) 417–431 MR956596
- [10] **H Zieschang**, *Über einfache Kurven auf Vollbrezeln*, Abh. Math. Sem. Univ. Hamburg 25 (1961/1962) 231–250 MR0149469

*Department of Mathematics, Peking University
Beijing 100871, China*

*Department of Mathematics, Indiana University
Bloomington IN 47405, USA*

*School of Science, Xi'an Jiaotong University
Xi'an 710049, China*

bjjiang@math.pku.edu.cn, shidawang@gmail.com,
zhangqiang.math@gmail.com

Received: 10 October 2010 Revised: 16 February 2011