Math. Ann. 311, 467–479 (1998)

Mathematische
Annalen

Bounds for fixed points on surfaces

Boju Jiang

Department of Mathematics, Peking University, Beijing 100871, China
(e-mail: jiangbj@sxx0.math.pku.edu.cn)

Received: 27 May 1997 / in revised form: 4 November 1997

Mathematics Subject Classification (1991): Primary 55M20; Secondary 20F32, 57M15, 57N05, 58C30, 58F20

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 $\mathrm{Fix} f := \{x \in X \mid x = f(x)\}$ that are invariant under homotopies of the map f (see [J1] for an introduction).

The fixed point set $\mathrm{Fix} f$ 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 path which is homotopic (relative to end-points) to its own f-image. Each fixed point class $\mathbf{F}$ is an isolated subset of $\mathrm{Fix} f$ hence its index $\mathrm{ind}(f, \mathbf{F}) \in \mathbb{Z}$ is defined. A fixed point class with non-zero index is called *essential*. The number of essential fixed point classes is called the *Nielsen number* $N(f)$ of f. It is a homotopy invariant of f, so that every map homotopic to f must have at least $N(f)$ fixed points.

Nielsen fixed point theory is related to dynamics through the theory of periodic points, i.e. the study of the fixed points of the iterates $\{f^n \mid n = 1, 2, \ldots\}$ of f. One would like to understand how the invariants of the map f^n will change when the power n changes.

In the asymptotic study of surface homeomorphisms [J3], the following *a priori* bound for the index of Nielsen fixed point classes plays an important role.

Theorem JG. (Cf. [JG, Theorem 4.1]) *Suppose X is a connected compact surface X with Euler characteristic $\chi(X) < 0$, and suppose $f : X \to X$ is a self-homeomorphism. Then the index of the Nielsen fixed point classes of f is bounded:*

Partially supported by NSFC.

(A) *every fixed point class* $\mathbf{F}$ *of* f *has index* $\mathrm{ind}(f, \mathbf{F}) \leq 1$;
(B) *almost every fixed point class* $\mathbf{F}$ *of* f *has index* $\mathrm{ind}(f, \mathbf{F}) \geq -1$, *in the sense that*

$$\sum(\mathrm{ind}(f, \mathbf{F}) + 1) \geq 2\chi(X),$$

where the summation is over all classes $\mathbf{F}$ *with* $\mathrm{ind}(f, \mathbf{F}) < -1$. *Hence*
(C) $|L(f) - \chi(X)| \leq N(f) - \chi(X)$, *where* $L(f)$ *and* $N(f)$ *are the Lefschetz number and the Nielsen number of* f *respectively.*

The purpose of the present note is to show that these index bounds are indeed valid for any selfmap of a compact surface with $\chi < 0$, thus answering an earlier question [JG, end of Sect. 4].

These index bounds have attracted much attention. In an algorithmic study, Joyce Wagner [W] has verified them for a class of selfmaps of compact surfaces with boundary.

By a geometric approach, Michael Kelly [K1] has established the following result:

Theorem K. *Let* X *be a connected compact surface with boundary. Suppose a selfmap* $f : X \to X$ *is minimal in the sense that it has the least number of fixed points possible among all maps homotopic to* f. *Then the index of fixed points of* f *is bounded:*

(A′) *every fixed point* x *of* f *has index* $\mathrm{ind}(f, x) \leq 1$;
(B′) *almost every fixed point* x *of* f *has index* $\mathrm{ind}(f, x) \geq -1$, *in the sense that*

$$\sum(\mathrm{ind}(f, x) + 1) \geq 2\chi(X),$$

where the summation is over all fixed points x *with* $\mathrm{ind}(f, x) < -1$. *Hence*
(C′) $|L(f) - \chi(X)| \leq MF[f] - \chi(X)$, *where* $MF[f]$ *is the minimum number of fixed points in the homotopy class of* f.

We have phrased Kelly's result parallel to ours, but neither one implies the other (although (C) is stronger than (C′)), because on surfaces it is not always possible to coalesce an essential fixed point class into a single fixed point [J2]. In a recent preprint [K2], he also proved (A).

Our approach is to reduce the study of surface maps to that of surface homeomorphisms and of graph maps, by means of the notion of mutant. (We shall see that if a map is not homotopic to any homeomorphism of the surface, it can be deformed into a non-surjective map, hence can be replaced by a map on a spine of the punctured surface.) Just as the former type was analyzed in [JG] by Thurston's theory of surface homeomorphisms, we treat the latter type using the corresponding Bestvina-Handel theory for graph maps.

The paper is organized as follows. The notion of mutant is introduced in Sect. 1. Section 2 prepares the reader with the Bestvina-Handel theory for graph maps. The technical core of the paper is Sect. 3 where the index bounds are established for graph maps. Section 4 completes our proof for surface maps. In Sect. 5, the index bounds are used to establish an equality for asymptotic invariants. Section 6 asks a question about index bounds for more general polyhedra.

Acknowledgements. The author thanks Mladen Bestvina and Warren Dicks for helpful discussions.

1. Mutants

In this section we introduce the notion of mutant, and show that the Nielsen fixed point invariants are invariants of mutants.

Definition. *Let $f : X \to X$ and $g : Y \to Y$ be selfmaps of compact connected polyhedra. We say g is obtained from f by commutation, if there exist maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $f = \psi \circ \varphi$ and $g = \varphi \circ \psi$. We say g is a mutant of f if there is a sequence $\{f_i : X_i \to X_i \mid i = 0, \dots, k\}$ of selfmaps of compact polyhedra such that $f_0 = f$, $f_k = g$, and for each i, either f_{i+1} is obtained from f_i by commutation, or $X_{i+1} = X_i$ and f_{i+1} is homotopic to f_i.*

It is easy to see that if f and g are of the same homotopy type (see [J1, Definition I.5.3]), then they are mutants of each other.

As an immediate consequence of the homotopy invariance and the commutativity property [J1, Theorems I.4.5 and I.5.2] of the indices of essential fixed point classes, we have:

Proposition. *Mutants have the same set of indices of essential fixed point classes, hence also the same Lefschetz number and Nielsen number.* □

2. Graph maps: preparations

In this section, we first explain our terminology concerning graph maps, then we gather the results that we need from the Bestvina-Handel theory.

There are two approaches to the Bestvina-Handel theory: the original Bestvina-Handel paper [BH] and the groupoid version of Dicks-Ventura [DV]. Although the latter is more convenient for quoting specific results (cf. the Remark at the end of this section), we shall use the former as our main reference because of our topological context. The terminology we use is also close to [BH].

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* $\alpha : X \to Y$ is a cellular map, i.e. 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 constant map; in the latter case we say that the edge e is *α-pretrivial.* A graph map $\alpha : X \to Y$ 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 $p : [0, 1] \to X$ 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 $\alpha : X \to Y$ induces a function $D\alpha$ on the set of oriented edges of X. It sends a non-pretrivial oriented edge e to the first oriented edge of $\alpha(e)$; if e is α-pretrivial we say $D\alpha(e) = 0$.

A *turn* in X is an unordered pair of distinct oriented edges of X starting at a common vertex. It *degenerates* under α if its $D\alpha$-image is no longer a turn.

Suppose $\alpha : X \to Y$ is a graph map that maps edges to edges. If a turn $\{e_1, e_2\}$ in X degenerates under α (i.e. $\alpha(e_1) = \alpha(e_2)$), we can identify the two edges e_1 and e_2 into a single edge e' to obtain a new graph X' and a graph map $\alpha' : X' \to Y$. This operation will be called a *fold*, or *folding the turn* $\{e_1, e_2\}$. Cf. [S, Sect. 3.3].

Suppose $\beta : Z \to Z$ is a selfmap of a graph 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 Lemma is a slight modification of [DV, Lemma I.2.5] (originally due to Stallings [S]).

The Folding Lemma. *Any graph map $\gamma : Y \to Z$ can be expressed as a composition of graph maps*

$$\gamma = \gamma' \circ \eta, \qquad \eta = \beta_m \circ \cdots \circ \beta_1 \circ \delta \circ \alpha,$$

where α collapses pretrivial edges, δ subdivides, the β_i fold turns, and $\gamma' : Y' \to Z$ is a graph immersion.

The projection $\eta : Y \to Y'$ induces a surjective homomorphism of the fundamental group on each component of Y. When γ is π_1-injective, η is a homotopy equivalence; otherwise $\chi(Y') > \chi(Y)$.

Proof. Collapsing all γ-pretrivial edges in Y, we obtain a graph Y_0 and a graph map $\gamma_0 : Y_0 \to Z$ with no γ_0-pretrivial edges. Subdividing Y_0 at all γ_0-preimages of vertices of Z, we get a graph Y_1 and a graph map $\gamma_1 : Y_1 \to Z$ which maps edges to edges.

If some turn in Y_1 degenerates under γ_1, we can fold this turn to obtain a graph map $\gamma_2 : Y_2 \to Z$ which still maps edges to edges, and Y_2 has fewer edges than Y_1. Repeating the preceding step as often as possible, we finally obtain a graph map $\gamma_m : Y_m \to Z$, $m \geq 1$, which maps edges to edges and degenerates no turn in Y_m. This is the desired graph immersion γ'.

The γ-pretrivial edges in Y generate a subgraph $W \subset Y$. The map α collapses each component of W to a point, so it induces a surjection of π_1 on each component of Y. It is a homotopy equivalence if and only if W is a forest, otherwise it increases the Euler characteristic χ. The subdivision map δ is a homeomorphism. A fold also induces a surjection of π_1 on each component. It is a homotopy equivalence if and only if the turn folded is open, i.e. if the two oriented edges involved terminate at distinct vertices; otherwise it increases χ. It follows that η is a homotopy equivalence if and only if γ is π_1-injective; otherwise $\chi(Y') > \chi(Y)$. □

The following theorem summarizes the results of Bestvina-Handel [BH] that we need. Any unexplained terminology and notation in its proof are taken from that paper.

Theorem BH. *Let X be a connected graph and $h : X \to X$ be a π_1-injective map. Then h is of the same homotopy type as a graph selfmap $\beta : Z \to Z$, where Z is a connected graph without vertices of valence* 1, *and there is a β-invariant proper subgraph Z_0, containing all vertices of Z. The map $\beta : (Z, Z_0) \to (Z, Z_0)$ 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) *If there exists an indivisible β-Nielsen path which does not lie in Z_0, it is unique.*

(c) *If p is an indivisible β-Nielsen path which does not lie in 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. Note that all results of [BH] remain true for π_1-injective selfmaps of connected graphs, i.e. they can be generalized from automorphisms to injective endomorphisms of free groups. The only place where surjectivity was used in that paper was in applying the Bounded Cancellation Lemma, of which the original Cooper-Thurston proof uses surjectivity. A stronger version of the Bounded Cancellation Lemma assuming only injectivity has since appeared as Lemma II.2.4 in the paper [DV].

According to [BH, Theorem 5.12], there exists a stable relative train track map $f : G \to G$ of the same homotopy type as the given map h. In view of [BH, Lemma 5.2], we can assume that the graph G has no vertices of valence 1.

Suppose m is the length of the maximal filtration. Then H_m is the top stratum. Let M_m be the corresponding transition submatrix, and $\lambda = \lambda_m$ be the Perron-Frobenius eigenvalue.

When $\lambda = 0$, M_m is the zero matrix, so f sends G into G_{m-1}. Regard $f : G \to G$ as $\beta : Z \to Z$, and let Z_0 be the union of G_{m-1} with all vertices of G. It is of Type 1 which would not appear if h is a homopoty equivalence.

When $\lambda = 1$, M_m is an irreducible permutation matrix. This means that for every edge e in H_m, exactly one edge e' of H_m occurs in $f(e)$ and occurs only once. Suppose $f(e) = q_1 \cdot e' \cdot q_3$, where q_1, q_3 are (possibly trivial) paths in G_{m-1}. Breaking each edge e of H_m into 3 edges e_1, e_2, e_3 by inserting 2 new vertices in its interior, we get a new graph Z. Up to a homotopy rel G_{m-1}, we may assume that f induces a graph map $\beta : Z \to Z$ which coincides with f on G_{m-1}, and such that $\beta(e_1) = q_1 \cdot e'_1$, $\beta(e_2) = e'_2$ and $\beta(e_3) = e'_3 \cdot q_3$. Let Z_0 be the subgraph obtained from Z by deleting all the middle edges $\{e_2 \mid e \in H_m\}$. Then β maps Z_0 into itself and permutes the unoriented edges in $Z \setminus Z_0$ according to the irreducible permutation matrix M_m. Thus $\beta : (Z, Z_0) \to (Z, Z_0)$ is of Type 2.

This leaves the case $\lambda > 1$ which means the top stratum H_m is exponentially growing. Take $f : G \to G$ as $\beta : Z \to Z$, and take Z_0 to be the union of G_{m-1} with all vertices of G. We show that it is of Type 3.

In fact, by [BH, Lemma 5.10], β expands edges of $Z \setminus Z_0$ by the factor $\lambda > 1$ with respect to a length function L supported on $Z \setminus Z_0$. Property (a) is the same as Property (RTT-i) of [BH]. Property (b) follows from [BH, Theorem 5.15]. Suppose p is an indivisible β-Nielsen path which does not lie in Z_0. It follows from [BH, Lemma 5.8] that both the initial and terminal tips of p must lie in $Z \setminus Z_0$, otherwise p would be a concatenation of β-Nielsen subpaths, contradicting its indivisibility. By [BH, Lemma 5.11(1)], p contains exactly one illegal turn in $Z \setminus Z_0$. It is the unique illegal turn in $Z \setminus Z_0$ according to [BH, Theorem 5.15], hence it degenerates under β. Thus Property (c) is true. □

Remark. For the convenience of the reader, we also include a proof based on the Dicks-Ventura paper [DV]. Any unexplained notation in it refers to that paper.

Alternative Proof of Theorem BH based on [DV]. It follows from [DV, Theorem IV.1.1] that such a map $\beta : (Z, Z_0) \to (Z, Z_0)$ exists. Let λ be the Perron-Frobenius eigenvalue of the matrix $[\beta/Z_0]$. Then β expands edges of $Z \setminus Z_0$ by a factor of λ with respect to a Perron-Frobenius pseudometric ℓ for (β, Z, Z_0).

When $\lambda = 0$ we are in Type 1. Otherwise $\lambda \geq 1$. Then Property (a) restates [DV, Proposition IV.1.5(ii)].

When $\lambda = 1$, $[\beta/Z_0]$ is a permutation matrix. This means that for every edge e in $Z \setminus Z_0$, exactly one edge e' of $Z \setminus Z_0$ occurs in $\beta(e)$ and occurs only once. Property (a) implies $\beta(e) = e'^{\pm 1}$. Hence β permutes the unoriented edges in $Z \setminus Z_0$. Since $[\beta/Z_0]$ is irreducible, it is a cyclic permutation matrix. Thus we get Type 2.

The case $\lambda > 1$ is of Type 3. Property (b) follows from [DV, Theorem IV.4.3(iv)]. Property (c) is a consequence of [DV, Proposition IV.3.2 and Lemma IV.3.4]. □

3. Index bounds for graph maps

The aim of this section is to prove the index bounds for graph selfmaps.

Lemma A. *Let X be a connected graph and $f : X \to X$ be a selfmap. Then f is a mutant of a graph map $g : Y \to Y$ such that Y is connected, $\chi(Y) \geq \chi(X)$ and g is π_1-injective.*

Proof. Suppose f is not π_1-injective. It follows from the Folding Lemma that there are graph maps $\alpha : X \to X_1$ and $\beta : X_1 \to X$ such that $f = \beta \circ \alpha$, X_1 is connected and $\chi(X_1) > \chi(X)$. Then $f_1 := \alpha \circ \beta : X_1 \to X_1$ is a mutant of f. Repeat this step as often as possible, we finally obtain a mutant which is π_1-injective. □

Theorem 1. *Let X be a connected graph and $f : X \to X$ be a graph map. Then*

(A) *every fixed point class* $\mathbf{F}$ *of* f *has index* $\mathrm{ind}(f, \mathbf{F}) \leq 1$;
(B) *almost every fixed point class* $\mathbf{F}$ *of* f *has index* $\mathrm{ind}(f, \mathbf{F}) \geq -1$, *in the sense that*

$$\sum_{\mathrm{ind}(f,\mathbf{F})<-1} (\mathrm{ind}(f, \mathbf{F}) + 1) \geq 2\chi(X).$$

Hence $|L(f) - \chi(X)| \leq N(f) - \chi(X)$.

Proof. By Lemma A, without loss of generality we may assume that f is π_1-injective.

By the Proposition in Sect. 1, it suffices to prove the bounds (A) and (B) for the graph selfmap $\beta : (Z, Z_0) \to (Z, Z_0)$ in Theorem BH.

Let Z_i, $i = 1, \ldots, n$ be the connected components of Z_0. Suppose the β-invariant ones are $i = 1, \ldots, k$, $k \leq n$. Denote $\beta_0 := \beta|_{Z_0} : Z_0 \to Z_0$ and $\beta_i := \beta|_{Z_i} : Z_i \to Z_i$ 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 $\sum_{1 \leq i \leq n,\ \chi(Z_i)<0} \chi(Z_i) > \chi(Z)$, hence $\chi(Z_i) > \chi(Z)$ for all $1 \leq i \leq n$.

So, working inductively (note that the Theorem is trivial if $\chi(Z) \geq 0$), we may assume that the Theorem is true for all $\beta_i : Z_i \to Z_i$, $i = 1, \ldots, k$.

We now investigate the fixed point classes of β and their indices, for the Types 1–3.

For Type 1, all β-Nielsen paths lie in Z_0, hence every fixed point class of β is a fixed point class $\mathbf{F}_i$ of some β_i, $i = 1, \ldots, k$, with $\mathrm{ind}(\beta, \mathbf{F}_i) = \mathrm{ind}(\beta_i, \mathbf{F}_i)$.

Type 2 can be divided into several subtypes. Recall that the transition submatrix M for the edges in $Z \setminus Z_0$ is an irreducible permutation matrix.

Type 2a: M is the 1×1 matrix 1, and $\beta(e) = e$ for the oriented edge e in $Z \setminus Z_0$.

By a small perturbation of β on e we may assume that no interior point of e is fixed, thus β has the same fixed points as β_0. There are two further subtypes:

Type 2a1: Both ends of e are in the same fixed point class $\mathbf{F}'_0$ of β_0. The fixed point classes of β are those of β_0, with the same index except that $\mathrm{ind}(\beta, \mathbf{F}'_0) = \mathrm{ind}(\beta_0, \mathbf{F}'_0) - 1$.

Type 2a2: The edge e joins two different fixed point classes $\mathbf{F}'_1$ and $\mathbf{F}'_2$ of β_0. The fixed point classes of β are those of β_0, with the same index, except the fixed point class $\mathbf{F}' = \mathbf{F}'_1 \cup \mathbf{F}'_2$ of β with $\mathrm{ind}(\beta, \mathbf{F}') = \mathrm{ind}(\beta_0, \mathbf{F}'_1) + \mathrm{ind}(\beta_0, \mathbf{F}'_2) - 1$.

Type 2b: M is the 1×1 matrix 1, and $\beta(e) = \bar{e}$ for the oriented edge e in $Z \setminus Z_0$.

Type 2c: M is a cyclic permutation matrix, hence β cyclically permutes the unoriented edges of $Z \setminus Z_0$.

For Type 2b and Type 2c, every β-Nielsen path must lie in Z_0, hence every fixed point class $\mathbf{F}_i$ of β_i, $i = 1, \ldots, k$, is a fixed point class of β with $\mathrm{ind}(\beta, \mathbf{F}_i) = \mathrm{ind}(\beta_i, \mathbf{F}_i)$. In Type 2b, β has an additional fixed point class consisting of a single fixed point in e, with index 1.

In Type 3, we first introduce some terminology and notation. For any subset $S \subset Z$, let $\omega(S)$ be the number of oriented edges e in $Z \setminus Z_0$ starting from a vertex in S; and let $\delta(S)$ be the number of oriented edges e in $Z \setminus Z_0$ starting from a vertex in S such that $D\beta(e) = e$ (hence e is expanded along itself by β). A *pre-fpc* is either a fixed point class $\mathbf{F}_i$ of some β_i, $i = 1, \ldots, k$, or an isolated fixed point of β in $Z \setminus Z_0$. As an isolated subset of Fix β, their indices are as follows: In the former kind $\text{ind}(\beta, \mathbf{F}_i) = \text{ind}(\beta_i, \mathbf{F}_i) - \delta(\mathbf{F}_i)$. In the latter kind, the index is ± 1 because $Z \setminus Z_0$ is a union of open edges.

There are two subtypes.

Type 3a: Either there is no indivisible β-Nielsen path which does not lie in Z_0, or the unique (by Property (c)) such path p has both ends in the same pre-fpc. Then the fixed point classes of β are just the pre-fpc's defined above.

Type 3b: The unique indivisible β-Nielsen path p which does not lie in Z_0 connects two different pre-fpc's. Then the fixed point classes of β are the pre-fpc's, except that the two pre-fpc's joined by p combines into a single fixed point class of β.

With the above information about fixed point classes, bounds (A) and (B) for β can be proved from the inductive hypothesis that they hold for all β_i. We shall omit the easier types and concentrate on Type 3b.

The inequality (A) is obvious except for the combined fixed point class. If the two pre-fpc's being combined are $\mathbf{F}'_i$ and $\mathbf{F}'_j$, then its index is $\text{ind}(\beta, \mathbf{F}'_i) + \text{ind}(\beta, \mathbf{F}'_j) \leq \text{ind}(\beta_i, \mathbf{F}'_i) + \text{ind}(\beta_j, \mathbf{F}'_j) - 2$ since both $\delta(\mathbf{F}'_i) \geq 1$ and $\delta(\mathbf{F}'_j) \geq 1$ by Property (c). If, on the other hand, at least one of the pre-fpc's is a fixed point in $Z \setminus Z_0$, its index must be -1. In either case (A) follows from the inductive hypothesis.

To prove (B), denote

$$S_i := \sum_{\substack{\mathbf{F}_i \subset Z_i \\ \text{ind}(\beta_i, \mathbf{F}_i) - \delta(\mathbf{F}_i) < -1}} (\text{ind}(\beta_i, \mathbf{F}_i) - \delta(\mathbf{F}_i) + 1)$$

for $1 \leq i \leq k$, and let $S_i := 0$ for $k < i \leq n$. Observe that regardless of the nature of the two pre-fpc's being combined, we always have

$$\sum_{\text{ind}(\beta, \mathbf{F}) < -1} (\text{ind}(\beta, \mathbf{F}) + 1) \geq -1 + \sum_{i=1}^{n} S_i.$$

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 $i > k$. So we can assume $i \leq k$.

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

$$\begin{aligned} S_i \ &\geq \sum_{\substack{\mathbf{F}_i \subset Z_i \\ \text{ind}(\beta_i, \mathbf{F}_i) < -1}} (\text{ind}(\beta_i, \mathbf{F}_i) + 1) - \sum_{\mathbf{F}_i \subset Z_i} \delta(\mathbf{F}_i) \\ &\geq 2\chi(Z_i) - \delta(Z_i) \geq 2\chi(Z_i) - \omega(Z_i). \end{aligned}$$

If $\chi(Z_i) = 0$, then Z_i has the homotopy type of the circle. For all $\mathbf{F}_i \subset Z_i$ we have $|\text{ind}(\beta_i, \mathbf{F}_i)| \leq 1$, so

$$S_i \geq -\sum_{\mathbf{F}_i \subset Z_i} \delta(\mathbf{F}_i) \geq 2\chi(Z_i) - \delta(Z_i) \geq 2\chi(Z_i) - \omega(Z_i).$$

If $\chi(Z_i) = 1$, then Z_i has the homotopy type of a point. There is a unique $\mathbf{F}_i \subset Z_i$ and $\text{ind}(\beta_i, \mathbf{F}_i) = 1$, so

$$S_i = \min\{2 - \delta(\mathbf{F}_i), 0\} \geq \min\{2 - \delta(Z_i), 0\} \geq 2\chi(Z_i) - \omega(Z_i).$$

Thus Claim 1 is proved.

By Property (c), there are two oriented edges e_1, e_2 both starting at a vertex v_p such that $D\beta(e_1) = D\beta(e_2)$. Suppose v_p is in the component Z_h. Then at least one of e_1, e_2 is not $D\beta$-invariant, so $\delta(Z_h) < \omega(Z_h)$.

Claim 2: $S_h > 2\chi(Z_h) - \omega(Z_h)$.

In fact, when $\chi(Z_h) \leq 0$, or when $\chi(Z_h) = 1$ and $\delta(Z_h) \geq 2$, then we see from the proof of Claim 1 that

$$S_h \geq 2\chi(Z_h) - \delta(Z_h) > 2\chi(Z_h) - \omega(Z_h).$$

So it remains to examine the case that $\chi(Z_h) = 1$ and $S_h = 0$. We need to show that $\omega(Z_h) \geq 3$.

If Z_h is the single vertex v_p, then $\omega(Z_h) \geq 3$ because v_p has valence ≥ 3 in Z. If otherwise Z_h is a nontrivial tree, there is another vertex v' having valence 1 in Z_h. But the valence of v' in Z is ≥ 2, so there must be another edge e' in $Z \setminus Z_0$ starting at v'. This shows $\omega(Z_h) \geq 3$ in any case. Hence $S_h = 0 > 2 - 3 \geq 2\chi(Z_h) - \omega(Z_h)$ as desired. Thus Claim 2 is proved.

It follows from Claims 1–2 that

$$\sum_{i=1}^{n} S_i > \sum_{i=1}^{n} (2\chi(Z_i) - \omega(Z_i)) = 2\chi(Z_0) - 2(\chi(Z_0) - \chi(Z)) = 2\chi(Z).$$

So, finally, we have

$$\sum_{\text{ind}(\beta,\mathbf{F})<-1} (\text{ind}(\beta, \mathbf{F}) + 1) \geq -1 + \sum_{i=1}^{k} S_i \geq 2\chi(Z).$$

This is the inequality (B) for β.

The inductive proof of the Theorem is now complete. □

Remark. Theorem 1 is easy under the following additional hypothesis:

(W) The graph map $f : X \to X$ is of the same homotopy type as a graph map $g : Y \to Y$ such that every fixed point class of g consists of a single point.

(The letter W for Wecken, because the property that every fixed point class reduces to a single point is often referred to as the Wecken property.)

Proof of Theorem 1 under hypothesis (W). Without loss we may assume that Y has no vertex of valence 1.

It is clear that if a fixed point class $\mathbf{F}$ of g is in the interior of an edge, then $|\operatorname{ind}(g,\mathbf{F})| \leq 1$. If $\mathbf{F}$ is a vertex v, then $\operatorname{ind}(g,\mathbf{F}) = 1 - \delta(v) \geq 1 - \omega(v)$, where $\delta(v)$ is the number of oriented edges e from v which is expanded along itself by g, and $\omega(v)$ is the valence of v in Y. Hence (A), $\operatorname{ind}(g,\mathbf{F}) \leq 1$; and (B),

$$\sum_{\operatorname{ind}(g,\mathbf{F})<-1} (\operatorname{ind}(g,\mathbf{F})+1) \geq \sum_{\delta(v)>2} (2-\omega(v)) \geq \sum_{v} (2-\omega(v)) = 2\chi(Y). \ \square$$

QUESTION 1. Is the hypothesis (W) valid for all graph selfmaps?

The hypothesis (W) means that all Nielsen paths of g are loops. Observe that the presence of Nielsen paths is not always avoidable in Bestvina-Handel theory, as shown in the examples [BH, Example 5.16] and [DV, Example IV.4.4]. But in these examples the indivisible Nielsen paths are all loops. Question 1 asks whether *open* Nielsen paths are always avoidable.

4. Surface maps

We now study the index bounds on surfaces. The following Lemma is the key.

Lemma B. *Let X be a compact connected surface with $\chi(X) < 0$, and $f : X \to X$ be a selfmap. Then either*

(1) X is a closed surface and f is homotopic to a self-homeomorphism of X; or
(2) f is a mutant of a graph selfmap $g : Y \to Y$, with $\chi(Y) \geq \chi(X)$.

Proof. If X has boundary, X deformation retracts to a connected graph Y, so every map $f : X \to X$ is of the same homotopy type as a map $g : Y \to Y$. So the Lemma is trivial in this case.

Suppose X is a closed surface. There are two cases.

(1) Every selfmap homotopic to f is surjective. Then the absolute degree $|d|$ of f is nonzero. The reader is referred to [Ep] for a treatment of the absolute degree.

By [E] (see the author abstract in Zbl. Math. for a correction in the nonorientable case), the map f is homotopic to a map $f' : X \to X$ which is the composition of a pinch and a branched covering. Since f' is a selfmap, the pinch and the branched cover are both trivial for Euler characteristic reasons. Hence f' is a self-homeomorphism.

(2) The map f is homotopic to some map f' that is not surjective. Suppose $x' \notin f'(X)$. Then $X \setminus \{x'\}$ deformation retracts to a connected graph X_1. So there is a map $\phi : X \to X_1$ such that f is homotopic to $\iota \circ \phi$, where $\iota : X_1 \to X$ is the inclusion. Thus f is a mutant of $f_1 := \phi \circ \iota : X_1 \to X_1$. Note that $\chi(X_1) = \chi(X) - 1$.

Since the inclusion ι is not π_1-injective, neither is f_1. By the Folding Lemma it is a mutant of a π_1-injective graph map $g : Y \to Y$ with $\chi(Y) \geq \chi(X_1)+1 = \chi(X)$. $\square$

Remark. There are other arguments for the case (1) above. For example, when X is orientable, it follows from [ZVC, Theorem 3.3.3] that $|d| = 1$, hence f is homotopic to a self-homeomorphism by [ZVC, Corollary 3.3.9].

Theorem 2. *Let X be a compact connected surface with $\chi(X) < 0$, and $f : X \to X$ be a selfmap. Then the conclusion of Theorem 1 holds true.*

Proof. In view of the Proposition, it suffices to consider the two cases listed in Lemma B.

Case (1) is taken care of by Theorem JG. Case (2) follows from Theorem 1. □

5. Applications to asymptotic invariants

In order to study the asymptotic behavior of the number of periodic orbits, the *asymptotic Nielsen number* $N^{\infty}(f)$ of f was introduced in [J3, Sect. 2]. It is defined as the growth rate of the sequence $\{N(f^n)\}$ of the Nielsen number of iterates of f, when the power n increases to infinity. Another invariant is the *asymptotic absolute Lefschetz number* $L^{\infty}(f)$.

$$N^{\infty}(f) := \limsup_{n\to\infty} N(f^n)^{1/n},$$

$$L^{\infty}(f) := \limsup_{n\to\infty} \left(\sum_{\mathbf{F}^{(n)}} \left| \operatorname{ind}(f^n, \mathbf{F}^{(n)}) \right| \right)^{1/n},$$

where the summation is taken over all fixed point classes $\mathbf{F}^{(n)}$ of f^n.

It was proved that

$$N^{\infty}(f) = L^{\infty}(f)$$

for a surface self-homeomorphism [J3, Corollary 3.5]. This is at the basis of the dynamical applications described in [J3, Sect. 4]. Now we extend it to surface selfmaps.

Theorem 3. *Let X be a compact connected surface with $\chi(X) < 0$, and $f : X \to X$ be a selfmap. Then*

$$L^{\infty}(f) = N^{\infty}(f).$$

Proof. In view of Theorem 2, the equality follows from [J3, Theorem 2.3]. □

Still another asymptotic invariant was introduced and used in [J3]. The *asymptotic irreducible Nielsen number* of $f : X \to X$ is the growth rate

$$NI^{\infty}(f) := \limsup_{n\to\infty} NI(f^n)^{1/n}.$$

Here $NI(f^n)$ is the *Nielsen number of irreducible n-orbits* of f, defined as follows. A fixed point class $\mathbf{F}^{(n)}$ of f^n is *reducible* if for a proper factor m of n and for some (hence every) point $x \in \mathbf{F}^{(n)}$ there exists a path w from x to $f^m(x)$ such that the loop $wf^m(w)f^{2m}(w)\cdots f^{n-m}(w)$ is contractible in X. Otherwise it is *irreducible.* The f-image of an essential irreducible $\mathbf{F}^{(n)}$ is again an essential irreducible class, so we can talk about f-orbits of such $\mathbf{F}^{(n)}$'s. $NI(f^n)$ is defined as the number of f-orbits of essential irreducible fixed point classes of f^n. It is a lower bound to the number of periodic orbits of f of least period n, and it is a mutant invariant of f.

We know [J3, Corollary 3.5] that

$$NI^{\infty}(f) = N^{\infty}(f)$$

for surface self-homeomorphisms. A natural question:

QUESTION 2. Is the equality $NI^{\infty}(f) = N^{\infty}(f)$ also true for surface selfmaps?

For this equality it suffices to prove that the number of reducible essential fixed point classes of f^n is uniformly bounded in n. As before, we only need to prove it for π_1-injective graph selfmaps.

In the light of Thurston's theory of surface homeomorphisms, we would expect that for graph maps there exists an upper bound, depending only on the Euler characteristic of the graph, to the number of reducible essential fixed point classes of f^n for any selfmap f and any n. But for a proof one may need the (yet unpublished) Bestvina-Feighn-Handel results on the iterates of graph maps.

6. A more general question

So far we have restricted our attention to surfaces and graphs. For general compact polyhedra we can introduce the following notions:

Definition. *A compact polyhedron X is said to have the Bounded Index Property (BIP) if there is an integer $B > 0$ such that for any map $f : X \to X$ and any fixed point class* $\mathbf{F}$ *of f, the index $|\operatorname{ind}(f, \mathbf{F})| \leq B$. X has the Bounded Index Property for Homeomorphisms (BIPH) if there is such a bound for all homeomorphisms $f : X \to X$.*

The simplest spaces without BIP are the spheres S^k, $k > 1$. They are simply-connected, so for any selfmap f there is at most one fixed point class, and its index equals the Lefschetz number $L(f) = 1 + (-1)^k \deg f$. It is well known that there are selfmaps of arbitrarily large degree.

Spaces without BIPH are also easy to construct. We give an example of closed manifold.

Example. Let S^3 be regarded as the space of unimodular quaternions, and let $X = S^3 \times S^3$. Let $f : X \to X$ be the homeomorphism given by

$$f(q_1, q_2) = (q_1^2 q_2, q_1 q_2), \qquad f^{-1}(q_1, q_2) = (q_1 q_2^{-1}, q_2 q_1^{-1} q_2).$$

Clearly X is simply-connected, so that the index of the unique fixed point class equals the Lefschetz number $L(f)$.

By the Künneth Theorem, the only nontrivial homology groups of X are $H_0(X) \cong H_6(X) \cong \mathbb{Z}$ and $H_3(X) \cong \mathbb{Z}\oplus\mathbb{Z}$. It is not hard to see that $\deg f = 1$, and the homology homomorphism $f_* : H_3(X) \to H_3(X)$ has matrix (with respect to the standard basis) $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. The eigenvalues of A are $\lambda_1 = (3+\sqrt{5})/2 = 2.618$ and $\lambda_2 = (3-\sqrt{5})/2 = 0.382$. Thus, $L(f^n) = 2 - \lambda_1^n - \lambda_2^n$ is unbounded when n gets large.

We have shown in this paper that graphs and surfaces with negative Euler characteristic have BIP. More generally, we can ask

QUESTION 3. Suppose a compact polyhedron X is aspherical (i.e. $\pi_i(X) = 0$ for all $i > 1$). Does X have BIP or BIPH?

Supporting evidences include results on orientation preserving self-homeomorphisms of geometric 3-manifolds [JWW], and on selfmaps of infrasolvmanifolds [MC].

References

[BH] Bestvina, M., Handel, M., Train tracks and automorphisms of free groups, Annals of Math. **135** (1992), 1–51

[DV] Dicks, W., Ventura, E., The group fixed by a family of injective endomorphisms of a free group, Contemporary Math. vol. 195, Amer. Math. Soc. Providence (1996),

[E] Edmonds, A.L., Deformation of maps to branched coverings in dimension two, Annals of Math. **110** (1979), 113–125, Author abstract: Zbl. Math. #424.57002

[Ep] Epstein, D.B.A., The degree of a map, Proc. London Math. Soc. **16** (1966), 369–383

[J1] Jiang, B., Lectures on Nielsen Fixed Point Theory, Contemporary Math. vol. 14, Amer. Math. Soc., Providence, (1983)

[J2] Jiang, B., Fixed points and braids, Invent. Math. **75** (1984), 69–74; II, Math. Ann. **272** (1985), 249–256

[J3] Jiang, B., Estimation of the number of periodic orbits, Pacific J. Math. **172** (1996), 151–185

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

[JWW] Jiang, B., Wang, S., Wu, Y.-Q., Homeomorphisms of 3-manifolds and the realization of Nielsen number, preprint MSRI, Berkeley, 1996

[K1] Kelly, M., A bound on the fixed-point index for surface mappings, Ergod. Th. & Dynam. Sys., toappear

[K2] Kelly, M., A bound on the fixed point index for a class of self-maps on two-complexes, preprint 1997

[MC] McCord, Ch.K., Estimating Nielsen numbers on infrasolvmanifolds, Pacific J. Math. **154** (1992), 345–368

[S] Stallings, J.R., Topology of finite graphs, Invent. Math. **71** (1983), 551–565

[W] Wagner, J., An algorithm for calculating the Nielsen number on surfaces with boundary, Trans. Amer. Math. Soc. toappear

[ZVC] Zieschang, H., Vogt, E., Coldewey, H.-D., Surfaces and Planar Discontinuous Groups, Lecture Notes in Mathematics vol.835, Springer-Verlag, Heidelberg, 1980