PACIFIC JOURNAL OF MATHEMATICS
Vol. 172, No. 1, 1996

ESTIMATION OF THE NUMBER OF PERIODIC ORBITS

BOJU JIANG

The main theme of this paper is to estimate, for self-maps $f : X \to X$ of compact polyhedra, the asymptotic Nielsen number $N^\infty(f)$ which is defined to be the growth rate of the sequence $\{N(f^n)\}$ of the Nielsen numbers of the iterates of f. The asymptotic Nielsen number provides a homotopy invariant lower bound to the topological entropy $h(f)$. To introduce our main tool, the Lefschetz zeta function, we develop the Nielsen theory of periodic orbits. Compared to the existing Nielsen theory of periodic points, it features the mapping torus approach, thus brings deeper geometric insight and simpler algebraic formulation. The important cases of homeomorphisms of surfaces and punctured surfaces are analysed. Examples show that the computation involved is straightforward and feasible. Applications to dynamics, including improvements of several results in the recent literature, demonstrate the usefulness of the asymptotic Nielsen number.

Introduction.

Motivated by dynamical problems, Nielsen theory of fixed points of self-maps $f : X \to X$ of compact polyhedra was generalized to study periodic points, i.e. solutions to $f^n(x) = x$, where f^n is the n-th iterate. See e.g. [**J1**, §III.4]. As the Nielsen number $N(f)$ is a homotopy invariant lower bound to the number of fixed points of f, the Nielsen number $N(f^n)$ is certainly a lower bound to the number of n-points (i.e. fixed points of the n-th iterate) for any map g homotopic to f.

However, generally speaking, the Nielsen numbers are notoriously difficult to compute. We will demonstrate that the asymptotic growth rate of the sequence $\{N(f^n)\}$ (when $n \to \infty$), which we denote by $N^\infty(f)$, is a more computationally accessible invariant than the sequence itself, yet one that is still useful for dynamics. Although the exact evaluation of $N^\infty(f)$ would be desirable, its estimation is a more realistic goal and, as we shall show, one that is sufficient for many applications.

For an asymptotic study, the first challenge is to develop a unified algebraic formulation for the Nielsen theory of all iterates of f so that we can easily relate the fixed point class data of various f^n. This is why we propose the

Nielsen theory for periodic orbits. The key idea is to work on the mapping torus T_f of f and to count periodic orbits rather than periodic points, then Nielsen equivalent f-orbits on X correspond to freely homotopic closed orbit curves on T_f. (This observation of [**J2**] can actually be traced back to the pioneering work of Fuller [**Fu**] in a different context.) The fixed point data of f^n are organized into the generalized Lefschetz number $L_\Gamma(f^n)$, a homotopy invariant living in the free abelian group generated by the set of conjugacy classes in $\Gamma = \pi_1(T_f)$.

For the sake of practical computation, we assume that a matrix representation ρ of Γ is given. The traces of the ρ-images of the generalized Lefschetz numbers constitute a sequence of complex numbers. The Lefschetz zeta function ζ_f is a generating function for this sequence which turns out to be a rational function easily computable for cellular maps. Our Lefschetz zeta function is the same as that of Fried [**F4**] using matrix representations of $\pi_1(T_f)$, rather than the earlier version in [**F1**] using abelian representations, so that non-abelian information can be better retained. This makes a difference in applications, as shown in §4.3.

Now every zero or pole of ζ_f supplies a convenient lower bound for the asymptotic Lefschetz number $L^\infty(f)$ of f, defined to be the growth rate of the sequence $\{\|L_\Gamma(f^n)\|\}$ of norms of the generalized Lefschetz numbers. On the other hand, the asymptotic Lefschetz number is identified with the asymptotic Nielsen number for some important classes of maps.

The sketch above, of the approach to the estimation of $N^\infty(f)$ that we will present in this paper, is given a more detailed exposition in [**J3**].

The structure of the paper is as follows. §1 establishes the basic Nielsen theory of periodic orbits and introduces the Nielsen numbers, the Lefschetz numbers and the Lefschetz zeta function. §2 defines the asymptotic invariants, discusses the conditions for their equality and their relation to the topological entropy, and provides methods for their estimation. §3 analyses the case of homeomorphisms of compact aspherical surfaces and proposes a theory for homeomorphisms of punctured surfaces which often arise in recent 2-dimensional dynamical systems theory. The examples in §4 serve to illustrate our theory. Some open problems are given in §5.

1. Nielsen theory for periodic orbits.

We first give a brief account of the invariants of Nielsen fixed point theory in §1.1. To simplify the algebra involved, we shall work with the natural semi-flow on the mapping torus described in §1.2. The notion of periodic orbit classes is introduced in §1.3. The Lefschetz numbers and Nielsen numbers are then defined in §1.4, and their invariance shown in §1.5. When a matrix representation of the fundamental group of the mapping torus is given, we introduce in §1.6 the associated Lefschetz zeta function. Since this will be our main tool for asymptotic estimates, an analysis of our requirement on the representation is given in §1.7. Finally, in §1.8 we introduce the relative invariants.

1.1. Nielsen theory for fixed points. The basis of Nielsen fixed point theory is the notion of a fixed point class.

Let X be a compact connected polyhedron, $f : X \to X$ be a map. The fixed point set $\operatorname{Fix} f := \{x \in X \mid x = f(x)\}$ 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 $\operatorname{Fix} f$ hence its index $\operatorname{ind}(\mathbf{F}, f) \in \mathbb{Z}$ 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. (Cf. [**J1**, p.19].)

Pick a base point $v \in X$ and a path w from v to $f(v)$. Let $G := \pi_1(X, v)$ and let $f_G : G \to G$ be the composition

$$\pi_1(X,v) \xrightarrow{f_*} \pi_1(X, f(v)) \xrightarrow{w_*} \pi_1(X,v).$$

Two elements $g, g' \in G$ are said to be *f_G-conjugate* if there is an $h \in G$ such that $g' = f_G(h)gh^{-1}$. (There are two definitions of f_G-conjugacy in the literature, related by an inversion. The one we use here is the original one of [**R**] and [**We**] which turns out to be more convenient than the other one used in [**J1**, p. 26].) Thus G splits into f_G-conjugacy classes. Let G_f denote the set of f_G-conjugacy classes, and $\mathbb{Z}G_f$ denote the abelian group freely generated by G_f. We use the bracket notation $a \mapsto [a]$ for both projections $G \to G_f$ and $\mathbb{Z}G \to \mathbb{Z}G_f$, where $\mathbb{Z}G$ is the integral group ring of G.

For every $x \in \operatorname{Fix} f$, its *$G$-coordinate* $\operatorname{cd}_G(x, f) \in G_f$ is defined as follows: Pick a path c from v to x. The f_G-conjugacy class in G of the loop $w(f \circ c)c^{-1}$, which is evidently independent of the choice of c, is called the G-coordinate of x. (This also differs from the definition in [**J1**, p. 26] by an inversion.) Two fixed points are in the same fixed point class if and only if they have the same G-coordinates. The *G-coordinate* $\operatorname{cd}_G(\mathbf{F}, f)$ of a fixed point class $\mathbf{F}$ is then defined to be the common G-coordinate of its members.

The *generalized Lefschetz number* is defined ([**R**], [**We**], cf. [**FH**]) as

$$L_G(f) := \sum_{\mathbf{F}} \operatorname{ind}(\mathbf{F}, f) \cdot \operatorname{cd}_G(\mathbf{F}, f) \quad \in \mathbb{Z}G_f, \tag{1.1}$$

the summation being over all (essential) fixed point classes $\mathbf{F}$ of f. When all fixed points of f are isolated, we also have

$$(1.1') \qquad L_G(f) = \sum_{x \in \operatorname{Fix} f} \operatorname{ind}(x,f) \cdot \operatorname{cd}_G(x,f) \quad \in \mathbb{Z}G_f.$$

The Nielsen number $N(f)$ is the number of non-zero terms in $L_G(f)$, and the indices of the essential fixed point classes appear as the coefficients in $L_G(f)$.

The invariant $L_G(f)$ used to be called the Reidemeister trace because it can be computed as an alternating sum of traces on the chain level ([**R**], [**We**]).

Let $p : \widetilde{X}, \tilde{v} \to X, v$ be the universal covering. The deck transformation group is identified with G. Let $\tilde{f} : \widetilde{X} \to \widetilde{X}$ be the lift of f such that the reference path w lifts to a path from $\tilde{v}$ to $\tilde{f}(\tilde{v})$. Then for every $g \in G$ we have $\tilde{f} \circ g = f_G(g) \circ \tilde{f}$ (cf. [**J1**, pp. 24–25]).

Assume that X is a finite cell complex and $f : X \to X$ is a cellular map. Pick a cellular decomposition $\{e_j^d\}$ of X, the base point v being a 0-cell. It lifts to a G-invariant cellular structure on the universal covering $\widetilde{X}$. Choose an arbitrary lift $\tilde{e}_j^d$ for each e_j^d. They constitute a free $\mathbb{Z}G$-basis for the cellular chain complex of $\widetilde{X}$. The lift $\tilde{f}$ of f is also a cellular map. In every dimension d, the cellular chain map $\tilde{f}$ gives rise to a $\mathbb{Z}G$-matrix $\widetilde{F}_d$ with respect to the above basis, i.e. $\widetilde{F}_d = (a_{ij})$ if $\tilde{f}(\tilde{e}_i^d) = \sum_j a_{ij}\tilde{e}_j^d$, $a_{ij} \in \mathbb{Z}G$. Then we have the Reidemeister trace formula

$$(1.2) \qquad L_G(f) = \sum_d (-1)^d \left[\operatorname{tr} \widetilde{F}_d\right] \quad \in \mathbb{Z}G_f.$$

Remark. The base point v and the path w serve as a reference frame for the G-coordinate. (When v is a fixed point and w is the constant path, the G-coordinate of v is $[1] \in \mathbb{Z}G_f$.) A change of the reference path w would affect the homomorphism f_G, hence also the f_G-conjugacy relation in G and the set G_f where the G-coordinates live. This develops into a considerable mess when we apply the above theory to all the iterates f^n of f, as we are then forced to deal with infinitely many different sets G_{f^n} at the same time. In order to simplify the algebra, we propose the following alternative approach to the coordinates of fixed points.

1.2. The mapping torus. The *mapping torus* T_f of $f : X \to X$ is the space obtained from $X \times \mathbb{R}_+$ by identifying $(x, s+1)$ with $(f(x), s)$ for all $x \in X$, $s \in \mathbb{R}_+$, where $\mathbb{R}_+$ stands for the real interval $[0, \infty)$. On T_f there is a natural semi-flow ("sliding along the rays")

$$\varphi : T_f \times \mathbb{R}_+ \to T_f, \qquad \varphi_t(x,s) = (x, s+t) \text{ for all } t \geq 0.$$

A point $x \in X$ and a positive number $\tau > 0$ determine the *time-τ orbit curve* $\varphi_{(x,\tau)} := \{\varphi_t(x)\}_{0\le t\le\tau}$ in T_f. We may identify X with the cross-section $X \times 0 \subset T_f$, then the map $f : X \to X$ is just the return map of the semi-flow φ.

Take the base point v of X as the base point of T_f. Let $\Gamma := \pi_1(T_f, v)$. By the van Kampen Theorem, Γ is obtained from G by adding a new generator z represented by the loop $\varphi_{(v,1)}w^{-1}$, and adding the relations $z^{-1}gz = f_G(g)$ for all $g \in G$:

$$\Gamma = \langle\, G, z \mid gz = zf_G(g) \text{ for all } g \in G \,\rangle. \tag{1.3}$$

Remark. Note that the homomorphism $G \to \Gamma$ induced by the inclusion $X \subset T_f$ is not necessarily injective. Its kernel equals $\cup_{m>0} \ker(f_G^m)$, the union of the kernels of all iterates of $f_G : G \to G$. This fact can be proved by a topological argument similar to that of [**J2**, §3].

Notation. Let Γ_c denote the set of conjugacy classes in Γ. Theoretically, it is better to regard Γ_c as the set of free homotopy classes of closed curves in T_f, so that it is independent of the base point. Let $\mathbb{Z}\Gamma$ be the integral group ring of Γ, and let $\mathbb{Z}\Gamma_c$ be the free abelian group with basis Γ_c. We use the bracket notation $\alpha \mapsto [\alpha]$ for both projections $\Gamma \to \Gamma_c$ and $\mathbb{Z}\Gamma \to \mathbb{Z}\Gamma_c$.

1.3. Periodic orbit classes. We intend to study the periodic points of f, i.e. the fixed points of the iterates of f.

We shall call $\mathrm{PP}\, f := \{\, (x,n) \in X \times \mathbb{N} \mid x = f^n(x) \,\}$ the *periodic point set* of f, where $\mathbb{N}$ denotes the set of natural numbers. A fixed point x of f^n is called an *n-point* of f, and its f-orbit $\{x, f(x), \dots, f^{n-1}(x)\}$ an *n-orbit* of f. The latter is called a *primary* n-orbit if it consists of n distinct points, i.e. if n is the least period of the periodic point x.

A fixed point class $\mathbf{F}^n$ of f^n will be called an *n-point class* of f. Recall from [**J1**, Proposition III.3.3] that $f(\mathbf{F}^n)$ is also an n-point class, and $\mathrm{ind}(f(\mathbf{F}^n), f^n) = \mathrm{ind}(\mathbf{F}^n, f^n)$. Thus f acts as an index-preserving permutation among its n-point classes. We define an *n-orbit class* of f to be the union of an orbit of this action. In other words, two points $x, x' \in \mathrm{Fix}\, f^n$ are said to be in the same n-orbit class of f if and only if some $f^i(x)$ and some $f^j(x')$ are in the same n-point class of f. The set $\mathrm{Fix}\, f^n$ splits into a disjoint union of n-orbit classes.

On the mapping torus T_f, observe that $(x,n) \in \mathrm{PP}\, f$ if and only if the time-n orbit curve $\varphi_{(x,n)}$ is a closed curve. The free homotopy class of the closed curve $\varphi_{(x,n)}$ will be called the *Γ-coordinate* of (x,n), written $\mathrm{cd}_\Gamma(x,n) = [\varphi_{(x,n)}] \in \Gamma_c$.

It follows from [**J2**, §3] that periodic points $(x,n), (x',n') \in \mathrm{PP}\, f$ have the same Γ-coordinate if and only if $n = n'$ and x, x' belong to the same

n-orbit class of f. Thus we can equivalently define $x, x' \in \operatorname{Fix} f^n$ to be in the same n-orbit class if and only if they have the same Γ-coordinate, and define the Γ-*coordinate* of an n-orbit class $\mathbf{O}^n$ as the common Γ-coordinate of its members, written $\operatorname{cd}_\Gamma(\mathbf{O}^n)$.

Remark. The notion of Γ-coordinate has great algebraic advantage over that of G-coordinate (cf. Remark in §1.1). Ordinary conjugacy classes have replaced the awkward skew-conjugacy classes. The Γ-coordinates $\operatorname{cd}_\Gamma(\mathbf{O}^n)$ are independent of the choice of the base point, and for any n they all live in the same set Γ_c.

An important notion in the Nielsen theory for periodic orbits is that of reducibility. Suppose m is a factor of n and $m < n$. When the n-orbit class $\mathbf{O}^n$ contains an m-orbit class $\mathbf{O}^m$ then $\operatorname{cd}_\Gamma(\mathbf{O}^n)$ is the (n/m)-th power of $\operatorname{cd}_\Gamma(\mathbf{O}^m)$, because for $x \in \mathbf{O}^m$ the closed curve $\varphi_{(x,n)}$ is the closed curve $\varphi_{(x,m)}$ traced n/m times. This motivates the definition that the n-orbit class $\mathbf{O}^n$ is *reducible to period* m if $\operatorname{cd}_\Gamma(\mathbf{O}^n)$ has an (n/m)-th root, and that $\mathbf{O}^n$ is *irreducible* if $\operatorname{cd}_\Gamma(\mathbf{O}^n)$ is *primary* in the sense that it has no nontrivial root.

This notion of reducibility is consistent with that introduced in [**J1**]. An n-orbit class $\mathbf{O}^n$ is reducible to period m if and only if every n-point class $\mathbf{F}^n \subset \mathbf{O}^n$ is reducible to period m in the sense of [**J1**, Definition III.4.2].

1.4. Lefschetz numbers and n-orbit Nielsen numbers. Every n-orbit class $\mathbf{O}^n$ is an isolated subset of $\operatorname{Fix} f^n$. Its *index* is $\operatorname{ind}(\mathbf{O}^n, f^n)$, the index of $\mathbf{O}^n$ with respect to f^n. An n-orbit class $\mathbf{O}^n$ is called *essential* if its index is non-zero.

For each natural number n, we define the (generalized) *Lefschetz number* (with respect to Γ)

$$(1.4) \qquad L_\Gamma(f^n) := \sum_{\mathbf{O}^n} \operatorname{ind}(\mathbf{O}^n, f^n) \cdot \operatorname{cd}_\Gamma(\mathbf{O}^n) \quad \in \mathbb{Z}\Gamma_c,$$

the summation being over all n-orbit classes $\mathbf{O}^n$ of f. When every fixed point of f^n is isolated, we also have

$$(1.4') \qquad L_\Gamma(f^n) = \sum_{(x,n) \in \mathrm{PP}\, f} \operatorname{ind}(x, f^n) \cdot [\varphi_{(x,n)}] \quad \in \mathbb{Z}\Gamma_c.$$

Let $N_\Gamma(f^n)$ be the number of non-zero terms in $L_\Gamma(f^n)$. It is the number of essential n-orbit classes, and will be called the *Nielsen number of n-orbits.* Clearly it is a lower bound for the number of n-orbits of f.

Let $NI_\Gamma(f^n)$ be the number of non-zero primary terms in $L_\Gamma(f^n)$. It is the number of irreducible essential n-orbit classes, and will be called the *Nielsen number of irreducible n-orbits.* It is a lower bound for the number of primary n-orbits.

The indices of the essential n-orbit classes appear as the coefficients in $L_{\Gamma}(f^n)$. Another numerical invariant derived from $L_{\Gamma}(f^n)$ is its norm. We give a general definition here:

Notation. For any set S let $\mathbb{Z}S$ denote the free abelian group with the specified basis S. The *norm* in $\mathbb{Z}S$ is defined by

$$(1.5)\quad \left\|\sum_i k_i s_i\right\| := \sum_i |k_i| \in \mathbb{Z} \qquad \text{when the } s_i\text{'s in } S \text{ are all different.}$$

The norm $\|L_{\Gamma}(f^n)\|$ is the sum of absolute values of the indices of all the (essential) n-orbit classes. It equals $\|L_G(f^n)\|$, the sum of absolute values of the indices of all the (essential) n-point classes, because any two n-point classes contained in the same n-orbit class must have the same index. Hence $\|L_{\Gamma}(f^n)\| \geq N(f^n) \geq N_{\Gamma}(f^n)$.

Corresponding to the trace formula (1.2), we have the following trace formula:

Theorem 1.1. *Let $\widetilde{F}_d$ be the $\mathbb{Z}G$-matrices defined before* (1.2). *Then*

$$(1.6)\qquad L_{\Gamma}(f^n) = \sum_d (-1)^d \left[\operatorname{tr}(z\widetilde{F}_d)^n\right] \quad \in \mathbb{Z}\Gamma_c,$$

where $z\widetilde{F}_d$ is regarded as a $\mathbb{Z}\Gamma$-matrix.

Proof. Applying the theory of §1.1 to the iterates f^n of f, $n \geq 1$, we get

$$(1.7)\qquad L_G(f^n) := \sum_{\mathbf{F}^n} \operatorname{ind}(\mathbf{F}^n, f^n) \cdot \operatorname{cd}_G(\mathbf{F}^n, f^n) \quad \in \mathbb{Z}G_{f^n},$$

the summation being over all fixed point classes $\mathbf{F}^n$ of f^n. (The reference path for f^n is taken to be the path $w^{(n)} := w(f \circ w) \cdots (f^{n-1} \circ w)$ from v to $f^n(v)$.)

By definition, for $(x, n) \in \mathrm{PP}\, f$ and for any path c in X from v to x, the Γ-coordinate of (x, n) is the conjugacy class in Γ of the loop $c\varphi_{(x,n)}c^{-1} \sim \varphi_{(v,n)}(f^n \circ c)c^{-1} \sim z^n w^{(n)}(f^n \circ c)c^{-1}$. So

$$(1.8)\qquad \operatorname{cd}_{\Gamma}(x, n) = z^n \operatorname{cd}_G(x, f^n).$$

Now from (1.4) and (1.7) we see

$$(1.9)\qquad L_{\Gamma}(f^n) = z^n L_G(f^n).$$

On the other hand, the trace formula (1.2) gives

$$(1.10)\qquad L_G(f^n) = \sum_d (-1)^d \left[\operatorname{tr} \widetilde{F}_d^{(n)}\right] \quad \in \mathbb{Z}G_{f^n},$$

where $\widetilde{F}_d^{(n)}$ is the matrix of $\tilde{f}^n$. Since $\tilde{f} \circ g = f_G(g) \circ \tilde{f}$ for all $g \in G$, we have

$$\widetilde{F}_d^{(n)} = f_G^{n-1}\widetilde{F}_d \cdot f_G^{n-2}\widetilde{F}_d \cdots f_G\widetilde{F}_d \cdot \widetilde{F}_d. \tag{1.11}$$

Hence by (1.9–11) we obtain

$$\begin{aligned} L_\Gamma(f^n) &= \sum_d (-1)^d \left[\operatorname{tr}(z^n \widetilde{F}_d^{(n)})\right] \\ &= \sum_d (-1)^d \left[\operatorname{tr}(z\widetilde{F}_d)^n\right] \quad \in \mathbb{Z}\Gamma_c. \end{aligned}$$

□

Remark. Occasionally in applications we may use a homomorphism from Γ to a more convenient group Γ', which determines an obvious homomorphism $\mathbb{Z}\Gamma_c \to \mathbb{Z}\Gamma'_c$. Let $L_{\Gamma'}(f^n)$ be the image of $L_\Gamma(f^n)$. Let $N_{\Gamma'}(f^n)$ be the number of non-zero terms in $L_{\Gamma'}(f^n)$, and let $NI_{\Gamma'}(f^n)$ be the number of non-zero primary terms. Then $N_{\Gamma'}(f^n)$ etc. are lower bounds for $N_\Gamma(f^n)$ etc. respectively. This technique is similar to that for fixed points developed in [**J1**, §III.2].

1.5. Invariance properties. The following basic invariance properties are similar (with similar proofs) to that for fixed points (cf. [**J1**, §§I.4–5]).

Homotopy invariance. *Suppose $f \simeq f' : X \to X$ via a homotopy $\{f_t\}_{0\le t\le 1}$. The homotopy gives rise to a homotopy equivalence $T_f, v \simeq T_{f'}, v$ in a standard way. If we identify $\Gamma' = \pi_1(T_{f'}, v)$ with $\Gamma = \pi_1(T_f, v)$ via this homotopy equivalence, then $L_\Gamma(f'^n) = L_\Gamma(f^n)$ for all n, hence also $N_\Gamma(f'^n) = N_\Gamma(f^n)$ and $NI_\Gamma(f'^n) = NI_\Gamma(f^n)$.*

Commutativity. *Suppose $f : X \to Y$ and $g : Y \to X$. Then $T_{g\circ f}$ and $T_{f\circ g}$ are homotopy equivalent in a standard way. If we identify $\Gamma = \pi_1(T_{g\circ f})$ with $\Gamma' = \pi_1(T_{f\circ g})$ in this way, then $L_\Gamma((g \circ f)^n) = L_\Gamma((f \circ g)^n)$ for all n, hence also $N_\Gamma((g \circ f)^n) = N_\Gamma((f \circ g)^n)$ and $NI_\Gamma((g \circ f)^n) = NI_\Gamma((f \circ g)^n)$.*

Homotopy type invariance. *Suppose $h : X \to X'$ is a homotopy equivalence. Suppose $f : X \to X$ and $f' : X' \to X'$ are maps such that the diagram*

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ X' & \xrightarrow{f'} & X' \end{array}$$

commutes up to homotopy. Then $T_{f'}$ is homotopy equivalent to T_f, and when $\Gamma' = \pi_1(T_{f'})$ is suitably identified with $\Gamma = \pi_1(T_f)$, we have $L_\Gamma(f'^n) = L_\Gamma(f^n)$ for all n, hence also $N_\Gamma(f'^n) = N_\Gamma(f^n)$ and $NI_\Gamma(f'^n) = NI_\Gamma(f^n)$.

1.6. Twisted Lefschetz numbers and Lefschetz zeta function. Let R be a commutative ring with unity. Let $\mathrm{GL}_l(R)$ be the group of invertible $l \times l$ matrices in R, and $\mathcal{M}_{l\times l}(R)$ be the algebra of $l \times l$ matrices in R.

Suppose a representation $\rho : \Gamma \to \mathrm{GL}_l(R)$ is given. It extends to a representation $\rho : \mathbb{Z}\Gamma \to \mathcal{M}_{l\times l}(R)$. We define the *$\rho$-twisted Lefschetz number*

$$(1.12)\quad L_\rho(f^n) := \mathrm{tr}\,(L_\Gamma(f^n))^\rho = \sum_{\mathbf{O}^n} \mathrm{ind}(\mathbf{O}^n, f^n)\cdot \mathrm{tr}\,(\mathrm{cd}_\Gamma(\mathbf{O}^n))^\rho \quad \in R$$

for every $n \in \mathbb{N}$, the summation being over all n-orbit classes $\mathbf{O}^n$ of f. It is well defined because matrices in a conjugacy class have the same trace. When all fixed points of f^n are isolated, we have

$$(1.12')\qquad L_\rho(f^n) = \sum_{(x,n)\in \mathrm{PP}\, f} \mathrm{ind}(x, f^n)\cdot \mathrm{tr}\,(\varphi_{(x,n)})^\rho \quad \in R.$$

It has the trace formula

$$\begin{aligned} L_\rho(f^n) &= \sum_d (-1)^d\, \mathrm{tr}\left(\mathrm{tr}(z\widetilde{F}_d)^n\right)^\rho \\ (1.13)\qquad &= \sum_d (-1)^d\, \mathrm{tr}\left((z\widetilde{F}_d)^\rho\right)^n \quad \in R \end{aligned}$$

where for a $\mathbb{Z}\Gamma$-matrix A, its ρ-image A^ρ means the block matrix obtained from A by replacing each element a_{ij} with the $l\times l$ R-matrix a_{ij}^ρ.

We now define the (ρ-twisted) *Lefschetz zeta function* of f to be the formal power series

$$(1.14)\qquad \zeta_\rho(f) := \exp \sum_n L_\rho(f^n)\frac{t^n}{n}.$$

It has constant term 1, so it is in the multiplicative subgroup $1 + tR[[t]]$ of the formal power series ring $R[[t]]$.

Clearly $\zeta_\rho(f)$ enjoys the same invariance properties as that of $L_\Gamma(f^n)$. As to its computation, we obtain from (1.13) the following determinant formula:

Theorem 1.2. *$\zeta_\rho(f)$ is a rational function in R.*

$$(1.15)\qquad \zeta_\rho(f) = \prod_d \det\left(I - t(z\widetilde{F}_d)^\rho\right)^{(-1)^{d+1}} \quad \in R(t),$$

where I stands for suitable identity matrices.

Proof.

$$\zeta_\rho(f) = \exp \sum_d (-1)^d \sum_n \mathrm{tr}\left((z\widetilde{F}_d)^\rho\right)^n \frac{t^n}{n}$$

$$= \prod_d \det\left(I - t(z\widetilde{F}_d)^\rho\right)^{(-1)^{d+1}}.$$

□

By (1.12), (1.14) and the homotopy invariance, we have the

Twisted version of the Lefschetz fixed point theorem. *Let $f : X \to X$ be a map and $\rho : \pi_1(T_f) \to \mathrm{GL}_l(R)$ be a representation. If f is homotopic to a fixed point free map g, then $L_\rho(f) = 0$. If f is homotopic to a periodic point free map g, then $\zeta_\rho(f) = 1$.*

Remark 1. When $R = \mathbb{Q}$ and $\rho : \Gamma \to \mathrm{GL}_1(\mathbb{Q}) = \mathbb{Q}$ is trivial (sending everything to 1), then $L_\rho(f) \in \mathbb{Z}$ is the ordinary Lefschetz number $L(f)$, and $\zeta_\rho(f)$ is the classical Lefschetz zeta function $\zeta(f) := \exp \sum_n L(f^n)t^n/n$ introduced by Weil (cf. [**Bt**]).

Remark 2. Our Lefschetz zeta function is essentially the same as the twisted Lefschetz function of David Fried. He first introduced it in [**F1**] using f-invariant abelianizations of $\pi_1(X)$, and showed in [**F2**] that it is a certain Reidemeister torsion of the mapping torus T_f. Then in the paper [**F4**] he adopted the Reidemeister torsion approach, with respect to a flat vector bundle (which is equivalent to a matrix representation of the fundamental group).

Example. (Recipe for surfaces with boundary).

Let X be a surface with boundary, and $f : X \to X$ be a map. Suppose $\{a_1, \cdots, a_r\}$ is a free basis for $G = \pi_1(X)$. Then X has the homotopy type of a bouquet X' of r circles which can be decomposed into one 0-cell and r 1-cells corresponding to the a_i's, and f has the homotopy type of a cellular map $f' : X' \to X'$. By the homotopy type invariance of the invariants, we can replace f with f' in computations. The homomorphism $f_G : G \to G$ induced by f and f' is determined by the images $a'_i := f_G(a_i)$, $i = 1, \cdots, r$. By (1.3), the fundamental group $\Gamma = \pi_1(T_f)$ has a presentation

$$\Gamma = \langle a_1, \cdots, a_r, z \mid a_i z = z a'_i,\ i = 1, \cdots, r \rangle. \tag{1.16}$$

As pointed out in [**FH**], the matrices of the lifted chain map $\tilde{f}'$ are

$$\widetilde{F}_0 = (1), \tag{1.17}$$

$$\widetilde{F}_1 = D := \left(\frac{\partial a'_i}{\partial a_j}\right),$$

where D is the Jacobian matrix in Fox calculus (see [**Bi**, §3.1] for an introduction). Then, by (1.6), in $\mathbb{Z}\Gamma_c$ we have

$$L_\Gamma(f) = [z] - \sum_{i=1}^{r}\left[z\frac{\partial a'_i}{\partial a_i}\right], \tag{1.18}$$

$$L_{\Gamma}(f^n) = [z^n] - [\mathrm{tr}(zD)^n]. \tag{1.19}$$

When a representation $\rho : \Gamma \to \mathrm{GL}_l(R)$ is given, by (1.13) and (1.15) we have

$$L_\rho(f) = \mathrm{tr}\, z^\rho - \mathrm{tr}(zD)^\rho \quad \in R, \tag{1.20}$$

$$\zeta_\rho(f) = \frac{\det(I - t(zD)^\rho)}{\det(I - tz^\rho)} \quad \in R(t). \tag{1.21}$$

1.7. A closer look at the representation ρ. A practical difficulty in the use of $L_\rho(f^n)$ and $\zeta_\rho(f)$ is to find a useful ρ which was assumed to be a group homomorphism $\Gamma \to \mathrm{GL}_l(R)$. Can we weaken the assumption on ρ?

Observe from (1.8) that the Γ-coordinate of an n-orbit class can be written as $z^n g$ for some $g \in G$, whereas a general element of Γ has the form $z^k g z^{-l}$ with $g \in G$ and $k, l \geq 0$. The definition (1.12) only requires that $\mathrm{tr}(\mathrm{cd}_\Gamma(\mathbf{O}^n))^\rho \in R$ be well defined, so ρ need to behave well only on a subset of Γ, not on the whole Γ. This motivates the following approach.

Definition. Let Γ_+ be the *monoid* defined by the presentation (1.3)

$$\Gamma_+ := \mathrm{Monoid}\langle G, z \mid gz = z f_G(g) \text{ for all } g \in G \rangle. \tag{1.22}$$

In other words, as a set,

$$\Gamma_+ = \{ z^n g \mid n \geq 0, g \in G \}. \tag{1.23}$$

The letter z is regarded as a symbol so that Γ_+ is in one-one correspondence with $\mathbb{Z}_+ \times G$, where $\mathbb{Z}_+$ is the monoid of non-negative integers. And the multiplication in Γ_+ is defined by

$$(z^n a)(z^m b) := z^{n+m}\left(f_G(a) b\right). \tag{1.24}$$

The obvious projection $\eta : \Gamma_+ \to \Gamma$, $z^n g \mapsto z^n g$ is a monoid homomorphism which will often be omitted in notations. Beware that η is not necessarily monomorphic.

Lemma 1.3. *Suppose $z^n a, z^n b \in \Gamma_+$ project to conjugate elements in Γ, where $n > 0$. Then there exist $0 \leq r < n$ and $h \in G$ such that in Γ_+ we have*

$$z^n b = h^{-1} z^{n-r} a z^r h. \tag{1.25}$$

Proof. Suppose $z^n b = \gamma^{-1} z^n a \gamma$ for some $\gamma \in \Gamma$. This γ can be written in the form $\gamma = z^k c z^{-l}$ with $c \in G$ and $k, l \geq 0$. So in Γ we have

$$z^n f_G^l(b) = z^n(z^{-l} b z^l) = z^{-l}(z^n b) z^l = z^{-l}(\gamma^{-1} z^n a \gamma) z^l$$

$$= (c^{-1}z^{-k})z^n a(z^k c) = (c^{-1}z^n)(z^{-k}az^k)c = z^n f_G^n(c^{-1})f_G^k(a)c,$$

hence $f_G^l(b) = f_G^n(c^{-1})f_G^k(a)c$.

By the Remark in §1.2, we can find some $m > 0$ such that

$$f_G^{l+m}(b) = f_G^{n+m}(c^{-1})f_G^{k+m}(a)f_G^m(c) \quad \in G.$$

Increasing m if necessary, we may assume $l + m = pm$ and $k + m = qm + r$, where $p, q > 0$ and $0 \leq r < n$.

Let $h = f_G^r(a^{-1})f_G^{n+r}(a^{-1})\cdots f_G^{(q-1)n+r}(a^{-1})f_G^m(c)f_G^{(p-1)n}(b)\cdots f_G^n(b)b \in G$. Then we have

$$\begin{aligned}
&f_G^n(h^{-1})f_G^r(a)h \\
&\quad = [f_G^n(b^{-1})\cdots f_G^{pn}(b^{-1})f_G^{n+m}(c^{-1})f_G^{qn+r}(a)\cdots f_G^{n+r}(a)]\cdot f_G^r(a) \\
&\quad\quad \cdot \left[f_G^r(a^{-1})f_G^{n+r}(a^{-1})\cdots f_G^{(q-1)n+r}(a^{-1})f_G^m(c)f_G^{(p-1)n}(b)\cdots f_G^n(b)b\right] \\
&\quad = f_G^n(b^{-1})\cdots f_G^{pn}(b^{-1})f_G^{n+m}(c^{-1})f_G^{k+m}(a)f_G^m(c)f_G^{(p-1)n}(b)\cdots f_G^n(b)b \\
&\quad = f_G^n(b^{-1})\cdots f_G^{pn}(b^{-1})f_G^{l+m}(b)f_G^{(p-1)n}(b)\cdots f_G^n(b)b = b.
\end{aligned}$$

Thus, in Γ_+ we get

$$\begin{aligned}
z^n b = z^n f_G^n(h^{-1})f_G^r(a)h &= h^{-1}z^n f_G^r(a)h \\
&= h^{-1}z^{n-r}z^r f_G^r(a)h = h^{-1}z^{n-r}az^r h,
\end{aligned}$$

as required. □

Theorem 1.4. *Suppose a monoid representation $\rho : \Gamma_+ \to \mathcal{M}_{l\times l}(R)$ is given. In other words, suppose we have a group representation $\rho : G \to \mathrm{GL}_l(R)$ and a matrix $z^\rho \in \mathcal{M}_{l\times l}(R)$ satisfying the condition*

$$g^\rho z^\rho = z^\rho(f_G(g))^\rho \qquad \text{for any } g \in G. \tag{1.26}$$

Extend ρ to a ring homomorphism $\rho : \mathbb{Z}\Gamma_+ \to \mathcal{M}_{l\times l}(R)$. Then the theory of §1.6 *works.*

Proof. The basis of §1.6 is the definition (1.12) of $L_\rho(f^n)$. So it suffices to show that for any $z^n a, z^n b \in \Gamma_+$ that are conjugate in Γ, we have $\mathrm{tr}(z^n a)^\rho = \mathrm{tr}(z^n b)^\rho$.

Let r and h be as in Lemma 1.3, and write a^ρ, b^ρ, h^ρ as A, B, H respectively. Then

$$\mathrm{tr}(z^n b)^\rho = \mathrm{tr}(h^{-1}z^{n-r}az^r h)^\rho = \mathrm{tr}(H^{-1}Z^{n-r}AZ^r H)$$

$$= \operatorname{tr}(Z^{n-r}AZ^r) = \operatorname{tr}(Z^nA) = \operatorname{tr}(z^na)^\rho.$$

□

Remark. If z^ρ is invertible, ρ will give a group representation $\Gamma \to \mathrm{GL}_l(R)$. The point of the theorem is that we do not require z^ρ to be an invertible matrix.

Example. (Free abelian group).

Suppose G is a (multiplicative) free abelian group with basis $\{g_1, \cdots, g_r\}$, and the homomorphism $f_G : G \to G$ is given by the $r \times r$ integral matrix $A = (a_{ij})$ such that $f_G(g_i) = g_1^{a_{i1}} \cdots g_r^{a_{ir}}$.

Every element $g = g_1^{v_1} \cdots g_r^{v_r} \in G$ corresponds to an integer row-vector $v(g) := (v_1, \cdots, v_r)$. Clearly $v(f_G(g)) = v(g) \cdot A$ for any $g \in G$.

Then the assignments

$$g^\rho = \begin{pmatrix} 1 & v(g) \\ 0 & I \end{pmatrix}, \qquad z^\rho = \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix} \tag{1.27}$$

define a monoid representation $\rho : \Gamma_+ \to \mathcal{M}_{(r+1)\times(r+1)}(\mathbb{Z})$. The verification of the condition (1.26) is trivial.

1.8. Relative invariants mod a subpolyhedron. Let X be a compact connected polyhedron as before, and A be a subpolyhedron. Let $f : X, A \to X, A$ be a self-map of the pair.

A fixed point x of f is *related to* A if there is a path c such that $c \simeq f \circ c : I, 0, 1 \to X, x, A$, where $\simeq$ means homotopic. A fixed point class $\mathbf{F}$ of f will be called a *fixed point class on* $X \setminus A$ if it is not related to A. The number of essential fixed point classes of f on $X \setminus A$ is called the *Nielsen number of the complement*, denoted $N(f; X \setminus A)$. It is a lower bound for the number of fixed points of f on $X \setminus A$, and it is invariant under homotopy of maps $X, A \to X, A$ ([**Z**], cf. [**S**, §2.3]). Obviously $N(f; X \setminus A) \leq N(f)$.

Under the mapping torus point of view, a fixed point x of f is related to A if and only if the corresponding closed orbit curve $\varphi_{(x,1)}$ in T_f is freely homotopic to a closed curve in $T_{f|A}$, the mapping torus of the restriction $f|A : A \to A$ naturally regarded as a subspace of T_f.

The Nielsen theory of periodic orbits for X developed above has a natural relative version for $X \setminus A$. A free homotopy class of closed curves in T_f (i.e. an element of Γ_c) will be called *related to* A if it contains a closed curve in $T_{f|A} \subset T_f$. An *n-orbit class of* f *on* $X \setminus A$ is defined to be an n-orbit class of f whose coordinate is not related to A. The *Nielsen number of the complement* $N_\Gamma(f^n; X \setminus A)$ is the number of essential n-orbit classes of f on $X \setminus A$. The *Lefschetz number of the complement* $L_\Gamma(f^n; X \setminus A) \in \mathbb{Z}\Gamma_c$ is obtained from $L_\Gamma(f^n)$ by deleting the terms related to A. Clearly $\|L_\Gamma(f^n; X \setminus A)\| \leq \|L_\Gamma(f^n)\|$.

2. Asymptotic Nielsen numbers and topological entropy.

The asymptotic behavior of the number of periodic orbits is more important than that number for a specific period n. The former is also often easier to estimate. In §2.1 several asymptotic invariants are defined as growth rates of the Nielsen numbers and Lefschetz numbers. Sufficient conditions for these invariants to be equal are given in §2.2. In §2.3 we propose our method of lower estimation for the asymptotic absolute Lefschetz number via twisted Lefschetz zeta functions. §2.4 provides a method of upper estimation. §2.5 is devoted to the relation between the asymptotic Nielsen number and the topological entropy. The final section §2.6 is an aside discussing the growth rates of some Nielsen type numbers.

2.1. Asymptotic invariants. The *growth rate* of a sequence $\{a_n\}$ of complex numbers is defined by

$$\mathrm{Growth}_{n\to\infty} a_n := \max\left\{1,\ \limsup_{n\to\infty} |a_n|^{1/n}\right\} \tag{2.1}$$

which could be infinity. Note that $\mathrm{Growth}a_n \geq 1$ even if all $a_n = 0$. When $\mathrm{Growth}a_n > 1$, we say that the sequence *grows exponentially.*

We define the *asymptotic Nielsen number* of f to be the growth rate of the Nielsen numbers

$$N^\infty(f) := \mathrm{Growth}_{n\to\infty} N(f^n) = \mathrm{Growth}_{n\to\infty} N_\Gamma(f^n), \tag{2.2}$$

where the second equality is due to the obvious inequality $N_\Gamma(f^n) \leq N(f^n) \leq n \cdot N_\Gamma(f^n)$. And we define the *asymptotic irreducible Nielsen number* of f to be the growth rate of the irreducible Nielsen numbers

$$NI^\infty(f) := \mathrm{Growth}_{n\to\infty} NI_\Gamma(f^n). \tag{2.3}$$

We also define the *asymptotic absolute Lefschetz number*

$$L^\infty(f) := \mathrm{Growth}_{n\to\infty} \|L_\Gamma(f^n)\|. \tag{2.4}$$

All these asymptotic numbers enjoy the invariance properties of §1.5.

The following proposition ensures that these asymptotic invariants are finite positive numbers.

Proposition 2.1.

$$NI^\infty(f) \leq N^\infty(f) \leq L^\infty(f) < \infty. \tag{2.5}$$

Proof. The first two inequalities are from the obvious fact

$$NI_\Gamma(f^n) \leq N_\Gamma(f^n) \leq \|L_\Gamma(f^n)\|.$$

The last one is by Proposition 2.6 below. □

Remark. These asymptotic invariants have obvious generalizations to the relative setting of §1.8.

2.2. Conditions for the equalities $NI^\infty(f) = N^\infty(f) = L^\infty(f)$. To compare $NI^\infty(f)$ with $N^\infty(f)$, we need the following definition.

Definition. An n-orbit class $\mathbf{O}^n$ and all n-point classes contained in it will be called *essentially irreducible* if it is essential and it does not contain any essential m-orbit class for any $m < n$.

Clearly every irreducible essential n-orbit class is essentially irreducible, but not vice versa.

Theorem 2.2. *A sufficient condition for the equality $NI^\infty(f) = N^\infty(f)$ is that f has the following Property of Essential Irreducibility:*

(EI) The number E_n of essentially irreducible n-point classes that are reducible is uniformly bounded in n.

Proof. The case $N^\infty(f) = 1$ is trivial. We assume $N^\infty(f) = 1 + a > 1$. Let E be a bound for E_n.

Let $S_n := \sum_{m \le n} N_{\mathrm{r}}(f^m)$. Then by [**FLP**, p. 185, Lemma 1], $\mathrm{Growth}_{n\to\infty} S_n = N^\infty(f) = 1 + a$, hence $S_n \le (1 + \frac{5}{4}a)^n$ for sufficiently large n.

We have

$$\begin{aligned} NI_{\mathrm{r}}(f^n) &\ge N_{\mathrm{r}}(f^n) - E_n - \sum_{\substack{m|n \\ m<n}} N_{\mathrm{r}}(f^m) \ge N_{\mathrm{r}}(f^n) - E - S_{n/2} \\ &\ge N_{\mathrm{r}}(f^n)\left(1 - \frac{E + S_{n/2}}{N_{\mathrm{r}}(f^n)}\right). \end{aligned}$$

Pick a subsequence $\{n_j\}$ such that $\lim_{j\to\infty} N_{\mathrm{r}}(f^{n_j})^{1/n_j} = N^\infty(f)$, so that $N_{\mathrm{r}}(f^{n_j}) \ge (1 + \frac{3}{4}a)^{n_j}$ for sufficiently large j. Then $S_{n_j/2}/N_{\mathrm{r}}(f^{n_j}) \le (1 + \frac{1}{4}a)^{-n_j/2}$, so the quantity in the big parentheses approaches 1 when $j \to \infty$. Hence the conclusion. □

Theorem 2.3. *A sufficient condition for the equality $N^\infty(f) = L^\infty(f)$ is that f has the following Property of Bounded Index:*

(BI) The maximum absolute value B_n of the indices of n-point classes $\mathbf{F}^n$ is uniformly bounded in n.

Proof. Suppose B is a bound for B_n. Then $\|L_\Gamma(f^n)\| = \sum_{\mathbf{F}^n} \|\,\mathrm{ind}(\mathbf{F}^n, f^n)\,\| \leq BN(f^n)$. Hence $L^\infty(f) \leq N^\infty(f)$. □

Note that both properties (EI) and (BI) are invariant under homotopy. Both are satisfied in many important cases, e.g. when X is a torus of any dimension, or when f is a homeomorphism of a surface X with $\chi(X) < 0$.

2.3. Lower estimation of $L^\infty(f)$ via $\zeta_\rho(f)$.

Proposition 2.4. *Let $R = \mathbb{C}$ and let $\rho : \Gamma_+ \to \mathcal{M}_{l\times l}(\mathbb{C})$ be a monoid representation. Suppose $\{\mu_n\}$ is a sequence such that for every $n \in \mathbb{N}$,*

$$|\,\mathrm{tr}(z^n g)^\rho| \leq \mu_n \qquad \textit{for all } g \in G, \tag{2.6}$$

and $\mu := \mathrm{Growth}_{n\to\infty}\mu_n$. Let w be a zero or a pole of the rational function $\zeta_\rho(f) \in \mathbb{C}(t)$. Then

$$L^\infty(f) \geq \frac{1}{\mu|w|}. \tag{2.7}$$

Proof. Note that $w \neq 0$ because $\zeta_\rho(f)(0) = 1$. We know from complex analysis and (1.14) that $\mathrm{Growth}L_\rho(f^n)$ is the reciprocal of the radius of convergence of the function $\log\zeta_\rho(f)$, hence $\mathrm{Growth}L_\rho(f^n) \geq 1/|w|$.

On the other hand, according to §1.6, the Γ-coordinates of n-orbit classes are in the form $[z^n g]$ with $g \in G$. So we can assume $L_\Gamma(f^n) = \sum_i k_i[z^n g_i]$, where the $[z^n g_i]$'s are different conjugacy classes in Γ. Then $L_\rho(f^n) \in \mathbb{C}$ are bounded by $|L_\rho(f^n)| = |\sum_i k_i\,\mathrm{tr}(z^n g_i)^\rho| \leq \sum_i |k_i| \cdot |\,\mathrm{tr}(z^n g_i)^\rho| \leq \mu_n \sum_i |k_i| = \mu_n\|L_\Gamma(f^n)\|$. Hence $\mathrm{Growth}L_\rho(f^n) \leq \mu \cdot L^\infty(f)$.

So we get the formula (2.7). □

Example 1. For homomorphisms of free abelian groups, the representation ρ defined by (1.27) satisfies the assumption of Proposition 2.4 where μ equals the spectral radius of the matrix A. More precisely,

$$\mu = \max\{1, |\lambda_1|, \cdots, |\lambda_r|\}, \tag{2.8}$$

where $\lambda_1, \cdots, \lambda_r$ are the eigenvalues of A.

Example 2. (Maps of the circle).

Let $f : S^1 \to S^1$ be a self-map of the circle and let $d \in \mathbb{Z}$ be its degree. The fundamental group $G = \pi_1(S^1)$ is the infinite cyclic group generated by a, and the homomorphism induced by f is $f_G : G \to G$, $a \mapsto a^d$. By (1.16), the fundamental group $\Gamma = \pi_1(T_f)$ has a presentation $\Gamma = \langle a, z \mid az = za^d\rangle$.

According to (1.27) with $r = 1$, a representation ρ of Γ_+ into $\mathcal{M}_{2\times 2}(\mathbb{Z})$ is defined by specifying

$$a^\rho = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad z^\rho = \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix}. \tag{2.9}$$

An easy computation shows that for any $d \in \mathbb{Z}$, we have

(2.10)
$$D^\rho = \left(\frac{\partial a^d}{\partial a}\right)^\rho = \begin{pmatrix} d & d(d-1)/2 \\ 0 & d \end{pmatrix}, \qquad (zD)^\rho = \begin{pmatrix} d & d(d-1)/2 \\ 0 & d^2 \end{pmatrix}.$$

Thus

$$\zeta_\rho(f) = \frac{(1-dt)(1-d^2t)}{(1-t)(1-dt)} = \frac{1-d^2t}{1-t}. \tag{2.11}$$

Hence, by (2.7) and (2.8), we get

$$L^\infty(f) \geq \max\{1, |d|\}. \tag{2.12}$$

Since the trace of a unitary matrix is bounded by its dimension, we get the very useful

Corollary 2.5. *Suppose $\rho : \Gamma \to \mathrm{U}(l)$ is a unitary representation. Let w be a zero or a pole of the rational function $\zeta_\rho(f)$. Then*

$$L^\infty(f) \geq \frac{1}{|w|}. \tag{2.13}$$

2.4. Upper estimation of $L^\infty(f)$. In practice, the initial data of our lower estimation in the last section is the knowledge of the $\mathbb{Z}G$-matrices $\{\widetilde{F}_d\}$ provided by a cellular map, which enables us to compute the Lefschetz zeta function. There is also a simple way to derive an upper bound from the same data.

We first extend the notation (1.5).

Notation. For a matrix $A = (a_{ij})$ in $\mathbb{Z}S$, its *matrix of norms* is defined to be the matrix

$$\|A\| := (\|a_{ij}\|) \tag{2.14}$$

which is a matrix of non-negative integers. (In what follows, the set S will be G or Γ.)

Proposition 2.6. *Let $\Phi_d := \|\widetilde{F}_d\|$ for every dimension d. Then*

$$L^{\infty}(f) \le \max_d\{\text{spectral radius of } \Phi_d\}. \tag{2.15}$$

Proof. By (1.6) and the definitions,

$$\begin{aligned}\|L_{\mathrm{r}}(f^n)\| &\le \sum_d \|[\operatorname{tr}(z\widetilde{F}_d)^n]\| \le \sum_d \|\operatorname{tr}(z\widetilde{F}_d)^n\| \le \sum_d \operatorname{tr}\|(z\widetilde{F}_d)^n\| \\ &\le \sum_d \operatorname{tr}\|z\widetilde{F}_d\|^n \le \sum_d \operatorname{tr}\|\widetilde{F}_d\|^n = \sum_d \operatorname{tr}\Phi_d^n.\end{aligned}$$

Hence

$$\begin{aligned}L^{\infty}(f) &= \operatorname{Growth}_{n\to\infty}\|L_{\mathrm{r}}(f^n)\| \\ &\le \operatorname{Growth}_{n\to\infty}\sum_d \operatorname{tr}\Phi_d^n \\ &= \max_d\{\operatorname{Growth}_{n\to\infty}\operatorname{tr}\Phi_d^n\} \\ &= \max_d\{\text{spectral radius of } \Phi_d\}.\end{aligned}$$

□

The use of this Proposition will be illustrated by the examples in §4.

2.5. Topological entropy. The most widely used measure for the complexity of a dynamical system is the topological entropy. (See [**Wa**] for an introduction.) For the convenience of the reader, we include its definition.

Let $f : X \to X$ be a self-map of a compact metric space. For given $\epsilon > 0$ and $n \in \mathbb{N}$, a subset $E \subset X$ is said to be *(n,ϵ)-separated* under f if for each pair $x \neq y$ in E there is $0 \le i < n$ such that $d(f^i(x), f^i(y)) > \epsilon$. Let $s_n(\epsilon, f)$ denote the largest cardinality of any (n,ϵ)-separated subset E under f. Thus $s_n(\epsilon, f)$ is the greatest number of orbit segments $\{x, f(x), \dots, f^{n-1}(x)\}$ of length n that can be distinguished one from another provided we can only distinguish between points of X that are at least ϵ apart. Now let

$$\begin{aligned}h(f,\epsilon) &:= \limsup_{n\to\infty}\frac{1}{n}\log s_n(\epsilon, f), \\ h(f) &:= \lim_{\epsilon\to 0} h(f,\epsilon).\end{aligned} \tag{2.16}$$

The number $0 \le h(f) \le \infty$, which is easily seen to be independent of the metric d used, is called the *topological entropy of f*.

If $h(f,\epsilon) > 0$ then, up to resolution $\epsilon > 0$, the number $s_n(\epsilon, f)$ of distinguishable orbit segments of length n grows exponentially with n. So $h(f)$

measures the growth rate in n of the number of orbit segments of length n with arbitrarily fine resolution.

A basic relation between periodic points and topological entropy is proved by Ivanov [**I**]. We present a different proof.

Theorem 2.7. *Let $f : X \to X$ be a self-map of a compact connected polyhedron. Then*

$$h(f) \geq \log N^{\infty}(f). \tag{2.17}$$

Proof. Let $\delta > 0$ be such that every loop in X of diameter $< 2\delta$ is contractible. Let $\epsilon > 0$ be a smaller number such that $d(f(x), f(y)) < \delta$ whenever $d(x, y) < 2\epsilon$. Let $E_n \subset X$ be a set consisting of one point from each essential n-point class. Thus $|E_n| = N(f^n)$. By the definition of $h(f)$, it suffices to show that E_n is (n, ϵ)-separated.

Suppose it is not so. Then there would be two n-points $x \neq y \in E_n$ such that $d(f^i(x), f^i(y)) \leq \epsilon$ for $0 \leq i < n$ hence for all $i \geq 0$. Pick a path c_i from $f^i(x)$ to $f^i(y)$ of diameter $< 2\epsilon$ for $0 \leq i < n$ and let $c_n = c_0$. By the choice of δ and ϵ, $f \circ c_i \simeq c_{i+1}$ for all i, so $f^n \circ c_0 \simeq c_n = c_0$. This means x, y in the same n-point class, contradicting the construction of E_n. □

Theorem 2.7 is remarkable in that it does not require smoothness of the map and it provides a common lower bound for the topological entropy of all maps in a homotopy class.

Example 1. (Linear maps on tori).

Let $T^k := \mathbb{R}^k/\mathbb{Z}^k$ be the k-dimensional torus. Let f be an automorphism of T^k defined by an integer matrix A. Then

$$h(f) = \sum_{|\lambda_i|>1} \log |\lambda_i| \tag{2.18}$$

where $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A. (Cf. [**Wa**, p. 203] or [**B1**, Corollary 16].) Note that $\prod_{|\lambda_i|>1} |\lambda_i|$ is the spectral radius of the endomorphism f^* induces by f on the cohomology ring $H^*(T^k, \mathbb{R})$ which is the exterior algebra of the linear space $H^1(T^k, \mathbb{R})$. So, according to [**MP**], it is a lower bound for the entropy of all continuous maps homotopic to f. Thus f has the minimal entropy in its homotopy class.

On the other hand, $N(f^n) = |L(f^n)|$ and $L(f^n) = \det(I - A^n) = \prod_i(1 - \lambda_i^n)$, so that

$$N^{\infty}(f) = \operatorname{Growth}_{n\to\infty} \prod_i |1 - \lambda_i^n| = \begin{cases} 1 & \text{if } L(f) = 0, \\ \prod_{|\lambda_i|>1} |\lambda_i| & \text{otherwise.} \end{cases} \tag{2.19}$$

Observe that f has both Properties (EI) and (BI), so $N^{\infty}(f) = NI^{\infty}(f) = L^{\infty}(f)$. Hence, if $L(f) \neq 0$ then $h(f) = \log L^{\infty}(f) = \log N^{\infty}(f) = \log NI^{\infty}(f)$. If $L(f) = 0$, then all $L(f^n) = N(f^n) = 0$ and $\log L^{\infty}(f) = \log N^{\infty}(f) = \log NI^{\infty}(f) = 0$, but $h(f)$ may be positive.

Example 2. (Pseudo-Anosov maps).

Let X be a compact surface with $\chi(X) < 0$. Let f be a pseudo-Anosov homeomorphism with stretching factor $\lambda > 1$. Then

$$h(f) = \log \lambda = \log N^{\infty}(f) \tag{2.20}$$

is the minimal entropy in the homotopy class of f ([**FLP**, p. 194] and [**I**]).

2.6. Growth rate of Nielsen type numbers. In Nielsen theory for periodic points, it is well known that $N(f^n)$ is often very poor as a lower bound for the number of fixed points of f^n. A good homotopy invariant lower bound $NF_n(f)$, called the Nielsen type number for n-th iterate, is defined in [**J1**, Definition III.4.8]. Consider any finite set of periodic orbit classes $\{\mathbf{O}^{k_j}\}$ (of varied periods k_j) such that every essential periodic m-orbit class, $m \mid n$, contains at least one class in the set. Then $NF_n(f)$ is the minimal sum $\sum_j k_j$ for all such finite sets.

The definition of $NF_n(f)$ is rather complicated, and if we count periodic orbits instead of periodic points, a good lower bound can be defined in a simpler way. The *Nielsen type number for n-orbits* $NO_n(f)$ is defined to be the total number of essentially irreducible m-orbit classes for all $m \mid n$.

When we count primary n-points, a good lower bound $NP_n(f)$, called the Nielsen type number of least period n, is defined in [**J1**] to be n times the number of irreducible essential n-orbit classes. If we count primary n-orbits, the good bound should be $NI_{\Gamma}(f^n)$ defined in §1.4.

The following proposition says that as far as asymptotic growth rate is concerned, these Nielsen type numbers are no better than the Nielsen numbers.

Proposition 2.8. *For any map* $f : X \to X$,

$$\mathrm{Growth}_{n\to\infty} NF_n(f) = \mathrm{Growth}_{n\to\infty} NO_n(f) = N^{\infty}(f), \tag{2.21}$$

$$\mathrm{Growth}_{n\to\infty} NP_n(f) = NI^{\infty}(f). \tag{2.22}$$

Proof. Let S_n be as in the proof of Theorem 2.2. By the definition of $NF_n(f)$ we see $N_{\Gamma}(f^n) \leq N(f^n) \leq NF_n(f) \leq \sum_{m|n} N(f^m) \leq \sum_{m|n} m \cdot N_{\Gamma}(f^m) \leq nS_n$. Similarly for $NO_n(f)$ we have $N_{\Gamma}(f^n) \leq NO_n(f) \leq \sum_{m|n} N_{\Gamma}(f^m) \leq S_n$. Hence the formula (2.21). The second formula follows from the equality $NP_n(f) = n \cdot NI_{\Gamma}(f^n)$. $\square$

3. Periodic orbit classes of surface homeomorphisms.

The results of §§2.2–3 provide us with a method of asymptotic estimation for maps $f: X \to X$ that have both Properties (EI) and (BI). In §3.1 we show that self-homeomorphisms of aspherical surfaces have both these Properties and the asymptotic Nielsen number coincides with the largest stretching factor in the Thurston canonical form. §3.2 is devoted to the development of a Nielsen theory for self-homeomorphisms of punctured surfaces which is very useful in applications.

3.1. Compact aspherical surfaces. Let X be a compact connected aspherical surface and let $f: X \to X$ be a homeomorphism. The main result 3.7 of this section is easy when X is the disc, the annulus, the Möbius strip, the torus or the Klein bottle. So we shall assume $\chi(X) < 0$.

Thurston Theorem ([**T**]). *Every homeomorphism $f: X \to X$ is isotopic to a homeomorphism φ such that either*

(1) *φ is a periodic map, i.e. $\varphi^m = id$ for some m; or*

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

(3) *φ is a reducible map, i.e. there is a system of disjoint simple closed curves $\gamma = \{\gamma_1, \cdots, \gamma_k\}$ in $\operatorname{int} X$ such that γ is invariant by φ (but the γ_i's may be permuted) and γ has a φ-invariant tubular neighborhood U such that each component of $X \setminus U$ has negative Euler characteristic and on each (not necessarily connected) φ-component of $X \setminus U$, φ satisfies (1) or (2).*

The φ above is called the Thurston canonical form of f. In (3) it can be chosen so that some iterate φ^m is a generalized Dehn twist on U. Such a φ, as well as the φ in (1) or (2), will be called *standard* [**JG**, §3.1]. A key observation is that if φ is standard, so are all iterates of φ.

For the convenience of the reader, we list the following information about the fixed point classes of a standard φ (cf. [**JG**, Lemmas 3.6 and 3.4]). The superscripts '+' and '−' indicate that φ preserves or reverses the local orientation at the fixed point class.

Lemma 3.1. *Every fixed point class of a standard φ is connected. The possible types of fixed point classes are listed below, with a description of their local behavior.*

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

- $(a)^{+}$ *$x \in \operatorname{int} M$, φ is conjugate to a rotation in a neighborhood of x;* $\operatorname{ind}(x, \varphi) = 1$.
- $(b)^{+}$ *$x \in \operatorname{int} M$ is a fixed point of an annular flip-twist;* $\operatorname{ind}(x, \varphi) = 1$.

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

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

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

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

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

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

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

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

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

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

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

The possible values of $\operatorname{ind}(A,\varphi)$ *are* $1, -1$ *or* 0.

(4)$^+$ *Fixed subsurface B of M with $\chi(B) \le 0$. The possible forms for a component C of ∂B:*

(a) *$C \subset \operatorname{int} M$; on the outside of B, C is a type $(p_C, 0)^+$ invariant boundary component of some pseudo-Anosov piece;*

(b) *$C \subset \operatorname{int} M$; on the outside of B, C is a boundary component of an annular twist;*

(c) *$C \subset \partial M$.*

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

Moreover, a fixed point class $\mathbf{F}$ *is related to a boundary component $C \subset \partial X$ if and only if* $\mathbf{F}$ *intersects C.*

When we talk about the type of an n-point class $\mathbf{F}^n$ of a standard φ, we mean the type of $\mathbf{F}^n$ as a fixed point class of φ^n.

Definition. An n-point class $\mathbf{F}^n$ is *special* if it is either of type (1c)$^+$ with $p \ge 3$ and $k = 0$, or of types (2c)$^+$, (2d)$^+$ or (4)$^+$.

Corollary 3.2. *Almost all essential n-point classes has index ± 1. More* *precisely,*

(i) *If an n-point class* $\mathbf{F}^n$ *is special then* $\operatorname{ind}(\mathbf{F}^n, \varphi^n) \leq -1$. *Otherwise* $-1 \leq \operatorname{ind}(\mathbf{F}^n, \varphi^n) \leq 1$.

(ii) *The number of special n-point classes is at most* $-2\chi(X)$, *and*

$$\sum_{special\ \mathbf{F}^n} |\operatorname{ind}(\mathbf{F}^n, \varphi^n) + 1| \leq -2\chi(X).$$

(iii) $\|L_\Gamma(\varphi^n)\| + 2\chi(X) \leq N(\varphi^n) \leq \|L_\Gamma(\varphi^n)\|$.

(iv) *Suppose* $A \subset \partial X$ *is a union of* $a \geq 0$ *boundary circles of* X. *Then*

$$N(\varphi^n; X \setminus A) \geq \|L_\Gamma(\varphi^n)\| + 2\chi(X) - 2a.$$

(v) *If* X *is orientable and* φ *preserves orientation, then*

$$N(\varphi^n; X \setminus A) \geq \|L_\Gamma(\varphi^n)\| + 2\chi(X).$$

Proof. (i) is clear. (ii) follows from the proof of [**JG**, Theorem 4.1]. (iii) follows from (i) and (ii). (iv) uses the last statement of Lemma 3.1 and the fact that each boundary circle can intersect at most 2 fixed point classes of φ^n, so $N(\varphi^n; X \setminus A) \geq N(\varphi^n) - 2a$. In the orientation preserving case (v), observe that if $\mathbf{F}^n$ intersects ∂X then $\mathbf{F}^n$ is special. □

Corollary 3.3. *Special n-point classes of* φ *are stable under iteration, and are the only ones that can contain an inessential periodic point class of lower period. More precisely:*

Let $\mathbf{F}^n$ *be an n-point class of* φ. *Let* n' *be a multiple of* n *and* $\mathbf{F}^{n'}$ *be the* n'*-point class containing* $\mathbf{F}^n$. *Then*

(i) $\operatorname{ind}(\mathbf{F}^n, \varphi^n) \geq \operatorname{ind}(\mathbf{F}^{n'}, \varphi^{n'})$.

(ii) *If* $\mathbf{F}^n$ *is special, then* $\mathbf{F}^n$ *and* $\mathbf{F}^{n'}$ *are equal as subsets of* X *and*

$$\operatorname{ind}(\mathbf{F}^n, \varphi^n) = \operatorname{ind}(\mathbf{F}^{n'}, \varphi^{n'}).$$

(iii) *If* $\operatorname{ind}(\mathbf{F}^n, \varphi^n) = 0 > \operatorname{ind}(\mathbf{F}^{n'}, \varphi^{n'})$ *then* $\mathbf{F}^{n'}$ *is special and* φ^n *reverses the local orientation at* $\mathbf{F}^n$.

Proof. Clear from Lemma 3.1. □

Lemma 3.4. *If an n-point class* $\mathbf{F}^n$ *of* φ *is reducible to period* m, *then it contains some* m*-point class* $\mathbf{F}^m$.

Remark. The reducibility of $\mathbf{F}^n$ to period m means it "contains" some (possibly empty) m-point class. The point of this Lemma is that it indeed contains some non-empty m-point class.

Proof. It clearly suffices to prove the case $m = 1$. We want to show $\mathbf{F}^n$ contains a fixed point of φ.

By [**J1**, Lemma III.4.6], the reducibility of $\mathbf{F}^n$ to period 1 implies there is a path c from some $x \in \mathbf{F}^n$ to $\varphi(x)$ such that the loop $c(\varphi \circ c) \cdots (\varphi^{n-1} \circ c)$ is contractible. Now $(\varphi \circ c) \cdots (\varphi^{n-1} \circ c)(\varphi^n \circ c)$ is also contractible, hence $c \simeq \varphi^n \circ c$. Applying [**JG**, Lemma 3.4] to the standard map φ^n, we find a path γ in $\mathbf{F}^n$ homotopic to c. It follows that $\varphi(\mathbf{F}^n) = \mathbf{F}^n$, and $\gamma(\varphi \circ \gamma) \cdots (\varphi^{n-1} \circ \gamma)$ is contractible in $\mathbf{F}^n$ (because it is contractible in X and $\pi_1(\mathbf{F}^n)$ injects into $\pi_1(X)$).

According to Lemma 3.1, $\mathbf{F}^n$ is either a point, or a circle, or an arc, or a subsurface B of X with $\chi(B) \leq 0$. In the first or the third case, φ certainly has a fixed point on $\mathbf{F}^n$. In the second case φ is a rotation or a reflection on the circle, no path γ of the above type can exist unless $\varphi|_{\mathbf{F}^n}$ has a fixed point.

It remains to consider the case that $\mathbf{F}^n$ is a subsurface B and $\varphi|_B : B \to B$ is a periodic map. Equip B with a hyperbolic (or Euclidean) metric such that φ is an isometry. Let γ_0 be a shortest path of the above type, i.e. from some point to its φ-image and $\beta_0 := \gamma_0(\varphi \circ \gamma_0) \cdots (\varphi^{n-1} \circ \gamma_0)$ contractible. This β_0 must be a smooth closed geodesic because otherwise γ_0 can be shortened. But a smooth closed geodesic cannot be contractible unless it degenerates to a point. Hence γ_0 is a point, a fixed point of φ. □

Corollary 3.5. *Every homeomorphism $f : X \to X$ has Properties (EI) and (BI), hence $NI^\infty(f) = N^\infty(f) = L^\infty(f)$.*

Proof. Via isotopy we may replace f with a standard φ. By Lemma 3.4 and Corollary 3.3(iii) every reducible essential n-point class contains some essential periodic point class of lower period, except possibly the special ones. By Corollary 3.2(ii) we have Property (EI) with $E = -2\chi(X)$. (When X is orientable and φ preserves orientation, we can even take $E = 0$.)

On the other hand, by Corollary 3.2(i),(ii), f has Property (BI) with $B = 1 - 2\chi(X)$.

Then apply Theorems 2.2 and 2.3. □

Lemma 3.6. *Suppose φ is standard and λ is the largest stretching factor of the pseudo-Anosov pieces ($\lambda := 1$ if there is no pseudo-Anosov piece). Then*

$$h(\varphi) = \log \lambda \qquad \text{and} \qquad N^\infty(\varphi) = \lambda.$$

Proof. Let U be the open regular neighborhood of the k reducing curves in the Thurston theorem, and $\{M_j\}$ be the components of $X \setminus U$. Let λ_j be the stretching factor of φ_j if φ_j is pseudo-Anosov and $\lambda_j = 1$ otherwise. Thus $\lambda = \max_j \lambda_j$.

The topological entropy of a periodic map is 0. By (2.20) we see $h(\varphi_j) = \log \lambda_j$ for all j. Since the topological entropy of a Dehn twist is 0, $h(\varphi|U) = 0$. So the first conclusion follows from the fact that $h(\varphi) = \max\{h(\varphi_j), h(\varphi|U)\}$ (cf. [**Wa**, Theorem 7.5]).

To prove the second conclusion, we need an inequality

$$N(\varphi_j) - 2k \leq N(\varphi) \leq \sum_j N(\varphi_j) + 2k. \tag{3.1}$$

Let $\mathbf{F}$ be a fixed point class of φ. Observe from Lemma 3.1 that if $\mathbf{F} \subset M_j$, then $\operatorname{ind}(\mathbf{F}, \varphi) = \operatorname{ind}(\mathbf{F}, \varphi_j)$. So if $\mathbf{F}$ is counted in $N(\varphi)$ but not counted in $\sum_j N(\varphi_j)$, it must intersect U. But we see from Lemma 3.1 that a component of U can intersect at most 2 essential fixed point classes of φ. Hence the second inequality.

Let $\mathbf{F}_j$ be a fixed point class of φ_j and let $\mathbf{F}$ be the fixed point class of φ containing $\mathbf{F}_j$. If $\mathbf{F}_j$ makes a contribution to $N(\varphi_j) - N(\varphi)$, $\mathbf{F}$ must intersect U. Via Lemma 3.1 we check that each component of U can contribute at most 2 to $N(\varphi_j) - N(\varphi)$. Hence the first inequality. Thus (3.1) is proved.

Applying (3.1) to φ^n, we have

$$N(\varphi_j^n) - 2k \leq N(\varphi^n) \leq \sum_j N(\varphi_j^n) + 2k.$$

Taking the growth rate in n, we get

$$N^\infty(\varphi) = \max_j N^\infty(\varphi_j). \tag{3.2}$$

But according to (2.20), $N^\infty(\varphi_j) = \lambda_j$. Hence the second conclusion. $\square$

The above results are summarized in

Theorem 3.7. *Let X be a compact connected surface with $\chi(X) < 0$, and let $f : X \to X$ be a homeomorphism. Let $A \subset \partial X$ be a union of boundary circles such that $f(A) = A$. Then*

$$NI^\infty(f) = NI^\infty(f; X \setminus A) = N^\infty(f) = N^\infty(f; X \setminus A) = L^\infty(f) = \lambda,$$

where λ is the largest stretching factor of the pseudo-Anosov pieces in the Thurston canonical form of f ($\lambda := 1$ if there is no pseudo-Anosov piece).

3.2. Punctured surfaces. Let X be a connected compact surface and let P be a nonempty finite set of points (punctures) in the interior of X. Assume that $\chi(X) - |P| < 0$ where $|P|$ denotes the cardinality of P. Let $f : X, P \to X, P$ be a homeomorphism.

Let Y be the compactification of $X \setminus P$ by blowing up each point of P into its circle of unit tangent vectors. Then Y is a compact surface with $\chi(Y) = \chi(X) - |P| < 0$. The added circles form a set $Q \subset \partial Y$ and $Y \setminus Q = X \setminus P$.

There always exists a homeomorphism $f' : X, P \to X, P$ that is piecewise linear near P and isotopic to f rel P. We can even require that the connecting isotopy is supported in any given small neighborhood of P. See [**E**, Appendix]. We shall call such an f' a local rectification of f. (Indeed f' can be further made smooth near P if one prefers.)

Let $g : Y, Q \to Y, Q$ be the blow-up of $f' \setminus P$, i.e. the homeomorphism extending $f' : Y \setminus Q \to Y \setminus Q$ to Q according to the piecewise differential of f' at P (cf. [**B2**, §2]). Then the homotopy class of g, hence the isotopy class of g also (see [**E**]), is independent of the local rectification f'. Since $G := \pi_1(Y) = \pi_1(Y \setminus Q) = \pi_1(X \setminus P)$, we can identify the automorphism $g_G : \pi_1(Y) \to \pi_1(Y)$ with $f_G : \pi_1(X \setminus P) \to \pi_1(X \setminus P)$.

The relative Nielsen number $N(g; Y \setminus Q)$ is thus independent of the local rectification f'. We define the *punctured Nielsen number* of f to be $N(f \setminus P) := N(g; Y \setminus Q)$. It is a lower bound for the number of fixed points of $f \setminus P$ because if g is the blow up of a rectification f' of f in a sufficiently small neighborhood of P, then every fixed point of f' that is not a fixed point of f must be a fixed point of g related to Q.

There is a direct definition of the punctured Nielsen number $N(f \setminus P)$ without using rectifications. Two fixed points x, x' of f on $X \setminus P$ are said to be in the same punctured fixed point class of f if there is a path c in $X \setminus P$ such that $c \simeq f \circ c : I, 0, 1 \to X \setminus P, x, x'$. A fixed point x of f on $X \setminus P$ is said to be related to P if there is a path c in X such that $c \simeq f \circ c : I, I \setminus \{1\}, 0, 1 \to X, X \setminus P, x, P$. The punctured Nielsen number $N(f \setminus P)$ is defined to be the number of essential punctured fixed point classes of f that are unrelated to P.

The equivalence of the two definitions is not difficult to see. Let U be a regular neighborhood of P in X and let V be a smaller one such that $f(V) \subset U$. Then every fixed point of f on $V \setminus P$ is related to P. Let f' be a rectification of f on V and let g be its blow-up. This $N(g; Y \setminus Q)$ is easily identified with the second definition.

The definition using rectifications is more convenient in computations because it explicitly involves the ordinary Nielsen theory on Y.

But the direct definition sometimes gives us more insight. For example, we can see $N(f \setminus P) \geq N(f) - |P|$. In fact, consider an essential fixed point class of f that does not intersect P. It must be a disjoint union of punctured fixed point classes of f that are unrelated to P, so at least one of these latter classes must be essential, hence the inequality. Similarly, we can see that if $P \subset P'$ and $f : X, P', P \to X, P', P$, then $N(f \setminus P) + |P| \leq N(f \setminus P') + |P'|$.

Remark. $N(f \setminus P)$ is not the same as $N(f; X \setminus P)$. The former is often a much better lower bound for the number of fixed points of f on $X \setminus P$. See the examples in §4.2. Note that the coordinates of the fixed point classes of $f \setminus P$ are not in $\pi_1(T_f)$, but in $\Gamma := \pi_1(T_g) = \pi_1(T_g \setminus T_{g|Q}) = \pi_1(T_f \setminus T_{f|P})$, the fundamental group of the complement of the link $T_{f|P}$ in T_f.

We now turn to the punctured invariants for periodic orbits of f.

Definition. $N_\Gamma(f^n \setminus P) := N_\Gamma(g^n; Y \setminus Q)$, a lower bound for the number of n-orbits of f on $X \setminus P$.

$NI_\Gamma(f^n \setminus P) := NI_\Gamma(g^n; Y \setminus Q)$, a lower bound for the number of primary n-orbits of f on $X \setminus P$.

$N(f^n \setminus P) := N(g^n; Y \setminus Q)$, a lower bound for the number of n-points of f on $X \setminus P$.

$L_\Gamma(f^n \setminus P) := L_\Gamma(g^n; Y \setminus Q)$, the sum of absolute values of the indices of n-orbits of f on $X \setminus P$.

The asymptotic invariant is also defined.

Definition. $N^\infty(f \setminus P) := \mathrm{Growth}_{n\to\infty} N_\Gamma(f^n \setminus P)$.

Theorem 3.8. *$N^\infty(f \setminus P)$ is the common growth rate of various punctured Nielsen numbers:*

$$\begin{aligned} N^\infty(f \setminus P) &= \mathrm{Growth}_{n\to\infty} N(f^n \setminus P) = \mathrm{Growth}_{n\to\infty} NI_\Gamma(f^n \setminus P) \\ &= \mathrm{Growth}_{n\to\infty} \|L_\Gamma(f^n \setminus P)\| = \lambda, \end{aligned}$$

where λ is the largest stretching factor of the pseudo-Anosov pieces in the Thurston canonical form of the punctured homeomorphism $f : X \setminus P \to X \setminus P$ ($\lambda := 1$ if there is no pseudo-Anosov piece).

Proof. Easy from Theorem 3.7 and the definitions. □

To compare $N^\infty(f \setminus P)$ with the topological entropy $h(f)$, we need a lemma.

Lemma 3.9. *There exists a finite regular branched cover $\widehat{X}, \widehat{P} \to X, P$ with branching set P such that every homeomorphism $f : X, P \to X, P$ lifts to a homeomorphism $\hat{f} : \widehat{X}, \widehat{P} \to \widehat{X}, \widehat{P}$.*

Proof. We need the following fact from group theory: Suppose G is a free group of finite rank and $g_1, \cdots, g_k \in G$ are all $\neq 1$. Then there exists a normal subgroup $K \subset G$ with finite index in G, which is invariant under any automorphism of G, and such that all $g_1, \cdots, g_k \notin K$. (See [**LS**, pp. 143, 195–196].)

Let Y, Q be the blow-up of X, P as before. Then X, P is obtained from Y, Q by shrinking every component of Q to a point. Let $G = \pi_1(X \setminus P) = \pi_1(Y)$. Let $k = |P|$ and $g_1, \cdots, g_k \in G$ be elements represented by the circles in Q. Since $\chi(Y) = \chi(X) - |P| < 0$, no boundary curve of Y is contractible in Y, hence every $g_i \neq 1$. Now take a normal subgroup K guaranteed by the above algebraic fact.

Let $q : \widehat{Y} \to Y$ be the regular covering of Y such that $q_* \pi_1(\widehat{Y}) = K$, and let $\widehat{Q} = q^{-1}(Q)$. Let $p : \widehat{X}, \widehat{P} \to X, P$ be obtained from $q : \widehat{Y}, \widehat{Q} \to Y, Q$ by shrinking every component of $\widehat{Q}$ and Q to a point. Then $p : \widehat{X}, \widehat{P} \to X, P$ is a finite regular branched cover. Every point of P is a branching point because a small circle around it cannot be lifted to a circle in $\widehat{X}$. On $X \setminus P$, the covering $p : \widehat{X} \setminus \widehat{P} \to X \setminus P$ also has $p_* \pi_1(\widehat{X} \setminus \widehat{P}) = K$.

A homeomorphism $f : X, P \to X, P$ restricts to $f \setminus P : X \setminus P \to X \setminus P$ which induces an automorphism of G. By the invariance property of $K \subset G$, the homeomorphism $f \setminus P$ lifts to a homeomorphism $\hat{f} \setminus \widehat{P} : \widehat{X} \setminus \widehat{P} \to \widehat{X} \setminus \widehat{P}$. Then compactify to get the required lifting $\hat{f} : \widehat{X} \to \widehat{X}$. $\square$

The following is the analogue of Theorem 2.1 for punctured surfaces.

Theorem 3.10. *For any homeomorphism $f : X, P \to X, P$,*

$$h(f) \geq \log N^{\infty}(f \setminus P).$$

Proof. Let $f' : X, P \to X, P$ and $g : Y, Q \to Y, Q$ be as before. Let $\psi : Y, Q \to Y, Q$ be the standard form of g, and let $\varphi : X, P \to X, P$ be the blow-down of ψ, i.e. shrinking every component of Q back again to a point.

First consider the simpler case that no component C of Q is a 1-prong boundary component of a pseudo-Anosov piece of ψ. Then φ itself is a standard map isotopic to f. So by Theorems 2.1 and 3.7,

$$\begin{aligned} h(f) &\geq \log N^{\infty}(f) = \log \lambda_{\varphi} = \log \lambda_{\psi} \\ &= \log N^{\infty}(g) = \log N^{\infty}(g; Y \setminus Q) = \log N^{\infty}(f \setminus P). \end{aligned}$$

For the general case, let $\widehat{X}, \widehat{P} \to X, P$ be the finite branched cover in Lemma 3.9, and lift f to a homeomorphism $\hat{f} : \widehat{X}, \widehat{P} \to \widehat{X}, \widehat{P}$. When blown up, $\widehat{Y}, \widehat{Q} \to Y, Q$ is an honest regular cover. Let $\hat{g}, \hat{\psi} : \widehat{Y}, \widehat{Q} \to \widehat{Y}, \widehat{Q}$ be the lifts of g, ψ. Now $\hat{\psi}$ is a standard map. If a component $\widehat{C}$ of $\widehat{Q}$ projects to a boundary component C of a pseudo-Anosov piece of ψ, then $\widehat{C}$ has more prongs than C does because every point of P is a branching point. Thus $\hat{\psi}$ belongs to the simpler case already proved, so $h(\hat{f}) \geq \log \lambda_{\hat{\psi}}$. But we know $h(\hat{f}) = h(f)$ by [**B1**, Theorem 17], and $\lambda_{\hat{\psi}} = \lambda_{\psi}$ by construction. Hence by

Theorem 3.7,

$$h(f) \geq \log \lambda_\psi = \log N^\infty(g) = \log N^\infty(f \setminus P).$$

□

4. Examples of asymptotic estimates.

To illustrate our method of estimation, we study some surface homeomorphisms arising in the recent literature of dynamical systems theory. §4.1 improves a well known result of Handel. §4.2 uses an abelian representation to estimate the growth rate of periodic orbits. §4.3 displays the power of non-abelian representations when abelianization does not work. For more applications to dynamics, see [**J3**].

4.1. Orientation reversing homeomorphisms of surfaces. The following theorem strengthens a result of Handel [**H**]:

Theorem 4.1. *If $f : X \to X$ is an orientation reversing homeomorphism of a compact oriented surface of genus g, and if f has orbits with $g+2$ distinct odd periods, then the number of primary n-orbits grows exponentially in n, hence $h(f) > 0$.*

Proof. It is shown in [**H**] that the punctured homeomorphism $f : X \setminus P \to X \setminus P$, where P is the union of the known orbits, has at least one pseudo-Anosov piece in its Thurston canonical form. According to Theorem 3.8, this guarantees $N^\infty(f \setminus P) > 1$, hence $h(f) > 0$ by Theorem 2.7. □

In [**H**] periodic points are discussed under the assumption that f is differentiable at the periodic points in question, which is now deleted, and the conclusion is that f has orbits with infinitely many distinct periods.

4.2. Orientation preserving embeddings of the disk. Let $X = D^2$ be a disk in the plane $\mathbb{R}^2$, and let $f : X \to X$ be an orientation preserving embedding. Suppose $P = \{x_1, \cdots, x_r\}$ is a finite set in the interior of D^2 such that $f(P) = P$. Then $N^\infty(f \setminus P)$ can be estimated once we know the induced automorphism $f_G : G \to G$ where $G := \pi_1(X \setminus P)$. (That f is an embedding rather than a homeomorphism is only a technical problem. By slightly enlarging the disk, we can extend the embedding to a homeomorphism of D^2 so that all the additional periodic points arising from this extension are related (on $D^2 \setminus P$) to the boundary of D^2, hence do not affect the asymptotic Nielsen number by Theorem 3.7.)

Consider the map $f : X, P \to X, P$ studied in [**GST**], where $r = 3$ and G is the free group on 3 standard generators a_1, a_2, a_3. We do not quote the description of the map f given there in terms of a braid, but only point out that $f_G : G \to G$ is easily seen to be

$$\begin{cases} a_1' := f_G(a_1) = a_1 a_3 a_1^{-1}, \\ a_2' := f_G(a_2) = a_1, \\ a_3' := f_G(a_3) = a_3^{-1} a_2 a_3. \end{cases} \tag{4.1}$$

In [**GST**] it is shown that there exist primary periodic orbits of every period n. We shall give a lower bound to the growth rate of the number of such orbits.

The Jacobian matrix D in Fox calculus is readily calculated.

$$D = \begin{pmatrix} 1 - a_1 a_3 a_1^{-1} & 0 & a_1 \\ 1 & 0 & 0 \\ 0 & a_3^{-1} & -a_3^{-1} + a_3^{-1} a_2 \end{pmatrix}. \tag{4.2}$$

According to (1.16), we have $\Gamma = \langle a_1, a_2, a_3, z \mid a_i z = z a_i',\ i = 1, 2, 3 \rangle$. An obvious way to get a representation of Γ is to abelianize. Thus we obtain a U(1) representation ρ by letting all $a_i \mapsto a$ and $z \mapsto 1$, where a is a unimodular complex number.

Thus

$$(zD)^\rho = \begin{pmatrix} 1 - a & 0 & a \\ 1 & 0 & 0 \\ 0 & a^{-1} & 1 - a^{-1} \end{pmatrix}, \tag{4.3}$$

so that by (1.21), for the blow-up $g : Y, Q \to Y, Q$ of a local rectification of f,

$$\zeta_\rho(g) = \frac{\det(I - t(zD)^\rho)}{\det(I - tz^\rho)} = 1 - (1 - a - a^{-1})t + t^2. \tag{4.4}$$

Take $a = -1$, then we get the zeta function $\zeta_\rho(g) = 1 - 3t + t^2$ and its smallest root is $r = (3 - \sqrt{5})/2$. Hence, by Corollary 2.5, Theorem 3.8 and Theorem 3.10, we get the estimates

$$\begin{gathered} NI^\infty(f \setminus P) = N^\infty(f \setminus P) = L^\infty(f \setminus P) = L^\infty(g) \geq (3 + \sqrt{5})/2, \\ h(f) \geq \log \frac{3 + \sqrt{5}}{2}. \end{gathered} \tag{4.5}$$

To obtain an upper bound by Proposition 2.6, note from (1.17) that

$$\|\widetilde{F}_0\| = (1), \qquad \|\widetilde{F}_1\| = \|D\| = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix}. \tag{4.6}$$

The characteristic polynomial of $\|D\|$ is $(\lambda-1)(\lambda^2-3\lambda+1)$, so (2.14) gives

$$L^\infty(f\setminus P)\le \text{spectral radius of } \|D\| = (3+\sqrt{5})/2. \tag{4.7}$$

Now that the lower and upper estimates coincide, we have

$$NI^\infty(f\setminus P) = N^\infty(f\setminus P) = L^\infty(f\setminus P) = (3+\sqrt{5})/2. \tag{4.8}$$

Since X is the disk, all $N(f^n)=1$ and $N^\infty(f)=1$, so the punctured Nielsen numbers do give better estimates.

Another example on the disk is Smale's horseshoe. In [**F3**] it was shown that for the horseshoe embedding $f: D^2\to D^2$ there is a 5-orbit $P=\{p_1,\cdots,p_5\}$. Using this orbit as punctures, the argument of [**F3**, §4] can be adapted to show that, for some representation $\rho:\Gamma\to \mathrm{U}(1)$, $\zeta_\rho(f)=1+t-t^2+t^3+t^4$. Thus by Corollary 2.5 we get $N^\infty(f\setminus P)>1.72$ and $h(f)>\log 1.72$.

4.3. A homeomorphism of the torus. The following example is taken from [**LM**, Example 2] where it was shown that $h(f)>0$.

Let X be the torus T^2 represented as $\mathbb{R}^2/\mathbb{Z}^2$, and let the three points $Q_1=(0,0)$, $Q_2=(\frac13,\frac13)$ and $Q_3=(\frac23,\frac23)$ constitute the puncture set P. Let $D_1,D_2: X,P\to X,P$ be diffeomorphisms of the form

$$D_1:\begin{cases} x'=x\\ y'=y+B_1(x)\end{cases} \qquad D_2:\begin{cases} x'=x+B_2(y)\\ y'=y\end{cases} \tag{4.9}$$

where x,y are in the unit interval I, and B_1,B_2 are smooth functions on I with $B_1\left(\frac23\right)=1$, $B_1(x)=0$ for x outside of the open interval $(\frac12,\frac56)$, $B_2\left(\frac13\right)=1$, $B_2(y)=0$ for y outside of the open interval $(\frac16,\frac12)$. Let $f=D_2\circ D_1: X,P\to X,P$.

Choose the point $(\frac12,\frac12)$ to be the base point in X. The fundamental group $G=\pi_1(X\setminus P)$ is a free group of rank 4 with generators a_2,a_3,b_1,b_2, where b_1 is represented by the loop $\{(\frac12+t,\frac12)\}_{t\in I}$; b_2 by the loop $\{(\frac12,\frac12+t)\}_{t\in I}$; a_2 is represented by the square loop with sides on the lines $y=\frac12$, $x=\frac16$, $y=\frac16$ and $x=\frac12$; a_3 represented by the square loop with sides on the lines $y=\frac12$, $x=\frac56$, $y=\frac56$ and $x=\frac12$.

It is clear that D_1,D_2 act on G by

$$D_1:\begin{cases} a_2\mapsto a_2\\ a_3\mapsto b_2a_3b_2^{-1}\\ b_1\mapsto a_3^{-1}b_1\\ b_2\mapsto b_2,\end{cases} \qquad D_2:\begin{cases} a_2\mapsto b_1a_2b_1^{-1}\\ a_3\mapsto a_3\\ b_1\mapsto b_1\\ b_2\mapsto b_2b_1a_2b_1^{-1}.\end{cases} \tag{4.10}$$

So the automorphism $f_G : G \to G$ induced by f is

$$
(4.11) \qquad f_G : \begin{cases} a_2 \mapsto a_2' = b_1 a_2 b_1^{-1} \\ a_3 \mapsto a_3' = (b_2 b_1 a_2 b_1^{-1})\, a_3\, (b_2 b_1 a_2 b_1^{-1})^{-1} \\ b_1 \mapsto b_1' = a_3^{-1} b_1 \\ b_2 \mapsto b_2' = b_2 b_1 a_2 b_1^{-1}. \end{cases}
$$

It is routine to compute the Fox calculus Jacobian matrix D.

$$
(4.12) \qquad D = \begin{pmatrix} b_1 & 0 & 1 - a_2' & 0 \\ (1 - a_3')\, b_2 b_1 & b_2' & (1 - a_3')\, b_2\, (1 - a_2') & 1 - a_3' \\ 0 & -a_3^{-1} & a_3^{-1} & 0 \\ b_2 b_1 & 0 & b_2 - b_2' & 1 \end{pmatrix}.
$$

Now

$$
(4.13)\ \Gamma = \langle a_2, a_3, b_1, b_2, z \mid a_2 z = z a_2', a_3 z = z a_3', b_1 z = z b_1', b_2 z = z b_2' \rangle.
$$

The abelianization does not work here, because the nature of (4.11) would force both a_2, a_3 to become 1, so that any U(1) representation can only give the trivial estimate $N^\infty(f \setminus P) \geq 1$. Thus we have to look for non-abelian representations of Γ. Fortunately a simple representation of Γ into the multiplicative group of unimodular quaternions works.

(4.14)

$$
\rho : \Gamma \to \mathrm{Sp}(1); \quad z \mapsto i, \quad a_2 \mapsto -1, \quad a_3 \mapsto -1, \quad b_1 \mapsto j, \quad b_2 \mapsto k.
$$

Under this representation, the matrix zD becomes

$$
(4.15) \qquad (zD)^\rho = \begin{pmatrix} k & 0 & 2i & 0 \\ 2 & j & -4j & 2i \\ 0 & i & -i & 0 \\ 1 & 0 & -2j & i \end{pmatrix}.
$$

But quaternions form a skew-field, not a field. We have to identify Sp(1) with SU(2) via the correspondence

$$
(4.16) \quad 1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad j \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad k \mapsto \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.
$$

Thus $(zD)^\rho$ becomes an 8×8 complex matrix. One then calculates that

$$
(4.17) \qquad \det(1 - t(zD)^\rho) = (1 + t)^2 (1 + t^2)[(1 - t + t^2)^2 + 3t^2],
$$

$$\zeta_\rho = (1+t)^2[(1-t+t^2)^2+3t^2], \tag{4.18}$$

from which follow the estimates

$$N^\infty(f\setminus P) > 2.29, \qquad h(f) > \log 2.29 \tag{4.19}$$

by Corollary 2.5.

For the upper bound, we have

$$\|\widetilde{F}_0\| = (1), \qquad \|\widetilde{F}_1\| = \|D\| = \begin{pmatrix} 1\,0\,2\,0 \\ 2\,1\,4\,2 \\ 0\,1\,1\,0 \\ 1\,0\,2\,1 \end{pmatrix}. \tag{4.20}$$

The characteristic polynomial of $\|D\|$ is $(\lambda^2+1)(\lambda^2-4\lambda+1)$. Then (2.14) gives

$$L^\infty(f\setminus P) \le \text{spectral radius of } \|D\| = 2+\sqrt{3} < 3.74. \tag{4.21}$$

So the estimate we get is

$$2.29 < N^\infty(f\setminus P) < 3.74. \tag{4.22}$$

5. Questions.

It is clear from our discussion that among the asymptotic invariants, $NI^\infty(f)$ is most interesting geometrically, and $L^\infty(f)$ is most manageable algebraically. So, the conditions for the equalities $NI^\infty(f) = N^\infty(f) = L^\infty(f)$ and the extent to which they can fail are worth further study. (In fact, the author knows of no counter-example to the equality $NI^\infty(f) = N^\infty(f)$.) For example,

Question 5.1. Is it true that a self-homeomorphism (or self-map) f of an aspherical compact polyhedron X always has Properties (EI) and (BI), or at least $NI^\infty(f) = N^\infty(f) = L^\infty(f)$?

A question related to the Entropy Conjecture of Shub is the following:

Question 5.2. For smooth maps of compact manifolds, is it always true that

$$h(f) \ge \log L^\infty(f)?$$

Another natural question is

Question 5.3. Give conditions for $\log N^{\infty}(f)$ to be the best lower bound for $h(f)$ of all maps homotopic to f. In other words, in the inequality

$$\inf\{h(g) \mid g \simeq f : X \to X\} \geq \log N^{\infty}(f)$$

when does the equality hold? (The example of torus maps in §2.5 shows that the equality may fail even for very nice maps.)

References

[Bi] J.S. Birman, *Braids, Links, and Mapping Class Groups*, Ann. Math. Studies, vol. 82, Princeton Univ. Press, Princeton, 1974.

[Bt] R. Bott, *On the shape of a curve*, Advances in Math., **16** (1975), 23–38.

[B1] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Tran. Amer. Math. Soc., **153** (1971), 401–413.

[B2] R. Bowen, *Entropy and the fundamental group*, The Structure of Attractors in Dynamical Systems, (N.G. Markley et al., eds.), Lecture Notes in Math., vol. 668, Springer-Verlag, Berlin, Heidelberg, New York, 1978, 21–29.

[E] D.B.A. Epstein, *Curves on 2-manifolds and isotopies*, Acta Math., **115** (1966), 83–107.

[FH] E. Fadell and S. Husseini, *The Nielsen number on surfaces*, Topological Methods in Nonlinear Functional Analysis, (S.P. Singh et al., eds.), Contemp. Math., vol. 21, Amer. Math. Soc., Providence, 1983, 59–98.

[FLP] A. Fathi, F. Laudenbach and V. Poénaru, *Travaux de Thurston sur les surfaces, Séminaire Orsay*, Astérisque, vol. 66–67, Soc. Math. France, Paris, 1979.

[F1] D. Fried, *Periodic points and twisted coefficients*, Geometric Dynamics, (J. Palis Jr. ed.), Lecture Notes in Math., vol. 1007, Springer-Verlag, Berlin, Heidelberg, New York, 1983, 261–293.

[F2] ———, *Homological identities for closed orbits*, Invent. Math., **71** (1983), 419–442.

[F3] ———, *Entropy and twisted cohomology*, Topology, **25** (1986), 455–470.

[F4] ———, *Lefschetz formulas for flows*, The Lefschetz Centennial Conference, Part III: Proceedings on Differential Equations, (A. Verjovsky ed.), Contemp. Math., vol. 58. III, Amer. Math. Soc., Providence, 1987, 19–69.

[Fu] F.B. Fuller, *The treatment of periodic orbits by the methods of fixed point theory*, Bull. Amer. Math. Soc., **72** (1966), 838–840.

[GST] J.-M. Gambaudo, S. van Strien and C. Tresser, *Vers un ordre de Sarkovskii pour les plongements du disque préservant l'orientation*, C. R. Acad. Sci. Paris, Série I, **310** (1990), 291–294.

[H] M. Handel, *The entropy of orientation-reversing homeomorphisms of surfaces*, Topology, **21** (1982), 291–296.

[I] N.V. Ivanov, *Entropy and the Nielsen numbers*, Soviet Math. Dokl., **26** (1982), 63–66.

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

[J2] ———, *A characterization of fixed point classes*, Fixed Point Theory and its Applications, (R.F. Brown ed.), Contemp. Math., vol. 72, Amer. Math. Soc., Providence, 1988, 157–160.

[J3] ———, *Nielsen theory for periodic orbits and applications to dynamical systems*, Nielsen Theory and Dynamical Systems, (Ch. McCord ed.), Contemp. Math., vol. 152, Amer. Math. Soc., Providence, 1993.

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

[LM] J. Llibre and R.S. MacKay, *Rotation vectors and entropy for homeomorphisms of the torus isotopic to the identity*, Ergod. Th. Dyn. Sys., **11** (1991), 115–128.

[LS] R.C. Lyndon and P.E. Schupp, *Combinatorial Group Theory*, Springer, New York, 1977.

[MP] M. Misiurewicz and F. Przytycki, F. *Entropy conjecture of tori*, Bull. Ac. Pol. Sci., Ser. Math., **25** (1977), 575–578.

[R] K. Reidemeister, *Automorphismen von Homotopiekettenringen*, Math. Ann., **112** (1936), 586–593.

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

[Wa] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York, 1982.

[We] F. Wecken, *Fixpunktklassen*, II, Math. Ann., **118** (1942), 216–234.

[Z] X.-Z. Zhao, *A relative Nielsen number for the complement*, Topological Fixed Point Theory and Applications, (B. Jiang ed.), Lecture Notes in Math., vol. 1411, Springer-Verlag, Berlin, Heidelberg, New York, 1989, 189–199.

Received May 28, 1993 and revised October 22, 1993. Partially supported by NSFC. The author is also grateful to the hospitality of UCLA and Heidelberg University while this paper was being written up, and the partial support by NSF Grant DMS88-57452 and Deutsche Forschungs Gemeinschaft.

PEKING UNIVERSITY
BEIJING 100871, CHINA