Ergod. Th. & Dynam. Sys. (1988), **8**, 585–596
Printed in Great Britain

Entropy and semi-conjugacy in dimension two

MICHAEL HANDEL
Department of Mathematics and Computer Science, Lehman College, Bronx, New York 10468, *USA*

(*Received* 18 *May* 1987, *revised* 3 *December* 1987)

Abstract. We prove that if a diffeomorphism f of a closed surface is homotopic to and has the same topological entropy as a pseudo-Anosov homeomorphism g, then f is semi-conjugate to g. As part of the proof, a necessary and sufficient condition is given for a pseudo-orbit of a pseudo-Anosov homeomorphism g to be shadowed by an actual orbit of g.

0. *Introduction*

We showed in [H] that if $f: M^2 \to M^2$ is a pseudo-Anosov diffeomorphism and if $g: M \to M$ is homotopic to f, then there is a closed subset $Y \subset M$ such that $g \mid Y$ is semi-conjugate to f; i.e. there is a map $\pi: Y \to M$ such that $\pi \circ g \mid Y = f \circ \pi$. In this paper we give a sufficient condition for Y to be all of M and hence for g to be semi-conjugate to f. In general the pre-images of points under the semi-conjugacy will be disconnected.

THEOREM 0.1. *Let $f: M^2 \to M^2$ be a pseudo-Anosov diffeomorphism of a closed surface and let $g: M^2 \to M^2$ be a homeomorphism that is homotopic to f. If g has the same topological entropy as f (it is necessarily at least as big) then g is semi-conjugate to f.*

The introduction of topological entropy in this context is due to John Smillie who showed that if g has periodic points that are not Nielsen equivalent to periodic points of f (or equivalently if Y does not contain all the periodic points of g) then the topological entropy of g is strictly bigger than that of f.

Let $N_n(g)$ be the number of distinct Nielsen classes represented by the fixed point set $\mathrm{Fix}(g^n)$ of g^n. As a corollary of Theorem 0.1, we show (Corollary 3.6) that if the exponential growth rate of $N_n(g)$ equals that of $N_n(f)$, then g is semi-conjugate to f.

In the final section of the paper, we consider extensions of Theorem 0.1 to the more general setting used by Franks for Anosov diffeomorphisms and Fathi for pseudo-Anosov homeomorphisms.

I would like to thank Albert Fathi for his helpful comments and careful reading of an earlier manuscript for this paper.

1. *Notation and definitions*

Throughout this paper M is a closed surface, f is a pseudo-Anosov diffeomorphism of M and g is a homeomorphism of M that is homotopic to f. Following A. B. Katok, we say that the f-orbit of x is *globally shadowed* by the g-orbit of y if there are lifts $\tilde{f}$, $\tilde{g}$, $\tilde{x}$ and $\tilde{y}$ to the universal cover $\tilde{M}$ of M such that $\tilde{D}(\tilde{f}^n(\tilde{x}), \tilde{g}^n(\tilde{y})) < K$ for all n and some fixed K where $\tilde{D}$ is an equivariant metric on $\tilde{M}$. Define $Y = \{y \in M$: the g-orbit of y is globally shadowed by the f-orbit of some $x \in M\}$ and let $\pi: Y \to M$ be the map which takes y to its associated x. We showed in **[H]** that Y is a closed subset of M, that π is a surjective map and that π semi-conjugates $g \mid Y$ onto f.

Our proof takes place in the universal cover $\tilde{M}$ of M and uses the invariant stable and unstable foliations (with singularities) on $\tilde{M}$ that are lifted from the invariant stable and unstable foliations (with singularities) that characterize f. Readers that are not familiar with pseudo-Anosov diffeomorphisms should consult **[T]**, **[F–L–P]** or **[H–T]** as required. For any point $\tilde{p} \in \tilde{M}$, we denote the *unstable leaf containing* $\tilde{p}$ by $\tilde{\mu}(\tilde{p})$ and the *stable leaf containing* $\tilde{p}$ by $\tilde{\sigma}(\tilde{p})$. When the unstable leaf through $\tilde{p}$ contains a singularity we assume that $\tilde{\mu}(\tilde{p})$ contains that singularity and each of the finitely many leaves emanating from that singularity; similarly for $\tilde{\sigma}(\tilde{p})$. The transverse measures for the foliations determine equivariant pseudo-metrics $\tilde{D}_s: \tilde{M} \times \tilde{M} \to [0, \infty)$ and $\tilde{D}_u: \tilde{M} \times \tilde{M} \to [0, \infty)$ satisfying $\tilde{D}_u(\tilde{f}\tilde{x}_1, \tilde{f}\tilde{x}_2) = \lambda \tilde{D}_u(\tilde{x}_1, \tilde{x}_2)$ and $\tilde{D}_s(\tilde{f}^{-1}\tilde{x}_1, \tilde{f}^{-1}\tilde{x}_2) = \lambda \tilde{D}_s(\tilde{x}_1, \tilde{x}_2)$ for all $\tilde{x}_1, \tilde{x}_2 \in \tilde{M}$ and all lifts $\tilde{f}$ of f where $\lambda > 1$ is the expansion constant for f. We will use the *equivariant metric* $\tilde{D}$ on $\tilde{M}$ defined by $\tilde{D} = \tilde{D}_u + \tilde{D}_s$.

Since Y is unchanged when f and g are replaced by f^k and g^k for any $k > 0$, we may assume that f fixes each singularity of the stable and unstable foliations, that f maps each stable and unstable half-leaf that initiates at a singularity to itself, and that $\lambda > 4$.

An *efficient path* (see figure 1.1) between $\tilde{x}$ and $\tilde{y}$ is a path that is made up of segments of stable and unstable leaves, that has the minimum possible number of such segments and that is shortest in the $\tilde{D}$-metric.

Choose $K > 0$ so that for any pair $\tilde{f}$ and $\tilde{g}$ of equivariantly homotopic lifts of f and g, and for all $\tilde{x} \in \tilde{M}$, $\tilde{D}(\tilde{f}(\tilde{x}), \tilde{g}(\tilde{x})) < K$. It follows that if $\tilde{D}_u(\tilde{x}_1, \tilde{x}_2) > K$ then

$$\tilde{D}_u(\tilde{g}(\tilde{x}_1), \tilde{g}(\tilde{x}_2)) > \tilde{D}_u(\tilde{x}_1, \tilde{x}_2) + K$$

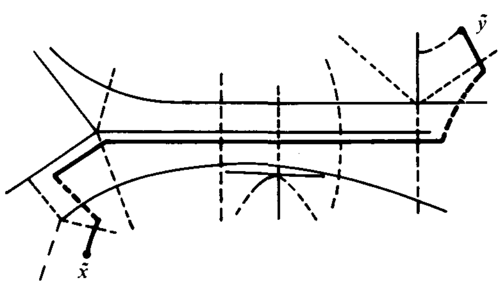

FIGURE 1.1

https://doi.org/10.1017/S0143385700004703 Published online by Cambridge University Press

and

$$\tilde{D}_u(\tilde{f}(\tilde{x}_1), \tilde{g}(\tilde{x}_2)) > \tilde{D}_u(\tilde{x}_1, \tilde{x}_2) + 2K;$$

similarly if $\tilde{D}_s(\tilde{x}_1, \tilde{x}_2) > K$ then

$$\tilde{D}_s(\tilde{g}^{-1}(\tilde{x}_1), \tilde{g}^{-1}(\tilde{x}_2)) > \tilde{D}_s(\tilde{x}_1, \tilde{x}_2) + K$$

and

$$\tilde{D}_s(\tilde{f}^{-1}(\tilde{x}_1), \tilde{g}^{-1}(\tilde{x}_2)) > \tilde{D}_s(\tilde{x}_1, \tilde{x}_2) + 2K.$$

If $\tilde{z}$ is a singularity of the foliations and if W^u is a component of $\tilde{M} - \tilde{\sigma}(\tilde{z})$ then W^u is called an *unstable wedge based at* $\tilde{z}$. See figure 1.2. For each unstable wedge W^u, we define E^u, an *unstable envelope based at* $\tilde{z}$ to be the component of $\{\tilde{x} \in W^u: \tilde{D}_u(\tilde{x}, \tilde{\sigma}(\tilde{z})) > K\}$ that intersects the unstable half-leaf in W^u originating at $\tilde{z}$. Note that if $\tilde{g}$ is the lift of g that is equivariantly homotopic to the lift $\tilde{f}$ of f that fixes $\tilde{z}$, and if E^u is an unstable envelope based at $\tilde{z}$, then $\tilde{g}(E^u) \subset E^u$. Stable wedges and envelopes are defined similarly.

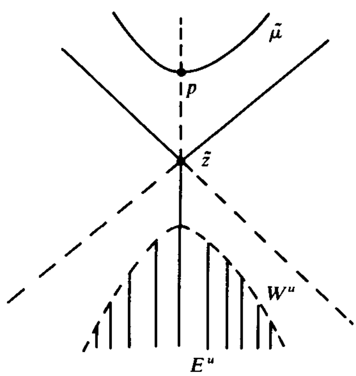

FIGURE 1.2

Remark 1.1. If $\tilde{x} \in \tilde{M}$ and $\tilde{\sigma}$ is a stable leaf such that $\tilde{D}_u(\tilde{x}, \tilde{\sigma}) > K$, then for all $n \geq 0$, $\tilde{g}^n(\tilde{x})$ is in the component of $\tilde{M} - \tilde{f}^n(\tilde{\sigma})$ that is the image under $\tilde{f}^n$ of the component of $\tilde{M} - \tilde{\sigma}$ that contains $\tilde{x}$. If $\tilde{\sigma} \subset W^u$ intersects the unstable half-leaf originating at $\tilde{z}$, then $\tilde{f}^n(\tilde{\sigma}) \subset E^u$ for all sufficiently large n. It follows that if $\tilde{x} \in \{W^u: \tilde{D}_u(\tilde{x}, \tilde{\sigma}(\tilde{z})) > K\}$, then $\tilde{g}^n(\tilde{x}) \in E^u$ for all sufficiently large n.

The *stable variation* along a path ρ is the maximum $\tilde{D}_s$-distance between two points on ρ. Note that if ρ has a stable variation less than K, then so does $\tilde{g}(\rho)$.

2. *An equivalent condition for semi-conjugacy*

The construction in [H] shows that the homeomorphism g is semi-conjugate to f if for each $y \in M$, there exists $x \in M$ such that the g-orbit of y is globally shadowed by the f-orbit of x. For Anosov diffeomorphisms, x can always be found by the standard shadowing techniques. In the pseudo-Anosov category, these techniques break down and the inability to solve the shadowing problem can be detected by the asymptotic behavior of the orbit of a lift $\tilde{y}$ of y with respect to some lift $\tilde{g}_0$ of g as follows.

PROPOSITION 2.1. *If the g-orbit of y is not globally shadowed by some f-orbit, then there is a singularity $\tilde{z}$ and disjoint stable and unstable envelopes E^s and E^u based at $\tilde{z}$ such that $\tilde{g}^{-n}(\tilde{y}) \in E^s$ and $\tilde{g}^n(\tilde{y}) \in E^u$ for all sufficiently large n, where $\tilde{g}$ is the lift of g that is equivariantly homotopic to the lift $\tilde{f}$ of f that fixes $\tilde{z}$.*

Proof of Proposition 2.1. Choose lifts $\tilde{y}$ of y, $\tilde{g}_0$ of g and $\tilde{f}_0$ of f such that $\tilde{g}_0$ and $\tilde{f}_0$ are equivariantly homotopic. We search for an $\tilde{f}_0$-orbit that shadows the $\tilde{g}_0$-orbit of $\tilde{y}$; the construction fails only when we encounter the desired lift $\tilde{g}$.

We first find a stable leaf $\tilde{\sigma}$ such that $\tilde{D}_u(\tilde{g}_0^n(\tilde{y}), \tilde{f}_0^n(\tilde{\sigma})) < K$ for all $n > 0$. If $\tilde{\mu}(\tilde{y})$ is non-singular, choose an orientation for $\tilde{\mu}(\tilde{y})$ and denote this oriented leaf by $\tilde{\mu}_0(\tilde{y})$. If $\tilde{\mu}(\tilde{y})$ contains a singularity $\tilde{z}_0$, then either $\tilde{g}_0^n(\tilde{y}) \in N_K(\tilde{f}_0^n(\tilde{\sigma}(\tilde{z}_0)))$ for all $n > 0$ and we're done, or (Remark 1.1) $\tilde{g}_0^n(\tilde{y}) \in \tilde{f}_0^n(E^u)$ for some unstable envelope E^u based at $\tilde{z}_0$ and for all sufficiently large n. In this latter case denote the oriented unstable half-leaf initiating at $\tilde{z}_0$ and intersecting E^u by $\tilde{\mu}_0(\tilde{y})$.

If the stable leaf $\tilde{\sigma}(\tilde{p})$ through $\tilde{p} \in \tilde{\mu}_0(\tilde{y})$ is non-singular, then each $\tilde{f}_0^n\tilde{\sigma}(\tilde{p})$ separates $\tilde{M}$ into two components which we label + and − according to the orientation on $\tilde{f}_0^n(\tilde{\mu}_0(\tilde{y}))$ induced from that on $\tilde{\mu}_0(\tilde{y})$. We define a function $T:\{\tilde{p} \in \tilde{\mu}_0(\tilde{y}): \tilde{\sigma}(\tilde{p}) \text{ is non-singular}\} \to \{+, 0, -\}$ as follows. If $\tilde{D}_u(\tilde{g}_0^n(\tilde{y}), \tilde{f}_0^n(\tilde{\sigma}(\tilde{p}))) < K$ for all $n > 0$ (i.e. if $\tilde{\sigma}(\tilde{p})$ is the desired stable leaf) then $T(p) = 0$. Otherwise, (Remark 1.1), $\tilde{g}_0^n(\tilde{y})$ is either contained in the + component of the complement of $\tilde{f}_0^n(\tilde{\sigma}(\tilde{p}))$ for all sufficiently large n or it is contained in the − component for all sufficiently large n; in the former case $T(p) = +$ and in the latter case $T(p) = -$. Note that $T^{-1}(-)$ and $T^{-1}(+)$ are open subsets of the domain, that the image of T contains both + and − and that if the domain is ordered according to the orientation on $\tilde{\mu}_0(\tilde{y})$ and the range is ordered by $+ < 0 < -$, then T is order preserving.

Let $\tilde{q}$ be the intersection of the closures of $T^{-1}(-)$ and $T^{-1}(+)$. If $\tilde{\sigma}(\tilde{q})$ is not the desired stable leaf then $\tilde{\sigma}(\tilde{q})$ contains a singularity $\tilde{z}_1$ and (Remark 1.1) there is an unstable wedge W_1^u based at $\tilde{z}_1$ that is disjoint from $\tilde{\mu}(\tilde{y})$ such that $\tilde{g}_0^n(\tilde{y}) \in \tilde{f}_0^n(E_1^u)$ for all sufficiently large n, where E_1^u is the unstable envelope for W_1^u. See figure 2.1. Let $\tilde{\mu}_1$ be the oriented unstable half leaf in W_1^u whose negative end converges to $\tilde{z}_1$. There is a map $T_1:\{\tilde{p} \in \tilde{\mu}_1: \tilde{\sigma}(\tilde{p}) \text{ is non-singular}\} \to \{+, 0, -\}$ as above. Iterating this argument we either find the desired stable leaf $\tilde{\sigma}$ or we find

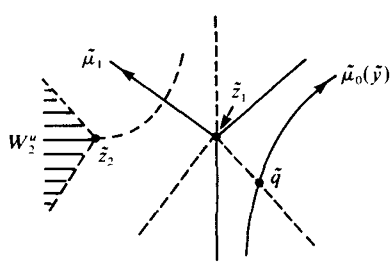

FIGURE 2.1

sequences $\{\tilde{z}_i\}$ of singularities, $\{E_i^u\}$ of unstable envelopes based at $\tilde{z}_i$ and $\{W_i^u \supset E_i^u\}$ of unstable wedges based at $\tilde{z}_i$ such that $\tilde{g}_0^n(\tilde{y}) \in \tilde{f}_0^n(E_i^u)$ for all sufficiently large n and such that $\tilde{z}_i \notin W_j^u$ for $j > i$. In particular, note that $\tilde{D}_u(\tilde{y}, \tilde{z}_i) < K$ for all i.

Let W_i^s be the stable wedge based at $\tilde{z}_i$ that contains $\tilde{y}$ and let E_i^s be the stable envelope based at $\tilde{z}_i$ that is contained in W_i^s. Then $W_i^s \cap W_i^u = \emptyset$ for all i. Since $\tilde{D}_u(\tilde{y}, \tilde{z}_i) < K$ for all i and since the set of singularities is discrete, there exists $i > 0$ such that $\tilde{D}_s(\tilde{y}, \tilde{z}_i) > K$. In this case $\tilde{y} \in W_i^s$ and the conclusions of the proposition are satisfied by $\tilde{z} = \tilde{z}_i$, $E^s = E_i^s$ and $E^u = E_i^u$.

A similar argument produces an unstable leaf $\tilde{\mu}$ such that $\tilde{D}_s(\tilde{g}_0^n(\tilde{y}), \tilde{f}_0^n(\tilde{\mu})) < K$ for all $n < 0$. If $\tilde{\mu}$ and $\tilde{\sigma}$ intersect then their intersection is a point $\tilde{x}$ whose $\tilde{f}_0$-orbit shadows the $\tilde{g}_0$-orbit of $\tilde{y}$. If $\tilde{\mu}$ and $\tilde{\sigma}$ are disjoint, choose $p \in \tilde{\mu}$ so that $\tilde{D}_u(\tilde{\sigma}(\tilde{p}), \tilde{\sigma}) \leq \tilde{D}_u(\tilde{\sigma}(\tilde{q}), \tilde{\sigma})$ for all $\tilde{q} \in \tilde{\mu}$. Then $\tilde{\sigma}(\tilde{p})$ contains a singularity $\tilde{z}$ and $\tilde{\sigma}$ is contained in an unstable wedge W^u that is disjoint from $\tilde{\mu}$. Let W^s be the stable wedge that contains $\tilde{\mu}$ and let E^s and E^u be the envelopes determined by W^s and W^u. The conclusions of the proposition are satisfied by $\tilde{z}$, E^s and E^u. □

3. *Computing entropy*

In this section we assume that g is not semi-conjugate to f and show that the entropy of g is strictly greater than that of f. We begin with some preliminary definitions and lemmas.

Since g is not semi-conjugate to f there is an orbit of g that is not globally shadowed by an orbit of f. Choose $\tilde{g}, \tilde{f}, \tilde{z}, \tilde{y}, E^s$, and E^u as in Proposition 2.1; recall that $K > \sup\{\tilde{D}_u(\tilde{f}(\tilde{x}), \tilde{g}(\tilde{x})) : \tilde{x} \in \tilde{M}\}$ and that after replacing g by an iterate of itself (and hence λ by a power of itself) if necessary, we may assume that $\lambda > 4$. Let $c_1 = 1/(\lambda - 1) < \frac{1}{3}$ and note that $1 + \lambda + \lambda^2 + \cdots + \lambda^{n-1} < c_1\lambda^n$ for all $n > 0$.

LEMMA 3.1. *There exists* $0 < c_2 < 1$ *such that if* $\tilde{D}_u(\tilde{a}, \tilde{b}) > K$ *then* $\tilde{D}_u(\tilde{g}^n(\tilde{a}), \tilde{g}^n(\tilde{b})) > c_2\lambda^n\tilde{D}_u(\tilde{a}, \tilde{b})$ for all $\tilde{a}, \tilde{b} \in \tilde{M}$.

Proof of Lemma 3.1. Repeated application of the fact that

$$\tilde{D}_u(\tilde{g}(\tilde{b}), \tilde{g}(\tilde{a})) > \tilde{D}_u(\tilde{f}(\tilde{b}), \tilde{f}(\tilde{a})) - 2K = \lambda\tilde{D}_u(\tilde{b}, \tilde{a}) - 2K$$

leads to

$$\tilde{D}_u(\tilde{g}^n(\tilde{b}), \tilde{g}^n(\tilde{a})) \geq \lambda^n(\tilde{D}_u(\tilde{a}, \tilde{b}) - 2Kc_1).$$

Choose $c_2 = \frac{1}{3}$. □

For any two points $\tilde{a}$ and $\tilde{b}$ in $\tilde{M}$, let $u(\tilde{a}, \tilde{b})$ be the number of unstable segments in an efficient path between $\tilde{a}$ and $\tilde{b}$.

LEMMA 3.2. *There exists* $C > 0$ *such that* $u(\tilde{g}(\tilde{a}), \tilde{g}(\tilde{b})) < u(\tilde{a}, \tilde{b}) + C$ *for all* $\tilde{a}, \tilde{b} \in \tilde{M}$.

Proof of Lemma 3.2. It is clear from the definition that

$$u(\tilde{g}(\tilde{a}), \tilde{g}(\tilde{b})) \leq u(\tilde{g}(\tilde{a}), \tilde{f}(\tilde{a})) + u(\tilde{f}(\tilde{a}), \tilde{f}(\tilde{b})) + u(\tilde{f}(\tilde{b}), \tilde{g}(\tilde{b})).$$

The lemma therefore follows from the fact that there is a uniform bound to the distance between $\tilde{f}(\tilde{x})$ and $\tilde{g}(\tilde{x})$ and the fact that $\tilde{f}$ carries efficient paths to efficient paths. □

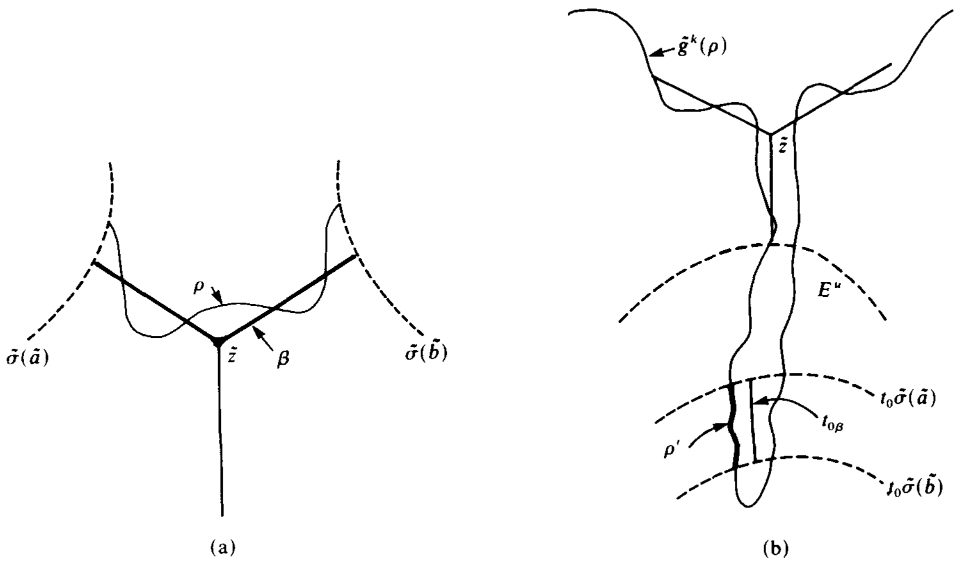

FIGURE 3.1

LEMMA 3.3. *Given $T > K$, let $\beta \subset \tilde{\mu}(\tilde{z})$ be the union of the two T-length initial segments of the unstable half leaves contained in the frontier of the stable wedge containing E^s (see figure 3.1). Denote the endpoints of β by $\{\tilde{a}, \tilde{b}\}$. Then for all sufficiently large T, there exists $k > 0$ and a covering translation t_0 such that if ρ:*

(*) *is a path connecting $\tilde{\sigma}(\tilde{a})$ to $\tilde{\sigma}(\tilde{b})$ with stable variation less than K, with $\tilde{D}_s(\tilde{x}, \beta) < K+1$ for each $\tilde{x} \in \rho$ and with interior disjoint from $\tilde{\sigma}(\tilde{a}) \cup \tilde{\sigma}(\tilde{b})$*

then there is a subpath $\rho' \subset \tilde{g}^k(\rho)$ such that $t_0^{-1}\rho'$ satisfies (*). Moreover, ρ' and the two endpoints of $\tilde{g}^k(\rho)$ are contained in three distinct unstable envelopes for $\tilde{z}$.

Proof of Lemma 3.3. For sufficiently large n, $\tilde{g}^{n-1}(E^s) \cap E^u$ is non-empty. Thus if T is sufficiently large, $\tilde{g}^{n-1}(E^s) \cup E^u$ separates $\tilde{g}^n(a)$ from $\tilde{g}^n(b)$ and since $\rho \cap \tilde{g}^{-1}(E^s) = \phi$, $\tilde{g}^n(\rho) \cap E^u$ is non-empty. A similar argument allows us to choose $n > 0$ so that for all sufficiently large T, $\tilde{g}^n(\rho) \cap E^u$ contains two points $\tilde{p}, \tilde{q} \in E^u$ satisfying $\tilde{D}_u(\tilde{p}, \tilde{q}) > K$.

As m increases $\tilde{D}_u(\tilde{g}^m(\tilde{p}), \tilde{g}^m(\tilde{q}))$ grows exponentially in m while $u(\tilde{g}^m(\tilde{p}), \tilde{g}^m(\tilde{q}))$ grows linearly in m. For sufficiently large m an efficient path γ' between $\tilde{g}^m(\tilde{p})$ and $\tilde{g}^m(\tilde{q})$ contains arbitrarily long unstable leaf segments and hence unstable leaf segments that are nearly parallel to translates of β. In particular, there is a covering translation t_0 and an unstable leaf segment $\beta' \subset N_1(t_0\beta) \cap \gamma'$ connecting $t_0\tilde{\sigma}(\tilde{a})$ to $t_0\tilde{\sigma}(\tilde{b})$. Since γ' is efficient, the stable leaves through the endpoints of β' separate $\tilde{g}^m(\tilde{p})$ from $\tilde{g}^m(\tilde{q})$ and there is an arc $\rho' \subset \tilde{g}^{m+n}(\rho)$ that connects these stable leaves but is otherwise disjoint from them. Let $k = m+n$. The stable variation of ρ' is bounded by the stable variation of $\tilde{g}^k(\rho)$ and is therefore less than K. Since the unstable leaf that contains β' separates $\tilde{g}^m(\tilde{p})$ from $\tilde{g}^m(\tilde{q})$, there exists $\tilde{x}_0 \in \tilde{g}^k(\rho)$ with $D_s(\tilde{x}_0, \beta')=0$. It follows that $\tilde{D}_s(\tilde{x}, \beta') < K$ and hence that $\tilde{D}_s(\tilde{x}, t_0\beta) < K+1$ for all $\tilde{x} \in \rho'$. □

We now assume that T is chosen as in Lemma 3.3. We say that a path $\rho \subset \tilde{M}$ and a covering translation t are an *admissible pair* if $t^{-1}(\rho)$ satisfies (*).

We say that a *path* ρ *is admissible* if it is part of an admissible pair. Note that if (ρ, t) is an admissible pair, then there is a subpath $\rho' \subset \tilde{g}(\rho)$ such that $(\rho', \tilde{g}t\tilde{g}^{-1})$ is an admissible pair. If α is an arc containing ρ, we say that the admissible pair (ρ, t) is α*-forced* if the stable leaves through the endpoints of ρ each separate the endpoints of α; we say that (ρ, t) is α*-unforced* if there is a singularity $\tilde{z}_i$ such that ρ and the two endpoints of α are contained in three distinct unstable envelopes for $\tilde{z}_i$. Thus Lemma 3.3 states that for any admissible pair $(\rho$, identity), there is a subpath $\rho' \subset \tilde{g}^k(\rho)$ such that (ρ', t_0) is a $\tilde{g}^k(\rho)$-unforced admissible pair.

For any path α, let $F_n(\alpha)$ [respectively $U_n(\alpha)$] be the maximum cardinality of a collection $\{(\rho(i), t(i))\}$ of $\tilde{g}^n(\alpha)$-forced [respectively $\tilde{g}^n(\alpha)$-unforced] admissible pairs where each $\rho(i) \subset \tilde{g}^n(\alpha)$ and the $t(i)$'s are distinct.

As a corollary to the (second paragraph of the) proof of Lemma 3.3, we have the following estimate on $F_n(\rho)$.

COROLLARY 3.4. *There exists $c_3 > 0$ such that $F_n(\rho) > c_3\lambda^n$ for all admissible paths.*

After replacing g by an iterate of itself if necessary, we may assume that the constant k of Lemma 3.3 equals 1, that $c_3\lambda > 4$ and that $c_1 + (2/c_3\lambda) < 1$. Choose $1 > d > 0$ so that $[d\lambda^j + 1] < c_3\lambda^j/2$ for all $j \geq 1$ where [ ] denotes the greatest integer function. Note that

$$\sum_{i=1}^{n} [d\lambda^{i-1}+1] \leq 1 + \sum_{i=1}^{n-1} (c_3\lambda^i/2) \leq 1 + c_1c_3\lambda^n/2 = (c_1 + 2/c_3\lambda^n)(c_3\lambda^n/2) < c_3\lambda^n/2.$$

The following lemma shows that for any admissible path $\rho \subset \tilde{M}$, the exponential growth rate of $U_n(\rho)$ is strictly larger than $\log(\lambda)$.

LEMMA 3.5. *Let ρ be an admissible path. Then $U_n(\rho) > B_n$ where (B_n) is a sequence satisfying $B_{n+1} > (\lambda + d)B_n$.*

Proof of Lemma 3.5. If $(\rho(i,j), t(i,j))$ is a $\tilde{g}^i(\rho)$-unforced admissible pair for some $i \geq 1$, then there is a unique singularity $\tilde{z}(i,j)$ such that $\tilde{g}^{n-i}(\rho(i,j))$ and the two endpoints of $\tilde{g}^n(\rho)$ are contained in three distinct envelopes of $\tilde{f}^{n-i}\tilde{z}(i,j)$ for all $n \geq i \geq 1$. The first step in the proof is an inductive construction $(i \geq 1)$ of a collection $P_i(\rho) = \{(\rho(i,j), t(i,j)) \colon 1 \leq j \leq [d\lambda^{i-1}+1]\}$ of $\tilde{g}^i(\rho)$-unforced admissible pairs such that for each $n \geq 1$, all the elements of $\{\tilde{f}^{n-i}(\tilde{z}(i,j)) \colon 1 \leq i \leq n; 1 \leq j \leq [d\lambda^{i-1}+1]\}$ are distinct. See figure 3.2. In particular, for all $1 \leq i, k \leq n$, $1 \leq j \leq [d\lambda^{i-1}+1]$, and $1 \leq l \leq [d\lambda^{k-1}+1]$, $\tilde{g}^{n-i}(\rho(i,j))$ and $\tilde{g}^{n-k}(\rho(k,l))$ do not intersect a common stable leaf.

There is no loss in assuming that ρ satisfies (*). Define $P_1(\rho) = \{(\rho(1,1), t(1,1))\}$ to be the admissible pair produced by Lemma 3.3; note that $\tilde{z}(1,1) = \tilde{z}$. We now assume that $P_1(\rho), \ldots, P_n(\rho)$ have been defined. Corollary 3.4 states that

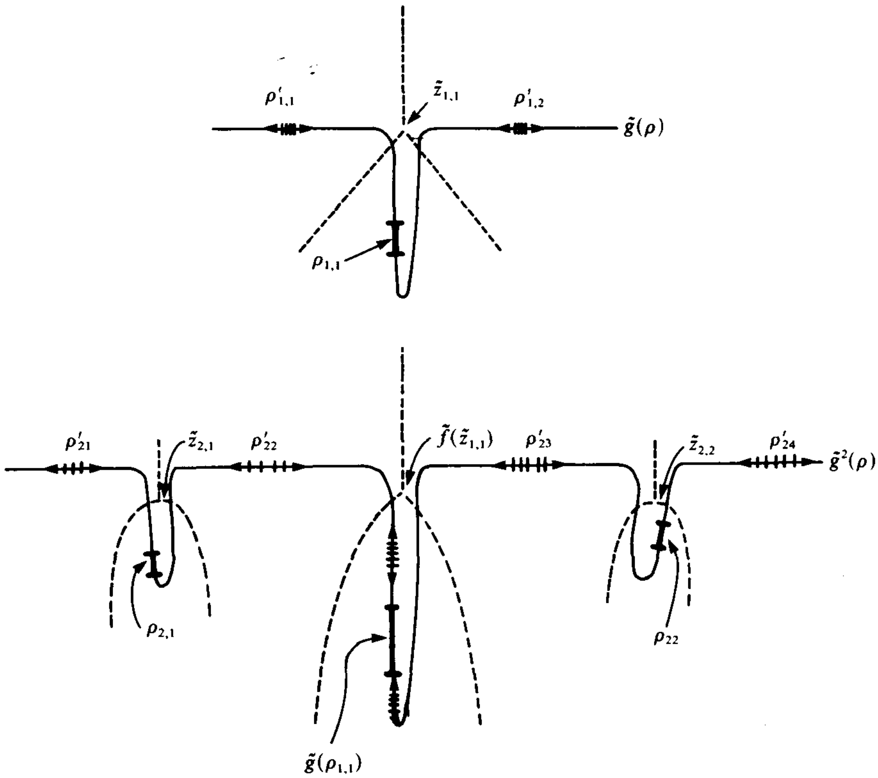

FIGURE 3.2

$F_n(\rho) > c_3\lambda^n$. Since

$$\sum_{i=1}^{n} [d\lambda^{i-1}+1] < c_3\lambda^n/2,$$

there are $[d\lambda^n+1] < c_3\lambda^n/2$ $\tilde{g}^n(\rho)$-forced admissible pairs $\{(\rho'(n,j), t'(n,j))\}$ such that

$$\{t'(n,j)\tilde{z}\} \cap \{\tilde{f}^{n-i}(\tilde{z}(i,l)) \colon 1 \le i \le n;\ 1 \le l \le [d\lambda^{i-1}+1]\} = \emptyset.$$

Applying Lemma 3.3 to $t'(n,j)^{-1}\rho'(n,j)$, we obtain a $\tilde{g}t'(n,j)^{-1}\tilde{g}^n(\rho)$-unforced admissible pair $(\alpha(j), t_0)$. Define

$$(\rho(n+1,j), t(n+1,j)) = (t''(n,j)\alpha(j), t''(n,j)t_0)$$

where $t''(n,j)$ is the covering translation $\tilde{g}t'(n,j)\tilde{g}^{-1}$, and note that $\tilde{z}(n+1,j) = \tilde{f}t'(n,j)\tilde{z}$. This completes the construction of $P_{n+1}(\rho)$.

We next define collections Q_n $(n \ge 1)$ of $\tilde{g}^n$ (ρ)-unforced admissible paths by $Q_0(\rho) = \rho$ and $(n \ge 1) Q_n(\rho) = \bigcup P_{i(k)}(P_{i(k-1)}(\ldots P_{i(1)}(\rho)))$ where the union is taken over all ordered k-tuples $(1 \le k \le n)$ of positive integers $i(1), \ldots, i(k)$ whose sum equals n and where P_i of a union is the union of the P_i's. Thus for each $(\beta, t) \in Q_n(\rho)$ there is a sequence of admissible paths

$$\tau_{i(1)} \in P_{i(1)}(\rho),\ \tau_{i(2)} \in P_{i(2)}(\tau_{i(1)}), \ldots, \tau_{i(k)} = \beta \in P_{i(k)}(\tau_{i(k-1)}).$$

If (β', t') is another element of $Q_n(\rho)$ with corresponding τ_j''s and if $\tau_{i(1)} \neq \tau'_{j(1)}$, then $\beta \subset \tilde{g}^{n-i(1)}\tau_{i(1)}$ and $\beta' \subset \tilde{g}^{n-j(1)}\tau'_{j(1)}$ do not intersect a common stable leaf. In particular, $t \neq t'$. A similar argument shows that all the elements of $Q_n(\rho)$ determine distinct covering translations.

Note that $Q_n(\rho)$ can also be defined recursively by

$$Q_n(\rho) = \bigcup_{j=0}^{n-1} Q_j \circ P_{n-j}(\rho);$$

it follows that

$$\text{Card}\,(Q_n) = \sum_{j=0}^{n-1} (\text{Card}\,(Q_j) \times [d\lambda^{n-j-1} + 1]).$$

Define $B_n = \sum_{j=0}^{n-1} (\text{card}\, Q_j)\, d\lambda^{n-j-1}$. Then $U_n(\rho) \geq \text{card}(Q_n) \geq B_n$ and

$$\begin{aligned} B_{n+1} &= \lambda B_n + d(\text{card}\, Q_n) \\ &\geq \lambda B_n + dB_n \\ &= (\lambda + d)B_n. \end{aligned}$$ □

Remark. The covering translations that occur in the admissible pairs counted in the proof of Lemma 3.5 are independent of the exact choice of ρ and depend only on the fact that ρ satisfies $(*)$.

The following corollary shows that the exponential growth rate of $N_n(g)$ is strictly greater than $\log \lambda$. Since there exists $\delta > 0$ such that $N_n(g)$ is an (n, δ)-separated set for all $n > 0$, the topological entropy of g is strictly greater than $\log \lambda$, the topological entropy of f.

COROLLARY 3.6. *There exists $c > 0$ such that $N_n(g) > c(\lambda + d)^n$.*

Proof of Corollary 3.6. Choose an arc α that satisfies $(*)$. Let (ρ, t) be one of the admissible pairs counted in the proof of Lemma 3.5 and let $\tilde{h} = t^{-1}\tilde{g}^n$; recall that $\rho \subset \tilde{g}^n(\alpha)$. It suffices to show that $\tilde{h}$ has fixed points and that there is a uniform bound (i.e. a bound independent of n) to the number of such $\tilde{h}$ that can be mutually conjugate.

There is a subarc $\alpha_1 = (t^{-1}\rho)$ of $\tilde{h}(\alpha)$ that satisfies $(*)$. Similarly, for all $k \geq 1$ there are subarcs α_k of $\tilde{h}^k(\alpha)$ that satisfy $(*)$. In particular, there exists a constant K_0 and $\tilde{x}_k \in \alpha$ such that $\tilde{D}(\tilde{h}^k(\tilde{x}_k), \tilde{x}_k) < K_0$ for all $k \geq 1$.

We claim that there exists $K_1 > 0$ such that $\tilde{D}(\tilde{h}^i(\tilde{x}_k), \tilde{x}_k) < K_1$ for $0 \leq i \leq k$ and all $k \geq 1$. Since $\tilde{D}_s(\tilde{h}^i(\tilde{x}_k), \tilde{x}_k)$ is uniformly bounded for all $i \geq 1$ and all $\tilde{x}_k$, it suffices to find $K_2 > K_0$ so that if $\tilde{D}_u(\tilde{h}^i(\tilde{x}_k), \tilde{x}_k) > K_2$ then $\tilde{D}_u(\tilde{h}^{i+1}(\tilde{x}_k), \tilde{x}_k) > K_2$. The existence of K_2 follows from the fact that

$$\begin{aligned} \tilde{D}_u(\tilde{h}^{i+1}(\tilde{x}_k), \tilde{x}_k) &\geq \tilde{D}_u(\tilde{h}^{i+1}(\tilde{x}_k), \tilde{h}(\tilde{x}_k)) - \tilde{D}_u(\tilde{h}(\tilde{x}_k), \tilde{x}_k) \\ &\geq c_2\lambda^n \tilde{D}_u(\tilde{h}^i(\tilde{x}_k), \tilde{x}_k) - C_1 \end{aligned}$$

where C_1 is greater than $\tilde{D}_u(\tilde{h}(\tilde{x}), \tilde{x})$ for all $\tilde{x} \in \alpha$ and $\tilde{D}_u(\tilde{h}^i(\tilde{x}_k), \tilde{x}_k) > K$.

It follows immediately that the entire $\tilde{h}$-orbit of any accumulation point $\tilde{x}$ of the set $\{\tilde{h}^{[k/2]}(\tilde{x}_k): k \geq 1\}$ is contained in the K_1-neighborhood of α. The Brouwer Translation Theorem (see for example [**F2**]) implies that Fix $(\tilde{h})$ is non-empty.

Moreover, for any fixed point $\tilde{y}$ of $\tilde{h}$, $\tilde{D}(\tilde{h}^i(\tilde{y}), \tilde{h}^i(\tilde{x}))$ is uniformly bounded for all $i \in \mathbb{Z}$. Thus $\tilde{D}(\tilde{y}, \tilde{x}) < 2K$ and Fix $(\tilde{h})$ is contained in the $2K + K_1$ neighborhood of α. As there are only finitely many covering translations that do not move this neighborhood of α entirely off of itself, there is a uniform bound to the number of $\tilde{h}$'s that are mutually conjugate. □

4. *Extensions and counterexamples*

Fathi [**F1**] generalized the results of [**H**] by replacing the hypothesis that $g \simeq f$ with the hypothesis that g is a homeomorphism of a compact metric space N for which there exists a non-nullhomotopic map $\alpha : N \to M$ satisfying $\alpha \circ g \simeq f \circ \alpha$. The conclusions are the same as in [**H**] except that $\pi : Y \to M$ is homotopic to $\alpha \mid Y$ rather than the inclusion of Y into M. In this section we consider the extent to which Theorem 0.1 applies to this context. In particular, we assume throughout this section that f, g and α are as above and that the entropy of g equals that of f.

Fathi also showed that if N is a closed surface and g is pseudo-Anosov, then $Y = N$ and in fact $\pi : N \to M$ is a branched covering. (His proof uses the fact that a pseudo-Anosov homeomorphism supports a unique measure of maximal entropy and the support of this measure is the entire surface.) Theorem 0.1 therefore implies that if N is a closed surface and g is homotopic to a pseudo-Anosov homeomorphism, then g is semi-conjugate to f. As Example 4.1 shows, the hypothesis that g determines a pseudo-Anosov mapping class is essential.

If $\tilde{\alpha} : \tilde{N} \to \tilde{M}$, $\tilde{g} : \tilde{N} \to \tilde{N}$ and $\tilde{f} : \tilde{M} \to \tilde{M}$ are lifts to the universal covers such that $\tilde{\alpha} \circ \tilde{g}$ and $\tilde{f} \circ \tilde{\alpha}$ are equivariantly homotopic and if α is homotopic to a semi-conjugating map π, then for all $\tilde{y} \in \tilde{N}$ there exists $\tilde{x} \in \tilde{M}$ and $C > 0$ such that $\tilde{D}(\tilde{\alpha}\tilde{g}^n(\tilde{y}), \tilde{f}^n(\tilde{x})) < C$ for all $n \in \mathbb{Z}$. We will use this as a criterion for proving that for certain g and α, there are no maps $\pi \simeq \alpha$ that semi-conjugate g to f.

Example 4.1. Fix a hyperbolic structure on M and let Λ^s and Λ^u be the geodesic laminations determined by the stable and unstable foliations F^s and F^u determined by f. There is a standard construction (see for example [**H–T**] or [**M**]) of a homeomorphism $h : M \to M$ that preserves Λ^s and Λ^u and that is semi-conjugate to f by a map $p \simeq$ identity. Choose a singularity z of F^s and F^u. We may assume that $p^{-1}(z) \subset \text{Fix}(h)$ is a hexagon (or a $2n$-gon for $n > 3$ if z is a more complicated singularity) and that for each point v in a stable [respectively unstable] leaf initiating at z, $p^{-1}(v)$ is an arc of an unstable [respectively stable] leaf. See figure 4.1.

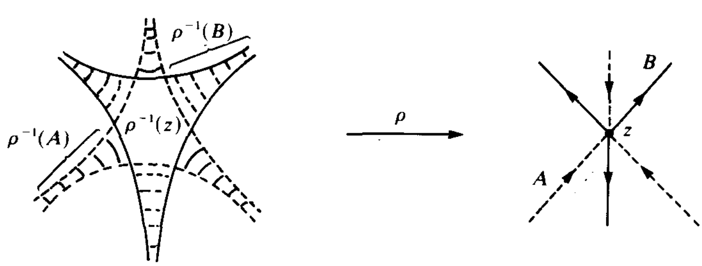

FIGURE 4.1

https://doi.org/10.1017/S0143385700004703 Published online by Cambridge University Press

Let M' be the surface obtained from M by removing the interior of $p^{-1}(z)$ and let N be the closed surface obtained by gluing together two copies M_1' and M_2' where $\partial M_1'$ and $\partial M_2'$ are identified by a homeomorphism that carries $\text{cl}\,(p_1^{-1}(A)) \cap \partial M_1'$ to $\text{cl}(p_2^{-1}(B)) \cap \partial M_2'$ where A and B are non-adjacent stable and unstable leaf segments initiating at z. See figure 4.2 where C is the image in N of $\text{cl}\,(p_1^{-1}(A)) \cap \partial M_1'$.

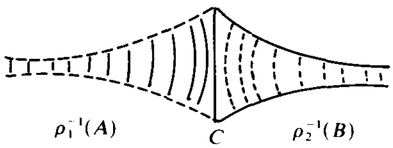

FIGURE 4.2

Then $h_1 : M_1' \to M_1'$ and $h_2 : M_2' \to M_2'$ fit together to define a homeomorphism $H : N \to N$ and $p_1 : M_1' \to M$ and $p_2 : M_2' \to M$ fit together to define a map $\alpha : N \to M$ such that $f \circ \alpha = \alpha \circ H$. There is a foliation (see figure 4.2) of $U = p_1^{-1}(A) \cup p_2^{-1}(B) \cup C$ where the leaves are given by

$$\{p_1^{-1}(v) \colon v \in A\} \cup \{p_2^{-1}(w) \colon w \in B\} \cup \{C\}.$$

Let X be a vector field on N supported on a neighborhood of C in U such that X is transverse to the foliation and points in the direction of $p_2^{-1}(B)$. Define $g = \phi \circ H$ where ϕ is the time one map of the flow generated by X.

Since g and H agree on their non-wandering sets, the entropy of g equals that of f. On the other hand, there are lifts

$$\tilde{z} \in \tilde{M}, \tilde{f} : \tilde{M} \to \tilde{M}, \tilde{C} \subset \tilde{N}, \tilde{H} : \tilde{N} \to \tilde{N}, \tilde{g} : \tilde{N} \to \tilde{N} \text{ and } \tilde{\alpha} : \tilde{N} \to \tilde{M}$$

such that $\tilde{f}$ fixes $\tilde{z}$, $\tilde{H}$ fixes $\tilde{C}$, $\tilde{g}$ is equivariantly homotopic to $\tilde{H}$ and $\tilde{\alpha} \circ \tilde{H} = \tilde{f} \circ \tilde{\alpha}$. Let $\tilde{A}, \tilde{B} \subset \tilde{M}$ be the lifts of A and B that initiate at $\tilde{z}$. Then for each $\tilde{y} \in \tilde{C}$ there exists $N > 0$ such that $\{\alpha\tilde{g}^n(\tilde{y}) \colon n > N\}$ agrees with an $\tilde{f}$-orbit in $\tilde{B}$ and $\{\alpha\tilde{g}^{-n}(\tilde{y}) \colon n > N\}$ agrees with an $\tilde{f}$-orbit in $\tilde{A}$. In particular,

$$\sup \{\tilde{D}(\tilde{f}^n(\tilde{x}), \alpha\tilde{g}^n(\tilde{y})) \colon n \in \mathbb{Z}\} = \infty \text{ for all } \tilde{x} \in \tilde{M}$$

and g is not semi-conjugate to f by a map that is homotopic to α. □

Our final example shows that Theorem 0.1 may fail in the general context even if α is a homotopy equivalence.

Example 4.2. Let $N = M \times [0, 1]$ and let G_t be an isotopy of f to itself such that $G_{1/2}$ contains a fixed point P that is not Nielsen equivalent to any fixed point of f. Choose a function $u : N \to [0, 1]$ such that $u^{-1}(0) = M \times \{0, 1\} \cup P \times \{\frac{1}{2}\}$ and let $H(y, t)$ be the time one map of the flow generated by the vector field $u \cdot \partial/\partial t$. Then $g(y, t) = H \circ (G_t(y), t)$ is a homeomorphism of N with non-wandering set $M \times \{0, 1\} \cup P \times \{\frac{1}{2}\}$ and therefore with entropy equal to that of f. Define α equal to the projection of N onto M. Choose lifts $\tilde{\alpha}, \tilde{f}, \tilde{g}$ and $\tilde{P}$ such that $\tilde{\alpha} \circ \tilde{g}$ and $\tilde{f} \circ \tilde{\alpha}$ are equivariantly homotopic and such that $\tilde{g}$ fixes $\tilde{P} \times \{\frac{1}{2}\}$. The Brouwer translation theorem and the choice of $\tilde{P}$ imply that no orbit of $\tilde{f}$ is bounded. Since

$\tilde{\alpha}\tilde{g}^n(\tilde{P}\times\{\frac{1}{2}\})=\tilde{P}$ for all $n\in\mathbb{Z}$, g is not semi-conjugate to f by any map that is homotopic to α. □

REFERENCES

[F1] A. Fathi. Homotopical stability of pseudo-Anosov maps. Preprint.

[F2] A. Fathi. An orbit closing proof of Brouwer's lemma on translation arcs. Preprint.

[F-L-P] A. Fathi, F. Laudenbach & V. Poenaru. Travaux de Thurston sur les surfaces. *Asterisque* **66–67** (1979).

[Fr] J. Franks. Anosov diffeomorphisms. *Proceedings of the Symposium in Pure Mathematics* **14** (1968) 61-94.

[H] M. Handel. Global shadowing of pseudo-Anosov diffeomorphisms. *Ergod. Th. & Dynam. Sys.* **5** (1985), 373-377.

[H-T] M. Handel & W. Thurston. New proofs of some results of Nielsen. *Adv. Math.* **56** (1985), 173-191.

[M] R. T. Miller. Nielsen's viewpoint on geodesic laminations. *Adv. in Math.* **45** (1982), 189-212.

[T] W. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. Preprint.