Arnold conjecture for surface homeomorphisms

S. Matsumoto

Department of Mathematics, College of Science and Technology, Nihon University, 1-8 Kanda-Surugadai, Chiyoda-ku, Tokyo, 101-8308, Japan

Received 19 December 1997; received in revised form 25 May 1998

Abstract

The purpose of this paper is to show the Arnold conjecture for homeomorphisms of closed and oriented surfaces. © 2000 Published by Elsevier Science B.V. All rights reserved.

Keywords: Fixed point index; Rotation vector; Homotopy translation arc; Chain recurrence

AMS classification: 58F20; 58F11; 58F05

1. Introduction

In dimension two, the Arnold conjecture concerns the fixed points of area preserving diffeomorphisms isotopic to the identity with vanishing mean rotation vector. It was first solved by Floer [6] and Sikorav [16] using variational methods. But since the problem is purely topological, it is natural to ask for a geometric proof, not involved in analysis of infinite-dimensional spaces. This was carried out by Franks in [7] for $C^1$ diffeomorphisms.

The purpose of these notes is to remark that, with some modifications, Franks’ argument can be made applicable even for homeomorphisms. However it is almost impossible to make our remark understandable by scattering around the points of changes of the argument in [7]. Also a result of Handel [11] exposed in [7] plays a crucial role in this development, and additional accounts of it including fundamental facts on proper homotopy theory are indispensable for our purpose. These reasons have determined us to write down this self-contained notes.

Let $M$ be a closed oriented surface, and let $f$ be a homeomorphism of $M$, isotopic to the identity. Given an isotopy $f_t : id \simeq f$, one can specify a lift $\tilde{f}$ of $f$ to the universal covering space $\tilde{M}$ of $M$ by lifting the isotopy $f_t$ so as to start at the identity of $\tilde{M}$ and fixing $\tilde{f}$ to be the end of the lifted isotopy. A lift of $f$ so obtained is called admissible.
Thus admissible lifts of $f$ correspond one to one to a homotopy class of isotopies of $M$ joining the identity to $f$.

If the surface $M$ is the 2-torus $T^2$, then any lift of $f$ is admissible. On the contrary, if the surface $M$ has genus $> 1$, then as is well known [10] the isotopy $f_t$ joining the identity to $f$ is unique up to homotopy. Therefore there is exactly one admissible lift, which we call the canonical lift of $f$.

Given any lift $\tilde{f}$ of $f$, the projected image of the fixed points of $\tilde{f} : \tilde{M} \to \tilde{M}$ to the surface $M$ is denoted by $\text{Fix}(\tilde{f})$. The fixed point set $\text{Fix}(f)$ of $f$ is decomposed into a disjoint union of such subsets $\text{Fix}(\tilde{f} : f)$, where $\tilde{f}$ runs through the set of lifts of $f$, admissible or not. If the genus of $M$ is greater than one and if the lift $\tilde{f}$ is the canonical lift of a homeomorphism $f : M \to M$ isotopic to the identity, then the set $\text{Fix}(f ; \tilde{f})$ is denoted by $\text{Fix}_c(f)$ and is called the contractible fixed point set. Notice that for $p \in \text{Fix}_c(f)$, the locus of $p$ by the isotopy joining the identity to $f$ forms a loop which is contractible in $M$.

If a homeomorphism $f$ of a closed oriented surface $M$ isotopic to the identity keeps the Lebesgue probability measure $m$ invariant, and if an admissible lift $Qf$ of $f$ is specified, then the mean rotation vector $R(m ; Qf)$ is defined as a homology class in $H_1(M; \mathbb{R})$. If further the genus of $M$ is greater than one, then since the admissible lift is unique, this class depends only on the homeomorphism $f$ and is denoted by $R(m ; f)$. For more details, see [7,14] or Section 5 of this paper.

Here is a main result of this paper.

**Theorem 1.** Assume that $f$ is a homeomorphism of a closed oriented surface $M$, isotopic to the identity and keeping the Lebesgue measure $m$ invariant. If the mean rotation vector $R(m ; \tilde{f})$ of an admissible lift $\tilde{f}$ vanishes, then the set $\text{Fix}(f ; \tilde{f})$ is nonempty and if further it is a finite set, then it contains at least two points of fixed point index one.

If the surface $M$ is the 2-sphere $S^2$, then the theorem asserts the existence of at least two fixed points of index one for any orientation and area preserving homeomorphism $f : S^2 \to S^2$, provided the fixed points are finite in number. If the surface $M$ has genus $> 1$, then the theorem concerns the contractible fixed point set $\text{Fix}_c(f)$ for an area preserving homeomorphism $f : M \to M$ isotopic to the identity, with vanishing mean rotation vector $R(m ; f)$.

In any case this theorem implies, via the Lefschetz and Nielsen fixed point theory [12], the Arnold conjecture which asserts under the same condition as Theorem 1 the existence of at least three fixed points in $\text{Fix}(f ; \tilde{f})$ if the genus $g$ of $M$ is nonzero, and $2g + 2$ if further all the fixed points of $\text{Fix}(f ; \tilde{f})$ are isolated and of index $\pm 1$.

Note that by a result of Pelikan and Slaminka [15], the fixed point index of an isolated fixed point of an area and orientation preserving homeomorphism of a surface is always $\leq 1$. Therefore in the sequel, we aim to find out two contractible fixed points of positive index.

The plan of these notes is as follows. In Section 2, we prepare necessary facts about the homotopy of a homeomorphism of $\mathbb{R}^2$, relative to a finite union of orbits. The complement is uniformized by a Fuchsian group of the first kind, and the hyperbolic geometry will play
an important role. Section 3 is devoted to the proof of the Handel fixed point theorem, concerning homeomorphisms of $\mathbb{R}^2$.

In Section 4, we generalize a fixed point theorem due to Franks [7] to homeomorphisms of $M$. This is done by adding a piece of Nielsen and Thurston theory to the original argument of [7]. Section 5 is devoted to the proof of Theorem 1.

2. Homeomorphisms of $\mathbb{R}^2$ and hyperbolic geometry

In this section, we prepare fundamental facts about the homotopy of a homeomorphism of $\mathbb{R}^2$ relative to a finite union of its orbits. We closely follow [11]. Throughout this section, $\varphi$ is to be an orientation preserving homeomorphism of $\mathbb{R}^2$. Let $O$ be a discrete infinite subset of $\mathbb{R}^2$ which is invariant by $\varphi$. Thus any compact subset of $\mathbb{R}^2$ intersects $O$ in a finite set of points. Let us endow the complement $S = \mathbb{R}^2 \setminus O$ with a complete hyperbolic structure. Thus $S$ is isometric to $\mathbb{H}^2/\Gamma$, where $\mathbb{H}^2$ is the Poincaré upper half-plane and $\Gamma$ is an infinitely generated Fuchsian group. We identify $S = \mathbb{H}^2/\Gamma$ and denote the canonical projection by $\pi : \mathbb{H}^2 \to S$. Denote by $\text{Cl}(\mathbb{H}^2)$ the union of $\mathbb{H}^2$ and the circle at infinity $\partial \mathbb{H}^2$. The space $\text{Cl}(\mathbb{H}^2)$ is endowed with a topology so that it is homeomorphic to a compact disc.

We further assume that $\Gamma$ is of the first kind, i.e., the limit set is the whole circle at infinity $\partial \mathbb{H}^2$. It is well known that such a hyperbolic structure exists on the surface $S$.

Let $A_v (v = 0, 1, 2, \ldots)$ be an arbitrary family of closed discs in $\mathbb{R}^2$ such that
(1) the boundary $\lambda_v = \partial A_v$ is a closed geodesic in $S$,
(2) $A_v \subset \text{Int} A_{v+1}$, $\forall v$,
(3) $A_0$ contains exactly two points of $O$,
(4) $A_{v+1} \setminus A_v$ ($\forall v$) contains exactly one point of $O$, and
(5) the set $O$ is contained in the union $\bigcup_v A_v$.

For any such family of discs, we can show the following.

(6) $\bigcup_v A_v = \mathbb{R}^2$.

The proof goes as follows. Any boundary point of $R = \bigcup_v A_v$ is an accumulation point of distinct geodesics $\lambda_v$. Therefore the boundary of $R$, if nonempty, is a union of complete geodesics of $S$. The same is true for the inverse image $\pi^{-1}(R \cap S)$, which is connected since the inclusion $R \cap S \hookrightarrow S$ is a homotopy equivalence.

Now the closure in $\text{Cl}(\mathbb{H}^2)$ of the inverse image $\pi^{-1}(R \cap S)$, being invariant by the action of $\Gamma$, must contain the limit set of $\Gamma$ which coincides with $\partial \mathbb{H}^2$. This shows that the inverse image must be the whole of $\mathbb{H}^2$, completing the proof of (6).

Notice that (6) implies that the family of the closed geodesics $\lambda_v$ are divergent in $\mathbb{R}^2$, i.e., any compact set of $\mathbb{R}^2$ intersects only finitely many of the closed geodesics.

Let us make some definitions useful in the sequel. A simple curve $c : \mathbb{R} \to \mathbb{H}^2$ is called a proper cross cut if the limits $\lim_{t \to \infty} c(t)$ and $\lim_{t \to -\infty} c(t)$ exist in $\partial \mathbb{H}^2$ and are distinct. The limits are called end points of $c$, and the subset of $\partial \mathbb{H}^2$ they form is denoted by $\partial c$. We fix once and for all a base point $x_0$ of $\mathbb{H}^2$, and consider only those proper cross cuts which
do not pass through $x_0$. The connected component of the complement $\mathbb{H}^2 \setminus c$ which does not contain $x_0$ is denoted by $D(c)$.

A family $\mathcal{C} = \{c_j\}_{j=1,2,\ldots}$ of disjoint proper cross cuts is called a proper chain if, denoting $D_j = D(c_j)$, the following conditions are satisfied.

1. $D_{j+1} \subset D_j$ ($\forall j$).
2. $\bigcap_j D_j = \emptyset$.
3. $\text{diam}(\partial c_j) \to 0$.

Notice that conditions (2) and (3) are equivalent to saying that the intersection $\bigcap_j \text{Cl}(D_j)$ is a singleton in $\partial \mathbb{H}^2$. This point is called the impression of the proper chain $\mathcal{C}$.

Given a point $a \in \partial \mathbb{H}^2$, $\mathcal{C}$ is called a proper chain surrounding $a$ if $\forall j$, $a$ is contained in the closure of $D_j$ in $\text{Cl}(\mathbb{H}^2)$ and is not an end point of $c_j$. Of course if $\mathcal{C}$ is a proper chain surrounding $a$, then the impression of $\mathcal{C}$ is $a$, but the converse is not always true.

Now let us embark upon the study of fundamental properties of a proper curve and geodesic of $S$. In the following proposition by a curve or geodesic is meant a mapping from $\mathbb{R}$. As usual a curve or a homotopy of curves is called proper if the inverse image of a compact subset of $S$ is compact. Notice that an arbitrary lift to the universal covering space $\mathbb{H}^2$ of a proper curve or homotopy is again proper, although the converse is not true in general.

**Proposition 2.1.**

1. Let $\tilde{c}$ be an arbitrary lift to the universal covering space $\mathbb{H}^2$ of a proper curve $c$ in $S$. Then the limits $\lim_{t \to -\infty} \tilde{c}(t)$ and $\lim_{t \to \infty} \tilde{c}(t)$ exist (in $\partial \mathbb{H}^2$).
2. Let $c$ be a proper curve in $S$, properly homotopic to a proper geodesic $\gamma$. If $c$ is simple, then $\gamma$ is also simple.
3. Let $c_i$ be a proper curve in $S$, properly homotopic to a proper geodesic $\gamma_i$ ($i = 1, 2$). If $\gamma_1$ and $\gamma_2$ intersect, then $c_1$ and $c_2$ intersect.
4. Let $c$ be a proper simple curve in $S$ such that no component of the complement $S \setminus c$ is homeomorphic to an open disc. Then $c$ is properly homotopic to a proper geodesic.
5. If two proper geodesics in $S$ are properly homotopic, then they coincide (up to parametrization).

**Proof.** (1) We are concerned only with the case $t \to \infty$. First of all, assume that $c$ tends to a cusp in $S$ as $t \to \infty$. Let $\delta_t$ be horocycles in $S$ which converge to the cusp, $t_i$ the maximal value of $t$ such that $c(t) \in \delta_t$, and $\tilde{\delta}_t$ a lift of $\delta_t$ which passes through $\tilde{c}(t_i)$. Then $\tilde{c}(t_i, \infty)$ is contained in the horoball surrounded by $\tilde{\delta}_t$. This shows that all the $\tilde{\delta}_t$’s are horocycles at the same point of $\partial \mathbb{H}^2$, completing the proof.

Assume that $c$ does not tend to a cusp as $t \to \infty$. Choose a number $v_0$ so that $c(0)$ is contained in $\text{Int}(A_{v_0})$. Then the restriction of $c$ to $[0, \infty)$ intersects the curve $\lambda_v = \partial A_v$ for any $v \geq v_0$. Furthermore since the curve $c$ is proper, there exists $t_v > 0$ such that $c(t_v)$ lies in $A_{v+1} \setminus A_v$ and that $c([t_v, \infty))$ is contained in $S \setminus A_v$. Clearly we have $t_v \uparrow \infty$. 

194  
Choose the base point of $\mathbb{H}$ to be $\tilde{c}(0)$. Let us choose a lift $\tilde{\lambda}_v$ of $\lambda_v$ for any $v \geq v_0$ so that the family $\{\tilde{\lambda}_v\}$ forms a proper chain. For $v = v_0$, there exists a lift $\tilde{\lambda}_{v_0}$ which separates $\tilde{c}([t_{v_0}, \infty))$ from $\tilde{c}(0)$. That is, the domain $D(\tilde{\lambda}_{v_0})$ contains $\tilde{c}([t_{v_0}, \infty))$. After the lift $\tilde{\lambda}_v$ is chosen, put $\tilde{\lambda}_{v+1}$ to be a lift which separates $\tilde{\lambda}_v$ from $\tilde{c}([t_{v+1}, \infty))$. Then $\tilde{c}([t_{v+1}, \infty))$ is contained in $D(\tilde{\lambda}_{v+1})$, and $D(\tilde{\lambda}_{v+1})$ is contained in $D(\tilde{\lambda}_v)$.

The proof will be complete once we show that $\tilde{\lambda}_v$ constitutes a proper chain. If not, $\tilde{\lambda}_v$ must converge to a geodesic in $\mathbb{H}^2$. But then there is a compact set $K$ in $\mathbb{H}^2$ which intersects infinitely many $\tilde{\lambda}_v$, and thus the projected image of $K$ in $S$ must intersect infinitely many $\lambda_v$. A contradiction.

(2)-(3) Let $h : c \simeq \gamma$ be a proper homotopy, and choose an arbitrary lift $\tilde{h} : \tilde{c} \simeq \tilde{\gamma}$. An argument analogous to the one above shows that $\lim_{t \to -\infty} \tilde{\gamma}(t) = \lim_{t \to -\infty} \tilde{c}(t)$ and $\lim_{t \to \infty} \tilde{\gamma}(t) = \lim_{t \to \infty} \tilde{c}(t)$. The details are left to the readers.

Thus if $c$ is simple, the two end points of any lift of $c$ are not separated in $\partial \mathbb{H}^2$ by the two end points of another lift of $c$. This shows that $\gamma$ is simple. The same observation shows also (3).

(4) Let $\tilde{c}$ be a lift of $c$. Once the two end points of $\tilde{c}$ are shown to be distinct, then there exists a geodesic $\tilde{\gamma}$ with the same end points. Clearly there exists a homotopy $\tilde{h}$ joining $\tilde{c}$ to $\tilde{\gamma}$ which keeps the end points at infinity fixed. Consider the projected image $h$ of $\tilde{h}$ which is a homotopy joining $c$ to the projected image $\gamma$ of $\tilde{\gamma}$. Dividing the argument into two cases according as $c$ tends to a cusp or not as $t \to \infty$, the observation we have made in the proof of (1) clearly shows that $h$ is a proper homotopy. Notice that $\tilde{h}$ is so chosen that it does not approach $\partial \mathbb{H}^2$ except near the end points of $\tilde{c}$.

Assume for contradiction that the end points of $\tilde{c}$ are the same. Notice that the family $\{g(\tilde{c}) \mid g \in \Gamma\}$ is disjoint and divergent, i.e., any compact subset of $\mathbb{H}^2$ intersects finitely many elements of the family. Let $D$ be the connected component of $\mathbb{H}^2 \setminus \tilde{c}$ whose closure intersects $\partial \mathbb{H}^2$ at one point.

First let us consider the case where there exists $g \in \Gamma$ such that $g(\tilde{c})$ is contained in $D$. By the divergency of the family, there exists an outermost one, which we denote again by $g(\tilde{c})$. Let $A$ be the domain bounded by $\tilde{c}$ and $g(\tilde{c})$. Then any translate of $A$ by a non-trivial element of $\Gamma$ is disjoint from $A$. That is, the image by $\pi$ of the closure of $A$ is an embedded open annulus in $S$. Clearly the embedding is proper. That is, $S$ must be an annulus. A contradiction.

In the remaining case, the domain $D$ is disjoint from any of its translates. Thus the projected image $\pi(D)$ is a component of $S \setminus c$ homeomorphic to an open disc, contrary to the hypothesis on $c$.

(5) The proof is left to the reader. \qed

Recall that $S$ is the complement of a discrete infinite subset $\mathcal{O}$ invariant by an orientation preserving homeomorphism $\varphi$ of $\mathbb{R}^2$. The homeomorphism $\varphi$ induces a homeomorphism of $S$, denoted by the same letter. For any proper simple geodesic $\gamma$, the image $\varphi(\gamma)$ satisfies the condition of (4). Therefore by (4) and (5), there exists a unique geodesic properly homotopic to $\varphi(\gamma)$, which is denoted by $\varphi_\gamma$. Given a lift $\tilde{\psi}$ (respectively $\tilde{\gamma}$) of $\varphi$ (respectively $\gamma$), let us denote by $\tilde{\varphi}_\gamma$ (respectively $\tilde{\varphi}_\gamma$) the lift of $\tilde{\varphi}$ (respectively $\tilde{\varphi}$) obtained by lifting a proper
homotopy from \( \varphi(\gamma) \) to \( \varphi_{\ast}\gamma \) so as to start at \( \hat{\varphi}(\hat{\gamma}) \). Thus \( \hat{\varphi}_{\ast}\gamma \) is a geodesic in \( \mathbb{H}^2 \) with the same end points as \( \hat{\varphi}(\hat{\gamma}) \).

Also the following notations will be used in the next section. For a finite union \( \bigcup_i \gamma_i \) of disjoint proper simple geodesics \( \gamma_i \), we denote

\[
\varphi_{\ast}\left( \bigcup_i \gamma_i \right) = \bigcup_i \varphi_{\ast}\gamma_i .
\]

and if \( \bigcup_i \gamma_i \) bounds a domain \( W \), we denote by \( \hat{\gamma} \) the domain bounded by \( \varphi_{\ast}\gamma_i \). (When the boundary of \( W \) is a single proper simple geodesic, \( \hat{\gamma} \) is determined by the orientation of the boundary in an obvious manner.)

Let \( \hat{\psi} : \mathbb{H}^2 \to \mathbb{H}^2 \) be an arbitrary lift of \( \psi : S \to S \).

**Proposition 2.2.** The lift \( \hat{\psi} \) extends to a homeomorphism of \( \text{Cl}(\mathbb{H}^2) \).

**Proof.** In order to show the proposition, we only need to prove that \( \hat{\psi} \) has a continuous extension to \( \partial \mathbb{H}^2 \), since the same conclusion for \( \hat{\psi}^{-1} \) implies that the extension is a homeomorphism of \( \text{Cl}(\mathbb{H}^2) \). This is equivalent to the following claim.

**Claim.** For any point \( a \in \partial \mathbb{H}^2 \), there exists a proper chain \( C = \{ c_j \} \) surrounding \( a \) such that the image \( \hat{\psi}(C) = \{ \hat{\psi}(c_j) \} \) is a proper chain.

Once the claim is shown, \( \hat{\psi} \) is extended to \( \partial \mathbb{H}^2 \) by mapping the point \( a \) to the impression of \( \hat{\psi}(C) \). The continuity at \( a \) follows from the condition that \( C \) is a proper chain surrounding \( a \).

Let us embark upon the proof of the claim. First of all let \( \hat{\gamma} : [0, \infty) \to \mathbb{H}^2 \) be a geodesic such that \( \hat{\gamma}(t) \to a \) as \( t \to \infty \). There are trichotomy by the property of the geodesic \( \gamma = \pi \circ \hat{\gamma} : [0, \infty) \to S \). Notice that it is independent of the choice of the geodesic \( \hat{\gamma} \).

**Case 1.** The geodesic \( \gamma \) is a nonproper curve in \( S \) (possibly a closed curve).

**Case 2.** The geodesic \( \gamma \) tends to a cusp of \( S \).

**Case 3.** None of the above.

**Case 1.** In this case it is easy to find a family of simple closed geodesics which cut \( S \) into a union of discs and once punctured discs. Then the geodesic \( \gamma \) cannot stay in a disc or a once punctured disc. On the other hand, \( \gamma \) is nonproper. Thus there is a simple closed geodesic \( \alpha \) such that \( \gamma \) intersects \( \alpha \) at infinitely many parameter values \( t_i \to \infty \). Let \( \hat{\alpha}_i \) be the lift of \( \alpha \) which passes through \( \hat{\gamma}(t_i) \). Then it is easy to show that \( C = \{ \hat{\alpha}_i \} \) is a proper chain at \( a \). Now let \( \varphi_{\ast}\alpha \) be the simple closed geodesic freely homotopic to \( \varphi(\alpha) \). Then any lift of \( \varphi(\alpha) \) lies in an \( r \)-neighborhood of a lift of \( \varphi_{\ast}\alpha \) for some fixed \( r > 0 \). That is, for any \( \hat{\varphi}(\hat{\alpha}_i) \) there exists a lift of \( \varphi_{\ast}\alpha \), say \( \hat{\varphi}_{\ast}\hat{\alpha}_i \), such that \( \rho(\hat{\varphi}(\hat{\alpha}_i)(t), \hat{\varphi}_{\ast}\hat{\alpha}_i(t)) < r \) for some parametrizations, where \( \rho \) denotes the hyperbolic metric on \( \mathbb{H}^2 \). The difference of the hyperbolic metric and the closed disc topology on \( \text{Cl}(\mathbb{H}^2) \) shows that \( \hat{\psi}(\hat{\alpha}_i) \) is a proper cross cut and that \( \partial \hat{\varphi}(\hat{\alpha}_i) = \partial \hat{\varphi}_{\ast}\hat{\alpha}_i \).

Since the geodesics \( \hat{\varphi}_{\ast}\hat{\alpha}_i \)'s are lifts of a fixed simple closed geodesic \( \varphi_{\ast}\alpha \), they are divergent, showing that \( \text{diam}(\partial \hat{\varphi}_{\ast}\hat{\alpha}_i) = \text{diam}(\partial \hat{\varphi}(\hat{\alpha}_i)) \to 0 \). Thus \( \hat{\psi}(C) = \{ \hat{\psi}(\hat{\alpha}_i) \} \) satisfies
condition (3) for being a proper chain. The other conditions are obvious from the corresponding properties of $C$ since $\bar{\psi}$ is a homeomorphism.

Case 2. Given a point $a \in \partial \mathbb{H}^2$ in the claim, one can choose a geodesic $\gamma : \mathbb{R} \to \mathbb{H}^2$ tending to $a$ as $t \to \infty$ in such a way that $\gamma = \pi \circ \bar{\gamma}$ tends to a cusp $x$ of $S$ as $t \to \infty$ and that $\gamma$ is simple and proper in $S$. Let $\delta_i$ be a sequence of horocycles in $S$ converging to the cusp $x$, and let $\delta_i$ be the lift of $\delta_i$ which intersects $\bar{\gamma}$ at a single point. Thus $\delta_i$ is a horocycle at $a$.

Let $T$ be a parabolic element in $\Gamma$ which keeps $a$ fixed. Then for any $i > 0$, the images $T^i(\bar{\gamma})$ and $T^{-i}(\bar{\gamma})$ are geodesics terminating at $a$ and they converge to $a$ from both sides as $i \to \infty$. For any $i$, form a proper cross cut $\beta_i$ by concatenating subarcs of $T^{-i}(\bar{\gamma})$, $\delta_i$ and $T^i(\bar{\gamma})$ in such a way that $C = \{\beta_i\}$ forms a proper chain surrounding $a$.

Let us show that $\bar{\psi}(C)$ is a proper chain. First of all $\psi(\gamma)$ is a simple proper curve in $S$ properly homotopic to the geodesic $\psi_2 \gamma$. Thus by Proposition 2.1, its lift $\bar{\psi}(\gamma)$ has distinct end points $b = \lim_{t \to \infty} \bar{\psi}(\gamma(t))$ and $d = \lim_{t \to -\infty} \bar{\psi}(\gamma(t))$.

The transformation $T_1 = \varphi T \bar{\psi}^{-1}$ is a parabolic element corresponding to the cusp $\psi(x)$. It is not difficult to show that $T_1$ keeps the point $b$ fixed. Therefore the curve $\bar{\psi}(\beta_i)$ is a proper cross cut whose end points are $T_1^{-i}(d)$ and $T_1^i(d)$. Since $\text{diam}(\partial \bar{\psi}(\beta_i)) \to 0$, the family $\bar{\psi}(C) = \{\bar{\psi}(\beta_i)\}$ is a proper chain, the other conditions being obvious from the fact that $\bar{\psi}$ is a homeomorphism.

Case 3. We assume that $\bar{\gamma} : [0, \infty) \to \mathbb{H}^2$ is a geodesic tending to a point $a \in \partial \mathbb{H}^2$, and that $\gamma = \pi \circ \bar{\gamma}$ is a proper geodesic in $S$ which does not converge to a cusp. The proper geodesic $\gamma$ intersects $\lambda_i = \partial A_i$ for any large $i$ and we can construct as before a proper chain $C = \{\lambda_i\}$ surrounding $a$ which consists of a lift of $\lambda_i$. To show that $\bar{\psi}(C)$ is a proper chain, let $\varphi_2 \lambda_i$ be the simple closed geodesic freely homotopic to $\varphi(\lambda_i)$. Then as before, $\bar{\psi}(\lambda_i)$ lies in an $\varepsilon_i$-neighborhood of a lift, say $\varphi_2 \lambda_i$, of $\varphi_2 \lambda_i$. Therefore $\bar{\psi}(\lambda_i)$ is a proper cross cut with the end points $\partial \bar{\psi}(\lambda_i) = \partial \varphi_2 \lambda_i$. Thus what is left is to show that $\text{diam}(\partial \varphi_2 \lambda_i) \to 0$.

Assume the contrary. Then $\varphi_2 \lambda_i$ must converge to a geodesic, say $\bar{\lambda}$. It is well known [2, (5.3.8)] that given any two points in the limit set of $\Gamma$ which in our case is $\partial \mathbb{H}^2$, there exists a hyperbolic element in $\Gamma$ whose fixed points are arbitrarily near the given points. In particular there exists a geodesic $\bar{\alpha}$ in $\mathbb{H}^2$ which intersects $\bar{\lambda}$ such that the projected image $\alpha = \pi \circ \bar{\alpha}$ is a closed geodesic in $S$. Then for any sufficiently large $i$, the geodesic $\varphi_2 \lambda_i$, and hence the curve $\psi(\lambda_i)$ intersects $\alpha$. That is, for any large $i$, $\lambda_i$ intersects the closed curve $\psi^{-1}(\alpha)$. A contradiction. ∎

3. Handel’s fixed point theorem

As in the previous section let $\psi$ be an orientation preserving homeomorphism of $\mathbb{R}^2$. Choose a set of nonperiodic points $\{x_1, \ldots, x_r\}$, each point from a distinct orbit, and let us denote the union of the orbits by $O = O(x_1, \ldots, x_r)$. We assume that $O$ is discrete in $\mathbb{R}^2$.

As before the surface $S = \mathbb{R}^2 \setminus O$ is uniformized by a Fuchsian group of the first kind. Let $\gamma$ be a proper simple geodesic in $S$ joining the cusps $\psi^i(x_1)$ and $\psi^{i+1}(x_1)$ (i.e., $\lim_{t \to -\infty} \gamma(t) = \psi^i(x_1)$ and $\lim_{t \to \infty} \gamma(t) = \psi^{i+1}(x_1)$). The geodesic $\gamma$ is called a homo-
topy translation arc if for any nonzero \( v \), we have \( y \cap \varphi_v^k y = \emptyset \). It is called forward proper if \( S^+(y) = \bigcup_{v \geq 0} \varphi_v^k y \) diverges in \( S \), i.e., any compact subset of \( S \) intersects but finitely many \( \varphi_v^k y \), and backward proper if \( S^-(y) = \bigcup_{v \leq 0} \varphi_v^k y \) diverges in \( S \). Notice that this condition is equivalent to saying that any compact subset of \( \mathbb{R}^2 \) intersects finitely many \( \varphi_v^k y \).

Now we shall state Handel’s fixed point theorem.

**Theorem 3.1.** Assume the following:

1. There exists \( N > 0 \) with the following property. For any \( 1 \leq i \leq r \), there exist a forward proper homotopy translation arc \( \gamma_i \) joining \( \varphi^N(x_i) \) and \( \varphi^{N+1}(x_i) \) and a backward proper homotopy translation arc \( \delta_i \) joining \( \varphi^{-N-1}(x_i) \) and \( \varphi^{-N}(x_i) \) such that all the \( S^+(\gamma_i) \)’s and the \( S^-(\delta_i) \)’s are mutually disjoint.

2. For any \( 1 \leq i \leq r \), there exists an oriented simple curve \( l_i \) in \( \mathbb{R}^2 \) joining \( \varphi^{-N}(x_i) \) and \( \varphi^N(x_i) \) whose interior does not intersect \( \bigcup_j (S^+(\gamma_j) \cup S^-(\delta_j)) \) with the following property: for any \( 1 \leq i \leq r \), \( l_i \) intersects \( l_{i+1} \) transversely at one point with positive intersection number, i.e., \( l_i \cdot l_{i+1} = 1 \), where \( l_{r+1} = l_1 \).

Then the fixed point set \( \text{Fix}(\varphi) \) is nonempty, and if \( \text{Fix}(\varphi) \) is discrete, there exists a fixed point of positive index.

Notice that condition (2) implies that \( r \geq 3 \).

The remainder of this section is devoted to the proof of Theorem 3.1. First of all the set \( S^+(\gamma_i) \) together with the forward orbit \( O^+(\varphi^N(x_i)) \) form a proper half-curve in \( \mathbb{R}^2 \). Since \( \mathcal{O} \) is discrete, there exists a simply connected neighbourhood of the half-curve in \( \mathbb{R}^2 \) which is disjoint from \( \mathcal{O} \setminus O^+(\varphi^N(x_i)) \). Then the boundary of the neighbourhood is a proper curve in \( S \). Let \( w_i \) be the geodesic in \( S \) properly homotopic to this curve. The geodesic \( w_i \) separates \( O^+(\varphi^N(x_i)) \) from the remaining points of \( \mathcal{O} \). Denote by \( W_i \) the component of \( S \setminus w_i \) which contains \( S^+(\gamma_i) \). The uniqueness of the geodesic \( w_i \) implies that the image \( \varphi_x w_i \) is contained in \( W_j \). In other words, \( \varphi_x W_i \) is contained in the interior of \( W_i \). Likewise, starting from \( S^-(\delta_i) \cup O^- (\varphi^{-N}(x_i)) \), we construct a geodesic \( a_i \) separating \( O^- (\varphi^{-N}(x_i)) \) from the remaining points of \( \mathcal{O} \) and a domain \( A_i \) bounded by \( a_i \) and containing \( S^- (\delta_i) \). Then the image \( \varphi_x A_i \) contains \( A_i \) in its interior. Notice that the closures of the \( W_i \)'s and the \( A_j \)'s are mutually disjoint.

Let

\[
R = S \setminus \bigcup_i (\text{Cl}(W_i) \cup \text{Cl}(A_i)), \quad \partial_+ R = \bigcup_i w_i.
\]

The subsurface \( \text{Cl}(R) \) is complete with geodesic boundaries and finitely many punctures. Let \( \mathcal{G} \) be the set of all the simple oriented geodesic segments contained in \( \text{Cl}(R) \) joining two points of distinct components of \( \partial_+ R \), and the shortest in their homotopy classes relative to \( \partial_+ R \). Of course in each homotopy class, there is exactly one element of \( \mathcal{G} \).

Now choose an arbitrary \( t \in G \). Assume the initial point \( t(0) \) is contained in \( w_j \). Let \( \alpha \) be a curve in \( \text{Cl}(W_j) \) joining the cusp \( \varphi^N(x_j) \) and \( t(0) \) whose interior does not intersect \( S^+(\gamma_j) \). Likewise, assuming \( t(1) \in w_j \), define a curve \( \omega \) joining \( t(1) \) to the cusp \( \varphi^N(x_j) \). Concatenating \( \alpha, t \) and \( \omega \), we obtain a simple proper curve in \( S \). Let \( \tilde{t} \) be the geodesic properly homotopic to this curve. Clearly \( \tilde{t} \) is defined uniquely.
Consider the image geodesic \( \varphi_2 \hat{t} \). Since \( \hat{t} \) does not intersect \( A_i \) and \( A_j \) is contained in \( \varphi_2 A_i \), \( \varphi_2 \hat{t} \) does not intersect \( A_j \). Some of the components of \( \varphi_2 \hat{t} \cap \text{Cl}(R) \) have end points in different components of \( \partial_+ R \). Then it is homotopic, relative to \( \partial_+ R \), to an element, say \( t' \) of \( \mathcal{G} \), with orientation induced from \( t \). In this case denote \( t' < t \). Since \( t' \) joins points of distinct components of \( \partial_+ R \), there is at least one \( t' \in \mathcal{G} \) such that \( t' < t \).

Also the geodesic segment \( \varphi_2 \hat{t} \setminus (\varphi_2 W_i \cup \varphi_2 W_j) \) is denoted by \( \varphi_2 t' \).

A subset \( T \) of \( \mathcal{G} \) fitted if the following conditions are satisfied.

(1) Any distinct elements \( t_1 \) and \( t_2 \) of \( T \) do not intersect unless \( t_1 = -t_2 \) (\( -t_2 \) is \( t_2 \) with the reversed orientation).

(2) If \( t_1 \in T \) and \( t_2 < t_1 \), then \( t_2 \in T \).

Notice that there are finitely many punctures in \( R \). Therefore condition (1) implies that \( T \) is a finite set.

Henceforth all the lemmas in this section are under the assumption of Theorem 3.1.

**Lemma 3.2.** There exists a fitted family.

**Proof.** Consider the \( \varphi_2 \)-orbit of all the backward proper homotopy arcs \( \delta_i \). Since the \( S^+(\delta_i) = \bigcup_{v<0} \varphi_2 \delta_i \) are mutually disjoint, the geodesics \( \varphi_2^r \delta_i \) are mutually disjoint for any \(-\infty < v < \infty \) and \( 1 \leq i \leq r \). In particular if \( v \geq 2 \), the geodesic \( \varphi_2^v \delta_i \) is disjoint from any \( A_j \). If \( v \geq 2N+1 \), \( \varphi_2^v \delta_i \) joins cusps in \( W_i \). Thus each component of \( \varphi_2^v \delta_i \cap \text{Cl}(R) \) joining points of different components of \( \partial_+ R \), with orientation induced from that of \( \delta_i \), is homotopic to an element of \( \mathcal{G} \). Form a family \( T \) by taking all such elements of \( \mathcal{G} \). Then since the \( \varphi_2^v \delta_i \) are mutually disjoint, they satisfy condition (1) of a fitted family.

Condition (2) is obvious by the very definition.

Therefore what is left is to show that \( T \) is nonempty. That is, for some \( v \geq 2N+1 \), \( \varphi_2^v \delta_i \cap \text{Cl}(R) \) admits a component joining points of different components of \( \partial_+ R \), equivalently \( \varphi_2^v \delta_i \) intersects some \( W_j \) (\( j \neq i \)).

Consider the closure \( c_i' \) in \( \mathbb{R}^2 \) of \( S^-(\varphi_2^{2N} \delta_i) = \bigcup_{1 \leq v \leq 2N} \varphi_2^v \delta_i \). \( c_i' \) is a proper half-cycle, starting at \( \psi^0(x_i) \) and ending in \( A_i \). Denote by \( c_i \) the proper infinite subarc of \( c_i' \) starting at a point of \( \partial W_i \) and contained in \( \mathbb{R}^2 \setminus W_i \). The union \( c_i \cup W_i \) divides \( \mathbb{R}^2 \) into two regions. One can assume that \( c_i \) is disjoint from \( W_j \) for any \( j \neq i \), for otherwise there is nothing to prove. Now by condition (2) of Theorem 3.1, \( c_i \cup W_i \) separates \( A_{i+1} \) and \( W_{i+1} \).

Consider the closure \( d_{i+1} \) in \( \mathbb{R}^2 \) of \( \bigcup_{1 \leq v < 2N} \varphi_2^v \gamma_{i+1} \). By the same reason as for \( \varphi_2^v \delta_i \), \( d_{i+1} \) is disjoint from \( W_i \) and joins \( \psi^N(x_{i+1}) \) and \( \psi^{-N}(x_{i+1}) \). Therefore \( d_{i+1} \) must intersect \( c_i \). That is, for some \( v' \) and \( v'' \), \( \varphi_2^v \delta_i \) intersects \( \varphi_2^{v''} \gamma_{i+1} \). This implies that for \( v \geq v' - v'' \), \( \varphi_2^v \gamma_i \) intersects \( S^+(\gamma_{i+1}) \), and hence \( W_{i+1} \), showing the lemma.

Given a fitted family \( T \), the directed graph of \( T \), denoted by \( \Gamma(T) \), is defined by making elements of \( T \) vertices of \( \Gamma(T) \) and drawing a directed edge from \( t \) to \( t' \) if \( t' < t \).

**Lemma 3.3.** For any \( t \in T \), there exist at least two distinct directed paths in \( \Gamma(T) \) of the same length starting at \( t \).
subcurve of $t$ to $n$. Part of $t$ to $n$. Geodesics properly homotopic to $\partial$ curve in $C$ in $\mathbb{R}^2$ which intersects each of the $A_i$’s and $R_j$’s in a connected subarc, the cyclic order of the $A_i$’s and $R_j$’s along $C$ (with appropriate orientation) is given by $\ldots, A_i, W_{i-1}, A_{i+1}, W_i, \ldots$.

Then it follows that for some $k$, $A_k$ and $W_k$ are separated by $t \cup W_i \cup W_j$. Thus $t$ intersects some $\varphi^{-1}\gamma_k$, and hence $(\varphi^v)\tilde{\gamma}^i$ intersects $\gamma_k$. That is, $(\varphi^v)\tilde{\gamma}^i \cap R$ has at least two components joining distinct components of $\partial R$. They are homotopic to distinct elements $t'$ and $t''$ of $T$.

On the other hand, by Proposition 2.1(5), we have $(\varphi^v)\tilde{\gamma} = (\varphi^v)\tilde{\gamma}^i$, since they are geodesics properly homotopic to $\varphi^v(\tilde{t})$. It is not difficult to show that there is one path from $t$ to $t'$ and another from $t$ to $t''$, both of length $v$. □

Corollary 3.4. There exist two intersecting distinct directed cycles in $\Gamma(T)$.

Proof. Lemma 3.3 and the finiteness of $T$ implies the existence of a cycle in $T$. Negating the conclusion, there is defined in an obvious way a partial order among the cycles. Any vertex of a lowest cycle does not satisfy the previous lemma. A contradiction. □

Lemma 3.5. For some $t \in T$, there exists an oriented path from $t$ to $-t$.

Proof. By the previous lemma, there exist two distinct cycles intersecting at some point $t \in T$. Iterating the cycles if necessary, one may assume that the cycles are of the same period, say $n$. That is, $(\varphi^n)\tilde{\gamma}^i \cap \text{Cl}(R)$ has at least two components, say $t_1$ and $t_2$. Homotopic to $t$ relative to $\partial_+ R$. We assume that $t_1$ and $t_2$ appear in this order in $(\varphi^n)\tilde{\gamma}^i$. Let $s$ be the subcurve of $\varphi^n\tilde{\gamma}^i$ joining the terminal point of $t_1$ and the initial point of $t_2$, and let $u$ be the curve in $\partial_+ R$ joining the terminal points of $t_1$ and $t_2$. If $s \cap \text{Cl}(R)$ contains an oriented arc homotopic to $-t$, we are done. If not, the curves $s$, $t_2$ and $u$ form a simple closed curve $c$. Let $D$ be the disc in $\mathbb{R}^2$ bounded by $c$. Notice that $\varphi^n\partial_+ R$ does not intersect $c$, and therefore $D$.

In case $t_1$ is contained in $D$, since the initial point of $\varphi^n\tilde{\gamma}^i$ lies on $\varphi^n\partial_i R_i$ for some $i$, the part of $\varphi^n\tilde{\gamma}^i$ which precedes $t_1$ must go out of $D$, and hence intersects $u$. This implies that some component of $\varphi^n\tilde{\gamma}^i \cap \text{Cl}(R)$ is homotopic to $-t$ relative to $\partial_+ R$. In the opposite case, the part of $\varphi^n\tilde{\gamma}^i$ which succeeds $t_2$ intersects $u$ and we get the same conclusion. □

Proof of Theorem 3.1. We assume that the fixed point set of $\varphi$ is discrete (possibly empty). By the previous lemma, there exists $t \in T$ such that $\varphi^n t \cap \text{Cl}(R)$ has a component homotopic to $-t$. Let $w_i$ (respectively $w_j$) be the component of $\partial_+ R$ which contains the initial (respectively terminal) point of $t$. Let $\tilde{R}$ be a lift of $R$ to the universal covering space $\mathbb{H}^2$ of $S = \mathbb{R}^2 \setminus \mathcal{O}$. Let $\tilde{t}$ be a lift of $t$ contained in $\text{Cl}(\tilde{R})$ and let $\tilde{w}_i$ (respectively $\tilde{w}_j$)
be the lift of \( w_i \) (respectively \( w_j \)) which contains the initial (respectively terminal) point of \( i \). Let \( \widetilde{S}_i \) (respectively \( \widetilde{S}_j \)) be the component of \( \mathbb{H}^2 \setminus w_i \) (respectively \( \mathbb{H}^2 \setminus w_j \)) which does not contain \( \tilde{R} \). Then by the assumption on \( t \) there exists a lift \( \tilde{\psi} \) of \( \psi^n \) such that \( \tilde{\psi}_2 \tilde{w}_i \) (respectively \( \tilde{\psi}_2 \tilde{w}_j \)) is contained in \( \widetilde{S}_i \) (respectively \( \widetilde{S}_j \)).

By Proposition 2.2, the lift \( \tilde{\psi} \) extends to \( \text{Cl} \mathbb{H}^2 \). Notice that by the above property, \( \tilde{\psi} \) does not have a fixed point in \( \partial \mathbb{H}^2 \). Therefore there must be a fixed point \( \tilde{p} \) of \( \tilde{\psi} \) in \( \mathbb{H}^2 \).

Thus \( p \) is a periodic point of \( \psi \). If \( p \) is not a fixed point of \( \psi \), then a version of Brouwer’s fixed point theorem asserts the existence of a fixed point of \( \psi \) of positive index. (See, e.g., [1].) The proof is done in this case.

Assume, on the other hand, \( p \) is a fixed point of \( \psi \). Then there exists a lift \( \tilde{\psi} \) of \( \psi \) which keeps the point \( \tilde{p} \) fixed. Then since the two lifts \( \tilde{\psi} \) and \( \tilde{\psi}^n \) of \( \psi^n \) keep the same point \( \tilde{p} \) fixed, we have \( \tilde{\psi} = \tilde{\psi}^n \). Since there are no fixed points of \( \tilde{\psi} \) in \( \partial \mathbb{H}^2 \), there are no fixed points of \( \tilde{\psi} \) in \( \mathbb{H}^2 \). This shows that the fixed points of \( \tilde{\psi} \) in \( \mathbb{H}^2 \), being discrete, are finite in number. By the Lefschetz fixed point theorem there exists a fixed point of \( \tilde{\psi} \) of positive index in \( \mathbb{H}^2 \).

4. Rotation vectors and fixed points

Throughout this section we are concerned with an oriented surface of finite type \( \Sigma \) with negative Euler number, i.e., \( \Sigma = M \setminus P \) where \( M \) is a closed oriented surface and \( P \) is a finite subset of it, possibly empty, and a homeomorphism \( g \) of \( \Sigma \) isotopic to the identity by an isotopy \( g_t \mathbin{\overset{\text{Id}}{\sim}} g \). The isotopy \( g_t \) is unique up to homotopy since the Euler number of \( \Sigma \) is assumed to be negative. Thus the canonical lift \( \tilde{g} \) and the contractible fixed point set \( \text{Fix}_c(g) \) are defined.

Let us endow \( \Sigma \) with a complete hyperbolic metric with finite area, and denote by \( \pi : \mathbb{D}^2 \to \Sigma \) the universal covering map, where \( \mathbb{D}^2 \) denotes the Poincaré disc.

For any point \( x \in \Sigma \), \( \delta(x) \) denotes the locus of \( x \) by the isotopy \( g_t \), i.e., the path given by \( t \mapsto g_t(x) \). For any periodic point \( p \) in \( \Sigma \) of period, say \( m \), \( R(p) = R(p; g) \) denotes the homology class
\[
(1/m) \left[ \delta(p) \circ \delta(g(p)) \circ \cdots \circ \delta(g^{m-1}(p)) \right],
\]
where \( \circ \) denotes the concatenation. This is a class in \( H_1(\Sigma; \mathbb{R}) \), well-defined, independent of the choice of the isotopy \( g_t \).

We are going to establish a fixed point theorem for \( g \) which takes different forms according as the genus of \( \Sigma \) vanishes or not. Most of this section is devoted to the case where the genus is positive. We assume this for a while.

Given a finite subset \( C \) of \( H_1(\Sigma; \mathbb{Z}) \), let us define its graph \( \Gamma(C) \) as follows. The vertices of \( \Gamma(C) \) are elements of \( C \) and two vertices are joined by an edge if the intersection number of their images in \( H_1(M; \mathbb{Z}) \) is nonvanishing. A subset \( C \) is called filling if the following conditions are satisfied.

1. The subset \( C \) contains the origin 0 in the interior of its convex hull in \( H_1(\Sigma; \mathbb{R}) \).
2. No two elements of \( C \) lie on a straight line passing through 0.
3. The graph \( \Gamma(C) \) is connected.
Notice that condition (1) implies that \( C \) is a generating set of \( H_1(\Sigma, \mathbb{R}) \). Also, condition (3) implies that the image of any element of \( C \) is nonzero in \( H_1(M, \mathbb{R}) \).

**Lemma 4.1.** Let \( \lambda_j \) (\( 1 \leq j \leq q \)) be (possibly nonsimple) closed curves of \( \Sigma \) such that the set \( \{[\lambda_j]\} \) is filling. Then any component of \( \Sigma \setminus (\bigcup_j \lambda_j) \) is either a disc or a once punctured disc.

**Proof.** Let \( \Omega \) be a component of \( \Sigma \setminus (\bigcup_j \lambda_j) \). In way of contradiction assume first that \( \Omega \) has nonzero genus. Then \( \Omega \) must contain simple closed curves \( \alpha \) and \( \beta \) whose intersection number in \( M \) is nonzero. By condition (1), the class \( [\alpha] \) is a linear combination of the \( [\lambda_j] \)'s. But then, since \( \beta \) does not intersect the \( \lambda_j \)'s, the intersection number of \( \beta \) and \( \alpha \) must be zero. A contradiction.

Assume next that \( \Omega \) is not a possibly punctured disc. Then \( \Omega \) contains a simple curve \( \gamma \), essential in \( M \). Since the graph of \( \{[\lambda_j]\} \) is connected, the set \( \bigcup_j \lambda_j \) is also connected. This implies that \( \gamma \) is nonseparating, i.e., represents a nonzero class in \( H_1(M; \mathbb{R}) \). Since the classes \( [\lambda_j] \) generate \( H_1(M; \mathbb{R}) \), the class \( [\gamma] \) must have nonzero intersection number with some \( [\lambda_j] \), contrary to the assumption that \( \gamma \) does not intersect any \( \lambda_j \).

Finally assume that \( \Omega \) has more than one puncture. Then a curve in \( \Omega \) joining two punctures represents a nonzero class of \( H_1(M; \mathbb{R}) \). Considering the intersection of \( H_1(M; \mathbb{R}) \) and \( H_1(\Sigma; \mathbb{R}) \), an argument similar to the above yields a contradiction. \( \square \)

Our first fixed point theorem is the following.

**Theorem 4.2.** Let \( \Sigma \) be a complete finite area hyperbolic surface of genus > 0, and let \( g \) be a homeomorphism isotopic to the identity. Assume there exist periodic points \( p_j \in \Sigma \) of period \( m_j \) (\( 1 \leq j \leq q \)) such that the set \( \{m_j R(p_j)\} \) is filling. Then the contractible fixed point set \( \text{Fix}_c(g) \) is nonempty, and if further it is finite, it contains a fixed point of positive index.

Henceforth assuming that \( \text{Fix}_c(g) \) is at most finite, let us show the existence of a fixed point of positive index.

Since the set \( \{m_j R(p_j)\} \) is filling, each of its elements represents a nonzero element of \( H_1(M, \mathbb{Z}) \) and therefore there exists an oriented closed geodesic \( \lambda_j \) in \( \Sigma \) which is homotopic to \( \delta(p_j) \circ \delta(g(p_j)) \circ \cdots \circ \delta(g^{m_j-1}(p_j)) \). The geodesic \( \lambda_j \) may not be prime, but for distinct \( i \) and \( j \), the curves \( \lambda_i \) and \( \lambda_j \) cannot trace the same geodesic by condition (2). Therefore any edge of the one-dimensional simplicial complex \( \bigcup_j \lambda_j \) admits a well-defined orientation. The figure of \( \bigcup_j \lambda_j \) may differ according to the choice of the hyperbolic structure of \( \Sigma \), but we do not care about this.

Let \( \Omega_1, \ldots, \Omega_r \) be the connected components of \( \Sigma \setminus (\bigcup_j \lambda_j) \), which are either discs or once punctured discs. Form a directed graph \( G \) as follows. The vertices of \( G \) are the components \( \Omega_1, \ldots, \Omega_r \). Choose an edge \( e \) of the set \( \bigcup_j \lambda_j \) and let \( \Omega_i \) and \( \Omega_k \) be two components which contain the edge \( e \) in its boundary. (It may happen that \( \Omega_i = \Omega_k \).) Draw an oriented edge from \( \Omega_i \) to \( \Omega_k \) if a curve in \( \Omega_i \cup \Omega_k \cup e \) starting at a point of \( \Omega_i \)
and ending at a point of \( \Omega_k \) intersects \( e \) transversely at one point with positive intersection number.

Under the hypotheses of Theorem 4.2, we obtain the following lemma.

**Lemma 4.3.** The graph \( G \) does not contain a directed cycle, or a directed path joining two punctured discs.

**Proof.** Suppose the contrary. Then a directed cycle or a directed path joining punctured discs gives birth to a homology class \( a \in H_1(\Sigma; \mathbb{Z}) \) such that \( a \cdot [\lambda_j] > 0 \) for any \( j \) and \( a \cdot [\lambda_j] > 0 \) for some \( j \), where the intersection number is a pairing of elements of \( H_1(\Sigma; \mathbb{Z}) \) and \( H_1(\Sigma; \mathbb{Z}) \). Now recall that the origin \( 0 \) is in the interior of the convex hull of \([\lambda_j]\)'s. As is well known and easy to show, it is possible to express \( 0 \) as \( 0 = \sum j t_j [\lambda_j] \) with all the coefficients \( t_j \) positive. A contradiction. \( \square \)

Because no directed cycle is included in \( G \), there is a maximal directed path in \( G \). Then one of the end points, say \( i \), is a disc without puncture. The boundary of \( i \) is formed by subarcs of the geodesics \( \lambda_1, \ldots, \lambda_r \) in this order. The orientation of \( \lambda_1, \ldots, \lambda_r \) is concurrent since \( i \) is an end point of a maximal directed path.

Here for simplicity, we change the notation: the domain \( i \) is denoted henceforth by \( \Omega_i \), and the closed oriented geodesics \( \lambda_1, \ldots, \lambda_r \) by \( \mu_1, \ldots, \mu_r \). The corresponding periodic points are denoted by \( \pi_1, \ldots, \pi_r \), their periods by \( m_1, \ldots, m_r \). This will cause no confusion, since the other periodic points will not appear any more.

Denote by \( \tilde{g} \) the canonical lift of \( g \). Let \( \tilde{\Omega} \) be a lift of \( \Omega \) and \( \tilde{\lambda}_i \) be the lift of \( \lambda_i \) intersecting the polygon \( C(\tilde{\Omega}) \) along its edge. Thus \( \tilde{\lambda}_i \) is a complete oriented geodesic in \( \mathbb{D}^2 \). Recall that the geodesic \( \lambda_i \) is homotopic to the curve \( \delta(p_i) \circ \delta(g(p_i)) \circ \cdots \circ \delta(g^{m_i-1}(p_i)) \). Lifting the homotopy so as to start at \( \tilde{\lambda}_i \), we obtain a lift of the curve \( \delta(p_i) \circ \delta(g(p_i)) \circ \cdots \circ \delta(g^{m_i-1}(p_i)) \). Choose and fix one lift \( x_i \) of the periodic point \( \pi_i \) which lies on this curve.

In what follows we are going to prove that the points \( x_i \) satisfy the assumption of Theorem 3.1 with respect to the homeomorphism \( \tilde{g} \), which is sufficient for our purpose. First of all there exists a hyperbolic element \( h_i \in \pi_1(\Sigma) \) such that \( \tilde{g}^{m_i}(x_i) = h_i(x_i) \). Clearly we have \( h_i(\tilde{\lambda}_i) = \tilde{\lambda}_i \). Since \( \tilde{g} \) is the canonical lift, \( \tilde{g} \) commutes with any deck transformation of \( \mathbb{D}^2 \). This shows that for any \( k \) and \( 0 \leq r \leq m_i - 1 \), we have \( \tilde{g}^{m_i k + r}(x_i) = h_i^k(\tilde{g}^r(x_i)) \). Thus the \( \alpha \)-limit point \( \alpha(x_i) \) and the \( \omega \)-limit point \( \omega(x_i) \) by the canonical lift \( \tilde{g} \) are the fixed points of a hyperbolic transformation \( h_i \) which has \( \tilde{\lambda}_i \) as an axis.

Since the oriented geodesics \( \tilde{\lambda}_i \) bound the polygon \( \tilde{\Omega} \) with concurrent orientation, they satisfy the following property: the geodesic \( \tilde{\lambda}_i (1 \leq i \leq r) \) intersects the geodesic \( \tilde{\lambda}_{i+1} \) transversely with positive intersection number, where the convention \( \tilde{\lambda}_{r+1} = \tilde{\lambda}_1 \) is used. Therefore the only thing left is to show condition (1) of Theorem 3.1, since condition (2) will be obvious by the above observation, after (1) is established.

Let \( O = \mathcal{O}(x_1, \ldots, x_r) \) and \( S = \mathbb{D}^2 \setminus O \) as before. Recall that the surface \( S \) is equipped with a complete hyperbolic structure. We aim to show the existence of forward (respectively backward) proper homotopy translation arc for \( x_i \) in \( S \). The argument will be
divided into two steps. In the first step, we consider the surface $S_i = \mathbb{D}^2 \setminus O(x_i)$ instead of $S$ and show that $\tilde{g}$ admits a forward and backward homotopy translation arc in $S_i$. Of course this will be much simpler than dealing with the space $S$. The second step is to use this fact to find out a forward proper homotopy translation arc for $x_i$ in $S$.

In the first step we are solely concerned with a single orbit $O(x_i)$, and study the dynamics of a homeomorphism which $\tilde{g}$ induces on a certain annulus. Here we use a piece of Nielsen–Thurston theory, together with Brouwer’s theorem. Let us begin by preparing the prerequisites needed in this development. For more details the reader may consult [12].

For any continuous map $f$ of a compact manifold $X$ to itself, two fixed points $p$ and $q$ are said to be Nielsen equivalent if they are the projected image of fixed points of the same lift of $f$ to the universal covering, or equivalently if there exists an arc $\gamma$ from $p$ to $q$ which is homotopic to its image $f \gamma$ relative to the end points.

The fixed point set $\text{Fix}(f)$ is divided into a disjoint union of Nielsen equivalence classes $F_j$. Classes $F_j$ are easily shown to be open and closed in $\text{Fix}(f)$, and therefore they are finite in number. The fixed point index $\text{Ind}(F_j; f) \in \mathbb{Z}$ can be defined, and if $F_j$ is a finite set, coincides with the sum of the usual indices of the points in $F_j$. The total sum of the index $\sum_j \text{Ind}(F_j; f)$ is equal to the Lefshetz number of $f$.

Given a homotopy $f_t$ such that $f_0 = f$ and a Nielsen class $F_j$ of $f$ with nonzero index, there exists a path $\{x_t\}$ in $X$ such that $x_0 \in F_j$, $f_t(x_t) = x_t$. The Nielsen class $F'_j$ for $f_t$ containing $x_t$ is independent of the choice of the path $\{x_t\}$ and we have

$$\text{Ind}(F_j; f) = \text{Ind}(F'_j; f_t).$$

We say $F_j$ is homotopic to $F'_j$.

Another prerequisite is the following Brouwer’s translation theorem. For the proof, see, e.g., [3,5,8,9].

**Theorem 4.4.** Let $f$ be a fixed point free orientation preserving homeomorphism of the plane $\mathbb{R}^2$. Then for any point $x \in \mathbb{R}^2$, there exists a domain $O$ containing $x$ and bounded by a proper curve $b$ and its translate $f(b)$ such that $b \cap f(b) = \emptyset$ and $O \cap f(O) = \emptyset$.

The domain $O$ is called a translation domain for $x$.

Now let us embark upon the proof of condition (1) of Theorem 3.1. Since $\tilde{g}$ is the canonical lift, it extends to the identity on the circle at infinity $\partial \mathbb{D}^2$. Define a closed annulus $A$ by $A = (\text{Cl}(\mathbb{D}^2) \setminus \text{Fix}(h_i))/h_i$. Let $\hat{g}$ be the pushdown of $\tilde{g}$ to $A$. Of course $\hat{g}$ is the identity on the boundary.

Denote by $\text{Fix}_c(\hat{g})$ the Nielsen class in $\text{Fix}(\hat{g})$ which corresponds to the lift $\tilde{g}$. Since by the assumption the contractible fixed points of $g : \Sigma \to \Sigma$ are finite, $\text{Fix}_c(\hat{g})$ is discrete in $\text{Int}(A)$ and if nonempty in $\text{Int}(A)$ accumulates at any point of the boundary. Since $\hat{g}$ is isotopic to the identity, the invariance of the index of a Nielsen class under the homotopy implies $\text{Ind}(\text{Fix}_c(\hat{g}); \hat{g}) = 0$.

Denote by $Y = \{y_1, \ldots, y_m\}$ the projected image of the orbit $O(x_i)$. The number $m_i$ is the period of the periodic point $p_i \in \Sigma$.

The case $m_i = 1$ will be treated separately at the end of the proof. Assume $m_i > 1$. Then for any small $\epsilon > 0$ there exists a modification $g'$ of $\hat{g}$ such that $g' = \hat{g}$ on a neighbourhood
of $\text{Fix}(\hat{g})$, $\text{Fix}(g') = \text{Fix}(\hat{g})$ and $g'(B_\varepsilon(y)) = B_\varepsilon(y)$, where $B_\varepsilon(y)$ denotes the open ball of radius $\varepsilon$ centered at $y$. Consider the closed subsurface $C = A \setminus \bigcup_i B_\varepsilon(y_i)$, and denote by $g_0$ the restriction of $g'$ to $C$.

Now any Nielsen class of the homeomorphism $\hat{g}$ of the annulus $A$ divides into a union of Nielsen classes of the homeomorphism $g_0$ of the subsurface $C$. In particular $\text{Fix}_c(\hat{g})$ is the union of Nielsen classes $F_j$ $(1 \leq j \leq q)$ of $g_0$. We have

$$\sum_j \text{Ind}(F_j; g_0) = \text{Ind}(\text{Fix}_c(\hat{g}); \hat{g}) = 0.$$

Let $\partial_+ A$ and $\partial_- A$ be the two boundary components of $A$. Denote by $F_{\pm}$ the Nielsen class of $g_0$ which contains $\partial_\pm A$. They are distinct classes contained in $\text{Fix}_c(\hat{g})$, as is easily shown considering the lift of $g_0$ to $\mathbb{D}^2$. Since each Nielsen class of $g_0$ is open in $\text{Fix}_c(\hat{g})$ and $\text{Fix}_c(\hat{g})$ is discrete in $\text{Int}(A)$, $\text{Fix}_c(\hat{g}) \setminus (F_+ \cup F_-)$ is a finite set.

Now let $f_0 : C \to C$ be the Thurston normal form of $g_0$. That is, there is an isotopy from $g_0$ to $f_0$ which keeps the boundary components pointwise fixed, and there exists a family of disjoint compact subsurfaces $P_1, \ldots, P_s$ of negative Euler number with the following properties.

1. The boundary components $\partial_\pm A$ of $C$ is disjoint from any $P_i$.
2. The other boundary components are contained in the union $\bigcup_i P_i$.
3. There is a permutation $\sigma$ of $\{1, 2, \ldots, s\}$ such that $f_0(P_i) = P_{\sigma(i)}$.
4. For each $i$, if $(f_0)^{n_i}(P_i) = P_i$ for some $n_i$, $(f_0)^{n_i}|_{P_i}$ is either periodic or pseudo Anosov.
5. The complement $C \setminus (\bigcup_i P_i \cup \partial_- A \cup \partial_+ A)$ consists of disjoint annuli, and contains no periodic points of $f_0$.

The last condition about the nonexistence of periodic points is achieved by adding to $f_0$ if necessary a dynamically simple flow in the annuli tending from one boundary component to the other.

Let us study the Nielsen equivalence class of the fixed points of $f_0$ and their indices. If a component $P_i$ is mapped by $f_0$ to another component, there are no fixed points in $P_i$ and we are not interested in such a component.

Suppose a component $P_i$ is kept invariant by $f_0$ and the restriction of $f_0$ to it is a pseudo Anosov homeomorphism. Then any fixed point in $P_i$ is isolated and it is well known that any two of interior fixed points are not mutually Nielsen equivalent in $P_i$. Moreover a simple topological observation based upon the $\pi_1$ injectivity of $P_i$ into $C$ shows that they are not Nielsen equivalent even in $C$. See, e.g., [13] for more details. The index of an interior fixed point is nonzero. Two fixed points in $P_i$ are Nielsen equivalent in $C$ if and only if they lie on the same boundary component. The index of a Nielsen class contained in a boundary component is negative.

Next if the restriction of $f_0$ to an invariant component $P_i$ is periodic of period $> 1$, then the fixed points are isolated, lie in the interior of $P_i$, and have index 1. Again any two of them are not Nielsen equivalent to each other in $C$.

Finally if $f_0$ is the identity when restricted to a component $P_i$, then the index of $P_i$ is equal to the Euler number of $P_i$ and is negative.
Two fixed points from different components can be Nielsen equivalent. However this happens only for a pair of boundary fixed points of two neighbouring Anosov components, or a combination of a full fixed point component (including ∂⁺A) and boundary fixed points of some adjacent Anosov components. In any case the index of such a Nielsen class is negative.

Extend \( f_0 \) to a homeomorphism \( f \) of the annulus \( A \), and consider the Nielsen class \( \text{Fix}_c(f) \) homotopic to \( \text{Fix}_c(h) \), which also has vanishing index. Again it consists of a disjoint union of Nielsen classes \( F' \) of \( f_0 \), whose indices sum up to zero. Let us denote the Nielsen class which contains \( \partial⁺A \) by \( F'_⁺ \).

The above observation shows that if \( \text{Ind}(F'_⁺; f_0) = 0 \), then \( \partial⁺A = F'_⁺ \). As is easily shown the converse also holds. Recall that all the other Nielsen classes have nonzero index.

Case 1. There exists a Nielsen class of \( f_0 \) of nonzero index in \( \text{Fix}_c(f) \).

In this case since the sum of the indices are 0, there must be a Nielsen class \( F' \) of positive index, which is of course not \( F'_⁺ \). \( F' \) is homotopic to a Nielsen class of \( g_0 \) contained in \( \text{Fix}_c(\tilde{g}) \). Since \( F'_⁺ \) is homotopic to \( F'_⁺ \), \( F' \) must be homotopic to a Nielsen class other than \( F'_⁺ \), thus to a finite set. This shows the existence of fixed points of \( g_0 \), hence of \( \tilde{g} \), of positive index. This gives birth to a contractible fixed point of \( g \) of positive index. The proof of Theorem 4.2 is complete in this case.

Case 2. \( \text{Fix}_c(f) \) does not admit a Nielsen class of nonzero index.

In this case, the above observation shows that \( \text{Fix}_c(f) = \partial⁺A \cup \partial₋A \).

Let us show that \( \tilde{g} \) admits a forward and backward proper homotopy translation arc \( \alpha \) in \( S_i = \mathbb{D}^2 \setminus \partial \mathcal{O}(x_i) \) such that \( \tilde{g}^m \alpha = (h_i)_\gamma \alpha \). (Recall that \( h_i \) is a deck transformation satisfying \( h_i(x_i) = \tilde{g}^m(x_i) \).)

We shall endow \( S_i \) with an especially nice hyperbolic structure. It will help a lot in the future when we induce a forward proper translation arc in \( S \) from that in \( S_i \). Let us consider a finite area complete hyperbolic structure on \( \text{Int}(A) \setminus \{y_1, \ldots, y_m\} \) and consider its lift to \( S_i \). Then \( S_i \) is uniformized by a Fuchsian group of the first kind, and the transformation \( h_i \) is an isometry of this hyperbolic structure.

Later we need to consider a hyperbolic structure on \( S \). Of course geodesics of \( S \) and \( S_i \) are different even if they are properly homotopic in \( S \). When we deal with the hyperbolic structure on \( S \), we continue to use the former notation \( \tilde{g}_\gamma \), etc. for a simple proper geodesic \( \gamma \). But when the space \( S_i \) is concerned, the same notation is confusing and we use the notation \( \tilde{g}_i \gamma \), etc.

Since \( f \) is isotopic to \( \tilde{g} \) relative to \( \{y_1, \ldots, y_m\} \), there is a lift \( \tilde{f} \) of \( f \) which is isotopic to \( \tilde{g} \) relative to \( \partial \mathcal{O}(x_i) \). The equality \( \text{Fix}_c(f) = \partial⁺A \cup \partial₋A \) implies that the homeomorphism \( \tilde{f} \) of \( \mathbb{D}^2 \) is fixed point free. Therefore by Brouwer’s translation theorem, there exists a translation domain \( O \) containing the point \( x_i \) and bounded by a proper curve \( b \) and its translate \( \tilde{f}(b) \).

By some abuse let us denote the restriction of \( \tilde{f} \) or \( \tilde{g} \) to \( S_i \) by the same letter. Let \( \beta \) be a proper geodesic in \( S_i \), properly homotopic to \( b \). Then since \( b \cap \tilde{f}(b) = \emptyset \), we have \( \beta \cap \tilde{f}(b) = \emptyset \). Since \( \tilde{f} \) is isotopic to \( \tilde{g} \) in \( S_i \), \( \tilde{f}_i \beta \) coincides with \( \tilde{g}_i \beta \). Let \( \Omega \) be the domain in \( S_i \) bounded by \( \beta \) and \( \tilde{g}_i \beta \). It is punctured at \( x_i \) and satisfies \( \Omega \cap \tilde{g}_i \Omega = \emptyset \), where \( \tilde{g}_i \Omega \) is defined as in Section 2, just before Proposition 2.2.
Let us show at this point that $\bigcup_i \text{Cl} \left( \tilde{g}_i^\gamma \Omega \right) = S_i$. We consider the space $S_i$ to be the quotient of $\mathbb{H}^2$ by a Fuchsian group $\Gamma_i$, the canonical projection being denoted by $\pi_i : \mathbb{H}^2 \to S_i$. Denote

$$ R = \bigcup \text{Cl} \left( \tilde{g}_i^\gamma \Omega \right). $$

Then any boundary point of $R$ is accumulated by points on disjoint geodesics $\tilde{g}_i^\gamma \beta$, showing that the boundary of $R$ consists of a disjoint union of simple complete geodesics. The same is true for the inverse image $\pi_i^{-1}(R)$ which is $\Gamma_i$-invariant and connected because the inclusion of $R$ into $S_i$ is a homotopy equivalence. Since the Fuchsian group $\Gamma_i$ is of the first kind, we have $\pi_i^{-1}(R) = \mathbb{H}^2$, showing that $\bigcup_i \text{Cl} \left( \tilde{g}_i^\gamma \Omega \right) = S_i$.

Now the geodesic $\alpha$ in $Q \cup \tilde{g}_i \Omega \cup \tilde{g}_i \beta$ joining $x_i$ and $\tilde{g}(x_i)$ is uniquely determined. Of course $\alpha$ intersects $\tilde{g}_i \beta$ transversely at one point. The fact that $\bigcup_i \text{Cl} \left( \tilde{g}_i^\gamma \Omega \right) = S_i$ implies that $\alpha$ is a forward and backward proper homotopy translation arc for $\tilde{g}$ in $S_i$.

Since $h_i$ is an isometry, we have $h_i(\alpha) = (h_i)_* \alpha$. Let us show that $(h_i)_* \alpha = \tilde{g}_i^m \alpha$. For this purpose it clearly suffices to show that the homotopy translation arc for $x_i$ in $S_i$ is unique. Suppose for contradiction that $\gamma$ is a homotopy translation arc different from $\alpha$. Since $\gamma$ and $\alpha$ are distinct, $\gamma$ is not contained in $Q \cup \tilde{g}_i \Omega \cup \tilde{g}_i \beta$. One may assume that $\gamma \cap \tilde{g}_i^\gamma \Omega \neq \emptyset$ for some $\nu > 1$. Choose such a largest $\nu$. Then a component $c$ of $\gamma \cap \text{Cl} \left( \tilde{g}_i^\gamma \Omega \right)$ and a curve in the boundary component $\tilde{g}_i^\gamma \beta$ joining the end points of $c$ bounds a disc punctured at $\tilde{g}_i^\gamma(x_i)$. Therefore the curve $\tilde{g}_i^\gamma \gamma$, which starts at $\tilde{g}_i^\gamma(x_i)$ and cannot intersect $c$, intersects $\tilde{g}_i^\beta$. That is, $\gamma$ intersects $\beta$. This shows that for some $\eta > 0$, $\gamma$ intersects $\tilde{g}_i^{-\eta} \Omega$. Take such a largest $\eta$. Then $\tilde{g}_i^{-\eta} \gamma$ and $\tilde{g}_i^\gamma \gamma$ intersect in $Q$. A contradiction.

Now we have shown that $h_i(\alpha) = (h_i)_* \alpha = \tilde{g}_i^m \alpha$. The discreteness of the Fuchsian group which uniformize the surface $S$ implies $[2, (5.1.2)]$ that for the points $x_1, \ldots, x_r$ in $\mathbb{H}^2$, their $\alpha$-limit points and $\alpha$-limit points, being the fixed points of hyperbolic elements $h_1, \ldots, h_r$, are all distinct. This implies that for large $n > 0$, $h_i^n(\alpha)$ does not intersect any other orbit $O(x_j)$ ($j \neq i$).

Let us show furthermore that for any $\nu > 0$, both curves $\tilde{g}_i^\gamma h_i^n(\alpha)$ and $\tilde{g}_i^\nu(h_i^n(\alpha))$ are contained in $S$ and are properly homotopic in $S$. Using this we shall show later that the geodesic $\gamma_i$ in $S$ which is properly homotopic to $h_i^n(\alpha)$ is a forward proper translation arc for $\tilde{g}$ in $S$, that is, $S_+ (\gamma_i) = \bigcup_{\nu > 0} \tilde{g}_i^\nu \gamma_i$ diverges in $S$. Recall that $\tilde{g}_i^\nu \gamma_i$ is a geodesic in $S$ which is homotopic to $\tilde{g}^\nu(h_i^n(\alpha))$ in $S$.

Proceeding with the details, let $H$ be a proper homotopy in $S$, between $h_i(\alpha)$ and $\tilde{g}_i^m(\alpha)$. Its image is contained in a compact subset $K$ of $\mathbb{H}^2$. Recall that the $\omega$-limit points and $\alpha$-limit points for the points $x_1, \ldots, x_r$ are all distinct. Therefore for any sufficiently large $N > 0$, $h_i^N(K)$ does not intersect the orbit of $x_j$ for $j \neq i$. Now the map $h_i^N \circ H$ is a homotopy between $h_i^{N+1}(\alpha)$ and $h_i^N(\tilde{g}_i^m(\alpha))$ whose image does not intersect the orbit $O(x_j)$ ($j \neq i$). Since $h_i$ and $\tilde{g}_i$ commute, we have $h_i^N(\tilde{g}_i^m(\alpha)) = \tilde{g}_i^m(h_i^N(\alpha))$. This shows that $h_i(h_i^N(\alpha))$ is properly homotopic to $\tilde{g}_i^m(h_i^N(\alpha))$ in $S$.

Denoting by $\simeq$ the equivalence relation by proper homotopy in $S$, we further obtain

$$ h_i^N(\alpha) = h_i(h_i^{N+1}(\alpha)) \simeq \tilde{g}_i^m(h_i^{N+1}(\alpha)) \simeq \tilde{g}_i^m(h_i^N(\alpha)). $$
This way we can show that $\tilde{g}^{qm_i}(h^N_1(\alpha))$ is contained in $S$ and $h^N_1(h^N_1(\alpha)) \simeq \tilde{g}^{qm_i}(h^N_1(\alpha))$ for any $q > 0$.

Also for any large $N$ and for any $0 \leq r \leq m_i - 1$, the curve $\tilde{g}_r^N h^N_1(\alpha)$ does not intersect $O(x_j)$ ($j \neq i$), and defines a proper homotopy class of curves in $S$. A homotopy between $\tilde{g}_r^N \alpha$ and $\tilde{g}^r(\alpha)$ post-composed by $h^N_1$ yields a proper homotopy in $S$ between $\tilde{g}_r^N h^N_1(\alpha)$ and $\tilde{g}^r(h^N_1(\alpha))$.

Consider the geodesic $\gamma_i$ in $S$ which is properly homotopic to $h^N_1(\alpha)$ for some large fixed $N$. Then for $v = qm_i + r$ we have

$$\tilde{g}_v^N \gamma_i \simeq \tilde{g}^v(h^N_1(\alpha)) \simeq \tilde{g}^v(\tilde{g}^{qm_i}(h^N_1(\alpha))) \simeq \tilde{g}^v(h^{N+r}_1(\alpha)) = \tilde{g}_v^N h^{N+r}_1(\alpha) \simeq \tilde{g}_v^N h^N_1(\alpha).$$

Since the $\tilde{g}_v^N h^N_1(\alpha)$ are disjoint for distinct values of $v$, the same is true for the geodesics $\tilde{g}_v^N h^N_1(\alpha)$, showing that $\gamma_i$ is a translation arc. Now the family $\tilde{g}_v^N h^N_1(\alpha)$ diverges in $S_i$, and hence in $S$. Homotoping these curves to geodesics in $S$, the family $\tilde{g}_v^N \gamma_i$ is also divergent, i.e., $\gamma_i$ is a forward proper translation arc in $S$.

The backward proper homotopy translation arc $\delta_i$ is constructed in a similar way. Also it is clear that homotopy translation arcs thus constructed are mutually disjoint.

Finally let us consider the case where the period $m_i = 1$. In the annulus $A$, we have a fixed point $y_1$. An essential simple closed curve based at $y_1$ is unique up to homotopy. The forward proper homotopy translation curve is constructed from this curve by an analogous method.

The rest of this section is devoted to the proof of the following theorem for the zero genus case.

**Theorem 4.5.** Let $\Sigma$ be a complete finite area hyperbolic surface of genus zero and let $g$ be a homeomorphism of $\Sigma$ which is isotopic to the identity. If there exists a periodic point $p$ of period $m > 1$ such that $R(p) = 0$ in $H_1(\Sigma; \mathbb{R})$, then the contractible fixed point set $\text{Fix}_c(g)$ is nonempty, and if further it is finite, it contains a fixed point of positive index.

**Proof.** Assume that the contractible fixed point set $\text{Fix}_c(g)$ is at most finite. If the curve $\delta(p) \circ \cdots \circ \delta(g^{m-1}(p))$ is contractible, then the canonical lift $\tilde{g}$ of $g$ admits periodic points, and thus a version of the Brouwer fixed point theorem [1] asserts the existence of fixed points of $\tilde{g}$ of positive index, completing the proof.

If not, there is a geodesic $\lambda$ of $\Sigma$ which is homotopic to the curve $\delta(p) \circ \cdots \circ \delta(g^{m-1}(p))$. Since $\Sigma$ is of genus zero, any connected component of the complement $\Sigma \setminus \lambda$ is either a disc or a punctured disc, possibly with more than one puncture. An argument similar to the proof of Theorem 4.2 can be applied to show the existence of a disc component whose boundary geodesics are concurrently oriented. The rest of the argument is the same as before. □

5. **Proof of Theorem 1**

This section contains a version of the Arnold conjecture for homeomorphisms of compact surfaces possibly with boundary and the proof of Theorem 1 as its application.
First of all we expose quickly the definition and some properties of rotation vectors. See [7,14] for more details. Let \( f \) be a continuous map of a compact metric space \( X \) homotopic to the identity. A lift \( \tilde{f} \) of \( f \) to the universal covering space \( \tilde{X} \) is called admissible if there exist a homotopy joining the identity to \( f \) whose lift to \( \tilde{X} \) joins the identity of \( \tilde{X} \) to \( \tilde{f} \). An admissible lift of \( f \) corresponds in a bijective way to a homotopy class of homotopies joining the identity of \( X \) and \( f \).

Fix once and for all an admissible lift \( \tilde{f} \) of \( f \) and a corresponding isotopy \( f_t \) joining the identity of \( X \) to \( f \). Given an \( f \)-invariant probability measure \( \mu \), let us define a homology class \( \mathcal{R}(\mu; \tilde{f}) \in H_1(X; \mathbb{R}) \) as follows. First of all recall that the homology group \( H_1(X; \mathbb{R}) \) is isomorphic to the group \( \text{Hom}([X, S^1], \mathbb{R}) \), where \([X, S^1]\) denotes the abelian group of homotopy classes of continuous maps from \( X \) to \( S^1 \). Choose a class \([v]\) and for any point \( x \in X \) consider the map \( v \circ \delta(x) : [0, 1] \to S^1 \), where \( \delta(x) : [0, 1] \to X \) is defined by \( \delta(x)(t) = f_t(x) \). The difference of the boundary values of an arbitrary lift of the map \( v \circ \delta(x) \) is a well defined continuous function of \( x \), denoted by \( \Delta(x; v) \). The assignment

\[
[X, S^1] \ni [v] \mapsto \int_X \Delta(x; v) \, d\mu \in \mathbb{R}
\]

is well defined independent of the choice of \( v \) in its homotopy class, and thus it defines a class in \( H_1(X; \mathbb{R}) \), denoted by \( \mathcal{R}(\mu; \tilde{f}) \) and called the rotation vector of the admissible lift \( \tilde{f} \) with respect to \( \mu \). Routine computation shows that this class is actually independent of the homotopy \( f_t \) in the right homotopy class. If the space \( X \) is a manifold, with boundary, and if the map \( h \) preserves the Lebesgue probability measure \( m \), then the rotation vector \( \mathcal{R}(m; \tilde{f}) \) is called the mean rotation vector. This way we obtain a mapping \( \mathcal{R}(\cdot; \tilde{f}) : \mathcal{M}(f) \to H_1(X; \mathbb{R}) \) from the space \( \mathcal{M}(f) \) of the \( f \)-invariant probability measures of \( X \), which is easily shown to be affine and continuous.

Returning to our main subject, let \( N \) be a compact oriented surface possibly with boundary. As is well known [4], two homeomorphisms of \( N \) are homotopic if and only if they are isotopic. Given a homeomorphism \( h \) of \( N \) isotopic to the identity by an isotopy \( h_t : \text{id} \simeq h \) and keeping the Lebesgue probability measure invariant, one has the mean rotation vector \( \mathcal{R}(m; h) \) of the admissible lift corresponding to the isotopy \( h_t \).

Besides this, if \( N \) has exactly one boundary component and if \( h \) is a homeomorphism of the interior \( \Sigma \) of \( N \) isotopic to the identity and keeping the Lebesgue measure \( m \) invariant, then the mean rotation vector \( \mathcal{R}(m; \tilde{h}) \) is defined as a class of \( H_1(\Sigma; \mathbb{R}) \) in the following way. Let us denote by \( M \) the closed surface obtained from \( N \) by collapsing the boundary to one point. Then the homomorphism \( h \), as well as the isotopy \( h_t \), is extended to \( M \), and thus the mean rotation vector is defined as a class of \( H_1(M; \mathbb{R}) \). On the other hand, there is an obvious identification of \( H_1(\Sigma; \mathbb{R}) \) with \( H_1(M; \mathbb{R}) \). This yields the definition of the mean rotation vector \( \mathcal{R}(m; \tilde{h}) \) in \( H_1(\Sigma; \mathbb{R}) \).

In case \( N \) has more than one boundary component, the mean rotation vector is not usually defined for a homeomorphism \( h \) of the interior \( \Sigma \) of \( N \), since \( h \) can be very wild near the boundary.

If the compact surface \( N \) is either a torus or an annulus, then any lift of the homeomorphism \( \tilde{h} \) is admissible. On the contrary if \( N \) has negative Euler number, then
an admissible lift is unique and is called the canonical lift of \( h \). In the latter case we denote \( \mathcal{R}(\cdot, \tilde{h}) \) by \( \mathcal{R}(\cdot, h) \). Recall that the projected image of the fixed point set of a lift \( \tilde{h} \) of \( h \) is denoted by \( \text{Fix}(h; \tilde{h}) \), and if \( N \) has negative Euler number and if \( \tilde{h} \) is the canonical lift, this set is denoted by \( \text{Fix}_c(h) \).

Theorem 5.1. Let \( N \) be a compact oriented surface possibly with boundary and \( \Sigma \) a once punctured closed oriented surface. Let \( h \) be a homeomorphism of \( N \) (respectively \( \Sigma \)) isotopic to the identity by an isotopy \( h_t : \text{id} \simeq h \), which corresponds to an admissible lift \( \tilde{h} \). Assume that \( h \) keeps the Lebesgue probability measure \( m \) invariant, with vanishing mean rotation vector of \( \tilde{h} \), \( \mathcal{R}(m; \tilde{h}) = 0 \). Then the set \( \text{Fix}(h; \tilde{h}) \) is nonempty, and if it is finite, it contains a fixed point of positive index.

Let us remark first of all that Theorem 5.1 implies Theorem 1 which asserts the existence of two fixed points of index one in the set \( \text{Fix}(f; \tilde{f}) \) for a homeomorphism \( f \) of a closed oriented surface \( M \) under the same assumption as Theorem 5.1. First of all applying Theorem 5.1 to the homeomorphism \( f \), one obtains a fixed point \( p \) of positive index in the set \( \text{Fix}(f; \tilde{f}) \). Now the locus of \( p \) by the isotopy \( f_t \) which corresponds to the admissible lift \( \tilde{f} \) is contractible in \( M \). Therefore by modifying the isotopy, one may assume that \( f_t \) keeps the point \( p \) invariant for any \( t \).

Let \( \Sigma = M \setminus \{p\} \), \( h = f|_\Sigma \), \( h_t = f_t|_\Sigma \), and \( \tilde{h} \) the admissible lift of \( h \) which corresponds to \( h_t \). Then the homeomorphism \( h : \Sigma \to \Sigma \) also satisfies the vanishing condition of the mean rotation vector, since \( H_1(\Sigma; \mathbb{R}) \) is identified with \( H_1(M; \mathbb{R}) \). Thus Theorem 5.1 applied to \( h \) implies the existence of a second fixed point of positive index in the set \( \text{Fix}(f; \tilde{f}) \).

The index of these two fixed points must actually be one since in general the index of an isolated fixed point of an area and orientation preserving homeomorphism is known to be less than 2 [15].

The rest of this section is devoted to the proof of Theorem 5.1. The case where the surface \( N \) (respectively \( \Sigma \)) has nonnegative Euler number will be treated at the end of this section. In the negative Euler number case the mean rotation number \( \mathcal{R}(m; h) \) is defined for the canonical lift of homeomorphism \( h \), which we assume to be zero. The argument is divided into two subcases according to the genus of \( N \) (respectively \( \Sigma \)).

Assume for the moment that \( N \) (respectively \( \Sigma \)) has negative Euler number and positive genus. Henceforth even when we are dealing with the compact surface \( N \), its interior is also denoted by \( \Sigma \).

Denote by \( D_x \) the Dirac measure at \( x \in \Sigma \). If the limit

\[
\mu(x) = \lim_{N \to \infty} (1/N) \sum_{i=1}^{N-1} D_{h^i(x)}
\]

exists as a measure of \( \Sigma \), it is called the asymptotic measure of \( x \). Notice that if \( p \) is a periodic point, then \( \mu(p) \) always exists and \( \mathcal{R}(\mu(p); h) \) coincides with \( \mathcal{R}(p; h) \) defined in Section 4.
The following proposition concerns approximation of the Lebesgue measure $m$ and will be useful later.

**Proposition 5.2.** For any $\delta > 0$, there exist recurrent points $x_i \in \Sigma$ for which the asymptotic measure $\mu(x_i)$ exists and $k_i > 0$ $(1 \leq i \leq n)$ such that

$$
\sum k_i = 1 \text{ and } \left| \mathcal{R}(m; h) - \sum k_i \mathcal{R}(\mu(x_i); h) \right| < \delta.
$$

**Proof.** Notice that the ergodic decomposition theorem as well as the Poincaré recurrence theorem are applicable to the homeomorphism $h$ on the open surface $\Sigma$. Therefore the Lebesgue measure $m$, viewed as a measure of $\Sigma$, can be arbitrarily approximated by a weighted sum $\sum k_i \mu_i$ of finitely many ergodic measures $\mu_i$ with $\sum k_i = 1$. Now $\mu_i$-a.e. point $x_i$ is recurrent, admits $\mu(x_i)$ and satisfies $\mu(x_i) = \mu_i$. This completes the proof of the proposition. \qed

We shall use the following generalized version of the concept of chain recurrence. Define $\Sigma_0 = \Sigma \setminus \text{Fix}_c(h)$, and consider any distance function $d$ obtained by an arbitrary complete Riemannian metric of $\Sigma_0$. Denote by $B_r(x)$ the open disc in $\Sigma_0$ of radius $r > 0$ and centered at $x$.

Let $E$ be the set of positive valued continuous functions $\varepsilon$ defined on $\Sigma_0$. A sequence $(x_1, \ldots, x_n)$ of points of $\Sigma_0$ is called an $\varepsilon$-chain if $d(h(x_i), x_{i+1}) < \varepsilon(h(x_i))$ $(1 \leq i \leq n - 1)$, an $\varepsilon$-cycle at $x_1$ if further $x_n = x_1$. A point $x \in \Sigma_0$ is called $E$-recurrent if for any $\varepsilon \in E$, there exists an $\varepsilon$-cycle at $x$. The homeomorphism $h$ is called $E$-transitive if given any two points $x, y$ of $\Sigma_0$ and any $\varepsilon \in E$, there exists an $\varepsilon$-chain from $x$ to $y$.

**Proposition 5.3.** Any point of $\Sigma_0$ is $E$-recurrent, and the homeomorphism $h$ is $E$-transitive.

**Proof.** The first half is proven using the fact that $h$ preserves a Lebesgue probability measure and therefore any point of $\Sigma_0$ is nonwandering. The second half follows from the connectedness of $\Sigma_0$. The details are left to the reader. \qed

For any small function $\varepsilon$ and any $\varepsilon$-cycle $Z = (x_1, \ldots, x_n)$, define the rotation vector $\mathcal{R}(Z; h)$ of $Z$ as the homology class, divided by the period of $Z$, of a closed curve obtained by concatenating in an obvious way the loci of points of $Z$ by the isotopy $h_t$ and small paths joining $x_i$ with $x_{i+1}$.

Roughly the plan of the proof is as follows. First we find out for any $\varepsilon \in E$ $\varepsilon$-cycles $Z_j$ for which the rotation vectors $\mathcal{R}(Z_j; h) \in H_1(\Sigma; \mathbb{R})$ is filling. Next we shall modify the homeomorphism $h$ slightly, away from the contractible fixed point set $\text{Fix}_c(h)$, without creating new contractible fixed points so that the $\varepsilon$-cycles become honest periodic orbits for the new homeomorphism $g$.

The following lemma, due to Franks [7], solves the first half of the above plan. Let us denote by $SH(\Sigma; \mathbb{R})$ the space of homological directions $(H_1(\Sigma; \mathbb{R}) \setminus \{0\})/\mathbb{R}_+$. 
Lemma 5.4. For any $\varepsilon \in E$, the classes of the rotation vectors of all the $\varepsilon$-cycles form a dense subset of $SH_1(\Sigma; \mathbb{R})$.

Proof. Consider an arbitrary class $c \in H_1(\Sigma; \mathbb{R})$ represented by a simple closed curve $\gamma$ contained in $\Sigma_0$. Of course all such classes constitute a dense subset of $SH_1(\Sigma; \mathbb{R})$. Let $U$ be a tubular neighbourhood of $\gamma$ whose closure is contained in $\Sigma_0$. Then there exists a homeomorphism $\varphi$ of $\Sigma$ isotopic to the identity, keeping $\gamma$ invariant, which is a small rotation when restricted to $\gamma$ and is the identity outside $U$. One may further assume that $\varphi$ keeps the Lebesgue measure $m$ invariant, $\mathcal{R}(m; \varphi) = \alpha c$ for some $\alpha > 0$ and that $d(x, \varphi(x)) < \varepsilon(x)$.

Now it follows directly from the definition of the rotation vector that

$$\mathcal{R}(m; \varphi h) = \mathcal{R}(m; \varphi) + \mathcal{R}(m; h) = \mathcal{R}(m; \varphi) = ac.$$ 

Notice that an orbit of $\varphi h$ is an $\varepsilon$-chain for $h$. Let $x_i$ be the points in Proposition 5.2 for the homeomorphism $\varphi h$, which we may assume to be contained in $\Sigma_0$. Thus $\sum_{i} k_i \mathcal{R}(\mu(x_i); \varphi h)$ is arbitrarily near $ac$.

Since $x_i$ is a recurrent point for $\varphi h$, one can construct from its orbit by $\varphi h$ an $\varepsilon$-cycle $Z_i$ of $h$ at $x_i$ such that $\mathcal{R}(Z_i; h)$ is arbitrarily near $\mathcal{R}(\mu(x_i); \varphi h)$. On the other hand by Proposition 5.3 any point $x_i$ is joined to a base point $x_0$ by an $\varepsilon$-chain (for $h$) and conversely the base point $x_0$ to any point $x_i$. Concatenating these $\varepsilon$-chains and appropriate iterates of the $\varepsilon$-cycles $Z_i$, one gets the desired $\varepsilon$-cycle $Z$ whose rotation vector $\mathcal{R}(Z; h)$ represents an element of $SH_1(\Sigma; \mathbb{R})$ which is arbitrarily near the class $[c]$.

By virtue of Theorem 4.2, the following proposition implies Theorem 5.1.

Proposition 5.5. There exists a homeomorphism $g$ of $\Sigma$ isotopic to $h$ with the following properties.

1. We have $\text{Fix}_c(g) = \text{Fix}_c(h)$ and $g = h$ in a neighbourhood of this set.
2. The homeomorphism $g$ admits periodic points $p_j$ of period $m_j$ ($1 \leq j \leq q$) such that the set $\{m_j \mathcal{R}(p_j; g)\}$ is filling.

Proof. Let $\tilde{h}$ be the canonical lift of $h$ to the universal covering $\pi : \mathbb{D}^2 \to \Sigma$, and let $\tilde{d}$ be the lift to $\pi^{-1}(\Sigma_0)$ of the metric $d$ of $\Sigma_0$ defined above. The positive valued continuous function $\rho$ on $\Sigma_0$ is defined by $\rho(x) = \tilde{d}(\tilde{x}, \tilde{h}^{-1}(\tilde{x}))$. This is well defined independent of the choice of the lift $\tilde{x}$ of $x$.

Choose $\varepsilon \in E$ so that $\varepsilon(x) < \rho(x)$ for any $x \in \Sigma$. Define $\varepsilon_1 \in E$ so that any $C^1$ curve $\gamma : [0, 1] \to \Sigma_0$ satisfying $\|\gamma'(t)\| < \varepsilon_1(\gamma(t))$ (for all $t$) has length $< \varepsilon(\gamma(0))$. The existence of such an $\varepsilon_1$ is left to the reader. Next $\varepsilon_2 \in E$ is defined by

$$\varepsilon_2(x) = \frac{1}{2} \inf \{ \varepsilon_1(y) \mid y \in B_{\varepsilon_1(x)}(x) \}.$$ 

Choose any filling subset $C = \{c_1, \ldots, c_q\}$ of $H_1(\Sigma; \mathbb{R})$. (The existence is clear.) By Lemma 5.4 there exists an $\varepsilon_2$-cycle $Z_i = (x_i^q)$ of period $m_i$ such that the classes of $\mathcal{R}(Z_i; h)$ and $c_i$ are arbitrarily near in $SH_1(\Sigma; \mathbb{R})$. Then the set $\{m_i \mathcal{R}(Z_i; h)\}$ is also
filling, since this is an open condition. We may assume that the points \( x_i \) in the \( \varepsilon_2 \)-cycles \( Z^j \) are mutually distinct for any \( i \) and \( v \).

For any point \( x_i \), let \( \delta^j_v : [0, 1] \to \Sigma \) be the minimal geodesic such that \( \delta^j_v(0) = h(x_i) \) and \( \delta^j_v(1) = x_{i+1} \). Of course the length of \( \delta^j_v \) is smaller than \( \varepsilon_2(h(x_i)) \) and \( \delta^j_v \) is contained in \( \Sigma_0 \). Consider the product \( \Sigma_0 \times [0, 1] \) and the curve \( \hat{\delta}^j_v \) in it defined by \( \hat{\delta}^j_v(t) = (\delta^j_v(t), t) \). By modifying slightly the curves if necessary, one may assume that they are disjoint. Choose a small tubular neighbourhoods of \( \hat{\delta}^j_v \) which are mutually disjoint. Define a vector field \( X = (Y_t, \partial/\partial t) \) such that \( X \neq (0, \partial/\partial t) \) outside the union of the tubular neighbourhoods and \( X \) is tangent to \( \hat{\delta}^j_v \). An appropriate choice of the vector field guarantees that \( \|Y_t(x)\| < \varepsilon_1(x) \). Define a homeomorphism \( \psi \) of \( \Sigma \) by mapping the initial point of the orbit of \( X \) to its terminal point. Of course \( \psi \) maps the point \( h(x_i) \) to \( x_i \). By the choice of \( \varepsilon_1 \) one has \( d(x, \psi(x)) < \varepsilon(x) < \rho(x) \). This implies no new creation of the contractible fixed points for the composite map \( g = \psi \circ h \). Also the modification was done outside a neighbourhood of \( \text{Fix}_c(h) \). Now the \( \varepsilon_2 \)-chain \( Z^j \)'s become periodic orbits for \( g \) with the same rotation vectors, which are filling. This shows the proposition. \( \square \)

Let us turn to the case of negative Euler number and zero genus. The same argument as above shows the existence of \( \varepsilon_2 \)-cycles \( Z^j \) for the homeomorphism \( h \) whose rotation vectors contain 0 in the interior of its convex hull. Then 0 can be expressed as a linear combination of these rotation vectors with integer coefficients. One may further assume that the \( \varepsilon_2 \)-cycles starts at the same base point. Thus concatenating these \( \varepsilon_2 \)-cycles, one obtains an \( \varepsilon_2 \)-cycle whose rotation vector vanishes. Then as before one can modify \( h \) slightly to a new homeomorphism for which the \( \varepsilon_2 \)-cycle becomes a periodic orbit. Theorem 5.1 is obtained by applying Theorem 4.5.

Finally consider the case where the surface has nonnegative Euler number. If the surface is the sphere \( S^2 \), then the Lefschetz fixed point theorem implies the existence of fixed points of positive index. If the surface is an open disc, by considering the one point compactification, one gets an area preserving homeomorphism \( h \) of the sphere. Now the Pelikan–Slaminka theorem implies the existence of a fixed point of index one in the open disc.

The case where the surface has vanishing Euler number is completely similar to the above genus zero hyperbolic case. That is, one can modify the homeomorphism \( h \) by another one \( g \) with a periodic orbit whose rotation vector vanishes. This is done away from the set \( \text{Fix}(h; \tilde{h}) \) in a way not to create a new fixed point. Since in this case there is no distinction between the homology and homotopy, the locus of the periodic orbit by the isotopy \( h_t \) is contractible, and therefore yields a periodic orbit for the corresponding lift \( \tilde{h} \), of period > 1. The theorem follows from the Brouwer fixed point theorem.

Acknowledgements

Hearty thanks are due to the referee, who pointed out mistakes of various characters, which existed densely in the previous version.
References