Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Master Thesis

Forcing relations for periodic orbits of surface homeomorphisms

Arnaud Maret
supervised by prof. Will J. Merry

Abstract. This paper surveys an abstract framework introduced by Boyland in 1989 to describe forcing relations between periodic orbits in iterated surface dynamics as a 2-dimensional analogue of the Sharkovskii Theorem. The starting point is the classification of isotopy classes of surface homeomorphisms by Thurston and Nielsen. This result is presented together with an introduction to pseudo-Anosov maps. It is used to derive a stability property of periodic orbits after isotopic perturbations. The forcing is based on the comparison of selected characteristics of the orbits. In dimension two, a periodic orbit is specified by a braid corresponding to the time evolution of the orbit in the suspension. A strategy to describe the dynamical preorder on braids in an algebraic manner is outlined along with some considerations on the antisymmetry of the forcing relation.

June 28, 2018

Pour ceux qui ne se baignent jamais une fois dans le même fleuve

Introduction

Dynamical systems are a branch of mathematics that models deterministic evolution. The general abstract setting consists in a space that contains all the possible configurations of the system and a rule that specifies how the system can evolve from one configuration to another. Mathematically, a discrete dynamical system is a metric space X together with a continuous map T from X to itself. The forward evolution of X in time is associated with iterations of the map T. After n identical intervals of time have elapsed, the space X is in the configuration $T^n(X)$ where $T^n := T \circ \cdots \circ T$ is the map T composed n times with itself. Backward iterates are considered only under the assumption that T is invertible. When the rule that defines the evolution is assimilated to the action of the real numbers rather than the integers, then it is referred as continuous dynamics instead of discrete time dynamics.

The consecutive states of a point $x \in X$ form the orbit of x. Concretely, if $T\colon X \to X$ denotes a discrete dynamical system, the T-orbit $o(x, T)$ of a point x is the set of images of x under all iterations of T. When T is invertible, negative iterates are also accounted for in the orbit of x. A point x is said to be periodic if it comes back to its initial position after finitely many iterations. The smallest integer $p \geq 1$ such that $T^p(x) = x$ is called the period of x. If $p = 1$, then x is called a fixed point of T.

Understanding the orbit structure of a system is fundamental in dynamical systems theory. A natural strategy to further this goal assigns coordinates to periodic orbits which are then regarded up to an equivalence relation that identifies orbits with the same coordinates. An appropriate choice of coordinates provides a fairly rich framework to examine the implications of the existence of a periodic orbit. This is referred as a dynamical forcing relation at the level of periodic orbits and is the main motivation of this survey.

Historically, the order relation was designed by Philip Boyland as a generalization of the long-known Sharkovskii dynamical ordering of the natural integers seen as admissible periods for continuous transformations of the interval. This paper focuses on a two-dimensional analogue of the Sharkovskii dynamical ordering. Namely, spaces of interest are surfaces whose dynamics are regulated by homeomorphisms. A fair amount of work on surface dynamics is due to Thurston and Nielsen among which a concise classification of surface automorphisms up to isotopy. This material is presented in Chapter 1 together with some considerations on foliations and topological entropy.

The Thurston-Nielsen theory studies isotopy classes of homeomorphisms and provides a canonical representative in each class. This approach addressed the following question: what type of periodic orbits is present in the appropriate sense in every representative of an isotopy class ? When such a list can be made explicit, it is natural to search for a minimal representative that only exhibits the dynamics present in every element of the class. Isotopy stability and minimal representatives are investigated in Chapter 2.

Chapter 3 is about the dynamical forcing order itself. The abstract framework is presented and then studied in the particular case of surfaces. It is immediately clear that specifying orbits with their length only, as it is done in the one-dimensional case, does not lead to a valuable forcing relation. A richer invariant is the isotopy class of the map relative to the orbit. Geometrically, it can be seen as an isotopy class of a collection of strings representing the time evolution of the orbit in the suspension. Such a choice of coordinates equips the forcing order with an interesting algebraic structure. Chapter 3 ends with the exposition of some clues towards a fully algebraic formulation of the forcing defined dynamically.

Among every surface, the disk has been most investigated. Reasons for that are a purely algebraic description of the set of coordinates (not to be confused with an algebraic description of the forcing order) and various combinatorial encodings of disk automorphisms. The forcing relations between periodic orbits of length 3 are particularly well understood. What makes period 3 special relies in the fact that any disk homeomorphism with an orbit of length 3 descends from an Anosov (or hyperbolic) toral automorphism. Those in turn descend from linear transformations of the plane. These results are surveyed in the final chapter.

The reader is assumed to have basic knowledge in dynamical systems and differential topology. Being familiar with the notions of surface, homotopy and isotopy, and universal covering is appreciable. Standard results in geometry and topology are used without any further reference.

None of the results presented in this paper is claimed to be original. The references provided do not always refer to the earliest publication but sometimes to the source used by the author. Most of the material studied in this paper takes its source from a survey article by Boyland [4], the introductory book on mapping class groups by Farb and Margalit [8], the famous *Travaux de Thurston sur les surfaces* by Fathi, Laudenbach and Poénaru [10], and a last article by Matsuoka [24].

Acknowledgements

The author is particularly thankful towards his supervisor Prof. Will J. Merry who agreed to conduct this master thesis on low-dimensional dynamics. Precious time was spent proofreading the present paper. His office door was always open for helpful conversations.

Thanks to P. Boyland, T. Hall and J. Smillie for sharing unpublished personal work. This survey also benefited from a fruitful online conversation with L. Mosher.

Notation and terminology

The set of natural integers is denoted by $\mathbb{N}$, the set of integers by $\mathbb{Z}$, the set of real numbers by $\mathbb{R}$ and the set of complex numbers by $\mathbb{C}$. The natural logarithm on the set of positive real numbers $\mathbb{R}_+$ is written log. The complex modulus in $\mathbb{C}$ is denoted $|\cdot|$. The real and imaginary parts of a complex number z are written respectively $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$. We write $\mathbb{R}^2$ for the 2-dimensional Euclidian plane with Euclidian norm $\|\cdot\|$. The unit interval $[0, 1]$ is usually abbreviated by I.

The cardinality of a finite set E is denoted $\sharp E$. If A is a subset of a set B, we write $A \subset B$. The Cartesian product of any abstract set $n \geq 1$ times with itself is denoted $E^{\times n}$.

If X is a topological space and $Y \subset X$ is a subspace, then $\overline{Y}$ stands for the closure of Y and $\mathrm{int}(Y)$ its interior. The topological boundary of Y is denoted by $\partial Y := \overline{Y} \setminus \mathrm{int}(Y)$. If (X, d) denotes a metric space, the open ball of center $x \in X$ and radius $r > 0$ is written $B(x, r) := \{y \in X : d(x, y) < r\}$.

The set of path connected components in X is written $\pi_0(X)$. The fundamental group of X at a point $x_0 \in X$ is denoted by $\pi_1(X; x_0)$ and the nth (singular) homology group by $H_n(X)$. The group law inside $\pi_1(X; x_0)$ is defined to be the concatenation of loops and is denoted with $*$.

Let $T\colon X \to X$ be a dynamical system. The set of fixed points of T is denoted $\mathrm{Fix}(T)$. The set of periodic points of T is written $\mathrm{Per}(T)$. A subspace $Y \subset X$ is *T-invariant* if $T(Y) \subset Y$. If $Y \subset X$ is an invariant subspace, then the *restriction of T to Y* is the map $T\restriction_Y\colon Y \to Y$ defined by $x \mapsto T(x)$. If $S\colon Y \to Y$ denotes a second dynamical system, we say that T is *semi-conjugate* to S if there exists a continuous map $\eta\colon Y \to X$ with dense image and such that the following diagram commutes.

$$\begin{tikzcd} Y \arrow[r, "S"] \arrow[d, "\eta"'] & Y \arrow[d, "\eta"] \\ X \arrow[r, "T"'] & X \end{tikzcd}$$

If in addition η is a homeomorphism, then T and S are said to be *conjugate*.

In any category, the symbol $\cong$ refers to two isomorphic objects (e.g. isomorphic groups or homeomorphic topological spaces).

Contents

Chapter 1

The Thurston-Nielsen classification

We are interested in studying periodic orbits of surface homeomorphisms. The combined work of Jakob Nielsen (1890-1959) and William Thurston (1946-2012, Fields medallist in 1982) provides a concise classification of surface homeomorphisms. In order to understand this classification, we start with a short recap on surfaces and the relation of isotopy between surface homeomorphisms.

1.1 Surfaces: generalities and hyperbolic metric

A *surface* is a 2-manifold. A surface is therefore allowed to have boundary components. The manifold boundary of a surface M is the complement of $\mathrm{int}(M)$ in M. It must not be confused with the topological boundary. When M denotes a surface, then ∂M will conventionally refer to the manifold boundary. If a compact surface has no boundary, then it is called a *closed surface*.

Standing assumptions. The following assumptions will prevail throughout the present paper. Unless otherwise indicated, all the surfaces are assumed to be compact, connected and orientable. Moreover, any homeomorphism of a surface is assumed to be orientation-preserving.

Two homeomorphisms f_1 and f_2 of a surface M are said to be *isotopic* if they are homotopic via a family of homeomorphisms of M. Precisely, f_1 is isotopic to f_2 if there is a continuous map $h_\bullet \colon M \times I \to M$ such that the maps $h_t \colon M \to M$ for $t \in I$, defined by $h_t(x) := h_\bullet(x,t)$, are all homeomorphisms of M with $h_0 = f_1$ and $h_1 = f_2$. We say that h_t is an isotopy from f_1 to f_2 and we write $f_1 \simeq f_2$. Observe that $h_{1-t} \colon M \to M$ defines an isotopy from f_2 to f_1.

A continuous map $\gamma \colon I \to M$ is called without distinction an *arc* or a *path*. Under the same terminology we refer to the map γ or its image inside M. The arc γ has two endpoints $\partial\gamma = \{\gamma(0), \gamma(1)\}$. Two arcs γ_1 and γ_2 sharing the same endpoints are *homotopic with fixed endpoints* if there is a homotopy $h_t \colon I \to M$ such that $h_0 = \gamma_1$, $h_1 = \gamma_2$, $h_t(0) = \gamma_1(0) = \gamma_2(0)$ and $h_t(1) = \gamma_1(1) = \gamma_2(1)$. The notions of arc, path and homotopy with fixed endpoints persist when M is replaced by any topological space.

Sometimes it is useful to allow a compact surface M to have a finite number of punctures. A *punctured surface* is obtained from a compact surface by removing finitely many points from its interior. A punctured surface shall not be confused with a marked surface. A *marked surface* is a surface M together with a finite collection of points $X \subset M$. If f is a homeomorphism of M with $f(X) = X$, then we say that f is a homeomorphism of the marked surface and we write $f\colon (M, X) \to (M, X)$.

Naturally, a punctured surface ceases to be compact. One way to compactify a punctured surface is to *blow-up* the holes. The following procedure is presented by Matsuoka in [24] and by Boyland in [4, Section 1.6].

Definition 1.1 (blow-up)**.** Let M be a surface and $X \subset \operatorname{int}(M)$ be a finite collection of points. The *blow-up of M at X* is the compact surface $\widehat{M}$ obtained from $M \setminus X$ by attaching a boundary circle at each point $x \in X$. There is a canonical projection $\pi\colon \widehat{M} \twoheadrightarrow M$ that acts trivially on $M \setminus X$ and projects every boundary circle to the corresponding point of X.

Let $f\colon M \to M$ be a homeomorphism of M that fixes X. If f is extendible to a homeomorphism $\widehat{f}\colon \widehat{M} \to \widehat{M}$ in the sense that the following diagram commutes,

$$\begin{tikzcd}
\widehat{M} \arrow[r, "\widehat{f}"] \arrow[d, "\pi"'] & \widehat{M} \arrow[d, "\pi"] \\
M \arrow[r, "f"'] & M,
\end{tikzcd}$$

then we call $\widehat{f}$ the *blow-up of f at X*.

Proposition 1.1. *Let M denote a surface. Let $f\colon M \to M$ be a homeomorphism of M and $X \subset \operatorname{int}(M)$ denote a finite invariant set for f. If f is smooth and non-singular at every point in X, then f has the blow-up $\widehat{f}\colon \widehat{M} \to \widehat{M}$ at X. Furthermore, if f has no fixed point in X, then* $\operatorname{Fix}(\widehat{f}) = \operatorname{Fix}(f)$.

Proof. Blowing-up is a local transformation. Therefore, it is sufficient to consider the case of a smooth homeomorphism $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ at 0 with $f(0) = 0$. The blow-up of $\mathbb{R}^2$ at 0 in polar coordinates is expressed as $\widehat{\mathbb{R}^2} := [0, +\infty) \times S^1$ where $S^1 \subset \mathbb{R}^2$ denotes the unit circle. After expressing $f(r, \theta) = f(r\cos(\theta), r\sin(\theta))$ in polar coordinates, we define $\widehat{f}\colon \widehat{\mathbb{R}^2} \to \widehat{\mathbb{R}^2}$ by

$$\begin{cases} \widehat{f}(r, \theta) = f(r, \theta), & \text{if } r > 0, \\ \widehat{f}(0, \theta) = (0, Df_0(\theta)). \end{cases}$$

Here Df_0 denotes the map induced by the differential of f at 0 on angles. If arg denotes the angle coordinate of a vector, then

$$Df_0(\theta) := \arg Df_0 \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

Said differently, $Df_0(\theta)$ denotes the angle between the horizontal basis vector and the image of the vector $(\cos(\theta); \sin(\theta))$ under the differential of f at 0. Since f is non-singular at 0 by assumption, $\widehat{f}$ is a homeomorphism of $\widehat{\mathbb{R}^2}$. $\square$

Connected, compact and orientable surfaces have been classified. This result is traditionally attributed to August Ferdinand Möbius (1790-1868).

Theorem 1.2 (Möbius)**.** *Any closed, connected and orientable surface is homeomorphic to the connect sum of a 2-dimensional sphere with $g \geq 0$ tori.*

Any compact, connected and orientable surface is obtained from a closed surface by removing $b \geq 0$ open disks with disjoint closure.

In the statement of Theorem 1.2, g is called the *genus* of the surface and b is the number of boundary components. If M is a punctured surface, then we denote the number of punctures by $n \geq 0$. Theorem 1.2 tells us that, up to homeomorphism, a surface is determined by the three integers (g, b, n). Therefore, the following quantity, known as the *Euler characteristic* of the surface M, is a topological invariant:

$$\chi(M) := 2 - 2g - b - n.$$

Throughout this paper, the following notations concerning usual surfaces and other spaces are adopted. Let $S^1 = \{z : |z| = 1\}$ denote the unit circle and S^2 the 2-dimensional sphere $(g = 0, b = 0, n = 0)$. The 2-dimensional disk $(g = 0, b = 1, n = 0)$ will be denoted by $D^2 = \{z : |z| \leq 1\}$ and T^2 will stand for the genus-1 torus $(g = 1, b = 0, n = 0)$. Finally, we write A for the annulus $(g = 0, b = 2, n = 0)$ obtained from the sphere S^2 by removing two open disks with disjoint boundaries. Observe that the surfaces listed above are exactly the surfaces (without puncture) of nonnegative Euler characteristic. The punctured surfaces with nonnegative Euler characteristic are precisely the one time punctured disk $D^2 \setminus \{\star\}$ $(g = 0, b = 1, n = 1)$, the one time punctured sphere $S^2 \setminus \{\star\} \cong \mathbb{R}^2$ $(g = 0, b = 0, n = 1)$ and the two times punctured sphere $\mathbb{R}^2 \setminus \{\star\}$ $(g = 0, b = 0, n = 2)$.

Most of our work later will take place in the universal cover rather than on the surface itself. A recap on universal covers is given in Subsection 1.5.3. The 2-sphere being simply connected is its own universal cover. The 1-torus can be constructed from a square fundamental domain. Copying the square across each edge, we observe that the universal cover of T^2 is the Euclidian plane $\mathbb{R}^2$. For closed surfaces of higher genus, advantage can be taken of the existence of a hyperbolic structure to identify their universal cover with the hyperbolic plane $\mathbb{H}^2$.

Recall that the hyperbolic plane $\mathbb{H}^2$ can be either modelled by the upper half plane $\{z = x + iy : y > 0\}$ with the metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}$$

or by the unit disk $\{z : |z| < 1\}$ with the metric

$$ds^2 = 4\frac{dx^2 + dy^2}{(1 - x^2 - y^2)^2}.$$

The group $\mathrm{Isom}^+(\mathbb{H}^2)$ of orientation-preserving isometries of $\mathbb{H}^2$ can be identified with the group $\mathrm{PSL}(2; \mathbb{R})$ of 2×2 matrices with real entries and determinant 1 modulo its center $\{\pm I_2\}$ (I_2 stands for the identity matrix of dimension 2).

The group isomorphism is given by

$$\begin{aligned} \mathrm{PSL}(2;\mathbb{R}) &\longrightarrow \mathrm{Isom}^+(\mathbb{H}^2) \\ \pm\begin{pmatrix} a & b \\ c & d \end{pmatrix} &\longmapsto \left(z \mapsto \frac{az+b}{cz+d} \right). \end{aligned}$$

In the upper half plane model, the geodesics are exactly the vertical rays and the upper half circles centred on the real axis. In particular, any two points are joined by a unique geodesic arc.

Proposition 1.3. *An orientation-preserving isometry of $\mathbb{H}^2$ that fixes two distinct points is the identity.*

Proof. An isometry f of $\mathbb{H}^2$ that fixes two distinct points also fixes the geodesic γ through these two points. Up to composing with another orientation-preserving isometry, we can assume that γ is a vertical ray. Any point outside of γ is either mapped to itself or to its symmetric with respect to γ because the distance from this point to γ is conserved by f. Therefore, f is either the identity or the reflection through γ. Since we assumed that f preserves the orientation, f is the identity. $\square$

Any surface of negative Euler characteristic admits a hyperbolic metric [8, Theorem 1.12]. Recall that a hyperbolic metric on a surface M is a complete, finite-area Riemannian metric with constant curvature -1 and such that ∂M is totally geodesic (i.e. geodesics in ∂M are geodesics in M). In particular, if M is closed and admits a hyperbolic metric, then its universal cover $\widetilde{M}$ is isometric to $\mathbb{H}^2$. If $\partial M \neq \emptyset$, then $\widetilde{M}$ is isometric to a totally geodesic subspace of $\mathbb{H}^2$.

For instance, the genus-2 torus admits a regular octagon as fundamental domain. The octagon is for instance obtained by cutting along the marked curves on the surface in Figure 1.1. The universal cover is then obtained by copying the fundamental domain across each edge. This process suggests that the universal cover of the genus-2 torus is the open disk model of $\mathbb{H}^2$.

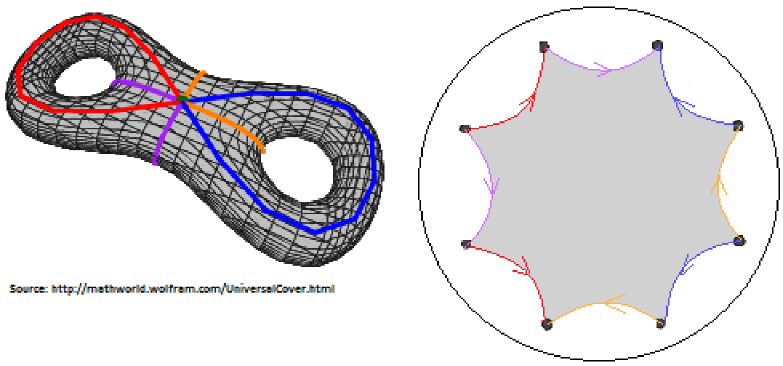

Figure 1.1: The left-hand side picture represents a regular hyperbolic octagon in $\mathbb{H}^2$. If we identify the coloured edges, we obtain a fundamental domain for the genus-2 torus.

A standard result in Riemannian geometry states that any homotopy class of closed curves in a compact manifold M contains a length-minimizing geodesic.

A *closed curve* in a surface M is identified with the image of a continuous map $S^1 \to M$. Under the further assumption that M is a surface equipped with a hyperbolic metric, then this geodesic is uniquely determined. Indeed, assuming there were two geodesic representatives, since they must stay at bounded distance of each other, their corresponding lift in $\mathbb{H}^2$ would coincide. One should however also consider the case where the geodesic representative of a homotopy class is a multiple of a simple closed geodesic. Recall that a closed curve is called *simple* if the mapping $S^1 \to M$ that defines it is injective. Farb and Margalit give a detailed proof of the uniqueness in [8, Proposition 1.3]. Consequently, any isometry of a hyperbolic surface isotopic to the identity must be the identity. The following two results are part of Theorem 3.19 in [10].

Lemma 1.4. *Let M denote a compact surface equipped with a hyperbolic metric. If $f\colon M \to M$ is an isometry of M isotopic to the identity, then f is the identity.*

Proof. By assumption, f fixes every homotopy class of closed curves in M. Since f maps geodesics to geodesics, it fixes the unique geodesic representative in every class. The desired conclusion follows after lifting f to an isometry of the universal cover of M that is a totally geodesic subspace of $\mathbb{H}^2$ and applying Proposition 1.3. □

Proposition 1.5. *Let M denote a compact surface equipped with a hyperbolic metric ρ. The group* $\mathrm{Isom}(M)$ *of isometries of M is finite.*

Proof. If we equip the set M^M of all maps $M \to M$ with the topology of pointwise convergence, then we obtain a compact topological space by Tychonov Theorem. The strategy is to prove that $\mathrm{Isom}(M)$ is discrete and closed (therefore compact) in M^M. Since a discrete compact space is necessarily finite, the desired conclusion follows.

We first observe that the subspace topology induced on $\mathrm{Isom}(M)$ from the pointwise topology coincides with the topology of uniform convergence. Let $d_\rho(\cdot,\cdot)$ be the distance function associated to the metric ρ. Let $f \in \mathrm{Isom}(M)$ and $\varepsilon > 0$. Using the compactness of M, we can find a finite subset $X \subset M$ such that for any $x \in M$ there is $z \in X$ such that $d_\rho(x, z) < \varepsilon/3$. Let

$$U := \{g \in \mathrm{Isom}(M) : d_\rho(g(z), h(z)) < \varepsilon/3, \forall z \in X\}.$$

By definition, U is open in the topology of pointwise convergence. Moreover, for any $g \in U$ and $x \in M$, we can find $z \in M$ such that $d_\rho(x, z) < \varepsilon/3$. The triangle inequality gives

$$\begin{aligned} d_\rho(g(x), h(x)) &\leq d_\rho(g(x), g(z)) + d_\rho(g(z), h(z)) + d_\rho(h(z), h(x)) \\ &= d_\rho(x, z) + d_\rho(g(z), h(z)) + d_\rho(x, z) \\ &< \varepsilon. \end{aligned}$$

Therefore, U is contained in the open ball centred at h with radius ε in the topology of uniform convergence. The reverse inclusion between the topologies is obvious. In particular, we deduce that $\mathrm{Isom}(M)$ is closed, and hence compact, in M^M.

Furthermore, Lemma 1.4 implies that the connected component of the identity in $\mathrm{Isom}(M)$ reduces to the identity itself. Therefore, $\mathrm{Isom}(M)$ is discrete. □

As a corollary of Proposition 1.5, we deduce that if f is an isometry of a surface M with respect to some hyperbolic metric on M, then there is an integer $n \geq 1$ such that f^n is the identity. Such an f is said to have *finite-order* (see Definition 1.3).

1.2 Mapping class group

Given a surface M, let $\mathrm{Homeo}(M)$ stand for the topological group of homeomorphisms of M endowed with the compact-open topology. Since M is assumed to be compact, $\mathrm{Homeo}(M)$ is metrizable via the metric

$$d(f, g) := \sup_{x \in M} d_M(f(x), g(x)).$$

where d_M is any metric on M. The subgroup of orientation-preserving homeomorphisms is denoted $\mathrm{Homeo}^+(M)$ and $\mathrm{Homeo}^+(M, \partial M)$ stands for the subgroup of $\mathrm{Homeo}^+(M)$ constituted of the elements whose restriction to ∂M is the identity.

A convenient property of the compact-open topology implies that two elements of $\mathrm{Homeo}(M)$ are in the same connected component if and only if there are isotopic. The same property holds for $\mathrm{Homeo}^+(M, \partial M)$ where isotopies are required to preserve the orientation and fix ∂M pointwise. The connected component of the identity in $\mathrm{Homeo}(M)$ is denoted $\mathrm{Homeo}_0(M)$. Observe that it is a normal subgroup of $\mathrm{Homeo}(M)$.

Another handy property of surfaces with negative Euler characteristic asserts the contractibility of the connected component of the identity in the group of homeomorphisms. It is a result due to Mary-Elizabeth Hamstrom and dates back to the 60s [13, Theorems 5.1 & 5.2].

Theorem 1.6 (Hamstrom)**.** *Let M denote a compact surface up to a finite number of interior punctures. If the Euler characteristic of M is negative, then* $\mathrm{Homeo}_0(M)$ *and* $\mathrm{Homeo}_0(M, \partial M)$ *are contractible.*

It is common in the literature to say that an isotopy h_t is a *deformation* of a second isotopy h'_t with $h'_0 = h_0$ and $h'_1 = h_1$, if the corresponding arcs in $\mathrm{Homeo}(M)$ are homotopic with fixed endpoints. As a consequence of Theorem 1.6, we observe that if $\chi(M) < 0$, then isotopies between maps in $\mathrm{Homeo}_0(M)$ and $\mathrm{Homeo}_0(M, \partial M)$ are unique up to deformation.

If we look at the homeomorphisms of a surface only up to isotopy, then we get a fundamental object known as the mapping class group of the surface.

Definition 1.2 (mapping class group)**.** The *mapping class group* of a potentially punctured surface M is defined as the quotient group

$$\begin{aligned} \mathrm{MCG}(M) :&= \mathrm{Homeo}^+(M, \partial M)/\mathrm{Homeo}_0(M, \partial M) \\ &= \mathrm{Homeo}^+(M, \partial M)/\text{isotopy} \\ &= \pi_0(\mathrm{Homeo}^+(M, \partial M)). \end{aligned}$$

The elements of the mapping class group are denoted by $[f]$ where f lies in $\mathrm{Homeo}^+(M, \partial M)$ and represents the class. The group law is the composition: $[f][g] := [f \circ g]$.

When M has nonempty boundary, there are two different common approaches in the literature. Some authors (see for instance Hall [11]) define the mapping class group without restricting to homeomorphisms that fix the boundary pointwise. Definition 1.2 as stated in the present paper is in the tradition of Farb and Margalit [8]. The careful reader has observed that isotopies are required to fix the boundary pointwise too.

It is sometimes useful to consider only classes of isotopic homeomorphisms that leave a finite set $X \subset \mathrm{int}(M)$ invariant, i.e. homeomorphisms of the marked surface (M, X). In this context, we also require that all isotopies leave X invariant. Such an isotopy is said to be *relative to* X. The group of isotopy classes relative to X is denoted by $\mathrm{MCG}(M \text{ rel } X)$. It corresponds to the definition of the mapping class group of the punctured surface $M \setminus X$.

If we consider orientation-preserving homeomorphisms of compact surfaces only, then we don't have to distinguish between isotopies and homotopies in the definition of mapping class group. This is a consequence of a theorem due to Reinhold Baer (1902-1979) [8, Theorem 1.12]:

Theorem 1.7 (Baer)**.** *Let M denote any compact surface and f be homeomorphism of M. Any homeomorphism g of M that is homotopic to f is also isotopic to f except if f is an orientation-reversing homeomorphism of either the 2-dimensional disk D^2 or the annulus A.*

Since we concentrate on the study of orientation-preserving transformations, Theorem 1.7 gives the following reformulation of the mapping class group of a surface M:

$$\mathrm{MCG}(M) = \mathrm{Homeo}^+(M, \partial M)/\text{homotopy}.$$

Considerations about differentiability have been intentionally left aside for the present paper. However, when it comes to blowing-up homeomorphisms, then differentiability is a handy property. The following theorem provides a useful shortcut in various places. A more elaborated discussion with further references is given in [8, Theorem 1.13].

Theorem 1.8. *Let M denote a compact surface and $X \subset \mathrm{int}\, M$ be a finite collection of points. Let f denote a homeomorphism of the marked surface (M, X). Then for every $x \in X$ there is an open neighbourhood $x \in U_x \subset \mathrm{int}(M)$ such that there is $f' \in [f] \in \mathrm{MCG}(M \text{ rel } X)$ satisfying the two following conditions:*

1. *f' is smooth and non-singular at every x in X,*
2. *f and f' coincide outside of $\bigcup_{x \in X} U_x$.*

As a first example of computation, we prove that the mapping class groups of the disk and the marked disk are trivial. The proof is constructive and known as the *Alexander trick* after the topologist James W. Alexander.

Lemma 1.9 (Alexander trick)**.** $\mathrm{MCG}(D^2) = \mathrm{MCG}(D^2 \text{ rel } \{\star\}) = 0.$

Proof. Given $f\colon D^2 \to D^2$ a homeomorphism that restricts to the identity on ∂D^2, we construct the following isotopy $h_t\colon D^2 \to D^2$ from f to the identity:

$$h_t(z) := \begin{cases} (1-t) f\left(\frac{z}{1-t}\right), & \text{if } 0 \le |z| < 1-t, \\ z, & \text{if } 1-t \le |z| \le 1. \end{cases}$$

In plain words, at the level t, we apply f to the smaller disk $(1-t)D^2$ and the identity outside. This construction is called the *Alexander trick*. The same argument works for the punctured disk $D^2 \setminus \{\star\} \subset \mathbb{C}$ after identifying $\star$ with the origin $0 \in \mathbb{C}$. $\square$

Given a surface M, the triviality of $\mathrm{MCG}(M)$ is equivalent to the connectedness of $\mathrm{Homeo}^+(M, \partial M)$. The Alexander trick actually provides a stronger conclusion: if we apply the homotopy h_t simultaneously to every element in $\mathrm{Homeo}^+(D^2, \partial D^2)$, we eventually obtain that $\mathrm{Homeo}^+(D^2, \partial D^2)$ is contractible and hence simply connected.

Corollary 1.10. $\mathrm{MCG}(S^2) = 0$.

Proof. Let $f\colon S^2 \to S^2$ be a homeomorphism. Since S^2 is simply-connected, any simple closed curve γ in S^2 is homotopic to its image $f(\gamma)$. Hence, up to an isotopy, assume without loss of generality that $f(\gamma) = \gamma$. Cut S^2 along γ to obtain two copies of the unit disk. Lemma 1.9 gives the desired conclusion. $\square$

The Thurston-Nielsen classification establishes the existence of a certain type of homeomorphism in each class of $\mathrm{MCG}(M)$. In the case of the torus T^2, we can understand the classification from an explicit description of its mapping class group. The following examples of mapping class groups computations are inspired from Farb and Margalit [8, Section 2.2].

To understand the structure of $\mathrm{MCG}(T^2)$, we first need to study the mapping class group of the annulus. It turns out that $\mathrm{MCG}(A)$ can be identified with the group of integers $\mathbb{Z}$.

Lemma 1.11. $\mathrm{MCG}(A) \cong \mathbb{Z}$.

Proof. The universal cover $\widetilde{A}$ of the annulus can be identified with the strip $\mathbb{R} \times [0,1]$. Given a class $[f] \in \mathrm{MCG}(A)$, recall that f has to fix ∂A pointwise. Therefore, any lift of f to the universal cover acts by translation of an integer on both $\mathbb{R} \times \{0\}$ and $\mathbb{R} \times \{1\}$. Let $\widetilde{f}$ denote the lift of f that fixes the origin. Therefore, $\widetilde{f}$ reduces to the identity on $\mathbb{R} \times \{0\}$. Define a map $\psi\colon \mathrm{MCG}(A) \to \mathbb{Z}$ by $\psi([f]) = \widetilde{f}((0,1))$.

First we give an equivalent definition of ψ to see that it is a well-defined map. Let $\widetilde{\gamma}\colon I \to \widetilde{A}$ be the arc defined by $\widetilde{\gamma}(t) := (0,t)$ and γ be its projection down to A. The concatenation $f(\gamma) * \gamma^{-1}$ is a closed loop in A based at $\gamma(0)$. In this setting, we find the quantity $\psi([f])$ again by looking at $[f(\gamma)*\gamma^{-1}] \in \pi_1(A) \cong \mathbb{Z}$. With the point of view presented in this reformulation, it becomes clear that ψ is a well-defined map. Moreover, since compositions of maps of A are lifted to the compositions of the underlying lifts in $\widetilde{A}$ that fix the origin $(0,0) \in \widetilde{A}$, we conclude that ψ is a well-defined homomorphism of groups.

Secondly, we claim that ψ is surjective. Consider the transformation of $\mathbb{R} \times [0,1]$ given by the matrix

$$F_n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

Note that F_n is equivariant with respect to the group of covering translations and therefore descends to a homeomorphism f_n of A. By construction we have $\psi([f_n]) = n$.

Finally, we prove that ψ is injective. Assume $\psi([f]) = 0$. Then f admits a lift $\widetilde{f}$ that fixes both boundary components of $\widetilde{A}$. Since the strip is a convex region, the straight-line homotopy between $\widetilde{f}$ and the identity map $id_{\widetilde{A}}$ on $\widetilde{A}$ is well defined. Furthermore, as $\widetilde{f}$ is the identity on $\partial\widetilde{A}$, it commutes with any covering translation of $\widetilde{A}$. Therefore, the straight-line homotopy between $\widetilde{f}$ and $id_{\widetilde{A}}$ is equivariant and descends to a homotopy between f and the identity in A that fixes ∂A pointwise. $\square$

Remark (the importance of fixing the boundary)**.** When computing the mapping class group of a surface, we identify isotopic homeomorphisms for which the isotopy can be chosen to fix the boundary pointwise. Let f denote an element of $\mathrm{Homeo}^+(A, \partial A)$. The argument presented to prove the injectivity statement in Lemma 1.11 applies to build an isotopy between f and the identity in A. However, in general, this isotopy fixes at most one of the boundary component of A (and can always be chosen to fix one boundary component). The point of this remark is the following: without the requirement of working with isotopies that restrict to the identity on the boundary, there would be only one isotopy class inside $\mathrm{Homeo}^+(A, \partial A)$.

In the case of the torus, we can identify the homotopy classes of oriented simple closed curves in T^2 with the primitive elements of the fundamental group $\pi_1(T^2) \cong \mathbb{Z}^2$, i.e. $(\pm 1, 0), (0, \pm 1)$ and all (p, q) with p, q co-prime. Given a primitive element (p, q), we project the real line $q \cdot y = p \cdot x$ to a simple closed curve in T^2 using the usual covering map $\mathbb{R}^2 \twoheadrightarrow \mathbb{R}^2 / \mathbb{Z}^2 \cong T^2$. Denote this simple closed curve by $\gamma_{(p,q)}$ (see Figure 1.2).

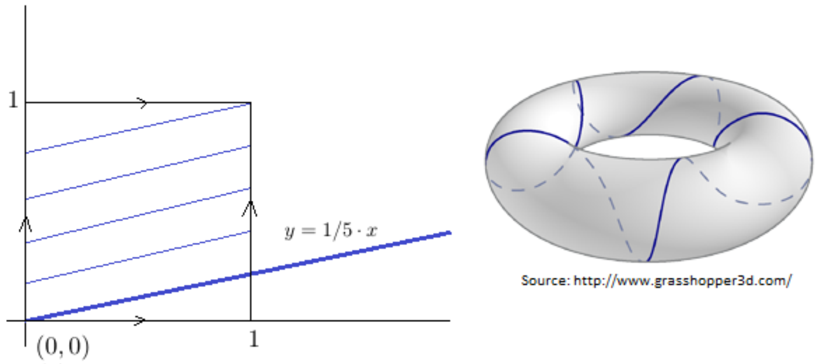

Figure 1.2: Projection of the real line $y = 1/5 \cdot x$ to a closed curve in the torus.

Conversely, given a simple closed curve γ in T^2, we can assume that, up to homotopy, $\gamma(0)$ is the projection of the origin $(0, 0) \in \mathbb{R}^2$. Let $\widetilde{\gamma} \colon I \to \mathbb{R}^2$ be a lift of γ. As $\widetilde{\gamma}(1)$ projects to $\gamma(0) = \gamma(1)$ in T^2, it must be an integer point $(m, n) \in \mathbb{Z}^2 \subset \mathbb{R}^2$. Note that $\widetilde{\gamma}$ is homotopic to the line segment joining the origin and (m, n) in $\mathbb{R}^2$. Since γ is assumed to be simple, (m, n) is thus a primitive element of $\mathbb{Z}^2$. The homotopy between $\widetilde{\gamma}$ and the line segment descends to a homotopy between γ and $\gamma_{(m,n)}$ as required.

The following lemma shows that one can identify the mapping class group of the torus with the group $\mathrm{SL}(2; \mathbb{Z})$ of 2×2 matrices with integer entries and determinant 1.

Lemma 1.12. $\mathrm{MCG}(T^2) \cong \mathrm{SL}(2;\mathbb{Z})$.

Proof. A homeomorphism $f: T^2 \to T^2$ generates an isomorphism $f_\sharp: H_1(T^2) \to H_1(T^2)$ on the first homology group of T^2. Recall that $H_1(T^2) \cong \mathbb{Z}^2$ and thus $f_\sharp$ is an element of $\mathrm{GL}(2;\mathbb{Z})$, a.k.a. the set of invertible 2×2 matrices with integer entries. In particular, $\det f_\sharp = \pm 1$ and since f is orientation-preserving, we must have $\det f_\sharp = 1$, i.e. $f_\sharp \in \mathrm{SL}(2;\mathbb{Z})$. Moreover, homotopic maps generate the same morphism in homology. Hence we have a well-defined morphism

$$\psi\colon\ \mathrm{MCG}(T^2) \to \mathrm{SL}(2;\mathbb{Z}).$$

We first claim that ψ is surjective. Indeed, any $B \in \mathrm{SL}(2;\mathbb{Z})$ descends to a homeomorphism f_B of T^2. Using our identification of closed loops in T^2 with rational lines in $\mathbb{R}^2$, we conclude $\psi([f_B]) = B$.

To see injectivity, assume that $\psi([f])$ is the identity on $\mathbb{Z}^2$. Let α and β be the simple closed curves in T^2 associated to the elements $(1,0)$, respectively $(0,1)$, in $\pi_1(T^2) \cong H_1(T^2)$. By assumption, $f(\alpha)$ and $f(\beta)$ are homotopic to α, respectively β. We can extend the homotopy between α and $f(\alpha)$ to a homotopy of T^2. Without loss of generality, we can assume that this homotopy fixes α pointwise.

The idea is now to cut T^2 along α and restrict our study to the annulus A. The map f induces a homeomorphism $\bar{f}$ of A as f fixes ∂A (identified with two copies of α) pointwise. Since β is homotopic to $f(\beta)$, both are continuous arcs in A (see Figure 1.3). Therefore, $[\bar{f}]$ represents the zero-translation in $\mathrm{MCG}(A) \cong \mathbb{Z}$. We conclude that $\bar{f}$ is homotopic to the identity on A and thus f is homotopic to the identity on T^2.

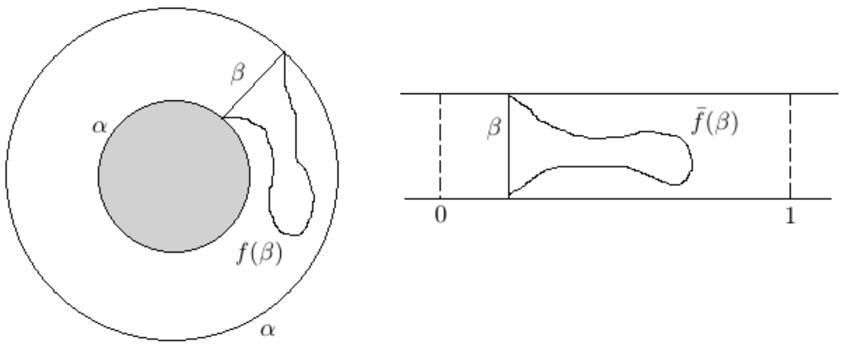

Figure 1.3: The curve β and its image $\bar{f}(\beta)$ in the annulus A and their lift to the universal cover of A identified with the strip $\mathbb{R} \times [0,1]$.

$\square$

Using the identification $\mathrm{MCG}(T^2) \cong \mathrm{SL}(2;\mathbb{Z})$, the elements of $\mathrm{MCG}(T^2)$ can be classified into three categories according to the trace of the corresponding element in $\mathrm{SL}(2;\mathbb{Z})$. Recall that the characteristic polynomial of a matrix $A \in \mathrm{SL}(2;\mathbb{Z})$ is given by

$$c_A(x) := \det(A - xI_2) = x^2 - \mathrm{trace}(A)x + 1.$$

Here, again, I_2 denotes the identity matrix of dimension 2. The Cayley-Hamilton Theorem says that $c_A(A) = 0$ as a linear transformation. Let $[f] \in \mathrm{MCG}(T^2)$. We denote by $A = \psi([f])$ the corresponding element in $\mathrm{SL}(2;\mathbb{Z})$. Let $f_A \in [f]$ be the transformation of the torus $T^2 \cong \mathbb{R}^2 / \mathbb{Z}^2$ generated by A. We distinguish three cases:

1. if $|\mathrm{trace}(A)| \in \{0, 1\}$, then A has two complex eigenvalues. If $\mathrm{trace}(A) = 0$, then Cayley-Hamilton implies $A^2 = -I_2$ and thus $A^4 = I_2$. If $\mathrm{trace}(A) = \pm 1$, then similarly we get $A^2 = \pm A - I_2$ and hence $A^6 = I_2$. In both cases, there exists some integer $n \geq 1$ such that f_A^n is the identity map. In this case we say that $[f]$ has *finite-order*.
2. if $|\mathrm{trace}(A)| = 2$, then the eigenvalues of A are ± 1. Thus A has a rational eigenvector. It follows that f_A fixes the corresponding isotopy class of closed curves in T^2. In this case we say that $[f]$ is *reducible*.
3. if $|\mathrm{trace}(A)| > 2$, then A has two irrational eigenvalues $0 < |\lambda|^{-1} < 1 < |\lambda|$. The two families of parallel lines in $\mathbb{R}^2$ with slopes $|\lambda|$ and $|\lambda|^{-1}$ project to two dense transverse line fields in T^2 along which f_A is expanding, respectively contracting. In this case we say that $[f]$ is *Anosov*.

The dynamics of the finite-order case are very simple. A common family of examples of reducible torus homeomorphisms are the projections of the linear transformations of $\mathbb{R}^2$ given by the matrices

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}.$$

These transformations fix any horizontal real line and moreover fix the real line $y = 0$ pointwise. As in the proof of Lemma 1.12, we can cut along the corresponding pointwise fixed closed curve on T^2 and restrict our study to the annulus A. If we identify A with $S^1 \times I$, then every circle $S^1 \times \{t\}$ is fixed and is pointwise fixed for $t = 0$ and $t = 1$.

Example 1.1. In terms of dynamics, the richest case is the Anosov case. For instance, consider the homeomorphism $f\colon T^2 \to T^2$ defined as the projection of the linear transformation of $\mathbb{R}^2$ given by the matrix

$$H_A := \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

The matrix H_A has eigenvalues $\lambda_{1,2} = (3 \pm \sqrt{5})/2$ with corresponding eigenvectors

$$v_1 = \begin{pmatrix} 2 \\ 1 + \sqrt{5} \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ 1 - \sqrt{5} \end{pmatrix}.$$

Note that $|\lambda_2| < 1 < |\lambda_1|$ and therefore, v_1 gives the direction of expansion and v_2 gives the direction of contraction:

$$\begin{cases} \left\| H_A^k v_1 \right\| = |\lambda_1|^k \left\| v_1 \right\| \xrightarrow{k \to +\infty} +\infty, \\ \left\| H_A^k v_2 \right\| = |\lambda_2|^k \left\| v_2 \right\| \xrightarrow{k \to +\infty} 0. \end{cases}$$

Locally around a fixed point $x \in T^2$, the dynamics of f look like Figure 1.4. In the literature, the point x is called a *hyperbolic fixed point* or sometimes a *saddle point*.

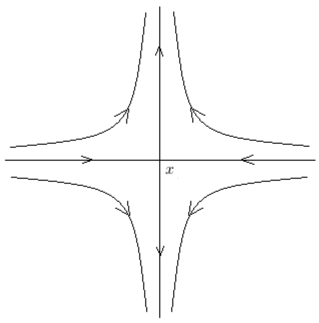

Figure 1.4: The transformation f contracts and expands respectively along two transverse directions around x.

1.3 The Thurston-Nielsen classification Theorem

The classification for the genus-1 torus contains the main features of the classification for higher genus surfaces. However, the notion of Anosov homeomorphisms is too restrictive when working with higher genus surfaces. The correct notion is called *pseudo-Anosov* and allows *inter alia* to have a finite set of singularities where more than two directions of contraction and expansion are permitted. The classification theorem establishes the existence of a representative in each isotopy class of homeomorphisms which is constructed from pseudo-Anosov and finite-order components glued together. We start by giving a precise definition of each type of components described above.

Definition 1.3 (finite-order)**.** A homeomorphism f of a surface M has *finite-order* if it has finite-order as an element of the group $\mathrm{Homeo}^+(M)$, i.e. if there exists some integer $n \geq 1$ such that f^n is the identity map. A class in $\mathrm{MCG}(M)$ has *finite-order* if it has finite-order as an element of the group $\mathrm{MCG}(M)$.

If $[f] \in \mathrm{MCG}(M)$ has finite-order, then there is an integer $n \geq 1$ such that f^n is isotopic to the identity. If M admits an hyperbolic metric, then we can furthermore isotope f to a homeomorphism φ of M of finite-order. This result is proved in [8, Theorem 7.1] with the help of Teichmüller theory. It should be regarded as a converse of Proposition 1.5.

Theorem 1.13. *Let M be a closed surface up to a finite number of punctures. Assume that M has negative Euler characteristic. If $[f] \in \mathrm{MCG}(M)$ has finite-order, then f is isotopic to a finite-order isometry $\varphi \in \mathrm{Homeo}^+(M)$ of M with respect to some hyperbolic metric on M.*

1.3.1 Pseudo-Anosov classes

In order to define pseudo-Anosov homeomorphisms, we need the notion of measured foliation of a surface. We first define singular foliations for closed surfaces.

Definition 1.4 (foliation)**.** Let M denote a closed surface and $X \subset M$ be a finite set of points (possibly empty). A *foliation $\mathcal{F}$ of M with singularities X* is a partition $\{F_a\}$ of $M \setminus X$ in disjoint path-connected subsets together with an atlas (ϕ_i, U_i) of M such that the following conditions hold.

1. For any point $p \in U_i \setminus X$, the chart ϕ_i takes the subsets F_a that intersect U_i to horizontal lines in $\mathbb{R}^2$. Moreover, for overlapping U_i and U_j the transition map $\psi_{ij} = \phi_j \circ \phi_i^{-1}$ takes horizontal lines to horizontal lines (see Figure 1.5). Formally, there exist maps ψ_{ij}^1, ψ_{ij}^2 such that

$$\psi_{ij}(x, y) = (\psi_{ij}^1(x, y), \psi_{ij}^2(y))$$

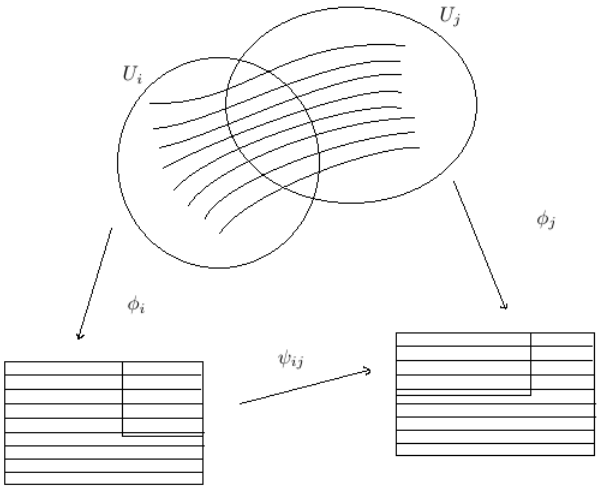

Figure 1.5: Local representation of a foliation away from singularities.

2. For any point $p \in U_i \cap X$ the chart ϕ_i takes the subsets F_a that intersect U_i to a k-pronged saddle with $k \geq 3$ (see for instance Figure 1.6).

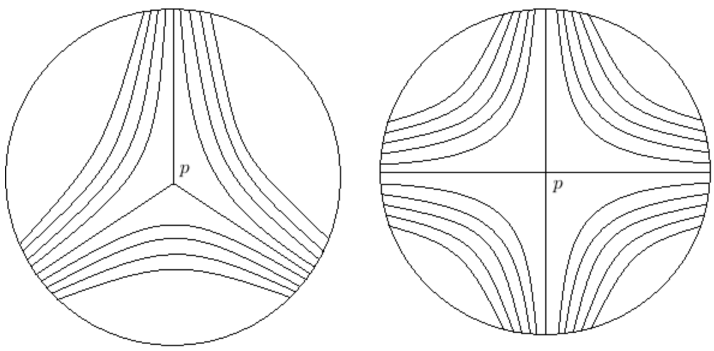

Figure 1.6: Local representation of a foliation at a 3-pronged saddle on the left-hand side and at a 4-pronged saddle on the right-hand side.

The subsets F_a are called the *leaves* of the foliation and the points of X are called the *singular points* or *singularities* of the foliation. A point in $M \setminus X$ is called a *regular point*. If $X = \emptyset$, then $\mathcal{F}$ is called a *regular foliation*.

The topology of the surface constrains the shape and the number of singularities in a foliation of a closed surface. This is an arithmetic condition that relates the Euler characteristic to the number of prongs.

Theorem 1.14 (Euler-Poincaré formula)**.** *Let M be a closed surface. If $p(x)$ denotes the number of unstable (or stable) half-leaves emanating from $x \in M$, then*

$$2\chi(M) = \sum_{x \in M} 2 - p(x).$$

A similar statement holds for compact surfaces. We refer the curious reader to [10, Proposition 5.1]. Observe moreover that for all but finitely many points $p(x) = 2$. Hence the sum on the right-hand side in the conclusion of Theorem 1.14 is a finite sum. As long as 1-pronged singularities are not permitted, we have $p(x) \geq 2$ for every point x. In particular, the Euler-Poincaré formula implies that a foliation of the genus-1 torus is necessarily regular.

For instance, given any line ℓ in $\mathbb{R}^2$, the set of parallel lines to ℓ is a foliation $\widetilde{\mathcal{F}}$ of $\mathbb{R}^2$ that projects to a foliation $\mathcal{F}$ of T^2 via the standard covering $\mathbb{R}^2 \twoheadrightarrow T^2$. If the slope of ℓ is rational, then every leaf is a simple closed curve in T^2. If the slope of ℓ is irrational, then every leaf is dense in T^2.

We can assign a positive real number to each smooth arc $\gamma \colon I \to \mathbb{R}^2$ transverse to $\widetilde{\mathcal{F}}$ by considering the total variation of γ in the direction perpendicular to ℓ. Let this quantity be denoted by $\mu(\gamma)$. If $\nu \colon \mathbb{R}^2 \to \mathbb{R}$ measures the signed distance to the line ℓ, then we can write

$$\mu(\gamma) = \int_\gamma |d\nu| = |\nu(\gamma(1)) - \nu(\gamma(0))|.$$

In particular, μ is invariant under isotopies that preserve the two endpoints of a transverse arc in the same leaves. Note furthermore that the 1-form $d\nu$ is preserved by translations and thus projects to a 1-form ω on T^2. Integration against ω gives a transverse measure on T^2. The notion of transverse measure can be formally defined for foliations of higher genus surfaces.

Definition 1.5 (transverseness)**.** Let $\mathcal{F}$ be a (singular) foliation of a surface M. An arc γ is *transverse* to $\mathcal{F}$ if it is transverse to $\mathcal{F}$ at every point in its interior. If the image of γ contains singular points, then γ is considered transverse if it passes through different sectors every time it meets a singularity. An isotopy $\gamma_t \colon I \to M$ between two transverse arcs is *leaf-preserving* if γ_t is transverse to $\mathcal{F}$ for each t and the two endpoints of all γ_t remain in the same two leaves at all time.

A *transverse measure* μ on a foliation $\mathcal{F}$ assigns a positive real number to each arc transverse to $\mathcal{F}$ so that μ is invariant under leaf-preserving isotopies and is locally induced by $|dy|$ on $\mathbb{R}^2$. A *measured foliation* of M is a foliation together with a transverse measure. Two measured foliations are *transverse* if their leaves are transverse away from singularities (see Figure 1.7). In particular, they have the same set of singularities.

For instance, two non-parallel lines ℓ_1 and ℓ_2 in $\mathbb{R}^2$ generate two regular transverse measured foliations of $\mathbb{R}^2$ that descend to regular transverse measured foliations of T^2.

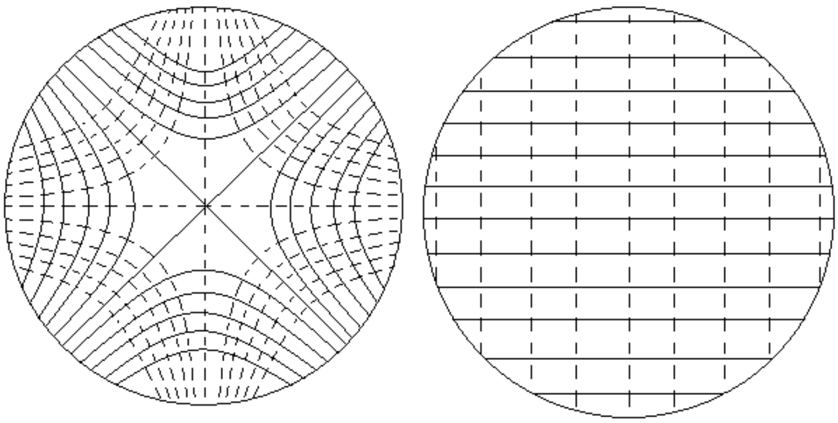

Figure 1.7: Local representation of two transverse regular foliations away from singularities on the right-hand side and two transverse singular foliations around a 4-pronged singularity on the left-hand side.

By definition, a transverse measure μ on a foliation $\mathcal{F}$ assigns a length to transverse arcs only. If γ is a closed curve in M, then we set

$$\mu(\gamma) := \sup \sum_i \mu(\alpha_i)$$

where $\alpha_1, \ldots, \alpha_k$ are disjoint arcs of γ transverse to $\mathcal{F}$ and where the supremum is taken over all such finite collections of arcs.

Note that there is a natural action of $\mathrm{Homeo}(M)$ on the set of measured foliations of M: let $f \in \mathrm{Homeo}(M)$ and $(\mathcal{F}, \mu)$ be a measured foliation of M, then

$$f \cdot (\mathcal{F}, \mu) := (f(\mathcal{F}), f_\star(\mu)),$$

where $f_\star(\mu)(\gamma) := \mu(f^{-1}(\gamma))$ is the pushforward of μ (γ being any arc transverse to $\mathcal{F}$), defines a measured foliation of M.

The existence of a transverse measure on a foliation, as hackneyed as it may seem, greatly constrains the shape of the leaves. The following theorem is analogous to the well-known Poincaré Recurrence Theorem (see Theorem 1.18) in ergodic theory. Its proof relies on the existence of a *good atlas* (see [8, Theorem 14.14]).

Theorem 1.15 (Poincaré Recurrence Theorem for foliations)**.** *Let $(\mathcal{F}, \mu)$ be a measured foliation of a compact surface M. If ℓ denotes an infinite leaf (or an infinite half-leaf) of $\mathcal{F}$, then any arc γ transverse to $\mathcal{F}$ that intersects ℓ must intersect ℓ infinitely many times.*

In the torus case, let f_A be the homeomorphism of the torus generated by a linear transformation $A \in \mathrm{SL}(2;\mathbb{Z})$ with two distinct real eigenvalues $0 < \lambda_2 < 1 < \lambda_1$ and corresponding eigenvectors v_1 and v_2. Let $(\mathcal{F}_1, \mu_1)$ and $(\mathcal{F}_2, \mu_2)$ be the two transverse measured foliations of T^2 generated by the two lines of direction v_1 and v_2 in $\mathbb{R}^2$. The homeomorphism f_A preserves both foliations. The pushforward measures are obtained by multiplying μ_1 and μ_2 by $\lambda_1 = \lambda_2^{-1}$

respectively $\lambda_2 = \lambda_1^{-1}$. In other words, f_A stretches in the direction of v_1 and contracts in the direction of v_2. With $\lambda := \lambda_1 > 1$ we have

$$\begin{cases} f_A \cdot (\mathcal{F}_1, \mu_1) = (\mathcal{F}_1, \lambda^{-1} \cdot \mu_1) \\ f_A \cdot (\mathcal{F}_2, \mu_2) = (\mathcal{F}_2, \lambda \cdot \mu_2). \end{cases}$$

Remark (transverse foliations of non-closed surfaces)**.** We have defined the notion of singular foliation for closed surfaces only. However, this theory can be generalized to surfaces with punctures and boundary.

At a puncture, a foliation is allowed to have the form of a regular point or a k-pronged singularity ($k \geq 3$). The only extra configuration is the allowance for a puncture to have the shape of a 1-pronged singularity. A local description of the leaves near a 1-pronged singularity is outlined in Figure 1.8.

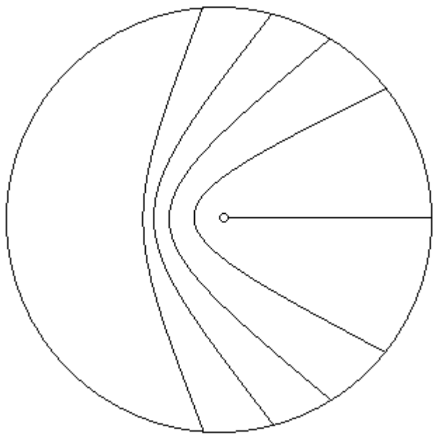

Figure 1.8: Local representation of a 1-pronged singularity at a puncture.

Concerning surfaces with boundary we allow the following new arrangement at boundaries. Leaves are allowed to be either parallel or transverse to any boundary component. Singular points may lie on the boundary and 1-pronged singularities are also permitted on the boundary.

Two transverse foliations are permitted to share common arcs on the boundary. A leaf that coincides with a boundary component is called *peripheral.* If two singularities of two transverse foliations coincide in the interior of the surface, they alternate on the boundary.

In analogy to Anosov transformations of the torus, we define the concept of pseudo-Anosov homeomorphisms for a general surface.

Definition 1.6 (pseudo-Anosov)**.** A homeomorphism ϑ of a surface M is *pseudo-Anosov (pA)* if there exists a pair of transverse measured foliations $(\mathcal{F}^u, \mu_u)$ and $(\mathcal{F}^s, \mu_s)$ of M and a number $\lambda > 1$ such that

$$\begin{cases} \vartheta \cdot (\mathcal{F}^u, \mu_u) = (\mathcal{F}^u, \lambda \cdot \mu_u), \\ \vartheta \cdot (\mathcal{F}^s, \mu_s) = (\mathcal{F}^s, \lambda^{-1} \cdot \mu_s). \end{cases}$$

The measured foliations $(\mathcal{F}^u, \mu_u)$ and $(\mathcal{F}^s, \mu_s)$ are called respectively the *unstable foliation* and the *stable foliation.* The number $\lambda > 1$ is the *stretch factor* of

ϑ. A class in $\mathrm{MCG}(M)$ is called *pseudo-Anosov* if its elements are isotopic to a pseudo-Anosov homeomorphism of M.

If M has nonempty boundary, then we require furthermore that each component of ∂M is a peripheral leaf of both $\mathcal{F}^u$ and $\mathcal{F}^s$ and contains singularities of the two foliations.

The notion of singular foliation for closed surfaces carries naturally to the notion of singular foliation for punctured surfaces. And so does the notion of a pA homeomorphism. For instance, we can remove a finite orbit of a pA homeomorphism from a closed surface to obtain a pA homeomorphism of a punctured surface. In the literature, authors sometimes define the notion of a pseudo-Anosov homeomorphism of a compact surface M *relative to a finite set* $X \subset M$ (see for instance [4, Chapter 7]). It coincides with our definition of pseudo-Anosov homeomorphism for the punctured surface $M \setminus X$.

The notion of pA homeomorphism for surfaces with boundary is not as well established in the literature as it is for closed surfaces. Part of the reason is that some results that hold for pA maps of closed surfaces do not carry to surfaces with boundary (see for instance Theorem 2.20). The definition presented here corresponds to [10, Exposé 11]. Whatever convention we follow, we naturally expect our definition of pA maps for surfaces with boundary to be coherent with the definition relative to a finite set after blowing-down to a point every boundary component.

1.3.2 Reducible classes

Before proceeding to the statement of the Thurston-Nielsen Classification Theorem, we need to formalize the notion of "gluing together" finite-order and pA components. This is the family of reducible homeomorphisms.

Definition 1.7 (reducibility)**.** Let M denote a surface. A homeomorphism f of M is called *reducible* if there exists an integer $n \geq 1$ and a nonempty collection $\{c_1, \ldots, c_n\}$ of isotopy classes of simple closed curves in $\mathrm{int}(M)$ with pairwise disjoint representatives such that

1. $\{f(c_1), \ldots, f(c_n)\} = \{c_1, \ldots, c_n\}$,
2. every component of M cut along pairwise disjoint representatives has negative Euler characteristic.

The collection $\{c_1, \ldots, c_n\}$ is called a *reduction system* for $[f] \in \mathrm{MCG}(M)$. A class of maps in $\mathrm{MCG}(M)$ is *reducible* if its elements are isotopic to a reducible homeomorphism of M (analogously to pA classes).

In the reducible case, we understand the dynamics of a class $[f] \in \mathrm{MCG}(M)$ via the following procedure. We choose pairwise disjoint representatives $\gamma_i \in c_i$ and a homeomorphism φ isotopic to f such that

$$\{\varphi(\gamma_1), \ldots, \varphi(\gamma_n)\} = \{\gamma_1, \ldots, \gamma_n\}.$$

There is an integer $k \geq 1$ such that φ^k fixes every connected component of $M \setminus \cup\gamma_i$. Studying the dynamics of φ then comes down to the independent examination of the restrictions of φ^k to the connected components of $M \setminus \cup\gamma_i$. In the torus case for instance, reducibility implies the existence of an invariant simple closed curve. It allows us to restrict our attention to a homeomorphism of the annulus instead (see the proof of Lemma 1.12).

Boyland defines the family of reducible classes in the mapping class group in terms of simple closed curves instead of isotopy classes of simple closed curves [4, Chapter 7]. To get an equivalent definition one must then further require that the reduction system (constituted of simple closed curves instead of classes) comes equipped with an invariant open tubular neighbourhood. This tubular neighbourhood corresponds to an accumulation of small isotopic deformations of each simple closed curve.

Similarly as for the pA case, the literature sometimes defines reducibility for a compact surface M *relative to a finite set* $X \subset M$ (see for instance [24, Definition 5.1]). It has the extra requirement that the reduction system must lie in $\mathrm{int}(M) \setminus X$. Therefore it corresponds to our notion of reducibility for the punctured surface $M \setminus X$.

Note that reducible and finite-order classes overlap. For instance, consider a rotation of a six-times-punctured annulus as illustrated by Figure 1.9. The map f has finite-order 6 and the isotopy class of the simple closed curve γ is in itself a reduction system for $[f]$. Hence $[f]$ has finite-order and is reducible. We can build a similar example for a closed surface by transforming the six punctures and the boundary components in order to get a genus-7 closed surface similarly as [8, Figure 13.1]. However, the pA classes are never reducible nor finite-order.

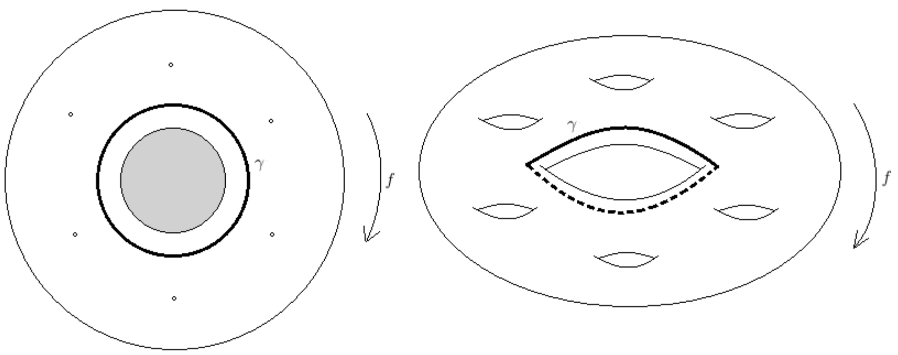

Figure 1.9: A six times punctured annulus on the left-hand side under the action of a rotation of $-\pi/3$. The curve γ is invariant under the rotation. The same configuration also applies for a genus-7 torus as illustrated on the right-hand side.

1.3.3 The classification

We now have defined all the different types exhibited in the Thurston-Nielsen classification Theorem and we are ready to state it formally.

Theorem 1.16 (Thurston-Nielsen classification Theorem)**.** *Let M denote a surface. Every class in* $\mathrm{MCG}(M)$ *is either reducible, pseudo-Anosov or has finite-order. Moreover, if a class is pseudo-Anosov, then it is neither periodic nor reducible. Pseudo-Anosov representatives are unique up to topological conjugacy in the eponymous classes.*

Farb and Margalit give a modern proof of Theorem 1.16 in [8, Theorem 13.2] with the help of Teichmüller theory. They also recall the ideas of Thurston's original proof. Historically, it is interesting to note that Thurston did not publish his proof. The first complete published proof was the work of Bers in 1978 who used Teichmüller theory rather than foliations as Thurston did. Farb and Margalit trace the history of Theorem 1.16 in more details and give further references in [8].

If a class $[f] \in \mathrm{MCG}(M)$ is reducible and φ denotes a reducible representative, then there is an integer $k \geq 1$ such that φ^k fixes every connected component of M cut along a reduction system. We can apply Theorem 1.16 again to each of these components. Iterating this methodology sufficiently many times and on every component, we eventually obtain that $[f]$ has a representative that breaks into finitely many components of type finite-order or pA only.

Remark (classification for punctured surfaces)**.** If we intend to adapt the classification for punctured surfaces, then it is most convenient to proceed as follows. Let $f\colon M \to M$ be a homeomorphism of a compact surface M and $X \subset \mathrm{int}(M)$ be a finite invariant set for f. The classification is usually stated by saying that the isotopy class of f in $\mathrm{MCG}(M \text{ rel } X)$ is either reducible relative to X, pA relative to X or has finite-order. In other words, f is isotopic relative to X to some homeomorphism φ of M that is reducible relative to X, pA relative to X or has finite-order.

The representative φ provided by Theorem 1.16 (or by the previous Remark) is traditionally called a *canonical representative* of $[f] \in \mathrm{MCG}(M \text{ rel } X)$. We adhere to this terminology thereafter. With a slight abuse of notation, we allow ourself to write $\varphi \in [f]$ even if φ does not necessarily fixes the boundary components of M.

1.4 Dynamics of pseudo-Anosov homeomorphisms

Theorem 1.16 essentially tells us that any infinite-order and irreducible class must be pA. We shall see that similarly to the case of the torus, pA classes have the richest dynamics. For the sake of brevity, we only state the properties of such maps that will come in handy in our study of forcing relations later. A deeper study of pA homeomorphisms from which ours is inspired can be found in either [8, Chapter 14] or [10].

Standing assumptions. The following assumptions will prevail throughout Section 1.4. Unless otherwise noted, $\vartheta\colon M \to M$ denotes a pA homeomorphism of a surface M. The underlying transverse foliations of M are denoted by $(\mathcal{F}^s, \mu_s)$ for the stable foliation and $(\mathcal{F}^u, \mu_u)$ for the unstable foliation. The corresponding stretch factor of ϑ is expressed by the number $\lambda > 1$.

A primary observation shows that ϑ^n is also pA for any $n \geq 1$ with respect to the same foliations and with stretch factor λ^n.

Recall that ϑ contracts the leaves of the stable foliation by a factor λ with respect to μ_u and *vice versa* expands the leaves of the unstable foliation. In particular, the leaves of $\mathcal{F}^s$ or $\mathcal{F}^u$ are never closed except if the leaf coincides with a boundary component. Specifically, any non-peripheral leaf is at least half-infinite. Moreover, there is always some iteration ϑ^n of ϑ that fixes all the singularities. Therefore a single leaf of either $\mathcal{F}^s$ or $\mathcal{F}^u$ can never connect two singularities except if they lie on the same boundary component.

An application of the Poincaré Recurrence Theorem for foliations (Theorem 1.15) in the context of pA homeomorphisms immediately gives:

Lemma 1.17. *Any non-peripheral leaf of either $\mathcal{F}^s$ or $\mathcal{F}^u$ is dense in M.*

Lemma 1.17 together with the Baire Category Theorem imply the existence of a dense orbit for ϑ. A detailed proof is given in [8, Theorem 14.17] or [10, Corollaire 9.19]. In particular, a pA homeomorphism is topologically transitive. Recall that a dynamical system $T\colon X \to X$ is *topologically transitive* if for every pair of nonempty open subsets $U, V \subset X$, there is an integer $k \geq 0$ such that $T^k(U) \cap V \neq \emptyset$. In a stronger manner, it is true that the set of periodic points of ϑ is dense in M. This is mainly a consequence of the Poincaré Recurrence Theorem in its general setting.

Theorem 1.18 (Poincaré Recurrence Theorem)**.** *Let $\mathcal{M}$ be a finite measure space and $T\colon \mathcal{M} \to \mathcal{M}$ be a measure preserving transformation. If $A \subset \mathcal{M}$ has positive measure, then almost every point in A returns to A infinitely many times.*

The proof Theorem 1.18 is rather simple and can be found in any introductory book about ergodic theory (see for instance [19, Theorem 4.1.19]).

Proposition 1.19. *The set of periodic points of ϑ is dense in M.*

Proof. We introduce the notion of standard rectangle. A *standard rectangle* is a simply connected and nonempty region $U \subset M$ such that ∂U is the union of four arcs that alternate between being subarcs of leaves of $\mathcal{F}^s$ and $\mathcal{F}^u$. In local coordinates, U simply looks like Figure 1.10.

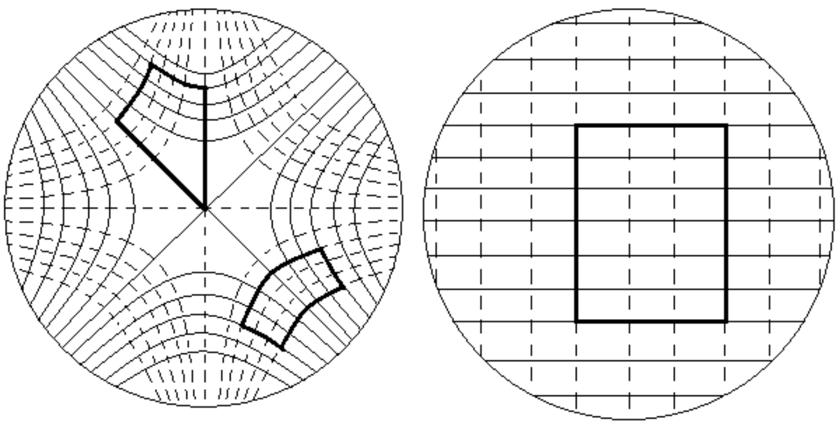

Figure 1.10: Two standard rectangles in local coordinates near a 4-pronged singularity on the left-hand side. Another standard rectangle in local coordinates on the right-hand side where leaves of the stable and unstable foliations are mapped to vertical respectively horizontal lines in $\mathbb{R}^2$ (i.e. away from singularities).

Let $U \subset M$ be a standard rectangle that does not contain any singularity of the foliations. To establish the desired conclusion, it is sufficient to prove that

$\overline{U}$ contains a periodic point. Let $V \subset U$ be a strictly smaller standard rectangle. Consider the area measure $\mu := \mu_s \cdot \mu_u$. This measure is left invariant by ϑ (as $\lambda \cdot \lambda^{-1} = 1$). Therefore, the Poincaré Recurrence Theorem applies and we can find $x_1 \in V$ such that $\vartheta^n(x_1) \in V$ for infinitely many $n \geq 1$. Fix some $k \geq 1$ for which $\vartheta^k(x_1) \in V$.

Let J be the intersection of the stable leaf that goes through x_1 with $\overline{U}$. Up to increasing the value of k, we may assume $\vartheta^k(J) \subset U$. Thus $\vartheta^k(J)$ is a subarc of a stable leaf that crosses U. Using a translation along unstable leaves we can bring back $\vartheta^k(J)$ into J. The restricted map composed with this translation gives a continuous map $J \to J$ which has a fixed point x_0 (it can be seen as a consequence of the Intermediate Value Theorem).

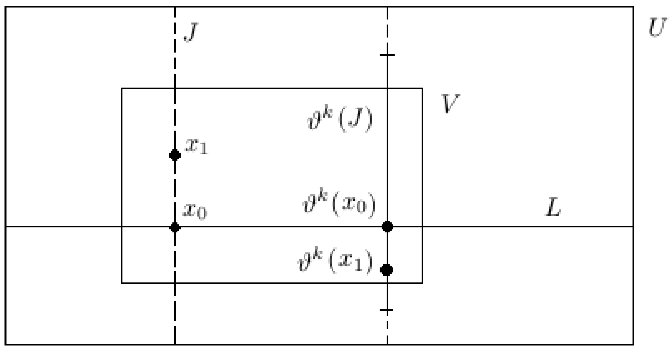

Figure 1.11: Construction described in the proof of Proposition 1.19.

Let L be the intersection of the unstable leaf that goes through x_0 with $\overline{U}$. Again, up to increasing k, we can assume that $L \subset \vartheta^k(L)$. A standard argument in one-dimensional analysis shows that ϑ^k must have fixed point in L. $\square$

It is part of the study of forcing relations on periodic orbits in isotopy classes to realize what kind of fixed points remain through an isotopy. The well-known Lefschetz-Hopf Theorem is a powerful tool to prove existence of fixed points. Its statement requires to define the notion of index of a fixed point.

Definition 1.8 (fixed point index)**.** Let $U \subset \mathbb{R}^2$ be an open subset and $f : U \to \mathbb{R}^2$ a continuous map. Let $x_0 \in U$ be an isolated fixed point of f. Let $x_0 \in V \subset U$ be an open disk such that $f(x) \neq x$ for all $x \in V \setminus \{x_0\}$. Consider the map $g^V : \partial V \to S^1$ defined by

$$x \mapsto \frac{x - f(x)}{\|x - f(x)\|}.$$

The *index of the fixed point* x_0 of f is defined as the integer $\mathrm{ind}(x_0, f) := g^V_\sharp(1) = \mathrm{trace}(g^V_\sharp)$ where $g^V_\sharp$ denotes the induced homomorphism on $H_1(S^1) \cong \mathbb{Z}$.

The index of a fixed point is independent of the choice of the open disk V in the statement of Definition 1.8 (see [19, Proposition 8.4.1]). Geometrically, $\mathrm{ind}(x_0, f)$ corresponds to the number of complete turns of the vector $(x, f(x))$ when x describes a small oriented circle around x_0. For a homeomorphism f of a surface M and an isolated fixed point $x_0 = f(x_0)$, we define $\mathrm{ind}(x_0, f)$ locally in a chart.

Theorem 1.20 (Lefschetz-Hopf Theorem)**.** *Let f denote a homeomorphism of a compact surface M. We define its Lefschetz number as the integer*

$$\Lambda_f := \sum_{k=0}^{+\infty} (-1)^k \mathrm{trace}(f_\sharp | H_k(M; \mathbb{Q}))$$

where $f_\sharp | H_k(M; \mathbb{Q})$ denotes the induced map on the kth homology group of M with rational coefficients. The number Λ_f is invariant under homotopy. If f only has isolated fixed points, then

$$\Lambda_f = \sum_{x \in \mathrm{Fix}(f)} \mathrm{ind}(x, f).$$

In particular, $\Lambda_f \neq 0$ implies the existence of a fixed point.

The following proposition tells us that the index of a periodic point x_0 of the pA homeomorphism ϑ (seen as a fixed point of some iterated map) depends only on the action of ϑ on the half-leaves leaving from x_0. The theorem is stated in [24] and the proof is inspired from [20, Proposition 2.7].

Proposition 1.21. *Let $n \geq 1$ be an integer. Let $x_0 \in \mathrm{int}(M)$ denote an isolated interior fixed point of ϑ^n. If $p(x_0)$ denotes the number of unstable (or stable) half-leaves emanating from x_0, then*

$$\mathrm{ind}(x_0, \vartheta^n) = \begin{cases} 1 - p(x_0), & \textit{if } \vartheta^n \textit{ preserves each half-leaf,} \\ 1, & \textit{if } \vartheta^n \textit{ rotates the half-leaves.} \end{cases}$$

Proof. First we assume that $\theta := \vartheta^n$ fixes every half-leaf. Locally, the unstable and stables half-leaves leaving from x_0 separate a small disk around x_0 into $2p(x_0)$ hyperbolic sectors (see Figure 1.12). Note that if x_0 is a regular point of the foliation, then we assimilate it as a 2-pronged singularity. As x moves along one of the $2p(x_0)$ arcs, the vector $(x, \theta(x))$ describes an angle of

$$-\pi + \frac{\pi}{p(x_0)}.$$

Therefore, when x completes its journey around the disk, the vector $(x, \theta(x))$ has described a total angle of

$$2p(x_0)\left(-\pi + \frac{\pi}{p(x_0)}\right) = 2\pi(1 - p(x_0)).$$

Hence $\mathrm{ind}(x_0, \theta) = 1 - p(x_0)$.

If θ rotates the half-leaves, then the tip of the vector $(x, \theta(x))$ never crosses the hyperbolic sector containing x. Therefore, the vector $(x, \theta(x))$ describes a single turn and $\mathrm{ind}(x_0, \theta) = 1$.

Note that only these two cases may occur. As θ maps half-leaves to half-leaves, it maps hyperbolic sectors to hyperbolic sectors. If θ fixes one half-leaf, it must thus fix both neighbour leaves (because θ is always assumed to be orientation-preserving). An inductive argument shows that θ has to fix all the half-leaves. If θ rotates one of the half-leaf, then a similar reasoning shows that it acts by the same rotation on the other half-leaves. □

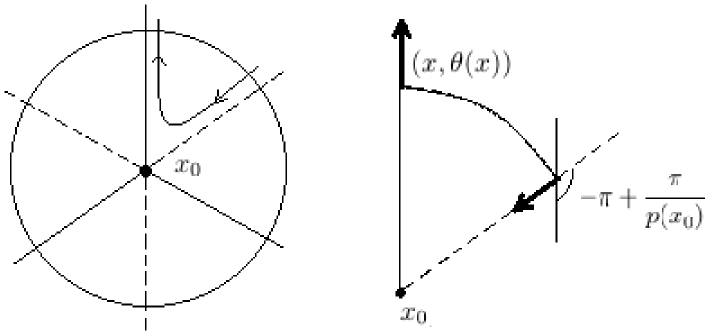

Figure 1.12: Configuration in the proof of Proposition 1.21 in the case of a 3-*pronged* singularity.

Remark. We deduce from Proposition 1.21 that an interior periodic point of ϑ has non-zero index if and only if it is not a one-pronged singularity. Furthermore, if ϑ has a periodic point on the boundary ∂M, then it may have index zero. Note nevertheless that a periodic point on ∂M is necessarily a singularity of either $\mathcal{F}_s$ and $\mathcal{F}_u$.

For convenience we introduce the notion of *fixed point index of a function.* It is reasonable to expect that in the case where the function only has finitely many fixed points, then the fixed point index of the function is the sum of the indices of its fixed points. The following construction and the consecutive definition are inspired from Górniewicz [24, Definition 7.1].

Let $U \subset \mathbb{R}^2$ be an open subset and $f\colon U \to \mathbb{R}^2$ be a continuous function such that the set of fixed points $\mathrm{Fix}(f) \subset U$ is compact. We identify the sphere S^2 with the Alexandroff compactification (one-point compactification) of $\mathbb{R}^2$. Consider the following diagram.

$$\begin{tikzcd} S^2 \arrow[r, "\iota_1"] & (S^2, S^2 \setminus \mathrm{Fix}(f)) & (U, U \setminus \mathrm{Fix}(f)) \arrow[l, "\iota_2"'] \arrow[d, "id - f"] \\ & & (\mathbb{R}^2, \mathbb{R}^2 \setminus \{0\}) \end{tikzcd}$$

Here ι_1 and ι_2 denote the respective inclusions. Apply the functor $H_2(-)$ to the diagram. The excision axiom tells us that ι_2 generates an isomorphism on the homology level.

$$\begin{tikzcd} \mathbb{Z} \cong H_2(S^2) \arrow[r, "(\iota_1)_\sharp"] & H_2(S^2, S^2 \setminus \mathrm{Fix}(f)) \arrow[r, "(\iota_2)_\sharp^{-1}"] & H_2(U, U \setminus \mathrm{Fix}(f)) \arrow[d, "(id - f)_\sharp"] \\ & & H_2(\mathbb{R}^2, \mathbb{R}^2 \setminus \{0\}) \cong \mathbb{Z} \end{tikzcd}$$

Definition 1.9 (fixed point index of a function)**.** In the notation above and in analogy to Definition 1.8, we define the *fixed point index of f* to be the integer

$$\operatorname{ind}(f) := \Big((id - f)_\sharp \circ (\iota_2)_\sharp^{-1} \circ (\iota_1)_\sharp\Big)(1).$$

If $U = U_1 \cup U_2$ splits into two disjoint open subsets, then

$$\operatorname{ind}(f) = \operatorname{ind}(f\restriction_{U_1}) + \operatorname{ind}(f\restriction_{U_2}).$$

In particular, if f only has finitely many fixed points $\{x_1, \ldots, x_n\} \subset U$, then $\operatorname{ind}(f) = \sum_{i=1}^{n} \operatorname{ind}(x_i, f)$. The fixed point index of a function is invariant under homotopy provided that at every level t the set of fixed points $\operatorname{Fix}(h_t)$ of the homotopy $h_t \colon U \to \mathbb{R}^2$ is compact in U. This is a consequence of the homotopy axiom in homology. For more properties of the fixed point index, the reader may consult [24, Chapter 2, Section 7].

Another consequence of the existence of a transverse measure invariant under leaf-preserving isotopies for both the stable and unstable foliations says that the action of ϑ on the free homotopy classes of M has no periodic orbits. In other words:

Proposition 1.22. *Let γ be a simple closed curve in M. If γ is freely homotopic to its image $\vartheta(\gamma)$, then γ is homotopic to a point or a boundary component.*

The proof of Proposition 1.22 relies on the positivity of the following quantity. If c denotes a homotopy class of simple closed curves not homotopic to a point or a boundary component, then

$$I(\mathcal{F}^s, c) := \inf\{\mu_s(\gamma) : \gamma \in c\} > 0.$$

The same result does not hold for the unstable foliation and the associated transverse measure. The argument that shows the positivity of $I(\mathcal{F}^s, c)$ is involving and cases must be distinguished according to how $\gamma \in c$ superposes on the leaves of $\mathcal{F}^s$. The details are provided in [10, Proposition 5.9].

In the context of Proposition 1.22, observe that $I(\mathcal{F}^s, \vartheta(c)) = \lambda \cdot I(\mathcal{F}^s, c)$. Therefore, if, as in Proposition 1.22, c and $\vartheta(c)$ designates the same homotopy class, then $I(\mathcal{F}^s, \vartheta(c)) = I(\mathcal{F}^s, c)$ and hence $I(\mathcal{F}^s, c) = 0$ because $\lambda > 1$. This is a contradiction.

Another handy property of pA mapping classes and simple closed curves is the following proposition. It is a consequence of a more elaborate result (see for instance [8, Theorem 14.24]). If a and b denote two homotopy classes of simple closed curves not homotopic to a point or a boundary component, then we define

$$\iota(a, b) := \min\{\sharp\alpha \cap \beta : \alpha \in a, \beta \in b\}.$$

Proposition 1.23. *Let a and b denote two homotopy classes of simple closed curves not homotopic to a point or a boundary component. Then there is some integer $n \geq 1$ such that $\iota(\vartheta^n(a), b) > 0$.*

To extend the notion of measured foliations to surfaces with boundary, different configurations must be evaluated. However, it is possible to define a pA homeomorphism of a surface M with boundary from the definition of a pA homeomorphism of a punctured surface by blowing up the punctures. The following construction is inspired from Jiang and Guo [18, Section 2.1].

Let M denote a surface. The tangent bundle of M is written $TM := \bigsqcup_{x\in M} T_xM$ where T_xM denotes the tangent space at x. If $v\colon M \to TM$, $x \mapsto v_x \in T_xM$, denotes a smooth vector field on M, then for any $x \in M$ there is a unique curve γ_x in M defined on a maximal open interval such that γ_x is solution of the system

$$\left|\begin{array}{l} \frac{d\gamma}{dt}(t) = v_{\gamma(t)} \\ \gamma(0) = x. \end{array}\right.$$

We say that a vector field v is *complete* if every such curve γ_x is defined for $t \in \mathbb{R}$. We write $\exp(tv)(x) := \gamma_x(t)$. A complete vector field on M generates for every $t \in \mathbb{R}$ a diffeomorphism $\exp(tv)\colon M \to M$. The curves $\{\exp(tv) : t \in \mathbb{R}\}$ are called the *flow curves of* v. Two distinct flow curves never intersect because γ_x is uniquely determined by v.

The first step consists in describing the dynamics of ϑ near an interior point of M as the flow generated by some vector field. Around a regular point, a pA homeomorphism has the behaviour of a hyperbolic (fixed) point. If $\lambda > 1$ denotes the stretch factor of ϑ, then we introduce the vector field $V\colon \mathbb{C} \to T\,\mathbb{C}$ defined by

$$V_z := \log\lambda \cdot \bar{z}.$$

The unique solution of

$$\left|\begin{array}{l} \frac{d\gamma}{dt}(t) = V_{\gamma(t)} \\ \gamma(0) = z = x + iy \end{array}\right.$$

is given by $\gamma_z(t) = \lambda^t x + i\lambda^{-t} y$. The time-1 map $exp(V)\colon M \to M$ of V should be regarded as an Anosov transformation of the genus-1 torus.

Let $p \geq 2$ be an integer. We are about to describe the dynamics of ϑ at a p-pronged singularity. Consider the diffeomorphism $\Phi_p\colon \mathbb{C} \to \mathbb{C}$ defined by

$$\Phi_p(z) := z^{p/2}$$

and its inverse $\Phi_p^{-1}(z) := z^{2/p}$. Let $v := (\Phi_p^{-1})_\star V$ be a vector field on $\mathbb{C}$. Explicitly, v has the following form:

$$v_{\Phi_p^{-1}(z)} = \frac{d}{dt}(\Phi_p^{-1} \circ \gamma_z(t))\bigg|_{t=0} = \frac{2\log\lambda}{p} \cdot \bar{z} \cdot z^{2/p-1}$$

and therefore

$$v_z = \frac{2\log\lambda}{p} \cdot \bar{z}^{p/2} \cdot z^{1-p/2} = \frac{2\log\lambda}{p} \cdot z \cdot \left(\frac{z}{|z|}\right)^{-p}.$$

In polar coordinates, v has thus the following form:

$$v_{re^{i\theta}} = \frac{2\log\lambda}{p} \cdot re^{i(1-p)\theta}.$$

By varying the values of the modulus r and the argument θ we obtain the direction of the flow curves associated to v at different points. We observe that the rays $re^{ik\pi/p}$ for $k = 0, \ldots, 2p-1$ are fixed by v seen as a transformation from $\mathbb{C}$ to $\mathbb{C}$. They divide the complex plane into $2p$ hyperbolic sectors as expected. See Figure 1.13. This description of the dynamics of ϑ as the flow curves of v can be used to give an explicit description of the blow-up of ϑ at an interior point.

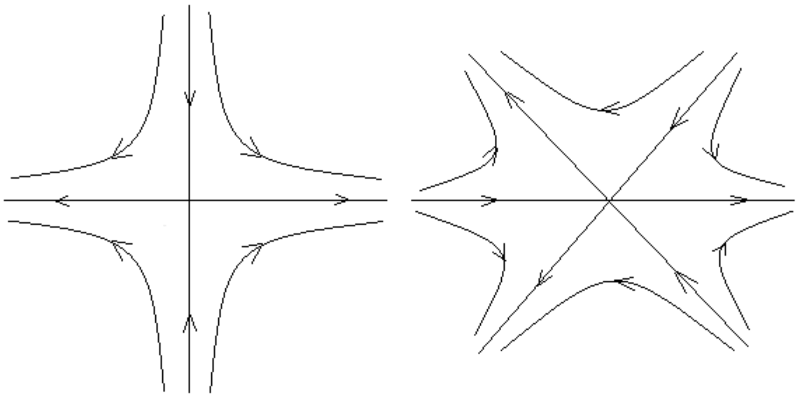

Figure 1.13: An illustration of the flow curves of v for $p = 2$ on the left-hand side and for $p = 3$ on the right-hand side.

Proposition 1.24. *A pA homeomorphism ϑ of a surface M has the blow-up $\widehat{\vartheta}\colon \widehat{M} \to \widehat{M}$ at any point in the interior of M. Moreover, $\widehat{\vartheta}$ is a pA homeomorphism of $\widehat{M}$.*

Proof. Assume we want to blow-up ϑ at a p-pronged point x in $\mathrm{int}(M)$ with $p \geq 2$. It is a local process. Therefore we identify x with the origin in the complex plane and describe ϑ with the vector field v constructed previously. We start by specifying a transformation of the complex plane that maps $\mathbb{C} \setminus D^2$ homeomorphically to $\mathbb{C} \setminus \{0\}$. Consider

$$\begin{aligned} \Psi\colon\ \mathbb{C} \setminus D^2 &\longrightarrow \mathbb{C} \setminus \{0\} \\ z &\longmapsto z - \frac{z}{|z|} \\ re^{i\theta} &\longmapsto (r-1)e^{i\theta}. \end{aligned}$$

and its inverse $\Psi^{-1}(z) := z + \frac{z}{|z|}$. To obtain a vector field on $\mathbb{C} \setminus D^2$, we take $v' := (\Psi^{-1})_\star v$. Explicitly,

$$v'_{re^{i\theta}} = \frac{2\log\lambda}{p}\Big((r-1)e^{i(1-p)\theta} + e^{i(\theta-\pi/2)}\sin(p\theta)\Big).$$

In particular, v' can be smoothly extended to $\mathbb{C} \setminus \mathrm{int}(D^2)$ with $v' \restriction_{\partial D^2}$ tangent to D^2. The flow curves of v' describe the local dynamics of the blow-up $\widehat{\vartheta}$ around the boundary circle corresponding to x. See Figure 1.14.

The leaves of $\mathcal{F}^s$ and $\mathcal{F}^u$ around x in M are precisely the level sets of $\mathrm{Re}\,\Phi_p$ and $\mathrm{Im}\,\Phi_p$ respectively. The foliations around the boundary circle corresponding to x in $\widehat{M}$ are defined to be the level sets of $\mathrm{Re}\,(\Phi_p \circ \Psi)$ and $\mathrm{Im}\,(\Phi_p \circ \Psi)$ and correspond to the lifts of $\mathcal{F}^s$ and $\mathcal{F}^u$ in $\widehat{M}$ up to an extra leaf around the boundary circle. We equip the lifted foliations with transverse measures by precomposing μ_s and μ_u with Ψ. It follows from the construction that $\widehat{\vartheta}$ is a pA homeomorphism of $\widehat{M}$ with respect to these transverse measured foliations of $\widehat{M}$. $\square$

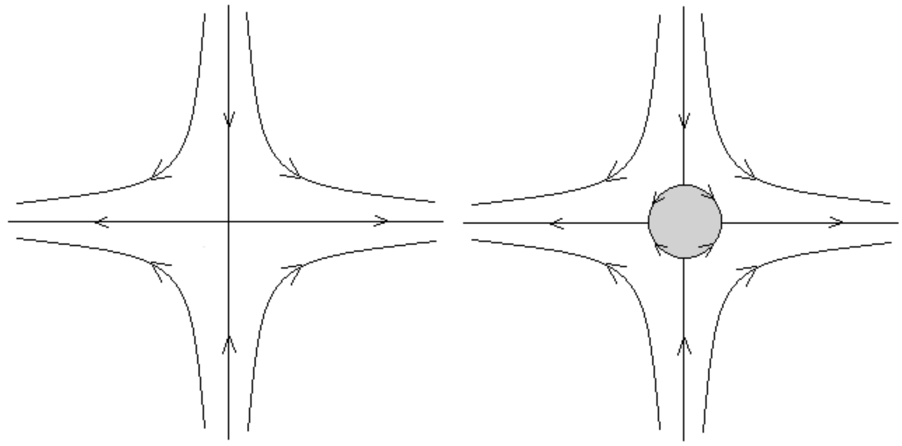

Figure 1.14: The dynamics around a regular point of ϑ on the left-hand side and the dynamics after blowing-up the point on the right-hand side.

1.5 The geometry of measured foliations

1.5.1 Closed 1-form

In Definition 1.4, a foliation is presented in rather descriptive words. One can also approach foliations with closed 1-forms. The construction results in a more general notion of foliation where singularities can have a shape different than a $(k \geq 3)-$saddle point. For a deeper study of foliations on surfaces see Nikolaev [25] or Farb and Margalit [8, Section 11.2]. Our present digression is inspired by these authors' work.

Let $v\colon M \to TM$ denote a smooth vector field on M. Given a smooth function $f\colon M \to \mathbb{R}$, we define its *Lie derivative in the direction of the vector field* v as the map

$$L_v f := \frac{d}{dt}\left(f \circ \exp(tv)\right)\Big|_{t=0}.$$

Note that we don't need completeness of v to define the Lie derivative; we only need each flow curve to be defined on some non-degenerate interval $(-\varepsilon, \varepsilon)$ around 0 (which is always the case). The Lie derivative has the following property:

$$\frac{d}{dt}\left(f \circ \exp(tv)\right) = L_v f \circ \exp(tv).$$

It follows immediately from a change of variable in the derivative on the left-hand side in the definition of Lie derivative. In particular, if $L_v f = 0$, then f is constant along each flow curve $\exp(tv)$.

Let $\phi\colon U \to V \subset \mathbb{R}^2$ be a local chart of M where U is assumed to be contractible. We first explain how to build a foliation $\mathcal{F}_U$ of U from a smooth differential closed 1-form $\omega\colon TM \to \mathbb{R}$ that vanishes at at most finitely many points $p \in M$. Recall that closeness of ω means $d\omega = 0$ where d denotes the differential operator. The Poincaré Lemma tells us that locally, or more precisely in any contractible open subspace of M, every closed 1-form is exact. As a matter of fact, there is some smooth function $f\colon U \to \mathbb{R}$ such that $df = \omega\restriction_U$.

A *critical point* of f is a point $p \in U$ for which $(df)_p$ does not have maximal rank. Since f takes real values, the critical points of f are exactly the points $p \in U$ for which $(df)_p = 0$. In other words, the critical points of f coincide with a subset of the zeros of ω of which there are at most finitely many by assumption.

We would like to build a local non-trivial vector field $v \colon U \to TU$ such that

$$L_v f = 0$$

and such that v vanishes only at critical points of f. Having restricted our attention to a contractible chart U, we can give a rather intuitive construction for such a vector field. The map $f \circ \phi^{-1} \colon V \to \mathbb{R}$ is differentiable at every point of V with differential $D(f \circ \phi^{-1}) \colon \mathbb{R}^2 \to \mathbb{R}$. For $(x, y) \in V$, we write

$$D(f \circ \phi^{-1})_{(x,y)} =: \begin{pmatrix} \alpha_1(x,y) & \alpha_2(x,y) \end{pmatrix}.$$

Define $X \colon V \to \mathbb{R}^2$ by

$$X(x,y) := \begin{pmatrix} -\alpha_2(x,y) \\ \alpha_1(x,y) \end{pmatrix}$$

so that $D(f \circ \phi^{-1})_{(x,y)} X(x,y) = 0$ for all $(x,y) \in V$. To define a vector field on U from the vector field X on V we fix an identification $\psi_p \colon \mathbb{R}^2 \xrightarrow{\cong} T_pU$ for every $p \in U$. For $p \in U$, let $v_p := (\psi_p \circ X)(\phi(p)) \in T_pU$.

$$\begin{tikzcd}
V \arrow[r, "X"] & \mathbb{R}^2 \arrow[dr, "\psi_p"] & \\
U \arrow[u, "\phi"] \arrow[r, dashed, "v"] & TU & T_pU \arrow[l, "\supset"']
\end{tikzcd}$$

It defines a vector field $v \colon U \to TU$. For all $p \in U$, we have

$$D(f \circ \phi^{-1})_{\phi(p)} \psi_p^{-1}(v_p) = D(f \circ \phi^{-1})_{\phi(p)} X(\phi(p)) = 0.$$

and hence $(df)_p(v_p) = 0$. In other words $L_v f = 0$ as required. Note finally that $v_p = 0$ if and only if $X(\phi(p)) = 0$ or equivalently $(df)_p = 0$. Such a vector field v has the property that f is constant along each flow curve $\exp(tv)$.

Define the singularities of $\mathcal{F}_U$ to be the zeros of $\omega\restriction_U = df$ and the leaves of $\mathcal{F}_U$ to be the images in M of all the flow curves. Equivalently, one can also define the leaves of $\mathcal{F}_U$ to be the connected components of the level sets $f^{-1}(y)$ of f (for y in the image of f).

Recall that we ultimately aim to build a foliation $\mathcal{F}$ of M starting from the closed 1-form ω. Gluing together a bunch of local foliations obtained via the previous construction requires one to make non-trivial adjustments. To avoid any further technicalities, we content ourself with the Fröbenius Theorem in differential topology that gives a sufficient condition for a subbundle E of TM to arise from a foliation of M. This condition is fulfilled when E is defined to be the kernel of the closed form ω, i.e. for every $p \in M$ let $E_p := \ker \omega(p)$. A transverse measure for $\mathcal{F}$ is obtained by integrating against the 1-form ω:

$$\mu(\gamma) := \int_\gamma |\omega|$$

for any arc γ transverse to $\mathcal{F}$. If $\omega = df$ is exact, then μ simply measures the variation of $\gamma\colon [a, b] \to M$ with respect to f:

$$\mu(\gamma) = |f(\gamma(a)) - f(\gamma(b))|.$$

It is immediate that such a μ is invariant under leaf-preserving isotopies.

Conversely, let $(\mathcal{F}, \mu)$ denote a measured foliation of a surface M in the sense of Definition 1.5. In some chart $(x, y)\colon U \to \mathbb{R}^2$ in a neighbourhood of a regular point, μ is induced by the closed form $\omega = dy$ so that

$$\mu(\gamma) = \int_\gamma |dy|.$$

The level sets of $y\colon U \to \mathbb{R}$ precisely correspond to the leaves of $\mathcal{F}$. One must be careful at this point and remark that taking $\omega = -dy$ serves the same purpose. In conclusion, we can always describe a measured foliation, at least locally and away from singularities, in terms of a closed 1-form.

1.5.2 Singular Euclidian metric

If we are given two transverse measured foliations $(\mathcal{F}^1, \mu_1)$ and $(\mathcal{F}^2, \mu_2)$ of a surface M, then the two transverse measures generate a singular Euclidian metric

$$d\mu^2 := d\mu_1^2 + d\mu_2^2.$$

Locally in a chart $(x, y)\colon U \to \mathbb{R}^2$ around a regular point, $d\mu$ is simply the standard flat Euclidian metric $d\mu^2 = dx^2 + dy^2$. By a *singular Euclidian metric* on M, we mean a metric that is flat outside a finite number of points around which we glue together flat rectangles making the singular point look like a cone with concentrated curvature or in some sense a Dirac mass. In other words, if X denotes the finite set of singularities of $\mathcal{F}_1$ and $\mathcal{F}_2$, then a singular Euclidian metric on M is a flat metric on the punctured surface $M \setminus X$. An in-depth work on singular Euclidian metrics on surfaces has been written by Troyanov [29].

At a singularity, the foliations have the shape of a p-pronged saddle with $p \geq 3$. Switching to polar coordinates, the neighbourhood of a singularity can be seen as a cone with total angle $p\pi$ defined as

$$V_p := \{(r, \theta) : r \geq 0 \text{ and } \theta \in \mathbb{R}/(p\pi\,\mathbb{Z})\}$$

with the metric $ds^2 = dr^2 + r^2 d\theta^2$. This makes explicit the fact that $d\mu$ corresponds to a singular norm on the tangent bundle with zeros being the singularities of the foliations. Since by assumption $p \geq 2$, we visualize V_p as a saddle shaped surface in the Euclidian space $\mathbb{R}^3$ as illustrated in Figure 1.15.

The cone V_p is isometric to $\mathbb{C}$ with the metric $ds^2 = |z|^{p-2}|dz^2|$. Observe that a small neighbourhood of a regular point is isometric to some neighbourhood of the origin in $\mathbb{C}$ with metric $ds^2 = |dz^2|$. A precise formulation of the isometry is given by Troyanov in [29, §1, Proposition 1]. The isometry sends a point $(r, \theta) \in V_p$ to a complex number with argument $2\theta/(p-2)$. The length form around a singularity is thus given by

$$|d\mu| = |z|^{(p-2)/2}|dz| = |z|^{(p-2)/2}\sqrt{dx^2 + dy^2}.$$

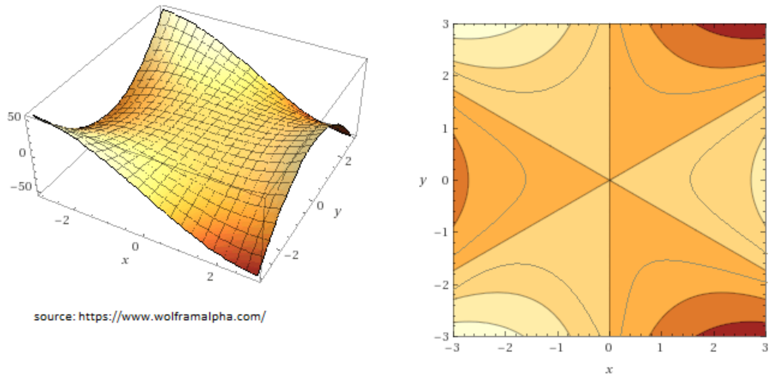

Figure 1.15: A 3-pronged saddle embedded in $\mathbb{R}^3$ on the left and the corresponding diagram of a singularity of a foliation on the right-hand side. The precise function plotted for the illustration is $(x, y) \mapsto x^3 + 3xy^2$ for $-3 \leq x, y \leq 3$.

Troyanov proved that given any arc γ in M, there is a unique shortest arc among all arcs homotopic to γ with fixed endpoints (see [29, §3, Proposition 1] or [8, Subsection 11.3.1]). The following result establishes *geodesic convexity* in the neighbourhood of singularities. The proof is transcribed from [28, Theorem 8.1]. The same result holds for any point in a Riemannian manifold. It is a standard result in Riemannian geometry due to Whitehead.

Proposition 1.25. *Around any singularity of the foliations, there is a neighbourhood U such that any two points in U can be joined by a unique shortest arc that lies inside U.*

Proof. Let s be a singularity of the foliations and V be a ball around s isometric to a ball of radius r centred at the origin in the complex plane with metric $d\mu^2 = |z|^{p-2}dz^2$ and so that s is identified with the origin. Observe that in the μ-metric the radius of the circle $|z| = r$ is $r^{p/2}$. Let $U := \{z : |z|_\mu < r \cdot 2^{-2/p}\}$ so that any two points in U can be joined by an arc (namely the concatenation of the two radii from the points to the origin) shorter in μ-metric than any arc that leaves the ball of radius r around the origin.

Let z_1 and z_2 be two points in U. Without loss of generality, we assume $z_1 \neq 0$. Further, up to a rotation, we assume that z_1 lies on the positive real axis (which is one of the half-leaves emanating from the singularity). Let γ be an arc than connects z_1 to z_2.

First assume that $|\arg \gamma(t)| \leq 2\pi/p$, i.e. γ never leaves the two sectors adjacent to the half-leaf through z_1. The transformation $\Phi_p(z) = z^{p/2}$ maps both sectors to closed half planes. The arc γ is mapped to the arc $\Phi_p(\gamma)$ that is longer in the Euclidian metric than the segment joining $\Phi_p(z_1)$ to $\Phi_p(z_2)$. In this case, the shortest arc between z_1 and z_2 is given by the pre-image of the segment joining $\Phi_p(z_1)$ to $\Phi_p(z_2)$.

Now assume that γ leaves one of the two adjacent sectors for the first time at a point ζ. The previous argument shows that the subarc of γ connecting z_1 to ζ is longer than the sum of the two radii from the origin to z_1 and ζ. Therefore, if $z_2 = 0$, then the shortest arc from z_1 to $z_2 = 0$ is the radius from the origin to z_1. Similarly, if z_2 lies in the interior of a sector non-adjacent to z_1, then the shortest arc from z_1 to z_2 is the concatenation of the two radii. Both cases are illustrated by Figure 1.16. $\square$

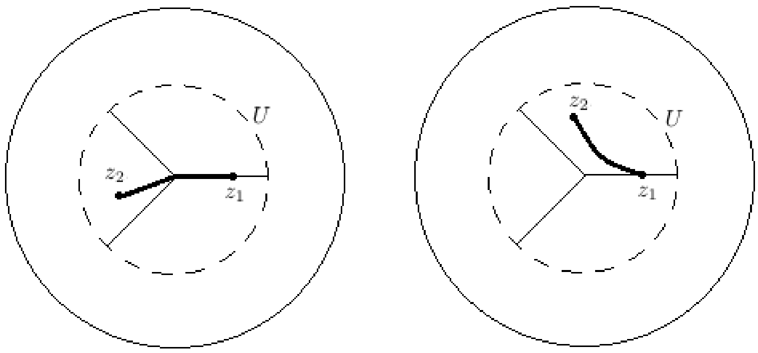

Figure 1.16: Local description for $p = 3$. The two main situations are illustrated. On the left-hand side, the shortest path is the concatenation of two radii that makes an angle at least $2\pi/p$. On the right-hand side, the shortest path is the preimage under Φ_p of the segment joining $\Phi_p(z_1)$ and $\Phi_p(z_2)$.

The length of any rectifiable arc γ in M is written

$$\ell_\mu(\gamma) := \int_\gamma |d\mu|.$$

Let $d_\mu(\cdot, \cdot)$ denotes the shortest distance between two points with respect to the length form $d\mu$. Given any homotopy class of simple closed curves c, let $L_\mu(c)$ denotes the length of the shortest representative in the class c.

Given any Riemannian metric ρ on M, we would like to compare it to μ, i.e. find positive constants k and K such that

$$k \leq \frac{L_\rho(c)}{L_\mu(c)} \leq K$$

for any homotopy class of simple closed curves c where L_ρ is defined analogously to L_μ. This comparison criteria would be immediate if μ were a regular metric as the corresponding norms on the tangent bundles would be comparable. Here we need to approach the question carefully as the norm associated to μ on TM is a singular norm. The following procedure was originally presented in [10, Lemma 9.22] and repeated in [8, Theorem 14.23].

Let us consider open balls $B(s_i, r)$ in the μ-metric around the singularities $\{s_i\}$ of the foliations with $r > 0$ small enough so that the balls are pairwise disjoint and such that any geodesic (with respect to any of the two metrics) joining two points on the boundary $\partial B(s_i, r)$ lies entirely in $B(s_i, r)$. The existence of such an r is guaranteed by Proposition 1.25 and Whitehead Theorem.

Outside of the balls $B(s_i, r/2)$, the length functions ℓ_μ and ℓ_ρ are comparable as it is always the case for Riemannian metrics. In other words, there exist constants k' and K' such that

$$k' \leq \frac{\ell_\rho(\gamma)}{\ell_\mu(\gamma)} \leq K'$$

for any rectifiable arc γ in $M \setminus \bigcup_i B(s_i, r/2)$.

To estimate length of curves near singularities, we first estimate the distances. If two points in $\partial B(s_i, r)$ are sufficiently close, then the geodesic arc joining them lies entirely outside $B(s_i, r/2)$ and previous inequality between lengths holds. Moreover, any complementary set of an open neighbourhood of the diagonal $\partial B(s_i, r) \times \partial B(s_i, r)$ defines a compact set on which the quotient of the distance functions d_μ and d_ρ is positive and continuous. Thus we can find constants k'', K'' such that

$$k'' \leq \frac{d_\rho(x, y)}{d_\mu(x, y)} \leq K''$$

for all $x, y \in \partial B(s_i, r)$. With $k := \min\{k', k''\}$ and $K := \max\{K', K''\}$, we obtain

$$k \leq \frac{L_\rho(c)}{L_\mu(c)} \leq K$$

for any homotopy class of simple closed curves c. To sum up, the metric generated by transverse measured foliations is not Riemannian but can always be compared to any Riemannian metric in the above sense.

1.5.3 Universal cover

Most of our work later takes place in the universal cover rather than in the surface itself. We thus explain now how a foliation of a surface lifts to a foliation of its universal cover and how the lifted metric arising from transverse measured foliations behaves.

Given a surface M we denote its universal cover by $\widetilde{M}$. Let $p \colon \widetilde{M} \twoheadrightarrow M$ denote the covering map. Recall that $\widetilde{M}$ is simply connected and every point $x \in M$ admits an open neighbourhood $x \in U_x \subset M$ such that $p^{-1}(U_x)$ is a disjoint union of open sets $V_\alpha \subset \widetilde{M}$ whose images $p(V_\alpha)$ are all homeomorphic to U_x. A homeomorphism $\tau \colon \widetilde{M} \to \widetilde{M}$ that satisfies $p \circ \tau = p$ is called a *covering translation* (or *deck transformation*) of $\widetilde{M}$;

$$\begin{tikzcd} \widetilde{M} \arrow[r, "\tau"] \arrow[dr, "p"'] & \widetilde{M} \arrow[d, "p"] \\ & M. \end{tikzcd}$$

Covering translations of $\widetilde{M}$ form a subgroup of $\mathrm{Homeo}(\widetilde{M})$ denoted $\mathrm{Homeo}_p(\widetilde{M})$ isomorphic to $\pi_1(M)$.

If ρ denotes a Riemannian metric on M, then $\widetilde{\rho} := p^\star \rho$ is a Riemannian metric on $\widetilde{M}$. The metric $\widetilde{\rho}$ is invariant under covering translations. Moreover, $p \colon (\widetilde{M}, \widetilde{\rho}) \to (M, \rho)$ is a local isometry.

If M admits a measured foliation $(\mathcal{F}, \mu)$, then we can lift it to a measured foliation $(\widetilde{\mathcal{F}}, \widetilde{\mu})$ of $\widetilde{M}$. The leaves and singularities of $\widetilde{\mathcal{F}}$ are defined as the connected components of the preimages via the covering map of the leaves and singularities of $\mathcal{F}$. Note that $\widetilde{\mathcal{F}}$ may have infinitely many singularities but only finitely many up to covering translations. In terms of closed 1-form, if

$\mathcal{F}$ is locally described by $\omega\colon TU \to \mathbb{R}$ and $p^{-1}(U) = \bigcup_\alpha V_\alpha$, then $\widetilde{\mathcal{F}}$ is locally described in each V_α by the closed 1-form $\widetilde{\omega}\colon TV_\alpha \to \mathbb{R}$ defined as $\widetilde{\omega} := p^\star\omega$.

The transverse measure $\widetilde{\mu}$ defines a pseudo-distance $\widetilde{d}(\cdot,\cdot)$ on $\widetilde{M} \times \widetilde{M}$ defined to be the minimum length of an arc connecting $\widetilde{x}$ and $\widetilde{y}$ measured by $\widetilde{\mu}$. In particular, if $\widetilde{x}$ and $\widetilde{y}$ lie in the same leaf inside some chart defining the foliation, then $\widetilde{d}(\widetilde{x},\widetilde{y}) = 0$. Conversely, if $\widetilde{d}(\widetilde{x},\widetilde{y}) = 0$, then $\widetilde{x}$ and $\widetilde{y}$ lie on the same leaf. It follows from the construction of $\widetilde{\mathcal{F}}$ that the pseudo-metric $\widetilde{d}$ is *equivariant*, i.e. $\widetilde{d}(\tau\widetilde{x},\tau\widetilde{y}) = \widetilde{d}(\widetilde{x},\widetilde{y})$ for any points $\widetilde{x},\widetilde{y} \in \widetilde{M}$ and for any covering translation τ. With a slight abuse of language, we refer to $\widetilde{d}$ as a *distance* function.

In the context of pA homeomorphisms (or simply given two transverse measured foliations), we can add the distances generated by the two measures to get a proper distance function on the universal cover. Let ϑ be a pA homeomorphism of a surface M. Let $\widetilde{\vartheta}$ be a lift of ϑ to the universal cover $\widetilde{M}$. The homeomorphism $\widetilde{\vartheta}$ makes the following diagram commute.

$$\begin{tikzcd}
\widetilde{M} \arrow[r, "\widetilde{\vartheta}"] \arrow[d, "p"'] & \widetilde{M} \arrow[d, "p"] \\
M \arrow[r, "\vartheta"'] & M
\end{tikzcd}$$

Note that the other lifts of ϑ are exactly $\{\widetilde{\vartheta}\tau : \tau \in \mathrm{Homeo}_p(\widetilde{M})\}$. A lift of ϑ is uniquely determined if we fix the image of a point. Concretely, we chose $x \in M$ and two lifts $\widetilde{x}$ of x and $\widetilde{y}$ of $y := \vartheta(x)$. Then there is a unique lift $\widetilde{\vartheta}$ of ϑ that makes the above diagram commute and satisfies $\widetilde{\vartheta}(\widetilde{x}) = \widetilde{y}$.

Let $(\mathcal{F}^s, \mu_s)$ and $(\mathcal{F}^u, \mu_u)$ be the stable and unstable foliations of M with respect to ϑ. We lift them to transverse measured foliations $(\widetilde{\mathcal{F}^s}, \widetilde{\mu}_s)$ and $(\widetilde{\mathcal{F}^u}, \widetilde{\mu}_u)$ of $\widetilde{M}$. We adopt the following convention in the notation for the distance functions associated to the transverse measures. Let $\widetilde{d}_s$ be the equivariant pseudo-distance generated by $\widetilde{\mu}_u$ and $\widetilde{d}_u$ be the equivariant pseudo-distance generated by $\widetilde{\mu}_s$. We have permuted the indices u and s so that the function $\widetilde{d}_s$ measures distances in the direction of the stable leaves and $\widetilde{d}_u$ in the direction of the unstable leaves. Doing so, we keep the same notation as Handel in [14] and [15]. The dynamics of ϑ carry to the lift $\widetilde{\vartheta}$ so that if $\lambda > 1$ denotes the stretch factor of ϑ, then

$$\widetilde{d}_s(\widetilde{\vartheta}(\widetilde{x}), \widetilde{\vartheta}(\widetilde{y})) = \lambda^{-1} \cdot \widetilde{d}_s(\widetilde{x},\widetilde{y}),$$
$$\widetilde{d}_u(\widetilde{\vartheta}(\widetilde{x}), \widetilde{\vartheta}(\widetilde{y})) = \lambda \cdot \widetilde{d}_u(\widetilde{x},\widetilde{y}).$$

If $\partial\widetilde{M} = \emptyset$, the function $\widetilde{d}\colon \widetilde{M} \times \widetilde{M} \to [0,\infty)$ defined by $\widetilde{d} := \widetilde{d}_s + \widetilde{d}_u$ defines an equivariant distance on $\widetilde{M}$. In particular, $\widetilde{d}_s(\widetilde{x},\widetilde{y}) = 0$ and $\widetilde{d}_u(\widetilde{x},\widetilde{y}) = 0$ implies $\widetilde{x} = \widetilde{y}$.

If $\widetilde{M}$ has nonempty boundary, then stable and unstable leaves may coincide on the boundary. Therefore, we can have $\widetilde{d}_s(\widetilde{x},\widetilde{y}) = 0$ and $\widetilde{d}_u(\widetilde{x},\widetilde{y}) = 0$ for distinct points $\widetilde{x}$ and $\widetilde{y}$ on the same boundary component. In this case, the function $\widetilde{d} = \widetilde{d}_s + \widetilde{d}_u$ is an equivariant distance on $\mathrm{int}(\widetilde{M})$.

1.6 Topological entropy

The topological entropy is an important invariant in dynamical systems that attempts to measure the complexity of a system in a single number. We assume that the reader is familiar with the notion of entropy and its various basic properties. For a detailed introduction one may consult for instance Katok and Hasselblatt [19, Chapter 3].

The topological entropy of a map f defined on a compact metric space is a nonnegative real number denoted by $h_{top}(f) \in [0, +\infty]$. If we deal with non-compact metric spaces the common definitions of topological entropy in terms of open covers or spanning sets do not apply any more. In our context, the universal cover of a surface is in general non-compact. Hence we need a definition of entropy for non-compact metric spaces in order to relate the entropy of a surface homeomorphism to the geometry of the universal cover. The results in this section are treated in more details in [10].

Definition 1.10 (entropy of non-compact space maps)**.** Let X be a metric space (not necessarily compact) and $f\colon X \to X$ be a continuous map. For a compact subspace $K \subset X$, we define

$$h_K(f) := \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \begin{cases} \operatorname{span}_K(n, \epsilon), \\ \operatorname{sep}_K(n, \epsilon), \end{cases}$$

where $\operatorname{span}_K(n, \epsilon)$ is the minimum size of an (n, ϵ)-spanning set for K and $\operatorname{sep}_K(n, \epsilon)$ is the maximum size of an (n, ϵ)-separated set for K. Set

$$h_X(f) := \sup_{K \subset X} h_K(f) \in [0, +\infty].$$

If X is compact in Definition 1.10, then we recover the classical topological entropy, i.e. $h_{top}(f) = h_X(f)$. This is a consequence of the Lebesgue Number Lemma.

The next proposition relate the topological entropy of a continuous map f of a surface M to the dynamics of a lift $\widetilde{f}$ to the universal cover $\widetilde{M}$.

Proposition 1.26. *Let M be a surface and $\widetilde{M}$ be its universal cover with covering map $p\colon \widetilde{M} \twoheadrightarrow M$. Assume $\widetilde{M}$ is equipped with a metric $\widetilde{d}$. Let $f\colon M \to M$ be a continuous map and $\widetilde{f}\colon \widetilde{M} \to \widetilde{M}$ be a lift of f. Then the following estimation of $h_{top}(f)$ holds:*

$$h_{top}(f) \geq \limsup_{n \to +\infty} \frac{1}{n} \log \widetilde{d}\left(\widetilde{f}^n(\widetilde{x}), \widetilde{f}^n(\widetilde{y})\right)$$

for any points $\widetilde{x}$ and $\widetilde{y}$ in $\widetilde{M}$.

Proof. Let $\widetilde{x}$ and $\widetilde{y}$ be any two points in $\widetilde{M}$. Let $n \geq 1$ be an integer and $\epsilon > 0$. Let $\widetilde{\alpha}\colon I \to \widetilde{M}$ be a path from $\widetilde{x}$ to $\widetilde{y}$ and $\{\widetilde{z}_1, \ldots, \widetilde{z}_\ell\} \subset \widetilde{\alpha}(I)$ be an $(n+1, \epsilon)$-spanning set of minimal size for the compact subspace $\widetilde{\alpha}(I) \subset \widetilde{M}$. In particular,

$$\widetilde{f}^n(\widetilde{\alpha}) \subset \bigcup_{i=1}^{\ell} B(\widetilde{f}^n(\widetilde{z}_i), \epsilon).$$

Therefore,

$$\widetilde{d}(\widetilde{f}^n(\widetilde{x}), \widetilde{f}^n(\widetilde{y})) \leq \operatorname{diam}(\widetilde{f}^n(\widetilde{\alpha})) \leq 2l\epsilon.$$

Taking limits, we obtain

$$h_{\widetilde{\alpha}}(\widetilde{f}) := h_{\widetilde{\alpha}(I)}(\widetilde{f}) \geq \limsup_{n\to+\infty} \frac{1}{n} \log \widetilde{d}(\widetilde{f}^n(\widetilde{x}), \widetilde{f}^n(\widetilde{y})).$$

Let $\alpha := p(\widetilde{\alpha}) \subset M$ and $y := p(\widetilde{y})$. To conclude we need to compare $h_{\widetilde{\alpha}}(\widetilde{f})$ and $h_\alpha(f) := h_{\alpha(I)}(f)$. Since both $\widetilde{\alpha}(I)$ and $\alpha(I)$ are compact there exists $m \geq 1$, such that for ϵ sufficiently small, $p^{-1}(B(y,\epsilon)) \cap \widetilde{\alpha}(I)$ can be covered by m balls of radius 2ϵ. Using uniform continuity of f on M, we can build, from an (n,ϵ)-spanning set in $\alpha(I)$, an $(n, 2\epsilon)$-spanning set in $\widetilde{\alpha}(I)$ that is m times larger. Hence

$$\operatorname{span}_{\alpha(I)}(n,\epsilon,f) \geq \operatorname{span}_{\widetilde{\alpha}(I)}(n,2\epsilon,\widetilde{f}) \cdot m$$

and thus $h_\alpha(f) \geq h_{\widetilde{\alpha}}(\widetilde{f})$. The conclusion follows. $\square$

Proposition 1.26 applies to give a lower bound for the entropy of a pA homeomorphism and at the same time shows the strength of working in the universal cover.

Corollary 1.27. *Let M be a surface and $\vartheta\colon M \to M$ a pA homeomorphism of M with stretch factor $\lambda > 1$. Then the following inequality holds:*

$$h_{top}(\vartheta) \geq \log \lambda > 0.$$

It turns out that the lower bound provided by Corollary 1.27 is sharp.

Proposition 1.28. *Let M be a surface and $\vartheta\colon M \to M$ a pA homeomorphism of M. If $\lambda > 1$ denotes the stretch factor of ϑ, then*

$$h_{top}(\vartheta) = \log \lambda > 0.$$

As a direct consequence of Proposition 1.28 we observe that a pA homeomorphism determines a unique stretch factor. For a detailed proof of Proposition 1.28 see [10, Proposition 10.13]. The proof of the reverse inequality ($h_{top}(\vartheta) \leq \log(\lambda)$) relies on the construction of a Markov partition for ϑ and the conclusion that ϑ is a factor of some subshift of finite type.

The Thurston-Nielsen classification (Theorem 1.16) tells us that irreducible classes of surface homeomorphisms are either pA or have finite-order. Corollary 1.27 establishes that pA transformations have positive entropy. A finite-order transformation has an iteration which is the identity map and therefore has zero topological entropy. What is more, a canonical reducible representative has non-zero topological entropy if and only if it has at least one pA component.

Chapter 2

Isotopy stability of periodic orbits

In two consecutive papers published in 1985 [14] and 1987 [15], Michael Handel first proved that a pA homeomorphism ϑ of a closed surface M minimizes the topological entropy in its isotopy class $[\vartheta] \in \mathrm{MCG}(M)$. Two years later he refined the previous result by showing that only maps to which ϑ is semi-conjugate reach the minimal entropy. Recall that conjugacy is the notion of isomorphism in the category of dynamical systems. The semi-conjugacy built by Handel identifies periodic points that are preserved by an isotopic perturbation of ϑ. This chapter aims to state and prove Handel's Theorem about unremovability of periodic orbits.

Standing assumptions. The following assumptions will prevail throughout Chapter 2. Unless otherwise noted, M denotes a **closed** surface with universal cover $\widetilde{M}$ and covering map $p\colon \widetilde{M} \twoheadrightarrow M$. Further, ϑ denotes a pA homeomorphism of M with stretch factor $\lambda > 1$ and $g \in [\vartheta] \in \mathrm{MCG}(M)$ is isotopic to ϑ. The existence of ϑ equips M with a transverse stable foliation $(\mathcal{F}^s, \mu_s)$ and an unstable measured foliation $(\mathcal{F}^u, \mu_u)$. Moreover, $\widetilde{d} = \widetilde{d}_s + \widetilde{d}_u$ denotes the equivariant distance on $\widetilde{M}$ generated by the transverse measured foliations on $\widetilde{M}$ obtained by lifting the stable and unstable foliations on M.

2.1 Global shadowing and Nielsen equivalence

Let $h_t\colon M \to M$ be an isotopy from ϑ to g with $h_0 = \vartheta$ and $h_1 = g$. If we fix a lift $\widetilde{\vartheta}$ of ϑ, then the isotopy h_t lifts uniquely to an isotopy $\widetilde{h}_t\colon \widetilde{M} \to \widetilde{M}$ with $\widetilde{h}_0 = \widetilde{\vartheta}$ and such that the following diagram commutes.

$$\begin{tikzcd}
 & & \widetilde{M} \arrow[d, "p"] \\
I \times \widetilde{M} \arrow[urr, dashed, "\widetilde{h}_\bullet"] \arrow[r, "id_I \times p"'] & I \times M \arrow[r, "h_\bullet"'] & M
\end{tikzcd}$$

The map $\widetilde{g} := \widetilde{h}_1$ is a lift of g uniquely determined by h_t and $\widetilde{\vartheta}$. We refer to $\widetilde{g}$ as the lift of g *equivariantly homotopic* to $\widetilde{\vartheta}$. In other words, the isotopy h_t induces a one-to-one correspondence between the lifts of ϑ and the lifts of g:

$$\{\widetilde{\vartheta} : p \circ \widetilde{\vartheta} = \vartheta \circ p\} \overset{h_t}{\longleftrightarrow} \{\widetilde{g} : p \circ \widetilde{g} = g \circ p\}.$$

The notion used by Handel to build a semi-conjugacy between ϑ and g is called the global shadowing of orbits. Before introducing this concept, we complete our standing assumptions for Chapter 2 as follows.

Standing assumptions. Unless otherwise noted, $\widetilde{\vartheta}$ denotes an arbitrarily fixed lift of ϑ and $\widetilde{g}$ is the unique lift of g equivariantly homotopic to $\widetilde{\vartheta}$ with respect to the choice of some isotopy h_t.

Remark. If $\tau \in \mathrm{Homeo}_p(\widetilde{M})$ denotes a covering translation, then $\widetilde{\vartheta} \circ \tau$ is another lift of ϑ and hence there is $\sigma \in \mathrm{Homeo}_p(\widetilde{M})$ such that $\widetilde{\vartheta} \circ \tau = \sigma \circ \widetilde{\vartheta}$. By continuity of the isotopy, we must have $\widetilde{h}_t \circ \tau = \sigma \circ \widetilde{h}_t$ for every $t \in I$. In particular $\widetilde{g} \circ \tau = \sigma \circ \widetilde{g}$.

The compactness of M, together with $\widetilde{g}$ being equivariantly homotopic to $\widetilde{\vartheta}$, implies the following lemma.

Lemma 2.1. *There exists $K > 0$ such that*

$$\sup_{\widetilde{x} \in \widetilde{M}} \left\{ \widetilde{d}\left(\widetilde{\vartheta}(\widetilde{x}), \widetilde{g}(\widetilde{x})\right), \widetilde{d}\left(\widetilde{\vartheta}^{-1}(\widetilde{x}), \widetilde{g}^{-1}(\widetilde{x})\right) \right\} \leq K.$$

Since $\widetilde{d}$ is an equivariant distance, the same K works as an upper bound if we replace $\widetilde{\vartheta}$ by another lift of ϑ and $\widetilde{g}$ by the corresponding equivariantly homotopic lift of g in the conclusion of Lemma 2.1.

2.1.1 Global shadowing

Definition 2.1 (global shadowing)**.** Let f_1 and f_2 be two isotopic homeomorphisms of M and $\delta > 0$ be a real number. We say that the f_1-orbit of $x \in M$ *globally δ-shadows* the f_2-orbit of $y \in M$ if there exist

1. equivariantly homotopic lifts $\widetilde{f_1}$ of f_1 and $\widetilde{f_2}$ of f_2,
2. a lift $\widetilde{x}$ of x and a lift $\widetilde{y}$ of y,

such that

$$\sup_{k \in \mathbb{Z}} \widetilde{d}\left(\widetilde{f_1}^k(\widetilde{x}), \widetilde{f_2}^k(\widetilde{y})\right) \leq \delta.$$

We write $(x, f_1) \sim^{\delta} (y, f_2)$. We say furthermore that the f_1-orbit of $x \in M$ *globally shadows* the f_2-orbit of $y \in M$ if $(x, f_1) \sim^{\varepsilon} (y, f_2)$ for some $\varepsilon > 0$. We write

$$(x, f_1) \sim (y, f_2).$$

Observe that as $\widetilde{d}$ is an equivariant distance, if the boundedness condition for global shadowing is verified for a pair of equivariant lifts, then it is verified for any pair of equivariant lifts. Moreover, the notion of global shadowing does not depend on the choice of the original metric on M in its definition. However it is convenient for computations to use the distance $\widetilde{d}$ associated to the transverse foliations.

The notion of global shadowing can be defined in terms of arcs in M. This equivalence is enlightened in the next proposition. Let us first introduce some notation. If ρ denotes a Riemannian metric (or a singular Euclidian metric) on M and ℓ is the corresponding length function, then for an arc γ in M we write

$$L(\gamma) := \inf\{\ell(\beta) : \beta \text{ is homotopic to } \gamma \text{ with fixed endpoints}\}.$$

Proposition 2.2. *Let f_1 and f_2 be two isotopic homeomorphisms of M and $h_t \colon M \to M$ be an isotopy from f_1 to f_2. Let x and y be two points in M. Then the following are equivalent:*

1. $(x, f_1) \sim (y, f_2)$,
2. *there exists an arc $\gamma \colon I \to M$ from x to y and $\delta > 0$ such that*

$$\sup_{k\in\mathbb{Z}} L\left((h_t^k \circ \gamma)(t)\right) < \delta.$$

Observe that if Proposition 2.2 holds for some Riemannian metric (generating the function L), then it holds for any Riemannian metric and also for the singular Euclidean metric $d\mu^2 = d\mu_s^2 + d\mu_u^2$. This is a consequence of the comparison criteria developed previously (see Proposition 1.25).

Proof. We start by fixing $\widetilde{f_1}$ and $\widetilde{f_2}$ two equivariant lifts of f_1 and f_2 via some lifted isotopy $\widetilde{h}_t \colon \widetilde{M} \to \widetilde{M}$.

First assume that $(x, f_1) \sim (y, f_2)$. By assumption, there is $\delta > 0$ such that

$$\sup_{k\in\mathbb{Z}} \widetilde{d}\left(\widetilde{f_1}^k(\widetilde{x}), \widetilde{f_2}^k(\widetilde{y})\right) < \delta$$

for some lifts $\widetilde{x}$ of x and $\widetilde{y}$ of y. Choose an arc $\widetilde{\gamma}$ connecting $\widetilde{x}$ and $\widetilde{y}$. For any $k \in \mathbb{Z}$ we introduce

$$\widetilde{\alpha}_k(t) := (\widetilde{h}_t^k \circ \widetilde{\gamma})(t).$$

The arc $\widetilde{\alpha}_k$ connects $\widetilde{f}_1^k(\widetilde{x})$ to $\widetilde{f}_2^k(\widetilde{y})$. Since $\widetilde{M}$ is simply connected, $\widetilde{\alpha}_k$ is homotopic with fixed endpoints to the geodesic arc $\widetilde{\gamma}_k$ from $\widetilde{f}_1^k(\widetilde{x})$ to $\widetilde{f}_2^k(\widetilde{y})$. Therefore

$$L\left((h_t^k \circ \gamma)(t)\right) = L(p \circ \widetilde{\alpha}_k) \le \ell(p \circ \widetilde{\gamma}_k) < \delta.$$

To prove the converse, assume there is an arc γ in M from x to y and $\delta > 0$ such that

$$\sup_{k\in\mathbb{Z}} L\left(\alpha_k\right) < \delta$$

where $\alpha_k(t) := (h_t^k \circ \gamma)(t)$. For $k \in \mathbb{Z}$ let β_k be an arc from x to y homotopic with fixed endpoints to α_k such that $\ell(\beta_k) < \delta$. Lift γ to an arc $\widetilde{\gamma}$ starting at some lift $\widetilde{x}$ of x. Let $\widetilde{y} := \widetilde{\gamma}(1)$. Lift α_k to the arc $\widetilde{\alpha}_k$ in $\widetilde{M}$ that starts at $\widetilde{f}_1^k(\widetilde{x})$. By uniqueness of lift, $\widetilde{\alpha}_k(t) = (\widetilde{h}_t^k \circ \widetilde{\gamma})(t)$. Therefore $\widetilde{\alpha}_k(1) = \widetilde{f}_2^k(\widetilde{y})$.

Similarly, we lift β_k to an arc $\widetilde{\beta}_k$ in $\widetilde{M}$ that starts at $\widetilde{f}_1^k(\widetilde{x})$. Because of the existence of a homotopy with fixed endpoints between β_k and α_k, we must have $\widetilde{\beta}_k(1) = \widetilde{\alpha}_k(1) = \widetilde{f}_2^k(\widetilde{y})$. Recall that if we are working with a Riemannian metric, then the covering map p is locally an isometry. Hence, by decomposing the distance into sufficiently many pieces via the triangle inequality, we eventually obtain

$$\widetilde{d}\left(\widetilde{f_1}^k(\widetilde{x}), \widetilde{f_2}^k(\widetilde{y})\right) \leq \ell(\beta_k) < \delta.$$

Since k is arbitrary, the desired conclusion follows. □

Global shadowing defines an equivalence relation on the set of pairs (x, f) where x is a point in M and f varies among all the homeomorphisms in the same isotopy class. Furthermore, if $f_1 = f_2 = \vartheta$, then the pA character of ϑ implies that each point lies alone in its class. This is the purpose of the following lemma.

Lemma 2.3. $(x, \vartheta) \sim (y, \vartheta) \Rightarrow x = y$.

Proof. Assuming $(x, \vartheta) \sim (y, \vartheta)$, there is $\delta > 0$ such that

$$\sup_{k \in \mathbb{Z}} \widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}), \widetilde{\vartheta}^k(\widetilde{y})\right) \leq \delta$$

for some lifts $\widetilde{x}$ of x and $\widetilde{y}$ of y. For every $k \geq 0$ we have

$$0 \leq \lambda^k \cdot \widetilde{d_u}(\widetilde{x}, \widetilde{y}) = \widetilde{d_u}\left(\widetilde{\vartheta}^k(\widetilde{x}), \widetilde{\vartheta}^k(\widetilde{y})\right) \leq \delta$$

and hence $\widetilde{d_u}(\widetilde{x}, \widetilde{y}) = 0$. Similarly, for all $k \leq 0$ we have

$$0 \leq \lambda^k \cdot \widetilde{d_s}(\widetilde{x}, \widetilde{y}) = \widetilde{d_s}\left(\widetilde{\vartheta}^{-k}(\widetilde{x}), \widetilde{\vartheta}^{-k}(\widetilde{y})\right) \leq \delta$$

and thus $\widetilde{d_s}(\widetilde{x}, \widetilde{y}) = 0$. Therefore $\widetilde{d}(\widetilde{x}, \widetilde{y}) = 0$ and the desired conclusion follows. □

Remark (global shadowing classes for surfaces with boundary)**.** Lemma 2.3 shows that on a closed surface, the global shadowing equivalence classes associated to ϑ contain a single element. However, if M has nonempty boundary, then $\widetilde{d}$ is not a distance strictly speaking. In particular, the final step in the proof of Lemma 2.3 does not hold. The correct adjustment to make, as noted by Handel in a final remark in [14, Remark 2.4], consists in refining the conclusion of Lemma 2.3 by specifying that any two points on the same boundary component lie in the same equivalence class.

The next elementary lemma shows that global shadowing is preserved at the limit under the assumption of compactness for M.

Lemma 2.4. *Assume $x_n \to x$ and $y_n \to y$ are two converging sequences in M with $(x_n, \vartheta) \sim (y_n, g)$ for all $n \geq 1$, then $(x, \vartheta) \sim (y, g)$.*

Proof. For $n \geq 1$, let

$$\delta_n = \sup_{k \in \mathbb{Z}} \widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}_n), \widetilde{g}^k(\widetilde{y}_n)\right) < +\infty.$$

Assume *ab absurdo* that the collection of δ_n is unbounded. In other words, $\sup_{n \geq 1} \delta_n = +\infty$. In particular, we can find some integer $m \geq 1$ sufficiently large and $N \geq 1$ such that for all $k \geq N$

$$\frac{K+1}{\lambda - 1} \leq \widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}_m), \widetilde{g}^k(\widetilde{y}_m)\right) \leq \delta_m,$$

where $\lambda > 1$ is the stretch factor of ϑ and K refers to the upper bound provided by Lemma 2.1. Note that for any two points $\widetilde{w}$ and $\widetilde{z}$ in $\widetilde{M}$, the triangle inequality gives

$$\begin{aligned}\widetilde{d}\left(\widetilde{\vartheta}(\widetilde{w}),\widetilde{g}(\widetilde{z})\right) &\geq \widetilde{d}\left(\widetilde{\vartheta}(\widetilde{w}),\widetilde{\vartheta}(\widetilde{z})\right) - \widetilde{d}\left(\widetilde{\vartheta}(\widetilde{z}),\widetilde{g}(\widetilde{z})\right)\\ &\geq (\lambda+\lambda^{-1})\widetilde{d}(\widetilde{w},\widetilde{z}) - K.\end{aligned}$$

Hence, if $\widetilde{d}(\widetilde{w},\widetilde{z}) \geq (K+1)/(\lambda-1)$, then

$$\widetilde{d}\left(\widetilde{\vartheta}(\widetilde{w}),\widetilde{g}(\widetilde{z})\right) \geq \widetilde{d}(\widetilde{w},\widetilde{z}) + 1.$$

Therefore, for all $k \geq N$

$$\widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}_m),\widetilde{g}^k(\widetilde{y}_m)\right) \geq \widetilde{d}\left(\widetilde{\vartheta}^N(\widetilde{x}_m),\widetilde{g}^N(\widetilde{y}_m)\right) + k - N.$$

This is a contradiction since $\widetilde{d}(\widetilde{\vartheta}^k(\widetilde{x}_m),\widetilde{g}^k(\widetilde{y}_m))$ is not larger than δ_m for every k. We conclude that the collection of δ_n is bounded:

$$\sup_{n\geq 1}\delta_n < +\infty.$$

For any $k \in \mathbb{Z}$ and $n \geq 1$, the triangle inequality again gives

$$\begin{aligned}\widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}),\widetilde{g}^k(\widetilde{y})\right) \leq \widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}),\widetilde{\vartheta}^k(\widetilde{x}_n)\right) &+ \widetilde{d}\left(\widetilde{\vartheta}^k(\widetilde{x}_n),\widetilde{g}^k(\widetilde{y}_n)\right)\\ &+ \widetilde{d}\left(\widetilde{g}^k(\widetilde{y}_n),\widetilde{g}^k(\widetilde{y})\right).\end{aligned}$$

Both $d(\widetilde{\vartheta}^k(\widetilde{x}),\widetilde{\vartheta}^k(\widetilde{x}_n))$ and $d(\widetilde{g}^k(\widetilde{y}),\widetilde{g}^k(\widetilde{y}_n))$ converge to zero as n grows. The remaining term $d(\widetilde{\vartheta}^k(\widetilde{x}_n),\widetilde{g}^k(\widetilde{y}_n))$ is bounded by $\sup_{n\geq 1}\delta_n < +\infty$ for every $n \geq 1$. Hence $(x,\vartheta) \sim (y,g)$ as required. $\square$

There exists an equivalent description of global shadowing for periodic orbits of pA homeomorphisms. This notion is called Nielsen equivalence. In contrary to global shadowing, the concept of Nielsen equivalence does not involve directly the metric $\widetilde{d}$ in its definition.

2.1.2 Nielsen equivalence

Definition 2.2 (Nielsen equivalence)**.** Let $n \geq 1$ be an integer. Let f_1 and f_2 be two isotopic homeomorphisms of M. Furthermore, let x be a fixed point of f_1^n and y be a fixed point of f_2^n. We say that the f_1-orbit of x is *Nielsen equivalent* to the f_2-orbit of y if there exist

1. a covering translation $\tau \in \mathrm{Homeo}_p(\widetilde{M})$,
2. equivariantly homotopic lifts $\widetilde{f_1}$ of f_1 and $\widetilde{f_2}$ of f_2
3. a lift $\widetilde{x}$ of x and a lift $\widetilde{y}$ of y,

such that $\widetilde{f_1}^n(\widetilde{x}) = \tau\widetilde{x}$ and $\widetilde{f_2}^n(\widetilde{y}) = \tau\widetilde{y}$. We write

$$(x, f_1^n) \overset{\text{NE}}{\sim} (y, f_2^n).$$

Observe that if x is a fixed point of f_1^n and y is a fixed point of f_2^n, then for every lift $\widetilde{x}$ of x and $\widetilde{y}$ of y, there exist two covering translations τ and σ such that $\widetilde{f_1^n}(\widetilde{x}) = \tau\widetilde{x}$ and $\widetilde{f_2^n}(\widetilde{y}) = \sigma\widetilde{y}$. Nielsen equivalence describes the situation where one can find a pair of lifts with $\tau = \sigma$. If $f_1 = f_2 =: f$, then Nielsen equivalence defines an equivalence relation among the elements of $\mathrm{Fix}(f^n)$. Similarly as for global shadowing, Nielsen equivalence for periodic orbits has a reformulation in terms of arcs.

Proposition 2.5. *Let $n \geq 1$ be an integer. Let f_1 and f_2 be two isotopic homeomorphisms of M and $h_t\colon M \to M$ be an isotopy from f_1 to f_2. Furthermore, let x be a fixed point of f_1^n and y be a fixed point of f_2^n. Then the following are equivalent:*

1. $(x, f_1^n) \overset{\mathrm{NE}}{\sim} (y, f_2^n)$,
2. *there is an arc $\gamma\colon I \to M$ from x to y such that $(h_t^n \circ \gamma)(t)$ is homotopic to $\gamma(t)$ with fixed endpoints.*

Proof. Without lost of generality, up to replacing the functions by their iterates, we assume that $n = 1$.

First assume that $(x, f_1) \overset{\mathrm{NE}}{\sim} (y, f_2)$. Hence there exist equivariant lifts $\widetilde{f}_1$ of f_1 and $\widetilde{f}_2$ of f_2, lifts $\widetilde{x}$ of x and $\widetilde{y}$ of y and a covering translation τ such that

$$\widetilde{f}_1(\widetilde{x}) = \tau\widetilde{x} \text{ and } \widetilde{f}_2(\widetilde{y}) = \tau\widetilde{y}.$$

Let $\widetilde{\gamma}$ be any arc in $\widetilde{M}$ from $\widetilde{x}$ to $\widetilde{y}$. Let $\widetilde{h}_t$ be the lifted isotopy from $\widetilde{f}_1$ to $\widetilde{f}_2$. The arc $\widetilde{\alpha}(t) := (\widetilde{h}_t \circ \widetilde{\gamma})(t)$ in $\widetilde{M}$ connects $\tau\widetilde{x}$ to $\tau\widetilde{y}$. So does the arc $\tau\widetilde{\gamma}$. Since $\widetilde{M}$ is simply connected, there is a homotopy with fixed endpoints between $\widetilde{\alpha}$ and $\tau\widetilde{\gamma}$ that descends to the desired homotopy in M.

To prove the converse, assume that there is an arc $\gamma\colon I \to M$ from x to y such that $(h_t \circ \gamma)(t)$ is homotopic to $\gamma(t)$ with fixed endpoints. Let $\widetilde{f}_1$ be any lift of f_1 and τ be the covering translation such that $\widetilde{f}_1(\widetilde{x}) = \tau\widetilde{x}$ for some lift $\widetilde{x}$ of x. Let $\widetilde{f}_2$ be the lift of f_2 equivariantly homotopic to $\widetilde{f}_1$ with respect to the lifted isotopy $\widetilde{h}_t\colon \widetilde{M} \to \widetilde{M}$.

We write $\alpha(t) := (h_t \circ \gamma)(t)$. Lift γ to an arc $\widetilde{\gamma}$ in $\widetilde{M}$ that starts at $\widetilde{x}$. Let $\widetilde{y} := \widetilde{\gamma}(1)$. Lift α to an arc $\widetilde{\alpha}$ in $\widetilde{M}$ that starts at $\tau\widetilde{x}$. By uniqueness of lifts, we have $\widetilde{\alpha}(t) = (\widetilde{h}_t \circ \widetilde{\gamma})(t)$.

If we write $\widetilde{y}_1 := \widetilde{\alpha}(1)$, then we have $\widetilde{y}_1 = \widetilde{h}_1(\widetilde{y}) = \widetilde{f}_2(\widetilde{y})$. On the other hand, $p \circ \widetilde{\alpha}$ and γ are homotopic with fixed endpoints. Hence the arcs $\tau^{-1}\widetilde{\alpha}$ and $\widetilde{\gamma}$ in $\widetilde{M}$ have the same endpoints. In other words, $\tau^{-1}\widetilde{y}_1 = \widetilde{y}$ and therefore $\widetilde{f}_2(\widetilde{y}) = \tau\widetilde{y}$ as required. □

Before enlightening the relation between global shadowing and Nielsen equivalence, we reformulate Proposition 1.22 in terms of covering translations. Recall that we identified the fundamental group $\pi_1(M)$ with the group of covering translations $\mathrm{Homeo}_p(\widetilde{M})$ with the following procedure. Given a loop γ in M, we chose a base point y of γ and a lift $\widetilde{y} \in \widetilde{M}$ of y. The loop γ lifts uniquely to a path $\widetilde{\gamma}$ in $\widetilde{M}$ initiating from $\widetilde{y}$. Define $\Phi\colon \pi_1(M) \to \mathrm{Homeo}_p(\widetilde{M})$ to be the canonical isomorphism that sends $[\gamma]$ to the covering translation τ with $\widetilde{\gamma}(1) = \tau\widetilde{y}$.

We fix $x \in M$ and we identify $\pi_1(M)$ with $\pi_1(M;x)$. Let $\widetilde{x} \in \widetilde{M}$ be a lift of x. Since $\widetilde{M}$ is simply-connected, all the paths from $\widetilde{x}$ to $\widetilde{\vartheta}(\widetilde{x})$ in $\widetilde{M}$ are homotopic. We choose one of them and project it to a path δ from x to $\vartheta(x)$ in M. Consider the action ψ of ϑ on $\pi_1(M;x)$ given by $\psi([\gamma]) := [\delta^{-1} * \vartheta(\gamma) * \delta]$. The action Ψ of ϑ on $\mathrm{Homeo}_p(\widetilde{M})$ defined by $\Psi(\tau) = \widetilde{\vartheta}^{-1}\tau\widetilde{\vartheta}$ is compatible with ψ in the sense that it makes the following diagram commute.

$$\begin{tikzcd}
\pi_1(M) \arrow[r, "\Phi"] \arrow[d, "\psi"'] & \mathrm{Homeo}_p(\widetilde{M}) \arrow[d, "\Psi"] \\
\pi_1(M) \arrow[r, "\Phi"'] & \mathrm{Homeo}_p(\widetilde{M})
\end{tikzcd}$$

In the world of covering translations, Proposition 1.22 reads:

Proposition 2.6. *If a covering translation $\tau \in Homeo_p(\widetilde{M})$ commutes with $\widetilde{\vartheta}$, i.e. if $\tau = \Psi(\tau)$, then τ is the identity.*

We are now ready to prove the equivalence between global shadowing and Nielsen equivalence for periodic orbits of pA homeomorphisms. It was originally stated and proved by Handel in [14, Lemma 1.7].

Proposition 2.7. *Let $n \geq 1$ be an integer. Let x be a fixed point of ϑ^n and y be a fixed point of g^n. Then*

$$(x, \vartheta^n) \overset{\mathrm{NE}}{\sim} (y, g^n) \Leftrightarrow (x, \vartheta) \sim (y, g).$$

Proof. The direct implication follows immediately from the equivariance of the distance $\widetilde{d}$. This direction does not require a pA assumption. It is also a straightforward consequence of the reformulations of both notions in terms of arcs in M.

To prove the return implication, we start by fixing two lifts $\widetilde{x}$ of x and $\widetilde{y}$ of y. There exist two covering translations τ and σ such that $\widetilde{\vartheta}^n(\widetilde{x}) = \tau\widetilde{x}$ and $\widetilde{g}^n(\widetilde{y}) = \sigma\widetilde{y}$. Let $\delta > 0$ be a positive number for which $(x, \vartheta) \sim^{\delta} (y, g)$. For any $k \in \mathbb{Z}$, we have by definition

$$\widetilde{d}\left(\left(\sigma^{-1}\widetilde{\vartheta}^n\right)^k(\widetilde{x}), \widetilde{y}\right) = \widetilde{d}\left(\left(\sigma^{-1}\widetilde{\vartheta}^n\right)^k(\widetilde{x}), \left(\sigma^{-1}\widetilde{g}^n\right)^k(\widetilde{y})\right) \leq \delta.$$

As a compact subspace of $\widetilde{M}$ only contains finitely many lifts of $\widetilde{x}$, there is an integer $m \neq 0$ such that $(\sigma^{-1}\widetilde{\vartheta}^n)^m(\widetilde{x}) = \widetilde{x}$. Thus

$$\begin{aligned}
\left(\sigma^{-1}\widetilde{\vartheta}^n\right)^m(\sigma^{-1}\tau)(\widetilde{x}) &= \left(\sigma^{-1}\widetilde{\vartheta}^n\right)^{m+1}(\widetilde{x}) \\
&= \left(\sigma^{-1}\widetilde{\vartheta}^n\right)(\widetilde{x}) \\
&= (\sigma^{-1}\tau)(\widetilde{x}) \\
&= (\sigma^{-1}\tau)\left(\sigma^{-1}\widetilde{\vartheta}^n\right)^m(\widetilde{x}).
\end{aligned}$$

Since $(\sigma^{-1}\widetilde{\vartheta}^n)^m$ is a lift of ϑ^{mn}, Proposition 2.6 tells us that necessarily $\sigma^{-1}\tau$ has to be the identity and hence $\sigma = \tau$. The desired conclusion follows. $\square$

Proposition 2.7 together with Lemma 2.3 immediately implies the following corollary. Again we emphasize that if M has boundary, then the conclusion of the following corollary must be modified as in Lemma 2.3.

Corollary 2.8. *Let $n \geq 1$ be an integer. Let x and y be two fixed points of ϑ^n. Then*

$$(x, \vartheta^n) \overset{\text{NE}}{\sim} (y, \vartheta^n) \Rightarrow x = y.$$

2.2 Unremovability for pseudo-Anosov maps

The dynamics of pA homeomorphisms often result in minimal properties among its isotopy class. For instance, a pA homeomorphism has the least number of interior periodic orbits for every period in its isotopy class. The way this minimal characteristic is proved provides an argument for the existence of a closed subset $Y \subset M$ such that ϑ is semi-conjugate to the restricted map $g\restriction_Y$.

The following proposition was stated and proved by Handel in [14, Lemma 2.1]. Farb and Margalit give a detailed reformulation of Handel's work in [8, Theorem 14.20]. Their proof of the existence statement relies on the Lefschetz-Hopf Fixed Point Theorem (Theorem 1.20) applied to the sphere S^2 constructed from two copies of $\mathbb{H}^2 \cup \partial\,\mathbb{H}^2$ glued on the boundary. Strictly speaking, this reasoning only applies to closed surfaces whose universal cover be can identified with $\mathbb{H}^2$ (i.e. when the genus is greater than 1). Our proof of the existence statement follows a general argument by Jiang [17, Theorem 4.8].

Proposition 2.9. *Let $n \geq 1$ be an integer. Let $x \in M$ be a ϑ-periodic point of period n. There exists a g-periodic point $y \in M$ with period n such that*

$$(x, \vartheta^n) \overset{\text{NE}}{\sim} (y, g^n).$$

Proof. We first prove the existence of a fixed point y of g^n with $(x, \vartheta^n) \overset{\text{NE}}{\sim} (y, g^n)$. The assertion about the minimal period of y will be treated subsequently.

Recall that for any $n \geq 1$, the homeomorphism ϑ^n is the pA representative of its class. Therefore, without loss of generality, it is sufficient to prove the existence of a fixed point of g^n in the case $n = 1$.

Let x denote a fixed point of ϑ and $\widetilde{x}$ be a lift of x. Let $\widetilde{\vartheta}$ denote the unique lift of ϑ with $\widetilde{\vartheta}(\widetilde{x}) = \widetilde{x}$. Let $\widetilde{g}$ be the unique lift of g equivariantly homotopic to $\widetilde{\vartheta}$. It is sufficient to prove the existence of a fixed point $\widetilde{y}$ for $\widetilde{g}$. Indeed, any fixed point $\widetilde{y}$ of $\widetilde{g}$ descends to fixed point $y = p(\widetilde{y})$ of g in M. By construction, it follows immediately that $(x, \vartheta) \overset{\text{NE}}{\sim} (y, g)$.

Choose any homotopy $h_t \colon M \to M$ from ϑ to g. Let $\widetilde{h}_t \colon \widetilde{M} \to \widetilde{M}$ be the lifted homotopy from $\widetilde{\vartheta}$ to $\widetilde{g}$. We write $\mathbb{F}_\vartheta := p(\mathrm{Fix}(\widetilde{\vartheta})) \subset \mathrm{Fix}(\vartheta)$. The set $\mathbb{F}_g$ is defined similarly. Corollary 2.8 implies that $\mathbb{F}_\vartheta = \{x\}$.

Consider the *fat homotopy* associated to h_t defined by

$$\begin{aligned} H_\bullet \colon M \times I &\to M \times I \\ (z, t) &\mapsto H_t(z) := (h_t(z), t). \end{aligned}$$

The advantage of $H_\bullet$ is that it has the nature of a dynamical system on the compact space $M \times I$. Lift it to the fat homotopy $\widetilde{H}_\bullet \colon \widetilde{M} \times I \to \widetilde{M} \times I$ with $\widetilde{H}_0 := (\widetilde{\vartheta}, 0)$. In particular $\widetilde{H}_1 = (\widetilde{g}, 1)$. Analogically we introduce $\mathbb{F} :=$

$(p \times id)(\mathrm{Fix}(\widetilde{H}_\bullet))$. Note that $\mathbb{F} \subset M \times I$ is a compact subset. Both $\mathbb{F}_\vartheta \times \{0\} = (M \times \{0\}) \cap \mathbb{F}$ and $\mathbb{F}_g \times \{1\} = (M \times \{1\}) \cap \mathbb{F}$ are closed subsets of $\mathbb{F}$.

The next step consists in proving that $\mathbb{F}$ is a continuum of fixed points of $H_\bullet$ in $M \times I$ that intersects every level. In particular, it will imply that the level $(M \times \{1\}) \cap \mathbb{F} = \mathbb{F}_g \times \{1\}$ is nonempty and hence we conclude that $\mathrm{Fix}(\widetilde{g})$ is nonempty as required.

Assume *ab absurdo* that there exists a closed connected component S of $\mathbb{F}$ that contains $\mathbb{F}_\vartheta \times \{0\}$ but not $\mathbb{F}_g \times \{1\}$. Consider the t-level inclusion $\iota_t \colon M \to M \times I$ that maps z to $\iota_t(z) := (z, t)$. We define the level sets of S as $S_t := \iota_t^{-1}(S)$. By construction, S is an isolated set of fixed points of $H_\bullet$. Let $S \subset W \subset M \times [0,1)$ be an open neighbourhood of S with $S = W \cap \mathrm{Fix}\, H_\bullet$. Let $W_t := \iota_t^{-1}(W)$ be the t-slice of W. We emphasize that W is chosen such that $W_1 = \emptyset$. See Figure 2.1.

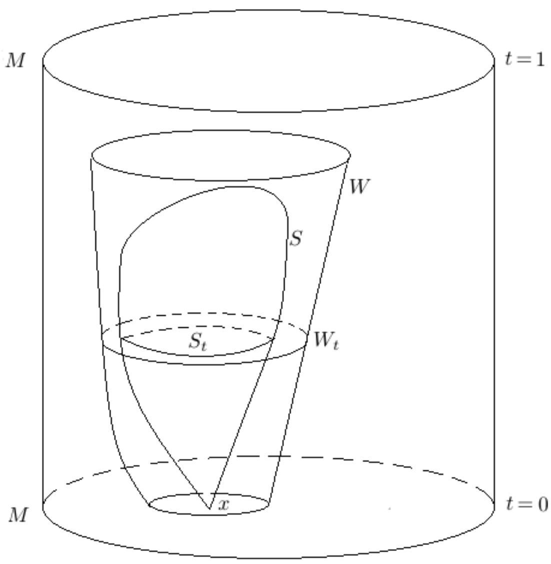

Figure 2.1: The suspension manifold $M \times I$ and the subspaces $x \in S \subset W \subset M \times [0,1)$.

The desired contradiction follows from a careful study of the indices of the fixed points along the fat homotopy H_t. We claim that the quantity $\mathrm{ind}(H_\bullet\restriction_{W_t})$ introduced in Definition 1.9 is independent of $t \in I$. Indeed, we have

$$\mathrm{ind}(H_\bullet\restriction_{W_t}) = \mathrm{ind}\left((H_\bullet \circ \iota_t)\restriction_{\iota_t^{-1}(W)}\right) = \mathrm{ind}\left((\iota_t \circ H_\bullet)\restriction_W\right).$$

All the maps $\iota_t \circ H_\bullet$ are homotopic with $\mathrm{Fix}((\iota_t \circ H_\bullet)\restriction_W) = S \cap (M \times \{t\}) \subset M \times I$ being compact. By homotopy invariance of the fixed point index, the quantity $\mathrm{ind}\left((\iota_t \circ H_\bullet)\restriction_W\right)$ does not depend on $t \in I$. The claim follows.

At the level 0, $\mathrm{ind}(H_\bullet\restriction_{W_0}) = \mathrm{ind}(x, \vartheta)$ and hence is non-zero by Proposition 1.21. At the level 1, $\mathrm{ind}(H_\bullet\restriction_{W_1}) = \mathrm{ind}(g\restriction_{W_1}) = 0$ because $W_1 = \emptyset$. This is a contradiction. We conclude that $\mathbb{F}$ is indeed a continuum of fixed points that has nonempty intersection at every level. In particular $\widetilde{g}$ has a fixed point and the existence statement follows.

It remains to prove the assertion about the minimal period of y. Let $x \in M$ be a ϑ-periodic point of period n. Suppose that $y = g^n(y)$ satisfies $(x, f^n) \overset{\text{NE}}{\sim} (y, g^n)$. Assume *ab absurdo* that there is $m < n$ with $g^m(y) = y$. Note that m must divide n. Let $\widetilde{y}$ be a lift of y. There is a unique lift $\widetilde{g}$ of g such that $\widetilde{g}^n(\widetilde{y}) = \widetilde{y}$. Let τ be the covering translation such that $\tau\widetilde{g}^m(\widetilde{y}) = \widetilde{y}$. Then $(\tau\widetilde{g}^m)^{n/m}$ is another lift of g^n that fixes $\widetilde{y}$. Therefore $\widetilde{g}^n = (\tau\widetilde{g}^m)^{n/m}$.

Let $\widetilde{\vartheta}$ be the unique lift of ϑ with $\widetilde{\vartheta}^n(\widetilde{x}) = \widetilde{x}$ for some fixed lift $\widetilde{x}$ of x. The maps $\widetilde{\vartheta}^n$ and $(\tau\widetilde{\vartheta}^m)^{n/m}$ are both lifts of ϑ^n. Hence there is a covering translation σ such that $\sigma\widetilde{\vartheta}^n = (\tau\widetilde{\vartheta}^m)^{n/m}$. By construction, $(\tau\widetilde{\vartheta}^m)^{n/m}$ is homotopic to $\widetilde{\vartheta}^n$:

$$(\tau\widetilde{\vartheta}^m)^{n/m} \simeq (\tau\widetilde{g}^m)^{n/m} = \widetilde{g}^n \simeq \widetilde{\vartheta}^n.$$

Therefore σ must be the identity and thus $\widetilde{\vartheta}^n = (\tau\widetilde{\vartheta}^m)^{n/m}$. For any $k \geq 0$ we have

$$\widetilde{\vartheta}^n\left((\tau\widetilde{\vartheta}^m)^k(\widetilde{x})\right) = (\tau\widetilde{\vartheta}^m)^{k+n/m}(\widetilde{x}) = (\tau\widetilde{\vartheta}^m)^k(\widetilde{\vartheta}^n(\widetilde{x})) = (\tau\widetilde{\vartheta}^m)^k(\widetilde{x}).$$

Corollary 2.8 implies that $\widetilde{x}$ and $(\tau\widetilde{\vartheta}^m)(\widetilde{x})$ project to the same element in M. In other words we have $\vartheta^m(x) = x$. This is a contradiction as x has least period $n > m$. We conclude that y also has least period n as required. □

The following reformulation of Proposition 2.9 in terms of global shadowing extends to non-periodic points.

Corollary 2.10. *For every $x \in M$, there exists $y \in M$ with*

$$(x, \vartheta) \sim (y, g).$$

Moreover, if x is ϑ-periodic with period n, then y can be chosen to be g-periodic with period n.

Proof. If x is a periodic point of ϑ, then Proposition 2.9 applies and the desired conclusion follows from Lemma 2.7.

Otherwise, Proposition 1.19 tells us that the density of periodic points for pA homeomorphisms implies that x is a limit of ϑ-periodic points x_n. By the above, for every $n \geq 1$, we can find $y_n \in M$ such that $(x_n, \vartheta) \sim (y_n, g)$. By the compactness of M we can assume (up to passing to a subsequence) that the sequence y_n converges to a point $y \in M$. The desired conclusion follows from Lemma 2.4. □

The association between the orbits of ϑ and g described in Corollary 2.10 has various consequences. The next proposition corresponds to Theorem 14.20 in [8].

Corollary 2.11. *For every $n \geq 1$, the pA homeomorphism ϑ has the minimal number of period-n orbits among the maps in its isotopy class $[\vartheta] \in \mathrm{MCG}(M)$.*

Proof. Let $n \geq 1$. For every ϑ-periodic point $x \in M$ of period n, Corollary 2.10 asserts the existence of a g-periodic point $y \in M$ with period n such that $(x, \vartheta) \sim (y, g)$. This association is injective in the sense that if $(x_1, \vartheta) \sim (z, g)$ and $(x_2, \vartheta) \sim (z, g)$, then $x_1 = x_2$ by Lemma 2.3. We emphasize that without extra precaution, Lemma 2.3 requires M to be a closed surface. We conclude that g has at least as many periodic orbits of period n than ϑ. □

By taking a closer look at the proof of Proposition 2.9, we observe that we actually obtained a stronger conclusion than just the Nielsen equivalence. The periodic point y constructed is connected to x by the continuum of fixed points $\mathbb{F}$. Therefore there is arc γ_0 that connects x to y such that $\gamma_0(t)$ is a fixed point of h_t^n for every t where h_t is the chosen isotopy from ϑ to g. By construction $(x, \vartheta^n) \overset{\text{NE}}{\sim} (\gamma_0(t), h_t^n)$. Hence using the part of the proof of Proposition 2.9 about the minimal period, we conclude that $\gamma_0(t)$ has minimal period n under h_t. This is true for every t.

Recall that by Proposition 2.5, $(x, \vartheta^n) \overset{\text{NE}}{\sim} (y, g^n)$ is equivalent to the existence of an arc γ connecting x to y such that $(h_t^n \circ \gamma)(t)$ is homotopic to $\gamma(t)$. This condition is naturally verified for the arc γ_0 constructed above. This underlying stronger relation is similar to what Boyland called *strong Nielsen equivalence* in [4]. This phenomenon is more commonly called unremovability of periodic orbits (see for instance [24, Definition 5.8]) or isotopy stability.

Definition 2.3 (unremovability)**.** Let $f_1\colon M \to M$ and $f_2\colon M \to M$ be two isotopic homeomorphisms with respect to the isotopy $h_t\colon M \to M$. Let $n \geq 1$ be an integer. An n-periodic point x_1 of f_1 is *connected* to an n-periodic point x_2 of f_2 by the isotopy h_t, if there is an arc $\gamma\colon I \to M$ connecting x_1 to x_2 and an isotopy $h'_t\colon M \to M$ that is a deformation of h_t such that $\gamma(t)$ is an n-periodic point of h'_t for every $t \in I$. Recall that h'_t being a deformation of h_t means that the corresponding arcs in $\mathrm{Homeo}(M)$ are homotopic with fixed endpoints.

An n-periodic point x of a homeomorphism $f\colon M \to M$ is *unremovable* (or *isotopy stable*) if for any homeomorphism $f'\colon M \to M$ isotopic to f and any isotopy from f to f', x is connected to an n-periodic point of f'.

The standard scheme to prove unremovability has two requirements. First of all, some iterate of f must have non-zero fixed point index in the sense of Definition 1.9. Boyland called such an f *essential.* The existence part in the proof of Proposition 2.9 shows that the same iterate of f' has a fixed point. The second requirement is called *uncollapsibility.* It ensures that the corresponding periodic point of f' has the same minimal period.

Definition 2.4 (collapsibility)**.** A periodic point x of a homeomorphism f of M is said to *collapse* to a periodic point y of f if

1. $n := \sharp o(x, f) > \sharp o(y, f)$ and n is a multiple of $\sharp o(y, f)$,
2. $(x, f^n) \overset{\text{NE}}{\sim} (y, f^n)$.

For instance, the second part of the proof of Proposition 2.9 shows that the periodic points of a pA homeomorphism are uncollapsible. To enlighten the stronger conclusion that results from the proof of Proposition 2.9, we reformulate it into the following theorem.

Theorem 2.12 (Unremovability)**.** *Let φ be a canonical representative in a class of homeomorphisms in* $\mathrm{MCG}(M)$*. The periodic points of φ in the interior of any pseudo-Anosov component are unremovable.*

Furthermore, if f denotes a homoemorphism of M isotopic to φ and N is a pseudo-Anosov component of φ, then different periodic points in $\mathrm{int}(N)$ *are connected to different periodic points of f.*

The proof of the second statement in Theorem 2.12 relies on an argument similar to the proof of Corollary 2.11.

Remark (unremovability for surfaces with boundary)**.** Recall that our standing assumptions specify that we are working with a closed surface M. The proof of Proposition 2.9 strongly relies on the fact that the index of a fixed point of ϑ is non-zero. This might not be the case for a fixed point on a boundary component. See Remark after Proposition 1.21. However, the statement of Theorem 2.12 specifies that periodic points are chosen in the interior of a pA component and therefore holds for surfaces with boundary.

We can use the conclusion of Corollary 2.10 to build a semi-conjugacy between ϑ and some restriction of g. The next theorem is the main conclusion in Handel's first paper on the topic [14, Theorem 2].

Theorem 2.13 (Handel)**.** *Let ϑ be a pA homeomorphism of a closed surface M. Let g be a homeomorphism of M homotopic to ϑ. Then there exists a closed g-invariant subset $Y \subset M$ such that ϑ is semi-conjugate to $g\restriction_Y$ via a surjective map $\eta\colon Y \twoheadrightarrow M$ homotopic to the inclusion $\iota\colon Y \hookrightarrow M$.*

Proof. We define Y to be

$$Y := \{y \in M : \exists x \in M \text{ with } (x, \vartheta) \sim (y, g)\}.$$

It has been seen in the proof of Corollary 2.11 that for every $y \in Y$ there is exactly one $x \in M$ such that $(x, \vartheta) \sim (y, g)$. Thus we define $\eta\colon Y \to M$ as follows. We set $\eta(y_0) = x_0$ if and only if $(x_0, f) \sim (y_0, g)$. Corollary 2.10 gives the surjectivity of η. Lemma 2.4 proves simultaneously the closeness of Y in M and the continuity of η. It follows from the definition of global shadowing that

$$(x, \vartheta) \sim (y, g) \Rightarrow (\vartheta(x), \vartheta) \sim (g(y), g).$$

Hence $g(Y) \subset Y$ and $\vartheta \circ \eta = \eta \circ g\restriction_Y$.

To see that η is homotopic to the inclusion ι, we lift η to a map $\widetilde{\eta}\colon \widetilde{Y} \to \widetilde{M}$ between the universal covers of Y and M. Given $\widetilde{y} \in \widetilde{Y}$ and $y := p(\widetilde{y})$, then by construction we have $(\eta(y), \vartheta) \sim (y, g)$. In particular, there is $\delta = \delta(\widetilde{y}) > 0$ such that $\widetilde{d}(\widetilde{y}, \widetilde{\eta}(\widetilde{y})) < \delta$. By Lemma 2.4, the same δ actually works for every $\widetilde{y}$. Therefore $\widetilde{\eta}$ stays at bounded distance from the inclusion $\widetilde{\iota}\colon \widetilde{Y} \hookrightarrow \widetilde{M}$. As required, since $\iota \circ p = p \circ \widetilde{\iota}$, we conclude that η is homotopic to ι. $\square$

When working with non-closed surfaces, one can adapt the conclusion of Theorem 2.13 by replacing M with its interior $\operatorname{int}(M)$ to get a valid statement as observed by Handel [14, Remark 2.4].

The existence of a semi-conjugacy between ϑ and some restriction of g has the following consequence at the level of topological entropy. It can be regarded as a coherent minimal property of pA homeomorphisms.

Corollary 2.14. *A pA homeomorphism minimizes the topological entropy among the homeomorphisms in its isotopy class.*

Proof. Theorem 2.13 provides a closed g-invariant set Y such that ϑ is semi-conjugate to $g\restriction_Y$. It is a standard result in dynamical systems that the existence of a semi-conjugacy implies

$$h_{top}(\vartheta) \leq h_{top}(g\restriction_Y).$$

Moreover, the restriction of a dynamical system to a closed invariant subspace always has a lower topological entropy than the original system. Hence we conclude that

$$h_{top}(\vartheta) \leq h_{top}(g\restriction_Y) \leq h_{top}(g).$$

Since in our standing assumptions ϑ is any pA homeomorphism of M and g any element in the isotopy class of ϑ, the desired conclusion follows. $\square$

A natural question at this stage is the following. Under what conditions on g do we have $Y = M$ in Theorem 2.13 ? Or in other words, when does there exist a g-periodic orbit that is not shadowed by any ϑ-periodic orbit ? This is in some sense a converse to Corollary 2.10. Handel answered this question in a second paper published in 1987 [15].

2.3 Entropy as a sufficient condition for conjugacy

The sufficient condition described by Handel in [15] involves topological entropy. Under the assumption that $h_{top}(g) = h_{top}(\vartheta)$, Handel proved that one has $Y = M$. In this section, we outline Handel's original arguments.

The precise statement is the following: if the topological entropy of g reaches the lower bound guaranteed by Corollary 2.14, namely the topological entropy of ϑ, then ϑ is semi-conjugate to g. A convenient quantity, easily relatable to the entropy, is the number of Nielsen equivalent classes. For every $n \geq 1$ and any homeomorphism f of M, let $\mathrm{pnt}(f, n)$ denote the number of distinct Nielsen equivalence classes represented by an element of $\mathrm{Fix}(f^n)$ and $\mathrm{pnt}^\infty(f)$ denote the exponential growth rate of $\mathrm{pnt}(f, n)$:

$$\mathrm{pnt}^\infty(f) := \limsup_{n\to+\infty} \frac{\max\{0, \log \mathrm{pnt}(f, n)\}}{n}.$$

The exponential growth of $\mathrm{pnt}(f, n)$ is related to the topological entropy of f as presented in the following lemma. Its proof is transcribed from [17, Theorem 2.7].

Lemma 2.15. *Given a homeomorphism $f\colon M \to M$, then $h_{top}(f) \geq \mathrm{pnt}^\infty(f)$.*

Proof. For an integer $n \geq 1$, let $N_n(f) \subset \mathrm{Fix}(f^n)$ be a set of representatives of Nielsen equivalence classes with $\sharp N_n(f) = \mathrm{pnt}(f, n)$. We prove that for any $n \geq 1$, $N_n(f)$ is an (n, δ)-separated set for some $\delta > 0$. The desired conclusion follows directly from this property.

Let $\varepsilon > 0$ be such that every loop in M of diameter less than 2ε is contractible. By compactness of M, let $0 < \delta < \varepsilon$ such that if $d(x, y) < 2\delta$, then $d(f(x), f(y)) < \varepsilon$. Let $n \geq 1$ be an integer. Suppose *ab absurdo* that there exist two distinct points x and y in $N_n(f)$ such that $d(f^i(x), f^i(y)) \leq \delta$ for every $i = 0, \dots n-1$. Since x and y are fixed points of f^n, the previous inequality holds for any integer i. It follows from the definition of the distance d that there is a path γ_i from $f^i(x)$ to $f^i(y)$ with diameter less than 2δ for $i = 0, \dots, n-1$. Set $\gamma_n := \gamma_0$. Observe that by construction $f \circ \gamma_i$ has diameter less than 2ε.

Since we assumed $\delta < \varepsilon$, the two arcs $f \circ \gamma_i$ and γ_{i+1} are homotopic with fixed endpoints. Therefore

$$f^n \circ \gamma_0 \simeq \gamma_n = \gamma_0.$$

Using Proposition 2.5, we deduce that x and y lie in the same Nielsen equivalence class of fixed points of f^n. This is a contradiction. $\square$

Remark (the pA case)**.** In the context of pA homeomorphisms, Lemma 2.15 is sharp in the sense that if $f = \vartheta$, then $h_{top}(\vartheta) = \log(\lambda) = \mathrm{pnt}^\infty(\vartheta)$ [10]. Moreover, if ϑ is pseudo-Anosov relative to a finite set $X \subset M$, then $h_{top}(\vartheta) = \mathrm{pnt}^\infty(\vartheta \text{ rel } X)$ where $\mathrm{pnt}^\infty(\vartheta \text{ rel } X)$ is the exponential growth rate of the number of Nielsen classes in the punctured surface $M \setminus X$ [4, Theorem 7.2].

Recall that our objective is to explain how we get a strict inequality at the entropy level if we assume that ϑ is not semi-conjugate to g or in other words, under the assumption that there is a g-periodic orbit which is not globally shadowed by any ϑ-periodic orbit. The strategy is to find a real number $\varepsilon > 0$ such that $\mathrm{pnt}(g, n) > c(\lambda + \varepsilon)^n$ for some positive constant c and for any sufficiently large integer n. Lemma 2.15 would immediately imply that $h_{top}(g) > h_{top}(\vartheta)$ as required.

The main axis of Handel's proof aims to relate the number of Nielsen classes with the geometry of the foliations in $\widetilde{M}$. Another way to think of $\mathrm{pnt}(g, n)$ is to count the number of covering translations τ for which the maps $\tau^{-1}\widetilde{g}^n$ are not conjugate and have a nonempty set of fixed points. By definition two fixed points of a pair of distinct such maps lie in different Nielsen classes. To outline the procedure suggested by Handel to find sufficiently many covering translations with the above property, we start with some terminology.

Let $K > 0$ be the positive number provided by Lemma 2.1 that serves as an upper bound for

$$\sup_{\widetilde{x} \in \widetilde{M}} \left\{ \widetilde{d}\left(\widetilde{\vartheta}(\widetilde{x}), \widetilde{g}(\widetilde{x})\right), \widetilde{d}\left(\widetilde{\vartheta}^{-1}(\widetilde{x}), \widetilde{g}^{-1}(\widetilde{x})\right) \right\} \leq K.$$

For a point $\widetilde{x} \in \widetilde{M}$, we write $\omega_s(\widetilde{x})$ and $\omega_u(\widetilde{x})$ for the stable leaf, respectively the unstable leaf, through $\widetilde{x}$. If $\widetilde{z} \in \widetilde{M}$ denotes a singularity of the foliations, then the connected components of $\widetilde{M} \setminus \omega_s(\widetilde{z})$ are called the *unstable wedges* based at $\widetilde{z}$. The *stable wedges* based at $\widetilde{z}$ are analogously the connected components of $\widetilde{M} \setminus \omega_u(\widetilde{z})$. If W^u denotes an unstable wedge based at $\widetilde{z}$, then the connected component of

$$E^u := \{\widetilde{x} \in W^u : \widetilde{d}_u(\widetilde{x}, \omega_s(\widetilde{z})) > K\}$$

that intersects $\omega_u(\widetilde{z})$ is called the *unstable envelope* based at $\widetilde{z}$ and is denoted by E^u_0. Note that E^u intersects $\omega_u(\widetilde{z})$ because the branches $\omega_u(\widetilde{z})$ emanating from $\widetilde{z}$ are half-infinite. Moreover, at most one connected component of E^u intersects $\omega_u(\widetilde{z})$ as one can follow subarcs of $\omega_u(\widetilde{z})$ to connect two points in E^u. The *stable envelope* based at $\widetilde{z}$ is defined analogically (see Figure 2.2).

For convenience, we are going to make a few extra assumptions about ϑ justified by the following property:

Lemma 2.16. *Let $n \geq 1$ be an integer. Let x and y be two points in M. Then*

$$(x, \vartheta) \sim (y, g) \Leftrightarrow (x, \vartheta^n) \sim (y, g^n).$$

Proof. The direct implication follows directly from the definition of global shadowing. To prove the return, assume that $(x, \vartheta^n) \sim (y, g^n)$. Applying the triangle inequality, we obtain

$$\begin{aligned}\widetilde{d}(\widetilde{\vartheta}^{m+1}(\widetilde{x}), \widetilde{g}^{m+1}(\widetilde{y})) &\leq \widetilde{d}(\widetilde{\vartheta}^{m+1}(\widetilde{x}), \widetilde{\vartheta}(\widetilde{g}^m(\widetilde{y}))) + \widetilde{d}(\widetilde{\vartheta}(\widetilde{g}^m(\widetilde{y})), \widetilde{g}^{m+1}(\widetilde{y})) \\ &\leq (\lambda + \lambda^{-1}) \cdot \widetilde{d}(\widetilde{\vartheta}^m(\widetilde{x}), \widetilde{g}^m(\widetilde{y})) + K\end{aligned}$$

for any integer m and any lift $\widetilde{x}$ of x and $\widetilde{y}$ of y in $\widetilde{M}$. Hence if $(x, \vartheta^n) \sim^\delta (y, g^n)$ for $\delta > 0$, then $(x, \vartheta) \sim^\varepsilon (y, g)$ for

$$\varepsilon := (\lambda + \lambda^{-1})^n \cdot \delta + K \cdot \sum_{i=0}^{n-1} (\lambda + \lambda^{-1})^i.$$

$\square$

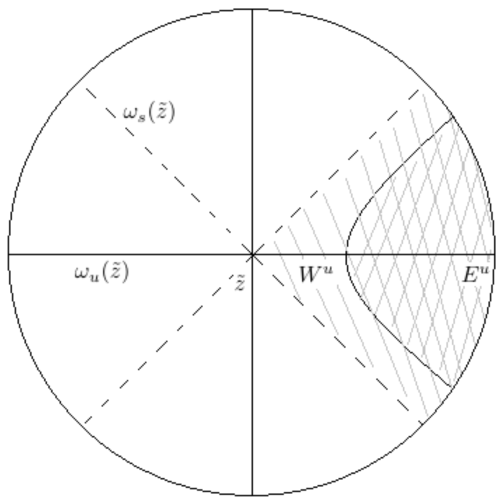

Figure 2.2: Local illustration of an unstable wedge and the corresponding unstable envelope based at a 4-pronged singularity.

Invoking Lemma 2.16, we observe that the set Y defined in Theorem 2.13 to build a semi-conjugacy between ϑ and $g\restriction_Y$ does not change if we replace ϑ and g by some positive iterates ϑ^n and g^n. Therefore, up to replacing ϑ and g by some positive iterates, we can further complete our standing assumptions without loss of generality as follows.

Standing assumptions. The following assumptions will prevail for the remainder of Section 2.3. We assume that the pA homeomorphism ϑ satisfies the following three properties:

- ϑ fixes all the singularities of the stable and unstable foliations of M,
- ϑ fixes all the half-leaves emanating from the singularities of the stable and unstable foliations,
- $\lambda > 4$, where λ denotes the stretch factor of ϑ.

Lemma 2.17. *Let $\widetilde{x}$ and $\widetilde{y}$ be two points in $\widetilde{M}$. Then*

$$\widetilde{d}_u(\widetilde{x},\widetilde{y}) > K \Rightarrow \begin{cases} \widetilde{d}_u(\widetilde{g}(\widetilde{x}),\widetilde{g}(\widetilde{y})) > d_u(\widetilde{x},\widetilde{y}) + K, \\ \widetilde{d}_u(\widetilde{\vartheta}(\widetilde{x}),\widetilde{g}(\widetilde{y})) > d_u(\widetilde{x},\widetilde{y}) + 2K, \\ \forall n \geq 1, d_u(\widetilde{g}^n(\widetilde{x}),\widetilde{g}^n(\widetilde{y})) > \lambda^n/3 \cdot d_u(\widetilde{x},\widetilde{y}), \end{cases}$$

$$\widetilde{d}_s(\widetilde{x},\widetilde{y}) > K \Rightarrow \begin{cases} \widetilde{d}_s(\widetilde{g}^{-1}(\widetilde{x}),\widetilde{g}^{-1}(\widetilde{y})) > d_s(\widetilde{x},\widetilde{y}) + K, \\ \widetilde{d}_s(\widetilde{\vartheta}^{-1}(\widetilde{x}),\widetilde{g}^{-1}(\widetilde{y})) > d_s(\widetilde{x},\widetilde{y}) + 2K, \\ \forall n \geq 1, d_s(\widetilde{g}^{-n}(\widetilde{x}),\widetilde{g}^{-n}(\widetilde{y})) > \lambda^n/3 \cdot d_s(\widetilde{x},\widetilde{y}). \end{cases}$$

Proof. For instance, if we assume $\widetilde{d}_u(\widetilde{x},\widetilde{y}) > K$, then

$$\begin{aligned} \widetilde{d}_u(\widetilde{g}(\widetilde{x}),\widetilde{g}(\widetilde{y})) &\geq \widetilde{d}_u(\widetilde{\vartheta}(\widetilde{x}),\widetilde{\vartheta}(\widetilde{y})) - \widetilde{d}_u(\widetilde{\vartheta}(\widetilde{x}),\widetilde{g}(\widetilde{x})) - \widetilde{d}_u(\widetilde{\vartheta}(\widetilde{y}),\widetilde{g}(\widetilde{y})) \\ &\geq \lambda \widetilde{d}_u(\widetilde{x},\widetilde{y}) - 2K \\ &> \widetilde{d}_u(\widetilde{x},\widetilde{y}) + K. \end{aligned}$$

Observe that the last inequality holds because (by assumption) $\lambda > 4$. □

Assume that $\widetilde{z}$ denotes a singularity of the foliations that is fixed by $\widetilde{\vartheta}$, then $\widetilde{g}(E_0^u) \subset E_0^u$ where E_0^u denotes the unstable envelope based at $\widetilde{z}$. This a consequence of Lemma 2.17. Furthermore, for every point $\widetilde{x} \in E^u$, $\widetilde{g}^n(\widetilde{x}) \in E_0^u$ for sufficiently large n. Indeed, for every integer $n \geq 0$, $\widetilde{g}^n(\widetilde{x})$ lies in the component of $M \setminus \widetilde{\vartheta}^n(\omega_s(\widetilde{z}))$ that is the image under $\widetilde{\vartheta}^n$ of the component of $M \setminus \omega_s(\widetilde{z})$ containing $\widetilde{x}$. Yet, $\widetilde{\vartheta}^n(\omega_s(\widetilde{z}))$ lies in E_0^u for sufficiently large n. What is more is the following equivalent reformulation of the assertion that ϑ is not semi-conjugate to g. Its proof being essentially a technical argument, we omit it here. Full details are provided by Handel in [15, Proposition 2.1].

Proposition 2.18. *Let y be a g-periodic point whose orbit is not globally shadowed by any f-orbit. Then there is a singularity $\widetilde{z}$ of the foliations, a stable wedge E_0^s and an unstable wedge E_0^u based at $\widetilde{z}$ such that $\widetilde{g}_z^n(\widetilde{y}) \in E_0^u$ and $\widetilde{g}_z^{-n}(\widetilde{y}) \in E_0^s$ for sufficiently large n, where $\widetilde{y}$ is a lift of y and $\widetilde{g}_z$ is the lift of g equivariantly homotopic to the lift $\widetilde{\vartheta}_z$ of ϑ that fixes $\widetilde{z}$.*

Recall that we work under the assumption that there exists a g-orbit which is not globally shadowed by any ϑ-orbit. Since no relevant constraints on $\widetilde{\vartheta}$ and $\widetilde{g}$ have been imposed, we assume from now on that $\widetilde{\vartheta} = \widetilde{\vartheta}_z$ and $\widetilde{g} = \widetilde{g}_z$.

To obtain the desired estimation on $\mathrm{pnt}(g,n)$, Handel suggested to count a certain type of distinct paths in $\widetilde{M}$. Let $T > K$ be a real number and $\beta \subset \omega_u(\widetilde{z})$ be the union of two segments with length T emanating from $\widetilde{z}$ and contained in the frontier of the stable wedge based at $\widetilde{z}$ that contains E_0^s. Let $\widetilde{a}$ and $\widetilde{b}$ be the endpoints of β. A path γ in $\widetilde{M}$ is said to be *admissible* if there is a covering translation τ such that $\tau^{-1}\gamma$ satisfies the following properties:

- $\tau^{-1}\gamma$ connects $\omega_s(\widetilde{a})$ and $\omega_s(\widetilde{b})$,
- the interior of $\tau^{-1}\gamma$ is disjoint from $\omega_s(\widetilde{a})$ and $\omega_s(\widetilde{b})$,
- $\tau^{-1}\gamma$ has stable variation less than K,
- $\tau^{-1}\gamma$ stays at $\widetilde{d}_s$-distance at most $K+1$ of β.

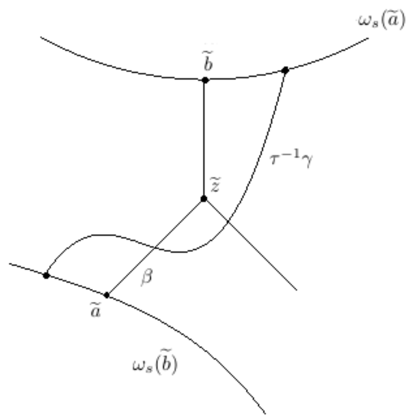

Figure 2.3: The arc $\tau^{-1}\gamma$ in the neighbourhood of β where γ is an admissible arc.

The pair (γ, τ) is called an *admissible pair* (see Figure 2.3). If α is an arc in $\widetilde{M}$, we say that the admissible pair (γ, τ) is α-unforced if $\gamma \subset \alpha$ and if there is a singularity $\widetilde{z}_0$ such that γ and the endpoints of α are contained in three different unstable envelopes based at $\widetilde{z}_0$. The maximal cardinality of a set of $\widetilde{g}^n(\alpha)$-unforced admissible pairs $\{(\gamma_i, \tau_i)\}$ with $\tau_i \neq \tau_j$ for $i \neq j$ is denoted $U_n(\alpha)$.

Proposition 2.19. *Let α be an arc in $\widetilde{M}$ such that (α, id) is an admissible pair and $n \geq 1$ an integer. If there is a constant $\varepsilon > 0$ such that $U_n(\alpha) > c_0(\lambda + \varepsilon)^n$ for some constant $c_0 > 0$, then* $\mathrm{pnt}(g, n) > c(\lambda + \varepsilon)^n$ *for some constant $c > 0$.*

Proof. Let (γ, τ) be a $\widetilde{g}^n(\alpha)$-unforced admissible pair and $\widetilde{h} := \tau^{-1}\widetilde{g}^n$. We prove that $\widetilde{h}$ has a nonempty set of fixed points and that there is an upper bound independent of n for the number of such $\widetilde{h}$ that are conjugate. The previous digression shows how this implies the desired conclusion about $\mathrm{pnt}(g, n)$.

To see that $\widetilde{h}$ has a fixed point, first recall that by construction $\gamma_1 := \tau^{-1}\gamma \subset \widetilde{h}(\alpha)$ and (γ_1, id) is an admissible pair. Inductively, we find a subarc $\gamma_k \subset \widetilde{h}^k(\alpha)$ such that (γ_k, id) is an admissible pair. In particular, there is $\widetilde{x}_k \in \gamma_k$ and a constant K_0 such that $\widetilde{d}(\widetilde{h}^k(\widetilde{x}_k), \widetilde{x}_k) < K_0$ for all $k \geq 1$. We apply Lemma 2.17 to find a potentially larger constant K_1 such that $\widetilde{d}(\widetilde{h}^i(\widetilde{x}_k), \widetilde{x}_k) < K_1$ for all $0 \leq i \leq k$ and every $k \geq 1$. We deduce that the $\widetilde{h}$ orbit of any accumulation point of the set $\{\widetilde{h}^\kappa(\widetilde{x}_k) : k \geq 1\}$, where $\kappa = \lfloor k/2 \rfloor$, lies at distance at most K_1 of α. We conclude the existence of fixed point for $\widetilde{h}$ with the help of the Brouwer Translation Theorem [27].

Let $\widetilde{x}$ be an accumulation point of $\{\widetilde{h}^\kappa(\widetilde{x}_k) : k \geq 1\}$. If $\widetilde{y}$ denotes a fixed point of $\widetilde{h}$, then $\widetilde{d}(\widetilde{h}^i(\widetilde{y}), \widetilde{h}^i(\widetilde{x}))$ is uniformly bounded for all $i \in \mathbb{Z}$. Therefore, every fixed point of $\widetilde{h}$ lies at bounded distance of α. In particular, there is at most finitely many such $\widetilde{h}$ that are conjugate and this quantity is uniformly (that is independent of n) bounded. $\square$

It remains to prove the existence of an arc α such that (α, id) is an admissible pair and for which $U_n(\alpha)$ has an exponential growth larger than λ. The construction is provided in details by Handel in [15, Lemma 3.5]. The arguments being again essentially geometric and technical considerations, we omit them in the present paper. This is precisely the last remaining step before being able to conclude that under the assumption of the existence of a g-orbit not globally shadowed by any ϑ-orbit, the expected strict inequality at the entropy level holds.

Theorem 2.20 (Handel)**.** *Let ϑ be a pA homeomorphism of a closed surface M. Let g be a homeomorphism of M homotopic to ϑ. If g has the same topological entropy as ϑ, then ϑ is semi-conjugate to g.*

Handel's Theorem is presented as Theorem 7.6 (b) in [4] by Boyland. References are given to [15] for the proof. In a remark after the statement of the theorem, Boyland says that "the analogous result with boundary (...) is more technical but is understood". No references are provided in [4]. In [3, Lecture 8], Boyland uses a boundary version of Handel's Theorem when studying dynamics of the punctured disk. Nevertheless, in [5, Theorem 3.2], Boyland investigates a generalization of Theorem 2.13 for surfaces with boundary. This is presumably the first step towards a similar generalization of Theorem 2.20. The following remark aims to add some precisions about how to handle Theorem 2.20 when M is not a closed surface.

Remark (Handel's theorem for non-closed surfaces)**.** Handel's Theorem holds in the context of marked surfaces. Or at least there is a consensus that a slight modification of the proof in the closed case gives the result for marked surfaces. The generalized statement is the following. If ϑ is pA relative to a finite ϑ-invariant set $X \subset M$ and g is homotopic to ϑ relative to X and has the same topological entropy, then ϑ is semi-conjugate to g. However, the same statement does not hold if M has nonempty boundary. The following counter-argument is due to Lee Mosher[1].

It involves the notion of rotation number for circle homeomorphisms. If $T\colon S^1 \to S^1$ denotes an orientation-preserving homeomorphism and $\widetilde{T}\colon \mathbb{R} \to \mathbb{R}$ is a lift of T to the universal cover $\mathbb{R}$ of S^1, then

$$\rho(T) := \lim_{k \to +\infty} \frac{\widetilde{T}^k(\widetilde{x}) - \widetilde{x}}{k} \pmod 1$$

is a well-defined element of $\mathbb{R} / \mathbb{Z}$ called the *rotation number* of T. For further considerations about the rotation number and circle homeomorphisms, we refer the reader to [19, Chapter 11].

Assume for simplicity that ∂M has a unique component γ. It follows from our definition of a pA transformations for surfaces with boundary that ϑ has a finite nonempty invariant set in γ. Therefore, $\vartheta \restriction_\gamma$ seen as a homeomorphism of S^1 has a rational rotation number. We can perturb ϑ by isotopy to obtain a second map ϑ' (not necessarily pseudo-Anosov any more) so that $\vartheta' \restriction_\gamma$ can be given any rotation number and without raising the entropy. Since rotation numbers are preserved under semi-conjugacy (in either direction), ϑ and ϑ' are not semi-conjugate.

[1] `https://mathoverflow.net/questions/300932/handels-theorem-for-surfaces-with-boundary`

2.4 Nielsen equivalence and finite-order

The notion of Nielsen equivalence presented in Definition 2.2 naturally applies to periodic points of a finite-order homeomorphism. Let f be such a map and $n \geq 1$ be the least integer for which f^n is the identity (n is called the *period* of n). The periodic points of f with a period less than n are called *branch periodic points.* If the period is equal to n, then it is called a *regular periodic point.* The following Nielsen equivalence relations are presented by Boyland in Lemma 1.1 of [5].

Proposition 2.21. *Let f be a finite-order homeomorphism of a surface M with period $n \geq 1$.*

1. *If x_1 and x_2 denote two regular periodic points of f, then*
$$(x_1, f^n) \overset{\text{NE}}{\sim} (x_2, f^n).$$
2. *A regular periodic point is collapsible to any branch periodic point in the sense of Definition 2.4.*
3. *If there exist two branch periodic points y_1 and y_2 of f with common period $k < n$ such that*
$$(y_1, f^k) \overset{\text{NE}}{\sim} (y_2, f^k),$$
then $\chi(M) \geq 0$.

Proof. The first two assertions are obvious. We can simply lift f^n to the identity in $\widetilde{M}$. To prove the third assertion, assume that y_1 and y_2 are two Nielsen equivalent branch periodic points. Proposition 2.5 establishes the existence of an arc γ connecting y_1 to y_2 in M such that $f^k \circ \gamma$ is homotopic to γ with fixed endpoints. Since f has finite-order, we can average any Riemannian metric ρ on M and define
$$\rho' := \sum_{i=1}^{n} (f^i)_\star \rho$$
so that f becomes an isometry of (M, ρ'). *Ab absurdo*, suppose that $\chi(M) < 0$. It follows that $\widetilde{M}$ is isometric to some totally geodesic subspace of $\mathbb{H}^2$. Lift γ to an arc $\widetilde{\gamma}$ joining two lifts $\widetilde{x}$ and $\widetilde{y}$ of x and y in $\widetilde{M}$. Fix some lift $\widetilde{f}$ of f. By uniqueness of lift, $\widetilde{f}^k \circ \widetilde{\gamma}$ is the unique arc lifting $f^k \circ \gamma$ that emanates from $\widetilde{x}$. The existence of a homotopy with fixed endpoints between γ and $f^k \circ \gamma$ implies that $\widetilde{f}^k$ fixes both $\widetilde{x}$ and $\widetilde{y}$. Proposition 1.3 thus implies that $\widetilde{f}^k$ is the identity and hence $k = n$ by minimality of n. This is a contradiction. Therefore $\chi(M) \geq 0$ as required. $\square$

Chapter 3

Dynamical forcing relations

In his survey article *Topological methods in surface dynamics* [4] published in 1994, Boyland described a general frame to study forcing relations between periodic orbits of dynamical systems. It was built in analogy with the conclusion of the long-known Sharkovskii Theorem (see Theorem 3.1).

3.1 Abstract setting

A continuous transformation of a compact and non-degenerate interval is called an *interval map*. For simplicity, we restrict thereafter interval maps to self-transformations of the unit interval $I = [0, 1]$. In a Russian article published in 1964 [26], Oleksandr Sharkovskii classified interval maps according to the set of admissible periods. This classification introduces a different total ordering of the natural integers known as the *Sharkovskii order* and defined as follows:

$$3 \rhd 5 \rhd 7 \rhd \ldots \rhd 2 \cdot 3 \rhd 2 \cdot 5 \rhd 2 \cdot 7 \rhd \ldots \rhd 2^2 \cdot 3 \rhd 2^2 \cdot 5 \rhd 2^2 \cdot 7 \rhd \ldots \rhd 2^3 \rhd 2^2 \rhd 2 \rhd 1.$$

First are listed the odd integers in increasing order, then twice the odd integers and so on going over all the powers of two. Finally, we list the pure powers of 2 in decreasing order. The smallest element in the Sharkovskii order is 1 and the maximal element is 3. Sharkovskii's Theorem is traditionally stated in the following manner in the literature.

Theorem 3.1 (Sharkovskii's Theorem)**.** *Let f be an interval map. If f has a periodic point of period $p \geq 1$, then f also has a periodic point of period q for all $q \lhd p$.*

Moreover, for every $p \in \mathbb{N} \cup \{2^\infty\}$, there is an interval map that has a periodic point of period p and no periodic point of period $q \rhd p$ (if $p = 2^\infty$, there exists an interval map that has a periodic orbit of period 2^n for every $n \geq 0$, but no periodic orbit of other period).

A detailed proof of Theorem 3.1 is given in [19, Theorem 15.3.2.]. The statement of Theorem 3.1 suggests that the Sharkovskii order on $\mathbb{N}$, defined in algebraic terms, derives from a dynamical ordering of the periodic orbits. Historically, the study of interval dynamics indicated the existence of a forcing relation between periodic orbits. For instance, in 1975, Thien-Yien Li and James A. Yorke who were studying chaos on the interval and were not aware of Sharkovskii's work, proved independently that an interval map that has a periodic orbit of length 3 also has periodic orbits of any length [21].

Inspired by the structure of the conclusion of Theorem 3.1, Boyland developed a work environment to study forcing relations between periodic orbits applicable to any family of dynamical systems. He suggested the following procedure.

Firstly, we choose a metric space X on which dynamics are worth of interest. Secondly, we fix a family of dynamical systems among which we intend to study forcing. Formally, let Maps $\subset \mathcal{C}^0(X)$ be a subset of the continuous transformations of X. In order to make valuable conclusions in higher dimensions, it is preferable not to consider all the dynamical systems on X at once in Maps.

The forcing relation is based on the comparison of selected characteristics referred commonly as the *specification* of the periodic orbits (e.g. period, topological time evolution). Mathematically, we choose a set Spec of all the possible specifications and to every $f \in$ Maps and every $x \in \mathrm{Per}(f)$ we specify the orbit of x with an element of Spec:

$$\mathrm{spec}\colon o(x, f) \mapsto \mathrm{spec}(o(x, f)) \in \mathrm{Spec}.$$

Terminology. The four objects (X, Maps, Spec, spec) constitute *Boyland's framework* for the study of forcing among periodic orbits.

Given $f \in$ Maps, we denote by $\mathrm{spec}(f)$ the set of specifications of all the periodic orbits of f:

$$\mathrm{spec}(f) := \{\mathrm{spec}(o(x, f)) : x \in \mathrm{Per}(f)\} \subset \mathrm{Spec}.$$

We define a dynamical relation on the set of specifications in the following manner.

Definition 3.1 (forcing relation)**.** Let s_1 and s_2 be two elements of Spec. We say that s_1 *dominates* s_2 and we write $s_1 \succeq s_2$ if and only if every map that exhibits the specification s_1 also exhibits s_2, i.e.

$$s_1 \succeq s_2 \iff \forall f \in \mathrm{Maps}, s_1 \in \mathrm{spec}(f) \Rightarrow s_2 \in \mathrm{spec}(f).$$

The following lemma is a straightforward consequence of Definition 3.1.

Lemma 3.2. *The relation $\succeq$ is a preorder on Spec, i.e. it is a reflexive and transitive relation:*

- *reflexivity: $\forall s \in$ Spec, $s \succeq s$,*
- *transitivity: $\forall s_1, s_2, s_3 \in$ Spec, if $s_1 \succeq s_2$ and $s_2 \succeq s_3$, then $s_1 \succeq s_3$.*

Depending on the space X, the choice of Maps $\subset \mathcal{C}^0(X)$ and the way to specify periodic orbits, the relation $\succeq$ on Spec might also be *antisymmetric*, i.e. for all $s_1, s_2 \in$ Spec, if $s_1 \succeq s_2$ and $s_2 \succeq s_1$, then $s_1 = s_2$. If the antisymmetry property is verified, then $\succeq$ becomes a *partial order* on Spec.

Example 3.1. In dimension one, if we take $X = I = [0, 1]$, Maps $= \mathcal{C}^0(I)$ and Spec $= \mathbb{N}$ with $\mathrm{spec}(o(x, f)) = \sharp o(x, f)$, then the induced dynamical order on $\mathbb{N}$ is the Sharkovskii order. This is a consequence of Theorem 3.1. Note that the Sharkovskii order on $\mathbb{N}$ is not only antisymmetric (the proof of the antisymmetry relies on the second part of the statement of Theorem 3.1), but also *connected* in the sense that any two elements are comparable.

In dimension one, we can thus fully describe the order relation on the set of specifications with the Sharkovskii order defined on the natural integers. The structure of the forcing relation therefore results from a well understood algebraic reordering of $\mathbb{N}$. As we will see thereafter, in the higher dimension case where X is a surface, little is known about the algebraic structure of the set of specifications.

In a model for which we do not expect the preorder relation on Spec to be a total order, it is of main interest to understand the structure of the genealogy sets. Given a specification $s \in \text{Spec}$, we define its *genealogy set* to be

$$\begin{aligned}\mathcal{G}(s) &:= \{\widetilde{s} \in \text{Spec} : s \succeq \widetilde{s}\} \\ &= \bigcap \{\text{spec}(f) : f \in \text{Maps}, s \in \text{spec}(f)\}.\end{aligned}$$

When studying interval dynamics it is sufficient to specify a periodic orbit by its length. In dimension two, we need to extract further characteristics in the specification; the period only is not sufficient to serve as base for an interesting forcing relation. For instance, for any positive integer k, we can build a rotation map of the genus-1 torus whose set of admissible periods is the singleton $\{k\}$. Hence we should not expect any period to force any other period in general. A more sophisticated specification associates to every periodic orbit a *braid type*.

3.2 Braids and periodic orbits

Since the 1980's people have brought the notion of braids from geometry and pure algebra to the study of surface dynamics. This section aims to present how periodic orbits and the intuitive notion of braids are related. The case of the unit disk D^2 provides a simple framework that visually highlights the underlying connection between orbits and braids. The specification of a periodic orbit by its braid type is due to Boyland. In Boyland's work, a braid type is an element of the mapping class group. In this paper, we introduce the notion of braid type in topological terms before explaining its connection with the mapping class group following the works of Matsuoka [24, Chapter 1, Section 5] and Farb and Margalit [8, Chapter 9].

3.2.1 The topological approach

Definition 3.2 (topological braid)**.** Let M denote a surface. Let $n \geq 1$ be an integer and $X_n \subset \text{int}(M)$ be a finite set of points with cardinality n. An *n-braid based at X_n* is a collection of n paths $\gamma_i \colon I \mapsto M \times I$ called *strands* such that

1. $\{\gamma_i(0) : i = 1, \dots, n\} = X_n \times \{0\}$ and $\{\gamma_i(1) : i = 1, \dots, n\} = X_n \times \{1\}$,
2. $\gamma_i(t) \in \text{int}(M) \times \{t\}$ for all $i = 1, \dots, n$ and for all $t \in I$,
3. the images $\gamma_i(I)$ are disjoint for $i = 1, \dots, n$.

Thereafter, a *braid* will also refer, somewhat imprecisely, to the union of the images $\gamma_i(I)$ seen as a disconnected subspace of $M \times I$. See Figure 3.1 for an example.

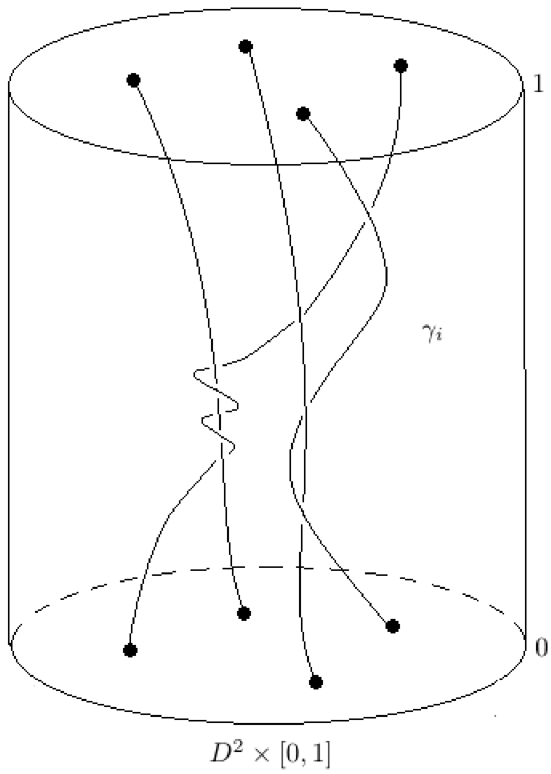

Figure 3.1: Example of a 4-braid based at four distinct points of D^2.

It is convenient to look at braids up to continuous deformations. Intuitively, a braid is defined by the overlapping of its strands. Stretching a strand for instance should not affect the nature of the braid.

Definition 3.3 (isotopy of braids)**.** Let $n \geq 1$ be an integer. We say that two n-braids based at the same finite set X_n are *isotopic* if one can be obtained from the other via a continuous deformation through a family of n-braids based at X_n.

In particular, we assume that no two strands cross each other during the deformation. A weaker notion of deformation allows the strings to cross. It is known as *homotopy* of n-braids. The set of equivalence classes of n-braids based at some finite set X_n under isotopy is denoted $B_n(M, X_n)$. The set $B_n(M, X_n)$ can be equipped with a group structure:

- the group law is the concatenation of the braids,
- the identity braid is the braid with vertical strands,
- the inverse of a braid is given by the reflection through the plane determined by $M \times \{0\}$.

Our definition of braid group is dependent upon a choice of a finite set of points in the considered surface. To get rid of this dependence we need a notion of free isotopy between n-braids that allows base points to vary.

Definition 3.4 (free isotopy of braids & braid types)**.** Two n-braids, potentially based at two different finite sets of cardinality n, are *freely isotopic* if one can be obtained from the other via a continuous deformation through a family of n-braids. The set of equivalence classes of n-braids under free isotopy is denoted $BT_n(M)$. The elements of $BT_n(M)$ are called the *braid types*.

There is an intuitive way to associate a braid type to periodic orbit. Let M denote a surface and $f: M \to M$ be a homeomorphism. Assume that $X \subset M$ is a finite invariant set for f. Since f is bijective, X is a finite union of periodic orbits of f. One way to generate a braid is to look at the time evolution of every point of X in the suspension manifold $M \times I$ where the level 0 represents the initial configuration in M and the level 1 represents the final configuration of the points in $f(M) = M$. This observation is valid whenever M transforms continuously from M to $f(M)$ or in other words, when f is isotopic to the identity. Under this assumption, it makes sense to look at the time evolution of the periodic orbits. Said differently, we switch from discrete dynamics to continuous dynamics where orbits are curves.

The time evolution of every point in X defines a strand in $M \times I$. More precisely, we chose once for all a *preferred isotopy* $h_t^f : M \to M$ from the identity to f. The union of all trajectories $\bigcup_{x \in X} \{(h_t^f(x), t) : 0 \leq t \leq 1\} \subset M \times I$ forms a braid based at X in the sense of Definition 3.2 called the *time evolution braid of* X.

Remark (dependence upon a choice of the isotopy)**.** Observe that *a priori* the braid type associated to the time evolution braid depends on the choice of the preferred isotopy h_t^f. However if h'_t is a deformation of h_t^f, then the two braids describing the respective time evolution of X according to h_t^f and h'_t are isotopic. Therefore, if the connected component $\mathrm{Homeo}_0(M)$ of the identity inside $\mathrm{Homeo}(M)$ is simply connected, then any two paths from the identity to f are homotopic and thus the free isotopy class of the time evolution braid is independent of the choice of a preferred isotopy. From Theorem 1.6 we know that this is the case among others for surfaces with negative Euler characteristic.

Definition 3.5 (braid type for periodic orbits)**.** Let $x \in M$ be a periodic point of f. The *braid type* associated to $o(x, f)$ is the braid type determined by the time evolution braid of $o(x, f)$ in $M \times I$. We denote it by $bt(o(x, f))$.

If $x \in \mathrm{Per}(f)$ is a periodic point for f with period $n \geq 1$, then $bt(o(x, f))$ is an element of $BT_n(M)$. Definition 3.5 naturally extents to any finite invariant set. The action of f on any finite invariant set $X \subset M$ can be identified with a permutation of the elements of X (given that we fixed an ordering of these points). If X is a periodic orbit, then this permutation is cyclic. In general, X is a finite union of periodic orbits and thus the corresponding permutation is the composition of the cyclic permutations associated to each periodic orbit.

For a matter of precision we introduce $CBT_n(M) \subset BT_n(M)$ the set of *cyclic n-braid types* defined to be the braid types whose representatives permute cyclically the elements of the finite set they are based on. The braid type associated to a period-n orbit of $f: M \to M$ therefore lies in $CBT_n(M)$.

3.2.2 Boyland's approach

In [4], Boyland used a slightly different approach to define the braid type of a periodic orbit. We will see that these approaches, if not always rigorously equivalent, are nevertheless similar.

If M denotes a surface and $n \geq 1$ is an integer, let

$$C_n(M) := \{X \subset M : \sharp X = n\}$$

be the collection of subsets of M with cardinality n. Another way to describe $C_n(M)$ is to look at the space

$$M^{\times n} \setminus \bigcup_{i<j} \{(x_1, \ldots, x_n) : x_i = x_j\} \subset M^{\times n}$$

up to permutations. According to this identification we equip $C_n(M) \subset M^{\times n}$ with the subspace topology. The underlying relation between $C_n(M)$ and the braid group is given by the following isomorphism.

Lemma 3.3. *Let $n \geq 1$ be an integer and $X_n \subset \mathrm{int}(M)$ be a finite collection of interior points. We have the following isomorphism of groups:*

$$B_n(M, X_n) \cong \pi_1(C_n(M); X_n).$$

Proof. Let b denote an n-braid based at X_n. The intersection of b with any slice $M \times \{t\} \subset M \times I$ defines an element of $C_n(M)$. Let $c_b \colon I \to C_n(M)$ be the loop in $C_n(M)$ defined by $c_b(t) := b \cap (M \times \{t\})$.

Conversely, a loop c in $C_n(M)$ based at X_n defines a n-braid $b_c := \{(c(t), t) : t \in I\}$. Note that the continuity of the loop c in $C_n(M)$ is equivalent to the continuity of the n induced loops in M. $\square$

Strictly speaking, since braids were defined to be based at interior points of M, the proof of Lemma 3.3 holds for closed surfaces only. However, M and $\mathrm{int}(M)$ are homotopy equivalent and therefore have the same fundamental group. Hence, the proof can be adjusted by replacing M with $\mathrm{int}(M)$.

Let us consider the case $n = 1$ in Lemma 3.3. First observe that $C_1(M) = M$. Given a loop c in M based at some point $x \in M$, we look at it as a continuous deformation of x into itself. Since M is compact, this deformation can be extended to an isotopy $h_t \colon M \to M$ of M that fixes ∂M such that $h_0 = id$ and $c(t) = h_t(x)$. This is a standard result in differential topology. The reader will find a precise statement in [8, Proposition 1.11]. If we write $\varphi_c := h_1$, then we obtain a homomorphism of groups

$$\begin{aligned} \mathrm{Push} \colon \pi_1(M; x) &\longrightarrow \mathrm{MCG}(M \text{ rel } \{x\}) \\ [c] &\longmapsto [\varphi_c]. \end{aligned}$$

It is not obvious to prove that Push is a well-defined homomorphism of groups, as there is no canonical choice for the extension of isotopy. Specifically, we have to prove that given two extended isotopies from c, their endpoints are connected in $\mathrm{Homeo}^+(M \text{ rel } \{x\}, \partial M)$. Moreover, if c_1 and c_2 are two homotopic loops in M based at x, then we would like the homotopy between c_1 and c_2 (that fixes x) to lift to a homotopy between φ_{c_1} and φ_{c_2}.

However, assuming for now that Push is well defined, we can make the following observation. Given a loop c based at x, the homeomorphism φ_c is by construction isotopic to the identity. We emphasize however that this isotopy might not fix x. If $\mathrm{MCG}(M \text{ rel } \{x\}; id)$ denotes the subgroup of $\mathrm{MCG}(M \text{ rel } \{x\})$ composed of the classes whose representatives are isotopic to the identity (without necessarily fixing x), then, assuming that Push is well defined, we get a surjective homomorphism

$$\mathrm{Push}\colon \pi_1(M; x) \twoheadrightarrow \mathrm{MCG}(M \text{ rel } \{x\}; id).$$

Indeed, given a class $[f] \in \mathrm{MCG}(M \text{ rel } \{x\}; id)$, there is an isotopy $h_t\colon M \to M$ from the identity to f. The images $h_t(x)$ define a loop in M based at x whose image under Push is $[f]$ as required.

The convenient assumption to make at this stage, so that *inter alia* Push becomes well defined, consists in restricting our attention to surfaces with negative Euler characteristic. As it was previously mentioned in Theorem 1.6, the connected component of the identity in the group of homeomorphisms of such surfaces is simply connected. For further considerations about the point-pushing map, we content ourself with the following theorem known as the *Birman exact sequence.*

Theorem 3.4 (Birman Exact Sequence). *Let M denote a compact surface up to a finite number of punctures and let $x \in M$. If $\pi_1(\mathrm{Homeo}^+(M, \partial M); id) = 0$, then the point-pushing map is well defined and the following sequence is exact:*

$$1 \longrightarrow \pi_1(M; x) \xrightarrow{Push} \mathrm{MCG}(M \text{ rel } \{x\}) \xrightarrow{Forget} \mathrm{MCG}(M) \longrightarrow 1.$$

A proof of Theorem 3.4 consists in applying the long exact sequence in homotopy to the fibre bundle

$$\mathrm{Homeo}^+(M \text{ rel } \{x\}, \partial M) \longrightarrow \mathrm{Homeo}^+(M, \partial M) \xrightarrow{ev_x} M,$$

where ev_x is the evaluation at $x \in M$. The tail of the sequence immediately gives the conclusion of Theorem 3.4. For a complete proof, the reader may consult [8, Theorem 4.6]. The Birman Exact Sequence can be generalized when we consider homeomorphisms that fix more than one point.

Theorem 3.5 (Extended Birman Exact Sequence). *Let M denote a compact surface up to a finite number of punctures. Let $n \geq 1$ be an integer and $X_n \subset \mathrm{int}(M)$ be a finite collection of n points. If $\pi_1(\mathrm{Homeo}^+(M, \partial M); id) = 0$, then the following sequence is exact:*

$$1 \longrightarrow \pi_1(C_n(M); X_n) \xrightarrow{Push} \mathrm{MCG}(M \textit{ rel } X_n) \xrightarrow{Forget} \mathrm{MCG}(M) \longrightarrow 1.$$

In analogy to the case $n = 1$, if we fix some $X_n \in C_n(M)$ for $n \geq 1$, then we can build a surjective group homomorphism by extending loops to isotopies:

$$\mathrm{Push}\colon \pi_1(C_n(M); X_n) \twoheadrightarrow \mathrm{MCG}(M \text{ rel } X_n; id).$$

It is well defined and corresponds to the homomorphism Push provided by Theorem 3.5. Lemma 3.3 states that the braid group $B_n(M, X_n)$ is isomorphic to $\pi_1(C_n(M); X_n)$. Therefore, Push induces a surjective group homomorphism

$$\Theta\colon B_n(M, X_n) \twoheadrightarrow \mathrm{MCG}(M \text{ rel } X_n; id).$$

Despite an involving construction, Θ is an intuitive transformation. The image of a braid under Θ can be optically seen by pulling a rubber surface representing M up the braid. It induces a visual deformation of M whose final step is the desired homeomorphism. Theorem 3.5 tells us that Θ is an isomorphism whenever $\text{Homeo}_0^+(M, \partial M)$ has a trivial fundamental group (or simply when $\text{MCG}(M)$ is trivial).

Remark (convention about the boundary)**.** It is of utmost importance to remember in this context that we only consider homeomorphisms and isotopies that fix the boundary pointwise in the definition of the mapping class group. The authors who do not impose any behaviour on the boundary may similarly define a surjective homomorphism

$$B_n(M, X_n) \twoheadrightarrow \text{Homeo}^+(M \text{ rel } X_n; id)/\text{Homeo}_0^+(M \text{ rel } X_n)$$

but it fails to be injective in many more cases. Obviously, if $\partial M = \emptyset$, then the two constructions are identical and injectivity is verified whenever $\chi(M) < 0$. Among the people who opted for this contrasting approach were Matsuoka [24] and Hall [11].

As we are interested in braid types, we would like Θ to descend to a transformation of $BT_n(M)$. The following procedure is suggested by several authors. In the definition of $BT_n(M)$, we looked at n-braids up to free isotopies, allowing the base set to vary. Since the notions of homotopy and isotopy of loops in $C_n(M)$ are equivalent, two n-braids are freely isotopic if and only if their corresponding loops in $C_n(M)$ are freely homotopic. Therefore $BT_n(M)$ is the space of free loops of $C_n(M)$.

Let $X_n \in C_n(M)$ be a finite subset of M. We consider the forgetful mapping

$$\text{Forget}\colon \pi_1(C_n(M); X_n) \longrightarrow \{\text{free homotopy classes of loops}\}$$

that sends a homotopy class of loops with fixed base set to its free homotopy class. A standard result in topology tells us that since $C_n(M)$ is path-connected, Forget is a surjective map.

Moreover, $\text{Forget}([c_1]) = \text{Forget}([c_2])$ if and only if $[c_1]$ and $[c_2]$ are in the same conjugacy class of $\pi_1(C_n(M); X_n)$, i.e. if γ represents the loops described by X_n along the free isotopy from c_1 to c_2, then $[c_1] = [\gamma * c_2 * \gamma^{-1}] = [\gamma] \cdot [c_2] \cdot [\gamma^{-1}]$ in $\pi_1(C_n(M); X_n)$. Therefore, Forget induces a one-to-one correspondence between the conjugacy classes of $\pi_1(C_n(M); X_n)$ and the free homotopy classes of loops in $C_n(M)$. Consequently we have the following identification:

$$B_n(M, X_n)/\text{conj} = BT_n(M).$$

Here G/conj is the notation we adopt for the set of conjugacy classes in a any abstract group G.

The surjective group homomorphism Θ described before descends to a surjective map

$$\overline{\Theta}\colon BT_n(M) = B_n(M, X_n)/\text{conj} \longrightarrow \text{MCG}(M \text{ rel } X_n; id)/\text{conj}.$$

Moreover, if Θ is injective then so is $\overline{\Theta}$. In this case, we have a one-to-one correspondence between braid types and the conjugacy classes of the mapping class group $\text{MCG}(M \text{ rel } X_n; id)$. This relation suggests the following alternative definition of the braid type for periodic orbits. It was the original approach taken by Boyland in [4] to define the specification of a periodic orbit.

Definition 3.6 (braid type for periodic orbits *bis*)**.** Let f and g be two homeomorphisms of M isotopic to the identity. Let $x \in M$ be an f-periodic point and $y \in M$ be a g-periodic point. We say that the periodic orbits $o(x,f)$ and $o(y,g)$ *have the same braid type* if there is a homeomorphism ξ of M isotopic to identity with $\xi(o(x,f)) = o(y,g)$ such that f is isotopic to $\xi^{-1} \circ g \circ \xi$ relative to $o(x,f)$. The equivalence class of $o(x,f)$ is denoted $[o(x,f)]$ and is called the *braid type* of $o(x,f)$.

If we want to compare braid types of different maps it is convenient to build a reference model. For every integer $n \geq 1$, we fix a finite set $X_n \in C_n(M)$. Given a homeomorphism f of M isotopic to the identity and $x \in \mathrm{Per}(f)$ with period n, we chose a homeomorphism $\xi = \xi(f,x)$ of M isotopic to the identity with $\xi(X_n) = o(x,f)$ and we consider the conjugacy class $[\xi^{-1} \circ f \circ \xi] \in \mathrm{MCG}(M \text{ rel } X_n; id)$.

$$\begin{tikzcd}
(M, X_n) \arrow[r, "\xi \simeq id"] \arrow[d, "\xi^{-1} \circ f \circ \xi"'] & (M, o(x,f)) \arrow[d, "f"] \\
(M, X_n) \arrow[r, "\xi"] & (M, o(x,f))
\end{tikzcd}$$

Lemma 3.6. *Let f and g be two homeomorphisms of M isotopic to the identity. Let further $x \in \mathrm{Per}(f)$ and $y \in \mathrm{Per}(g)$. Then $[o(x,f)] = [o(y,g)]$ if and only if the classes $[\xi(x,f)^{-1} \circ f \circ \xi(x,f)]$ and $[\xi(y,g)^{-1} \circ g \circ \xi(y,g)]$ are conjugate in MCG(M rel $X_n; id$).*

Proof. Assume first $[o(x,f)] = [o(y,g)]$. By definition, there is a homeomorphism ξ of M isotopic to identity with $\xi(o(x,f)) = o(y,g)$ such that f is isotopic to $\xi^{-1} \circ g \circ \xi$ relative to $o(x,f)$. Hence $\xi(x,f)^{-1} \circ f \circ \xi(x,f)$ is isotopic to $\xi_0^{-1} \circ g \circ \xi_0$ relative to X_n, where $\xi_0 := \xi \circ \xi(x,f)$. Naturally, $[\xi_0^{-1} \circ g \circ \xi_0]$ and $[\xi(y,g)^{-1} \circ g \circ \xi(y,g)]$ are conjugate in $\mathrm{MCG}(M \text{ rel } X_n; id)$. The converse is proved by similar arguments. □

Therefore, Lemma 3.6 allows one to look at a braid type $[o(x,f)]$ as a well-defined element of $\mathrm{MCG}(M \text{ rel } X_n; id)/\mathrm{conj}$. A careful study of the various constructions presented in this section shows that $\overline{\Theta}(bt(o(x,f)) = [o(x,f)]$.

In conclusion, we see that the two notions of braid type introduced in Definitions 3.5 and 3.6 are directly related by the surjective map $\overline{\Theta}$. Moreover, if M is such that $\overline{\Theta}$ is an injective map, then both definitions of braid types are equivalent. This is the case *inter alia* when M has negative Euler characteristic (Theorem 1.6) or when M is the unit disk (Lemma 1.9).

3.3 Forcing for surface dynamics

The perspective of specifying a periodic orbit of a surface homeomorphism by its braid type leads to the following framework. Let M denote a surface on which the dynamics take place. We set

$$\mathrm{Maps} := \mathrm{Homeo}^+(M, \partial M; id),$$

where again $\mathrm{Homeo}^+(M, \partial M; id)$ designates the group of orientation-preserving homeomorphisms of M, isotopic to the identity and whose restriction to ∂M is the identity. The set of specifications is chosen to be

$$\mathrm{Spec} := CBT(M) := \bigcup_{n \geq 1} CBT_n(M).$$

A period-n orbit $o(x, f)$ of $f \in \mathrm{Homeo}^+(M, \partial M; id)$ is specified by its braid type as in Definition 3.5:

$$\mathrm{spec}(o(x, f)) := bt(o(x, f)) \in CBT_n(M) \subset CBT(M).$$

The collection of braid types of all the periodic orbits of $f \in \mathrm{Homeo}^+(M, \partial M; id)$ is denoted by

$$\mathrm{spec}(f) := bt(f) := \{bt(o(x, f)) : x \in \mathrm{Per}(f)\} \subset CBT(M).$$

It follows from the definition of the pre-order $\succeq$ on $CBT(M)$ (Definition 3.1) that for any braid type $\beta \in CBT(M)$, the genealogy set of β can be written as

$$\begin{aligned} \mathcal{G}(\beta) &= \{\gamma \in CBT(M) : \beta \succeq \gamma\} \\ &= \bigcap \left\{ bt(f) : f \in \mathrm{Homeo}^+(M, \partial M; id), \beta \in bt(f) \right\}. \end{aligned}$$

The use of braid types to specify orbits is justified by two main results on the forcing. Both achievements are originally due to Boyland [4]. Firstly, one can reduce the analysis of the genealogy sets to the study of the dynamics of a single map. Secondly, the order relation on $CBT(M)$ is antisymmetric. These two results are presented in the following subsections.

3.3.1 The genealogy sets

Let $\beta \in CBT_n(M)$ be a braid type. The goal is to express the genealogy set of β as the braid types of a unique representative map φ_β. The representative map must in some sense optimize the number of periodic orbits to exhibit exactly the braid types present in every map exhibiting β. The adjective originally used by Boyland to qualify φ_β was *condensed* representative.

We know from the previous chapter that a pA homeomorphism not only minimizes the number of interior periodic orbits in its isotopy class but also satisfies some unremovability property of Nielsen equivalence classes. However, the periodic orbits of two Nielsen equivalent periodic points do not necessarily have the same braid type.

Indeed, let $\vartheta \colon M \to M$ be a pA homeomorphism of the surface M (relative to the empty set) isotopic to the identity with respect to the preferred isotopy $h_t^\vartheta \colon M \to M$ that deforms the identity into ϑ. Let g be a homeomorphism of M isotopic to ϑ and $h_t^g \colon M \to M$ be the preferred isotopy that deforms the identity into g. Let $h_t \colon M \to M$ be the isotopy from ϑ to g that makes the following diagram commute.

$$\begin{tikzcd} (M,\vartheta) \arrow[r, "h_{1-t}^\vartheta"] \arrow[dr, "h_t"'] & (M, id) \arrow[r, "h_t^\vartheta"] \arrow[d, "h_t^g"] & (M,\vartheta) \arrow[dl, "h_t"] \\ & (M, g) & \end{tikzcd}$$

Let $n \geq 1$ be an integer. Assume that $x \in M$ is an interior n-periodic point of ϑ and $y \in M$ is the n-periodic point of g provided by Proposition 2.9 such that

$$(x, \vartheta^n) \overset{\text{NE}}{\sim} (y, g^n).$$

By Proposition 2.5, there is an arc $\gamma\colon I \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$ and such that $\gamma(s)$ is homotopic with fixed endpoints to the arc $\alpha(s) := (h_s^{-n} \circ \gamma)(s)$. Let $H_t\colon I \to M$ denotes an arc homotopy with $H_0 = \gamma$ and $H_1 = \alpha$.

Let b_x be an n-braid based at $o(x, \vartheta)$ that represents the type $bt(o(x, \vartheta))$. Similarly let b_y be an n-braid based at $o(y, g)$ that represents the type $bt(o(y, g))$. Nielsen equivalence is sufficient to build a braid homotopy $\mathbb{H}_s\colon I \to M \times I$ between b_x and b_y. The first $(n-1)$ strings (i.e. for $i = 0, \ldots, n-2$) of $\mathbb{H}_s$ are defined by the time evolution of the points $((h_s)^i \circ \gamma)(s)$:

$$\zeta_i(s)\colon t \mapsto \Big((h^g_{ts} \circ h^\vartheta_{t(1-s)} \circ (h_s)^i \circ \gamma)(s), t\Big) \in M \times \{t\} \subset M \times I.$$

The string $\zeta_i(s)$ connects $((h_s)^i \circ \gamma)(s)$ in level 0 to $((h_s)^{i+1} \circ \gamma)(s)$ in level 1. The $(n-1)$th string should connect $((h_s)^{n-1} \circ \gamma)(s)$ in level 0 to $\gamma(s)$ in level 1. Observe that under no further assumption $\gamma(s)$ is generally a different point from $((h_s)^n \circ \gamma)(s)$. Therefore the definition of $\zeta_{n-1}(s)$ requires to use the homotopy H_t to adapt the trajectory and reach the correct final point:

$$\begin{aligned} \zeta_{n-1}(s)\colon t &\mapsto \Big((h^g_{ts} \circ h^\vartheta_{t(1-s)} \circ (h_s)^{n-1} \circ H_t)(s), t\Big) \in M \times \{t\} \subset M \times I \\ 1 &\mapsto ((\underbrace{h^g_s \circ h^\vartheta_{(1-s)}}_{=h_s} \circ (h_s)^{n-1} \circ H_1)(s), 1) = (\gamma(s), 1). \end{aligned}$$

With this adjustment for the last string, $\mathbb{H}_s$ is a well-defined n-braid based at $\{((h_s)^i \circ \gamma)(s) : i = 0, \ldots, n-1\}$. Let us now check that $\mathbb{H}_s$ defines a continuous deformation from b_x to b_y as required. We have

$$\begin{cases} \zeta_i(0)\colon t \to \left(h^\vartheta_t\left(\vartheta^i(x)\right), t\right), & \text{for } i = 0, \ldots, n-1, \\ \zeta_i(1)\colon t \to \left(h^g_t\left(g^i(y)\right), t\right), & \text{for } i = 0, \ldots, n-1, \end{cases}$$

and therefore $\mathbb{H}_0 = b_x$ and $\mathbb{H}_1 = b_y$ as required. See Figure 3.2. We emphasize that $\mathbb{H}_s$ is generally not an isotopy of braids since nothing prevents the points $((h_s)^i \circ \gamma)(s)$ to coincide for different values of i. Hence, Nielsen equivalence is not sufficient to have equality of the corresponding braid types.

However, we know from Theorem 2.12 that the interior periodic points of ϑ are unremovable. Therefore, we can assume that the arc γ connects x to y via a continuous collection of n-periodic points $\gamma(s)$ for h_s. In particular $((h_s)^n \circ \gamma)(s) = \gamma(s)$ for every s. Under this new assumption on γ, the strings of $\mathbb{H}_s$ given by

$$\zeta_i(s)\colon t \mapsto \Big((h^g_{ts} \circ h^\vartheta_{t(1-s)} \circ (h_s)^i \circ \gamma)(s), t\Big), \text{ for } i = 0, \ldots, n-1$$

are disjoint. This is a consequence of the minimal period of $\gamma(s)$ under h_s. Note that henceforth there is no necessary adjustment in the definition of ζ_{n-1}. We conclude that $\mathbb{H}_s$ is an isotopy of braids and hence

$$bt(o(x, \vartheta)) = bt(o(y, g)).$$

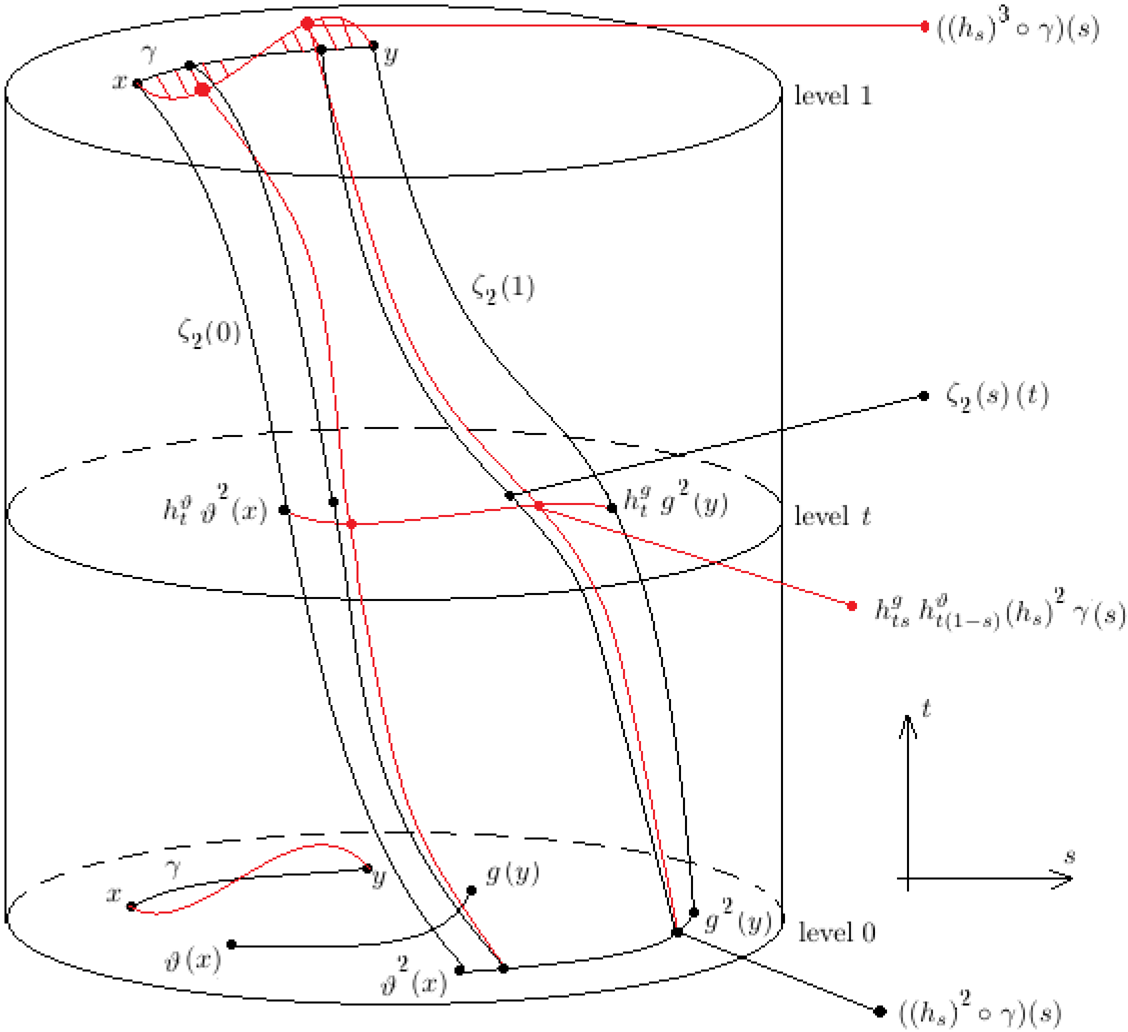

Figure 3.2: Braid homotopy generated by two period-3 Nielesen equivalent periodic points x and y on the disk. The variable t represents the vertical evolution and the parameter s the evolution inside the levels. The initial trajectory of the string ζ_2 (in red) must be adapted via the homotopy between γ and $((h_s)^2\circ\gamma)(s)$ to get the correct trajectory in black.

This means that not only the periodic points of ϑ are unremovable but also that all braid types exhibited by ϑ persists under isotopy. This is summarized in the following theorem proven by Matsuoka in [24, Theorem 5.10].

Theorem 3.7 (Isotopy stability of braid types)**.** *Let $f\colon M \to M$ be a homeomorphism of M and $X \subset \mathrm{int}(M)$ be a finite f-invariant set. Assume that the canonical representative $\varphi \in [f] \in \mathrm{MCG}(M \text{ rel } X)$ has a pA component. Let S denote the union of the interiors of all pA components of φ. Then there is an injective map $\iota\colon \mathrm{Per}(\varphi) \cap S \to \mathrm{Per}(f)$ such that for any $x \in \mathrm{Per}(\varphi) \cap S$*

1. *x and $\iota(x)$ have the same period,*
2. *$bt(o(x, \varphi)) = bt(o(\iota(x), f))$.*

Proof. If $X = \emptyset$, then $\iota(x)$ is defined to be the periodic point of f connected to x in the sense of Theorem 2.12. The condition about the braid types follows from the previous digression.

If $X \neq \emptyset$, the strategy is to blow-up M at X to reduce to the previous case. Let $\widehat{M}$ denotes the blow-up of M at X and $\pi\colon \widehat{M} \to M$ be the canonical projection. Since φ is a canonical homeomorphism in the sense of Theorem 1.16, it has the blow up $\widehat{\varphi}\colon \widehat{M} \to \widehat{M}$. This is a consequence of Proposition 1.24. Moreover, if φ is pA relative to X, then $\widehat{\varphi}$ is pA relative to the empty set.

For $x \in X$, we choose a small open disk $U_x \subset M$ centred at x. Let $U := \bigcup_{x\in X} U_x$ and $\widehat{U} := \pi^{-1}(U)$. Theorem 1.8 allows us to isotope f relative to X to a homeomorphism f' of M such that

- f' is smooth and non-singular at every $x \in X$,
- f and f' coincide away from U.

Proposition 1.1 tells us that f' has the blow-up $\widehat{f}\colon \widehat{M} \to \widehat{M}$. By construction, φ and f' are isotopic relative to X and thus $\widehat{\varphi}$ is isotopic to $\widehat{f}$ relative to the empty set. The desired conclusion mostly follows by applying the case $X = \emptyset$ to $\widehat{\varphi}$ and $\widehat{f}$. Let ι_0 be the underlying injective map between the periodic points of $\widehat{\varphi}$ and $\widehat{f}$.

Let x be an n-periodic point of φ inside S. If $x \in X$, set $\iota(x) := x$. If $x \notin X$, let $\iota(x) := \iota_0(x) =: y$. If $y \notin \widehat{U}$, then y is the desired periodic point of f. Assume *ab absurdo* that $y \in \widehat{U}$. Then $y \in \pi^{-1}(U_z)$ for some $z \in X$. Let $\gamma_z := \partial\widehat{M} \cap \pi^{-1}(U_z)$ be the boundary circle of $\widehat{M}$ corresponding to z. Up to shrinking U_z we can assume that $(y, \widehat{f})$ is $\widehat{f}$-globally shadowed by some point on γ_z. By construction, x and y are connected by the homotopy from $\widehat{\varphi}$ to $\widehat{f}$. Hence $(x, \widehat{\varphi})$ is $\widehat{\varphi}$-globally shadowed by some point on γ_z. This contradicts Lemma 2.3. $\square$

Theorem 3.7 establishes a property of isotopy stability for braid types present inside pA components. The notions of pseudo-Anosov and finite-order inherited from the Thurston-Nielsen Classification descend reasonably to braid types. The correspondence between the topological definition of a braid type and its counterpart as conjugacy class of the mapping class group suggests the following definition. Recall that given an integer $n \geq 1$ and any finite set $X_n \subset \mathrm{int}(M)$, there is a surjective group homomorphism:

$$\Theta\colon B_n(M, X_n) \twoheadrightarrow \mathrm{MCG}(M \text{ rel } X_n; id).$$

Definition 3.7 (pA, finite-order and reducible braid types)**.** An n-braid $b \in B_n(M, X_n)$ based at X_n is called *finite-order, pseudo-Anosov (pA) or reducible* if the isotopy class $\Theta(b) \in \mathrm{MCG}(M \text{ rel } X_n; id)$ is finite-order, pseudo-Anosov relative to X_n or reducible relative to X_n, respectively. A braid type is called *finite-order, pseudo-Anosov (pA) or reducible* if it has a finite-order, pseudo-Anosov (pA) or reducible representative, respectively.

Let $\beta = [b] \in BT_n(M)$ denote a braid type whose representative b is an n-braid based at X_n. We choose a canonical representative φ_β in $\Theta(b) \in \mathrm{MCG}(M \text{ rel } X_n; id)$ in the sense of the Thurston-Nielsen Classification. Observe that φ_β is uniquely determined by β up to topological conjugacy. Indeed, if $[b_1] = [b_2]$ in $BT_n(M)$, then the braid isotopy between b_1 and b_2 generates a homeomorphism h of M that sends the base set of b_1 to the base set of b_2. If φ_1 and φ_2 denote the canonical representatives in $\Theta(b_1)$ and $\Theta(b_2)$ respectively, then φ_1 is isotopic to $h \circ \varphi_2 \circ h^{-1}$ relative to the base set of b_1.

At this point, it is important to observe that without any further assumptions, it is not generally true that $\beta \in bt(\varphi_\beta)$. By construction there is an homeomorphism f of M isotopic to the identity that fixes X_n such that $\beta \in bt(f)$. The canonical homeomorphism φ_β leaves by assumption the set X_n invariant. However, depending on the choice of the preferred isotopy from the identity to φ_β, the isotopy class of the time evolution braid of X_n under φ_β is not necessarily the same as $[b]$. If $\beta \in bt(\varphi_\beta)$, then we say that *$\varphi_\beta$ represents β*.

Nevertheless, under the assumption that the braid type associated to the time evolution braid is independent on the choice of a preferred isotopy between the identity and φ_β (e.g.. when $\chi(M) < 0$), then φ_β represents β. Indeed, if we choose the preferred isotopy to be the conjugation of the preferred isotopy between the identity and f with the isotopy from f to φ_β that fixes X_n, then it follows straightforwardly that $\beta \in \varphi_\beta$.

If $\partial M = \emptyset$ and β is a pseudo-Anosov braid type represented by the canonical homeomorphism φ_β, then we can use the property of unremovability for braid types provided by Theorem 3.7 to rewrite the genealogy set of β as the set of braid types exhibited by φ_β:

$$\mathcal{G}(\beta) = \bigcap_{\beta \in bt(f)} bt(f) = bt(\varphi_\beta).$$

In plain words, the above relation means that the braid types dominated by β are exactly the braid types exhibited by φ_β.

Bolstered by this reformulation of the genealogy set as the set of braid types of a single map for pA braid types, Boyland built a *condensed representative* that serves the same purpose for every family of braid types. The condensed representative is a refinement of the canonical reducible representative from Theorem 1.16. Boyland outlines the procedure to condense the dynamics of a reducible representative in [4, Section 7.5] and gives further references to his own work for the details [5]. The general idea is to collapse periodic orbits with the same braid type.

The first manipulation consists in reducing to the minimum the number of periodic orbits on the boundary of a pA component. Figure 3.3 illustrates the reduction of four fixed points on a circle boundary component to a single fixed point. The second operation aims to merge Nielsen equivalent periodic points in different components of the reducible representative. In particular, a pA

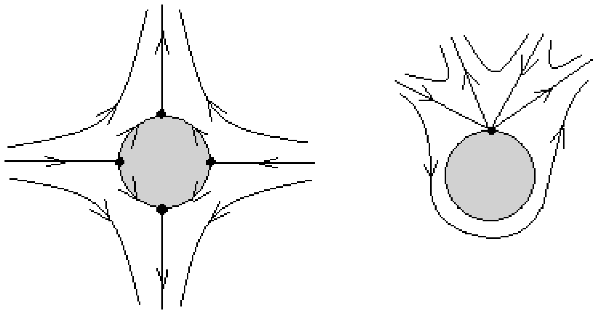

Figure 3.3: On the left, four fixed points lying on the same boundary component. After appropriate reduction, we obtain a single fixed point as illustarted on the right.

homeomorphism of a closed surface is already condensed. The technicalities arise with periodic points inside the tubular neighbourhood of reducing curves. For any formal consideration about the condensed representative, we content ourself with the following theorem (see [5, Theorem 2.4]).

Theorem 3.8 (Boyland, condesed representative)**.** *Each homeomorphism of a compact surface is isotopic to a condensed homeomorphism. Each braid type exhibited by a condensed homeomorphism is uncollapsible and unremovable (i.e. present in every isotopic map). Further, each braid type corresponds to a single periodic orbit.*

Strictly speaking, Definition 2.4 only introduces the concept of collapsibility for periodic points. Theorem 3.8 requires a notion of collapsibility for braid types. The following proposition provides a natural definition of collapsibility for braids. It is again originally due to Boyland [3, Lemma 4.1] and [5, Lemma 2.1].

Proposition 3.9. *Let f be a homeomorphism of a surface M isotopic to the identity and $x \in \mathrm{int}(M)$ be an interior f-periodic point with period $n \geq 1$. Then the following are equivalent.*

1. *There is an f-periodic point y such that x collapses to y in the sense of Definition 2.4.*

2. *If b_x denotes the time evolution n-braid generated by $o(x, f)$ in $M \times I$, then there an integer $1 \leq k < n$ that divides n and a k-braid c in $M \times I$ such that $n = mk$ and b_x is homotopic to c^m. Here c^m denotes the pseudo-braid obtained by copying m times each of the k strings of c.*

3. *There an integer $1 \leq k < n$ that divides n, a covering translation σ and lifts $\widetilde{f}$ and $\widetilde{x}$ of f and x such that $\widetilde{x}$ has period n/k for $\sigma \widetilde{f}^k$.*

The second assertion in Proposition 3.9 is the definition of collapsibility for n-braids proposed by Boyland in [3]. The same definition is stated in terms of lifting classes in [5]. Both notions are equivalent. Since we are principally concerned with braid types, we omit the definition involving lifting classes in this paper.

Proof. We prove that (1) $\Rightarrow$ (2), (2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (1). First assume that there is a periodic point y such that x collapses to y. Let $k := \sharp o(y, f)$ and $m := n/k$. By assumption $1 < m \leq n$. Let b_y denote the time evolution k-braid generated by $o(y, f)$. To see that b_x is homotopic to b_y^m we use an argument similar to the digression that precedes Theorem 3.7.

Let's assume the second assertion. Let h_t be the preferred isotopy from the identity to f. We lift h_t to an isotopy $\widetilde{h}_t$ between the identity in $\widetilde{M}$ and some lift $\widetilde{f} := \widetilde{h}_1$ of f. The ith string of b_x $(i = 1, \dots, n)$ projected down to M is an arc that connects $f^{i-1}(x)$ to $f^i(x)$ parametrized by $\beta_i(t) := h_t(f^{i-1}(x))$. Choose a lift $\widetilde{x}$ of x. The arc $\widetilde{\beta}_i(t) := \widetilde{h}_t(\widetilde{f}^{i-1}(\widetilde{x}))$ connects $\widetilde{f}^{i-1}(\widetilde{x})$ to $\widetilde{f}^i(\widetilde{x})$ and therefore lifts β_i. See Figure 3.4.

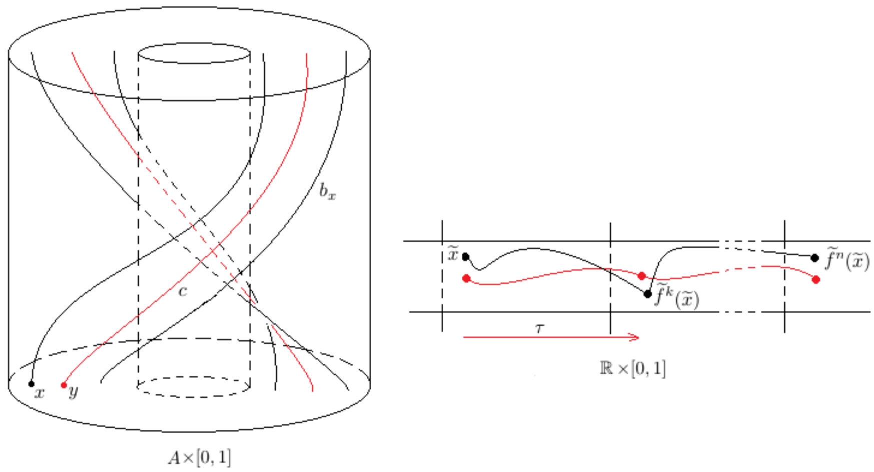

Figure 3.4: Illustration of the situation in the case where M is the annulus and x has period 4.

Let y be the base point of c that is mapped to x under the homotopy between b_x and c^m. We project c down to M to get a loop γ based at y. Let τ be the covering translation associated to γ. From the existence of a homotopy between b_x and c^m, we deduce that τ commutes with $\widetilde{f}^k$ (recall that $k = n/m$). Moreover, $\widetilde{f}^n(\widetilde{x}) = \tau^m \widetilde{x}$. Therefore, with $\sigma := \tau^{-1}$, we obtain $(\sigma \widetilde{f}^k)^m(\widetilde{x}) = \widetilde{x}$ as required.

Finally, assume that the third assertion holds. By assumption, $\sigma \widetilde{f}^k$ is a homeomorphism of $\mathbb{R}^2$ that has a periodic point with period n/k. Brouwer's Lemma on Translation Arcs states that any orientation-preserving homeomorphism h of $\mathbb{R}^2$ for which $\mathrm{Per}(h) \setminus \mathrm{Fix}(h) \neq \emptyset$ has a nonempty set of fixed points [9, Corollary 2.4]. The same result holds for the 2-sphere S^2 [9, Lemma 2.2]. If M is closed, then we can identify $\widetilde{M}$ with either S^2 or $\mathbb{R}^2$. If $\partial M \neq \emptyset$, then $\mathrm{int}(M)$ has universal cover $\mathbb{R}^2$. Since we assumed that $x \in \mathrm{int}(M)$ and $\sigma \widetilde{f}^k$ has a periodic point with period $n/k > 1$, then $\sigma \widetilde{f}^k$ has a fixed point $\widetilde{y}$ in $\widetilde{\mathrm{int}(M)}$. The point y defined to be the projection of $\widetilde{y}$ inside M satisfies the desired conclusion. $\square$

We write φ^c_β for the condensed homeomorphism provided by Theorem 3.8 inside the isotopy class $\Theta(b) \in \mathrm{MCG}(M \text{ rel } X_n; id)$ corresponding to the braid type $\beta = [b]$. A corollary of the existence of a condensed representative is the following analogue of the pA genealogy sets.

Theorem 3.10 (Boyland)**.** *For any braid type $\beta \in CBT(M)$, if the condensed homeomorphism φ^c_β represents β (i.e. if $\beta \in bt(\varphi^c_\beta)$), then*

$$\mathcal{G}(\beta) = bt(\varphi^c_\beta).$$

Theorem 3.10 facilitates the study of the algebraic structure of the forcing order on the set of braids. To determine if a braid type β dominates another braid type γ, it is sufficient to check if γ is the braid type of some periodic orbit of φ^c_β. Markov partitions are a powerful tool to enumerate the periodic orbits of φ^c_β. In the pA case, Bestvina and Handel built an algorithm to find a Markov partition for $\varphi^c_\beta = \varphi_\beta$. See [4, Chapter 10].

3.3.2 The antisymmetry

In a preprint that goes back to 1989 [3], Boyland proved that the antisymmetry of the forcing relation is a consequence of a result due to Brunovsky [6] in bifurcation theory.

Theorem 3.11 (Boyland)**.** *Let M denote a compact surface. Then the relation $(CBT(M), \succeq)$ is a partial order.*

In 1997, Los [22, Lemma 5.10] proposed a different approach to prove the partial order. His arguments rely on the topological aspects of the braid types studied in a combinatorial manner. We do not present a complete proof of the antisymmetry in this paper. However, after discussing Boyland's and Los' proofs, we explain how the antisymmetry is equivalent to a weaker statement about pA braid types.

Boyland's approach is constructive and features geometrical arguments essentially. Given β_1 and β_2 two distinct braid types with $\beta_1 \succeq \beta_2$, Boyland's strategy is to build a homeomorphism that exhibits β_2 but not β_1 in order to conclude that the reverse domination relation does not hold. Start with a homeomorphism f that exhibits both β_1 and β_2. Secondly, isotope f to a simpler homeomorphism f_0 whose periodic points are exactly a finite collection of fixed points. The existence of such a homeomorphism is proved in [5, Proposition 1.6]. Finally, invoke Brunovsky's Theorem (see [3, Lecture 8] or [10]) to find an isotopy h_t from f to f_0 such that periodic orbits of f with period less than a fixed upper bound disappear definitely along h_t at different parameter values (i.e. different times). Let t_1 and t_2 be the suprema of the set of time values t for which $\beta_1 \in bt(h_t)$ and $\beta_2 \in bt(h_t)$ respectively. Since $\beta_1 \succeq \beta_2$, we must have $t_1 < t_2$. In particular, there a time parameter $\tau \in (t_1, t_2)$ such that $\beta_2 \in bt(h_\tau)$ but $\beta_1 \notin bt(h_\tau)$.

Los' argument distinguishes various cases. Assume two braid types β_1 and β_2 dominate each other. Los gives a different justification according to whether β_1 and β_2 are reducible or not in the sense of Definition 3.6. Precisely, he first claims that the reducible case comes down to the irreducible case. He then discusses the situations of pA braid types. However, the antisymmetry is presented as a side result in [22] by Los and a negligible amount of details are presented. We suggest the following argument to justify the reduction to irreducible braid types.

Suppose that either β_1 or β_2 is reducible. The corresponding canonical homeomorphism is reducible in the sense of Definition 1.7. Since it is assumed to be isotopic to the identity, it fixes any reducing system pointwise and therefore each irreducible component is invariant. In particular, both β_1 and β_2 must be associated to the same irreducible component. Otherwise, we massage by isotopy one of the corresponding components (to the identity map for instance) while leaving the other one untouched. In the process we get rid of one braid type but not the other one. Therefore, it is enough to deal with the case where both β_1 and β_2 are irreducible braid types.

The case where both braid types are irreducible and of finite-order type is qualified as "easy" and details are omitted again. Strictly speaking, Los' paper is about the dynamics of the unit disk. Yet the group of topological n-braids $B_n(D^2, X_n)$, where $X_n \subset \mathrm{int}(D^2)$ is collection of n points, is torsion-free for any $n \geq 0$ (see Section 4.1). In particular, there exist no cyclic braid types of finite-order except braid types associated to fixed points. In this instance, it is indeed easy to conclude the antisymmetry for finite-order braid types. However, in a final remark after Lemma 5.10 in [22], Los claims that his proof of the antisymmetry "is exactly the same" for any other surface. If torsion-freeness is the implicit justification, then we shall investigate it. The following proposition is presented as a corollary of Theorem 1.13 by Farb and Margalit [8, Corollary 7.3].

Proposition 3.12. *Let M denote a potentially punctured surface with negative Euler characteristic. If M has nonempty boundary, then* $\mathrm{MCG}(M)$ *is torsion-free.*

The proof of Proposition 3.12 relies on Theorem 1.13 and on the notion of *Dehn twists*. A Dehn twist is an element of $\mathrm{Homeo}^+(A, \partial A)$ (where A denotes the closed annulus) that applies a full rotation of 2π to one of the boundary component while fixing the other one. If we identify a regular neighbourhood of a closed curve in M with the annulus, then we can observe Dehn twists inside homeomorphisms of M. The crucial property in the context of Proposition 3.12 is that Dehn twists about components of ∂M have infinite order. The existence of the group isomorphism Θ between the mapping class group and the braid group for surfaces with negative Euler characteristic gives the following corollary.

Corollary 3.13. *Let M denote a surface with negative Euler characteristic and nonempty boundary. For any $n \geq 0$, if $X_n \subset \mathrm{int}(M)$ denotes a collection of n points in M, then $B_n(M, X_n)$ is torsion-free.*

On the other hand, if M is a closed surface with negative Euler characteristic, then the canonical representative φ_β of any finite-order braid type β is, using Theorem 1.13, isotopic to an isometry ψ of M with respect to some hyperbolic metric on M. We emphasize that the isotopy is relative to a base set of β in M. Since by assumption φ_β is isotopic to the identity, then ψ is also isotopic to the identity and Lemma 1.4 therefore implies that ψ is the identity. We deduce that φ_β represents the identity element in the corresponding mapping class group.

In conclusion, the antisymmetry for irreducible braid types of finite-order is a consequence of torsion-freeness and hyperbolicity for surface of negative Euler characteristic. It remains to deal with pA braid types. A fruitful strategy here consists in introducing an invariant quantity in the sense that it preserves the preorder relation on braid types. A close look at the definition of the order relation $\succeq$ on $CBT(M)$ shows that any quantity attached to braid types that is defined as the infimum of a similar quantity for homeomorphisms would preserve $\succeq$. For instance, if we define the *entropy* of a braid type $\beta \in CBT(M)$ as $h(\beta) := \inf\{h_{top}(f) : \beta \in bt(f)\}$, then it follows from the definition of the forcing on $CBT(M)$ that

$$\beta_1 \succeq \beta_2 \Rightarrow h(\beta_1) \geq h(\beta_2).$$

The unremovability property in Theorem 3.8 is derived similarly as what was achieved previously in this paper for pA homeomorphisms. In particular, every orbit of a condensed homeomorphism inside the interior of a pA component is globally shadowed by an orbit of any isotopic map [4, Theorem 7.7(c)]. Hence, there is an analogue of Theorem 2.13 that holds for condensed maps [5, Theorem 3.2]. Analogously to the pA case, we deduce that a condensed map minimizes the topological entropy in its isotopy class. Subsequently, if β denotes a braid type represented by φ_β^c, then

$$h(\beta) = h_{top}(\varphi_\beta^c).$$

A common result in dynamical systems states that given a continuous transformation T of some compact metric space X, if we can cover X with finitely many closed T-invariant subspaces, then the topological entropy of T is the maximum of the topological entropies of the restricted maps to the invariant subspaces covering X (see for instance [19, Proposition 3.1.7]). Since the topological entropy of a pA map is given by the logarithm of its stretch factor and the topological entropy of a finite-order map is zero, the next result is an intuitive consequence of the construction of condensed homeomorphisms. See [4, Theorem 7.7(a)] for extra details.

Proposition 3.14. *The topological entropy of a condensed homeomorphism is the logarithm of the largest stretch factor of any pA component. If it has no pA component, then it has zero topological entropy.*

In the context of braid types, we formulate the following corollary of Proposition 3.14. It can be however proved without involving condensed homeomorphisms as in [10, Exposé 10].

Corollary 3.15. *If $\beta \in CBT(M)$ denotes a braid type represented by the canonical homeomorphism φ_β, then $h(\beta) > 0$ if and only if φ_β has a pA component.*

As mentioned before, the entropy preserves the order relation on the braid types. To go further in the study of the relation between the preorder on $CBT(M)$ and the entropy of braid types, it is natural to investigate under what conditions on two distinct braid types $\beta_1 \succeq \beta_2$ we have a strict inequality at the level of entropy. As it was pointed out by Boyland [3, Theorem 9.3 (a)], β_1 being pA is a sufficient condition to get a strict inequality between entropies.

Surprisingly, this statement alone implies the antisymmetry of the dynamical order relation on $CBT(M)$. And that is how Los established the partial order in 1997.

For convenience, we only present the equivalence between the antisymmetry and the statement about pA braid types for surface with negative Euler characteristic. This assumption is common in the literature. Recall that if $\chi(M) < 0$, then our definition of braid types for periodic orbits is independent on the choice of a preferred isotopy and equivalent to Boyland's approach (i.e. $\overline{\Theta}$ is bijective). In particular, any braid type $\beta = [b] \in CBT(M)$ is represented by the canonical representative in $\Theta(b)$.

Theorem 3.16. *Let M denote a surface with negative Euler characteristic. Then the following are equivalent.*

1. *The dynamical preorder $\succeq$ on $CBT(M)$ is antisymmetric.*
2. *If β_1 and β_2 are distinct irreducible elements in $CBT(M)$ such that β_1 is pA and $\beta_1 \succeq \beta_2$, then $h(\beta_1) > h(\beta_2)$.*

Proof. First assume that the order relation on $CBT(M)$ is antisymmetric. Let β_1 and β_2 be two distinct irreducible braid types such that β_1 is pA and $\beta_1 \succeq \beta_2$. Since β_2 is irreducible, it is either pA or has finite-order. In the latter case Corollary 3.15 tells us that $h(\beta_1) > 0 = h(\beta_2)$ and the conclusion follows. Therefore, we can assume that β_2 is pA as well. Since by assumption $\beta_1 \succeq \beta_2$, then $h(\beta_1) \geq h(\beta_2)$. Assume *ab absurdo* that $h(\beta_1) = h(\beta_2)$.

As we assumed that $\chi(M) < 0$, for $i = 1, 2$, we can write $\beta_i = bt(o(x_i, \varphi_{\beta_i}))$ where φ_{β_i} is pA relative to $o(x_i, \varphi_{\beta_i})$ and x_i is a periodic point of φ_{β_i}. As noticed before, $h(\beta_i) = h_{top}(\varphi_{\beta_i})$ for $i = 1, 2$. Since we assumed that $\beta_1 \succeq \beta_2$, there is by definition a periodic point y of φ_{β_1} such that $\beta_2 = bt(o(y, \varphi_{\beta_1}))$. Let $Y := o(y, \varphi_{\beta_1})$.

We fix a homeomorphism $h\colon (M, o(x_2, \varphi_{\beta_2})) \to (M, Y)$ isotopic to the identity. Consider the map

$$\varphi_2 := h \circ \varphi_{\beta_2} \circ h^{-1}\colon (M, Y) \to (M, Y).$$

The map φ_2 is by definition conjugate to φ_{β_2} and thus $h_{top}(\varphi_2) = h_{top}(\varphi_{\beta_2})$. If we write $\varphi_1 := \varphi_{\beta_1}$ to emphasize that φ_1 is a homeomorphism of M relative to Y, then our absurd assumption implies $h_{top}(\varphi_1) = h_{top}(\varphi_2)$. Observe that φ_2 is pA relative to Y. Since β_2 is the braid type corresponding to the time evolution of Y in both φ_1 and φ_2, both maps are isotopic relative to Y. This is immediate if we switch to Boyland's definition of braid types. All hypothesis of Handel's Theorem (Theorem 2.20) are verified. We deduce the existence of a continuous and surjective map d of M isotopic to the identity, that makes the following diagram commute.

$$\begin{tikzcd}
(M, Y) \arrow[r, "\varphi_1"] \arrow[d, "d"'] & (M, Y) \arrow[d, "d"] \\
(M, Y) \arrow[r, "\varphi_2"'] & (M, Y)
\end{tikzcd}$$

Recall that both φ_1 and φ_2 fix Y and thus so does d. Therefore, if $X := o(x_1, \varphi_{\beta_1})$, then $d(X)$ is a periodic orbit of φ_2 and hence $(h^{-1} \circ d)(X)$ is a periodic orbit of φ_{β_2}. Since d acts as a conjugacy map, it follows that

$$bt(o((h^{-1} \circ d)(x_1), \varphi_{\beta_2})) = bt(o(x_1, \varphi_{\beta_1})) = \beta_1.$$

Therefore, $\beta_1 \in bt(\varphi_{\beta_2})$ and thus $\beta_2 \succeq \beta_1$ because we know from Theorem 3.10 that $bt(\varphi_{\beta_2})$ is exactly the set of braid types dominated by β_2. As we assumed the antisymmetry of $\succeq$, it follows that $\beta_1 = \beta_2$. This is a contradiction and hence $h(\beta_1) > h(\beta_2)$ as desired.

To prove the converse, we assume the second assertion. Let β_1 and β_2 be two cyclic braid types in $CBT(M)$ such that $\beta_1 \succeq \beta_2$ and $\beta_2 \succeq \beta_1$. In particular, we have $h(\beta_1) = h(\beta_2)$. As it was enlightened previously, we can reduce to the case where both β_1 and β_2 are irreducible. Since we have equality at the entropy level, either both are pA or both have finite-order. The finite-order case follows again from the previous digression about torsion-freeness. If β_1 and β_2 are pA, then the second assertion implies $\beta_1 = \beta_2$ as required. $\square$

Chapter 4

Forcing on the disk

When it comes to forcing for periodic orbits, among every surface, the unit disk has been studied the most. As a historical notice, in his first preprint on the subject [3], back in 1989, Boyland principally dealt with the case of the disk. He did however generalise his results later on. Of the key advantages in working with the disk is a purely algebraic description of the braid group due to Artin [2]. Furthermore, authors like Los [22] and Hall [11, 12] developed combinatorial tools to describe dynamics on the disk.

4.1 Algebraic structure of the braid group

Let $n \geq 1$ be an integer and $X_n \subset \operatorname{int}(D^2)$ be a finite collection of points. The Extended Birman Exact Sequence (Theorem 3.5) describes a close relation between the braid group and the mapping class group of the corresponding marked surface. It requires a contractibility condition on the group of orientation- and boundary-preserving homeomorphisms provided by the Alexander trick (Lemma 1.9) in the case of the disk. The Extended Birman Exact Sequence thus applies and gives the exactness of the following sequence:

$$1 \longrightarrow \pi_1(C_n(D^2); X_n) \xrightarrow{Push} \mathrm{MCG}(D^2 \text{ rel } X_n) \xrightarrow{Forget} \mathrm{MCG}(D^2) \longrightarrow 1.$$

Another consequence of the Alexander trick is the triviality of the mapping class group of the disk. Further, as we identified n-braids based at X_n with homotopy classes of loops based at X_n in $C_n(D^2)$ (Lemma 3.3), we obtain the following group isomorphism:

$$B_n(D^2, X_n) \cong \mathrm{MCG}(D^2 \text{ rel } X_n).$$

Recall that a topological n-braid is defined to be a collection of n strings $\{\gamma_i\}$ inside the suspension cylinder $D^2 \times I$. The braid configuration can be projected from inside the cylinder to a vertical plane. In practice, it corresponds to the choice of a viewing angle from which we observe the cylinder. A *braid diagram* is the image obtained by projecting the images $\gamma_i(I) \subset D^2 \times I$ on a fixed vertical plane where we make sure to indicate at each crossing of strands which one is passing over the other one as illustrated in Figure 4.1. The opportunity is taken to emphasize that the braid diagram of a given braid depends on the choice of the vertical plane (i.e. the viewing point).

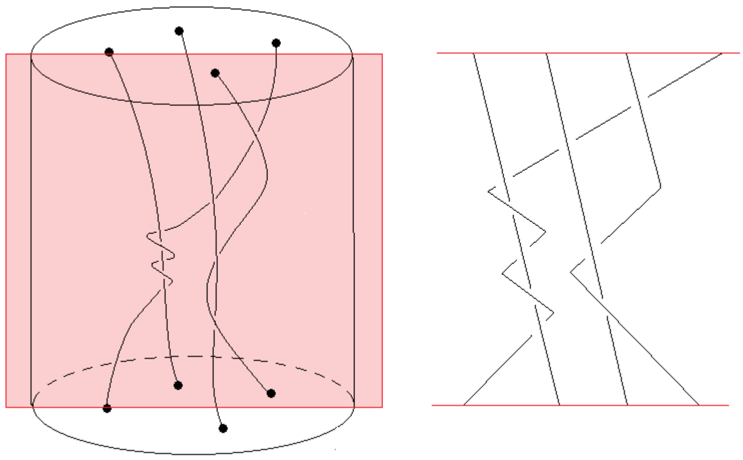

Figure 4.1: Braid diagram corresponding to the 4-braid introduced in Figure 3.1. The vertical plane is the red slice that splits the cylinder. It corresponds to the choice of a frontal angle of observation.

It is convenient to fix a model for the positions of the points of X_n inside the disk. Identify $D^2 = \{z \in \mathbb{C} : |z| \leq 1\}$ and then let $X_n = \{x_1, \ldots, x_n\}$ where, for $i = 1, \ldots, n$, we define $x_i := 2i/(n+1) - 1$. In other words, we choose the x_i's to be equidistant on the intersection of the real axis with the interior of D^2. Having fixed X_n, we adopt the following abbreviations: let $B_n := B_n(D^2, X_n)$, $BT_n := BT_n(D^2)$ and $\mathrm{MCG}_n := \mathrm{MCG}(D^2 \text{ rel } X_n)$.

For $i = 1, \ldots, n-1$, let $\sigma_i \in B_n$ be the isotopy class of the n-braid whose ith strand overcrosses the $(i+1)$th strand once and all other strands are vertical. Figure 4.2 illustrates the braid diagram of σ_i. The group B_n is generated

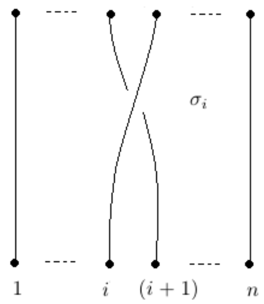

Figure 4.2: Braid diagram of the n-braid σ_i.

by the σ_i's. Intuitively every class of n-braids has a representative for which strands cross each other at different heights. This representative can readily be expressed as a finite concatenation of the σ_i's and their inverses. More precisely, Artin gave the following purely algebraic description of the braid group [2]:

$$B_n \cong \left\langle \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_i\sigma_j = \sigma_j\sigma_i, \text{ for } |i-j| > 1, \\ \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}, \text{ for all } i. \end{array} \right\rangle$$

The group on the right-hand side is called *Artin's braid group* with $n-1$ generators. When representing a word in the symbols σ_i's as a braid diagram, we adopt the convention of reading the word from left to right and drawing the corresponding string crossings from bottom to top. The binding relations in the quotient can be outlined with the corresponding braid diagrams as illustrated in Figure 4.3.

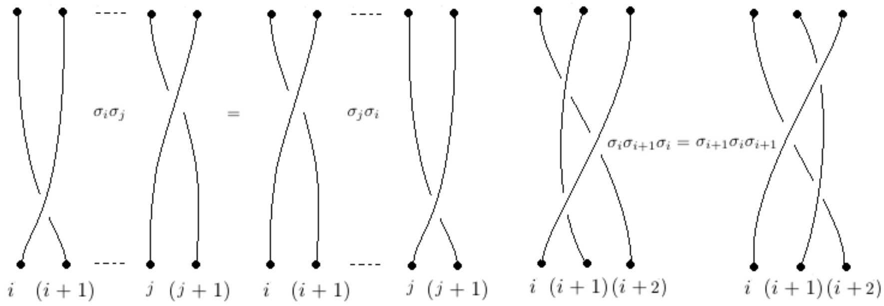

Figure 4.3: Illustration of the two binding relations in Artin's algebraic description of B_n.

Among the algebraic properties of the braid group we mention torsion-freeness. It has been seen that this characteristic came in handy when dealing with the antisymmetry of the order relation. There are different approaches to prove torsion-freeness. One of them adapts the argument of Proposition 3.12. We refer the curious reader to [1, Section 2.5] where three distinct proofs are presented.

Furthermore, the braid group is not commutative when $n \geq 3$. Its centre is an infinite cyclic group generated by a full twist. The *full twist*, illustrated in Figure 4.4, is the isotopy class in MCG_n corresponding to the following n-braid:

$$\theta_n := (\sigma_1\sigma_2 \dots \sigma_{n-1})^n \in B_n.$$

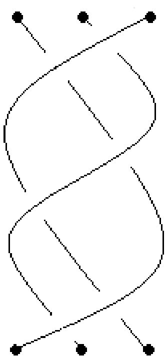

Figure 4.4: The full twist with three stings. It corresponds to the element $\theta_3 = (\sigma_1\sigma_2)^3 \in B_3$.

By a direct computation or by contemplating the braid diagram, one can convince himself that θ commutes with every generator of B_n. What is more, θ is the only generator of the center of the braid group. Details are provided in [8, Section 9.2]. If $Z(G)$ denotes the center of an abstract group G, then $Z(B_n) = \langle\theta\rangle$ holds.

We carefully emphasized in various places already that without requiring the isotopies to fix the boundary pointwise in the definition of the mapping class group (Definition 1.2), the following group homomorphism is not always injective:

$$\Theta' \colon B_n \twoheadrightarrow \mathrm{Homeo}^+(D^2 \text{ rel } X_n)/\mathrm{Homeo}_0^+(D^2 \text{ rel } X_n).$$

The group on the right-hand side is an alternative definition of the mapping class group. Since no boundary behaviour is imposed, it can be identified with the mapping class group (in the sense of Definition 1.2) of the open disk $\mathrm{int}(D^2) \cong S^2 \setminus \{\star\}$ relative to X_n:

$$\mathrm{Homeo}^+(D^2 \text{ rel } X_n)/\mathrm{Homeo}_0^+(D^2 \text{ rel } X_n) \cong \mathrm{MCG}(\mathrm{int}(D^2) \text{ rel } X_n).$$

The kernel of the homomorphism Θ' is the infinite cyclic group generated by the full twist, a.k.a. the center of B_n [8, Section 9.2]. Therefore, the following group isomorphism holds:

$$B_n/Z(B_n) \cong \mathrm{MCG}(\mathrm{int}(D^2) \text{ rel } X_n).$$

Observe furthermore that if $Y_{n+1} \subset S^2$ denotes a collection of $n+1$ points on the sphere, then $\mathrm{MCG}(\mathrm{int}(D^2) \text{ rel } X_n)$ can be further identified with an index n subgroup of $\mathrm{MCG}(S^2 \text{ rel } Y_{n+1})$ consisting of the elements that fix one designated marked point.

We content ourself with the previous considerations on the algebraic properties of the braid group in the present survey. For a deeper study, one may consult [8, Section 9.2].

4.2 Homeomorphisms of the disk

It has been enlightened that the dynamical partial order on braid types is directly connected with the Thurston-Nielsen classification of mapping classes. If the surface is the genus-1 torus, then $\mathrm{MCG}(T^2)$ is identified with $\mathrm{SL}(2;\mathbb{Z})$ and the classification comes down to computing the trace of the associated matrix. Yet, there is a standard way to descend a homeomorphism of the torus to a homeomorphism of the disk and thus take advantage of the well understood classification of toral mapping classes.

4.2.1 Construction from the torus

The following procedure is described in these terms by Boyland [4]. We think of the torus T^2 as the quotient $(\mathbb{R}/\mathbb{Z})^2$. Every linear transformation $C \in \mathrm{SL}(2;\mathbb{Z})$ descends to a homeomorphism $\overline{C}$ of T^2. Let

$$S := \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

and $\overline{S}$ denote the associated toral homeomorphism. The transformation $\overline{S}$ has four fixed points; namely the projections of $(0,0),(1/2,0),(0,1/2)$ and $(1/2,1/2)$. Geometrically, $\overline{S}$ is a rotation of the torus of angle π about a straight line axis that goes through all four fixed points. Therefore, the quotient $T^2/\overline{S}$ is the sphere S^2.

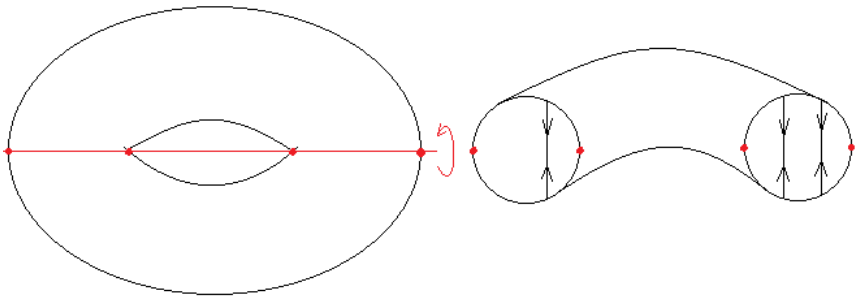

Figure 4.5: The rotation of π around the axis in red on the left. Both circle components are then collapsed to obtain the quotient.

Given $C \in \mathrm{SL}(2;\mathbb{Z})$, since it always commutes with S as matrices, $\overline{C}$ descends to a homeomorphism C^* of the sphere. To further obtain a homeomorphism of the disk, a natural idea would be to blow-up the sphere around a fixed point of C^* (for instance the projection of the origin). The lack of smoothness prevents such an operation in general. As we would like to keep the dynamical properties of C^*, it is not an option to alter it by isotopy to gain smoothness.

$$\begin{tikzcd}
\mathbb{R}^2 \arrow[r] \arrow[d, "{C,S}"'] & T^2 \arrow[r] \arrow[d, "{\overline{C},\overline{S}}"'] & T^2/\overline{S} \cong S^2 \arrow[d, "C^*"'] \\
\mathbb{R}^2 \arrow[r] & T^2 \arrow[r] & T^2/\overline{S} \cong S^2
\end{tikzcd}$$

To circumvent this difficulty, we apply the blow-up before taking the two quotients. Let $\widehat{\mathbb{R}^2}$ denote the blow-up of $\mathbb{R}^2$ at $\mathbb{Z}^2$. For $C \in \mathrm{SL}(2;\mathbb{Z})$, let $\widehat{C}$ denote the blow-up of C. The quotient of $\widehat{\mathbb{R}^2}$ by the integer lattice defines a genus-1 torus with one boundary component denoted T_1^2. The map $\widehat{C}$ descends to a homeomorphism $\overline{C}'$ of T_1^2. Finally, we quotient T_1^2 by the action of $\overline{S}'$. It defines a sphere with one boundary component or said differently a disk. Since S and C still commute after the blow-up, the matrix C descends to a homeomorphism C' of the disk.

$$\begin{tikzcd}
\mathbb{R}^2 \arrow[d, "{C,S}"'] & \widehat{\mathbb{R}^2} \arrow[l] \arrow[r] \arrow[d, "{\widehat{C},\widehat{S}}"'] & T_1^2 \arrow[r] \arrow[d, "{\overline{C}',\overline{S}'}"'] & T_1^2/\overline{S}' \cong D^2 \arrow[d, "C'"'] \\
\mathbb{R}^2 & \widehat{\mathbb{R}^2} \arrow[l] \arrow[r] & T_1^2 \arrow[r] & T_1^2/\overline{S}' \cong D^2
\end{tikzcd}$$

Example 4.1. Consider the linear transformation of Example 1.1 defined as

$$H_A := \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Seen as a toral homeomorphism, $\overline{H}_A$ is an Anosov transformation because trace$(H_A) > 2$. It has a periodic orbit of length 3 given by $(1/2, 1/2) \mapsto (0, 1/2) \mapsto (1/2, 0)$ identified with their projections down on the torus. As it will be later enlightened, the braid type associated to this periodic orbit plays a fundamental role in the understanding of the partial order on BT_3.

As an Anosov homeomorphism, $\overline{H}_A$ admits, among other dynamical properties, a Markov partition. This structure descends to the disk. In particular, if H'_A denotes the disk transformation associated to H_A, then H'_A has a dense orbit and the set of H'_A-periodic points is dense in D^2.

4.2.2 Line diagrams

The isomorphism between the braid group and the mapping class group shows that braids record the dynamical behaviour of each isotopy class about the marked points. Another method to record dynamical features with geometrical objects consists in looking at the images of the line segments joining consecutive marked points. Recall that we fixed $X_n \subset \text{int}(D^2)$ to be a collection of n consecutive points x_i on the real axis. For $i = 1, \ldots, n-1$, let α_i be the oriented horizontal line segment from x_i to x_{i+1}.

Definition 4.1 (line diagram)**.** Let $[f] \in \text{MCG}_n$. The *line diagram* $\mathcal{L}_{[f]}$ of $[f]$ is the set of homotopy classes relative to X_n of the directed images under f of the line segments α_i's. Associated to the line diagram is the permutation generated by f on the set X_n.

For a matter of precision, an *arc homotopy relative to* X_n means a homotopy with fixed endpoints such that the interior of any intermediate arc never passes through a point of X_n. An example of line diagram is given in Figure 4.6.

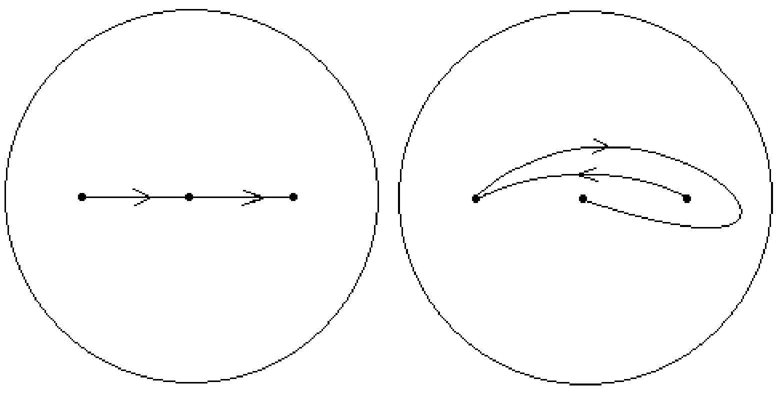

Figure 4.6: The action of the mapping class corresponding to $\sigma_1\sigma_2^{-1}$ in MCG_3. On the left-hand side is the inital configuration and on the right-hand side are two arc representatives of the homotopy classes under the action of $\sigma_1\sigma_2^{-1}$.

The line diagram does not totally determine a class in MCG_n. In other words, different elements of MCG_n are associated to the same line diagram. However, the line diagram determines a unique isotopy class in MCG_n up to multiplication by an iterate of the full twist. Recall that the full twist is the unique generator of the center of MCG_n. That is, there is a one-to-one correspondence between the set of line diagrams and MCG_n modulo its center. This result is presented in the following proposition and was proved by Hall in [15, Lemma 3.8].

Proposition 4.1. *Let $[f]$ and $[g]$ be two elements of MCG_n. If $\mathcal{L}_{[f]} = \mathcal{L}_{[g]}$, then f and g are isotopic relative to X_n (but not necessarily relative to ∂D^2).*

To prove Proposition 4.1 two intermediate results are needed. The first one is a theorem due to David Epstein [7]. It is a statement about extension of arc homotopy to ambient boundary-fixing isotopy of the surface.

Theorem 4.2 (Epstein)**.** *Let M denote a surface with nonempty boundary and $X \subset \mathrm{int}(M)$ be a finite collection of points. Let γ_1 and γ_2 be two arcs in M with common endpoints in ∂M. If γ_1 and γ_2 are homotopic relative to X, then there is a homeomorphism $f \in Homeo^+(M, \partial M)$ such that $\gamma_1 = f \circ \gamma_2$ and $[f]$ is the identity element in $\mathrm{MCG}(M \text{ rel } X)$.*

The second statement necessary to prove Proposition 4.1 is a computation of the mapping class group of the marked sphere. Recall that the braid group B_n modulo its center was identified with an index n subgroup of $\mathrm{MCG}(S^2 \text{ rel } Y_{n+1})$ where $Y_{n+1} \subset S^2$ is a collection of $n+1$ points. The following lemma is proved in various places in the literature (see for instance [8, Proposition 2.3]).

Lemma 4.3. *Let x and y be two distinct points on the sphere and $Y := \{x, y\}$. Then $\mathrm{MCG}(S^2 \text{ rel } Y) \cong \mathbb{Z}/2\mathbb{Z}$.*

Proof. A surjective group homomorphism $\mathrm{MCG}(S^2 \text{ rel } Y) \to \mathbb{Z}/2\mathbb{Z}$ is given by the action of a mapping class on the two marked points. It remains to prove that a homeomorphism f of S^2 that fixes both x and y is isotopic to the identity relative to Y.

Let γ be the geodesic arc that connects x to y. Since γ and $f \circ \gamma$ share endpoints and lie on the sphere, there are homotopic relative to Y. Hence, we can assume, without loss of generality, that f fixes γ pointwise: $\gamma = f \circ \gamma$. Cut S^2 along γ and descend f to a homeomorphism of the disk that fixes the boundary. The Alexander trick gives the desired conclusion. $\square$

Proof of Proposition 4.1. Let α_i denote the horizontal line segment from x_i to x_{i+1} for $i = 1, \ldots, n-1$. Because of the group structure on MCG_n and the definition of the line diagram, it is sufficient to prove that if $f \circ \alpha_i$ is homotopic to α_i relative to X_n for all $i = 1, \ldots, n-1$, then f is isotopic to the identity relative to X_n.

Using Theorem 1.8, we can further assume that f is smooth and non-singular at each x_i. Therefore, f lifts to a homeomorphism $\widehat{f}$ of the blow-up $\widehat{D}^2$ of D^2 at X_n. Let $\pi\colon \widehat{D}^2 \twoheadrightarrow D^2$ denote the canonical projection. It is sufficient to prove that $\widehat{f}$ is isotopic to the identity in $\widehat{D}^2$ relative to every boundary circle except $\pi^{-1}(\partial D^2)$. This isotopy then descends to an isotopy between f and the identity in D^2 relative to X_n.

By assumption $f(x_i) = x_i$ for all $i = 1, \ldots, n-1$. Therefore, $\widehat{f}$ fixes each component of $\partial\widehat{D}^2$. Up to an isotopy around each boundary component, we can assume that $\widehat{f}$ fixes the boundary pointwise. For $i = 1, \ldots, n-1$, let $\widehat{\alpha}_i := \overline{\pi^{-1}(\mathrm{int}(\alpha_i))}$ be a lift of α_i that connects the corresponding boundary circles. By assumption $\widehat{f} \circ \widehat{\alpha}_i$ is homotopic to $\widehat{\alpha}_i$ with fixed endpoints.

Write $M^{(n)} := \widehat{D}^2$ and $\varphi^{(n)} := \widehat{f}$. With the help of Epstein's Theorem (Theorem 4.2), up to an isotopy, we can further assume that $\varphi^{(n)}$ fixes $\widehat{\alpha}_{n-1}$ pointwise. Similarly as in the proof of Lemma 4.3, we cut $M^{(n)}$ along $\widehat{\alpha}_{n-1}$ to obtain a surface $M^{(n-1)}$ with n boundary components (one less than $M^{(n)}$). The homeomorphism $\varphi^{(n)}$ lifts to a homeomorphism $\varphi^{(n-1)}$ of $M^{(n-1)}$ such that:

- $\varphi^{(n-1)} \circ \widehat{\alpha}_i \simeq \widehat{\alpha}_i$ relative to their endpoints, for $i = 1, \ldots, n-1$,
- $\varphi^{(n-1)}$ fixes $\partial M^{(n-1)}$ pointwise,
- any isotopy of $M^{(n-1)}$ relative to $\partial M^{(n-1)} \setminus \pi^{-1}(\partial D^2)$ lifts to an isotopy of $M^{(n)}$ relative to $\partial M^{(n)} \setminus \pi^{-1}(\partial D^2)$.

Inductively, after $n-1$ iterations of the previous procedure, we eventually end up with a surface $M^{(1)}$ homeomorphic to the annulus A and a homeomorphism $\varphi^{(1)}\colon A \to A$ that fixes both boundary components pointwise. Forgetting the boundary, Lemma 4.3 implies that $\varphi^{(1)}$ is isotopic to the identity. Moreover, the proof of Lemma 4.3 indicates that this isotopy can be chosen relative to the inside boundary component of A (i.e. $\partial M^{(1)} \setminus \pi^{-1}(\partial D^2)$). Such an isotopy generates an isotopy $\varphi^{(2)} \simeq id$ relative to $\partial M^{(2)} \setminus \pi^{-1}(\partial D^2)$. Going back every step of the induction, we eventually obtain an isotopy $\varphi^{(n)} \simeq id$ relative to $\partial M^{(n)} \setminus \pi^{-1}(\partial D^2)$. This gives the desired conclusion. □

Example 4.2 (continuation of Example 4.1)**.** We identify the 3-periodic orbit of H'_A described in Example 4.1 with the points of X_3 in the following manner: $x_1 = (0, 1/2)$, $x_2 = (1/2, 1/2)$ and $x_3 = (1/2, 0)$. The line diagram associated to the isotopy class of H'_A relative to X_3 is computed by first looking at the images of the line segments in the torus and then projecting to the disk. The line diagram is illustrated in Figure 4.6. Up to full twists, the 3-braid corresponding to the time evolution of X_3 under H'_A is $\sigma_1\sigma_2^{-1} \in B_3$.

Unlike the braid corresponding to the time evolution of X_n, the line diagram misses some information about the dynamics of the mapping class; namely the presence of full twists. However, the line diagram remains a useful symbolic representation.

4.2.3 Fat representative and Markov encoding

Starting from a line diagram, Hall described a construction to obtain what he called a fat one-dimensional representative [12]. Analogously to Boyland's condensed representative, Hall's fat one-dimensional representative has a minimal orbit structure that is easily determined with the help of a Markov encoding of the dynamics. The description outlined here sketches the careful work of Hall.

The first step consists in transforming the horizontal line segments between consecutive points in X_n into Markov boxes. Visually, every line segment is fattened up to a rectangle. Each box is foliated with vertical stable leaves and horizontal unstable leaves. The Markov boxes are connected together with boxes around each point of X_n. The Markov boxes are labelled $M_1, \ldots, M_{n-1}$ and the boxes around the marked points are labelled $A_1, \ldots, A_n$. The heights and lengths are adjusted so that everything fits inside the unit disk. Denote by R the rectangle defined as the union of all boxes. An illustration is provided in Figure 4.7.

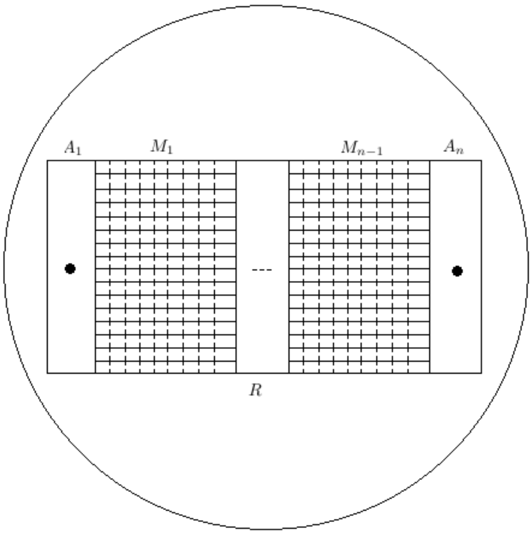

Figure 4.7: Configuration of the rectangle R where the $n-1$ horizontal line segments have been fattened-up to foliated Markov boxes $M_1, \ldots, M_{n-1}$. To complete the rectangle, n boxes $A_1, \ldots, A_n$ are added around each marked point.

Proposition 4.1 says that a line diagram $\mathcal{L}$ determines an isotopy class $[f]_{\mathcal{L}}$ relative to X_n. Inside $[f]_{\mathcal{L}}$, we chose a representative φ that satisfies the following properties:

1. $\varphi(R) \subset \operatorname{int}(R)$,
2. φ preserves the foliations and uniformly contracts, respectively expands, along stable, respectively unstable, leaves,
3. each box A_i is an attractor for φ^n and φ^n is a contraction on every A_i,
4. the boundary ∂D^2 is a repeller and is fixed pointwise by φ,
5. the set of images under φ of the Markov boxes collapsed along stable leaves lies in $\mathcal{L}$ (to ensure $\mathcal{L}_{[\varphi]} = \mathcal{L}$).

Such a homeomorphism φ is called a *fat representative* and has the following properties. The periodic points of φ are either the fixed points on ∂D^2, the points in X_n or some points in $M_1 \cup \cdots \cup M_{n-1}$. It is a consequence of the repeller character of ∂D^2 and the attractor character of the points in X_n. Furthermore, by the Brouwer Fixed Point Theorem, φ has at least one fixed point in R.

Example 4.3 (continuation of Example 4.2)**.** Start with the line diagram corresponding to H'_A illustrated in Figure 4.6. The action of a fat representative is illustrated in Figure 4.8. Contemplating Figure 4.8, it seems feasible to encode the dynamics of the fat representative in the Markov boxes in terms of a shift map as it is commonly done with horseshoes.

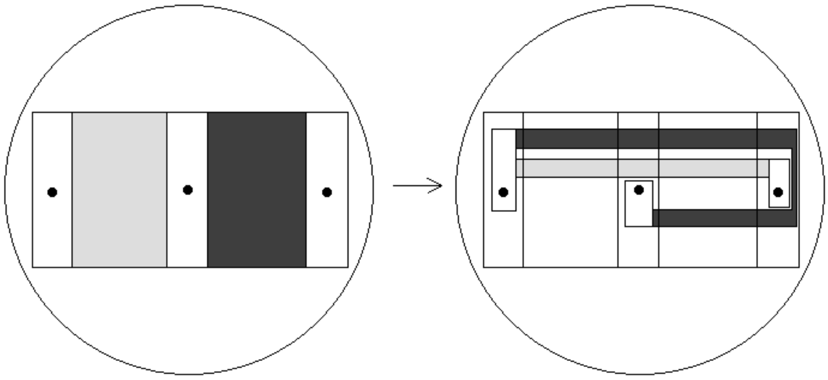

Figure 4.8: The action of a fat representative in the isotopy class of H'_A relative to X_3.

The central object used to define an encoding of a fat representative is the so-called transition matrix.

Definition 4.2 (transition matrix)**.** The *transition matrix* of a fat representative φ is an $(n-1)\times(n-1)$ matrix E with nonnegative integer entries defined as follows: for any pair of indices $i, j \in \{1, \ldots, n-1\}$, the entry E_{ij} is the number of connected components of $M_i \cap \varphi(M_j)$.

For instance, the transition matrix of the fat representative illustrated in Example 4.3 is nothing but the initial matrix H_A used to construct the disk homeomorphism H'_A.

The transition matrix encodes the dynamics of the associated fat representative in the sense that it can be used to build a conjugacy with a subshift of finite type. We recall here the definition of such a system.

Definition 4.3 (subshift of finite type)**.** Let Σ_n be set of two-sided infinite sequences on n symbols. The dynamical system $\sigma\colon \Sigma_n \to \Sigma_n$ defined by shifting the sequence to the left by one place is called the *full shift*.

Let B denote an $n \times n$ matrix with entries being only zeros or ones. Define a subspace $\Sigma(B) \subset \Sigma_n$ by

$$\Sigma(B) := \left\{ (a_i)_{i=-\infty}^{+\infty} \in \Sigma_n : B_{a_i a_{i-1}} = 1, \forall i \in \mathbb{Z} \right\}.$$

The subspace $\Sigma(B)$ is σ-invariant. The system σ_B defined as the restriction of σ to $\Sigma(B)$ is called a *subshift of finite type*.

The transition matrix defined earlier has nonnegative integer entries and thus not necessarily only zeros or ones. It is however possible to perform a sequence of operations to a transition matrix to obtain a (larger) matrix with entries $\{0, 1\}$ and that still records the key dynamical data. The following procedure is again described by Hall in [11].

Start with an $m \times m$ matrix E with nonnegative integer coefficients. Assume E contains some entry larger than 1. Let c_1 be the index of the leftmost column that does not only contains zeros and ones. Let a_1 be the largest element in the c_1th column. Consider the block matrix of size $m \times a_1$ in which the ith line start with E_{ic_1} ones followed by zeros. Replace the c_1th column of E with this block matrix. Add sufficiently many copies of the last line in this new matrix to get a square matrix $E^{(1)}$. Repeat until the resulting matrix $E^{(\infty)}$ only has entries $\{0, 1\}$ (despite the notation, this algorithm terminates after finitely many repetitions).

A direct computation shows that from an eigenvector of E we can build an eigenvector of $E^{(1)}$ by repeating the last entry as many times as the last line was copied when building $E^{(1)}$. These two eigenvectors are associated to the same eigenvalue. The extra eigenvalues of $E^{(1)}$, compared to the set of eigenvalues of E, are only zeros corresponding to the copies of the last line.

Example 4.4 (continuation of Example 4.3)**.** As previously observed, the transition matrix associated to H'_A is the matrix H_A:

$$H_A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

This matrix has an entry larger than 1. We perform the operations described above. First we replace the second column with a 2×2 block matrix:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Finally we copy the last line once to obtain

$$H_A^{(1)} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Observe that the eigenvalues of $H_A^{(1)}$ are the two eigenvalues of H_A plus 0 (see Example 1.1). Geometrically, the extension of the matrix H_A in the second column corresponds to a vertical splitting of the second Markov box. The transition matrix corresponding to the three Markov boxes illustrated in Figure 4.9 is precisely the matrix $H_A^{(1)}$.

If φ denotes a fat representative with transition matrix E, then the *subshift of finite type associated to* E (and thus φ) is defined to be the subshift of finite type associated to the matrix $E^{(\infty)}$. It is denoted $\sigma_E := \sigma_{E^{(\infty)}}$ and acts on the space of sequences $\Sigma(E) := \Sigma(E^{(\infty)})$.

It has been seen in Example 4.4 that the matrix $E^{(\infty)}$ corresponds to the transition matrix with respect to a larger collection of Markov boxes (obtained by splitting vertically some of the original boxes). Let M denote the union of all Markov boxes including the split boxes. It is a standard exercise in dynamical systems to prove that the dynamics of the restriction of φ to $\bigcap_{i \in \mathbb{Z}} \varphi^i(M)$ are conjugate to the subshift of finite type $\sigma_E \colon \Sigma(E) \to \Sigma(E)$.

Among all subshifts of finite type, the ones associated to matrices that have a power with positive entries are of particular interest. One can directly check that the set of periodic orbits of such a shift is dense and further that there

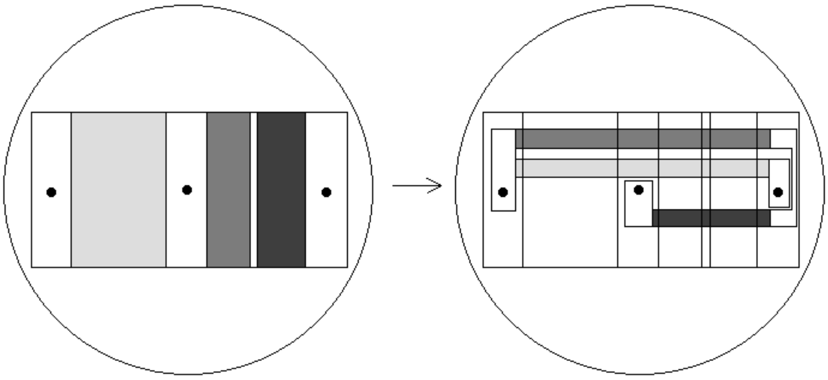

Figure 4.9: Figure 4.8 altered by a splitting of the second Markov box.

exists a dense orbit. A more elaborated argument shows that the topological entropy of such a shift is the logarithm of the spectral radius of its associated matrix [10, Exposé 10].

It is thus of prime interest to understand what fat representatives have transition matrix with a power having only positive entries. Such matrices are called *Perron-Fröbenius* as they satisfy the hypothesis of the Perron-Fröbenius Theorem.

Theorem 4.4 (Perron-Fröbenius)**.** *Let B denote a Perron-Fröbenius matrix, i.e. B is a square matrix with nonnegative entries and has a power with positive entries. Then the following properties hold:*

1. *B has an eigenvalue $\lambda > 1$ that equals the spectral radius of B,*
2. *there is an eigenvector of B associated to λ and an eigenvector of B^t (B transposed) associated to λ with positive entries,*
3. *the eigenspace associated to λ is one-dimensional.*

As a corollary of Theorem 4.4, observe that a subshift of finite type associated to a Perron-Fröbenius matrix has positive topological entropy. Indeed, by Theorem 4.4, the spectral radius of a Perron-Fröbenius matrix is larger than 1. This assertion about entropy suggests that among fat representatives with positive entropy we might find Perron-Fröbenius transition matrices. Of course, pseudo-Anosov homeomorphisms have positive topological entropy and are thus natural candidates to test. The next result is due to Hall [12, Lemma 4].

Proposition 4.5. *If φ is a fat representative of a pseudo-Anosov isotopy class relative to X_n, then its transition matrix E is Perron-Fröbenius.*

The proof of Proposition 4.5 requires the following elementary lemma. A proof is provided by Hall in [11, Lemma 1.9].

Lemma 4.6. *Let B denote a square matrix with nonnegative entries and such that trace$(B) > 0$. Then B is Perron-Fröbenius if and only if for every pair of indices (i, j), there is an integer $k = k(i, j) \geq 1$ such that $(B^k)_{ij} > 0$.*

Proof of Proposition 4.5. Since φ has a fixed point inside one of the Markov boxes, there is an index $i \in \{1, \dots, n-1\}$ such that $E_{ii} > 0$. Hence $\mathrm{trace}(E) > 0$. By Lemma 4.6, it is sufficient to prove that given any pair of indices $i, j \in \{1, \dots, n-1\}$, there is an integer $k \geq 1$ such that $(E^k)_{ij} > 0$. Equivalently, it is sufficient to find an integer $k \geq 1$ such that $M_i \cap \varphi^k(M_j) \neq \emptyset$.

Fix an index $j \in \{1, \dots, n-1\}$. For any $r = 1, \dots, n-1$, let Γ_r be a simple closed curve in $A_r \cup M_r \cup A_{r+1}$ that is the boundary of a disk containing the marked points x_r and x_{r+1}. Let m be the smallest nonnegative integer such that $\varphi^m(x_j) = x_1$. Since φ^n lies in a pA mapping class, Proposition 1.23 establishes the existence of an integer $s \geq 1$ such that $\varphi^{ns+m}(\Gamma_j) \cap \Gamma_{n-1} \neq \emptyset$.

Let $k := ns + m$. Now observe that $\varphi^k(\Gamma_j) \subset \varphi^k(A_j \cup M_j \cup A_{j+1})$ and $\varphi^k(\Gamma_j) \cap A_1 \neq \emptyset$. Since $\varphi^k(\Gamma_j) \cap \Gamma_{n-1} \neq \emptyset$, then $\varphi^k(\Gamma_j)$ must intersect every Markov box M_i for $i = 1, \dots, n-2$. By changing Γ_{n-1} for Γ_1 and choosing m such that $\varphi^m(x_j) = x_n$, the same argument provides an integer k' for which $\varphi^{k'}(\Gamma_j)$ intersects every Markov box M_i for $i = 2, \dots, n-1$. The desired conclusion follows. $\square$

As a matter of fact, Proposition 4.5 admits a converse. A fat representative is said to be *minimal* if it has a minimal number of periodic orbits exhibiting each braid type compared to to any isotopic homeomorphism. Hall gives an equivalent condition for being minimal that can be directly checked on the line diagram [12, Theorem 6]. In the process, Hall deduces that a minimal fat representative whose transition matrix is Perron-Fröbenius lies in a pseudo-Anosov mapping class relative to X_n. The idea is to use the positive eigenvectors provided by Theorem 4.4 to define the lengths and heights of the foliated Markov boxes. With suitable identifications of the sides, the quotient space becomes a sphere that can be then blown-up to get a disk. The reader is invited to consult [12] for more details. In the present paper, we content ourself with the following example.

Example 4.5 (continuation of Example 4.4)**.** Observe that in the case of a 2×2 transition matrix with nonnegative integer entries, having a trace greater than 2 is sufficient to be Perron-Fröbenius. Regarding toral homeomorphisms, these correspond to Anosov transformations.

In the case of interest, the transition matrix H_A has trace 3 and thus descends to an Anosov homeomorphism $\overline{H}_A$ of the torus T^2. As explained earlier, the map $\overline{H}_A$ further descends to a homeomorphism H_A^* of the sphere $S^2 \cong T^2/\overline{S}$ where $\overline{S}$ is the toral rotation generated by the linear transformation

$$S = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The transverse foliations of T^2 descend to transverse foliations of S^2 with 1-pronged singularities at the four fixed points of $\overline{S}$ (in coherence with the Euler-Poincaré formula; Theorem 1.14). Let $X \subset S^2$ denote the projections of the four fixed points of $\overline{S}$. The map H_A^* is a pA homeomorphism of S^2 relative to X. Let $x_0 \in S^2$ denote the projection of the origin $(0,0)$. By Proposition 1.24, H_A^* has blow-up $\widehat{H}_A^* \colon \widehat{S}^2 \to \widehat{S}^2$ at x_0. Since the resulting space is the unit disk, we obtain a pA homeomorphism $\widehat{H}_A^*$ of the unit disk relative to $X \setminus \{x_0\}$. The corresponding transverse foliations of the unit disk have three

1-pronged singularities at the points of $X \setminus \{x_0\}$ and a 3-pronged singularity on the boundary (again in coherence with the Euler-Poincaré formula).

Because of the Anosov character of H_A, there is no need to worry about when to apply the blow-up. In particular, the homeomorphisms $\widehat{H}_A^*$ and H'_A lie the same conjugacy class of isotopic homeomorphisms relative to X_3. Moreover, this class is pseudo-Anosov and therefore the associated braid type $[\sigma_1\sigma_2^{-1}]$ is pseudo-Anosov in the sense of Definition 3.7.

In conclusion, through the Examples 1.1 to 4.5, it has been seen that the Anosov toral homeomorphism associated to the linear transformation

$$H_A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

descends to a pseudo-Anosov disk homeomorphism relative to a period-3 orbit. With the help of a conjugation, this orbit can be identified with the model X_3. The corresponding 3-braid is $\sigma_1\sigma_2^{-1} \in B_3$.

4.3 Algebraic structure of the forcing order

The algebraic criterion to determine whether a braid type dominates another one is due to Handel [16]. It applies to the restricted case of pseudo-Anosov 3-braids and extends a prior result of Matsuoka [23]. We start by giving two algebraic criteria to test if a given braid is pseudo-Anosov. The following proposition appears in various places in the literature and is considered as a standard result. The reader is invited to consult [4, Section 8.2] for further references.

Proposition 4.7. *If f is a finite-order orientation-preserving homeomorphism of the disk, then f is conjugate to a rotation. More precisely, if $f^q = id$, then f is conjugate to $R_{p/q} \colon (r, \theta) \mapsto (r, \theta + p/q)$ for some integer $0 \leq p < q$.*

When dealing with the sphere or any genus 0 surface, the Jordan Curve Theorem helps to determine whether a mapping class is irreducible. The following proposition is also considered as a well-known result.

Proposition 4.8. *Let f be an orientation-preserving homeomorphism of the disk. Assume that f has a periodic orbit $o(x, f)$ of prime length q. Then f lies in an irreducible mapping class relative to $o(x, f)$.*

Proof. Assume $\{\gamma_1, \ldots, \gamma_m\}$ are pairwise disjoint representatives of a reducing system. Each closed curve γ_i must enclose at least two points of the orbits so that the interior of γ_i has negative Euler characteristic (seeing the marked points as punctures). Further, no curve γ_i can enclose all q points, since its exterior also has negative Euler characteristic. In particular, such a configuration is impossible when $q = 2$. If $q > 2$, then $m \geq 2$ must hold. By definition of a reducing system, f permutes the reducing curves. This implies the existence of a non-trivial factor for q and therefore provides the desired contradiction since q is assumed to be a prime number. $\square$

Proposition 4.7 provides a characterisation of finite-order mapping classes on the disk. By incorporating Proposition 4.8 we obtain the following classification of irreducible cyclic braid types.

Corollary 4.9. *Let q be a prime number. The cyclic braid type associated to a periodic orbit of length q is either pseudo-Anosov or is the braid type associated to a periodic orbit of $R_{p/q}$ for some nonnegative integer $p \not\equiv 0 \pmod{q}$ (because of full twists, p might be larger than q).*

There is an algebraic criterion to determine if a q-braid is pseudo-Anosov or finite-order. It involves the concept of exponent sum of a braid.

Definition 4.4 (exponent sum)**.** Given a q-braid $b \in B_q$, we define the *exponent sum* of b as the sum of the exponents of the generators in a word that represents b. The binding relations in Artin's description of B_q preserve the exponent sum and hence it is a well-defined quantity. Furthermore, it is conjugacy invariant and thus it is well-defined at the level of braid types. The exponent sum of a braid type β is written $es(\beta)$.

Let $\alpha_{p/q}$ be the braid type associated to q-periodic points under $R_{p/q}$. By drawing the corresponding braid diagram, we observe that $\alpha_{1/q}$ is represented by $\sigma_1 \dots \sigma_{q-1}$ and hence $es(\alpha_{1/q}) = q - 1$. Since $R_{p/q} = (R_{1/q})^p$, then $es(\alpha_{p/q}) = p(q-1)$. Corollary 4.9 can be refined as follows.

Lemma 4.10. *Let q be a prime number. Given a cyclic braid type β associated to a q-periodic orbit, then either $es(\beta) = p(q-1)$ for some integer $p \not\equiv 0 \pmod{q}$ and β is reducible or β is pseudo-Anosov.*

This provides another way to prove that the homeomorphism H'_A in Example 4.5 lies in a pseudo-Anosov class relative to X_3 since the associated time-evolution 3-braid $\sigma_1\sigma_2^{-1}$ has exponent sum 0.

As mentioned in introduction, Handel's algebraic description of the forcing order concerns BT_3. The 3-braids σ_1 and σ_2^{-1} play a central role in Handel's work. They are the associated time-evolution braids of X_3 under the action of L'_1 and L'_2 respectively, where L'_1 and L'_2 are defined as the disk homeomorphisms generated by the following two Anosov transformations:

$$L_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } L_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

A direct computation shows that $H_A = L_2 L_1$ where H_A is the matrix of Example 4.5. This relation is expected to hold as the 3-braid corresponding to the isotopy class of H_A relative to X_3 is precisely $\sigma_1\sigma_2^{-1}$. Recall that the braid words are conventionally read from left to right and therefore L_2 is multiplied on the left in order for L_1 to act first.

Now assume L is any matrix inside $SL(2;\mathbb{Z})$ with a trace larger than 2. The matrix L can be decomposed as follows:

$$L = L_2^{i_r} L_1^{j_r} \dots L_2^{i_1} L_1^{j_1},$$

where the indices $i_1, \dots, i_r$ and $j_1, \dots, j_r$ are nonnegative integers. This is a standard result concerning families of generators in $SL(2;\mathbb{Z})$. Therefore, the isotopy class relative to X_3 of the disk homeomorphism L' corresponding to L is associated with the 3-braid $\sigma_1^{j_1}\sigma_2^{-j_1} \dots \sigma_1^{j_r}\sigma_2^{-j_r}$. Lemma 4.10 implies that L' lies in a pseudo-Anosov class since the exponent sum of the associated braid is 0. Hence, like it was primarily observed for the matrix H_A in Example 4.5, any

Anosov transformation of the torus descends to a pseudo-Anosov isotopy class relative to X_3 of disk homeomorphisms.

Reciprocally, Matsuoka established the following algebraic characterisation of pseudo-Anosov braid types in BT_3 [24]. The proof relies on the converse of the previous statement, namely that any pseudo-Anosov disk homeomorphism relative to three interior points descends from an Anosov toral homeomorphism [10, Exposé 13]. This precise property makes period 3 special.

Theorem 4.11 (Matsuoka)**.** *A braid type in BT_3 is pseudo-Anosov if and only if it is represented by a cyclic word consisting solely of the generators σ_1 and σ_2^{-1} (and not their inverses) and has at least one of each of these generators.*

Theorem 4.11 is not directly connected with the forcing order on braid types. However, nine years after Matsuoka published his characterisation of pseudo-Anosov 3-braid types, Handel, starting from the characterisation, described the forcing dynamical order on BT_3 in the same algebraic language. He proved the following theorem in 1995.

Theorem 4.12 (Handel)**.** *Let β_1 and β_2 denote two pseudo-Anosov braid types in BT_3. Then $\beta_1 \succeq \beta_2$ if and only if the cyclic word in σ_1 and σ_2^{-1} contained in β_2 is obtained from that of β_1 by deleting generators.*

For instance, $[(\sigma_1\sigma_2^{-1})^2] \succeq [\sigma_1\sigma_2^{-1}]$ and $[\sigma_1^3\sigma_2^{-1}] \succeq [\sigma_1\sigma_2^{-1}]$. The natural next step would be to look for an analogous algebraic description for braid types based on a larger set. From the words of Handel himself, if such an explicit description exists, then "it will not be nearly as simple as the one presented here" [16].

References

[1] Juan González-Meneses. Basic results on braid groups. *Annales mathématiques*, 18(1):15–59, 2011.

[2] Emil Artin. Theory of braids. *Annals of Mathematics*, 48(1):101–126, 1947.

[3] Philip Boyland. Notes on dynamics of surface homeomorphisms. 1989.

[4] Philip Boyland. Topological methods in surface dynamics. *Topology and its Applications*, 58:223–298, 1994.

[5] Philip Boyland. Isotopy stability for dynamics on surfaces. 1999.

[6] Pavol Brunovsky. On one parameter families of diffeomorphisms I and II. *Comment. Math. Univ. Carolin.*, 11:559–582, 1970.

[7] David Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.

[8] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, 2012.

[9] Albert Fathi. An orbit closing proof of Brouwer's Lemma on translation arcs. *L'enseignement Mathématique*, 33(1):315–322, 1987.

[10] Albert Fathi, François Laudenbach, and Valentin Poénaru. *Travaux de Thurston sur les Surfaces*, volume 66 of *Astérisque*. Société mathématique de France, Paris, 1979.

[11] Toby Hall. *Periodicity in chaos: the dynamics of surface automorphisms.* PhD thesis, University of Cambridge, 1991.

[12] Toby Hall. Fat one-dimensional representatives of pseudo-anosov isotopy classes with minimal periodic orbit structure. *Nonlinearity*, 367(7):367–384, 1994.

[13] Mary-Elizabeth Hamstrom. Homotopy groups of the space of homeomorphisms on a 2-manifold. *Illinois J. Math*, 10(4):563–573, 1966.

[14] Michael Handel. Global shadowing of pseudo-anosov homeomorphisms. *Ergodic Theory and Dynamical Systems*, 5:373–377, 1985.

[15] Michael Handel. Entropy and semi-conjugacy in dimension two. *Ergodic Theory and Dynamical Systems*, 8:585–596, 1987.

[16] Michael Handel. The forcing partial order on the three times punctured disk. *Ergodic Theory and Dynamical Systems*, 17:593–610, 1995.

[17] Boju Jiang. *Lectures On Nielsen Fixed Point Theory*, volume 14 of *Contemporary Mathematics*. American Mathematical Society, Providence, Rhode Island, 1982.

[18] Boju Jiang and Jianhan Guo. Fixed points of surface diffeomorphisms. *Pacific Journal of Mathematics*, 160(1):67–90, 1993.

[19] Anatole Katok and Boris Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*, volume 54 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, Cambridge, 1995.

[20] Erwan Lanneau and Jean-Luc Thiffeault. On the minimum dilatation of pseudo-Anosov homeomorphisms on surface of small genus. *Annales de L'Institut Fourier*, 61(1):105–144, 2011.

[21] Tien-Yien Li and James A. Yorke. Period three implies chaos. *The American Mathematical Monthly*, 82(10):985–992, 12 1975.

[22] Jérôme Los. On the forcing relations for surface homeomorphisms. *Publications mathématiques de l'I.H.É.S.*, 85:5–61, 1997.

[23] Takashi Matsuoka. Braids of periodic points and a 2-dimensional analogue of Sharkovskii's ordering. *Dynamical Systems and Nonlinear Oscillations*, 1:58–72, 1986.

[24] Takashi Matsuoka. Periodic points and braid theory. *Handbook of Topological Fixed Point Theory*, pages 171–216, 2005.

[25] Igor Nikolaev. *Foliations on Surfaces*, volume 41 of *A Series of Modern Surveys in Mathematics*. Springer, Berlin, 2001.

[26] Oleksandr Sharkovskii. Co-existence of cycles of a continuous mapping of the line into itself. *Ukrainian Mathematical Journal*, 16:61–71, 1964.

[27] Edward E. Slaminka. A Brouwer translation theorem for free homeomorphisms. *Transactions of the American Mathematical Society*, 306(1):277–291, 1988.

[28] Kurt Strebel. *Quadratic differentials*, volume 5 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin Heidelberg, 1984.

[29] Marc Troyanov. Les surfaces Euclidiennes à singularités coniques. *L'Enseignement Mathématique*, 32(2):79–94, 1986.