May 8, 2025

THURSTON'S DEFORMATION RETRACTION OF TEICHMÜLLER SPACE

INGRID IRMER

Abstract. In [24], a short, simple and elegant construction of a mapping class group-equivariant deformation retraction of Teichmüller space of a closed compact surface was given. The preprint [24], which unfortunately is not online, has not been broadly accepted. The purpose of this paper is to go through the construction in detail and resolve any questions that have arisen in the literature and in personal communications.

Contents

1. Introduction

One reason for studying the Teichmüller space of a surface is because it is a contractible space on which the mapping class group acts properly discontinuously. On the other hand, it is clear that Teichmüller space is not the lowest dimensional space on which the mapping class group acts properly discontinuously; it is known that there exist nontrivial mapping class group-equivariant deformation retractions, for example Theorem 2.7 of [5].

When the surface has one or more punctures, an explicit mapping class group-equivariant deformation retraction of Teichmüller space was given in [7] and [22]. The dimension of the image of this deformation retraction was shown to be equal to the virtual cohomological dimension of the mapping class group; this is a homological invariant that provides a lower bound on the dimension of the image of such a deformation retraction. The existence of a puncture was a crucial ingredient in all these constructions; informally speaking, a puncture or some form of marked point is needed, relative to which coordinates defining a cell decomposition are defined.

Thurston's construction in [24] resolved the problem of a missing basepoint by using curve lengths to parametrise Teichmüller space, constructing a mapping class group-equivariant deformation retraction of the Teichmüller space of a closed, compact surface. The image of this deformation retraction is the so-called *Thurston spine* $\mathcal{P}_g$. This is a CW complex contained in $\mathcal{T}_g$ consisting of the set of points representing hyperbolic surfaces that are cut into polygons by the set of shortest geodesics (also known as the systoles).

Section 3 discusses the construction from [24] in detail. A number of questions about this construction have been raised, for example [11]. Section 4 resolves all these questions. This paper aims to preserve the elegant simplicity of the construction from [24], while providing the technical details. Also for its historical significance, the author is of the opinion that this material, based on a talk given by Thurston in the 1980s, should be readily available. Surveys of other ground-breaking contributions to the study of Teichmüller space made by or strongly influenced by Thurston can be found in [18], [19] and [20].

Acknowledgements. The author would like to thank Stavros Garoufalidis, Scott Wolpert and Don Zagier for helpful discussions and comments.

2. Definitions and conventions

The purpose of this section is to supply detailed definitions and background for the rest of the paper. A good reference for much of this background is the textbook [15].

Let $\mathcal{S}_g$ be a closed, compact, connected, orientable surface of genus g.

A *marking* of $\mathcal{S}_g$ is a diffeomorphism $f : \mathcal{S}_g \to M$, where M is a closed, orientable, hyperbolic surface with genus $g \geq 2$, and $\mathcal{S}_g$ is a closed, orientable, topological surface of genus g. The Teichmüller space $\mathcal{T}_g$ is the set of pairs (M, f) modulo the equivalence relation $(M, f) \sim (N, h)$ if $f \circ h^{-1}$ is isotopic to an isometry. A topology on $\mathcal{T}_g$ is usually assumed that makes it homeomorphic to $\mathbb{R}^{6g-6}$. More information can be found in Section 10.6 of [6].

The group of isotopy classes of orientation preserving diffeomorphisms from $\mathcal{S}_g \to \mathcal{S}_g$ is known as the *mapping class group* Γ_g of $\mathcal{S}_g$. The mapping class group acts on $\mathcal{T}_g$ by changing the marking, namely $\Gamma_g \times \mathcal{T}_g \to \mathcal{T}_g$ is given by $\gamma \times (M, f) \mapsto (M, f \circ \gamma^{-1})$. The quotient of $\mathcal{T}_g$ by this action is called the *moduli space* of $\mathcal{S}_g$ and will be denoted by $\mathcal{M}_g$.

It will be assumed that the genus $g \geq 2$ in order to ensure that all surfaces admit a hyperbolic structure. Once a point in Teichmüller space is chosen, by an abuse of notation, S_g will be used to denote the surface $\mathcal{S}_g$ endowed with the corresponding marked hyperbolic structure.

A *curve* on S_g is assumed to be a closed, embedded, nonoriented, nontrivial isotopy class of maps of S^1 to S_g. The length of the curve will be defined to be the length of its geodesic representative. When there is no possibility of confusion, the image of a particular representative of the isotopy class, such as a geodesic, will also be referred to as a curve. An individual curve will be denoted by a lowercase c, sometimes with a subscript, whereas a finite set of curves will be denoted by an uppercase C.

A set of curves on a surface is said to *fill* the surface if the complement of the geodesic representatives is a union of polygons. The Thurston spine, $\mathcal{P}_g$, is the set of points at which the set of shortest curves (these are called *systoles*) fill the surface. The systole function $f_{\text{sys}} : \mathcal{T}_g \to \mathbb{R}$ is the function whose value at the point $x \in \mathcal{T}_g$ is given by the length of the systoles at x.

Whenever a metric is required, the Weil-Petersson metric will be assumed.

Definition 1 (Systole stratum $\text{Sys}(C)$)**.** *For a fixed set of curves C, the systole stratum $\text{Sys}(C)$ is the set of points of $\mathcal{T}_g$ on which the set of systoles is exactly C.*

Following Thurston, the term "stratum" is used to mean a decomposition into locally closed subsets, with each point contained in a neighbourhood that intersects only finitely many strata. This stratification can also be extended to the metric completion of $\mathcal{T}_g$ with respect to the Weil-Petersson metric, where a stratum of noded surfaces is labelled by the multicurve that has been pinched to obtain the noded surface.

A systole stratum is a semi-analytic subset of $\mathcal{T}_g$. It is the solution to a system of analytic equations, stating that certain geodesics (the systoles) have the same length in addition to a locally finite set of inequalities ensuring that these geodesics are shorter than all others. The local finiteness is well-known, and follows for example from the collar lemma. As a consequence of Lojasiewicz's theorem, [14], a neighbourhood of $\mathcal{P}_g$ in $\mathcal{T}_g$ admits a triangulation compatible with the stratification. $\mathcal{P}_g$ is therefore a simplicial complex.

Tangent cones and cone of increase. The piecewise-smooth structure behind many of the arguments in this paper make it convenient to define tangent cones, analogously to definitions given for polyhedra in [4] or [2].

The tangent cone to a simplex T at point p of $\mathcal{T}_g$ is the set of $v \in T_p\mathcal{T}_g$ such that $v = \dot{\gamma}(0)$ for a smooth oriented path $\gamma(t)$ with $\gamma(0) = p$ and $\gamma(\epsilon)$ in T for sufficiently small $\epsilon > 0$. In other words, it is the set of 1-sided limits of tangent vectors to the simplex. When $p \in \mathcal{P}_g$ is on the boundary of more than one simplex of $\mathcal{P}_g$, the tangent cone to $\mathcal{P}_g$ at p is the union of the tangent cones of the simplices with p on the boundary. Tangent cones of strata are defined similarly. For a triangulation compatible with the stratification, the tangent cone to $\text{Sys}(C)$ at p is the union of the tangent cones of the simplices with interior contained in $\text{Sys}(C)$ with p on the boundary. Lojasiewicz's theorem also applies to level sets within strata, which can therefore also be triangulated, and for which tangent cones can be defined.

As f_{sys} is only piecewise-smooth, the notion of gradient is replaced by the *cone of increase*. By local finiteness, at a point $p \in \text{Sys}(C)$, f_{sys} can only be increasing or stationary in a direction in which the lengths of all the curves in C are increasing or stationary. The *cone of increase* of f_{sys} at p is given by the tangent cone at p of the intersection $I(C, x) := \{x \in \mathcal{T}_g \mid L(c)(x) \geq f_{\text{sys}}(p)\ \forall c \in C\}$.

A *length function* is an analytic map $\mathcal{T}_g \to \mathbb{R}^+$, the simplest example of which is the map whose value at the point $x \in \mathcal{T}_g$ is given by the length of c at x. More general length functions are positive linear combinations of lengths of curves. Length functions were shown to be convex along earthquake paths in [12] and strictly convex on Weil-Petersson geodesics in [25].

As f_{sys} is not smooth, it cannot be a Morse function. There is however a sense in which it behaves just like a Morse function.

Definition 2 (Topological Morse function). *Let M be an n-dimensional topological manifold. A continuous function $f : M \to \mathbb{R}^+$ is a topological Morse function if the points of M consist of regular points and critical points. When $p \in M$ is a regular point, there is an open neighbourhood U containing p, where U admits a homeomorphic parametrisation by n parameters, one of which is f. When p is a critical point, there exists a $k \in \mathbb{Z}$, $0 \leq k \leq n$, called the index of p, and a homeomorphic parametrisation of U by parameters $\{x_1, \ldots, x_n\}$, such that everywhere on U, f satisfies*

$$f(x) - f(p) = \sum_{i=1}^{n-k} x_i^2 - \sum_{n-k+1}^{i=n} x_i^2$$

It was shown in [1] that f_{sys} is a topological Morse function.

Topological Morse functions were first defined in [17], where it was shown that, when they exist, they can be used in most of the same ways as their smooth analogues for constructing cell decompositions of manifolds and computing homology. While Morse theory was not mentioned explicitly in [24], the deformation retraction is reminiscent of the way in which a Morse function can be used to construct a cell complex homotopy equivalent to a manifold with boundary.

3. Thurston's deformation retraction

This section describes Thurston's deformation retraction onto the Thurston spine in detail. References are [24] and Chapter 3 of [8].

Thurston constructed a Γ_g-equivariant isotopy ϕ_t from $\mathcal{T}_g$ into a regular neighbourhood of $\mathcal{P}_g$. Referring to ϕ_t as an isotopy has led to some confusion, as ϕ_t is defined for all $t > 0$. Although ϕ_t can be defined for all $t > 0$, it is only the restriction of ϕ_t to a compact interval $[0, T]$ that will be needed. The construction relies on the next proposition.

Proposition 3 (Proposition 0.1 of [24]). *Let C be any collection of curves on a surface that do not fill. Then at any point of $\mathcal{T}_g$, there are tangent vectors that simultaneously increase the lengths of all the geodesics representing curves in C.*

Remark 4. *It is important to note that Proposition 3 implies that all critical points of f_{sys} are contained in $\mathcal{P}_g$.*

The proof of Proposition 3 given in [24] uses Lipschitz maps, and is explained in detail in [21]. A different proof will be given here, illustrating how the convexity of length functions constrains the differential topology of $\mathcal{M}_g$. There is no claim to originality. Results similar to Proposition 3 have been proven using a variety of techniques; the first instance of which the author is aware can be found in Lemma 4 of [3]. Wolpert has also pointed out that it follows from Riera's formula, [23].

Proof. Let $C = \{c_1, \dots, c_n\}$. The length of a curve c will be denoted by $L(c)$. Let $L(c)_x$ be the level set of $L(c)$ passing through a point x of $\mathcal{T}_g$.

Since the curves in C do not fill, the intersection $N(x) := \cap_{j=1,\dots,n} L(c_j)_x$ is not compact. This is because the intersection must be invariant under the action of a subgroup of Γ_g generated by Dehn twists around curves disjoint from the curves in C.

A length function $\sum_{i=1}^n a_i L(c_i)$ with each $a_i \in \mathbb{R}^+ \cup \{0\}$ and not uniformly zero cannot have a minimum in $\mathcal{T}_g$. This is because such a minimum must be a unique point by strict convexity, but $N(x)$ is not compact for any $x \in \mathcal{T}_g$.

It is always possible to find a point $w \in \mathcal{T}_g$ at which the lemma holds. This can be done by finding a point q in the metric completion of $\mathcal{T}_g$ with respect to the Weil-Petersson metric, with the property that a curve c is pinched at q, where c has nonzero geometric intersection number with each of the curves in C. Choosing w sufficiently close to q will ensure that the lemma holds at w.

Suppose the proposition breaks down at $y \in \mathcal{T}_g$. Along a path γ from w to y, there must be a point $z \in \mathcal{T}_g$ at which the lemma first breaks down. At z, there exists therefore a nontrivial subset G_z of $\{\nabla L(c_i) \mid c_i \in C\}$ that spans a proper subspace of $T_z\mathcal{T}_g$, and whose elements are not contained in a halfspace of this subspace.

The existence of G_z implies that it is possible to find $a_1, \dots, a_n \in \mathbb{R}^+ \cup \{0\}$ not all zero such that the sum

$$\sum_{i=i}^{n} a_i \nabla L(c_i)(z)$$

is zero. By strict convexity of length functions along Weil-Petersson geodesics, this implies that the length function

$$L = \sum_{i=i}^{n} a_i L(c_i)$$

has a local—and hence global—minimum at z. The proposition follows by contradiction. □

For any $\epsilon > 0$, an open subset $\mathcal{P}_{g,\epsilon}$ of $\mathcal{T}_g$ is defined to be the subset of $\mathcal{T}_g$ consisting of hyperbolic structures such that the set of geodesics whose length is within ϵ of the shortest length fill the surface. It is not hard to see that each $\mathcal{P}_{g,\epsilon}$ is open, its projection to $\mathcal{M}_g$ has compact closure, and the intersection of $\mathcal{P}_{g,\epsilon}$ over all positive ϵ is the subcomplex $\mathcal{P}_g$. It follows that for any regular neighbourhood $\mathcal{N}_g$ of $\mathcal{P}_g$, there is an ϵ such that $\mathcal{P}_{g,\epsilon} \subset \mathcal{N}_g$.

Thurston's choice of vector field. Recall that the Weil-Petersson metric used here is invariant under the action of the mapping class group. At a point x of $\mathcal{T}_g \setminus \mathcal{P}_g$, let $C(x)$ be a set of shortest geodesics. The notion "shortest" will be made precise later; $C(x)$ can contain more curves than just the systoles at x. If the geodesics in $C(x)$ do not fill the surface, by Proposition 3, it is possible to define a Γ_g-equivariant vector field X_C with the property that the length of every curve in C is increasing in the direction of X_C. For nonfilling C, Thurston suggested a vector field with unit length that maximises the sum

$$\sum_{c \in C} \log X_C(L(c)) \tag{1}$$

As the logarithms are assumed to be real, this is a shorthand way of saying that $X_C(x)$ determines a smooth choice of direction in which the length of each curve in C is increasing. For a point x very close to $\mathcal{P}_g$ the curves in $C(x)$ might also fill, depending on how the notion of "set of shortest curves" is defined. For simplicity, X_C is defined to be zero when the curves in C fill. The construction will only require a vector field that is nonzero outside of some regular neighbourhood of $\mathcal{P}_g$ that can be made arbitrarily small.

Denote the cardinality of a set S by $|S|$. For an $\epsilon > 0$ define $U_C(\epsilon)$ to be the set containing every point x of $\mathcal{T}_g$ representing a hyperbolic structure for which C is the set of curves of length less than $f_{\text{sys}}(x) + |C|\epsilon$. Then $\mathcal{U} := \{U_{C_i} \mid C_i \text{ is a finite set of curves on } \mathcal{S}_g\}$ is a cover of $\mathcal{T}_g$ when ϵ is sufficiently small.

Paradoxically, local finiteness of $\mathcal{U}$ is a consequence of the fact that the number $n(L, x)$ of simple curves of length less than or equal to L at a point $x \in \mathcal{T}_g$ grows faster than linearly, [16]. For $x \in \mathcal{T}_g$, the cardinality of a set $C(x)$ of curves of length less than $f_{\text{sys}}(x) + |C|\epsilon$ is either uniformly bounded depending on ϵ, or infinite, in which case U_C is not in $\mathcal{U}$.

This unusual construction of a cover was presumably made because it has the property that if two sets U_{C_1} and U_{C_2} intersect, either $C_1 \subset C_2$ or $C_2 \subset C_1$. This choice is not in any way canonical, and different choices are clearly also allowable here.

Note that for every point x not on $\mathcal{P}_g$, there is an ϵ such that for some set U_{C_i} containing x, the curves in C_i do not fill.

Let $\{\lambda_{C_i}\}$ be a partition of unity subordinate to the covering $\{U_{C_i}\}$. The partition of unity is chosen in such a way as to be invariant under the action of Γ_g on the sets of geodesics $\{C\}$. For example, it could be defined as a function of geodesic lengths. The vector field X_ϵ is constructed by using the partition of unity $\{\lambda_{C_i}\}$ to average over the vector fields $\{X_{C_i}\}$. Note that this averaging process does not create zeros. For a point x in the intersection of the open sets U_{C_i}, $i = 1, \ldots, k$, there is at least one shortest or equal shortest curve c in the intersection of the sets C_i. Any vector field X_{C_i}, $i = 1, \ldots, k$ evaluated at x has the property that if it is nonzero, it increases the length of c at x. It follows that X_ϵ can only be zero at x if every vector field being averaged over at x is zero.

Denote by K a subset of $\mathcal{T}_g$ that is compact modulo the action of the mapping class group. The goal is now to construct an isotopy ϕ_t of $\mathcal{T}_g$ with the property that for any ϵ there is a $T(\epsilon)$ such that taking $t > T(\epsilon)$ ensures that for any K, $\phi_t(K)$ is contained within $\mathcal{P}_{g,\epsilon}$. This is done by using the flow generated by $X_{\epsilon'(t)}$ where $\epsilon'(t) > 0$ is small. The parameter $\epsilon'(t)$ can be decreased as time goes on.

The proof that $X_{\epsilon'(t)}$ generates a flow defined for all $t \in [0, \infty)$ is the same as the proof of the standard result that a smooth, compactly supported vector field is complete; see for example Theorem 9.16 of [13]. In particular, first recall that the α-thick part of $\mathcal{T}_g$, call it $\mathcal{T}_g^\alpha$, defined as $\mathcal{T}_g^\alpha = \{x \in \mathcal{T}_g \mid f_{\text{sys}}(x) \geq \alpha\}$ is invariant under the flow and compact modulo the action of Γ_g. This means that by the existence and uniqueness theorem of ODEs, around each point p in $\mathcal{T}_g^\alpha$ there is a neighbourhood $\mathcal{U}_p$ and an $\epsilon > 0$ such that the flow is defined on $\mathcal{U}_p \times \big[0, \epsilon(p)\big)$. By compactness, there is a nonzero uniform lower bound ϵ of $\epsilon(p)$ on $\mathcal{T}_g^\alpha$. Consequently, it is always possible to flow for a time ϵ longer. This concludes the proof that $X_{\epsilon'(t)}$ generates a flow defined for all $t \in [0, \infty)$.

The next lemma gives control over the rate at which f_{sys} increases along the flowlines of X_ϵ outside of the thick part of $\mathcal{T}_g$. Recall that $C_i(x)$ is the set of curves with length at x less than $f_{\text{sys}}(x) + |C_i|\epsilon$ defined above, and ϵ_M denotes the Margulis constant.

Lemma 5. *When $X_{\epsilon'}$ is Thurston's vector field, i.e. it is constructed using Equation (1), for sufficiently small ϵ' there is an upper bound on the time needed for any flowline γ of $X_{\epsilon'}$ to enter $\mathcal{T}_g^{\epsilon_M}$.*

Proof. At a point x at which $f_{\text{sys}}(x) < \epsilon_M$, the systoles at x are pairwise disjoint, and therefore have cardinality at most $3g-3$. Suppose $\epsilon' < \frac{\epsilon_M}{3g-3}$. For a point $x \in \mathcal{T}_g$ at which the width of a collar of a geodesic of length at most $f_{\text{sys}}(x)+(3g-3)\epsilon'$ is greater than $f_{\text{sys}}(x)+(3g-3)\epsilon'$, any choice of $C_i(x)$ has cardinality at most $3g-3$. Since the curves in $C_i(x)$ are pairwise disjoint, it follows from Riera's formula, [23], that the gradients of the lengths of any two curves in $C_i(x)$ make an angle less than $\frac{\pi}{2}$. In the worst case, the Weil-Petersson gradients $\{\nabla L(c_1), \dots, \nabla L(c_k)\}$ make angles pairwise close to $\frac{\pi}{2}$. When $f_{\text{sys}(x)}$ is arbitrarily small relative to ϵ', for a systole c at x, this does not give a uniform bound away from $\frac{\pi}{2}$ on the angle between $\nabla L(c)(x)$ and $X_{C_i(x)}$. After flowing for a time at most $(3g-3)(3g-4)\epsilon'$, either a bound can be obtained, because in the worst case scenario, $C_i(x)$ will then have cardinality 1, or f_{sys} is large enough that the curves in C_i intersect. Consequently, after a time t at most $(3g-3)^2\epsilon'$, either $f_{\text{sys}} \circ \gamma(t) > \epsilon'$ or $f_{\text{sys}} \circ \gamma(t)$ is larger than a constant $k(\epsilon')$ at which the curves in $C_i(\gamma(t))$ can intersect.

Both the ϵ'-thick part of $\mathcal{T}_g$, $\mathcal{T}_g^{\epsilon'}$, and the $k(\epsilon')$-thick part $\mathcal{T}_g^{k(\epsilon')}$ are compact modulo the action of Γ_g. In the closure of $\mathcal{T}_g^{\epsilon'} \setminus \mathcal{T}_g^{\epsilon_M}$, there is therefore a uniform lower bound on the rate at which f_{sys} is increasing along a flowline. Similarly for the closure of $\mathcal{T}_g^{k(\epsilon')} \setminus \mathcal{T}_g^{\epsilon_M}$. $\square$

In [24], the following claim was made.

Claim 6. *For any $\epsilon > 0$ there is a $T(\epsilon)$ such that flowing for $t > T(\epsilon)$ ensures that any K is carried inside - and remains within - $\mathcal{P}_{g,\epsilon}$.*

In [24], the reference to "flowing" refers to the flow generated by the specific vector field X_ϵ. In Section 5, it will be discussed what flows satisfy this claim and what flows do not. For the moment Claim 6 will be assumed.

Back to the construction of ϕ_t. For any $t \in [0, \infty)$, the isotopy ϕ_t takes a point to its image at time t under the flow. Denote by $\mathcal{I}_t$ the image of $\mathcal{T}_g^{\epsilon_M}$ under ϕ_t.

Suppose $\epsilon'(t)$ has been chosen small enough to ensure that the zeros of $X_{\epsilon'(t)}$ are contained in the interior of $\phi_t(\mathcal{T}_g^{\epsilon_M})$. The boundary of $\phi_t(\mathcal{T}_g^{\epsilon_M})$ is similar to the boundary of a level set of f_{sys} such as $\mathcal{T}_g^{\epsilon_M}$ in the sense that $X_{\epsilon'(t)}$ points inward at every point of $\phi_t(\mathcal{T}_g^{\epsilon_M})$. One way of proving this is to use that ϕ_t gives a flowline-preserving diffeomorphism of a regular neighbourhood of $\partial\mathcal{T}_g^{\epsilon_M}$ onto a regular neighbourhood of $\partial\phi_t(\mathcal{T}_g^{\epsilon_M})$. A point on a flowline outside $\mathcal{T}_g^{\epsilon_M}$ is mapped to a point on a flowline outside of $\partial\phi_t(\mathcal{T}_g^{\epsilon_M})$, and vice versa. Since the flowlines determine a foliation of the regular neighbourhood of $\partial\mathcal{T}_g^{\epsilon_M}$, this is also the case for $\partial\phi_t(\mathcal{T}_g^{\epsilon_M})$. This implies that there are no places where a flowline is tangent to $\partial\phi_t(\mathcal{T}_g^{\epsilon_M})$, which would need to exist for some value of t if $X_{\epsilon'(t)}$ were to transition from pointing inwards to pointing outwards.

Suppose ϵ is small enough to ensure that $\mathcal{P}_{g,\epsilon}$ is contained in a regular neighbourhood $\mathcal{N}_g$ of $\mathcal{P}_g$. Choose t^* such that the isotopy ϕ_{t^*} maps every choice of K into $\mathcal{P}_{g,\epsilon}$. Existence of such a t^* is guaranteed by Claim 6. A deformation retraction mapping K onto $\mathcal{P}_g$ is obtained by taking a composition of ϕ_{t^*} with a deformation retraction that arises from the deformation retraction of $\mathcal{N}_g$ onto $\mathcal{P}_g$.

The existence of this second deformation retraction will now be shown. For ease of notation, it will be shown that $\mathcal{I}_{t^*} = \phi_{t^*}(\mathcal{T}_g^{\epsilon_M})$ deformation retracts onto $\mathcal{P}_g$. An identical argument works with the α-thick part of $\mathcal{T}_g$ in place of $\mathcal{T}_g^{\epsilon_M}$, for any α small enough such that the α-thick part of $\mathcal{T}_g$ contains $\mathcal{P}_g$. As the α-thick subsets are an exhaustion of $\mathcal{T}_g$ by sets compact modulo the action of Γ_g, this will then give the required deformation retraction of $\mathcal{T}_g$.

First note that the boundary of $\mathcal{I}_{t^*}$ is connected. This is because, as shown in Proposition 12.10 of [6], $\partial\mathcal{T}_g^{\epsilon_M}$ is connected and by Proposition 3 there are no critical points of f_{sys} between $\partial\mathcal{I}_{t^*}$ and $\partial\mathcal{T}_g^{\epsilon_M}$. By construction, the set $\mathcal{I}_{t^*}$ has $\mathcal{P}_g$ in the interior, because a flowline is prevented from actually reaching $\mathcal{P}_g$ by the fact that for any ϵ, X_ϵ is zero at points sufficiently close to $\mathcal{P}_g$. Consequently, $\mathcal{I}_{t^*}$ is a connected subset of $\mathcal{N}$ with a connected boundary that separates $\partial\mathcal{N}$ from $\mathcal{P}_g$. The deformation retraction of the regular neighbourhood $\mathcal{N}$ onto $\mathcal{P}_g$ then gives the required deformation retraction of $\mathcal{I}_{t^*}$ onto $\mathcal{P}_g$.

This concludes the proof of the existence of a Γ_g-equivariant deformation retraction of $\mathcal{T}_g$ onto $\mathcal{P}_g$, modulo the proof of the Claim 6.

Remark 7. *In subsequent work, pre-images of points on $\mathcal{P}_g$ under the deformation retraction will be used to study possible further deformation retractions. For a point x in the interior of a locally top-dimensional cell of $\mathcal{P}_g$, this deformation retraction can be performed in such* *a way that it is not difficult to see that the pre-image of x is a ball of codimension equal to the dimension of the cell, and intersecting $\mathcal{P}_g$ in the single point x. This will now be explained.*

For small enough ϵ, $\mathcal{P}_{g,\epsilon}$ can be treated as a tubular neighbourhood of $\mathcal{P}_g$. As a semi-analytic subset of $\mathcal{T}_g$, $\mathcal{P}_g$ has properties in common with embedded submanifolds, that guarantee existence of an analogue of tubular neighbourhoods. The technicalities are discussed in detail in [9]. It follows that there exists a deformation retraction of $\mathcal{P}_{g,\epsilon}$ onto $\mathcal{P}_g$ for which the pre-image of x is a ball $B(x)$ of codimension equal to the dimension of the cell. This ball intersects $\mathcal{P}_g$ only in the point x. The pre-image of x in $\mathcal{T}_g$ under the deformation retraction of $\mathcal{T}_g$ to $\mathcal{P}_g$ is then the ball obtained as the union of $B(x)$ with the portion of the flowlines of X_ϵ obtained by starting on a point of $\partial B(x)$ and flowing backwards in time. Note that by construction, flowing $\partial B(x)$ backwards in time does not introduce intersections with $\mathcal{P}_g$.

4. Questions raised in the literature

In the last two pages of [11], a list of questions about the construction in [24] was made. All but one of these questions were answered in the exposition above. The final (and main) objection, given on page 14 is as follows. Point 4 on page 13 of [11] states "Use the flow defined by the vector field X_ϵ to deform points of $\mathcal{T}_g$ into a neighbourhood $P_{B\epsilon}$ of P [the Thurston spine]." Below is stated "Certainly, there is no problem to deform any compact subset K of $\mathcal{T}_g - P_{B\epsilon}$ into $P_{B\epsilon}$ in a fixed time, but we need to deform the whole space....If it can be shown that points in $P_{B\epsilon}$ cannot be flowed out of $P_{B\epsilon}$, then it is fine, and Step (4) is valid. In summary, for this method in [24] to succeed, we need vector fields whose flows increase the number of geodesics whose lengths are close to the systole of the surface, rather than only increasing their lengths simultaneously." The objection here is towards Claim 6.

The number of curves at x of length close to $f_{\text{sys}}(x)$ is not the right measure of complexity to use in this context. Consider for example a stratum on which the systoles intersect but do not fill. At any point on the boundary of this stratum the number of systoles is larger than in the interior. A vector field whose flow increases the number of curves with length close to f_{sys} would have a zero inside the stratum. The statement "In summary, for this method in [24] to succeed, we need vector fields whose flows increase the number of geodesics whose lengths are close to the systole of the surface, rather than only increasing their lengths simultaneously" is false.

Instead of working with the *number* of geodesics whose lengths are close to that of the systoles, a topologically more reasonably function to work with is defined as follows.

Definition 8 (Gap function g_{sys}). *The gap function $g_{\text{sys}} : \mathcal{T}_g \to \mathbb{R}_+$ is the function whose value at the point x is given by the smallest real number r such that the set of curves with lengths within r of f_{sys} fill $\mathcal{S}_g$.*

The gap function is zero only on $\mathcal{P}_g$, and near $\mathcal{P}_g$ behaves like a measure of distance from $\mathcal{P}_g$.

While it is possible to construct vector fields (for example, the choice made by Thurston, see Proposition 6.3 of [10]) satisfying the three conditions listed at the beginning of Section 5, and for which g_{sys} is strictly decreasing near $\mathcal{P}_g$, even this more reasonable sufficient condition is not essential for Thurston's construction.

Introducing g_{sys} has led to the question, why not just construct a vector field that decreases g_{sys}? One reason is that g_{sys} is only well-behaved on a neighbourhood of $\mathcal{P}_g$; globally it is quite degenerate. For example, there are points at which there is a choice of the longest curve in a shortest filling set. At points like these, the cone of increase of g_{sys} consists of the unions of cones of increase for the different choices. The cone of increase can become disconnected when the zero vector is removed. Unlike f_{sys}, g_{sys} is not a topological Morse function.

5. The claim

Assuming the flow generated by Thurston's vector field, Claim 6 states that for any $\epsilon' > 0$ there is a $T(\epsilon')$ such that flowing for $t > T(\epsilon')$ ensures that any K is carried inside — and remains within — $\mathcal{P}_{g,\epsilon'}$. It was shown in Proposition 6.3 of [10] that on a neighbourhood of $\mathcal{P}_g$, the vector field not only increases f_{sys} but also decreases g_{sys}. As already discussed in Section 4, this implies Claim 6.

The purpose of this section is to discuss what vector fields (other than Thurston's specific choice) generate flows to which the same statement also applies. Perhaps one reason this claim has been controversial is that it was not understood which vector fields it applies to and which ones it does not apply to. Thurston merely gave an example of a vector field that works, mentioned that his choice was not unique, and left it up to the audience or reader to do the calculus needed to verify the details. In order to resolve this confusion, a smooth vector field X'_ϵ will now be constructed that does not satisfy an analogue of the claim. Please note that X'_ϵ is different from Thurston's vector field X_ϵ.

(1) The function f_{sys} is increasing in the direction of X'_ϵ.
(2) The zeros of X'_ϵ are contained in a neighbourhood of $\mathcal{P}_g$ that shrinks onto $\mathcal{P}_g$ as ϵ approaches zero.
(3) X'_ϵ is mapping class group-equivariant.

for which the flow of X'_ϵ does not satisfy Claim 6.

Example 9. *In the final section of* [10] *an example of a 3-dimensional stratum* $\text{Sys}(\{c_1, c_2, c_3, c_4\})$ *in* $\mathcal{P}_2$ *was given, where the set of curves* $C = \{c_1, c_2, c_3, c_4\}$ *is shown in Figure 1.*

The set C *is minimal in the sense that removing any curve gives a set that does not fill. Consequently,* $\text{Sys}(C)$ *is not on the boundary of any larger dimensional stratum of* $\mathcal{P}_g$*. It was shown in Section 3 of* [10] *that, away from isolated critical points, for any point of* $\text{Sys}(C)$ *the gradients of the lengths of curves in* C *are linearly independent and there is an open cone of directions in which* f_{sys} *is increasing. It follows that each proper subset of* C *realises a*

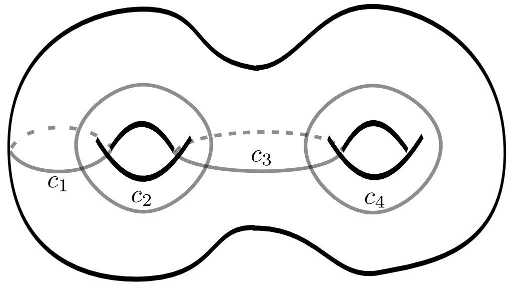

FIGURE 1. The set of systoles in the stratum in Example 9 is given by $\{c_1, c_2, c_3, c_4\}$.

stratum adjacent to $\mathrm{Sys}(C)$.

Let c_1^* *be a curve that intersects* c_1 *but is disjoint from every curve in the set* $C \setminus \{c_1\}$.

For $c_1 \in C$, *denote by* V *the vector field of unit length in the direction in which the twist parameter around* c_1^* *is increasing, and suppose* X_ϵ *is any vector field constructed as in the previous section. Also recalling the partition of unity in the previous section, define* λ *to be* $\sum \lambda_{C'}$ *where the sum is taken over all subsets* C' *of* $C \setminus \{c_1\}$. *As* $U_{\{c_2,c_3,c_4\}}$ *is open in* $\mathcal{T}_g$, *there are points in* $U_{\{c_2,c_3,c_4\}}$ *at which* $\nabla L(c_1)$ *has nonzero projection onto* V. *If there is a point* x *at which* $L(c_1)$ *is increasing in the direction of* V, *define* $X'_\epsilon = X_\epsilon + r\lambda V$, *where* $r \in \mathbb{R}^+$ *is chosen large enough to ensure that in a neighbourhood of* x, $L(c_1)$ *is increasing in the direction of* X'_ϵ *faster than the lengths of any of the other curves in* C. *If there is no point* x *in* $U_{\{c_2,c_3,c_4\}}$ *at which* $L(c_1)$ *is increasing in the direction of* V, *define* $X'_\epsilon = X_\epsilon - r\lambda V$, *where* $r \in \mathbb{R}^+$ *is chosen large enough to ensure that in a neighbourhood of* x, $L(c_1)$ *is increasing in the direction of* X'_ϵ *faster than the lengths of any of the other curves in* C.

Since the lengths of the curves $C \setminus \{c_1\}$ *are stationary along* V, *the vector field* X'_ϵ *increases both* f_{sys} *and the difference in length between the curves* $\{c_2, c_3, c_4\}$ *and* c_1, *increasing* ϵ *in the definition of* $\mathcal{P}_{g,\epsilon}$. *Moreover, the zeros of* X'_ϵ *coincide with the zeros of* X_ϵ. *The vector field* X'_ϵ *is mapping class group-equivariant whenever* $r \in \mathbb{R}^+$ *is chosen consistently.*

References

[1] H. Akrout. Singularités topologiques des systoles généralisées. *Topology*, 42(2):291–308, 2003.
[2] A. Barvinok. Lattice points, polyhedra, and complexity. In *Geometric combinatorics*, volume 13 of *IAS/Park City Math. Ser.*, pages 19–62. American Mathematical Society, Providence, RI, 2007.
[3] L. Bers. Nielsen extensions of Riemann surfaces. *Annales Academiæ Scientiarum Fennicæ, Series A. I. Mathematica*, 2:29–34, 1976.
[4] M. Brion and M. Vergne. Lattice points in simple polytopes. *Journal of the American Mathematical Society*, 10(2):371–392, 1997.
[5] S. Broughton. The equisymmetric stratification of the moduli space and the Krull dimension of mapping class groups. *Topology and its applications*, 37:101–113, 1990.
[6] B. Farb and D. Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[7] J. Harer. The virtual cohomological dimension of the mapping class group of an orientable surface. *Inventiones Mathematicae*, 84:157–176, 1986.
[8] J. Harer. The cohomology of the moduli space of curves. In *Theory of moduli (Montecatini Terme, 1985)*, volume 1337 of *Lecture Notes in Mathematics*, pages 138–221. Springer Berlin, 1988.
[9] K. Hofmann. *Triangulation of locally semi-algebraic spaces*. ProQuest LLC, Ann Arbor, MI, 2009. Thesis (Ph.D.)–University of Michigan.
[10] I. Irmer. The Morse-Smale property of the Thurston spine. *SIGMA*, 21(027), 2025.
[11] L. Ji. Well-rounded equivariant deformation retracts of Teichmüller spaces. *L'Enseignement Mathématique*, 60(02):109–129, 2013.
[12] S. Kerckhoff. The Nielsen realization problem. *Annals of Mathematics*, 117(2):235–265, 1983.
[13] J. Lee. *Introduction to smooth manifolds, 2nd Edition*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, 2013.
[14] S. Lojasiewicz. Triangulation of semi-analytic sets. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, 18(4):449–474, 1964.
[15] B. Martelli. An introduction to geometric topology. arXiv:1610.02592, 2022.
[16] M. Mirzakhani. Growth of the number of simple closed geodesics on hyperbolic surfaces. *Annals of Mathematics. Second Series*, 168(1):97–125, 2008.
[17] M. Morse. Topologically non-degenerate functions on a compact n-manifold M. *Journal d'Analyse Mathématique*, 7:189–208, 1959.
[18] K. Ohshika and A. Papadopoulos. A glimpse into Thurston's work. In *In the tradition of Thurston—geometry and topology*, pages 1–58. Springer, Cham, [2020] ©2020.
[19] K. Ohshika and A. Papadopoulos. Introduction. In *In the tradition of Thurston II. Geometry and groups*, pages 1–5. Springer, Cham, [2022] ©2022.
[20] K. Ohshika and A. Papadopoulos. Introduction. In *In the tradition of Thurston III—geometry and dynamics*, pages 1–5. Springer, Cham, [2024] ©2024.
[21] A. Papadopoulos and G. Théret. Shortening all the simple closed geodesics on surfaces with boundary. *Proceedings of the American Mathematical Society*, 138(138):1775–1784, 2010.
[22] R. Penner. The decorated Teichmüller space of punctured surfaces. *Communications in Mathematical Physics*, 113(2):299–339, 1987.
[23] G. Riera. A formula for the Weil-Petersson product of quadratic differentials. *Journal d'Analyse Mathématique*, 95:105–120, 2005.
[24] W. Thurston. A spine for Teichmüller space. Preprint, 1985.
[25] S. Wolpert. Geodesic length functions and the Nielsen problem. *Journal of Differential Geometry*, 25(2):275–296, 1987.

SUSTech International Center for Mathematics, Southern University of Science and Technology, Shenzhen, China

Department of Mathematics, Southern University of Science and Technology, Shenzhen, China

Email address: ingridmary@sustech.edu.cn