DIVERGENCE OF GEODESICS IN TEICHMÜLLER SPACE AND THE MAPPING CLASS GROUP

MOON DUCHIN AND KASRA RAFI

Abstract. We show that both Teichmüller space (with the Teichmüller metric) and the mapping class group (with a word metric) have geodesic divergence that is intermediate between the linear rate of flat spaces and the exponential rate of hyperbolic spaces. For every two geodesic rays in Teichmüller space, we find that their divergence is at most quadratic. Furthermore, this estimate is shown to be sharp via examples of pairs of rays with exactly quadratic divergence. The same statements are true for geodesic rays in the mapping class group. We explicitly describe efficient paths "near infinity" in both spaces.

1. Introduction

The volume of a ball in Teichmüller space grows exponentially fast as function of its radius, as in the case of hyperbolic space. In this paper, we show that despite this, the "circumference" of the ball grows only quadratically. To be precise, for two geodesic rays γ_1 and γ_2 in a proper geodesic space X with a common basepoint $x \in X$, let their *divergence* be the infimal length of all paths connecting $\gamma_1(t)$ to $\gamma_2(t)$ which maintain distance at least t from the basepoint:

$$\operatorname{div}(\gamma_1, \gamma_2, t) = \operatorname{dist}_{X \setminus B_t(x)}\big(\gamma_1(t), \gamma_2(t)\big),$$

where $B_t(x)$ is the open ball of radius t about x. Note that div may be infinite in this generality.

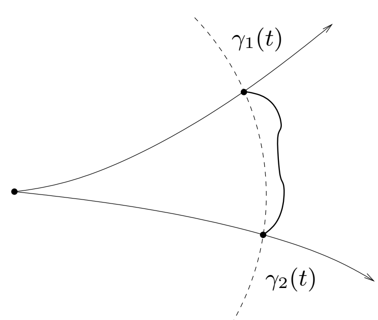

Figure 1. A path connecting $\gamma_1(t)$ to $\gamma_2(t)$ that stays outside the open ball of radius t centered at the basepoint.

This definition is formulated to realize the trichotomy that positive, zero, and negative curvature correspond to sublinear, linear, and superlinear divergence of geodesics, respectively; the correspondence is discussed further below. The main

Date: October 16, 2008.

results of this paper provide an upper bound for the divergence of geodesics in Teichmüller space and the mapping class group for certain surfaces of finite type.

Throughout this article, we compare non-negative functions $f(t)$ and $g(t)$ using the symbols $\stackrel{+}{\asymp}, \stackrel{\cdot}{\asymp}, \asymp, \stackrel{+}{\succ}, \stackrel{\cdot}{\succ}, \succ, \stackrel{+}{\prec}, \stackrel{\cdot}{\prec}, \prec$ to denote equality or inequality with respect to an additive constant, a multiplicative constant, or both, respectively, where the constants depend on the topology of S only. For example, $f(t) \asymp g(t)$ means that there are constants c_1 and c_2 depending only on the topology of S such that

$$\frac{1}{c_1} f(t) - c_2 \le g(t) \le c_1 f(t) + c_2 \quad \forall t.$$

We may refer to this as the order of a function; for instance if $f(t) \asymp t^2$ we may say that f is on the order of t^2. Note that sometimes the distinction between $\stackrel{+}{\prec}$ and $\stackrel{\cdot}{\prec}$ is crucial, as in many applications of Theorem 2.3 below. Note also that although much of the exposition is streamlined by this notation, it is sometimes necessary to pay attention to the constants, as when considering functions $f(t) \asymp t^2$ and $g(t) \asymp t^2$ and trying to prove that $f(t) - g(t) \asymp t^2$.

With this equivalence relation, any two linear functions (respectively, polynomial of degree n) are identified. Having $f \asymp 1$ means the function is bounded above.

Theorem A. *Let S be a surface of genus g with p punctures, such that $3g+p > 4$. Let X be either the Teichmüller space $\mathcal{T}(S)$ with the Teichmüller metric or the mapping class group $\mathrm{Mod}(S)$ with a word metric from a finite generating set. For any pair of geodesic rays $\gamma_1(t)$,$\gamma_2(t)$ with a common basepoint $x \in X$,*

$$\mathrm{div}(\gamma_1, \gamma_2, t) \prec t^2.$$

To accomplish the quadratic upper bound, we explicitly construct paths that travel through chains of product regions. In the Teichmüller case, this path travels through the *thin part* of $\mathcal{T}(S)$, which is stratified into regions which have a product structure, up to additive distortion. Estimating the length of the path uses a combinatorial formula for distance in Teichmüller space (Theorem 2.4). In the case of the mapping class group, we use the quasi-isometrically embedded copies of $\mathbb{Z}^2$ generated by pairs of Dehn twists about disjoint curves.

This theorem does not provide a quadratic or even linear lower bound for all divergence rates in either space X, and indeed none is possible since in both cases there are non-diverging pairs that pass every threshold separation. That is, based at every point in Teichmüller space or the mapping class group, and for arbitrarily large $M > 0$, there are pairs of rays with $\limsup \mathrm{div}(\gamma_1, \gamma_2, t) = M$. One constructs these examples in Teichmüller space from pairs of quadratic differentials with the same underlying topology (see [Mas75]); in the mapping class group, one uses the many undistorted copies of $\mathbb{Z}^2$.

On the other hand, the upper bound is sharp.

Theorem B. *For each X as above, there exist pairs of rays γ_1, γ_2 at every basepoint such that $\mathrm{div}(\gamma_1, \gamma_2, t) \asymp t^2$.*

Examples realizing the quadratic rate are obtained from axes of pseudo-Anosov mapping classes. The lower bound on the divergence is furnished by means of *quasi-projection* theorems to these axes; in Teichmüller space, this is an immediate application of a result of Minsky, while in the mapping class group an analogous

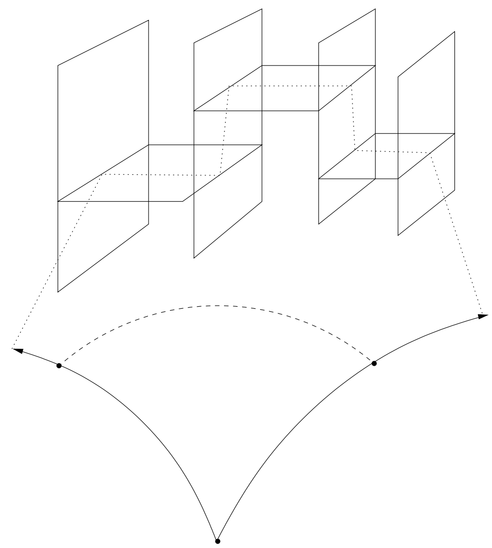

FIGURE 2. An efficient path chains through product regions in the space.

result is obtained using hyperbolicity of the curve complex. (See [Beh06] for a different treatment of divergence in the case of the mapping class group.)

We remark that when $3g + p = 4$ (that is, when S is either a once-punctured torus or a four-times-punctured sphere), the Teichmüller metric on $\mathcal{T}(S)$ has negative curvature and the main theorem clearly does not hold. In higher complexity, Teichmüller space is not hyperbolic ([MW95], [Min96a]). However, there is a long-standing analogy between the geometry of Teichmüller space and that of a hyperbolic space ([Ber78], [Ker80], [Mas81], [Wol75], [Min96b]). This paper provides another point of view from which the Teichmüller space is quite different from a hyperbolic space.

These results should also be regarded as showing that $\mathcal{T}(S)$ and $\mathrm{Mod}(S)$ have *intermediate divergence* in the following sense. A proper geodesic space X is said to have *exponential divergence* if there exists a threshold value $D > 0$ such that for all rays γ_1, γ_2,

$$\mathrm{div}(\gamma_1, \gamma_2, t_0) > D \text{ for some } t_0 \implies \mathrm{div}(\gamma_1, \gamma_2, t) \succ e^t.$$

Gromov hyperbolicity (δ–hyperbolicity) is then equivalent to exponential divergence ($D = 3\delta$ will suffice; see [BH99]). On the other hand, in flat spaces, every two geodesic rays have a linear divergence function.

A geodesic space can be said to have *intermediate divergence* if no rays diverge faster than $f(t)$, and some diverge at the rate $f(t)$, for a function growing superlinearly but subexponentially.

In [Ger94],[Ger94b], Gersten attributes to Gromov the expectation that there should be no nonpositively curved spaces of intermediate divergence. However, Gersten constructed such an example, giving a finite CAT(0) 2–complex whose universal cover possesses two geodesic rays which diverge quadratically and such that no pair of rays diverges faster than quadratically. Furthermore, in those papers, Gersten introduced a variation on divergence which appeals directly to the metric with no reference to geodesics. (He measures the lengths of paths staying far from a basepoint x between pairs of points at fixed distance from x, with a slightly weaker equivalence relation on rates of growth than the one given here.) The advantage of this approach is that it produces a quasi-isometry invariant, so that the divergence of a finitely generated group can be discussed without specifying a generating set. Gersten and Kapovich-Leeb ([Ger94b],[KL98]) study divergence in 3-manifold groups and find that some (namely graph-manifold groups) have quadratic divergence in the same sense as above. Here, we show that the same behavior occurs in the spaces $\mathcal{T}(S)$ and $\mathrm{Mod}(S)$, which have been extensively studied in their own right. Note that our Theorem A implies quadratic divergence in Gersten's coarse sense, for the word metric with respect to any finite generating set.

By contrast, symmetric spaces have a gap in the possible orders of $\mathrm{div}(-,-,t)$ between linear and exponential rates; there, quadratic divergence never occurs for any pair of rays [Gro93]. Thus this paper also highlights the limitations of the long-standing analogies between $\mathcal{T}(S)$ and symmetric spaces.

Acknowledgments. The authors would like to thank the Fields Institute for hosting the workshop at which these results were first conceived. We also thank Howard Masur, Yair Minsky, Saul Schleimer, and Dylan Thurston for very helpful conversations during the conference. In particular, Saul Schleimer observed that the upper-bound arguments for Teichmüller space would essentially carry over to the mapping class group. Thanks also to Misha Kapovich, Rohit Thomas, and an anonymous referee for helpful comments on the paper.

2. Background

Let S be an orientable, connected topological surface of genus g with p punctures. Throughout this paper, we assume that $3g+p>4$. In this section, we review some background material and establish a few lemmas.

The space of curves and arcs. Let Y be an essential subsurface of S (if it is a proper subsurface, then it has boundary, and we write ∂Y for that set of curves). By a *curve* in Y we mean a non-trivial, non-peripheral, simple closed curve in Y and by an *essential arc* we mean a simple arc, with endpoints on the boundary of Y, that cannot be pushed to the boundary of Y. In case Y is not an annulus, the homotopy class of an arc is considered relative to the boundary of Y; when Y is an annulus, the homotopy class of an arc is considered relative to the endpoints of the arc. We will use D_α to denote the Dehn twist about a simple closed curve α.

Let $\mathcal{C}(Y)$ be the set of all homotopy classes of curves and essential arcs on the surface Y. We define a distance on $\mathcal{C}(Y)$ as follows: for $\alpha,\beta\in\mathcal{C}(Y)$, define $d_Y(\alpha,\beta)$ to be equal to one if $\alpha\neq\beta$ and α and β can be represented by disjoint curves or arcs. The metric on $\mathcal{C}(Y)$ is the maximal metric having the above property. Thus, $d_Y(\alpha,\beta)=n$ if $\alpha=\gamma_0,\gamma_1,\ldots,\gamma_n=\beta$ is some shortest sequence of curves or arcs on S such that successive γ_i are disjoint. Note that the notation $\mathcal{C}(Y)$ is often used for the complex of curves on Y (see [MM99] for definitions). Instead, we use $\mathcal{C}(Y)$ to denote the zero-skeleton of the complex of curves and arcs, with distance induced by the one-skeleton.

In [MM99] Masur and Minsky show that $\mathcal{C}(S)$ with the above metric is δ–hyperbolic in the sense of Gromov, and consequences of this are discussed further below. The curve complex is well known (and often proved) to be connected. The isometry group of $\mathcal{C}(S)$ is the mapping class group $\mathrm{Mod}(S)$ of S, as shown by Ivanov, Korkmaz, and Luo.

Subsurface projections. For α a curve in S, we define the *subsurface projection* of α to the essential subsurface Y as follows: Let

$$f : \bar{S} \to S$$

be a regular covering of S such that $f_*(\pi_1(\bar{S}))$ is conjugate to $\pi_1(Y)$ (this is called the Y–cover of S). Since S admits a hyperbolic metric, $\bar{S}$ has a well-defined boundary at infinity, and we use the same notation $\bar{S}$ to denote the cover with its boundary added when appropriate. Let $\bar{\alpha}$ be the lift of α to $\bar{S}$. Components of $\bar{\alpha}$ that are essential arcs or curves on $\bar{S}$, if any, form a subset of $\mathcal{C}(\bar{S})$. The surface $\bar{S}$ is homeomorphic to Y. We call the corresponding subset of $\mathcal{C}(Y)$ the *subsurface projection* of α to Y and will denote it by α_Y. If there are no essential arcs or curves in $\bar{\alpha}$, then α_Y is the empty set; otherwise we say that α *intersects* Y *essentially.* This projection only depends on α up to homotopy.

Let α and α' be curves in S that intersect a subsurface Y essentially. We define the *projection distance* between α and α' to be the maximum distance in $\mathcal{C}(Y)$ between the elements of the projections α_Y and α'_Y, and denote it by $d_Y(\alpha, \alpha')$. If Y is an annulus whose core is the curve γ, then we may denote $\mathcal{C}(Y)$ by $\mathcal{C}(\gamma)$ and $d_Y(\alpha, \alpha')$ by $d_\gamma(\alpha, \alpha')$.

Many of the results in the Masur-Minsky papers on the geometry of the curve complex proceed by analyzing the structure of $\mathcal{C}(S)$ in terms of subsurfaces Y by means of subsurface projection. For instance, here is a useful result from [MM99].

Theorem 2.1 (Theorem 3.1 in [MM99])**.** *There is a constant M_0 such that if the projection distance to some subsurface Y satisfies*

$$d_Y(\alpha, \beta) \geq M_0,$$

then any geodesic in $\mathcal{C}(S)$ connecting α to β intersects the 1–neighborhood of ∂Y.

Markings. Following Thurston, a *pants decomposition* on S is a maximal collection of disjoint curves; these $3g - 3 + n$ curves are called the *pants curves* of that decomposition. The *Fenchel-Nielsen coordinates* on $\mathcal{T}(S)$ are obtained by assigning a length and a twist coordinate to each curve in a pants decomposition. Alternatively, a second set of curves may be chosen, transverse to the first, so that lengths of all $6g - 6 + 2n$ curves give approximate coordinates on $\mathcal{T}(S)$. Along these lines, we define a *marking* on S to be a set of pairs of curves $\mu = \{(\alpha_1, \beta_1), \ldots, (\alpha_m, \beta_m)\}$ such that $\{\alpha_1, \ldots, \alpha_m\}$ is a pants decomposition of S, the curve β_i is disjoint from α_j when $i \neq j$, and intersects α_i once (respectively, twice) if the surface filled by α_i and β_i is a once-punctured torus (respectively, a four-times-punctured sphere). The pants curves α_i are called the base curves of the marking μ. For every i, the corresponding β_i is called the transverse curve to α_i in μ. When the distinction between the base curves and the transverse curves is not important, we represent a marking as simply a set of $2m$ curves. Denote the space of all markings on S by $\mathcal{M}(S)$.

Still following [MM00], there are two types of *elementary moves* in $\mathcal{M}(S)$.

(1) Twist: Replace β_i by β'_i, where β'_i is obtained from β_i by a Dehn twist or a half twist around α_i.

(2) Flip: Replace the pair (α_i, β_i) with (β_i, α_i); also "clean up" by, for $j \neq i$, replacing β_j with a curve β_j' that does not intersect β_i (the new base curve) in such a way that $d_{\alpha_j}(\beta_j, \beta_j')$ is as small as possible (see [MM00] for details).

In the first move, a twist can be positive or negative. A half twist is allowed when α_i and β_i intersect twice. Masur-Minsky analyzed the geometry of the curve complex by finding efficient paths of markings through elementary moves. The following theorem is a version of their result in which the high powers of Dehn twists are rearranged to appear consecutively; this adaptation facilitates computations of Teichmüller distance through changes of marking.

Theorem 2.2 ([Raf07])**.** *There exists a constant* K *(depending on* S*) such that for any two markings* η_1, η_2 *on* S*, there is a path of markings*

$$\eta_1 = \mu_1, \ldots, \mu_n = \eta_2,$$

where μ_i *and* μ_{i+1} *differ by an elementary move except that, for each* α *with* $d_\alpha(\eta_1, \eta_2) \geq \mathsf{K}$*, there is a unique index* i_α *such that*

$$\mu_{i_\alpha+1} = D_\alpha^p \mu_{i_\alpha} \quad \textit{with} \quad |p| \stackrel{+}{\asymp} d_\alpha(\eta_1, \eta_2).$$

This path is efficient:

$$n \asymp \sum_{Y \subseteq S} \Big[d_Y(\eta_1, \eta_2) \Big]_{\mathsf{K}},$$

where the sum is over non-annulus subsurfaces Y *and* $[N]_{\mathsf{K}} := \begin{cases} N, & N \geq \mathsf{K} \\ 0, & N < \mathsf{K}. \end{cases}$

Product regions in Teichmüller space. Let $\mathcal{T}(S)$ denote the Teichmüller space of S equipped with the Teichmüller metric. A point $\tau \in \mathcal{T}(S)$ is a hyperbolic metric on S (constant curvature -1). Minsky has shown that the thin part of Teichmüller space has a product-like structure, as we now describe. Let Γ be a set of disjoint curves on S and let $\mathrm{Thin}_\epsilon(\Gamma)$ denote the set of points in Teichmüller space such that all curves from Γ are short in hyperbolic length:

$$\mathrm{Thin}_\epsilon(\Gamma) := \Big\{ \tau \in \mathcal{T}(S) \,\Big|\, l_\tau(\gamma) < \epsilon \text{ for all } \gamma \in \Gamma \Big\}.$$

These cover the ϵ*–thin part* of Teichmüller space, which consists of those metrics with some short curve:

$$\mathrm{Thin}_\epsilon := \Big\{ \tau \in \mathcal{T}(S) \,\Big|\, \mathrm{injrad}(\tau) < \epsilon \Big\}.$$

The ϵ*–thick part* is the complement of Thin_ϵ. Let $\mathrm{Prod}(\Gamma)$ denote the product space

$$\mathrm{Prod}(\Gamma) := \mathcal{T}(S \setminus \Gamma) \times \prod_{\gamma \in \Gamma} \mathbb{H}_\gamma,$$

where $S \setminus \Gamma$ is considered as a surface of lower complexity and each $\mathbb{H}_\gamma$ is a horoball in the copy of the hyperbolic plane parametrizing the Fenchel-Nielsen coordinates corresponding to a short curve γ (the x–coordinate in $\mathbb{H}_\gamma$ represents the twist parameter along γ in σ and the y–coordinate represents the reciprocal of the length of γ in σ; see [Min96a]). Endow $\mathrm{Prod}(\Gamma)$ with the sup metric. Minsky has shown, for small enough ϵ, that $\mathrm{Thin}_\epsilon(\Gamma)$ is well-approximated by $\mathrm{Prod}(\Gamma)$.

Theorem 2.3 (Product regions [Min96a]). *The Fenchel-Nielsen coordinates on $\mathcal{T}(S)$ give rise to a natural homeomorphism $\pi\colon \mathcal{T}(S) \to \mathrm{Prod}(\Gamma)$. There exists an $\epsilon_0 > 0$ sufficiently small that this homeomorphism restricted to $\mathrm{Thin}_{\epsilon_0}(\Gamma)$ distorts distances by a bounded additive amount.*

Note that $\mathcal{T}(S \setminus \Gamma) = \prod_Y \mathcal{T}(Y)$, where the product is over all connected components Y of $S \setminus \Gamma$. Let π_0 denote the coordinate factor of π mapping to $\mathcal{T}(S \setminus \Gamma)$, let π_Y denote the coordinate factor mapping to $\mathcal{T}(Y)$, and, for $\gamma \in \Gamma$, let π_γ denote the coordinate factor mapping to $\mathbb{H}_\gamma$.

Short markings and Teichmüller distance. Let σ be a point in the Teichmüller space $\mathcal{T}(S)$ of S. A *short marking* on σ is a marking whose curves are chosen greedily to be as short as possible with respect to hyperbolic length. That is, let α_1 be a simple closed curve of minimal length in σ, α_2 a shortest curve disjoint from α_1, and so on, to form a pants decomposition of S (the *Bers constant* gives an upper bound on the lengths of curves in a greedily chosen pants decomposition). Then, let the transverse curve β_i be the shortest curve intersecting α_i and disjoint from α_j, $i \neq j$. There are only finitely many choices in this process. The following distance formula relates the Teichmüller distance between two points σ_1 and σ_2 to the combinatorics of short markings in σ_1 and σ_2. Let ϵ_0 be as before. Define Γ_{12} to be the set of curves that are ϵ_0–short in both σ_1 and σ_2, and, for $i = 1, 2$, define Γ_i to be the set of curves that are ϵ_0–short in σ_i but not in σ_{3-i}. Let η_1 and η_2 be short markings on σ_1 and σ_2, respectively.

Theorem 2.4 ([Raf07]). *For sufficiently large K, the distance in $\mathcal{T}(S)$ between σ_1 and σ_2 is given by the following formula:*

$$d_{\mathcal{T}}(\sigma_1, \sigma_2) \asymp \sum_Y \Big[d_Y(\eta_1, \eta_2)\Big]_{\mathsf{K}} + \sum_{\alpha \notin \Gamma_{12}} \log \Big[d_\alpha(\eta_1, \eta_2)\Big]_{\mathsf{K}} + \tag{1}$$
$$+ \max_{\alpha \in \Gamma_{12}} d_{\mathbb{H}_\alpha}(\sigma_1, \sigma_2) + \max_{\substack{\alpha \in \Gamma_i \\ i=1,2}} \log \frac{1}{l_{\sigma_i}(\alpha)},$$

where $[N]_{\mathsf{K}} := \begin{cases} N, & N \geq \mathsf{K} \\ 0, & N < \mathsf{K}, \end{cases}$ *as before.*

For the rest of the paper, we fix ϵ_0 so that Minsky's product regions theorem and Theorem 2.4 hold.

Subsurface distances and word length in the mapping class group. Consider the mapping class group $\mathrm{Mod}(S)$ of the surface S. We fix a finite set $\mathbf{A}$ of curves in S such that the set $\{D_\alpha : \alpha \in \mathbf{A}\}$ of Dehn twists around curves in $\mathbf{A}$ generates $\mathrm{Mod}(S)$. (It follows that the curves in $\mathbf{A}$ fill S.) Equip $\mathrm{Mod}(S)$ with the word metric corresponding to this generating set and denote the word length of an element $h \in \mathrm{Mod}(S)$ by $|h|$.

Definition 2.5 (Subsurface distance). Let Y be a subsurface of S and μ_1, μ_2 be two sets of curves on S. We define

$$d_Y(\mu_1, \mu_2) = \max_{\alpha_1 \in \mu_1,\, \alpha_2 \in \mu_2} d_Y(\alpha_1, \alpha_2),$$

where $d_Y(\alpha_1, \alpha_2)$ is as defined above. For $h \in \mathrm{Mod}(S)$, let

$$d_Y(h) = d_Y\big(\mathbf{A}, h(\mathbf{A})\big),$$

$\mathbf{A}$ as above.[1] When Y is an annulus whose core curve is γ, we may write $d_\gamma(h)$ instead of $d_Y(h)$.

Next, we need the following theorem of Masur and Minsky which relates the word length of a mapping class h to the subsurface distances corresponding to h. The sum in the statement of the theorem is broken into two parts (summing over annular and non-annular subsurfaces) to highlight the comparison between this case and that in Theorem 2.4. The theorem essentially says that word length is comparable to the sum of the very large subsurface projections.

Theorem 2.6 (Word length in the mapping class group [MM00])**.** *There is a constant K_0 such that for every threshold $\mathsf{K} \ge \mathsf{K}_0$ there exists c such that for every $h \in \operatorname{Mod}(S)$,*

$$\frac{|h|}{\mathsf{c}} \le \sum_\alpha \Big[d_\alpha(h)\Big]_{\mathsf{K}} + \sum_Y \Big[d_Y(h)\Big]_{\mathsf{K}} \le \mathsf{c}\,|h|, \tag{2}$$

where the first sum is over all curves on S, the second sum is over all subsurfaces of S that are not an annulus or a pair of pants and $[N]_{\mathsf{K}} := \begin{cases} N, & N \ge \mathsf{K} \\ 0, & N < \mathsf{K}. \end{cases}$

We will sometimes refer to such a constant c as the word-length constant (for a threshold K), taking c_0 to be the value corresponding to K_0.

We also need the following few simple lemmas.

Lemma 2.7. *For $h_1, h_2, h_3 \in \operatorname{Mod}(S)$ and any subsurface Y of S, we have the following triangle inequality.*

$$d_Y\big(h_1(\mathbf{A}), h_3(\mathbf{A})\big) \overset{+}{\prec} d_Y\big(h_1(\mathbf{A}), h_2(\mathbf{A})\big) + d_Y\big(h_2(\mathbf{A}), h_3(\mathbf{A})\big)$$

Proof. The set $\mathbf{A}$ has bounded diameter in $\mathcal{C}(S)$, therefore its projection to any subsurface Y also has bounded diameter. The same is true for $h_i(\mathbf{A})$, $i = 1, 2, 3$. But d_Y satisfies the triangle inequality in $\mathcal{C}(Y)$. Therefore, the above inequality is also true with the additive error of at most $3\operatorname{diam}(\mathbf{A})$ in $\mathcal{C}(S)$. □

One may think of d_α as measuring the relative twisting between two curves around α.

Lemma 2.8. *For any curve α in S and any mapping class $h \in \operatorname{Mod}(S)$,*

$$d_\alpha(D_\alpha^n\, h) \overset{+}{\succ} n - \mathsf{c}\,|h|,$$

where c is the word-length constant.

Proof. Using Lemma 2.7, we have

$$\begin{aligned} d_\alpha(D_\alpha^n\, h) = d_\alpha\big(\mathbf{A}, D_\alpha^n\, h(\mathbf{A})\big) &\overset{+}{\succ} d_\alpha\big(h(\mathbf{A}), D_\alpha^n\, h(\mathbf{A})\big) - d_\alpha\big(\mathbf{A}, h(\mathbf{A})\big) \\ &\overset{+}{\succ} n - d_\alpha(h) \end{aligned}$$

Here, the last estimate is valid because $h(\mathbf{A})$ fills, so it contains a curve β such that $d_\alpha(\beta, D_\alpha^n \beta) \overset{+}{\asymp} n$. Finally, Theorem 2.6 implies that $d_\alpha(h) \overset{+}{\prec} \mathsf{c}\,|h|$ and the lemma follows. □

[1]Since $\mathbf{A}$ fills S, $h(\mathbf{A})$ fills S as well for every h; therefore, $h(\mathbf{A})$ intersects every subsurface essentially and the projection of $h(\mathbf{A})$ to Y is always non-empty.

Lemma 2.9. *If α and β are disjoint, then $d_\alpha(D^n_\beta)$ has an upper bound that is independent of n.*

Proof. Let $\bar{S}$ be the annular cover of S with respect to α, let $\bar{\alpha}$ be the lift of α that is a closed curve and $\bar{\mathbf{A}}$ be the set of lifts of the elements of $\mathbf{A}$ which intersect $\bar{\alpha}$. Let γ be a curve in S intersecting α but not β such that $d_\alpha(\mathbf{A}, \gamma) \asymp 1$ (the last condition can be obtained after applying an appropriate power of D_α to γ). Let $\bar{\gamma}$ be a lift of γ to $\bar{S}$ that intersects $\bar{\alpha}$.

Since S is hyperbolic, $\bar{S}$ has a well-defined boundary at infinity. The lifts of β are arcs with endpoints on the boundary; a Dehn twist around β in S lifts to shearing along these arcs in $\bar{S}$. Since $\bar{\gamma}$ is disjoint from all lifts of β, no amount of shearing of $\bar{\mathbf{A}}$ along lifts of β can change the intersection number of $\bar{\mathbf{A}}$ with $\bar{\gamma}$. That is,

$$d_\alpha(\gamma, \mathbf{A}) = d_\alpha(\gamma, D^n_\beta \mathbf{A}) \asymp 1,$$

and therefore

$$d_\alpha(\mathbf{A}, D^n_\beta \mathbf{A}) \leq d_\alpha(\gamma, \mathbf{A}) + d_\alpha(\gamma, D^n_\beta \mathbf{A}) \asymp 1. \qquad \square$$

Lemma 2.10. *For all words $h \in \mathrm{Mod}(S)$, all integers n, m, and all curves α and β such that $i(\alpha, \beta) = 0$, we have*

$$|D^n_\alpha D^m_\beta h| \overset{+}{\succ} \frac{\max(|n|, |m|)}{\mathsf{c}} - |h|$$

for the word-length constant c.

Proof. Without loss of generality, $n \geq m \geq 0$. We have

(Theorem 2.6)	$\lvert D^n_\alpha D^m_\beta h\rvert \overset{+}{\succ} d_\alpha(D^n_\alpha D^m_\beta h)/\mathsf{c}$
(D_α and D_β commute)	$= d_\alpha(D^n_\beta D^m_\alpha h)/\mathsf{c}$
(Lemma 2.9)	$\overset{+}{\asymp} d_\alpha(D^n_\alpha h)/\mathsf{c}$
(Lemma 2.8)	$\overset{+}{\succ} n/\mathsf{c} - \lvert h\rvert.$

This completes the proof. $\square$

Below, we will need to construct paths outside of the ball $B_t(e)$ in $\mathrm{Mod}(S)$. To initiate, it will be necessary to push away from the ball; the following lemma says that for any Dehn twist, a power of either it or its inverse accomplishes this. Let

$$\phi \colon \mathbb{N} \to \mathrm{Mod}(S)$$

be a geodesic ray in $\mathrm{Mod}(S)$ based at the identity.

Lemma 2.11 (Pushing off)**.** *There is a constant $d \in \mathbb{N}$ such that, for any curve $\alpha \in \mathbf{A}$ and any nonnegative integers $n, m, t \in \mathbb{N}$, the magnitudes $|D^n_\alpha \phi(dt)|$ and $|D^{-m}_\alpha \phi(dt)|$ are not both less than or equal to t.*

Proof. Let c be the word-length constant from Theorem 2.6 and set $d > c + 2$. Let $h = \phi(dt)$ and suppose, for contradiction, that

$$|D^n_\alpha h| \leq t \quad \text{and} \quad |D^{-m}_\alpha h| \leq t$$

for some m, n. Then $|D^{m+n}_\alpha| \leq 2t$. But $d_\alpha(D^{m+n}_\alpha) \geq m + n - 1$. Using the word-length constant from Theorem 2.6, we obtain

$$m + n \leq 2\mathsf{c}\, t + 1. \tag{3}$$

Since D_α is a generator for $\mathrm{Mod}(S)$, $|D_\alpha^n| \le n$. Thus

$$t \ge |D_\alpha^n h| \ge |h| - |D_\alpha^n| \ge dt - n.$$

That is, $n \ge (d-1)t$. Similarly, $m \ge (d-1)t$ and

$$m + n \ge 2(d-1)t > 2(\mathsf{c}+1)t = 2\mathsf{c}\,t + 2t > 2\mathsf{c}\,t + 1,$$

which is a contradiction. □

Consequences of hyperbolicity. Recall that a space is called is *δ–hyperbolic* for some $\delta > 0$ if every geodesic triangle is δ–thin: each side is contained in a δ–neighborhood of the union of the other two sides. Also, if a geodesic is regarded as an isometric embedding from (a subinterval of) $\mathbb{R}$ into X, then a *Q–quasi-geodesic* replaces the equality with a coarse equality with additive and multiplicative constant Q. We collect here some standard consequences of hyperbolicity for use later in the paper. (See [BH99] for a reference on hyperbolic spaces and quasi-geodesics.)

For the following three lemmas, let X be a δ–hyperbolic space and let $L \subset X$ be a Q–quasi-geodesic line, ray, or segment. Suppose $a, b \in X$ and $\bar{a}, \bar{b} \in L$ are such that $d(a, \bar{a})$ and $d(b, \bar{b})$ realize the distance from a and b, respectively, to L.

Lemma 2.12 (Thin quadrilaterals)**.** *There exist constants $M_1 = M_1(\delta, Q)$ and $M_2 = M_2(\delta, Q)$ such that if $d(\bar{a}, \bar{b}) > M_1$, then any geodesic from a to b intersects the M_2–neighborhood of L.*

Lemma 2.13 (Bounded shadows)**.** *There exists a constant $M_3 = M_3(\delta, Q)$ such that for any geodesic segment I from a to $\bar{a}$, the closest-point projection of I to L has diameter $\le M_3$.*

Lemma 2.14 (Bounded projection)**.** *There exists a constant $M_4 = M_4(\delta, Q)$ such that $d(\bar{a}, \bar{b}) \le M_4 \cdot d(a, b) + M_4$.*

Below, we will write Proj_L for the closest-point projection to a quasi-geodesic L. Since $\mathrm{Proj}_L(a)$ is of bounded diameter, this is a coarsely well-defined map.

Recalling from above the Masur-Minsky result that $\mathcal{C}(S)$ is δ–hyperbolic, we will reserve the notation δ for this particular hyperbolicity constant.

3. Divergence in Teichmüller space

As before, let $\mathcal{T}(S)$ denote the Teichmüller space of S equipped with the Teichmüller metric. For a quadratic differential q on a Riemann surface of topological type S, let $[q]$ be the corresponding point of $\mathcal{T}(S)$. (That is, $[q]$ is the hyperbolic metric in the conformal class of q; for more background on Teichmüller space and quadratic differentials, see for instance [Str80], [Abi80], or [Hub06].) Take $\sigma \in \mathcal{T}(S)$ and two quadratic differentials q_1 and q_2 such that $[q_1] = [q_2] = \sigma$. Let $q_1(t)$ and $q_2(t)$ be the images of q_1 and q_2, respectively, under the time-t Teichmüller geodesic flow. The maps

$$t \mapsto [q_i(t)], \qquad i = 1, 2$$

from $[0, \infty)$ to $\mathcal{T}(S)$ are geodesic rays in $\mathcal{T}(S)$ emanating from σ. We want to show that for all σ, q_1, q_2 as above and $t > 0$, there is a path from $[q_1(t)]$ to $[q_2(t)]$ in $\mathcal{T}(S)$ with length on the order of t^2 that stays outside of $B_t(\sigma)$, the open ball of radius t in $\mathcal{T}(S)$ centered at σ (see Fig 1).

We will repeatedly use the same argument to show that a point in $\mathcal{T}(S)$ maintains distance at least t from σ: we show that the value $\epsilon(t) = e^{-2t}$ is small enough that any point in $\tau \in \mathcal{T}(S)$ with an ϵ–short curve satisfies $d_{\mathcal{T}}(\tau, \sigma) \geq t$. Then we carry out the appropriate sequence of elementary moves while maintaining some ϵ–short curve at all times.

Constructing a path for the upper bound. To begin the progress from one ray to the other, we push off from the ball $B_t(\sigma)$ of radius t around σ so that the subsequent moves are guaranteed to stay far from σ. It suffices to construct, for sufficiently large t, a path between $\sigma_1 = [q_1(3t)]$ and $\sigma_2 = [q_2(3t)]$ whose length is of order t^2, while controlling the distance from σ. The path will follow a sequence of elementary moves, maintaining a sufficiently short curve at all times in order to ensure that we stay outside of $B_t(\sigma)$.

Fix $\epsilon = e^{-2t}$ and suppose α on S and $\tau \in \mathcal{T}(S)$ satisfy $l_\tau(\alpha) \leq \epsilon$. In [Wol79], it is shown that for a K–quasi-conformal map between Riemann surfaces, the hyperbolic lengths of curves are changed by at most a multiplicative factor of K. Consequently the Teichmüller distance is bounded below by the ratio of hyperbolic lengths for any particular curve.

$$d_{\mathcal{T}}(\sigma, \tau) \geq \frac{1}{2} \log \frac{l_\sigma(\alpha)}{l_\tau(\alpha)} \geq \frac{1}{2} \log \frac{\epsilon_1}{\epsilon} \geq t.$$

That is, this value of ϵ has the property described above that if any curve on a surface τ is ϵ–short, then $\tau \notin B_t(\sigma)$.

We first give the argument under the assumption that the path starts and ends in the ϵ_0–thick part (that is, $\sigma_1, \sigma_2 \notin \mathrm{Thin}_{\epsilon_0}$); we will treat the general case last. Recall that ϵ_0 is chosen as in the product regions theorem.

Let $\mu_1, \ldots, \mu_n$ be the sequence of markings described in Theorem 2.2 such that $\eta_1 = \mu_1$ is a short marking on σ_1 and $\eta_2 = \mu_n$ is a short marking on σ_2. Note that the condition that σ_1 and σ_2 are in the ϵ_0–thick part of Teichmüller space implies that the length of η_i in σ_i is bounded independent of t.

For an elementary move on a marking, let the *associated base curve* be either the twisting curve, if the move is a twist, or the (initial) base curve in the flipped pair, if the move is a flip. For the sequence of markings $\{\mu_i\}$, let α_i be the associated base curve for the move $\mu_i \to \mu_{i+1}$ and let γ_i be any base curve of the marking μ_i which is different from α_i (the complexity condition $3g + p > 4$ implies that there are at least two base curves in every marking). Note that γ_i and γ_{i+1} have intersection number zero. For any marking μ_i and M sufficiently large, the set

$$B_i = \{\tau \in \mathcal{T}(S) : l_\tau(\mu_i) \leq M\} \subset \mathcal{T}(S)$$

is nonempty and has bounded diameter (this follows, again, from the Wolpert formula). Fix such an M, which is suppressed in the notation B_i from here on.

Since we can assume t is sufficiently large, we will proceed taking $t > \max\{\operatorname{diam} B_i\}$ and $t > \log \frac{1}{\epsilon_1}$, where ϵ_1 is the injectivity radius of σ.

Let $\tau^1 = \sigma_1$, $\tau^n = \sigma_2$ and, for $1 < i < n$, choose τ^i to be a point in B_i.

Let $\tau^i_{\gamma_i}$ be the point of $\mathcal{T}(S)$ that has the same Fenchel-Nielsen coordinates as τ^i except that γ_i (which has bounded length in τ^i) is pinched to have length equal to ϵ. Let $\tau^i_{\gamma_{i-1}}$ and $\tau^i_{\gamma_{i-1}\gamma_i}$ be defined similarly. Consider the piecewise geodesic path

P in $\mathcal{T}(S)$ defined by connecting up the points in the sequence below.

$$\sigma_1 = \tau^1 \xrightarrow{p_1} \tau^1_{\gamma_1} \xrightarrow{e_1} \tau^2_{\gamma_1} \xrightarrow{p_2} \tau^2_{\gamma_1\gamma_2} \xrightarrow{\bar{p}_1} \tau^2_{\gamma_2} \xrightarrow{e_2} \tau^3_{\gamma_2} \xrightarrow{p_3} \tau^3_{\gamma_2\gamma_3} \xrightarrow{\bar{p}_2} \tau^3_{\gamma_3} \longrightarrow \cdots$$
$$\cdots \longrightarrow \tau^{n-1}_{\gamma_{n-2}\gamma_{n-1}} \xrightarrow{\bar{p}_{n-2}} \tau^{n-1}_{\gamma_{n-1}} \xrightarrow{e_{n-1}} \tau^n_{\gamma_{n-1}} \xrightarrow{\bar{p}_{n-1}} \tau^n = \sigma_2.$$

There are three kinds of steps:

(p_i) *pinches* the next curve, γ_i, by shortening it to length ϵ;
(e_i) applies an *elementary move* while keeping the curve γ_i short;
($\bar{p}_i$) *releases* the previous curve, γ_i, by restoring it to its length before pinching.

We may think of p_i, $\bar{p}_i$ and e_i as paths in $\mathcal{T}(S)$ which are subpaths of P. We denote the lengths of these paths by $|p_i|$, $|\bar{p}_i|$ and $|e_i|$ respectively. We will bound the lengths of these subpaths and show that they stay outside $B_t(\sigma)$ in order to complete the proof of Theorem A for $\mathcal{T}(S)$.

Except along p_1 and $\bar{p}_{n-1}$, at every point in the path P, at least one curve has length equal to ϵ; therefore, by the choice of ϵ, these points are outside of $B_t(\sigma)$. The lengths of the subpaths p_1 and $\bar{p}_{n-1}$ are (up to an additive error) equal to t and they have one point at distance $3t$ from σ; therefore these subpaths are also outside of $B_t(\sigma)$.

We will estimate the lengths of the p_i and $\bar{p}_i$ by showing that

$$d_{\mathcal{T}}(\tau^i, \tau^i_{\gamma_i}) \stackrel{+}{\asymp} d_{\mathcal{T}}(\tau^i, \tau^i_{\gamma_{i-1}}) \stackrel{\cdot}{\asymp} t,$$

and

$$d_{\mathcal{T}}(\tau^i_{\gamma_{i-1}}, \tau^i_{\gamma_{i-1}\gamma_i}) \stackrel{+}{\asymp} d_{\mathcal{T}}(\tau^i_{\gamma_{i-1}\gamma_i}, \tau^i_{\gamma_i}) \stackrel{\cdot}{\asymp} t,$$

To estimate the distance from τ^i to $\tau^i_{\gamma_i}$, consider first pinching the curve γ_i in τ^i to obtain a new metric $\tau \in \mathcal{T}(S)$, where γ_i has length ϵ_0. The path from τ^i to τ has bounded length because both τ^i and τ are in B_i, which has bounded diameter. Now letting $\Gamma = \{\gamma_i\}$, we note that $\tau, \tau^i_{\gamma_i}$ are both in $\mathrm{Thin}_{\epsilon_0}(\Gamma)$ and Theorem 2.3 applies. But their projections to $S \setminus \Gamma$ are identical. Therefore, their distance, up to an additive error, is equal to the distance in $\mathbb{H}_{\gamma_i}$ between $\pi_{\gamma_i}(\tau^i_{\gamma_i})$ and $\pi_{\gamma_i}(\tau)$, their projections to that factor. The points τ and $\tau^i_{\gamma_i}$ have the same twisting parameters around γ_i (the x–coordinates of their projections to $\mathbb{H}_{\gamma_i}$ are the same), but different length parameters (the y–coordinates are $1/\epsilon_0$ and $1/\epsilon$ respectively). Therefore

$$d_{\mathcal{T}}(\tau^i_{\gamma_i}, \tau) \stackrel{+}{\asymp} d_{\mathbb{H}_{\gamma_i}}\big(\pi_{\gamma_i}(\tau^i_{\gamma_i}), \pi_{\gamma_i}(\tau)\big) = \frac{1}{2}\log\frac{\epsilon_0}{\epsilon} \stackrel{+}{\asymp} t.$$

Next, we estimate the lengths of the e_i by showing that

$$d_{\mathcal{T}}(\tau^i_{\gamma_i}, \tau^{i+1}_{\gamma_i}) \stackrel{+}{\asymp} \begin{cases} 1 & \text{if } \mu_i \text{ and } \mu_{i+1} \text{ differ by an elementary move.} \\ \log p & \text{if } \mu_{i_\alpha+1} = D^p_\alpha \mu_{i_\alpha}, \quad |p| \stackrel{+}{\asymp} d_\alpha(\eta_1, \eta_2) \geq \mathsf{K}. \end{cases}$$

If the elementary move from μ_i to μ_{i+1} is a simple twist or flip, then the length of the segment (e_i) is bounded. This is because $\pi_{\gamma_i}(\tau^i_{\gamma_i})$ and $\pi_{\gamma_i}(\tau^{i+1}_{\gamma_i})$ are within bounded distance, and therefore

$$\begin{aligned} d_{\mathcal{T}}(\tau^i_{\gamma_i}, \tau^{i+1}_{\gamma_i}) &\stackrel{+}{\asymp} d_{S\setminus\gamma_i}\big(\pi_0(\tau^i_{\gamma_i}), \pi_0(\tau^{i+1}_{\gamma_i})\big) \\ &\stackrel{+}{\asymp} d_{S\setminus\gamma_i}\big(\pi_0(\tau^i), \pi_0(\tau^{i+1})\big) \\ &\stackrel{+}{\prec} d_{\mathcal{T}}(\tau^i, \tau^{i+1}) \asymp 1. \end{aligned}$$

But there can also be high powers of twists: for every α where $d_\alpha(\mu_1, \mu_n)$ is large there is an index i_α where $\mu_{i_\alpha+1} = D^p_\alpha \mu_{i_\alpha}$, as in Theorem 2.2. Here the length of (e_i) is on the order of $\log d_\alpha(\mu_1, \mu_n)$. Therefore the total length of P is

$$\sum_i |p_i| + |\bar{p}_i| + |e_i| \asymp n\, t + \sum_\alpha \log \Big[d_\alpha(\mu_1, \mu_n) \Big]_{\mathsf{K}}.$$

But Theorem 2.2 and Theorem 2.4 imply

$$n \asymp \sum_{Y \subseteq S} \Big[d_Y(\mu_1, \mu_n) \Big]_{\mathsf{K}} \leq d_{\mathcal{T}}(\sigma_1, \sigma_2) \leq 6t \ .$$

It follows that the length $|P|$ is at most on the order of

$$(6t)\, t + 6t \stackrel{\cdot}{\asymp} t^2.$$

This tells us in particular that the elementary moves (which move within a single product region) contribute negligibly to the total length of the path when the quadratic rate is realized; the pinch-and-release steps (which pivot from one product region to the next) account for the whole length of the path, asymptotically.

Above, we assumed that the starting and ending points of the path were in the ϵ_0–thick part. This condition was used in the definition of τ^i (for example we have assumed that $\tau^1 \in B_1$). This assumption is not always true. However, we can modify the beginning and the end of the path above to accommodate the general case as follows:

Let τ be a point in $\mathcal{T}(S)$ with the same Fenchel-Nielsen coordinates as σ_1 except that the length of γ_1 is ϵ. Let $\tau^1_{\gamma_1}$ be the point obtained by increasing the lengths of other short curves (lengths less than ϵ_0) to a moderate length. We will show that these paths $P_1 = [\sigma_1, \tau]$ and $P_2 = [\tau, \tau^1_{\gamma_1}]$ are outside of $B_t(\sigma)$ and have length of order t; the rest of the calculation reverts to the arguments above.

Note that there is a lower bound on the σ_1–lengths of all curves. That is, again using [Wol79] and recalling that ϵ_1 is the injectivity radius at σ,

$$\frac{1}{2} \log \frac{l_\sigma(\gamma)}{l_{\sigma_1}(\gamma)} \leq 3t \quad \forall \gamma \quad \Longrightarrow \quad l_{\sigma_1}(\gamma) \geq \frac{l_\sigma(\gamma)}{e^{6t}} \geq \epsilon_1 e^{-6t} \ .$$

Since there is no twisting or other change in the marking, it follows from Theorem 2.4 that both of the paths P_1 and P_2 have length at most on the order of t (the only contribution to distance comes from the length ratios). The path P_2 stays outside of $B_t(\sigma)$ because along this path the curve γ_1 has length ϵ. To see that the path P_1 is outside of $B_t(\sigma)$ we take two cases. If the length of γ_1 in σ_1 is less than ϵ, then it remains less than ϵ along this path and we are done. If $\epsilon_0 \geq l_{\sigma_1}(\gamma_1) \geq \epsilon$, then the length of this path is no more than $\frac{1}{2} \log \frac{\epsilon_0}{\epsilon} \stackrel{+}{\asymp} t$ by the product regions theorem (Theorem 2.3). But $d_{\mathcal{T}}(\sigma, \sigma_1) = 3t$, which completes the argument.

Applying the same modifications to the end of path P, we obtain the desired result for the general case.

Example realizing the quadratic rate. To see that the quadratic estimate is sharp, we must furnish an example of a pair of rays whose divergence rate is exactly quadratic. We will use a *quasi-projection* result of Minsky ([Min96b]) which shows that intervals which are far (relative to their length) from a cobounded geodesic (segment, ray, or line) project to sets whose diameter is uniformly bounded above. Recall that an ϵ–cobounded geodesic in $\mathcal{T}(S)$ is one which stays in the ϵ–thick part. (Note that part of the proof entails that Proj, the closest-point projection, is coarsely well-defined in the following setting.)

Theorem 3.1 (Quasi-projection for $\mathcal{T}(S)$ [Min96b])**.** *For every $\epsilon > 0$ there are constants b_1, b_2 depending on ϵ and the topology of S such that the following holds. Let G be an ϵ–cobounded geodesic in $\mathcal{T}(S)$, suppose $\tau \in \mathcal{T}(S)$ satisfies $d(\tau, G) > b_1$, and let $r = d(\tau, G) - b_1$. Then*

$$\operatorname{diam}\Big(\operatorname{Proj}_G \left(B_r(\tau) \right) \Big) \le b_2.$$

It is straightforward to replace the geodesic in the statement of the theorem with a quasi-geodesic. Also note that it suffices to check a bounded geometry condition for the endpoints of the geodesic (in $\mathcal{T}(S)$ or on the Thurston boundary) to ensure that it stays in the thick part [Raf05].

To apply this theorem, we may for instance choose q and $q' = -q$ to point in opposite directions along the axis of a pseudo-Anosov mapping class. (Coboundedness is guaranteed because pseudo-Anosov axes project to closed curves in moduli space.)

Proposition 3.2. *For any ϵ–cobounded geodesic G in $\mathcal{T}(S)$ and any point $\sigma \in G$, let σ_t and σ_t' be the points on G at distance t from σ. Then for any path P in $\mathcal{T}(S)$ from σ_t to σ_t' which maintains distance at least t from σ,*

$$|P| \succ t^2.$$

Proof. Take G_0 to be the subsegment of G of length t, centered at σ. Then any path P between σ_t and σ_t' outside $B_t(\sigma)$ maintains distance at least $t/2$ from G_0 at all times.

Now if P is a path connecting σ_t and σ_t' in $\mathcal{T}(S)$ of length $|P|$, we can divide it into pieces of length $t/2$, taking $P = I_1 \cup \cdots \cup I_n$ for successive pieces of length $t/2$ (with I_n possibly shorter). The number of these intervals, n, satisfies

$$2|P| \le nt \le 2|P| + t.$$

Each I_i maintains distance at least $t/2$ from G_0. Since $|I_i| \le t/2$ and t is large, I_i is covered by two balls of radius $t/2 - b_1$, centered on the endpoints of the interval. Thus Minsky's quasi-projection theorem assures that $|\operatorname{Proj}_{G_0}(I_i)| < 2b_2$ for each i. If the endpoints of I_i are x_{i-1} and x_i and $\bar{x}_{i-1}$ and $\bar{x}_i$ are closest-point projections to G, we see that $d(\bar{x}_{i-1}, \bar{x}_i) \le 2b_2$. But then $t = d(\bar{x}_0, \bar{x}_n) \le 2nb_2$, and since $nt \asymp |P|$, we obtain $|P| \succ t^2$. □

4. The mapping class group

Constructing a path for the upper bound. Consider two infinite distinct geodesic rays

$$\phi : \mathbb{N} \to \operatorname{Mod}(S) \quad \text{and} \quad \psi : \mathbb{N} \to \operatorname{Mod}(S)$$

in $\operatorname{Mod}(S)$ emanating from a common point (without loss of generality, the identity element) in $\operatorname{Mod}(S)$. We want to show that, for all $t \in \mathbb{N}$, there is a path from $\phi(t)$ to $\psi(t)$ in $\operatorname{Mod}(S)$, such that no point in this path is within distance t of the origin, and whose length is on the order of t^2. We construct this path by traveling iteratively through chained copies of $\mathbb{Z}^2$, each copy generated by Dehn twists about a pair of disjoint curves. However, first we need to move far enough from the identity, to points $\phi(dt)$ and $\psi(dt)$, so that the first few steps of the sequence are sure not to backtrack near the identity. This is accomplished by taking d from Lemma 2.11.

Let $B_t = B_t(e)$ be the ball of radius t in $\mathrm{Mod}(S)$ about the identity. The segments $[\phi(t), \phi(dt)]$ and $[\psi(t), \psi(dt)]$ stay outside B_t and their lengths are of order t. To prove Theorem A for $\mathrm{Mod}(S)$, it is sufficient to build a path P between the two points $u = \phi(dt)$ and $v = \psi(dt)$ that stays outside of B_t and has length on the order of t^2.

The path from u to v will involve high powers of Dehn twists arranged in "switch moves" from one twist flat to the next. We fix the exponent $m = m(t)$ to be larger than $(3d\mathsf{c} + \mathsf{c})t$ (but of order t).

Lemma 4.1 (Switch moves)**.** *For any two curves $\alpha, \beta \in \mathbf{A}$, any $w \in \mathrm{Mod}(S)$ with $|w| \leq 3dt$, and m chosen as above, there is a path $(\star)$ from $D_\alpha^m w$ to $D_\beta^m w$ staying outside of B_t and of length $\prec t$.*

Proof. For the curves α, β, we fix a chain of curves

$$\{\gamma_i^{\alpha\beta}\} = \left\{\gamma_1^{\alpha\beta}, \ldots, \gamma_k^{\alpha\beta}\right\}, \text{ with } k = k(\alpha, \beta), \tag{4}$$

having the property that

$$\alpha - \gamma_1^{\alpha\beta} - \cdots - \gamma_k^{\alpha\beta} - \beta$$

is a path in the curve complex, that is, each adjacent pair of curves is disjoint on S. When there is no ambiguity about α and β, we denote these curves simply by $\{\gamma_i\}$. Define $(\star)$ to be the following path from $D_\alpha^m w$ to $D_\beta^m w$:

$(\star)$

$$D_\alpha^m\, w \xrightarrow{D_{\gamma_1}^m} D_{\gamma_1}^m\, D_\alpha^m\, w \xrightarrow{D_\alpha^{-m}} D_{\gamma_1}^m\, w \xrightarrow{D_{\gamma_2}^n} D_{\gamma_2}^m\, D_{\gamma_1}^m\, w \to \cdots \to D_{\gamma_k}^m\, D_\beta^m\, w \xrightarrow{D_{\gamma_k}^{-m}} D_\beta^m\, w$$

Every word visited by this path has the form $D_{\gamma_i}^m D_{\gamma_j}^{m'} w$ for some $0 \leq m' \leq m$ and by Lemma 2.10 has a word length at least $m/\mathsf{c} - 3dt \geq t$. Therefore, the subpath $(\star)$ stays outside of B_t. Also, an upper bound for $k = k(\alpha, \beta)$ depends on the choice of $\mathbf{A}$ only. Therefore the length of each subpath $(\star)$ (which is $2km$) is of order t. □

To illustrate how the switch moves work we consider the following example. Suppose that $\alpha - \gamma_1 - \gamma_2 - \beta$ is a path in the curve complex from α to β and denote the associated Dehn twists by a, g, h and b, respectively. Note that $\langle a, g\rangle$, $\langle g, h\rangle$ and $\langle h, b\rangle$ are free abelian subgroups of $\mathrm{Mod}(S)$. Let w be the word

$$w = h^{-m}\ b^m\ g^{-m}\ h^m\ a^{-m}\ g^m.$$

After cancellation, w is equivalent to $b^m a^{-m}$. Assuming that a, g, h and b are all in the generating set of $\mathrm{Mod}(S)$, we can also consider w as path along the edges of the Cayley graph of $\mathrm{Mod}(S)$ connecting a^m to b^m (first follow edges marked by the generator g for m steps, then follow edges marked a^{-1} for m steps, and so on). Note that, in the process of carrying out the word, we stay far from the identity in the Cayley graph: for any decomposition $w = vw'$, the word $w'a^m$ contains a high power of at least one of a, g, h or b, so its word-length is large. This path starts from a^m, travels through the quasi-flat $\langle a, g\rangle$ to g^m, through $\langle g, h\rangle$ to h^m and through $\langle h, b\rangle$ to b^m. (On the other hand, the shorter path connecting a^m to b^m corresponding to the word $b^m a^{-m}$ would go through the identity.)

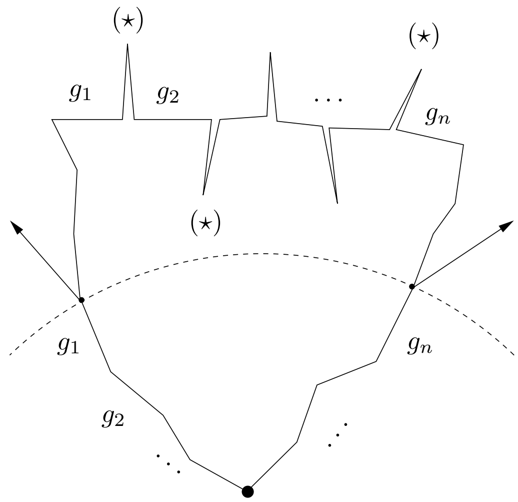

FIGURE 3. The path in the Cayley graph pushes off from the ball B_t and moves through a chain of Dehn twist flats in $\mathrm{Mod}(S)$.

Now we will build a path P from u to v using these switch moves. Let $n = |u\,v^{-1}|$ and write $u\,v^{-1} = g_n \cdots g_2 g_1$, where $g_i = D_{\alpha_i}^{\pm 1}$, $\alpha_i \in \mathbf{A}$, is a generator. For $0 \le r \le n$, let $v_r = g_r \cdots g_1\, v$, so that $v_0 = v$ and $v_n = u$.

$$v \xrightarrow{g_1^{\pm m}} g_1^{\pm m} v \xrightarrow{g_1} g_1^{\pm m} v_1 \xrightarrow{(\star)} g_2^m v_1 \xrightarrow{g_2} g_2^m v_2 \xrightarrow{(\star)} g_3^m v_2 \longrightarrow \cdots$$
$$\cdots \longrightarrow g_{n-1}^m v_{n-1} \xrightarrow{(\star)} g_n^{\pm m} v_{n-1} \xrightarrow{g_n} g_n^{\pm m} v_n \xrightarrow{g_n^{\mp m}} v_n = u$$

The signs $\pm$ are chosen by Lemma 2.11, which implies that the segments $[v, g_1^m v]$ and $[v, g_1^{-m} v]$ do not both intersect B_t. We choose the segment for the beginning of P that is disjoint from B_t. Similarly, the sign for the power at the last step is chosen so that the path from v_n to $g_n^{\pm m} v_n$ is outside B_t. Note that $n \le 2dt$ and $|v_i| \le |v| + n \le 3dt$; therefore, by Lemma 2.8 and the assumption on m,

$$d_{\alpha_j}(g_j^m v_i) \overset{+}{\succ} m - 3d\mathsf{c}t \ge \mathsf{c}t.$$

By Theorem 2.6, we get $|g_j^m v_i| \ge t$, which confirms that all of P stays outside of B_t.

We have shown that the length of path P, which contains $n-1$ "switch move" subpaths, is of order t^2. This finishes the proof of Theorem A for $\mathrm{Mod}(S)$.

Example realizing the quadratic rate. We now prove Theorem B for the case of $\mathrm{Mod}(S)$ by first establishing a quasi-projection theorem. Note that closest-point projection is not in general well-behaved in the mapping class group itself, so the theorem is stated in terms of projection in the curve complex. We map from $\mathrm{Mod}(S)$ to $\mathcal{C}(S)$ by the coarse map $g \mapsto g\mathbf{A}$. Note that the word-length formula (Theorem 2.6) ensures that this map is coarsely Lipschitz.

Distance contraction results for the mapping class group with arguments based on subsurface projection appear with various formulations in the literature (including Masur-Minsky, Behrstock, and forthcoming work of Brock-Masur-Minsky). A statement is given and proved here in the generality which we will require, providing a strong parallel with Theorem 3.1.

A geodesic line, ray, or segment in $\mathrm{Mod}(S)$ is called *E–cobounded* if for every proper subsurface $Y \subset S$ and for every two elements $a, b \in G$, we have $d_Y(a\mathbf{A}, b\mathbf{A}) \le E$.

Theorem 4.2 (Quasi-projection for $\mathrm{Mod}(S)$). *For every $E > 0$ there exist constants B_1 and B_2 depending on E and the topology of S such that the following holds. Let $G \subset \mathrm{Mod}(S)$ be an E–cobounded geodesic, suppose $g \in \mathrm{Mod}(S)$ satisfies $d(g, G) > B_1$, and let $R = \frac{d(g,G)}{B_1}$. Then $\mathcal{G} = G\mathbf{A}$ is a quasi-geodesic in $\mathcal{C}(S)$ and*

$$\mathrm{diam}\Big(\mathrm{Proj}_{\mathcal{G}}\left(B_R(g)\mathbf{A}\right)\Big) \le B_2.$$

As before, the theorem could be stated for quasi-geodesics with the same argument. Note that for quasi-projection to a segment, it suffices that the endpoints $a, b \in \mathrm{Mod}(S)$ satisfy $d_Y(a\mathbf{A}, b\mathbf{A}) \le E$ in order for the conclusion to obtain.

For a concrete application, consider a pseudo-Anosov element w and its axis $\{w^n\}$ in $\mathrm{Mod}(S)$. These axes are known to be cobounded by work of Masur-Minsky. We then apply the quasi-projection theorem exactly as above to show that, for a high power m, any path P connecting w^{-m} to w^m outside the ball of radius $t := |w^m| \asymp |m|$ must have length at least on the order of t^2. (Since projection to the curve complex coarsely contracts distances, this agrees with the desired inequality.)

The rest of this section is devoted to proving Theorem 4.2. The constants in the rest of this section will depend on the choice of E, which is chosen once and for all. Keeping this in mind we can write, for example,

$$d_S(\mathbf{A}, \mathcal{G}) \asymp 1.$$

Proof of Theorem 4.2. Let $h \in B_R(g)$. Let $\alpha = g(\mathbf{A})$ and $\beta = h(\mathbf{A})$ (these are sets of curves, and they fill S). Let $\overline{\alpha}$ and $\overline{\beta}$ be closest-point projections of α and β to $\mathcal{G}$, respectively. Choose $\overline{g}, \overline{h} \in G$ so that

$$\overline{g}(\mathbf{A}) \cap \overline{\alpha} \neq \emptyset \quad \text{and} \quad \overline{h}(\mathbf{A}) \cap \overline{\beta} \neq \emptyset,$$

which is possible because $\mathcal{G} = G\mathbf{A}$. Our goal is to show that $d_S(\overline{\alpha}, \overline{\beta}) \le B_2$, so assume for contradiction that

$$d_S(\overline{\alpha}, \overline{\beta}) > B_2.$$

Given any K, c as in Theorem 2.6, we can define $\mathcal{Y}$ to be the set of proper subsurfaces $Y \subsetneq S$ with $d_Y(\alpha, \overline{\alpha}) \ge \mathsf{K}$. By the theorem's hypothesis, we have $|g^{-1}\overline{g}| = d(g, \overline{g}) \ge d(g, G) = B_1 R$ and so

$$\sum_{Y \in \mathcal{Y}} d_Y(\alpha, \overline{\alpha}) + d_S(\alpha, \overline{\alpha}) \ge \frac{|g^{-1}\overline{g}|}{\mathsf{c}} \ge \frac{B_1 R}{\mathsf{c}}.$$

We will proceed in two cases. First assume that

$$d_S(\alpha, \overline{\alpha}) \ge \frac{B_1 R}{2\mathsf{c}}. \tag{5}$$

If $B_2 > M_1$, then any geodesic connecting α to β passes through the M_2–neighborhood of $\mathcal{G}$, by Lemma 2.12. Hence for large enough B_1, we have

$$d_S(\alpha, \beta) \ge d_S(\alpha, \overline{\alpha}) - M_2 \ge \frac{B_1 R}{2\mathsf{c}} - M_2 \ge \frac{B_1 R}{3\mathsf{c}}.$$

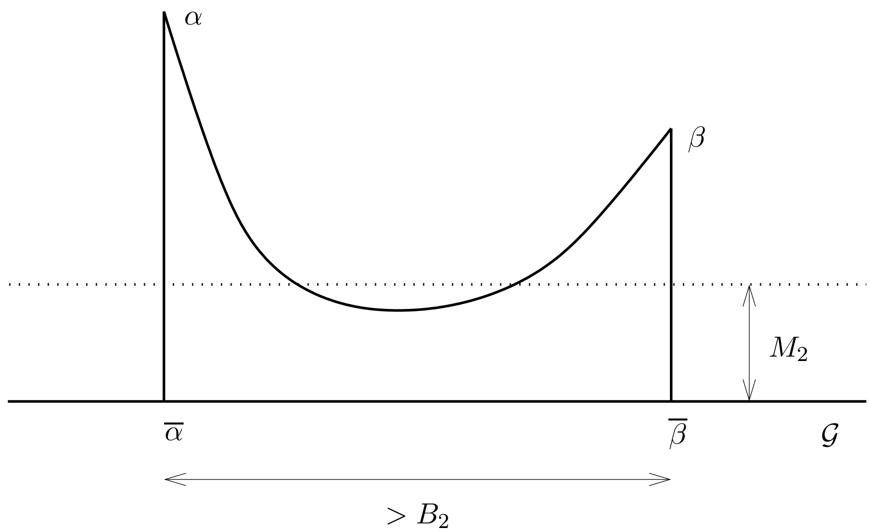

FIGURE 4. We assume for contradiction that the projections of α and β to $\mathcal{G}$ are far apart in $\mathcal{C}(S)$.

Enlarging B_1 again if necessary, we can assume $d_S(\alpha, \beta) \ge \mathsf{K}_0$. Theorem 2.6 implies:

$$|h^{-1}g| \ge \frac{1}{\mathsf{c}} d_S(\alpha, \beta) \ge \frac{B_1 R}{3\mathsf{c}\mathsf{c}_0} > R,$$

for B_1 sufficiently large. But this contradicts the assumption that $d(g, h) \le R$.

In the second case, the assumption of Equation (5) is not true, so we have

$$\sum_{Y \in \mathcal{Y}} d_Y(\alpha, \overline{\alpha}) \ge \frac{B_1 R}{2\mathsf{c}}, \tag{6}$$

where again $\mathcal{Y}$ is chosen relative to the threshold K.

Next we show that the subsurfaces with large projection distance between α and $\overline{\alpha}$ do not also have large distance between β and $\overline{\beta}$.

Claim. *For every $Y \in \mathcal{Y}$, $d_Y(\beta, \overline{\beta}) \le M_0$.*

Proof. If $\mathsf{K} > M_0$, then Theorem 2.1 implies that any geodesic $[\alpha, \overline{\alpha}]$ intersects the 1–neighborhood of ∂Y for every $Y \in \mathcal{Y}$. Applying Lemma 2.13 and Lemma 2.14 we can conclude that

$$d_S(\overline{\alpha}, \mathrm{Proj}(\partial Y)) \le M_3 + 2M_4.$$

If $d_Y(\beta, \overline{\beta}) > M_0$ for any particular $Y \in \mathcal{Y}$, we similarly get

$$d_S(\overline{\beta}, \mathrm{Proj}(\partial Y)) \le M_3 + 2M_4.$$

But then

$$d_S(\overline{\alpha}, \overline{\beta}) \le 2M_3 + 4M_4.$$

For B_2 larger than this final constant, this is a contradiction. □

Note that $\overline{g}, \overline{h} \in G$ and G is a E–cobounded geodesic. Therefore, for every $Y \in \mathcal{Y}$,

$$d_Y(\overline{\alpha}, \overline{\beta}) = d_Y(\overline{g}(\mathbf{A}), \overline{h}(\mathbf{A})) \le E.$$

Using the triangle inequality for projection distance (Lemma 2.7) we get

$$d_Y(\alpha,\beta) \stackrel{+}{\succ} d_Y(\alpha,\overline{\alpha}) - d_Y(\overline{\alpha},\overline{\beta}) - d_Y(\overline{\beta},\beta) \geq d_Y(\alpha,\overline{\alpha}) - M_0 - E.$$

By choosing K large enough, we can ensure for $Y \in \mathcal{Y}$ that $d_Y(\alpha,\beta)$ is much larger than these additive errors in order to get a multiplicatively coarse equality

$$d_Y(\alpha,\beta) \stackrel{\cdot}{\asymp} d_Y(\alpha,\overline{\alpha})$$

as well as ensuring that

$$d_Y(\alpha,\beta) \geq \mathsf{K}_0.$$

We apply these, as well as Theorem 2.6 one last time, to see that

$$\begin{aligned} \mathsf{c}_0\,|h^{-1}g| &\stackrel{+}{\succ} \sum_{Y\in\mathcal{Y}} \Big[d_Y(g(\mathbf{A}),h(\mathbf{A}))\Big]_{\mathsf{K}_0} \\ &\stackrel{+}{\asymp} \sum_{Y\in\mathcal{Y}} \Big[d_Y(\alpha,\beta)\Big]_{\mathsf{K}_0} \\ &\stackrel{\cdot}{\asymp} \sum_{Y\in\mathcal{Y}} d_Y(\alpha,\overline{\alpha}) \geq \frac{B_1 R}{2\mathsf{c}}, \end{aligned}$$

where the final inequality comes from (6). We have shown that

$$|h^{-1}g| \succ \frac{B_1 R}{2\mathsf{c}\mathsf{c}_0}.$$

Again, choosing B_1 large enough provides the contradiction. □

References

[Abi80] W. Abikoff. *The real analytic theory of Teichmüller space*, volume 820 of *Lecture Notes in Mathematics*. Springer, Berlin, 1980.

[Beh06] J. Behrstock. Asymptotic geometry of the mapping class group and Teichmüller space. *Geometry & Topology*, 10:1523–1578, 2006.

[Ber78] L. Bers. An extremal problem for quasiconformal mappings and a theorem by Thurston. *Acta Math.*, 141(1-2):73–98, 1978.

[BH99] M.R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[Ger94] S.M. Gersten. Quadratic divergence of geodesics in CAT(0) spaces. *Geom. Funct. Anal.*, 4(1):37–51, 1994.

[Ger94b] S.M. Gersten. Divergence in 3-manifold groups. *Geom. Funct. Anal.* 4(6):633–647, 1994.

[Gro93] M. Gromov. *Geometric group theory, Vol. 2: Asymptotic invariants of infinite groups.* Niblo and Roller, eds. LMS Lecture Note Ser., vol. 182, Cambridge Univ. Press, Cambridge, 1993.

[Hub06] J. Hubbard. *Teichmüller theory and applications to geometry, topology and dynamics.* Matrix Editions, Ithaca, NY, 2006.

[Iva97] N.V. Ivanov. Automorphism of complexes of curves and of Teichmüller spaces. *Internat. Math. Res. Notices*, 1997(14):651–666, 1997.

[KL98] M. Kapovich and B. Leeb. 3-manifold groups and nonpositive curvature. *Geom. Funct. Anal.*, 8(5):841–852, 1998.

[Ker80] S.P. Kerckhoff. The asymptotic geometry of Teichmüller space. *Topology*, 19(1):23–41, 1980.

[Mas75] H.A. Masur. On a class of geodesics in Teichmüller space. *Ann. of Math. (2)*, 102(2):205–221, 1975.

[Mas81] H.A. Masur. Transitivity properties of the horocyclic and geodesic flows on moduli space. *J. Analyse Math.*, 39:1–10, 1981.

[Min96a] Y.N. Minsky. Extremal length estimates and product regions in Teichmüller space. *Duke Math. J.*, 83(2):249–286, 1996.

[Min96b] Y.N. Minsky. Quasi-projections in Teichmüller space. *J. Reine Angew. Math.*, 473:121–136, 1996.

[MM99] H.A. Masur and Y.N. Minsky. Geometry of the complex of curves. I. Hyperbolicity. *Invent. Math.*, 138(1):103–149, 1999.

[MM00] H.A. Masur and Y.N. Minsky. Geometry of the complex of curves. II. Hierarchical structure. *Geom. Funct. Anal.*, 10(4):902–974, 2000.

[MW95] H.A. Masur and M. Wolf. Teichmüller space is not Gromov hyperbolic. *Ann. Acad. Sci. Fenn. Ser. A I Math.*, 20(2):259–267, 1995.

[Pap95] P. Papasoglu. On the subquadratic isoperimetric inequality. *Geometric Group Theory*, 149–157, de Gruyter, 1995.

[Raf07] K. Rafi. A combinatorial model for the Teichmüller metric. Geom. Funct. Anal., 17(3):936–959, 2007.

[Raf05] K. Rafi. A characterization of short curves of a geodesic in Teichmüller space. *Geometry & Topology*, 9:179–202, 2005.

[Sh85] H. Short, ed. MSRI notes on hyperbolic groups. *Group theory from a geometrical viewpoint*, 3–63, World Sci. Publ., River Edge, NJ, 1991.

[Str80] K. Strebel. *Quadratic differentials*, volume 5 of *A series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 1980.

[Wol75] S.A. Wolpert. Noncompleteness of the Weil-Petersson metric for Teichmüller space. *Pacific J. Math.*, 61(2):573–577, 1975.

[Wol79] S.A. Wolpert. The length spectra as moduli for compact Riemann surfaces. *Ann. of Math. (2)*, 109(2):323–351, 1979.

E-mail address: mduchin@umich.edu
URL: http://www.math.lsa.umich.edu/∼mduchin/

E-mail address: rafi@math.uchicago.edu
URL: http://www.math.uchicago.edu/∼rafi/