

# LARGE-SCALE RANK OF TEICHMÜLLER SPACE

---

ALEX ESKIN, HOWARD MASUR, and KASRA RAFI

## Abstract

Suppose that  $\mathcal{X}$  is either the mapping class group equipped with the word metric or Teichmüller space equipped with either the Teichmüller metric or the Weil–Petersson metric. We introduce a unified approach to study the coarse geometry of these spaces. We show that for any large box in  $\mathbb{R}^n$  there is a standard model of a flat in  $\mathcal{X}$  such that the quasi-Lipschitz image of a large sub-box is near the standard flat. As a consequence, we show that, for all these spaces, the geometric rank and the topological rank are equal. The methods are axiomatic and apply to a larger class of metric spaces.

## 1. Introduction

In this article we study the large-scale geometry of several metric spaces: the Teichmüller space  $\mathcal{T}(S)$  equipped with the Teichmüller metric  $d_{\mathcal{T}}$ , the Teichmüller space equipped with the Weil–Petersson metric  $d_{WP}$ , and the mapping class group  $\text{Mod}(S)$  equipped with the word metric  $d_W$ . (Brock [8] showed that the Weil–Petersson metric is quasi-isometric to the pants complex.) Even though the definitions of distance in these spaces are very different, they share a key feature, namely, an inductive structure. That is, they are a union of product regions associated to lower-complexity surfaces with the gluing pattern given by the curve complex.

Let  $S$  be a possibly disconnected surface of finite hyperbolic type, and let  $\mathcal{X} = \mathcal{X}(S)$  be a metric space that is one of the metric spaces mentioned above. One major goal in understanding the large-scale geometry of a metric space is to prove the quasi-isometric rigidity of the space—that any quasi-isometry is a bounded distance from an isometry. The usual starting point is to understand its flats. By a flat here we mean a quasi-isometric image of Euclidean space. In a follow-up paper [14], we will use the description of flats to investigate the quasi-isometric rigidity of  $\mathcal{T}(S)$ .

In this article, we analyze quasi-Lipschitz maps from a *large box*  $B \subset \mathbb{R}^n$  into  $\mathcal{X}$ . A *quasi-Lipschitz map* is a usual Lipschitz map where one allows an additive error.

DUKE MATHEMATICAL JOURNAL

Vol. 166, No. 8, © 2017 DOI [10.1215/00127094-0000006X](https://doi.org/10.1215/00127094-0000006X)

Received 17 September 2013. Revision received 26 August 2016.

First published online 28 March 2017.

2010 *Mathematics Subject Classification*. Primary 32G15; Secondary 20F65.

Our goal is to give a description of the image of such a map on a *large sub-box*  $B' \subset B$ . We show that the image of  $B'$  looks like a *standard flat* up to a small linear error. A standard flat is a product of *preferred paths* associated to disjoint subsurfaces of  $S$  (see Definitions 4.4 and 7.1). One can think of a preferred path as a generality of a Teichmüller geodesic. Our main theorem is the following.

**THEOREM A (Image of a box is locally standard)**

*Let  $\mathcal{X}$  be quasi-isometric to one of the following spaces: Teichmüller space  $\mathcal{T}(S)$  equipped with the Teichmüller metric  $d_{\mathcal{T}}$ , Teichmüller space with the Weil–Petersson metric  $d_{WP}$  (or pants graph), or the mapping class group  $\text{Mod}(S)$  equipped with the word metric  $d_W$ . For all  $K, C$  and for all  $R_0, \epsilon_0$ , there exists  $R_1$  such that if  $B$  is a box of size at least  $R_1$  and  $f: B \rightarrow \mathcal{X}$  is a  $(K, C)$ -quasi-Lipschitz map, then there is a sub-box  $B' \subset B$  of size  $R' \geq R_0$  such that  $f(B')$  lies inside an  $O(\epsilon_0 R')$ -neighborhood of a standard flat in  $\mathcal{X}$ .*

The implied constants in the theorem only depend on  $K, C$ , and the topology of the surface.

As a corollary, we determine the large-scale rank of the space  $\mathcal{X}$ . For a surface  $X$  of genus  $g$  with  $b$  boundary components, the *complexity* is defined to be  $\xi(X) = 3g - 3 + b$ . A subsurface  $X$  of  $S$  is *essential* if  $\xi(X) > 0$ . Thrice-punctured spheres ( $\xi(X) = 0$ ) are never essential. Annuli ( $\xi(X) = -1$ ) are essential only in the cases of the mapping class group and Teichmüller space with the Teichmüller metric.

Define the *topological rank* of  $\mathcal{X}$ ,  $\text{rank}_{\text{top}}(\mathcal{X})$ , to be the largest integer  $r$  so that there are pairwise-disjoint essential subsurfaces  $W_1, \dots, W_r$  in  $S$ . The dimension of a standard flat in  $\mathcal{X}$  is at most  $\text{rank}_{\text{top}}(\mathcal{X})$ .

**THEOREM B (Geometric rank)**

*For every  $K$  and  $C$ , there is a constant  $R_2$  so that if  $B$  is a box of size at least  $R_2$  in  $\mathbb{R}^n$  and  $f: B \rightarrow \mathcal{X}$  is a  $(K, C)$ -quasi-isometric embedding, then  $n \leq \text{rank}_{\text{top}}(\mathcal{X})$ . Furthermore, the theorem is sharp. For  $n = \text{rank}_{\text{top}}(\mathcal{X})$ , there is a quasi-isometric embedding of a Euclidean  $n$ -dimensional half-space into  $\mathcal{X}$ .*

Define the *geometric rank* of  $\mathcal{X}$ ,  $\text{rank}_{\text{geo}}(\mathcal{X})$ , to be the largest integer  $n$  so that there exist  $K, C$  such that, for any  $R$ , there is a  $(K, C)$ -quasi-isometric embedding  $f$  of a ball  $B \subset \mathbb{R}^n$  of radius  $R$  into  $\mathcal{X}$ . Also, let  $g$  be the genus of  $S$ , let  $p$  be the number of punctures of  $S$ , and let  $c$  be the number of components of  $S$ .

**COROLLARY C**

*The topological and the geometric ranks of  $\mathcal{X}$  are equal. Namely, if  $\mathcal{X}$  is either  $(\text{Mod}(S), d_W)$  or  $(\mathcal{T}(S), d_{\mathcal{T}})$ , then*

$$\text{rank}_{\text{geo}}(\mathcal{X}) = \text{rank}_{\text{top}}(\mathcal{X}) = 3g + p - 3c,$$

and if  $\mathcal{X}$  is  $(\mathcal{T}(S), d_{\text{WP}})$ , then

$$\text{rank}_{\text{geo}}(\mathcal{X}) = \text{rank}_{\text{top}}(\mathcal{X}) = \left\lfloor \frac{3g + p - 2c}{2} \right\rfloor.$$

*Remark 1.1*

In the case of  $\text{Mod}(S)$ , there are quasi-isometric embeddings of all of Euclidean space, of dimension the geometric rank, into  $\text{Mod}(S)$ . One can achieve this via Dehn twists about disjoint simple closed curves (see [15]). The same is true for Teichmüller space with the Weil–Petersson metric. One takes a maximal collection of disjoint essential surfaces and a bi-infinite quasi-geodesic in each. In [7, Theorem 1.3], B. Bowditch proves that there is a quasi-isometric embedding of a Euclidean  $n$ -dimensional half-space into  $(\mathcal{T}(S), d_{\mathcal{T}})$  if and only if  $n \leq \text{rank}_{\text{top}}(\mathcal{T}(S), d_{\mathcal{T}})$ . In [7, Theorem 1.4] he shows that there are quasi-isometric embeddings of the entire Euclidean space of dimension  $\text{rank}_{\text{top}}(\mathcal{T}(S), d_{\mathcal{T}})$  into Teichmüller space if and only if the genus of  $S$  is at most 1 or  $S$  is a closed surface of genus 2.

*Remark 1.2*

Theorems A and B hold for a larger class of metric spaces than is discussed above. Essentially, one needs a mapping-class-group action and a distance formula similar to (11) (see Masur–Schleimer [21] for examples of such distance formulas). The definition of an *essential surface* has to be modified to mean any type of surface that appears in the associated distance formula.

Many such spaces, for example, the arc complex and the disk complex, are known to be Gromov-hyperbolic and therefore have geometric rank 1 (see [21]). Hence, the corollary is already known. Others are typically not Gromov-hyperbolic, such as the Hatcher–Thurston complex and the separating curve complex, and our results prove that the geometric rank and topological rank are equal. These complexes have been omitted to simplify the exposition. After this article was completed, the axiomatic approach suggested here was formalized by Behrstock, Hagen, and Sisto [2], who introduced the notion of hierarchically hyperbolic spaces. In fact, they give a different proof of the rank theorem.

*History*

The idea of studying the rank of these objects was introduced by Brock and Farb [9]. In the case when  $\mathcal{X}$  is the pants graph, Corollary C was first proved in that paper when the surface is the twice-punctured torus. They also showed that the topological rank is bounded above by the geometric rank and conjectured Corollary C. Corollary C was then proved for the mapping class group with the word metric or

Teichmüller space with the Weil–Petersson metric by Behrstock and Minsky [4] and Hamenstädt [16]. The rank statement is used to prove the quasi-isometric rigidity of  $\text{Mod}(S)$  by Behrstock, Kleiner, Minsky, and Mosher [3] and by Hamenstädt [16]. The case of Teichmüller space with the Teichmüller metric has not been studied previously. Note that the map that sends  $\text{Mod}(S)$  to the orbit of a point in  $\mathcal{T}(S)$  is not a quasi-isometry or even a quasi-isometric embedding because of the thin regions in  $\mathcal{T}(S)$  which locally look like products of horoballs. Unlike [4], which uses asymptotic cones, our approach uses the local behavior of quasi-Lipschitz maps.

### *Main tools*

To prove our theorems, we further develop some tools that already exist in the literature. The first one is the idea of coarse differentiation. This was introduced in the context of geometric group theory by Eskin, Fisher, and Whyte [12], [13] and used to prove the quasi-isometric rigidity of lattices in Sol and in the quasi-isometry classification of lamplighter groups (see references in those papers for its use in other contexts). The statement they used is similar to Theorem 2.8 below, which holds for quasi-Lipschitz maps between more general metric spaces. However, since we are mostly concerned with maps where the domain is a subset of  $\mathbb{R}^n$ , we prove the following statement, which is cleaner and easier to use. For a box  $B$  in  $\mathbb{R}^n$ , let  $\underline{B}$  be a central sub-box of  $B$  with one-third the diameter.

### THEOREM D (Coarse differentiation)

*Let  $\mathcal{Z}$  be a complete metric space. For every  $K, C, \epsilon_0, \theta_0$ , and  $R_0$  there is  $L_0$  so that the following holds. For  $L \geq L_0$ , let  $f : B \rightarrow \mathcal{Z}$  be a  $(K, C)$ -quasi-Lipschitz map, where  $B$  is a box of size  $L$  in  $\mathbb{R}^n$ . Then there are  $R \geq R_0$  and a subdivision  $\mathcal{B}_R$  of  $\underline{B}$  to sub-boxes of size  $R$  so that the proportion of boxes  $B' \in \mathcal{B}_R$  where  $f|_{B'}$  is  $\epsilon_0$ -efficient at scale  $R$  is at least  $(1 - \theta_0)$ .*

The precise definition of a *box* and *efficient* will be given in the next section. Roughly what it means is that for any line there is a partition of its points of a certain scale such that the sum of the distances between the images of successive partition points is bounded above up to a small additive linear error by the distance between the image of the endpoints. Thus, a certain reverse triangle inequality holds. This notion is very different from quasi-geodesic, where errors are allowed to be multiplicative. Here they are additive errors. One can also think of *efficiency* as meaning that the image of every line is a line up to a *sublinear error*. One should think of the above theorem as a coarse version of Rademacher’s theorem that if  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is Lipschitz, then  $f$  is differentiable almost everywhere. In Theorem D the sub-box  $B'$  is an analogue of a point of differentiability.

The importance of efficiency lies first of all in the fact that in a product space the projection of an efficient path to a factor is still efficient. The corresponding statement for quasi-geodesics is false. Furthermore, efficient maps into hyperbolic spaces are easy to control (see Proposition 2.14).

This becomes especially relevant because of the construction of Bestvina, Bromberg, and Fujiwara [5], which we make use of. They embed the mapping class group into a product of finitely many hyperbolic spaces. Their construction is axiomatic and can be adapted easily to embed any of our spaces  $\mathcal{X}$  into a product of finitely many hyperbolic spaces. The  $L^1$ -metric on this product induces a metric on the space  $\mathcal{X}$ , and this is the metric with respect to which we apply the coarse differentiation theorem. Thus, the fact that the projections of efficient paths are still efficient becomes useful in this context. Note that the notion of efficiency is not preserved under quasi-isometry, and the choice of metric here is essential. The conclusion of this discussion, Theorem 4.9, will be that efficient paths fellow-travel paths with the same endpoints that have nice properties. These latter paths, which we call *preferred paths*, will play the role of geodesics.

The power then of Theorem D lies in the fact that one can add the assumption of efficiency for free, just by replacing  $B$  with a sub-box  $B'$ . Altogether, this will mean that, on large boxes, the image of every line fellow-travels a preferred path.

Finally, we use the realization theorem of Behrstock, Minsky, Mosher, and Kleiner [3]. They provide a description of the image of the mapping class group in the product of curve complexes. We adopt it to provide a description of the image of  $\mathcal{X}$ . This is necessary to translate back the information obtained in each hyperbolic factor to information in  $\mathcal{X}$ .

### *Outline of the article*

Section 2 is devoted to the development of coarse differentiation theory and to the discussion of efficient maps. The main result is Theorem 2.5, as discussed above. We also establish the basic properties of efficient maps and prove that efficient paths in Gromov-hyperbolic spaces stay close to geodesics—an analogue of the Morse lemma.

In Section 3, we discuss the combinatorial model for each of the spaces considered in the article. The three seemingly different metric spaces above have very similar models. Namely, Teichmüller space equipped with the Weil–Petersson metric is quasi-isometric to the pants graph (see [8]). The mapping class group is quasi-isometric to the marking graph by the work of Masur and Minsky [20], and Teichmüller space equipped with the Teichmüller metric is quasi-isometric to the space of augmented markings by the work of Rafi [26] and Durham [11].

The advantage of this approach is that we can measure the relative complexity of two points  $x, y \in \mathcal{X}$  from the point of view of a subsurface  $W$ . This is the distance

in the curve complex of  $W$  between the projections of  $x$  and  $y$  to  $W$ . The curve complex of every surface  $W$  is known to be Gromov-hyperbolic. We then define a coarse metric on each of these combinatorial models using a distance formula which is the sum over relative complexity from the point of view of different subsurfaces. Since we work in the category of spaces up to quasi-isometries, the distance needs to be defined only up to a multiplicative error.

In Section 4, we introduce the notion of preferred paths. These are paths whose projections to every curve complex are quasi-geodesics, and they replace the notion of geodesics in our spaces. The main statement in the section is Theorem 4.9, which shows that an efficient path stays near a preferred path joining its endpoints. Hence, the outcome of the coarse differentiation theorem is indeed a box where straight lines are mapped to straight lines up to the first order. This is the key tool for the rest of the article. The proof uses the construction in [5], which allows one to embed  $\mathcal{X}$  into a product of hyperbolic spaces. The projection of the efficient paths into each factor stays near a geodesic in that factor. We then use this and a consistency theorem (Theorem 3.2) to build the preferred path in  $\mathcal{X}$  tracing the given efficient path.

Section 5 establishes some properties of preferred paths. The main ones are fellow-traveling properties that say that, under certain conditions, preferred paths that begin and end near the same point fellow-travel in the middle (Propositions 5.7 and 5.10). These statements are used in the succeeding sections to build big boxes with the required properties. In Section 6, the main inductive step is proved (Theorem 6.1), and in Section 7 we assemble the proofs of the main theorems.

#### *Treatment of constants*

Suppose that  $\mathcal{Y}$  and  $\mathcal{Z}$  are geodesic metric spaces. We say a map  $f: \mathcal{Y} \rightarrow \mathcal{Z}$  is *quasi-Lipschitz* if there are constants  $K$  and  $C$  so that

$$d_{\mathcal{Z}}(f(x_1), f(x_2)) \leq K d_{\mathcal{Y}}(x_1, x_2) + C.$$

We fix constants  $K$  and  $C$  once and for all. We also fix an upper bound for the complexity of the surface  $S$  and the dimension  $n$ . When we say a constant is *uniform*, we mean that its value depends only on  $K$ ,  $C$ , the topology of  $S$ , and the value of  $n$ . Similarly, we will use terms like *quasi-isometric embedding* or *quasi-isometry* to mean that the associated constants are the same as those for  $K$  and  $C$  fixed above.

To simplify the presentation, we try to avoid naming uniform constants when they are used locally inside of an argument. Instead, we adopt the following notation. Let  $a$  and  $b$  represent various quantities, and let  $M$  and  $C'$  be uniform constants. We say that  $a$  is less than  $b$  up to a multiplicative error,  $a \stackrel{*}{\prec} b$ , if  $a \leq Mb$ . We say that  $a$  and  $b$  are comparable,  $a \stackrel{*}{\asymp} b$ , if we have both  $a \stackrel{*}{\prec} b$  and  $b \stackrel{*}{\prec} a$ .

Using a similar notation when the error is additive or both additive and multiplicative, we say that  $a \overset{+}{\prec} b$  if  $a \leq b + C'$  and  $a \prec b$  if  $a \leq M a + C'$ . Again,  $a \overset{+}{\succ} b$  if we have both  $a \overset{+}{\prec} b$  and  $b \overset{+}{\prec} a$ , and  $a \succ b$  if we have both  $a \prec b$  and  $b \prec a$ . Also, we often use the notation  $a = O(b)$  to mean  $a \overset{*}{\prec} b$ . For example,

$$a \overset{+}{\prec} b + O(c) \iff a \leq b + M c + C',$$

for uniform constants  $M$  and  $C'$ .

Using this notation we may write

$$a \overset{+}{\succ} b \text{ and } b \overset{+}{\succ} c \implies a \overset{+}{\succ} c.$$

Here, the additive error in the last inequality is the sum of the additive errors in the first two inequalities and hence is still a uniform constant. That is, different occurrences of  $\overset{+}{\succ}$  have different implied constants. But as long as we use statements of this type a uniformly bounded number of times, all of the implied constants are still uniform.

**2. Coarse differentiation**

Being *differentiable* means that, to first order, lines are mapped to lines. In this section, we introduce a notion of coarse differentiability for a quasi-Lipschitz map. But first, we need a coarse notion of a straight line that behaves better than a quasi-geodesic in a product space. Points along a line (a geodesic) satisfy the reverse triangle inequality. We emulate this by introducing the notion of  $\epsilon$ -efficient paths. The points along an efficient path satisfy the reverse triangle inequality up to a small multiplicative error.

*Definition 2.1*

Let  $Z$  be a complete metric space, let  $\gamma: [a, b] \rightarrow Z$  be a  $(K, C)$ -quasi-Lipschitz (not necessarily continuous) path, and let  $R > 0$  be such that  $|b - a| \overset{*}{\prec} R$ . Let  $0 < \epsilon < 1$ . An  $r$ -partition of  $[a, b]$  is a set of times  $a = t_0 < t_1 < \dots < t_m = b$  so that  $(t_{i+1} - t_i) \leq r$ . Let  $z_i = \gamma(t_i)$ . We define the *coarse length of  $\gamma$  with steps  $r$*  to be

$$\Delta(\gamma, r) = \inf_{r\text{-partitions}} \sum_{i=0}^{m-1} d_Z(z_i, z_{i+1}). \tag{1}$$

We say that  $\gamma$  is  $\epsilon$ -efficient at *scale  $R$*  if

$$\Delta(\gamma, \epsilon R) \leq d_Z(\gamma(a), \gamma(b)) + O(\epsilon R). \tag{2}$$

Here the constant in the implied bound  $O$  can depend on  $K, C$  but not on  $\epsilon$  or  $R$ . We will always assume that  $\epsilon R \geq 1$ . Note that, by the triangle inequality,  $\Delta(\gamma, \epsilon R) \geq d_Z(\gamma(a), \gamma(b))$ . Hence, (2) is essentially a reverse triangle inequality.

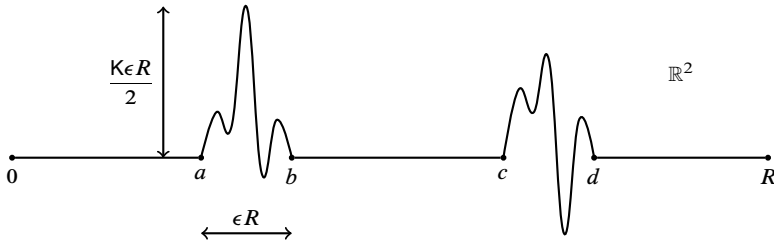


Figure 1. We modify the segment  $[0, R]$  by adding spikes.

*Example 2.2*

We now review a few examples of paths that are or are not  $\epsilon$ -efficient. (The scale is always the length of the domain.)

As a first example, consider a geodesic segment  $\gamma : [0, R] \rightarrow \mathbb{R}^2$  along the  $x$ -axis, which we can think of as the graph of the function which is identically 0. Then modify this graph  $\gamma$  by adding a spike (see Figure 1). That is, replace  $\gamma|_{[a,b]}$  for  $(b - a) \leq \epsilon R$  with the graph of a map  $h : [a, b] \rightarrow \mathbb{R}$ , where  $h(a) = h(b) = 0$  and  $h$  is, say,  $K$ -Lipschitz. Note that  $h$  could take a value as large as  $K\epsilon R/2$ .

This new path is still  $\epsilon$ -efficient at scale  $R$ , since an  $\epsilon R$ -partition could miss the interval  $[a, b]$  and the sum of distances given by such a partition is exactly equal to  $R$ . In fact, one can add many such spikes along disjoint intervals of lengths less than  $\epsilon R$ .

As a second example, consider a path  $\gamma : [0, R] \rightarrow \mathcal{X}$  where the image has diameter  $\epsilon^2 R$ . Since there is an  $\epsilon R$ -partition where the number of intervals is of order  $\frac{1}{\epsilon}$ , we have

$$\Delta(\gamma, \epsilon R) \stackrel{*}{\asymp} \frac{1}{\epsilon}(\epsilon^2 R) = \epsilon R.$$

Hence, (2) holds and  $\gamma$  is efficient. An instance of this example in  $\mathbb{H}$  is the horocyclic segment in the line  $y = 1$  from  $(-\frac{R}{2}, 1)$  to  $(\frac{R}{2}, 1)$  where  $\epsilon^2 R \geq \log R$ . For a given  $\epsilon$ , we can always choose  $R$  large enough so that this holds. The diameter of this segment is  $\log R$ , and the path is efficient. That is, from our point of view, moving along a horocycle is essentially the same as the constant map. Note that horocycles are not quasi-geodesics, and thus, efficient paths do not have to be quasi-geodesics. However, as we shall see (Lemma 2.12), all efficient paths in  $\mathbb{H}$  are contained in an  $O(\epsilon R)$ -neighborhood of a geodesic, in this case the constant map.

For a further example note that, for  $\epsilon$  small enough, a semicircle of radius  $R$  in  $\mathbb{R}^2$  is a quasi-geodesic but it is not  $\epsilon$ -efficient at scale  $R$ . As  $\epsilon$  goes to zero,  $\Delta(\gamma, \epsilon R)$  converges to  $\pi R$ , which is much larger than  $2R$ .



We now establish some elementary properties of efficient paths. First, for a subset  $Z' \subset Z$  the closest point projection to  $Z'$  picks out the not necessarily unique closest point.

LEMMA 2.3

Consider a map  $\gamma : [a, b] \rightarrow Z$ .

- (1) Suppose that  $\gamma$  is  $\epsilon$ -efficient at scale  $R$  and that  $k$  is a uniformly bounded integer. Then for  $k$  points  $a \leq s_i \leq b$

$$\sum_{i=0}^{k-1} d_Z(\gamma(s_i), \gamma(s_{i+1})) = d_Z(\gamma(a), \gamma(b)) + O(\epsilon R).$$

- (2) For  $[c, d] \subset [a, b]$ , if  $\gamma$  is  $\epsilon$ -efficient at scale  $R$ , then so is  $\gamma' = \gamma|_{[c,d]}$ .
- (3) Assume that  $Z = Z_1 \times \dots \times Z_l$  is equipped with the  $L^1$ -metric, and let  $\gamma_i$  be the projection of  $\gamma$  to  $Z_i$ . If  $\gamma$  is  $\epsilon$ -efficient at scale  $R$ , then so is every  $\gamma_i$ .
- (4) If  $Z'$  is a subset of  $Z$  and  $\gamma(t)$  is an  $\epsilon^2$ -efficient path at scale  $R$  that is contained in an  $O(\epsilon^2 R)$ -neighborhood of  $Z'$ , then the composition of  $\gamma$ , followed by the closest point projection  $Z'$ , defines an  $\epsilon$ -efficient path at scale  $R$ .

*Proof*

Let  $t_0, \dots, t_m$  be an  $\epsilon R$ -partition of  $[a, b]$  which is within  $\epsilon R$  of the infimum in the definition of  $\Delta(\gamma, \epsilon R)$ . Add the points  $s_j$  to the partition. This will increase the sum in (1) by at most  $O(\epsilon R)$ . In fact, if  $t_i \leq s_j \leq t_{i+1}$ , then  $|t_i - t_{i+1}| \leq \epsilon R$  and, since  $\gamma$  is quasi-Lipschitz,

$$d_Z(\gamma(t_i), \gamma(s_j)) + d_Z(\gamma(s_j), \gamma(t_{i+1})) + 2C = O(\epsilon R).$$

Since the number of points  $s_j$  is uniformly bounded, adding all times  $s_j$  to the partition will increase the sum by at most  $O(\epsilon R)$ . Now, removing all  $t_i$ 's will only decrease the sum, and hence part (1) of the lemma holds.

To see part (2), as above, let  $t_0, \dots, t_m$  be a set of times where the sum in (1) is within  $\epsilon R$  of the infimum and so that the times  $c$  and  $d$  are included in the set  $\{t_i\}$ . Let  $z_i = \gamma(t_i)$ . Letting  $c = t_j$  and  $d = t_k$  we have

$$\begin{aligned} \Delta(\gamma, \epsilon R) &\geq \sum_{i=1}^m d_Z(z_i, z_{i+1}) - O(\epsilon R) \\ &\geq d_Z(\gamma(a), \gamma(c)) + \sum_{i=j}^{k-1} d_Z(z_i, z_{i+1}) + d_Z(\gamma(d), \gamma(b)) - O(\epsilon R) \\ &\geq d_Z(\gamma(a), \gamma(c)) + \Delta(\gamma', \epsilon R) + d_Z(\gamma(d), \gamma(b)) - O(\epsilon R). \end{aligned}$$

Also, by definition,

$$\Delta(\gamma, \epsilon R) \leq d_{\mathcal{Z}}(\gamma(a), \gamma(b)) + O(\epsilon R).$$

Hence,

$$\begin{aligned} \Delta(\gamma', \epsilon R) &\leq d_{\mathcal{Z}}(\gamma(a), \gamma(b)) - d_{\mathcal{Z}}(\gamma(a), \gamma(c)) - d_{\mathcal{Z}}(\gamma(d), \gamma(b)) + O(\epsilon R) \\ &\leq d_{\mathcal{Z}}(\gamma(c), \gamma(d)) + O(\epsilon R). \end{aligned}$$

This finishes the proof of part (2).

We prove part (3) for  $l = 2$ . The general case is similar. Consider the partition  $t_0, \dots, t_m$  such that the sum in (1) for  $\Delta(\gamma, \epsilon R)$  is within  $\epsilon R$  of the infimum. Let  $x_i = \gamma_1(t_i)$  and  $y_i = \gamma_2(t_i)$ . Since  $\mathcal{Z}$  is equipped with the  $L^1$ -metric, we have

$$\begin{aligned} \sum_{i=1}^m d_{\mathcal{Z}_1}(x_i, x_{i+1}) + d_{\mathcal{Z}_2}(y_i, y_{i+1}) \\ \leq d_{\mathcal{Z}_1}(\gamma_1(a), \gamma_1(b)) + d_{\mathcal{Z}_2}(\gamma_2(a), \gamma_2(b)) + O(\epsilon R). \end{aligned} \tag{3}$$

But, by the triangle inequality, we have

$$\sum_{i=1}^m d_{\mathcal{Z}_2}(y_i, y_{i+1}) \geq d_{\mathcal{Z}_2}(\gamma(a), \gamma(b)). \tag{4}$$

Subtracting (4) from (3) we obtain

$$\Delta(\gamma_1, \epsilon R) \leq \sum_{i=1}^m d_{\mathcal{Z}_1}(x_i, x_{i+1}) \leq d_{\mathcal{Z}_1}(\gamma_1(a), \gamma_1(b)) + O(\epsilon R).$$

That is,  $\gamma_1$  is  $\epsilon$ -efficient at scale  $R$ . The same holds for  $\gamma_2$ .

To see the last part, again let  $t_1, \dots, t_m$  be an almost optimal subdivision. (Recall that, in this case,  $\gamma$  is  $\epsilon^2$ -efficient.) Choose a subpartition  $s_1, \dots, s_l$  so that

$$|s_{i+1} - s_i| \stackrel{*}{\asymp} \epsilon R.$$

Then,  $l \stackrel{*}{\lesssim} \frac{1}{\epsilon}$ . Also, let  $\gamma'$  be the path obtained from composing  $\gamma$  with the closest point map to  $\mathcal{Z}'$ . Let  $z_i = \gamma(s_i)$  and  $z'_i = \gamma'(s_i)$ . Then

$$\begin{aligned} \Delta(\gamma', \epsilon R) &\leq \sum_{i=1}^l d_{\mathcal{Z}}(z'_i, z'_{i+1}) \\ &\leq \sum_{i=1}^l d_{\mathcal{Z}}(z'_i, z_i) + d_{\mathcal{Z}}(z_i, z_{i+1}) + d_{\mathcal{Z}}(z_{i+1}, z'_{i+1}) \end{aligned}$$

$$\begin{aligned} &\leq \Delta(\gamma, \epsilon R) + l \cdot O(\epsilon^2 R) \leq \Delta(\gamma, \epsilon^2 R) + O(\epsilon R) \\ &\leq d_{\mathcal{Z}}(z_1, z_l) + O(\epsilon R) \leq d_{\mathcal{Z}}(z'_1, z'_l) + O(\epsilon^2 R) + O(\epsilon R) \\ &\leq d_{\mathcal{Z}}(z'_1, z'_l) + O(\epsilon R). \end{aligned}$$

Above, in the third line, we used the fact that the  $s_i$ 's are a subpartition of an almost optimal partition. In the fourth line, we used the fact that the pairs  $z_1, z'_1$  and  $z_l, z'_l$  are  $O(\epsilon^2 R)$ -close. This finishes the proof.  $\square$

*Definition 2.4*

A box in  $\mathbb{R}^n$  is a product of intervals, namely,  $B = \prod_{i=1}^n I_i$ , where  $I_i$  is an interval in  $\mathbb{R}$ . We say a box  $B$  is of size  $R$  if, for every  $i$ ,  $|I_i| \stackrel{*}{\geq} R$  and if the diameter of  $B$  is less than  $R$ . Note that if  $B$  is of size  $R$  and of size  $R'$ , then  $R \stackrel{*}{\asymp} R'$ .

Throughout the article,  $\mathbb{R}^n$  is always equipped with the Euclidean metric. A map  $f : B \rightarrow \mathcal{Z}$  from a box of size  $R$  in  $\mathbb{R}^n$  to a metric space  $\mathcal{Z}$  is called  $\epsilon$ -efficient if, for any geodesic  $\gamma : [a, b] \rightarrow B$ , the path  $f \circ \gamma$  is  $\epsilon$ -efficient at scale  $R$ .

Let  $B = \prod_{i=1}^n I_i$  be a box of size  $L$  in  $\mathbb{R}^n$ . By a central sub-box  $\underline{B}$  we mean a sub-box of  $B$  with one-third the diameter and such that the distance from the face of  $\underline{B}$  with the  $j$ th coordinate fixed to the corresponding face of  $B$  is  $|I_j|/3$ . For any constant  $0 < R < L$ , let  $\mathcal{B}_R$  be a subdivision of  $\underline{B}$  into boxes of size  $R$ . That is,

- (1) boxes in  $\mathcal{B}_R$  are of size  $R$ ;
- (2) they are contained in  $\underline{B}$ , and hence, their distance to the boundary of  $B$  is at least  $\frac{L}{3}$ ;
- (3) they have disjoint interiors; and
- (4) their union is  $\underline{B}$ .

For any metric space  $\mathcal{Z}$ , we prove that any quasi-Lipschitz maps from  $B$  to  $\mathcal{Z}$  are coarsely differentiable almost everywhere in a central box of comparable size. This is Theorem D from the Introduction, which we repeat.

**THEOREM 2.5 (Coarse differentiation)**

*For every  $K, C, \epsilon_0, \theta_0$ , and  $R_0$ , there is  $L_0$  depending on these constants so that the following holds. For  $L \geq L_0$ , let  $f : B \rightarrow \mathcal{Z}$  be a  $(K, C)$  quasi-Lipschitz map where  $B$  is a box of size  $L$  in  $\mathbb{R}^n$ . Then there is  $R \geq R_0$  so that the proportion of boxes  $B' \in \mathcal{B}_R$  where  $f|_{B'}$  is  $\epsilon_0$ -efficient at scale  $R$  is at least  $(1 - \theta_0)$ .*

*Remark 2.6*

Note that the size of the error,  $O(\epsilon_0 R)$ , depends on the size  $R$  of the boxes. An  $\epsilon_0$ -efficient map from a much larger box is allowed to have a much larger error. What we control is the size of the error as a proportion of the size of the box.

Theorem 2.5 is stronger than what we need, as we will need only one efficient box. However, this more general statement will be useful in the upcoming paper [14] proving the rigidity of Teichmüller space.

2.1. *Choosing the correct scales*

We first prove a much coarser differentiation statement. In a sense, the statement of Theorem 2.5 is a direct analogue of Rademacher’s theorem, but the proof of Rademacher’s theorem is a direct analogue of the proof of Theorem 2.8 below.

In what follows we assume that  $\epsilon$  is always of the form  $\epsilon = \frac{1}{N}$  for a large positive integer  $N$ .

Definition 2.7

A family  $\mathcal{F}$  of geodesics in  $\mathbb{R}^n$  is called *locally finite* if, for any compact subset  $B$  of  $\mathbb{R}^n$ , only finitely many geodesics in  $\mathcal{F}$  intersect  $B$ .

Let  $B = \prod_{i=1}^n I_i$  be a box with diameter precisely  $L$ . Define  $\mathcal{F}_B$  to be the collection of restrictions of paths in  $\mathcal{F}$  to  $B$  that are long. More precisely, let

$$\mathcal{F}_B = \left\{ \gamma \mid \gamma = \gamma' \cap B, \gamma' \in \mathcal{F}, |\gamma| \geq \frac{2L}{5} \right\}.$$

For  $\gamma \in \mathcal{F}_B$  we say a set of points  $G(\gamma, r)$  along  $\gamma$  is an *r-grid for  $\gamma$*  if they subdivide  $\gamma$  into segments of size exactly  $r$ , except perhaps for the two end segments, which may have a size less than  $r$ . An *r-grid* is an example of an *r-partition*. An *r-grid  $G(r)$*  is a collection of *r-grids* for every segment in  $\mathcal{F}_B$ . Since the ratio of lengths of any two segments in  $\mathcal{F}_B$  is bounded by  $\frac{5}{2}$ , by taking into account the endpoints, the ratio of the number of points in any *r-grid* in two different such segments is bounded by 3. When an *r-grid*  $G(\gamma, r) = \{p_1, \dots, p_k\}$  is fixed, we define

$$\bar{\Delta}(\gamma, r) = \sum_{i=1}^{k-1} d_Z(f(p_i), f(p_{i+1})).$$

This is essentially the same as the definition of  $\Delta$ , except that the sum here is over a fixed *r-grid* instead of an infimum over all *r-partitions*. The infimum could a priori be much smaller. Given a scale  $R$  and a segment  $\gamma \in \mathcal{F}_B$  with an  $\epsilon R$ -grid  $G(\gamma, \epsilon R)$ , we define  $\mathcal{F}(\gamma, R)$  to be the set of *all* subsegments of  $\gamma$  of length  $R$  that start and end at points in  $G(\gamma, \epsilon R)$ . This makes sense since  $\epsilon = \frac{1}{N}$ . We also define

$$\mathcal{F}_B(R) = \bigcup_{\gamma \in \mathcal{F}_B} \mathcal{F}(\gamma, R).$$

**THEOREM 2.8**

Let  $\mathcal{F}$  be a locally finite family of geodesics in  $\mathbb{R}^n$ . For any  $K, C, R_0, \epsilon = \frac{1}{N}$ , and  $\theta > 0$ , there exists a constant  $L_0$  such that the following holds. Let  $L > L_0$ , let  $B \subset \mathbb{R}^n$  be a box of size  $L$ , and let  $f : B \rightarrow \mathcal{Z}$  be a  $(K, C)$ -quasi-Lipschitz map. Then, there exist a scale  $R \geq R_0$  and an  $\epsilon R$ -grid  $G(\epsilon R)$  such that, for at least a fraction  $(1 - \theta)$  of segments  $\gamma \in \mathcal{F}_B(R)$ ,

$$\overline{\Delta}(\gamma, \epsilon R) \stackrel{*}{\prec} d_{\mathcal{Z}}(f(a), f(b)) + \epsilon R, \tag{5}$$

where  $a, b$  are the endpoints of  $\gamma$ .

Note that (5) implies that  $\gamma'$  is  $\epsilon$ -efficient at scale  $R$  since a grid is an example of a partition. Using grids is more suitable for our proof. Effectively, the theorem states that  $f|_B$  is nearly affine at scales  $R$ , up to an error of  $O(\epsilon R)$ .

*Remark 2.9*

In the theorem,  $R$  depends on  $\epsilon, \theta, K, C$ , and also on  $B$ . However, the proof will find  $R$  as one of finitely many values as long as  $\epsilon, \theta, K$ , and  $C$  are fixed. In fact, the ratio  $L/R$  is bounded: we have  $R \geq \epsilon^{-M} L$  for  $M \stackrel{*}{\asymp} \frac{K}{\epsilon\theta}$ .

We begin with the following lemma, which is in some sense the heart of the matter.

**LEMMA 2.10**

Suppose that an interval  $I$  is divided into  $N_0$  segments  $I_j = [c_j, d_j]$  of length  $r$  with  $c_{j+1} = d_j$  and that  $\gamma$  is a path defined on  $I$ . Suppose that each segment  $I_j$  is further subdivided into  $\frac{1}{\epsilon}$  segments, each of length  $\epsilon r$ . Let  $\theta$  be the proportion of the segments  $I_j$  for which (5) fails. Then

$$\overline{\Delta}(\gamma, \epsilon r) - \overline{\Delta}(\gamma, r) \stackrel{*}{\succ} \epsilon r \theta N_0.$$

*Proof*

On each segment  $I_j$  for which (5) fails, we have

$$\overline{\Delta}(\gamma|_{I_j}, \epsilon r) - d_{\mathcal{Z}}(\gamma(d_j), \gamma(c_j)) \geq \epsilon r.$$

We sum this up over all such intervals. The number of intervals is  $\theta N_0$ . Adding the intervals for which (5) holds only increases the difference. □

*Proof of Theorem 2.8*

Pick  $r_0 \geq \max\{R_0, C\}$  ( $C$  is the additive error in the definition of a quasi-Lipschitz map), and inductively let

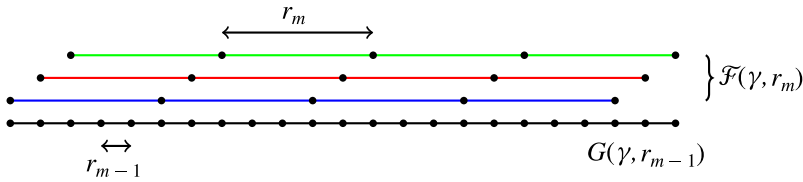


Figure 2. The segments in  $\mathcal{F}(\gamma, r_m)$  can be grouped into nonoverlapping segments where each group essentially covers  $\gamma$ . The choice of  $r_m$ -grid  $G(\gamma, r_m)$  maximizes the proportion of nonoverlapping subsegments of  $\gamma$  that are not  $\epsilon$ -efficient at scale  $r_{m-1}$ .

$$r_m = \frac{r_{m-1}}{\epsilon} = Nr_{m-1}.$$

Let  $M$  be a large positive integer (to be determined below), and let  $L_1 = r_M$ . Choose an arbitrary  $r_0$ -grid  $G(r_0)$  for  $\mathcal{F}_B$ . Recall that  $\mathcal{F}_B(r_1)$  is the set of all segments of length  $r_1$  with endpoints in  $G(r_0)$ . Let  $\theta_1$  be the fraction of segments in  $\mathcal{F}_B(r_1)$  for which (5) does not hold with  $R = r_1$ . If  $\theta_1 \leq \theta$ , then we are done with  $R = r_1$ . Thus, assume  $\theta_1 > \theta$ .

For  $\gamma \in \mathcal{F}_B$ , we choose an  $r_1$ -grid  $G(\gamma, r_1) \subset G(\gamma, r_0)$  as follows. Note that any grid  $G(\gamma, r_1)$  is essentially a *decomposition* of  $\gamma$  into segments of length  $r_1$  (except for the subsegments at the ends). That is, we are choosing a nonoverlapping collection of segments in  $\mathcal{F}(\gamma, r_1)$  so that the next segment starts where the previous segment ends. There are at most  $N$  choices depending on where one chooses the closest point to the endpoints of  $\gamma$  (see Figure 2). We choose  $G(\gamma, r_1)$  among all such grids, which maximizes the proportion  $\theta_1(\gamma)$  of these nonoverlapping subsegments of  $\gamma$  that do not satisfy (5) with  $R = r_1$ . Any subsegment of  $\gamma$  of length  $r_1$  will occur in some grid. Furthermore, as we have observed, by the definition of  $\mathcal{F}_B$  and taking into account endpoints, for  $\gamma, \gamma' \in \mathcal{F}_B$ , the ratio of the number of  $r_1$ -segments in  $\gamma$  and in  $\gamma'$  is at most 3. These two facts imply that the average

$$\frac{\sum_{\gamma} \theta_1(\gamma)}{|\mathcal{F}_B|} \geq \frac{\theta_1}{3} > \frac{\theta}{3}.$$

For  $m = 1, \dots, M$ , we proceed in the same way. If  $\theta_m \leq \theta$ , then we are done. Otherwise, for every  $\gamma$ , we choose the grid  $G(\gamma, r_m)$  where the proportion  $\theta_m(\gamma)$  of segments that do not satisfy (5) with  $R = r_{m+1}$  is maximum (see Figure 2). Again we have

$$\frac{\sum_{\gamma} \theta_m(\gamma)}{|\mathcal{F}_B|} \geq \frac{\theta_m}{3} > \frac{\theta}{3}.$$

We show that if  $M$  is large enough, then this contradicts the assumption that  $f$  is quasi-Lipschitz. First, note that Lemma 2.10 applied with  $r = r_m$  and  $r_{m-1} = \epsilon r_m$

gives

$$\overline{\Delta}(\gamma, r_{m-1}) - \overline{\Delta}(\gamma, r_m) \succ^* \epsilon r_m \theta_m(\gamma) |G(\gamma, r_m)| \overset{*}{\asymp} \epsilon L \theta_m(\gamma).$$

After iterating this over  $m$  as  $m$  goes from  $M$  down to 1 we get

$$\overline{\Delta}(\gamma, r_0) - \overline{\Delta}(\gamma, r_M) \succ^* \epsilon L \sum_{m=1}^M \theta_m(\gamma).$$

Using the fact that  $f$  is quasi-Lipschitz and that  $r_0 > C$ , we have

$$\Delta(\gamma, r_0) \leq \frac{L}{r_0} (Kr_0 + C) \overset{*}{\prec} KL.$$

Hence,

$$KL \succ^* \epsilon L \sum_{m=1}^M \theta_m(\gamma),$$

and thus,

$$\frac{K}{\epsilon} \overset{*}{\succ} \sum_{m=1}^M \theta_m(\gamma).$$

Averaging over all geodesics  $\gamma \in \mathcal{F}_B$  gives

$$\frac{K}{\epsilon} \overset{*}{\succ} \sum_{m=1}^M \left( \frac{1}{|\mathcal{F}_B|} \sum_{\gamma \in \mathcal{F}_B} \theta_m(\gamma) \right) \overset{*}{\succ} \sum_{m=1}^M \frac{\theta_m}{3} \geq \frac{M\theta}{3}.$$

Choosing  $M$  large enough, we obtain a contradiction. Hence, for some  $m$ ,  $\theta_m \leq \theta$  and we are done. □

*Proof of Theorem 2.5*

Let  $\epsilon_0 < 1$ ,  $K$ ,  $C$ ,  $\theta_0$ , and  $R_0$  be given. Choose a family  $\mathcal{F}$  of geodesics in  $\mathbb{R}^n$  as follows. Pick a finite set of vectors  $\mathcal{V}$  in the unit sphere  $S^{n-1} \subset \mathbb{R}^n$  that is  $(\epsilon_0)^2$ -dense in  $S^{n-1}$  with size  $|\mathcal{V}| \overset{*}{\asymp} \frac{1}{(\epsilon_0)^2}$ . For a direction  $\vec{v} \in \mathcal{V}$ , let  $\mathcal{F}_{\vec{v}}$  be a family of parallel lines in the direction  $\vec{v}$  where the distance between nearby lines is comparable to 1. Then

$$\mathcal{F} = \bigcup_{\vec{v} \in \mathcal{V}} \mathcal{F}_{\vec{v}}$$

is a locally finite family of geodesics in  $\mathbb{R}^n$ .

Let  $\epsilon = \epsilon_0^2$ , and let  $\theta$  be much larger than  $\theta_0 \epsilon^{n+2}$ , with the amount to be determined below. Apply Theorem 2.8 to obtain the constant  $L_0$ . Assume that a box  $B$  of size  $L \geq L_0$  and a  $(K, C)$ -quasi-Lipschitz map  $f : B \rightarrow \mathcal{Z}$  are given. Let  $R$  be the scale, and let  $\mathcal{G}(\epsilon R)$  be the grid given by Theorem 2.8 for this finite family. Then Theorem 2.8 says that at least a proportion  $(1 - \theta)$  of the segments in  $\mathcal{F}$  are  $\epsilon$ -efficient at scale  $R$ .

Let  $\mathcal{B}_R$  be a collection of disjoint sub-boxes of  $B$  giving a decomposition of a central box in  $B$  as in the statement of Theorem 2.5. Let  $B' \in \mathcal{B}$  be a box that contains a geodesic  $\beta$  that is not  $\epsilon_0$ -efficient at scale  $R$ . Let  $\vec{v}$  be the direction closest to the direction of  $\beta$ . Let

$$\mathcal{NE}(B', \vec{v}) \subset \mathcal{F}_B(R)$$

be the set of geodesic segments in  $\mathcal{F}_B(R)$  that are in the direction of  $\vec{v}$ , intersect  $B'$ , and are not  $\epsilon_0$ -efficient at scale  $R$ .

*Claim 2.11*

Every geodesic in  $\mathcal{F}_{\vec{v}}$  that intersects an  $\epsilon R$ -neighborhood of  $\beta$  contains a segment in  $\mathcal{NE}(B', \vec{v})$ .

*Proof*

Assume that  $\gamma \in \mathcal{F}_{\vec{v}}$  intersects an  $\epsilon R$ -neighborhood of  $\beta$ . Condition (2) of the description of  $\mathcal{B}_R$  implies that  $|\gamma \cap B| \geq \frac{L}{3}$ . Choose a grid  $G(\epsilon R)$  for  $\gamma \cap B$ . The subsegments of length  $R$  of  $\gamma$  that start and end in  $G(\epsilon R)$  are included in  $\mathcal{F}_{B,R}$ . Since the difference between the direction of  $\gamma$  and  $\beta$  is at most  $\epsilon = \epsilon_0^2$ ,  $\beta$  is contained in an  $O(\epsilon R)$ -neighborhood of  $\gamma$ . Also, the length of  $\beta$  is less than the diameter of  $B'$ , which is less than  $R$ . Hence, there is a segment  $\gamma_0 \in \mathcal{F}_B(\gamma, R)$  of length  $R$  such that  $\beta$  is included in an  $O(\epsilon R)$ -neighborhood of  $\gamma_0$  (refer to Figure 3). We show that the

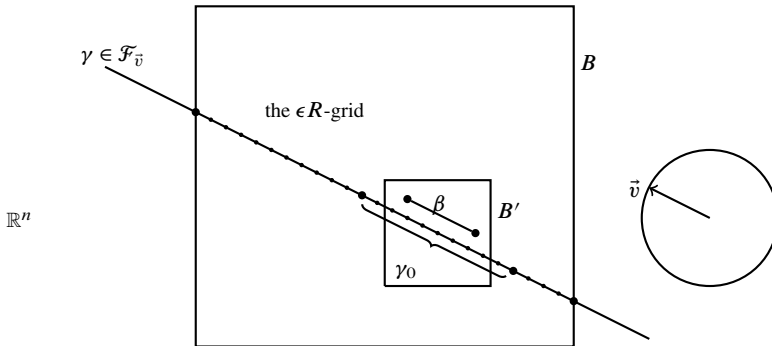


Figure 3. The arc  $\gamma$  is in an  $\epsilon R$ -neighborhood of  $\beta_0$ .



assumption that  $\beta$  is not  $\epsilon_0$ -efficient at scale  $R$  implies that  $\gamma_0$  is not  $\epsilon$ -efficient at scale  $R$ , which is what we claimed.

Assume, for contradiction, that  $\gamma_0$  is  $\epsilon$ -efficient at scale  $R$ . Then every subsegment of  $\gamma_0$  is also efficient at scale  $R$  by Lemma 2.3(2). Choose a subsegment  $\gamma_1$  of  $\gamma_0$  so that the endpoints of  $\gamma_1$  and  $\beta$  are  $O(\epsilon R)$ -close. We now apply Lemma 2.3(4) with  $Z' = f(\gamma_0)$  to conclude that  $\beta$  is  $\epsilon_0$ -efficient, which is a contradiction. This proves the claim.  $\square$

We continue with the proof of the theorem. Let  $\mathcal{F}_{B'}(R)$  be the subset of  $\mathcal{F}_B(R)$  consisting of segments that intersect  $B'$ . In every direction  $\vec{v} \in \mathcal{V}$  there are at most  $O(\frac{R^{n-1}}{\epsilon})$  segments in  $\mathcal{F}_{B'}(R)$ . This is because a cross section of  $B'$  perpendicular to  $\vec{v}$  has area  $O(R^{n-1})$ , the grid has size  $\epsilon R$ , and the segments have length  $R$ . The number of geodesics in  $\mathcal{F}_{\vec{v}}$  that intersect an  $\epsilon R$ -neighborhood of  $\beta$  is on the order of  $(\epsilon R)^{n-1}$  (which is the area of a cross section of an  $\epsilon R$ -neighborhood of  $\beta$  perpendicular to  $\vec{v}$ ). That is,

$$\frac{|\mathcal{N}\mathcal{E}(B', \vec{v})|}{|\mathcal{F}_{B'}(R)|} \succ^* \frac{1}{|\mathcal{V}|} \cdot \frac{(\epsilon R)^{n-1}}{\frac{R^{n-1}}{\epsilon}} \prec^* \epsilon^{n+2},$$

since  $|\mathcal{V}| \prec^* \frac{1}{\epsilon^2}$ . Note that a definite proportion of segments in  $\mathcal{F}_B(R)$  intersect some box  $B' \in \mathcal{B}$ , and each segment in  $\mathcal{F}_B(R)$  intersects at most a uniform number of boxes. Hence,

$$|\mathcal{F}_B(R)| \prec^* \sum_{B' \in \mathcal{B}} |\mathcal{F}_{B'}(R)|.$$

Define

$$\mathcal{N}\mathcal{E}(B') = \bigcup_{\vec{v} \in \mathcal{V}} \mathcal{N}\mathcal{E}(B', \vec{v}) \quad \text{and} \quad \mathcal{N}\mathcal{E} = \bigcup_{B' \in \mathcal{B}} \mathcal{N}\mathcal{E}(B').$$

Assume that the proportion of boxes  $B'$  that contain a nonefficient segment is larger than  $\theta_0$ . Since the sizes of  $\mathcal{F}_{B'}(R)$  are comparable for every  $B'$ , we have

$$\frac{|\mathcal{N}\mathcal{E}|}{|\mathcal{F}_B(R)|} \prec^* \sum_{B' \in \mathcal{B}} \frac{|\mathcal{N}\mathcal{E}(B')|}{|\mathcal{F}_{B'}(R)|} \succ^* \theta_0 \epsilon^{n+2}.$$

We have shown that there is a uniform constant  $c$  (only depending on  $K, C$ , and the dimension  $n$ ) such that at least a proportion  $c\theta_0\epsilon^{n+2}$  of segments are not  $\epsilon$ -efficient at scale  $R$ . If we chose  $\theta < c\theta_0\epsilon^{n+2}$ , then we have a contradiction to Theorem 2.8. The contradiction finishes the proof.  $\square$

2.2. *Efficient maps into a hyperbolic space*

The following is the first use of efficient paths when the target is Gromov-hyperbolic and is similar to the familiar Morse argument.

LEMMA 2.12

Suppose that  $\mathcal{X}$  is a Gromov-hyperbolic space and that  $\gamma: [a, b] \rightarrow \mathcal{X}$  is  $(K, C)$ -quasi-Lipschitz and  $\epsilon$ -efficient at scale  $R$ . Then  $\gamma$  stays in an  $O(\epsilon R)$ -neighborhood of a geodesic  $\ell$  joining  $\gamma(a)$  and  $\gamma(b)$ .

*Proof*

Before beginning the proof we remark that we do not actually need that the space  $\mathcal{X}$  is Gromov-hyperbolic; we only need that there is a contraction map to  $\gamma$ .

We begin the proof. To simplify the choice of constants, assume that  $\gamma$  is continuous. Recall that we are assuming  $\epsilon R \geq 1$ . Because  $\mathcal{X}$  is hyperbolic there are uniform constants  $B, R_0$ , and  $c_0 > 0$  such that if a point is distance  $L \geq R_0$  from  $\ell$ , then the closest point projection to  $\ell$  of a ball of radius  $c_0 L$  about the point has diameter at most  $B$ . Fix a large but uniform constant  $D_0 > \max(B, C)$  to be determined later. Suppose, for  $M > D_0$ , that the path  $\gamma$  leaves an  $M\epsilon R$ -neighborhood of  $\ell$ . We can find times  $c, d \in [a, b]$  so that, at these times,  $\gamma$  is distance  $D_0\epsilon R$  from  $\ell$ ; for  $t \in [c, d]$ ,  $\gamma(t)$  is at least  $D_0\epsilon R$  away from  $\ell$ , so that in between  $c$  and  $d$  the path  $\gamma$  travels to some point at maximum distance  $M\epsilon R$  from  $\ell$ . We wish to bound  $M$ .

By Lemma 2.3,  $\gamma' = \gamma|_{[c,d]}$  is still  $\epsilon$ -efficient at scale  $R$ . Let

$$c = t_1 < \dots < t_k = d$$

be a partition so that, for  $z_i = \gamma(t_i)$ ,

$$\Delta(\gamma', \epsilon R) \succ \sum_i d_{\mathcal{X}}(z_{i+1}, z_i). \tag{6}$$

(In fact, by the definition of  $\Delta$ , we can choose  $t_i$  so that the two sides are equal. However, we are about to modify the partition  $t_i$ .) Now since  $c_0$  is a fixed constant,  $t_{i+1} - t_i \leq \epsilon R$ , and the map  $\gamma$  is quasi-Lipschitz, we can remove some of the times  $t_i$ , so that for  $D_0$  large enough

$$\frac{c_0}{2} D_0 \epsilon R \leq d_{\mathcal{X}}(z_{i+1}, z_i) \leq c_0 D_0 \epsilon R.$$

Let  $N$  be the number of points in the new partition. Note that after removing points from the partition, by the triangle inequality, the right-hand side of (6) decreases, and so we can assume that

$$\Delta(\gamma', \epsilon R) \geq \frac{N c_0 D_0 \epsilon R}{2}. \tag{7}$$

We also have

$$N \geq \frac{M}{c_0 D_0}. \tag{8}$$

The contraction property implies that the projection map  $\pi : \mathcal{X} \rightarrow \ell$  satisfies

$$d_{\mathcal{X}}(\pi(z_i), \pi(z_{i+1})) \leq B.$$

By using these projected points to  $\ell$ , and since  $\gamma(c)$  and  $\gamma(d)$  are at a distance  $D_0 \epsilon R$  from  $\ell$ , we have

$$d_{\mathcal{X}}(\gamma(c), \gamma(d)) \leq 2D_0 \epsilon R + NB.$$

On the other hand,

$$\Delta(\gamma', \epsilon R) \succ \sum d_{\mathcal{X}}(z_{i+1}, z_i) \geq \frac{Nc_0 D_0 \epsilon R}{2}.$$

From the assumption that  $\gamma'$  is  $\epsilon$ -efficient at scale  $R$ , the above two inequalities give

$$\frac{Nc_0 D_0 \epsilon R}{2} \leq D_0 \epsilon R + NB + O(\epsilon R).$$

Now we can choose  $D_0$  large in terms of  $B$  so that the above inequality implies that  $N$  is uniformly bounded in terms of  $B$  as well. Hence,  $M$  is also uniformly bounded by (8). This finishes the proof.  $\square$

We now consider an efficient map from a box to a hyperbolic space. First we need the following lemma.

LEMMA 2.13

*Given  $n$  and  $N$ , there is  $\sigma = \sigma(n, N) > 0$  such that, for each  $L$ , if  $\{C_i\}$  is a collection of  $N$  convex bodies in  $\mathbb{R}^n$  that cover a ball  $B$  of radius  $L$ , then some  $C_i$  contains a ball of radius  $\sigma L$ .*

*Proof*

For a convex set  $C$ , let  $R = R(C)$  be the out-radius of  $C$ , that is, the radius of the smallest ball that contains it. Let  $r = r(C)$  be the in-radius, that is, the radius of the largest ball contained in the set, and let  $w = w(C)$  be the width, that is, the minimum distance between supporting hyperplanes.

From [18, Theorem 1] we have, for some  $\kappa = \kappa(n) > 0$  and any convex set  $C$ , that

$$\text{Vol}(C) \leq \kappa R^n \int_0^{\arcsin \frac{w}{2R}} \cos^n \theta \, d\theta \leq \kappa R^n \arcsin \frac{w}{2R} \leq \kappa R^n \arcsin \frac{cr}{2R}. \tag{9}$$

The last inequality follows from the Steinhagen inequality, which states that there is a constant  $c = c(n) > 0$  such that

$$w \leq cr.$$

Since the convex sets  $C_i$  cover the ball of radius  $L$ , for some  $c' > 0$ , there is some  $C = C_i$  with

$$\text{Vol}(C) \geq \frac{c'L^n}{N}.$$

This implies that  $R = R(C) \geq c''L$  for some constant  $c'' = c''(N) > 0$ .

We will show that  $r = r(C) \geq \sigma L$  for a uniform constant  $\sigma$  by arguing in two cases. Assume that

$$\arcsin \frac{cr}{2R} \geq \frac{\pi}{4}.$$

Then  $\frac{cr}{2R} \geq \frac{\sqrt{2}}{2}$  and so

$$cr \geq \sqrt{2}R \geq \sqrt{2}c''L,$$

and we are done by taking  $\sigma = \frac{\sqrt{2}c''}{c}$ . Now assume that

$$\arcsin \frac{cr}{2R} \leq \frac{\pi}{4}$$

so that

$$\arcsin \frac{cr}{2R} \leq \frac{cr}{R}.$$

But then (9) gives

$$c'L^n/N \leq \text{Vol}(C) \leq \kappa R^n \frac{cr}{R} = \kappa c R^{n-1} r,$$

so

$$r \geq \frac{c'}{cN\kappa} \left(\frac{L}{R}\right)^{n-1} L,$$

and again we are done by taking  $\sigma = \frac{c'}{c(c'')^{n-1}N\kappa}$ . □

**PROPOSITION 2.14**

*Suppose that  $\mathcal{Z}$  is a Gromov-hyperbolic space and that  $f : B \rightarrow \mathcal{Z}$  is an  $\epsilon$ -efficient map at scale  $R$  from a box of size  $R$  in  $\mathbb{R}^n$  to  $\mathcal{Z}$ . Then, there is a sub-box  $B' \subset B$  with  $|B'| \stackrel{*}{\asymp} |B|$ , so that the image  $f(B')$  lies in an  $O(\epsilon R)$ -neighborhood of a line  $\ell'$  in  $\mathcal{Z}$ .*

*Proof*

By taking a sub-box, we assume that  $B = [0, R]^n$ , and we let the  $\ell_i$ 's be the edges of the box  $B$ . Given a line  $\ell \subset B$ , denote by  $\ell'$  a geodesic in  $\mathcal{Z}$  joining the image of its endpoints.

We first prove by induction on  $n$  that, for any  $q \in B$ ,  $f(q)$  is within  $O(\epsilon R)$  of a point in  $\ell'_i$ , the geodesic corresponding to the image of one of the edges. We start with  $n = 2$  and with  $\ell_1, \ell_2, \ell_3, \ell_4$  being the four edges of  $B$  arranged in counterclockwise order. We have by Lemma 2.12 that each point of  $f(\ell_i)$  is within  $O(\epsilon R)$  of  $\ell'_i$ . Now let  $\ell_q$  be the line through  $q$  parallel to  $\ell_1$ , and let  $\ell'_q$  be the corresponding geodesic in the image. Take the rectangle with sides  $\ell_q$  and  $\ell_1$  and subsegments  $m_2 \subset \ell_2$  and  $m_4 \subset \ell_4$ . Lemma 2.12 implies that the endpoints of  $m'_2$  are within  $O(\epsilon R)$  of  $\ell'_2$ . The same holds for  $m'_4$  and  $\ell'_4$ . The quadrilateral bounded by  $\ell'_1, m'_2, \ell'_q, m'_4$  is  $O(1)$ -thin, which implies that  $f(q)$  is within  $O(\epsilon R)$  of one of the other three sides and therefore within  $O(\epsilon R)$  of one of the  $\ell'_i$ 's.

Now suppose that the statement is true for boxes in  $\mathbb{R}^{n-1}$  and  $B \subset \mathbb{R}^n$ . Take again the geodesics  $\ell_i$  that correspond to the edges of the box  $B$  and any point  $q \in B$ . It lies on a face  $B_q^{n-1}$  parallel to a face of  $B$ . Let  $\tau_i$  be the edges of  $B_q^{n-1}$ . By induction,  $f(q)$  lies within  $O(\epsilon R)$  of some  $\tau'_i$ . Since each  $\tau_i$  itself lies in an  $(n - 1)$ -dimensional face of  $B$ , again by induction, each point of  $f(\tau_i)$  lies within  $O(\epsilon R)$  of the union of the  $\ell'_i$ 's. Thus,  $f(q)$  is within  $O(\epsilon R)$  of some  $\ell'_i$ , completing the induction step.

Now, fix any  $n$  and one of the geodesics  $\ell'_i$ . If  $n + 1$  points  $q_1, \dots, q_{n+1}$  span an  $n$ -simplex  $\Lambda$  and are such that each  $f(q_j)$  is within  $O(\epsilon R)$  of  $\ell'_i$ , then the image under  $f$  of every point of  $\Lambda$  is within  $O(\epsilon R)$  of  $\ell'_i$ . By the Carathéodory theorem, the convex hull of the set of points mapped within  $O(\epsilon R)$  of  $\ell'_i$  is the union of such simplices, and therefore, the convex hull is a convex set of points mapped within  $O(\epsilon R)$  of  $\ell'_i$ . We conclude by Lemma 2.13 that there is a box  $B'$  with  $|B'| \stackrel{\approx}{\asymp} |B|$  consisting of points mapped within  $O(\epsilon R)$  of one of the  $\ell'_i$ 's.  $\square$

**3. Combinatorial model**

Let  $S$  be a possibly disconnected surface of finite hyperbolic type. Define the complexity of  $S$  to be

$$\xi(S) = \sum_W (3g_W + p_W - 4), \tag{10}$$

where the sum is over all connected components  $W$  of  $S$ ,  $g_W$  is the genus of  $W$ , and  $p_W$  is the number of punctures.

Let  $(\mathcal{T}(S), d_{\mathcal{T}})$  represent Teichmüller space equipped with the Teichmüller metric, let  $(\mathcal{T}(S), d_{WP})$  represent Teichmüller space equipped with the Weil–Petersson metric, and let  $(\text{Mod}(S), d_W)$  represent the mapping class group equipped with the word metric. Here one chooses a finite generating set and builds the Cayley graph.

A vertex of the graph is an element  $g$  of  $\text{Mod}(S)$ . One connects  $g_1, g_2$  if there is a generator  $h$  such that  $g_1 = g_2h$ . We construct combinatorial models for these spaces.

Let  $\mathcal{P}(S)$  be the pants complex of  $S$ . A vertex of  $\mathcal{P}(S)$  is a maximum number of disjoint essential nonhomotopic simple closed curves. Two are connected with an edge if there is an elementary move that changes one pants decomposition to the other. In an elementary move, one changes only one curve, and the new curve intersects the original one a minimum number of times. Define a marking  $(P, \{\tau_\alpha\}_{\alpha \in P})$  to be a pants decomposition together with a transverse curve  $\tau_\alpha$  for each pants curve  $\alpha$ . The transverse curves are assumed to be disjoint from other curves in  $P$  and to intersect  $\alpha$  minimally. Again this can be made into a graph. There are two types of moves that define the edges: either a Dehn twist about a pants curve applied to a transverse curve or a replacement of a pants curve with its transversal and an introduction of a new transversal (see [20] for more details). The marking complex is denoted by  $\mathcal{M}(S)$ .

An augmented marking  $(P, \{\tau_\alpha\}_{\alpha \in P}, \{\ell_\alpha\}_{\alpha \in P})$  is a marking together with a positive real number  $\ell_\alpha$  (the length of  $\alpha$ ) associated to every pants curve  $\alpha$ . The length of each curve is assumed to be less than the Bers constant for the surface  $S$ . The space of augmented markings is denoted by  $\mathcal{AM}(S)$  (see [26] and also [11] for a slightly different definition and extensive discussion of  $\mathcal{AM}(S)$ ). We will use these spaces as combinatorial models for, respectively,  $(\mathcal{T}(S), d_{WP})$ ,  $(\text{Mod}(S), d_W)$ , and  $(\mathcal{T}(S), d_{\mathcal{T}})$ . Assume that  $\mathcal{X} = \mathcal{X}(S)$  is one of these model spaces. Later in this section we will equip  $\mathcal{X}$  with a coarse metric.

### 3.1. Curve complex

Let  $W$  be an essential subsurface of  $S$ . We always assume that a subsurface is connected (unless specified otherwise) and that the embedding  $W \subset S$  induces an injective map  $\pi_1(W) \rightarrow \pi_1(S)$ . We also exclude the cases where  $W$  is a thrice-punctured sphere or an annulus going around a puncture.

Let  $\mathcal{C}(W)$  be the curve graph of  $W$  with metric  $d_{\mathcal{C}(W)}$ . This is a graph where the vertices are free homotopy classes of nontrivial nonperipheral simple closed curves (henceforth, simply referred to as *curves*) and edges are pairs of curves intersecting minimally (see [20] for the precise definition and discussion). We make a special definition for the case of annuli. For an annulus  $A$ ,

- $\mathcal{C}(A)$  is a horoball in  $\mathbb{H}^2$  when  $\mathcal{X} = \mathcal{AM}(S)$ ;
- $\mathcal{C}(A)$  is  $\mathbb{Z}$  when  $\mathcal{X} = \mathcal{M}(S)$ ; and
- $\mathcal{C}(A)$  is a point when  $\mathcal{X} = \mathcal{P}(S)$ .

The curve complex of every subsurface is Gromov-hyperbolic in all cases. This is clear when  $W$  is an annulus and is a theorem of Masur and Minsky (see [19, Theorem 1.1]) in other cases.

For every subsurface  $W$  of  $S$ , there is a coarsely defined projection map (see [20] for a general discussion and [11] for the case of augmented markings)

$$\pi_W : \mathcal{X} \rightarrow \mathcal{C}(W).$$

We sketch the definition here. Assume first that  $W$  is not an annulus. Given  $x \in \mathcal{X}$  (recall that in all three cases  $x$  contains a pants decomposition which we denote by  $P_x$ ), choose any pants curve  $\gamma \in P_x$  that intersects  $W$ . If  $\gamma \subset W$ , then choose the projection to be  $\gamma$ . If  $\gamma$  is not contained in  $W$ , then  $\gamma \cap W$  is a collection of arcs with endpoints on  $\partial W$ . Choose one such arc, and perform a surgery using this arc and a subarc of  $\partial W$  to find a point in  $\mathcal{C}(W)$ . The choice of different arcs or different choices of intersecting pants curves determines a set of diameter 2 in  $\mathcal{C}(W)$ ; hence, the projection is coarsely defined.

For annuli  $A$ , the definition is slightly different. When  $\mathcal{X} = \mathcal{P}(S)$  the projection map is trivially defined since  $\mathcal{C}(A)$  is just a point. When  $\mathcal{X} = \mathcal{M}(S)$ , consider the annular cover  $\tilde{A}$  of  $S$  associated to  $A$ . Identify the space of arcs in  $\tilde{A}$  (homotopy classes of arcs connecting different boundaries of  $\tilde{A}$  relative to their endpoints) with  $\mathbb{Z}$  by identifying some arc  $\omega_0$  with zero and sending every other arc  $\omega$  to the signed intersection number between  $\omega$  and  $\omega_0$ . Define  $\pi_A(x)$  by lifting the pants deposition  $P_x$  and transverse curves  $\tau_\alpha$  to  $\tilde{A}$ . At least one of these curves lifts to an arc connecting different boundaries of  $\tilde{A}$ , and different ones have bounded intersection number. Hence, the map is coarsely defined. We refer to this number as the *twisting number* of  $x$  around  $\alpha$  and denote it by  $\text{twist}_\alpha(x)$ , which is a coarsely defined integer.

Now consider the case  $\mathcal{X} = \mathcal{AM}(S)$ . Let  $B$  be the Bers constant of the surface  $S$ . For an annulus  $A$ , we identify  $\mathcal{C}(A)$  with the subset  $H \subset \mathbb{R}^2$  of all points in  $\mathbb{R}^2$  where the  $y$ -coordinate is larger than  $1/B$ . Note that, for an augmented marking  $x = (P, \{\tau_\alpha\}, \{l_\alpha\})$ , the twisting number  $\text{twist}_\alpha(x)$  can still be defined as above. If the core curve of  $A$  is in  $P_x$ , we define

$$\pi_A(x) = (\text{twist}_\alpha(x), 1/l_\alpha),$$

otherwise

$$\pi_A(x) = (\text{twist}_\alpha(x), 1/B).$$

Also, for subsurfaces  $U$  and  $V$  we have a projection map

$$\pi_{U,V} : \mathcal{C}(U) \rightarrow \mathcal{C}(V),$$

which is defined on the subset of  $\mathcal{C}(U)$  consisting of curves that intersect  $V$ . Here  $U$  is nonannular; for an annulus  $A$ , elements of  $\mathcal{C}(A)$  cannot be projected to other subsurfaces. When the context is clear, we denote all these projection maps simply by  $\pi$ . By construction, all projection maps are quasi-Lipschitz.

3.2. Distance formula

For  $x, y \in \mathcal{X}$ , define the  $W$ -projection distance between  $x$  and  $y$  to be

$$d_W(x, y) = d_{\mathcal{E}(W)}(\pi_W(x), \pi_W(y)).$$

In fact, when  $W$  is an annulus with core curve  $\alpha$ , we sometimes denote this distance with  $d_\alpha(x, y)$ . We define the distance in  $\mathcal{X}$  using these projection distances. For a threshold  $T > 0$  large enough, define

$$d_{\mathcal{X}}(x, y) = \sum_{W \in \mathcal{W}_T(x, y)} d_W(x, y), \tag{11}$$

where  $\mathcal{W}_T(x, y)$  is the set of subsurfaces with  $d_W(x, y) \geq T$ . This is not a real metric, since the distance between different points may be zero and the triangle inequality does not hold. However, it is symmetric, and the triangle inequality holds up to a multiplicative error. That is, for  $x, y, z \in \mathcal{X}$ ,

$$d_{\mathcal{X}}(x, y) + d_{\mathcal{X}}(y, z) \stackrel{*}{\succ} d_{\mathcal{X}}(x, z).$$

Also, changing the threshold changes the metric by only uniform additive and multiplicative constants. That is, for  $T' \geq T$  we have (see [20], [25])

$$\sum_{W \in \mathcal{W}_T(x, y)} d_W(x, y) \asymp \sum_{W \in \mathcal{W}_{T'}(x, y)} d_W(x, y). \tag{12}$$

Even though this is not a metric, it makes sense to say  $\mathcal{X}$  is quasi-isometric to another metric space. In fact, in the category of metric spaces up to quasi-isometry, this notion of distance is completely adequate. We fix a threshold  $T$  once and for all so that  $d_{\mathcal{X}}(x, y)$  is a well-defined number for all  $x, y \in \mathcal{X}$ . The threshold  $T$  needs to be large enough so that the statements in the rest of this section hold.

There is a coarsely defined map

$$(\mathcal{T}(S), d_{WP}) \rightarrow \mathcal{P}(S)$$

sending a Riemann surface  $X$  to the shortest pants decomposition in  $X$ , which is, by [8], a quasi-isometry. Hence,  $\mathcal{P}(S)$  with the above metric is a combinatorial model for the Weil–Petersson metric.

A point in  $(\text{Mod}(S), d_W)$  can be coarsely represented as a marking (see [19]). That is, there is a coarsely defined map

$$(\text{Mod}(S), d_W) \rightarrow \mathcal{M}(S),$$

which can be defined by, for example, fixing a point  $x_0 \in \mathcal{M}(S)$  and sending a mapping class  $\phi \in \text{Mod}(S)$  to the marking  $\phi(x_0)$ . It is shown in [20] that this map is a quasi-isometry.



A point in  $(\mathcal{T}(S), d_{\mathcal{T}})$  can be coarsely represented as an augmented marking (see [26]). That is, there is a coarsely defined map

$$(\mathcal{T}(S), d_{\mathcal{T}}) \rightarrow \mathcal{A}\mathcal{M}(S)$$

defined as follows. A point in  $X$  in Teichmüller space is mapped to the augmented marking  $x = (P, \{\tau_{\alpha}\}, \{l_{\alpha}\})$ , where  $P$  is the shortest pants decomposition in  $X$ ,  $\alpha \in P$ ,  $\tau_{\alpha}$  is the shortest transverse curve to  $\alpha$  in  $X$ , and  $l_{\alpha}$  is the hyperbolic length of  $\alpha$  in  $X$ . It follows from [25] that this map is a quasi-isometry (again, see [11] for more details in this case).

By  $(\mathcal{X}(S), d_{\mathcal{X}})$  we denote one of the model spaces above. When the context is clear, we use  $\mathcal{X}$  instead of  $\mathcal{X}(S)$ . However, often we need to talk about  $\mathcal{X}(W)$  when  $W$  is a subsurface of  $S$ . For example, if  $\mathcal{X}(S)$  is the space of pants decompositions of  $S$ , then  $\mathcal{X}(W)$  is the space of pants decompositions of  $W$ .

### 3.3. Bounded projection, consistency, and realization

In this section, we review some properties of the projection maps. We will also derive a coarse characterization of the image of the curve complex projections of points in  $\mathcal{X}$  similar to [3]. We start with a theorem from [20].

**THEOREM 3.1** (Bounded geodesic image theorem)

*There exists a constant  $M_0$  so that the following holds. Assume that  $V \subsetneq U$  are subsurfaces of  $S$  and  $\beta_1, \dots, \beta_k$  is a geodesic in  $\mathcal{C}(U)$ . Then either there is some  $\beta_j$  that is disjoint from  $V$  or  $d_V(\beta_1, \beta_k) \leq M_0$ .*

Masur and Minsky proved this theorem, except in the case in which  $V$  is an annulus with core curve  $\alpha$  and  $\mathcal{C}(V)$  is a horoball. Then, applying Theorem 3.1, we have that either some  $\beta_j$  is disjoint from  $\alpha$  or

$$\text{twist}_{\alpha}(\beta_1) \stackrel{\pm}{\asymp} \text{twist}_{\alpha}(\beta_k).$$

Also,

$$\pi_V(\beta_i) = (\text{twist}_{\alpha}(\beta_i), 1/B).$$

Hence,

$$d_{\alpha}(\beta_1, \beta_k) \stackrel{\pm}{\asymp} \log |\text{twist}_{\alpha}(\beta_1) - \text{twist}_{\alpha}(\beta_k)| = O(1).$$

Thus, Theorem 3.1 holds in the horoball case as well.

The other important property of the projection maps is the consistency and realization result of [3]. First we recall that for subsurfaces  $U, V$  the notation  $U \pitchfork V$

means that  $U \cap V \neq \emptyset$  and neither is contained in the other. Consider the following consistency condition on a tuple:

$$z \in \prod_U \mathcal{C}(U).$$

Denote the coordinate of  $z$  in  $\mathcal{C}(U)$  with  $z_U$ . For a constant  $M$ , we say  $z$  is  $M$ -consistent if

(1) whenever  $U \pitchfork V$ ,

$$\min(d_U(z_U, \partial V), d_V(z_V, \partial U)) \leq M;$$

(2) if  $V \subsetneq U$ , then

$$\min(d_U(z_U, \partial V), d_V(z_V, z_U)) \leq M.$$

To any  $z \in \mathcal{X}$ , the tuple of projections of  $z$  is a tuple  $\mathbf{z}$  so that  $z_U = \pi_U(z)$ . The following, in the case in which  $\mathcal{X}$  is either  $\mathcal{M}(S)$  or  $\mathcal{P}(S)$ , is [3, Theorem 4.3]. However, it holds true for  $\mathcal{AM}(S)$  as well.

**THEOREM 3.2 (Consistency and realization)**

*The tuples that are consistent are essentially those that are tuples of projections. More precisely, there is a constant  $M_1$ , so that*

- (1) *for  $z \in \mathcal{X}$ , the tuple of projection  $\mathbf{z}$  of  $z$  is  $M_1$ -consistent;*
- (2) *if a tuple  $\mathbf{z}$  is  $M$ -consistent for some uniform  $M$ , then there is a realization  $z \in \mathcal{X}$  so that*

$$\forall U \quad d_U(z, z_U) = O(1).$$

*Proof*

As mentioned before, this is known for  $\mathcal{P}(S)$  and  $\mathcal{M}(S)$ . We verify the theorem in the case of  $\mathcal{AM}(S)$ .

First we check part (1). For any  $z \in \mathcal{X}$ , let  $P_z$  be the associated pants decomposition. The nonannular projections of  $z$  are the same as projections of  $P_z$ , and by the  $\mathcal{P}(S)$ -case of Theorem 3.2, these projections are consistent. Let  $A$  be an annulus, and let  $U$  be any other surface intersecting  $A$ . As in Theorem 3.1, the consistency still holds because

- the image of  $\pi_{U,A}$  is always on the boundary of the horocycle  $\mathcal{C}(A)$ ,
- the distance in  $\mathcal{C}(A)$  between two points on the boundary is the log of the twisting difference, and
- the twisting distance is bounded as a consequence of the consistency theorem for  $\mathcal{M}(S)$ .

That is, the consistency constant for  $\mathcal{AM}(S)$  is no larger than that of  $\mathcal{M}(S)$ .

To see part (2) we need to construct an augmented marking from a consistent tuple  $\mathbf{z}$ . Use the realization part of Theorem 3.2 for  $\mathcal{P}(S)$  to construct a pants decomposition  $P_0$  so that, for every subsurface  $U$  that is not an annulus,  $d_U(P_0, z_U) = O(1)$ . Still, for some curves  $\alpha$ , with corresponding annulus  $A$ , the projection of  $P$  to  $\mathcal{C}(A)$  may not be close to  $z_\alpha$ . We claim that there is a uniform bound so that for intersecting curves  $\alpha$  and  $\beta$  it cannot be the case that the distance of the projection of  $P$  to the corresponding annulus for both  $\alpha$  and  $\beta$  is more than this bound.

To prove the claim, let  $U$  be the surface they fill. Then  $U$  intersects either  $\alpha$  or  $\beta$  (say,  $\alpha$  without loss of generality). By the construction of  $P_0$ ,

$$d_U(P_0, z_U) = O(1) \xrightarrow{\pi_{U,A} \text{ is quasi-Lipschitz}} d_\alpha(P_0, z_U) = O(1),$$

and, by the second condition in the consistency of the coordinates of  $\mathbf{z}$ ,

$$d_\alpha(z_U, z_\alpha) = O(1).$$

Now, the triangle inequality implies that  $d_\alpha(P_0, z_\alpha) = O(1)$ . We have proved the claim.

We also note that, for any such curve  $\alpha$  (where  $d_\alpha(P_0, z_\alpha)$  is large) and for every subsurface  $U$  intersecting  $\alpha$ ,

$$d_U(\alpha, P_0) \stackrel{\pm}{\asymp} d_U(\alpha, z_U) = O(1).$$

This is because if  $z_U$  is far from  $\alpha$  in  $\mathcal{C}(U)$ , then the projection of  $z_U$  to  $\mathcal{C}(\alpha)$  is defined and is near  $z_\alpha$  ( $\mathbf{z}$  is constant). But, as above,  $d_U(P_0, z_U) = O(1)$ , which implies  $d_\alpha(P_0, z_U) = O(1)$ . This is a contradiction.

Let  $\alpha$  be the multicurve consisting of all the curves above. Since,  $d_U(P_0, \alpha) = O(1)$  for every nonannular subsurface  $U$ ,  $\alpha$  can be extended to a pants decomposition  $P$  with  $d_U(P_0, P) = O(1)$  for every nonannular subsurface  $U$ . That is,

$$d_V(P, z_V) = O(1), \quad \text{for every subsurface } V \text{ intersecting } P.$$

We now complete  $P$  into an augmented marking. For a curve  $\beta \in P$ , write  $z_\beta \in \mathcal{C}(\beta)$  as

$$z_\beta = (t_\beta, l_\beta),$$

where  $t_\beta$  is an integer and  $l_\beta$  is a real number less than the Bers constant. Let  $\tau_\beta$  be a curve intersecting  $\beta$  minimally that is disjoint from other curves in  $P$  with  $\text{twist}_\beta(\tau_\beta) \stackrel{\pm}{\asymp} t_\beta$ . (This can always be achieved by applying Dehn twists around  $\beta$ .) Now,

$$x = (P, \{\tau_\beta\}_{\beta \in P}, \{l_\beta\}_{\beta \in P})$$

is the desired augmented marking. This is because, for  $\beta \in P$ , the projections to  $\mathcal{C}(\beta)$  are close to  $z_\beta$  by construction and, for every other subsurface  $V$ , the projection of  $x$  to  $V$  is the same as the projection of  $P$  to  $V$ .  $\square$

The following statements will be useful later.

**COROLLARY 3.3**

*Let  $U$  and  $V$  be two subsurfaces where  $\partial V$  intersects  $U$ . For any  $x \in \mathcal{X}$ , if*

$$d_U(x_U, \partial V) > M_1, \quad \text{then} \quad d_V(x_V, x_U) = O(1).$$

*Proof*

There are two cases. If  $V \subsetneq U$ , then this is immediate from part (2) of the consistency condition.

Otherwise,  $\partial U$  intersects  $V$  and has a defined projection to  $\mathcal{C}(V)$ . Since  $x_U$  is disjoint from  $\partial U$ ,  $d_V(x_U, \partial U) = O(1)$ . On the other hand the assumption of the corollary and part (1) of the consistency condition say that  $d_V(x_V, \partial U) = O(1)$ . The corollary follows from the triangle inequality.  $\square$

Given  $x, y, z \in \mathcal{X}$  and a subsurface  $W$ , let  $\eta_W$  be the center of the triangle  $(x_W, y_W, z_W)$  guaranteed by the hyperbolicity of  $\mathcal{C}(W)$ . That is,  $\eta_W$  is  $\delta_W$ -close to all three geodesics  $[x, y]_W$ ,  $[y, z]_W$ , and  $[x, z]_W$ , where  $\delta_W$  is the hyperbolicity constant of  $\mathcal{C}(W)$ .

**LEMMA 3.4**

*The set  $\{\eta_W\}$  is  $O(1)$ -consistent.*

*Proof*

Let  $U, V$  be arbitrary domains which are not disjoint. We can assume that  $U \not\subset V$ , and hence  $\pi_U(\partial V)$  is defined. We know that the projections of  $x, y, z$  to  $U$  and  $V$  are themselves  $M_1$ -consistent.

Consider the triangle  $\Delta_U$  with vertices  $x_U, y_U, z_U$  in  $\mathcal{C}(U)$ . Note that if  $\partial V$  is uniformly close to all three edges of  $\Delta_U$ , then  $\partial V$  is uniformly close to  $\eta_U$  and we are done. Hence, we can, without loss of generality, assume that no point in  $[x, y]_U$  is near  $\partial V$  in  $\mathcal{C}(U)$ . In fact, since  $\eta_U$  is  $\delta$ -close to  $[x, y]_U$ , we can assume every point in the convex hull of  $x_U, y_U$ , and  $\eta_U$  is more than  $M_1$  away from  $\partial V$ .

This implies that  $x_U, y_U$ , and  $\eta_U$  have defined projections to  $V$ , and Theorem 3.1 implies that their projections are a bounded distance from one another. That is,

$$d_V(\eta_U, x_U) = O(1), \quad d_V(\eta_U, y_U) = O(1), \quad \text{and} \quad d_V(x_U, y_U) = O(1).$$

On the other hand, because  $d_U(x_U, \partial V)$  and  $d_U(y_U, \partial V)$  are both larger than  $M_1$ , by Corollary 3.3

$$d_V(x_V, x_U) \leq M_1 \quad \text{and} \quad d_V(y_V, y_U) \leq M_1.$$

By the triangle inequality,

$$d_V(x_V, y_V) = O(1).$$

From the definition of  $\eta_V$ , we have

$$d_V(\eta_V, [x, y]_V) \leq \delta_V.$$

Again, using the triangle inequality, we get

$$d_V(\eta_U, \eta_V) = O(1). \tag{13}$$

This is the consistency condition when  $V \subset U$ .

Thus, assume that  $U \pitchfork V$ . Since  $\eta_U$  and  $\partial U$  are disjoint,  $d_V(\eta_U, \partial U) = O(1)$ . This and (13) imply that

$$d_V(\partial U, \eta_V) = O(1),$$

which is the required consistency condition in this case. □

*Definition 3.5*

Since the tuple of centers is consistent, it has a realization  $\eta$ . We call  $\eta$  the *center* of the triangle with vertices  $x$ ,  $y$ , and  $z$ .

*3.4. Product regions*

For every subsurface  $W$  of  $S$ , we have a projection map

$$\phi_W: \mathcal{X} \rightarrow \mathcal{X}(W)$$

defined by Theorem 3.2. Namely, since the projections of a point  $x \in \mathcal{X}$  to subsurfaces of  $S$  are consistent, the projections to subsurfaces of  $W$  are also consistent and hence can be realized by a point in  $\mathcal{X}(W)$ . For points  $x, y \in \mathcal{X}$ , we define

$$d_{\mathcal{X}(W)}(x, y) = d_{\mathcal{X}(W)}(\phi_W(x), \phi_W(y)).$$

The subsurface  $W$  is allowed to be an annulus, in which case  $\mathcal{X}(W) = \mathcal{C}(W)$ .

For a curve system  $\alpha$ , let  $\mathcal{X}_\alpha$  be the set of points in  $x \in \mathcal{X}$  where  $\alpha$  is a subset of the pants decomposition  $P_x$  associated to  $x$ . Consider a point  $x \in \mathcal{X}_\alpha$ . Since every pants curve in  $x$  is disjoint from  $\alpha$ , the projection of  $x$  to any subsurface intersecting  $\alpha$  is a distance at most 2 from the projection of  $\alpha$ . Therefore, for a sufficiently large

threshold  $T$  and for any  $x, y \in \mathcal{X}_\alpha$ , the set  $\mathcal{W}_T(x, y)$  consists only of subsurfaces disjoint from  $\alpha$ , each of which is contained in some component  $W$  of  $S \setminus \alpha$ . Therefore, the map

$$\Phi: \mathcal{X}_\alpha \rightarrow \prod_W \mathcal{X}(W), \quad \text{where } \Phi = \prod_W \phi_W,$$

is a quasi-isometry. (The image is coarsely surjective) Here the product space is equipped with the  $L^1$ -metric. A version of this theorem for Teichmüller space was first proved by Minsky [22] and is known as the *product regions theorem*. We see that the fact that  $\Phi$  is a quasi-isometry is essentially immediate from the distance formula. However, the proof of the distance formula in [25] used Minsky’s product regions theorem.

There is also a projection map  $\phi_{\mathcal{X}_\alpha}: \mathcal{X} \rightarrow \mathcal{X}_\alpha$ . For  $x \in \mathcal{X}$  and each component  $W$  of  $S \setminus \alpha$ , take  $\phi_W(x)$  and then take the union over all  $W$ ’s. We then add the curves  $\alpha$  to find a point in  $\mathcal{X}_\alpha$ . We define  $d_{\mathcal{X}_\alpha}$  to mean the distance between projections to  $\mathcal{X}_\alpha$ . That is, for  $x, y \in \mathcal{X}$ , we define

$$d_{\mathcal{X}_\alpha}(x, y) := d_{\mathcal{X}}(\phi_{\mathcal{X}_\alpha}(x), \phi_{\mathcal{X}_\alpha}(y)).$$

Note that the projection of  $\phi_{\mathcal{X}_\alpha}(x)$  to  $\mathcal{X}(W)$  is close to  $\phi_W(x)$ , because  $\phi_W$  was defined using the consistency result. Therefore, we have

$$d_{\mathcal{X}_\alpha}(x, y) \asymp \sum_W d_{\mathcal{X}(W)}(x, y), \tag{14}$$

where the sum is over components of  $S \setminus \alpha$ .

We finish with an estimate of  $d_{\mathcal{X}}$  using the projection distances  $d_{\mathcal{X}_\alpha}$ .

LEMMA 3.6

Suppose  $x, y \in \mathcal{X}$  and  $\alpha_1, \dots, \alpha_k$  is a geodesic in  $\mathcal{C}(S)$  joining  $\alpha_1 \in P_x$  to  $\alpha_k \in P_y$ . Then

$$d_{\mathcal{X}}(x, y) < \sum_{i=1}^k d_{\mathcal{X}_{\alpha_i}}(x, y).$$

*Proof*

Let  $M_0$  be the constant in Theorem 3.1. Then any subsurface  $U$  with  $d_U(x, y) \geq M_0$  is disjoint from some  $\alpha_j$  and so is a subset of  $S \setminus \alpha_j$ . Thus, it appears as a term in some  $d_{\mathcal{X}(W)}(x, y)$ , where  $W$  is a component of  $S \setminus \alpha_j$ . We are done by (14).  $\square$

**4. Efficient paths are nearly geodesics**

In this section, we show that efficient paths are nearly geodesics in the space  $\mathcal{X}$ . To do this, we use the construction of Bestvina, Bromberg, and Fujiwara [5], which gives a

quasi-isometric embedding of the mapping class group to a finite product of Gromov hyperbolic spaces. Their construction is completely axiomatic and works, essentially without modifications, for any of our spaces  $\mathcal{X}$ . We first review their construction. We remark first that their notation differs from ours. We will explain the differences.

4.1. *A quasitree of curve complexes*

We summarize some statements in [5]. Fix a threshold  $K$ . Let  $\mathbf{Y}$  be a collection of subsurfaces of  $S$  with the property that if  $V, W \in \mathbf{Y}$ , then

- $V \pitchfork W$ , and
  - every curve in  $W$  intersects  $V$ . Hence, the domain of  $\pi_{W,V}$  is  $\mathcal{C}(W)$ .
- Starting from the function, for each  $Y \in \mathbf{Y}$ ,  $d_Y : \mathbf{Y} \setminus \{Y\} \times \mathbf{Y} \setminus \{Y\} \times \mathbf{Y} \setminus \{Y\} \rightarrow [0, \infty]$  (in their notation the function is  $d_Y^{\mathcal{C}}$ ), one can define a new function  $D_Y : \mathbf{Y} \setminus \{Y\} \times \mathbf{Y} \setminus \{Y\} \times \mathbf{Y} \setminus \{Y\} \rightarrow [0, \infty]$  (in their notation it is  $d_Y$ ), which satisfies a list of conditions, the chief of which from our point of view is

$$D_U(Z, W) \leq d_U(Z, W) \stackrel{\pm}{\leq} D_U(Z, W).$$

Define (see [5, Definition 3.6])  $\mathbf{P}_K(\mathbf{Y})$  to be a graph whose vertices are elements of  $\mathbf{Y}$ , and two vertices  $V, W$  are connected with an edge if for all  $U \in \mathbf{Y}$

$$D_U(\partial V, \partial W) \leq K.$$

It is shown in [5] for  $K$  sufficiently large that  $\mathbf{P}_K(\mathbf{Y})$  is quasi-isometric to a tree.

Define  $\mathcal{C}(\mathbf{Y})$  to be the space obtained from  $\mathbf{P}_K(\mathbf{Y})$  by attaching a copy of  $\mathcal{C}(W)$  for every vertex  $W \in \mathbf{Y}$  as follows. The vertex set of  $\mathcal{C}(\mathbf{Y})$  is the union of vertex sets of  $\mathcal{C}(W)$ ,  $W \in \mathbf{Y}$ . If  $V$  and  $W$  are joined by an edge in  $\mathbf{P}_K(\mathbf{Y})$ , then we join the vertex  $\pi_V(\partial W)$  in  $\mathcal{C}(V)$  to  $\pi_W(\partial V)$  in  $\mathcal{C}(W)$ . We also include the edges of  $\mathcal{C}(W)$  for each  $W$ . It follows from [5, Theorem E], using the fact that each space  $\mathcal{C}(W)$  is hyperbolic, that the resulting space  $\mathcal{C}(\mathbf{Y})$  is also hyperbolic.

Furthermore, Bestvina, Bromberg, and Fujiwara [5] show the following when  $S$  is connected:

- the subsurfaces of  $S$  can be decomposed into finitely many disjoint subsets  $\mathbf{Y}^1, \dots, \mathbf{Y}^k$ , each having the transversality property mentioned above; and
- there is a finite-index subgroup  $\Gamma$  in  $\text{Mod}(S)$  which fixes each  $\mathbf{Y}^i$ .

When  $S$  is disconnected, we decompose the subsurfaces of each component as above and let  $\mathbf{Y}^1, \dots, \mathbf{Y}^k$  be the list of all such collections.

We can assume that  $\mathbf{Y}^j$  contains only essential subsurfaces; that is, the thrice-punctured spheres are always excluded, and in the case in which  $\mathcal{X}$  is the pants graph, annuli are also excluded. Let

$$\mathbf{C} = \prod_{j=1}^k \mathcal{C}(\mathbf{Y}^j)$$

be equipped with the  $L^1$ -metric. Thus,  $\mathbf{C}$  is a product of finitely many hyperbolic spaces.

We define a projection map  $\Psi^j : \mathcal{X} \rightarrow \mathcal{C}(\mathbf{Y}^j)$  as follows. For  $x \in \mathcal{X}$ , choose a subsurface  $W_j \in \mathbf{Y}^j$  that minimizes

$$\max_{\alpha \in P_x} i(\alpha, \partial W_j),$$

where the minimum is over  $W_j \in \mathbf{Y}^j$ . Define

$$\Psi^j(x) = x_{W_j}.$$

Recall that  $x_{W_j}$  is the projection of  $x$  to the curve complex  $\mathcal{C}(W_j)$ . That is,  $x_{W_j}$  is a point in  $\mathcal{C}(W_j)$  and hence is a point in  $\mathcal{C}(\mathbf{Y}^j)$ .

*Remark 4.1*

Our definition of the projection is slightly different from that in [5]. There, the action of the mapping class group is used to define the projection. However, in the case in which  $\mathbf{Y}^j$  consists of annuli and  $\mathcal{X}$  is the augmented marking space,  $x_{W_j}$  is a point in the horoball  $H \subset \mathbb{H}^2$  and not a curve. In particular, the action of the mapping class group is not coarsely transitive. But in the other two cases, the two definitions match.

We claim that, for large  $K$ , the consistency condition (Section 3.3) shows that this map is coarsely well defined. That is, up to a bounded distance in  $\mathbf{C}$ , the image is independent of the choice of  $W_j$ .

First we note that, since the finite-index subgroup  $\Gamma$  acts preserving each  $\mathcal{C}(\mathbf{Y}^j)$ , the above minimum is uniformly bounded by a constant independent of  $x$ . We need to check that distinct choices of  $W_j$  and  $V_j$  give points at a bounded distance in  $\mathcal{C}(\mathbf{Y}^j)$ . First we check that we can choose  $K$  so that for all  $U \in \mathbf{Y}^j$

$$D_U(\partial W_j, \partial V_j) \leq d_U(\partial W_j, \partial V_j) \leq K. \tag{15}$$

Choose an  $\alpha \in x$  that intersects  $U$ . We have  $i(\alpha, \partial W_j) = O(1)$ , where the implied bound in  $O(1)$  is universal and does not depend on  $K$ . Therefore,

$$d_U(\alpha, \partial W_j) = O(1),$$

and similarly for  $\alpha$  and  $\partial V_j$ . The triangle inequality says that  $W_j$  and  $V_j$  are connected by an edge in  $\mathbf{P}_K(\mathbf{Y}^j)$  for  $K$  sufficiently large. To show that  $x_{W_j}$  and  $x_{V_j}$  are close in  $\mathcal{C}(\mathbf{Y}^j)$ , we need that

$$d_{V_j}(x_{V_j}, \partial W_j) = O(1) \quad \text{and} \quad d_{W_j}(x_{W_j}, \partial V_j) = O(1).$$



This holds since, for every curve  $\alpha \in \mathcal{X}$ ,

$$i(\alpha, \partial W_j) = O(1) \quad \text{and} \quad i(\alpha, \partial V_j) = O(1).$$

We often denote  $\Psi^j(x)$  by  $x_j$ . Now define a map

$$\Psi: \mathcal{X} \rightarrow \mathbf{C} \quad \text{with} \quad \Psi = \prod_j \Psi^j.$$

The following result is proved in [5] for the mapping class group.

LEMMA 4.2

There is  $K' > K$  so that, for every  $x, y \in \mathcal{X}$ ,

$$d_{\mathbf{C}}(\Psi(x), \Psi(y)) \stackrel{*}{\asymp} \sum_{W \in \mathcal{W}_{K'}(x,y)} d_W(x, y).$$

This follows from the lower bound of Bestvina, Bromberg, and Fujiwara [5, Theorem 4.13]; we have replaced  $D_W$  with  $d_W$ , and in their theorem, the multiplicative error is  $\frac{1}{2}$ . In view of (15), the terms on the right-hand side are larger only additively compared to the terms in [5, Theorem 4.13]. Hence, choosing a larger threshold ensures that the sum above does not have any extra terms. Also, increasing every term in the above sum by a uniform additive amount increases the sum by at most a uniform multiplicative amount. The proof in [5] uses only the hyperbolicity of each curve complex and the consistency condition detailed in Section 3.3 and works verbatim in our case. Hence, we omit the proof. As a consequence we have the following theorem, which is also proved in [5] for the mapping class group. We give a proof here because our projection maps  $\psi^j$  are defined differently from those in [5].

THEOREM 4.3

For  $K$  sufficiently large, the map  $\Psi$  is a quasi-isometric embedding from  $\mathcal{X}$  into  $\mathbf{C}$ .

*Proof*

For  $x \in \mathcal{X}$ , the map  $\Psi$  is defined by

$$\Psi(x) = (x_{W_1}, \dots, x_{W_k}),$$

for some subsurface  $W_j \in \mathbf{Y}^j$ ,  $1 \leq j \leq k$ . We show that  $\Psi^j$  is quasi-Lipschitz, which is essentially the same as the proof that  $\Psi^j$  is coarsely well defined. If  $d_{\mathcal{X}}(x, x') = O(1)$ , then

$$d_U(x, x') = O(1) \quad \forall U.$$

We have  $i(x, \partial W_j) = O(1)$ , and so

$$d_U(x, \partial W_j) = O(1) \quad \text{and similarly} \quad d_U(x', \partial W'_j) = O(1).$$

Together this gives

$$d_U(\partial W_j, \partial W'_j) = O(1),$$

which implies for large  $K$  that  $W_j$  and  $W'_j$  are connected by an edge in  $\mathbf{P}_K(\mathbf{Y}^j)$ . We also know that  $x$  is close to  $x'$ , which has a bounded intersection with  $\partial W'_j$ . Hence,

$$d_{W_j}(x_{W_j}, \partial W'_j) = O(1), \quad \text{and similarly} \quad d_{W'_j}(x_{W'_j}, \partial W_j) = O(1).$$

Therefore,

$$d_{\mathcal{E}(\mathbf{Y}^j)}(\Psi^j(x), \Psi^j(x')) = O(1).$$

This means that the maps  $\Psi^j$  are quasi-Lipschitz and so is  $\Psi$ .

We need to find a lower bound for the distance between  $\Psi(x)$  and  $\Psi(y)$ . By Lemma 4.2, there is  $K' > K$  so that

$$d_C(\Psi(x), \Psi(y)) \stackrel{*}{\asymp} \sum_{W \in \mathcal{W}_{K'}(x,y)} d_W(x, y),$$

and since the distance formula works for any threshold,

$$d_{\mathcal{X}}(x, y) \stackrel{*}{\asymp} \sum_{W \in \mathcal{W}_{K'}(x,y)} d_W(x, y).$$

Hence,

$$d_C(\Psi(x), \Psi(y)) \stackrel{*}{\asymp} d_{\mathcal{X}}(x, y).$$

This finishes the proof. □

#### 4.2. Preferred paths and efficient paths

Since the space  $\mathcal{X}$  is not hyperbolic, a quasi-geodesic connecting two points in  $\mathcal{X}$  may not be well behaved. Instead, we define a notion of preferred path connecting two points in  $\mathcal{X}$ .

##### Definition 4.4

Given  $x, y \in \mathcal{X}$ , we say a quasi-geodesic  $\omega: [a, b] \rightarrow \mathcal{X}$  is a *preferred path* connecting  $x$  to  $y$  if

- $\omega(a) = x, \omega(b) = y$ ; and

- for every subsurface  $U$ , the map

$$\omega_U = \pi_U \circ \omega : [a, b] \rightarrow \mathcal{C}(U)$$

is an unparameterized quasi-geodesic.

LEMMA 4.5

For any  $x, y \in \mathcal{X}$  there is a preferred path connecting  $x$  to  $y$ .

*Proof*

In the case of the mapping class group and the pants complex, a resolution of a hierarchy is a preferred path (see [20]). In the case in which  $\mathcal{X}$  is Teichmüller space with the Teichmüller metric, such a path is constructed in [25, Theorem 5.7].  $\square$

*Remark 4.6*

It is known that a Teichmüller geodesic is not always a preferred path. (There may be backtracking in annuli.) It is unknown if a Weil–Petersson geodesic or a geodesic in the mapping class group is a preferred path.

Now let  $U$  be a subsurface of  $S$ , and let  $x_U$  and  $y_U$  be the projections of  $x$  and  $y$  to  $\mathcal{C}(U)$ . Denote a geodesic segment in  $\mathcal{C}(U)$  connecting  $x_U$  to  $y_U$  by  $[x, y]_U$ . Given  $\kappa > 0$ , let

$$\mathcal{G}(x, y, \kappa) = \{z \in \mathcal{X} \mid \forall U, d_U(z, [x, y]_U) \leq \kappa\}.$$

This notion was introduced in [3], where they call it the hull. In a sense, this set is the union of all points in all preferred paths.

LEMMA 4.7

There is a constant  $\kappa_0$  depending only on the topology of  $S$  and the constant involved in the definition of a preferred path so that, for any preferred path  $\omega : [a, b] \rightarrow \mathcal{X}$  and any  $a \leq t \leq b$ ,

$$\omega(t) \in \mathcal{G}(x, y, \kappa_0).$$

*Proof*

Since  $\mathcal{C}(U)$  is Gromov-hyperbolic and the projection of  $\omega$  is an unparameterized quasi-geodesic, it stays in a uniform neighborhood of the geodesic connecting its endpoints.  $\square$

LEMMA 4.8

There exists  $\kappa_0$  depending only on topology and the constant  $M_0$  in Theorem 3.1 such that if  $x, y \in \mathcal{X}$  and  $w, z \in \mathcal{G}(x, y, \kappa_0)$ , then the following statements hold.

- For any subsurface  $U$ ,

$$d_U(w, z) \stackrel{+}{\prec} d_U(x, y).$$

In fact,  $[w, z]_U$  is contained in a uniform neighborhood of  $[x, y]_U$ .

- (Convexity) If  $w, z \in \mathcal{G}(x, y, \kappa_0)$ , then  $\mathcal{G}(w, z, \kappa_0)$  is contained in a uniform neighborhood of  $\mathcal{G}(x, y, \kappa_0)$ .

*Proof*

Since  $\mathcal{C}(U)$  is Gromov-hyperbolic, if both  $z_U$  and  $w_U$  are close to  $[x, y]_U$ , then so is  $[z, w]_U$ . Hence the length of  $[z, w]_U$  is less than that of  $[x, y]_U$ . This proves the first statement.

We prove the second statement. Let  $\kappa_0$  be the maximum hyperbolicity constant for any  $\mathcal{C}(U)$ . (In fact, all curve complexes are known to be uniformly hyperbolic (see [1], [6], [10], [17], [23]).) Given any  $p \in \mathcal{G}(z, w, \kappa_0)$  let  $\eta$  be the center (see Definition 3.5) of points  $x, y, p$  given by Lemma 3.4. For any subsurface  $U$ , the distance between  $\eta_U$  and  $[x, y]_U$  is bounded by the hyperbolicity constant of  $\mathcal{C}(U)$ . Hence, for  $\kappa_0$  large enough,  $\eta \in \mathcal{G}(x, y, \kappa_0)$ .

We need to show that  $d_{\mathcal{X}}(p, \eta)$  is uniformly bounded. For every subsurface  $U$ ,  $p_U$  is within  $\kappa_0$  of  $[z, w]_U$ , and by the first part,  $[z, w]_U$  is contained in a uniform neighborhood of  $[x, y]_U$ . That is, the distance between  $p_U$  and  $[x, y]_U$  is uniformly bounded, and hence the distance between  $p_U$  and  $\eta_U$  is also uniformly bounded. This gives a bound on  $d_{\mathcal{X}}(p, \eta)$  by the distance formula. □

We now fix  $\kappa_0$  so that the above two lemmas hold and simply write  $\mathcal{G}(x, y)$ . The following is the main theorem of this section and states that efficient paths follow preferred paths.

THEOREM 4.9

Let  $\gamma: [0, R] \rightarrow \mathcal{X}$  be an  $\epsilon$ -efficient path at scale  $R$  connecting  $x = \gamma(0)$  to  $y = \gamma(R)$ . Then, the image of  $\gamma$  stays in an  $O(\epsilon R)$ -neighborhood of  $\mathcal{G}(x, y)$ . Moreover, it stays in an  $O(\epsilon R)$ -neighborhood of a preferred path connecting  $x$  to  $y$ .

We now show that taking the center of three points (see Definition 3.5) and projecting to  $\mathcal{C}(Y^j)$  commute.

LEMMA 4.10

For any  $x, y, z \in \mathcal{X}$ , let  $\eta$  be the center of the triangle  $\Delta$  with vertices  $x, y, z$ . Let  $x_j, y_j, z_j$ , and  $\eta_j$ , respectively, denote  $\Psi^j(x), \Psi^j(y), \Psi^j(z)$ , and  $\Psi^j(\eta)$ , the projections of these points to  $\mathcal{C}(\mathbf{Y}^j)$ . Then  $\eta_j$  is near the center of the triangle  $(x_j, y_j, z_j)$  in  $\mathcal{C}(\mathbf{Y}^j)$ .

*Proof*

First we claim that, for every  $W \in \mathbf{Y}^j$ ,  $d_W(x, x_j) = O(1)$ . To prove the claim let  $x_j$  be a curve  $x_V$  in a surface  $V \in \mathbf{Y}^j$ . We have that

$$x_V \text{ is disjoint from } \partial V \implies d_W(\partial V, x_V) = O(1)$$

and

$$i(x, \partial V) = O(1) \implies d_W(\partial V, x) = O(1).$$

The claim follows from the triangle inequality.

In [5, Lemma 4.15], it is shown that a geodesic in  $\mathcal{C}(\mathbf{Y}^j)$  connecting  $x_j$  to  $y_j$  is a bounded Hausdorff distance from a union of geodesics  $[x_j, y_j]_W$ , where the union is over the subsurfaces for which  $d_W(x_j, y_j)$  is large. The same holds for the geodesic connecting  $x_j$  to  $\eta_j$ .

But, as a consequence of the above claim, the geodesic  $[x_j, y_j]_W$  is a bounded Hausdorff distance from  $[x, y]_W$ , and  $[x_j, \eta_j]_W$  is a bounded Hausdorff distance from  $[x, \eta]_W$ . Also, by assumption, we know that  $[x, \eta]_W$  is contained in a bounded neighborhood of  $[x, y]_W$ . Therefore,  $[x_j, \eta_j]_W$  is contained in a bounded neighborhood of  $[x_j, y_j]_W$ . In particular, if  $d_W(x_j, \eta_j)$  is large, then so is  $d_W(x_j, y_j)$ .

That is, every subsurface that appears in the geodesic connecting  $x_j$  to  $\eta_j$  also appears in the geodesic connecting  $x_j$  to  $y_j$ , and the portion of the geodesic  $[x_j, \eta_j]$  that is in  $W$  stays near the geodesic  $[x_j, y_j]$ . Thus,  $\eta_j$  is itself close to  $[x_j, y_j]$ .

The same holds for  $x_j, z_j$  and  $y_j, z_j$ . Since  $\mathcal{C}(\mathbf{Y}^j)$  is Gromov-hyperbolic and  $\eta_j$  is close to all three geodesics, it is near the center of the triangle. □

We now prove the theorem.

*Proof of Theorem 4.9*

Let  $\gamma_j = \Psi^j \circ \gamma$  be the projection of the path  $\gamma$  to  $\mathcal{C}(\mathbf{Y}^j)$ . By Lemma 2.3, each  $\gamma_j$  is still  $\epsilon$ -efficient. Since  $\mathcal{C}(\mathbf{Y}^j)$  is hyperbolic, by Lemma 2.12,  $\gamma_j(t)$  is within an  $O(\epsilon R)$ -distance of a geodesic  $[x_j, y_j]$  connecting  $x_j = \Psi_j(x)$  to  $y_j = \Psi_j(y)$ . Let  $z = \gamma(t)$ , and let  $\eta$  be the center of  $x, y, z$ . From the construction, we have  $\eta \in \mathcal{G}(x, y)$ . We estimate the distance between  $\eta$  and  $z$ .

By Lemma 4.10,  $\eta_j$  is the center of triangle  $(x_j, y_j, z_j)$ . The distance from  $z_j$  to  $[x_j, y_j]$  is, up to an additive error, the distance from  $z_j$  to the center  $\eta_j$ . Therefore,

$$d_{\mathcal{C}(Y^j)}(z_j, \eta_j) \stackrel{\pm}{\asymp} d_{\mathcal{C}(Y^j)}(z_j, [x_j, y_j]) = O(\epsilon R).$$

It follows, since  $\Psi$  is coarsely Lipschitz and the metric in  $\mathbf{C}$  is the  $L^1$ -metric, that

$$d_{\mathcal{X}}(z, \eta) \stackrel{*}{\asymp} \sum_j d_{\mathcal{C}(Y^j)}(z_j, \eta_j) = O(\epsilon R).$$

This finishes the proof of the first statement of the theorem.

We prove the second statement, namely, that  $\gamma$  stays in an  $O(\epsilon R)$ -neighborhood of a preferred path connecting  $x$  to  $y$ . Let  $\eta(t)$  be the center of the triangle with vertices  $x$ ,  $y$ , and  $\gamma(t)$ . The issue is that  $\eta(t)$  may not trace a preferred path, since the  $\epsilon$ -efficient path  $\gamma$  is allowed to backtrack up to  $O(\epsilon R)$ . We proceed therefore as follows.

For a time  $t$  and a subsurface  $W$ , let  $\eta_W(t)$  be the projection of  $\eta(t)$  to  $W$ . Consider the geodesic  $[x, y]_W$  in  $\mathcal{C}(W)$ . Let  $s_W \in [0, t]$  be a time where  $d_W(x_W, \eta_W(s_W))$  is maximized (see Figure 4), and define

$$\omega_W(t) = \eta_W(s_W).$$

Note that  $\omega_W(t)$  is an unparameterized quasi-geodesic, since it stays close to the geodesic  $[x, y]_W$  and does not backtrack.

We prove, for every  $t$ , that the set  $\{\eta_W(s_W)\}$  is  $M_1$ -consistent, where  $M_1$  is the constant from Theorem 3.2. Pick two intersecting surfaces  $U$  and  $V$ . Suppose first that  $U \pitchfork V$ . If  $d_U(x, y) = O(1)$ , then  $d_U(\eta(s_U), \eta(s_V)) = O(1)$ . Now since  $\eta_U(s_V)$  and  $\eta_V(s_V)$  are consistent, so are  $\eta_U(s_U)$  and  $\eta_V(s_V)$ . Similarly, if  $d_V(x, y) = O(1)$ , then we are done.

Otherwise, as we have seen before, we can assume without loss of generality that  $d_V(\partial U, y) = O(1)$  and  $d_U(x, \partial V) = O(1)$ . Now, if  $d_U(\eta(s_U), \partial V)$  is large enough,

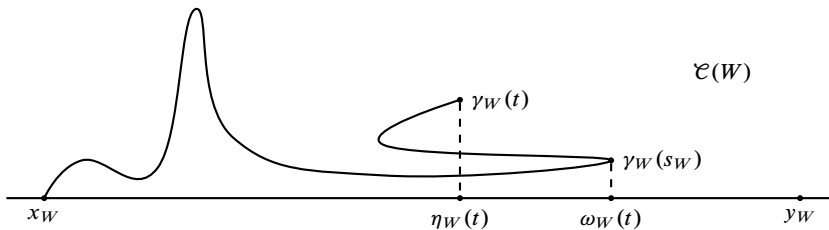


Figure 4. The point  $\omega_W(t)$  is defined to be the point  $\eta_W(s)$  that is farthest along in  $[x, y]_W$  for  $s \in [0, t]$ .

then the segment  $[\eta_U(s_U), y]_U$  has a distance more than  $M_1$  from  $\partial V$  and by the bounded geodesic image theorem (see Theorem 3.1) we have  $d_V(\eta(s_U), y) \leq M_0$ . But  $\eta_V(s_V)$  is between  $\eta_V(s_U)$  and  $y$  and hence is close to both. By the triangle inequality, we have

$$d_V(\partial U, \eta_V(s_V)) = O(1).$$

Next assume that  $V \subsetneq U$  and that  $d_U(\eta(s_U), \partial V)$  is large. In  $\mathcal{C}(U)$ , one of the segments  $[x, \eta(s_U)]_U$  or  $[\eta(s_U), y]_U$  has a distance at least  $M_1$  from  $\pi_U(\partial V)$ . Assume first that it is  $[x_U, \eta_U(s_U)]_U$ . But  $\eta_U(s_V)$  is contained in this segment and is also far from  $\pi_U(\partial V)$ . Then Theorem 3.1 implies that

$$d_V(\eta(s_U), \eta(s_V)) = O(1).$$

The remaining case is when  $[\eta(s_U), y]_U$  has a distance of at least  $M_1$  from  $\partial V$ . Then

$$d_V(\eta(s_U), y) = O(1).$$

But  $\eta_V(s_V)$  is contained in the interval  $[\eta_V(s_U), y]$ . Hence, it is close to both, and in particular,

$$d_V(\eta(s_U), \eta(s_V)) = O(1).$$

Now, using Theorem 3.2(2), there is a point  $\omega(t) \in \mathcal{X}$  where the projection of  $\omega(t)$  to  $W$  is close to  $\omega_W(t)$  (hence justifying the notation).

Finally, we need to prove

$$d_{\mathcal{X}}(\gamma(t), \omega(t)) = O(\epsilon R).$$

Let  $\omega_j(t)$  be the projection of  $\omega(t)$  to  $\mathcal{C}(\mathbf{Y}^j)$ . We observe that  $\omega_j(t)$  is near the point  $\eta_j(s)$ ,  $s \in [0, t]$ , that is farthest along in  $[x_j, y_j]$ . This is because a geodesic in  $\mathcal{C}(\mathbf{Y}^j)$  is a union of geodesics in subsurfaces  $W_1, \dots, W_k$  appearing in natural order, and since all these subsurfaces intersect, if a point  $z_j$  is ahead of  $z'_j$  along  $[x_j, y_j]$ , then the projection of  $z_j$  is ahead of the projection of  $z'_j$  in every subsurface. Therefore, there is  $s \in [0, t]$  so that

$$d_{\mathcal{C}(\mathbf{Y}^j)}(\omega_j(t), \eta_j(s)) = O(1).$$

From Lemma 2.3(1) we have

$$\begin{aligned} & d_{\mathcal{C}(\mathbf{Y}^j)}(x_j, \gamma_j(s)) + d_{\mathcal{C}(\mathbf{Y}^j)}(\gamma_j(s), \gamma_j(t)) + d_{\mathcal{C}(\mathbf{Y}^j)}(\gamma_j(t), y_j) \\ & \leq d_{\mathcal{C}(\mathbf{Y}^j)}(x_j, y_j) + O(\epsilon R). \end{aligned}$$

But  $\mathcal{C}(\mathbf{Y}^j)$  is Gromov-hyperbolic, so the projection to  $[x_j, y_j]$  is distance-decreasing. Furthermore,  $d_{\mathcal{C}(\mathbf{Y}^j)}(\eta_j, [x_j, y_j]) = O(1)$ . Hence,

$$d_{\mathcal{C}(\mathbf{Y}^j)}(x_j, \eta_j(s)) + d_{\mathcal{C}(\mathbf{Y}^j)}(\eta_j(s), \eta_j(t)) + d_{\mathcal{C}(\mathbf{Y}^j)}(\eta_j(t), y_j) \leq d_{\mathcal{C}(\mathbf{Y}^j)}(x_j, y_j) + O(\epsilon R).$$

But we know  $\eta_j(t)$  comes before  $\eta_j(s)$ . Therefore,

$$d_{\mathcal{C}(\mathbf{Y}^j)}(\eta_j(s), \eta_j(t)) = O(\epsilon R),$$

and hence,

$$d_{\mathcal{C}(\mathbf{Y}^j)}(\omega_j(t), \eta_j(t)) = O(\epsilon R).$$

Since this is true for every  $j$ , we also have

$$d_{\mathcal{X}}(\omega(t), \eta(t)) = O(\epsilon R).$$

This finishes the proof. □

### 5. Behavior of preferred paths

In this section, we analyze preferred paths more carefully to obtain more control over their behavior. In Proposition 5.5 we show (up to taking a subsurface) that if a preferred path is making progress in a subsurface, then it has to stay close to the set of points in  $\mathcal{X}$  that contain the boundary of that subsurface in their pants decomposition. This is analogous to the main results in [24] for Teichmüller geodesics. At the end of the section we prove two fellow-traveling results (Propositions 5.7 and 5.10) for preferred paths. We start by proving a few lemmas.

The following lemma gives a bound on the thickness of  $\mathcal{G}(x, y)$  in terms of projection distances  $d_{\mathcal{X}(W)}(x, y)$ .

LEMMA 5.1

*For any  $D > 0$ , if  $d_{\mathcal{X}(W)}(x, y) \leq D$  for every proper subsurface  $W$ , then any point  $z \in \mathcal{G}(x, y)$  is within a distance  $O(D)$  of any preferred path  $\gamma$  joining  $x, y$ .*

*Proof*

Let  $\gamma$  be a preferred path connecting  $x$  to  $y$ , and let  $z \in \mathcal{G}(x, y)$ . Assume first that  $S$  is connected, and let  $\gamma(t)$  be a point with

$$d_{\mathcal{C}(S)}(z, \gamma(t)) = O(1).$$

Such a point exists because  $d_S(z, [x, y]_S) = O(1)$  by the definition of  $\mathcal{G}(x, y)$  and  $\gamma$  connects  $x$  to  $y$ .



Let  $\alpha_1, \dots, \alpha_k$  be the geodesic in  $\mathcal{C}(S)$  connecting a pants curve in  $z$  to a pants curve in  $\gamma(t)$  with  $k = O(1)$ . For every  $\alpha_i$  and every subsurface  $U$  of  $S - \alpha_i$ , we have (Lemma 4.8)

$$d_U(z, \gamma(t)) \stackrel{+}{\lesssim} d_U(x, y).$$

Hence, using the distance formula, we get

$$d_{\mathcal{X}_{\alpha_i}}(z, \gamma(t)) \stackrel{*}{\lesssim} d_{\mathcal{X}_{\alpha_i}}(x, y) \leq D.$$

Now, by Lemma 3.6

$$d_{\mathcal{X}}(z, \gamma(t)) \stackrel{*}{\lesssim} \sum_i d_{\mathcal{X}_{\alpha_i}}(z, \gamma(t)) = O(D). \quad \square$$

If  $S$  is not connected, then writing  $S = W_1 \cup \dots \cup W_k$  as a disconnected union, for any time  $t$ , as above gives

$$\begin{aligned} d_{\mathcal{X}}(z, \gamma(t)) &= \sum_i d_{\mathcal{X}(W_i)}(z, \gamma(t)) \\ &\stackrel{*}{\lesssim} \sum_i d_{\mathcal{X}(W_i)}(x, y) = O(D). \end{aligned}$$

In preparation for the next lemma, we recall a result of Rafi and Schleimer [27]. They give the following definition.

*Definition 5.2*

Given a pair of points  $x, y \in \mathcal{X}(S)$ , thresholds  $T_1 \geq T_0 > 0$ , and a subsurface  $W$ , a collection  $\Omega$  of subsurfaces  $W_i \subsetneq W$  is an antichain in  $W$  for  $x$  and  $y$  if the following statements hold.

- If  $W_i \in \Omega$ , then  $d_{\mathcal{C}(W_i)}(x, y) \geq T_0$ .
- If  $d_V(x, y) \geq T_1$ , then  $V \subset W_i$  where  $W_i \in \Omega$ .
- If  $W_i, W_j \in \Omega$ , then  $W_i$  is not a proper subsurface of  $W_j$ .

The size of  $\Omega$  is a lower bound for the distance in the curve complex.

LEMMA 5.3 (see [27])

There is a constant  $A = A(T_0, T_1)$  such that

$$|\Omega| \leq Ad_W(x, y).$$

We show that if  $\gamma$  is moving in some  $\mathcal{X}(W)$  but is not close to  $\mathcal{X}_{\partial W}$ , then this happens because  $\gamma$  is really moving in a subsurface of  $W$ .

LEMMA 5.4

For all sufficiently large  $M, D$  and any subsurface  $W$  of  $S$ , if  $\gamma: [a, b] \rightarrow \mathcal{X}$  is a preferred path connecting  $x$  to  $y$  such that for all  $t$

$$d_{\mathcal{X}}(\gamma(t), \mathcal{X}_{\partial W}) \geq M \quad \text{and} \quad d_{\mathcal{X}(W)}(x, y) \geq D,$$

then there is a proper subsurface  $V \subsetneq W$  such that

$$d_{\mathcal{X}(V)}(x, y) \stackrel{*}{\asymp} d_{\mathcal{X}(W)}(x, y).$$

*Proof*

Let  $M_0$  be the constant for Theorem 3.1, and let  $M_1$  be from Theorem 3.2. We will introduce additional uniform constants: a threshold constant  $T$  for the distance formula and  $M, D$  depending on  $T$  which will satisfy the conclusions of the lemma. All will depend solely on  $M_0, M_1$ , and the uniform constants coming from the definition of preferred paths.

We claim we can choose  $M$  large enough such that, for all  $x, y$ ,

$$d_W(x, y) \leq M.$$

Begin by choosing  $M, T$  so that

$$M > T > \max(2K(M_0 + C), M_1),$$

where  $K, C$  are the constants that arise from the fact that the projection of a preferred path to a surface is a  $(K, C)$ -quasi-geodesic.

Suppose now by contradiction that  $d_W(x, y) > M$ . Choose  $z \in \gamma$  whose projection to  $\mathcal{C}(W)$  is at the midpoint of the quasi-geodesic  $\pi_W \circ \gamma$ . From our assumption, we know that both

$$d_W(x, z) \geq \frac{M}{2} \quad \text{and} \quad d_W(z, y) \geq \frac{M}{2}.$$

Let  $\alpha = \partial W$ , and recall there is a projection map  $\phi_{\mathcal{X}_\alpha}: \mathcal{X} \rightarrow \mathcal{X}_\alpha$ . Let  $w = \phi_{\mathcal{X}_\alpha}(z) \in \mathcal{X}_{\partial W}$ . By the assumption of the lemma,

$$d_{\mathcal{X}}(z, w) \geq M.$$

By consistency,

$$d_U(z, w) \leq M_1$$

for all  $U \subsetneq W$  or  $U$  disjoint from  $W$ . Since  $M > T$ , the distance formula with threshold  $T$  says that there is a subsurface  $U$  such that either  $W \subsetneq U$  or  $U \pitchfork W$  and

$$d_U(z, w) \geq T.$$

Consider first the possibility that  $W \subsetneq U$  and  $d_U(z, \partial W) \geq T$ . Since the  $(K, C)$ -quasi-geodesic  $[x, y]_U$  in  $\mathcal{C}(U)$  passes through  $z_U$ , either  $[x, z]_U$  or  $[z, y]_U$  stays a distance at least  $\frac{T}{K} - C > M_0$  from  $\partial W$ . Then, by Theorem 3.1, either  $d_W(x, z) \leq M_0 < \frac{M}{2}$  or  $d_W(y, z) \leq M_0 < \frac{M}{2}$ , which is a contradiction.

Consider the other possibility that  $U \cap W$  with

$$d_U(z, \partial W) \geq T.$$

By the first consistency condition we have that

$$d_W(z, \partial U) \leq M_1.$$

For  $M$  again sufficiently large compared to  $M_1$ , the assumptions that  $d_W(x, z) \geq \frac{M}{2}$  and  $d_W(y, z) \geq \frac{M}{2}$  and the triangle inequality now imply both

$$d_W(x, \partial U) \geq \frac{M}{4}$$

and

$$d_W(y, \partial U) \geq \frac{M}{4}.$$

Again, the first consistency condition implies for  $M$  large enough that

$$d_U(x, \partial W) = O(1)$$

and

$$d_U(y, \partial W) = O(1),$$

which together imply that

$$d_U(x, y) = O(1)$$

by the triangle inequality. This in turn implies  $d_U(x, z) = O(1)$  and  $d_U(y, z) = O(1)$ , and using the triangle inequality one more time, we conclude that

$$d_U(z, \partial W) = O(1).$$

The implied constants here are uniform, not depending on  $M, T$ . Thus, choosing  $T$  larger than this constant we again have a contradiction. We have proved the claim.

Thus, we can assume we can choose  $M$  so that

$$d_W(x, y) \leq M$$

for all  $x, y$ . By the distance formula,

$$d_{\mathcal{X}(W)}(x, y) \stackrel{*}{\asymp} \sum_{U \in \mathcal{W}_M(x, y)} d_U(x, y).$$

Here  $\mathcal{W}_M(x, y)$  is the collection of subsurfaces  $U \subset W$  where  $d_U(x, y) > M$ . Since  $d_W(x, y) < M$ ,  $W$  itself is not in the sum. Consider the antichain  $\Omega$  in  $W$  for  $x$  and  $y$ . Then Lemma 5.3 applied with  $T_0 = T_1 = M$  implies that  $\Omega = \{V_1, \dots, V_k\}$  where  $k = O(M)$ . Each subsurface in  $\mathcal{W}_M(x, y)$  is a subset of some  $V_i$ . Hence, for  $V$  equal to some  $V_i$  and for  $D$  sufficiently large compared to  $M$ , we have

$$d_{\mathcal{X}(W)}(x, y) \stackrel{*}{\asymp} \sum_{U \in \mathcal{V}_M(x, y)} d_U(x, y),$$

where  $\mathcal{V}_M(x, y)$  is the collection of subsurfaces  $U \subseteq V$  where  $d_U(x, y) \geq M$ . That is,

$$d_{\mathcal{X}(V)}(x, y) \stackrel{*}{\asymp} d_{\mathcal{X}(W)}(x, y). \quad \square$$

PROPOSITION 5.5

There exists a constant  $D_0$  such that, given a subsurface  $W \subset S$  and a preferred path  $\gamma: [a, b] \rightarrow \mathcal{X}$  connecting  $x$  to  $y$  where

$$D = d_{\mathcal{X}(W)}(x, y) \geq D_0,$$

there are a subsurface  $V \subset W$  and a subinterval  $[c, d] \subset [a, b]$  so that

- $d_{\mathcal{X}(V)}(\gamma(c), \gamma(d)) \stackrel{*}{\asymp} D$ ;
- for  $t \in [c, d]$ ,  $d_{\mathcal{X}}(\gamma(t), \mathcal{X}_{\partial V}) = O(1)$ .

*Proof*

We use induction on the complexity of subsurfaces. If  $\gamma$  does not come within  $M$  of  $\mathcal{X}_{\partial W}$ , then we apply Lemma 5.4. Let  $W' \subsetneq W$  be a subsurface such that

$$d_{\mathcal{X}(W')}(x, y) \stackrel{*}{\asymp} D.$$

Since  $W'$  has lower complexity than  $W$ , Proposition 5.5 applies by induction. That is, there is a subsurface  $V \subset W' \subset W$  with the desired properties.

Thus, assume that  $\gamma$  does in fact come within  $M$  of  $\mathcal{X}_{\partial W}$ , and let  $z_1$  and  $z_2$  be points in  $\gamma$  marking the first and the last times  $\gamma$  is within  $M$  of  $\mathcal{X}_{\partial W}$ . We have either

$$d_{\mathcal{X}(W)}(x, z_1) \stackrel{*}{\asymp} D, \quad d_{\mathcal{X}(W)}(z_1, z_2) \stackrel{*}{\asymp} D, \quad \text{or} \quad d_{\mathcal{X}(W)}(z_2, y) \stackrel{*}{\asymp} D.$$

If  $d_{\mathcal{X}(W)}(z_1, z_2) \stackrel{*}{\asymp} D$ , then we are done after taking  $V = W$ . In the other two cases (say,  $d_{\mathcal{X}(W)}(x, z_1) \stackrel{*}{\asymp} D$ ), the path connecting  $x$  and  $z_1$  does not come close to  $\mathcal{X}_{\partial W}$  but travels a large distance in  $\mathcal{X}_W$ . Hence, we again can apply Lemma 5.4 and induction to finish the proof. □

5.1. Steady progress

Consider a preferred path that stays near the space  $\mathcal{X}_{\partial W}$  for some subsurface  $W$ . Sometimes it is desirable for  $\gamma$  to make steady progress in the curve complex of  $W$ . We make this notion precise.

Definition 5.6

Suppose that  $\gamma: [a, b] \rightarrow \mathcal{X}_{\partial W}$  is a preferred path connecting  $x = \gamma(a)$  and  $y = \gamma(b)$ . Let  $L = d_{\mathcal{X}(W)}(x, y)$ , and let  $a = t_0 < t_1 < t_2 < t_3 < t_4 < t_5 = b$  such that each  $y_i = \gamma(t_i)$  satisfies

$$L \stackrel{*}{\succ} d_{\mathcal{X}(W)}(y_i, y_{i+1}) \stackrel{+}{\succ} L/5.$$

For a constant  $C_0$ , we say that  $\gamma$  makes  $C_0$ -steady progress in  $W$  if, for  $i = 0, \dots, 4$ ,

$$d_W(y_i, y_{i+1}) \geq C_0.$$

Note that if  $\gamma$  makes  $C_0$ -steady progress in  $W$  for some  $C_0$ , then it makes  $C'_0$ -steady progress in  $W$  for  $C'_0 < C_0$ .

We first prove a fellow-traveling proposition that indicates one use of this concept. Then in a following lemma we show that we can find  $W$  in which we make steady progress. The fellow-traveling proposition states that if the endpoints of two preferred paths are close compared to their lengths and the first one makes steady progress in  $W$ , then the middle part of the second one also stays near  $\mathcal{X}_{\partial W}$ . Namely, let  $\gamma$  be a preferred path joining  $x, y \in \mathcal{X}_{\partial W}$  that makes  $C_0$ -steady progress in  $W$ . Let  $x = y_0, \dots, y_5 = y$  be as in Definition 5.6, and let  $z$  be a point in  $\gamma$  between  $y_2$  and  $y_3$  (see Figure 5) so that

$$d_{\mathcal{X}}(z, y_2) \stackrel{+}{\succ} D/10 \quad \text{and} \quad d_{\mathcal{X}}(z, y_3) \stackrel{+}{\succ} D/10.$$

PROPOSITION 5.7

There are uniform constants  $c_0, c_1, C_0, D_0$  so that, in the setup of the above paragraph, the following holds. Assume that  $D := d_{\mathcal{X}(W)}(x, y) \geq D_0$ , and suppose that  $\gamma'$  is a preferred path joining  $x'$  and  $y'$  with

$$d_{\mathcal{X}}(x, x') \leq c_0 D \quad \text{and} \quad d_{\mathcal{X}}(y, y') \leq c_0 D.$$

Then there is a subsegment of  $\gamma'$  with length comparable to  $D$  that stays in a bounded neighborhood of  $\mathcal{X}_{\partial W}$ . In fact, for  $z'$  on  $\gamma'$ , if

$$d_{\mathcal{X}}(z', z) \leq c_1 D, \quad \text{then} \quad d_{\mathcal{X}}(z', \mathcal{X}_{\partial W}) = O(1).$$

Proof

Consider the projection  $\pi_W(\gamma)$  to  $\mathcal{C}(W)$ . We have  $d_W(x, y_1) \stackrel{*}{\succ} C_0$ . Even though the

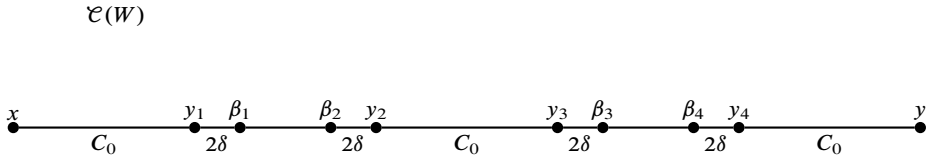


Figure 5. The projections of  $y_0, \dots, y_5$  in  $\mathcal{C}(W)$  are at least  $C_0$  apart. For  $i = 1, \dots, 4$ , the  $\beta_i$  in  $[x, y]_W$  is  $2\delta$  away from the projection of  $y_i$  in the indicated direction.

distance in  $\mathcal{X}$  between  $y_1$  and  $x$  is (up to additive error) at least  $D/5 \gg c_0 D$ , the projection of  $x'$  to  $\mathcal{C}(W)$  may be to the right of  $\pi_W(y_1)$  along  $[x, y]_W$ . However, this cannot happen if we travel a few steps toward  $y$  along  $[x, y]_W$ . We make this precise.

Pick a curve  $\beta_1$  along  $[x, y]_W$  that is a distance  $3\delta \geq 6$  toward  $y$  from the projection of  $y_1$  to  $W$ .

*Claim 5.8*

There is  $c_0$  sufficiently small so that, for any  $z'' \in \mathcal{X}$ ,  $d_{\mathcal{X}}(x, z'') \leq c_0 D$  implies that the closest point projection of  $z''_W$  to  $[x, y]_W$  lies on  $[x, \beta_1]$ .

We remind the reader that  $\delta$  is the hyperbolicity constant for  $\mathcal{C}(W)$ . We can assume that  $\delta \geq 2$ .

*Proof*

We know that  $d_{\mathcal{X}(W)}(x, y_1) \stackrel{\dagger}{\succ} D/5$  and that the distance in  $\mathcal{X}(W)$  is the sum of subsurface projections to subsurfaces in  $\mathcal{W}_T(x, y_1)$ . The boundary of any such subsurface is near a curve in  $[x, y_1]_W$ . For  $T$  larger than  $M_0$ , by Theorem 3.1 the projection of  $[\beta_1, y]_W$  to any subsurface  $U \in \mathcal{W}_T(x, y_1)$  has uniformly bounded diameter  $M_0$ . (Every curve in  $[\beta_1, y]_W$  intersects  $U$  since the distance of the projection  $y_1$  to  $W$  from  $\beta_1$  is at least 3.) Arguing by contradiction, suppose, for  $c_0$  small, that  $z''$  satisfies  $d_{\mathcal{X}}(x, z_1) \leq c_0 D$ , and yet the closest projection of  $z''_W$  to  $[x, y]_W$  lies on  $[\beta_1, y]_W$ . Then every curve in the geodesic  $[z'', y]_W$  also intersects  $U$  since it is at least a distance  $2\delta \geq 4$  from  $y_1$ . Hence,

$$d_U(x, z'') \stackrel{\dagger}{\succ} d_U(x, y_1).$$

Therefore,  $d_{\mathcal{X}(W)}(x, z'') \stackrel{*}{\succ} d_{\mathcal{X}(W)}(x, y_1) \stackrel{*}{\succ} D/5$ , and we have a contradiction for  $c_0$  sufficiently small. □

Similarly we find

- a curve  $\beta_4$  near the projection of  $y_4$  so that, for any  $z''$ , if

$$d_{\mathcal{X}(W)}(z'', y) \leq c_0 D,$$

then the closest point projection of  $z''_W$  to  $[x, y]_W$  lies on  $[\beta_4, y]_W$ ;

- curves  $\beta_2$  and  $\beta_3$ , respectively, near the projections of  $y_2$  and  $y_3$  to  $\mathcal{C}(W)$  so that, for any  $z_2, z_3 \in X$ ,

$$d_W(z_2, \beta_2) \leq \delta, \quad d_W(z_3, \beta_3) \leq \delta \implies d_{\mathcal{X}}(z_2, z_3) \stackrel{*}{\succ} D. \quad (16)$$

Now, by the claim, the closest point projection of  $x'_W$  to  $[x, y]_W$  lies on  $[x, \beta_1]_W$ , and the closest point projection of  $y'_W$  to  $[x, y]_W$  lies on  $[\beta_4, y]_W$ . It follows from the hyperbolicity of the curve complex that the  $2\delta$ -neighborhood of the path  $[x', y']_W$  has to contain  $[\beta_2, \beta_3]_W$ . In particular,

$$d_W(x', y') \stackrel{\perp}{\succ} C_0.$$

For  $C_0 > M_0$  this means that the path  $[x', y']$  passes within a distance 1 of  $X_{\partial W}$ . Let  $z'_1$  and  $z'_4$  be the first and the last times, respectively, in the path  $[x', y']$  that this occurs. A  $\delta$ -neighborhood of the geodesic  $[z'_1, z'_4]_W$  must also contain  $[\beta_2, \beta_3]_W$ .

This means that there are points  $z'_2$  and  $z'_3$  along  $[z'_1, z'_4]$  whose projection to  $\mathcal{C}(W)$  are  $\delta$ -close to  $\beta_2$  and  $\beta_3$ , respectively. Thus, by (16),

$$d_{\mathcal{X}}(z'_2, z'_3) \stackrel{*}{\succ} D \implies d_{\mathcal{X}}(z'_1, z'_4) \stackrel{*}{\succ} D.$$

This is the desired subsegment of  $\gamma'$ . To see the last assertion of the theorem, note that if  $d_{\mathcal{X}}(z', z) \leq c_1 D$  for  $c_1$  small enough, then  $z'$  is indeed in the segment  $[z'_2, z'_3]$ . This finishes the proof. □

The next lemma says that we can find subsurfaces where there is steady progress. The constant  $D_0$  appears in Proposition 5.5, Lemma 5.9, and Proposition 5.7. This means that we choose  $D_0$  large enough such that all three statements hold.

LEMMA 5.9

We can choose  $D_0$  such that, for any surface  $W$ , if  $\gamma$  is a preferred path joining  $x, y \in \mathcal{N}_{O(1)}(\mathcal{X}_{\partial W})$  and

$$D := d_{\mathcal{X}(W)}(x, y) \geq D_0,$$

then, for some interval  $[c, d] \subset [a, b]$  and some subsurface  $V \subset W$ , we have that

- $d_{\mathcal{X}(V)}(\gamma(c), \gamma(d)) \stackrel{*}{\succ} D$ ; and
- $\gamma|_{[c, d]}$  makes  $C_0$ -steady progress in  $V$ .

*Proof*

We first note that, by definition, a quasi-geodesic in a lowest-complexity subsurface

makes steady progress; otherwise, since it is a quasi-geodesic, it would have to make progress in the curve complex of *some* proper subsurface, but there are none. The proof is now by induction on complexity. If  $\gamma$  does not make steady progress in  $W$ , then for some  $i$ ,  $d_W(y_i, y_{i+1}) = O(1)$ . If so, we use the antichains (Lemma 5.3) and argue as in case 1 in the proof of Lemma 5.4 to conclude that there exists a subsurface  $V \subsetneq W$  where

$$d_{\mathcal{X}(V)}(y_i, y_{i+1}) \stackrel{*}{\succ} D.$$

Here the implied constant depends only on  $C_0$  and not on  $D$ . Now, an induction on the complexity of  $W$  implies the lemma. (Replace  $\gamma$  with the preferred path connecting  $y_i$  to  $y_{i+1}$ , and replace  $W$  with  $V$ .) □

We now prove a second fellow-traveling proposition. Recall that  $\delta$  is the hyperbolicity constant for the curve complex of  $S$ .

PROPOSITION 5.10

*Assume that  $S$  is connected, and let  $d_S(\cdot, \cdot)$  denote distance in  $\mathcal{C}(S)$ , the curve complex of  $S$ . There are constants  $C_1 < C_2$  with the following property. For two pairs of points  $x, y$  and  $x', y'$ , suppose that*

$$d_S(x, x') \leq C_1\delta \quad \text{and} \quad d_S(y, y') \leq C_1\delta.$$

*Suppose that  $z' \in \mathcal{G}(x', y')$  is such that*

$$d_S(z', x) \geq C_2 \quad \text{and} \quad d_S(z', y) \geq C_2.$$

*Then*

$$z' \in \mathcal{G}(x, y).$$

*Proof*

We need to show that, for any subsurface  $U$ ,

$$d_U(z', [x, y]_U) = O(1).$$

If  $U = S$ , then we know from the hyperbolicity of  $\mathcal{C}(S)$  that  $z'_S$  is within  $2\delta$  of  $[x', x]_S \cup [x, y]_S \cup [y, y']_S$ . Note that different paths connecting two points in  $\mathcal{C}(S)$  are  $O(\delta)$  apart in the metric  $d_S$ . For  $C_2$  large compared to  $C_1$ ,  $x'_S$  is much closer to  $x_S$  than  $z'_S$  is, so  $z'_S$  is far from the path  $[x, x']_S$ . Similarly,  $z_S$  is far from  $[y, y']_S$ . Thus, it has to be near  $[x, y]_S$ .

Now assume that  $U \neq S$ . If the distance in  $\mathcal{C}(S)$  from  $\partial U$  to  $z'_S$  is at least 3, then without loss of generality, we can assume that  $\partial U$  is at least a distance 2 from



$[z', y]_S$ . (Otherwise,  $\partial U$  would be at least a distance 2 from  $[x, z']_S$ .) Therefore, by Theorem 3.1  $d_U(z', y) \leq M_0$ . That is,  $z'_U$  is within  $M_0$  of  $y_U$  and, hence, within  $M_0$  of  $[x, y]_U$ .

If  $\partial U$  is within 3 of  $z'_S$  in  $\mathcal{C}(S)$ , then for  $C_2$  large compared to  $C_1$  it is a distance at least 2 from  $[x, x']_S$  and  $[y, y']_S$ . Therefore,

$$d_U(x, x') \leq M_0 \quad \text{and} \quad d_U(y, y') \leq M_0.$$

Hence,  $[x, y]_U$  is within  $O(M_0)$  of  $[x', y']_U$ . But we know that  $z'_U$  is within  $O(1)$  of  $[x', y']_U$ , and therefore, it is within  $O(M_0)$  of  $[x, y]_U$ . This finishes the proof.  $\square$

### 6. Local structure of efficient maps

We want to prove the following result. We assume that  $S$  is connected.

#### THEOREM 6.1

For all  $R_0, \epsilon_0$  there are  $R_1 \geq R_0$  and  $\epsilon_1 < \epsilon_0$  so that if  $B$  is a box in  $\mathbb{R}^n$  with  $|B| = R \geq R_1$  and  $f : B \rightarrow \mathcal{X}$  is an  $\epsilon R$ -efficient map with  $\epsilon \leq \epsilon_1$ , then there is a sub-box  $B' \subset B$  with  $R' = |B'| \geq R_0$  such that one of the following statements holds.

- $R' \succ^* \sqrt[3]{\epsilon^2} R$ , and for some curve  $\alpha$ ,  $f(B')$  lies within an  $O(\sqrt[3]{\epsilon} R')$ -neighborhood of  $\mathcal{X}_\alpha$ .
- $R' \succ^* \sqrt[3]{\epsilon} R$ , and there exist  $x, y$  so that  $f(B')$  lies within an  $O(\sqrt[3]{\epsilon} R')$ -neighborhood of a preferred path in  $\mathcal{X}$  joining  $x$  to  $y$ .

#### Proof

We can assume that the diameter  $f(R)$  is at least  $\sqrt[3]{\epsilon} R$ ; otherwise, the first case holds by taking  $B' = B$ ,  $R' = R$ , and any curve  $\alpha \in \pi_S \circ f(B)$ . Thus, choose  $x = f(p)$  and  $y = f(q)$ , where  $p, q \in B$  such that

$$d_{\mathcal{X}}(x, y) \geq \sqrt[3]{\epsilon} R.$$

Identify the geodesic segment  $[p, q]$  with an interval in  $\mathbb{R}$ . By Theorem 4.9,  $f([p, q])$  stays in the  $O(\epsilon R)$ -neighborhood of a preferred path  $\gamma$  joining  $x, y$ .

Now suppose, for some proper subsurface  $W \subsetneq S$ , that

$$d_{\mathcal{X}(W)}(x, y) \geq \sqrt[3]{\epsilon^2} R.$$

We now claim that the first conclusion of the theorem holds. That is, there are a sub-box  $B'$  of size

$$R' \succ^* d_{\mathcal{X}(W)}(x, y) \succ^* \sqrt[3]{\epsilon^2} R$$

and a curve  $\alpha$  such that  $f(B')$  lies within an  $O(\sqrt[3]{\epsilon} R')$ -neighborhood of  $\mathcal{X}_\alpha$ . We prove the claim.

It follows from Proposition 5.5 and Lemma 5.9 that we can find a subinterval  $[d_1, d_2] \subset [a, b]$ , a subsurface  $V$ , and  $C_0$  such that

- (1)  $d_{\mathcal{X}(V)}(\gamma(d_1), \gamma(d_2)) \succ R'$ ;
- (2)  $\gamma([d_1, d_2])$  stays in an  $O(1)$ -neighborhood of  $\mathcal{X}_{\partial V}$ ; and
- (3) the path  $\gamma|_{[d_1, d_2]}$  makes  $C_0$ -steady progress in  $V$ .

Let  $x_i = \gamma(d_i)$ . Let  $s_i \in [p, q]$  so that

$$d_{\mathcal{X}}(f(s_i), x_i) = O(\epsilon R).$$

For small  $c$  consider any two points  $t_1, t_2$  at distance  $cR'$  from  $s_1, s_2$ , respectively, and set  $y_i = f(t_i)$ . For  $c$  small enough we have

$$\begin{aligned} d_{\mathcal{X}}(x_i, y_i) &\leq d_{\mathcal{X}}(x_i, f(s_i)) + d_{\mathcal{X}}(f(s_i), y_i) \\ &\leq O(\epsilon R) + KcR' \leq c_0R', \end{aligned}$$

where  $c_0$  is the constant given by Proposition 5.7. Let  $d \in [d_1, d_2]$  be such that  $x = \gamma(d)$  is the midpoint of  $\gamma|_{[d_1, d_2]}$ . Let  $p$  be any point such that

$$d_{\mathcal{X}}(f(p), x) = O(\epsilon R).$$

By Proposition 5.7, there is  $c_1$  so that all  $y \in \mathcal{G}(y_1, y_2)$  that satisfy

$$d_{\mathcal{X}}(y, x) \leq c_1R'$$

also satisfy

$$d_{\mathcal{X}}(y, \mathcal{X}_{\partial V}) = O(1).$$

For  $c$  small, a box of size  $R'$  centered at  $p$  is mapped under  $f$  within a distance  $O(\epsilon R) = O(\sqrt[3]{\epsilon}R')$  of such  $y$  and so the image of the box lies within  $O(\sqrt[3]{\epsilon}R')$  of  $\mathcal{X}_{\alpha}$  for  $\alpha = \partial V$ . This proves the claim.

We continue the proof of the theorem. By the first part of the argument we can assume that, for all  $W \subsetneq S$ ,

$$d_{\mathcal{X}(W)}(x, y) \prec \sqrt[3]{\epsilon^2}R. \tag{17}$$

This and Lemma 3.6 imply that  $\gamma$  makes  $C'_0$ -steady progress in the entire surface  $S$  for some  $C'_0 \succ 1/\sqrt[3]{\epsilon}$ . For  $\epsilon$  small enough, this implies that it makes  $C_0$ -steady progress where  $C_0$  is the fixed constant of Proposition 5.7. Let  $C_1$  be the constant of Proposition 5.10. For a small but fixed  $c > 0$ , take a  $cR'$ -neighborhood of  $p$  and a  $cR'$ -neighborhood of  $q$  where now

$$R' = d_{\mathcal{X}}(x, y) \succ \sqrt[3]{\epsilon}R.$$

Let  $p', q'$  be any points in these neighborhoods, and let  $x' = f(p')$  and  $y' = f(q')$ . By Theorem 4.9 we can find a preferred path  $\gamma'$  joining  $x', y'$  within  $O(\epsilon R)$  of  $f([p', q'])$ . Since the map  $f$  is quasi-Lipschitz it follows, for  $c$  sufficiently small, that

$$d_{\mathcal{X}}(x', y') \succ^* R'.$$

Choose any point  $\hat{p}$  in the middle third of  $[p', q']$ . There is  $z' \in \mathcal{G}(x', y')$  whose projection to  $\mathcal{C}(S)$  is at least  $C_1$ -far from  $x'_S$  and  $y'_S$ , and

$$d_{\mathcal{X}}(f(\hat{p}), z') = O(\epsilon R).$$

By Proposition 5.10, we know that  $z' \in \mathcal{G}(x, y)$  and so  $f(\hat{p})$  is within  $O(\epsilon R)$  of  $\mathcal{G}(x, y)$ , and by Lemma 5.1, any point of  $\mathcal{G}(x, y)$  is within a distance  $O(\sqrt[3]{\epsilon^2} R)$  of  $\gamma$ .

We have shown that any point in the middle third of any segment starting near  $p$  and ending near  $q$  is mapped to a point that is in an  $O(\epsilon R)$ -neighborhood of  $\gamma$ . But such a path covers a box of size  $R'$ . Thus, there is a box of size  $R'$  which maps within  $O(\sqrt[3]{\epsilon^2} R) = O(\sqrt[3]{\epsilon} R')$  of a preferred path. We are done.  $\square$

**7. Proof of main theorems**

We are ready to prove Theorems A and B. We first prove a version of Theorem A for efficient maps. Then, we use coarse differentiation to finish the proof.

*Definition 7.1*

Let  $\alpha$  be a (possibly empty) curve system. For every connected component  $W$  of  $S \setminus \alpha$  (including annuli if  $\mathcal{X}$  is not  $\mathcal{P}(S)$ ), let  $\omega_W : I_W \rightarrow \mathcal{X}(W)$  be a preferred path. Consider the box  $B = \prod_W I_W \subset \mathbb{R}^n$ , where  $n$  is the number of components of  $S \setminus \alpha$ . Then

$$F : B \rightarrow \mathcal{X}_\alpha = \prod_W \mathcal{X}(W), \quad \text{where } F = \prod_W \omega_W,$$

is a quasi-isometric embedding because each  $\gamma_W$  is a quasi-geodesic and the product space is equipped with the  $L^1$ -metric. We call this map a *standard flat* in  $\mathcal{X}$ .

**THEOREM 7.2**

*Let  $S$  be a surface with complexity  $\xi = \xi(S)$ . For given  $\epsilon_0 < 1$  and  $R_0$ , let*

$$\epsilon_\xi = \epsilon_0^{(6^\xi)} \quad \text{and} \quad R_\xi = \frac{R_0}{\epsilon_\xi}.$$

Assume that  $f : B \rightarrow \mathcal{X}$  is an  $\epsilon_\xi$ -efficient map at scale  $R_\xi$ , where  $B$  is a box of size  $R_\xi$  in  $\mathbb{R}^n$ . Then, there is a box  $B' \subset B$  of size  $R' \geq R_0$  so that the image  $f(B')$  lies inside the  $O(\epsilon_0 R')$ -neighborhood of a standard flat in  $\mathcal{X}$ .

*Proof*

We prove the theorem by induction on the complexity  $\xi = \xi(S)$  of the surface  $S$  (see (10)). If  $\xi = 0$ , then

$$S = \prod_{i=1}^m S_i,$$

where each  $S_i$  is either a once-punctured torus or a four-times-punctured sphere. When  $\mathcal{X}$  is the pants complex,  $\mathcal{X}(S_i)$  is quasi-isometric to the Farey graph; when  $\mathcal{X}$  is the augmented marking space,  $\mathcal{X}(S_i)$  is isometric to a copy of the hyperbolic plane; and when  $\mathcal{X}$  is the marking complex,  $\mathcal{X}(S_i)$  is a graph whose vertices are the edges  $(\alpha, \beta)$  of the Farey graph and two vertices  $(\alpha, \beta)$  and  $(\alpha', \beta')$  are connected by an edge if the pairs share a curve, say,  $\alpha = \alpha'$  and  $\beta, \beta'$  are related by a Dehn twist about  $\alpha$ . The latter is known to be quasi-isometric to a tree. Hence, in all cases,  $\mathcal{X}(S_i)$  is a Gromov-hyperbolic space. That is,  $\mathcal{X}$  is a product of hyperbolic spaces.

In this case,  $R_\xi = R_0$  and  $\epsilon_\xi = \epsilon_0$ . Let  $f_i : B \rightarrow \mathcal{X}(S_i)$  be the projection of  $f$  to  $\mathcal{X}(S_i)$ . By Lemma 2.3,  $f_i$  is still  $\epsilon_0$ -efficient at scale  $R_0$ . Applying Proposition 2.14 to  $f_1 : B \rightarrow \mathcal{X}(S_1)$ , we obtain a sub-box  $B_1$  where  $f_1(B_1)$  lies in an  $O(\epsilon_0 R_0)$ -neighborhood of a quasi-geodesic in  $\mathcal{X}(S_1)$ . But in this case, quasi-geodesics are also preferred paths. Now we apply Proposition 2.14 to  $f_2 : B_1 \rightarrow \mathcal{X}(S_2)$  to obtain a box  $B_2$  so that  $f_2(B_2)$  lies in an  $O(\epsilon_0 R_0)$ -neighborhood of a preferred path in  $\mathcal{X}(S_2)$ . Continuing this way, we find a box  $B_m$  where the image of every  $f_i$  lies in an  $O(\epsilon_0 R_0)$ -neighborhood of a preferred path in  $\mathcal{X}(S_i)$ . This means that  $f(B_m)$  lies in an  $O(\epsilon_0 R_0)$ -neighborhood of a standard flat in  $\mathcal{X}$ . Note that  $B_m$  has the same size as  $B$  (within a uniform multiplicative error). This proves the base case of the induction.

Assume now that  $\xi$  is nonzero. Apply Theorem 6.1. If the second condition holds, then we are done, for in this case

$$R' \succ^* \sqrt[3]{\epsilon_\xi} R_\xi = \frac{\sqrt[3]{\epsilon_\xi}}{\epsilon_\xi} R_0 \geq R_0,$$

since a preferred path is itself a standard flat and  $\sqrt[3]{\epsilon_\xi} = \epsilon_{\xi-1}^2 \leq \epsilon_0$ .

Otherwise, we have a box  $\bar{B}$  of size

$$\bar{R} \succ^* \sqrt[3]{\epsilon_\xi^2} R_\xi \geq R_{\xi-1}$$

that maps to an  $O(\sqrt[3]{\epsilon_\xi} \bar{R})$ -neighborhood of  $\mathcal{X}_\alpha$  for some curve  $\alpha$ . The map  $f$  is  $\epsilon_{\xi-1}^2$ -efficient at scale  $\bar{R}$ , because reducing the size of the box by some factor (in this

case  $\sqrt[3]{\epsilon_\xi^{-2}}$  only makes the efficiency constant increase by the same factor and

$$\frac{\epsilon_\xi}{\sqrt[3]{\epsilon_\xi^2}} = \sqrt[3]{\epsilon_\xi} = \epsilon_{\xi-1}^2.$$

Composing  $f$  with the closest point projection map to  $\mathcal{X}_\alpha$  and using the fact that  $\mathcal{X}_\alpha$  is quasi-isometric to  $\mathcal{C}(\alpha) \times \mathcal{X}(S \setminus \alpha)$ , we obtain a map

$$\bar{f}: \bar{B} \rightarrow \mathcal{C}(\alpha) \times \mathcal{X}(S \setminus \alpha).$$

By Lemma 2.3(4),  $\bar{f}$  is  $\epsilon_{\xi-1}$ -efficient at scale  $\bar{R}$ .

Projecting to the second factor, we have an  $\epsilon_{\xi-1}$ -efficient map from a box of size  $R_{\xi-1}$  to  $\mathcal{X}(S \setminus \alpha)$ , which by the inductive assumption has a sub-box  $B_0$  of size at least  $R_0$  that stays in an  $O(\epsilon_0 R_0)$ -neighborhood of a standard flat in  $\mathcal{X}(S \setminus \alpha)$ . Now projecting to the first factor and applying Proposition 2.14 ( $\mathcal{C}(\alpha)$  is hyperbolic), we find a sub-box  $B'$  of size at least  $R_0$  that stays in an  $O(\epsilon_0 R_0)$ -neighborhood of a line in  $\mathcal{C}(\alpha)$ . That is,  $f(B')$  stays in an  $O(\epsilon_0 R_0)$ -neighborhood of a standard flat in  $\mathcal{X}$ . This finishes the proof. □

*Proof of Theorem A*

Let  $R_1$  be large enough so that the box  $B$  is guaranteed (by Theorem 2.5) to have a sub-box  $B_\xi$  of size at least  $R_\xi$ , where the restriction of  $f$  to  $B_\xi$  is  $\epsilon_\xi$ -efficient at scale  $R_\xi$ . Note that we are not using the full force of the theorem; we need only one efficient sub-box. Apply Theorem 7.2 to  $f: B_\xi \rightarrow \mathcal{X}$  to obtain the theorem. □

*Proof of Theorem B*

Pick  $1/\epsilon_0$  much larger than  $K$  and  $R_0$  much larger than  $C/\epsilon_0$ . Apply Theorem A to obtain an  $R_2$  so that given a box of size at least  $R_2$  there is a sub-box  $B'$  of size  $R' > R_0$  whose image is in an  $O(\epsilon_0 R')$ -neighborhood of a standard flat  $F: \mathbb{R}^m \rightarrow \mathcal{X}$ . Taking a composition of  $f$ , the closest point projection to the image of  $F$ , and then  $F^{-1}$ , we obtain a map  $G: B' \rightarrow \mathbb{R}^m$  with the property that, for  $p, q \in B'$ ,

$$d_{\mathbb{R}^m}(G(p), G(q)) \stackrel{*}{\asymp} d_{\mathbb{R}^n}(p, q) \pm O(\epsilon_0 R'). \tag{18}$$

We show that there is no such map if  $n$  is bigger than  $\text{rank}_{\text{top}}(\mathcal{X}) \geq m$ . The proof is similar to the proof of the nonexistence of quasi-isometries between  $\mathbb{R}^n$  and  $\mathbb{R}^m$ . Consider a net of  $O(\epsilon_0 R')^n$  points in  $B'$  that are pairwise  $K_1 \epsilon_0 R'$  apart, where  $K_1$  is much larger than the constants involved in (18). Then, by the choice of  $K_1$ , the image points are at least a distance  $\epsilon_0 R'$  apart and are contained in a ball of radius  $O(R')$  in  $\mathbb{R}^m$ . The number of points in such a net is of order  $O(R' \epsilon_0)^m$ . Choosing  $R'$  large enough (which can be done by choosing  $R_0$  large) we obtain a contradiction.

To see the second assertion, we note that if, for every subsurface  $W_i$ ,  $\mathcal{C}(W_i)$  contains an infinite geodesic, then the product of these geodesics is a quasi-isometric image of  $\mathbb{R}^n$ . This fails when  $W_i$  is an annulus and  $\mathcal{X}$  is the augmented marking space. (A horoball does not contain a bi-infinite geodesic.) In this case, we choose an  $\epsilon_0$ -thick point  $X$  in  $\mathcal{T}(S)$  and a pants decomposition of curves of length at most some fixed  $B$ . The point  $x \in \mathcal{X}$  associated to  $X$  is uniformly close to the product region associated to  $P$  (see Section 3.4). Consider an infinite ray for every horoball associated to a curve in  $P$ . The product of these rays is a quasi-isometric image of an orthant in  $\mathbb{R}^n$ .  $\square$

*Acknowledgments.* We are extremely grateful to the referee for a careful reading and numerous useful suggestions and corrections. We are also grateful to Saul Schleimer for helpful suggestions. We are grateful to Brian Bowditch for pointing out an error in a previous version of this article.

Eskin's work was supported in part by National Science Foundation (NSF) grants DMS 0905912 and DMS 1201422. Masur's work was supported in part by NSF grant DMS-0905907. Rafi's work was supported in part by NSF grant DMS-1007811.

## References

- [1] T. Aougab, *Uniform hyperbolicity of the graphs of curves*, *Geom. Topol.* **17** (2013), 2855–2875. [MR 3190300](#). (1552)
- [2] J. Behrstock, M. F. Hagen, and A. Sisto, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, to appear in *Geom. Topol.*, preprint, [arXiv:1412.2171v4](#) [math.GT]. (1519)
- [3] J. Behrstock, B. Kleiner, Y. N. Minsky, and L. Mosher, *Geometry and rigidity of mapping class groups*, *Geom. Topol.* **16** (2012), 781–888. [MR 2928983](#). (1520, 1521, 1541, 1542, 1551)
- [4] J. Behrstock and Y. N. Minsky, *Dimension and rank for mapping class groups*, *Ann. of Math. (2)* **167** (2008), 1055–1077. [MR 2415393](#). [DOI 10.4007/annals.2008.167.1055](#). (1520)
- [5] M. Bestvina, K. Bromberg, and K. Fujiwara, *Constructing group actions on quasi-trees and applications to mapping class groups*, *Publ. Math. Inst. Hautes Études Sci.* **122** (2015), 1–64. [MR 3415065](#). (1521, 1522, 1546, 1547, 1548, 1549, 1553)
- [6] B. H. Bowditch, *Uniform hyperbolicity of the curve graphs*, *Pacific J. Math.* **269** (2014), 269–280. [MR 3238474](#). [DOI 10.2140/pjm.2014.269.269](#). (1552)
- [7] ———, *Large-scale rank and rigidity of the Teichmüller metric*, to appear in *J. Topol.*, preprint, 2016. (1519)
- [8] J. Brock, *The Weil-Petersson metric and volumes of 3-dimensional hyperbolic convex cores*, *J. Amer. Math. Soc.* **16** (2003), 495–535. [MR 1969203](#). (1517, 1521, 1540)

- [9] J. BROCK and B. FARB, *Curvature and rank of Teichmüller space*, Amer. J. Math. **128** (2006), 1–22. [MR 2197066](#). (1519)
- [10] M. CLAY, K. RAFI, and S. SCHLEIMMER, *Uniform hyperbolicity of the curve graph via surgery sequences*, Algebr. Geom. Topol. **14** (2014), 3325–3344. [MR 3302964](#). (1552)
- [11] M. DURHAM, *The augmented marking complex of a surface*, J. Lond. Math. Soc. **94** (2016), 933–969. [DOI 10.1112/jlms/jdw065](#). (1521, 1538, 1539, 1541)
- [12] A. ESKIN, D. FISHER, and K. WHYTE, *Coarse differentiation of quasi-isometries, I: Spaces not quasi-isometric to Cayley graphs*, Ann. of Math. (2) **176** (2012), 221–260. [MR 2925383](#). (1520)
- [13] ———, *Coarse differentiation of quasi-isometries, II: Rigidity for Sol and lamplighter groups*, Ann. of Math. (2) **177** (2013), 869–910. [MR 3034290](#). [DOI 10.4007/annals.2013.177.3.2](#). (1520)
- [14] A. ESKIN, H. MASUR, and K. RAFI, *Rigidity of Teichmüller space*, preprint, [arXiv:1506.04774v1](#) [math.GT]. (1517, 1528)
- [15] B. FARB, A. LUBOTZKY, and Y. MINSKY, *Rank-1 phenomena for mapping class groups*, Duke Math. J. **106** (2001), 581–597. [MR 1813237](#). (1519)
- [16] U. HAMENSTÄDT, *Geometry of the mapping class group, III: Quasi-isometric rigidity*, preprint, [arXiv:math/0512429v2](#) [math.GT]. (1520)
- [17] S. HENSEL, P. PRZYTYCKI, and R. C. H. WEBB, *1-slim triangles and uniform hyperbolicity for arc graphs and curve graphs*, J. Eur. Math. Soc. (JEMS) **17** (2015), 755–762. [MR 3336835](#). (1552)
- [18] M. A. HERNÁNDEZ CIFRE, G. SALINAS, and S. SEGURA GOMIS, *Two optimization problems for convex bodies in the  $n$ -dimensional space*, Beitr. Algebra Geom. **45** (2004), 549–555. [MR 2093025](#). (1535)
- [19] H. A. MASUR and Y. N. MINSKY, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. **138** (1999), 103–149. [MR 1714338](#). [DOI 10.1007/s002220050343](#). (1538, 1540)
- [20] ———, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. **10** (2000), 902–974. [MR 1791145](#). (1521, 1538, 1539, 1540, 1541, 1551)
- [21] H. A. MASUR and S. SCHLEIMMER, *The geometry of the disk complex*, J. Amer. Math. Soc. **26** (2013), 1–62. [MR 2983005](#). [DOI 10.1090/S0894-0347-2012-00742-5](#). (1519)
- [22] Y. N. MINSKY, *Extremal length estimates and product regions in Teichmüller space*, Duke Math. J. **83** (1996), 249–286. [MR 1390649](#). (1546)
- [23] P. PRZYTYCKI and A. SISTO, *A note on acylindrical hyperbolicity of mapping class groups*, preprint, [arXiv:1502.02176v1](#) [math.GT]. (1552)
- [24] K. RAFI, *A characterization of short curves of a Teichmüller geodesic*, Geom. Topol. **9** (2005), 179–202. [MR 2115672](#). (1556)
- [25] ———, *A combinatorial model for the Teichmüller metric*, Geom. Funct. Anal. **17** (2007), 936–959. [MR 2346280](#). (1540, 1541, 1546, 1551)
- [26] ———, *Hyperbolicity in Teichmüller space*, Geom. Topol. **18** (2014), 3025–3053. [MR 3285228](#). (1521, 1538, 1541)

- [27] K. RAFI and S. SCHLEIMER, *Covers and the curve complex*, *Geom. Topol.* **13** (2009), 2141–2162. [MR 2507116](#). [DOI 10.2140/gt.2009.13.2141](#). (1557)

*Eskin*

Department of Mathematics, University of Chicago, Chicago, Illinois, USA;  
[eskin@math.uchicago.edu](mailto:eskin@math.uchicago.edu)

*Masur*

Department of Mathematics, University of Chicago, Chicago, Illinois, USA;  
[masur@math.uchicago.edu](mailto:masur@math.uchicago.edu)

*Rafi*

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada;  
[rafi@math.toronto.edu](mailto:rafi@math.toronto.edu)