Hyperbolicity of the complex of free factors

Mladen Bestvina and Mark Feighn*

January 31, 2013

Abstract

We prove that the complex of free factors of a free group of finite rank is hyperbolic.

Contents

1 Introduction 2
2 Review 6
3 Detecting boundedness in the free factor complex \( \mathcal{F} \) 13
4 More on folding paths 17
5 Loops with long illegal segments 27
6 Projection to a folding path 36
7 Hyperbolicity 53
8 Fellow Travelers and Symmetry 55
9 Thin Triangles 57

*Both authors gratefully acknowledge the support by the National Science Foundation.
1 Introduction

The complex of free factors of a free group $F$ of rank $n$ is the simplicial complex $F$ whose vertices are conjugacy classes of proper free factors $A$ of $F$, and simplices are determined by chains $A_1 < A_2 < \cdots < A_k$. The outer automorphism group $Out(F)$ acts naturally on $F$, which can be thought of as an analog of the Bruhat-Tits building associated with $GL_n(\mathbb{Z})$. This complex was introduced by Hatcher and Vogtmann in [20] where it is shown that it has the homotopy type of the wedge of spheres of dimension $n - 2$. They defined this complex in terms of sphere systems in $\#_n S^1 \times S^2$ and used variants in their work on homological stability [19, 21, 18].

There is a very useful analogy between $F$ and the curve complex $C$ associated with a compact surface (with punctures) $\Sigma$. The vertices of $C$ are isotopy classes of essential simple closed curves in $\Sigma$, and simplices are determined by pairwise disjoint curves. The curve complex was introduced by Harvey [17] and was classically used by Harer in his work on duality and homological stability of mapping class groups [15, 16]. The key result here is that the curve complex is homotopy equivalent to the wedge of spheres.

More recently, the curve complex has been used in the study of the geometry of mapping class groups and of ends of hyperbolic 3-manifolds. The fundamental result on which this work is based is the theorem of Masur and Minsky [26] that the curve complex is hyperbolic. In the low complexity cases when $\mathcal{C}$ is a discrete set one modifies the definition of $\mathcal{C}$ by adding an edge when the two curves intersect minimally. In the same way, we modify the definition of $F$ when the rank is $n = 2$ by adding an edge when the two free factors (necessarily of rank 1) are determined by a basis of $F$, i.e. whenever $F = \langle a, b \rangle$, then $\langle a \rangle$ and $\langle b \rangle$ span an edge. In this way $F$ becomes the standard Farey graph. The main result in this paper is:

**Main Theorem.** The complex $F$ of free factors is hyperbolic.

The statement simply means that when the 1-skeleton of $F$ is equipped with the path metric in which every edge has length 1, the resulting graph is hyperbolic.

There are variants of the definition that give rise to quasi-isometric complexes. For example, one can take the complex of partial bases, where vertices are conjugacy classes of elements that are part of a basis, and simplices correspond to compatibility, i.e. subsets of a basis. The $Aut(F)$-version of this complex was used in [10] to study the Torelli group. As another example, $F$ is quasi-isometric to the nerve of the cover $\{U(A)\}$ of the thin part of Outer space, where for a conjugacy class of proper free factors $A$, the
set \( U(A) \) consists of those marked graphs whose \( A \)-cover has core of volume \(< \epsilon \), for a fixed small \( \epsilon > 0 \).

Our proof is very much inspired by the Masur-Minsky argument, which uses Teichmüller theory. Bowditch \([7]\) gave a somewhat simpler argument. In the remainder of the introduction, we give an outline of the hyperbolicity of the curve complex which follows \([26]\) and \([7]\), and where we take a certain poetic license.

The proof starts by defining a coarse projection \( \pi : \mathcal{T} \to \mathcal{C} \) from Teichmüller space. To a marked Riemann surface \( X \) one associates a curve with smallest extremal length. To see that this is well defined one must argue that short curves intersect a bounded number of times (in fact, at most once), and then one uses the inequality

\[
d_{\mathcal{C}}(\alpha, \beta) \leq i(\alpha, \beta) + 1
\]

where \( i \) denotes the intersection number. Interestingly, the entire argument uses only this inequality to estimate distances in \( \mathcal{C} \) (and only for bounded intersection numbers).

Teichmüller space carries the Teichmüller metric, and any two points are joined by a unique Teichmüller geodesic. If \( t \mapsto X_t \) is a Teichmüller geodesic, consider the (coarse) path \( \pi(X_t) \) in \( \mathcal{C} \). One observes:

(i) The collection of paths \( \pi(X_t) \) is (coarsely) transitive, i.e. for any two curves \( \alpha, \beta \) there is a path \( \pi(X_t) \) that connects \( \alpha \) to \( \beta \) (to within a bounded distance).

Next, for any Teichmüller geodesic \( \{X_t\} \) one defines a projection \( \mathcal{C} \to \{X_t\} \). Essentially, for a curve \( \alpha \) the projection assigns the Riemann surface \( X_t \) on the path in which \( \alpha \) has the smallest length. The key lemma is the following (see \([26\, Lemma\, 5.8]\)), proved using the intersection number estimate above:

(ii) If \( \alpha, \beta \) are adjacent in \( \mathcal{C} \) and \( X_\alpha, X_\beta \) are their projections to \( X_t \), then \( \pi(X_\alpha) \) and \( \pi(X_\beta) \) are at uniformly bounded distance in \( \mathcal{C} \).

Consequently, one has a (coarse) Lipschitz retraction \( \mathcal{C} \to \pi(X_t) \) for every Teichmüller geodesic \( X_t \). It quickly follows that the paths \( \pi(X_t) \) are reparametrized quasi-geodesics (this means that they could spend a long time in a bounded set, but after removing the corresponding subintervals the resulting coarse path is a quasi-geodesic with uniform constants, after possibly reparametrizing).

The final step is:
(iii) Triangles formed by three projected Teichmüller geodesics are uniformly thin.

Hyperbolicity of $\mathcal{C}$ now follows by an argument involving an isoperimetric inequality (see [7, Proposition 3.1] and our Proposition 7.1).

Our argument follows the same outline. In place of Teichmüller metric and Teichmüller geodesics we use the Lipschitz metric on Outer space and folding paths. There are technical complications arising from the non-symmetry of the Lipschitz metric and the non-uniqueness of folding paths between a pair of points in Outer space. Our projection from $\mathcal{F}$ to a folding path comes in two flavors, left and right, and we have to work to show that the two are at bounded distance from each other when projected to $\mathcal{F}$. Similarly, we have to prove directly that projections of folding paths fellow travel, even when the two have opposite orientations. The role of simple closed curves is played by simple conjugacy classes in $\mathbb{F}$, i.e. nontrivial conjugacy classes contained in some proper free factor.

The first two hints that Outer space has some hyperbolic features was provided by Yael Algom-Kfir’s thesis [1] and by [5]. Algom-Kfir showed that axes of fully irreducible automorphisms are strongly contracting. In the course of our proof we will generalize this result (see Proposition 7.2) which states that all folding paths are contracting provided their projections to $\mathcal{F}$ travel with definite speed. In [5] a certain non-canonical hyperbolic $Out(\mathbb{F})$ complex was constructed. It is also known that fully irreducible automorphisms act on $\mathcal{F}$ with positive translation length [23, 5].

Below is a partial dictionary between Teichmüller space and Outer space relevant to this work.

| Surfaces | Free groups |
|----------|-------------|
| curve complex $\mathcal{C}$ | complex of free factors $\mathcal{F}$ |
| simple closed curve | free factor, or a simple conjugacy class |
| intersection number $i(\alpha, \beta)$ | number of times a loop in a graph crosses an edge |
| Teichmüller space | Outer space |
| Teichmüller distance | Lipschitz distance |
| Teichmüller geodesic | folding path |
| Teichmüller map | optimal map |
| quadratic differential | train track structure on the tension graph |
| horizontal curve | legal loop |
| vertical curve | illegal loop |
The paper is organized as follows. In Section 2 we review the basic notions about Outer space, including the Lipschitz metric, train tracks, and folding paths. Section 3 proves the analog of the inequality $d_C(\alpha, \beta) \leq i(\alpha, \beta) + 1$, using the Whitehead algorithm. Sections 4 and 5 contain some additional material on folding paths, including the formula for the derivative of a length function, as well as the key fact that a simple class which is largely illegal must lose a fraction of its illegal turns after a definite progress in $F$. In Section 6 we define the (left and right) projections of a free factor to a folding path and establish that images in $F$ of folding paths are reparametrized quasi-geodesics. The key technical lemma in that section is Proposition 6.10 (legal and illegal) establishing that the two projections are at bounded distance when measured in $F$. We end this section with a very useful method of estimating where the projections lie in Lemma 6.16. In Section 7 we recall the argument that for hyperbolicity it suffices to establish the Thin Triangles condition, and we also derive the contraction property of folding paths, measured in $F$. In Section 8 we prove the Fellow Travelers property (which of course follows from the Thin Triangles property), in both parallel and anti-parallel setting. Finally, in Section 9 we establish the Thin Triangles property.

The proofs of the three main technical statements in the paper, namely Proposition 5.4 (surviving illegal turns), Proposition 6.10 (legal and illegal), and Proposition 6.13 (closing up to a simple class), should be omitted on the first reading.

**Acknowledgments.** We thank the American Institute of Mathematics and the organizers and participants of the Workshop on Outer space in October 2010 for a fruitful exchange of ideas. We particularly thank Michael Handel for telling us a proof of Lemma 6.1. We thank Saul Schleimer for an inspiring conversation and Yael Algom-Kfir for her comments on an earlier version of this paper. We heartily thank the anonymous referee for a very careful reading and many suggestions we feel greatly improved the exposition. We also thank the referee for pointing out an error in one of two arguments we gave to finish the proof of Proposition 5.4.

Since the first version of this paper appeared on the arXiv, there have been some very interesting developments. Lee Mosher and Michael Handel [13] have proved that the free splitting complex $S$ for $F$ is hyperbolic. Ilya Kapovich and Kasra Rafi [24] have shown that the hyperbolicity of the complex of free factors follows from the hyperbolicity of the free splitting complex. Arnaud Hilion and Camille Horbez [22] have proved that $S$ is...
hyperbolic from the point of view of the sphere complex.

2 Review

In this section we review some definitions and collect standard facts about Outer space, Lipschitz metric, train tracks, and folding paths.

**Outer space.** A graph is a cell complex $G$ of dimension $\leq 1$. The rose $R_n$ of rank $n$ is the graph with one 0-cell (vertex) and $n$ 1-cells (edges). A vertex of $G$ not of valence 2 is topological. The closure in $G$ of a component of the complement of the set of topological vertices is a topological edge. In particular, a topological edge may be a circle. A marking of a graph $G$ is a homotopy equivalence $g : R_n \to G$. A metric on $G$ is a function $\ell$ that to each edge $e$ assigns a positive number $\ell(e)$. We often view the graph $G$ as the path metric space in which each edge $e$ has length $\ell(e)$. The Unprojectivized Outer space $\mathcal{X}$ is the space of equivalence classes of triples $(G, g, \ell)$ where $G$ is a finite graph with no vertices of valence $\leq 2$, $g$ is a marking of $G$, and $\ell$ is a metric on $G$. Two triples $(G, g, \ell)$ and $(G', g', \ell')$ are equivalent if there is a homeomorphism $h : G \to G'$ that preserves edge-lengths and commutes with the markings up to homotopy. Outer space $\mathcal{X}$ is the space of projective classes of such triples, i.e. modulo scaling the metric. Equivalently, $\mathcal{X}$ is the space of triples as above where the metric is normalized so that the volume $\text{vol}(G) := \sum \ell(e) = 1$. Assigning length 0 to an edge is interpreted as a metric on the graph with that edge collapsed, and in this way $\mathcal{X}$ becomes a complex of simplices with missing faces (the missing faces correspond to collapsing nontrivial loops), which then induces the simplicial topology on $\mathcal{X}$. Outer space was introduced by Culler and Vogtmann [9], who showed that $\mathcal{X}$ is contractible. We will usually suppress markings and metrics and talk about $G \in \mathcal{X}$. It is sometimes convenient to pass to the universal cover and regard $G \in \hat{\mathcal{X}}$ as an action of $\mathbb{F}$ on the tree $\tilde{G}$.

We find the following notation useful. If $z$ is a nontrivial conjugacy class, it can be viewed as a loop in the rose $R_n$, and via the marking can be transported to a unique immersed loop in any marked graph $G$. This loop will be denoted by $z|G$. The length of this loop, i.e. the sum of the lengths of edges crossed by the loop, counting multiplicities, is denoted by $\ell(z|G)$. Note that if $z$ is not simple, then $\ell(z|G) \geq 2 \text{vol}(G)$ and in fact $z|G$ must cross every edge at least twice.
Morphisms between trees and train tracks. Recall that a morphism between two $\mathbb{R}$-trees $S, T$ is a map $\tilde{\phi} : S \to T$ such that every segment $[x, y] \subset S$ can be partitioned into finitely many subintervals on which $\tilde{\phi}$ is an isometric embedding. For simplicity, in this paper we work only with simplicial metric trees. A direction at $x \in S$ is a germ of nondegenerate segments $[x, y]$ with $y \neq x$. The set $D_x$ of directions at $x$ can be thought of as the unit tangent space; a morphism $\tilde{\phi} : S \to T$ determines a map $D_{\tilde{\phi}_x} : D_x \to D_{\tilde{\phi}(x)}$, thought of as the derivative. A turn at $x$ is an unordered pair of distinct directions at $x$. A turn $\{d, d'\}$ at $x$ is illegal (with respect to $\tilde{\phi}$) if $D_{\tilde{\phi}_x}(d) = D_{\tilde{\phi}_x}(d')$. Otherwise the turn is legal. There is an equivalence relation on $D_x$ where $d \sim d'$ if and only if $d = d'$ or $\{d, d'\}$ is illegal. The equivalence classes are gates. The collection of equivalence classes at each $x$ is called the illegal turn structure on $S$ induced by $\tilde{\phi}$. If at each $x \in S$ there are at least two gates, the illegal turn structure is called a train track structure on $S$. This is equivalent to the requirement that $\tilde{\phi}$ embeds each edge of $S$ and has at least two gates at every vertex. A path in $S$ is legal if it makes only legal turns.

If $S \xrightarrow{\tilde{\phi}} T \xrightarrow{\tilde{\psi}} U$ is a composition of morphisms then there are two illegal turn structures of interest on $S$: one induced by $\tilde{\psi}$ and the other induced by the composition. In this situation, we will sometimes refer to the second of these as the pullback illegal turn structure on $S$ via $\tilde{\phi}$. Note that an illegal turn in the first structure is also illegal in the second. In particular, if the second structure is a train track then so is the first.

If $S$ and $T$ are equipped with abstract train track structures (equivalence relation on $D_x$ for every vertex $x$ with at least two gates), we say that a morphism $\tilde{\phi} : S \to T$ is a train track map if on each edge $\tilde{\phi}$ is an embedding and legal turns are sent to legal turns. In particular, legal paths map to legal paths.

We also extend this terminology to maps between graphs. If $\phi : \Delta \to \Sigma$ is a map between connected metric graphs such that the lift $\tilde{\phi} : \tilde{\Delta} \to \tilde{\Sigma}$ is a morphism of trees, then we also say that $\phi$ is a morphism and we can define the notion of legal and illegal turns on $\tilde{\Delta}$, which descends to $\Delta$. If there are at least two gates at each point, we have a train track structure on $\Delta$. If $\Delta$ and $\Sigma$ are equipped with abstract train track structures, the map $\phi$ is a train track map if it sends edges to legal paths and legal turns to legal turns. When the graphs $\Delta$ or $\Sigma$ are not connected we work with components separately.

\[\text{The standard usage of the term train track map is to self-maps of a graph, but this natural extension of the terminology should not cause any confusion.}\]
Lipschitz metric and optimal maps. Let $G$ and $G'$ be two points in $\hat{X}$. The homotopy class of maps $h : G \to G'$ such that $hg \simeq g'$ (with $g,g'$ markings for $G,G'$) is called the difference of markings. If $G$ and $G'$ are in $\hat{X}$, i.e. if they have volume 1, the Lipschitz distance $d_X(G,G')$ between $G$ and $G'$ is the log of the minimal Lipschitz constant over all differences of markings maps $G \to G'$. The Lipschitz distance is an asymmetric metric on $\hat{X}$ inducing the correct topology. For more information, see [11, 2, 4].

The basic fact, that plays the role of Teichmüller’s theorem in Teichmüller theory, is the following statement, due to Tad White. For a proof see [11] or [4].

**Proposition 2.1.** Let $G,G' \in \hat{X}$. There is a difference of markings map $\phi : G \to G'$ with the following properties.

- $\phi$ sends each edge of $G$ to an immersed path (or a point) with constant speed (called the slope of $\phi$ on that edge).
- The union of all edges of $G$ on which $\phi$ has the maximal slope $\lambda = \lambda(\phi)$, is a subgraph of $G$ with no vertices of valence 1. This subgraph is called the tension graph, denoted $\Delta = \Delta(\phi)$.
- $\phi$ induces a train track structure on $\Delta$.

The last bullet says that the map $\lambda \Delta \to G'$ induced by $\phi$ is a morphism immersing edges and each vertex has at least two gates. Note that the Lipschitz constant of $\phi$ is $\lambda$ and so, in the case that $G,G' \in X$, we have that $d_X(G,G') \leq \log \lambda$. In fact equality holds since the presence of at least two gates at each vertex of $\Delta$ guarantees that it contains legal loops, whose length gets stretched by precisely $\lambda$ so there can be no better map homotopic to $\phi$. We call any $\phi$ satisfying Proposition 2.1 an optimal map. A morphism that induces a train track structure is an example of an optimal map. Unfortunately, unlike Teichmüller maps, optimal maps are not uniquely determined by $G,G'$.

A legal loop can be constructed in $\Delta$ by starting with an edge and extending it inductively to longer legal edge paths until some oriented edge is repeated. In fact, this guarantees the existence of a “short” legal path. We will say a loop in $\Delta$ is a candidate if it is either embedded, or it forms the figure 8, or it forms a “dumbbell”. Every candidate determines a conjugacy class that generates a free factor of rank 1. Thus graph $\Delta$ admits a legal candidate. See [11].

**Folding at speed 1.** Now assume that $S,T$ represent points of Unprojectivized Outer space $\hat{X}$ (i.e. they are universal covers of marked metric
graphs), and \( \tilde{\phi} : S \to T \) is an equivariant morphism. Equivariantly subdivide \( S \) so that \( \tilde{\phi} \) embeds each edge and the inverse image under \( \phi \) of a vertex of \( T \) is a vertex of \( S \). Choose some \( \epsilon > 0 \) smaller than half of the length of any edge in \( S \). Then for \( t \in [0, \epsilon] \) define the tree \( S_t \) as the quotient of \( S \) by the equivalence relation: \( u \sim_t v \) if and only if there is a vertex \( x \) with \( d(x, u) = d(x, v) \leq t \) and \( \tilde{\phi}(u) = \tilde{\phi}(v) \). Then \( S_t \) represents a point in \( \hat{X} \) and \( \tilde{\phi} \) factors as \( S_t \tilde{\phi}_0 \to S_t \tilde{\phi}_t \to T \) for some equivariant morphism \( \tilde{\phi}_{t\infty} \) and the quotient map \( \tilde{\phi}_0 : S \to S_t \), which is also an equivariant morphism. The trees \( S_t, t \in [0, \epsilon] \) form a path in \( \hat{X} \), and \( S_0 = S \). We say that \( S_t \) is the path obtained from \( S \) by folding all illegal turns at speed 1 with respect to \( \tilde{\phi} \).

If \( \tilde{\phi} \) induces only one gate at some vertex \( v \in S \), then \( S_t \) will have a valence 1 vertex for \( t > 0 \). In that case we always pass to the minimal subtree of \( S_t \). When \( \tilde{\phi} \) induces a train track structure on \( S \), \( S_t \) is automatically minimal (if \( S \) is). For simplicity we state Proposition 2.2 in the train track situation only.

**Proposition 2.2.** Let \( \tilde{\phi} : S \to T \) be an equivariant morphism between two trees in \( \hat{X} \) inducing a train track structure on \( S \). There is a (continuous) path \( S_t \) in \( \hat{X} \), \( t \in [0, \infty) \), and there are equivariant morphisms \( \tilde{\phi}_{st} : S_s \to S_t \) for \( s \leq t \) so that the following holds:

1. \( S_0 = S, S_t = T \) for \( t \) large;
2. \( \tilde{\phi}_{tt} = Id, \tilde{\phi}_{su} = \tilde{\phi}_{tu}\tilde{\phi}_{st} \) for \( s \leq t \leq u \);
3. \( \tilde{\phi}_{0t} = \tilde{\phi} \) and \( \tilde{\phi}_{tt'} = Id \) for large \( t < t' \);
4. each \( \tilde{\phi}_{st} \) isometrically embeds edges and induces at least two gates at every vertex of \( S_s \);
5. for \( s < t, t' \) the illegal turns at vertices of \( S_s \) with respect to \( \tilde{\phi}_{st} \) coincide with those with respect to \( \tilde{\phi}_{st'} \), so \( S_s \) has a well-defined train track structure; and
6. for every \( s < t \) there is \( \epsilon > 0 \) so that \( S_{s+\tau}, \tau \in [0, \epsilon] \) is obtained from \( S_s \) by folding all illegal turns at speed 1 with respect to \( \tilde{\phi}_{st} \).

Moreover, this path is unique.

**Proof.** Uniqueness is clear from the definition of folding at speed 1. There can be no last time \( s \) so that two paths satisfying the above conditions agree (including the maps \( \tilde{\phi}_{tt'} \)) up to \( S_s \) but no further, by item (6).
There are three methods to establish existence, and they will be only sketched.

2.2. A. Stallings’ Method. This works when $S$ and $T$ can be subdivided so that $\tilde{\phi}$ is simplicial and all edge lengths are rational (or fixed multiples of rational numbers). In our applications we can arrange that this assumption holds. Then we may subdivide further so that all edge lengths are equal. The path $S_t$ is then obtained exactly as in the Stallings’ beautiful paper [29], by inductively identifying any pair of edges with a common vertex that map to the same edge in $T$. This operation of elementary folding can be performed continuously to yield a 1-parameter family of trees, i.e. a path, between the original tree and the folded tree. Putting these paths together gives the path $S_t$.

2.2. B. Via the vertical thickening of the graph of $\tilde{\phi}$. This method is due to Skora [28], who built on the ideas of Steiner. Skora’s preprint was never published; the interested reader may find the details in [8] and [12]. Consider the graph of $\tilde{\phi}$ as a subset of $S \times T$ and define the “vertical $t$-thickening” of it as

$$W_t = \{ (x,y) \in S \times T \mid d(\tilde{\phi}(x),y) \leq t \}$$

Next, consider the decomposition $D_t$ of $W_t$ into the path components of the sets $W_t \cap S \times \{y\}$, $y \in T$. Let $S_t = W_t/D_t$ be the decomposition space with the metric defined as follows. A path in $W_t$ is linear if its projection to both $S$ and $T$ has constant speed (possibly speed 0). A piecewise linear path $\gamma$ in $W_t$ is taut if the preimages $\gamma^{-1}(\ell)$ of leaves in $D_t$ are connected. Then define the distance in $S_t$ as the length of the projection to $T$ of any piecewise linear taut path connecting the corresponding leaves in $W_t$. In this way $S_t$ becomes a metric tree. The morphisms $\tilde{\phi}_{st}$ are induced by inclusion $W_s \hookrightarrow W_t$.

2.2. C. Via integrating the speed 1 folding direction. Starting at $S$ consider the path $S_t$, $t \in [0, \epsilon]$, obtained by folding all illegal turns at speed 1. Now extend this path by folding all illegal turns of $S_t$ at speed 1. Continue in this way inductively, and show that either $T$ is reached in finitely many steps, or there is a well defined limiting tree, from which folding can proceed. This is the approach taken in [11], to which the reader is referred for further discussion.

One possible approach is as follows. Say $S_t$ is defined for $t \in [0, t_0)$ with $S_0 = S$. To define the limiting tree $S_{t_0}$, note that for each conjugacy class, the length along the path is nonincreasing and thus converges. The limiting length function defines a tree, representing a point in compactified Outer
space. The lengths of conjugacy classes are bounded below by their values in \( T \), so the limiting tree \( S_{t_0} \) is free simplicial and thus represents a point in Outer space. We may view the tree \( S_{t_0} \) as the equivariant Gromov-Hausdorff limit of the path \( S_t \). The maps \( S \to S_t \), viewed as subsets of \( S \times S_t \) via their graphs, subconverge to a morphism \( S \to S_{t_0} \), and similarly by a diagonal argument one constructs morphisms \( S_t \to S_{t_0} \) that compose correctly. To show uniqueness of such morphisms, one uses Gromov-Hausdorff limits and the fact that the only (equivariant) morphism \( S_{t_0} \to S_{t_0} \) is the identity.

The path \( S_t, t \in [0, \infty) \) from Proposition 2.2 is the path induced by \( \tilde{\phi} \).

The following three lemmas are stated for clarity; their proofs are left as exercises. (The first is immediate from 2.2.B and implies the other two.)

**Lemma 2.3.** Let \( S_t, t \in [0, \infty) \), be the path in \( \hat{X} \) induced by \( \tilde{\phi} : S \to T \) and \([s_1, s_2]\) be a path in \( S_0 \). Suppose \( \tilde{\phi}(s_1) = \tilde{\phi}(s_2) \) and set \( h \) equal to the outradius of \( \tilde{\phi}([s_1, s_2]) \) with respect to \( \tilde{\phi}(s_1) \), i.e.

\[
h = \max_{s \in [s_1, s_2]} d_T(\tilde{\phi}(s_1), \tilde{\phi}(s))
\]

Then \( s_1 \) and \( s_2 \) are identified by time \( h \) but not before, i.e. \( \tilde{\phi}_0t(s_1) = \tilde{\phi}_0t(s_2) \) iff \( t \geq h \).

**Lemma 2.4.** Let \( S_t \) be the path in \( \hat{X} \) induced by \( S \to T \) and \( R \to S \) be an equivariant morphism with \( R \in \hat{X} \). The path \( R_t \) induced by the composition \( R \to S \to T \) and the path induced by \( R \to S_{t_0} \) agree for \( t \in [0, t_0] \).

If \( A \) is a nontrivial finitely generated subgroup of \( \mathbb{F} \) and \( T \in \hat{X} \), denote by \( A|T \) the minimal \( A \)-invariant subtree of \( T \).

**Lemma 2.5.** Let \( R, S, T \in \hat{X} \), \( R \to S \) and \( S \to T \) be equivariant morphisms, and \( A \) be a finitely generated subgroup of \( \mathbb{F} \). Suppose \( S_t \) is induced by \( S \to T \) and \( R_t \) is induced by the composition \( R \to S \to T \). If \( A|R \to A|S \) is an isomorphism, then so is \( A|R_t \to A|S_t \).

If \( S_t \) is a path in \( \hat{X} \) induced by the morphism \( \tilde{\phi} : S \to T \), then \( A|S_t \) is a path in the Unprojectivized Outer space \( \hat{X}(A) \) of \( A \). It is locally obtained by folding all illegal turns of the induced morphism \( A|S_t \to T \) and passing to minimal subtrees. For \( t < t' \), there is a generalized morphism \( A|S_t \to A|S_{t'} \), i.e. every segment in the domain has a partition into finitely many subsegments with the property that on each subsegment the map is an isometric embedding or degenerate. We will not need generalized morphisms in this paper.
Folding paths. Suppose $S \to T$ is an equivariant morphism between trees in $\hat{X}$ that induces a train track structure on $S$. The induced path $S_t$, $t \in [0, \infty)$, is a folding path in $\hat{X}$. The projection of $S_t$ to $X$ is a folding path in $X$. A folding path $G_t$ in $X$ can be parametrized by arclength, so that $d_X(G_t, G_t') = t - t'$ for $t' \leq t$.

Notation 2.6. From now on, we switch to thinking of points of Outer space as finite graphs when discussing folding paths. We have several meanings for folding path. To avoid confusion, we will use the following notation:

1. $\hat{G}_t$, $t \in [0, \omega)$, denotes a folding path in $\hat{X}$, i.e. a path in $\hat{X}$ induced by an equivariant morphism $S \to T$ giving a train track structure on $S$. If $\omega \in [0, \infty)$ is minimal such that $\hat{G}_\omega = T$, we also refer to $\hat{G}_t$, $t \in [0, \omega]$, as a folding path. We call $t$ the natural parameter.

2. $\bar{G}_t$, $t \in [0, \omega)$ (or $\bar{G}_t$, $t \in [0, \omega]$) denotes a folding path in $X$ obtained by projecting a path $\hat{G}_t$ as in (1). So, $\bar{G}_t = \hat{G}_t \cdot \operatorname{vol}(\hat{G}_t)$.

3. $G_t$, $t \in [0, L]$, denotes a path as in (2), but reparametrized in terms of arclength with respect to $d_X$. So, $G_{t(s)} = G_s$ where

$$t(s) = \frac{\operatorname{vol}(\hat{G}_0)}{\operatorname{vol}(G_s)}$$

For all $t \leq t' \in [0, L]$ there is a morphism $e^{t'-t}G_t \to G_t$.

Unless otherwise noted, a folding path $G_t$ without further adjectives or decorations will mean a folding path as in (3).

Proposition 2.7 ([11]). Let $G, \Sigma \in \mathcal{X}$. There is a geodesic in $\mathcal{X}$ from $G$ to $\Sigma$ which is the concatenation of two paths, the first is a (reparametrized) linear path in a single simplex of $\mathcal{X}$, and the second is a folding path in $\mathcal{X}$ parametrized by arclength.

Proof. Fix an optimal map $\phi : G \to \Sigma$ and let $\Delta = \Delta(\phi) \subset G$ be its tension graph with maximal slope $\lambda = \lambda(\phi)$. If $\Delta = G$ then the rescaled map $\lambda \phi : \lambda G \to \Sigma$ is a morphism and satisfies the requirement that it induces a train track structure on $G$. Proposition 2.2 gives a folding path $\hat{G}_t$ in $\hat{X}$ from $\lambda G$ to $\Sigma$ and morphisms $\hat{\phi}_{st} : \hat{G}_s \to \hat{G}_t$. To see that $\hat{G}_t$ projects to a geodesic $G_t$ in $X$, note that $\phi_{su} = \hat{\phi}_{tu} \hat{\phi}_{st}$ for $s < t < u$ implies $d_X(G_s, G_u) = d_X(G_s, G_t) + d_X(G_t, G_u)$.

Now suppose $\Delta \neq G$. Denote by $e_1, \ldots, e_k$ the (topological) edges of $G$ outside of $\Delta$. For each tuple $x = (x_1, x_2, \ldots, x_k)$ of lengths in the cube
[0, \ell_1] \times [0, \ell_2] \times \cdots \times [0, \ell_k], \text{ where } \ell_i \text{ is the length of } e_i \text{ in } G, \text{ denote by } \mu(x) \text{ the smallest maximal slope among maps } G \to \Sigma \text{ that are homotopic to } \phi \text{ rel } \Delta, \text{ where } G \text{ is given the metric } x \text{ outside } \Delta \text{ (so } \mu(x) = \infty \text{ if some loop is assigned length 0). Among all } x \text{ in the cube with } \mu(x) = \lambda \text{ choose one with the smallest sum of the coordinates, say } x_0. \text{ Denote by } \hat{G}' \in \hat{X} \text{ the graph } G \text{ with the metric } x_0 \text{ outside of } \Delta. \text{ (Some edges of } \hat{G}' \text{ may get length } 0 \text{ in which case its projection } G' \text{ to } X \text{ is on the boundary of the original simplex.)}

Let } \phi_0 : \hat{G}' \to \Sigma \text{ be a map homotopic to } \phi \text{ rel } \Delta, \text{ linear on edges, and with the maximal slope } \Delta(\phi_0) = \hat{G}'. \text{ Indeed, } \Delta(\phi_0) = \hat{G}' \text{ (otherwise some edge length can be reduced contradicting the choice of } x_0) \text{ and } \phi_0 \text{ induces at least two gates at every vertex (otherwise } \phi_0 \text{ may be perturbed so that the tension graph becomes a proper subgraph, see the proof of Proposition 2.1 e.g. in [4]).}

Since we have maps } G \to G' \text{ and } G' \to \Sigma \text{ with slopes 1 and } \lambda \text{ respectively, we also have } d_X(G, \Sigma) = d_X(G, G') + d_X(G', \Sigma).

\section{Detecting boundedness in the free factor complex } F

In this section we define a coarse projection } \pi : \mathcal{X} \to F \text{ and prove an analog of the inequality } d_C(\alpha, \beta) \leq 1 + i(\alpha, \beta) \text{ (see Lemma 3.2). An immediate consequence is that } \pi \text{ is coarsely Lipschitz (see Corollary 3.5).}

Recall that a nontrivial conjugacy class } x \text{ in } F \text{ is } simple \text{ if any (equivalently some) representative is contained in a proper free factor. If } x \text{ is a simple class, denote by } \hat{x} \text{ the conjugacy class of a smallest free factor containing a representative of } x. \text{ A proper connected subgraph } P \text{ of a marked graph } G \text{ that contains a circle defines a vertex } \hat{P} \text{ of } F.

\begin{lemma}
\label{lem:3.1}
Let } G \text{ be a marked graph. If } P, Q \subset G \text{ are proper connected subgraphs defining free factors } \hat{P}, \hat{Q} \text{ then } d_F(\hat{P}, \hat{Q}) \leq 4.
\end{lemma}

\begin{proof}
If } \text{rank}(F) = 2 \text{ then } d_F(\hat{P}, \hat{Q}) \leq 1 \text{ (using the modified definition of } F). \text{ Now assume } \text{rank}(F) \geq 3. \text{ Enlarge } P, Q \text{ to connected graphs } P', Q' \text{ that contain all but one edge of } G. \text{ Thus their intersection contains a circle } R, \text{ so we have a path in } F \text{ given by subgraphs } P, P', R, Q', Q \text{ and we see } d_F(\hat{P}, \hat{Q}) \leq 4.
\end{proof}

If } G \text{ is a marked graph, define } \pi(G) := \{ \hat{P} \mid P \text{ is a proper, connected, noncontractible subgraph of } G\}
The induced multi-valued, $\text{Out}(F)$-equivariant map $X \to F$, still called $\pi$, given by $G \mapsto \pi(G)$ is coarsely defined in that, by Lemma 3.1, the diameter of each $\pi(G)$ is bounded by 4. We refer to $\pi$ as the coarse projection from $X$ to $F$.

Lemma 3.2. Let $G$ be a marked graph and $x$ a simple class in $F$. If $x|G$ crosses an edge $e$ $k$ times, then the distance in $F$ between $\dot{x}$ and some free factor represented by a subgraph of $G$ is $\leq 6k + 9$.

Proof. This proof will use the fact, due to Reiner Martin [25], that if the Whitehead graph of a simple class $x$ is connected then it has a cut point. The classical fact, due to Whitehead [31], is the analogous statement for the special case that $x$ is primitive. Stallings’ paper [30] is a good modern reference for Whitehead’s result and the reader is directed there for the definitions of Whitehead graph and Whitehead automorphism.

First assume that $e$ is nonseparating. By collapsing a maximal tree in $G$ that does not contain $e$ we may assume that $G$ is a rose. Let $a_1, a_2, \ldots, a_m, c$ be the associated basis with $c$ corresponding to $e$ and set $A = \langle a_1, \ldots, a_m \rangle$. Thus $c^{\pm 1}$ appears in the cyclic word for $x$ $k$ times. If the Whitehead graph of $x$ is disconnected, consider a 1-edge blowup $\tilde{G}$ of $G$ so that $x$ realized in $\tilde{G}$ is contained in a proper subgraph. In this case $d_F(\dot{x}, A) \leq 5$ by Lemma 3.1 (4 for the distance between $A$ and the free factor determined by the image of $x$, and 1 more to get to $\dot{x}$). If the Whitehead graph is connected then it has a cut point [31, 25]. Let $\phi$ be the associated Whitehead automorphism. If the special letter is some $a_i^{\pm 1}$ then the free factor $A$ is $\phi$-invariant. If the special letter is $c^{\pm 1}$ then $d_F(A, \phi(A)) \leq 6$ ($d_F(A, \langle c \rangle) \leq 3$ and $\langle c \rangle$ is fixed by $\phi$). But there are at most $k$ automorphisms of the latter kind in the process of reducing $x$ until its Whitehead graph is disconnected. Thus $d_F(\dot{x}, A) \leq 6k + 5$.

Now assume $e$ is separating. By collapsing a maximal tree on each side of $e$ we may assume that $G$ is the disjoint union of two roses $R_A$ and $R_B$ connected by $e$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ be the bases determined by $R_A$ and $R_B$ respectively. Notice that the assumption about $e$ means that there are $k$ times when the cyclic word for $x$ changes from the $a_i$’s to the $b_j$’s or vice versa ($k$ is necessarily even here). If the Whitehead graph of $x$ with respect to $a_1, \ldots, a_n, b_1, \ldots, b_m$ is disconnected, we see as above that $d_F(\dot{x}, A) \leq 5$ where $A = \langle a_1, \ldots, a_n \rangle$. Otherwise there is a cut point and let $\phi$ be the associated Whitehead automorphism. If the special letter is some $a_i^{\pm 1}$ then $A$ is $\phi$-invariant. Likewise, $\phi(A)$ is conjugate to $A$ if the special

\footnote{an element of some basis for $F$.}
letter is $b_j^{±1}$ and all the $a_i^{±1}$ are on one side of the cut. If they are not on one side of the cut, then the subgraph spanned by the $a_i^{±1}$'s is disconnected and we may consider the associated 1-edge blowup $\tilde{R}_A$ of $R_A$. Let $\tilde{G}$ be the 1-edge blowup of $G$ obtained by attaching $e \cup R_B$ to $\tilde{R}_A$ along either of the two vertices. The blowup edge $e'$ can be crossed by $x$ only if it is immediately followed or preceded by $e$ (but not both). Thus $x$ crosses $e'$ at most $k$ times. If $e'$ is nonseparating then by the first paragraph $d_F(\hat{x}, P) \leq 6k + 5$ for some subgraph $P \subset \tilde{G}$ and so $d_F(\hat{x}, (b_1, \ldots, b_m)) \leq 6k + 9$. If $e'$ is separating replace $A$ by a smaller free factor $A'$ and continue. 

We introduce the following notation. Suppose $G, G'$ are marked graphs, $A$ is a proper free factor of $F$, and $x$ is a simple class in $F$. Then:

- $d_F(G, G') := \sup\{d_F(A, A') \mid A \in \pi(G), A' \in \pi(G')\}$
- $d_F(A, G') := \sup\{d_F(A, A') \mid A' \in \pi(G')\}$
- $d_F(G, x) := d_F(G, \hat{x})$
- $d_F(A, x) := d_F(A, \hat{x})$

For example, Lemmas 3.1 and 3.2 combine to give the following:

**Lemma 3.3.** Let $x$ be a simple class in $F$ and $G$ a marked graph so that $x|G$ crosses some edge $\leq k$ times. Then $d_F(G, x) \leq 6k + 13$.

**Remark 3.4.** Let $z$ be a conjugacy class in $F$ and $G \in \mathcal{X}$ a marked graph. Since $\text{vol}(G) = 1$, there is an edge of $G$ that is crossed at most $[\ell(z|G)]$ times by $z|G$.

**Corollary 3.5.** If $d_X(G, G') \leq \log K$ then $d_F(G, G') \leq 12K + 32$.

**Proof.** Let $z$ be a candidate that realizes $d_X(G, G')$. Thus $\ell(z|G) < 2$, $z|G$ crosses some edge once and $\ell(z|G') < 2K$, so $z|G'$ crosses some edge $< 2K$ times. Therefore

$$d_F(G, G') \leq d_F(G, z) + d_F(z, G') \leq 19 + (12K + 13) = 12K + 32$$

In most of the paper we will be concerned with showing that distances in $F$ are bounded above. We will use the obvious terminology: In Corollary 3.5 we showed that the distance in $F$ between projections of graphs from $\mathcal{X}$ are bounded above as a function of the distance in $\mathcal{X}$.
**Convention 3.6 (Bounded distance).** When we say a distance in $\mathcal{F}$ is bounded without any variables, we mean bounded above by a universal constant that depends only on the rank $n$ of $\mathcal{F}$.

Note that Corollary 3.5 says that $\pi : \mathcal{X} \to \mathcal{F}$ is coarsely Lipschitz: If $d_{\mathcal{X}}(G, G') \leq N$ with $N$ an integer, then $d_{\mathcal{F}}(G, G') \leq CN$ for a universal $C > 0$. Indeed, choose a geodesic from $G$ to $G'$ and apply Corollary 3.5 $N$ times to pairs of points at distance $\leq 1$.

By the injectivity radius $\text{injrad}(G)$ of a metric graph $G$ we mean the length of a shortest embedded loop in $G$. If $A$ is a finitely generated subgroup of $\mathcal{F}$ and $\hat{G} \in \hat{\mathcal{X}}$ we denote by $A|\hat{G}$ the core of the covering space of $\hat{G}$ corresponding to $A$. Thus there is a canonical immersion $A|\hat{G} \to \hat{G}$. We adopt the following convention: Unless otherwise specified, the metric and illegal turn structures on $A|\hat{G}$ are the ones obtained by pulling back via $A|\hat{G} \to \hat{G}$.

**Corollary 3.7.** Suppose $A$ is a proper free factor of $\mathcal{F}$ and $G \in \mathcal{X}$. If $\text{injrad}(A|G) < k + 1$ then $d_{\mathcal{F}}(A, G) \leq 6k + 14$.

**Proof.** This follows immediately from Remark 3.2 and Lemma 3.3.

Coarse paths in $\mathcal{F}$ obtained from folding paths by projecting will play a crucial role.

**Corollary 3.8.** The collection of projections to $\mathcal{F}$ of folding paths in $\mathcal{X}$ is a coarsely transitive family: for any two proper free factors $A, B$ of $\mathcal{F}$ there is a folding path $G_t, t \in [0, L]$, such that $A \in \pi(G_0)$ and $B \in \pi(G_L)$. The same is true for the subcollection consisting of folding paths induced by morphisms that satisfy the rationality condition from the Stallings method of folding in $\mathcal{X}$.

**Proof.** Let $A, B$ be two free factors. Choose $G, \Sigma \in \mathcal{X}$ so that some subgraph of $G$ represents $A$ and some subgraph of $\Sigma$ represents $B$, and so that $G$ is a rose, and apply Proposition 2.7 to obtain a geodesic from $G$ to $\Sigma$ that is the concatenation of two paths, the first a linear path from $G$ to $G'$ in a single simplex, and the second a folding path from $G'$ to $G$. The initial linear path keeps the underlying graph a rose and its coarse projection is constant. Thus, the desired folding path is the second path from $G'$ to $\Sigma$.

To achieve rationality, choose $\Sigma$ to be a rose with rational edge lengths. Let $\phi : G \to \Sigma$ be an optimal map after adjusting the metric so that $\Delta(\phi) = G$. If the vertex of $G$ maps to the vertex of $\Sigma$, rationality is automatic. Otherwise the vertex of $G$ maps to a point in the interior of some edge.
Perturb $\phi$ so that this point is rational, and adjust the edge lengths in $G$ so that the perturbed map is optimal, with the same train track structure. The new map satisfies rationality. \hfill \Box

4 More on folding paths

We now discuss folding in more detail. Let $\hat{G}_t$, $t \in [0, \omega]$, be a folding path in $\hat{X}$ (from now on we replace trees by quotient graphs) with the natural parametrization. (See Notation 2.6 to recall our conventions.) So for $s < t$ we have maps $\phi_{st} : \hat{G}_s \to \hat{G}_t$ that have slope 1 on each edge, immerse each edge, and induce train track structures.

Unfolding. Traversing a folding path in reverse is unfolding. The main result of this subsection is Theorem 4.2 giving local and global pictures of both folding and unfolding. This result will only be used for the proofs of technical Propositions 5.4 (surviving illegal turns) and 6.10 (legal and illegal) which should be skipped in a first reading. In fact, Theorem 4.2 is obvious in the case of Stallings’ rational paths (see 2.2A) and so could be avoided altogether (see Corollary 3.8).

**Definition 4.1.** An (abstract) widget $W$ of radius $\epsilon$ is a metric graph that is a cone on finitely many, but at least 2, points all the same distance $\epsilon$ from the cone point. A widget has a canonical morphism to $W \to [0, \epsilon]$ that sends the cone point to $\epsilon$ and the other vertices to 0. There is also a canonical path, parametrized by $[0, \epsilon]$, of finite trees from $W$ to $[0, \epsilon]$ that locally folds all legal turns of the morphism with speed 1. See Figure 1. An (abstract) gadget of radius $\epsilon$ is a union of finitely many widgets of radius $\epsilon$. The union is required to be disjoint except that widgets are allowed to meet in vertices and is also required to be a forest. There is a canonical path, parametrized by $[0, \epsilon]$, of forests obtained by folding each widget.

**Figure 1:** An abstract widget of radius $\epsilon$.

Let $\hat{G} \in \hat{X}$. A widget (resp. gadget) of radius $\epsilon$ in $\hat{G}$, is an embedding
in \( \hat{G} \) of an abstract widget (resp. abstract gadget) of radius \( \epsilon \). There is a natural path in \( \hat{X} \) starting with \( \hat{G} \) and parametrized by \([0, \epsilon]\) given by folding the gadget. See Figure 2. An (abstract) widget, resp. gadget is a widget, resp. gadget, of some radius.

**Theorem 4.2.** Let \( \hat{G}_t, t \in [0, \omega] \) be a folding path in \( \hat{X} \) with its natural parametrization. There is a partition \( 0 = t_0 < t_1 < \cdots < t_N = \omega \) of \([0, \omega]\) such that the restriction of \( \hat{G}_t \) to each \([t_i, t_{i+1}]\) is given by folding a gadget in \( \hat{G}_{t_i} \).

In the context of Theorem 4.2 we say that, as we traverse \([t_i, t_{i+1}]\) in reverse, we are unfolding a gadget. The analogous result holds for the induced path \( G_t, t \in [0, L] \), in \( X \) parametrized by arclength and we use the same terminology. For example, we say \([0, L]\) has a finite partition into subintervals such the restriction of \( G_t \) to each subinterval is given by folding (unfolding) a gadget.

**Proposition 4.3** is an immediate consequence of Theorem 4.2.

**Proposition 4.3.** Let \( G_t, t \in [0, L] \), be a folding path in \( X \). There is a partition of \([0, L]\) into finitely many subintervals so that the restriction of \( G_t \) to each subinterval is a (reparametrized) linear path in a simplex of \( X \).

In particular, a folding path in \( X \) changes an open simplex only at discrete times. Outside these times, illegal turns all belong to vertices with 2 gates, and one gate is a single direction.

The restriction of a folding path to a simplex of \( X \) need not be linear. This can happen, for example, if an illegal turn becomes legal.

The rest of this subsection is devoted to proving Theorem 4.2. It is clear from our description of folding in 2.2.C that, for each \( t_* \in [0, \omega] \), there is \( \epsilon > 0 \) such that the restriction of \( \hat{G}_t \) to \([t_*, t_* + \epsilon]\) is given by folding a gadget in \( \hat{G}_{t_*} \). To complete the proof of Theorem 4.2 we will show that, for each \( t_* \in (0, \omega] \), there is \( \epsilon > 0 \) such that the restriction of \( G_t \) to \([t_* - \epsilon, t_*]\) is given by folding a gadget in \( \hat{G}_{t_* - \epsilon} \).

Let \( N \) be the closed \( \epsilon \)-neighborhood of a vertex \( v \) in \( \hat{G}_{t_*} \) of valence \( \geq 3 \). We will describe the preimage \( N_\epsilon \) of \( N \) in \( \hat{G}_{t_* - \epsilon} \) for small enough \( \epsilon > 0 \). As long as \( \epsilon \) is small enough and \( N_\epsilon \rightarrow N \) is not injective, we will find a connected gadget of radius \( \epsilon \) in \( N_\epsilon \) so that \( N_\epsilon \rightarrow N \) folds this gadget. The gadget needed to complete the proof will be the disjoint union of these connected gadgets of radius \( \epsilon \), one for each such vertex of \( G_{t_*} \). We will also equip \( N \) and \( N_\epsilon \) with height functions. For convenience, set \( \phi_\epsilon := \phi_{t_* - \epsilon, t_*} \).

First we describe the height functions. Assume that \( \epsilon \) is small enough so that \( N \) is a cone on a finite set with cone point \( v \). The height function
on $N$ is the morphism $N \rightarrow [-\epsilon, \epsilon]$ given as follows. If the length in $N$ of $[v, w]$ is $\epsilon$ and the direction at $v$ determined by $[v, w]$ has more than one preimage in $N_\epsilon$, then map $[v, w]$ isometrically to $[0, \epsilon]$; otherwise map $[v, w]$ isometrically to $[0, -\epsilon]$. The height function $h$ on $N_\epsilon$ is the composition $N_\epsilon \rightarrow N \rightarrow [-\epsilon, \epsilon]$. Now we describe $N_\epsilon$ for small enough $\epsilon$, and justify this description immediately after that. In $N_\epsilon$, the preimage of $v$ is the set of points of height $0$. The set $h^{-1}([0, \epsilon])$ is a gadget with widgets the closures of the components of $h^{-1}((0, \epsilon))$. Each of the height $\epsilon$ vertices has a unique direction in $\hat{G}_{t*, -\epsilon}$ not in the gadget, and we draw this direction upwards. Height $0$ vertices may have additional directions not contained in the gadget, and we draw those downwards. See Figure 2. All illegal turns in $N_\epsilon$ appear at vertices of height $\epsilon$ and these vertices have two gates in $\hat{G}_{t*, -\epsilon}$ (all downward directions form one gate and the single upward direction is the other gate). In particular, all turns at the height $0$ vertices are legal. After the widgets are folded, in $\hat{G}_{t*}$, the height $0$ vertices get identified to $v$, each widget contributes an upward direction at $v$, and the downward directions at $v$ come from downward directions in $N_\epsilon$ based at height $0$ vertices. Some pairs of directions may be illegal in $\hat{G}_{t*}$, but in that case they have to come from directions in $N_\epsilon$ that don’t form a turn (i.e. that are based at different vertices).

Now we justify our description of $N_\epsilon$. First choose $\epsilon_0$ small enough so that:

1. the closed $\epsilon_0$-neighborhood of $v$ is a cone on a finite set with cone point $v$ and

2. the cardinality of the preimage of $v$ in $N_\epsilon$ and the number of directions

\[ \begin{array}{c}
\text{Figure 2: An example of } N_\epsilon \text{ with 3 widgets, 3 vertices at height } \epsilon \text{ and 5 vertices at height 0. The union of the 3 widgets is a gadget.}
\end{array} \]
based at points in the preimage of \( v \) is independent of \( 0 < \epsilon \leq \epsilon_0 \);

Note that it is possible to choose \( \epsilon_0 \) satisfying (2) because, since all \( \hat{\phi}_{st} \) are surjective, the cardinality of \( \hat{\phi}_{\epsilon}^{-1}(v) \) is non-increasing as \( \epsilon \) decreases, and similarly the number of directions based at \( \hat{\phi}_{\epsilon}^{-1}(v) \) is non-increasing.

Choose \( 0 < \epsilon_1 < \epsilon_0 \) so that the closed \( \epsilon_1 \)-neighborhood \( N' \) in \( N_{\epsilon_0} \) of the preimage of \( v \) contains no valence \( \geq 3 \) vertices of \( \hat{G}_{t_0-\epsilon_0} \) other than the preimages of \( v \). We claim that, for \( 0 < \epsilon \leq \epsilon_1 \), \( N' \) has the description given above. It is enough to consider the case \( \epsilon = \epsilon_1 \) because our description of \( N' \) is stable under decreasing \( \epsilon \).

By definition, \( N' \) is the preimage in \( \hat{G}_{t_0-\epsilon_1} \) of \( N \). It is equal to the closed \( \epsilon_1 \)-neighborhood of the preimage of \( v \). (Indeed, it is clear that the neighborhood is in \( N' \). If \( w \in N' \) does not map to \( v \), then since \( \hat{G}_{t_0-\epsilon_1} \) has a train track structure there is a direction at \( w \) whose image in \( N \) points toward \( v \) and there is a legal path of length \( |h(w)| \) starting at this direction. The endpoint of the path then maps to \( v \).) Similarly, \( N' \) is the preimage in \( \hat{G}_{t_0-\epsilon_0} \) of \( N \). In particular, \( N' \to N_{\epsilon_1} \) is surjective.

\( N' \) is a disjoint union of cones on points of height \( \pm \epsilon_1 \) with cone points the preimages of \( v \). Notice that all turns in \( N' \) are legal (or else (2) fails). \( N' \to N_{\epsilon_1} \) is an embedding off the points of height \( \pm \epsilon_1 \) in \( N' \) (otherwise there would be a path in \( N_{\epsilon_1} \) between distinct directions at preimages of \( v \) whose image has outradius with respect to \( v \) that is less than \( \epsilon_1 \), by Lemma 2.3 these distinct directions would then be identified before time \( t_0 \), and we contradict (2)). In fact, \( N' \to N_{\epsilon_1} \) is an embedding off points of height \( \epsilon \) (otherwise there would be a path \( \sigma \) in \( N_{\epsilon_1} \) that is cone on a pair of distinct points in the preimage of \( v \) with cone point of height \( -\epsilon_1 \), contradicting the definition of \( h \) since the directions at the ends of \( \sigma \) are identified in \( \hat{G}_{t_0} \)).

We see that \( N_{\epsilon_1} \) is the obtained from \( N' \) by identifying some pairs of points of height \( \epsilon \). The picture is completed with a few observations.

- \textit{\( N_{\epsilon_1} \) is connected:} By Lemma 2.3, any two points in the the preimage of \( v \) in \( N_{\epsilon_1} \) are connected by a path in \( N_{\epsilon_1} \) (or else they aren’t identified in \( \hat{G}_{t_0} \)). So, the preimage of \( v \) is contained in a single component and from the picture we’ve developed so far we see that \( N_{\epsilon_1} \) is connected.

- \textit{\( N_{\epsilon_1} \) is a tree:} A loop \( \sigma \) in \( N_{\epsilon_1} \) has homotopically trivial image in \( N \). Since \( \hat{\phi}_{\epsilon_1} \) is a homotopy equivalence, \( \sigma \) is homotopically trivial.

- \textit{Every point \( w \) in \( N_{\epsilon_1} \) of height \( \epsilon_1 \) has at least two downward directions:} Otherwise, there is a point \( v' \) in the preimage of \( v \) and a path \([v', w] \) of increasing height with the property that the induced direction at \( v' \)
is not identified by \( \dot{\phi}_{\epsilon_1} \) with any other direction. This contradicts the definition of positive height.

This completes the proof of Theorem 4.2.

Fix \( \epsilon \leq \epsilon_1 \). We introduce a little more terminology for later use. By construction, every point in \( \hat{G}_{t*} \) at distance \( \epsilon \) from \( v \) has a unique direction pointing away from \( v \). If \( d \) is a direction at \( v \), there is then a unique corresponding direction \( d' \) pointing away from \( v \) and based at a point at distance \( \epsilon \) from \( v \) determined by “\( d \) points to \( d' \)”. If the height of \( d' \) is \( \epsilon \) then we denote \( d' \) by \( d^{\epsilon} \); if the height of \( d' \) is \( -\epsilon \) then we denote \( d' \) by \( d^{-\epsilon} \). We say that \( d^{\epsilon} \) points up and \( d^{-\epsilon} \) points down. The direction \( d^{\pm \epsilon} \) has a unique lift \( \tilde{d}^{\pm \epsilon} \) to \( \hat{G}_{t*} \). We say \( \tilde{d}^{\epsilon} \) points up and \( \tilde{d}^{-\epsilon} \) points down. See Figure 6.

Unfolding a path. Given an immersed path \( \gamma \) in \( \hat{G}_\omega \), one may try to “lift” it along the folding path, i.e. to find immersed paths \( \gamma_t \) in \( \hat{G}_t \) that map to \( \gamma \) (up to homotopy rel endpoints). This is always possible, since it is clearly possible locally. At discrete times new illegal turns may appear inside the path. Note that at discrete times the lifts are not unique, when an endpoint of the path coincides with the vertex of an illegal turn. Figure 6 illustrates the nonuniqueness of lifts.

![Diagram of an illegal turn](image)

Figure 3: The figure illustrates the ambiguity in lifting paths under unfolding which we see is parametrized by the preimage of \( v \).

To get uniqueness, we can remove the end of the path that lifts nonuniquely. Thus we may have to remove segments\(^3\) at the ends whose size grows at speed 1. Now suppose there are illegal turns in the path. As we unfold, each illegal turn makes the length of the path grow with speed 2, and the illegal

\(^3\)We use “segment” synonymously with “immersed path”, but usually to connote a smaller piece of something.
turn closest to an end moves away from the end at speed 1. We deduce that lifting is unique between the first and last illegal turns along the path $\gamma$, including the germs of directions beyond these turns. We call this the **unfolding principle**.

In particular, this applies to illegal turns themselves: if a loop $z|\hat{G}_\omega$ has two occurrences of the same illegal turn, pulling back these turns produces two occurrences of the same path (most of the time a neighborhood of a single illegal turn, but see Figure 4). But note that distinct illegal turns

![Figure 4: A path unfolds to a path with two illegal turns.](image)

might pull back to the same illegal turn, see Figure 5.

![Figure 5: In this folding path the number of illegal turns grows from 1 to 2.](image)

**Illegality.** If a graph $G$ is equipped with a track structure, by the *illegality of $G$* we mean

$$m(G) = \sum_v \sum_{\Omega_v} (|\Omega_v| - 1)$$
where the sum is over all vertices $v$ of $G$ and all gates $\Omega_v$ at $v$. Thus a gate that contains $k \geq 1$ directions contributes $k - 1$ to the count. The right derivative of the function

$$t \mapsto \text{vol}(\hat{G}_t)$$

at $t = t_0$ is $-m(\hat{G}_0)$, the negative of the illegality of $\hat{G}_t$. If $G$ is marked and $z$ is a conjugacy class in $\mathcal{F}$ then $k(z|G)$ is by definition the number of illegal turns in $z|G$, i.e. the number of illegal turns in the illegal turn structure on $z|G$ induced by $z|G \to G$.

For the rest of this section, $G_t$, $t \in [0, L]$, is a folding path in $\mathcal{X}$ parametrized by arclength. We will sometimes abbreviate $m(G_t)$ by simply $m_t$, $z|G_t$ by $z_t$ and $k(z_t)$ by $k_t$.

**Lemma 4.4.** (1) Let $z$ be a conjugacy class in $\mathcal{F}$. The the right derivative of the length function $t \mapsto \ell(z_t)$ at $t = 0$ is

$$\frac{d}{dt}\ell(z_t)|_{t=0^+} = \ell(z_0) - 2\frac{k_0}{m_0}$$

(2) Let $\sigma_0$ be a nondegenerate immersed path in $G_0$ whose initial and terminal directions are in illegal turns of $G_0$. Then, for small $t$, there is a corresponding path $\sigma_t$ whose initial and terminal directions are in illegal turns of $G_t$. The right derivative at 0 of the length $L_t$ of $\sigma_t$ is $L_0 - 2\frac{k_0}{m_0}$, where now $k_0 - 1$ is the number of illegal turns in the interior of $\sigma_0$. If the initial and terminal directions of $\sigma_0$ are in the same gate, then the same is true for $\sigma_t$.

**Proof.** (1): Let $\hat{G}_t$ be the naturally parametrized by path in $\hat{X}$ so that $G_{s(t)} = \hat{G}_t = \hat{G}_t/\text{vol}(\hat{G}_t)$ and $\text{vol}(\hat{G}_0) = 1$ (see Notation 2.6). For small $t \geq 0$, $\text{vol}(\hat{G}_t) = 1 - m_0t$, $\ell(z|\hat{G}_t) = 1 - 2k_0t$, and $s(t) = -\log(\text{vol}(\hat{G}_t))$. The proof of the derivative formula in (2) is identical.

We will sometimes abuse notation and write $\ell'(z_{t_0})$ for $\left.\frac{d}{dt}\ell(z|G_t)\right|_{t=t_0^+}$.

**Corollary 4.5.** Let $z$ and $w$ be conjugacy classes in $\mathcal{F}$ and suppose $m_t$ and $k_t = k(z_t)$ are constant for $t \in [0, \epsilon)$.

(1) The length $\ell(z_t) = ae^t + b$ on $[0, \epsilon)$ where $a = \ell(z_0) - 2\frac{k_0}{m_0}$ and $b = 2\frac{k_0}{m_0}$.

(2) If $\ell'(z_0) \geq \ell'(w_0)$ then $\ell(z_t) - \ell(w_t)$ is nondecreasing on $[0, \epsilon)$.

(3) If the average length $A$ of a maximal legal segment in the loop $z_0$ is $> 2/m_0$ the loop grows in length on $[0, \epsilon)$, and if it is $< 2/m_0$ it shrinks.
Proof. (1) and (2) follow easily from Lemma 4.4. For (3), \( k_0 A = \ell(z_0) \) and so, by (1), \( a = \ell(z_0)(1 - 2/m_0 A) \).

**Corollary 4.6.** Let \( z \) be a conjugacy class in \( F \). The length \( \ell(z_t) \) is piecewise exponential on \( [0, L] \), i.e. \( [0, L] \) has a finite partition such that the restriction of \( \ell(z_t) \) to each subinterval is as in Corollary 4.5(1).

**Proof.** Since we may subdivide \( [0, L] \) as \( 0 = s_0 < s_1 < \cdots < s_k = L \) so that, on each \( [s_i, s_{i+1}) \), \( m_t \) and \( k_t \) are constant, the corollary follows directly from Corollary 4.5(1) and Proposition 4.3.

We will say that a segment has endpoints illegal turns if the path has been infinitesimally extended at each end, i.e. directions have been specified at the initial and terminal endpoints, and these directions together with the segment determine illegal turns. If the segment is degenerate then this means that the added directions form an illegal turn.

**Corollary 4.7.** Let \( \sigma_L \) be an immersed path in \( G_L \) with endpoints illegal turns and, for \( t \in [0, L] \), let \( \sigma_t \) be the immersed path in \( G_t \) with endpoints illegal turns obtained by applying the Unfolding Principle to \( \sigma_L \). Corollaries 4.5 and 4.6 hold if \( \ell(z_t) \) is replaced by the length of \( \sigma_t \).

**Proof.** Since the derivative formula in Lemma 4.4(2) is obtained from that in Lemma 4.4(1) by such a replacement, proofs of corresponding statements are identical.

**Corollary 4.8.** A legal segment \( \sigma_0 \) of length \( L_0 \geq 2 \) inside \( z_0 \) gives rise to a legal segment \( \sigma_t \) of length \( L_t \geq 2 + (L_0 - 2)e^t \) inside \( z_t \) for \( t \in [0, L] \). In particular, a legal segment of length \( L_0 \geq 3 \) grows exponentially.

**Proof.** \((L_t - 2)' \geq L_t - 2 \) using Lemma 4.4(2).

Similar considerations control the lengths of topological edges \( e \) of \( G_t \) that are not involved in any illegal turn, i.e. the directions determined by the ends of \( e \) aren’t in any illegal turns.

**Lemma 4.9.** Suppose that \( e_0 \) is an edge of \( G_0 \) that is not involved in an illegal turn. Then for small \( t \), there is a corresponding edge \( e_t \) of \( G_t \) not involved in an illegal turn. Further, the length \( L_t \) of \( e_t \) satisfies \( L_t = L_0 e^t \).

**Proof.** For small \( t \), the morphism \( e^t G_0 \to G_t \) restricted to \( e_0 \) is an isometric embedding with image an edge of \( G_t \).
Lemma 4.10. Suppose $z$ is a conjugacy class such that $k(z_0)$ and $\ell(z_L)$ are bounded. Then $\ell(z_t)$ is bounded for all $t \in [0, L]$.

Proof. By Corollary 4.5(3) and Corollary 4.6, if $\ell(z_t)/k(z_t) \geq \ell(z_t)/k(z_0) \geq 2 \geq 2/m_t$, then $\ell(z_t)$ grows. Therefore, $\ell(z_t) \leq \max\{2k(z_0), \ell(z_L)\}$ for all $t \in [0, L]$.

By a surface relation we mean a conjugacy class that, with respect to some rose, crosses every edge twice and has a circle as its Whitehead graph (equivalently, attaching a 2-cell to the rose via the curve results in a surface).

Lemma 4.11. Suppose $m_t \geq m$, for all $t \in [0, L]$, and let $w$ be a conjugacy class in $\mathbb{F}$. Assume $k(w_0) = m$. If $\ell(w_L) \leq K$ then either

(i) there is a simple class $u$ such that:

- $k(u_0)$ is bounded; and
- $\ell(u_t)$ is bounded by a function of $K$ for all $t \in [0, L]$

(in particular $d_F(G_0, G_L)$ is bounded by a function of $K$); or

(ii) $w$ is a surface relation.

Moreover, if $m_t > m$ for all $t$ then (i) holds.

Proof. Arguing as in Lemma 4.10 we have $\ell(w_t) \leq \max\{2, K\}$ (loops of length $> 2$ with no more illegal turns than the illegality grow under folding). So we may take $u = w$ provided $w$ is simple. See Lemma 3.3 and Remark 3.4 for the parenthetical remark. We now consider four cases.

Case 1. $\ell(w_0) < 2$. Then $w$ is simple as $w_0$ crosses some edge at most once.

Case 2. $\ell(w_0) = 2$. Then either $w$ is simple or $w_0$ crosses every edge exactly twice. In the latter case, collapse a maximal tree in $G_0$ – with respect to the resulting rose the Whitehead graph of $w$ is either a circle (and then $w$ is a surface relation) or the disjoint union of at least two circles (and then $w$ is simple).

Case 3. $2 < \ell(w_0) < 2 + \text{injrad}(G_0)$. Then either $w$ is simple or $w_0$ crosses every edge at least twice. Assume the latter. Under our assumption the edges crossed more than twice form a forest. Collapse a maximal tree that contains this forest and argue as in Case 2.

Case 4. $\ell(w_0) \geq 2 + \text{injrad}(G_0)$. Choose a conjugacy class $v$ with $\ell(v_0) = \text{injrad}(G_0)$. We now claim that $\ell(v_t) \leq \ell(w_t) - 2$ for all $t$. This is
clearly true at $t = 0$. In fact, this condition persists in that there is no last time $t_0 < L$ where it is true. Indeed, Lemma 4.4(1) shows that

$$\ell'(v_{t_0}) \leq \ell(v_{t_0}) \leq \ell(w_{t_0}) - 2 \leq \ell'(w_{t_0})$$

and so the inequality continues to hold for $t > t_0$ (see Corollary 4.5(2)). Thus $v$ is a simple class with both $\ell(v_t)$ bounded. We may take $u = v$.

For the moreover part, we have $\ell'(w_t) \geq \ell(w_t) - 2\frac{m}{m+1}$ for all $t$. Thus if $\ell(w_0) < 2$ then $w$ is simple and the statement follows with $u = w$. If $\ell(w_0) \geq 2$, then we claim $d_X(G_0, G_L)$ is bounded (see Corollary 3.5). Indeed, by Corollary 4.5 we have $\ell(w_t) \geq ae^t + b$ where $a = \ell(w_0) - 2\frac{m}{m+1} \geq \frac{2}{m+1}$ and $b \geq 2$. In particular, $K \geq \frac{2}{m+1}e^L + 2$. We may take $u = v$ such that $\ell(v_0) = injrad(G_0)$.

We also have the following variant.

**Lemma 4.12.** Suppose in addition to the hypotheses of Lemma 4.11 that, for some illegal turn in $w_t G_0$, one of the two edges $e$ forming the turn is nonseparating and has length a definite fraction $p > 0$ of $injrad(G_0)$. Then either:

1. $w$ is a simple class with both $\ell(w_t)$ bounded.
2. $w$ is a surface relation and any class $z$ such that $z_0$ contains a segment $S = e \cdots e$ that closes up (by identifying the two copies of $e$) to $w_0$ fails to be simple.

**Proof.** Referring to the proof of Lemma 4.11 in Cases 1 and 4, (i') holds; so assume we are in Cases 2 or 3. In fact we are free to assume $\ell(w_0) < 2 + p injrad(G_0)$, for otherwise the argument of Case 4 shows that for $v$ with $\ell(v_0) = injrad(G_0)$ we have $\ell(v_t) \leq \frac{\ell(w_t) - 2}{p}$ for any $t$ and (i') follows (and this time the bound also depends on $p$). Now the forest consisting of the edges crossed by $w_0$ more than twice (assuming all edges are crossed at least twice) does not include $e$, and we may collapse a maximal tree that contains this forest but does not contain $e$. Now $z_0$ can be thought of as $e \cdots e \cdots = e Ae B$ with the subpath $S = e Ae$ giving $w$. Since $e$ is not collapsed, the Whitehead graph of $z$ in the rose contains the Whitehead graph of $w$, which is a 1-manifold. So if $w$ is not simple, neither is $z$. \qed
5 Loops with long illegal segments

The key result of this section is Proposition 5.4 (surviving illegal turns) which is a generalization of [3, Lemma 2.10]. Before stating Proposition 5.4, we need a bit of terminology and some preliminary lemmas to be used in the proof. In this section, $G_t$, $t \in [0, L]$, is a folding path in $X$ parametrized by arclength. Consider a conjugacy class $z$ in $F$ and the induced path of loops $z_t := z|G_t$. The illegal turns along $z_t$ are folding as $t$ increases, but at discrete times an illegal turn may become legal, or several illegal turns may collide and become one (e.g. see Figure 4). We say that a consecutive collection of illegal turns along $z$ survives to $G_L$ if none of them become legal nor do they collide with a neighboring illegal turn in the collection, for any $t \in [0, L]$. In particular, each illegal turn in the collection in $z_t$ unfolds to a single illegal turn in the collection in $z_t'$ for $t' \leq t$. We will call the portion of $z_t$ spanned by our collection the good portion of $z_t$. The turns in the collection in $z_t$ are, in order, $\tau_{t,1}, \tau_{t,2}, \ldots$ and have vertices $p_{t,1}, p_{t,2}, \ldots$. In particular, if $t' < t$ then $\tau_{t,i}$ unfolds to $\tau_{t',i}$. The image of $\tau_{t,i}$ in $G_t$ is $\tilde{\tau}_{t,i}$. In this context, the Unfolding Principle gives the implication $\tilde{\tau}_{t,i} = \tilde{\tau}_{t,j} = \tilde{\tau}_{t',i}$.

Suppose that a consecutive collection of illegal turns along $z$ survives to $G_L$. For each $t \in [0, L]$ denote by $T_t$ the set of turns that occur in the given consecutive collection, i.e. $T_t = \{\tilde{\tau}_{t,1}, \tilde{\tau}_{t,2}, \ldots\}$. Let $D_t$ denote the set of directions in $G_t$ that occur in a turn in $T_t$. Of course, $D_t$ is partitioned into equivalence classes with respect to the relation “being in the same gate”, but we consider a finer equivalence relation generated by $d \sim d'$ if $\{d, d'\} \in T_t$. We will call the equivalence classes subgates. Each subgate, say at the vertex $v$, gives rise to a “Whitehead graph”: the vertices of the graph are the directions in the subgate, and an edge is drawn between $d$ and $d'$ if $\{d, d'\} \in T_t$.

Lemma 5.1. In the situation above, let $d_1, d_2, \ldots, d_k$ be the vertices along an embedded closed curve in the Whitehead graph of a subgate (that is, $d_i \neq d_j$ for $i \neq j$) at $G_t$. Then for any $t' < t$ there is an induced embedded closed curve in the Whitehead graph of a subgate of $G_{t'}$.

Specifically, each turn $\{d_i, d_{i+1}\}$ (taken mod $k$, so including $\{d_k, d_1\}$) unfolds to a turn in $T_{t'}$; Lemma 5.1 says that these turns also form an embedded closed curve in a subgate (in particular, all are based at the same vertex).
Proof of Lemma 5.1 Let $\hat{G}_{s(t)}$ be the path in $\hat{X}$ giving rise to $G_t$ (see Notation 2.6). As combinatorial graphs, we identify $G_t$ and $\hat{G}_{s(t)}$ (they differ only by homothety). In particular, a direction in one is naturally identified with a direction in the other and we will use the same names for two such directions. If we set $s := s(t)$ and $s' := s(t')$, it suffices to argue in the case $s' = s - \epsilon$ for small $\epsilon > 0$; for the conclusion clearly holds in the limit. To that end, we use the description of $N_\epsilon$ developed in Section 4. So, let $v$ be the vertex in $\hat{G}_s$ that is the base of the directions $d_i$ in our subgate, $N$ be the $\epsilon$-neighborhood in $\hat{G}_s$ of $v$, and $N_\epsilon$ be the preimage of $N$ in $\hat{G}_{s'}$. Set $d_i^\epsilon = d_i^\epsilon$ or $d_i^{-\epsilon}$ depending on whether or not $d_i$ points up or down and similarly for $\tilde{d}_i^\epsilon$. For $i \neq j$, let $[d_i^\epsilon, d_j^\epsilon]$ denote the path in $\hat{G}_s$ from the base of $d_i^\epsilon$ to the base of $d_j^\epsilon$ extended infinitesimally by the outgoing (germs of) directions $d_i^\epsilon$ and $d_j^\epsilon$. Similarly, $[\tilde{d}_i^\epsilon, \tilde{d}_j^\epsilon]$ denotes the unique immersed path lifting $[d_i^\epsilon, d_j^\epsilon]$. Note that if $z|\hat{G}_s$ contains the illegal turn $\{d_i, d_j\}$ then it also contains the (infinitesimally extended) path $[d_i^\epsilon, d_j^\epsilon]$ and $z|\hat{G}_{s'}$ contains $[d_i^\epsilon, d_j^\epsilon]$. If $d_i^\epsilon$ points down then it is supported by a widget $W$ if $d_i^\epsilon$ is in the downward direction from some $x \in W$ of height 0. In this case, we also say that $d_i^\epsilon$ is supported by $x$. If $\tilde{d}_i^\epsilon$ points up, then it is supported by the unique widget at which it is based. See Figure 6.

![Figure 6](image_url)

Figure 6: An upward and a downward direction supported by the leftmost widget.

By hypothesis, $[\tilde{d}_i^\epsilon, \tilde{d}_{i+1}^\epsilon]$ contains a unique illegal turn that we abusively denote $\{\tilde{d}_i^\epsilon, \tilde{d}_{i+1}^\epsilon\}$. Recall that all illegal turns in $G_{s'}$ have height $\epsilon$ and are in a widget. Note that if $[\tilde{d}_i^\epsilon, \tilde{d}_{i+1}^\epsilon]$ (or indeed any path between height $\pm \epsilon$ vertices) contains its unique illegal turn in $W$ then it falls into one of the following three cases:
• \( \tilde{d}_i^* \) and \( \tilde{d}_{i+1}^* \) point up in distinct widgets adjacent to \( W \)
• \( \tilde{d}_i^* \) and \( \tilde{d}_{i+1}^* \) point down and are supported by \( W \)
• one of \( \tilde{d}_i^* \) or \( \tilde{d}_{i+1}^* \) points down and is supported by \( W \) and the other points up in a widget adjacent to \( W \)

We now make the following observations.

(1) If \( \tilde{d}_i^* \) points up and is supported by the widget \( W \), then all directions pointing upward at height 0 in \( W \) are mapped to \( d_i^* \). If \( \tilde{d}_i^* \) points down, then it is in a unique downward direction from a height 0 point and this direction maps to \( d_i^* \).

(2) The directions \( \tilde{d}_i^* \) are all distinct. Indeed, by hypothesis the \( d_i^* \) (hence the \( d_i^* \)) are distinct, and \( \tilde{d}_i^* \) is the unique lift of \( d_i^* \).

(3) It is not possible for a single widget to support both an upward \( \tilde{d}_i^* \) and a downward \( \tilde{d}_j^* \). Indeed, this would force \( \{d_i, d_j\} \) to be illegal and then \( [\tilde{d}_i^*, \tilde{d}_j^*] \) would have a height 0 illegal turn. However, all illegal turns occur at height \( \epsilon \).

(4) It is not possible for adjacent widgets to both support upward \( \tilde{d}_i^* \) and \( \tilde{d}_j^* \). This is because this would force an illegal turn at the common height 0 vertex formed by directions that map to \( d_i \) and \( d_j \).

(5) It is not possible for downward \( \tilde{d}_i^* \) and \( \tilde{d}_j^* \) to be supported by the same height 0 vertex. Indeed, this would then force an illegal turn based at this height 0 vertex.

Recall that we want to prove \( \{\tilde{d}_1^*, \tilde{d}_2^*, \tilde{d}_3^*, \ldots, \tilde{d}_k^*, \tilde{d}_1^*\} \) gives rise to an embedded closed curve in the Whitehead graph of a subgate at \( G_s' \). To do this we prove two things.

• (there is a loop) The turns \( \{\tilde{d}_i^*, \tilde{d}_{i+1}^*\} \) are all based at the same vertex \( w \), i.e. the base of the illegal turn crossed by \( [\tilde{d}_i^*, \tilde{d}_{i+1}^*] \) is independent of \( i \).

• (the loop is embedded) For \( i \neq j \), \( \tilde{d}_i^* \) and \( \tilde{d}_j^* \) are not in the same direction from \( w \).

To see there is a loop, suppose we have three consecutive directions \( \tilde{d}_{i-1}^*, \tilde{d}_i^*, \tilde{d}_{i+1}^* \) that determine two illegal turns in \( N_\epsilon \) not based at the same
vertex. There are two cases. First suppose $\tilde{d}_i^*$ points down and is supported by some height 0 vertex $x$. Paths from $\tilde{d}_i^*$ to $\tilde{d}_{i\pm 1}^*$ lead through two distinct widgets each containing $x$. Since our directions $\tilde{d}_1^*, \ldots, \tilde{d}_k^*$ are cyclically ordered and $N_\epsilon$ is a tree, there must be some $j$ with $j \neq i \neq j + 1$ so that $[\tilde{d}_j^*, \tilde{d}_{j+1}^*]$ passes through $x$. (Indeed, otherwise all $\tilde{d}_j^*$, for $j \neq i$, lie in the same component of $N_\epsilon \setminus \{w\}$. Since $\tilde{d}_{i-1}^*$ and $\tilde{d}_{i+1}^*$ lie in distinct components, this is a contradiction.) The path $[\tilde{d}_j^*, \tilde{d}_{j+1}^*]$ cannot terminate at $\tilde{d}_i^*$ by (2), and by (5) it cannot terminate in any downward direction supported by $x$, but must continue to another (adjacent) widget. By (3) the widgets containing $x$ do not support upward $\tilde{d}_j^*$ and $\tilde{d}_{j+1}^*$, so the path crosses two illegal turns, contradiction. The other case is that $\tilde{d}_i^*$ is upward, say based at a height $\epsilon$ vertex $x$ inside a widget $W$. Then there are distinct widgets $W_+$ and $W_-$ adjacent to $W$ so that $[\tilde{d}_j^*, \tilde{d}_{j+1}^*]$ crosses an illegal turn in $W_\pm$. Again there must be some $j$ so that $[\tilde{d}_j^*, \tilde{d}_{j+1}^*]$ either crosses $x$ (if $W_+ \cap W_- = \emptyset$) or crosses the intersection point $W_+ \cap W_-$ (if there is one, and then this point is in $W$ as well). In the latter case the path does not terminate at any direction supported by this point by (3) nor at an upward direction supported by $W_\pm$ by (4). Thus this path has two illegal turns, contradiction. In the former case, $[\tilde{d}_j^*, \tilde{d}_{j+1}^*]$ must have at least 3 illegal turns: one at $x$ and one on each side of $x$, by (3) and (4).

We have established the first item. To see that the loop is embedded, suppose that there are $i \neq j$ such that $\tilde{d}_i^*$ and $\tilde{d}_j^*$ are in the same direction from $w$. Each of these directions is contained in a path between height $\pm \epsilon$ vertices with one illegal turn, that illegal turn being based at $w$. Considering the three cases listed above for such paths, $\tilde{d}_i^*$ and $\tilde{d}_j^*$ have to either be both downward and supported by the same height 0 vertex of the widget $W$ containing $w$, contradicting (5), or one is downward and the other upward supported in the same widget adjacent to $W$, contradicting (3), or they are upward and supported by widgets adjacent to $W$ and to each other, contradicting (4).

Given a finite, simple connected graph $G$, we may consider the components $C_i$ of $G$ cut open along its set $CV$ of cut vertices, i.e. $C_i$ is the closure in $G$ of a component of $G \setminus CV$. Let $\sim$ be the equivalence relation on the edges of $G$ generated by $e \sim e'$ if there is an embedded loop containing $e$ and $e'$.

**Lemma 5.2.** Suppose $G$ is a finite, simple, connected graph. Let $\{C_i\}$ be the components of $G$ cut open along its cut vertices.

\[^5\text{no multiple edges, no edges that are loops}\]
(1) $e \sim e'$ if and only if they are in the same $C_i$.

(2) The cardinality of $\{C_i\}$ is at most the number of edges of $G$ with equality iff $G$ is a tree.

Proof. (1): Since a circle has no cut vertex, if $e$ and $e'$ are contained in an embedded loop, then they are in the same $C_i$. If $e$ and $e'$ are in the same $C_i$ then there is an edge path $e_1, \ldots, e_p$ in $C_i$ from $e$ to $e'$. The common vertex between $e_i$ and $e_{i+1}$ is not a cut vertex of $C_i$ (indeed, $C_i$ has no cut vertices). It follows that $e_i$ and $e_{i+1}$ are contained in an embedded loop in $C_i$. Hence $e = e_1 \sim e_2 \sim \cdots \sim e_p = e'$.

(2): If $G$ is a tree then $\{C_i\}$ is the set of closed edges of $G$. In particular, we have equality. In general, consider the bipartite tree whose vertex set is the disjoint union of $\{C_i\}$ and the set $\{v\}$ of cut points of $G$ and whose edges are given by the relation $v \in C_i$. For the equality, note that, since $G$ is simple, if $G$ is not tree then some $C_i$ contains more than one edge.

We will consider the equivalence relation on $T_t$ generated by $\bar{\tau}_{t,i} \sim \bar{\tau}_{t,j}$ if there is an embedded loop in the Whitehead graph $WG_t$ of a subgate containing $\bar{\tau}_{t,i}$ and $\bar{\tau}_{t,j}$. By Lemma 5.2(1), equivalent turns are represented by edges in the same component of $WG_t$ cut open along its cut vertices. According to the Lemma 5.1 unfolding equivalent turns produces equivalent turns, and in particular the turns have the same vertex. Recall that $m_t = m(G_t)$ is the illegality of $G_t$, see Page 22.

Lemma 5.3. (1) $|T_t/\sim| \leq m_t$ with equality iff subgates coincide with gates and the Whitehead graph of subgates are trees.

(2) $|T_t/\sim| \leq |T_t|$ with equality iff Whitehead graphs of subgates are trees.

Proof. The first item follows from definitions and the second is a consequence of Lemma 5.2(2).

We are now ready for the main result of this section. Let $\check{m}$ denote the maximal possible number of illegal turns for any train track structure on any element of $X$.

Proposition 5.4 (surviving illegal turns). Let $z$ be a simple class and $G_t$, $t \in [0, L]$, a folding path in $X$ parametrized by arclength. Assume that $M = 2\check{m} + 1$ consecutive illegal turns of $z|G_0$ survive to $G_L$ and that the legal segments between them in $z|G_L$ have bounded size. Then $d_F(G_0, G_L)$ is bounded.
We saw above that two illegal turns in our consecutive collection in $z|G_t$ that give the same element of $T_t$ also give the same element of $T_{t'}$ for $t' \leq t$, i.e. we saw $\tau_{t,i} = \tau_{t,j}$ implies $\tau_{t',i} = \tau_{t',j}$. However, distinct illegal turns might unfold to the same illegal turn, i.e. $\tau_{t',i} = \tau_{t',j}$ does not imply $\tau_{t,i} = \tau_{t,j}$. So $|T_{t'}| \leq |T_t|$. By partitioning $[0,L]$ into a bounded number of subintervals and renaming, we may assume in proving Proposition 5.4 that $|T_t|$ is constant on $[0,L]$. Likewise, we may assume that $|T_{t}/\sim|$ is constant (in general $|T_{t}/\sim|$ may decrease under unfolding when a new circle is formed). The proof of Proposition 5.4 breaks into two cases:

1. Some subgate contains a circle in its Whitehead graph at time $L$.

2. There are no circles in Whitehead graphs of subgates at time $L$.

In Lemma 5.5, resp. Lemma 5.6, we show that Proposition 5.4 holds in Case 1, resp. Case 2. So once we have proved these next two lemmas, we will also have proved Proposition 5.4.

In Lemma 5.5 we prove a little more.

**Lemma 5.5.** Suppose that, in addition to the hypotheses of Proposition 5.4, some subgate contains a circle in its Whitehead graph at time $L$. Then, there is a partition $0 = t_0 < t_1 < \cdots < t_N = L$ of $[0,L]$ into boundedly many subintervals and simple classes $u_1, u_2, \ldots, u_N$ such that

- $k(u_i|G_{t_{i-1}})$ is bounded (recall that $k(u_i|G_{t_{i-1}})$ is the number of illegal turns in $u_i|G_{t_{i-1}}$); and
- $\ell(u_i|G_t)$ is bounded for all $t \in [t_{i-1}, t_i]$.

In particular $d_F(G_0, G_L)$ is bounded.

*Proof.* In light of Lemma 4.10, we only need to prove there are simple $u_i$ such that:

- $k(u_i|G_{t_{i-1}})$; and
- $\ell(u_i|G_{t_i})$ is bounded.

Since some subgate contains a circle in its Whitehead graph at time $L$, by Lemma 5.1 the same is true at every $t$. Choose distinct illegal turns $\tau_{L,i}, \tau_{L,j}$ in $z_L$ that are equivalent in $T_L$ so that the number of illegal turns between in the good portion of $z_L$ is smaller than the number of equivalence classes. (This is possible because circles in Whitehead graphs have more than one edge.) Let $[p_{t,i}, p_{t,j}]$ denote the resulting segment in the good portion of $z_t$.
and let $\sigma_t$ be the loop obtained by closing up our segment, i.e. by identifying $p_{t,i}$ and $p_{t,j}$. We refer to $\sigma_t$ as a monogon because it is immersed in $G_t$ except possibly at the point $p_{t,i} = p_{t,j}$. In particular, the conjugacy class $w(t)$ in $\mathbb{F}$ represented by $\sigma_t$ is nontrivial. Of course, $w(t)$ is also represented by the immersed circle $w(t)\mid G_t$ which is obtained by tightening $\sigma_t$. By construction and Lemma 5.3(1), $k(w(t)\mid G_t) \leq |T_t/\sim| < m_t$. In particular, $w(L)\mid G_L$ has bounded length. We claim that $w(0) = w(L)$. Once this claim is established, the last sentence of Lemma 4.11 (with $m$ equal to the constant $|T_t/\sim|$) completes Case 1.

To prove our claim, we must show that $\sigma_0$ and $\sigma_L$ determine the same conjugacy class. In folding $z_0$ to $z_L$, maximal arcs in the directions of $\tau_{0,k}$ are identified in $G_L$, i.e. they have the same image which is an immersed arc $\alpha_k$ in $G_L$. If we tighten the image of $[p_{0,i}, p_{0,j}]$ in $G_L$, the result is the image of $[p_{L,i}, p_{L,j}]$ extended at its ends by $\alpha_i$ and $\alpha_j$. The claim follows since $\alpha_i = \alpha_j$. Indeed, $\tau_{0,i}$ and $\tau_{0,j}$ are equivalent, and so in the same subgate, and so in the same gate. We see $\alpha_i = \alpha_j$ at least if $L$ is small enough and that this condition persists.

Lemma 5.6. Suppose that, in addition to the hypotheses of Proposition 5.4, no subgate contains a circle in its Whitehead graph at time $L$. Then $d_F(G_0, G_L)$ is bounded.

Proof. Since there are no circles in Whitehead graphs of subgates at time $L$, $s := |T_L/\sim| = |T_L| \leq m_L$. By our assumptions that $|T_t/\sim|$ and $|T_t|$ are constant on $[0, L]$, there are no circles for any $t$. In particular, $s \leq m_t$ for all $t \in [0, L]$. See Lemma 5.3

First assume that there are two occurrences of the same illegal turn in the consecutive collection at time $L$ that are separated by $< s - 1$ illegal turns, i.e. there are $\tau_{L,i}$ and $\tau_{L,j}$ with $\tau_{L,i} = \tau_{L,j}$ and $0 < j - i < s - 1$. Closing up gives a curve with $< s$ illegal turns, so again the conclusion follows by arguing as in Case 1.

So from now on we assume that this does not happen, i.e. all $s$ illegal turns occur repeatedly in a cyclic order in the consecutive collection along $z_t$. If it so happens that one of these illegal turns at $t = 0$ involves an edge $e$ which is nonseparating and has length a definite fraction $p$ of $\text{injrad}(G_0)$, we argue using Lemma 4.12 as follows. Consider the loop in $G_0$ obtained by closing up the segment that starts with $e$ and ends at the next occurrence of the same illegal turn. (Such a segment exists since there are $M > \tilde{m}$ illegal turns in our collection.) If the edge following this segment is $e$, we can appeal to Lemma 4.12 to deduce the conclusion of the lemma (because $z$ is simple).
If the edge following the segment is not $e$, then the last edge $\bar{e}$ of the segment is $e$ with the opposite orientation and closing up forces cancellation. If the tightened loop has length $< 2$, it is simple and its image in $G_L$ is bounded, so the conclusion follows (cf. the first sentence of the proof of Lemma 4.11). If the tightened loop has length $\geq 2$, then the original segment from $e$ to $\bar{e}$ has length $\geq 2 + 2p \text{injrad}(G_0)$ and the same argument as in Case 4 of Lemma 4.11 (using Lemma 4.4(2) instead of Lemma 4.4(1); see also proof of Lemma 4.12) shows that for $\ell(v|G_0) = \text{injrad}(G_0)$ we must have $\ell(v|G_t)$ bounded.

To summarize, the conclusion of the lemma holds whenever the following condition is satisfied at $t = 0$:

$$(\star) \text{ There is a nonseparating edge } e \text{ in } G_t \text{ such that:}$$

(i) $e$ is in the good portion of $z_t$;
(ii) $\ell_G(e) \geq p \text{injrad}(G_t)$; and
(iii) $e$ is involved in some turn in our collection $T_t$.

So now all that remains is to reduce to the case where $(\star)$ is satisfied at $t = 0$.

As a warmup, first consider the case where $m(G_t) = s$ for all $t$. In particular, every edge that is involved in an illegal turn of $G_t$ is involved in an illegal turn of our collection. Let $\beta \in [0, L]$ be the first time that a nonseparating edge of length $\geq \frac{1}{3n-3} \text{injrad}(G_\beta)$ is involved in an illegal turn of $G_\beta$. If there is no such $\beta$ then we claim that the conclusion of the lemma holds. Indeed, if $v$ is a conjugacy class with $\ell(v_0|G_0) = \text{injrad}(G_0)$ where $v_0 := v|G_0$ then there is a nonseparating edge $e_0$ of length $\geq \frac{1}{3n-3} \text{injrad}(G_0)$ in $v_0$. By Lemma 4.9, there is a corresponding edge $e_t$ for small $t$. Note that

$$\ell(e_t) = \ell(e_0)e_t^\ell(e_0) \geq \frac{\ell(v_0)e_t}{3n-3} \geq \frac{\ell(v_0)}{3n-3} \geq \frac{1}{3n-3} \text{injrad}(G_t)$$

and so $e_t$ is not involved in any illegal turns. In fact, we see that there can be no first time where $\ell(e_t) < \frac{1}{3n-3} \text{injrad}(G_t)$. We have $\ell(v_t) \leq e_t^\ell(e_0) \leq e_t \ell(e_0) = (3n-3)\ell(e_t) < 3n-3$ is bounded.

If there is such a $\beta$, then for the same reason the conclusion holds for $G_t$, $t \in [0, \beta]$. Also, the conclusion holds for $G_t$, $t \in [\beta, L]$ since $(\star)$ holds at $t = \beta$.

It remains to consider the case when $m_t$ is perhaps sometimes $> s$. Let $G_t$, $t \in [0, L']$ be the path in $X$ starting at $G_0$ obtained as in 2.2.C by folding only the $s$ illegal turns in our collection $T_t$ (and then normalizing
and reparametrizing by arclength). Note that by the Unfolding Principle each illegal turn $\tau_{0,i}$ our collection at time 0 comes with a pair of legal paths that get identified in the folding process. Folding only these turns amounts to identifying these paths. There are induced morphisms $G_0 \to e^{-L'}G_{L'}' \to e^{-L}G_L$. (In particular, $G_I'$ is a folding path.) We can scale the second of these morphisms to obtain $\phi : G_{L'}' \to G_L$. By the special case $m(G_I') = s$, there are simple classes $v$ and $u$ and a partition of $[0,L']$ into two subintervals such that $\ell(v|G_I')$ is bounded on the first and $\ell(u|G_I')$ is bounded on the second. Note that $\phi$ is an immersion on the segment spanned by our collection of turns, i.e. this segment is now legal in the train track structure induced by $\phi$. Let $x|G_{L'}'$ be obtained by identifying consecutive occurrences of the same oriented element of our collection (possible since $M > 2\tilde{m}$). So, $\ell(x|G_{L'}')$ is legal with respect to $G_{L'}' \to G_L$. Also $k(x|G_0)$ and $\ell(x|G_L)$ are bounded. If $\ell(x|G_L) < 2$, then $x$ is simple and the conclusions of the lemma hold. If $\ell(x|G_{L'}') \geq 2$, then $d_{\mathcal{X}}(G_{L'}',G_L) = L - L'$ is bounded. Indeed, since

$$\ell(w|G_L) = \ell(w|G_{L'}')e^{L-L'} \geq 2e^{L-L'}$$

is bounded, so is $L - L'$. In particular, $\ell(u|G_L)$ is bounded. $\Box$

We have completed the proof of Proposition 5.4 $\Box$

**Definition 5.7.** An immersed path or a loop in a metric graph $G$ equipped with a train track structure is illegal if it does not contain a legal segment of length 3 (in the metric on $G$).

**Lemma 5.8.** Let $z$ be a simple class and $G_t$, $t \in [0,L]$, a folding path in $\mathcal{X}$ parametrized by arclength. Assume $z_t$ is illegal for all $t$. Then either $\ell(z_t) < \ell(z_0)/2$ or $d_{\mathcal{F}}(G_0,G_L)$ is bounded.

**Proof.** There are two cases. First suppose that the average distance between consecutive illegal turns in $z_0$ is $\geq 1/\tilde{m}$. Then, by Proposition 5.4 (surviving illegal turns), after a bounded progress in $\mathcal{F}$ the loop $z$ must lose at least $1/M$ of its illegal turns. Repeating this a bounded number of times, we see that, after bounded progress in $\mathcal{F}$, $k_t \leq k_0/6\tilde{m}$. Thus either the length of $z_t$ is less than $1/2$ of the length of $z_0$ or the average distance between illegal turns at time $t$ is $\geq 3$, so there is a legal segment.

Now suppose the average distance between illegal turns in $z_0$ is $< 1/\tilde{m}$. By Lemma 4.4

$$\ell'(z_0) = \ell(z_0) - 2\frac{k_0}{m_0}$$

We are assuming that $k_0/m_0 > \ell(z_0)$ so the above derivative is $< -\ell(z_0)$. Thus in this case the length of $z$ decreases exponentially, until either half
the length is lost after a bounded distance in $X$, or the average distance between illegal turns becomes $\geq \frac{1}{\hat{m}}$, when the above argument finishes the proof.

\begin{remark}
One source of asymmetry between legality and illegality is that a long legal segment gets predictably longer under folding, while a long illegal segment may not get longer under unfolding. For example, take a surface relation inside a subgraph where folding amounts to an axis of a surface automorphism. But the lemma above implies that an illegal segment inside a loop representing a simple class will get predictably longer under unfolding after definite progress in $F$.
\end{remark}

6 Projection to a folding path

We thank Michael Handel for pointing out the technique for proving the following lemma (see [14, Proposition 8.1]).

\begin{lemma}
Let $G \in X$ be a metric graph with a train track structure and $A < F$ a free factor. Suppose $A|G$ satisfies:
\begin{itemize}
  \item there is an illegal loop $a$ in $A|G$ (see Definition 5.7),
  \item there is an immersed legal segment in $A|G$ of length $3(2n-1)$, $n = \text{rank}(F)$.
\end{itemize}
Then $d_{\mathcal{F}}(A,G)$ is bounded.
\end{lemma}

\begin{proof}
If the injectivity radius of $A|G$ is $\leq 3(2n-1)$ the conclusion follows from Corollary 3.7.

Choose a complementary free factor $B$ to $A$ and add a wedge of circles to $A|G$ representing $B$ to get a graph $H$. Extend $A|G \to G$ to a homotopy equivalence (difference of markings) $H \to G$, which is an immersion on each 1-cell. Pull back the metric to $H$ and consider the folding path in $\hat{X}$ induced by $H \to G$. Let $H'$ be the first graph on this folding path with injectivity radius $3(2n-1)$. Now give $H'$ the pullback illegal turn structure via $H' \to G$. Since $A|G \to H'$ is an immersion, the interior of the legal segment in $A|G$ embeds in $H'$. Since $H'$ has at most $2n-2$ topological vertices, the interior of the legal segment meets at most $2n-2$ topological vertices. So there is a legal segment of length 3 inside one of the topological edges of $H'$. Thus $a|H'$ does not cross this topological edge and hence $d_{\mathcal{F}}(a,H') \leq 1 + 4$ (1 bounds the distance between $\hat{a}$ and the free factor given by the subgraph of $H'$ traversed by $a$ and 4 coming from Lemma 3.1). Since $d_{\mathcal{F}}(a,A) \leq 1$ and
$d_F(H', G)$ is bounded by Corollary 3.7 applied to a shortest loop in $H'$, the statement follows. \hfill \Box

The following is a slight generalization.

**Lemma 6.2.** Let $G \in \mathcal{X}$ be a metric graph with a train track structure and $A < \mathbb{F}$ a free factor. Suppose $A|G$ satisfies:

- there is a loop $a$ in $A|G$ with the maximal number of pairwise disjoint legal segments of length 3 bounded by $N$,

- there is an immersed legal segment in $A|G$ of length $3(2n - 1)$, $n = \text{rank}(\mathbb{F})$.

Then $d_F(A, G)$ is bounded as a function of $N$.

**Proof.** The proof is similar. With $H'$ defined as in the proof of Lemma 6.1 there is an edge in $H'$ which is crossed by $a|H'$ at most $N$ times. Hence, by Lemma 3.3 $d_F(H', a) \leq 6N + 13$. \hfill \Box

For the rest of this section, let $G_t, t \in [0, L]$, be a folding path in $\mathcal{X}$ parametrized by arclength and let $A$ be the conjugacy class of a proper free factor. A folding path has a natural orientation given by the parametrization. We will think of this orientation as going left to right.

Set

$$I := (18\bar{m}(3n - 3) + 6)(2n - 1)$$

where recall that $\bar{m}$ is the maximal possible number of illegal turns in any $G \in \mathcal{X}$ (so $\bar{m}$ is some linear function of the rank). The number $I$ comes from Proposition 6.10 (legal and illegal) which will say that if $A|G_0$ has a long (i.e. of length $> 3$) legal segment and $A|G_L$ has a long (i.e. of length $> I$) illegal segment then $d_F(G_0, G_L)$ is bounded. Recall that a segment is illegal if it does not contain a legal subsegment of length 3. This motivates the following definitions.

**Definition 6.3.** Denote by left$_{G_t}(A)$ (or just left($A$) if the path $G_t$ is understood) the number

$$\inf\{t \in [0, L] \mid A|G_t \text{ has an immersed legal segment of length 3}\}$$

The left projection Left$_{G_t}(A)$ of $A$ to the path $G_t$ is $G_{\text{left}(A)}$.

Denote by right$_{G_t}(A)$ (or just right($A$) if $G_t$ is understood) the number

$$\sup\{t \in [0, L] \mid A|G_t \text{ has an immersed illegal segment of length } I\}$$


The right projection \( \text{Right}_{G_t}(A) \) of \( A \) to the path is \( G_{\text{right}(A)} \). If the above sets are empty, we interpret \( \inf \) as \( L \) and \( \sup \) as 0.

We make analogous definitions for any simple class \( a \), so e.g. \( \text{left}_{G_t}(a) \) is the first time \( a | G_t \) contains a legal segment of length 3 (here wrapping around is allowed, i.e. \( a | G_t \) may be legal and of length < 3). In all cases, we may suppress subscripts if the path is understood. Note that the first set displayed above is closed under the operation of increasing \( t \). Clearly, \( \text{left}(A) \leq \text{right}(A) \).

We also generalize these definitions from free factors to marked graphs in the obvious way. If \( H \in \mathcal{X} \), \( \text{left}(H) := \min \text{left}(\pi(H)) \), \( \text{Left}(H) := G_{\text{left}(H)} \), \( \text{right}(H) := \max \text{right}(\pi(H)) \), and \( \text{Right}(H) := G_{\text{right}(H)} \).

**Proposition 6.4.** Suppose \( B < A \) are free factors. Then:

- \( \text{left}(A) \leq \text{left}(B) \), \( \text{right}(A) \geq \text{right}(B) \); and
- either \( d_X(\text{Left}(A), \text{Left}(B)) = e_{\text{left}(B) - \text{left}(A)} \) is bounded or the distance in \( F \) between \( A \) and \( \{G_t \mid t \in [\text{left}(A), \text{left}(B)]\} \) is bounded.

**Proof.** The first bullet is clear since we are taking \( \inf \) and \( \sup \) over smaller sets. Denote by \( \text{left}'(A) \) the first time along the folding path that \( A | G_t \) has a legal segment of length \( 3(2n - 1) \). Then \( \text{left}(A) \leq \text{left}'(A) \), and \( \text{left}'(A) - \text{left}(A) \) is bounded by Corollary 6.5. If \( \text{left}(B) \leq \text{left}'(A) \) we are done, so suppose \( \text{left}'(A) < \text{left}(B) \). It follows from Lemma 6.1 that the set of \( G_t \)'s for \( t \in [\text{left}'(A), \text{left}(B)] \) has a bounded projection in \( F \), and the projection is close to \( A \).

Given a constant \( K > 0 \), we will say that a coarse path \( \gamma : [\alpha, \omega] \rightarrow F \) is a reparametrized quasi-geodesic if there is a subdivision \( \alpha = t_0 < t_1 < \cdots < t_m = \omega \) such that \( \text{diam}_x(\gamma([t_i, t_{i+1}])) \leq K \), \( m \leq d_F(\gamma(\alpha), \gamma(\omega)) \), and \( |i - j| \leq d_F(\gamma(t_i), \gamma(t_j)) + 2 \) for all \( i, j \). In particular, a map \( [0, m] \rightarrow F \) given by mapping \( x \in [0, m] \) to an element of \( \gamma(t_{[x]}) \) is a quasi-geodesic with constants depending only on \( K \). A collection \( \{\gamma_i\}_{i \in I} \) of reparametrized quasi-geodesics is uniform if the \( K \) appearing in the definition of \( \gamma_i \) is independent of \( i \) and is, in fact, a function of the rank \( n \) of \( F \) alone. A coarse Lipschitz function \( f : X \rightarrow Y \) between metric spaces is one that satisfies \( d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2) + K \) for all \( x_1, x_2 \in X \). A function \( f : X \rightarrow A \subseteq X \) is a coarse retraction if \( d(a, f(a)) \leq K \) for all \( a \in A \). In all these cases, \( f \) is allowed to be multivalued with the bound of \( K \) on the diameter of a point image.

**Corollary 6.5.** For any folding path \( G_t \) the projection

\[ \mathcal{F} \rightarrow \pi(G_t) \]
is a coarse Lipschitz retraction with constants depending only on $\text{rank}(\mathcal{F})$. Consequently, the collection of paths $\{\pi(G_t)\}$ is a uniform collection of reparametrized quasi-geodesics in $\mathcal{F}$.

**Proof.** That the map is coarsely Lipschitz follows from Proposition 6.4. To prove that it is a coarse retraction, we need to argue that $\pi(\text{Left}(G_{t_0}))$ is bounded distance from $\pi(G_{t_0})$ for $t_0 \in [0, L]$. Let $a$ be the conjugacy class of a legal candidate in $G_{t_0}$, so $\text{left}(a) \leq t_0$ and $d_\mathcal{F}(a, G_{t_0})$ is bounded. By Proposition 6.4 again, it is enough to argue that $d_\mathcal{F}(\text{Left}(a), G_{t_0})$ is bounded. Let $t'$ be the smallest parameter such that $a_{t'}$ is legal. Then $\ell(a_{t'}) \leq 2$ so $d_\mathcal{F}(a, G_{t'})$ is bounded. Now note that $\text{left}(a) = t'$ since for $t < t'$ any legal segment of length 3 in $a_t$ would force $\ell(a_{t'}) > 3$.

The argument for the second part is from [7]. Let $G_t$ be a folding path so that $\pi(G_t)$ is a coarse path joining free factors $A$ and $B$. Choose a geodesic $C_i$, $i = 0, \ldots, m$ of free factors joining $C_0 = A$ and $C_m = B$ in $\mathcal{F}$. Consider the coarse projection $D_i$ of $C_i$ to $\pi(G_t)$. By Proposition 6.4 the diameter of the segment bounded by $D_i$ and $D_{i+1}$ is uniformly bounded. Now the $D_i$’s may not occur monotonically along $\pi(G_t)$. To fix this, let $i_1 < i_2 < \cdots < i_k$ be the sequence defined inductively by $i_1 = 0$ and $i_{j+1}$ is the smallest index such that $D_{i_{j+1}}$ occurs after $D_{i_j}$ in the order on $G_t$ given by $t$. Then by construction the interval between $D_{i_j}$ and $D_{i_{j+1}}$ has uniformly bounded diameter and the number $k$ is bounded by $m = d_\mathcal{F}(A, B)$. Call a subdivision satisfying these properties a *admissible*. To ensure the last property $|i - j| \leq d_\mathcal{F}(\gamma(t_i), \gamma(t_j)) + 2$ take an admissible subdivision with minimal $k$. \(\square\)

**Definition 6.6.** Given $A \in \mathcal{F}$, $H \in \mathcal{X}$, and a folding path $G_t$, $\pi(\text{Left}(A))$ is the projection of $A$ to $\pi(G_t)$ and $\pi(\text{Left}(H))$ is the projection of $H$ to $\pi(G_t)$.

**Definition 6.7.** For $\kappa > 0, C > 0$ we say a folding path $G_t$ makes $(\kappa, C)$-definite progress in $\mathcal{F}$ if for any $D > 0$ and $s < t$, $d_\mathcal{X}(G_s, G_t) > D \kappa + C$ implies $d_\mathcal{F}(G_s, G_t) > D$.

**Corollary 6.8.** For any folding path $G_t$ the projection

\[
\mathcal{F} - R(G_t) \to \{G_t\}
\]

\[A \mapsto \text{Left}(A)\]

where $R(G_t)$ is the set of free factors at a certain bounded distance from $G_t$ measured in $\mathcal{F}$, is coarsely Lipschitz (with respect to the path metric in $\mathcal{F} - R(G_t)$).
Moreover, the projection is coarsely defined and coarsely Lipschitz on all of $\mathcal{F}$ provided $G_t$ makes $(\kappa,C)$-definite progress in $\mathcal{F}$ (with constants depending on $\kappa,C$).

**Lemma 6.9.** Let $G_t$, $t \in [0,L]$, be a folding path and $A$ a free factor. The length of any illegal segment contained in a topological edge of $A|G_L$ is less than

$$\frac{3}{2}m \cdot \text{edgelength}(A|G_0) + 6$$

where $m = \max\{m_t \mid t \in [0,L]\}$ and $\text{edgelength}(A|G_0)$ is the maximal length of a topological edge in $A|G_0$.

**Proof.** Fix an illegal segment of length $\ell_L$ in the interior of a topological edge of $A|G_L$. We will assume that the endpoints are illegal turns and argue

$$\ell_L \leq \frac{3}{2}m \cdot \text{edgelength}(A|G_0)$$

After adding $< 3$ on each end we recover any illegal path. By the Unfolding Principle our path lifts to an illegal segment bounded by illegal turns inside some topological edge of $A|G_t$, whose length will be denoted $\ell_t$. In particular, $\ell_0 \leq \text{edgelength}(A|G_0)$. In order to obtain a contradiction, assume $(\diamondsuit)$ fails. Let $t_0$ be the first time the right derivative of $\ell_t$ is nonnegative (if such $t_0$ does not exist then $\ell_L \leq \ell_0$ and $(\diamondsuit)$ holds, contradiction). Thus $\ell_{t_0} \leq \ell_0$ and the average length of a maximal legal segment inside the path is $\geq 2/m \ell_{t_0} \geq 2/m$ by Corollaries 4.5(3) and 4.7. Since $\ell_L \geq \frac{3}{2}m \ell_{t_0}$, the average length of a legal segment of our path is guaranteed to be $\geq 3$, contradicting the hypothesis that our segment is illegal. Thus $(\diamondsuit)$ holds and the lemma follows. \qed

Recall that the number $I$ used in the next proposition was defined on Page 37.

**Proposition 6.10** (legal and illegal). Let $G_t$, $t \in [0,L]$, be a folding path and $A$ a free factor. Assume that $A|G_0$ has a legal segment of length 3, and that $A|G_L$ has an illegal segment of length $I$. Then $d_{\mathcal{F}}(G_0,G_L)$ is bounded.

**Proof.** Since a legal segment of length 3 grows to a legal segment of length $> 12(3n - 3)(2n - 1)$ in bounded time (Corollary 4.8), by replacing $G_0$ with $G_t$ for a bounded $t$, we may assume that $A|G_0$ has a legal segment of length $12(3n - 3)(2n - 1)$. In order to obtain a contradiction, assume the distance $d_{\mathcal{F}}(G_0,G_L)$ is large. Let $\tau \in [0,L]$ then be chosen so that $d_{\mathcal{F}}(G_0,G_\tau)$, $d_{\mathcal{F}}(G_\tau,G_L)$ and $d_{\mathcal{F}}(G_\tau,A)$ are all large.
Wedge $A|G_{\tau}$ onto a rose representing a complementary free factor to $A$ in order to obtain a graph $H' \in \mathcal{X}$ and a difference of markings morphism $H' \to G_{\tau}$ extending $A|G_{\tau} \to G_{\tau}$ and which is an isometric immersion on every edge. In particular, $H' \to G_{\tau}$ induces a train track structure on $H'$. If $H'$ has bounded injectivity radius then $d_{\mathcal{X}}(A,G_{\tau})$ is bounded, contradicting the choice of $G_{\tau}$. So suppose the injectivity radius of $H'$ is large and fold $H' \to G_{\tau}$ until a graph $H''$ is reached which is the last time there is an edge $E''$ of length 4. Cf. [14, Proposition 8.1].

In particular, $\text{vol}(H'') \leq 4(3n - 3)$. We continue by folding with speed 1 the subset of those illegal turns of $H'' \to G_{\tau}$ that don’t involve $E''$. Since $H'' \to G_{\tau}$ induces a train track structure on $H''$, so does our subset. For small $t$, we obtain a graph $H''_t$ where the image $E''_t$ of $E''$ (perhaps no longer topological) still has length 4 and a morphism $H''_t \to G_{\tau}$ inducing a train track structure. We continue folding all illegal turns not involving $E''_t$ (as in 2.2.C) until we obtain morphism $H \to G_{\tau}$ inducing a train track structure and isometrically immersing both the image $E$ in $H$ of $E''$ and its complement. The only illegal turns of $H \to G_{\tau}$ involve $E$. In fact, since our illegal turns give a train track structure, the only illegal turns involve the topological edge containing $E$. We will now use $E$ for the name of this topological edge. After folding from $H''_t$ to $H$, some edge lengths may now be $> 4$. But since $\text{vol}(H) \leq \text{vol}(H'')$, edge lengths in $H$ are at most $4(3n - 3)$ (and $E$ still has length at least 4).

We may assume that the complement of $E$ does not have a valence 1 vertex. Indeed, assuming otherwise, with respect to the train track structure induced by $H \to G_{\tau}$ there are two possibilities for the illegal turns (which recall must involve $E$), see Figure 7. In the left picture, the length of $E$ stays constant under folding. We continue folding until the separating edge folds in with $E$, and this is our new $H$. The right picture is impossible: $E$ is a “monogon” (of length at least 4) and the folding towards $G_{\tau}$ stops before the loop degenerates. But this means that $G_{\tau}$ has volume $> 2$, a contradiction.

![Figure 7](image)

Figure 7: Two possibilities when $E$ is a loop attached to a separating edge. The square represents the remainder of the graph.
We will also assume for concreteness that the complement of $E$ is connected, and denote by $B$ the free factor determined by it. When the complement is disconnected, there are two free factors determined by the components. The changes are straightforward and left to the reader.

We have morphisms $H \to G_\tau \to G_L$ and now also bring in the pullback illegal turn structure via $H \to G_\tau$. To distinguish between the two structures, terms like $p$-legal and $p$-illegal will refer to this pullback, i.e. the one induced by $H \to G_L$. Terms like $i$-legal and $i$-illegal will refer to the structure induced by $H \to G_\tau$. The same terminology will be applied to turns in $\hat{H}_t$ (resp. $\hat{H}_t$ and $K_t$) with respect to $\hat{\psi}_s : \hat{H}_s \to \hat{G}_s$ (resp. $\hat{H}_t \to \hat{G}_t$ and $K_t \to \hat{G}_t$) constructed below. Note that $H$ may have $p$-illegal turns in the interior of topological edges. (Consider that perhaps $H = H'$.) By construction, $i$-illegal turns must involve $E$.

There are two cases.

**Case 1.** $E$ contains a $p$-legal segment of length 3. As $G_\tau$ folds toward $G_L$, we will use the technique of 2.2.C to fold $H$ and produce a new path (though usually not a folding path) in $X$. To describe this path, it is convenient to view the folding path $G_t$ as in Proposition 2.2, i.e. without rescaling and folding with speed 1. So, let $\hat{G}_{s(t)}$, $t \in [\tau, L]$, be the folding path $\hat{X}$ induced by the morphism $G_\tau \to e^{L-\tau}G_L$ with natural parameter $s$.

We claim that, for $s \in [s(\tau), s(L)]$, there is a path $\hat{H}_s$ in $\hat{X}$ that starts at $H$ and satisfies:

1. $\hat{H}_s = B|\hat{G}_s \cup \hat{E}_s$, where $\hat{E}_s$ is a topological edge containing a $p$-legal segment of length at least $3 \text{vol}(\hat{G}_s)$,

2. the immersion $B|\hat{G}_s \to \hat{G}_s$ extends to a morphism (difference of markings) $\hat{\psi}_s : \hat{H}_s \to \hat{G}_s$ inducing a train track structure on $\hat{H}_s$. In particular, $\hat{\psi}_s$ is an isometric immersion on $\hat{E}_s$,

3. $A|\hat{G}_s \to \hat{G}_s$ factors through $\hat{\psi}_s$. In particular, $A|\hat{G}_s \to \hat{H}_s$ is an isometric immersion.

By construction (1–3) hold for $\hat{H}_{s(\tau)} := H$ and $\hat{E}_{s(\tau)} := E$. Following 2.2.C, assume $\hat{H}_s$ has been defined on a subinterval $J$ of $[s(\tau), s(L)]$ containing $s(\tau)$. If $J = [s(\tau), s_0]$ with $s_0 \neq s(L)$, then we can, for small time $\epsilon > 0$, fold all $p$-illegal turns of $\hat{H}_{s_0}$ at speed 1 (see Page 8) thereby extending the path to $[s(\tau), s_0 + \epsilon]$. We see that (1–3) hold for $s \in [s_0, s_0 + \epsilon]$. Indeed, $\hat{H}_{s_0} \to \hat{G}_s$ factors through $\hat{H}_{s_0} \to \hat{H}_s$ and so $\hat{H}_s \to \hat{G}_s$ is a morphism. By Lemma 4.8, $\hat{H}_s$ has a topological edge $\hat{E}_s$ containing a $p$-legal segment of length at least $3 \text{vol}(\hat{G}_s)$ and whose complement has core representing $B$. 

}\text{42}
Since \(B|\hat{H}_{s_0}, A|\hat{H}_{s_0}\), and the interior of \(\hat{E}_{s_0}\) contain no \(i\)-illegal turn, the same is true at \(s\). There must be an \(i\)-illegal turn of \(\hat{H}_{s_0}\) involving both \(\hat{E}_{s_0}\) and an edge in \(B|\hat{H}_{s_0}\) (or else \(\hat{H}_s\) has a monogon as above which has been ruled out).

We move to the case \(J = [s(\tau), s_0]\). As in \[\text{4.2}\] C, we may define a limit tree \(\hat{H}_{s_0} \in \hat{X}\). By Lemma \[\text{4.8}\] \(\hat{H}_{s_0}\) has a topological edge containing a \(p\)-legal segment of length at least \(3\ \text{vol}(G_{s_0})\) and whose complement has core representing \(B\). The limit of these morphisms is a morphism, so (2) and (3) also hold. Finally, \(\hat{E}_{s_0}\) can't be a loop connected to \(B|\hat{H}_{s_0}\) by a separating edge (or else the same would have been true at smaller \(s\).

Set \(H_t := H_s/\text{vol}(\hat{G}_s)\) and define the image of \(\hat{E}_s\) in \(H_s\) to be \(E_s\). Reverting to our original parametrization, we now have our original path \(G_t, t \in [\tau, L]\), in \(X\) and a new path \(H_t, t \in [\tau, L]\), in \(\hat{X}\) such that, for each \(t\), 
\[H_t = B|H_t \cup E_t, \] there is a morphism \(\psi_t : H_t \to G_t, A|G_t \text{ isometrically embeds in } H_t, \text{ and } E_t \text{ contains a } p\text{-legal segment of length at least 3}.\]

We need one more modification to control the length of \(E_t\). Define \(K_t \in \hat{X}\) as follows. If the length of \(E_t\) is \(\leq 4\), \(K_t := H_t\). If the length of \(E_t\) in \(H_t\) is \(> 4\), define \(K_t\) to be the graph obtained by folding \(i\)-illegal turns of \(H_t \to G_t\) until the length of \(E_t\) is 4. Since \(A|G_t\) and \(B|G_t\) are immersed in \(H_t\), the only effect is to fold pieces of the end of \(E_t\) into \(B|G_t\). In particular, the analogues of (1-3) hold for \(K_t\), except it is possible that \(E_t\) no longer has a \(p\)-legal segment of length 3.

By keeping in mind that the length of \(E_t\) in \(K_t\) is at most 4 and applying Lemma \[\text{6.9}\] to \(G_t, t \in [\tau, L]\), and \(B\), the length of any \(p\)-illegal segment contained in a topological edge of \(K_t\) is bounded by 
\[\frac{3}{2} \hat{m} \cdot \text{edgelength}(B|G_{\tau}) + 6 \leq \frac{3}{2} \hat{m} \cdot 3 \text{ edgelength}(K_{\tau}) + 6 \leq (18\hat{m}(3n-3)+6) (2n-1)\]

Since the number of topological vertices of \(K_t\) is \(\leq 2n - 2\), a \(p\)-illegal segment in \(K_t\) of length \(I = (18\hat{m}(3n-3)+6)/(2n-1)\) meets some topological vertex of \(K_t\) twice. We see that, for \(t \in [\tau, L]\), either the \text{injrad}(K_t) is bounded by \(I\) (in which case \(d_{\mathcal{F}}(G_t, B)\), and hence \(d_{\mathcal{F}}(G_t, G_{\tau})\), is bounded), or there are no \(p\)-illegal segments in \(K_t\) of length \(I\) and hence the same holds for \(A|G_t\). Applying this to \(t = L\) we see that \(d_{\mathcal{F}}(G_{\tau}, G_L)\) is bounded, contradicting the choice of \(G_{\tau}\).

**Case 2.** \(E\) doesn’t contain a \(p\)-legal segment of length 3. In particular, the interior of \(E\) crosses a \(p\)-illegal turn. Let \(\hat{G}_s(t), t \in [0, \tau]\), be the folding path in \(\hat{X}\) giving rise to \(G_t\) and ending at \(G_{\tau}\). We will produce a path
\( \hat{H}_{s(t)}, t \in [0, \tau] \) (usually not a folding path) in \( \hat{X} \) ending at \( H \) and, for each \( s \in [s(0), s(\tau)] \), satisfying:

1. \( \hat{H}_s = B|\hat{G}_s \cup \hat{E}_s \), where \( \hat{E}_s \) is a single edge,

2. the immersion \( B|\hat{G}_s \to \hat{G}_s \) extends to a morphism (difference of markings) \( \psi_s : H_s \to G_s \) which is an isometric immersion on \( \hat{E}_s \),

3. \( A|\hat{G}_s \to \hat{G}_s \) factors through \( \hat{\psi}_s \). In particular, \( A|\hat{G}_s \to \hat{H}_s \) is an isometric immersion.

Note that (1–3) hold for \( \hat{H}_{s(\tau)} = H \). Let \( 0 = s_0 < s_1 < \cdots < s_N = s(\tau) \) be a partition of \( [0, s(\tau)] \) so that the restriction of \( \hat{G}_{t_i} \) to each \( [s_i, s_{i+1}] \) is given by folding a gadget. Assume \( \hat{H}_s \) has been defined for \( s \in [s_i, s_N] \) satisfying (1–3). We now work to extend \( \hat{H}_s \) over \( [s_{i-1}, s_N] \) still satisfying (1–3).

We first define \( \hat{H}_s \), \( s \in (s_{i-1}, s_i] \), via the following local operations. \( \hat{H}_s \) is defined as \( B|\hat{G}_s \) with an edge \( \hat{E}_s \) attached, and we specify the attaching points. Consider first the case that a direction \( e \) of an end of \( \hat{E}_{s_i} \) forms a \( \hat{i} \)-illegal turn with a direction \( b \) in \( B|\hat{G}_{s_i} \) (such a direction is then unique).

Intuitively, as \( s \) decreases, \( B|\hat{G}_s \) unfolds and we choose to fold \( b \) and \( e \) with speed 1. A more elaborate description follows.

Let \( \hat{\phi} = \hat{\phi}_{s_{i_{-1}}} : \hat{G}_s \to \hat{G}_{s_i} \) be the folding morphism. It induces a morphism \( \hat{\phi}_B : B|\hat{G}_s \to B|\hat{G}_{s_i} \). Let \( \epsilon = s_i - s \) and let \( \hat{N} \) be the \( \epsilon \)-neighborhood in \( B|\hat{H}_{s_i} = B|\hat{G}_{s_i} \) of the vertex \( v \) of \( e \). \( N(B) \) is a subset of the \( \epsilon \)-neighborhood \( N \) of \( v \) in \( G_{s_i} \). \( N_\epsilon \) denotes the preimage in \( \hat{G}_s \) of \( N \) and \( \hat{N}_\epsilon \) is the preimage of \( \hat{N} \) in \( B|\hat{G}_s \). Using the language of widgets, we attach the end of \( \hat{E}_s \) corresponding to \( e \) to the base of \( \hat{b}^* \) in \( \hat{N}_\epsilon \). (To recall notation, see Figure 6.)

To define \( \hat{E}_s \), delete a length \( \epsilon \) segment from the end of \( \hat{E}_{s_i} \). \( B|\hat{G}_s \to \hat{G}_s \) now extends to a morphism \( \hat{\psi}_s : \hat{H}_s \to \hat{G}_s \). Figure illustrates the diagram

\[
\begin{array}{ccc}
\hat{N}_\epsilon & \longrightarrow & \hat{N} \\
\downarrow & & \downarrow \\
N_\epsilon & \longrightarrow & N
\end{array}
\]

with ends of \( \hat{E}_s \) and \( \hat{E}_{s_i} \) attached to (resp.) \( \hat{N}_\epsilon \) and \( \hat{N} \).

Now suppose that the direction \( e \) does not form an \( \hat{i} \)-illegal turn with any direction in \( B|\hat{G}_{s_i} \). Then there is a natural way to construct the attaching point in \( \hat{G}_s \) by watching \( \hat{G}_{s_i} \) unfold to \( \hat{G}_s \). In terms of widgets, we attach the end of \( \hat{E}_s \) corresponding to \( e \) to the point in \( B|\hat{G}_s \) closest to \( \hat{e}^* \). Figure 9
Figure 8: The thickened segments represent $\tilde{b}^s$ and $b$. The curved segments represent ends of $\tilde{E}_s$ and $\tilde{E}_{s_i}$.

illustrates the diagram

$\tilde{N}_c \longrightarrow \tilde{N}$

$\downarrow \quad \downarrow$

$N_c \longrightarrow N$

with $\tilde{E}_s$ and $\tilde{E}_{s_i}$ attached.

There is a unique homotopy class of paths in $\hat{G}_s$ connecting the images of attaching points. The map $\psi_s$ is defined so that it isometrically immerses $\hat{E}_s$ to the immersed path the above homotopy class.

Now suppose $\hat{H}_s$ is defined for $s \in (s_{i-1}, s_i)$ and we want to define $\hat{H}_{s_{i-1}}$. Let $\sigma$ be a conjugacy class in $F$. By construction, for $s \in (s_{i-1}, s_i)$, $\ell(\sigma|G_s) = \ell(\sigma|G_{s_i}) + (2x - y)(s_i - s)$ where $x$ is the number of $p$-illegal turns crossed by $\sigma|G_s$ that are $i$-legal and $y$ is the number of $i$-illegal turns crossed. Note that $x$ and $y$ are constant on $(s_{i-1}, s_i)$. In particular, $\lim_{s \to s_{i-1}} \ell(\sigma|\hat{H}_s)$ exists, thereby defining $\hat{H}_{s_{i-1}}$. That $\hat{H}_{s_{i-1}}$ is in $\hat{X}$ follows from the existence of the limiting morphism $\hat{H}_{s_{i-1}} \to \hat{G}_{s_{i-1}}$. Finally, note that $\hat{E}_s$ doesn’t degenerate to a point in this limit. Indeed, $\hat{E}_s(\tau)$ crosses a $p$-illegal turn and this property persists by the Unfolding Principle (as $s$ decreases, $p$-illegal turns in $\hat{E}_s$ move away from the endpoints of $\hat{E}_s$ which balances any loss at the ends of $\hat{E}_s$ due to $i$-illegal turns).
Figure 9: To decide where to attach \( \hat{E}_s \) mimic what happens in \( \hat{G}_s \). The thickened lines represent \( \hat{E}_s \) and \( \hat{E}_s \).

Set \( H_s = \hat{H}_s / \text{vol}(\hat{G}_s) \) and revert to our original parametrization. We now have a path \( H_t, t \in [0, \tau] \). Define \( K_t \) exactly as before, i.e. if \( \ell(E_t) > 4 \), then fold \( i \)-illegal turns of \( H_t \rightarrow G_t \) until \( E_t \) has length 4.

A \( p \)-legal segment of length \( > 3 \cdot 4(3n - 3) \) interior to an edge of \( K_s \) would produce an edge in \( B|G_s \), hence also in \( B|G_\tau \), of length \( > 3 \cdot 4(3n - 3) \). We would then have an edge of length \( > 4(3n - 3) \) in \( K_\tau \), contradiction. A \( p \)-legal segment in \( K_s \) of length \( > 12(3n - 3)(2n - 1) \) is then forced meet least \( 2n - 1 \) topological vertices which implies \( \text{injrad}(K_s) \leq 12(3n - 3)(2n - 1) \). By assumption, \( A|G_0 \), hence also \( K_0 \), has a \( p \)-legal segment of length \( > 12(3n - 3)(2n - 1) \). Arguing as at the end of Case 1, we get the contradiction that \( d_X(G_0, G_\tau) \) is bounded.

**Corollary 6.11.** The image in \( \mathcal{F} \) of the interval \([\text{Left}_{G_t}(A), \text{Right}_{G_t}(A)]\) has bounded diameter.

**Proof.** The endpoints have bounded \( \mathcal{F} \)-distance by Proposition 6.10 (legal and illegal), and therefore the whole interval projects to a bounded set by Corollary 6.5.

Corollary 6.11 says that the projection of \( A \in \mathcal{F} \) to \( \pi(G_t) \) is bounded distance from \( \pi([\text{Left}_{G_t}(A), \text{Right}_{G_t}(A)]) \). (Recall Definition 6.6) We will, in Lemma 6.16 see a way to estimate where this projection lies. To prove Lemma 6.16 we first need a simple lemma about cancelling paths.
in a graph and then a general fact that in a different form appears in [1]
Proposition 5.10 and claim on p. 2218.

If \( X \) is an edge-path in a graph, then \([X]\) denotes the path obtained from
\( X \) by tightening, i.e. \([X]\) is the immersed edge-path homotopic rel endpoints
to \( X \). If the endpoints of \( X \) coincide and the resulting closed path is not
null-homotopic, then \([X]\) denotes the loop obtained from \( X \) by tightening,
i.e. \([X]\) is the immersed circle freely homotopic to \( X \).

**Lemma 6.12.** Let \( V \) be an immersed edge path in a graph \( G \). Suppose
\( V \) represents an immersed circle, i.e. \( V \) begins and ends at a vertex \( P \) and has
distinct initial and terminal directions. Let \( W \) be a nontrivial initial edge
subpath of \( V \) ending at a vertex \( Q \) and let \( V' = [W^{-1}W] \) (so that \( V' \) also
represents an immersed circle). Also, let \( X \) and \( Y \) be immersed edge paths
in \( G \) starting at \( P \) and \( Q \) respectively. Suppose that \( WY \) is immersed in \( G \).
Then:

1. the maximal common initial subpath of \( X \), \( WY \), and \( WV'Y \) has the
   form \( V^NW' \) for some \( N \geq 0 \) and some initial subpath \( W' \) of \( V \).

2. The maximal initial subpath of \( X \) and \( VX \) has the form \( V^NW' \) for some
   \( N \geq 0 \) and some initial subpath \( W' \) of \( V \).

The proof of Lemma 6.12 is left to the reader. Note that (2) follows from
(1) applied to \( X \), \( VX \), and \( VX \).

If \( a \) and \( b \) are conjugacy classes in \( F \) and \( G \in X \), \( z \) is any class (not necessarily
simple), and \( K > 0 \). If \( \ell(z|G) \geq C_nK\ell(z|H) \) then there is a class \( u \) in \( F \)
such that:

- \( \ell(u|H) < 2; \)
- \( d_F(H, u) \) is bounded; and
- \( u|G \) and \( z|G \) share a segment of length \( K \).

In particular, \( u \) is simple.

If \( \ell(z|H) < 2 \) then \( u = z \) works, i.e. \( z \) satisfies the conclusions of the
lemma. So, in the proof of Proposition 6.13 we assume \( \ell(z|H) \geq 2. \)
Fix a map $\phi : H \to G$ so that each edge is immersed (or collapsed) and each vertex has at least two gates (e.g. first change the metric on $H$ as in Proposition 2.7 so that the tension graph is all of $H$, but in the rest of the proof we use the original metric).

In the proof we will not keep track of exact constants, but will talk about “long segments in common with $z|G$”. For example, suppose an edge path $A$ in $H$ is the concatenation $A = BC$ of two sub-edge paths, and we look at $[\phi(A)]$, which is the tightening of the composition $[\phi(B)][\phi(C)]$. If $[\phi(A)]$ contains a long segment in common with $z|G$, then so does $[\phi(B)]$ or $[\phi(C)]$ (or both), but “long” in the conclusion means about a half of “long” in the assumption. The number of times this argument takes place will be bounded, and at the end the length can be taken as large as we want by choosing the original length (i.e. $C_n$) large.

Represent $z|H$ as a composition of $\sim \ell(z|H)$ paths $z_i$ where each $z_i$ is either an edge, or a combinatorially long (but of length $\leq 1$) path contained in the thin subgraph (union of immersed loops of small length). Thus the loop $z|G$ is obtained by tightening the composition of the paths $[\phi(z_i)]$. In the process of tightening, everything must cancel except for a (possibly degenerate) segment $\sigma_i \subset [\phi(z_i)]$ in each path, and at least one $\sigma_i$ must have length $\geq \sim C_n K$. So we conclude that, for some $z_i$, $[\phi(z_i)]$ contains a long segment in common with $z|G$. There are now two cases, depending on whether $z_i$ is contained in the thin part or is an edge. Lemmas 6.14 and 6.15 will prove in turn that in each case the conclusions of the proposition hold. After proving these lemmas, we will have completed the proof of Proposition 6.13.

**Lemma 6.14.** Assume that, in addition to the hypotheses of Proposition 6.13, there is an edge $e$ so that $\phi(e)$ contains a segment of length $\sim C_n K$ in common with $z|G$. Then the conclusions of Proposition 6.13 hold.

**Proof.** Start extending $e$ to a legal edge path until an edge is repeated. There are several possibilities.

*Type 0.* The first repetition is $e$ itself, i.e. we have $e...e$. Then identifying the $e$’s gives a legal loop $u$ that crosses each edge at most once, and $u$ works.

*Type 1.* The first repetition is either $e^{-1}$ or another edge with reversed orientation, i.e. $e...e^{-1}$ or $e..a..a^{-1}$. Schematically we picture this as a monogon. Note that there are two ways to traverse the monogon starting with $e$ and ending with $e^{-1}$, both legal.

*Type 2.* The first repetition is an edge $a$ different from $e$ and with the same orientation, i.e. $e..a..a$. We picture this as a spiral.
A monogon or spiral has its tail and its loop. In $e..e^{-1}$ the tail is $e$ and the loop is represented by the edge path between $e$ and $e^{-1}$; in $e..a..a^{-1}$, the tail is $e..a$ and the loop is represented by the edge path between $a$ and $a^{-1}$; and in $e..a..a$ the tail is $e..a$ and the loop is represented by the edge path between the $a$'s.

We can also extend $e$ in the opposite direction until an edge repeats, and so we have three subcases.

Subcase 1. Type 1-1, i.e. we have Type 1 on both sides. Here we have a morphism to $H$ from a graph $Y$ as in Figure 10 whose induced illegal turns form a subset of those indicated.

![Figure 10: Type 1-1.](image)

If there is an edge $b$ (different from $e$) crossed by both monogons then by switching from one copy of $b$ to the other we may form a legal loop $u$ that crosses $e$ once and all other edges at most twice. Indeed, if at least one copy of $b$ is in a loop of a monogon, then there is a legal segment in $Y$ of the form $b..e..b$ that crosses $e$, $b$, and $b^{-1}$ only as indicated. If both copies of $b$ are in tails, then there is either $b..e..b$ as above or $b..e..b^{-1}..b$. In the latter case, choose the first $b$ as close as possible to $e$ (to guarantee all edges in our legal loop are crossed at most twice), and this loop works. If there is no such $b$, the loop $u$ that traverses both monogons once is legal, crosses each edge at most twice, and crosses some edge once, so $u$ works.

Subcase 2. Type 1-2, i.e. we have a monogon on one side and a spiral on the other.

If some edge $b$ different from $e$ is crossed by both the spiral and the monogon, we can form a legal loop $u$ that crosses $e$ once as in Subcase 1, and we are done. Otherwise, let $u$ be the loop that crosses both the spiral and the monogon, so it has (potentially) one illegal turn. We claim that either the loop $v$ of the spiral or $u$ works. Indeed, $u$ crosses some edge once and all edges at most twice. Write $u$ as $Ev'E^{-1}v^{-1}$, where $E$ is the edge-path formed by the two tails and $v'$ is the loop of the monogon. Schematically, $u$ can be drawn as in Figure 12.
Now consider the image $\phi(u)$. To see how much cancellation occurs, let $Z$ be the maximal common initial segment of $[\phi(E)]$ and $[\phi(vE)]$. By Lemma 6.12(2), $Z$ has the form $[\phi(v)]^NW'$ for some initial segment $W'$ of $[\phi(v)]$. If $Z$ shares a long segment with $z|G$ then $v$ works. Otherwise, $u|G$ shares a long segment with $z|G$, and so works.

Subcase 3. Type 2-2, i.e. we have two spirals.

The first case is that the two spirals do not contain any edges in common except for $e$. Then the loop $u$ that crosses both spirals can be written in the form $u = Ev'E^{-1}v^{-1}$, pictured as a bigon (see Figure 14). The argument is now similar to Subcase 2. Consider the maximal common initial segment $V$ (resp. $V'$) of $\phi(E)$ and $[\phi(vE)]$ (resp. $\phi(E^{-1})$ and $[\phi(v'E^{-1})]$). If either $V$ or $V'$ shares a long segment with $z|G$ then $v$ or $v'$ works. Otherwise, $u$ works.
The second case is that some edge $b$, other than $e$, occurs on both spirals and we can construct a loop that crosses $e$ only once by jumping from one $b$ to the other. If this loop is legal, we can take it for $u$. Otherwise, it has one illegal turn, and there are two possibilities.

Suppose first that $b$ in the tail of a spiral whose loop is $v$. There is then a segment of the form $b..e..b^{-1}..v..b$ where the only illegal turn is the initial point of $v$. Let $u$ be the loop obtained by identifying the first and last $b$’s. Exactly as in Subcase 2, either the cancelling segments of $\phi(u)$ share a long segment with $z|G$ (in which case $v$ works) or else $u|G$ shares a long segment with $z|G$.

Secondly, suppose that $b$ appears in the loop $v$ of a spiral. Here there is a segment of the form $b..e..w$ where $w$ is an initial segment of $v$ ending with $b$ and the only illegal turn is the initial vertex of $w$. Let $u$ be the loop obtained by identifying the $b$’s in $b..e..w$. Let $u'$ be the loop obtained by identifying the first and last $b$’s of $b..e..vw$. By Lemma 6.12(1), either the cancelling segments in one of $\phi(u)$ and $\phi(u')$ shares a long segment with $z|G$ (in which case $v$ works) or else one of $u|G$ and $u'|G$ shares a long segment with $z|G$.

Finally, by choosing the first $b$ of our segment as close as possible to $e$, we guarantee that the loop we produce crosses each edge at most twice and some edge once and so works.

**Lemma 6.15.** Suppose that, in addition to the hypotheses of Proposition 6.13,
there is no edge as in Case 1, but there is a path $w$ in the thin part of $H$ of length $\leq 1$ such that $[\phi(w)]$ contains a segment of length $\sim C_nK$ in common with $z|G$. Then the conclusions of Proposition 6.13 hold.

Proof. First, if necessary, concatenate $w$ with a combinatorially bounded path in the thin part so that its endpoints coincide. After this operation $[\phi(w)]$ still has a long piece in common with $z$ since $\phi$-images of edges do not. If taking $u = w$ does not work, then the path $[\phi(w)]$ has the form $VUV^{-1}$ with $V$ having a long piece in common with $z|G$. Choose a combinatorially short loop $a$ in the thin part, based at the endpoints of $w$. We aim to show that either $a$ or $aw$ works.

Write $[\phi(a)] = ABA^{-1}$ so that $B = [[B]]$. The loop $[[\phi(aw)]]$ is obtained by tightening the loop $ABA^{-1}UVU^{-1}$ which has at most two illegal turns (the two occurrences of $\{A,U\}$). See Figure 15.

![Figure 15: $\phi(aw)$](image)

There are two cases. If the maximal common initial segment of $A$ and $U$ is a proper segment of $A$ then our loop becomes immersed after cancelling the copies of this common initial segment at the two illegal turns. Here $aw$ works.

If $U = AU'$, then after cancelling the common $A$’s, we are left with $BU'VU'^{-1}$. Since the initial and terminal directions of $B$ are distinct, one of the turns $\{B,U\}$ and $\{B^{-1},U\}$ is legal. Using Lemma 6.12[2] again, either a power of $B$ has a long segment in common with $U$ (in which case $a$ works) or not (in which case $aw$ works).

We have completed the proof of Proposition 6.13.

Lemma 6.16. Let $G_t$, $t \in [0,L]$, be a folding path in $\mathcal{X}$ parametrized by arclength, $H \in \mathcal{X}$, and $z$ a class in $\mathcal{F}$. For some $\tau \in [0,L]$, assume that $\ell(z|G_\tau) \geq \ell(z|H)$. 

52
(i) If \( z \mid G \tau \) is legal then either \( \text{left}(H) \leq \tau \) or \( \text{d}_{F}(\text{Left}(H), G \tau) \) is bounded.
(Recall Definition 6.6.)

(i') If \( z \mid G \tau \) has no immersed illegal segment of length \( I \) then either \( \text{left}(H) \leq \tau \) or \( \text{d}_{F}(\text{Left}(H), G \tau) \) is bounded.

(ii) If \( z \) is simple and \( z \mid G \tau \) is illegal then either \( \text{d}_{F}(\text{Right}(H), G \tau) \) is bounded or \( \text{right}(H) \geq \tau \).

**Proof.** We first prove (i). Let \( C_{n} \) be the constant from Proposition 6.13 (closing up to a simple class). Since \( z \mid G \tau \) is legal, by replacing \( \tau \) by \( \tau + e^{3C_{n}} \) we may assume that \( \ell(z \mid G \tau) \geq 3C_{n} \ell(z \mid H) \). Proposition 6.13 then provides a simple class \( u \) with \( \ell(u \mid H) < 2 \) and \( \text{d}_{F}(H, u) \) bounded so that \( u \mid G \tau \) has a segment of length 3 in common with \( z \mid G \tau \); in particular \( u \mid G \tau \) contains a legal segment of length 3 and so \( \text{left}_{G \tau}(u) \leq \tau \). We are done by Lemma 3.3 and Corollary 6.5.

The proof of (i') is similar. Because \( z \mid G \tau \) has no illegal segment of length \( I \), its length grows at an exponential rate under folding. Also, the property of not having an illegal segment of length \( I \) is stable under folding. Therefore we may assume that \( \ell(z \mid G \tau) \geq C_{n}I \ell(z \mid H) \) and conclude that \( u \mid G \tau \) produced by Proposition 6.13 shares a length \( I \) segment with \( z \mid G \tau \) and so has a legal segment of length 3.

The proof of (ii) is also analogous. Using Lemma 5.8 by moving left from \( G \tau \) a bounded amount in \( F \) we may assume \( \ell(z \mid G \tau) \geq C_{n}I \ell(z \mid H) \). Then Proposition 6.13 provides a simple class \( u \) with \( \ell(u \mid H) < 2 \) and \( \text{d}_{F}(H, u) \) bounded so that \( u \mid G \tau \) has a segment of length \( I \) in common with \( z \mid G \tau \). In particular, this segment is illegal.

We summarize the conclusions in Lemma 6.16(i) and (i') [resp. (ii)] by saying that the projection of \( H \) to \( G_{i} \) is coarsely to the left [resp. right] of \( G_{\tau} \), measured in \( F \). (Recall Corollary 6.11)

**7 Hyperbolicity**

The following proposition provides a blueprint for proving hyperbolicity of \( F \). In the case of the curve complex the same blueprint was used by Bowditch.

**Proposition 7.1.** \( F \) is hyperbolic if and only if the following holds for projections of folding paths. There is \( C > 0 \) so that:
(i) (Fellow Travel) Any two projections $\pi(G_t)$ and $\pi(H_t)$ of folding paths that start and end “at distance 1” (coarsely interpreted) are in each other’s Hausdorff $C$-neighborhood.

(ii) (Symmetry) If $\pi(G_t)$ goes from $A$ to $B$ and $\pi(H_t)$ from $B$ to $A$ then the two projections are in each other’s Hausdorff $C$-neighborhood.

(iii) (Thin Triangles) Any triangle formed by projections of three folding paths is $C$-thin.

More precisely, in (i), if $G$ and $H$ are the initial points of the two paths, the hypothesis means that there exist adjacent free factors $A$ and $B$ such that $A \in \pi(G)$ and $B \in \pi(H)$, and similarly for terminal points.

**Proof.** It is clear that (i)-(iii) are necessary for hyperbolicity, since projections of folding paths form a uniform collection of reparametrized quasi-geodesics and in hyperbolic spaces quasi-geodesics stay in bounded neighborhoods of geodesics. The converse is due to Bowditch [7] (a variant was used earlier by Masur-Minsky [26]). Here is a sketch. We will show that any loop $\alpha$ in $F$ of length $L$ bounds a disk of area $\sim L \log L$. (Think of bounded length loops as bounding disks of area 1.) Subdivide $\alpha$ into $3 \times 2^N$ segments of size $\sim 1$ and think of it as a polygon. Subdivide it into triangles in a standard way: a big triangle in the middle with vertices $2^N$ segments apart, then iteratively bisect remaining polygonal paths. Represent each diagonal by the image of a folding path — up to bounded Hausdorff distance the choices are irrelevant. Using Thin Triangles, each triangle of diameter $D$ can be filled with a disk of area $\sim D$. Adding the areas of all the triangles gives $\sim N \times 2^N \sim L \log L$. 

Proposition 7.2 generalizes Algom-Kfir’s result [1].

**Proposition 7.2.** Let $H, H' \in X$ with $d_X(H, H') \leq M$ and let $G_t$ be a folding path such that $d_X(H, G_t) \geq M$ for all $t$. Then the distance between the projections of $H$ and $H'$ to $G_t$ is uniformly bounded in $F$.

**Proof.** Denote by $G_1$ the leftmost of left$\gamma_t(z)$ and by $G_2$ the rightmost of right$\gamma_t(z)$ as $z|H$ varies over candidates in $H$. Then the interval $[G_1, G_2]$ is bounded in $F$ by Lemma 5.3 Proposition 6.4 and Corollary 6.11. Let $z_1|H$ be a candidate that realizes the distance to $G_1$, so $\ell(z_1|G_1) \geq e^M \ell(z_1|H)$ and $\ell(z_1|H') \leq e^M \ell(z_1|H)$. Combining these inequalities gives $\ell(z_1|G_1) \geq \ell(z_1|H')$, so Lemma 6.16(ii) shows Right$\gamma_t(H')$ is coarsely to the right of $G_1$, measured in $F$. In the same way one argues that Left$\gamma_t(H')$ is coarsely to the left of $G_2$, measured in $F$. The claim follows. 

54
Corollary 7.3. A folding line that makes \((\kappa, C)\)-definite progress in \(\mathcal{F}\) is strongly contracting in \(\mathcal{X}\) (with the constants depending on \(\kappa\) and \(C\)). \(\square\)

This simply means that, in the situation of Proposition 7.2, projections of \(H\) and \(H'\) to \(G_t\) are at a uniformly bounded distance in \(\mathcal{X}\) (depending on \(\kappa\) and \(C\)), measured from left to right.

Note that a folding line that makes definite progress in \(\mathcal{F}\) is necessarily in a thick part of \(\mathcal{X}\) (i.e. the injectivity radius of \(G_t\) is bounded below by a positive constant). The converse does not hold (but recall that it does hold in Teichmüller space, and Corollary 7.3 is the direct analog of Minsky’s theorem [27] that Teichmüller geodesics in the thick part are strongly contracting).

Remark 7.4. One can avoid the use of the technical Proposition 5.4(surviving illegal turns) in the proof of Proposition 7.2.

8 Fellow Travelers and Symmetry

We will fix constants \(C_1, C_2,\) and \(D\) from Proposition 6.4 and Proposition 7.2, so that:

- If \(B < A\) are free factors at \(\mathcal{F}\)-distance \(\geq C_1\) from a folding path \(G_t\) then the \(\mathcal{X}\)-distance along \(G_t\) between left\((A)\) and left\((B)\) is bounded by \(D\),

- The \(\mathcal{F}\)-diameter of the projection of a path of length \(M\) to any folding path at distance \(\geq M\) is always \(\leq C_2\),

- \(C_1 > C_2\).

Proposition 8.1. Fix \(C\) sufficiently large. Suppose \(G_t\) and \(H_\tau\) are two folding paths, \(d_\mathcal{F}(G_t, H_\tau) \geq C\) for all \(t, \tau\), but the initial points and the terminal points are at \(\mathcal{F}\)-distance \(\leq 10C\). Then the projections of the two paths to \(\mathcal{F}\) are uniformly bounded in diameter.

The same holds if the initial point of \(G_t\) is \(10C\)-close to the terminal point of \(H_\tau\) and the terminal point of \(G_t\) is \(10C\)-close to the initial point of \(H_\tau\).

Proof. Subdivide \(H_\tau\) into a minimal number of segments whose \(\mathcal{F}\)-diameter is bounded by \(C_1\). Say the subdivision points are \(s_0 < s_1 < s_2 < \cdots < s_m\). Let \(G_{t_i} = \text{left}_{G_t}(H_{s_i})\). When \(C > 2C_1\) we have that the distance, measured from left to right, between \(G_{t_i}\) and \(G_{t_{i+1}}\) along \(G_t\) is \(\leq C_1 D\) (here \(C > 2C_1\) is needed so that interpolating free factors are also far from \(G_t\)). The \(\mathcal{F}\)-distance between \(G_{t_0}\) and the initial point of \(G_t\), and also between \(G_{t_m}\)
and the terminal point of $G_t$ is bounded (recall Corollary 6.5). Further, $d_X(G_{t_0}, G_{t_m}) \leq mC_1D$ as long as $G_{t_0}$ is to the left of $G_{t_m}$ (if it is to the right, the whole path $G_t$ is $F$-bounded). So the projection of $[G_{t_0}, G_{t_m}]$ to $H_\tau$ is bounded by $mC_2$, as long as the $X$-distance between $G_t$ and $H_\tau$ is bounded below by $C_1D$ (if not then the $F$-distance between $G_t$ and $H_\tau$ is bounded, contradiction when $C$ is sufficiently large, see Corollary 3.5). So,

$$mC_2 + 2(\text{const}) \geq (m - 1)C_1$$

and since $C_1 > C_2$ this implies that $m$ is bounded above. The claim follows.

The proof is analogous in the “anti-parallel” case.

Fellow Traveler and Symmetry properties are now immediate.

**Proposition 8.2.** Let $G_t$ and $H_\tau$ be folding paths whose initial points are at $F$-distance $\leq R$ and the same holds for terminal points. Then $\pi(G_t) \subset F$ and $\pi(H_\tau) \subset F$ are in each other’s bounded Hausdorff neighborhoods, the bound depending only on $R$.

The same holds when the initial point of $\pi(G_t)$ is $R$-close to the terminal point of $\pi(H_\tau)$ and the terminal point of $G_t$ is $R$-close to the initial point of $H_\tau$.

**Proof.** Let $C > R$ be a sufficiently large constant as in Proposition 8.1. If $\pi(G_t)$ is not contained in the Hausdorff $C$-neighborhood of $\pi(H_\tau)$ there is a subpath $[G_{t_1}, G_{t_2}]$ such that no point of it is $C$-close to $\pi(H_\tau)$, but the endpoints $G_{t_1}, G_{t_2}$ are within $10C$. Then there is a subpath $[H_{\tau_1}, H_{\tau_2}]$ of $H_\tau$ whose endpoints are within $10C$ of the endpoints of $[G_{t_1}, G_{t_2}]$ (but notice that we don’t know in advance if the orientations are parallel or anti-parallel). Now by Proposition 8.1 the set $\pi([G_{t_1}, G_{t_2}])$ is in a bounded Hausdorff neighborhood of $\pi(H_\tau)$. By the same argument $\pi(H_\tau)$ is contained in a bounded Hausdorff neighborhood of $\pi(G_t)$.

**Remark 8.3.** Note that, in the situation of Proposition 8.2 any $G_{t_0}$ is bounded $F$-distance from its projection to $H_\tau$. Indeed, if $H_{\tau_0}$ is bounded $F$-distance from $G_{t_0}$ then from Corollary 6.5 it follows that the projection of $G_{t_0}$ is bounded $F$-distance from $H_{\tau_0}$.

**Proposition 8.4.** Let $G_t$ and $H_\tau$ be folding paths whose initial points are at $F$-distance $\leq R$ and the same holds for terminal points. There is a uniform bound, depending only on $R$, to $d_F(\text{Left}_{G_t}(A), \text{Left}_{H_\tau}(A))$ for any free factor $A$.

The same holds in the anti-parallel case.
Figure 16: The “ahead” and the “behind” cases.

Proof. For notational simplicity, we may assume that $A = \langle a \rangle$ is cyclic. First suppose $G_t$ and $H_\tau$ are parallel. Modulo interchanging the two paths, we can assume that the projection of $\text{Left}_{G_t}(a)$ to $H_\tau$ is “ahead” of $\text{Left}_{H_\tau}(a)$ (i.e., when $\text{Left}_{G_t}(a)$ is projected to $H_\tau$, it is coarsely right of $\text{Left}_{H_\tau}(a)$ measured in $\mathcal{F}$) and the projection of $\text{Left}_{H_\tau}(a)$ to $G_t$ is “behind” (defined analogously) $\text{Left}_{G_t}(a)$. By Lemma 6.16 if $\ell(a|\text{Left}_{G_t}(a)) \leq \ell(a|\text{Right}_{H_\tau}(a))$ then the projection of $\text{Left}_{G_t}(a)$ to $H_\tau$ is behind $\text{Right}_{H_\tau}(a)$, and the claim is proved. If $\ell(a|\text{Left}_{G_t}(a)) \geq \ell(a|\text{Right}_{H_\tau}(a))$ then the projection of $\text{Right}_{H_\tau}(a)$ to $G_t$ is ahead of $\text{Left}_{G_t}(a)$, and we are done again.

Now suppose $G_t$ and $H_\tau$ are anti-parallel. There are two subcases to consider, see Figure 16. In the first case $\text{Left}_{G_t}(A)$ is ahead the projection to $G_t$ of $\text{Left}_{H_\tau}(A)$. This is a symmetric condition with respect to interchanging $G_t$ and $H_\tau$ (by Remark 8.3). Say $\ell(a|\text{Left}_{G_t}(a)) \leq \ell(a|\text{Left}_{H_\tau}(a))$. Then the projection of $\text{Left}_{G_t}(a)$ to $H_\tau$ is ahead of $\text{Left}_{H_\tau}(a)$ by Lemma 6.16 and the claim follows.

The “behind” case is similar, but we consider right projections. Say $\ell(a|\text{Right}_{G_t}(a)) \leq \ell(a|\text{Right}_{H_\tau}(a))$. Then the projection of $\text{Right}_{G_t}(a)$ to $H_\tau$ is behind $\text{Right}_{H_\tau}(a)$, and the claim again follows.

9 Thin Triangles

Proposition 9.1. Triangles in $\mathcal{F}$ made of images of folding paths are uniformly thin. More precisely, if $A, B, C$ are three free factors coarsely joined by images of folding paths $AB$, $AC$, $BC$ and $\hat{C}$ is the projection of $C$ to $AB$, then $A\hat{C}$ is in a bounded Hausdorff neighborhood of $AC$ and $B\hat{C}$ is in a bounded Hausdorff neighborhood of $BC$.

We will consider points $U, V, W \in \mathcal{X}$ and folding paths $H_t$ from $V$ to $W$
and $G_t$ from $U$ to $W$. Denote by $P$ the rightmost of $\text{Right}_{G_t}(z)$ as $z$ ranges over candidates in $V$. By Lemma 3.3, $P$ has bounded $\mathcal{F}$-distance from the projection of $V$ to $G_t$. See Figure 17.

![Figure 17: A thin triangle.](image)

We will prove that $\pi(VP) \cup \pi(PW)$ and $\pi(VW)$ are contained in uniform Hausdorff neighborhoods of each other (by $VP$ we mean a folding path from $V$ to $P$ etc). The basic idea is that $VP \cup PW$ behaves like a folding path and the claim is an instance of the Fellow Traveler Property.

**Claim 1.** $d_X(V, W) \geq d_X(V, P) + d_X(P, W) - C$ for a universal constant $C$.

To prove the claim, let $v|V$ be a candidate for $d_X(V, P)$. By definition of $P$, $v|P$ has only bounded length illegal subsegments. It follows that after removing the 1-neighborhood of each illegal turn, a definite percentage of the length of $v|P$ remains, and hence

$$\ell(v|W) \geq e^{d_X(P, W)} \epsilon \ell(v|P) = e^{d_X(V, P) + d_X(P, W)} \epsilon \ell(v|V)$$

for a fixed $\epsilon > 0$. Thus

$$d_X(V, W) \geq \log \frac{\ell(v|W)}{\ell(v|V)} \geq d_X(V, P) + d_X(P, W) + \log \epsilon$$
By $Q$ denote the point on $VW$ such that $d_X(V, Q) = d_X(V, P)$. From Claim 1 we see that $d_X(Q, W) \leq d_X(P, W) \leq d_X(Q, W) + C$.

**Claim 2.** $d_F(P, Q)$ is bounded.

This time let $v|V$ be a candidate for $d_X(V, W)$. Thus

$$\ell(v|Q) = e^{d_X(V, Q)} \ell(v|V) = e^{d_X(V, P)} \ell(v|V) \geq \ell(v|P)$$

Let $Q'$ be the point along $QW$ with $d_X(Q, Q') = \log(3(2n - 1)C_n)$ (if such a point does not exist, both $P$ and $Q$ are uniformly close to $W$ and we are done). Then $\ell(v|Q') = 3(2n - 1)C_n \ell(v|Q) \geq 3(2n - 1)C_n \ell(v|P)$, so Proposition 6.13 (closing up to a simple class) implies that there is a simple class $p$ such that $p|P$ is of length $< 2$ and $p|Q'$ has a legal segment of length $3(2n - 1)$. Now $p|Q'$ cannot contain many disjoint legal segments of length 3, for otherwise $p$ would grow to a much longer length along $Q'W$ than along $PW$. Now Claim 2 follows from Lemma 6.2.

**Proof of Proposition 9.1.** By Proposition 8.2 we are free to replace a folding path by another whose endpoints project close to the endpoints of the original, and we are allowed to reverse orientations. By Proposition 8.4 these replacements affect the projections by a bounded amount, so that the projection $\hat{C}$ of $C$ to $AB$ is coarsely well-defined, independently of the choice of a folding path whose projection coarsely connects $A$ and $B$ with either orientation. In particular, we may assume that $U, V, W$ project near $A, C, B$ respectively and we may consider folding paths $UW$ and $VW$ that end at the same graph $W$. The above discussion then shows that $B\hat{C}$ is contained in a bounded Hausdorff neighborhood of $BC$. Making analogous choices of the folding paths, the claim about $A\hat{C}$ follows similarly.

The following fact shows that for the purposes of this paper the collection of folding paths can be replaced by the larger collection of geodesic paths.

**Proposition 9.2.** Let $V, P, W \in \mathcal{X}$ so that $d_X(V, P) + d_X(P, W) = d_X(V, W)$ and let $V' \in \mathcal{X}$ be such that $d_X(V, V') + d_X(V', W) = d_X(V, W)$, $d_F(V, V')$ is bounded, and there is a folding path $G_t$ from $V'$ to $W$ (see Proposition 2.7). Then the $F$-distance between $P$ and some $G_t$ is uniformly bounded. Moreover, if $H_\tau$ is a geodesic in $\mathcal{X}$ from $V$ to $P$, then the induced correspondence $\tau \mapsto t$ can be taken to be monotonic with respect to $\tau$. Consequently, the set of projections to $\mathcal{F}$ of geodesics in $\mathcal{X}$ is a uniform collection of reparameterized quasi-geodesics in $\mathcal{F}$.

The proof is a variant of the discussion above.
Proof. Let $P'$ be the point on $G_t$ with $d_X(V,P') = d_X(V,P)$ (if such a point does not exist we assign $V'$ to $P$). We need to argue that $d_F(P,P')$ is bounded. Let $v|V$ be a candidate realizing $d_F(V,W)$. Thus $v|G_t$ is legal for all $t$ and $\ell(v|P) = \ell(v|P')$. Let $Q'$ be a point on $G_t$ with $d_X(P',Q') = \log(3(2n - 1)C_n)$ (if such a point does not exist, both $P$ and $P'$ are close to $W$ and we are done). Then $\ell(v|Q') = 3(2n - 1)C_n \ell(v|P)$, so Proposition 6.13 (closing up to a simple class) implies that there is a simple class $p$ such that $\ell(p|P) < 2$ and with $p|Q'$ containing a legal segment of length $3(2n - 1)$. Now $p|Q'$ cannot contain many disjoint legal segments of length 3, for otherwise $p$ would grow to a much longer length along $Q'W$ than along $PW$. Lemma 6.2 implies that $d_F(P,Q')$, and therefore $d_F(P,P')$, is bounded. \hfill \Box

Theorem 9.3. $F$ is $\delta$-hyperbolic. Images of geodesic paths in $X$ are in uniform Hausdorff neighborhoods of geodesics with the same endpoints.

Furthermore, an element of $Out(F)$ has positive translation length in $F$ if and only if it is fully irreducible. The action of $Out(F)$ on $F$ satisfies Weak Proper Discontinuity (see \cite{6}).

Proof. Since we have checked Conditions (i), (ii), and (iii) in Proposition 6.1, $F$ is hyperbolic. The second statement is a consequence of Proposition 9.2 and last the two statements follow from:

- There are coarsely well-defined, Lipschitz maps from $F$ to the hyperbolic complexes $X$ constructed in \cite{5} Sections 4.4.1 and 4.4.2.

- Given a fully irreducible element $f$ of $Out(F_n)$, there is an $X$ on which $f$ has positive translation length (\cite{5} Main Theorem).

- Further, for every $x \in X$ and every $C > 0$ there is $N > 0$ such that $\{g \in Out(F) \mid d_X(x,xg) \leq C, d_X(xf^N,xf^Ng) \leq C\}$ is finite (\cite{5} Section 4.5).

\hfill \Box

References

[1] Yael Algom-Kfir. Strongly contracting geodesics in outer space. Geom. Topol., 15(4):2181–2233, 2011.

[2] Yael Algom-Kfir and Mladen Bestvina. Asymmetry of outer space. Geom. Dedicata, 156:81–92, 2012.
[3] M. Bestvina, M. Feighn, and M. Handel. Laminations, trees, and irreducible automorphisms of free groups. *Geom. Funct. Anal.*, 7(2):215–244, 1997.

[4] Mladen Bestvina. A Bers-like proof of the existence of train tracks for free group automorphisms. *Fund. Math.*, 214(1):1–12, 2011.

[5] Mladen Bestvina and Mark Feighn. A hyperbolic Out($F_n$)-complex. *Groups Geom. Dyn.*, 4(1):31–58, 2010.

[6] Mladen Bestvina and Koji Fujiwara. Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.*, 6:69–89 (electronic), 2002.

[7] Brian H. Bowditch. Intersection numbers and the hyperbolicity of the curve complex. *J. Reine Angew. Math.*, 598:105–129, 2006.

[8] Matt Clay. Contractibility of deformation spaces of $G$-trees. *Algebr. Geom. Topol.*, 5:1481–1503 (electronic), 2005.

[9] Marc Culler and Karen Vogtmann. Moduli of graphs and automorphisms of free groups. *Invent. Math.*, 84(1):91–119, 1986.

[10] Matthew B. Day and Andrew Putman. The complex of partial bases for $F_n$ and finite generation of the Torelli subgroup of Aut($F_n$). arXiv:1012.1914.

[11] Stefano Francaviglia and Armando Martino. Metric properties of Outer space. *Pub. Mat.*, 55:433–473, 2011.

[12] Vincent Guirardel and Gilbert Levitt. Deformation spaces of trees. *Groups Geom. Dyn.*, 1(2):135–181, 2007.

[13] Michael Handel and Lee Mosher. The free splitting complex of a free group I: Hyperbolicity. arXiv:1111.1994.

[14] Michael Handel and Lee Mosher. Subgroup classification in Out($F_n$). arXiv:0908.1255.

[15] John L. Harer. Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math. (2)*, 121(2):215–249, 1985.

[16] John L. Harer. The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.*, 84(1):157–176, 1986.
[17] W. J. Harvey. Boundary structure of the modular group. In Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), volume 97 of Ann. of Math. Stud., pages 245–251. Princeton Univ. Press, Princeton, N.J., 1981.

[18] Allan Hatcher, Karen Vogtmann, and Natalie Wahl. Erratum to: “Homology stability for outer automorphism groups of free groups [Algebr. Geom. Topol. 4 (2004), 1253–1272 (electronic) by Hatcher and Vogtmann. Algebr. Geom. Topol., 6:573–579 (electronic), 2006.

[19] Allen Hatcher and Karen Vogtmann. Cerf theory for graphs. J. London Math. Soc. (2), 58(3):633–655, 1998.

[20] Allen Hatcher and Karen Vogtmann. The complex of free factors of a free group. Quart. J. Math. Oxford Ser. (2), 49(196):459–468, 1998.

[21] Allen Hatcher and Karen Vogtmann. Homology stability for outer automorphism groups of free groups. Algebr. Geom. Topol., 4:1253–1272, 2004.

[22] Arnaud Hilion and Camille Horbez. The hyperbolicity of the sphere complex via surgery paths. arXiv:1210.6183.

[23] Ilya Kapovich and Martin Lustig. Geometric intersection number and analogues of the curve complex for free groups. Geom. Topol., 13(3):1805–1833, 2009.

[24] Ilya Kapovich and Kasra Rafi. On hyperbolicity of free splitting and free factor complexes. arXiv:1206.3626.

[25] Reiner Martin. Non-uniquely ergodic foliations of thin type, measured currents, and automorphisms of free groups. PhD thesis, UCLA, 1995.

[26] Howard A. Masur and Yair N. Minsky. Geometry of the complex of curves. I. Hyperbolicity. Invent. Math., 138(1):103–149, 1999.

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

[28] Richard K. Skora. Deformations of length functions in groups. preprint, 1989.

[29] John R. Stallings. Topology of finite graphs. Invent. Math., 71(3):551–565, 1983.
[30] John R. Stallings. Whitehead graphs on handlebodies. In \textit{Geometric group theory down under (Canberra, 1996)}, pages 317–330. de Gruyter, Berlin, 1999.

[31] J. H. C. Whitehead. On certain sets of elements in a free group. \textit{Proc. London Math. Soc.}, 41:48–56, 1936.