

A Bers-like proof of the existence of train tracks for free group automorphisms

by

Mladen Bestvina (Salt Lake City, UT)

To Mike Davis on the occasion of his 60th birthday

Abstract. Using Lipschitz distance on Outer space we give another proof of the train track theorem.

1. Introduction. An elegant proof of Thurston's classification of surface homeomorphisms [Thu88] was given by Bers [Ber78]. Given a surface homeomorphism Φ the proof proceeds by studying the associated displacement function $\tilde{\Phi}$ on Teichmüller space \mathcal{T} with respect to Teichmüller metric, i.e. the function

$$\tilde{\Phi}(x) = d(x, \Phi(x)).$$

There are three possibilities:

- (elliptic) $\inf \tilde{\Phi} = 0$ and the infimum is realized. In this case $\tilde{\Phi}$ has a fixed point and it is not hard to show that Φ is isotopic to a homeomorphism of finite order.
- (hyperbolic) $\inf \tilde{\Phi} > 0$ and the infimum is realized. In this case Bers proceeds to show that Φ is isotopic to a pseudo-Anosov homeomorphism.
- (parabolic) The infimum is not realized. In this case Bers shows that Φ is reducible.

In this paper we carry out Bers' argument in the context of $\text{Out}(F_n)$ and Outer space. The role of Teichmüller metric is played by the Lipschitz metric, a direct analog of Thurston's Lipschitz metric on Teichmüller space [Thu]. This metric is not symmetric and one must carefully choose the order

2010 *Mathematics Subject Classification*: Primary 20E05.

Key words and phrases: automorphisms of free groups, train track representative, Outer space.

of the two points when measuring distance. The result is an alternative proof of the train track theorem [BH92].

THEOREM. *Every irreducible automorphism Φ has a topological representative which is a train track map.*

2. Outer space and Lipschitz metric. In this section we review basic definitions and set some notation.

A *graph* Γ is a finite cell complex of dimension ≤ 1 . A *core graph* is a graph Γ with all vertices of valence ≥ 2 . A *rose* in rank n is the wedge R_n of n circles. A *marking* of a graph Γ is a homotopy equivalence $f : R_n \rightarrow \Gamma$. A *metric* on Γ is an assignment ℓ of positive lengths $\ell(e)$ to the edges e of Γ such that the sum is 1. If α is an immersed loop in Γ , we define the *length* of α with respect to the metric ℓ as the sum $\ell(\alpha)$ of the lengths of the edges of Γ crossed by α , with multiplicity. If α is not immersed, we define $\ell(\alpha)$ as the length of the immersed loop homotopic to α (or 0 if α is nullhomotopic). We may view Γ as a geodesic metric space with each edge e having length $\ell(e)$. A *direction* at $x \in \Gamma$ is a germ of isometric embedding $d : [0, \epsilon) \rightarrow \Gamma$ with $d(0) = x$. Thus most points of Γ have two directions, and the number of directions at a vertex is the valence. Directions can be viewed as analogs of unit tangent vectors. If $\phi : \Gamma \rightarrow \Gamma'$ is a map which is linear on edges and $\phi(x) = x'$, then ϕ induces a map ϕ_* from the set of directions at x to the set of directions at x' (unless the slope of ϕ is 0 on an edge containing x).

Recall that Culler–Vogtmann’s *Outer space* \mathcal{X}_n in rank n [CV86] is the set of equivalence classes of triples (Γ, f, ℓ) where

- Γ is a core graph,
- $f : R_n \rightarrow \Gamma$ is a marking, and
- ℓ is a metric on Γ .

Two such triples (Γ, f, ℓ) and (Γ', f', ℓ') are equivalent if there is a homeomorphism $\phi : \Gamma \rightarrow \Gamma'$ such that

- $\phi f \simeq f'$, and
- for all loops α in Γ , $\ell(\alpha) = \ell'(\phi(\alpha))$.

If Γ and Γ' have no vertices of valence 2, then ϕ must induce a bijection between the edges of Γ and Γ' , and $\ell(e) = \ell'(\phi(e))$ for every edge $e \subset \Gamma$. Every triple (Γ, f, ℓ) is equivalent (for $n \geq 2$) to some (Γ', f', ℓ') such that Γ' has no vertices of valence 2, by “unsubdividing” and assigning the sum of the lengths of the subdivision edges to the newly created edges.

There are three ways of defining the topology on \mathcal{X}_n , all yielding the same topology:

- \mathcal{X}_n can be decomposed into open simplices corresponding to fixing the marking and varying lengths on Γ . This gives \mathcal{X}_n the structure of a

“complex of simplices with missing faces”. Take the weak topology with respect to the collection of these simplices with missing faces.

- There is an embedding $\mathcal{X}_n \hookrightarrow \mathbb{R}^{\mathcal{S}}$ where \mathcal{S} is the set of nontrivial conjugacy classes in F_n , or equivalently, the set of immersed loops in R_n . The embedding is given by

$$[(\Gamma, f, \ell)] \mapsto (\alpha \mapsto \ell(f(\alpha))).$$

Now take the subspace topology.

- A neighborhood of $[(\Gamma, f, \ell)]$ is determined by $\epsilon > 0$ and consists of classes $[(\Gamma', f', \ell')]$ such that there is a map $\phi : \Gamma \rightarrow \Gamma'$ with $\phi f \simeq f'$ and such that ϕ is $<(1 + \epsilon)$ -Lipschitz.

The group $\text{Out}(F_n)$ acts on \mathcal{X}_n on the right as follows. Let $\Phi \in \text{Out}(F_n)$. We may view Φ as a homotopy equivalence (defined up to homotopy) $\Phi : R_n \rightarrow R_n$. Then

$$[(\Gamma, f, \ell)]\Phi = [(\Gamma, f\Phi, \ell)].$$

The third definition of the topology on \mathcal{X}_n can be promoted to a (non-symmetric) metric.

Let $[(\Gamma, f, \ell)], [(\Gamma', f', \ell')] \in \mathcal{X}_n$. Consider maps $\phi : \Gamma \rightarrow \Gamma'$ so that

- $\phi f \simeq f'$, and
- ϕ is linear on edges.

Call any map ϕ that satisfies these two conditions a *difference of markings*. Let $\sigma(\phi)$ denote the maximal slope of ϕ .

Observe that if α is any loop in Γ then

$$\ell'(\phi(\alpha)) \leq \sigma(\phi)\ell(\alpha).$$

The following result is due to Tad White (unpublished). A proof appears in [FM], but for completeness we include a proof.

PROPOSITION 2.1. *Let $\phi_0 : \Gamma \rightarrow \Gamma'$ be a difference of markings. Then*

$$\inf\{\sigma(\phi) \mid \phi \simeq \phi_0 : \Gamma \rightarrow \Gamma' \text{ is a difference of markings}\} = \sup_{\alpha} \frac{\ell'(\phi_0(\alpha))}{\ell(\alpha)}$$

and moreover both inf and sup are realized.

Proof. That $\inf \geq \sup$ follows from the observation just before the proposition. Arzelà–Ascoli implies that \inf is realized. Let $\phi : \Gamma \rightarrow \Gamma'$ be a difference of markings that realizes \inf . Denote by $\Delta = \Delta(\phi)$ the union of all edges of Γ on which ϕ has slope equal to $\sigma(\phi)$. We may also assume that Δ is minimal possible. Therefore Δ is a core graph (any edges with a valence 1 vertex can be removed by a small homotopy of ϕ , by moving the image of the valence 1 vertex in the direction that lowers the slope on the edge of Δ containing it).

Let v be a vertex of Δ and consider a *turn*, i.e. pair of distinct directions $\{d, d'\}$ out of v in Δ . We say that this turn is *legal* if $\phi_*(d) \neq \phi_*(d')$, and it is *illegal* if $\phi_*(d) = \phi_*(d')$. A path or a loop in Δ is *legal* if it crosses only legal turns. Note that the loop $\phi(\alpha)$ is immersed if and only if α is legal. Also observe that there is an equivalence relation on the set of directions (within Δ) out of each vertex $v \in \Delta$, where $d \sim d'$ if and only if $\phi_*(d) = \phi_*(d')$. A turn $\{d, d'\}$ is legal if and only if $d \approx d'$. We will call equivalence classes *gates*.

If Δ contains a vertex with only one gate, then a small homotopy as above would reduce Δ . Therefore there are at least two gates at each vertex. It follows that every legal edge path in Δ can be extended in a legal fashion. In particular, Δ admits a legal loop α . By construction $\ell'(\phi(\alpha)) = \sigma(\phi)\ell(\alpha)$, and thus α realizes the sup and equality holds as claimed. ■

REMARK 2.2. The legal loop α can be chosen to cross each edge of Δ at most twice. Therefore, the sup can be calculated by taking the maximum over a finite collection of loops in Γ . Furthermore, $\ell'(\phi_0(\alpha))$ does not depend on ϕ_0 , only on the homotopy class of ϕ_0 , so the quantity in the statement can be easily calculated.

To simplify notation, we will replace $[(\Gamma, f, \ell)]$ with Γ . Denote by $\sigma(\Gamma, \Gamma')$ the quantity in the statement of Proposition 2.1.

DEFINITION 2.3. Let $\Gamma, \Gamma' \in \mathcal{X}_n$. A difference of markings $\phi : \Gamma \rightarrow \Gamma'$ is *optimal* if $\sigma(\phi) = \sigma(\Gamma, \Gamma')$. The *tension (sub)graph* $\Delta = \Delta_\phi \subset \Gamma$ with respect to an optimal map ϕ is the union of the edges on which the slope of ϕ equals $\sigma(\phi)$.

The tension graph is equipped with a *train track structure* as in the proof of Proposition 2.1: a turn $\{d, d'\}$ is legal if $\phi_*(d) \neq \phi_*(d')$, and illegal otherwise. Directions (in Δ) at a vertex break up into equivalence classes, called *gates*, so that $\{d, d'\}$ is illegal if and only if d, d' are in the same gate.

DEFINITION 2.4. $d(\Gamma, \Gamma') = \log \sigma(\Gamma, \Gamma')$.

PROPOSITION 2.5.

- $d(\Gamma, \Gamma') \geq 0$ with equality only when $\Gamma = \Gamma'$.
- $d(\Gamma, \Gamma'') \leq d(\Gamma, \Gamma') + d(\Gamma', \Gamma'')$.
- $\text{Out}(F_n)$ acts on \mathcal{X}_n by isometries: $d(\Gamma \cdot \Phi, \Gamma' \cdot \Phi) = d(\Gamma, \Gamma')$.

Proof. If $\sigma(\Gamma, \Gamma') < 1$ then the volume of the image (i.e. the sum of the lengths of the images of edges) of an optimal map $\phi : \Gamma \rightarrow \Gamma'$ is < 1 , contradicting the fact that ϕ is surjective. Likewise, if $\sigma(\Gamma, \Gamma') = 1$ then ϕ must be an isometry.

The second statement follows by composing optimal maps and homotoping rel vertices to a map linear on edges. The third statement is obvious. ■

But note that in general $d(\Gamma, \Gamma') \neq d(\Gamma', \Gamma)$. See [AKB]. It can also be shown that $d : \mathcal{X}_n \times \mathcal{X}_n \rightarrow [0, \infty)$ is continuous.

3. Trichotomy. Fix an automorphism $\Phi \in \text{Out}(F_n)$ and consider the *displacement function*

$$\tilde{\Phi} : \mathcal{X}_n \rightarrow [0, \infty) \quad \text{given by} \quad \Gamma \mapsto d(\Gamma, \Gamma \cdot \Phi).$$

There are three possibilities:

- (elliptic) $\inf \tilde{\Phi} = 0$ and it is realized.
- (hyperbolic) $\inf \tilde{\Phi} > 0$ and it is realized.
- (parabolic) $\inf \tilde{\Phi}$ is not realized.

We consider these cases separately.

3.1. Φ is elliptic. An example is pictured below.

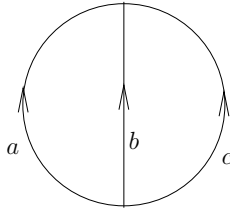


Fig. 1. $\Phi(a) = \bar{b}$, $\Phi(b) = \bar{c}$, $\Phi(c) = \bar{a}$. All three edges have length $1/3$, the tension graph is all of Γ and all nondegenerate turns are legal. Φ has order 6. Bar denotes the edge with opposite orientation.

Here we are assuming that Φ has a fixed point, i.e. $\Gamma \cdot \Phi = \Gamma$ for some $\Gamma \in \mathcal{X}_n$. If f is the marking of Γ , then $f\Phi \simeq \phi f$ for an isometry $\phi : \Gamma \rightarrow \Gamma$. Since (for $n \geq 2$) isometries of Γ have finite order, it follows that for some $k > 0$ we have $\phi^k = \text{id}$ and hence $f = \phi^k f \simeq f\Phi^k$, so Φ^k is (homotopic to) the identity, i.e. Φ has finite order.

3.2. Φ is hyperbolic. Suppose that $\inf \tilde{\Phi}$ is realized on $\Gamma \in \mathcal{X}_n$. Let $\log \lambda = d(\Gamma, \Gamma \cdot \Phi) = \inf \tilde{\Phi} > 0$, so $\lambda > 1$. Let $\Delta \subseteq \Gamma$ be the tension graph with its train track structure, with respect to an optimal map ϕ .

An example of a hyperbolic automorphism is given in Figure 2.

PROPOSITION 3.1. *After an arbitrarily small perturbation of Γ that preserves the condition that $d(\Gamma, \Gamma \cdot \Phi) = \log \lambda$ there is an optimal map $\phi : \Gamma \rightarrow \Gamma \cdot \Phi$ such that*

- $\phi(\Delta) \subseteq \Delta$,
- ϕ sends edges of Δ to legal paths, and
- ϕ_* sends legal turns to legal turns.

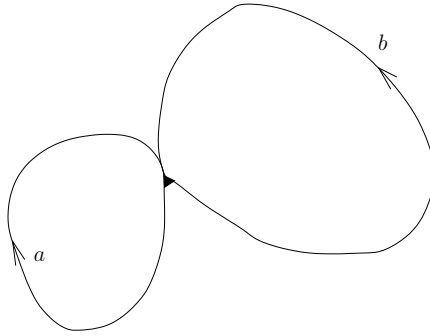


Fig. 2. $\Phi(a) = ab$, $\Phi(b) = bab$. Lengths of edges and $\sigma(\phi) = \lambda$ are computed from the equations $\ell(a) + \ell(b) = 1$, $\lambda\ell(a) = \ell(a) + \ell(b)$ and $\lambda\ell(b) = \ell(a) + 2\ell(b)$, i.e. $\lambda = (3 + \sqrt{5})/2$, $\ell(a) = (3 - \sqrt{5})/2$, $\ell(b) = (\sqrt{5} - 1)/2$. The tension graph is all of Γ . The gates are indicated by the little triangle at the vertex: they are $\{\bar{a}, \bar{b}\}$, $\{a\}$, $\{b\}$, where we adopt the convention that a represents the initial direction of the edge a , while \bar{a} represents the terminal direction of a , and similarly for b .

A 2-gate vertex v is allowed to be mapped to a non-vertex, with the directions in each gate mapping to one of two directions out of $\phi(v)$, i.e. the two directions out of a non-vertex are regarded as forming a legal turn.

Proof of Proposition 3.1. Fix an optimal map $\phi : \Gamma \rightarrow \Gamma \cdot \Phi$. The complexity of Δ is the pair $(\text{rank } H_1(\Delta), -\text{rank } H_0(\Delta))$. In the moves that follow the complexity is never increased, and is often decreased. Note that if Δ is a core graph and $\Delta' \subset \Delta$ is a proper core subgraph, then the complexity of Δ' is strictly smaller than that of Δ .

Suppose that all vertices of Δ have ≥ 2 gates (in particular, Δ is a core graph) and $\phi(\Delta) \not\subseteq \Delta$, and let e be an edge of Δ with $\phi(e) \not\subseteq \Delta$. Perturb the metric on Γ by scaling by $\mu > 1$ on the edges of Δ and scaling down on the edges in the complement of Δ , maintaining volume 1. Denote the new metric graph by Γ' . The tension graph Δ' for the new pair $\Gamma' \rightarrow \Gamma' \cdot \Phi$ (using the same map made linear on edges) is contained in Δ and does not contain e . Note that the slope on some edge must be λ by the minimality assumption. Continuing in this way we obtain a perturbation of Γ where $\phi(\Delta) \subseteq \Delta$. If during the process we encounter Δ with a 1-gate vertex, we perturb the map as in the proof of Proposition 2.1 with the effect that Δ is replaced by a smaller graph.

If ϕ maps an edge e of Δ over an illegal turn, first perturb by folding the illegal turn (this means identify initial segments of small length $\epsilon > 0$ in the two edges and rescale the metric so that the volume is 1; the map ϕ naturally induces a map on the quotient). After homotoping rel vertices to a map linear on edges, we see that an edge (induced by e) drops out of Δ and complexity decreases. Again it may be necessary to remove 1-gate vertices.

Now suppose that ϕ_* maps a legal turn to an illegal turn. Perturb by folding the illegal turn. This converts the legal turn to an illegal turn and it either lowers the number

$$\sum_v [\max(0, G(v) - 2)]$$

where $G(v)$ is the number of gates at v , or else it introduces a 1-gate vertex. In the latter case the subsequent perturbation of ϕ lowers the complexity of Δ . At the end of the process we have ϕ and Δ satisfying the conclusion. ■

COROLLARY 3.2. *Let Γ realize $\inf \tilde{\Phi} = \log \lambda$. Then $d(\Gamma, \Gamma \cdot \Phi^k) = k \log \lambda$ for any $k = 1, 2, \dots$*

Proof. After perturbing as in the proof of Proposition 3.1, the statement follows by observing that for a legal loop α the loop $\phi(\alpha)$ is also legal, and iterating we find that the length of $\phi^k(\alpha)$ is equal to $\lambda^k \ell(\alpha)$. By continuity, this is true before perturbing as well. ■

REMARK 3.3. It is easy to see that if Φ is any automorphism and $\phi : \Gamma \rightarrow \Gamma \cdot \Phi$ an optimal map satisfying the conclusions of Proposition 3.1, then $\tilde{\Phi}$ achieves minimum at Γ . Indeed, as above we have $d(\Gamma, \Gamma \cdot \Phi^k) = kd(\Gamma, \Gamma \cdot \Phi)$ for $k > 0$ and if $\Gamma' \in \mathcal{X}_n$ then

$$\begin{aligned} d(\Gamma, \Gamma \cdot \Phi^k) &\leq d(\Gamma, \Gamma') + d(\Gamma', \Gamma' \cdot \Phi^k) + d(\Gamma' \cdot \Phi^k, \Gamma \cdot \Phi^k) \\ &\leq d(\Gamma, \Gamma') + kd(\Gamma', \Gamma' \cdot \Phi) + d(\Gamma', \Gamma). \end{aligned}$$

The first and last terms are independent of k so by dividing by k and taking the limit as $k \rightarrow \infty$ we see that $d(\Gamma, \Gamma \cdot \Phi) \leq d(\Gamma', \Gamma' \cdot \Phi)$.

For another example of an automorphism of this type consider $\Phi : F_3 \rightarrow F_3$ given by $\Phi(a) = ab$, $\Phi(b) = bab$ and $\Phi(c) = cw$ where w is any word in a and b . Let Γ be the rose with the metric on a, b as in Figure 2 but scaled by $t > 0$, and let the length of c be $1 - t$. When $t > 0$ is sufficiently small (depending on w), Γ realizes the minimal displacement $\log \lambda$ with the same λ as in Figure 2.

3.3. Φ is parabolic. One example of a parabolic automorphism, with $\inf \tilde{\Phi} = 0$, is shown in Figure 3.

For another example, consider $\Phi(a) = ab$, $\Phi(b) = bab$, $\Phi(c) = cad$, $\Phi(d) = dcad$. On a and b take the metric from Figure 2 scaled by a parameter $t > 0$, and on c and d take the same metric (with c corresponding to a and d to b) scaled by $1 - t$. The tension graph consists of c and d , with slope converging to $\lambda = (3 + \sqrt{5})/2$ as $t \rightarrow 0$. In this example $\inf \tilde{\Phi} = \log \lambda > 0$ is not realized. (One way to see this is to show that the (combinatorial) length of $\Phi^k(c)$ grows like $k\lambda^k$.)

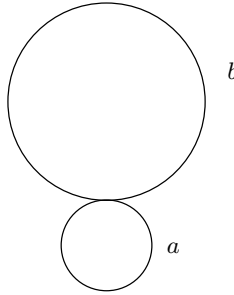


Fig. 3. $\Phi(a) = a$, $\Phi(b) = ab$. Let the length of a be $t > 0$ and the length of b be $1 - t$. Then the slope on a is 1 and the slope on b is $1/(1 - t)$. Thus as $t \rightarrow 0$ the displacement converges to 0. The tension graph is the loop b .

Fix a sequence Γ_k such that

$$d(\Gamma_k, \Gamma_k \cdot \Phi) \rightarrow D = \inf \tilde{\Phi}$$

but D is not realized.

For $\theta > 0$ denote by $\mathcal{X}_n(\theta)$ the subspace of \mathcal{X}_n consisting of all marked metric graphs Γ where every nontrivial loop has length $\geq \theta$ (the θ -thick part of \mathcal{X}_n). Then $\mathcal{X}_n(\theta)$ is an $\text{Out}(F_n)$ -invariant closed subset on which $\text{Out}(F_n)$ acts cocompactly.

PROPOSITION 3.4. *For any $\theta > 0$ there are only finitely many Γ_k with $\Gamma_k \in \mathcal{X}_n(\theta)$.*

Proof. Suppose not, and after passing to a subsequence assume that $\Gamma_k \in \mathcal{X}_n(\theta)$ for every k . By cocompactness, there are $\Psi_k \in \text{Out}(F_n)$ such that $\Gamma_k \cdot \Psi_k \rightarrow \Gamma_\infty$ (after taking a subsequence). Therefore, we have

$$\begin{aligned} d(\Gamma_\infty \cdot \Psi_k^{-1}, \Gamma_\infty \cdot \Psi_k^{-1} \Phi) & \\ & \leq d(\Gamma_\infty \cdot \Psi_k^{-1}, \Gamma_k) + d(\Gamma_k, \Gamma_k \cdot \Phi) + d(\Gamma_k \cdot \Phi, \Gamma_\infty \cdot \Psi_k^{-1} \Phi) \\ & = d(\Gamma_\infty, \Gamma_k \cdot \Psi_k) + d(\Gamma_k, \Gamma_k \cdot \Phi) + d(\Gamma_k \cdot \Psi_k, \Gamma_\infty) \end{aligned}$$

and hence

$$d(\Gamma_\infty \cdot \Psi_k^{-1}, \Gamma_\infty \cdot \Psi_k^{-1} \Phi) \rightarrow D$$

(since $d(\Gamma_k \Psi_k, \Gamma_\infty) \rightarrow 0$ and $d(\Gamma_\infty, \Gamma_k \Psi_k) \rightarrow 0$). In other words,

$$d(\Gamma_\infty \cdot \Psi_k^{-1} \Phi^{-1} \Psi_k, \Gamma_\infty) \rightarrow D.$$

Note that Arzelà–Ascoli implies that there are only finitely many $\Psi \in \text{Out}(F_n)$ such that $d(\Gamma_\infty \Psi, \Gamma_\infty) \leq D + 1$ (the set of e^{D+1} -Lipschitz maps $\Gamma_\infty \rightarrow \Gamma_\infty$ is compact and nearby maps are homotopic, so such maps represent only finitely many homotopy classes). It follows that, after taking a subsequence, $\Psi_k^{-1} \Phi^{-1} \Psi_k$ is a constant sequence and

$$d(\Gamma_\infty \cdot \Psi_k^{-1} \Phi^{-1} \Psi_k, \Gamma_\infty) = D,$$

i.e. the displacement of $\Gamma_\infty \cdot \Psi_k^{-1}$ under Φ is D , contradicting our assumption that inf is not realized. ■

For $\epsilon > 0$ the ϵ -small subspace Γ^ϵ of Γ is the union of all essential (not necessarily immersed) loops of length $\leq \epsilon$ (but note that Γ^ϵ may not be a subgraph). There is $\epsilon_n > 0$ such that for any $\Gamma \in \mathcal{X}_n$ the subspace Γ^{ϵ_n} is always proper (i.e. not equal to Γ , e.g. there is always an edge of length $\geq 1/(3n-3)$, assuming no valence 2 vertices, so $\epsilon_n < 1/(3n-3)$ works). Moreover, there is a bound B_n to the length of any chain of proper core subgraphs.

PROPOSITION 3.5. *For large k any optimal map $\phi : \Gamma_k \rightarrow \Gamma_k \cdot \Phi$ leaves a nonempty proper core subgraph invariant up to homotopy (and so ϕ is homotopic to a possibly nonoptimal map that maps this core subgraph to itself).*

Proof. Let $\theta = \epsilon_n / (e^{D+1})^{B_n}$. By Proposition 3.4 eventually $\Gamma_k \notin \mathcal{X}_n(\theta)$. Choose k so large that in addition the displacement of Γ_k is $< D+1$. Set $\delta_i = \epsilon_n / e^{(D+1)^i}$, $i = 0, 1, \dots, B_n$. Then

$$\Gamma_k^{\delta_0} \supset \Gamma_k^{\delta_1} \supset \dots \supset \Gamma_k^{\delta_{B_n}}$$

form a chain of homotopically nontrivial proper subspaces (not necessarily subgraphs, but abstract graphs) of length $B_n + 1$, so there must be some i such that $\Gamma_k^{\delta_i}$ and $\Gamma_k^{\delta_{i+1}}$ have the same core. By definition, an optimal map must send $\Gamma_k^{\delta_{i+1}}$ into $\Gamma_k^{\delta_i}$, so the common core is mapped to itself up to homotopy. ■

4. Train tracks. In this section we complete the proof of the train track theorem. We first make the standard definitions; the definition of a train track structure is as before, but does not require a metric on the graph.

DEFINITION 4.1. Let Γ be a marked graph with marking $f : R_n \rightarrow \Gamma$. We say that $\phi : \Gamma \rightarrow \Gamma$ represents $\Phi \in \text{Out}(F_n)$ if $\phi f \simeq f\Phi$. The automorphism Φ is *reducible* if there is some $\phi : \Gamma \rightarrow \Gamma$ that represents Φ and leaves a homotopically nontrivial (i.e. not a forest) proper subgraph invariant. Otherwise Φ is *irreducible*.

DEFINITION 4.2. A *train track structure* on a core graph Γ is an equivalence relation on the set of directions at every vertex of Γ with at least two equivalence classes (called *gates*) at every vertex. As before, a turn $\{d, d'\}$ is *illegal* if $d \sim d'$, and *legal* otherwise. An immersed loop or a path is *legal* if it takes only legal turns.

DEFINITION 4.3. Let $\Phi \in \text{Out}(F_n)$ be irreducible. A map $\phi : \Gamma \rightarrow \Gamma$ representing Φ is a *train track map* if it sends each edge to a nondegenerate immersed path and the following two equivalent conditions are satisfied:

- (i) There is a train track structure on Γ such that ϕ sends edges to legal paths and if v is a vertex of Γ then either
- $\phi(v) = w$ is a vertex and inequivalent directions at v map to inequivalent directions at w , or
 - $\phi(v)$ is not a vertex, v has two gates, and all directions in one gate fold to one direction at $\phi(v)$, while all directions in the other gate fold to the other direction at $\phi(v)$.
- (ii) There is a loop α in Γ such that every iterate $\phi^k(\alpha)$ is immersed, $k = 1, 2, \dots$

To prove that (i) implies (ii) it suffices to choose any legal loop for α , since (i) guarantees that legal loops map to legal loops. For the converse, first note that the iterates of α cross every edge of Γ (otherwise the union of the edges crossed by the iterates of α would be a homotopically nontrivial proper ϕ -invariant subgraph). Thus iterated images of edges are immersed. Define a train track structure on Γ by declaring $d \sim d'$ if $\phi_*^k(d) = \phi_*^k(d')$ for some $k \geq 1$. The reader may now easily check that the conditions in (i) hold. For example, any simplicial isomorphism $\phi : \Gamma \rightarrow \Gamma$ is a train track map.

REMARK 4.4. For a given $\phi : \Gamma \rightarrow \Gamma$ there may be more than one invariant (i.e. satisfying (i)) train track structure. For example, in Figure 2 we could take the gates to be $\{\bar{a}, \bar{b}\}$ and $\{a, b\}$. In the above paragraph we constructed the invariant train track structure with the *minimal* collection of illegal turns.

The discussion in Section 3 proves the following theorem.

THEOREM ([BH92]). *Every irreducible automorphism Φ is represented by a train track map $\phi : \Gamma \rightarrow \Gamma$.*

There is an additional caveat. In [BH92] train track maps always send vertices to vertices. This is very useful, and luckily it is not hard to achieve. Let ϕ_t , $t \in [0, 1]$, be a homotopy of $\phi = \phi_0$ that moves each $\phi(v)$ that is not a vertex to an endpoint of the edge containing $\phi(v)$. We also insist that during the homotopy the order of the images of vertices in the same edge does not change (i.e. there are no collisions, or “uncollisions”) until the very last moment when several images of vertices may arrive at the same vertex. It is not hard to see that the images of legal loops under ϕ_t are unchanged (and they are still legal loops). There may be edges that map to points under ϕ_1 ; collapse all such edges iteratively (i.e. after a collapse there may be new such edges that are then collapsed). We obtain a new map $\phi' : \Gamma' \rightarrow \Gamma'$ representing Φ . Any legal loop in Γ induces a loop in Γ' whose

ϕ' -iterates are immersed, so ϕ' is a train track map that sends vertices to vertices.

Finally, as in [BH92], one can put a metric on Γ' by solving linear equations e.g. as in Figure 2. The array of lengths is a positive left eigenvector of the transition matrix M for ϕ' whose ij -entry is the number of times $\phi'(e_j)$ crosses e_i in either direction. The irreducibility of Φ implies the irreducibility of M , so by Perron–Frobenius theory, M has a unique (positive) eigenvalue with an associated positive eigenvector.

REMARK 4.5. When $\phi : \Gamma \rightarrow \Gamma$ sends vertices to vertices and edges to nondegenerate immersed paths, finding an invariant train track structure is algorithmic (when it exists). One forms a (finite) directed graph whose vertices are the directions at the vertices of Γ and a directed edge from d to d' when $d' = \phi_*(d)$. Thus each vertex of the directed graph has one outgoing edge, and following outgoing edges eventually ends in a periodic orbit (i.e. a cycle). Then define $d \sim d'$ if d and d' are based at the same vertex and their forward iterates eventually coincide. For example, in Figure 2 we have $a \mapsto a$, $b \mapsto b$ and $\bar{a} \mapsto \bar{b} \mapsto \bar{b}$, so $\bar{a} \sim \bar{b}$ is the only nontrivial equivalence.

Along the same lines one can give a proof of the existence of relative train track maps [BH92] representing any given $\Phi \in \text{Out}(F_n)$. One works in a relative Outer space, where all graphs contain a fixed subgraph Γ_0 and on which $\phi : \Gamma_0 \rightarrow \Gamma_0$ partially representing Φ has already been constructed. The edges of Γ_0 are assigned length 0. The strategy of the absolute case applies here as well. The details may appear elsewhere.

Acknowledgements. The author is indebted to Ric Wade for reading the paper carefully and making helpful comments.

The author gratefully acknowledges the support by the National Science Foundation.

References

- [AKB] Y. Algom-Kfir and M. Bestvina, *Asymmetry of Outer space*, to appear.
- [Ber78] L. Bers, *An extremal problem for quasiconformal mappings and a theorem by Thurston*, Acta Math. 141 (1978), 73–98.
- [BH92] M. Bestvina and M. Handel, *Train tracks and automorphisms of free groups*, Ann. of Math. (2) 135 (1992), 1–51.
- [CV86] M. Culler and K. Vogtmann, *Moduli of graphs and automorphisms of free groups*, Invent. Math. 84 (1986), 91–119.
- [FM] S. Francaviglia and A. Martino, *Metric properties of Outer space*, arXiv:0803.0640.
- [Thu] W. P. Thurston, *Minimal stretch maps between hyperbolic surfaces*, arXiv:math.GT/9801039.

- [Thu88] W. P. Thurston, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. (N.S.) 19 (1988), 417–431.

Mladen Bestvina
Department of Mathematics
University of Utah
255 S 1400 E Rm 233
Salt Lake City, UT 84112, U.S.A.
E-mail: bestvina@math.utah.edu

*Received 19 January 2010;
in revised form 2 February 2011*