Annals of Mathematics, 117 (1983), 235–265

The Nielsen realization problem

By Steven P. Kerckhoff

Closed, oriented surfaces of genus $g \geq 2$ admit many hyperbolic (constant Gaussian curvature -1) metrics in contrast to Mostow's rigidity theorems in higher dimensions. Only special hyperbolic surfaces have non-trivial groups of isometries, but many different, non-isomorphic groups arise for different symmetric metrics. The group of isometries of a closed hyperbolic surface M_g^2 is always finite and the only isometry isotopic to the identity is the identity itself. As a result, hyperbolic surfaces with non-trivial groups of isometries have been a primary source for the construction of finite subgroups of the group of isotopy classes of diffeomorphisms of M_g^2, $\pi_0 \operatorname{Diff}(M_g^2)$. An old question, usually referred to as the *Nielsen Realization Problem*, is whether every such finite subgroup arises as a group of isometries of some hyperbolic surface. In this paper we answer the question in the affirmative.

THEOREM 5. *Every finite subgroup G of $\pi_0 \operatorname{Diff}(M_g^2)$ can be realized as a group of isometries of a hyperbolic surface.*

Remark. Theorem 5 has been known for G cyclic (Nielsen [14]), solvable (Fenchel [2]), and in numerous other special cases (see Zieschang [22]).

Since it is the generalization to higher genus of the classical modular group for genus 1, we refer to $\pi_0 \operatorname{Diff}(M_g^2)$ as the *modular group of genus* g and denote it by Mod_g. It is often called the *mapping class group* (we allow orientation-reversing maps however), and is known to be naturally isomorphic to the group of outer automorphisms of $\pi_1 M_g^2$. Mod_g acts properly discontinuously on T_g, the Teichmüller space of all hyperbolic metrics on M_g^2 (up to isometry isotopic to the identity), and the surfaces with a nontrivial group of isometries are distinguished as those with a non-trivial (necessarily finite) isotropy subgroup. This paper is primarily a study of the geometry of Teichmüller space, the results of which allow us to prove the following fixed point theorem which is equivalent to Theorem 5.

THEOREM 4. *Every finite subgroup $G \subset \mathrm{Mod}_g$ acting on T_g fixes some point in T_g.*

T_g is an open $6g - 6$ dimensional cell so that Theorem 4 follows for certain finite groups for topological reasons, but the general case relies on some strong convexity properties of T_g. There is substantial historical precedent for such an approach. In the genus 1 case, T_1 is the space of flat metrics of area 1 and is isometric to $\mathbf{H}^2$, 2-dimensional hyperbolic space. Mod_1 is $GL(2, \mathbf{Z})$ and acts by isometries on T_1 so that Theorem 4 is an easy consequence of the convexity properties of $\mathbf{H}^2$. Moreover, Kravetz [8] gave an analogous proof of Theorem 4 based on the false belief that T_g (with the Teichmüller metric) has negative curvature. This is now known to be false. (See [9] and [11].)

The convexity properties described here are not of a metric but rather of a function on T_g. It is shown that the geodesic length function of a closed curve is convex along certain paths in T_g. These paths are the traces of one-parameter families of deformations of hyperbolic structures which generalize the deformations gotten by twisting along simple closed geodesics. Called *earthquakes*, these deformations were invented (and named) by Thurston who also proved that any two points in T_g can be connected by an earthquake path.

The outline of the proof of Theorem 4 is briefly as follows: By taking sufficiently many simple closed curves $\{\gamma_i\} = \underline{\gamma}$, we insure that the sum of the geodesic lengths of the γ_i, $l_{\underline{\gamma}}$, realizes a minimum in T_g (Lemma 3.1). Now $l_{\underline{\gamma}}$ is shown to be strictly convex along earthquake paths (Theorem 1). Since every pair of points can be connected by such a path (Theorem 2), $l_{\underline{\gamma}}$ attains a unique minimum (Theorem 3). If $\underline{\gamma}$ is a G-invariant set of simple, closed curves, $l_{\underline{\gamma}}$ is G-invariant and thus its unique minimum is a fixed point for G.

Section 1A is a brief discussion of some geometric properties of $\mathbf{H}^2$ and its unit tangent bundle; Section 1B covers geodesic laminations. These sections are brief and purely expository, primarily stating known results which are needed in the paper. The reader unfamiliar with these subjects may wish to consult some of the references below for proofs. Section 2 develops the theory of earthquakes. Section 3 is the heart of the paper, where we discuss the geometric and analytic properties of the geodesic length function, especially along earthquake paths. Theorem 1 (the convexity result) is proved here. In Section 4, we give the formal proofs of Theorems 3 and 4 and indicate how the proof generalizes to general Fuchsian groups. Section 5 details some fairly immediate (and presumably well-known) restatements and consequences of Theorem 4 (e.g., Theorem 5), including some new information about Seifert fibered 3-manifolds. In the Appendix we give a proof of Thurston's Earthquake Theorem (Theorem 2), included here since no written source exists.

The results of this paper were previously announced in [7].

References. The theory of geodesic laminations is developed rather extensively in Chapters 8 and 9 of [17] so that no attempt is made here to reprove the basic theorems proved there. The reader may also find references on measured foliations (e.g. [1] and [18]) useful as they are analogous to geodesic laminations in a different context. (See the remark at the end of Section 1B.) Finally, a good general reference for background material on hyperbolic surfaces, Teichmüller spaces, and modular groups is the book *Discrete Groups and Automorphic Functions* [3].

Section I

I.A. A *hyperbolic surface* M^2 is a closed, oriented 2-dimensional manifold with a Riemannian metric of constant Gaussian curvature -1. Its universal cover also has constant negative curvature so it is isometric to the open unit disk endowed with the Poincaré metric (denoted by $\mathbf{H}^2$). Since the covering translations of the universal cover act by isometries, M^2 can be represented by $\mathbf{H}^2/\Gamma$, where Γ is a subgroup of the (orientation-preserving) isometry group of $\mathbf{H}^2$ (denoted by $I(\mathbf{H}^2)$) isomorphic to $\pi_1 M^2$ (called the Fuchsian group of M). Γ is well-defined up to conjugacy in $I(\mathbf{H}^2)$.

The Teichmüller space of genus g ($g \geq 2$), (T_g), is the space of marked hyperbolic surfaces, that is, hyperbolic surfaces together with a fixed isomorphism of $\pi_1 M$ to Γ where two surfaces are thought to be equivalent if there is an isometry between them respecting this isomorphism. Equivalently, T_g can be defined as the space of hyperbolic structures on a single underlying differentiable surface with two hyperbolic structures identified if there is an isometry between them isotopic to the identity or as the subset of discrete representations of $\pi_1 M$ into $I(\mathbf{H}^2)$ up to conjugacy. These definitions will be used interchangeably.

The *modular group* (Mod_g) or *mapping class group* of genus g is the group of diffeomorphisms of the surface of genus g up to isotopy ($\pi_0 \operatorname{Diff} M$). It is well-known that Mod_g is isomorphic to the group of outer automorphisms of $\pi_1 M$, Out π_1. The diffeomorphism group of M acts on T_g by pulling back metrics and the action descends to an action of Mod_g since the points of T_g are isotopy classes of metrics. Equivalently the automorphism group of $\pi_1 M$ acts on the space of discrete representations of $\pi_1 M$ into $I(\mathbf{H}^2)$ and the action descends to an action of Out π_1 on T_g. The action is properly discontinuous and, except for an element of order 2 in Mod_2, is faithful. The quotient space is the moduli space of genus g.

The action of $I(\mathbf{H}^2)$ on $\mathbf{H}^2$ extends continuously to the boundary circle of the open unit disk, called the circle at infinity, S^1_∞. Note that S^1_∞ can be identified with the circle of directions at a point in $\mathbf{H}^2$; directions at two different points are identified if the corresponding directed rays are asymptotic. A pair of distinct points on S^1_∞ determines a unique bi-infinite geodesic in $\mathbf{H}^2$. If a hyperbolic surface M is closed, every element of Γ is hyperbolic, i.e., it has two fixed points on S^1_∞, one attracting and one repelling. The geodesic corresponding to this pair of points projects to a closed geodesic on the surface which is the unique geodesic on M in its isotopy class.

If $f\colon M \to M'$ is a homotopy equivalence between two marked hyperbolic surfaces then by the work of Nielsen any lift of f to $\tilde{f}\colon \mathbf{H}^2 \to \mathbf{H}^2$ extends continuously to a homeomorphism on S^1_∞ and the extension is invariant under homotopy, depending only on the lift and on M and M'. (See also Mostow [12] or Thurston [17] for a proof in dimension ≥ 2.) The extension is equivariant so that the fixed endpoints of an element $\gamma \in \Gamma$ are taken to the endpoints of $\gamma' \in \Gamma'$ where $\tilde{f}_*(\gamma) = \gamma'$ ($\tilde{f}_*\colon \Gamma \to \Gamma'$ induced by $\tilde{f}$). This invariance on S^1_∞ of the extension of $\tilde{f}$ under homotopy (hence isotopy) suggests that the circle at infinity is likely to be important in the study of the action of Mod_g on T_g. For, although Mod_g does not act necessarily in a well-defined manner on a hyperbolic surface M, it does act on the circle at infinity of the universal cover of M (and hence on $T_1(M)$; see Section V). This was the beginning point for Nielsen's deep work on Mod_g and again becomes significant here (e.g., in the proof of Theorem 1).

$I(\mathbf{H}^2)$ also acts on $T_1(\mathbf{H}^2)$, the unit tangent bundle of $\mathbf{H}^2$. In fact the action is faithful and transitive so that $T_1(\mathbf{H}^2)$ can be identified with $I(\mathbf{H}^2)$. It follows that an element of $I(\mathbf{H}^2)$ is completely determined by its action on a single point of $T_1(\mathbf{H}^2)$. Topologically, $T_1(\mathbf{H}^2)$ is a trivial circle bundle over the disk.

The metric on $T_1(\mathbf{H}^2)$ is defined so that a path of parallel-translated vectors has the same length as the path projected onto $\mathbf{H}^2$ and the metric along the fiber comes from the standard one on S^1. This is the usual left invariant metric coming from the Lie group structure of $T_1\mathbf{H}^2 \approx \mathrm{PSL}(2, \mathbf{R})$. We will use $d(.,.)$ to denote either the metric in $\mathbf{H}^2$ or $T_1\mathbf{H}^2$ when the meaning is clear from the context.

The following lemma will be useful in understanding geodesic laminations. Given a geodesic l in $\mathbf{H}^2$ and a point x on it, we denote by x_l, the unit vector at x along l. (The choice between the two possibilities will always be evident from the context.)

LEMMA 1.1. *Given two disjoint geodesics l and l' in $\mathbf{H}^2$ and any two points x and x' on l and l' respectively, distance $\varepsilon < 1$ apart, then $d(x_l, x'_{l'}) < C\varepsilon$ for a universal constant C.*

Proof. We may fix l and x without loss of generality. Then for a fixed x', the worst case occurs when l and l' are asymptotic in one direction, so we consider only this case. The tangent vector $x'_{l'}$ determined by the asymptotic condition varies differentiably with x' and equals x when $x' = x$; so the lemma follows. $\square$

In Section II we will be studying deformations of $\mathbf{H}^2$ of the following sort: Let v be a tangent vector in $T_1(\mathbf{H}^2)$, w another tangent vector based at a point $x \in \mathbf{H}^2$ (usually distinct from the basepoint y of v.) Denote by $\mathfrak{S}_v(t)w$ the vector obtained by translating the geodesic from y to x distance t along the geodesic l determined by v. (See Figure 1.)

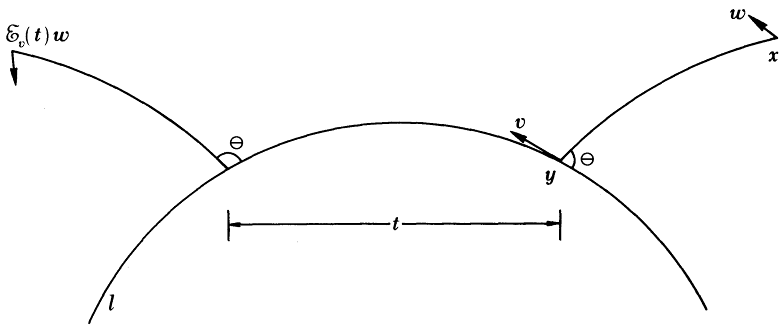

FIGURE 1

The dependence of $\mathfrak{S}_v(t)w$ is actually only on l, not on v, and we will say it is "the vector w translated distance t in the direction v" (or "along l"). Note that when $x = y$ this is just parallel translation.

The following lemma is the main control needed in the study of earthquakes in Section II.

LEMMA 1.2. *If* $d(v, v') < \varepsilon$, $v, v' \in T_1\mathbf{H}^2$, *and if* $w \in T_1\mathbf{H}^2$, *then*

i) $d(\mathfrak{S}_v(t)w, \mathfrak{S}_{v'}(t)w) < Kt\varepsilon$;

ii) $d(\mathfrak{S}_v(t)w, w) < Kt$

for all $t \leq T$, *where* K *is a constant depending only on* T *and the distance between the basepoints of* v *and* w.

Proof. We may fix w without loss of generality. Now $\mathfrak{S}_v(t)w$ depends differentiably on t and v. Moreover, $\mathfrak{S}_v(t)w = \mathfrak{S}_{v'}(t)w$ for all $v, v' \in T_1\mathbf{H}^2$ when $t = 0$ and for all $t \in \mathbf{R}$ when $v = v'$. The first inequality follows where K depends only on T and v (or v'). Since the circle of directions at the basepoint y of v is compact, the latter dependence can be reduced to one on y. Since $\mathbf{H}^2$ can be rotated by isometry, fixing the basepoint x of w, the dependence can further be reduced to one on $d(x, y)$.

The second inequality similarly follows from differentiability. □

I.B. A geodesic lamination on a hyperbolic surface M is a closed subset $\mathfrak{L} \subset \mathfrak{M}$ which is the union of geodesics and which is foliated in the following sense: There are open sets U_i covering $\mathfrak{L}$ with continuous maps ψ_i: $\mathfrak{L} \cap U_i \subset$

$U_i \to (0,1) \times B_i \subset \mathbf{R}^2$ taking $\mathfrak{L} \cap U_i$ to horizontal arcs $(0,1) \times y$, $y \in B_i$, in the plane such that the overlap maps preserve the horizontal property, i.e., $\psi_i \circ \psi_j^{-1}(x, y)$ is of the form $(f(x, y), g(y))$. The geodesic laminations of interest here will also be required to possess a positive Borel measure μ on its local leaf space B_i which is invariant under co-ordinate change. Equivalently, μ can be defined as a measure on arcs transverse to $\mathfrak{L}$ which is invariant under projection along leaves. Now μ is required to have all of $\mathfrak{L}$ as support and to be finite on compact arcs. The existence of the transverse measure on all of $\mathfrak{L}$ restricts the local behavior in that a local cross-section is either discrete or a Cantor set. A lamination will be called discrete when all its local cross-sections are discrete. We will drop the distinction in notation between a lamination with transverse measure and the measure itself and denote both by μ.

One example of a geodesic lamination is a simple closed geodesic ϕ. The measure μ is just the counting measure on ϕ times a positive real number. The closure of an infinite simple geodesic (if it does not asymptotically approach a closed geodesic) gives a lamination with Cantor set local cross-sections. The transverse measure may be defined as a limit (in the weak sense) as $d \to \infty$ of $1/d$ times the counting measure of segments of length d along a single leaf.

The transverse measure μ on a lamination induces a Stieltjes measure along any arc A transverse to it, defined as the total mass of μ along open sub-arcs. This induced measure allows one to integrate functions along A (denoted by $\int_A f\,d\mu$). In particular, the integral of the characteristic function of A is just the total mass of μ along A. To get a quantity dependent only on the isotopy class of A (endpoints fixed) we define the *intersection number*, $i(A, \mu)$, of A with μ to be the infimum of $\int_{A'} d\mu$ over all arcs A' isotopic to A fixing endpoints. Similarly we define the intersection number of a simple closed curve ϕ with μ, $i(\phi, u)$ to be the infimum of $\int_{\phi'} d\mu$ over all closed curves ϕ' isotopic to ϕ. It is not hard to see that the infimum is realized in each case by the unique geodesic in the isotopy class if we use the convention that a subarc of μ has zero intersection number with μ.

Another useful quantity to consider along an arc A is the angle θ which A makes with the leaves of μ, $0 \leq \theta < \pi$, measured counterclockwise from A to μ. Assume that A is a geodesic arc transverse to μ and define the *total angle* as $\int_A \theta\, d\mu$.

To see that the integral exists, it suffices to show that for any $\delta > 0$, there is an $\varepsilon > 0$ such that $|\theta(x) - \theta(x')| < \delta$ if $d(x, x') < \varepsilon$ and $x, x' \in A \cap \mu$. Then the integral can be estimated uniformly well by its Riemann sums. That θ is sufficiently well behaved follows from Lemma 1.1. (Note that θ is strictly greater than zero along A so there is no problem with the discontinuity of θ at 0.) Similarly, the *total cosine* $\int_A \cos\theta\, d\mu$ of A with μ can be defined.

The total angle of A with μ divided by the total mass $i(A, \mu)$ will be called the *average angle* of A with μ and be denoted by $\theta(A, \mu)$. The vector $v_A(\mu)$ in $\mathbf{R}^2$ determined in polar co-ordinates by $(i(A, \mu), \theta(A, \mu))$ will be used to compare two laminations μ and μ'.

The set of all geodesic laminations with transverse measure (referred to simply as laminations from now on) on M will be denoted by $\mathcal{ML}$. Now $\mathcal{ML}$ is given the following topology which can be thought of as a C^1-topology weighted by the transverse measure: Given a finite set of open geodesic arcs $\{A_i\}$, an $(\{A_i\}, \varepsilon)$-neighborhood of μ consists of all $\nu \in \mathcal{ML}$ such that $|v_{A_i}(\nu) - v_{A_i}(\mu)| < \varepsilon$ for all i.

This topology has the following geometric meaning in terms of the laminations themselves. Suppose μ_i converges to μ in $\mathcal{ML}$. If $x \in M$ is a point on a leaf of μ, then there are points $x_i \in \mu_i$ such that $x_i \to x$. Conversely if $x_i \in \mu_i$ is a convergent sequence with $x \in M$ its limit, then either $x \in \mu$ or there is a neighborhood U of x such that the mass $\mu_i(A)$ across any geodesic arc A in U goes to zero as $i \to \infty$. Similar statements hold for tangent vectors, tangent to μ and μ_i.

The measure on μ can always be multiplied by a positive scalar which causes $\mathcal{ML}$ to be non-compact. Deleting the zero lamination and dividing out by scalar multiplication determines a space $\mathcal{PL}$ of projective classes of geodesic laminations which inherits the quotient topology from $\mathcal{ML}$. $\mathcal{PL}$ is compact; in fact:

Theorem A (Thurston [17]). *$\mathcal{PL}$ is homeomorphic to S^{6g-7} and $\mathcal{ML}$ is homeomorphic to $\mathbf{R}^{6g-6}$ (where g = genus of M).*

At first sight it is surprising that $\mathcal{ML}$ and $\mathcal{PL}$ are even finite dimensional manifolds since there are uncountably many arcs in M. But in Lemma 1.1 we have seen that the behavior of μ at one point approximately determines its behavior in a whole neighborhood which in turn controls its behavior for a considerable distance along the leaves. In fact, the proof of Theorem A shows that the intersection numbers with a finite number of simple closed curves provide local co-ordinates for $\mathcal{ML}$.

For the subset of $\mathcal{ML}$ determined by simple closed geodesics with a multiple of the counting measure, it is not hard to see, by cutting M into $2g - 2$ spheres with 3 geodesic boundary components, that a point is determined by $6g - 6$ local parameters (see Fathi et al. [1]). Although these discrete laminations are special, it is shown during the proof of Theorem A that, if we denote by S the set of isotopy classes of simple closed curves and embed $S \times \mathbf{R}_+$ in $\mathcal{ML}$ by sending (γ, r) to the geodesic isotopic to γ with r times the counting measure, then

Theorem B (Thurston [18]). *$S \times \mathbf{R}_+$ is dense in $\mathcal{ML}$; S is dense in $\mathcal{PL}$.*

This theorem allows one to extend ideas that are easily defined for simple closed geodesics to the entire space of laminations. (For example the *length* of a geodesic lamination is well-defined; see [17], Chapter 9.) One example of this principle is the object of the next section.

It should be pointed out that although $\mathcal{ML}$ and $\mathcal{PL}$ are defined in terms of a specific hyperbolic structure, there is a homeomorphism between $\mathcal{ML}(M)$ and $\mathcal{ML}(M')$, $M, M' \in T_g$, which can be defined using the circle at infinity of $\mathbf{H}^2$. The point is that there is a $1-1$ correspondence between pairs of points on S^1_∞ and geodesics in $\mathbf{H}^2$, and that any lift of a map from M to M' to $\mathbf{H}^2$ extends to S^1_∞, the extension depending only on the lift and on M and M'. The map on pairs of points on S^1_∞ induces one on laminations. (The transverse measure is equivalent to a measure on pairs of points so that can be carried over also.) The map induced on $S \times \mathbf{R}_+$ is just the identity and in general the corresponding laminations are isotopic on the underlying manifold.

Remark. For the reader familiar with the theory of measured foliations ([1], [18]), it should be pointed out that geodesic laminations are the analog of measured foliations in the context of hyperbolic geometry and that there is a homeomorphism from $\mathcal{MF}$, the space of measured foliations, to $\mathcal{ML}$. (This homeomorphism is also defined in terms of S^1_∞.) Thus Theorems A and B also follow from the comparable theorems about measured foliations.

Warning. In [17] it is proved only that $\mathcal{PL}$ is a sphere; its dimension is not actually computed. However, the dimension is easily seen to be $6g - 7$ by the comment above about parametrizing $S \times \mathbf{R}^+$ by $6g - 6$ parameters or by, in the terminology of [17], computing the number of independent parameters of any essentially complete train track.

Section II

The deformations studied in this section will be the primary tool in studying the length function on T_g. The prototype for this deformation, defined when the shearing lamination is a simple closed geodesic, dates back at least to Fenchel and Nielsen, but seems not to have been studied in depth until a recent rekindling of interest. It was Thurston who first realized that the deformations generalized to the form considered here.

Fix a hyperbolic surface M and choose any simple closed geodesic ϕ on M. Cut along ϕ to get a (possibly disconnected) surface with geodesic boundary. A new hyperbolic manifold M_t is formed by gluing the boundary components back with a left twist of distance t; i.e., the two images of any point on ϕ are separated by distance t along the image of ϕ in M_t. Notice that the notions of "left" and "right" twists depend only on the orientation of M (no orientation of ϕ is necessary). M_0 is the original manifold and M_t is clearly a homeomorphic hyperbolic surface.

There is a map τ_t from M_0 to M_t which is an isometry off ϕ but is not uniquely defined at ϕ and is discontinuous there. (It is sometimes useful to think of a point on ϕ as splitting into two copies of itself, one of which is moved distance t, the other fixed.) For the hyperbolic structure of M_t to determine a new point in T_g, not just in the moduli space, we require a homotopy class of maps from M_0 to M_t. This is done by requiring that any closed curve γ in M_0 be mapped to a curve in M_t homotopic to the curve determined by following $\tau_t(\gamma)$ until it hits ϕ, then running along ϕ distance t to the left, then following $\tau_t(\gamma)$, and so on. Thus the hyperbolic structure determined by a full left twist is distinct from M_0.

Remark. These deformations can also be described as 1-parameter families of amalgamations (HNN constructions or amalgamated free products) of the Fuchsian group(s) of the surface(s) with boundary. These are known ([3], Chapter 9) to change the lengths of geodesics real analytically. As a result, the question of differentiability of the more general deformations to be studied here is settled simply by controlling convergence. (See, for example, Corollary 3.4.)

The twist deformation can be defined for any $\gamma \in S$, $t \in \mathbf{R}_+$, and $M \in T_g$, and, more generally, given any $(\gamma, \mu) \in S \times \mathbf{R}_+ \subset \mathfrak{M}\mathfrak{L}$, $M \in T_g$, we define the *time t twist deformation* of M to be the new structure obtained from M by twisting left distance $t\mu$ along γ. In other words μ determines the speed of twisting. Finally, for any $\nu \in \mathfrak{M}\mathfrak{L}$ we want to determine a deformation by a limiting process, following Thurston.

Definition. The *left earthquake deformation of M at time t determined by ν* is the limit in T_g of the time t twist deformations of M for any sequence (γ_i, μ_i) in $S \times \mathbf{R}_+$ converging to ν in $\mathfrak{M}\mathfrak{L}$. It will be denoted by $\mathcal{E}_\nu(t)$ (where M is understood from the context).

Since M above is arbitrary, $\mathcal{E}_\nu(t)$ can be thought of as a map from T_g to itself. Furthermore, we will see below that for any M the sequence of twist *maps* determined by (γ_i, μ_i) above converges (in an appropriate sense) so that $\mathcal{E}_\nu(t)$ may be thought of as a *map* determining a new structure for each t. As a result, we will speak of left earthquake maps and deformations interchangeably (and equivalently) as left earthquakes.

As in the twist case, an earthquake is an isometry in the complement of ν. However, if ν has no atoms, $\mathcal{E}_\nu(t)$ is continuous; the image of a transverse arc looks qualitatively like the graph of the Cantor function. The issue of a homotopy class of maps from M to M_t is settled in this case by choosing the class of $\mathcal{E}_\nu(t)$. When ν has atoms, the same device as before is used to connect disconnected curves along the isolated leaves of ν.

Some Terminology. We note that twist deformations (and maps) are earthquakes, but we will save the adjective "twist" for the simple closed curve case. To avoid linguistic tedium and absurdity, we will use "to shear" in place of "to earthquake" in its various forms (verb, gerund, etc.). The lamination ν determining the earthquake will be referred to as the "shearing lamination."

Remark. The definition of earthquakes as limits may seem unnatural at first and, indeed, it is possible to define them directly from the shearing laminations ν themselves by requiring that the "total amount of shearing" along any transverse geodesic arc A in M be equal to $ti(A, \nu)$. However, showing that this rough notion can be made precise and well-defined seems to involve approximating the Cantor set local cross-sections B_i of ν by discrete cross-sections and applying (locally) the approximation techniques developed below. Thus nothing much seems to be gained by this approach, while a few of the proofs are somewhat harder with such a direct definition.

Of course, for the definition above to be of any use, it is necessary to show that the limits exist and are independent of the converging sequence. To this end, we will need a few lemmas which relate the closeness of simple closed curves in $\mathcal{ML}$ to the closeness of the effect of shearing along them.

Definition. If A is a geodesic arc transverse to $\gamma, \bar{\gamma} \in S \times \mathbf{R}_+$, we will say that γ and $\bar{\gamma}$ are *ε-close along δ-subarcs* of A if there are subarcs A_i, such that $A = \bigcup A_i$, whose lengths are bounded by δ such that $\sum_{i=1}^{n} |v_{A_i}(\gamma) - v_{A_i}(\bar{\gamma})| < \varepsilon$.

For any $\nu \in \mathcal{ML}$ and any transverse geodesic arc A and real numbers $\varepsilon, \delta > 0$ there is a neighborhood U of ν such that all $\gamma, \bar{\gamma} \in (S \times \mathbf{R}_+) \cap U$ are ε-close along δ-subarcs of A. To control the corresponding time t twist maps, we will first replace the leaves of γ and $\bar{\gamma}$ by a single leaf across each subarc (Lemma 2.1) so that there is a 1–1 correspondence between leaves of γ and of $\bar{\gamma}$. The desired estimates then follow from Lemma 1.2 and the triangle inequality (Lemma 2.2 and Proposition 2.3).

We will work in the universal cover, $\mathbf{H}^2$, where keeping track of the homotopy class of a map is easier. Twists along geodesics clearly lift to $\mathbf{H}^2$, and the same notation is used in this case. For example, if l is a geodesic in $\mathbf{H}^2$ with a transverse measure, $\mathcal{E}_l(t)$ denotes the time t twist along l (which is well-defined once a neighborhood of a point not on l is fixed).

Consider the following situation: Let $x, y \in \mathbf{H}^2$, $v \in T_1\mathbf{H}^2$ based at y, $\bar{A}$ be the geodesic arc from x to y. Suppose γ is a discrete lamination in $\mathbf{H}^2$ with equal measures on each leaf whose intersections with $\bar{A}$ are contained in a small subarc A. Let l be a single geodesic intersecting A with angle equal to the average angle $\theta(A, \gamma)$ and with mass equal to $\mu = i(A, \gamma)$. Let $\mathcal{E}_\gamma(t)$ and $\mathcal{E}_l(t)$ be the time t twist maps in $\mathbf{H}^2$ along γ and l respectively which fix a neighborhood of x.

LEMMA 2.1. *With notation as above, if the subarc A has length less than δ, then, for any $T \in \mathbf{R}_+$, the distance between the image vectors $\mathcal{E}_\gamma(t)v$ and $\mathcal{E}_l(t)v$ in $T_1\mathbf{H}^2$ is less than $Kt\mu\delta$, for all $t \le T$, where K is a constant depending only on $T\mu$ and $d(x, y)$.*

Proof. Consider $\mathcal{E}_\gamma(t)$ and denote by $l_1, \dots, l_n$ the images of the leaves of γ under $\mathcal{E}_\gamma(t)$. The image of $\bar{A}$ is a disconnected arc (still denoted $\bar{A}$). The point x is connected to $\mathcal{E}_\gamma(t)y$ by a staircase path, going along one component of $\bar{A} - A$ to l_1, along l_1 distance μ/n, along A to l_2, along l_2 distance μ/n and so on. (See Figure 2.) Call the successive components of $\bar{A}$, $A_0, A_1, \dots, A_n$, so that A_0 connects x to l_1, A_i connects l_i to l_{i+1}, $i = 1, \dots, n-1$, and A_n connects l_n to $\mathcal{E}_\gamma(t)y$. Let the length of A_i be denoted by δ_i.

Now alter $\mathcal{E}_\gamma(t)$ as follows: The shearing along l_n distance μ/n is replaced by shearing along l_{n-1} distance μ/n farther for a total distance of $2\mu/n$. The change in $\mathcal{E}_\gamma(t)v$ is less than $Kt\frac{\mu}{n}C\delta_{n-1}$ by Lemmas 1.1 and 1.2. Replacing the shearing along l_{n-1} by shearing a total distance of $3\mu/n$ along l_{n-2} changes $\mathcal{E}_\gamma(t)v$ by less than $Kt\frac{2\mu}{n}C\delta_{n-2}$. Continuing in this fashion until the map just shears distance μ along l_1 gives a total change of less than $KtC\sum_{i=1}^{n-1}\frac{i\mu}{n}\delta_{n-i}$ which is less than $KCt\mu\delta$ since $\frac{i\mu}{n} < \mu$, $i \le n-1$, and $\sum_{i=1}^{n-1}\delta_{n-i} = \delta$.

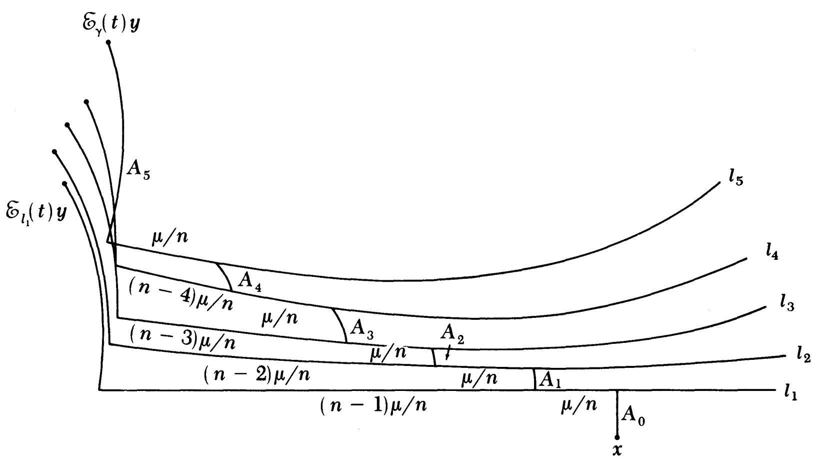

FIGURE 2

Note that the constants K from Lemma 1.2 depend on the distance between different points in each case, but that the distance between these points is bounded in terms of $T\mu$ and $d(x, y)$ so that a single, sufficiently large K will do.

To complete the proof, we must replace l_1 with l. Since by Lemma 1.1 the average angle $\theta(A, \gamma)$ differs from the angle of intersection of any particular leaf with A by at most $C\delta$ and the points of intersection are within δ of each other, Lemma 1.2 allows us to replace l_1 by l with a change bounded by $K(C+1)\delta t\mu$. Lemma 2.1 follows by the triangle inequality, if we let $\bar{K} = K(2C+1)$ and denote $\bar{K}$ by K. □

LEMMA 2.2. *If x, y, v, and $\bar{A}$ are as above and if l and $\bar{l}$ are geodesics in $\mathbf{H}^2$ with measures μ, $\bar{\mu}$ such that $l \cap \bar{A} = \bar{l} \cap \bar{A} = p$, $p \neq x, y$ and $|v_{\bar{A}}(l) - v_{\bar{A}}(\bar{l})| < \varepsilon$, then for any $T \in \mathbf{R}_+$, $d(\mathcal{E}_l(t)v, \mathcal{E}_{\bar{l}}(t)v) < Kt\varepsilon$, for all $t \leq T$, where K depends only on $d(x, y)$ and $T\mu$.*

Proof. Since $|\mu - \bar{\mu}| < \varepsilon$, we can assume that $\mu = \bar{\mu}$, with error bounded by $Kt\varepsilon$ (Lemma 1.2). Then, if the angle between l and $\bar{l}$ at p is θ, $\theta\mu < \varepsilon$, so, again by Lemma 1.2, $d(\mathcal{E}_l(t)v, \mathcal{E}_{\bar{l}}(t)v) < Kt\theta\mu < Kt\varepsilon$. □

In the proof of the proposition below, the effects of two twist maps on a fixed $v \in T_1\mathbf{H}^2$ are compared by successively changing the maps across small subarcs of an arc A. Control of the error for such a change on a small arc is given by the previous two lemmas. However, to use the triangle inequality to control the error over all of A, an observation is necessary.

When a twist map (in $\mathbf{H}^2$) is determined by shearing along several geodesics, the order of shearing is irrelevant; the rest of the geodesics move by isometry. Since Lemmas 2.1 and 2.2 are stated for two maps applied to the *same* tangent vector, it is necessary to shear *last* along the geodesics in the subarc where the change is being made. In this way the two maps across all of A differ only at the last stage, in the way they move the *image* of v under the rest of the twists, and Lemmas 2.1 and 2.2 apply. But, since v has been moved, the distance between x and the basepoint of v will differ in each case. However, a uniform choice of the constants in 2.1 and 2.2 is possible since everything takes place in a compact set. This having been said, no mention of the validity of the use of the triangle inequality or of the choice of constants is made in the proof below.

PROPOSITION 2.3. *Let $\nu \in \mathcal{ML}$ be any lamination (lifted to $\mathbf{H}^2$) and let $x, y \in \mathbf{H}^2$, A be the geodesic from x to y, $v \in T_1\mathbf{H}^2$ be based at y, where x, y do not lie on the atomic part of ν. Then for any $\varepsilon > 0$, $T > 0$, there is a neighborhood U of ν in $\mathcal{ML}$ such that for all $\gamma, \bar{\gamma} \in (S \times \mathbf{R}_+) \cap U$, $d(\mathcal{E}_\gamma(t)v, \mathcal{E}_{\bar{\gamma}}(t)v) < Kt\varepsilon$, for all $t \leq T$, K a constant depending only on $d(x, y)$ and $Ti(\nu, A)$.*

Proof. Choose U such that for any $\gamma, \bar{\gamma}$ as above, γ and $\bar{\gamma}$ are ε-close along δ-subarcs of A, where δ is chosen so that $\delta i(A, \mu) < \varepsilon$, for all $\mu \in U$. We can assume that neither x nor y is on a leaf of γ or $\bar{\gamma}$ by choosing U small enough that the measure across all geodesic arcs in a sufficiently small neighborhood around x and y is less than ε. (This is possible because ν is not atomic at x nor at y.) Therefore, these leaves can be ignored and the error incorporated into $Kt\varepsilon$. As in Lemma 2.1, replace γ and $\bar{\gamma}$ on each subarc of A by single geodesics which pass through the same point. By Lemma 2.1 and the triangle inequality, this changes $\mathcal{E}_\gamma(t)v$ and $\mathcal{E}_{\bar{\gamma}}(t)v$ by less than $Kt\delta i(A, \gamma)$ and $Kt\delta i(A, \bar{\gamma})$, respectively, which are both less than $Kt\varepsilon$ by the choice of δ.

Denote the new geodesics by $l_1, \dots, l_n$ and $\bar{l}_1, \dots, \bar{l}_n$ respectively. First shear along the l_i according to their measures. If we shear along $l_1, \dots, l_{n-1}$ and then l_n is replaced by $\bar{l}_n$, the change is less than $Kt\varepsilon_n$ by Lemma 2.2, where ε_n is the deviation of γ and $\bar{\gamma}$ on the last subarc. Similarly, we can then replace l_{n-1} by $\bar{l}_{n-1}$ and so on until all the l_i are replaced by $\bar{l}_i$. The total change is less than $\sum_{i=1}^n Kt\varepsilon_i = Kt\varepsilon$.

Letting $\bar{K} = 3K$ and denoting $\bar{K}$ by K complete the proof. □

To compare the images of a point in T_g under various twist maps, we need a notion of an ε-neighborhood in T_g. Fix a set of generators ϕ_i for $\pi_1 M$, continuously choose a representation in PSL(2, $\mathbf{R}$) for each point in T_g and fix $v \in T_1\mathbf{H}^2$. An ε-neighborhood of $M \in T_g$ consists of all $\bar{M} \in T_g$ such that $d(\phi_i(v), \bar{\phi}_i(v)) < \varepsilon$, for all i, where ϕ_i and $\bar{\phi}_i$ denote the elements in PSL(2, $\mathbf{R}$) corresponding to the various generators in the representations of M and $\bar{M}$ respectively. The induced topology is independent of the choices made.

PROPOSITION 2.4. *Given $M \in T_g$, $\nu \in \mathcal{ML}$, $T > 0$, $\delta > 0$, there is a neighborhood U of ν in $\mathcal{ML}$ such that for all $\gamma, \bar{\gamma} \in (S \times \mathbf{R}_+) \cap U$ and every fixed $t \leq T$, $\mathcal{E}_\gamma(t)$ and $\mathcal{E}_{\bar{\gamma}}(t)$ are in the same δ-neighborhood in T_g.*

Proof. Let x be the basepoint of v. Assume x is not on a leaf of ν in $\mathbf{H}^2$ after a small perturbation, if necessary, and lift the twist maps so as to fix a neighborhood of x. Connect x to $\phi_i(x)$ by geodesic arcs A_i, where the ϕ_i are the generators of $\pi_1 M$ acting on $\mathbf{H}^2$ (and $T_1\mathbf{H}^2$) and let $v_i = \phi_i(v)$. After shearing by γ and $\bar{\gamma}$, respectively, ϕ_i will take v_i to $\mathcal{E}_\gamma(t)v_i$ and $\mathcal{E}_{\bar{\gamma}}(t)v_i$. Choosing ε so that $\delta = K\varepsilon T$ and choosing U sufficiently small to satisfy Proposition 2.3 for all A_i simultaneously, we see that the proposition follows. □

The following corollaries are immediate:

COROLLARY 2.5. *Left earthquakes along ν are well-defined for all $\nu \in \mathcal{ML}$ and for all time $t \in \mathbf{R}_+$.*

COROLLARY 2.6. *The map* $\mathcal{ML} \times \mathbf{R}_+ \to T_g$ *which takes* (ν, t) *to the image of a base surface* M *under* $\mathcal{E}_\nu(t)$ *is continuous.*

To see that earthquakes $\mathcal{E}_\nu(t)$ are *maps* as well as curves in T_g, we need to see that the twist maps defined by a sequence of simple closed curves converging to ν in $\mathcal{ML}$ converge. Because ν may have an atomic part, where $\mathcal{E}_\nu(t)$ will not be defined, the definition of convergence will avoid such points. However, the behavior there is controlled by nearby points because of the nature of the maps. Again, to keep track of the homotopy class of the maps, we look at the maps lifted to $\mathbf{H}^2$ and consider a fundamental domain D for the hyperbolic surface.

Definition. Let $\nu \in \mathcal{ML}$, and let $\mathcal{E}_{\gamma_i}(t)$ be a sequence of twist maps, $\gamma_i \in S \times \mathbf{R}_+$, $\gamma_i \to \nu$. We say $\mathcal{E}_{\gamma_i}(t)$ *converges* if there are lifts to $\mathbf{H}^2$ such that for each t and every $x \in D$ not on an atomic part of ν, $\mathcal{E}_{\gamma_i}(t)x$ converges.

If x is on γ_i itself for some i, $\mathcal{E}_{\gamma_i}(t)x$ is interpreted as either "copy" of $\mathcal{E}_{\gamma_i}(t)x$. In particular the amount of shearing at x must be small.

PROPOSITION 2.7. *For every* $\nu \in \mathcal{ML}$ *and any* $\{\gamma_i\} \in S \times \mathbf{R}_+$ *such that* $\{\gamma_i\}$ *converges to* ν, *the twist maps* $\mathcal{E}_{\gamma_i}(t)$ *converge.*

Proof. Pick $p \notin \nu$ in $\mathbf{H}^2$ to be fixed for all i and choose a fundamental domain $D \subset \mathbf{H}^2$. Consider $x \in D$ not on the atomic part of ν. By Proposition 2.3, for each x and for all $\varepsilon > 0$, there is a neighborhood U of ν in $\mathcal{ML}$ where $d(\mathcal{E}_\gamma(t)x, \mathcal{E}_{\bar\gamma}(t)x) < KT\varepsilon$, for all $t \le T$, and all $\gamma, \bar\gamma \in (S \times \mathbf{R}_+) \cap U$. Convergence follows. □

The limit *map* $\mathcal{E}_\nu(t)$ of $\mathcal{E}_{\gamma_i}(t)$, $\gamma_i \to \nu$, will be called the time t earthquake *map*.

COROLLARY 2.8. $\mathcal{E}_\nu(t)$ *is an isometry off* ν *and is continuous off the atomic part of* ν.

Proof. Around any point $x \notin \nu$ there is a neighborhood V which misses ν and, for any $\varepsilon > 0$, there is a neighborhood U of ν in $\mathcal{ML}$ where the measure of $\mu \in U$ across every geodesic arc in V is less than ε. Thus the distortion of the metric in V becomes arbitrarily small and in the limit $\mathcal{E}_\nu(t)$ is an isometry there.

If x is on the diffuse part of ν then, for any $\varepsilon > 0$, there is a neighborhood W around x, depending on ε, where the measure of ν across all geodesic arcs in W is less than $\frac{1}{2}\varepsilon$. For U around ν small enough, the measure is less than ε for all $\mu \in U$. Thus, under $\mathcal{E}_{\gamma_i}(t)$, $\gamma_i \in U \cap (S \times \mathbf{R}_+)$, points in W are sheared less than $Kt\varepsilon$ by Lemma 1.2 and the triangle inequality, so that continuity follows. □

Remark. It should be pointed out that for each $x \in D$ not on the atomic part of ν, $\mathcal{E}_{\gamma_i}(t)x$ converges uniformly for $t \le T$ so $\mathcal{E}_\nu(t)x$ is continuous in t. Also if x is on the atomic part, it lies on a simple closed geodesic ϕ (contained in ν) with measure μ. Now $\mathcal{E}_\nu(t)$ will just be the $t\mu$-twist map along ϕ in a neighborhood of ϕ.

Section III

In this section we study the geodesic length function, particularly its behavior under earthquake deformations. The lengths of a finite number of simple closed curves completely determine a hyperbolic surface. Any unbounded sequence of points in T_g has the property that some geodesic is becoming infinitely long because either some length is becoming infinite or going to zero; in the latter case any curve intersecting the short one is becoming infinitely long. Therefore, any subset of T_g on which the length of every simple closed geodesic is bounded is a bounded subset [1].

Definition. A collection of simple closed curves $\underline{\gamma} = \{\gamma_i\}$ is said to *fill up* M^2 if, whenever they have the minimal number of pairwise intersections, $M^2 - \bigcup \gamma_i$ is a union of disks. Equivalently, for any $\nu \in \mathcal{ML} - 0$, $i(\nu, \gamma_i) \neq 0$ for some i.

LEMMA 3.1. *If $\underline{\gamma} = \{\gamma_i\}$ fills up M^2, then the function $l_{\underline{\gamma}}\colon T_g \to \mathbf{R}$ which assigns to a hyperbolic surface the sum of the geodesic lengths of the γ_i is proper. In particular it realizes a minimum in T_g.*

Proof. It suffices to show that the set $B_{\underline{\gamma}}(K) = \{x \in T_g | l_{\underline{\gamma}}(x) \leq K\}$ is bounded for any constant K. Since $\underline{\gamma}$ fills up M^2, any $\phi \in S$ can be homotoped to a curve in $\bigcup \gamma_i$ which covers no point in $\bigcup \gamma_i$ more than N times. N depends on ϕ but not on x. Therefore $l_\phi(x) \leq N l_{\underline{\gamma}}(x)$ so $l_\phi(x)$ is bounded on B_k for all $\phi \in S$. $\square$

The relationship between a geodesic curve on the initial surface M_0 of an earthquake path and the corresponding geodesic on M_t is quite complicated even in the case of a discrete shearing lamination. However, the first order variation of length is easily computed due to the fact that it is not affected by corners appearing in the geodesic. (See Lemma 3.2 below.)

Fix an isotopy class of $\gamma \in S$ and consider the time t earthquake along a closed geodesic ϕ. Assume $i(\phi, \gamma) \neq 0$. To measure the length of the geodesic in M_t isotopic to γ (call it $\gamma(t)$), look at the pre-image in M_0 of $\gamma(t)$ under the time t earthquake map. It is disconnected at each intersection with ϕ into geodesic arcs whose endpoints are distance t apart along ϕ. At time $t = 0$, it is the geodesic in M_0, and the arcs move continuously apart as t varies. Since the earthquake map is an isometry off the lamination, the sum of lengths of these arcs at time t is equal to $l(t)$, the length of $\gamma(t)$.

Look at the universal cover of M_0 and at a lift of γ (call it γ also). Note that γ intersects n lifts of ϕ (call them l_i) at points x_i with angle θ_i (measured counter-clockwise from γ to l_i). Let i run from 1 to $n+1$ with x_1 identified with x_{n+1} by γ. Consider any C^1-family $\mathcal{G}$ of geodesic arcs $A_i(t)$ from l_i to l_{i+1} whose time zero position is equal to γ and which map to a closed curve isotopic to γ in M_t. Denote the sum of the lengths of these arcs by $l_{\mathcal{G}}(t)$.

LEMMA 3.2.
$$\frac{dl_{\mathcal{G}_{(0)}}}{dt} = \sum_{i=1}^{n} \cos\theta_i.$$

Proof. The arc $A_i(t)$ is determined by points $a_i(t)$ on l_i and $b_{i+1}(t)$ on l_{i+1} which are parametrized by their directed distance from the initial points x_i, x_{i+1}. The condition that the arcs map to a closed curve in M is equivalent to the condition $b_{i+1}(t) - a_{i+1}(t) = t$. The first derivative of the lengths of the $A_i(t)$ at $t = 0$ is $-\cos\theta_i \dfrac{da_i}{dt} - \cos(\pi - \theta_{i+1})\dfrac{db_{i+1}}{dt}$. (This is an easy application of the first variation formula or can be derived in $\mathbf{H}^2$ by differentiating the side-angle formula for hyperbolic triangles ([17], Chapter 2).) The lemma follows from $\dfrac{db_{i+1}}{dt} = 1 + \dfrac{da_{i+1}}{dt}$ and $\cos(\pi - \theta_{i+1}) = -\cos\theta_{i+1}$.

COROLLARY 3.3. $\dfrac{dl_\gamma}{dt}(0) = \displaystyle\sum_{i=1}^{n} \cos\theta_i$ *along the earthquake path* $\mathcal{E}_\phi(t)$, $\phi \in S$.

COROLLARY 3.4. $\dfrac{dl_\gamma}{dt}(0) = \int_\gamma \cos\theta\, d\mu$ *along a general earthquake path* $\mathcal{E}_\mu(t)$, $\mu \in \mathcal{ML}$ *where θ is the function measuring the angle from γ to μ.*

Proof. Let $\phi_i \in S \times \mathbf{R}_+$ be such that $\phi_i \to \mu$ in $\mathcal{ML}$. Denote by $C_i(M)$, $M \in T_g$, the total cosine (see Section IB) between ϕ_i and γ on M. Corollary 3.3 says that $\dfrac{dl_\gamma}{dt}(0) = C_i(M)$ along $\mathcal{E}_{\phi_i}(t)$ when $\mathcal{E}_{\phi_i}(0) = M$. Similarly let $C(M) = \int_\gamma \cos\theta\, d\mu$ be the total cosine between μ and γ on M. Recall that this integral exists (Section IB) because for any ε there are sufficiently small subarcs of γ such that the difference between the average angle and any given angle is less than ε.

It follows that the difference between the integral and its Riemann sum over these subarcs is less than $i(\gamma, \mu)\varepsilon$. Note that since the angle between two geodesics is a continuous function of T_g, $C(M)$ is a continuous function on M.

By choosing N large enough, the ϕ_i $(i > N)$ can be taken to be ε-close to μ on these subarcs. Then $|C_i(M) - C(M)|$, $i > N$, is less than a constant times ε. By continuity of angles again, this estimate holds uniformly on compact subsets of T_g (with the subarcs smaller and N larger if necessary).

If $C_i(t) = C_i(\mathcal{E}_{\phi_i}(t))$ and $C(t) = C(\mathcal{E}_\mu(t))$, we have shown that $C_i(t) \to C(t)$ uniformly for $0 \leq t \leq T$ since $\mathcal{E}_{\phi_i}(t) \to \mathcal{E}_\mu(t)$ uniformly in T_g by Proposition 2.4. Therefore $\int_0^t C_i(s)\, ds \to \int_0^t C(s)\, ds$. If $l_i(t)$ denotes l_γ along $\mathcal{E}_{\phi_i}(t)$, then it follows that $l_i(t) \to l_\gamma(t)$, $l_\gamma(t) = l(0) + \int_0^t C(s)\, ds$ and $\frac{dl_\gamma}{dt}(t) = C(t)$. Corollary 3.4 is the case where $t = 0$. □

Having computed the first derivative of the length function, one can study the variation of this derivative along earthquake paths.

PROPOSITION 3.5. *For every intersection of $\gamma \in S$ with $\mu \in \mathcal{ML}$, the angle of intersection $\theta(t)$ from γ to μ is strictly decreasing as a function of t along the earthquake path $\mathcal{E}_\mu(t)$. Equivalently,* $\cos\theta(t)$ *is strictly increasing.*

The key point in the proof of Proposition 3.5 is that, although it is difficult to describe exactly the position of $\gamma(t)$, it is possible to describe the endpoints on S^1_∞ of any lift of $\gamma(t)$ to $\mathbf{H}^2$. Before beginning the proof it is useful to have a formula relating endpoints of geodesics and the angle between them.

Definition. The *cross-ratio*, $\chi(a, b, c, d)$, of four points a, b, c, d in the extended complex plane $C \cup \infty$ is defined to be $\dfrac{(a-c)(b-d)}{(a-d)(b-c)}$.

The cross-ratio is invariant under linear fractional transformations. Given two intersecting geodesics g_1 and g_2, translate them so that their point of intersection is the origin and their endpoints are ± 1 and $\pm e^{i\phi}$ respectively. Calculation shows that

$$\chi(1, e^{i\phi}, -e^{i\phi}, -1) = \cos^2(\phi/2) \tag{1}$$

where ϕ is the angle from g_1 to g_2.

Proof of Proposition 3.5. Isolate a point x of intersection between γ and μ and choose a lift of it to $\mathbf{H}^2$. Denote the leaf through x of the lifted lamination μ by l and assume that l and x are fixed under all deformations. (In the case when μ is discrete at l, fix one copy of x.) This can be done by isometries so that it does not affect the calculation of angles. Now $\theta(t)$ is the angle $\gamma(t)$ makes with l in the universal cover of M_t.

Although we cannot describe $\gamma(t)$ we can describe a curve in M_t homotopic to it. Since M_t is compact, the homotopy moves points at most a bounded distance so that such a curve will have the same endpoints as $\gamma(t)$. The image $\bar\gamma(t)$ of $\gamma(0)$ under the earthquake map is homotopic to $\gamma(t)$, if, as before, at discrete points of μ, where the image of $\gamma(0)$ is disconnected, the two images of the point of discontinuity are joined along the leaves of μ. The relationship between the endpoints of $\bar{\gamma}(t)$ and those of $\gamma(0)$ is contained in the following lemma:

LEMMA 3.6. *The endpoints of $\bar{\gamma}(t)$ are strictly to the left (as viewed from the point x) of the endpoints of $\gamma(0)$ for all $t > 0$.*

The proposition follows immediately from Lemma 3.6 since it implies that the cross-ratio $\chi(p_1, e_1(t), e_2(t), p_2)$ is strictly decreasing, where p_i and $e_i(t)$ are the endpoints of l and $\gamma(t)$ respectively, ordered so that $p_1, e_1(t), p_2, e_2(t)$ are in counter-clockwise order around S^1_∞. (Normalize with $p_i = \pm 1$ and compute directly.) From (1) it follows that $\cos^2(\phi(t)/2)$ is strictly decreasing where $\phi(t)$ is the angle from μ to $\gamma(t)$. Since $\theta(t) = \pi - \phi(t)$ and $\phi(t)$ is strictly increasing, $\theta(t)$ is strictly decreasing as was to be shown.

Proof of Lemma 3.6. First consider the discrete case, $\mu \in S \times \mathbf{R}_+$. Then $\bar{\gamma}(t)$ is the union of arcs coming from $\gamma(0)$ under the earthquake map and pieces of leaves of μ. One such arc A_0 passes through the point x. If A_0 is continued to a bi-infinite geodesic, its endpoints will be precisely those of $\gamma(0)$. Move along $\bar{\gamma}(t)$ in one direction until coming to the next arc A_1. If this arc is continued in the forward direction, its endpoint will be strictly to the left of the forward endpoint of A_0 (see Figure 3). This follows from the fact that the angles that A_1 and A_0 make with the leaf of μ joining them are the same so that, were they to intersect (even at infinity), they, together with the leaf of μ, would form a triangle whose angle sum is at least π. This is impossible by the Gauss-Bonnet Theorem.

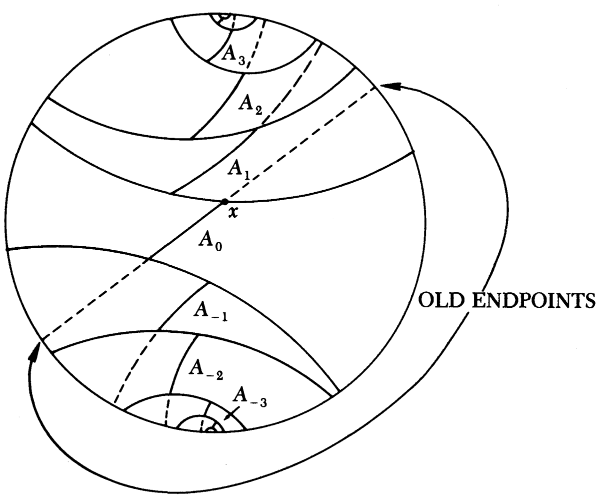

FIGURE 3

Similarly, the forward endpoint of the next arc, A_2, is to the left of the forward extension of A_1 and so on. In fact, at each stage, the entire remaining forward piece of $\gamma(t)$ is to the left of the forward extension of A_i. On the other hand, this piece of $\bar{\gamma}(t)$ and the remaining A_i's all lie in the half space determined by the corresponding leaf μ. These leaves form a nested set. Therefore, the forward endpoints of the A_i's converge to the forward endpoint of $\bar{\gamma}(t)$ which is then seen to be to the left of that of $\gamma(0)$ as claimed. The same argument shows that the other endpoint is to the left of that of $\gamma(0)$ also.

It follows, by taking a limit, that, in the general case, the endpoints of $\bar{\gamma}(t)$ are never to the right of those of $\gamma(0)$, but it is necessary to rule out the possibility that one of them is the same. To this end it suffices to estimate the change in the cross-ratio caused by shearing across a small interval purely in terms of the transverse measure and angles (hence independently of the number of leaves). Then the argument goes through in the limit.

By sending the unit disk to the upper half plane, the endpoints of the shearing geodesic l to 0 and ∞, we see that shearing distance t along l multiplies the quantity $(1/\chi) - 1$ by e^t where $\chi = (\infty, b, c, 0)$ and b, c are the endpoints of a geodesic crossing l. It follows that there is a constant C_θ, depending only on χ (and hence on the angle θ of intersection), such that shearing distance t decreases χ by at least $C_\theta t$ for small t. Since nearby leaves l_i intersect the geodesic in almost the same angle as l does, their cross-ratios χ_i with their endpoints will have bounded ratios with χ. It follows that shearing a total distance t along all of the l_i's increases χ by at least $\bar{C}_\theta t$ for some new constant $\bar{C}_\theta$ depending only on θ and the closeness of the l_i to l and not on the number of leaves or amount of shearing done on any individual leaf. This completes the proof of Lemma 3.6 and Proposition 3.5. □

Definition. A function $f\colon T_g \to \mathbf{R}_+$ is *convex along earthquake paths* if for any earthquake path $\mathcal{E}_\nu(t)$, $t \in (0, 1)$,

$$f \circ \mathcal{E}_\nu(t) \leq tf \circ \mathcal{E}_\nu(0) + (1 - t)f \circ \mathcal{E}_\nu(1).$$

It is strictly convex if the strict inequality holds.

THEOREM 1. *The geodesic length function, l_γ, of a simple closed curve γ is convex along any earthquake path $\mathcal{E}_\nu(t)$. It is strictly convex if and only if $i(\gamma, \nu) \neq 0$.*

Proof. If $i(\gamma, \nu) = 0$, l_γ is a constant which is convex but not strictly convex. If $i(\gamma, \nu) \neq 0$, the first derivative of l_γ along $\mathcal{E}_\nu(t)$ is strictly increasing since, by Proposition 3.5, each term in the integral expression for the derivative (Corollary 3.4) is strictly increasing. Thus $l_\gamma(t)$ is strictly convex. □

Remark. A quick look at the proofs above shows that Proposition 3.5 and Theorem 1 are true for any (not necessarily simple) closed curve. Proposition 3.5 is true for every leaf of a lamination crossing the leaves of ν (so is Theorem 1 once the length of a lamination is defined). These facts will be used in the appendix.

The following Earthquake Theorem shows that there are sufficient earthquake paths to make Theorem 1 useful. A proof can be found in the appendix to this paper.

THEOREM 2 (Thurston). *For every $x, y \in T_g$ there is a unique left earthquake path from x to y.*

Remark. Due to a striking formula of Wolpert [21] for the second order variation of the length function under twist maps, an alternative proof of the convexity result of this section is now possible. The point is that the second derivative is positive along this dense set of earthquake paths and can be bounded away from zero to get convexity in the limit. Moreover, the formula is linear so that, as in Corollary 3.4, it gives a formula for the second derivative along any earthquake path (where a finite sum is replaced by an integral) and shows that $l_\gamma(t)$ is C^2. It is possible to show that $l_\gamma(t)$ is C^∞ directly, but formulae for the higher order derivatives seem to be unknown.

Section IV

In this section we pull together the results of the previous sections to prove the main theorems of this paper. Then we indicate how the proofs can be adapted to the case of a general Teichmüller space.

THEOREM 3. *If a collection of simple, closed curves $\underline{\gamma}$ fills up M^2, then the length function $l_{\underline{\gamma}}$ has a unique minimum in T_g.*

Proof. By Lemma 3.1, $l_{\underline{\gamma}}$ attains a minimum in T_g. Suppose there were two minima, x and y. By the Earthquake Theorem (Theorem 2) there is an earthquake path from x to y. By Theorem 1, $l_{\underline{\gamma}}$ is the sum of functions which are convex along earthquake paths. Since $\underline{\gamma}$ fills up M^2, at least one of the curves in $\underline{\gamma}$ has non-zero intersection number with the lamination μ determining the earthquake path. It follows that at least one of the functions in the sum is strictly convex so that $l_{\underline{\gamma}}$ is strictly convex along the earthquake path from x to y. Thus x equals y and the minimum is unique. □

THEOREM 4. *Every finite subgroup G of* Mod_g, *acting on T_g, has a fixed point.*

Proof. Let $\underline{\gamma}$ be a collection of simple closed curves which fill up M^2. Then the orbit $G\underline{\gamma}$ of this collection also fills up the surface. Since $G\underline{\gamma}$ is G-invariant, the length function $l_{G\underline{\gamma}}$ is G-invariant. By Theorem 3, $l_{G\underline{\gamma}}$ has a unique minimum. Since $l_{G\underline{\gamma}}$ is G-invariant, this minimum is a fixed point. □

The case of a general, finitely generated Fuchsian group Γ (or, equivalently, of a hyperbolic manifold with finitely many geodesic boundary components, branch points, and cusps) can be dealt with in precisely the same way as those with a closed surface as quotient space once the proper definitions are established.

The Teichmüller space $T_{g,n,b}$ is the space of hyperbolic structures on a surface of genus g with n branch points and cusps and b geodesic boundary components of length 1. There are many different such spaces for fixed g, n, b depending on the orders $\nu_1, \ldots, \nu_k$ of the branch points and the number of cusps ($\nu_i = \infty$ in this case). Thus the notation $T_{g,n,b}$ assumes that "n" stands for a collection of n integers (some possibly infinite). In each case $T_{g,n,b}$ is an open ball of dimension $6g - 6 + 2(b + n)$. The boundary condition of fixed geodesic length is different from the one used in the complex analytic approach but is natural in our hyperbolic context.

We will refer to such hyperbolic surfaces with their singular structures as hyperbolic orbifolds, $M^2_{g,n,b}$, following Thurston ([17], Chapter 13). The modular group $\mathrm{Mod}_{g,n,b}$ can be defined topologically as the group of diffeomorphisms of a surface of genus g with b boundary components which fix n distinguished points up to isotopy through maps with the same property. $\mathrm{Mod}_{g,n,b}$ acts on $T_{g,n,b}$ as before by pulling back metrics.

$M^2_{g,n,b}$ is the quotient space of a convex subset of $\mathbf{H}^2$ by a Fuchsian group Γ and $T_{g,n,b}$ can be equivalently defined as the space of discrete, faithful representations (up to conjugacy) of Γ in $I(\mathbf{H}^2)$ having $M^2_{g,n,b}$ as quotient space. $\mathrm{Mod}_{g,n,b}$ is then the group of "allowable" outer automorphisms of Γ acting as before on the space of representations. The algebraic criterion for the quotient space to be correct and for the automorphisms to be "allowable" (i.e., induced from a homeomorphism of $M^2_{g,n,b}$) is essentially that the characterization of elements as hyperbolic, parabolic, elliptic of order n, and of boundary elements be preserved. Further restrictions are necessary when $b \neq 0$. For details on the algebraic definitions of allowable representations and automorphisms as well as on the equivalence of the two definitions of $T_{g,n,b}$ and $\mathrm{Mod}_{g,n,b}$, see [4].

The correct space, $\mathcal{ML}_0$, of measured laminations in this case consists of those on the hyperbolic orbifold which do not intersect the distinguished set (i.e., the singular set and the boundary). The space is closed because there are neighborhoods of the distinguished set with the property that any *simple* geodesic entering the neighborhood intersects the distinguished set. This can be proved fairly easily by looking at local models of neighborhoods of fixed sets of the distinguished elements. $\mathfrak{M}\mathfrak{L}_0$ is the space of measured laminations with compact support in the hyperbolic orbifold minus the distinguished set and is a Euclidean space of dimension $6g - 6 + 2(b + n)$ (see [17], Chapter 9). Simple closed curves are again dense in the corresponding projective space.

Left earthquakes are defined as before, and, since the laminations do not intersect the distinguished set, they preserve the orbifold structure. By the same proof as before one can show that every pair of points in $T_{g,n,b}$ can be connected by a unique left earthquake. Simple closed curves $\{\gamma_i\}$ fill up $M^2_{g,n,b}$ if every lamination intersects one of them or, equivalently, if their complement consists of disks, punctured disks with a single distinguished point or annuli with a single distinguished boundary component. The proofs of the paper then go through without modification.

Remark. It should be pointed out that the Realization Problem was previously known to be solved in the case $n + b > 0$ (see Zieschang [22] and Neumann-Raymond [13]). Furthermore, this case follows easily from the results of this paper $(n + b = 0)$ by topological considerations (compare Theorem 6 below). However, the point of the previous discussion was to indicate that the methods (hence the geometric information about lengths of geodesics) carry over to this case as well.

The case of a non-orientable surface cannot be proved directly in the same way because left earthquakes cannot be defined (since "left" cannot be defined). However, the theorems above which are proved for oriented surfaces allow orientation-reversing diffeomorphisms. A finite group G of isotopy classes of diffeomorphisms of a non-orientable surface can be lifted to a (larger) finite group H of isotopy classes of diffeomorphisms of the oriented double cover.

H contains a new element of order two which commutes (in the modular group) with the rest of the group. Since H can be realized as a group $\overline{H}$ of isometries, the isometry T representing this element commutes with all of $\overline{H}$. Therefore $\overline{H}$ descends to a group of isometries isomorphic to G acting on the quotient space under the action of T which is necessarily the original non-orientable surface.

Section V

This section is intended to point out some fairly immediate consequences or restatements of Theorem 4.

THEOREM 5. *Every finite subgroup G of* Mod_g *can be realized as a group of isometries of some hyperbolic structure on a surface of genus g.*

Proof. By Theorem 4, G has a fixed point, M, when acting on T_g. It follows that, for each $g \in G$, there is an isometry of M to itself in the isotopy class of g. The important point is that this isometry is unique because if there were two, then, by composing one with the inverse of the other, we could find an isometry of M isotopic to the identity but not equal to the identity. Such an isometry would have a lift to $\mathbf{H}^2$ which would commute with every element of $\pi_1 M$ (acting on $\mathbf{H}^2$). This would imply that every element of $\pi_1 M$ had the same endpoints, which is clearly absurd.

Similarly, the group of isometries $H \subset I(M)$ of M generated by choosing the unique isometry in each class of $g \in G$ is isomorphic to G because any word in H which represents the trivial word in G is an isometry isotopic to the identity, hence equal to the identity. □

For any manifold M there is a map $\pi\colon \mathrm{Diff}(M) \to \pi_0 \mathrm{Diff}(M)$ and a standard question in topology is whether or not it is possible to lift $\pi_0 \mathrm{Diff}(M)$ back into $\mathrm{Diff}(M)$, i.e., to choose a representative in $\mathrm{Diff}(M)$ for each element in $\pi_0 \mathrm{Diff}(M)$ so that any word in the lifted elements isotopic to the identity equals the identity. For example, this is made possible when $M = T^2$ by choosing the unique linear map in each class. Theorem 5 implies that the lifting problem is solvable for *finite* subgroups G when M is a surface of arbitrary genus. The general case is still far from being understood.

THEOREM 6. *The lifting problem for* $\pi\colon \mathrm{Diff}(M) \to \pi_0 \mathrm{Diff}(M)$ *is solvable for finite* $G \subset \pi_0 \mathrm{Diff}(M)$ *when M is a surface of genus g.*

It should be pointed out that, although Theorem 6 looks somewhat weaker than Theorem 5, it is, in fact, equivalent to it. For if we have a finite *group* of diffeomorphisms we can take any metric on M and average it over G to get an invariant metric. In particular, G acts as a group of *conformal* maps of the averaged Riemann surface. Conformal maps of the unit disk are just linear fractional transformations which in turn are precisely the isometries of $\mathbf{H}^2$. Similarly there is a unique hyperbolic surface in each conformal class so that G can be made to act by isometries on the hyperbolic surface corresponding to the averaged surface.

Finally, since the modular group has a purely group theoretic definition, one expects a group theoretic version of Theorem 4. We define a Fuchsian group to be a finitely generated, discrete subgroup Γ of $I(\mathbf{H}^2)$ which is not cyclic. The quotient space may have finite order fixed points, cusps, and boundary. (It is an orbifold, in the terminology of the previous section.) In these cases, the automorphism group of Γ, Aut Γ, is defined as the group of allowable automorphisms as discussed in the previous section. It is well-known that Fuchsian groups have trivial center (otherwise the whole group would have a common fixed point and thus be cyclic). In fact the centralizer of every element is cyclic.

Given a finite extension $\bar{\Gamma}$ of a group Γ by a finite group G,

$$1 \to \Gamma \to \bar{\Gamma} \to G \to 1,$$

there is a homomorphism from $\bar{\Gamma}$ to $\operatorname{Aut}(\Gamma)$ defined by sending $\bar{\gamma}$ to the automorphism of Γ induced by conjugation by $\bar{\gamma}$. It is not hard to see that this descends to a homomorphism $\phi\colon G \to \operatorname{Out}(\Gamma)$. Conversely, given a homomorphism ϕ from G to $\operatorname{Out}(\Gamma)$, there is a unique group extension inducing ϕ if the center of Γ is trivial. (The general theory here is that the obstruction to existence of such an extension lies in $H^3(G, Z(\Gamma))$ and uniqueness is measured by $H^2(G, Z(\Gamma))$, where $Z(\Gamma)$ is the center of Γ. See, e.g. [10].) We emphasize that in our case only allowable automorphism are considered.

Then Theorem 4 (actually its generalization as discussed in the previous section) can be restated as follows:

THEOREM 7. *An allowable extension* $\bar{\Gamma}$ *of a Fuchsian group by a finite group* G *is again a Fuchsian group if and only if the induced homomorphism* $\phi\colon G \to \operatorname{Out}(\Gamma)$ *is injective.*

Proof. If ϕ is injective, then G can be thought of as a finite subgroup of $\operatorname{Out}(\Gamma)$ and thus, by Theorem 5, can be realized as a group of isometries acting on $\mathbf{H}^2/\Gamma = M$. Hence $\bar{\Gamma}$ is isomorphic to the Fuchsian group associated to the quotient space M/G.

If ϕ is not injective, it is straightforward to show that some element of $\bar{\Gamma}$ has centralizer which is not cyclic. This is impossible in a Fuchsian group. □

Remark. Finite extensions of 2-dimensional Euclidean and spherical groups have been understood for some time so that Theorem 7 can be generalized to a theorem about all 2-dimensional geometric groups using known results (see [20]).

A similar argument can be used to conclude some new results about compact 3-dimensional Seifert fiber spaces M^3 (see [5] or [16] for a detailed description of these spaces). These manifolds are the union of circles which are the fibers of a "fibration" with finitely many "singular" fibers over a 2-manifold. A regular fiber generates a normal cyclic subgroup N of $\pi_1 M^3$ and the quotient group $\pi_1 M^3/N \approx \Gamma$ is a 2-dimensional geometric group (i.e., the fundamental group of a 2-dimensional orbifold). When Γ is hyperbolic (hence a Fuchsian group) we will say that M^3 is of *hyperbolic type*. In this case $N \approx \mathbf{Z}$ is a characteristic subgroup of $\pi_1 M^3$.

THEOREM 8. *Any 3-manifold $\overline{M}^3$ finitely covered by a Seifert fiber space M^3 of hyperbolic type is homotopy equivalent to a Seifert fiber space (also of hyperbolic type).*

Proof. The proof is essentially group theoretic diagram chasing, using Theorem 7, and the fact that N is characteristic.

We can assume that $\pi_1 M^3$ is a normal subgroup of $\pi_1 \overline{M}^3$ of finite index by looking at the intersection of the conjugates of $\pi_1 M^3$. The corresponding covering space is again a Seifert fiber space since the pre-images of the fibers of M^3 are again circles. Denote $\pi_1 M^3$ by π and $\pi_1 \overline{M}^3$ by $\bar{\pi}$. The hypotheses imply the existence of the following diagram:

$$\begin{array}{ccccccccc} & & 1 & & & & & & \\ & & \downarrow & & & & & & \\ & & \mathbf{Z} & & & & & & \\ & & \downarrow & & & & & & \\ 1 & \rightarrow & \pi & \rightarrow & \bar{\pi} & \rightarrow & G & \rightarrow & 1 \\ & & \downarrow & & & & & & \\ & & \Gamma & & & & & & \\ & & \downarrow & & & & & & \\ & & 1 & & & & & & \end{array}$$

where Γ is Fuchsian and G is finite.

G acts by outer automorphism on π. Since $N \approx \mathbf{Z}$ is a characteristic subgroup, the action descends to an action on $\pi/\mathbf{Z} \approx \Gamma$ which gives rise to a short exact sequence:

$$1 \rightarrow \Gamma \rightarrow \overline{\Gamma} \rightarrow G \rightarrow 1$$

where $\bar{\pi}$ maps onto $\overline{\Gamma}$ with kernel $\mathbf{Z}$. Assume that G acts on Γ with trivial kernel. Then, by Theorem 7, $\overline{\Gamma}$ is again Fuchsian. Thus $\bar{\pi}$ is an infinite cyclic extension of a Fuchsian group and so is isomorphic to the fundamental group of a Seifert fiber space of hyperbolic type. Since $\overline{M}^3$ is a $K(\bar{\pi}, 1)$ group, it is homotopy equivalent to such a fiber space.

In the case when the kernel K of the action of G on Γ is non-trivial, we look at the covering space $\tilde{M}$ whose fundamental group $\tilde{\pi}$ is a normal subgroup of $\bar{\pi}$ with quotient group $G' = G/K$. As above, $\tilde{\pi}$ maps onto $\tilde{\Gamma}$ which fits into the short exact sequence:

$$1 \rightarrow \Gamma \rightarrow \tilde{\Gamma} \rightarrow K \rightarrow 1.$$

Since K acts trivially on Γ and the center of Γ is trivial, $\tilde{\Gamma} \approx \Gamma \times K$. Then $\tilde{\pi}$ maps onto Γ with kernel F giving the diagram below (without the dotted arrows).

Standard diagram chasing allows the completion of the diagram.

$$
\begin{array}{ccccccccc}
 & & 1 & & 1 & & 1 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \to & \mathbf{Z} & \cdots\cdots > & F & \cdots\cdots > & K & \to & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \to & \pi & \longrightarrow & \tilde{\pi} & \longrightarrow & K & \to & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \to & \Gamma & \longrightarrow & \Gamma & \longrightarrow & 1 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 1 & & 1 & & & &
\end{array}
$$

Since $\tilde{\pi}$ is torsion free ($\tilde{M}^3$ is a $K(\tilde{\pi}, 1)$), F is a finite, torsion-free extension of $\mathbf{Z}$ and hence equal to $\mathbf{Z}$. (Hence K is cyclic and the map $\mathbf{Z} \to F$ is just $z \to kz$ for some $k \in \mathbf{Z}$. This can be interpreted geometrically as the regular fibers of M^3 being wrapped k times around the "fiber" of $\tilde{M}^3$ (if $\tilde{M}^3$ is Seifert fibered).) It follows that $\tilde{M}^3$ is homotopy equivalent to a Seifert fiber space. Using the first argument for G' acting on $\tilde{M}^3$, we conclude that $\overline{M}^3$ is also. $\square$

Remark. For sufficiently large 3-manifolds, homotopy equivalence (preserving the peripheral structure, if the boundary is non-empty) implies homeomorphism, so that if $\overline{M}^3$ is sufficiently large, it is homeomorphic to a Seifert fiber space. The case when $\overline{M}^3$ is sufficiently large has been known for some time. (See [6], [19].)

Remark. Quite recently Peter Scott [16] has shown that for Seifert fiber spaces with infinite π_1, homotopy equivalence implies homeomorphism so that the $\overline{M}^3$ above are actually homeomorphic to Seifert fiber spaces in general.

Finally, it should be pointed out that an independent proof of the fact that 3-manifolds finitely covered by Seifert fiber spaces of hyperbolic type are again Seifert fibered would give another solution to the Nielsen problem. Briefly, the argument is that, since Mod_g acts on the circle at infinity of $\mathbf{H}^2$, it acts (up to a choice of lifting) on the set of ordered, distinct triples, O, of points on S^1_∞. Now O is identified with $T_1(\mathbf{H}^2)$ (actually two copies of it), O/Γ is identified with $T_1(M^2)$, and the action of Mod_g is well-defined on O/Γ. Any finite $G \subset \mathrm{Mod}_g$ acts on O/Γ and the action is free by the Nielsen problem for cyclic groups. Thus $G \setminus O/\Gamma$ is finitely covered by $T_1(M^2)$ which is Seifert fibered. The extension of Γ by G is Fuchsian if and only if $G \setminus O/\Gamma$ is again Seifert fibered. In fact, the quotient 3-manifold is the unit tangent bundle of the quotient space M^2/G.

The argument above was pointed out to the author by Thurston, but seems to have been known before. (However, a precise origin was never determined.) It has been independently discovered more recently by Neumann and Raymond [13].

Appendix

In order for the convexity result of Section III to imply the existence of a unique minimum, it is necessary to connect every two points by an earthquake path. This fact is contained in the theorem below, due to Thurston, which says that there is a unique such path from x to y in T_g if one considers only *left* earthquakes. This theorem is a generalization of the statement that there is a unique "left" horocycle from x to y in $\mathbf{H}^2$ (which is equal to T_1). Note that in both cases the relation is not symmetric with respect to x and y.

The following proof is the same in outline, if not in detail, as that given by Thurston in a course at Princeton University during 1976–7. Since no written proof exists, the author feels obliged to prove it here.

THEOREM 2 (Thurston). *For every $x, y \in T_g$ there is a unique left earthquake from x to y.*

Proof. The idea of the proof is to fix a point x in T_g and consider the earthquake "exponential" map from x, $\mathcal{E}_x$: $\mathcal{ML} \to T_g$ which sends μ to the image of x under the time 1 earthquake map corresponding to μ. For any x, $\mathcal{E}_x$ will be shown to be a homeomorphism which clearly implies the theorem.

Since our study of $\mathcal{E}_x$ concentrates on the behavior of the 1-parameter subsets of $\mathcal{ML}$ in a fixed projective class, it is somewhat easier to pick one element from each projective class in a continuous fashion, identify this "section" with $\mathcal{PL}$, and think of $\mathcal{ML}$ parametrized (in "polar co-ordinates") by $\mathcal{PL}$ and a time co-ordinate $t \in [0, \infty)$. Now x is the time zero image of every lamination and the image of $\mu \times [0, \infty)$, $\mu \in \mathcal{PL}$, is the infinite earthquake path from x corresponding to μ, which we will call the ray in the direction μ. An example of such a section would be the subset of laminations whose length on x is 1, but any section will do. All laminations are assumed to be in this subset, denoted by $\mathcal{PL}$, unless stated otherwise.

Since both $\mathcal{ML}$ and T_g are homeomorphic to open $6g - 6$ dimensional balls, it suffices, by invariance of domain, to show that $\mathcal{E}_x$ is continuous, proper and $1 - 1$. That $\mathcal{E}_x$ is continuous is the content of Corollary 2.6.

Recall that for any finite set $\underline{\gamma} = \{\gamma_i\}$ of simple closed curves filling up M the length function $l_{\underline{\gamma}}$ is proper. Thus for any such set of curves and any compact set

C in T_g there is a constant K such that C is contained in the compact set $B_K(x)$ of all points y in T_g for which $l_{\gamma_i}(y) - l_{\gamma_i}(x) \leq K$ for all $\gamma_i \in \gamma$. To show that $\mathcal{E}_x$ is proper it suffices to show that $\mathcal{E}_x^{-1}(B_K(x))$ is compact for all K.

Clearly along any ray in the direction μ, such that $i(\mu, \gamma_i) \neq 0$, l_{γ_i} becomes arbitrarily long and, since the derivative of l_{γ_i} is strictly increasing, once the ray leaves B_K, it stays outside it. Since, for every μ, $i(\mu, \gamma_i) \neq 0$ for some i, every ray leaves B_K eventually. Since $\mathcal{PL}$ is compact and $\mathcal{E}_x$ is continuous, the time of departure is uniformly bounded for all μ. Thus $\mathcal{E}_x^{-1}(B_K(X))$ is compact.

To prove that $\mathcal{E}_x$ is $1 - 1$, first note that the ray in the direction of μ is embedded since for any $\gamma \in S$ such that $i(\gamma, \mu) \neq 0$, the quantity $\int_\gamma \cos\theta \, d\mu$ is an invariant of the surface and is strictly increasing along the ray. To see that two different rays μ and ν do not intersect requires two subcases.

In the case $i(\mu, \nu) \neq 0$ the average cosine of the angle θ from μ to ν, $\iint_M \cos\theta \, d\mu \times d\nu$, is strictly increasing along the ray in the direction ν since each term is strictly increasing. However, it is strictly decreasing along the ray μ. This follows from the facts that the angle τ from ν to μ is $\pi - \theta$ (so that $\cos\tau = -\cos\theta$) and that $\iint_M \cos\tau \, d\mu \times d\nu$ is strictly increasing along μ. Thus the rays are disjoint except at x.

In the case when $i(\mu, \nu) = 0$ we can assume, by adding leaves with weight zero, that μ and ν have the same leaves with different transverse measures. The idea in this case is to distinguish the rays by the length function of closed curves which are very nearly tangent ($\theta \sim 0$) to μ (and hence to ν) and which therefore have first derivatives very near their intersection numbers with μ and ν.

Suppose that $\mathcal{E}_\mu(t) = \mathcal{E}_\nu(s)$ for $s, t \neq 0$. By multiplying the measure of ν by a constant, we can assume that $s = t$. Let ϕ be a (not necessarily simple) closed geodesic such that

$$1 > \cos\theta > 1 - \varepsilon \tag{1}$$

for all angles of intersection of ϕ with μ (and ν). Since

$$\frac{dl_\phi}{dt} = \int_\phi \cos\theta \, d\mu \qquad \text{and} \qquad \frac{dl_\phi}{ds} = \int_\phi \cos\theta \, d\nu$$

it follows that

$$si(\phi, \nu)(1 - \varepsilon) < \Delta_s l_\phi < si(\phi, \nu) \qquad \text{and} \qquad ti(\phi, \mu)(1 - \varepsilon) < \Delta_t l_\phi < ti(\phi, \mu)$$

for all s, t where $\Delta_s l_\phi$ and $\Delta_t l_\phi$ are the changes in l_ϕ along $\mathcal{E}_\nu$ and $\mathcal{E}_\mu$ at times s and t respectively. Since $\Delta_s l_\phi = \Delta_t l_\phi$ when $\mathcal{E}_\mu(s) = \mathcal{E}_\nu(t)$, this and $t = s$ imply that $1-\varepsilon<\dfrac{i(\phi,\mu)}{i(\phi,\nu)}<\dfrac{1}{1-\varepsilon}$ for any such ϕ. Below it is shown that for any ε sufficiently small there always exists a ϕ satisfying (1) such that

$$\frac{i(\phi,\mu)}{i(\phi,\nu)}>\frac{1}{1-\varepsilon}, \tag{2}$$

implying a contradiction. Therefore, it will follow that $\mathcal{E}_\nu(s)\neq\mathcal{E}_\mu(t)$ when $s,t\neq 0$ as claimed.

First, suppose that μ and ν consist of only closed curves. Take any $\bar{\phi}$ satisfying (2) for some small ε (which exists since $\mu\neq\nu$) and consider the geodesic isotopic to the image of $\bar{\phi}$ under (left-handed) Dehn twists around each of the curves in μ. If the twists are of a sufficiently high order, then (1) will be satisfied. Since the intersection numbers remain unchanged, this is the required geodesic ϕ.

Now assume that μ has no closed leaves. Let A be a small transverse geodesic arc whose endpoints, e_1, e_2, are in the same connected component of μ such that $i(A,\mu)$ and $i(A,\nu)$ satisfy (2). Orient A so that it goes from e_1 to e_2 and let l be the portion of the leaf of μ leaving from e_2 to the left as viewed along A. We will see below that e_1 is the limit of points of positive intersection of l with A (where "positive" means crossing from right to left).

If we assume that e_1 is such a limit point, it is possible to go along l until it intersects A positively very close to e_1 (at p, say). The geodesic $\bar{\phi}$ from e_1 to itself, homotopic to the path from e_1 to e_2 along A, along l to p and back to e_1 along A, will have intersection numbers with μ and ν approximately equal to $i(A,\mu)$ and $i(A,\nu)$ respectively, the error depending only on how close p is to e_1. The geodesic $\bar{\phi}$ is very nearly tangent to μ, but probably has a corner at e_1. The angle, however, is small and l is long so the unique closed geodesic ϕ homotopic to $\bar{\phi}$ will be close to it and hence also very nearly tangent to μ. The intersection numbers with μ and ν are the same as those of $\bar{\phi}$.

By choice of A sufficiently small and l sufficiently long, ϕ can be made to satisfy (1) for as small an ε as desired. To see that A can be chosen arbitrarily small, note that if it is subdivided, at least one sub-interval still satisfies (2) with the same value of ε.

To see that l can be chosen arbitrarily long and p arbitrarily close to e_1, it is necessary to show that e_1 is, in fact, a limit of positive intersections of A with l. It is clearly a limit of positive and negative intersections since e_1 and e_2 are in the same connected component. (Strictly speaking, e_1 could be the lower boundary point of an interval in the Cantor set cross-section and hence approached only from below where the leaf through e_2 fails to hit A. In this case, replace e_1 with the upper boundary point of the interval without altering the intersection number.)

Suppose there is some point of negative intersection, since otherwise we are done. If any point of positive (resp. negative) intersection is the limit of points of negative (resp. positive) intersection, then every point of intersection is the limit of points of intersection of the opposite sign. (Just follow l from one point to the other; a small neighborhood of leaves of μ follows the same path.) In this case e_1 is the limit of positive intersections as claimed.

If no point is the limit of intersections of opposite sign, there are sub-intervals of positive and negative intersections. Replacing A with any such sub-interval satisfying (2), we have an interval with the required properties.

In the general case, when μ has both closed and non-closed leaves, it may be necessary both to take high order Dehn twists around closed leaves and to drag a small transverse arc along an infinite leaf. The existence of a closed curve ϕ satisfying (1) and (2) follows as before.

This completes the proof of Theorem 2. □

STANFORD UNIVERSITY, STANFORD, CALIFORNIA

BIBLIOGRAPHY

[1] A. FATHI et al., Travaux de Thurston sur les surfaces, Astérisque (Orsay Séminaire), 1979.
[2] W. FENCHEL, Estensioni di gruppi discontinui e transformazioni periodiche delle superficie, Rend. Accad. Naz. Lincei **5** (1948), 326–329.
[3] W. J. HARVEY (ed.), *Discrete Groups and Automorphic Functions*, Academic Press (London), 1977.
[4] W. J. HARVEY and C. MACLACHLAN, On the mapping class group and Teichmüller spaces, Proc. London Math. Soc. **30** (1975), 496–512.
[5] J. HEMPEL, *3-Manifolds*, Annals of Math Studies #86, Princeton Univ. Press (Princeton, N.J.), 1976.
[6] W. JACO and P. B. SHALEN, *Seifert Fibered Spaces in 3-Manifolds*, Mem. A. M. S., #220, (1979).
[7] S. P. KERCKHOFF, The Nielsen Realization Problem, Bull. A.M.S. **2** (New Series) (1980), 452–454.
[8] S. KRAVETZ, On the geometry of Teichmüller spaces and the structure of their modular groups, Ann. Acad. Sci. Fenn. **278** (1959), 1–35.
[9] M. LINCH, On metrics in Teichmüller space, thesis, Columbia University (1971).
[10] S. MACLANE, *Homology*, Springer-Verlag (Berlin), 1963.
[11] H. MASUR, On a class of geodesics in Teichmüller space, Ann. of Math. **102** (1975), 205–221.
[12] G. D. MOSTOW, *Strong Rigidity of Locally Symmetric Spaces*, Ann. of Math. Studies #78, Princeton Univ. Press, 1979.
[13] W. NEUMANN and F. RAYMOND, Automorphisms of Seifert manifolds, preprint.
[14] J. NIELSEN, Abbildungklassen Endliche Ordung, Acta Math. **75** (1943), 23–115.
[15] P. ORLIK, *Seifert Manifolds*, Lecture Notes in Mathematics, #291, Springer-Verlag (Berlin), 1972.

[16] P. Scott, There are no fake Seifert fiber spaces with infinite π_1, Ann. of Math. **117** (1983), 35–70.

[17] W. P. Thurston, The geometry and topology of 3-manifolds, notes (Princeton Univ.), to be published as: *Three Dimensional Geometry and Topology* in Lecture Note Series, Princeton Univ. Press, 1983.

[18] ______, On the geometry and dynamics of diffeomorphisms of surfaces, I, preprint.

[19] F. Waldhausen, Gruppen mit Zentrum und 3-dimensionale Mannigfaltigkeiten, Topology **6** (1947), 505–517.

[20] J. Wolf, *Spaces of Constant Curvature*, McGraw-Hill (New York), 1967.

[21] S. Wolpert, On Poincaré series and the symplectic deformation geometry of Riemann surfaces, preprint; published as: On the symplectic geometry of deformations of a hyperbolic surface, Ann. of Math. **117** (1983), 207–234.

[22] H. Zieschang, *Finite Groups of Mapping Classes of Surfaces*, Lecture Notes in Math., #875, Springer-Verlag, 1981.

(Received June 4, 1981)
(Revised August 3, 1982)