

ON THE GEOMETRY AND DYNAMICS OF DIFFEOMORPHISMS OF SURFACES

WILLIAM P. THURSTON

PREFACE

This article was widely circulated as a preprint, about 12 years ago. At that time the *Bulletin* did not accept research announcements, and after a couple of attempts to publish it, I gave up, and the preprint did not find a home. I very soon saw that there were many ramifications of this theory, and I talked extensively about it in a number of places. One year I devoted my graduate course to this theory, and notes of Bill Floyd and Michael Handel from that course were circulated for a while. The participants in a seminar at Orsay in 1976–1977 went over this material, and wrote a volume [FLP] including some original material as well. Another good general reference, from a somewhat different point of view, is a set of notes of lectures by A. Casson, taken by S. Bleiler [CasBlei].

There are by now several alternative ways to develop the classification of diffeomorphisms of surfaces described here. At the time I originally discovered the classification of diffeomorphism of surfaces, I was unfamiliar with two bodies of mathematics which were quite relevant: first, Riemann surfaces, quasiconformal maps and Teichmüller's theory; and second, Nielsen's theory of the dynamical behavior of surface at infinity, and his near-understanding of geodesic laminations. After hearing about the classification of surface automorphisms from the point of view of the space of measured foliations, Lipman Bers [Bers1] developed a proof of the classification of surface automorphisms from the point of view of Teichmüller theory, generalizing Teichmüller's theorem by allowing the Riemann surface to vary as well as the map. Dennis Sullivan first told me of some neglected articles by Nielsen which might be relevant. This point of view has been discussed by R. Miller, J. Gilman, M. Handel and me.

The analogous theory, of measured laminations and 2-dimensional train tracks in three dimensions, has been considerable development. This has been applied to reinterpret some of Haken's work, to classify incompressible surfaces in particular classes of 3-manifolds in papers by me, Hatcher, Floyd, Oertel and others in various combinations. Shalen, Morgan, Culler and others have developed the related theory of groups acting on trees, and its relation to measured laminations, to define and analyze compactifications of representation spaces of groups in $SL(2, \mathbb{C})$ and $SO(n, 1)$; this has many interesting applications, including the theory of incompressible surfaces in 3-manifolds.

Received by the editors March 3, 1987.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 58F15.

©1988 American Mathematical Society
0273-0979/88 \$1.00 + \$.25 per page

Recently, Bestvina and Handel have developed a classification of automorphisms of the free group analogous to the classification of automorphisms of the fundamental group of a surface.

Kerckhoff proved in his thesis that the modular group does not in general extend continuously to the Teichmüller boundary of Teichmüller space. Kerckhoff and I later proved that the modular group does not in general extend continuously to the Bers boundary for Teichmüller space. In particular, it follows that the measured foliation compactification defined here is in general different from these two compactifications.

The literature on this subject is now quite large, and I cannot even touch on all aspects of it here, such as algorithms, noncompact surfaces, handlebodies, measure theory, hyperbolic three-manifolds, etc.

There would be no simple stopping point if I began to incorporate the more recent developments in the original paper, so it is being published here in the original form.

The bibliography below is not exhaustive, and was not systematically assembled, so that inclusion or noninclusion of a paper should not be taken as an indication of quality or relevance of the content. Nonetheless, I hope that it will be useful.

REFERENCES

- [Anosov] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds with negative curvature*, Trudy Mat. Inst. Steklov. **90** (1967); English transl., Proc. Steklov. Inst. Math. **90** (1969).
- [ArnYoc] P. Arnoux and J. C. Yoccoz, *Construction de difféomorphismes pseudo-Anosov*, C. R. Acad. Sci. Paris Ser. I Math. **292** (1981), 75–78.
- [Bers1] L. Bers, *An extremal problem for quasiconformal mapping and a theorem by Thurston*, Acta Math. **141** (1978), 73–98.
- [Bers2] ———, *On boundaries of Teichmüller spaces and Kleinian groups*. I, Ann. of Math. (2) **91** (1970), 570–600.
- [Bers3] ———, *On iterates of hyperbolic transformations of Teichmüller space*, Amer. J. Math. **105** (1983), 1–11.
- [BirKid] J. Birman and M. Kidwell, *Fixed points of pseudo-Anosov diffeomorphisms of surfaces*, Adv. in Math. **46** (1982), 217–220.
- [Bonahon1] F. Bonahon, *Cobordism of automorphisms of surfaces*, Ann. Sci. École Norm. Sup. **16** (1983), 237–270.
- [Bonahon2] ———, *Bouts des variétés hyperboliques de dimension 3*, Ann. of Math. (2) **124** (1986), 71–158.
- [CasBlei] A. J. Casson, Notes by S. A. Bleiler, *Automorphisms of surfaces, after Nielsen and Thurston*, Preprint (1982).
- [CasLong] A. J. Casson and D. D. Long, *Algorithmic compression of surface automorphisms*, preprint.
- [CulMor] M. Culler and J. W. Morgan, *Group actions on \mathbf{R} -trees*, Proc. London Math. Soc. (to appear).
- [CulShal1] M. Culler and P. Shalen, *Varieties of group representation and splittings of 3-manifolds*, Ann. of Math. (2) **117** (1983), 43–45.
- [CulShal2] ———, *Bounded, separating surfaces in knot manifolds*, Invent. Math. **75** (1984), 537–545.
- [FatLau] A. Fathi and F. Laudenbach, *Diffeomorphismes pseudo-Anosov et décomposition de Heegaard*, C. R. Acad. Sci. Paris Ser. A-B **291** (1980), A423–A425.

- [**Fathi1**] A. Fathi, *The Poisson bracket on the space of measured foliations on a surface*, Duke J. Math.
- [**Fathi2**] ——, *Dehn twists and pseudo-Anosov diffeomorphisms*, preprint.
- [**FLP**] A. Fathi, F. Laudenbach and V. Poenaru et al., *Travaux de Thurston sur les surfaces*, Astérisque **66–67** (1979), 1–284.
- [**Fried1**] D. Fried, *Transitive Anosov flows and pseudo-Anosov maps*, Topology **22** (1983), 299–303.
- [**Fried2**] ——, *Flow equivalence, hyperbolic systems and a new zeta function for flows*, Comment. Math. Helv. **57** (1982), 237–259.
- [**Fried3**] ——, *Growth rate of surface homeomorphisms and flow equivalence*, preprint.
- [**GerKat**] M. Gerber and A. Katok, *Smooth models of Thurston's pseudo-Anosov maps*, Ann. Sci. École Norm. Sup. **15** (1982), 173–204.
- [**Gilman1**] J. Gilman, *On the Nielsen type and the classification for the mapping-class group*, Adv. in Math. **40** (1981), 68–96.
- [**Gilman2**] ——, *Determining Thurston classes using Nielsen types*, Trans. Amer. Math. Soc. **272** (1982), 669–675.
- [**Gard**] F. P. Gardiner, *Teichmüller theory and quadratic differentials*, Wiley, New York, 1987.
- [**Handel1**] M. Handel, *Global shadowing of pseudo-Anosov homeomorphisms*, Ergodic Theory and Dynamical Systems (to appear).
- [**Handel2**] ——, *Entropy and semi-conjugacy in dimension 2*, Ergodic Theory and Dynamical Systems (to appear).
- [**HanTh**] M. Handel and W. P. Thurston, *Anosov flows on new three-manifolds*, Invent. Math.
- [**HanTh**] ——, *New proofs of some results of Nielsen*, Adv. in Math. **56** (1985), 173–191.
- [**HatTh**] A. Hatcher and W. Thurston, *Incompressible surfaces in 2-bridge knot complements*, Invent. Math. **79** (1985), 225–246.
- [**Harer**] J. Harer, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. (2) **121** (1985), 215–249.
- [**HubMas**] J. Hubbard and H. Masur, *Quadratic differentials and foliations*, Acta Math. **142** (1979), 221–274.
- [**Kerckhoff1**] S. P. Kerckhoff, *The asymptotic geometry of Teichmüller space*, Topology **19** (1980), 23–41.
- [**Kerckhoff2**] ——, *The Nielsen realization problem*, Ann. of Math. (2) **117** (1983), 235–265.
- [**Kerckhoff3**] ——, *Earthquakes are analytic*, Comment. Math. Helv. **60** (1985), 17–30.
- [**KMS**] S. Kerckhoff, H. Masur and J. Smillie, *Ergodicity of billiard flows and quadratic differentials*, Ann. of Math. (2) **124** (1986), 293–311.
- [**Kra**] I. Kra, *On the Nielsen-Thurston-Bers type of some self-maps of Riemann surfaces*, Acta Math. **146** (1981), 231–270.
- [**Masur1**] H. Masur, *Interval exchange transformations and measured foliations*, Ann. of Math. (2) **115** (1982), 169–200.
- [**Masur2**] ——, *Ergodic actions of the mapping class group*, Proc. Amer. Math. Soc. **94** (1985), 455–459.
- [**Masur3**] ——, *Two boundaries of Teichmüller space*, Duke Math. J. **49** (1982), 183–190.
- [**Miller**] R. T. Miller, *Geodesic laminations from Nielsen's viewpoint*, Adv. in Math. **45** (1982), 189–212.
- [**MorShal1**] J. W. Morgan and P. Shalen, *Valuations, trees, and degenerations of hyperbolic structures. I*, Ann. of Math. (2) **120** (1984), 401–476.
- [**MorShal2**] ——, *Degenerations of hyperbolic structures. II: measured laminations, trees and 3-manifolds*, Ann. of Math. (2) **127** (1988), 403–456.
- [**MorShal3**] ——, *Degenerations of hyperbolic structures. III*, Ann. of Math. (2) **127** (1988).
- [**Mosher0**] L. Mosher, *Pseudo-Anosovs on punctured surfaces*, Princeton thesis, 1983.
- [**Mosher1**] ——, *The classification of pseudo-Anosovs*, Proceedings of Conference on Hyperbolic Geometry in Durham, Cambridge Univ. Press.

- [Mosher2] ——, *Topological invariants of measured foliations*, Preprint, 1988.
- [Nielsen1] J. Nielsen, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*. I, Acta Math. **50** (1927), 189–358.
- [Nielsen2] ——, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*. II, Acta Math. **53** (1929), 1–76.
- [Nielsen3] ——, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*. III, Acta Math. **58** (1932), 87–167.
- [Oertel] U. Oertel, *Measured laminations in 3-manifolds*, Trans. Amer. Math. Soc. **305** (1988), 531–573.
- [Pap1] A. Papadopoulos, *Diffeomorphismes pseudo-Anosov et automorphismes symplectiques de l'homologie*, Ann. Sci. École Norm. Sup. **15** (1982), 543–546.
- [Pap2] ——, *L'extension du flot de Fenchel-Nielsen au bord de Thurston de l'espace de Teichmüller*, C. R. Acad. Sci. Paris **302** (1986), 325–327.
- [Pap3] ——, *Geometric intersection functions and hamiltonian flows on the space of measured foliations on a surface*, Pacific J. Math. **124** (1986), 375–402.
- [Pap4] ——, *Sur le bord de Thurston de l'espace de Teichmüller d'une surface non compacte*, Math. Ann. (1988).
- [Pap5] ——, *On Thurston's boundary of Teichmüller space and the extension of earthquakes*, 1988 Preprint.
- [Rees] M. Rees *An alternative approach to the ergodic theory of measured foliations on surfaces*, Ergodic Theory and Dynamical Systems **1** (1981), 461–488.
- [Bass-Serre] J.-P. Serre (with H. Bass), *Trees*, Dynamical Systems, Springer-Verlag, New York, 1980.
- [Smale] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.
- [Strebel] K. Strebel, *Quadratic differentials*, Springer-Verlag, Berlin, Heidelberg, New York and Tokyo, 1984.
- [Sullivan] D. P. Sullivan, *Travaux de Thurston sur les groupes quasi-fuchsiens et les variétés hyperboliques de dimension 3 fibrées sur S^3* , Bourbaki Seminar (1979–80), 196–214, Lecture Notes in Math., vol. 842, Springer-Verlag, Berlin and New York, 1981.
- [Th0] W. P. Thurston, *The geometry and topology of three-manifolds*, Princeton Math. Dept., 1979.
- [Th1] ——, *Minimal stretch maps between hyperbolic surfaces*, Topology (to appear).
- [Veech1] W. Veech, *Gauss measures for transformations on the space of interval exchange maps*, Ann. of Math. (2) **115** (1982), 201–242.
- [Veech2] ——, *The Teichmüller geodesic flow*, Ann. of Math. (2) **124** (1986), 441–530.
- [Wald] F. Waldhausen, *On irreducible manifolds which are sufficiently large*, Ann. of Math. (2) **87** (1968), 56–88.
- [Wolp1] S. Wolpert, *On the symplectic geometry of deformations of a hyperbolic surface*.
- [Zieschang] H. Zieschang, *Finite groups of mapping classes of surfaces*, Springer-Verlag, Berlin and New York, 1980.

1. This paper is a description of some results about diffeomorphisms of surfaces, and the topology of surfaces. “Surface”, in this paper, will mean a compact, connected surface, unless otherwise noted. Most of the proofs are deferred to the future.

If M is a surface, we will denote by $\mathcal{S}(M)$ the set of all isotopy classes of two-sided simple closed curves on M , not bounding a disk and not isotopic to M . It is easy to see that, for most surfaces, elements of $\mathcal{S}(M)$ can be quite complicated (cf. Figure 1). In working with simple closed curves, one gets the sense of some geometric concept of “nearness” among them, not closely related to the homotopy class, but having to do with how many strands pass around in a certain direction. We will shortly formalize such a concept, defining a completion of $\mathcal{S}(M)$, to be denoted $\mathcal{PS}(M)$. It turns out that

$\mathcal{PS}(M)$ is homeomorphic to a sphere, and $\mathcal{S}(M)$ consists of a dense set of “rational” points on this sphere. From the coordinates of such a point, the corresponding simple closed curve can be drawn, rather mechanically, and quickly by a computer.

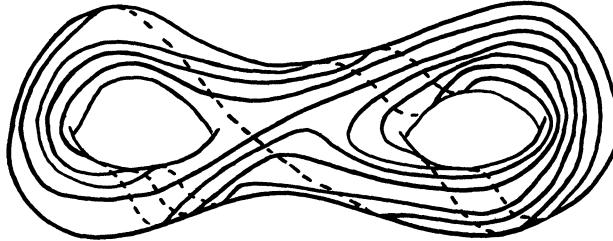


FIGURE 1. A typical simple closed curve on a surface is complicated, from the point of view of someone tracing out the curve.

$\mathcal{PS}(M)$ has two other interpretations. First, it is the space consisting of all equivalence classes of “measured foliations” on M (see §3). Second, it forms a boundary for the Teichmüller space of M , to which the action of the group of diffeomorphisms of M extends.

With the aid of this tool, a canonical representative is found for every isotopy class of diffeomorphisms of M , well-defined up to conjugacy by diffeomorphisms isotopic to the identity.

2. We denote by $M_{h,b,c}$ the surface obtained from the sphere S^2 by connected sum with h tori, b disks and c projective spaces. We will sometimes use M_h^2 to mean $M_{h,0,0}^2$.

The classification of surfaces says that every surface is an $M_{h,b,c}^2$; that $M_{x,y,z+1+2k}^2$ is synonymous with $M_{x+k,y,z+1}^2$; and that these are all the relations.

If α and β are two members of \mathcal{S} , we let $i(\alpha, \beta)$ denote their geometric intersection number, i.e., the minimum number of intersections of simple curves representing α and β .

PROPOSITION. (1) *For every $\alpha \in \mathcal{S}$ there is a $\beta \in \mathcal{S}$ such that $i(\alpha, \beta) \neq 0$.*

(2) *For every pair of distinct $\alpha, \alpha_2 \in \mathcal{S}$, there is a $\beta \in \mathcal{S}$ such that $i(\alpha_1, \beta) \neq i(\alpha_2, \beta)$.*

The map $i: \mathcal{S} \times \mathcal{S} \rightarrow \mathbf{R}$ gives a map $i_*: \mathcal{S} \rightarrow \mathbf{R}^{\mathcal{S}}$. In light of the proposition, $i_*(\mathcal{S}) \subset \mathbf{R}^{\mathcal{S}} - 0$. We regard $\mathbf{R}^{\mathcal{S}}$ as a topological vector space with the product topology. There is an associated projective space $\mathbf{P}(\mathbf{R}^{\mathcal{S}})$, with a projection

$$P: \mathbf{R}^{\mathcal{S}} - 0 \rightarrow \mathbf{P}(\mathbf{R}^{\mathcal{S}}).$$

$\mathbf{P}(\mathbf{R}^{\mathcal{S}})$ is given the quotient topology.

For fairly general reasons, $P \circ i_*(\mathcal{S}) \subset \mathbf{P}(\mathbf{R}^{\mathcal{S}})$ has a compact closure. We denote this closure $\mathcal{PS}(M_{h,b,c}^2)$.

It is immediate from the definition that the correspondence $M \mapsto \mathcal{PS}(M)$ is natural, so that a diffeomorphism ϕ of M induces a homeomorphism $\mathcal{PS}(\phi)$ of \mathcal{PS} . It is also immediate that $\mathcal{PS}(\phi)$ depends only on the isotopy class of ϕ .

We denote by $\mathcal{S}'(M)$ the set of isotopy classes of nonempty unions of disjoint two-sided simple closed curves on M , none of which bounds a disk and none of which are isotopic to ∂M . Geometric intersection number, and the space \mathcal{PS}' , are defined just as for \mathcal{S} .

THEOREM 1. *The space $\mathcal{PS}'(M)$ is equal to $\mathcal{PS}(M)$, and is topologically a sphere. If $\chi(M) < 0$ then the dimension of this sphere is $6h+2b+3c-7$. (A sphere of dimension -1 is thought of as the empty set \emptyset .) The seven exceptional cases when $M = M_{h,b,c}$ has Euler characteristic $\chi(M) > 0$ can be tabulated as follows:*

common name	h	b	c	$(M_{h,b,c})$
<i>Sphere</i>	0	0	0	\emptyset
<i>Projective space</i>	0	0	1	\emptyset
<i>Klein bottle</i>	0	0	2	S^0
<i>Disk</i>	0	1	0	\emptyset
<i>Moebius band</i>	0	1	1	\emptyset
<i>Annulus</i>	0	2	0	\emptyset
<i>Torus</i>	1	0	0	S^1

The spaces \mathcal{PS} may be interesting in the study of three-manifolds. For instance, every handlebody decomposition h of genus g of a three-manifold determines an element of \mathcal{S}' (the system of curves along which disks are attached), and hence, a point on \mathcal{PS} .

3. Next we define a type of foliation with singularities on M , which we call a “measured foliation”. In some neighborhood U of a regular point of a measured foliation \mathcal{F} there must be a “flow box” or a chart $\phi: U \rightarrow \mathbf{R}^2$, which sends the leaves of \mathcal{F} to horizontal lines in \mathbf{R}^2 . If two such neighborhoods U_i and U_j overlap, there should be a transition function ϕ_{ij} defined on $\phi_j(U_j)$ such that $\phi_{ij} \circ \phi_j = \phi_i$, on $U_i \cap U_j$, with the property that ϕ_{ij} is of the form

$$\phi_{ij}(x, y) = (f(x, y), c \pm y),$$

where c is a constant. In other words, the transition functions preserve the distances between y -coordinates of points. A finite number of singular points are permitted, where the singularities are “ p -pronged saddles”, with $p \geq 3$. Cf. Figure 2.

If M has boundary, we require that each boundary component of M have at least one singularity of the measured foliation. Near nonsingular points of ∂M , the boundary is a leaf. Cf. Figure 3.

A neighborhood of a p -pronged saddle may be obtained from p rectangles, glued together along edges. It is most natural to think of the singularity of the foliation as corresponding to a singularity in the differentiable structure of

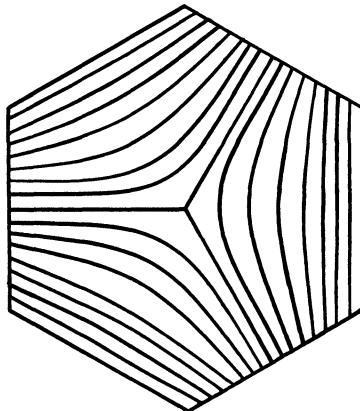


FIGURE 2. A neighborhood of a 3-pronged singularity of a measured foliation. This may be obtained by gluing together the left and right halves of the bottom edges of three rectangles, in an appropriate pattern.

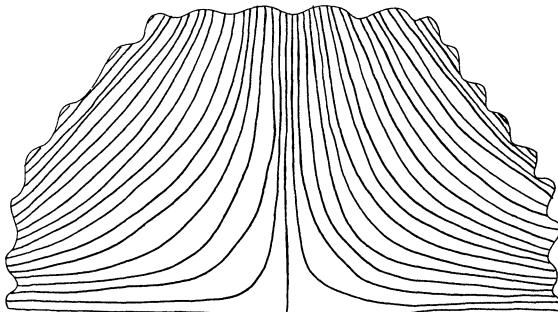


FIGURE 3. A neighborhood of a typical singularity of a measured foliation on ∂M .

the surface. This singularity has the type of a cone of cone angle $p\pi$, that is, a deleted neighborhood of the cone point has universal cover identified with the universal cover of a deleted neighborhood of the origin in \mathbf{R}^2 , and the whole angle around the cone point is $p\pi$. It makes sense to speak of a diffeomorphism, in this singular differentiable structure, as being a homeomorphism which is a diffeomorphism away from the singularities, and in a neighborhood of each singular point it is related to a diffeomorphism of \mathbf{R}^2 via the identification of the universal cover of the deleted neighborhood of the singularity with the universal cover of a deleted neighborhood of the origin in \mathbf{R}^2 . A singularity on the boundary of a surface, similarly, corresponds to a singularity in the differentiable structure of the surface which is like half a cone.

If \mathcal{F} is a measured foliation and a is a simple closed curve, we define $\int_a \mathcal{F}$ to be the total variation of the y -coordinates of a , as measured locally with respect to any chart. If $\beta \in \mathcal{S}$, we define $i(\mathcal{F}, \beta)$ to be the infimum of $\int_b \mathcal{F}$, where b is a simple closed curve representing β .

Two measured foliations \mathcal{F}_1 and \mathcal{F}_2 are *measure equivalent* if for all $\beta \in \mathcal{S}$, $i(\mathcal{F}_1, \beta) = i(\mathcal{F}_2, \beta)$.

Two measured foliations \mathcal{F}_1 and \mathcal{F}_2 are *projectively equivalent* if $p \circ i_* \mathcal{F}_1 = p \circ i_* \mathcal{F}_2$, i.e., if there is a $\lambda \in \mathbf{R}$ such that for all $\alpha \in \mathcal{S}$,

$$i(\mathcal{F}_1, \alpha) = \lambda i(\mathcal{F}_2, \alpha).$$

PROPOSITION. *For every measured foliation \mathcal{F} there is a $\beta \in \mathcal{S}$ such that $i(\mathcal{F}, \beta) \neq 0$.*

As before, i defines a map i_* from the set of measured foliations to \mathbf{R} . We define $\mathcal{PF} = \{p \circ i_*(\mathcal{F}) | \mathcal{F} \text{ is a measured foliation}\}$.

THEOREM 2. $\mathcal{PF} = \mathcal{PS}$. In other words, the topological sphere $\mathcal{PS}(M)$ can be identified with the set consisting of all projective equivalence classes of measured foliations on M .

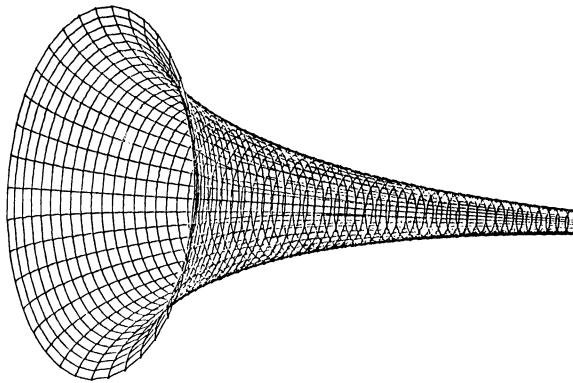


FIGURE 4. A portion of a pseudosphere, which extends to infinity at the thin end. A pseudosphere is a surface of revolution of Gaussian curvature -1 . The generating curve is a tractrix, which is characterized by the property that its tangent lines meet the x -axis a unit distance from the point of tangency.

4. To avoid special cases, we now restrict ourselves to the case $\chi(M) < 0$. Such a surface always has Riemannian metrics with curvature -1 . We will say that two Riemannian metrics g_1 and g_2 are isotopic if g_1 comes from g_2 via a diffeomorphism isotopic to the identity. If M is closed, the Teichmüller space $\mathcal{T}(M)$ can be defined as the set of isotopy classes of all Riemannian metrics of curvature -1 . If M has boundary, $\mathcal{T}(M)$ is the set of isotopy classes of Riemannian metrics of curvature -1 on the interior of M which are modelled on the thin end of a pseudosphere (cf. Figure 4) near a boundary component. Such a metric will be called a *hyperbolic* metric. An alternative characterization is that a hyperbolic metric is a complete metric with $K = -1$ and finite total volume. If $\gamma \in \mathcal{T}(M)$ and $\alpha \in \mathcal{S}$, define $i(\gamma, \alpha)$ to be

the infimum of the length of a simple closed curve, a , representing α . (It is classical that there is a unique closed geodesic representing α ; this geodesic has length $i(\gamma, \alpha)$.) Two hyperbolic metrics g_1 and g_2 are isotopic if for every $\alpha \in \mathcal{S}$, $i(g_1, \alpha) = i(g_2, \alpha)$. We endow $\mathcal{T}(M)$ with the minimum topology such that each function $i(\gamma, \alpha)$ is continuous. We now have a continuous map $p \circ i_* : \mathcal{T}(M) \rightarrow \mathbf{P}(\mathbf{R}^{\mathcal{S}})$.

THEOREM 3. *The closure of $p \circ i_*(\mathcal{T}(M))$ is $p \circ i_*(\mathcal{T}(M)) \cup \mathcal{PS}$. This gives a natural topology to $\mathcal{T}(M) \cup \mathcal{PS}$ in which it is homeomorphic to a disk $D^{6g+2b+3c-6}$.*

It has been known since Fricke that $\mathcal{T}(M)$ is homeomorphic to the interior of a disk. There are several interesting compactifications of $\mathcal{T}(M)$, but the action of the group of diffeomorphisms of M is not yet known to extend to the other boundaries.

We use the terminology $\overline{\mathcal{T}}(M)$ for the closed Teichmüller space $\mathcal{T}(M) \cup \mathcal{PS}(M)$.

COROLLARY. *If ϕ is any diffeomorphism of M , either*

- (a) ϕ fixes an element of the Teichmüller space, or
- (b) ϕ fixes a projective class of measured foliations.

In case (a), it follows (by definition) that there is a Riemannian metric g for which ϕ is isotopic to an isometry ϕ' . Such a ϕ' has finite order.

PROOF. Brouwer fixed point theorem on $\overline{\mathcal{T}}$.

For any subset $F \subset \mathcal{S}$, let $K_F \subset \mathbf{P}(\mathbf{R}^{\mathcal{S}})$ be the set consisting of points of the form $p(f)$ where $f \in \mathbf{R}^{\mathcal{S}}$ vanishes identically on F . If F is big enough, then $\mathcal{PS} \cap K_F = \emptyset$. There is a natural projection $\pi_F : (\mathbf{P}(\mathbf{R}^{\mathcal{S}}) - K_F) \rightarrow \mathbf{P}(\mathbf{R}^F)$ inherited from projection of the vector spaces. For big enough (but still finite) F , π_F/\mathcal{PS} is an embedding in $\mathbf{P}(\mathbf{R}^F)$. In other words, the preceding constructions would go through using an appropriately chosen finite subset of \mathcal{S} . This has the disadvantage that it is harder to see the map $\mathcal{PS}(\phi)$ induced by a diffeomorphism ϕ .

\mathcal{T} has a natural smooth structure, since the image of \mathcal{T} in \mathbf{R}^F , for big enough F , is a smooth submanifold. Unfortunately, the space \mathcal{PS} does not seem to have a natural smooth structure except in the cases when its dimension is ≤ 1 . In the 1-dimensional case ($M = M_{1,0,0}$ or $M_{1,1,0}$ or $M_{0,4,0}$) the map of $\mathbf{P}(\mathbf{R}^F)$ to itself obtained by squaring the coordinates in \mathbf{R}^F sends $\pi_F(\overline{\mathcal{T}})$ to a smooth submanifold, for sufficiently large F . This smooth structure is classical.

In general, however, \mathcal{PS} has a kind of structure, a piecewise integral projective structure, which is nicer for many purposes than a smooth structure. To define piecewise integral projective (or PIP for short), first define an integral projective k -simplex in \mathbf{P}^n to be a set of the form $p(\{\sum_{i=1}^k t_i a_i : t_i \geq 0\})$, where $\{a_i\}$ generates a maximal abelian rank k subgroup of $\mathbf{Z}^{n+1} \subset \mathbf{R}^{n+1}$. A PIP triangulation of a set K in \mathbf{P}^n is a cell division such that each cell is an integral projective simplex. A PIP homeomorphism between two subsets K and L is a homeomorphism h such that K admits a PIP triangulation for which h restricted to any simplex is a transformation coming from $\mathrm{GL}(n, \mathbf{Z})$. Abstract

definitions, not depending on embeddings, can also be obtained easily. For big enough F , $\pi_F(\mathcal{P})$ is a PIP submanifold of $\mathbf{P}(\mathbf{R}^F)$, and the transformations $\mathcal{P}(\phi)$ are PIP homeomorphisms. It is easy to see that the word problem is solvable in the group of PIP homeomorphisms of a compact manifold. Even better, the fixed point set of a PIP homeomorphism is effectively computable.

5. Case (b) of the corollary requires further analysis according to the nature of the projective class of foliations fixed by $\mathcal{PF}(\phi)$.

We will say that ϕ is *reducible* by Γ if Γ is a system of disjoint simple closed curves $\Gamma_1, \dots, \Gamma_n$ ($n \geq 1$), such that the Γ_i represent distinct elements of \mathcal{S} , and Γ is invariant by ϕ (but the Γ_i may be permuted). If ϕ is reducible, we may think of ϕ as coming from the simpler surface (possibly nonconnected) obtained by cutting along Γ . There is an upper bound, depending only on M , to the number of such cuts which can be made.

We will use the notation $\mathcal{F}_1 = \lambda \mathcal{F}_2$ if \mathcal{F}_1 and \mathcal{F}_2 agree as foliations, but for any arc α ,

$$\int_{\alpha} \mathcal{F}_1 = \lambda \int_{\alpha} \mathcal{F}_2.$$

Two measured foliations are *transverse* if they are transverse in the usual sense near regular points not on ∂M^2 . (Of course the two foliations must be tangent at nonsingular boundary points.) The typical picture near a singularity is given in Figure 5.

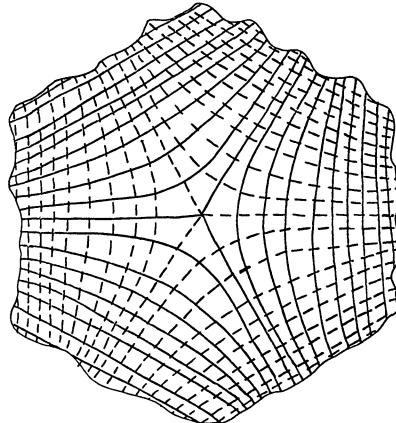


FIGURE 5. A pair of transverse measured foliations in a neighborhood of a common 3-prong singularity.

THEOREM 4. *For any diffeomorphism ϕ of a surface M^2 , ϕ is isotopic to a diffeomorphism ϕ' such that either:*

- (i) ϕ' fixes an element of \mathcal{F} , and has finite order; or
- (ii) There is a number $\lambda > 1$ and a pair of transverse measured foliations \mathcal{F}^s and \mathcal{F}^u such that $\phi'(\mathcal{F}^s) = 1/\lambda \mathcal{F}^s$ and $\phi'(\mathcal{F}^u) = \lambda \mathcal{F}^u$; or
- (iii) ϕ' is reducible by a system of curves Γ . Γ has a ϕ' -invariant tubular neighborhood $\mathcal{N}(\Gamma)$, such that on each (not necessarily connected) ϕ' -component of $M^2 - \mathcal{N}(\Gamma)$, ϕ' satisfies (i) or (ii).

The representative ϕ' of the isotopy class of ϕ is determined up to conjugacy by the condition that it satisfies either (i) or (ii), or (iii) together with a simple restriction on its behavior in $\mathcal{N}(\Gamma)$.

In case (ii), the projective class of \mathcal{F}^s and \mathcal{F}^u is uniquely determined.

In case (ii), we mean that ϕ' is a diffeomorphism with respect to the singular differentiable structure determined by the foliations \mathcal{F}^s and \mathcal{F}^u . It is a more delicate question whether ϕ' is topologically conjugate to a diffeomorphism.

Case (ii) is the typical case; we will call such a ϕ' a *pseudo-Anosov* diffeomorphism. The dynamics of $\overline{\mathcal{T}}(\phi)$, the induced map on the disk $\overline{\mathcal{T}}(M)$, may be described very simply when ϕ is isotopic to a pseudo-Anosov diffeomorphism. $\overline{\mathcal{T}}(\phi)$ has exactly two fixed points, represented by \mathcal{F}^s and \mathcal{F}^u . Every other point in $\overline{\mathcal{T}}(\phi)$ tends (rather quickly) toward \mathcal{F}^s under iterates of $\overline{\mathcal{T}}(\phi)$. This dynamical situation is equivalent to case (ii).

Here is a consequence of Theorem 5, using the theory of foliations of surfaces:

THEOREM 5. *For any diffeomorphism ϕ of M , there is a finite set $1 \leq \lambda_1 < \lambda_2 < \dots < \lambda_k$ of algebraic integers such that for any $\alpha \in \mathcal{S}$ there is a λ_i such that for any Riemannian metric g ,*

$$\lim_{n \rightarrow \infty} |i(g, \phi^n \alpha)|^{1/n} = \lambda_i.$$

ϕ is isotopic to a pseudo-Anosov diffeomorphism iff $k = 1$ and $\lambda_1 > 1$. λ_1 may be identified with the λ of Theorem 4 in case (ii).

The λ_i 's may be thought of as eigenvectors in various ways. The simplest is the following. We suppose, for simplicity, that ϕ is of type (ii). There is a finite set of singularities, $\{s_1, \dots, s_j\}$, of \mathcal{F}^u where $j \leq -2\chi(M)$. Let \tilde{M} be the unique 2-fold branched cover of M , branched along the singularities of odd multiplicity, for which the induced foliation $\tilde{\mathcal{F}}^u$ is transversely oriented. Now, \mathcal{F}^u defines a closed 1-form ω on \tilde{M} , which is nontrivial in cohomology. Let $\tilde{\phi}$ be a diffeomorphism of \tilde{M} covering ϕ . Then, $\phi\omega = \pm\lambda\omega$, so $\pm\lambda$ is an eigenvalue of the map induced by $\tilde{\phi}$ on homology.

One can obtain a bound for the degree of λ as follows. An elementary index formula shows that there are at most $2\chi(M)$ singularities of \mathcal{F}^u , hence at most $4 - 4h - 2c + b$ singularities in the interior of M . Then

$$\begin{aligned} \chi(\tilde{M})^2 &\geq 2\chi(M) - \#(\text{interior singularities}) \\ &\geq 8 - 8h - 4c - 3b. \end{aligned}$$

The form ω lies in $V = H^1(\tilde{M}^2; \partial\tilde{M}^2)$, whose dimension cannot exceed $6 - 8h - 4c - 2b$. Let T be the deck transformation for \tilde{M} over M , so $T^2 = \text{id}$. Let V_+ and V_- be the eigenspaces for T_* corresponding to eigenvalues $+1$ and -1 . Then V_+ is isomorphic to $H^1(M; \partial M)$ so its dimension is $2h + c$, and the dimension of V_- does not exceed $6 - 6h - 3c - 2b$. ω lies in V_- , by definition, and V_- is an integral subspace invariant by ϕ_* (since ϕ commutes with T), so λ is an eigenvector of a $d \times d$ integral matrix, where d is the dimension of $\mathcal{T}(M)$. (This is no accident; a neighborhood of ω in $\mathbf{P}(V_-)$ gives local coordinates

for $\mathcal{P}\mathcal{F}$.) Therefore λ is an algebraic integer of degree $\leq d$. The examples of Theorem 7 show this bound is sharp. If M is oriented, $\phi_- = \phi^*|V_-$ is a symplectic transformation. It follows that $\lambda + 1/\lambda$ is an algebraic integer of degree $\leq d/2$.

Another way to study λ is via the theory of Markov partitions for ϕ . With this theory one obtains λ as an eigenvector of a nonnegative integral matrix of bounded size. Hence the sequence of possible λ 's for a given surface is discrete. By studying Markov partitions it can also be shown that the number of possible conjugacy classes of pseudo-Anosov diffeomorphisms for any fixed λ is finite.

If ϕ is of type (i), then obviously $k = 1$ and $\lambda_1 = 1$. If ϕ is of type (iii), then $\lambda_1 = 1$, and the other λ_i 's depend on the ϕ -components of $M - (\Gamma)$.

REMARK. Most of what has been said since the beginning of §4 remains true when $\chi(M) \geq 0$, with the appropriate definitions. When $\chi(M) = 0$, $\mathcal{T}(M)$ can be identified with the set of isotopy classes of flat metrics of total area 1, and geodesic boundary. Theorem 3 remains true for the torus and Klein bottle, but is false for the annulus and Moebius band. When $\chi(M) > 0$, Theorem 3 is true but vacuous. The corollary to Theorem 3, as well as Theorem 4 and Theorem 5, are true in all cases.

Now let M be the universal covering of M , and let $\tilde{\phi}$ be any homeomorphism of M covering ϕ . If M is given a hyperbolic metric, then M is isometric to the interior of the unit disk, with the Poincaré metric. This gives M a compactification making it a disk, and $\tilde{\phi}$ extends to a homeomorphism of the disk. Nielsen analyzed this situation extensively, and described the possibilities for the dynamic behavior of the extension of $\tilde{\phi}$ to the boundary of the disk. In particular, this induced map $\partial\phi$ always has periodic points, with period $p \leq 2 - \chi(M_{g,b,c}) - b$.

This analysis can be derived from Theorem 4. The key fact needed is that if \mathcal{F} is a foliation on M induced from a measured foliation on M , each end of each leaf of \mathcal{F} converges to a point on the boundary of the disk. If ϕ is pseudo-Anosov, then a lift $\tilde{\phi}$ can have at most one fixed point. If there are no fixed points, $\partial\tilde{\phi}$ has one contracting and one expanding fixed point and positive iterates of $\tilde{\phi}$ carry any point in M arbitrarily close to the contracting fixed point. If $\tilde{\phi}$ has a fixed point x then the leaves of x in \mathcal{F}^u and \mathcal{F}^s must have a singularity at x , or no singularity at all. Let $p = 2$ if x is nonsingular, or the number of prongs of the singularity of \mathcal{F}^u if p is singular. $\partial\phi$ has p contracting and p expanding periodic points, which use the limits of the leaf of x in \mathcal{F}^u and \mathcal{F}^s , respectively. These points are permuted the same way that ϕ permutes the prongs of the singularity at x . An index argument (which Nielsen does rigorously) shows that if ψ is isotopic to ϕ and if $\tilde{\psi}$ is the lift of ψ corresponding to $\tilde{\phi}$, then $\tilde{\psi}$ must also have a fixed point. From this can be derived:

THEOREM 6. *If ϕ is a pseudo-Anosov diffeomorphism, ϕ has the minimum number of periodic points (for every period) in its isotopy class.*

On the other hand, Nielsen's analysis of the fixed point structure of $\partial\tilde{\phi}$ can be used to give another derivation of Theorem 4.

6. We will describe in detail an elementary construction for a large class of examples of diffeomorphisms in canonical form. For any two-sided simple closed curve α , a Dehn twist T_α about α is a diffeomorphism, supported in an annular neighborhood of α , which carries a short arc crossing α to a long arc winding once around α towards the right. This depends on an orientation of M , but not of α . We extend this notation to \mathcal{S}' : if $\alpha \in \mathcal{S}'$, then $T_\alpha \in \pi_0(\text{Diff } M)$ is the composition of T_{α_i} where α_i ranges over the components of α . The order of composition is irrelevant.

We will study the subgroup of $\pi_0(\text{Diff } M)$ generated by two of these generalized Dehn twists. If α and β are systems of curves representing members of \mathcal{S}' , we first isotope them so that $|\alpha \cap \beta| = i(\alpha, \beta)$, in which case α and β are *tight*. If α and β are tight, then α and β *fill up* M if each component of $M - (\alpha \cup \beta)$ is a disk or a half-open annulus whose boundary lies on ∂M .

Let the components of α and β be labelled α_j and β_k , $[j = 1, \dots, n; k = 1, \dots, m]$. There is an $n \times m$ geometric intersection matrix with entries $N_{j,k} = i(\alpha_j, \beta_k)$. It is easy to see that $\alpha \cup \beta$ is connected iff some power of NN^t is strictly positive. In this case NN^t has an eigenvector

$$NN^tV = \mu V$$

with all entries of V strictly positive. V is unique up to a positive multiple, and $\mu > 0$.

If we interchange the roles of α and β , and hence interchange N and N^t , note that the eigenvalue μ remains unchanged, since the m -vector $V^1 = \mu^{-1/2}N^tV$ is strictly positive with

$$N^tNV^1 = \mu V^1.$$

(The factor of $\mu^{-1/2}$ is chosen for the sake of symmetry so that $V = \mu^{-1/2}NV^1$.)

THEOREM 7. *Suppose α and β are tight, and $\alpha \cup \beta$ is connected. Let $G(\alpha, \beta)$ be the subgroup of $\pi_0(\text{Diff } M)$ generated by T_α and T_β . There is a representation ρ of $G(\alpha, \beta)$ in $\text{PSL}(2, \mathbf{R})$ given by*

$$T_\alpha = \begin{bmatrix} 1 & \mu^{1/2} \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad T_\beta = \begin{bmatrix} 1 & 0 \\ -\mu^{1/2} & 1 \end{bmatrix}.$$

The image $\rho(G(\alpha, \beta))$ is a discrete group. For some k -fold covering group $\widetilde{\text{PSL}(2, \mathbf{R})^k}$, ρ lifts to a representation ρ^k which is faithful. If α and β fill up M , then k is finite, and $g \in G$ is of type (i), (ii) or (iii) according to whether $\rho(g)$ is elliptic, hyperbolic or parabolic.

Full information without the hypothesis that $\alpha \cup \beta$ is connected can be derived from this statement, by looking at subsurfaces where the hypothesis is satisfied.

We describe the construction in detail when α and β fill up M ; otherwise, one can restrict to a subsurface filled up by α and β . If M has boundary, attach a disk to each boundary component. $\alpha \cup \beta$ divides M into cells; let C be the dual cell division. Each 2-cell of C is a square.

Each 2-cell S of C contains a unique intersection point of a component α_j of α and a component β_k of β . Construct a flat Riemannian metric with singularities on M by making S a rectangle whose sides transverse to α_j have length V_j , and whose sides transverse to β_k have length V_k^1 . The rectangles intersecting α_j string together to form a cylindrical neighborhood of α_j of thickness V_j and height $(NV^1)_j = \mu^{-1/2}(NN^1V)_j = \mu^{1/2}V_j$. Similarly, β_k has a cylindrical neighborhood of thickness V_k^1 and circumference $\mu^{1/2}V_k^1$. There is a positively oriented orthogonal frame field e , well-defined up to ± 1 , whose first vector is parallel to the α -curves, and second vector parallel to the β -curves. There is a unique affine transformation of M , with derivative (measured by e)

$$\begin{bmatrix} 1 & \mu^{1/2} \\ 0 & 1 \end{bmatrix}$$

which fixes the tops and bottoms of the α cylinders. One can see this by using coordinates $\{(x, y) | 0 \leq y \leq V(\alpha, \beta)_j\}$ for the universal cover of the α_j cylinders. The linear map $\begin{bmatrix} 1 & \mu^{1/2} \\ 0 & 1 \end{bmatrix}$ fixes the x -axis and sends $\begin{bmatrix} 0 \\ V_j \end{bmatrix}$ to $\begin{bmatrix} \mu^{1/2}V_j \\ V_j \end{bmatrix}$ which is identified to $\begin{bmatrix} 0 \\ V_j \end{bmatrix}$ in the cylinder. This affine map $\text{Af}(T_\alpha)$ is representative of T_α . Similarly, there is an affine map $\text{Af}(T_\beta)$ with derivative

$$\begin{bmatrix} 1 & 0 \\ \mu^{-1/2} & 1 \end{bmatrix},$$

fixing the tops and bottoms of the β cylinders, which represents T_β .

If ϕ is any element of the group of homeomorphisms generated by $\text{Af}(T_\alpha)$ and $\text{Af}(T_\beta)$ then ϕ is an affine map with constant derivative. If $d\phi$ is elliptic, there is a quadratic form invariant by $d\phi$, or in other words, a flat Riemannian metric such that ϕ is an isometry, and ϕ is type (i). If $d\phi$ is hyperbolic, the two eigenspaces of $d\phi$ define two transverse measured foliations of M and ϕ is type (ii). If $d\phi$ is parabolic, then ϕ has a one-dimensional eigenspace with eigenvalue 1. This eigenspace defines a foliation \mathcal{F} invariant by ϕ . All the vertices of C are fixed by ϕ , so each leaf of \mathcal{F} through a vertex must be fixed by ϕ . If ϕ is not the identity, these leaves cannot have a point of accumulation, hence they are closed. \mathcal{F} is nonsingular in the complement of these leaves, so each component of the complement must be an open cylinder or Moebius band. The latter is ruled out because it has no nontrivial affine automorphisms fixed on the boundary, and we see that ϕ is of the form T_γ for some $\gamma \in \mathcal{S}'$ (where γ has one or more components corresponding to each cylinder). In particular, ϕ is of type (iii) when $d\phi$ is parabolic. In any case, ϕ is isotopic to the identity iff ϕ is the identity, so Af extends to a homomorphism of $G(\alpha, \beta)$ to $\text{Homeo}(M)$. Therefore $\rho = (d \circ \text{Af})/\pm I$ is the desired representation.

The group $\text{PSL}(2, \mathbf{R})$ has an action on $\mathbf{R}^2/(x \sim -x)$ coming from the linear action of $\text{SL}(2, \mathbf{R})$ on \mathbf{R}^2 . The covering group $\text{PSL}(2, \mathbf{R})^k$ is the group of transformations of the k -fold covering of $\mathbf{R}^2/(x \sim -x)$ branched at 0, which cover transformations in $\text{PSL}(2, \mathbf{R})$. Each vertex of C is fixed by the group generated by $\text{Af}(G(\alpha, \beta))$, and a neighborhood of such a vertex looks like a k -fold branched covering of $\mathbf{R}^2/x \sim -x$. The action near such a vertex

defines a lifting $\tilde{\rho}^k$ of ρ . For each vertex, this representation $\tilde{\rho}^k$ is faithful, since whenever $\tilde{\rho}^k(g) = 1$, then $Af(G)$ is the identity near the vertex, hence everywhere.

It is not hard to write down a presentation for $G(\alpha, \beta)$ in terms of the cell-division C , and μ , but we omit this. However, we remark that $G(\alpha, \beta)$ is a free group iff $\mu \geq 4$, which is equivalent to $\text{Tr } \rho(T_\alpha T_\beta) \leq -2$.

COROLLARY. *There are pseudo-Anosov diffeomorphisms which act trivially on the homology of M .*

This answers negatively a conjecture of Nielsen [1].

PROOF OF COROLLARY. On M_g , let α and β be two null-homologous curves which fill up M . An easy way to generate such a pair is to begin with a standard null-homologous curve, α , and a complicated curve β_0 such that α and β_0 fill up M . Then α and $\beta = T_{\beta_0}(\alpha)$ satisfy the conditions. The group $G(\alpha, \beta)$ acts trivially on the homology of M . Since $\text{Tr}(\rho(T_\alpha T_\beta^{-1})) = 2 + \mu$, it follows that $T_\alpha \circ T_\beta^{-1}$ is isotopic to a pseudo-Anosov diffeomorphism.

A flat Riemannian metric on a surface with cone-like singularities having cone angle $k\pi$, k an integer greater than 2, determines a complex geodesic in the Teichmüller metric for Teichmüller space. This geodesic consists of Riemannian metrics conformally equivalent to a flat metric affinely equivalent to the given one. An equivalent statement to Theorem 7 is that $\mathcal{T}(G(\alpha, \beta))$ is a group of translations of some complex geodesic ϕ , where α and β fill up M . The Teichmüller metric induces a metric isometric to the hyperbolic plane on each complex geodesic; that is why $\text{PSL}(2, \mathbf{R})$, the group of isometries of the hyperbolic plane, is natural in Theorem 7.

Case (ii) of Theorem 4 is equivalent to the statement that $\mathcal{T}(\phi)$ is a translation of a real geodesic in Teichmüller space. Lipman Bers has found a nice analytic proof of the existence statements of Theorem 4, making use of this equivalence.

7. The nicest aspects of this theory I have been trying to sketch are not formal, but intuitive. If you draw pictures of a pseudo-Anosov diffeomorphism, you can understand geometrically what it *does*, something which has puzzled me for several years. The space \mathcal{PS} is actually quite concrete, and for a diffeomorphism ϕ , $\mathcal{PS}(\phi)$ can be calculated fairly readily. This induced transformation is piecewise projective, locally described by integral matrices, so it can be computed by hand (although a computer is handier); it is pleasant to see something of this abstract origin made very concrete. (In contrast, $\mathcal{T}(\phi)$ is given by hideously complicated transcendental expressions, in the usual coordinate systems.) And there is a great deal of natural geometric structure on \mathcal{PS} , relating to structure on M , beautiful to contemplate.

REFERENCES

1. J. Nielsen, *Surface transformation classes of algebraically finite type*, Danske Vid. Selsk. Math.-Phys. Medd. **68** (1944).

