Acta Math., 170 (1993), 121–149

Representation theoretic rigidity in PSL(2, **R**)

by

CHRISTOPHER BISHOP and TIM STEGER[1]

University of California
Los Angeles, CA, U.S.A.

University of Chicago
Chicago, IL, U.S.A.

1. Introduction

Let G be a connected simple Lie group with trivial center, let Γ be an abstract group, and let ι_1 and ι_2 be inclusions of Γ in G. Assume throughout that each of the images $\iota_j(\Gamma)$ is a *lattice* subgroup, meaning that $\iota_j(\Gamma)$ is discrete and that the G-invariant measure on $G/\iota_j(\Gamma)$ has total finite mass. We say that ι_1 and ι_2 are *equivalent* if there is some automorphism ϱ of G so that $\iota_2=\varrho\circ\iota_1$. If G is not isomorphic to $\mathrm{PSL}(2,\mathbf{R})$ then the Mostow rigidity theorem (see [18], [19], [16] and [24]) says that ι_1 and ι_2 are necessarily equivalent. Alternatively, this says that any isomorphism between lattice subgroups of G extends to an automorphism of the whole group. This remarkable result fails for $\mathrm{PSL}(2,\mathbf{R})$ (see Section 2). Nonetheless, taking $G=\mathrm{PSL}(2,\mathbf{R})$, we have

THEOREM 1. *Suppose that π_1 and π_2 are irreducible unitary representations of* $\mathrm{PSL}(2,\mathbf{R})$, *not in the discrete series. Then $\pi_1\circ\iota_1$ and $\pi_2\circ\iota_2$ are equivalent representations of Γ if and only if ι_1 and ι_2 are equivalent inclusions and π_1 and π_2 are equivalent representations of* $\mathrm{PSL}(2,\mathbf{R})$.

As usual, two unitary representations of a group are called equivalent if there is a unitary equivalence of the two representation spaces which intertwines the two group actions. The situation is entirely different for discrete series representations, as explained in Section 8. Theorem 1 for $\iota_1\sim\iota_2$ was proven in [6].

The central step in the proof of Theorem 1 is a certain analytic criterion for the equivalence of ι_1 and ι_2. Let $\mathrm{PSL}(2,\mathbf{R})$ act on the upper half plane $\mathbf{H}=\{\mathrm{Im}(z)>0\}$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z)=\frac{az+b}{cz+d},$$

[1] Both authors are partially supported by the NSF

let $d(\cdot,\cdot)$ denote hyperbolic distance on $\mathbf{H}$ and let

$$h(g) = \exp(-d(g(i),i)) \quad \text{for } g \in \mathrm{PSL}(2,\mathbf{R}).$$

One can show that h is in $L^{1+\varepsilon}(\mathrm{PSL}(2,\mathbf{R}))$ (with respect to Haar measure) for any $\varepsilon>0$ but is not in $L^1(\mathrm{PSL}(2,\mathbf{R}))$. Similarly $h\circ\iota_j$ is in $l^{1+\varepsilon}(\Gamma)$ but not in $l^1(\Gamma)$.

THEOREM 2. *Fix s, $0<s<1$. The lattice inclusions ι_1 and ι_2 are equivalent if and only if*

$$\sum_{\gamma\in\Gamma} h^s(\iota_1(\gamma))h^{1-s}(\iota_2(\gamma)) = \infty.$$

Indeed, the proof will also show

THEOREM 3. *If ι_1 and ι_2 are not equivalent, then there is some $\delta=\delta(s)>0$ so that*

$$\sum_{\gamma\in\Gamma} (h^s(\iota_1(\gamma))h^{1-s}(\iota_2(\gamma)))^{1-\delta} < \infty.$$

We call ι_1 and ι_2 geometrically conjugate if there is some homeomorphism β of $\mathbf{H}$ such that $\iota_2(\gamma)=\beta\circ\iota_1(\gamma)\circ\beta^{-1}$ for every $\gamma\in\Gamma$. Such a β will extend uniquely to a homeomorphism of $\mathbf{R}\cup\infty$, the boundary of $\mathbf{H}$. Moreover, the boundary homeomorphism is completely determined by ι_1 and ι_2, even though the interior homeomorphism is not. If dx is ordinary Lebesgue measure on $\mathbf{R}$ and if ι_1 and ι_2 are not equivalent, then dx and $\beta_*(dx)$ are mutually singular (see [18] or [1]). Also see [13] and [30].

THEOREM 4. *Suppose that ι_1 and ι_2 are geometrically conjugate and $\delta>0$ is as in Theorem 3. Then there is a set $E\subset\mathbf{R}$ such that $\dim(E)\leqslant 1-\delta$ and $\dim(\beta(E^c))\leqslant 1-\delta$.*

This answers a question from [31]. The equivalence classes of embeddings belonging to a fixed geometrical conjugacy class make up the Teichmüller space of a surface $\mathbf{H}/\Gamma$, so our results can be interpreted from the point of view of Teichmüller theory.

The paper is organized as follows. In Section 2 we review some well known facts on Fuchsian groups which we will need later. In Section 3 we prove Theorem 2 when the groups in question are cocompact and in Section 4 we consider the non-cocompact case. In Section 5 we prove Theorems 3 and 4. In Section 6 we review some basic facts about the unitary, irreducible representations of $\mathrm{PSL}(2,\mathbf{R})$ and in Section 7 we deduce Theorem 1 from Theorems 2 and 3. This section also contains a brief sketch of the argument in [6] which proves Theorem 1 in the case $\iota_1\sim\iota_2$. Finally in Section 8 we discuss why Theorem 1 must fail for discrete series representations and how to modify the theorem if $\mathrm{PSL}(2,\mathbf{R})$ is replaced by $\mathrm{SL}(2,\mathbf{R})$.

The reader who is primarily interested in Theorem 1 can read Sections 6, 7 and 8 first. Except for the statement of Theorem 3 they do not depend on the earlier sections at all. The first author thanks Mladen Bestvina for several helpful conversations concerning Dehn's work and the arguments in Sections 3 and 4.

2. Properties of lattices in PSL(2, R)

PSL$(2, \mathbf{R})$ is the quotient of SL$(2, \mathbf{R})$, the group of 2×2 real matrices with determinant one, by its two element center, $\{\pm I\}$. This group G acts on $\mathbf{H}=\{\mathrm{Im}(z)>0\}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z)=\frac{az+b}{cz+d}.$$

The action preserves the *hyperbolic* metric $ds^2=(dx^2+dy^2)/y^2$, and in fact G is the full group of orientation preserving isometries of $\mathbf{H}$. By a lattice subgroup of G we mean a discrete subgroup of G which acts by left translation on G so that a fundamental domain has finite Haar measure. The prime example of a lattice subgroup is PSL$(2, \mathbf{Z})$, the subgroup of matrices in PSL$(2, \mathbf{R})$ with integer entries. A lattice subgroup is necessarily finitely generated and of divergence type, i.e., $\sum_\Gamma h(\gamma)=\infty$. A nonidentity Möbius transformation of $\mathbf{H}$ is one of three types: elliptic if it has a fixed point in $\mathbf{H}$, parabolic if it has one fixed point on $\mathbf{R}\cup\{\infty\}$ and hyperbolic if it has 2 fixed points on $\mathbf{R}\cup\{\infty\}$. If a group G of Möbius transformations of $\mathbf{H}$ acts discontinuously (any compact set hits itself only finitely often) the group is called Fuchsian. If G has no elliptic elements then $R=\mathbf{H}/G$ is a Riemann surface and G is isomorphic to the fundamental group of R. If G is also a lattice then R is either a compact Riemann surface (in which case G has no parabolic elements and has a compact fundamental domain) or a compact surface with a finite number of points removed (in which case G is a free group, has parabolic elements and all fundamental domains are noncompact). For example, if G is the free group on two generators then $\mathbf{H}/G$ could either be a torus with one puncture or a sphere with three punctures. The corresponding lattice embeddings of the free group cannot be equivalent (the Möbius function conjugating them would also define a conformal mapping between the quotient surfaces). Thus Mostow rigidity fails for PSL$(2, \mathbf{R})$. One can also produce examples by varying the conformal structure on a single surface (different points of moduli space) or taking inequivalent sets of generators for the same Riemann surface (i.e., same points in moduli space but different points in Teichmüller space).

We can also consider PSL$(2, \mathbf{R})$ as acting on the unit disk $\mathbf{D}=\{|z|<1\}$ using the identification of $\mathbf{H}$ and $\mathbf{D}$ via the Möbius transformation

$$\tau(z)=\frac{z-i}{z+i}.$$

On the unit disk the function h defined in the introduction has a simple interpretation

$$h(g)=\exp(-d(g(0),0))=\frac{1-|g(0)|}{1+|g(0)|}\sim 1-|g(0)|=\mathrm{dist}(g(0), \mathbf{T}).$$

On the upper half plane we have

$$h(g)=\exp(-d(g(i),i))\sim \mathrm{Im}(g(i))$$

if $|g(i)|\leqslant C$.

What are the automorphisms of PSL$(2,\mathbf{R})$? Let G_0 be the full group of isometries of $\mathbf{H}$ with the hyperbolic metric. This includes PSL$(2,\mathbf{R})$ as a normal subgroup of index 2 (the orientation preserving isometries). The inner automorphisms of G_0 all leave PSL$(2,\mathbf{R})$ fixed and they constitute the entire automorphism group of PSL$(2,\mathbf{R})$. One direction of Theorem 2 is now clear. If the two embeddings are equivalent then there is a $g\in G_0$ such that $\iota_1(\gamma)=g\circ\iota_2(\gamma)\circ g^{-1}$ for every $\gamma\in\Gamma$. Thus,

$$h(\iota_1(\gamma))\sim h(\iota_2(\gamma))$$

and so

$$\sum_{\gamma\in\Gamma} h^s(\iota_1(\gamma))h^{1-s}(\iota_2(\gamma))\sim\sum_{\gamma\in\Gamma} h(\iota_1(\gamma)).$$

Since $\iota_1(\Gamma)$ is a lattice in PSL$(2,\mathbf{R})$ it is a Fuchsian group of divergence type, so the right hand side must be ∞. The difficult part of Theorem 2 is the converse.

Also of interest will be the automorphism group of Γ when Γ is the fundamental group of a compact Riemann surface of genus p with q points removed. If $q=0$ then Γ is called a surface group and is generated by $2p$ elements $a_1, b_1, ..., a_p, b_p$ and the single relation

$$a_1b_1a_1^{-1}b_1^{-1}\dots a_pb_pa_p^{-1}b_p^{-1}=1.$$

The Dehn–Nielsen theorem [7] states that the only automorphisms of Γ are those induced by homeomorphisms of the Riemann surface to itself.

If $q>0$ then Γ is a free group on $m=2p+q-1$ generators and there are many more automorphisms than those induced by surface homeomorphisms. Characterizing them is the same as characterizing all possible sets of generators. Suppose Γ is generated by $a_1, ..., a_m$ with no relations. Elements of the group can be thought of as reduced words in the generators, i.e., words in which no cancellation is possible. If we replace a generator a_j with one of the elements a_j^{-1}, a_ia_j or a_ja_i (for $i\neq j$) we obtain a new set of generators. These are called the elementary Nielsen transformations of the generators and any two minimal collection of generators can be transformed into one another by a finite series of such transformations and permutations (this is Nielsen's theorem [20], [15]). Therefore, unlike surface group automorphisms, automorphisms of the free group may have nothing to do with the punctured surface of which Γ is the fundamental group.

Another useful fact is Selberg's lemma [26]. It states that any finitely generated matrix group has a normal subgroup of finite index with no torsion (i.e., no elements of finite order). In particular, this means that any lattice Γ in PSL$(2,\mathbf{R})$ has a normal subgroup Γ' of finite index which has no elliptic elements (an elliptic element in a Fuchsian group must have finite order).

Now suppose Γ, ι_1 and ι_2 are as above and that $\Gamma'\subset\Gamma$ is a normal subgroup of finite index in Γ. We claim it suffices to prove Theorem 2 for Γ'. It is clear that the sum in Theorem 2 converges for Γ if and only if it does for Γ'. All we have to check is that if the embeddings ι_1 and ι_2 are equivalent when restricted to Γ' then they are equivalent on all of Γ. Since they are equivalent on Γ' there is an automorphism ϱ of $\mathrm{PSL}(2,\mathbf{R})$ such that $\iota_1=\varrho\circ\iota_2$ on Γ'. Thus we may assume $\iota_1=\iota_2$ on Γ', and we must show they are equal on Γ.

Now let $g\in\Gamma$ and $h\in\Gamma'$. Since Γ' is normal

$$\iota_1(ghg^{-1})=\iota_2(ghg^{-1}).$$

Therefore

$$\iota_2(g^{-1})\iota_1(g)\iota_1(h)=\iota_1(h)\iota_2(g^{-1})\iota_1(g).$$

For a fixed g this holds for every $h\in\Gamma'$ and this implies that $\iota_2(g^{-1})\iota_1(g)$ is the identity (the elements of $\mathrm{PSL}(2,\mathbf{R})$ commuting with any non-identity element form a one parameter abelian subgroup, so cannot contain all of $\iota_1(\Gamma)$). Thus $\iota_1(g)=\iota_2(g)$ for every $g\in\Gamma$, as required.

Thus it suffices to prove Theorem 2 for any normal, finite index subgroup of Γ. We may assume that either Γ is the fundamental group of a compact surface (homeomorphic to $\mathbf{H}/\iota_k(\Gamma)$ for $k=1,2$) or is a finitely generated free group (in which case $\mathbf{H}/\iota_k(\Gamma)$, $k=1,2$ are punctured compact surfaces, possibly different).

3. Proof of Theorem 2 for cocompact lattices

In this section we will prove Theorem 2 assuming Γ is a surface group, so assume $R_1=\mathbf{H}/\iota_1(\Gamma)$ and $R_2=\mathbf{H}/\iota_2(\Gamma)$ are both compact Riemann surfaces of genus p.

The Dehn–Nielsen theorem implies the existence of a homeomorphism, Φ, of $\overline{\mathbf{H}}$ to itself ($\overline{\mathbf{H}}$ includes the point at infinity) which intertwines the ι_1- and ι_2-actions of Γ. This map either preserves orientation or reverses it, so by composing with a reflection if necessary we may assume that Φ preserves orientation. Since composing ι_2, and consequently Φ, with a Möbius transformation doesn't alter the problem we may also assume Φ fixes any three points of our choice on $\mathbf{R}$. This will be convenient below. This homeomorphism can also be taken to be quasiconformal, but this is not important here. However, it is important to note that Φ is the lift to $\mathbf{H}$ of a homeomorphism $\varphi: R_1\to R_2$.

We also need some facts about the surfaces R_1 and R_2, which we take from [7]. On R_1 we can find $2p$ geodesics $\alpha_1,\alpha_2,\ldots,\alpha_{2p}$ such that each α_j is a simple closed curve, such that each α_j meets α_{j+1} in exactly one point, such that there are no other intersections

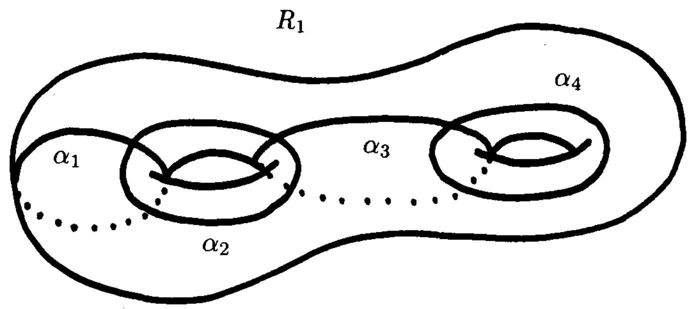

Fig. 1. Geodesics on R_1

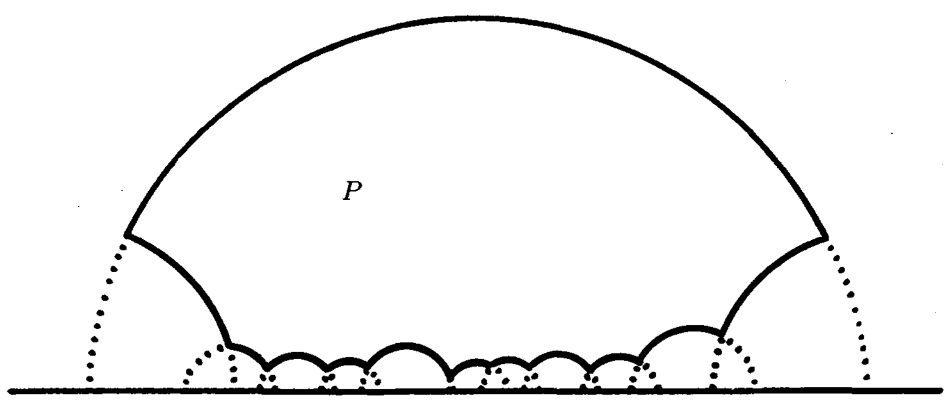

Fig. 2. A fundamental polygon

and such that $S_1=R_1\setminus(\alpha_1\cup...\cup\alpha_{2p})$ is simply connected. We can do this by first choosing any simple closed curves with these properties and then homotoping them to geodesics. See Figure 1. Let $\Psi_1:\mathbf{H}\to R_1$ denote the covering map induced by $\iota_1(\Gamma)$. A connected preimage $\Psi_1^{-1}(S_1)$ of S_1 is a hyperbolic polygon with $8p-4$ sides. It's a polygon since it's bounded by geodesics, and is convex because each interior angle is less than π (this holds on R_1 and Ψ preserves angles). Thus we obtain a tesselation $\mathcal{T}_1$ of the disk into identical convex polygons, and by $P\in\mathcal{T}_1$ we will mean one of these polygons. See Figure 2. Note that each side of the polygon can be extended to an infinite geodesic in $\mathbf{H}$ which consists entirely of edges in the tesselation. Also note that exactly four polygons meet at each vertex and adjacent angles at a vertex always sum to π.

On R_2 we choose geodesics $\beta_1,...,\beta_{2p}$ such that β_j is freely homotopic to $\varphi(\alpha_j)$ for each j. It then follows that the β's satisfy the same intersection relations as the α's (e.g., see [7, pp. 379–389]). Thus we can define S_2, Ψ_2 and $\mathcal{T}_2$ just as above. Note, however, that $\mathcal{T}_2$ is not necessarily the image of $\mathcal{T}_1$ under Φ.

By conjugating with Möbius transformations if necessary we may assume that $\Psi_k(i)\in S_k$ for $k=1,2$, i.e., that i is in the interior of some fundamental polygon for each tesselation. To evaluate the infinite sum in Theorem 2, we wish to group the elements of Γ into "generations". We could do this using minimum length representations of elements in terms of the generators of the group, but we will take a more geometrical approach based on the tesselation described above. After dividing Γ up into a finite number of pieces and composing with some Möbius transformations which fix i we may assume that we are only summing over the $\gamma\in\Gamma$ such that $|\iota_k(\gamma)(i)|\leqslant C$ for $k=1,2$. We will denote this subset of Γ by $\tilde{\Gamma}$. It will also be convenient to arrange that $\Phi(\infty)=\infty$ and that none of the edges of polygons under consideration are vertical. Thus each vertex of the tesselation corresponds to the crossing of semi-circles centered on the real axis (i.e., geodesics) which meet the axis at the points a, b, c, d labeled left to right. For each vertex v in the tesselation we single out the geodesic ray from v to b (the second from left) and call it a "red" ray. An edge of a polygon lying on a red ray is called a red edge.

Let us note that the choices of red rays are consistent between vertices, i.e., suppose a vertex v_1 lies on two geodesics in the tesselation, l_1 and l_2, that the red ray r_1 associated to v_1 lies along l_1 and that v_2 is another vertex of the tesselation that lies on this red edge. Then v_2's red ray r_2 is a subset of r_1. To see why, suppose not. Then there is a geodesic l_3 of the tesselation which intersects l_1 at v_2 and $r_2\subset l_3$. This means l_3's left endpoint lies to the right of l_1's left endpoint and its right endpoint lies to the right of l_1's. Thus l_3 must intersect l_2. But this is impossible since it means the 3 distinct geodesics on R_1 represented by these geodesics on $\mathbf{H}$ all meet each other, contradicting the way the $\{\alpha_j\}$ were chosen. Thus $r_2\subset r_1$. In particular this means that two red rays corresponding to distinct vertices are either disjoint or one is included in the other. If some geodesic were made up entirely of red ray, then that geodesic would have to intersect other geodesics at arbitrarily small angles, impossible since there are only $8p-4$ possible angles. Thus removing all the red rays from $\mathbf{H}$ leaves us with a dense, simply connected subdomain which contains no vertices of the tesselation.

Let P_0 be the polygon of $\mathcal{T}_1$ containing i. We can make the group into a graph by saying two elements γ, γ' are adjacent iff $\iota_1(\gamma)(P_0)$ and $\iota_1(\gamma')(P_0)$ share a non-red edge. Next we show that the graph is actually a tree, i.e., connected and without loops. Note that if the polygons P_1 and P_2 share a non-red edge e_0 and r_1 and r_2 are the red rays associated to the endpoints of e_0, then $r_1\cup e_0\cup r_2$ forms a path in $\mathbf{H}$ which separates P_1 and P_2. Thus if $P_1, P_2, \ldots, P_n$ is a path of distinct adjacent polygons (in the sense above) then P_n cannot be adjacent to P_1 since it is separated from P_1 by $r_1\cup e_0\cup r_2$. Thus there are no loops in the graph we have defined.

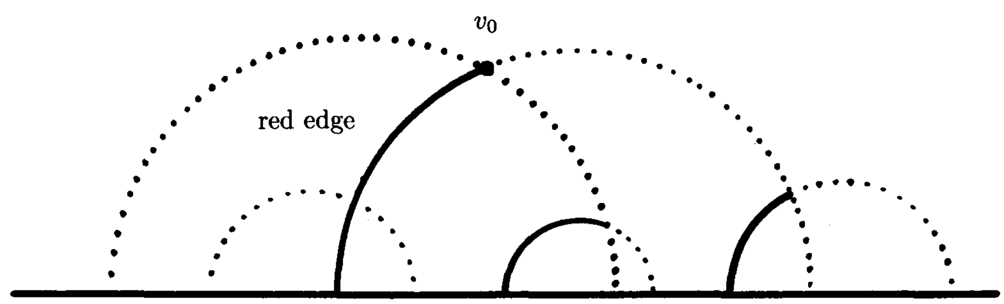

Fig. 3. The red edges

We label the identity in Γ as the 0th generation and say $\gamma \in \Gamma$ is in the nth generation if its tree distance to the identity is n. We denote the elements in the nth generation by $\mathcal{G}_n$. If $\gamma \in \mathcal{G}_n$ then $D(\gamma)$ is the collection of elements in $\mathcal{G}_{n+1}$ which are adjacent to γ (the "daughters" of γ).

Next we note that the tree structure induced by the other tesselation $\mathcal{T}_2$ is exactly the same. This is because Φ (thought of as a map from $\mathbf{R}$ to itself) is strictly increasing, so if we take two intersecting geodesics in $\mathbf{H}$, map them via Φ and then deform them into the corresponding geodesics, the order on the endpoints is not changed (in fact, the homotopy does not change the endpoints at all).

Given a polygon P in the nth generation it shares exactly one non-red edge e_0 with a polygon in the $(n-1)$st generation which we refer to as the "top edge" of P. We can arrange for this name to be geometrically as well as combinatorially justified in that P is in the bounded component of $\mathbf{H}\backslash l_0$ where l_0 is the geodesic containing e_0. We say a polygon P is "below" one of its edges e if P is in the bounded complementary component of $\mathbf{H}\backslash l$, l the geodesic containing e. First note that P can be below at most 2 of its edges. This is because any two such geodesics intersect (otherwise one separates the other from P) and there cannot be three such by an earlier argument. Thus there are at most two such edges and if there are two they must have a vertex in common (same argument). Therefore one of the two edges lies in the red ray associated to the common vertex. Now consider P_0, the polygon containing i. By breaking Γ into a finite number of pieces we need only consider polygons which are reached from P_0 by first passing through an edge of P_0 which P_0 is above (obvious meaning). For each of its daughters P_i, the edge shared with P_0 must be the unique non-red edge which P_i is below. Proceeding by induction gives the claim.

Now to each $\gamma \in \Gamma$ under consideration we will associate two open intervals $I_1(\gamma)$ and $I_2(\gamma)$ on the real line such that for $k=1,2$

$$\gamma, \gamma' \in \mathcal{G}_n,\ \gamma \neq \gamma' \Longrightarrow I_k(\gamma) \cap I_k(\gamma') = \varnothing, \tag{3.1}$$

$$\sum_{\gamma'\in D(\gamma)} |I_k(\gamma')| = |I_k(\gamma)|, \tag{3.2}$$

$$|I_k(\gamma)| \sim \mathrm{Im}(\iota_k(\gamma)(i)). \tag{3.3}$$

Thus, these intervals are arranged in generations. Moreover

$$\sum_{\tilde{\Gamma}} h^s(\iota_1(\gamma))h^{1-s}(\iota_2(\gamma)) \sim \sum_{\tilde{\Gamma}} |I_1(\gamma)|^s |I_2(\gamma)|^{1-s}.$$

Therefore it is enough to show the sum on the right converges if the embeddings ι_1 and ι_2 are not equivalent.

Now we will define the intervals and verify the three properties listed above. For convenience we take $k=1$, but of course $k=2$ is identical. Fix a $\gamma\in\mathcal{G}_n\subset\Gamma$ and let $P\in\mathcal{T}_1$, be the corresponding polygon. Let e_0 denote the edge which P shares with an $(n-1)$st generation polygon. Orient e_0 so that P is on its left and let v_0 be the terminal vertex (the endpoint in the counterclockwise direction). Let a be the endpoint on $\mathbf{R}$ of the red ray passing through v_0. Now let v_1 denote the other endpoint of e_0 and let b be the endpoint on $\mathbf{R}$ of the red ray passing through v_1. It's easy to check that $a<b$ since the rays don't intersect and can't have the same endpoint on $\mathbf{R}$ (otherwise the corresponding geodesics would be arbitrarily close and hence intersect). We define $I_1(\gamma)=(a,b)$. Note that $I_1(\gamma)$ is just the interior of the accumulation set on $\mathbf{R}$ of the points $\iota_1(\gamma')(i)$ for those γ' descended from γ in the tree.

Then (3.1) and (3.2) are clear from the definition and the fact that red rays corresponding to distinct vertices either don't intersect or lie on same geodesic. To check (3.3), note that

$$b-a \geqslant C\mathrm{Im}(v_1) \geqslant C\mathrm{Im}(\iota_1(\gamma)(i)),$$

since P has compact closure. To prove the other direction, we first note that

$$0 < \mathrm{Re}(v_0)-\mathrm{Re}(v_1) \leqslant C\mathrm{diam}(P) \leqslant C\mathrm{Im}(\iota_1(\gamma)(i)).$$

Next, let θ_0 be the interior angle of P at v_0 and note that the red ray going from v_0 to a is at at least angle θ_0 from the vertical (measured in the counter-clockwise direction) since it makes an angle of θ_0 with the other geodesic ray passing through v_0 whose endpoint is to the left of a (see Figure 4). This implies

$$\mathrm{Re}(v_0)-a \leqslant C(\theta_0)\mathrm{Im}(v_0) \leqslant C\mathrm{Im}(\iota_1(\gamma)(i)).$$

Thus since $b\leqslant\mathrm{Re}(v_1)$,

$$b-a \leqslant (\mathrm{Re}(v_1)-\mathrm{Re}(v_0))+(\mathrm{Re}(v_0)-a) \leqslant C\mathrm{Im}(\iota_1(\gamma)(i)),$$

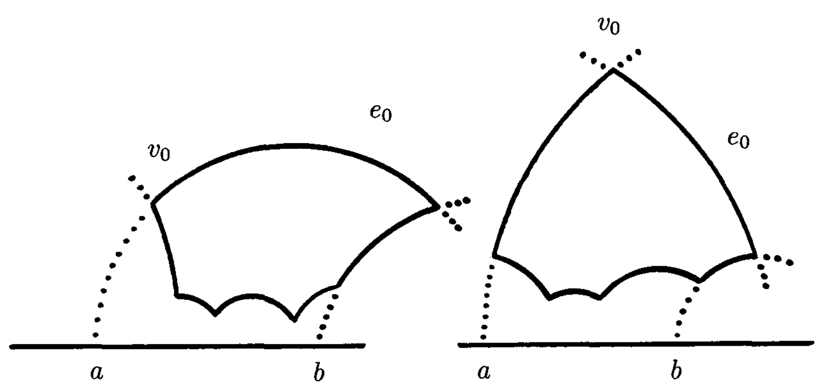

Fig. 4. The boundary intervals

as required.

We now show the sum converges. If we knew that the polygon P corresponding to $\gamma \in \Gamma$ was uniquely determined up to a Euclidean similarity by the sequence of numbers $\{|I_1(\gamma')|/|I_1(\gamma)| : \gamma' \in D(\gamma)\}$ then we would be done by the following argument: If $|I_1(\gamma')|/|I_1(\gamma)| = |I_2(\gamma')|/|I_2(\gamma)|$ for every $\gamma' \in D(\gamma)$ then by the the above hypothesis the corresponding polygons $P_1 \in \mathcal{T}_1$ and $P_2 \in \mathcal{T}_2$ are just a translate and dilate of each other. But this means there is a Möbius transformation mapping P_1 to P_2 which sends vertices and edges in P_1 to the corresponding vertices and edges in P_2. But this implies the surfaces R_1 and R_2 are conformal which contradicts our hypothesis that ι_1 and ι_2 are not conjugate embeddings. Thus the two sequences of numbers are not the same. Hence strict inequality must hold in Hölder's inequality applied to the two (finite) sequences

$$a_{\gamma'} = \left(\frac{|I_1(\gamma')|}{|I_1(\gamma)|}\right)^s, \quad b_{\gamma'} = \left(\frac{|I_2(\gamma')|}{|I_2(\gamma)|}\right)^{1-s}, \quad \gamma' \in D(\gamma)$$

with the conjugate indices $p = s^{-1}$ and $q = (1-s)^{-1}$. Thus there is an $\varepsilon > 0$ so that

$$\sum a_{\gamma'} b_{\gamma'} \leqslant (1-\varepsilon)\left(\sum a_{\gamma'}^p\right)^{1/p}\left(\sum b_{\gamma'}^q\right)^{1/q} \leqslant 1-\varepsilon.$$

So,

$$\sum_{D(\gamma)} |I_1(\gamma')|^s |I_2(\gamma')|^{1-s} \leqslant (1-\varepsilon)|I_1(\gamma)|^s |I_2(\gamma)|^{1-s}.$$

Moreover, given ι_1, ι_2 the sequences $\{a_{\gamma'}\}$ and $\{b_{\gamma'}\}$ which can arise in this way form a precompact family, so the ε can be chosen to depend only on ι_1 and ι_2. Therefore

$$\sum_{\tilde{\Gamma}} |I_1(\gamma)|^s |I_2(\gamma)|^{1-s} \leqslant C \sum_{n=0}^{\infty} (1-\varepsilon)^n < \infty.$$

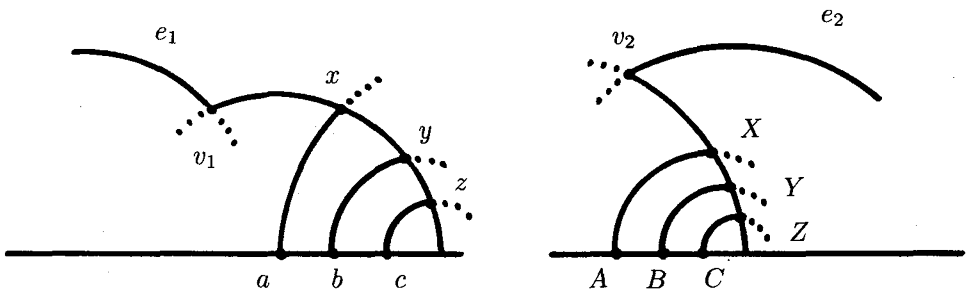

Fig. 5. The points a, b, c

We will not prove the claim in the previous paragraph. However, we will prove a slightly weaker result which is sufficient for our purpose. Let $D_1(\gamma)=D(\gamma)$ and $D_j(\gamma)=\{\gamma' \in D(\gamma'')$ for some $\gamma'' \in D_{j-1}(\gamma)\}$, i.e., $D_j(\gamma)$ are the jth generation descendants of γ.

LEMMA 3.1. *For any $\gamma \in \tilde{\Gamma}$, the numbers $\{|I_1(\gamma')| : \gamma' \in D_5(\gamma)\}$ determine the group $\iota_1(\Gamma)$ up to conjugation by a Euclidean similarity.*

The proof of Theorem 2 follows just as above, except that we break the sum over generations into 5 sums, depending on the value of the generation modulo 5.

To prove the lemma, recall that the sides of any polygon P of $\mathcal{T}_1$ are identified in pairs by elements of $\iota_1(\Gamma)$. The set of all such side pairings for P generate the whole group, so it is enough to show these Möbius transformations are determined by the data in the lemma. Consider the polygon P' of $\mathcal{T}_1$ corresponding to γ, let e_0' be its top edge, and let e_0 be the edge of P' diametrically opposite e_0'. We will apply the above considerations to the polygon P which lies below e_0.

Suppose τ is the Möbius transformation identifying two edges e_1 and e_2 of P. Orient e_1 and let v_1 be the initial endpoint so that P is on the right as we move along e_1 away from v_1. See Figure 5. Let r_1 be the half infinite ray starting from v_1 and which begins with the other edge of P with endpoint v_1. Let x, y and z be the first three vertices of $\mathcal{T}_1$ we come to along r_1 after leaving v_1. Let a, b and c denote the endpoints on $\mathbf{R}$ of the red rays passing through x, y and z respectively. There are several cases depending on where e_1 lies with respect to the top edge e_0, but in each case note that $a, b, c \in I_1(\gamma)$ and their positions can be computed from the numbers $\{|I_1(\gamma')| : \gamma' \in D_5(\gamma)\}$.

We orient e_2 so that P is on its left and define v_2 as above. Let r_2 be the transverse ray at v_0, but now chosen so that it does not contain an edge of P. We let X, Y, Z denote the first three vertices encountered and A, B, C the projections of these points onto $\mathbf{R}$ along red rays. Again we have $A, B, C \in I_1(\gamma)$ and their positions can be computed from the given information. But the side pairing map τ must map $a \to A$, $b \to B$ and $c \to C$, and so is completely determined by these points. This completes the proof of Theorem 2 in the cocompact case.

4. Proof of Theorem 2 for non-cocompact lattices

We now turn to case when Γ is a finitely generated free group, i.e., $\mathbf{H}/\iota_k(\Gamma)$, $k=1,2$, are finitely punctured compact Riemann surfaces. Unlike the previous case, the surface is not uniquely determined by the group. For example, the free group on two generators is the fundamental group of both a sphere with three punctures and a torus with one puncture. Thus different embeddings of Γ do not arise simply from homeomorphisms of a single surface. The isomorphism $\Phi=\iota_1^{-1}\circ\iota_2$ discussed in the previous section need not be induced by any homeomorphism of $\overline{\mathbf{H}}$ to itself. In the previous section we associated an interval on the boundary to each $\gamma\in\Gamma$. We will do this again here, but the intervals must be replaced by more general sets. This arises because there may be collections of generators for Γ which are side pairings of some polygon for $\iota_1(\Gamma)$ but not of any polygon for $\iota_2(\Gamma)$.

Although we lose some geometrical intuition, we gain combinatorial simplicity because the free group has no relations to worry about. For example, there is an obvious tree structure on the group given by reduced words in the generators.

A more serious problem is that Γ contains parabolic elements. When we get to the part of the proof were we want to say strict inequality must hold in Hölder's inequality we will not be able to take $\varepsilon>0$ uniformly for all $\gamma\in\Gamma$ because this fails for high powers of parabolic elements. Instead, we will have to argue that the convergence or divergence of the sum over Γ is unaffected if we simply drop such terms.

So suppose Γ is the free group on m generators. Then $R_1=\mathbf{H}/\iota_1(\Gamma)$ is a compact Riemann surface of genus $p\geqslant 0$ with $q>0$ points removed and $m=2p+q-1$. We can find m non-intersecting simple curves on R_1 with endpoints at the punctures which cut the surface into a simply connected domain S. These can be chosen to be geodesics, so S lifted to $\mathbf{H}$ becomes a $2m$ sided polygon with all of its vertices on the real line. We can choose m elements of $\iota_1(\Gamma)$ which identify the sides in pairs, and these elements give a set of generators $\{g_1, ..., g_m\}$ of Γ. See Figure 6.

We make Γ into a tree in the obvious way by saying γ and γ' are adjacent if there is a generator g_j such that $\gamma=\gamma' g_j^{\pm}$. Just as before we may assume there are no vertical edges in the tesselation and after breaking the group up into a finite number of pieces we need only sum over a "branch" $\tilde{\Gamma}$ which has been normalized so $\iota_k(\gamma)(i)\in[-1,1]\times(0,\frac{1}{2}]$ for $k=1,2$ and $\gamma\in\tilde{\Gamma}$ (for ι_1 this localization is clearly possible and for ι_2 it will be justified later). For each $\gamma\in\tilde{\Gamma}$ we have an associated polygon P with a distinguished "top" edge e_0 which is shared with a polygon of one lower generation. The edge e_0 has two endpoints

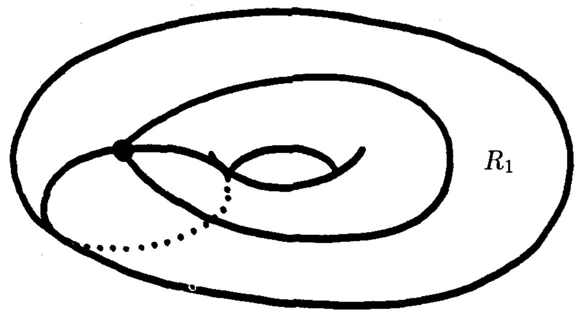

Fig. 6. Geodesics on R_1

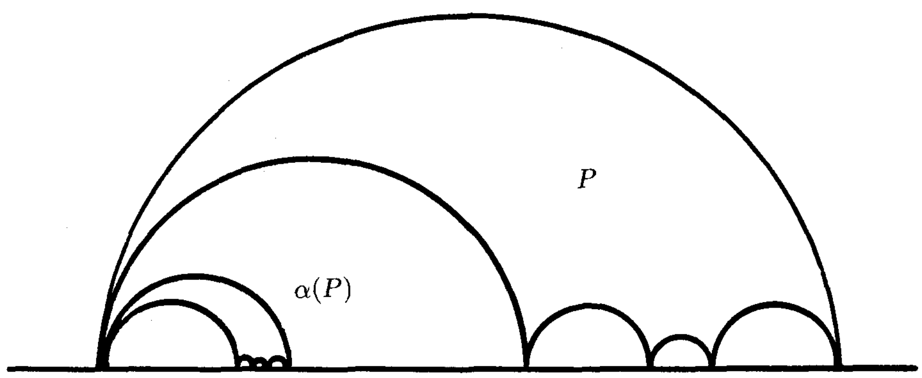

Fig. 7. A "bad" polygon and a parabolic α

a, b on $\mathbf{R}$ and we let $I_1(\gamma)=(a,b)$. Clearly (3.1) and (3.2) hold and

$$\operatorname{Im}(\iota_1(\gamma)(i)) \leqslant C|I_1(\gamma)|,$$

but the reverse inequality does not hold with a constant independent of γ. This is because if $\iota_1(\gamma)$ is a parabolic element we can have $|I_1(\gamma^n)| \sim n^{-1}$ but $\operatorname{Im}(\iota_1(\gamma^n)(i)) \sim n^{-2}$. See Figure 7. However,

$$\operatorname{Im}(\iota_1(\gamma)(i)) \sim \min_{\gamma' \in D(\gamma)} |I_1(\gamma')|. \tag{4.1}$$

We leave this as a simple exercise for the reader (one way to do this is to note that the (hyperbolic) angle subtended by each side of the polygon at the image of i is invariant).

We now wish to define $I_2(\gamma)$. It will not be an interval, but a finite union of intervals (the number of such intervals bounded independent of γ). For $\gamma \in \Gamma$ let $E_\gamma \subset \Gamma$ correspond to the reduced words in the $\{g_j\}$ which begin with the word representing γ. Then we define $I_2(\gamma)=\{x \in \mathbf{R}: x$ is in the Euclidean closure of $\iota_2(E_\gamma)\}$. If $\iota_2=\iota_1$ then this agrees with the more geometrical definition given in the last paragraph (it may be helpful to recall that our groups act on the left, so if $\gamma' \in E_\gamma$ then $I_1(\gamma') \subset I_1(\gamma)$). The point of this more elaborate definition is that the generators $\{g_1, ..., g_m\}$ may not correspond to side pairings of any polygon for $\iota_2(\Gamma)$ so that the geometrical definition doesn't make sense.

First we see why each $I_2(\gamma)$ is a finite union of intervals. It is enough to do this when γ is one of the generators $\{g_j\}$ or their inverses, since every $I_2(\gamma)$ is a Möbius image of one of these (depending on the last letter in the reduced word for γ; this is because if $\gamma' = \gamma \cdot \alpha$ then $I_2(\gamma') = \iota_2(\gamma)(I_2(\alpha))$). Repeating our earlier argument we can choose a new set of generators $\{h_1, ..., h_m\}$ for Γ which correspond to side pairings of some polygon for $\iota_2(\Gamma)$. Let $F_\gamma \subset \Gamma$ be the reduced words in the $\{h_j\}$ beginning with the word for γ. If we let $I_3(\gamma)$ be the closure of $\iota_2(F_\gamma)$ on $\mathbf{R}$, then our remarks above show it is an interval (i.e., a "base" for some polygon of which the $\{h_j\}$ are side pairings). Any $\gamma \in \Gamma$ now has a "g-representation" as a word in the $\{g_j\}$ and a "g-length" as well as a an "h-representation" and "h-length". Now choose $M \in \mathbf{N}$ so that each h_k can be represented as a word of length less than M in the $\{g_j\}$ and conversely. Thus for any element γ, "g-length"$\leqslant M\cdot$"h-length"$\leqslant M^2\cdot$"g-length". Thus if γ has "h-length" greater than M^2, it has "g-length" greater than M, so lengthening its "h-representation" by adding a generator to the right of the word cannot alter the first letter of its "g-representation". So if γ has "h-length" M^2+1 and g_j is a generator then either $F_\gamma \subset E_{g_j}$ or $F_\gamma \cap E_{g_j} = \varnothing$. Thus either $I_3(\gamma) \subset I_2(g_j)$ or $I_3(\gamma) \cap I_2(g_j) = \varnothing$ (except for endpoints). Thus $I_2(g_j)$ consists of the union of less than $(2m)^{M^2+1}$ sets of the form $I_3(\gamma)$, so is a finite union of intervals. The same argument shows that each interval $I_3(\gamma)$ where γ has "h-length" greater than M^2 is in exactly one of the sets $I_2(g_j^{\pm})$. Thus the sets $\{I_2(\gamma)\}$ satisfy (3.1) and (3.2). Conformal invariance (as in (4.1)) also implies

$$\operatorname{Im}(\iota_2(\gamma)(i)) \sim \min_{\gamma' \in D(\gamma)} |I_2(\gamma')|. \tag{4.2}$$

This argument also shows that the argument breaking Γ into a finite number of pieces $\tilde{\Gamma}$ which are "local" is correct.

Next we note that the embeddings ι_1 and ι_2 are determined by the lengths of these sets. To avoid problems with the parabolic elements we will define a second tree structure on Γ in which each element has infinitely many "daughters" and high powers of parabolics are entirely omitted. Fix γ and consider the corresponding polygon P for ι_1. Denote the "top" edge by e_0 (the unique edge shared with a polygon of lower generation). Denote its left and right endpoints by v_1 and v_2. The vertex v_1 corresponds to a puncture on R_1 and a simple loop on R_1 around this puncture (considered as an element of the fundamental group) corresponds to a parabolic element $\iota_1(\alpha)$ of $\iota_1(\Gamma)$. Then $\iota_1(\gamma\alpha\gamma^{-1})$ is parabolic with v_1 as its (unique) fixed point on $\mathbf{R}$ and (by taking α^{-1} if necessary) we may assume that $\gamma\alpha$ is a descendent of γ. See Figure 7. There is a corresponding parabolic element

β corresponding to v_2. By taking powers (if necessary) we may assume α and β have the same "g-length", say r. Let $\tilde{D}(\gamma)=D_r(\gamma)\setminus\{\gamma\alpha,\gamma\beta\}$, or more geometrically, $\tilde{D}(\gamma)$ are the elements corresponding to all the intervals $\{I_1(\gamma'):\gamma'\in D_r(\gamma)\}$ except for the ones on the far left and far right.

Now define

$$T_1(\gamma)=\bigcup_{k=0}^{\infty}\{\gamma\alpha^k\gamma^{-1}\gamma',\,\gamma\beta^k\gamma^{-1}\gamma':\gamma'\in\tilde{D}(\gamma)\}\cup\bigcup_{k=1}^{\infty}\{\gamma\alpha^k\beta,\,\gamma\beta^k\alpha\}.$$

The corresponding intervals $\{I_1(\gamma')\}$, $\gamma'\in T_1(\gamma)$, form a disjoint (except for endpoints) cover of $I_1(\gamma)$ and elements in $T_1(\gamma)$ can be enumerated so that they satisfy

$$\frac{1}{Cn^2}\leqslant|I_1(\gamma_n)|\leqslant\frac{C}{n^2}. \tag{4.3}$$

This is because $|I_1(\gamma\alpha^n\gamma^{-1}\gamma')|\sim|I_1(\gamma')|/n^2$ for any $\gamma'\sim\tilde{D}(\gamma)$, as we shall see in some computations below. A similar construction appears in Beardon's paper [3].

T_1 defines a new tree structure on a subset Γ (elements of the form $\gamma\alpha^k$ and $\gamma\beta^k$ are omitted). Let $T_j(\gamma)=\{\gamma''\in T_1(\gamma')$ for some $\gamma'\in T_{j-1}(\gamma)\}$ be the jth generation descendants of γ and let $T_\infty(\gamma)=\bigcup_j T_j(\gamma)$. Then

LEMMA 4.1. *For any s, $0<s<1$ there is a C independent of $\gamma\in\Gamma$ such that*

$$\sum_{E_\gamma}\operatorname{Im}(\iota_1(\gamma')(i))^s\operatorname{Im}(\iota_2(\gamma')(i))^{1-s}\leqslant C\sum_{T_\infty(\gamma)}|I_1(\gamma')|^s|I_2(\gamma')|^{1-s}.$$

LEMMA 4.2. *If ι_1 and ι_2 are not equivalent and $0<s<1$ then there exists $\varepsilon>0$ independent of $\gamma\in\Gamma$ such that*

$$\sum_{T_1(\gamma)}|I_1(\gamma')|^s|I_2(\gamma')|^{1-s}\leqslant(1-\varepsilon)|I_1(\gamma)|^s|I_2(\gamma)|^{1-s}.$$

These easily prove the sum in Theorem 2 converges; 4.1 reduces to showing the sum over the subtree converges and 4.2 proves this via the geometric formula. To prove Lemma 4.1, note that if $\gamma'\in\tilde{D}(\gamma)$ then $\operatorname{Im}(\iota_1(\gamma)(i))\leqslant C|I_1(\gamma')|$. Similarly, (4.2) implies $\operatorname{Im}(\iota_2(\gamma)(i))\leqslant C|I_2(\gamma')|$. Thus for each $\gamma\in E_\gamma$ we pick some element of $\gamma'\in\tilde{D}(\gamma)$. Since each γ' is chosen only once, the inequality in Lemma 4.1 follows immediately.

To prove Lemma 4.2 we will first show that if the infinite sequences $\{|I_1(\gamma')|\}$, $\{|I_2(\gamma')|\}$, $\gamma'\in T_1(\gamma)$ are equal then ι_1 and ι_2 are equivalent. Fix some $\gamma''\in\tilde{D}(\gamma)$ and let $I_1(\gamma'')=[x,y]$ and $I_2(\gamma'')=\bigcup_{j=1}^N[x_j,y_j]$. Our first step is to prove $N=1$ and $y-x=y_1-x_1$. We will assume that the left endpoint of $I_1(\gamma)$ is $\{0\}$ so that $\iota_1(\alpha)$ has the form

$$\tau_1(z)=\frac{z}{az+1}$$

for some $a>0$. An easy computation shows that $\iota_1(\alpha^n)$ has the form

$$\frac{z}{naz+1}.$$

Therefore

$$\iota_1(\alpha^n)(y)-\iota_1(\alpha^n)(x)\sim\frac{1}{n^2}.$$

Thus by our hypothesis

$$\sum_j |\iota_1(\alpha^n)(y_i)-\iota_2(\alpha^n)(x_i)|\sim\frac{1}{n^2},$$

and therefore each term is about this size. This is only possible if $\iota_2(\alpha)$ is also parabolic (otherwise they all decrease geometrically). By conjugating with a translation and reflection if necessary we may assume it fixes $\{0\}$ and has its pole on the negative axis, i.e.,

$$\iota_2(\alpha)=\tau_2(z)=\frac{z}{Az+1}$$

for some $A>0$. Thus

$$\tau_1^n(y)-\tau_2^n(x)=\sum_j \tau_2^n(y_j)-\tau_2^n(x_j)$$

$$\frac{y-x}{(nay+1)(nax+1)}=\sum_j \frac{y_j-x_j}{(nAy_j+1)(nAx_j+1)}.$$

Fixing everything except n, both sides are rational functions which agree at infinitely many points (the positive integers). Therefore, they are the same function. Since the intervals $[x_j,y_j]$ have disjoint interiors and both sides must have the same poles we see $N=1$, $ay=Ay_1$ and $ax=Ax_1$. Since $y-x=|I_1(\gamma'')|=|I_2(\gamma'')|=y_1-x_1$ we must have $x=x_1$ and $y=y_1$.

Thus for any $\gamma''\in\tilde{D}(\gamma)$ we have $I_1(\gamma'')=I_2(\gamma'')$. Now let P be the polygon associated to γ and suppose two edges e_1 and e_2 are identified by an element of the group g. As in the last section this Möbius transformation is completely determined by its action on three points. For example, if we take three vertices of the polygon with e_1 as its top edge, their positions and their images' positions are completely determined by $\{|I_1(\gamma')|:\gamma'\in T_1(\gamma)\}$. Details are left to the reader.

Thus if ι_1 and ι_2 are not equivalent, we can apply the strict inequality in Hölder's inequality and deduce that there is an $\varepsilon>0$ (depending on γ) such that

$$\sum_{T_1(\gamma)} |I_1(\gamma')|^s|I_2(\gamma')|^{1-s}\leqslant(1-\varepsilon)|I_1(\gamma)|^s|I_2(\gamma)|^{1-s}. \tag{4.4}$$

To remove the dependence of ε on γ we will use the following argument. Let P_0 be a fixed polygon for ι_1, with "top edge" e corresponding to the generator or inverse generator g_1, meaning that if $P_{00} \in T_1$ is the base polygon, the polygon containing i, and if f is the edge of P_{00} corresponding to e, then g_1 is that generator, or that inverse of a generator, which maps f to some other edge of P_{00}. We translate and dilate so that P_0's "base" on $\mathbf{R}$ is $[-1,1]$. Let P be any other polygon in the tesselation whose top edge also corresponds to g_1 and also rescale it so its base is $[-1,1]$. Then there is a Möbius transformation τ from P_0 to P which belongs to the one parameter family which fixes -1 and 1. Let P' be a "daughter" of P_0 and let P'' be the corresponding daughter of P. Then unless P' is the far left or far right daughter, $1/C \leqslant |\tau'(z_1)|/|\tau'(z_2)| \leqslant C$ for $z_1, z_2 \in P'$ with a constant depending only on the size of the leftmost and rightmost intervals of P_0 (Harnack's inequality). Thus up to Euclidean similarities, P'' is only a bounded distortion of P', i.e., the P'''s which arise in this way belong to a compact family of polygons. Therefore the vectors $\{I_1(\gamma')\}$ considered in (4.4) come from a compact family. In fact, we can think of generating them by taking a fixed collection of intervals, say $\{I_1(\gamma')\}$ for some fixed γ, rotating around $\iota_1(\gamma)(i)$ by a Möbius transformation, and dilating so the resulting set has total length 1. The "bad" case is when the endpoint of an interval is rotated to a point near ∞ and so one component of the vector is much larger than the others. The argument above says this case never arises for the elements in T_∞. The vectors $\{I_2(\gamma)\}$ occurring in (4.4) also arise from the same rotate and normalize construction although now we have finite unions of intervals rather than just intervals. However, if (4.4) holds only for small enough ε then the vector $\{I_2(\gamma')\}$ cannot degenerate unless $\{I_1(\gamma')\}$ does, so these vectors also belong to a compact family. The sum on the left of (4.4) varies continuously over this family, so an ε can be chosen which holds uniformly on the compact subset. Repeating this argument for each generator and its inverse, we see that any polygon which is not a leftmost or rightmost daughter of its "mother" belongs (up to Euclidean similarities) to a compact family. In particular, this holds for everything in $T_1(\gamma)$ (a special argument is needed for the elements of the form $\gamma\beta^k\alpha$ and $\gamma\alpha^k\beta$ but this is very similar, and uses the fact that the estimate on the derivative used above actually holds uniformly except near one of $I_1(\gamma)$'s endpoints).

This proves Lemma 4.2, so we have now completed the proof of Theorem 2 in all cases.

5. Proof of Theorems 3 and 4

First we prove Theorem 3. As in Theorem 2, it is enough to only consider surface groups and free groups. In the surface group case there is nothing to do because the inequality

$$\sum_{\gamma' \in D_4(\gamma)} |I_1(\gamma')|^s |I_2(\gamma')|^{1-s} \leqslant (1-\varepsilon)|I_1(\gamma)|^s |I_2(\gamma)|^{1-s}$$

(with a finite number of terms) implies

$$\sum_{\gamma' \in D_4(\gamma)} (|I_1(\gamma')|^s |I_2(\gamma')|^{1-s})^{1-\delta} \leqslant (1-\varepsilon/2)|I_1(\gamma)|^{(1-\delta)s} |I_2(\gamma)|^{(1-\delta)(1-s)}$$

if δ is small enough (depending only on the number of terms in the sum). By dividing the group into generations and using the geometric formula as before we see the sum in Theorem 3 converges.

In the free group case, Hölder's inequality is applied to an infinite sum so we need to be a little more careful. Let $\{a_n\}$ be an enumeration of the lengths $|I_1(\gamma')|$, $\gamma' \in T_1(\gamma)$. As noted in the previous section these can be selected so

$$\frac{1}{Cn^2} \leqslant a_n \leqslant \frac{C}{n^2}.$$

Let $\{b_n\}$ denote $|I_2(\gamma')|$ with the same ordering. We may also normalize so

$$\sum_{n=1}^{\infty} a_n = 1, \quad \sum_{n=1}^{\infty} b_n = 1,$$

and we assume that

$$\sum_n a_n^s b_n^{1-s} \leqslant (1-\varepsilon)\Big(\sum_n a_n\Big)^s \Big(\sum_n b_n\Big)^{1-s} = 1-\varepsilon \tag{5.1}$$

for some $\varepsilon > 0$. Now fix a large integer N and $0 < \delta \leqslant s/(s+3)$ small (both to be chosen below) and use Hölder's inequality with the conjugate indices $p = ((1-s)(1-\delta))^{-1}$ and $q = (s+\delta-s\delta)^{-1}$ to show

$$\begin{aligned} \sum_{n \geqslant N} (a_n^s b_n^{1-s})^{1-\delta} &\leqslant \Big(\sum_{n \geqslant N} a_n^{s(1-\delta)q}\Big)^{1/q} \Big(\sum_{n \geqslant N} b_n^{(1-s)(1-\delta)p}\Big)^{1/p} \\ &\leqslant C\Big(\sum_{n \geqslant N} n^{-2s(1-\delta)/(s+\delta-s\delta)}\Big)^{1/q} \Big(\sum_{n \geqslant N} b_n\Big)^{1/p} \\ &\leqslant C\Big(\sum_{n \geqslant N} n^{-3/2}\Big)^{1/q} \cdot 1 \\ &\leqslant CN^{-s/2}. \end{aligned}$$

Now choose N so $CN^{-s/2}\leqslant\varepsilon/4$. Thus

$$\sum_n (a_n^s b_n^{1-s})^{1-\delta}\leqslant\sum_{n<N}(a_n^s b_n^{1-s})^{1-\delta}+\varepsilon/4.$$

By choosing δ small enough (depending only on N) and using (5.1) the first term can be made less than $1-\varepsilon/2$. Thus

$$\sum_n (a_n^s b_n^{1-s})^{1-\delta}\leqslant\left(1-\frac{\varepsilon}{4}\right)\left(\left(\sum_n a_n\right)^s\left(\sum_n b_n\right)^{1-s}\right)^{1-\delta}.$$

Using the geometric formula now shows the sum in the theorem converges and this proves Theorem 3.

To prove Theorem 4 we group, as above, the intervals $\{I_i(\gamma)\}$ into nested generations and let $\mathcal{C}_n$ denote the collection of nth generation intervals such that $|I_1(\gamma)|\leqslant|I_2(\gamma)|$. Then by our estimate

$$\sum_{\mathcal{C}_n}|I_1(\gamma)|^{1-\delta}\leqslant\sum_{n\text{th generation}}|I_1(\gamma)|^{(1-\delta)/2}|I_2(\gamma)|^{(1-\delta)/2}\leqslant(1-\varepsilon)^n.$$

Let $E_n=\bigcup_{\mathcal{C}_n}I_1(\gamma)$, $E=\bigcap_{n=1}^\infty\bigcup_{k>n}E_k$, $F_n=\mathbf{T}\setminus E_n$ and $F=\bigcap_{n=1}^\infty\bigcup_{k>n}F_k$. Then $E\cup F=\mathbf{T}$ since every point is in either infinitely many E_n's or F_n's. Also E can be covered by intervals satisfying

$$\sum_{k>n}\sum|I_1(\gamma)|^{1-\delta}\leqslant\sum_{k>n}(1-\varepsilon)^k<C(1-\varepsilon)^n.$$

Thus $\dim(E)\leqslant 1-\delta$. The same calculation shows $\dim(\beta(F))\leqslant 1-\delta$. Since $E^c\subset F$ this proves Theorem 4.

6. Representation of PSL(2, R) and SL(2, R)

In this section we will review the definitions and basic properties of the irreducible, unitary representations of $G=\mathrm{SL}(2,\mathbf{R})$. This is primarily for the benefit of the reader not previously acquainted with the unitary representation theory of $\mathrm{SL}(2,\mathbf{R})$. Two unitary representations $(\pi,\mathcal{H})$ and $(\pi',\mathcal{H}')$ are called equivalent if there is a unitary map $J:\mathcal{H}\to\mathcal{H}'$ which intertwines the G-actions, i.e. $J\circ\pi=\pi'\circ J$. The representation $(\pi,\mathcal{H})$ is irreducible if there are no non-trivial closed invariant subspaces. The *unitary dual* of G, denoted $\widehat{G}$, is the set of equivalence classes of irreducible unitary representations of G. The unitary dual of $\mathrm{SL}(2,\mathbf{R})$ is known and consists of three parts: the discrete series, the principal series and the complementary series.

To describe this dual we will work with a conjugate version of G, namely

$$\mathrm{SU}(1,1)=\left\{\begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix}\in \mathrm{GL}(2,\mathbf{C}):|\alpha|^2-|\beta|^2=1\right\}.$$

Then G acts on the unit disk, $\mathbf{D}$, and unit circle, $\mathbf{T}$, by the Möbius transformations

$$g(z)=\frac{\alpha z+\beta}{\bar\beta z+\bar\alpha}.$$

Define

$$K=\{g\in G:g(0)=0\}=\left\{k(\theta)=\begin{pmatrix}e^{i\theta} & 0\\ 0 & e^{-i\theta}\end{pmatrix}:\theta\in\mathbf{R}\right\},$$

and note that any unitary representation $(\pi,\mathcal{H})$ of G is a representation of K by restriction. Say that $v\in\mathcal{H}$ is K-covariant or n-K-covariant if $\pi(k(\theta))v=e^{in\theta}v$ for some integer n. We say v is K-finite if it is the sum of finitely many K-covariant vectors. The K-finite vectors are dense in $\mathcal{H}$ as follows from the very well known representation theory of K.

If $(\pi,\mathcal{H})$ is a unitary representation and if $v_1,v_2\in\mathcal{H}$ then the function $m(g)=m_{v_1,v_2}(g)=\langle v_1,\pi(g)v_2\rangle$ is called a matrix coefficient of π. One way of understanding π is to understand something about its matrix coefficients. Matrix coefficients between pairs of K-covariant vectors are particularly simple and useful. Suppose v_j is n_j-K-covariant, $j=1,2$. Then

$$m(k(\theta_1)gk(\theta_2))=e^{-in_1\theta_1}e^{-in_2\theta_2}m(g). \tag{6.1}$$

We have $G=KA^+K$ where

$$A=\left\{a(R)=\begin{pmatrix}\cosh R/2 & \sinh R/2\\ \sinh R/2 & \cosh R/2\end{pmatrix}:R\in\mathbf{R}\right\},$$

and we set $A^+=\{a(R):R\geqslant 0\}$. In terms of the function h introduced in Section 1, $R=-\log h(g)$. Because of (6.1) above, a matrix coefficient, m, between two K-covariant vectors is determined by the values $m(a(R))$ for $R\geqslant 0$. The matrix $-I$ is in the center of G, so $\pi(-I)$ commutes with $\pi(g)$ for all $g\in G$. If π is irreducible then by Schur's lemma $\pi(-I)$ must be a scalar and that scalar must be either ± 1. This gives a division of $\widehat{G}$ into two pieces. The first piece, where $-I$ acts as $+1$, is the unitary dual of $\mathrm{PSL}(2,\mathbf{R})$.

The *discrete series*, $\widehat{G}_{\mathrm{disc}}$, consists of (equivalence classes of) irreducible representations whose matrix coefficients are in $L^2(G)$. The results of this paper do not apply to discrete series representations, but we include some information on them for completeness. For each discrete series representation there is an integer $k\geqslant 2$ such that the matrix coefficient between any two K-covariant vectors satisfies $m(a(R))\sim Ce^{-kR/2}$ as $R\to\infty$. For given k there are precisely two such discrete series representations $\pi_k^{\pm}$, one of which has n-K-covariant vectors for $n=k, k+2, k+4, \ldots$ and the other for $n=-k, -k-2, -k-4, \ldots$. The action of $\pi(-I)$ is by $+1$ or -1 according to whether k is even or odd. A more precise description of the discrete series is omitted (see [11], [14]).

The *spherical principal series* is indexed by $\{s\in\mathbf{C}: \mathrm{Re}(s)=\frac{1}{2}\}$. The representation π_s acts on $L^2(\mathbf{T})$ by

$$(\pi_s(g)F)(e^{i\varphi}) = |dg^{-1}(e^{i\varphi})/d(e^{i\varphi})|^s F(g^{-1}(e^{i\varphi})),$$

or more concretely by

$$\left(\pi_s\begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix}F\right)(e^{i\varphi}) = |-\bar\beta e^{i\varphi}+\alpha|^{-2s} F\left(\frac{\bar\alpha e^{i\varphi}-\beta}{-\bar\beta e^{i\varphi}+\alpha}\right).$$

The representations π_s and $\pi_{1-s}=\pi_{\bar s}$ are equivalent. The matrix coefficient $m(g)$ between any two K-covariant vectors satisfies

$$m(a(R)) = C_+e^{-sR}+C_-e^{-(1-s)R}+o(e^{-R/2}),$$

if $s\neq\frac{1}{2}$ and

$$m(a(R)) = CRe^{-R/2}+O(e^{-R/2}),$$

if $s=\frac{1}{2}$. The action of K on $L^2(\mathbf{T})$ is by $(\pi_s(k(\theta)F)(e^{i\varphi})=F(e^{-2i\theta}e^{i\varphi})$. Thus $\pi_s(-I)=\pi_s(k(\pi))$ is the identity and there are n-K-covariant vectors for all even integers n. In particular there is a K-invariant vector for each of these representations and this is why they are called *spherical* representations.

If we use the G-invariant measure $4(1-x^2-y^2)^{-2}\,dx\,dy$ on $\mathbf{D}$, then $L^2(\mathbf{D})$ is a unitary representation of G under the action by translation. The Laplace–Beltrami operator, $\Delta=(1-x^2-y^2)^2(d^2/dx^2+d^2/dy^2)$, commutes with the action of G and hence its eigenspaces are preserved by the action of G. More precisely, the generalized eigenspaces which occur in the spectral decomposition of $L^2(\mathbf{D})$ with respect to Δ are themselves representation spaces for representations of G. The representations obtained in this way constitute the principal spherical series with π_s being the representation on the generalized eigenspace of $\lambda=s^2-s$. The λ-eigenfunction of Δ corresponding to a function F in $L^2(\mathbf{T})$ is given by $(\mathcal{P}_sF)(z)=\langle F, \pi_s(g)1\rangle$, where $g(0)=z$. Since $1\in L^2(\mathbf{T})$ is fixed by K and since K is the stabilizer of the origin, the choice of g is immaterial. The map $\mathcal{P}_s$ intertwines the π_s action of G on $L^2(\mathbf{T})$ to the translation action on the λ-eigenspace of Δ. The map $J_s=\mathcal{P}_{1-s}^{-1}\mathcal{P}_s$ is unitary and intertwines the π_s and π_{1-s} actions on $L^2(\mathbf{T})$. The map J_s is called the *intertwining map* and is given by $J_s(1)=1$ and $J_s(e^{\pm in\varphi})=e^{\pm in\varphi}\prod_{j=0}^{n-1}(j+(1-s))/(j+s)$ for $n>0$.

With two exceptions, the representations of the *nonspherical principal series* are indexed by $\{s\in\mathbf{C}\colon \mathrm{Re}(s)=\frac{1}{2}, s\neq\frac{1}{2}\}$. The representation π'_s acts on $L^2(\mathbf{T})$ by

$$\pi'_s\begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix}F(e^{i\varphi})=(-\bar\beta e^{i\varphi}+\alpha)|-\bar\beta e^{i\varphi}+\alpha|^{2s-1}F\left(\frac{\bar\alpha e^{i\varphi}-\beta}{-\bar\beta e^{i\varphi}+\alpha}\right).$$

Again, π'_s and π'_{1-s} are equivalent. The matrix coefficient between any two K-covariant vectors satisfies

$$m(a(R))=C_+e^{-sR}+C_-e^{-(1-s)R}+o(e^{-R/2}),$$

as $R\to\infty$. The action of K is given by $\pi'_s(k(\theta))=e^{i\theta}F(e^{2i\theta}e^{i\varphi})$, so that there are n-K-covariant vectors for all odd n and $\pi'_s(-I)=-1$.

There are two representations of the nonspherical series corresponding to the index $s=\frac{1}{2}$. One of these acts by

$$\left(\pi'_{1/2}\begin{pmatrix}\alpha & \beta\\ \bar\alpha & \bar\beta\end{pmatrix}F\right)(e^{i\varphi})=(-\bar\beta e^{i\varphi}+\alpha)F\left(\frac{\bar\alpha e^{i\varphi}-\beta}{-\bar\beta e^{i\varphi}+\alpha}\right),$$

on the Hardy space $H^2(\mathbf{T})$ of boundary values of holomorphic functions. There are n-K-covariant vectors for $n=1,3,5\ldots$ and the matrix coefficient between any two K-covariant vectors satisfies $m(a(R))\sim Ce^{-R/2}$ as $R\to\infty$. The other representation, $\pi''_{1/2}$, can be given by the same formula as $\pi'_{1/2}$ but acting on $L^2(\mathbf{T})\ominus H^2(\mathbf{T})$ or can be given, in an alternative realization, by

$$\left(\pi''_{1/2}\begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix}F\right)(e^{i\varphi})=\overline{(-\bar\beta e^{i\varphi}+\alpha)}\,F\left(\frac{\bar\alpha e^{i\varphi}-\beta}{-\bar\beta e^{i\varphi}+\alpha}\right),$$

acting on the conjugate of $H^2(\mathbf{T})$. This representation has n-K-covariant vectors for $n=-1,-3,-5,\ldots$. Otherwise it is similar to $\pi'_{1/2}$. These two representations are called the *boundary of discrete series* representations.

The spherical and nonspherical principal series and make up $\widehat{G}_{\mathrm{princ}}$, the principal series part of the unitary dual. Principal series representations are characterized as having matrix coefficients which are not in $L^2(G)$ but which are in $L^{2+\varepsilon}(G)$ for every $\varepsilon>0$. The principal and discrete series are necessary and sufficient to decompose $L^2(G)$ (itself a unitary representation of G) as a direct integral of irreducible representations.

The *spherical complementary series* is indexed by $\{s\in(0,1)\colon s\neq\frac{1}{2}\}$. The representation π_s acts on the Sobolev space of functions on $\mathbf{T}$ with $\frac{1}{2}-s$ derivatives in L^2. Specifically, π_s acts on $\mathcal{H}_s=\{\sum_{n=-\infty}^{\infty}a_ne^{in\theta}\}$ with the norm

$$\left\|\sum a_ne^{in\theta}\right\|_s^2=|a_0|^2+\sum_{n=1}^{\infty}(|a_n|^2+|a_{-n}|^2)\prod_{j=0}^{n-1}\frac{j+(1-s)}{j+s}<\infty.$$

The action is given by exactly the same formula as for the spherical principal series namely

$$(\pi_s(g)F)(e^{i\varphi}) = |dg^{-1}(e^{i\varphi})/d(e^{i\varphi})|^s F(g^{-1}(e^{i\varphi})).$$

The action is unitary with respect to the norm defined above. As usual π_s and π_{1-s} are equivalent. The action of K is just as for the spherical principal series, there are n-K-covariant vectors for all even n, and the matrix coefficient between any two K-covariant vectors satisfies $m(a(R)) \sim Ce^{-(1-s)R}$ as $R \to \infty$ if we take $s \in (\frac{1}{2}, 1)$. The complementary series part of $\widehat{G}$, $\widehat{G}_{\text{comp}}$ consists of π_s for $\frac{1}{2} < s \leqslant 1$. All the equivalence classes in $\widehat{G}$ have now been described; i.e., $\widehat{G} = \widehat{G}_{\text{disc}} \amalg \widehat{G}_{\text{princ}} \amalg \widehat{G}_{\text{comp}}$.

7. Proof of Theorem 1

In this and the next section we will use the notation $\iota_1 \sim \iota_2$ to indicate that the embeddings are equivalent and similarly for representations $\pi_1 \sim \pi_2$. If $\iota_1 \sim \iota_2$ then Theorem 1 follows from [6, Proposition 2.5], so we will assume $\iota_1 \not\sim \iota_2$ and prove that $\pi_1 \circ \iota_1 \not\sim \pi_2 \circ \iota_2$. We first prove Theorem 1 in the case when $\pi_1 = \pi_2$. We denote our representation by π_s where s is as in the previous section (in particular, $\frac{1}{2} < s < 1$ if π_s is in the complementary series). We shall consider the representation on the Hilbert space $\mathcal{H}_s$.

Let $\mathbf{1}$ represent the constant function in $\mathcal{H}_s$ and let P be the orthogonal projection of $\mathcal{H}_s$ onto the constants (more abstractly, this is the projection onto the 1 dimensional subspace fixed by the circle group $K \subset \mathrm{PSL}(2, \mathbf{R})$). Let Γ be a group and let $\iota: \Gamma \to \mathrm{PSL}(2, \mathbf{R})$ be a lattice embedding. The first step is to write the projection P in terms of π_s and ι. If we were working with all of $\mathrm{PSL}(2, \mathbf{R})$ this would be easy since the projection is just given by averaging over all rotations of the circle. We want to write P as an average over elements of $\iota(\Gamma)$ and this requires more work. As before we consider

$$h(g) = \exp(-d(0, g(0)))$$

for $g \in \mathrm{PSL}(2, \mathbf{R})$. Now define

$$P^\iota_\varepsilon = C^s_\varepsilon \sum_{\gamma \in \Gamma} h^{s+\varepsilon}(\iota(\gamma))\, \pi_s(\iota(\gamma))$$

where

$$C^s_\varepsilon = \frac{\mathrm{vol}(\mathrm{PSL}(2, \mathbf{R})/\iota(\Gamma))}{\int_{\mathrm{PSL}(2,\mathbf{R})} \langle \pi_s(g)\mathbf{1}, \mathbf{1} \rangle h^{s+\varepsilon}(g)\, dg}.$$

The definition of P^ι_ε makes sense even though $h^{s+\varepsilon} \notin l^1(\Gamma)$ because the sum does exist in a principle value sense. More importantly to us, the sum converges absolutely when computing any matrix coefficient $\langle P^\iota_\varepsilon v, w \rangle$ and this is all we need. The construction of P^ι_ε is reminiscent of Patterson's construction in [22]. We now state the following facts which we will use.

LEMMA 7.1. *The projection P is the weak limit of the $\{P_\varepsilon^\iota\}$ as $\varepsilon \to 0$. (In other words, $\langle P_\varepsilon^\iota f, g\rangle \to \langle Pf, g\rangle$ for every $f, g \in \mathcal{H}_s$.)*

LEMMA 7.2. $\lim_{\varepsilon\to 0} C_\varepsilon^s = 0$.

LEMMA 7.3. *There exists a dense set $V \subset \mathcal{H}_s$ such that for $v_1, v_2 \in V$ and $g \in$ $\mathrm{PSL}(2, \mathbf{R})$ we have*

$$|\langle \pi_s(g)(v_1), v_2\rangle| \leqslant Ch^{1-\mathrm{Re}(s)}(g), \quad s \neq \tfrac{1}{2}, \tag{7.1}$$

$$|\langle \pi_s(g)(v_1), v_2\rangle| \leqslant Ch^{1/2}(g)d(0, g(0)) = Ch^{1/2}(g)\log(h^{-1}(g)), \quad s = \tfrac{1}{2}. \tag{7.2}$$

The constant C depends on v_1, v_2 and s but not on g.

The most difficult of these is Lemma 7.1, the proof of which is not yet published. An analogous statement, adapted to the case where Γ acts on a homogeneous tree instead of on the upper half-plane, may be found in [8], Chapter II, Sections 7–9. The proof given there is a modification of the second author's original proof of Lemma 7.1. A much improved proof, due to the second author and Michael Cowling, is based on the following inequality, which is a dual form of Proposition 1.1 of [6]. Let $\Gamma \subseteq G$ be any lattice subgroup of any unimodular locally compact group. Let ϕ and ψ be continuous compactly supported functions on G and let $\psi'(x) = \sum_{\gamma\in\Gamma} |\psi(x\gamma)|$. Let π be any unitary representation of G and let λ be the left translation representation of G on $L^2(G/\Gamma)$. Then

$$\|\pi|_\Gamma((\psi^* * \phi * \psi)|_\Gamma)\| \leqslant \|\psi'\|^2_{L^2(G/\Gamma)} \|(\lambda \otimes \pi)(\phi)\|.$$

If $\mathcal{H}$ is the space on which π acts, then $\lambda \otimes \pi$ acts on $L^2(G/\Gamma) \otimes \mathcal{H} \cong L^2(G/\Gamma, \mathcal{H})$, and one proves the inequality by calculating $\Psi^* \circ (\lambda \otimes \pi)(\phi) \circ \Psi$ where $\Psi: \mathcal{H} \to L^2(G/\Gamma, \mathcal{H})$ is given by

$$\Psi(v)(x) = \sum_{\gamma\in\Gamma} \psi(x\gamma)\pi(x\gamma)v.$$

One obtains more precise information by considering (cf. [6]) the orthogonal decomposition of $L^2(G/\Gamma)$ into the constants and the functions of integral zero. In the present context Γ should be the free group, G should be $\mathrm{PSL}(2, \mathbf{R})$, ϕ should be taken in a sequence of compactly supported approximations to $h^{s+\varepsilon}$, and ψ should be a more or less arbitrary but fixed positive bi-K-invariant function of small support. One key point is that $(\psi^* * h^{s+\varepsilon} * \psi)(x)$ is asymptotically equal to a constant multiple of $h^{s+\varepsilon}(x)$ as x tends to infinity, the constant depending on ε, but approaching a limit as ε decreases to 0.

Lemma 7.2 is just a computation involving the well known asymptotics of matrix coefficients for K-finite vectors which is fairly straightforward when s is real (the functions in the integral are positive) but requires a little more care otherwise (the oscillation in the two functions in the denominator cancels near infinity). Lemma 7.3 is also a standard result. We might point out that if Lemma 7.3 held for all vectors in $\mathcal{H}_s$ instead of just on a dense subspace then Theorem 1 would be immediate. This is because this and the assumption that $\pi\circ\iota_1\sim\pi\circ\iota_2$ would imply that the matrix coefficients of π are square integrable, contradicting the assumption that π is not in the discrete series.

Now suppose we have two inequivalent embeddings ι_1 and ι_2 of Γ as a lattice in $\mathrm{PSL}(2,\mathbf{R})$ and suppose that there is an intertwining operator J between the representations $\pi_s\circ\iota_1$ and $\pi_s\circ\iota_2$, i.e., $J\colon\mathcal{H}_s\to\mathcal{H}_s$ is unitary and

$$J^{-1}\circ(\pi_s\circ\iota_1)\circ J=\pi_s\circ\iota_2.$$

We will derive a contradiction. Observe (with $P^1_\varepsilon=P^{\iota_1}_\varepsilon$),

$$\begin{aligned}\langle P^1_\varepsilon(J(v_1)),J(v_1)\rangle&=C^s_\varepsilon\sum_{\gamma\in\Gamma}h^{s+\varepsilon}(\iota_1(\gamma))\langle\pi_s\circ\iota_1(\gamma)J(v_1),J(v_1)\rangle\\&=C^s_\varepsilon\sum_{\gamma\in\Gamma}h^{s+\varepsilon}(\iota_1(\gamma))\langle J^{-1}\circ(\pi_s\circ\iota_1(\gamma))J(v_1),v_1\rangle\\&=C^s_\varepsilon\sum_{\gamma\in\Gamma}h^{s+\varepsilon}(\iota_1(\gamma))\langle\pi_s\circ\iota_2(\gamma)v_1,v_1\rangle.\end{aligned}$$

Now take $v=v_1$ in Lemma 7.3 and note

$$\begin{aligned}|\langle P^1_\varepsilon(J(v_1)),J(v_1)\rangle|&\leqslant|C^s_\varepsilon|C\sum_{\gamma\in\Gamma}h^{\mathrm{Re}(s)+\varepsilon}(\iota_1(\gamma))h^{1-\mathrm{Re}(s)}(\iota_2(\gamma))\log(h^{-1}(\iota_2(\gamma)))\\&\leqslant|C^s_\varepsilon|C\sum_{\gamma\in\Gamma}h^{\mathrm{Re}(s)}(\iota_1(\gamma))h^{1-\mathrm{Re}(s)}(\iota_2(\gamma))\log(h^{-1}(\iota_2(\gamma)))\to 0\end{aligned}$$

since $C^s_\varepsilon\to 0$ by Lemma 7.2 and the sum is bounded by Theorem 3. (This is for the case $s=\frac{1}{2}$. For $\mathrm{Re}(s)\neq\frac{1}{2}$ Theorem 2 suffices.) Therefore

$$\langle P(J(v)),J(v)\rangle=0$$

for a dense set of vectors in $L^2(\mathbf{T})$. But this means $\langle w,\mathbf{1}\rangle=0$ for a dense set of $w\in L^2(\mathbf{T})$, an obvious contradiction. Thus if ι_1 and ι_2 are inequivalent embeddings of Γ, no intertwining operator can exist.

Given two different representations π_1 and π_2 we associate to them s_1 and s_2 as in the descriptions of principle and complementary series representations. Relabeling if necessary we may assume $\mathrm{Re}(s_1)\geqslant\mathrm{Re}(s_2)$. Then the proof given above goes through exactly as before (in fact if $\mathrm{Re}(s_2)<\mathrm{Re}(s_1)$ we don't even need Theorem 2).

For the sake of completeness, we will sketch the proof given in [6] for the case $\iota_1\sim\iota_2$ but $\pi_1\nsim\pi_2$. In this paragraph we let $G=\mathrm{PSL}(2,\mathbf{R})$ and $\Gamma\subset G$ a lattice subgroup. We let $\lambda: G \to L^2(G/\Gamma)$ be the quasi-regular representation given by $\lambda(y)f(x) = f(y^{-1}x)$ and note that we can write

$$\lambda = 1 \oplus \int_{\widehat{G}} n_\tau \tau d\mu(\tau)$$

where 1 denotes the action on constants and $\widehat{G}$ is the unitary dual of G. Given unitary representations (π, H_π), (ϱ, H_ϱ) of G let $\mathrm{Int}(\pi, \varrho)$ denote the space of all intertwining operators between π and ϱ, i.e., $V: H_\pi \to H_\varrho$ and $V \circ \pi = \varrho \circ V$. Since a unitary, nonscalar operator always has nontrivial invariant subspaces, we see that a representation π is irreducible iff $\mathrm{Int}(\pi, \pi)$ has dimension 1. Cowling and Steger show that there is always an injection $J \to Q_J$ from $\mathrm{Int}(\pi|_\Gamma, \varrho|_\Gamma)$ to $\mathrm{Int}(\pi, \lambda \otimes \varrho)$ given by the formula

$$[Q_J v](x) = \varrho(x) J \pi(x)^{-1}, \quad v \in H_\pi, x \in G/\Gamma.$$

If π is an irreducible, unitary representation of G which is not in the discrete series they prove that $\pi \not\subset \tau \otimes \pi$ for any nontrivial $\tau \in \widehat{G}$. This involves comparing the L^p properties of the matrix coefficients of π and $\tau \otimes \pi$ and using the Kunze–Stein phenomenon. This result implies that the injection $\mathrm{Int}(\pi|_\Gamma, \pi|_\Gamma) \to \mathrm{Int}(\pi, \lambda \otimes \pi)$ must actually map into $\mathrm{Int}(\pi, 1 \otimes \pi)$. Since π is irreducible on G this space has dimension 1, thus $\dim(\mathrm{Int}(\pi|_\Gamma, \pi|_\Gamma)) = 1$, and so $\pi|_\Gamma$ is also irreducible. Now suppose π_1 and π_2 are two representations not in the discrete series and that $\pi_1|_\Gamma \sim \pi_2|_\Gamma$. Since $\pi_1|_\Gamma$ and $\pi_2|_\Gamma$ are irreducible, the arguments mentioned also show (after relabeling if necessary) that $\pi_1 \not\subset \tau \otimes \pi_2$ for any nontrivial τ. Since $\mathrm{Int}(\pi_1|_\Gamma, \pi_2|_\Gamma)$ is nontrivial, and it injects into $\mathrm{Int}(\pi_1, \lambda \otimes \pi_2)$, we must have $\pi_1 \subset 1 \otimes \pi_2$. Since π_2 is irreducible, this implies $\pi_1 \sim \pi_2$, as required.

8. Theorem 1 for discrete series and SL(2, R)

Theorem 1 fails for discrete series representations. If π_1 and π_2 are in the discrete series, then $\pi_1 \circ \iota_1$ and $\pi_2 \circ \iota_2$ are square integrable representations of Γ, hence continuously reducible. Any square integrable representation of Γ is characterized up to unitary equivalence by a single real number, its continuous dimension, and the continuous dimension of $\pi_j \circ \iota_j$ is the product of the formal dimension of π_j and the volume of $G/\iota_j(\Gamma)$ (see [9, Theorem 3.3.2] or [12, Lemma 1]). Since the volumes of $G/\iota_1(\Gamma)$ and $G/\iota_2(\Gamma)$ are determined by their Euler characteristics, which in turn are determined by the abstract group Γ (and hence equal), the two representations $\pi_1 \circ \iota_1$ and $\pi_2 \circ \iota_2$ are equivalent if and only if π_1 and π_2 have the same formal dimension. The two discrete series representations π_n^+ and π_n^- have the same formal dimension and otherwise the formal dimensions of distinct discrete series representations are distinct.

Theorem 1 remains true for SL(2, **R**) with a few obvious exceptions. Suppose ι_1, ι_2 are embeddings $\Gamma \to \mathrm{SL}(2, \mathbf{R})$. If $\iota_1 \sim \iota_2$ then $\pi_1 \circ \iota_1 \sim \pi_2 \circ \iota_2$ iff $\pi_1 \sim \pi_2$ by the results in [6] sketched above. There is a 2 to 1 projection $P: \mathrm{SL}(2, \mathbf{R}) \to \mathrm{PSL}(2, \mathbf{R})$ and composing with this projection gives two embeddings $\hat{\iota}_1$ and $\hat{\iota}_2$ of Γ as lattices in PSL(2, **R**). If $\hat{\iota}_1 \nsim \hat{\iota}_2$ then the conclusion of Theorem 2 holds and the argument of Section 7 shows $\pi_1 \circ \iota_1 \nsim \pi_2 \circ \iota_2$ (the estimates and lemmas are same for SL(2, **R**) as they are for PSL(2, **R**)). Finally we must consider the case when $\hat{\iota}_1 \sim \hat{\iota}_2$ but $\iota_1 \nsim \iota_2$. That this is possible is easy to see. For example, if a free group, Γ, on m generators is embedded as a lattice in PSL(2, **R**), then it can be lifted to a lattice in SL(2, **R**) in 2^m ways, since each generator can be lifted in 2 ways and there are no relations to worry about (the resulting group is a lattice since it is a subgroup of finite index in the lattice $P^{-1}(\iota(\Gamma))$). A surface group can be lifted in 2^{2g} ways, where g is the genus of the surface. Here one needs to worry about whether the relation is lifted to the identity or its negative. However, since these surfaces are orientable they have even Euler characteristic, and this implies the latter case never occurs.

So suppose $\hat{\iota}_1 \sim \hat{\iota}_2$ but $\iota_1 \nsim \iota_2$. Note that $\Gamma' = \{\gamma \in \Gamma: \iota_1(\gamma) = \iota_2(\gamma)\}$ is a subgroup of index 2 in Γ (it contains all words in the generators with an even number of each generator). If π_1 and π_2 are distinct representations (unitary, irreducible, not in discrete series) then by [6] $\pi_1 \circ \iota_1|_{\Gamma'} \nsim \pi_2 \circ \iota_2|_{\Gamma'}$, and hence they are not equivalent on Γ. Now suppose $\pi_1 = \pi_2 = \pi$. If π is one of the representations of SL(2, **R**) that factors through PSL(2, **R**) then of course it restricts to equivalent representations on ι_1 and ι_2. So assume π does not factor through PSL(2, **R**). Thus it must be in the principal series and of the form

$$\pi_s'(g)F(e^{i\varphi}) = (-\bar{\beta}e^{i\varphi} + \alpha)|-\bar{\beta}e^{i\varphi} + \alpha|^{2s-1} F\left(\frac{\bar{\alpha}e^{i\varphi} - \beta}{-\bar{\beta}e^{i\varphi} + \alpha}\right)$$

as described in Section 6. Obviously $\pi \circ \iota_1 \sim \pi \circ \iota_2$ on Γ' and by the results of Cowling and Steger this restriction is irreducible, so $\dim(\mathrm{Int}(\pi|_{\iota_1(\Gamma')}, \pi|_{\iota_2(\Gamma')})) = 1$. If there were an intertwining map between $\pi \circ \iota_1$ and $\pi \circ \iota_2$, it could not be scalar since the operators $\pi \circ \iota_1(\gamma)$ and $\pi \circ \iota_2(\gamma)$ differ for some γ (here we are using the fact that $\iota_1(\gamma) = -\iota_2(\gamma)$ for some $\gamma \in \Gamma$ and the explicit formula for π_s). But this operator would also intertwine the representations restricted to Γ', so $\dim(\mathrm{Int}(\pi|_{\iota_1(\Gamma')}, \pi|_{\iota_2(\Gamma')})) > 1$! This is a contradiction so $\pi \circ \iota_1 \nsim \pi \circ \iota_2$. These remarks can be summarized as

THEOREM 5. *Suppose ι_1, ι_2 are lattice embeddings of Γ in* SL(2, **R**) *and that π_1, π_2 are unitary, irreducible representations of* SL(2, **R**) *not in the discrete series. Then $\pi_1 \circ \iota_1 \sim \pi_2 \circ \iota_2$ iff $\pi_1 \sim \pi_2$ and one of the following hold*:

(1) $\iota_1 \sim \iota_2$.

(2) π_1 *factors through* PSL(2, **R**) *and* $\hat{\iota}_1 \sim \hat{\iota}_2$.

References

[1] AGARD, S., A geometric proof of Mostow's rigidity theorem for groups of divergence type. *Acta Math.*, 151 (1983), 231–252.

[2] ALHFORS, L. V., *Conformal Invariants.* McGraw-Hill, New York, 1973.

[3] BEARDON, A. F., Inequalities for certain Fuchsian groups. *Acta. Math.*, 127 (1971), 221-258.

[4] — *The Geometry of Discrete Groups.* Springer-Verlag, New York, 1983.

[5] CARLESON, L., *Selected Problems on Exceptional Sets.* Van Nostrand, 1967.

[6] COWLING, M. & STEGER, T., Irreducibility of restrictions of unitary representations to lattices. *J. Reine Angew. Math.*, 420 (1991), 85–98.

[7] DEHN, M., *Papers on Group Theory and Topology.* Translated by John Stillwell. Springer-Verlag, New York, 1987.

[8] FIGÀ-TALAMANCA, A. & NEBBIA, C., *Harmonic Analysis and Representation Theory for Groups Acting on Homogeneous Trees.* London Math. Soc. Lecture Note Ser., 162. Cambridge University Press, Cambridge, 1991.

[9] GOODMAN, F., DE LA HARPE, P. & JONES, V., *Coxeter Graphs and Towers of Algebras.* MSRI Publications, 14. Springer-Verlag, New York, 1989.

[10] KEEN, L., Canonical polygons for finitely generated Fuchsian groups. *Acta. Math.*, 115 (1966), 1–16.

[11] KNAPP, A. W., *Representation Theory of Semisimple Groups.* Princeton University Press, Princeton, New Jersey, 1986.

[12] KUHN, G. & STEGER, T., Restrictions of the special representation of Aut(Tree_3) to two cocompact subgroups. To appear in *Rocky Mountain J. Math.*

[13] KUUSALO. T., Boundary mappings of geometric isomorphisms of Fuchsian groups. *Ann. Acad. Sci. Fenn. Ser. A I*, 545 (1973), 1–6.

[14] LANG, S., $SL_2(\mathbf{R})$. Addison-Wesley, Reading, Massachusetts, 1975.

[15] LYNDON, R. C. & SCHUPP, P. E., *Combinatorial Group Theory.* Springer-Verlag, New York, 1977.

[16] MARGULIS, G. A., Non-uniform lattices in semisimple algebraic groups, in *Lie Groups and Their Representations* (I.M. Gelfand, ed.). Wiley, New York, 1975.

[17] — Discrete subgroups of motions of manifolds of non-positive curvature. *Amer. Math. Soc. Transl.*, 109 (1977), 33–45.

[18] MOSTOW, G. D., Quasiconformal mappings in n-space and the rigidity of hyperbolic space forms. *Inst. Hautes Études Sci. Publ. Math.*, 34 (1968), 53–104.

[19] — *Strong Rigidity of Locally Symmetric Spaces.* Ann. of Math. Studies, 78. Princeton University Press, Princeton, New Jersey, 1973.

[20] NIELSEN, J., Die Isomorphismengruppe der freien Gruppen. *Math. Ann.*, 91 (1924), 169–209.

[21] — Untersuchen zur Topologie der geschlossenen zweiseitigen Flächen I, II, III. *Acta Math.*, 50 (1927), 189–358; 53 (1929), 1–76; 58 (1932), 87–167.

[22] PATTERSON, S. J., The limit set of a Fuchsian group. *Acta. Math.*, 136 (1976), 241–273.

[23] PENSAVALLE, C., Spherical series and generators of a free group. Preprint, Università di Sassari, 1992.

[24] PRASAD, G., Strong rigidity of $\mathbf{Q}$-rank 1 lattices. *Invent. Math.*, 21 (1973), 255–286.

[25] RAGHUNATHAN, M. S., *Discrete Subgroups of Lie Groups.* Springer-Verlag, New York, 1972.

[26] SELBERG, A., On discontinuous groups in higher-dimensional spaces, in *Contributions to Function Theory.* Tata Institute, Bombay, 1960.

[27] SERIES, C., The infinite word problem and limit sets in Fuchsian groups. *Ergodic Theory Dynamical Systems*, 1 (1981), 337–360.

[28] STEIN, E. M., *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey, 1970.
[29] SULLIVAN, D., The density at infinity of a discrete group of hyperbolic motions. *Inst. Hautes Études Sci. Publ. Math.*, 50 (1979), 171–202.
[30] TUKIA, P., A rigidity theorem for Möbius groups. *Invent. Math.*, 97 (1989), 405–431.
[31] — Hausdorff dimension and quasisymmetric mappings. *Math. Scand.*, 65 (1989), 152-160.
[32] ZIMMER, R. J., Strong rigidity for ergodic actions of semisimple Lie groups. *Ann. of Math.*, 112 (1980), 511-529.
[33] — Ergodic theory, group representations and rigidity. *Bull. Amer. Math. Soc.*, 6 (1982), 383–416.
[34] — *Ergodic Theory and Semisimple Groups.* Birkhäuser, Boston, 1984.

CHRISTOPHER BISHOP
Department of Mathematics
SUNY at Stony Brook
Stony Brook, NY 11794-3651
U.S.A.
E-mail address: bishop@math.sunysb.edu

TIM STEGER
Department of Mathematics
University of Georgia
Athens, GA 30602-7403
U.S.A.
E-mail address: steger@joe.math.uga.edu

Received April 3, 1990
Received in revised form January 2, 1992